



UNIVERSITY OF AGDER

# Mapping Extremist Forums using Text Mining

**Targeir Attestog, Samangi Perera**

**Supervisor**

Professor Ole-Christoffer Granmo

*This Master's Thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education. However, this does not imply that the University answers for the methods that are used or the conclusions that are drawn.*

University of Agder, 2013

Faculty of Engineering and Science

Department of Information and Communication Technology

# Abstract

Political opinions far from what is considered normal, a distorted view of reality, and hatred to certain other groups are spread amongst political extremists like Islamists and White Supremacists. Demonstrations and violence performed by some members of these groups are well-known from mass media and get a lot of attention. The Islamists and right-extremists exploit this benefit to spread their message to ordinary people. In online forums, young, curious people can read detailed information (or propaganda) from extremists. Which words do extremists then use to convince each other in addition to other curious readers that what they stand for is right?

The goal of this thesis is to first find algorithms or techniques for how to discover characteristic vocabulary in online extremist forums and words that frequently are used in the same forum message. Then we analyse the results to find patterns of what is typical vocabulary in the different forums. Mapping the extremists' habits of vocabulary usage can help us know better how extremists write in online extremist forums, and possibly also help us recognize them when they write on some other websites.

In this thesis, we find frequent and characteristic words by means of Global Term Frequency (GTF) and pairs of co-occurring words by means of odds ratio in different extremist forums. We compare normalized GTF (NGTF) of words in two forums to find out where they are used most. Words used in only one of two forums are found as well. We find the GTFs for words written by five of the ten most active authors in each forum, and we find words that one author writes, while the other of ten most active authors does not write.

From results we see that Islamists write most about religion, but also some politics. Some popular words are “allah”, “prophet”, “fasting”, and “hajj”. The right-extreme websites Stormfront and Vigrind discuss politics and argument for their own ideology and against the mainstream politics. Frequent words in Stormfront are “white”, “jews”, and “race”, in the Norwegian Vigrind website “jødene”, “tyskland”, and “krigen”. In the German right-extreme website Deutsche Stimme, “npd”, “Deutschland”, “partei”, and “volk” are frequent words. Both Islamists and right-extremists are preoccupied by family values.

Our results are useful for discovering topics that extremists write about in their online forums, topics that other people do not write about at all or write about with a different point of view.

# Preface

This thesis is the result of a master's thesis project at the Faculty of Engineering and Science at the University of Agder in Grimstad, Norway. After writing a pre-project report in the late autumn 2012, all the research and thesis writing were done during the spring semester in 2013. The project has the topic code IKT590 and is worth 30 ECTS credit points. It is meant to be the final project for a two-year master study program in Information and Communications Technology (ICT) with the study profile System Development.

The purpose of a master's thesis project is to immerse oneself in a specific subject area, and it is supposed to be an independent research.

The thesis was carried out under the supervision of professor Ole-Christoffer Granmo. We would like to thank him for the valuable guidance and advices throughout the project. The task of the completion of this thesis would have not been possible without his support. We would also like to thank associate professor Morten Goodwin for helping us with the pre-project report in the autumn. We also thank our families and friends for moral support while we were working with this research.

*Targeir Attestog*  
*Samangi Perera*  
University of Agder  
Grimstad, Norway  
2 June 2013

# Contents

List of Figures .....	vii
List of Tables .....	viii
Chapter 1: Introduction .....	1
1.1 Problem Discussion.....	2
1.2 Hypotheses .....	3
1.3 Contributions.....	4
1.4 Key Components of our Project .....	5
1.5 Structure of Report .....	5
Chapter 2: Former work.....	7
2.1 Relationship Discovery in Large Text Collections Using Latent Semantic Indexing	7
2.2 Names: A New Frontier in Text Mining .....	8
2.3 Anomaly Detection in Extremist Web Forums Using a Dynamical Systems Approach .....	9
2.4 US Domestic Extremist Groups on the Web: Link and Content Analysis.....	9
2.5 Applying Authorship Analysis to Extremist-Group Web Forum Messages.....	10
2.6 Exploring the Dark Side of the Web: Collection and Analysis of U.S. Extremist Online Forums.....	11
2.7 Authorship Attribution of Web Forum Posts .....	12
2.8 Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Board.....	13
2.9 Contributions to our Project .....	14
Chapter 3: Method .....	16
3.1 Online Forums.....	16
3.2 Getting Data .....	18
3.2.1 Download of Webpages.....	19
3.2.2 Extraction of Data from Webpages.....	21
3.3 Forum Text Mining .....	23
3.3.1 TF and IDF.....	23
3.3.2 GTF and NGTF.....	23
3.3.3 Forum Word Count Comparison .....	24
3.3.4 Word Colocation Analysis.....	25
3.5 Author Text Mining .....	26

3.6	Manual Analysis.....	27
Chapter 4: Results .....		28
4.1	Word Frequency Analysis of Ummah.....	29
4.2	Word Frequency Analysis of Islamic Web-Community.....	30
4.3	Word Frequency Analysis of Stormfront.....	31
4.4	Word Frequency Analysis of Swiss English Forum .....	37
4.5	Word Frequency Analysis of the CS:GO Forum in Steam .....	39
4.6	NGTF Comparison of Swiss English Forum and Islamic Web-Community.....	40
4.7	NGTF Comparison of Islamic Web-Community with Steam.....	43
4.8	NGTF Comparison of Swiss English Forum and Ummah.....	52
4.9	NGTF Comparison of Steam and Ummah .....	63
4.10	NGTF Comparison of Swiss English Forum and Stormfront.....	67
4.11	NGTF Comparison of Steam and Stormfront .....	72
4.12	NGTF Comparison of Stormfront and Ummah .....	80
4.13	Word Frequency Analysis of Vigrid .....	84
4.14	Word Frequency Analysis of Diskusjon.no .....	87
4.15	NGTF Comparison of Vigrid and Diskusjon.no .....	90
4.16	Word Frequency Analysis of Deutsche Stimme .....	98
4.17	Word Frequency Analysis of Turn to Islam (German) .....	101
4.18	NGTF Comparison of Deutsche Stimme and Turn to Islam (German) .....	106
4.19	Author Vocabulary Analysis.....	110
4.19.1	Ummah.....	110
4.19.2	Stormfront.....	124
4.20	Word Colocation Analysis .....	136
4.20.1	Steam.....	136
4.20.2	Islamic Web-Community.....	138
4.20.3	Deutsche Stimme .....	141
4.20.4	Turn to Islam.....	147
4.20.5	Ummah.....	152
4.20.6	Stormfront.....	158
Chapter 5: Discussion .....		165
Chapter 6: Conclusion.....		170
References.....		172

Appendix A: Code, Text and Word Statistics.....	176
---	-----

# List of Figures

Figure 1: Naïve Bayes formula from [4].....	2
Figure 2: Workflow for analysis of extremist forums .....	16
Figure 3: Example of HTML DOM node hierarchy from W3Schools [21].....	21
Figure 4: Definition of logarithmic IDF .....	23
Figure 5: Definition of NGTF.....	24
Figure 6: Joint odds ratio, formula from [24] .....	26
Figure 7: 30 most frequent words in Ummah .....	30
Figure 8: 30 most frequent words in Islamic Web-Community .....	31
Figure 9: 30 most frequent words in Stormfront .....	32
Figure 10: 30 most frequent words in Steam .....	40

# List of Tables

Table 1: Variants of zion/zionist/zionism (2631 occurrences in total) .....	33
Table 2: Variants of nazi (4880 occurrences in total) .....	33
Table 3: Top 30 words from Stormfront, sorted in descending order by GTF or NGTF .....	34
Table 4: Top 30 words from Stormfront, sorted in descending order by NGTF-IDF .....	35
Table 5: 20 most frequent words in Swiss English Forum .....	37
Table 6: 20 words with highest NGTF-IDF values in Swiss English Forum .....	38
Table 7: Top 30 words in Swiss English Forum (f1) and Islamic Web-Community (f2) sorted by f1/f2 NGTF .....	40
Table 8: Top 30 words in Swiss English Forum (f1) and Islamic Web-Community (f2) sorted by f2/f1 NGTF .....	42
Table 9: Top 30 words in Islamic Web-Community (f1) and Steam (f2) sorted by f1/f2 NGTF .....	43
Table 10: Top 30 words in Islamic Web-Community (f1) and Steam (f2) sorted by f2/f1 NGTF .....	46
Table 11: Words that are in Islamic Web-Community, but not in Steam.....	48
Table 12: Words that are in Steam, but not in Islamic Web-Community.....	50
Table 13: 30 words with highest f1/f2 NGTF ratios for words that are both in Swiss English Forum (f1) and Ummah (f2) .....	52
Table 14: 30 words that are both in Ummah (f2) and Swiss English Forum (f1), with highest f2/f1 ratios .....	55
Table 15: 30 words that are used in Swiss English Forum, but not in Ummah, sorted by GTF in descending order .....	59
Table 16: 26 words that are used in Ummah, but not in Swiss English Forum, sorted by GTF in descending order .....	61
Table 17: Words in Steam (f1) and Ummah (f2) with 30 highest f1/f2 NGTF ratios .....	64
Table 18: Words in Steam (f1) and Ummah (f2) with 30 highest f2/f1 NGTF ratios .....	66
Table 19: Words in Swiss English Forum (f1) and Stormfront (f2) with 30 highest f1/f2 NGTF ratios .....	67
Table 20: Words in Swiss English Forum (f1) and Stormfront (f2) with 30 highest f2/f1 NGTF ratios .....	68
Table 21: Top 30 words in Steam (f1) and Stormfront (f2) sorted by f1/f2 NGTF.....	72
Table 22: Top 30 words in Steam (f1) and Stormfront (f2) sorted by f2/f1 NGTF.....	77
Table 23: Top 30 words in Stormfront (f1) and Ummah (f2) sorted by f1/f2 NGTF.....	80
Table 24: Top 30 words in Stormfront (f1) and Ummah (f2) sorted by f2/f1 NGTF.....	81
Table 25: 30 words with highest GTFs in Vigrid .....	84
Table 26: 30 words with highest NGTF-IDF values .....	85
Table 27: 20 words with highest NGTF-IDF values .....	89
Table 28: Top 30 words in Vigrid (f2) and Diskusjon.no (f1) sorted by f2/f1 NGTF.....	90
Table 29: Top 30 words in Vigrid (f2) and Diskusjon.no (f1) sorted by f1/f2 NGTF.....	94
Table 30: GTFs and examples for Turn to Islam .....	101
Table 31: 25 non-stop words with highest f2/f1 NGTF ratios. Turn to Islam is f2, and Deutsche Stimme is f1. ....	106



Table 32: 30 words with highest f1/f2 NGTF ratios. Deutsche Stimme is f1, and Turn to Islam is f2.....	107
Table 33: GTFs and examples for "Anna" .....	110
Table 34: GTFs and examples for "AbuMubarak" .....	113
Table 35: GTFs and examples for "Abdul Haq" .....	117
Table 36: GTFs and examples for "hamza81" .....	119
Table 37: GTFs and examples for "Z-blade" .....	122
Table 38: GTFs and examples for "kazan188" .....	124
Table 39: GTFs and examples for "Mjodr" .....	125
Table 40: GTFs and examples for "WhiteRights" .....	127
Table 41: GTFs and examples for "revision" .....	130
Table 42: GTFs and examples for "Revalo" .....	131
Table 43: Co-occurring word pairs in Steam .....	138
Table 44: Co-occurring words in Islamic Web-Community .....	139
Table 45: Co-occurring words in Deutsche Stimme.....	141
Table 46: Co-occurring words in Turn to Islam .....	147
Table 47: Word co-occurrences in Ummah .....	152
Table 48: Word co-occurrences in Stormfront .....	158

# Chapter 1: Introduction

In the last years, extreme ideologies have had easier conditions for getting widespread due to new information and communication technology. The World Wide Web became open to ordinary people in 1994, and already the following year the White Supremacist forum Stormfront.org was launched [1].

Extremism is a political act or ideology which is far from the political centre of society. In a democracy, people that want to introduce an authoritarian society are often called extremists. The word may also refer to persons with points of view that are inconsistent with existing norms and moral, and it does not necessarily mean that these persons pose a threat to society. The opposite of extremism is moderation, which means to be close to the average norms in society. It is difficult to define extremism precisely, because the experience of what is extreme is subjective. Also important is that inside a group of extremists, like Islamists/jihadists, racists or nationalists, there are usually several individual opinions. One should keep in mind that people usually do not call themselves extremists. In some cases they may use the word radicals about themselves. "Extremism" and "extremist" are normally exonyms, words that are only used by people other than the ones that are called extremists. It is often used pejoratively [2] .

It is obviously advantageous for the society to know about how extremists think, both for revealing them if they do something illegal or dangerous (becoming terrorists), and for knowing what and how they think if we want to convince them that they are wrong (like with the terrorists in the movie "Air Force One"). If we know what extremists write in online forums, we can get an idea of how they think, maybe understand why they have opinions they have. Then we can easier find arguments to reason with them that other opinions can make sense, so that the extremists also understand how people outside their group think, and the same extremists may even change their point of view.

Analysing the vocabulary and semantics of text written by extremists is interesting, because we then get to know how they express themselves in their writing. We supposed at the start that we would find some words or expressions that are expected, like "allah" and "prophet" in Islamist forums, but also get some surprises. In our project, we hoped to learn about how people like Islamists and nationalists express themselves.

By understanding the semantic relationships between the words/terms found in online extremist forum messages, we were able to learn the vocabulary patterns used by authors with extreme opinions, which potentially can be helpful in the counter-extremism field in the future. Our goal was to find out which words and combinations of words were typically used by different kinds of extremists. Therefore, our goal was to explore which words and expressions are typical for extremist forums by means of text mining. We wrote our own Java applications for doing the mining, analysis, and interpretation of the data. We compared the vocabulary applied in "ordinary" forums with the words used in extremist forums. In the forums we also found words and expressions that appeared in the same messages (or even sentences), thus potentially being semantically related.

## 1.1 Problem Discussion

In this thesis, we will by means of text mining and text analysis techniques find characteristic vocabulary and words that are frequently used in the same message in online extremist forums.

The purpose of our thesis is to explore which words and expressions are typical for extremist forums by means of text mining. We analyse the text contents of Islamist and Neo-Nazi/White Supremacist forums and compare their vocabularies with the vocabularies in "ordinary" forums about politics and daily life. Through this exploration, our goal is to gain insight into the vocabulary that extremists typically use and find out how effective text mining is for that purpose. By means of comparisons of vocabularies in two different forums at a time, we find words that extremists use more than ordinary people in online forums, and words that extremists use and the others not. This is useful for finding out how extremists express themselves differently from other people. Another part of the project is to determine whether individual forum authors can be profiled based on their vocabulary usage, and whether it is possible to distinguish between different authors based on their use of words. The intent is to see how different or similar different authors' ways of writing are compared to each other. We only compare authors from the same forum with each other.

We analyse vocabularies of forums in English, Norwegian, and German. We have not found any examples of former research in Norwegian extremist websites, but we know from the Dark Web Forum Portal that at least a German Islamist forum has been analysed earlier [3]. We envision that the techniques developed and knowledge gained in this project may be of help when uncovering new extremists and their networks.

One naïve approach of solving the problem in this project would be to just pick out some random forum threads and see which words are written there. This does not work well, because ordinary people normally have biased assumptions for what extremists write about, and will therefore look for words and expressions they expect to find. In many cases such expected words will be found, while other important features of the text may be overlooked by human eyes.

Naïve Bayes Classifier does also not work well, because it assumes that all word frequencies (or probabilities) are independent of each other. In addition some pre-knowledge about word frequencies in forums is required. Such information we can get from a training set where one knows for certain what kind of forums different samples of text come from, an extremist forum or an ordinary forum.

$$v_{NB} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

**Figure 1: Naïve Bayes formula from [4]**

Let us say that the target value  $v_{NB}$  in Figure 1 is that a forum is written by extremists.  $P(v_{\text{extremist}}) = P(v_{\text{ordinary}}) = 0.5$  can then be examples for prior probabilities.  $P(v_j)$  is multiplied with the product of the frequency/probability  $P(\text{word } i | v_j)$  for each word  $i$  in the forum  $v_j$ . The maximum value of this product of the prior probability for forum type  $j$  and the probability for each word given the forum type  $j$ , decides whether it is an extremist forum or an ordinary forum. This means that we need knowledge in advance about the frequencies

(probabilities) of different words in previously examined forums written by extremists and ordinary people. For the research for this thesis we do not have such knowledge from a training set before we start the text mining, so the approach described in the formula in Figure 1 does not work well for us.

The formula shows also that the occurrences of two or more different words in a text are always independent of each other. It is evident that this is not true for online forums, because humans never write down random words in an arbitrary order. They write sentences with meaning, and then some specific words must be used together to make sense. For instance, the words “god” and “prophet” often occur in the same sentence in a religious text, so they seem to be somehow semantically interconnected to each other, and are hence not independent. An important part of our project is actually to find out which words tend to be in the same sentences or messages.

## 1.2 Hypotheses

As starting points for our research we have the following hypotheses:

**1. It is possible to map the characteristics of a forum by finding the most frequently occurring words and the combinations of two words that are most frequently used in the same message or sentence.**

We wish to find out which words are typical and characteristic for extremists in online forums. Stop words (and, or, so, etc.), words that are used in nearly all sentences, are therefore not interesting. Most of them are removed even before the counting of words and counting of word co-occurrences (co-locations) start.

Using global term frequency (GTF), the frequency of a word in a whole forum instead of the term frequency (TF) in single messages is more appropriate in our research, since forum messages are often very short, and counting of word appearances in each single message does not give any useful results (See subsections 3.3.1 and 3.3.2). We prioritize presenting the interesting words from the list of 30 most used words, but many other words further down in the list may also be worth mentioning.

We also look for words that tend to occurrence in the same forum message. Ideally, we should find words that are in the same sentence. Then we can find components of non-hyphenated compound words. An example of two such components can be given by “compound” and “word”, which together becomes “compound word”. We can also find names, like the pair “John” and “Smith”. Else there is a possibility of finding words that happen to be somehow semantically related, like “buy” and “tickets” in the sentence “I buy tickets for us.”.

**2. Characteristic vocabulary in the text which an author has written in a forum, can be used to distinguish a specific author from other authors in the same forum.**

A unique profile of an author can be created by using his/her vocabulary in the forum. We first find the GTF's of the words each of five of the most productive authors has written, and then we see which words are most frequent in each author's text. We then find words that are unique for each author in the group of ten most productive authors in a forum. If time allows, we can also use co-occurring words as author characteristics. Method is described in section 3.5.

### **3. It is possible to distinguish the characteristic vocabulary of an extremist forum from the choice of words in other kinds of forums (computer game forums, political forums, daily life forums).**

Distinguishing texts by extremists from texts by non-extremists should be possible, though subtle and challenging in some cases. The definition of extremism depends on a person's point of view. The word "extreme" means that something is completely different from what is thought of as usual. An extremist would likely not call himself extremist, because he or she thinks that his/her opinion is correct and that all other kinds of people are wrong. We compare normalized GTF values of words the two forums have in common (see subsection 3.3.3) to find words that are used much more in one of the forums. We also find the words that are only applied in one of the same two forums we compare. Ordinary forums may be ones that discuss politics, computer games or anything else from everyday life. We perform vocabulary analysis to make it more clear how extremists differ from other persons in how they express themselves on the Internet.

As outcome we expect that different kinds of extremists use specific words in their vocabulary, which are typical for them. We expect that Islamists write much about Allah and his prophet, Muhammad, and that they write many argumentations about why their religion, Islam, is better than other religions. Amongst Nazis/right-extremists we expect much glorification of Norwegian, English, German, and other European people and negative assertions about Jews, Roma (Gipsys), Africans, social democrats and so on. Short, Nazis will likely write racist commentaries about many kinds of non-European people and extreme, right-turned political opinions that are just loosely founded on reality. We expect that both groups will argue that they are right and others are wrong. There will likely also be some unexpected discussion topics.

## **1.3 Contributions**

We hope to contribute with results that show how different extremist groups argue for their opinions, that is, which words they use and how they put these words together to reason for what they think is right. There may be some extremist webpages with contents about everyday topics, like the weather, hobbies and social life. Extremists are namely humans like the rest of us. Nonetheless, we should be aware that extreme political or religious points of view can be hidden in such everyday discussions, but as mentioned, not necessarily.

We contribute with an analysis architecture, as shown in Figure 2, and described in chapter 3, which should be a good approach for finding significant and interesting texts written by extremists in their forums. This architecture has a workflow where we first download HTML document files from these online forums, extract forum message text data from these webpages, find word frequencies and co-occurring words on forum and author level, compare

extremist forums with ordinary forums, and analyse the statistical results that we finally end up with.

Note that we download and analyse text from ordinary forums as well, for comparing the contents of the extremist forums with them. We then find words that extremists use and other people not, and words that extremists use more than ordinary people. See details in subsection 3.3.3.

Finding pairs of co-occurring words can be useful for discovering compound words (like “compound word”), expressions with two words (like “thank you”), or names consisting of a first and last name (like “Jan Johansen”). The words do not have to be neighbours to have some kind of sensible connection to each other (“I do my **work**, and get paid for **it**.”). Our algorithm runs unfortunately slow and generates a lot of useless/noisy results. Still it also gives many useful results, and puts focus on word pairs that we otherwise would not have discovered. See details in subsection 3.3.4.

## 1.4 Key Components of our Project

The purpose of this thesis project is to figure out which words are frequently used in extremist forums. Included in this task is also finding out which words are used in the same forum messages. These words may potentially have semantic relations to each other. The vocabulary of the extremist forums is compared to the vocabulary of “ordinary” forums. One can in chapter 4 see which words are typical for extremists of the kinds we study, namely Islamists and white supremacists/neo-Nazis. Forums in the languages English, Norwegian, and German are analysed.

A part of the project is to find the typical words used by some of the most active authors and then test how different each author’s vocabulary is from the vocabularies of the other productive authors.

Our project follows the steps of collecting online forum text data, storing them structurally in a database, perform text mining for extracting useful statistical data for the analysis, and then perform manual analysis for finding the kind of information we are looking for. We finally interpret the data to get knowledge about what extremists write about, how they do that, and in some cases try out some theories for why they do that.

## 1.5 Structure of Report

In chapter 2 one can read about former work done in the field we have researched. In chapter 3 we describe how we have collected text from online forums and generated statistical information from all of it. We describe the process step by step with a figure of the workflow of the analysis architecture as a starting point. In chapter 4 we present the statistics of frequent words in the different forums and words that often occur together in the same forum message, which we refer to as the forum text mining step in our analysis architecture. In addition, the word counts in one forum are compared with the word counts in another forum. Words that exist in one forum, but not another, are also found. Chapter 4 includes also short discussions of the different concrete results by means of examples of how the frequent words are used. There is a description of the results from the author text mining step in the analysis architecture at the end of the of chapter 4. The analysis in chapter 4 as a whole is what call

the manual analysis step in our architecture. Chapter 5 includes a general discussion which shows how the statistics shown in chapter 4 can prove the different hypotheses from chapter 1. Chapter 5 concludes our thesis with short summarization of our results.

## Chapter 2: Former work

### 2.1 Relationship Discovery in Large Text Collections Using Latent Semantic Indexing

Bradford's paper "Relationship Discovery in Large Text Collections Using Latent Semantic Indexing" addresses the problem that users do not know where to focus their attention in finding important information in large collections of text. It presents the results of an experiment which has been tested using a counterterrorism analysis example where test database used for it contains 158,492 English-language news articles [5].

The analysis has focused on individual terrorists, terrorist groups, targets, and weapons as the entities of the interest concerned in such threat conditions. The main feature of this experiment is the application of the technique of Latent Semantic Indexing (LSI).

Latent Semantic Indexing is used to recognize associations between entities of interest. The text of the news articles is pre-processed using the LingPipe entity extraction software. LingPipe is a set of Java tools designed to carry out linguistic analysis on natural language data [5]. Named entities extracted by the LingPipe software are used in the creation of the LSI representation space. Thus, the LSI space consists of representation vectors for named entities with the other individual terms in the documents. The cosine of the angle between the representation vectors for any two entities in the LSI space gives significant information about the degree of association of those two entities which gives a direct measure of their degree of contextual association in the collection [5]. Therefore, identification of such relations helps to focus our concentration only on important information we are interested about.

Most of the LSI applications, the elements used to create the LSI space are the individual terms found in the documents, but in this experiment, the indexing units consist of both individual terms and series of multi-terms recognized as meaningful elements that are corresponding to named entities by the entity extraction software. However LingPipe itself is unable to directly recognize entities of particular type or interest based on the existing training data. Therefore, for this experiment to be carried out, they have recognized the entities by correlation of some known list of terrorists groups with the list of entities tagged as organizations by LingPipe. Similarly targets are identified with the correlation of the list of potential targets with the list of entities tagged by LingPipe as locations.

Results of the experiment are encouraging. For example, in Entity-Entity comparisons – terrorist groups versus targets matrix representation, the cosine value .5087 between the vectors corresponding to the *GSPC* (terrorist group) and *Strasbourg Cathedral* (target) is high compared to the other entries. According to this paper, Frankfurt cell of the GSPC certainly had planned an attack on the cathedral in Strasbourg during Christmas in 2000. This implies that there exists a correct association between those two entities. Also in the document collection there were no articles that directly connected the term GSPC with Strasbourg Cathedral. Hence, this is a good example that the LSI technique takes into account higher-



order relations in making associations [5]. Similarly, in Entity-entity comparisons – terrorist groups versus weapons yields some correct associations.

In Entity-Entity comparisons – names versus names matrix, some entries with high cosine values mean that they have close relations. They also correlate well with actual relations, where real names and aliases for the same person characteristically are closely related. For instance, the vectors representing the names “Ammar Saefi” and “Abdelrazzaq El Mizalli” have a very high cosine value (0.9675) and “Abdelrazzaq El Mizalli” is an alias used by “Ammar Saefi”. Therefore it is very important that this facilitates identification of possible use of aliases that can be utilized in counterterrorism activities.

In this experiment, conceptual descriptions of the entities are also used for the comparisons. For example it is shown that there are strong and close associations with the persons (names) and the documented description of the terrorist group GSPC than using the single term named entity GSPC.

In addition a composite vector of specific type that is representing the average of contexts of 14 different terrorists groups is created and compared with the general type of entities (organizations) identified by the LingPipe software. Results are impressive that, by taking ten highest cosine values of the organization entities that are most similar to the composite terrorist group are to be terrorists groups or elements of thereof. These identified entities are used for the expansion of the list of terrorists groups.

## 2.2 Names: A New Fronter in Text Mining

The article “Names: A New Fronter in Text Mining” by Patman and Thomson explains that name recognition and extraction is not easy [6]. Originally one tried to do this in a rule-based way with English and Western European pattern rules, with a first name, one or more middle names and then a last name [6]. This does not work well for the rest of the world. In China and Korea the family name appears first, and there are other kinds of patterns around the world. Another problem is that names written with Cyrillic letters, Chinese symbols, and characters from other alphabets may be interpreted to Latin letters in different ways. This makes it difficult for computer programs to figure out if two differently written, but somehow similar names really are the same. There may also be some issues when two quite similar names refer to two different persons. Sometimes there may be confusion because sometimes a name is written in full length (“Thomas Alva Edison”), with initials (“T. A. Edison”), or with a nickname (“Tom Edison”). This is of course dependent on culture [6].

Language Analysis Systems, Inc. (LAS) has implemented a lot of approaches in tools to handle cultural traits in names, like the three following ones [6]:

**Automatic Name Analysis** includes rules for names in different cultures. Names are compared to the different rules and the most probable culture is calculated. The culture assignment is then used by other technologies to find the most appropriate name-matching strategy.

**NameVariantGenerator** performs comparison between a name to 750 million words in a database after creating a regular expression based on the given name and its derived culture. One ends up with a list of other likely variants of the name.

NameHunter is a search engine that calculates how similar two names are, based on orthography, word order, and how many elements there are in the text string. Noisy data are allowed.

**MetaMatch** finds out which names are similar based on phonetics after each word has been assigned a culture.

NameGenderizer finds the most likely gender of a name (male/female) based on most frequent assignment of gender.

Name entity recognition was first just based on rules. Later they have been more probabilistic and statistical, which seems to work better [6].

For extracting names and then matching them, the procedure is like this:

Inside a document, find all possible names, establish coreference between name mentionings, find out if different names have similar contexts, find longest canonical names and compare them with shorter, similar names. Save them as variants of each other if contexts match.

Then across several documents with same culture/language, find canonical names that are similar to each and mark them as possible matches. Then examine if the surrounding descriptions in the text are overlapping each other. If the overlappings are many enough for one pair of similar names, then add them a list that tells they are variations of each other. If descriptions are not overlapping enough, keep them as possible match pair until they are recognized as variants of each other or the algorithm stops [6].

## **2.3 Anomaly Detection in Extremist Web Forums Using a Dynamical Systems Approach**

In “Anomaly Detection in Extremist Web Forums Using a Dynamical Systems Approach” by Steve Cramer [7], one has done dynamic anomaly detection with finite-time Lyapunov exponents. This is an unsupervised learning technique to find changes in data over time. In this case, one has found out which words are used more or less at a specific period of time than previously. This can then alarm new activity in a extremist forum, which in some cases may be a warning of a security threat. An analyst must anyway have a look at it to take a final decision of further actions. The algorithm is just meant to bring focus-of-attention to forum entries that stick out from what has been seen earlier [7]. In this research, clustering has just been “one-hop”, which means that a cluster consists of an author and his/her messages in the forum. A more ideal way would have been to cluster incrementally, so that messages are clustered according to time periods [7].

## **2.4 US Domestic Extremist Groups on the Web: Link and Content Analysis**

“US Domestic Extremist Groups on the Web: Link and Content Analysis” [8] describes a project which has been done as a part of the University of Arizona’s Dark Web study of US domestic extremist and hate groups which have become a threat for the American homeland

security. This study is carried out to understand how these groups use the Internet to transmit their ideas and policies, to identify proper methods to collect web pages from the groups, and to automate methodical procedures for analysing and visualizing their web site content. Their approach constitutes of three parts, namely “collection building”, “link analysis”, and “content analysis” [8].

In collection building they have proposed a recursive method to have a relevant, most known and high quality collection of US domestic extremist and hate groups’ websites. This method includes the four steps “identifying seed URLs”, “conducting out-link and in-link expansions”, “filtering the collection” and “automatically collecting and processing extremist web sites”. When identifying seed URLs, they have identified a group of domestic extremist websites and searched URLs listed on the websites of watchdog organizations. In the step of conducting out-link and in-link expansions, once they have recognized seed URLs, URL out-links and in-links are extracted automatically using link-analysis programs. In the filtering the collection step, unrelated website URLs are removed from the collection, and when the extremist websites are identified, the last step with automatic spidering using “SpidersRUs” is performed to download the contents of the websites that the URLs are pointing to.

Link analysis is based on the hyperlink structure and by analysing it, it facilitates to discover extremist and hate groups’ framework and their organizational structure and also expose concealed communities in the associations between diverse websites for the similar group and the connections with other extremist group websites [8].

Based on the link analysis results, it is shown that neo-confederate and white supremacist groups have a common interest in defending the southern border and a common resentment for Hispanic illegal immigrants [8]. Similarly, white supremacy/neo-Nazi and Christian identity groups have the same ideology and occasionally they are grouped as one.

In content analysis, the paper authors have developed an attribute-based coding scheme which contains some high-level attributes like communications, fundraising, sharing ideology, virtual community, etc. to capture the content systematically [8]. This coding system allows determining and detecting how different groups utilize sharing of resources and fund raising, etc. The results of the content analysis also seem to be impressive. It is shown that the attribute of the highest frequency among six extremist groups is sharing ideology suggests the fact that target of extremist and hate groups is to depict their own characterization of their intentions [8]. It is also observed that the websites maintained by Nation of Islam seem to use highly sophisticated techniques such as using multimedia resources for a higher level in communications attribute.

## **2.5 Applying Authorship Analysis to Extremist-Group Web Forum Messages**

Abbasi’s and Hsinchen’s paper “Applying authorship analysis to extremist-group Web forum messages” describes the approach of authorship identification of English and Arabic online message forums which has been done as a part of the University of Arizona’s Dark Web research [9]. This experiment is based on the English messages obtained from a US forum that belongs to the White Knights as well as Arabic messages from the Palestinian Al-Aqsa Martyrs group. The authorship identification process includes three parts such as collection,

extraction, and experimentation.

Several characteristics of Arabic language such as inflection, diacritics, word length, and elongation make the process of authorship attribution difficult [9]. Therefore in order to address these issues in their approach, after collecting the messages from the forums and storing them in text and HTML format, they have used elongation filter, clustering algorithm, and root dictionary into the Arabic extraction process since it is more complicated than the English extraction process. After extracting the feature values from each message, they are grouped into four feature sets as F1, F2, F3 and F4. F1 feature set includes lexical features, and F2 contains lexical and syntactic features. In F3, lexical, syntactic and structural features are included. Similarly, F4 feature set consists of all the above features as well as content-specific features. They have used two machine learning classifiers (C4.5 and SVM) for this experiment. Throughout the experiment, they have used 30 random samples of five authors and 20 messages per author in both data sets and performed a 30-fold cross validation with C4.5 and SVM classifiers [9].

The accuracy for different feature sets using C4.5 and Support Vector Machine (SVM) classifiers shows that an SVM classifier is to be better than C4.5 with the highest accuracy when applied to all the features for both English and Arabic, where the accuracy gets increased with the addition of more feature types. “The difference in accuracy between classifiers across Arabic messages was far greater than it was in English messages: SVM outperformed C4.5 by more than 20 percent on all feature set combinations” [9]. To evaluate and illustrate the statistical significance of the additional feature types used, pairwise t-tests were carried out across the samples. It is shown that all feature types significantly enhanced classification accuracy for Arabic and English, except for content-specific words [9].

Decision tree analysis (C4.5) is carried out to see the outcome of the individual features for both English and Arabic forum messages. The result shows that syntactical and structural characteristics that include technical structure features (for example font colour, font size, embedded images, and hyperlinks) also are quite robust feature types across languages and these technical structure features are more important in classifying messages from the Al-Aqsa Martyrs [7]. Also to evaluate feature usage of both English and Arabic messages, they have constructed a graph that consists of only lexical and structural features since they are mainly language-independent.

## **2.6 Exploring the Dark Side of the Web: Collection and Analysis of U.S. Extremist Online Forums**

The article “Exploring the Dark Side of the Web: Collection and Analysis of U.S. Extremist Online Forums” by Zhou et al. states that “contents in extremist online forums are invaluable data sources for extremism research” [10]. In the study it is based on, the authors have found a web mining approach for collecting and monitoring extremist forums. It identifies such forums from different sources, and ended up with 110 U.S. domestic extremist forums with more than 640,000 documents. The paper mentions that chat rooms are difficult to monitor, because conversations there are often not saved after chatting sessions. It is easier with forums where messages are saved to be read and responded to later.

Zhou et al. propose a combination of expert knowledge and automatic web mining techniques

to monitor and collect data from extremist forums. The approach has three steps: forum identification, forum collection and parsing, and forum analysis. Zhou et al. have got to know about extremist groups from the U.S. government agencies and terrorism research centers. Then they have used some of the large search engines to find extremist forums on the own websites of the groups and at ISP's (e.g. Google Forums, Yahoo forums). After that, web spiders designed for retrieving data from these forums have been run for downloading text and attachment files (multimedia). Many forums require membership. In these cases the researchers have created accounts and logged in manually. The spiders had some arbitrary delays in their executions to pretend that they were human webservers. This approach also included update of collection when a spider revisited a webpage [10].

The paper mentions that different pieces of software are used in the forums. One of them is vBulletin, which is based on PHP, and is used in the forums Stormfront and Ummah in our research. The research of the paper used a spider for each forum software. It also included spidering of password-protected forums. Extremist forums often ask membership applicants for motivation, so application did not work always the first time. The spiders log in with username and password while it surfs through forums and downloads its contents. It is also important to avoid downloading multiple views of same contents or avoid redownload from a duplicate link. The paper always mentions that extremists, at least Neo-Nazis are attracted to large forums. Small forums stay small [10].

## **2.7 Authorship Attribution of Web Forum Posts**

“Authorship Attribution of Web Forum Posts” by Pillay and Solorio presents how authorship attribution of web forum posts can be performed by using both supervised and unsupervised learning techniques [11]. In this paper they have focused on recognizing stylistic characteristics of the writings of different authors of web forum posts. For this they have obtained data from the forums of the “Chronicle of Higher Education”, an academic portal, and for the experiment they used data sets with the number of authors ranging from 5 to 100. After they have extracted stylometric features (lexical and syntactical) of the data set and by using some statistical language models, a feature vector has been produced. Then, clustering has been used for the data set (feature vector) to identify stylistically similar clusters and the resulting solution clusters have been analysed. For this they have used the CLUTO software package to cluster the datasets into similar clusters. The analysis of these resulting cluster solutions help to understand how authors are dispersed within the clusters. “The measure used to evaluate the quality of clustering results is based on the entropy and purity values produced by the clustering methods” [11]. It is said that when the entropy values are small and the purity values large, then the clustering solution is good [11].

The unique cluster identifiers of the each resulting cluster also have been used as a discerning feature during classification where different machine learning classifiers/algorithms are trained using the lexical features, syntactical features and clustering solutions resulted from clustering to classify and to identify the authors. This approach is done in order to minimize the complexity and to improve the predication accuracy of the classification [11]. They have performed classification on six data sets using the data mining software toolkit called Weka with 10-fold cross validation and the machine learning algorithms used for the classification are Bayesian networks, decision trees(C4.5) and naïve Bayes. The results of the classification are encouraging that it is shown that they were able to classify authors in a higher accuracy by using above different classifiers. Different accuracy levels were achieved

on data sets when using different classifiers. It is also observed that for most of their data sets were able to achieve higher accuracy when they have used Bayesian networks and decision trees than the naïve Bayes. It is also shown that classification models which use clustering solutions (cluster identifier) as a feature along with the other stylistic features were able to produce higher classification results than the other models which did not include it.

## **2.8 Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Board**

In the paper “Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Board” one can read about techniques to detect whether two or more aliases (nicknames) really refer to the same author [12]. It is also described how to fuse these techniques.

The paper says that reasons for a person to have different nicknames can be that he/she wants to hide himself from analysts, has forgotten the password, or wants to give positive commentaries to his/her own posts in the forums [12].

Following entity matching techniques for identifying matching aliases are mentioned in the paper [11]:

1. Field matching of alias names, which can be done character-, token- or phonetically based. The first technique is best suited for names. Phonetics is not so well suited, since names from all over the world do not fit with English pronunciation. Token-based techniques are used for organizational names. There are two character-based techniques: edit distance and Jaro-Winkler metric, which are described in the paper.
2. Text analysis, which means to examine the style of writing, which again means to find which words are most frequently used by the author. Emotion symbols like smileys are also included here. It is normal to use Latent Semantic Analysis (LSA) and Singular Value Decomposition (SVD) to find documents (forum entries) that resemble each other. Truncated SVD is used to remove components from the document vector that describe seldom-occurring words. Ways of doing this are to find all values larger than unity (Kaiser criteria) or to find the elbow point of singular values in descending sequence (Principal Component Analysis, PCA).
3. Graph (network) analysis, which means that if two nodes representing authors in a graph have links/edges to the same set of other nodes, then the given pair of nodes likely refers to the same author. This is Social Network Analysis where one has found links between an author and other authors. The links are based on who has commented whom.
4. Spatio-temporal matching, which means that one looks at the points of time the different forum entries have been written. If two aliases post messages at almost the same hours of the day, this correlation may be used in combination with other techniques to increase the likelihood that the two authors are the same. The authors of the paper do not know about any previous attempts of using posting time for this

purpose.

The paper then describes how to combine results from the different techniques to get an answer if two authors really are the same or not. One way is to use classification voting with the results from the different techniques. A variation is to weight these classifiers with for example confidence. Another technique is to use Dempster-Shafer theory, which is thoroughly described in the paper.

The paper has no real practical calculation examples, but it describes well for which purposes the techniques can be used, namely identifying authors with duplicate aliases.

## 2.9 Contributions to our Project

From “Relationship Discovery in Large Text Collections Using Latent Semantic Indexing” we have gotten some ideas about how words can be related to each other in an indirect way. Instead of implementing LSA/LSI, we instead find words that are together in the same message by means of odds ratio. Words that are in the same message, but not in the same sentence may somehow be indirectly related to each other. We also got some inspiration for how to identify different aliases used by the authors of different forum messages, which we after all really did not get use of, because all author nicknames are mentioned explicitly in our forums. Additionally, we got the chance to observe and examine the LingPipe software in the initial phase of our master’s thesis project.

From “Names: A New Frontier in Text Mining” we again got an idea about use of different applications for name entity recognition, for instance how to find full names of persons in a text collection. In our project we just find first and last names separately, just like other words. In the word co-occurrence analysis of the forum Stormfront in section 4.20.6, we find some pairs consisting of a first name and a last name of a person.

In “Anomaly Detection in Extremist Web Forums Using a Dynamical Systems Approach”, we were inspired to consider analysing the frequencies of word usage in forums even though the method in that paper is different from our approach. The paper analyses changes of content in websites over time, while we only analyse the words in the content downloaded at only one specific point in time.

From “US Domestic Extremist Groups on the Web: Link and Content Analysis” we primarily learnt how extremist groups utilize the Internet in order to achieve their different goals. Some extremist groups use for instance the Internet to share their ideas and policies. They seem to share similar ideologies, while other groups are isolated. Some extremist websites seem to use highly sophisticated tools and technologies such as use of multimedia. Also how to collect forum data automatically inspired us to go for a feasible automatic download solution since we became exhausted by downloading forum messages from the forum threads manually one by one in the pre-project. Also how to filter off unwanted and unrelated web content such as HTML tags, contents outside the forum messages, stop words in the messages and so on were all taken into our consideration for the data collection phase in our master’s thesis project. We used Java code for the word filtering.

From “Applying Authorship Analysis to Extremist-Group Web Forum Messages” we

understood the significance of analysing the content of the extremist websites belonging to other languages than English, because many researchers only examine websites in English. Therefore, we were inspired to look at and analyse the contents of the forum messages downloaded from German and Norwegian extremist news websites and forums.

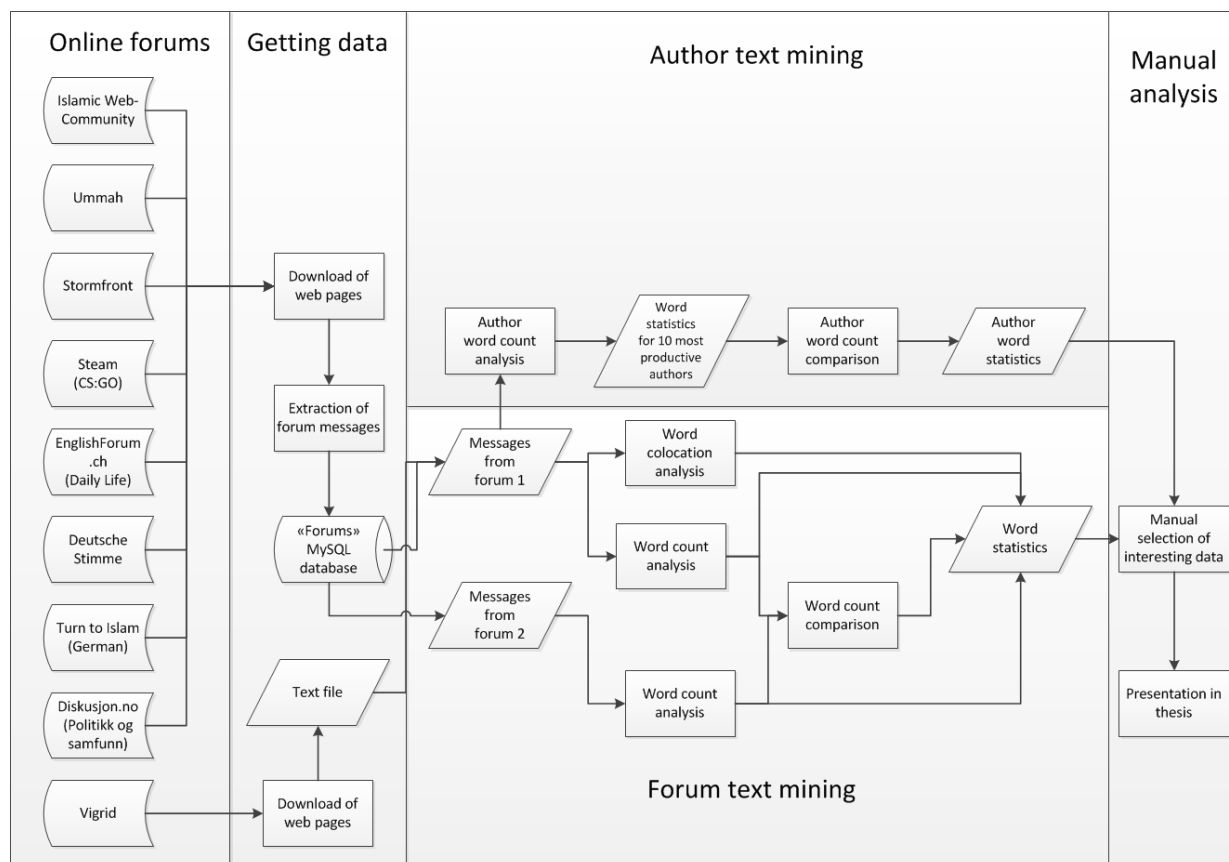
From “Exploring the Dark Side of the Web: Collection and Analysis of U.S. Extremist Online Forums” we understood the benefits of analysing forums instead of chat sessions, since chat sessions normally are not saved. Also we learnt about a mining approach for collecting and monitoring the extremist forums that uses different spidering techniques. We tried to use spiders for downloading websites, but we use instead use hard-coded webbots like described in chapter 3. We learnt about how important it is to focus on the methods of preventing re-download from a duplicate link or downloading multiple views of same contents. As described in chapter 3, we manually deleted duplicates of webpages that were downloaded by accident.

We learnt an approach for authorship attribution using both supervised and unsupervised learning from the study “Authorship Attribution of Web Forum Posts” which gave some inspiration to our project for the part of authorship vocabulary analysis. Our approach uses an unsupervised learning method, where we map the words written by the authors and frequencies of them in each author’s collection of forum messages. From the paper, we understand the importance of analysing the writing style of the authors, although we only consider the word frequencies.

“Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Board” helped us understand the different techniques used to find the different aliases used by a person in authorship attribution. Additionally, we got a chance to examine and get knowledge of different text analysis techniques focusing on the writing styles of different authors and their different applications of word patterns and emotion symbols like smileys, which has been already used in some cases in our author vocabulary analysis also. We were able to obtain some sort of knowledge about graph analysis which is very often used in social network analysis, which could have been useful for our author analysis, but because of limited time, should instead be a part of future work on this field.



# Chapter 3: Method



**Figure 2: Workflow for analysis of extremist forums**

In this chapter we describe our analysis architecture of the research behind this thesis. We divide the description into sections, like the sections in Figure 2: Workflow for analysis of extremist forums.

## 3.1 Online Forums

Before we start describing our method for the project, we first briefly present each of the forums/websites that we have downloaded, extracted statistical data from and analysed.

### Stormfront (stormfront.org)

Stormfront is an American white nationalist and supremacist forum site, which is actually known as the oldest major hate website. It appeared online for the first time in 1995 [1], short time after that the World Wide Web became accessible for everyone with an own computer and Internet connection. Members of the online forum are passionate about the purity of the white race, in which they seem to believe [1]. Although the forum is originally American, we have observed that there are texts in French, Italian, and Dutch in addition to English. We mostly ignore the parts of the forums that are not written in English.

## **Islamist forums**

We have downloaded pages from several Islamist forums, but decided to just analyse contents of

- Ummah ([ummah.com/forum](http://ummah.com/forum))
- Islamic Web-Community ([myiwc.com/forums](http://myiwc.com/forums))
- the German section of Turn to Islam ([turntoislam.com/forums/deutsch-german.48/](http://turntoislam.com/forums/deutsch-german.48/))

We downloaded much more text from Ummah than the others. All of the Islamist forums, especially the English ones, include some Arabic text. The German forum also included some English texts.

## **Steam Community – Counter-Strike: Global Offensive**

We have downloaded parts of the forum associated with the computer game “Counter-Strike: Global Offensive” (from the URL: [steamcommunity.com/app/730/discussions/](http://steamcommunity.com/app/730/discussions/)) for comparing the vocabulary of the each extremist forum with the words in a forum written by ordinary people.

## **Swiss English Forum ([englishforum.ch/daily-life/](http://englishforum.ch/daily-life/))**

We have also compared the vocabulary of the extremist forums with a daily life forum on a Swiss English-language website. Many of the forum writers seem to be foreigners living in Switzerland. The forum is associated with the English news website “The Local” [13] which in addition to Switzerland also has versions for France, Germany, Spain, Sweden, and Norway [14]. We could not find a similar forum site for Norway.

## **Vigrid ([vigrid.net](http://vigrid.net))**

As Norwegian research material we have downloaded text from the website [vigridtvedt.net](http://vigridtvedt.net), which is not really a forum, but includes a lot of articles commenting politics in a right-extreme fashion. After download it turned out that the site also is on the domain [vigrid.net](http://vigrid.net). This website is associated with the Norwegian political organization Vigrid, which in a way blends Norse mythology together with racial doctrine. We compare the contents of this website with the Norwegian political forum at [diskusjon.no](http://diskusjon.no). Vigrid.net claims that there is a white race supreme to other people, like the Stormfront forum does. In addition it does not believe that the mass extermination of Jews during the Second World War really found place [15].

## **Diskusjon.no – Politikk og samfunn ([diskusjon.no/index.php?showforum=57](http://diskusjon.no/index.php?showforum=57))**

As an ordinary Norwegian forum we selected to download text from the politics and society forum on the website [diskusjon.no](http://diskusjon.no). As the title implies, the forum participants discuss ideas and events in Norwegian politics.

## **Deutsche Stimme ([ds-aktuell.de](http://ds-aktuell.de))**

We have downloaded news articles with comments from the website of the German right-extreme newspaper Deutsche Stimme (which means “German Voice”). We wanted to

compare these articles with the contents of an ordinary, political forum in German, but due to download problems with that forum for our webbot, we just compared this Neo-Nazi online newspaper with the German part of the Islamist forum Turn to Islam.

### **Reasons for analysing the selected forums**

We can of course not guarantee that all authors in an extremist forum are political and/or religious extremists, but much of what is written there seems to support that many of them have extreme opinions, and that the forums are actually extremist forums. The Islamist forums are recognized as extremist forums by the Dark Web Project at the University of Arizona [3]. (One must request for access to the Dark Web Portal to visit this reference.)

The Stormfront forum we got to know by reading a paper. The Vigrid and Deutsche Stimme news sites were found by means of ordinary web search and recognized by our knowledge from mass media of the right-extreme, religious (Ásatrú) organization Vigrid and the German right-extreme nationalist party NPD, respectively. The sites themselves and online encyclopaedia articles confirmed that the sites and organizations belonged together. The Vigrid website seems to be up and running again now after it was closed down in 2009 [16].

Extremist forums are compared with ordinary forums for highlighting the words that make extremists differ from other people in vocabulary and discussion topics. Islamists are compared to right-extremists for detecting the differences in vocabulary and discussion topics for these two groups. There may also be some similarities. Details can be read in chapter 4.

## **3.2 Getting Data**

We first downloaded forums by means of PHP scripts written and customized specifically by us for this project. The code of these scripts is based on code examples in the book “Webbots, Spiders, and Screen Scrapers” [17]. The author of that book has written many useful, simplified PHP/CURL functions for webbot programming that we applied a lot in our PHP code. We programmed a webbot for each forum (except for Ummah). Much of the code in each program is taken from the code written by the author of the mentioned book, but we customized it for each of the forums. A webbot looks for and downloads webpages including forum threads and store them in text files on a personal computer’s hard drive. These webpages are downloaded as they are, namely HTML files. The webbot downloads forum threads that are accessible from the start page of a forum website, like “myiwc.com/forums” (an Islamist forum called Islamic Web-Community) or “stormfront.org/forum (a white supremacist forum called Stormfront). The webbot is specifically programmed to follow the links on the front-page that lead to forum pages which in their turn include lists of links to forum threads. The links can be recognized by first finding the HTML elements in which they are situated (for example “div” elements with a distinguishable and recognizable class or id attribute) or by simply finding all link attributes in the “a” HTML elements on a webpage. The problem with just finding all links on a web page and downloading all pages they are leading to is that one then often downloads pages that are not a part of the forum, sometimes even from outside the website. Therefore the webbot always checks whether the link belongs to the same domain, that is, it starts with the address of the front-page of the entire website, the root address. We customized the webbot code for each forum, because the structures of the forums were different, so we ended up with an own webbot for each forum.

### 3.2.1 Download of Webpages

Links to forum pages, which again include lists of links to forum threads, often have an address like “<http://myiwc.com/forums/forumdisplay.php?f=63>”. First the root address appears, then an argument represented by the name of a PHP script (“forumdisplay.php”) with a following URL argument which refers to a forum number, like “f=63”. In some forums there is no PHP script name in a forum page address, just a root address and an argument. The links leading to forum threads include the root address, in some forums the name of a PHP script for showing a forum thread (“showthread.php”), and always at the end a URL argument that consists of the letter t, an equal sign, and a thread number. An example of a thread link address is then “<http://myiwc.com/forums/showthread.php?t=3750>”.

By finding all forum threads by following the links from the front-page as described above, a webbot certainly collects enough data for some useful research in extremist forums. Often one only gets the newest threads, which mostly represent enough data for the analysis. The older threads may be hidden away in the sense that links to them are not visible on the front page. Another issue is that sometimes a thread may be so long that it needs to be divided into several web pages. Then one will miss the contents of a forum thread beyond the first page since each of our PHP webbots are programmed to just download the HTML files (web pages) that each thread link in each thread list points directly to. One could of course try to test by code if there exists more than one page for a forum thread by checking if there is a link to a last page of the forum thread either on the first page of the forum thread or on the forum thread list page. One can then see which page number that last page has.

We figured out that it was easier to download complete websites with “WebSite Downloader for Windows” [18]. This program can download complete websites by just giving the address of the front-page. A little problem was that this software downloaded too much data. Many of the pages were duplicates, so we had to delete several of them manually. Luckily, the file names of the duplicates followed patterns which made them filterable in the search feature in Windows Explorer. Duplicate HTML page files were recognizable by a forum message number after the name of the page itself in the filename. This peculiar duplicate download is caused by the fact that there exist links to the thread pages as well as the positions of the messages themselves on the pages in an online forum. We use WebSite Downloader for downloading the Islamist forum Ummah. Unfortunately it did not work so well for downloading the other forums, so we continued to use the other techniques we had found for downloading the other forums.

For downloading the Vigrid website, we use Java we wrote with import from the jsoup library [19] both for downloading and text extraction. It downloads the front-page and all webpages at [vigridtvedt.net](http://vigridtvedt.net) it links to. All the text in Vigrid is written in different “p” (paragraph) HTML elements on the different webpages. The Java program stores each webpage’s text in one of the lines of the text file `vigridtvedtnet.txt`. We just assume for Vigrid that a whole page is a message, because it is actually practically impossible (at least difficult) to code a webbot to find out where the messages start and end on a page at [vigrid.net](http://vigrid.net). All the text is namely as written in HTML paragraph elements, and these have no attributes for distinguishing them from each other. From all the other author forums we download contents

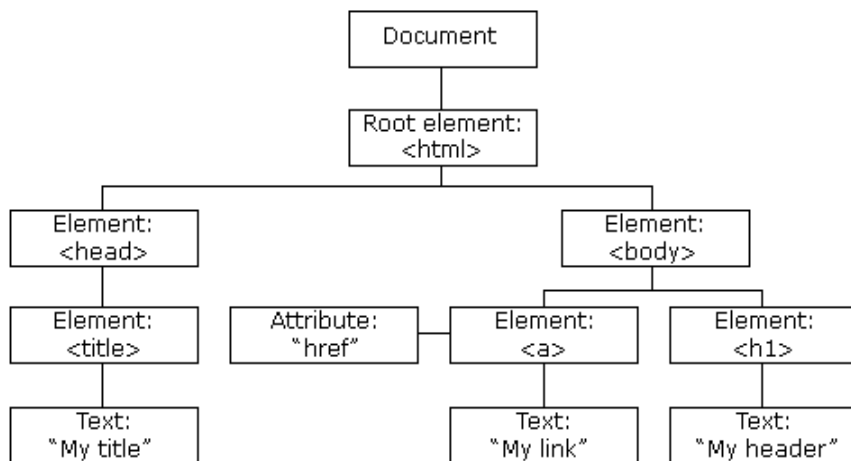
by means of PHP-coded webbots, which we have briefly described earlier in this section. We then only get the first page of each forum thread, which actually gives us enough data.

### 3.2.2 Extraction of Data from Webpages

Before extracting data from the downloaded HTML documents, one has to find out in which HTML elements the needed data are stored. This search for correct elements can be done in an easy way by using the plug-in Firebug for the web browser Firefox [20]. It is really meant for finding syntactical or logical errors in the code of a website, but it also works fine for finding the different HTML elements and attributes. Firebug shows the HTML elements in an expandable tree hierarchy instead of the “flat” listing of HTML code one sees when reading source code in text editors.

On the front-page, one must find manually find the links to the different forum thread lists. They are often in HTML elements that follow a common pattern like having an equal attribute value. A “class” attribute could for example have the value “forum-link” in all these HTML elements. After finding this pattern, we can write webbot code for going to those webpages. Then we must also find a HTML link pattern on these forum thread list pages, so that we can write proper code for finding the links to the forum threads and downloading the forum threads to the webbots, then one runs the code to download the pages, and finally one writes and executes the code for extracting the data to a MySQL database.

To extract data from the downloaded HTML files (like section, we use HTML DOM through the Java code library jsoup [19]. HTML DOM is a standard object model and programming interface for HTML, and can be used to get, change, add, or delete HTML elements [21]. We only use the get feature, because we are only interested in extracting data from the forum messages on each forum thread page. In HTML DOM, one refers to different objects as nodes. Among these nodes are HTML elements and their attributes. An example of a HTML DOM node hierarchy is shown in the figure below [21]:



**Figure 3: Example of HTML DOM node hierarchy from W3Schools [21].**

The text between the opening and closing tags of an HTML element is an own text node [22]. Text in HTML elements can be accessed (and manipulated) with methods. With HTML DOM, one can get the text between the tags of an HTML element by means of the text() method. What is nice with this feature is that if there are nested elements in the HTML document, then one can get all text from several elements nested into each other just by

requesting the text HTML DOM node of the outermost element with just one line of code. This could be a problem if one only wants the text from the outermost element, and not the inner nodes.

Text in attributes can be accessed with the method `attr(attribute_key)` where `attribute_key` must be for instance equal “href” to access the text in the href attribute of the current HTML element.

From each forum message in all forums, we extracted these data and stored them in a MySQL database table called “forum entries”:

- forum
- thread
- title
- message
- author
- publish date and time
- sequence number (not in database)

For the text analysis in our project we use plain text from the title and message body in the different forums. They are stored in the title and message attributes in the database table “forum entries” by means of the Java program which also extracts the text. In some few cases, there was no title, so we inserted a default title like “untitled” or “comment” into the “title” attribute in the database. The attribute “forum” was used for selecting the forum entries belonging to given forum for the forum mining and analysis, as described in the next section. The “author” was likewise used for selecting the text written by a given author for the author text mining and analysis. We did not use the attribute called “thread” which included the name of the forum thread a message belonged to.

We thought about using the date and time attribute for finding out how long it could take before a forum entry discussing a specific topic was answered, or how long a discussion could last. At the end, there was no use for this time attribute, since we decided to only care about what was written, not when it was written. We have kept the date/time attribute in the database just for showing curious readers how old a message is. The date formats in the different forums are not directly compatible with each other. Sometimes the date is only “X days ago”. We could then of course compute the date from the time the message was downloaded.

Finally, we did not use the sequence number, the place a message has in the sequence of all the entries in a forum thread. It could have been a good idea to use data about how late or early an author starts to write in a forum thread. We could then see whether an author was a discussion starter or a latecomer. Such a feature turned out to not be needed, so it is not even an attribute in the database table.

We did not store text from Vigrid in the database. We instead used a customized Java program for performing automatic analysis directly with the text from the different Vigrid webpages, which was all stored in a text file with one line for each webpage. A webpage is then regarded as a “forum message” in the analysis.

## 3.3 Forum Text Mining

For the following mining and analysis, we use titles and message texts stored in the MySQL database mentioned in the previous section. A title and a message combined are then defined as one document in the analysis. We analyse text from Vigrid with each line in the text file `vigridtvedtnet.txt` defined as one complete message or document. We find words that are typical for a forum by finding out which words that are much used in that collection of text from the forum, but it does not have to have used in every single thread. We also find pairs of words that tend to be in the same message

### 3.3.1 TF and IDF

For finding words that are typical for a document, which in this thesis corresponds a forum message, one finds the term frequency (TF) of a term (word or expression) in an entire document (which in our case is a forum message) and then multiplies it with inverse document frequency (IDF). In our thesis a term is always the same as a word. If a word is used in nearly all the forum entries and therefore has a high document frequency, the word is given less weight by a low inverse document frequency. The product of TF and IDF we call TF-IDF. IDF is basically the ratio of the total number of documents to the number of documents which include the term (word) we are interested in. When only the order of the TF-IDF values of the words in a forum is important, and not the values themselves, it is common to use the logarithm of what we spontaneously think IDF is [23]. We therefore define IDF as the logarithm of the real IDF in this project, which seems to be a normal definition in research. See formula in Figure 4.

$$\text{idf}(t,D) = \log \frac{|D|}{|\{d \in D \mid t \in d\}|}$$

**Figure 4: Definition of logarithmic IDF**

### 3.3.2 GTF and NGTF

Since the messages in a forum often are very short, it is more interesting and useful to apply the global term frequency (GTF), an expression we have invented in this project, instead of the term frequency (TF), which only counts words in single messages. GTF is the count of occurrences of a term (word) in all the text we have downloaded from a specific forum. TF values, as they are defined in the traditional definition of TF-IDF, are useless to us because many of the messages are very short. We are namely interested in how different words are used in a complete forum as a whole, not how they are used in each single message.

Analysing each single message would anyway be tedious, since a forum often has thousands of messages. If we want a GTF-IDF value, we multiply the GTF of a word with the IDF of the same word. IDF is then defined in the traditional way like described recently in this section and shown in Figure 4. IDF is not used so much in the analysis as first intended. We will explain why after introducing NGTF.



When comparing frequencies of words in two different forums with different sizes, we normalize the GTF value to get a value which we call normalized GTF or NGTF. NGTF is the GTF of the actual word divided by the maximum GTF value of any non-stop word in the forum, like shown in Figure 5.

$$\text{NGTF}(\text{word}) = \frac{\text{GTF}(\text{word})}{\text{GTF}(\text{most used word})}$$

**Figure 5: Definition of NGTF**

As a consequence of the definition, NGTF always has a value in the closed interval [0,1]. Although the theoretical minimum is 0, that value never appears in our result lists because we have no words in the analysis which are never used in the forum. We have also excluded all words that are used less than ten times in a forum, because they are practically unimportant for the analysis of a forum, where we find words which are used hundreds and even thousands of times.

We have implemented word counting with Java, so we get words with accompanying GTF and IDF values in a CSV (comma-separated values) or TSV (tab-separated values) file for each forum. We use formulas in Microsoft Excel 2010 to compute NGTF and NGTF-IDF values, because we then first can manually delete frequent stop words which are left after the removal of several stop words in a list in our Java word count program. Stop words that are left in the CSV/TSV file, are of course unimportant and uninteresting words that we did not think of in advance. We then assure that NGTF is defined with a non-stop word as the word with maximum frequency in the forum, as shown in Figure 5.

NGTF-IDF of a specific word is the product of NGTF and IDF of that word. NGTF-IDF is sometimes used for word ordering in the most-frequent-words tables for single forums in chapter 4 of this thesis. IDF is meant as a factor for giving stop words a lower rank in the list with NGTF-IDF. (N)GTF-IDF is not really suited for comparison. NGTF works better, because it is easier to use words for the comparison like “used more in forum 1 than in forum 2 (if the two forums were of the same size)” for NGTF.

### 3.3.3 Forum Word Count Comparison

By means of NGTF ratios we look for words that are much more used in one forum than in another forum it is compared to. By means of the words more frequent in one forum than another, we can discover which topics are more discussed in one forum than another. Words are stored together with NGTF ratios in a CSV/TSV file for each forum, ready for manual analysis.

NGTF value computation is necessary because two different forums often do not have the same size such that GTF values are not very suitable for comparison. Two forums are compared to each other by finding all words that are in both forums and then computing the ratio of the NGTF of each such word in forum 1 to the NGTF of the same word in forum 2, and vice versa. We refer to these ratios as f1/f2 NGTF ratio and f2/f1 ratio, respectively. The

f means of course “forum”. We can of course instead compare the GTF value of a word in one forum with the GTF value of the same word in another forum to find out how many times more a word is actually used in one forum than another. However, if a forum is extremely much larger than the other it is compared to, then the GTF ratio values will not fully make sense for comparison, because if a specific word is for instance used 1000 times in a forum A with totally 100000 words, and 1000 times in a forum B with totally 200000 words, we will not get a quite correct image of the situation if we say that a word is used with same frequency in both forums, such that the GTF ratio equals 1. This is namely true for absolute frequency (GTF), but not relative frequency, or what we have used: normalized frequency (NGTF). Therefore, normalization is necessary for simulating that the two forums are of the same size when comparing them. It then seems like the contents of the smaller forum are repeated until it reaches the size of the bigger one.

The comparative word analysis also includes finding all words that are in one of the forums, but not both. The importance of each of these unique words is ranked by GTF, so that we easier can find the words that better characterize the difference from the forum it is compared to. The important and interesting words are then normally the frequent ones and sometimes also the words down to the medium-frequent words. With these words we can find the topics that are discussed in one forum and absolutely not in another forum. We may get some few false-positives of such forum-unique words in the list because a word in one forum may be used less than 10 times in another forum. If a word is actually used for instance 9 times in one forum and 100 times in another forum, then it is at least used considerably much more in the latter forum, so the result is then not completely wrong.

Results can be seen in sections 4.1-4.18.

### 3.3.4 Word Colocation Analysis

The tendency of two specific words occurring together in the same forum message can be measured by a joint odds ratio as it is defined in Figure 6.  $X$  or  $Y$  in  $p_{XY}$  are there binary variables representing whether a word  $X$  or a word  $Y$  occurs in a forum message or not, respectively. The value 0 means that the word is not in a specific message, while 1 means that the word is in the message.  $p_{XY}$  is then the probability that the words are together in the same message in a forum. More precisely, it is the frequency of how often two specific words are in the same message.

Odds are the ratio of the probability of some event to happen to the same event not to happen. In our case, odds ratio of one word is the frequency of messages with that word divided by the frequency of messages without that word. This can be described briefly as

$$\text{odds}(Y) = p_{Y=1} / p_{Y=0}$$

where  $Y$  represents the existence of a specific word in a forum message.

Odds ratio is the odds of something to happen in one group divided by the odds of the same event to happen in another group [24]. In our application of the odds ratio, these two groups are actually the event of another specific word  $X$  to occur in the same message as word  $Y$ , and the event of the same other word  $X$  **not** to occur in the same message as word  $Y$ ,

respectively. In Figure 6, odds(Y) for the group where word X is in the message (X=1) is defined to the left of the big slash (the numerator of the joint odds ratio). Odds(Y) for the group where word X is **not** in the message (X=0) is defined to the right of the big slash (the denominator of the joint odds ratio).

The ratio of these two odds is in a sense the odds for word X, but more formally and precisely it is the definition of the joint odds ratio for X and Y. This can be simplified to the fraction to the right of the equal sign in Figure 6. The numerator is the product of the probabilities/frequencies of both words being and not being in the same message, respectively. The denominator is the product of the probabilities (frequencies) that only word X is in the message and that only word Y is in the message, respectively.

$$\frac{p_{11}/(p_{11} + p_{10})}{p_{10}/(p_{11} + p_{10})} \bigg/ \frac{p_{01}/(p_{01} + p_{00})}{p_{00}/(p_{01} + p_{00})} = \frac{p_{11}p_{00}}{p_{10}p_{01}}$$

**Figure 6: Joint odds ratio, formula from [24]**

For the probability values we can actually choose to use relative, normalized or absolute frequencies just as we want, since they will all give the same result, because of the possibility of reducing and expanding the fraction. Therefore we use absolute frequencies (GTFs) for simplicity in our Java program for calculating odds ratios. It is important that  $p_{10}$  and  $p_{01}$  in Figure 6 must not equal 0, or the odds ratio will be equal infinity [24]. To avoid infinite values, we set  $p_{XY}$ , where  $x \neq y$ , to 1 when these  $p_{XY}$  values really equal 0. The values for the odds ratio defined this way are then not formally correct, but we get a sensible order of the odds ratios of the different words when they are compared to each other.

The odds ratio of each word pair in a forum is computed for finding how closely the two words are related to each other in the sense of co-location. The algorithm for doing this unfortunately runs slow if we analyse thousands or even tens of thousands distinct words in a forum. It can take days to finish an execution if we need to find the odds ratios for all possible pairs of words. Therefore, some words that are used only a few times, are excluded for the analysis. How many words that we had to exclude is mentioned together with the results in section 4.20.

Then the ratio of the odds ratios of the same word pair in two different forums can be computed for comparing how much closer the same two words are related to each other in one forum than in another. This of course depends on whether the same colocation or co-occurrence word pairs are in both the forums to be compared.

## 3.5 Author Text Mining

We use GTF (and GTF-IDF) in the thesis with “global” defined as all the text a given author has written. (A document is then still defined as a forum message in IDF.)

We have analysed the vocabulary of five of the ten most productive authors. We have found the frequency (GTF) of all words in each author’s vocabulary and manually searched for words that seem to be characteristic for each author’s writing. One must notice that an author does not always use his/her own words. Sometimes they quote some other author(s) or a news

source outside the forum. We still consider the words in quotes as characteristic for an author since they seem to describe what the author is interested in.

To characterize each author even better, we find all words that are unique for each author relative to the 10 most-writing authors. We also find the words all of the ten most productive authors have in common in their vocabulary.

Results can be seen in section 4.19.

### **3.6 Manual Analysis**

After preparing lots of statistical data with word frequencies and odds ratios, we must analyse the results manually. A word that has a high GTF in a forum must be somehow important for that forum. The exception is of course stop words like “and”, “so”, “if”, etc. We mostly do not need to think about them in the manual analysis, because many of these words are filtered away from the GTF, GTF ratio and odds ratio lists. Sometimes there may be some stop words left in the list, which we did not think of removing in advance. If they were on the top of a GTF list, we deleted them, so that the NGTF values were properly computed. NGTF is namely a GTF value divided by the maximum GTF value in the forum. Since we do not want to include stop words in the analysis in the first place, it would be wrong to have the frequency of a stop word as the basis of NGTF frequency. Other stop words that are left in the list, we just ignore. If there are stop words from other languages than English in a list, they are only of interest for confirming that these languages are used in the accompanying forum.

Our programs for downloading websites and extracting frequency statistics generate much data. When we select words to present in the report, we have to think a little like journalists and ask ourselves the question “What is interesting for the reader, and what does he/she find important?”. By stating that, we of course do not mean that we manipulate the results such that they seem more interesting. What we do is to select the words that are relevant for presentation and sometimes also check if some of the seemingly irrelevant words also really are important to present. The results we found worth presenting, can be read in the following chapter, “Results and Analysis”.

## Chapter 4: Results

When we present words from the extremist forums and some ordinary forums in this thesis, we can just present some of the most interesting words that seem characteristic for each forum. We cannot present all words, because we have about 230 megabytes of plain text as raw data downloaded from the different online forums. What may be an issue is that our selection of words to present in this chapter may be influenced by our prejudices towards the different groups in the forums. We expected in the beginning of the project to find a lot of hate propaganda against Jews, Africans, and other originally non-European people in Nazi and white supremacist forums. We also expected them to discuss events from the Second World War. From Islamists we expected a lot of discussion about religion, and especially Islam. We can say that for both of these extremist groups, many of our pre-assumptions were somehow correct. However, there may of course be a possibility that we ignore some important words and discussion topics in the forums when we just look for what we expect. We have therefore also tried to look for words that are not expected, but at the same time also sensible for discussion in this thesis.

We first have a look at the words that are typical for each of the English extremist forums in sections 4.1-4.3

- Ummah,
- Islamic Web-Community
- Stormfront

and each of the ordinary forums in sections 4.4 and 4.5

- Swiss English Forum
- Steam (CS:GO)

Then we compare each of three extremist forums with each of the two ordinary forums in a pairwise manner in sections 4.6-4.11. We compare word frequencies in extremist forums with the corresponding word frequencies in ordinary forums to find out what extremists write more about than other people, and also what extremists write about and others not. We also compare in the opposite direction for trying to figure out which words make a forum non-extremist. In section 4.12 we compare the Islamist forum Ummah with the White Supremacist forum Stormfront to find differences in vocabulary and discussion topics for these two kinds of extremists.

As comparison to the English extremist forums, we analyse “Counter-Strike: Global Offensive” (CS:GO), an online team-based multiplayer game [25], and with a Swiss forum about daily life. The game forum is about a game where the goal is to fight terrorists, which are extremists that use violence to spread their message. One can also play as terrorists that plant bombs and fight against the antiterrorist police. However, ideologies of any kind are not discussed in the game forum, so it therefore serves well as an opposition to the extremist forum, such that a vocabulary comparison highlights the words and topics that should be typical for terrorists and not for computer game players. A lot of people spend time playing games every day, so it could be regarded as a daily life activity, although everyone does not do that. For covering other kinds of day-to-day topics, we have included a forum with the

self-descriptive name Swiss English Forum. It is associated with a Swiss news site in English [13]. It discusses only activities like school, parties, and shopping. There are some jokes about Swiss stereotypes, but no political discussion, and absolutely no extremist discussion. These two forums related to some kind of daily life activities are good for filtering out and highlighting words that are extremist-related from words related to daily life.

In sections 4.13 and 4.14 we present the most frequent words in the Norwegian extremist website Vigrid, and the ordinary, political forum Diskusjon.no before we compare these two with each other in section 4.15. Diskusjon.no is a political forum, so by comparing this with an right-extremist forum, we can figure out which words and topics are related to extremist politics, and which related to ordinary politics, like party politics.

Finally, we will present and compare the vocabularies of the German extremist websites Deutsche Stimme and Turn to Islam in sections 4.16-4.18. We compare right-extremists with Islamists again, like we do with Stormfront and Ummah in section 4.12.

Section 4.19 discusses author vocabularies in Ummah, and Stormfront. Section 4.20 looks into pairs of words frequently occurring in the same forum message.

When we refer to a line with a forum message sample in this chapter, it is a line in the file “forum\_entries.sql” unless we specify something else. If an example forum message is too long to be presented as a whole in this thesis, we mark places where text is removed with “(...)”.

## 4.1 Word Frequency Analysis of Ummah

Not surprisingly, the most frequent word in the Islamist forum Ummah is “allah”, the Arabic word for God. It is used 23132 times. Of course, there are some stop words that are more frequent than “allah”, but they are excluded from the analysis because they are not important. In comparison the word “people” is used 10467 times.

Although they are not stop words, some words are used in so many forum posts that they actually become less interesting for the analysis. The words “posted” and “originally” do not get so high rankings in list of words sorted in descending order by GTF-IDF compared to the rankings in GTF. The order does not change when GTF values are normalized. Each of the words “posted” and “originally” are slightly more frequent than “people”, likely because of the expression “posted originally by” which is used when mentioning which writer is being quoted in a forum message.

The word “allahu” is used 5959 times, and the word “akbar” is used 5616 times. The Arabic expression “Allahu akbar!” means “God is the greatest!”. This expression is therefore used maximum 5616 times, but not necessary that many times. The 343 remaining times “allahu” is applied, it is likely written together with other spelling variants of “akbar”, like the 12 times “akhbar” is written, or “allahu” is used together with completely other words.

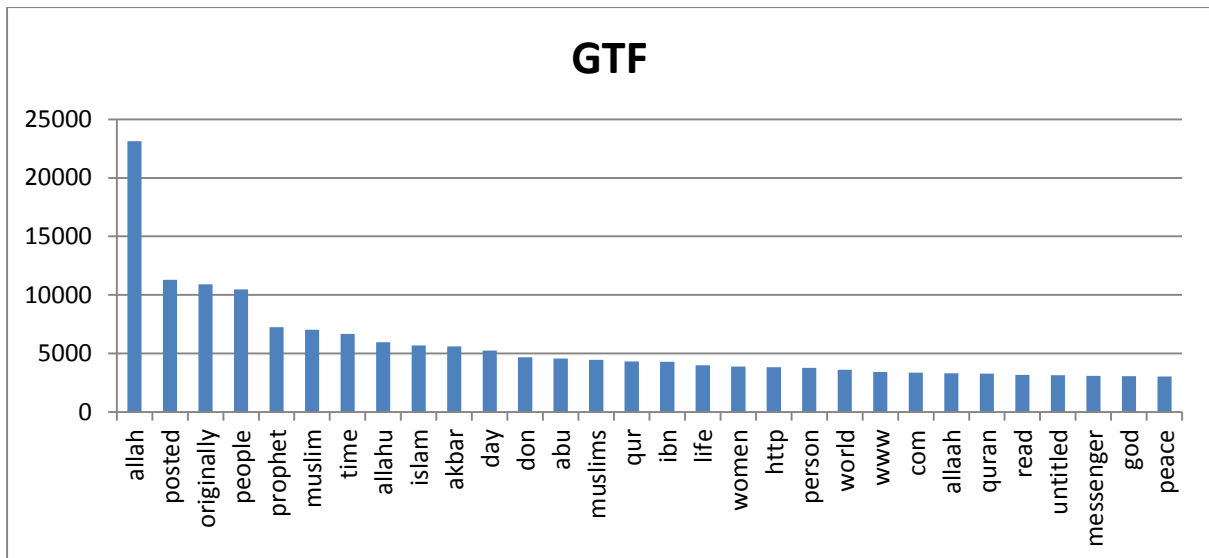


Figure 7: 30 most frequent words in Ummah

## 4.2 Word Frequency Analysis of Islamic Web-Community

Several of the words in Islamic Web-Community seem to be the same as the words in Ummah, which makes sense, since they are both Islamist forums. A lot of words are namely related to the religion Islam. As we see in Figure 8, “allah”, the Arabic name of God, is of course on the top, then follows “hajj”, the name of the pilgrimage to the holy city Mecca in Saudi Arabia. Among the top 30 words we also find “prophet”, “peace”, “salaama”, “akbar”, and “Islamic”, which all somehow seem related to the religious content. Many of the other words may seem to be technical terms related to how the forum itself works, like “php”, “style”, “nuke”, “www”, “site”. We have figured out that this is actually due to a small programming error, which resulted in extraction of the text in other HTML element(s), when the titles really should have been extracted. One can see an example of that in line 496:

***‘(Islamic Web-Community’, ‘f2-t52’, ‘Evolution style ©EX - Evolution Design All logos and trademarks in this site are property of their respective owner. The comments are property of their posters, all the rest © 2002 by Islamic Web-Community Web site Engine’s code is Copyright © 2002 by PHP-Nuke. All Rights Reserved. PHP-Nuke is Free Software released under the GNU/GPL license. :: Evolution style ©EX - Evolution Design :: phpb2 phpb2 style by phpb2.de :: PHP-Nuke theme by www.nukemods.com ::’, ‘Nokia has made available, via its website, a wide range of Ramadan related content for a number of phones (including all recent S60 devices). (...), ‘Salahudeen’, ‘20-08-2008, 22:21’)’***

The “title” is shown with bold font. As one can see, this can impossibly be the title for a message about Ramadan-related phone content. When we look at the original website, we see that the “title” really is text from the bottom of any webpage of the forum [26].

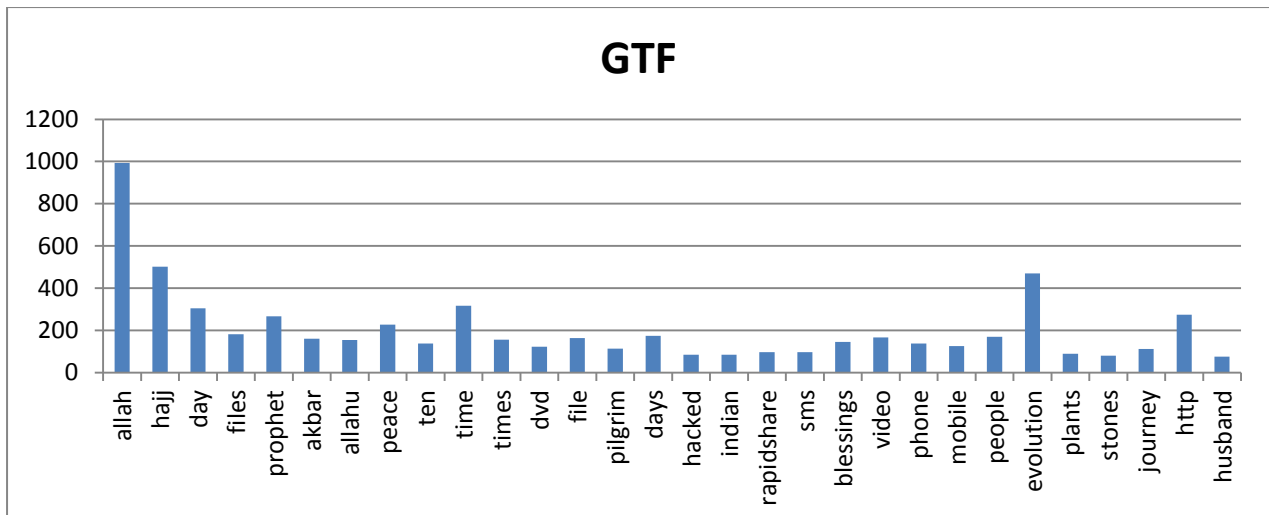


Figure 8: 30 most frequent words in Islamic Web-Community

### 4.3 Word Frequency Analysis of Stormfront

In the American white supremacist forum “Stormfront” the word “white” is mentioned 25769 times and is actually used more than the word “quote” (23285 times).

The plural “whites” is used 6980 times. It is also remarkable that “black” has been written 8564 times, and “jewish” 7137 times. Amongst the most popular words, we also find “race” (7016), “obama” (5652), “american” (4600), “hitler” (3776), and “negro” (3400). These words indicate that there should be much talk about races and nationalism in the Stormfront forum. Even the word “racist” is actually used 3267 times. Here is an example from line 47164:

*“(‘Stormfront’, ‘908090’, ‘Facebook banned me, I posted a racist comment on a news site’, ‘I posted a comment on a news site, like fox news or something, and facebook banned me for it??’, ‘Hellrazor777’, ‘08-20-2012, 01:19 AM’)”*

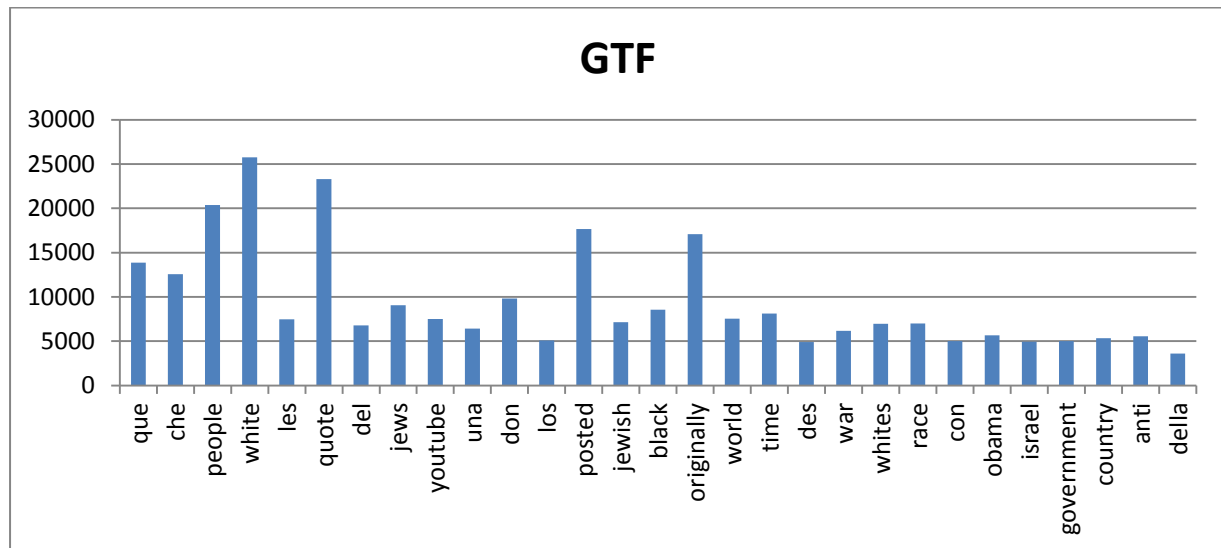
That the author admits he/she wrote a racist comment in the commentary field of a news article connected to Facebook, should imply that he/she does not regret it. From the example it does not look like the word “racist” is regarded as pejorative by that author.

Some everyday words are also applied in Stormfront, although they may be used for propaganda purposes: “family” (3268), “children” (3263), “money” (3426). “youtube”, the name of a video website, is actually mentioned 7509 times.

Some of the words in Stormfront are unintelligible if English is the only European language one understands. Among the 100 most occurring words, one can find words from French (“que” = “what”), Italian (“sono” = “I am”), Spanish (“por” = “for”), and Dutch (“het” = “the”, neuter form). Most of these are stop words. Something interesting is that when words are sorted by (normalized) GTF-IDF, then the non-English stop words are ranked higher than with (N)GTF, because they are seldom used. Else there is no noticeable difference in which words that are on the top when sorting by (N)GTF or (N)GTF-IDF, respectively, except for some small differences in the ordering. The appearance of some non-English words in the list of most frequent words suggests that Stormfront is today an international forum with mostly



English forum threads, but also a considerable amount of threads in other European languages.



**Figure 9: 30 most frequent words in Stormfront**

The list of 30 words with highest GTF includes the word “obama”. Barack Obama is the first coloured president in the USA, so it is likely to believe that white supremacists do not like him. The word “world” (7538) as well as the word “war” (6152) are both mentioned often. Many times it is very probable that they are mentioned together in the expression “world war” like in “World War 2” or “Second World War”. The word “anti” is used 5560 times, and has to be the prefix of different anti-ideologies like anti-Semitism, anti-racist, or another kind of opposition-expressing word like “anti-White climate” or “anti-German”. “anti” appears in the results because hyphens are removed from the analysis program before the counting of the words started. Hyphens are namely considered as punctuation characters by the regular expression “`\\p{Punct}`” [27] which we use recognizing the punctuations to be removed by the pre-processing feature of our word counting Java program. Punctuation characters must namely be removed from the text to avoid that they are included as parts of words like at the end of sentences. An example is “(...) **race.**”, from where would then get a word with the full stop included, not only the word “**race**”. The program would then consider “**race.**” as another word than “**race**”. This will influence the count of “**race**”. Hyphens are also removed, so “**anti-White**” becomes “**anti**” and “**White**”. This is a side effect we did not see in advance, but is not a big problem for our results.

As one can see in the tables below, which are extracted from the table of words in Stormfront, one is more likely to find a variant of the word “nazi” than of “zion”, “zionist” or “zionism”; actually it is 1.84 times more likely. Because of different languages there are many variants of the words “nazi” and “zionist”. “zionist” starts with S in some languages (like Norwegian). Zion is a mountain close to Jerusalem in Israel, and zionists are people who wants to have an independent Jewish state in the territory of Israel, which today is a reality. Nazi was a nickname for the national-socialist party that ruled alone in Germany from 1933 to 1945 and all the people that supported it or still supports it. We are mostly not going to consider several variants for the same word in this thesis. This is just a demonstration for how many different ways a word can be written.

**Table 1: Variants of zion/zionist/zionism (2631 occurrences in total)**

<b>Word</b>	<b>GTF</b>
zionists	459
zionistische	38
zionisten	16
zionist	1164
zionisme	11
zionism	253
zion	155
sionisti	49
sionistes	37
sioniste	51
sionistas	53
sionista	162
sionismo	111
sionisme	16
sion	45
sinonimo	11

**Table 2: Variants of nazi (4880 occurrences in total)**

<b>Word</b>	<b>G</b>
naziwhowantstokillsixmil	18
nazivaju	16
naziv	10
nazisti	53
nazista	89
nazismo	42

nazisme	29
nazism	80
nazis	11
nazioni	13
nazione	19
nazionalsocialisti	15
nazionalsocialista	37
nazionalsocialismo	10
nazionalit�	30
nazionalisti	29
nazionalista	30
nazionalismo	34
nazionali	23
nazionale	17
nazional	29
nazifascismo	10
nazies	38
nazie	48
nazi	24

**Table 3: Top 30 words from Stormfront, sorted in descending order by GTF or NGTF**

<b>Word</b>	<b>GTF</b>	<b>NGTF</b>	<b>NGTF-</b>
white	25769	1.000	1.099
quote	23285	0.904	0.993
people	20364	0.790	1.272
posted	17659	0.685	0.753
originally	17065	0.662	0.728

que	13863	0.538	1.753
che	12559	0.487	1.569
don	9815	0.381	0.792
jews	9044	0.351	0.872
black	8564	0.332	0.730
time	8111	0.315	0.725
world	7538	0.293	0.727
youtube	7509	0.291	0.842
les	7482	0.290	1.032
jewish	7137	0.277	0.750
race	7016	0.272	0.677
whites	6980	0.271	0.695
del	6798	0.264	0.897
una	6426	0.249	0.831
war	6152	0.239	0.715
obama	5652	0.219	0.594
anti	5560	0.216	0.569
country	5325	0.207	0.573
los	5086	0.197	0.760
news	5033	0.195	0.515
government	5020	0.195	0.574
con	5000	0.194	0.666
israel	4949	0.192	0.575
police	4949	0.192	0.555

**Table 4: Top 30 words from Stormfront, sorted in descending order by NGTF-IDF**

Word	GT	IDF	NGT	NGTF-IDF
------	----	-----	-----	----------

que	1386	3.258	0.538	1.753
che	1255	3.219	0.487	1.569
people	2036	1.609	0.790	1.272
white	2576	1.099	1.000	1.099
les	7482	3.555	0.290	1.032
quote	2328	1.099	0.904	0.993
del	6798	3.401	0.264	0.897
jews	9044	2.485	0.351	0.872
youtube	7509	2.89	0.291	0.842
una	6426	3.332	0.249	0.831
don	9815	2.079	0.381	0.792
los	5086	3.850	0.197	0.760
posted	1765	1.099	0.685	0.753
jewish	7137	2.708	0.277	0.750
black	8564	2.197	0.332	0.730
originall	1706	1.099	0.662	0.728
world	7538	2.485	0.293	0.727
time	8111	2.303	0.315	0.725
des	4914	3.784	0.191	0.722
war	6152	2.996	0.239	0.715
whites	6980	2.565	0.271	0.695
race	7016	2.485	0.272	0.677
con	5000	3.434	0.194	0.666
obama	5652	2.708	0.219	0.594
israel	4949	2.996	0.192	0.575
governm	5020	2.944	0.194	0.574
country	5325	2.773	0.207	0.573
anti	5560	2.639	0.216	0.569

della	3583	4.094	0.139	0.569
police	4949	2.890	0.192	0.555

## 4.4 Word Frequency Analysis of Swiss English Forum

Swiss English Forum [13] is, as the name proposes, a forum from Switzerland written in English. Here everything that has something to do with that small country in Central Europe is discussed. From the contents of the posts we see that many of the writers are immigrants in Switzerland. Some of them seem to have recently arrived.

The word “quote” in Swiss English Forum is the most frequent non-stop word found in this forum, and it is used 50988 times. “quote” is very often used since it appears in every message where another message is quoted, so the popularity of that word is absolutely no surprise. If you sort words by their global term frequency (GTF) in descending order, we get Table 5.

**Table 5: 20 most frequent words in Swiss English Forum**

Word	GTF
quote	50988
switzerland	37433
swiss	37075
people	29871
zurich	25886
time	25123
don	25105
german	16721
www	15869
english	15721
thanks	15573
basel	14443
http	14364
day	13900

help	13246
geneva	12159
buy	12121
looking	11953
car	11824
live	11238

Among these words, we have “swiss” and “switzerland” of course, since this is a Swiss forum, and we also find the name of three Swiss cities: Zürich, Basel and Geneva. “http” and “www” do not tell more than that some web addresses are mentioned in the forum. The word “don” is really “don’t”, another way of writing “do not”. The word is cut at the apostrophe because an apostrophe is a punctuation character according to the regular expression “\p{Punct}” [27], which our Java code replaces with a space before reading words from the forum. The expression “don’t” is really a stop word which has not been successfully removed. Else, the other words in Table 5 and Table 6 seem to be from daily life activities.

**Table 6: 20 words with highest NGTF-IDF values in Swiss English Forum**

Word	GTF	IDF	NGTF	NGTF-IDF
swiss	37075	1.386	0.727	1.010
people	29871	1.609	0.586	0.943
time	25123	1.792	0.493	0.883
don	25105	1.792	0.492	0.882
switzerland	37433	1.099	0.734	0.807
german	16721	2.303	0.328	0.755
www	15869	2.398	0.311	0.746
english	15721	2.303	0.308	0.710
zurich	25886	1.386	0.508	0.704
http	14364	2.485	0.282	0.700
quote	50988	0.693	1	0.693
thanks	15573	2.197	0.305	0.671
day	13900	2.303	0.273	0.628
basel	14443	2.197	0.283	0.622

car	11824	2.639	0.232	0.612
chf	10230	2.944	0.201	0.591
geneva	12159	2.398	0.238	0.572
help	13246	2.197	0.260	0.571
live	11238	2.565	0.220	0.565
looking	11953	2.398	0.234	0.562

The order and some of the words in Table 6 is a little different from Table 5, because the words are sorted by NGTF-IDF instead of NGTF. Nearly by just taking a glimpse at these words, we can see that they are used by the authors to express day-to-day activities, and nothing (at least very little) of extremist ideas, comments and vocabulary. There are some more daily-life-related words among the top words for NGTF-IDF than in the (N)GTF ranking. Hence, comparing the frequency of these words in an ordinary forum like this with an extremist forum, one can detect the difference between the normal text and extremists' text. Ordinary forums and extremist forums may use many of the same words, but to different extents. Stop words like “don(‘t)” are moved down in the list compared to the GTF list, so we do not see it among the words highest ranked by GTF-IDF. Unfortunately, this does not always work so well, especially not with non-English stop-words in a mainly English forum, like Stormfront.

## 4.5 Word Frequency Analysis of the CS:GO Forum in Steam

In the forum for the computer game “Counter-Strike: Global Offensive” (CS: GO) amongst the forums associated with the online computer game shop Steam, the most popular non-stop word is (not surprisingly) “game”, which is used 3816 times in the downloaded webpages. The word “comment” is used 7465, and the reason is that we assigned that word as title to untitled messages, so it is not really a good count of that word. In this forum, a thread has a title, not really the messages, so we assigned that thread title to the opening message of the thread and gave each of the other messages the title “comment”. Popular words in the forum are “play”, “players”, “people”, “competitive”, “cheaters”, “ban”, “matchmaking”, and “team”. “play” is applied 1737 times and “team” 885 times. The other recently mentioned words have values in between. All these words are somehow related to competitive online computer gaming.

In this forum we expected to find discussions about how to play the game “CS: GO” and how to solve technical issues in the game. Of course, there was some of that, but to our surprise, many forum authors complained about cheaters (hackers) in the game and complained about why these cheaters were not banned. In addition, somebody complained about themselves being banned without reason.

We will sometimes refer to this forum CS:GO game forum as “Steam”, although it is really just one of many forums related to the game distribution platform Steam.



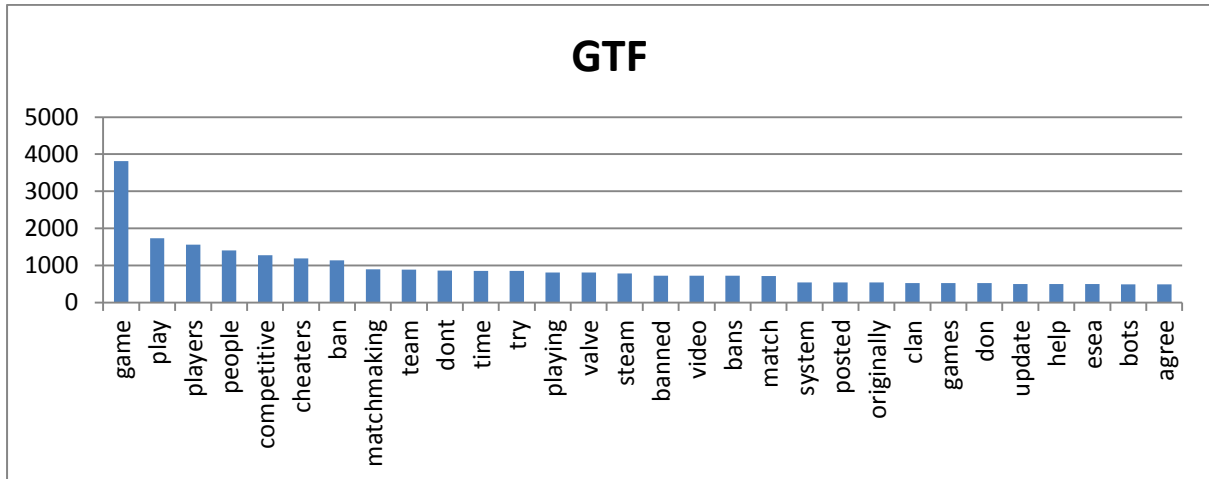


Figure 10: 30 most frequent words in Steam

## 4.6 NGTF Comparison of Swiss English Forum and Islamic Web-Community

We call this section and following “NGTF Comparison”, because we use the NGTF values for comparing two forums with each other. In the tables one can see NGTF ratios as comparison values. An NGTF ratio is the NGTF value in one forum divided by the NGTF value in the other forum. A word has then two different NGTF ratios, where one of them is the inverse of the other. To avoid confusion, we refer to one of the forums as f1 (forum 1) and the other as f2 (forum 2). An f1/f2 NGTF ratio of 100 for a given word tells for example that that word would have been used 100 times more in forum 1 if forum 1 and forum 2 both had the same amount of data. We also use the NGTF ratio that is higher than 1. The NGTF ratio 1 means of course that the NGTF value for the word is the same in both forums. In this section and the following we use NGTF ratios for comparison unless something else is specified.

When we have sorted the common words found in the forums Swiss English Forum (f1) and Islamic Web-Community (f2) by their f1/f2 NGTF ratios, the word “mountain” has the value 1, which means that it appears in the both forums the same amount of times relative to forum size (more precise: relative to frequency of most frequent word in each forum).

Table 7: Top 30 words in Swiss English Forum (f1) and Islamic Web-Community (f2) sorted by f1/f2 NGTF

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
quote	1888.444	36.778	9.194
don	836.833	16.298	10.532
english	827.421	16.117	11.258
pay	780.583	15.208	11.901
kids	717.5	13.979	11.807

thanks	677.087	13.188	9.373
car	656.889	12.791	10.026
probably	638.5	12.436	10.002
move	587.333	11.443	8.955
school	568.6	11.074	9.579
hand	506.909	9.87	8.149
seen	450.8	8.78	7.835
week	440.167	8.576	7.752
near	439.583	8.564	7.416
job	432.867	8.428	7.901
speak	387.75	7.553	7.551
live	362.516	7.06	5.858
experience	349.714	6.81	6.415
money	347.4	6.766	6.684
card	328.333	6.397	5.513
doesn	326.077	6.351	6.169
meet	322.294	6.277	5.611
city	321.813	6.268	6.267
else	319.444	6.22	6.156
look	317.276	6.18	5.789
friends	313.667	6.108	6.183
feel	312.706	6.09	5.977
didn	294.75	5.741	5.871
prices	294.643	5.738	4.964
start	289.412	5.636	5.027

The word “quote” is used 36.78 times more (in a normalized way) in Swiss English Forum (f1) than in Islamic Web-Community (f2). It seems therefore to be much more quoting in f1 than in f2. One particular detail is that the word “english” has an f1/f2 NGTF ratio of 11.258. This is natural, since “english” is in the name of the forum, and it is written in English. Many of the sentences with that word are about speaking English and different situations where one can use that language in Switzerland. English is namely not a mother tongue in Switzerland, where the official languages are German, French, Italian, and Romansh. Example from line 365622 in forum\_entries.sql: “looking to meet fellow **english** speaking ladies in lausanne”.

The main theme of mutual words, which are most frequent in f1, is daily life, more precisely objects and actions from daily life. Such words are “pay”, “thanks”, “car”, “school”, “speak”, “live”, “experience”, “money”, etc.

In the most frequent words in both forums by the f2/f1 NGTF ratios in descending order, one can see words related to religion, like “islamic” (73.1), “islam” (55.4), “muslim” (52.4). “lord” (22.7) is likely a reference to God or Allah. Surprisingly, some of the other top f2/f1 NGTF ratio words seem to be related to computer or information technology. “download” is actually the word that has the highest f2/f1 NGTF ratio (89.5). We also encounter “files” (46), “software” (35.6), “sms” (31), “file” (23.5), “web” (22), “php” (22), “applications” (21.6), “disc” (18.1), and “data” (17).

**Table 8: Top 30 words in Swiss English Forum (f1) and Islamic Web-Community (f2) sorted by f2/f1 NGTF**

<b>Word</b>	<b>GTF ratio (f2/f1)</b>	<b>NGTF ratio (f2/f1)</b>	<b>NGTF-IDF ratio (f2/f1)</b>
download	1.743	89.531	27.188
respective	1.708	87.727	8.006
islamic	1.423	73.087	7.074
greetings	1.272	65.307	12.511
islam	1.078	55.358	14.359
muslim	1.02	52.365	14.885
solar	1	51.335	25.613
released	0.921	47.298	4.567
files	0.897	46.039	20.294
theme	0.88	45.176	4.367
reserved	0.701	35.976	3.693
software	0.693	35.605	0
evolution	0.663	34.039	4.655
sms	0.604	31.007	16.705
holy	0.579	29.718	11.396
sake	0.546	28.055	12.561
obedience	0.529	27.174	13.362
design	0.51	26.201	3.063
portal	0.506	25.956	8.433
bah	0.5	25.673	12.715
whoever	0.489	25.087	12.253
portugal	0.48	24.641	5.467
encouraged	0.468	24.007	11.996
peace	0.464	23.837	8.731
file	0.458	23.524	9.564
lord	0.443	22.739	9.392
web	0.429	22.007	2.685
php	0.428	21.968	0
applications	0.421	21.616	10.354
pleased	0.417	21.437	9.157

The words that are unique to Swiss English Forum include many words that can be related to day-to-day life, but also to tourism. The words “swiss”, “switzerland”, “german”, “zurich”, “geneva”, “french”, “lausanne” are all in the Swiss forum, but not in the Islamist forum it is compared to. Of course “swiss” and “switzerland” are the most applied of the unique words because the forum is Swiss. German and French are the two most spoken native languages in Switzerland. Zürich, Geneva and Lausanne are cities in Switzerland. Throughout the list in the table above, one finds a lot of Swiss place names. We find words from more or less

serious topics in the text, from “police” (used 4051 times) and “permit” (3237 times) to “fun” (4284 times) and “luck” (3623 times).

When looking for an example for “police”, we found a funny message about the Swiss stereotype in line 112711. The title happens to include the word “fun”:

*“(EnglishForumCh', '10\_t1', 'What do Swiss people do for **fun**?', 'Quote: MotorsportsNorton Simple question, if you do not live in the bigger cities where there is quite a lot of life, what do Swiss people do for relaxation apart from sitting in their cellars counting their money. Apart from going to work and going for the occasional walk most villages seem like ghost towns in the evenings and weekends. curtain twitch, make notes and then call the **police** on neighbours who make a noise at 22.01 or shower at 0559 ?', 'rob1', 'Today, 12:11')”*

It looks like there are no words in “Islamic Web-Community” that do not exist in Swiss English Forum. That generated file of unique words is namely empty.

## 4.7 NGTF Comparison of Islamic Web-Community with Steam

When we have sorted the words found both in Islamic Web-Community (f1) and Steam (f2) by the normalized f1/f2 GTF ratios, we can see that the word “nuke” has the highest value (NGTF ratio 48.1, GTF ratio 12.5). This is because that word appears much more frequently in the Islamic Web-Community forum (f1) than in the Steam forum (f2). In other words, “nuke” is used 12.5 times more in Islamic Web-Community than in Steam, as shown in Table 9. If the two forums were of the same size, “nuke” would have been used 48.1 times more in f1 than in f2. In the table below there are words from Islam and computer technology which are used in both forums, but mostly in f1.

**Table 9: Top 30 words in Islamic Web-Community (f1) and Steam (f2) sorted by f1/f2 NGTF**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
nuke	12.536	48.173	5.770
style	7.245	27.841	3.693
live	3.100	11.914	5.400
sound	2.700	10.378	4.480
read	2.630	10.105	3.994
sin	2.600	9.990	4.7738
saying	2.389	9.180	4.747

message	2.385	9.164	3.434
client	2.000	7.685	4.705
remain	1.900	7.300	3.529
downloa	1.843	7.082	3.350
day	1.826	7.018	3.153
light	1.700	6.533	3.406
released	1.648	6.332	0.905
forum	1.558	5.986	2.776
click	1.543	5.929	2.676
power	1.410	5.418	3.480
rule	1.400	5.381	3.672
question	1.333	5.123	2.678
files	1.190	4.571	2.913
warning	1.114286	4.281552	2.476903
website	1.1	4.227038	2.351773
called	1.035714	3.979543	1.988028
page	1	3.842873	1.740171
friend	0.913043	3.509061	2.162211
perform	0.857143	3.293753	1.917249
http	0.820	3.143	1.117
main	0.810	3.110	1.760
ppl	0.761	2.923	2.380
giving	0.735	2.826	1.756

Generally, one would think of the word “nuke” as a nickname for nuclear weapons, but when we examine the text in the Islamic Web-Community, we see from the phrase “PHP-Nuke is a Free Software released under the GNU/GPL license.” that weapons are not the topic at all. Although that word is very common in this forum, it is not interesting for our analysis because PHP-Nuke is a web-based automated news publishing system, which seems to be used for the forum.

An example of “live” in line 217 is just an ordinary question for a person’s place of living: “(...) *i have a very close relative in saudi too, and another one is planning to move there soon. sister which city do you **live** in??*”.

This example is from an everyday chat where two women get to know each other online.

That the word “sin” is to be found more in Islamic Web-Community than in Steam, is not a surprise for us since the content of Islamic Web-community forum is related to the religion Islam where the devotees learn about sins, which are anything that goes against the commands of God (Allah). When we observe the text content of this forum, we see the phrase “*He who performs Hajj and avoids **sin** and vice in it, will return as **sin** free as the day his mother bore him.*?” (line 137 in forum\_entris.sql). Hajj is a pilgrimage to Mecca, which is very important to Muslims, and all Muslims are encouraged to travel on at least one such journey in his/her lifetime. Vice is another word for sin.

In line 546, we can see an occurrence of the word “sound”:

*“I will allow free discussion before intervening. However I disagree with John as there's a perspetion that Islam is being tergeted, and the most flagrant violations of all is Guantanamo prison camp, that nobody with **sound** mind can defend it.”*

Here, “sound” means “healthy”, not something that one can hear. In this case an imagined mind is healthy.

Here is an example of “sound” as something to be heard, from line 252. Here a sound recording to be played back on a computer is mentioned:

*“For those members and visitors that have **sound** in their PC. they can hear Eid Takbir. The takbir will be removed 3 days after Eid The takbir from Makka can be heard on first page and also in forums.”*

“Makka” is another way of writing “Mecca”, a holy city in Islam, and “Eid Takbir” is a prayer with several repetitions of the phrase “Allahu akbar” (“God is the greatest”) [28].

In line 94: “*Please try to attend and pass this **message** to your friends and relatives, so that more people can benefit from it. May Allah guide us all and help us to unite.*” is an example for the word “message”.

The word “power” is exemplified in line 459: “*Solar **power** remains expensive compared with making electricity from coal or natural gas, but it is bounding ahead, driven by quotas set by the states.*” This example of the application of “power” shows that electricity is discussed more in Islamic Web-Community than in the Steam forum. “power” refers namely to electric energy in the recent quote, while in the following example from line 551, the same word has the meaning of someone being able to have dominion over people and territory:

*“In Iraq, where there's an occupying **power** supported by the master of all imperisalists, the british, who fully know how to divide and **rule** and in Pakistan from time to time, that's the biggest laboratory of sysops of enemies of Islam.”*

In line 564:

*“We hope that with this new **warning** system we'll be more accountable on banning procedures, and every Member will know why he/she has been banned.”*

Warnings are sent from members of the forum to some forum administrator or moderator when someone writes anything inappropriate. Then that writer is banned from the Islamic Web-Community forum and gets an explanation of why.

**Table 10: Top 30 words in Islamic Web-Community (f1) and Steam (f2) sorted by f2/f1 NGTF**

<b>Word</b>	<b>GTF</b>	<b>NGTF</b>	<b>NGTF-</b>
game	112.2	29.2	11.8
play	62	16.1	9.3
update	35.7	9.3	6.8
banned	33	8.6	6.1
games	32.7	8.5	7.5
servers	28.5	7.4	6.3
player	24.1	6.3	4.8
posted	23.4	6.1	5.4
hackers	21.3	5.6	4.2
system	20.7	5.4	3.9
start	18.2	4.7	4.1
don	17.4	4.5	4.9
running	16.8	4.4	5
thanks	16.3	4.2	4.4
getting	16.1	4.2	4.3
happened	15.2	3.9	3.8
happen	15	3.9	3.9
support	15	3.9	4.3
money	14.1	3.7	4.2
check	13.6	3.5	3.8
ones	13.1	3.4	3.6
try	12.9	3.4	3.8
completely	12.7	3.3	3.3
server	12.3	3.2	3.5

tried	12	3.1	3.5
level	11.7	3	3.1
matter	11.7	3	3.4
fire	11.4	3	3.5
lost	10.6	2.8	2.8
lot	10.5	2.7	3.4

We can interpret from the results that the word “game” (29.2, 112.2) is found more in the Steam forum than in Islamic Web-Community, which should not be a surprise since Steam is a game forum. For example, in line number 75009 in forum\_entries.sql:

*“Haven”t ever found a cheater in CSGO. And I”ve played this **game** for almost 150 hours.”*

In line 75332 we find the word “play” as an action one does with a game:

*“Just played a game now that I couldn”t leave (worst thing ever) and was forced to **play** against 4 cheaters in the enemy team and one with speed+aimbot in my team.”*

In this case it does not look like the author had fun while playing because of cheaters ruining the gaming experience.

In line 74764 there is an example of someone getting banned from the game by accident because he/she is detected as a cheater by small technical error (bug) in the anti-cheat system of the game. This can happen when the player plays “too good”:

*“Well, if not, you will play with the bot till the end anyways and a person might get **banned** cause of a bug from Steam or the game.”*

In the following examples with “player”, the writers tell how the game should work in some special cases of multiplayer interaction.

In line 74145:

*“If a **player** teamkills than the one who died should be given the choice to forgive the kill or not, and if it”s forgiven then it doesn”t go against the one who teamkilled.”*

In line 66248:

*“Teams should be able to vote for a surrender at any time, not just if a **player** drops.”*

Hackers are players who use some self-written program code for getting illegal advantages in the game. Such advantages may be immortality and automatic aiming with the gun at enemies. Other players just get annoyed of them, like in line 73961:

*“Again as i have said i find **hackers** to be annoying and gay why would i join the crowd?”*

Following two examples show that fair players want the cheaters/hackers to be banned for good, so that others can continue playing fair. The only problem is that the automatic system in “CS:GO” sometimes detect some fair players as cheaters by accident when their activity seems somehow suspicious. Example in line 73665:

*“I would have banned **hackers** and grievers forever and go bankrupt.”*

and in line 73290:

*“To discovered and blatant **hackers** - eternal ban, give a chance to everyone else to play for real and improve (until they start blatant hacking,then ban them aswell) and if I may add, my own: 10.”*



From the above examples, we understand that when players are caught for cheating/hacking, they are supposed to be immediately banned from the game by means of an automatic system or by reporting from the other players. There are still problems that some cheaters/hackers are not discovered, and there are also some examples of fair players that get banned by accident.

Next, we consider the words that are in the Islamic Web-Community forum, but **not** in the Steam forum, and we look at the 30 words with highest GTF. It is important to note that some of these words may be used maximum 9 times or less in the Steam forum.

**Table 11: Words that are in Islamic Web-Community, but not in Steam**

<b>Word</b>	<b>GTF</b>
allah	993
hajj	502
evolution	470
web	276
prophet	267
design	248
property	238
phpbb2	234
peace	227
code	181
file	164
salaama	164
akbar	160
islamic	158
allahu	154
blessings	145
rest	145
phone	138
ten	137
greetings	131
gpl	129

license	129
comments	127
mobile	126
engine	123
respective	123
rights	123
dvd	122
owner	118

From the table above, we can see that the most frequently used unique word in Islamic Web-Community is “allah” which seems to be very often seen in Islamic extremist forums in general. Allah is the Arabic word for God and is used 993 times. The second-most popular word, Hajj, refers to a pilgrimage to Mecca, which all Muslims should do if they are able to. The word “prophet” has been used 267 times, which is very little compared to the use of “allah”, but still a considerable number of times. In Islam, a prophet is a person who has been assigned a mission by Allah (God) himself to tell how the true religion really is. “salaama” is an alternative way of writing “salama”, which is an Arabic word for peace. “akbar ” which is used 160 times, is a word very often found in extremist forums, but likely also other places. It is used in the phrase “Allahu Akbar” which means “God is the greatest”. This phrase is called the Takbir and is used (also) by ordinary Muslims in different kinds of situations: in prayers, when they are happy, to express approval, to praise a speaker, and to calm oneself down during extreme stress [28].

Line 534:

*“Magnify ones intention and get even MORE reward doing good deeds! Allah's Messenger (peace be upon him) said: "Deeds (their correctness and rewards) depend upon intentions, and every person gets but what he has intended.”*

This example shows that one can find the word “peace”. This is “peace” almost like in “rest in peace”, and is here meant as an expression for honouring the Messenger. “Allah’s Messenger” is here most likely Muhammad, the last prophet in Islam, and therefore also the last messenger of Allah/God.

The word “blessings” appears 145 times and for example in line 257:

*“may Allah bring great **blessings** and forgiveness for muslim ummah on the day of Arafah tomorrow Ameen.”*

The word “dvd” is found 122 times in this forum, like in line 170:

*“Read/watch/learn about Hajj When it comes to prayer and fasting, we tend to know much more about these pillars of the faith than Hajj. Most of us tend to put off finding more until we actually go. But who knows how long you have before going, so why wait? If books aren't your thing, find a video, an audio CD, a CD-rom or a **DVD** that can give you the basics.”*

We observe that the author of this message really wants to make the followers of Islam (and maybe also others) understand the importance of the Hajj pilgrimage and never miss it.

Then we consider the words that are in the Steam forum, but not in the Islamic Web-Community forum, and we look at the 30 most frequent words sorted in descending order by GTF in the table below.

**Table 12: Words that are in Steam, but not in Islamic Web-Community**

<b>Word</b>	<b>GTF</b>
players	1560
cheaters	1187
ban	1139
matchmaking	893
team	885
dont	857
playing	809
valve	806
steam	779
bans	720
match	710
originally	538
clan	528
esea	497
bots	492
played	453
maps	450
kick	448
bomb	422
fps	422
tell	412
runs	410
reconnect	403

settings	395
cheater	383
slow	382
fix	382
csgo	377
kill	376
restart	372

The word “matchmaking” is found 893 times in the Steam forum. In multiplayer video games, matchmaking is the process of connecting players together for online game sessions [29]. For instance in line 75374: *“Implement a mandatory weapons/mission course tutorial and a minimum number of deathmatch/arms race etc fully played before being able to que for competitive **matchmaking**.”* or in line 75353:

*“They don't like to make their hack obvious, they aim at the walls like regular players try to act like non hack players etc, And 10-10 It really is a matter of letting them win a little before he turns on the cheat's I have seen it so many times now in **matchmaking**, when you lead etc those scumbags becomes demigods and kill everyone.”*

The example hinted that all players should go through a tutorial and get some basic playing experience in other game modes before entering the matchmaking. The author also expresses his annoyance about cheaters in the matchmaking mode.

We can see the word “valve” has been used 806 times and for instance in line 75333:

*“I've encountered some users with their VAC Status bans on record. So let's report all these "players" and leave it to **valve** to do their work in order to make this game great.”*

Steam is developed by Valve Corporation, an American video game development and digital distribution company. Therefore, it is very natural to mention it in the forum of a game they have developed. Here, Valve is encouraged by a player to remove cheaters from the game.

The word “steam” is used 779 times and is for instance mentioned in line 75232:

*“Perhaps you could even pinpoint which setting is causing it and tell **steam** support.”*

In line 75058 one finds an example of how the word “clan” is used. There it seems to be a team of co-operating players:

*“CS:GO **Clan** League’, ‘Hello, I am a frequent player and user of the CS:GO community, and I have a small **clan** of 6 people that I belong too.”*

In computer gaming, a clan or a guild is an organized group of players that regularly play together in one specific or many different multiplayer games. Some people might say that a "clan" or "guild" becomes like a family, and that all opinions and decisions represent each single member in the "clan" or "guild”.

Also in line 74690 there is a good example:

*“Rather than just a pub servers, but also playing a **clan** war (lobby) Thank you, I hope you understand, and a nice evening!”*

Instead of playing on open, public CS:GO servers, some players prefer to play on servers where only members of their team/clan and other befriended teams/clans have access, such that they can compete against each other without other random players interfering.

The word “esea” is written 497 times, for instance like in line 74654:

*“Again i hate the cheaters all together and have switched on to playing more to ESEA. All the kills you have shown except one part which does seem questionable in esea is literally a good morning”*

ESEA (E-Sports Entertainment) is according to itself “the largest competitive video gaming community in North America, where thousands of players end up after getting sick of playing against cheaters and instead want to play against the best and get better” [30].

The word “fps” is used 422 times. It could have been first-person shooter (FPS), since that is the genre of “CS: GO”. This genre is centred on gun and projectile weapon-based combat through a first-person perspective, which means that the player experiences the action through the eyes of the protagonist [[31]]. However, in many of the examples we found, “fps” really means “frames per second” [32], how many images that are shown on the screen in each second, like used in line 74485:

*“how much fps do u have in-game? (min-max) taking your posts in mind i guess ist the hardware problem. check yout temps before and after game, also dont forget to clean your pc out of dust, clean coolers and add new thermal paste that should do the trick.’, ’, ’Dec 22, 2012 @ 6:20am ’”*

Line 74313:

*“I tried all of the above, except cleaning my pc, because it is a 1 month old laptop. It runs at 20FPS at 2880x1800 and around 30FPS at 1900x1200.”*

The word “csgo” is used 377 times. Counter-Strike: Global Offensive (CS:GO) is an online first-person shooting game developed by Valve Corporation and Hidden Path Entertainment [33] and the game which is discussed in the part of the Steam forum we have downloaded. We find examples of “csgo” usage in line 74291: *“Haven”t ever found a cheater in CSGO. And I”ve played this game for almost 150 hours.”* and in line 74301: *“There is nothing more left then to stop playing csgo and cry, I”ts not that good of a game anyway”.*

## 4.8 NGTF Comparison of Swiss English Forum and Ummah

**Table 13: 30 words with highest f1/f2 NGTF ratios for words that are both in Swiss English Forum (f1) and Ummah (f2)**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
switzerland	1970.2	893.8	124.4

swiss	662.1	300.4	52.2
geneva	405.3	183.9	54.6
cheers	289.4	131.3	62.4
shipping	176.8462	80.2	40.6
amazon	170.8	77.5	44.1
dutch	167.9	76.2	41.2
trains	166.7	75.6	40.2
christmas	161.9	73.5	32.8
german	154.8	70.2	25.5
beer	139.5	63.3	36.8
quote	132.1	59.9	8.7
vegan	121.7	55.2	30.2
restaurants	120.4	54.6	31.9
bike	109.4	49.6	27.5
downtown	102.3	46.4	26.9
cheaper	97.0	44.0	24.8
hypothetical	94.8	43.0	24.2
der	93.5	42.4	29.7
apartment	89.8	40.7	22.9
beef	88.6	40.2	25.0
flats	81.4	36.9	21.1
stores	78.9	35.8	20.6
shopping	76.0	34.5	16.0
salary	71.7	32.5	17.5
band	71.5	32.5	19.2
foreigners	70.8	32.1	19.3
tom	70.7	32.1	18.3
boots	68.8	31.2	20.0

lake	68.1	30.9	17.8
------	------	------	------

We now compare Swiss English Forum (f1) and Ummah (f2) with f1/f2 NGTF ratios. Look in table above for the needed ratios.

The word “switzerland” has an 893.8 times higher NGTF value in Swiss English Forum (f1) than in Ummah (f2), which is not a big surprise, since Swiss English Forum is a Swiss forum written in English. For instance, in line 365446, we find

*“Hi there, I am complitelt new here, in **Switzerland** since one month. I'm looking forward to meet some peoples, see and join the parties, thats the reason of my post- could anyone be so king, and maybe give me some tips about that? Which clubs are the best for international people, or just with kindly members ? (...).”*

From this and other similar comments we can understand that the forum is about day-to-day life events in Switzerland such as meeting new people and going to parties. At least some of the writers in the Swiss forum seem to be foreigners that have just come to Switzerland, want to know new people, and want to know how to behave and fit into the Swiss society.

The word “swiss” is used 300 times more in f1 than in f2 according to the NGTF ratio.

Example from f1 as shown in line 365413:

*“As British atheists attending a **Swiss** catholic wedding, we gave something very British.”*

“Swiss” is simply a word for telling that someone or something is from Switzerland. The excerpt mentioned a Catholic wedding where Swiss traditions are followed. In line 365407, the author has a nickname hinting that her husband is a Swiss, which can hint to a fact that she is not a Swiss herself:

*“Hello, Does anyone know what is traditionally given as a gift for a baby boy for his Christening? Many thanks!', 'Wife of a **swiss**'”.*

The word “geneva” has a 183.9 times higher NGTF value in f1 than in f2. Geneva is a city in Western Switzerland where indigenous people talk French. (The French name of the city is Genève.) Example from line 365293:

*“where can you buy a laptop that's not too expensive?!, 'Where are you ? In **geneva** ? If yes, it exists in **Geneva** one shop which recuperates old PC, reparate them and sell them very cheap : CASH computer, 13 rue Louis Favre in Geneva', 'David Demarest', '09.03.2013, 17:44')”*

The forum entry answers a question from another person about where to purchase a computer. The answer tells that in Geneva there is a shop where one can buy used and repaired computers for good prices. It is likely to believe that the message writer is not anglophone, rather francophone, since he/she uses the word “reparate” instead of “repair”. The person who asked the question may be a foreigner since he/she does not know where to buy a computer in Switzerland. This example also shows again that the forum is about daily life for foreigners that have just moved to the small country in Central Europe.

The word “cheers” has been used 131.3 times more in f1 than f2, and a good example of how this word is used, can be found in line 363076:

*“I rarely ever go in them unless I go clothes shopping, so didn't think to check. I bet they cost an arm and a leg in there though if you buy them individually... probably 50CHF a pan or more for a Tefal product. I'll check it out though, **cheers**!', 'Richdog', '17.02.2013, 17:33')”.*

Here, “cheers” seems to be a positively charged goodbye greeting almost like “See you soon!”. Also in line 112789 we find a good example of that word:

*“Are you good at putting CV's together?’, ‘Looking for someone to help me build a CV this afternoon! Anyone interested please message me. Cheers’, ‘mzigah’, ‘06.03.2013, 11:40’”.*

Here “cheers” seems to be used to thank someone in advance.

“shipping” has an 80 times higher NGTF in f1 than in f2. An illustrating instance of that word is found in line 363016:

*“Palmer's Cocoa Butter?’, ‘You can get Bio Oil and Cocoa Butter from www.feelunique.com (free shipping to Switzerland).’, ‘SteffieM’, ‘20.02.2013, 19:37’”.*

Here shipping means transporting items ordered from an online shop via sea, land or air. Since Switzerland do not border to any coast, the first option must be non-existing in that country.

The word “amazon” can be found 77 times more in Swiss English Forum than in Ummah. For instance in line 360997:

*“I would like a place with on online registry so guests can just shop online. I have tried amazon.com and amazon.de. The problem with amazon.com is that many items are not available to be shipped out of the US and if they can be shipped, the shipping fee is often more than the actual item.”*

The American website Amazon.com is one of the world’s largest online shops. It is most famous for selling books and ebooks, but also movies, computer games, toys, and clothes are available there. Amazon.de is the German branch of Amazon and works as an independent online shop in Germany, but people in Switzerland can also buy books and other stuff from there.

The word “boots” is found 31 times more in f1 and is exemplified in line 360272:

*“BTW most villages organise regular sales/swap for kids sports equipment which has been outgrown. Here we have one late autumn for ice-skates, skis, snowboards and boots, etc. then another in Spring for bikes, rollerskates and blades, etc.”*

Here “boots” likely means footwear for use in alpine skiing and snowboarding, designed in a way to be attachable to skis and snowboard, respectively.

Next, we have sorted the words found in the same both forums by the f2/f1 NGTF ratio. We remind that Ummah is f2, and Swiss English Forum is f1.

**Table 14: 30 words that are both in Ummah (f2) and Swiss English Forum (f1), with highest f2/f1 ratios**

Word	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
muslims	222.8	491.4	149.8
islam	73.8	162.6	45.5
muslim	68.9	151.9	40.9
companions	51.9	114.4	54.0
messenger	49.7	109.7	45.7



haram	47.1	103.8	45.6
sisters	42.5	93.7	34.4
Saudi	37.2	81.9	1
palestinian	34.73333	76.6	47.5
eid	33.2	73.0	32.3
masjid	31.7	69.9	35.7
Islamic	27.4	60.3	22.2
lord	27.2	60.0	26.7
prayers	26.0	57.2	26.3
jewish	23.2	51.1	28.1
originally	22.8	50.2	9.7
mercy	22.4	49.3	24.1
bush	21.3	47.0	29.6
deed	19.2	42.3	23.4
believer	19.1	42.2	22.3
apostle	18.2	40.0	24.8
brothers	17.7	38.9	17.6
umm	17.7	38.9	20.1
glory	17.5	38.5	22.5
commentary	17.1	37.6	21.5
palestine	17.0	37.4	23.4
sincere	16.1	35.6	19.7
quiz	16.1	35.5	16.7
christians	15.9	35.1	20.0
arabia	15.9	34.9	19.8

According to the results above, we can see that the word “muslims” has the highest f2/f1 NGTF ratio 491.4. This means in other words that “muslims” appears 491.4 times more in the Ummah forum than in Swiss English Forum, which is very obvious to happen since Ummah

is an Islamist forum and the other is a daily life forum. The word “islam” is used 162.6 times more in the same Islamist forum. We can for instance see that the words “muslims” and “islam” are both used in line 107831:

*“I have learnt a great deal from this discussion, i am sure it being in "Learn about **Islam**" will give **Muslims** as well as non-Muslims a thorough understanding of **Islam** and **Muslims**.’, 'Umm Layth', '30-06-06, 04:21 PM'”.*

The word “companions” has been written 114.4 times more in Ummah (f2) and for instance in line 107812 in our SQL file:

*“The Prophet was displeased and even removed the flag from him. You have to remember, the **companions** make mistakes too, as they are human.”*

Here the companions are the followers of Allaah who have being sent with the Prophet to help him in his mission to spread Islam by Allaah.

The word “messenger” is found 109.7 times more in f2 and is exemplified in line 107522:

*“Let the people cease to boast about their kin and kin-folk.” Abu Dawud, Book 36, 5097  
“Those who have experienced the beauty that is Al-Eeman are those who are content with Allah as their Rabb, with Al-Islam as their Way of Life, and with Muhammad (salla Allahu ‘alayhi wa sallam) as their **Messenger** and Prophet ”.*

Here Muhammad is considered as the last prophet and messenger in Islam, which is very central to that religion.

The word “haram” is used 103.8 times more in f2 than in f1. “haram” is an Arabic word which means sanctuary or holy site in Islam, and it can also refer to a forbidden action [34]. In line 95068 it is mentioned that the Prophet made temporary marriages forbidden:

*“(‘Ummah’, '293205-Points-of-Benefit', 'Re: Points of Benefit', '-- - Some of things which ‘Umar radiyallahu ‘anh was the first to do:- - - (...))In some books it is also recorded that he was the first to make mut’ah (temporary marriages) **haram**. However, this is definitely incorrect. It was made **haram** by the Prophet sallallahu ‘alaihi wa sallam at the time of the battle of Khaibar, and the hadith stating this is narrated by ‘Ali radiyallahu ‘anh.- (Shaikh Ebrahim Moos)', 'the\_middle\_road', '29-02-12, 10:34 PM'”*

The word “masjid” is used 69.9 times more in f2 than in f1. “masjid” is simply the Arabic and original word for mosque, and it means place of worship. Masjid Quba is first mosque in the history of Islam, and it still exists [35].

The word “lord” has a 60 times higher NGTF value in f2. In line 107215, we see that it refers to God or Allaah. The following example seems to be about angels that ask people whether they believe in God or not:

*“The Prophet of Allah turned to Umar and said, "Every one of you will be asked by the angels "who is your **lord**?" except for you, umar! You will ask the angels "Who is **YOUR Lord**!?" If it is a real hadith, then in my opinion it's the most beautiful hadith I've ever heard.”*

The word “jewish” has an f2/f1 NGTF ratio of 51.1. Here is a surprising example from line 103783:

*“While I admit there are some **Jewish** and even some Christians out their that would destroy any Muslim presence in Jerusalem and rebuild the temples... I like to point out that even during the heaviest fighting the Israel government has chosen **NOT** to destroy the mosque*

*and ordered its soldiers NOT to do so even when they have and had and today have the overwhelming military power to do so”.*

We read here that the Israeli government has never attempted to destroy the Muslim shrine called Dome of the Rock (often imprecisely referred to as a mosque) [36] in Jerusalem in order to rebuild the Jewish temple destroyed by the Romans nearly 2000 years ago. The author says that there are some Christians and Jews who want to do that, but Israeli government prevents that from happening.

“mercy” has been written 49 times more in f2 than in f1 and we see a nice example for that word in line 103036:

*“The key qualities every parent should strive to attain are epitomised by Prophet Muhammad (peace be upon him), who has complete: Love & **Mercy**: As we are all aware, his **mercy** is unparalleled and indeed, his treatment of children, and not just his own progeny, is an example to us all. The Prophet (peace be upon him) would kiss and embrace children often and he would take an active interest in their lives.”*

The author of this message wants to express his/her feelings for the prophet Muhammad and praises him for being kind to children and showing love and sympathy for everyone.

“apostle” is used 39 times more in f2 than in f1. In Christianity, Jesus had 12 disciples that are also called “apostles”. After Jesus’s resurrection and ascension they spread his teachings to the world. In the following example from line 89169, “Allah’s Apostle” is mentioned. We see from that forum quote that this must be the prophet Muhammad:

*“When he [the Prophet] was about to pass sentence, the man who [actually] had assaulted her stood up and said: **Apostle** of Allah, I am the man who did it to her. He [the Prophet] said to her: Go away, for Allah has forgiven you.”*

“glory” is used 38 times more in f2, and we see a good example in line 87101:

*“The good righteous deeds that last are `Subhan Allah (**glory** be to Allah)”, `Al-Hamdu Lillah (praise be to Allah)”, `La ilaha illallah (there is none worthy of worship except Allah)”, and `Allahu Akbar (Allah is Most Great)”.*

Here Subhan Allah is an Arabic saying which means “glory be to Allah” [37].

“palestine” has a 37 times higher NGTF value in f2 than in f1. Line 86909:

*“I learned what Israel had done in Lebanon – and what it continues to do in **Palestine** – with the full backing of the United States. And I learned what America itself was doing to Muslims. I learned about the Gulf War, and the depleted uranium bombs that killed thousands and caused cancer rates to skyrocket across Iraq. I learned about the American-led sanctions that prevented food, medicine, and medical equipment from entering Iraq, and how –according to the United Nations – over half a million children perished as a result.(...)”*

From this excerpt we can see how the author is furious and disappointed at the USA and the American authorities. Here it is mentioned what damages Americans allegedly did on Iraqi civilians during the Gulf War in Iraq in 1991, here referred as Muslims, since most of the population in that Middle Eastern country are Muslims.

“christians” is used 35 times more in f2 than in f1, and good example from f2 is to be found in line 85013:

*“For Muslims, that should not be a problem. We already have the answer. Although many Muslims are unaware of the fact, the Qur'an explicitly mentions the existence of extraterrestrial life. The existence of creatures of a spiritual nature, such as angels, in the universe, is accepted as a fact by all Muslims, as well as people of other religions, such as*

*Christians. The point that generates excitement among the public, and scientists is the question of whether material life forms like ourselves, which can be found by science, do actually exist outside the earth. (7) The objective of this article is to present evidence from the Qur'an for the existence in the universe, of MATERIAL life forms ("Life as we know it")*"

From here, we can see how this forum post's author wants to tell the public how finding proof for extra-terrestrial life is not so sensational, since the Quran, the holy book in Islam, seems to mention the existence of angels, "creatures of a spiritual nature". Christians are supposed to agree on the belief in the existence of angels, messengers from God. Still, there is also supposed to be evidence in the Quran for material living beings also other places than on Earth, according to an article which the author of the message above refers to.

Now we look at the words that are in Swiss English Forum, but not in Ummah. Here is an excerpt of that list which includes the 30 words with the highest GTF.

**Table 15: 30 words that are used in Swiss English Forum, but not in Ummah, sorted by GTF in descending order**

<b>Word</b>	<b>GTF</b>
zurich	25886
basel	14443
chf	10230
lausanne	8598
coop	7004
migros	5659
bern	5297
zürich	4889
expats	3713
zug	3614
nyon	3056
tram	2070
gallen	2030
ski	2027
canton	1979
skiing	1795

voucher	1761
recycling	1663
rolex	1582
microwave	1427
romandie	1419
fribourg	1398
expat	1396
brunch	1282
manor	1234
groceries	1216
commute	1210
zalando	1200
stucki	1200
sundays	1185

“zurich” has been written in Swiss English Forum as much as 25886 times, which should make sense, since Zürich is said to be the largest city in Switzerland. “basel” is used 14443 times, and Basel is also a city in Switzerland. It has the third largest population in that country. “chf” is used 10230 times, and is the official abbreviation for the Swiss franc, the currency used in Switzerland.

“lausanne” has been written 8598 times, which is natural, since Lausanne is a city in the French speaking part of Switzerland. “coop” which has been written 7004 times and “migros” which is written 5659 times are known to be Swiss supermarket chains . Migros is the largest retailing company and supermarket chain in Switzerland [38]. Both of these supermarket names appear in line 365527:

*“Hi everyone, Does anyone know where to find Palmer's Cocoa Butter to prevent pregnancy stretch marks\*? I've looked in **Coop**, **Migros**, various apothekes, etc... to no luck. Thanks! \*I realize all these lotions and potions may not actually do anything to thwart stretch marks, but figured it's worth trying..', 'mrbumby', '19.02.2013, 10:46'”*

“bern” which has been written 5297 times is known as the federal city (often called the capital) in Switzerland which has the fifth largest population amongst the cities in the country.

“expats” appears 3713 times and one example is shown in line 365643:

*“Do you know other expats here? I often hear English at the Badi but don't know if they live here or travel in from elsewhere.”*

“expat” is an informal, short version of the word “expatriate”. Expats are people who settle voluntarily in another country than their home country [39].

“zug” is used 3614 times and you can see it in line 115112:

*“I don't really see a reason why I would want to learn Swiss German High-German makes more sense. (offtopic: I swear there only live crazy people in Zug!)”*

Zug is a German-speaking city in Switzerland [40].

“trams”, collective transport vehicle on rails in the streets of a city, is mentioned 2070 times, like written in line 364338:

*“They are that nice, they even moved it right next to the HB so you can catch your train and trams home”*

“gallen” has been written 2030 times and for instance in line 360530:

*“meet someone to talk near st. gallen', 'Hi to everybody out there! I would like to meet an English native speaker living near St. Gallen.”*

St. Gallen is the name of a city in north-eastern Switzerland [41].

“ski” is used 2027 times and is exemplified in line 360515:

*“People here play tennis, ski, mountain-hiking, bicycling, etc etc.”*

Here a lot of different sport activities are mentioned. This supports our assumption that we are reading a forum about day-to-day life.

Now we look at the words used in Ummah, but not in Swiss English Forum.

**Table 16: 26 words that are used in Ummah, but not in Swiss English Forum, sorted by GTF in descending order**

Word	GTF
allah	23132
prophet	7251
allahu	5959
akbar	5616
abu	4565
qur	4330
ibn	4290
allaah	3326
quran	3299
untitled	3157
hadith	2905
muhammad	2712

ummah	2299
sunnah	1886
swt	1676
imam	1550
blessings	1459
ramadan	1447
sins	1415
reward	1376
surah	1338
narrated	1312
worship	1258
bukhari	1197
deeds	1192
scholars	1150

The word “ibn” is used 4290 times in Ummah. An example is given in line 75711:  
*“A vociferous critic of Wahhabism, Schwartz is a frequent contributor to National Review, The Weekly Standard, and other publications. Kathryn Jean Lopez: What is Wahhabism? Stephen Schwartz: Wahhabism is an extremist, puritanical, and violent movement that emerged, with the pretension of "reforming" Islam, in the central area of Arabia in the 18th century. It was founded by **Ibn** Abd al-Wahhab, who formed an alliance with the house of Saud, in which religious authority is maintained by the descendants of al-Wahhab and political power is held by the descendants of al-Saud”.*

This extraction is again from SQL file which seems to be presented in an interview. We can see that according to the comment, Wahhabism is a form of Islam which seems to be very conservative. The word “ibn” means “son” [42] and is a part of the name of Muhammad Ibn Abd al-Wahhab who is the founder of Wahhabi movement. Muhammad was in this case son of Abd al-Wahhab. He founded wahhabism in 1792, which was a reformed version of Islam that demanded simple and strict forms of faith and life [43]. The emirs in Darya in Central Arabia collaborated with him, organized riots, and occupied Mecca and Medina from the Ottoman Empire. This Arabic kingdom expanded and is in newer time known as Saudi Arabia [43].

The word “hadith” has been written 2905 times. In Islam, “hadith” means sayings or information that contains statements or acts of the prophet Muhammad [44]. For instance, in line 75730 from “forum\_entries.sql”:

*“I'm going to try keep this short and straight to the point and it is mainly for the kuffar. Islam is NOT as divided as Kuffar and some muslims think. There are Fundamental issues in islam and there are side issues. in Side issues, difference of opinon ARE allowed, As **hadith***

*of Muhammed (sallahualyhiwa sallam) provide evidence for and should not be used to cause disunity amongst muslims. And disunity caused by diffence of opinons in side issues is pure ignorance”.*

The word “ummah” is used 2299 times. For instance, in line 75765:

*“Assalamalikum May allah subhanhutalla nake it eas for all the muslims being tested around the globe for the beliefs and give them great reward for their efforts, and make Islam the dominant **ummah** that we once were.”*

“ummah” is the Arabic word for “people” which often refers to the world community of Muslims who share common ideology and culture. This word unsurprisingly is very common to be found in all kind of Islamic websites, and it is found in Quran, the holy book in Islam. The word is commonly used in Islamic countries [45].

The word “sunnah” which has been written 1886 times, has a meaning similar to the word “hadith” explained previously. Sunnah means “tradition” [46] and can be defined as the way of life based on Islamic norms and rules for Muslims practiced by the prophet Muhammad.

For instance, in line 76300:

*“(...) We know that holding on to the Prophet's **Sunnah** can save a person and bring him or her happiness both in this world and in the Hereafter. These are not idle words! Muslims are exhorted to first hold on to the Qur'an and to secondly hold on to the **Sunnah** of the holy Prophet (i.e. to follow the Shari'ah). The Qur'an is the word of God. The **Sunnah** of the Prophet is the application of Qur'an in life. That is why it is necessary to follow the **Sunnah** for a person who wants to follow the Qur'an and lead a life in accordance with it.”*

The word “swt” is used 1676 times, for instance in line 76312:

*“Allahu Akbar! Ibn Qayyim said in regards to time: ?Every gulp of air that goes out in a cause other than the cause of Allah (SWT) will turn to sorrow and regret on the Day of Judgment.? He also said, ?A sign that Allah (SWT) despises you is when you find yourself wasting your time with trivial matters; in this way you miss your chance of going to heaven”*

“swt” is an abbreviation for a very popular Arabic phrase “Subhanahu wa ta'ala” among Muslims which means “May He be Glorified and Exalted” [47]. This word can be seen very often after the word “Allah” in the Quran or other Islamic texts which show devotedness for Allah.

## **4.9 NGTF Comparison of Steam and Ummah**

“people” is used 1402 times in Steam and 10467 times in Ummah. Our selection from the Ummah forum is much bigger than from Steam, so we definitely need to normalize the frequencies to compare them. “people” has a normalized frequency of 0.367 in Steam and 0.452 in Ummah, which makes the difference less dramatic, so how often the word “people” is used in an arbitrary excerpt in Steam is not really so different from an arbitrary excerpt with the same length in Ummah.

In the list of words that are in both Steam (f1) and Ummah (f2) with an order sorted in descending order by the f1/f2 NGTF ratio, many of the words on the top are associated with the contents and technique of the computer game “Counter-Strike: Global Offensive” (CS:GO). f1/f2 NGTF ratios are mentioned in parentheses after every following mentioning of word from a forum.



“players” (630.4), “competitive” (480.3), “rounds” (160.9), “clan” (152.4), “game” (122.4), “shots” (78.3), “tutorial” (38.6) are examples of game-content related words. The f1/f2 NGTF ratios are very exaggerated (they become much larger) compared to the f1/f2 GTF ratios because the selection from the Ummah (f2) forum is much larger than from the Steam forum (f1). Players are very essential for a game, especially a multiplayer game, and this game seems to be of the competitive kind. Players cooperate in teams and compete against other teams. “clan” refers in f1 to small player teams, while it refers to different kinds of larger groups of people in f2.

**Table 17: Words in Steam (f1) and Ummah (f2) with 30 highest f1/f2 NGTF ratios**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
players	104.0	630.4	204.2
competitive	79.6	482.3	152.0
servers	33.3	201.6	90.8
rounds	26.5	160.9	68.7
clan	25.1	152.4	64.9
harsher	23.3	141.4	85.2
game	20.2	122.4	25.1
mac	18.6	112.8	45.5
mouse	17.9	108.5	57.1
settings	15.8	95.8	43.0
console	13.7	83.2	44.6
terribly	13.6	82.4	42.4
shots	12.9	78.3	47.0
thx	12.9	78.2	41.1
auto	11.8	71.2	37.5
max	11.5	69.7	32.8
undesirable	10.8	65.3	34.1
complains	10.7	64.8	34.6
damn	10.7	64.6	32.2
drivers	10.5	63.4	33.8
kick	9.7	59.0	32.1
hardware	9.7	58.9	32.0
menu	9.3	56.5	38.5
map	9.0	54.3	27.5
timer	8.2	49.9	34.5
isp	7.6	46.3	27.3
switched	7.6	46.2	26.4
score	7.1	42.9	25.0
tutorial	6.4	38.6	23.2
matched	6.4	38.6	23.2

From line 66190:

*“(‘Steam’, ‘9\_t29\_p1’, ‘CS:GO Clan League’, ‘Hello, I am a frequent player and user of the CS:GO community, and I have a small **clan** of 6 people that I belong too. We*

*(AVLNCANADA) were thinking about a concept of a small-league competitive mode. You sign up with your **clan**, and have games with other **clans** that register. The map is picked for you, and it plays like any other competitive match, except that you can track your progress, schedule games, and challenge other **clans**. No money rewards, just fun. Leaderboards, ranks, all that good stuff. What do you guys think? Thanks Chrome', 'ChromeXu', '19 hours ago')*”

“clan” is here even used more than once in the same message, also as the plural “clans”, which is not counted as “clan” in our statistics, but as an own word, since we have not implemented stemming or lemmatizing.

“CS:GO” is a shooting game, as we can see from the use of “shots”. Of course no one is hurt or killed for real in a computer game. It is another story with the extremist forum, where the discussed shots appear to be real. The gamers seem to be preoccupied by “score” in the game than the extremists (f1/f2 NGTF ratio: 42.86).

When we look at more technical terms, we see that “mac” has 112.8 times higher normalized frequency in f1 than in f2 (in reality used 18.6 times more). In the game forum, “mac” refers to Apple’s computer brand Mac(intosh), while in the extreme Islamist forum it is the company Federal Home Loan Mortgage Corporation with the nickname “Freddie Mac”, Mac trucks, MAC address in a computer network, and sometimes also Apple’s computer brand Mac. Other technical terms which are much more used in f1 are “auto” (which refers to some kind of automatic functionality), “mouse” (computer mouse), “hardware”, “controller”, “crashes”/“crashing”, and “isp” (Internet Service Provider).

The abbreviation “thx” for “thanks” has a 78.2 higher NGTF value in Steam, and is in reality used 12.9 times more. The probable reason is that the latter group does not have so many situations in its forums where thanking for an answer fits to the context. There may be more questions about how things work in Steam, while there are more utterings of personal meanings in Ummah. Hence, there are not so many suitable situations for thanking in the latter forum.

Noteworthy is that the word “complains” has a f1/f2 NGTF ratio of 64.8. “ban” has a corresponding value of 33.5. The plural “bans” is a unique word for Steam. Actually, there seems to be some kind of connection between “complains” and “bans” like in the following excerpts from a forum thread:

Line 66171:

*“(‘Steam’, ‘12\_t11\_p1’, ‘I hate this **bans** system’, ‘GOD WHY ?! Valve make something’, ‘damQworld’, ‘3 hours ago’)*”

Line 66172:

*“(‘Steam’, ‘12\_t11\_p1’, ‘Comment’, ‘IT WORKS FINE. Please stop all the **complains** about **bans**.’, ‘Toraqi’, ‘2 hours ago ’)*”

Line 66173:

*“(‘Steam’, ‘12\_t11\_p1’, ‘Comment’, ‘the only people who hate it are the ones getting **bans**’, ‘generatedname’, ‘2 hours ago ’)*”

Here “complains” is a misspelling of “complaints”, which shows that one cannot trust that forum authors always write correctly.

When we sort the words by f2/f1 NGTF, we find the words “life” (NGTF ratio 8.4, GTF ratio: 51.1), “www” (7.94, 48.2), “brother” (6.2, 37.3), “book” (6, 36.2), “god” (4.9, 29.5), “wife” (5.1), “country” (2.4, 14.7), “eat” (2.4, 14.7), “happy” (2.5, 15.1), “friends” (1.9, 11.8).

**Table 18: Words in Steam (f1) and Ummah (f2) with 30 highest f2/f1 NGTF ratios**

Word	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
life	51.1	8.4	4.9
www	48.2	7.9	4.5
brother	37.3	6.2	3.5
book	36.2	6.0	4.6
forum	32.4	5.3	3.2
day	31.5	5.2	2.9
wife	30.8	5.1	3.3
god	29.5	4.9	3.8
look	27.1	4.5	2.9
person	26.9	4.4	2.7
world	25.5	4.2	3.0
posted	21.0	3.5	1.3
originally	20.3	3.3	1.3
mercy	19.2	3.2	2.5
mind	17.6	2.9	2.1
month	17.2	2.8	2.3
learn	16.7	2.8	2.0
course	16.0	2.6	1.9
happy	15.1	2.5	1.9
power	15.0	2.5	2.1
little	14.7	2.4	1.8
country	14.7	2.4	2.0
eat	14.6	2.4	2.0
sent	13.5	2.2	1.8
act	12.9	2.1	1.8
killed	12.3	2.0	1.9
giving	12.2	2.0	1.5
head	12.2	2.0	1.7
friends	11.8	1.9	1.6
goes	11.7	1.9	1.5

For the forums Steam (f1) and Ummah (f2), the f2/f1 NGTF ratio values are understated because f2 is much bigger than f1. One can notice that many of the words are from daily life. The Steam forum is only centred around the game “CS:GO”, and nothing (or very little) from

the gamers' life outside the game is mentioned. At the same time, the messages in Ummah may have a lot of sometimes hidden references to their ideology, but the visible content still can sometimes be related to daily life situations. In the following example, the chase for making a career and trying to become wealthy is accused of being the reason that people give up values like (religious) faith, family, and friends (example from line 83211):

*“(Ummah', '160947-it-s-almost-mother-s-day-O-what-should-i-buy-my-mummy', 'it"s almost mother"s day :O what should i buy my mummy??', 'We live in a world full of people struggling to improve their lifestyle. A world filled with greed, selfishness, and conceit, where every man is more than willing to give up all his values, faith, family and **friends** to acquire wealth. We find ourselves lost and stuck in the midst of people unable to comprehend death, let alone life after death. To whom success is that of this world, where there is no creator to answer to nor lifestyle to adopt. A life similar to that of a dark road leading nowhere. Allah (SWT) has blessed us with the light of iman and the lifestyle of Rasulullah sallahu alaihi wasallam as the road to success. A life of tranquility, simplicity and preparation for the never-ending life. (...)', 'MMS', '23-02-08, 04:01 PM')”*

The words “please”, “post”, “hour”, “account”, “sign”, “care”, “follow”, “page”, “download”, “discussion” have approximately 1 as NGTF ratio, so they are equally significant in both forums.

These words seem to be mainly related to normal usage of an online forum.

When we look at which words that make Ummah different from Steam (Ummah words that are not in Steam), we find many religious words such as “allah”, “prophet”, “muslim”, but we also find some everyday words like “night”, “knowledge”, “woman”, “marriage”, “wrong”, “word”, “husband”, and so on.

Our analysis program could not find any words that make Steam (CS:GO) differ from Ummah.

## 4.10 NGTF Comparison of Swiss English Forum and Stormfront

**Table 19: Words in Swiss English Forum (f1) and Stormfront (f2) with 30 highest f1/f2 NGTF ratios**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
coop	538.8	272.3	96.1
swiss	403.0	203.7	43.1
switzerland	334.2	168.9	28.3
geneva	276.3	139.7	43.3
ski	135.1	68.3	33.6
sundays	118.5	59.9	36.0
commute	93.1	47.0	26.1
canton	76.1	38.5	18.7
vat	72.3	36.5	22.4

brands	70.8	35.8	19.7
gmo	69.3	35.0	23.4
baking	61.8	31.2	19.2
hubby	58.9	29.8	19.0
parcel	58.3	29.4	19.8
scooter	56.8	28.7	18.0
scanning	54.5	27.5	15.7
cigars	52.9	26.7	16.2
stove	52.7	26.6	17.3
haircut	51.1	25.8	15.7
sockets	50.9	25.7	18.2
salary	50.2	25.4	13.4
thx	50.0	25.3	16.5
saturdays	49.5	25.0	15.1
vouchers	49.2	24.9	15.6
shops	48.1	24.3	12.5
checkout	47.9	24.2	15.6
noise	47.1	23.8	15.2
baden	46.5	23.5	14.7
xmas	45.8	23.2	13.7
duplicate	44.0	22.2	13.6

The word “coop” is used 272.3 times more (in a normalized way) in Swiss English Forum (f1) than Stormfront (f2). In Switzerland, Coop is a chain of grocery stores. The same word is found in Stormfront, where “coop” simply is the same as “co-op”, abbreviation for co-operation. The words “swiss” (203.7 as f1/f2 NGTF ratio), “switzerland” (168.9), “geneva” (139.7), “canton” (38.5) are no surprises. “ski” with a f1/f2 GTF-IDF ratio of 68.3 may imply that they do not discuss skiing so much in the right-extreme Stormfront forum. Even the word “hobbit” is used 14.9 times more in the Swiss forum. This is most likely discussion about the new film called “The Hobbit”, which is based on the book with same name. The word “googling” is used 14 times more in f1 (not shown in table) and is a word for searching on the Internet, based on the name of the popular search engine Google. The word “foreigner” is actually used more in the Swiss forum, by a NGTF ratio of 12.6 (not in table above). Else, the words that are used much more in the Swiss forum, are in some category belonging to daily life or holidays. “holidays” is by the way used 11.4 times more than in Stormfront. The word “saturdays” is used 25 times more in f1, which indicates that the ordinary Swiss and other people living in Switzerland discuss activities for the weekends.

**Table 20: Words in Swiss English Forum (f1) and Stormfront (f2) with 30 highest f2/f1 NGTF ratios**

Word	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
jewish	419.8	831.7	248.9
una	338.2	668.6	249.8
obama	188.4	373.0	119.2
nationalist	127.3	251.5	104.5
del	121.4	240.3	104.1

ils	115.5	228.2	115.4
genocide	101.7	201.4	93.7
hitler	99.4	196.7	82.0
dat	92.6	183.2	92.1
muslims	89.3	176.8	72.9
jew	87.2	172.6	64.4
camp	73.9	146.1	75.1
politics	64.5	127.6	62.4
crimes	58.8	116.4	56.0
nationalism	54.3	107.4	51.6
arrested	52.9	104.6	46.2
dire	51.9	102.9	53.4
iran	51.2	101.3	46.5
nuclear	49.4	97.6	53.0
une	48.1	95.1	49.9
quando	47.5	93.8	50.2
mai	43.9	86.6	47.4
liberal	42.4	83.8	38.6
con	41.0	81.1	41.5
agenda	40.3	79.7	41.2
dans	39.8	78.7	44.3
officials	39.6	78.2	40.5
congress	38.7	76.6	40.7
indians	37.6	74.3	40.6
constitution	36.3	71.8	42.0

When we look at which words that are much more applied in Stormfront (f2) than in Swiss English Forum (f1), we see that “jewish” has the top ranking with 831.7 as NGTF ratio. “obama” (373) has also a high NGTF ratio. Then follow amongst others “nationalist” (251.5), “genocide” (201.4), “hitler” (196.7) and “jew” (172.6) give good clues that this could be a right-extremist forum. The word “muslims” has actually a f2/f1 NGTF ratio of 176.8. We also see the words “camp” (146.1) and “arrested” (104.6) which could mean almost anything, but since we know that Stormfront is a right-extreme forum, it makes sense to think about concentration camps, to where enemies of the Third Reich were brought (Jews and political opponents). Important words that are not included in the table are; “political” which is used 34 times more in f2, “white” 32.6, “violence” 32, “terrorist” 31.2, “minority” 30.8.

The following forum message from line 46815 has an example of the word “violence” and complains about mild judgment of African American youth crime, and the author claims that this is even tactics for destroying the white people by means of violence by letting the criminals soon out again:

*“(‘Stormfront’, ‘908007’, ‘Re: African youth crime concern’, ‘Something that is often overlooked is the Judicial System. The POLICE do their jobs and do it well, it is the SENTENCING that is letting us down. (...) THIS WEAK SENTENCING IS PART OF THE WAR AGAINST US. THEY WANT THE PEOPLE THAT DO THIS OUT AND ABOUT CAUSING MORE ATTACKS, VIOLENCE and DESTRUCTION OF US. I used to be*

*sceptical, annoyed at weak sentences and lament at all of this BUT NOW after I educate myself more, I see it as PART OF THE STRATEGY TO DESTROY US or at the very least to weaken us more. With weak sentencing, it will escalate into massive riots and **violence** in retaliation and this is what they want. So whilst the **violence** increases, just don't hold your breath for them to sentence these **violent** criminals to any meaningful prison time. Sorry for all the but this just frustrates me to no end, again that would be their desired reaction along with anger etc.', 'RougeWave', '08-19-2012, 11:38 PM')"*

Some white supremacists think there is some kind of conspiracy for exterminating white people, either by violence and killing or by blending races when a couple of a white and a black person get children together.

If we examine words that are used with the same normalized frequency in both forums, we find amongst others “passengers”, “partners”, “score”, “bring”, “quietly”, “international”, “red”, “business”, “hall”. We assume that both of the forums discuss the same international society, but with different points of view.

We show an example of usage of “international” in Stormfront, where that word seems to be associated with Jews in a negative way. The author of the following message thinks that an international movement of Jews hates all non-Jews.

Line 17413:

*“(‘Stormfront’, ‘900736’, ‘American NS’, ‘(...)Please wake up you non-Jew supporters of Israel, wake and see the **International** Jewish Supremacist for who and what he truly is, a hater of all things not Jewish, and a hater of all things not supportive of Jewish interests. Notice how the **International** Jewish Supremacist insists on Israel's sovereignty and right to self determination, while at the same time he stumps for One World Government as opposed to Nationalism.’, ‘hens are liars’, ‘07-17-2012, 07:53 AM’)”*

“international” is also used in Stormfront forum in a neutral way when mentioning for instance the international community, but internationalism in itself is never (at least not often) praised as something good. In Swiss English Forum, “international” can be used with positive connotation. In line 113262, German-speaking Swiss are praised as international in the sense open-minded to other cultures:

*“(‘EnglishForumCh’, ‘10\_t150’, ‘Hello! Questions about Romandie...’, ‘IMOH: 1 French speaking: grumpier and more stuck up until you really get to know them. 2. German Speaking: Friendlier, more **international**, funnier, and much easier to approach. Girls wise well it just depends where you pick someone up in both places so no generalisation just my opinion though.’, ‘Anthony1406’, ‘28.02.2013, 00:13’)”*

The same word can also be used in a more neutral context, but still there is often some hidden praising of being international, like in the following example from line 113339 where foreigners seem to prefer to let their children go to international schools rather than ordinary Swiss schools:

*“(‘EnglishForumCh’, ‘10\_t159’, ‘**International** school of Lausanne’, ‘And there is also a new **international** school due to open this year in Etoy, so you may also want to take a look at this is the already established schools don't have places for your children.’”*

*http://www.gemsworldacademy-etoy.com/ Although I have no idea how the school will run or know anything about the organisation.', 'stillme', '24.02.2013, 15:41')"*

Else we see that some Anglophone people in Switzerland seem to eat and like vegan and organic/ecological food, like shown in following excerpt from line 115685. This seems to be never discussed in Stormfront.

*"('EnglishForumCh', '11\_t11', 'Vegan-ecofriendly people around here?', 'Quote: mundoplath I would also like to discover **vegan** restaurants, and to find out if there are any organic farms in Zurich where I could buy fresh organic vegetables. Hi, welcome to the Forum. I doubt there is any farm in Zurich it being a City. The word in Switzerland is BIO rather than organic. The main supermarkets Coop-Micros have a good stock of Bio vegetables. Maybe what you want is a market that sells Bio products. Good luck with search for **Vegan** restaurants.', 'moggy', 'Yesterday, 13:56')"*

From amongst others the example above, we can see that some ordinary people (living in Switzerland) are concerned about the wellbeing of the environment. The right-extremists in Stormfront do not seem to discuss that so much, although renewable energy is discussed in an example we show somewhere else in this thesis. The following forum message discusses possible companionship with the American Green Party, but we see no clear sign of eco-friendliness from the right-extreme movement in line 8438:

*"('Stormfront', '898631', 'We need to make our movement appealing to other movements. USA Edition', '(...) what movements would be friendly to ours, if we were to gain influence in theres... Plausible groups that I have pointed out is The Libertarian and Patriot Movements, and to a lesser extant, the **eco** movement like the Green party, and The Tea Party. (...), 'TheTrojan', '07-06-2012, 06:19 PM')"*

The word “brunch”, describing a meal between breakfast and lunch, is also a unique word for the Swiss forum (f1). “ikea”, an international furniture store chain is also only mentioned there.

Unique words for f1 are names of Swiss cities and towns: Zürich, Basel, Bern, Fribourg, Dübendorf, Lugano, and Luzern. We also see that some few German words are used in the forum, like “Gemeinde” (local society).

Unique words for f2 are “romney”, “holocaust”, “stormfront”, “races”, “negroes”, “zionist”, “auschwitz”, “nationalists”, “washington”, “republican”, “multiculturalism”, “democratic”, “parliament”. With other words, it looks like the topics racism, American politics and even politics in general are discussed in f2, but not much in f1.



## 4.11 NGTF Comparison of Steam and Stormfront

**Table 21: Top 30 words in Steam (f1) and Stormfront (f2) sorted by f1/f2 NGTF**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
valve	80.6	544.3	147.3
cheaters	79.1	534.4	154.5
competitive	34.4	232.3	74.8
fps	23.4	158.3	54.4
ping	19.4	131.3	54.5
settings	18.0	121.2	48.2
hacker	14.1	95.2	40.7
thx	12.9	87.1	42.7
harsher	12.7	85.9	51.6
steam	12.6	84.8	32.1
elo	12.0	81.0	40.1
adjusted	11.7	78.8	40.8
hackers	11.6	78.6	33.9
aswell	11.2	75.4	36.0
servers	10.2	69.1	32.7
mouse	10.0	67.5	37.2
clan	9.8	66.0	30.2
accidental	8.8	59.1	29.8
console	8.6	58.0	29.1
undesirable	8.2	55.6	28.4
commands	7.4	49.8	26.4
peaks	6.8	45.9	26.0
players	6.6	44.6	19.2
critisize	6.5	43.6	26.5
maps	6.3	42.2	22.2
menu	6.2	41.9	22.3
workshop	6.2	41.7	22.3
punishments	6.1	41.1	25.3
tbh	5.9	40.0	22.8
def	5.8	38.8	22.3

The word “valve” is used 544.3 times more in Steam than in Stormfront. Valve is here not a device for opening and closing a pipe, but the name of the computer game company behind Steam and also the creator of the game “Counter-Strike: Global Offensive” (CS:GO). “cheaters” is used 534.4 times more in Steam, and refers to people who do not play fair in “CS:GO” and runs instead some special program code for giving themselves some illegal advantages in the game. “competitive”, 232.3 times more used, implies that “CS:GO” is a competitive game, which is correct. In Stormfront, “competitive” can refer to different things, both positively and negatively meant. In the first example, it is claimed that white people

should compete with other races for survival. White people should get more children and take well care of them according to “Riemann” in line 50294:

*“(‘Stormfront’, ‘908830’, ‘Re: Compete Better!’), ‘We do indeed need to focus on and hone our **competitive** survival skills. One thing that we can do about birth rates is to hold in high esteem and aid and abet those who do produce a fine family of offspring. This means we place a high value on homemaking as an activity rather than on a career trying to sell more widgets that no one needs..’, ‘Riemann’, ‘08-23-2012, 01:51 PM’)”*

In the following message from line 48347, crime is claimed to be a competitive sport among non-Americans:

*“(‘Stormfront’, ‘908378’, ‘Re: Arab paralympic trio arresed in NI’, ‘Perhaps it’s all part of **competitive** sport in their home countries? Anyone who commits crime abroad must be a couple of bombs short of a load.’, ‘Pensive’, ‘08-21-2012, 03:24 PM’)”*

Following words are much more used in Steam (f1) than in Stormfront (f2):

“cheaters“ (f1/f2 NGTF ratio 534.4, f1/f2 GTF ratio 79.1), “competitive” (232.3, 34.4), “fps” (158.3, 23.4), “settings” (121.2, 18), “hacker” (95.2, 14.1), “servers” (69.1, 10.2), “console” (58, 8.6), “commands” (49.8, 7.4), “players” (44.6, 6.6), “criticize” (43.6, 6.5), “maps” (42.2, 6.3), “workshop (41.7, 6.2), “punishments” (41.1, 6.1).

There is much more discussion and complaining about cheaters in Steam than in Stormfront. In Steam the discussion is much about players that cheat in the game with hacking, which means that they write and run program code that gives one’s playable character unforeseen and unintended advantages. One example is giving the character immortality. In the following message from line 66210, the player/author complains about player cheating with help from hacking:

*“(‘Steam’, ‘10\_t20\_p1’, ‘CS:GO biggest problem: **cheaters** everywhere!’), ‘You released a new game, you made it popular and made all efforts to remove 1.6 from tournaments, you earn good money out of it as many ppl buy it now, but you DONT DO ANYTHING against **cheaters**. Just played a game now that I couldn’t leave (worst thing ever) and was forced to play against 4 **cheaters** in the enemy team and one with speed+aimbot in my team. At least give us an option to leave when we play with **cheaters** if you DONT DO ANYTHING TO STOP THEM, invest some of the money you earn in some good programs to stop **cheaters** from playing. (...), ‘ADEL’, ‘Dec 2, 2012 @ 8:05am’)”*

When the Stormfront authors discuss “cheaters”, they write about tax cheaters and sport cheaters. The Jews are called cheaters, but it is a little unclear what kind of cheaters they are allegedly to be in line 18327:

*“(‘Stormfront’, ‘900955’, ‘The Viceroy’, ‘Quote: Originally Posted by The Viceroy If you could be born again, as any race you want, how many of you would choose to be born in to the Jewish master race? Ow, dellusions of grandeur... What is that master race you’re talking about? All I see is whinners, cowards, psychopaths, **cheaters**, perverts, paedophiles, liars, crooks, thieves, traitors, manipulators, ad libitum.’, ‘Anonymous’, ‘07-18-2012, 09:30 PM’)”*

Tax cheaters are not popular, and it seems to be even worse for racists when the tax cheaters are foreigners, at least in the example in line 43356:

*“(‘Stormfront’, ‘907165’, ‘Pakistanis among UK’s top tax **cheaters**’, ‘Quote: At least three Pakistanis top Britain’s most-wanted alleged tax-dodgers who are accused of fleeing the UK while owing the government over £765 million. (...), ‘Albiorix’, ‘08-15-2012, 09:48 PM’)”*

Although the following comment from line 50859 claims that cycling is mostly a sport for white people, it has some sense in it when it states that cheaters are cheaters, whatever skin colour they have. Lance Armstrong, a sport cyclist that used doping to achieve better and win in competitions.

*“(‘Stormfront’, ‘908960’, ‘Re: USADA to strip Lance Armstrong of 7 Tour titles’, ‘**Cheaters are cheaters. Their race is irrelevant. Armstrong deserves no sympathy from us. Cycling is a predominately white sport. He cheated other whites.**’, ‘ThinWhiteDuke’, ‘08-24-2012, 02:02 AM’)”*

The word “hacker” is sometimes used instead of “cheater” in Steam, because one needs to fiddle with computer code to cheat in “Counter-Strike: Global Offensive”, like mentioned in line 66230:

*“(‘Steam’, ‘10\_t7\_p1’, ‘Comment’, ‘Without wanting to criticize anything, yes there are a lot of hackers in game, but on 80% of the video you have shown with all the respect was face to face battles.(...) Only actual proof that exists of him being a **hacker** is the 45 kills.(...), ‘zPx’, ‘2 hours ago ’)”*

When “hacker” is mentioned in Stormfront, the discussion is about cybercriminals who exploit software vulnerabilities and security researchers who find and report the same vulnerabilities to software owners. Example is from line 46548:

*“(‘Stormfront’, ‘907937’, ‘Major iPhone security flaw uncovered by **hacker**’, ‘Major iPhone security flaw uncovered by **hacker** A major security flaw that has existed in Apple’s (AAPL) iPhone since the device was first released in 2007 has been revealed by a well-known **hacker**. The iOS security researcher, known publicly only as ‘pod2g’, on Friday published details about the vulnerability, which affects all versions of iOS through to the latest beta release of iOS 6. (...) iPhone SMS security flaw uncovered by **hacker**’, ‘Kommano88’, ‘08-19-2012, 09:41 AM’)”*

In Steam, players often use a map to figure out where they are and to where they are supposed to move in the landscape in a game level. In Stormfront, “map” refers to maps that show the real world or the verb “to map something”, which means something like creating an overview of or collecting data about something.

Line 28902:

*“(‘Stormfront’, ‘903548’, ‘An Attack On Iran’s Civilian Nuclear Facilities Is Also A Radiological Attack On U.S. Soldiers’, ‘Would we tolerate a radiological attack on our soldiers from a Muslim country? (...) Here is a **map** from Russia Today that shows some of our bases in the immediate area: The nuclear reactor in Bushehr is only one of the likely targets that will jeopardize our troops and thousands of civilians by exposing them to high levels of radiation. (...), ‘MichaelCross’, ‘07-29-2012, 05:14 PM’)”*

The word “workshop” in Steam refers to a tool called Steam Workshop, where players can create their own game levels which they can share with other players.

Line 66176:

*“('Steam', '26\_t13\_p1', 'lobby playing!!!', 'Hello! I wanted to suggest things that might set many like it. So, add maps Classic Competitive Matchmaking (clan war, lobby), to have available in multiple maps such de\_mirage\_go, de\_mill\_b2, de\_cobble\_b4, de\_prodigy\_csgo\_se. Or even better would be all the possible maps that can get the **workshop** would play Competitive Matchmaking Classic. Rather than just a pub servers, but also playing a clan war (lobby) Thank you, I hope you understand, and a nice evening!', 'YOGI', '23 hours ago')”*

In Stormfront, “workshop” is sometimes an actual place for getting things repaired, and other times a kind of course or presentation, where the participants can put what they have just learnt into practice in activities and conversations.

Line 61540:

*“('Stormfront', '911482', 'Europe?s Muslim population triples in 30 years', 'Quote: KARACHI: Europe?s Muslim population has tripled over the last 30 years. Many European see Islam as a threat and Muslim fundamentalism as a challenge to their values, particularly those related to women. These views were expressed by eminent scholars and speakers during a **workshop** on ?Multiculturalism in Europe in the light of International Developments after 9/11? (...)', 'White Dude', '09-05-2012, 06:57 PM')”*

“punishments” in Steam seems to refer to what happens to players that cheat. They get namely banned from the game.

Line 66252:

*“('Steam', '9\_t9\_p1', 'Comment', ' (...) Ill be playing 6-7 rounds with bot. Well, if not, you will play with the bot till the end anyways and a person might get banned cause of a bug from Steam or the game. This leaves more room for grievers abuse, so again, harsher **punishments** for griefing. 11. Harsher **punishments** for griefing in competitive matchmaking (...)', 'Kiccpe', '4 hours ago ')”*

In gaming context, a griefer is a player who deliberately annoys other players. According to the message above, such players should be punished, and it does not look they get harsh enough punishments. Cheaters get banned from the game for limited time periods. In Stormfront, the punishments can be something that different kinds of persons can get for bad actions and behaviour, everything from children that misbehave to criminals that do something illegal. Criminals should even be more severely punished according to an author in Stormfront in line 52759:

*“('Stormfront', '909394', 'Thetruthwillsetyoufree', '(...)John F. Tard believes in more prisons and severe **punishments**. He is tough on crime. That should learn those NAMs a lesson. (...). On top of that, prison is full of angry NAMs.', 'Adelaar', '08-27-2012, 02:17 PM')”*

In the other end of the list of words sorted by NGTF, where we see words that are much more used in Stormfront than in Steam, we find “white” ( f2/f1 GTF ratio 525.9, f2/f1 NGTF ratio 77.9), “youtube” (92.7, 13.7), “world” (53.1, 7.9), “look” (44.8, 6.6), “war” (43.9, 6.5), “life” (41.9, 6.2), “power” (37.9, 5.6), “control” (34.7, 5.1), “america” (32.5, 4.8), “killed” (24.8,

3.7). See Table 22 for more words. Many of these words seem to be flavoured by American right-extremists' way of seeing how the world works. They think namely that white-skinned people are supreme compared to others.

In Stormfront, “party” and “movement” refer to political groups of people, and it is natural that these words are frequently mentioned here. Simultaneously, “party” refers to a group of cooperating players in Steam. “movement” is in the game forum simply an action of moving a playable character from one place to another in a game level.

“youtube” is of course used to refer to video clips to be found on the website youtube.com by following links from both forums. They simply do it more often in Stormfront. “power” in Stormfront is used in the context of having power, ruling over something, or being a world power, like the United States are. In Steam, it is *electric* power that is mentioned.

Line 52356:

*“('Stormfront', '909298', 'How many more white folks do we need?', 'We have very little influence or **power** in the U.S. due to our numbers being small compared to the country's total population. (...)', 'Arnwald', '08-25-2012, 09:01 PM')”*

Line 66279:

*“('Steam', '35\_t36\_p1', 'This Ban system', 'I know we need it.. but i got a dc cause my internet its changed my ip adress and when it does this its went down for a half minutes... than i can't rejoin back to my left game... (...) if **power** went down and come back i can reconnect... but when internet freeze i can"T... i hope u guys going to solve this....', 'White\_Rabbit From PROland++', '3 hours ago')”*

World means the same in both of the forums: the planet Earth we live on. However, the word “world” is used in so many different contexts in Stormfront, such that one single example is not really illustrative. In line 66389 there is a player who claims to be the best best in the world at the game “CS: GO”:

*“('Steam', '26\_t25\_p1', 'I AM THE **WORLD** BEST', 'im the best in the **world** at this game', 'SS5/ lance', '10 hours ago')”*

In Stormfront, a war is a real war, at least some kind of conflict between humans. In Steam, war is a more innocent competition between players when they compete against each other in a game session.

Line 66400:

*“('Steam', '13\_t13\_p1', 'lobby playing!!!', 'Hello! I wanted to suggest things that might set many like it. So, add maps Classic Competitive Matchmaking (clan **war**, lobby), to have available in multiple maps such de\_mirage\_go, de\_mill\_b2, de\_cobble\_b4, de\_prodigy\_csgo\_se. Or even better would be all the possible maps that can get the workshop would play Competitive Matchmaking Classic. Rather than just a pub servers, but also playing a clan war (lobby) Thank you, I hope you understand, and a nice evening!', 'YOGI', '21 hours ago')”*

“killed” refers in Steam to killing of processes on a computer, which are of course something quite different from real people, while in Stormfront the same word refers to real murders.

Killing a process means to stop its execution by force. The word “kill” in Steam can also refer to murder of fictional game characters.

Line 66461:

*“(‘Steam’, ‘29\_t8\_p1’, ‘Comment’, ‘Unfortunately, this didn’t help. I verified the cache, renamed the blob, **killed** all but the necessary processes and tried every video setting, with no success. I experience that my mouse speed is terribly slow, also during the in-game menu. The mouse speed during the main menu seems to be normal, but even when selecting a map, the overall speed drops dramatically..’, ‘KeesKachel’, ‘Dec 22, 2012 @ 5:54am ’)”*

**Table 22: Top 30 words in Steam (f1) and Stormfront (f2) sorted by f2/f1 NGTF**

Word	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
white	525.9	77.9	16.4
que	184.8	27.4	18.6
black	138.1	20.5	9.0
race	100.2	14.8	7.6
youtube	92.7	13.7	8.4
country	75.0	11.1	6.3
che	69.0	10.2	6.4
sono	60.4	8.9	6.6
world	53.1	7.9	4.7
look	44.8	6.6	4.2
war	43.9	6.5	4.0
life	41.9	6.2	4.1
dat	38.6	5.7	4.9
power	37.9	5.6	4.1
control	34.7	5.1	3.6
little	34.3	5.1	3.5
real	33.2	4.9	3.4
posted	32.8	4.9	1.9
probably	32.6	4.8	3.4
america	32.5	4.8	2.9
pay	32.2	4.8	3.5
originally	31.7	4.7	1.8
course	31.3	4.6	3.3
plus	28.1	4.2	3.5
day	27.0	4.0	2.8
www	25.2	3.7	3.0
killed	24.8	3.7	2.8
pretty	23.7	3.5	2.7
half	23.0	3.4	2.7
third	22.5	3.3	2.6

The word “kill”, the stem of “killed”, has nearly the same normalized frequency in both forums, which means that the NGTF ratio of “kill” is approximately 1.

Line 69821:

*“(‘Steam’, ‘29\_t47\_p1’, ‘Comment’, ‘ok, for example: **kill** 100 enemies with Deagle ... I had aprox. 70 and now is 0 :( this hapens in all weapons except finished archvivments next... win 100 round in Italy/Dust ... I had 75 / 50 and now is 0 ...’, ‘vsn’, ‘43 minutes ago ’)”*

A word with NGTF ratio approximately equal to 1 is “community”, which in Steam refers to an online community and in Stormfront to a group of people with common opinons in real life. That the NGTF ratio equals 1 means that the NGTF value is the same in both forums. Another word with this attribute is “trade”, which means to buy and/or sell something. In the game platform Steam one can exchange in-game objects or games with each other. This is really not allowed to discuss in the forum, which is meant for topics about the game itself like how to complete difficult levels and how to solve technical issues. In Stormfront the same word describes trade of real objects and trade politics in general. The trade is not necessarily between members of the forum.

We could not find any words that are in the CS:GO forum at Steam, but not in Stormfront. However, apart from non-English stop words, we found amongst others these following words in Stormfront that are not in the CS:GO forum at Steam: “jewish”, “whites” “obama”, “israel”, “government”, “police”, “american”, “news”, “hitler”, “german”, “blacks”, “germany”, “jew”, “national”, “media”, “women”, “family”, “children”, “europe”, “british”, “president”, “negro”, etc. These mentioned words that are the most used among the Stormfront-unique words, summarize basically what the Stormfront writers love and hate. We can by means of many different examples in the text practically say for certain that they hate Jews, the current American president Barack Obama, Israel, different (news) media, and negroes (here: African Americans). President Obama is half European American and half Kenyan, not really a typical African American. To the white supremacists, he is still not white enough. They hate him both for having coloured skin and for standing for the same opinions as other members of the Democrat party do. In line 2103 is an example of Barack Obama somehow being scorned for being a sex symbol:

*“(‘Stormfront’, ‘897166’, ‘Re: Would You Want to Eat Dinner with this man.....??’, ‘Why are they calling him Barack instead of president **Obama**? I get the impression they are using his first name because they are playing to the segment of the white female population which views black males as sex symbols. I am not saying most white females are like this I am simply saying that they seem to be directing this advertisement to that particular segment that does view black males like this.’, ‘flugelrod’, ‘06-30-2012, 01:12 AM’)”*

The Stormfront authors seem to like white people, the American nation, white women, family, children, Europe, and Great Britain. Women are mostly spoken about in a neutral way, but there are some places where Jewish women are mentioned in a not so nice way, like in line 6461:

*“(‘Stormfront’, ‘898185’, ‘Re: Jews DO control the media’, ‘(...) Serious shows with leading Jewish characters began to appear in series like thirtysomething and Northern Exposure, which dealt with issues of tolerance, intermarriage, and assimilation. But in many of the programs that followed, particularly the sitcoms of the 1990s, Jewish men and especially Jewish **women** fell into stereotypical roles that Zurawik describes as “nebbishy boyfriends*

*lusting after non-Jewish **women**" or "Jewish-American princesses and smothering mothers." (...)', 'Revilo', '07-04-2012, 06:22 PM')"*

When the authors have even the tiniest suspicion that white women are abused by people belonging to non-European cultures, then they show their love to the women by verbally attacking foreign-cultured people on the forum. At least the author "realist06" does, when he/she accuses Pakistanis for sexually abused British girls, like in line 3649:

*"('Stormfront', '897549', 'Muslims in UK run underage sex ring consisting solely of white **women**', 'Quote: UK underage sex ring sparks racial tensions The story of Girl A, as she became known in court, is tragic by any measure, but it has also become explosive. Because there is no getting around it: The **girls** are white, and the men who used them as sex toys are Asian Muslims, mostly Pakistanis raised in Britain. And it's not just Rochdale -- roughly a dozen other cases of Asian Muslim men accused of grooming young white **girls** for sex are slowly moving to trial across northern England, involving up to several hundred **girls** in all. Read more: UK underage sex ring sparks racial tensions | Fox News', 'realist06', '07-01-2012, 08:28 PM')"*

In the following forum post, an author called "ConcernedMajority" in Stormfront is worried about western culture and family values and thinks that sitcoms (situation comedies) and music in addition to critical theory applied in education are to blame for that.

Line 3748:

*". 1935 Hitler. Regardless of how you feel about Hitler, he was right, look at how they train our **children** to hate their own culture in our schools, tv and how they corrupt their views on morality through Sitcoms, music, and Crititcal Theory applied in all levels of education. No, we can not sit down and relax when we get home, we must be better Fathers. 3. Create a culture filled home: We need to make sure there are pictures of our **family** up on the wall. Not just our immediate **family**, but our **family** history. We need stories regarding who we are, where we came from, and why we are who we are. The best way to teach is through pictures. (...)', 'ConcernedMajority', '07-01-2012, 11:05 PM')"*

We can deduce that family and children are important at least for the author "ConcernedMajority", and likely also for several other members of Stormfront, since we have not found any protest against this comment. On the contrary, they defend this point of view, like in line 3751:

*"('Stormfront', '897568', 'Re: Taking Back Our Nation By Building Stronger Families', 'When we take care of our families, we take care of our people. www.reclaimingournation.com', 'ConcernedMajority', '07-21-2012, 12:06 PM')"*

Of course it is correct to take care of one's family in ordinary people's eyes as well, but it does not seem normal to have hatred to other people as motivation for family moral. This seems to be an important feature that makes right-extremists differ from other people.



## 4.12 NGTF Comparison of Stormfront and Ummah

**Table 23: Top 30 words in Stormfront (f1) and Ummah (f2) sorted by f1/f2 NGTF**

Word	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
whites	465.3	417.7	137.1
van	222.9	200.1	107.3
los	211.9	190.2	99.2
negro	178.9	160.6	56.3
blacks	178.3	160.1	64.0
obama	152.8	137.1	54.9
para	143.5	128.8	73.5
par	122.1	109.6	62.9
nationalist	87.5	78.5	39.5
dat	74.0	66.5	41.5
pour	72.6	65.2	38.3
alla	64.2	57.6	36.6
diversity	60.8	54.6	30.2
quote	60.3	54.2	12.5
est	58.5	52.5	35.7
germans	58.2	52.2	31.0
poland	56.3	50.6	32.4
een	56.2	50.4	33.2
chicago	55.9	50.1	28.1
dan	53.1	47.7	28.4
white	50.8	45.6	10.9
stalin	47.6	42.7	27.8
candidate	47.4	42.5	25.7
marxist	47.0	42.2	24.8
mayor	46.8	42.0	24.9
multicultural	46.2	41.5	24.6
hitler	45.0	40.4	20.9
polish	43.8	39.3	24.9
nazis	43.2	38.7	21.7
riots	40.3	36.2	22.9

When examining words that are used in Stormfront as well as Ummah, we see that words related to racism, politics, and nationalities are used much more in Stormfront, while words associated with Islam and religions in general are used much more in Ummah. “whites” is used 417.8 times more in Stormfront, and “white” 45.6 times more. “negro”, a word for a dark-skinned person, is used 160.6 times more in Stormfront than in Ummah. The words “diversity” is used 54.6 times more. “multicultural” is used 41.4 times more in Stormfront, although it refers to the opposite of racism, it can be used in racist texts to write about why multiculturalism is wrong. The word “racial” is used 33.1 times more. Amongst the words

that are used much more in Stormfront than in Ummah, we find “obama”, “nationalist”, and “republican”, “marxist”, and “socialism”, words that are related to politics. We also see a pair of violence-related words: riots, and shooting. Else, we see some nationalities: Germans, Poland/Polish/Poles, Greeks, Indians (either people from India or Native Americans), Swedish.

**Table 24: Top 30 words in Stormfront (f1) and Ummah (f2) sorted by f2/f1 NGTF**

<b>Word</b>	<b>GTF ratio (f2/f1)</b>	<b>NGTF ratio (f2/f1)</b>	<b>NGTF-IDF ratio (f2/f1)</b>
ibn	390.0	434.5	149.3
qur	360.8	402.0	95.3
allahu	350.5	390.5	55.3
akbar	244.2	272.0	38.9
allah	204.7	228.0	48.1
quran	194.1	216.2	69.4
prophet	120.9	134.6	48.5
blessings	97.3	108.4	51.9
messenger	96.4	107.4	46.8
ahmad	81.5	90.7	45.9
narrated	77.2	86.0	39.2
companions	75.5	84.2	41.7
abu	74.8	83.4	30.2
salam	68.9	76.8	35.4
hijab	57.2	63.7	34.0
mubarak	53.9	60.1	28.1
muhammad	46.0	51.2	21.8
verses	45.3	50.4	28.0
recite	45.2	50.3	23.6
eid	44.2	49.2	23.7
sheikh	41.4	46.1	22.0
apostle	38.7	43.1	26.7
arabic	37.5	41.8	15.7
umm	32.8	36.5	19.4
permissible	30.1	33.5	19.0
sins	29.5	32.8	16.5
fasting	26.0	29.0	15.9
pious	25.6	28.5	16.3
imam	23.1	25.8	13.7
gaze	22.7	25.3	16.0

The word “ibn” is used 434 times more in Ummah than in Stormfront. This seems to be a part of different Arabic male names and has the meaning son [42]. Members of the Saudi Arabian royal family all have the last name Ibn Saud, which means “son of Saud”. The ancestor of that family is Saud ibn-Muhammed, sheik of Dariyya in the first half of the 18<sup>th</sup> century [48].

After “ibn”, we find the words “qur” from “qur’an”, “allahu”, “akbar”, “allah”, “quran”, “prophet”, “blessings”. Many of these words and also following are related to religion. “apostle”, “sins”, “fasting”, “pious”, “imam”, and “idols” are such words. “apostle” is another word for one of Jesus’s disciples. “sins” are actions of breaking rules of a religion. “fasting” is to eat little or nothing during a specified period of religious causes. In Islam, one eats only in the night during a fasting month called Ramadan. Being pious means to strictly follow the rules of a religion. “imam” is Sunni Islam’s parallel to “priest”. “idols” are gods other than the “official” one in Islam, Christianity, or Judaism. The language “arabic” is mentioned 41.8 times more in Ummah. Arabia is commonly known as the origin territory of Islam.

When we examine words that are in Ummah, but not in Stormfront, we find a lot of Arabic words, and at least some of them are related to the religion Islam. The word “ummah” (used 2299 times) refers to the international community of Muslims [49]. That is likely the background for the name of the forum Ummah and the reason that the word “ummah” is used in the forum Ummah, but not in Stormfront. Many of the unique words in Ummah seem to be Arabic words that we have not heard about. “makkah” (584 times) is most likely another way of spelling the name of Mecca, a holy city in Islam. “hadith” (2905 times) is a collection of religious scripts other than the Quran. They include stories about the prophet Muhammad, what he did, said, and quietly agreed in [44]. The word “ameen” (1005 times) is of course the same as the word “amen” known from the ending of prayers in Christianity. We also see English words associated with religion like “recitation” and “disbelievers”.

A surprising word is “breastfeeding” which is actually used 218 times in Ummah, and not in Stormfront. It is there because of discussions about midwives and new-born babies.

Line 83078:

*‘(‘Ummah’, ‘160909-Mothers-milk-how-to-overcome-any-problems-in-the-early-days-general-information%2Fpage2’, ‘Re: need advice on baby weightloss’, ‘akhi.. I was a midwife and **breastfeeding** counsellor a few years ago.. this link is very good. This is the association I did my counselling with.. Insha‘Allah it will also provide you with some very good reading and resources to talk to the midwives about <http://www.breastfeeding.asn.au/bfinfo/index.html>, ‘carol\_au’, ‘16-07-08, 09:27 PM’)’*

The Stormfront forum includes a lot of words from Italian, French, Spanish, and Dutch, and stop words from these languages are visible among the most used of the words that are in Stormfront, but not in Ummah, where English seems to be the discussion language with only some quotes from Arabic. We also find the word “romney” in Stormfront, and it is used 2406 times. Mitt Romney was the Republican candidate against Obama in the latest American presidential election in 2012. Of course “stormfront” is on this list. We also find “auschwitz” (WW2 concentration camp, 1029 times), “negroes” (plural of “negro”, 1198 times), liberals (opposite of conservatives in politics, 1062 times), “olympic” (1100 times), “nationalists” (931 times), “juifs” (French for “Jews”, 746 times), “guerra” (Spanish for “war”, 601 times), “reich” (like in “The Third Reich”, 627 times), Mexican (610 times), “republicans” (562 times), “racially” (adverb for “related to race”, 442 times), “allemagne” (French for “Germany”, 336 times), “memorial” (403 times), “immigrati” (Italian for “immigrants”, 385 times), supremacist (like in “white supremacist”, 371 times), “gunman” (179 times), “renaissance” (188 times), “neighborhood” (172 times), “slavic” (83 times). The Italian word “religione” is actually used 82 times. Still, words are more related to politics (the English part to American politics), nationalities, racism, and violence.

The word “olympics” is mentioned because of the success African people have at running in the Olympic games today and because of the famous African American Jesse Owens who won some gold medals for running at the Olympic Games in Berlin in 1936 [50]. There seems to be some controversy whether Adolf Hitler behaved rude to him or not, like mentioned in line 4792:

*“(‘Stormfront’, ‘897814’, ‘ProudkiwiNZL’, ‘Quote: Originally Posted by ProudkiwiNZL This damn story again. Even many lemmings know that Hitler never snubbed Jesse Owens, it was in fact the “great” FDR who ignored him. Just before every summer Olympics, the Jew media has to run a hundred, “Hitler snubbed Jesse Owens” stories. An excuse to further brainwash the sheeple. .’, ‘JamesMiller’, ‘07-03-2012, 12:45 AM’)”*

Voluntary neighbourhood work and physical training are connected to keeping the white human body in shape and adoring it. The following text would have been more innocent and normal if the word “white” had been dropped. The text implies namely with that race-connoted word that non-White bodies are not so great, like in line 3846:

*“(‘Stormfront’, ‘897585’, ‘(...)The White human body is absolutely stunning. It has been marveled by all the words greatest cultures. Our people have sculpted it, drawn it, painted it, and sought to adorn it. At the same time, seeing an unhealthy representation of this beautiful creation pains my soul. We have been given this beautiful gift of body, we must cherish it and exercise it. Exercise as a family, take family walks, do **neighborhood** service together. I will not go into all the benefits of exercise in this post, however among the many benefits you will find, happiness, well being, stronger heart, mind, reduced stress, and increased intelligence. We are a beautiful people, we were meant to be healthy, our bodies were designed to function a certain way. Obesity is not natural. The majority of obesity is a direct result from a lack of self control, laziness, and a lack of exercise. (...), ‘ConcernedMajority’, ‘07-02-2012, 01:31 AM’)”*

## 4.13 Word Frequency Analysis of Vigrid

In the analysis of this commentary website, we have for simplicity assumed that one page in the website is a “forum message”. In the most cases this is a correct assumption because one article often uses exactly one page. An obvious exception is the frontpage with summary of and links to the newest (and most important) articles. Most interesting words from Table 25 are “tyske” (“German”), “jødene” (“the Jews”), “tyskland” (“Germany”), “jødiske” (“Jewish”), “jøder” (“Jews”), “holoca” (most likely a misspelling of “holocaust”), “millioner” (“millions”, 844), “krigen” (“the war”).

**Table 25: 30 words with highest GTFs in Vigrid**

Word	Translation	GTF	NGTF	NGTF-IDF
tyske	German	1442	1.000	0.828
jødene	the Jews	1297	0.899	0.821
tyskland	Germany	1242	0.861	0.835
jødiske	Jewish	1218	0.845	0.705
jøder	Jews	1069	0.741	0.778
før	before	1067	0.740	0.478
bildet	the picture	1009	0.700	0.488
auschwitz	Auschwitz	986	0.684	1.014
hele	whole	914	0.634	0.454
holoca	Holoca(ust)	874	0.606	0.810
millioner	millions	842	0.584	0.688
krigen	the war	821	0.569	0.583
norge	Norway	809	0.561	0.647
igjen	again/left	748	0.519	0.418
alt	all/everything	744	0.516	0.400
norske	Norwegian	743	0.515	0.668
tyskerne	the Germans	721	0.500	0.612
hvite	white	719	0.499	0.743
europa	Europe	687	0.476	0.529
fikk	got	675	0.468	0.444
land	country	673	0.467	0.437
dag	day	648	0.449	0.434
ser	sees/watches/looks	638	0.442	0.362
viser	shows	636	0.441	0.381
verden	the world	633	0.439	0.406
flere	more/several	632	0.438	0.387
krig	war	627	0.435	0.526
store	big/large/great	620	0.430	0.406
usa	USA	601	0.417	0.580
gjennom	through	566	0.393	0.384

If we sort the words by (N)GTF-IDF, the order of the same words is “auschwitz”, “tyskland”, “holoca”, “tyske”, “jødene”, “jødiske”, “millioner”, “krigen”, “bildet”. Noticeable is that the word “hvite” (“white”) appears between “jøder” and “jødiske”. “hitler” is mentioned 558

times and appears right after “millioner” in the NGTF-IDF ranking. Adolf Hitler was the national-socialist (nazi) dictator leader of Germany from 1933 to 1945. The country names “england”, “russland” (Russia) and “usa” have high (N)GTF-IDF scores. These countries fought against Germany (and also Italy and Japan) during the Second World War.

**Table 26: 30 words with highest NGTF-IDF values**

Word	Translation	GTF	NGTF	NGTF-IDF
auschwitz	Auschwitz	986	0.684	1.014
blev	became	309	0.214	1.002
tyskland	Germany	1242	0.861	0.835
tyske	German	1442	1.000	0.828
jødene	Jews	1297	0.899	0.821
holoca	Holoca(ust)	874	0.606	0.810
jøder	Jews	1069	0.741	0.778
hvite	white	719	0.499	0.743
jødiske	Jewish	1218	0.845	0.705
millioner	millions	842	0.584	0.688
hitler	Hitler	558	0.387	0.669
norske	Norwegian	743	0.515	0.668
norge	Norway	809	0.561	0.647
tyskerne	the Germans	721	0.500	0.612
krigen	the war	821	0.569	0.583
usa	USA	601	0.417	0.580
waffen	weapon (in German)	326	0.226	0.569
russland	Russia	441	0.306	0.551
europa	Europe	687	0.476	0.529
krig	war	627	0.435	0.526
england	England	392	0.272	0.509
bildet	the picture	1009	0.700	0.488
før	before	1067	0.740	0.478
jødisk	Jewish	499	0.346	0.467
hele	whole	914	0.634	0.454
fikk	got	675	0.468	0.444
land	country	673	0.467	0.437
dag	day	648	0.449	0.434
jøden	the Jew	384	0.266	0.430
igjen	again/left	748	0.519	0.418

The word “waffen” is pulled forward from rank 83 with GTF to rank 19 with GTF-IDF. “Waffen” means “weapons” in German, but here it is most likely a reference to Waffen-SS, a division of the German military during the Second World War. From line 2396:

*“Waffen-SS er europeiske heltesoldater!”*

This sentence means “Waffen-SS are European hero soldiers”, so Vigrid thinks that the soldiers of the German Nazi regime were the good guys. Some other examples imply or sometimes even claim explicitly that the allied forces (from UK, USA and Soviet) fought

specifically for the Jews, like in line 2395, where it is even claimed that they fought against Judaism on the east front:

*“Tyskerne var den eneste siviliserte makt under den 2 verdenskrig. Bildet til høyre en verveplakat for **Waffen**-SS i Nederland hvor 50.000 soldater deltok i den europeiske forsvarskampen mot jødedommen på Østfronten. Til høyre noen av dem på Volkhov-fronten 1942/43.”*

The word “holoca” is most likely really “holoca\$h”, a word for Holocaust used by Vigrid. Holocaust was the Nazi regime’s organized mass murder of Jews during the Second World War. Vigrid claims that Jews lied about the genocide of their own people to get political and economic support. The following extract from line 1778 claims that Western leaders subject themselves to Jews who are supposed to be the Lord people according to the Vigrid author. These leaders must visit the synagogue in Yad Vashem and the Wailing Wall in Jerusalem, so that the Jews believe them. The author claims that the new world religion is Holoca\$h, holocaust, which he/she claims is exploited to give the Jews advantages:

*“Alle vestlige ledere av idag må være hengiven til jødernes sak og dyrke jødene som det utvalgte Herrefolket med enerett til å lede verden. Det er to pliktøvelser de må gjennom som minstekrav for at jødene skal tro dem. Den ene er et besøk i den nye verdensreligionens, **HoloCa\$h**, sentrale synagoge i Yad Vashem mens den andre er et besøk ved Klagemuren i Jerusalem. Her putter de en papirlapp inn i en sprekk i muren med sin bønn om å få tjene Jødene på en eller annen måte.”*

Noteworthy are also the words “drept” (“killed”, used 285 times), “kamp” (“fight”, used 273 times), “politisk” (“political”, used 249 times), “politiske” (used 242 times), “israel” (237 times), and “rase” (race, 239 times). The word “drept” is much used in the context of mentioning how many people were killed in different situations during the Second World War. The word “politisk” (sometimes “politiske”) hints that the texts on the websites are much about politics. “israel” is often mentioned in the texts about newer political history. Vigrid is against the existence of that state. It is according to Vigrid a compensation of holocaust, which Vigrid again does not believe happened. Vigrid even thinks that Jews want to conquer the world, at least somehow fight the white race, like in line 1735:

*“New York Times er eid og kontrollert av jøder og de liker ikke Schwartzers så godt. At de støtter negre som ledd i kampen mot hvite er en annen sak. NYT brakte et bilde av to politimenn som skjøt Obama i form av en ape i Jew Yorks gater. Da ble det rabalder og avisen måtte unnskylde seg. Tilsvarende med Bush var helt greit.”*

The quote says that the New York Times is owned and controlled by Jews. Jews allegedly support black people to fight against white people. The paper had to apologize after printing a caricature of Obama as an ape being shot in the streets of New York. When or if the same happened to the previous president Bush, there was no problem.

The writer of line 1602 in vigridtvedtnet.txt hopes that Iran will attack Israel. He/she says that Iran must attack. The USA, which Vigrid writers call “Jew\$A” allegedly withdrew its fleet from the Persian Gulf when Iran started a military drill there. The author guesses that the war would have started already if the USA had not withdrawn, and Iran will likely not get such an opportunity again:

*“Personlig tror jeg den eneste muligheten Iran har, etter at zionistene i mange år har ført en mangefrontet krig mot dem, er å slå til først med alt de har mot spesielt **Israel** og zionistenes baser rundt den Persiske Gulf og i Irak og Afghanistan. Lar de Jew\$A slå til først er det lite*

*trolig de har nok igjen til å svare med. Dette scenariet så nok Jew\$A som mest sannsynlig også siden de øyeblikkelig trakk hele den 5. flåten ut av den Persiske Gulf da Iran startet militærøvelsen sin. Hadde den blitt igjen i Gulfen tror jeg krigen hadde vært der allerede for en slik sjanse til å slå ut en hel flåtestyrke ville ikke Iran få igjen. Det ville vært av samme nivå som slaget i Teutoburger Wald da Herman Cheruskeren utraderte Romerrikets elitelegion.”*

The word “rase” is used when the website articles discusses races, the superiority of the white race, and that different races must not be mixed with each other. According to this racist ideology, this means that Jews and Africans are not allowed to get children with ethnic Europeans.

Here is an example that hints to a conspiracy theory which claims that Jews try to rule the world (from line 1552 in *vigriddtvednet.txt*, a file that includes much of the contents of *vigridd.net*): *“Takker for mailen din. Mitt engasjement kan deles i to deler: Del 1 og høyest prioritert er kampen mot den jødiske verdensmakten og del 2 Kampen for den nordiske rase og vår egen religion og kultur.”*

The writer of the message has as goals to fight against the Jewish world power and fight for the Nordic race and its religion and culture. *Vigridd* members believe in the Nordic gods like *Odin* and *Thor* [3], so the mentioned religion here is actually not Christianity.

Some stop words are removed manually from the top of the list, because some words from the stop word list could not be recognised due to issues with encoding of the Norwegian letters æ, ø and å.

The right-extreme organization *Vigridd* ceased officially to exist in 2009 (at least according to Wikipedia) [49], although the website *vigridd.net* still exists in April 2013 with the organization’s leader *Tore Wilhelm Tvedt* as editor (as it looked like) and with regular updates. *Tvedt* seems to have written many of the newer articles on the website.

## 4.14 Word Frequency Analysis of *Diskusjon.no*

*Diskusjon.no* is a website with all kinds of forums in Norwegian. For this thesis we have selected this forum about politics to be used as a text collection for comparison with the *Vigridd* website. The Norwegian politics and society forum is dominated by the following words:

*mars* (March, used 48390 times), *norge* (Norway, 37625), *våpen* (weapon, 29702), *innlegget* (the entry, 26766), *redigert* (edited, 25182), *helt* (hero, 21390), *norske* (Norwegian, 18423), *venstre* (name of a Norwegian, liberal party, 18110), *staten* (the state, 18009), *anarki* (anarchy, 16467), *høyre* (name of a Norwegian, conservative party, 16252), *februar* (February, 15026), *ønsker* (wishes, 14874), *kvinner* (women, 14301), *sosialistisk* (Socialist, 13662), *venstreparti* (Solististisk Venstreparti = a Norwegian socialist party, 13662), *politiet* (the police, 13474), *arbeiderparti* (Labour party, 13464). “*mars*” is most likely a month when many of the messages in the forum were posted, namely March, and has nothing to do with the planet. We also see the word “*februar*” (February).

The words “*innlegget*” (the entry) and “*redigert*” (edited) are not so interesting, because they are in the most cases used together in the phrase “*Dette innlegget har blitt redigert av*” (“This



entry is edited by”), and is used in all forum entries that have been edited after the first publishing. There is no point in finding a particular quote for “norske” (Norwegian) because that word is used simply to tell that something or someone is Norwegian.

Among these words we find the names of Norwegian political parties: Venstre, Høyre, Sosialistisk Venstreparti og Arbeiderparti(et). The word “staten” (the state) and “politiet” (the police) are related to the government.

The word “staten” obviously occurs when one discusses different aspects of the state, which can be the Norwegian state, or any other state it is compared to. Here is an entry from a forum thread where “anarki” (anarchy) is discussed as an alternative to governing Norway. In an anarchy, there is no government, and hence no state. The thread includes arguments for and against anarchy in Norway. The following entry says that freedom from the state is desirable, but other states would attack and occupy the territory, if the Norwegian state ceased to exist. If anarchy should work, according to the author, all states in the world must have ceased to exist, as described in line 13874:

*“(’Diskusjon’, ’100\_t38’, ’Ville et **anarki** ha fungert i Norge?’, ’**Anarki** i Norge i dag ville antageligvis ikke fungert, da andre stater i nærheten ville ha angrepet og drept deler av befolkningen for å få makt over området. Frihet(fra **staten**) er noe jeg ønsker og strever etter, men det er nok umulig å bare legge ned en state, man måtte lagt ned alle statene som finnes på jorden.’, ’Anonymous’, ’4. februar 2013 - 16:39’)”*

Weapons (våpen) are discussed in the context of whether the police should always be armed or not. Amongst others the author of the following post thinks that violence between the police and criminals will escalate if the police carry weapons all the time, instead of storing them in the car until there is need for them. Other people think the police should be armed all the time, so they can use them at once when they need to. The author of the following post from line 138289, suggests that the police stop violent people with tazers (electric shock pistols) first, instead of starting directly with ordinary guns.

*“(’Diskusjon’, ’100\_t42’, ’Gjør dere klar til økenede bruk av **våpen** ifra Politiet’, ’Legg merke til denne tråden viss politiet blir generelt bevæpnet, det kommer til å ta helt av dessverre, Sverige tilstander her kommer vi.. <http://www.dagbladet...rbund/24260556/> Jeg er helt enig at politiet skal ha **våpen** i bilen og at operasjonsleder/operasjonssentral skal være den som bestemmer bevæpning, men at de skal gå med **våpen** på seg 24/7 er ekstremt negativt å vil føre politiet mer vekk ifra oss sivile samt økende bruk av **våpen** unødvendig.. Hva med først ta ibruk tazer som faktisk er ganske effektiv selv mot ekstremt rusa mennesker, da nervesystem treffes. Dette innlegget har blitt redigert av DaniNichi: 8. november 2012 - 19:09’, ’Anonymous’, ’8. november 2012 - 19:07’)”*

“ønsker” is the third person singular in the present tense of the verb “å ønske” (to wish) or the plural of the noun “ønske” (wish). It is always used for describing what somebody wants. In the following message, the author wishes freedom to do what he/she wants without a small political party (Sosialistisk Venstreparti, SV) in the red-green government coalition introducing prohibition against for instance stripping.

Line 138383:

*“(’Diskusjon’, ’100\_t46’, ’SV-ledelsen vil forby stripping’, ’<http://www.abcnyhete...forby-stripping> Når skal dette formynder partiet få siste spiker i kista slik at vi normale liberale*

*mennesker som ønsker frihet kan leve uten frykt for at et mikro parti skal få igjennom diverse flere nisje forbud?', 'Anonymous', 'i dag, 13:28')"*

How women shall be treated in society is often discussed in the political forum at Diskusjon.no. The following excerpt is from an information forum thread. The message includes lists of different political cases, and for each case, one sees a list of which parties are for the case. Here is a case that the managing board in each Norwegian company must have at least 40 % women. We see which parties are for and against (“mot”) in line 138893:

*“(Diskusjon', '100\_t64', 'Det politiske landskap - Informasjonstråd', '(...)Minst 40% kvinner i bedriftenes styrer(kvotering) For: Det Norske Arbeiderparti, Sosialistisk Venstreparti, Venstre, Høyre, KRF Mot: Fremskrittspartiet, DLF Intet offisielt standpunkt: Uvisst: Demokratene, Pensjonistpartiet, Rød Valgallianse, Senterpartiet, Kristent Samlingsparti, Kystpartiet (...)', 'Anonymous', '23. desember 2005 - 15:21')"*

**Table 27: 20 words with highest NGTF-IDF values**

Word	Translation	GTF	NGTF	NGTF-IDF
kristelig	Christian	13068	0.232	1.460
rød	red	12870	0.229	1.438
folkeparti	People's Party	12672	0.225	1.416
valgallianse	alliance of parties	12672	0.225	1.416
fremskrittspartiet	the Progress Party	13068	0.232	1.299
senterpartiet	the Centre party	11484	0.204	1.283
sosialistisk	Socialist	13662	0.243	1.259
venstreparti	Socialist party	13662	0.243	1.259
arbeiderparti	Labour party	13464	0.239	1.172
venstre	Liberal Party (“left”)	18110	0.322	1.125
høyre	Conservative Party (“right”)	16252	0.289	1.118
norske	Norwegian	18423	0.327	1.054
dlf	The Liberal People's Party	13070	0.232	0.976
våpen	weapon	29702	0.528	0.946
demokratene	Democrats	8316	0.148	0.827
merknad	remark	6732	0.120	0.752
norge	Norway	37625	0.669	0.734
standpunkt	point of view	9901	0.176	0.717
intet	nothing	8913	0.158	0.711
kystpartiet	the Coast Party	7128	0.127	0.709

When we sort words by NGTF-IDF like in Table 27, we see that most of the top 20 words are names or parts of names of Norwegian political parties. We also see the word “våpen” (weapon) which is discussed in the political debate. The word “merknad” (remark, comment) is used before a short information remark in an information thread about the cases the different parties are for and against, like in line 138893 in forum\_entries.sql:

*“(Diskusjon', '100\_t64', 'Det politiske landskap - Informasjonstråd', '(...) Merknad: Det er her spørsmål om Norge skal innføre sanksjoner overfor Israel i Palestin/Israel konflikten. Diskuter konflikten her | Diskuter saken her Mandatfordeling (for/mot): 15-38. Det er ikke noe flertall for sanksjoner. (...)', 'Anonymous', '23. desember 2005 - 15:21')"*

“norge” is the Norwegian word for the country Norway, whose politics is discussed in this forums. No examples are necessary for that word.

“standpunkt” means stand or point of view of a person or a political party. This word is also much used in one or more information thread on the website like shown in line 138896, where a forum administrator has updated information about the FRP’s (Fremskrittspartiet) point of view:

“(’Diskusjon’, ’100\_t64’, ’Det politiske landskap - Informasjonstråd’, ’Har gjort litt "oppgradering"; - Informasjonstråd og diskusjon om og rundt FRP lenke lagt til - Setter inn fullt partinavn (fortsatt under utarbeidelse) - Oppdatert Fremskrittspartiets **standpunkt**. Nå er det lenker til nesten hvert eneste av deres **standpunkt**. (..), ’Anonymous’, ’23. desember 2005 - 15:23’)”

The word “intet” means “nothing”. No example is needed, since it is used in the same way as the English translation.

## 4.15 NGTF Comparison of Vigrid and Diskusjon.no

Table 28: Top 30 words in Vigrid (f2) and Diskusjon.no (f1) sorted by f2/f1 NGTF

Word	Translation	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
jødene	the Jews	6.6	255.7	37.1
jøder	Jews	5.4	209.7	35.0
bildet	the picture	5.1	197.9	22.0
tyskland	Germany	3.1	121.8	21.1
hitler	Hitler	2.8	110.0	30.3
verdenskrig	World War	2.0	76.5	16.6
vigrid	Vigrid	1.7	65.1	16.1
løgn	lie (untruth)	1.4	54.1	15.5
polen	Poland	1.2	48.7	15.8
britiske	British	1.1	43.8	14.7
boken	book	1.1	41.2	11.2
såkalte	so-called	1.1	41.2	10.3
øst	east	1.0	40.4	12.4
england	England	1.0	38.5	12.9
sannheten	the truth	1.0	38.2	10.5
røde	red	0.9	36.9	11.7
døde	dead	0.9	34.7	8.9
europa	Europe	0.9	33.7	7.6
holocaust	Holocaust	0.8	32.7	11.2
anti	anti	0.8	32.4	11.3
iran	Iran	0.8	29.6	12.5
engelske	English	0.7	28.8	10.0
artikkelen	the article	0.7	28.8	9.0

løgnene	the lies	0.7	28.4	9.2
militære	military	0.7	28.2	9.3
kjente	known	0.7	27.9	8.8
senere	later	0.7	27.4	7.0
sto	stood	0.7	26.7	7.1
fanger	prisoners	0.7	26.5	9.8
dens	its	0.7	25.6	8.3

In Vigrid, the word “jødene” (the Jews) has a 6.6 times higher GTF value and a 255.7 times higher NGTF value than in Diskusjon.no. (“jøder” GTF ratio 5.4, NGTF ratio 209.7). This means that the word “jødene”/“jøder” is much more used in Vigrid and is much more important than in the political forum at Diskusjon.no. The authors of Vigrid seem to be much more preoccupied or even more obsessed by Jews than ordinary people discussing politics at Diskusjon.no.

According to Vigrid, the Jews seem to be a kind of master people which controls nearly everything in the Western world, while they are victims of the Nazis according to Diskusjon.no.

Line 2512 in vigridtvedtnet.txt, a Vigrid writer claims to have proof that Jews influenced the Swedish press during the Second World War and lied among the situation in Norway. According to the Swedish press, Norwegians were starving and the Norwegian nationalist party Nasjonal Samling ruled the country tyrannically:

*“Etter den svenske presses utsagn og etter den oppfatning som den har skapt i det svenske folk, råder det i Norge alminnelig hungersnød og elendighet, foruten det frykteligste skrekkelige utøvd av Nasjonal Samling i sin alminnelighet og av Hirten i særdeleshets. (En koreansk professor har laget illustrasjonen til høyre for å vise hvor totalt jødene kontrollerer all media over hele den vestlige verden. I Sverige under 2. verdenskrig var det likedan med det jødiske Bonnierkonsernet i spissen).”*

In ine 139829 in forum\_entries.txt Jews are indirectly mentioned as victims of Nazis. The following excerpt mentions that some people that support the Nazis and defends Adolf Hitler’s and Heinrich Himmler’s warfare and treatment of the Jews, do not like to be called Neo-Nazis. The text also claims that one does not need to be racist when regarding other (non-Western) people as inferior. The author refers to this as being “culturally ethno-centric”, where one considers the western culture as superior, not the race of white-skinned people: *“(‘Diskusjon’, ‘101\_t33’, ‘Om debatt og personfokus - hva ønsker dere’, ‘(...) Man er ikke rasist om man synes at alle utlendinger, og spesielt innvandrere, er underlegne fjotter -- da man kan være kulturelt etnosentrisk, i stedet for rasist og komme til samme konklusjon om ikke-vestlige, basert på vestlig kultur heller enn en arisk rase. MEN ingen ønsker å bli stemplet som noe de ikke identifiserer seg med, slik som enkelte mennesker som støtter **vigrid**, forneker holocaust, støtter Nasjonal Samling, mener den ariske rase er overlegen, støtter Nazistene og forsvare **Hitler** og Himmlers krigføring og behandling av **jødene** -- men reagerer med vantro når man kaller dem for nynazist. (...), ‘Anonymous’, ‘6. mai 2010 - 02:25’)”*

“hitler” has a 2.8 times higher GTF value in Vigrid (NGTF ratio 110), which seems to be a good clue that this is a nazi website. As an example fo the use of the word “hitler” in

Diskusjon.no, one can look at the example for “jødene” above, where that dictator is implicitly criticized for his warfare and (bad) treatment of Jews.

Line 3158 in vigriddtvedtnet.txt praises Adolf Hitler for restoring German prosperity and honour after that people had been punished by the Versailles treaty. Hitler is therefore a hero to the person who wrote this in Vigridd and other examples imply that this is a general opinion on that website:

*“Tyskland bukket under av Versailles-diktatets tvang og straff og valgte til slutt Adolf **Hitler** til å rydde opp og gjenreise velstand og ære for det tyske folk. Det gjorde Hitler på en mesterlig måte.”*

The neo-nazi website also seems to be much more preoccupied by “tyskland” (Germany), “verdenskrig” (World War), “vigridd” (of course), “polen” (Poland), “england”, “iran”, “boken” (the book), “holocaust”, “palestina”, “sannheten” (the truth).

The word “tyskland” (Germany) in Vigridd is used as shown above, as a victim of the Versailles treaty of 1919 after Germany lost the First World War. Right-extremists are not alone in the opinion that Germany was treated unfair after that war. There are other examples of how unfair Germany was treated, also after its defeat in the Second World War. In line 161791 in forum\_entries.sql, the political situation of Germany today is compared to how it is in Norway. It refers to a URL where there is an image of Germany slightly to the left of Norway in politics, but also slightly more authoritarian:

*“(’Diskusjon’, ’20\_t49’, ’Verdens GeoPolitiske Spektrum’, ’Mens man faktisk kan utfra den gamle synsemetoden omtrent plassere alle Norges partier på en høyre-venstre akse, selv om noen ikke helt "passer" som f.eks. venstre, som vi ikke vet om vi bør sette til venstre eller til høyre, så komplisere det hele seg mye om vi begynner å lure på om f.eks. **Tyskland** ligger til høyre eller venstre for Norge? For å prøve å gi en rimelig skildring har jeg kikket på "tyngden" av Staten i økonomien, og på hvorvidt Staten bruker sine penger på venstre-saker (sosial sikkerhet:Pensjon, sykehus, skole, etc) eller på høyre-saker (Fysisk sikkerhet: Militære, Politi, Domstoler, etc) Konklusjonen er som følger: <http://postimage.org/image/8h1pv3cf9/> Nå skal det bemerkes at dette ikke gir ett fullt bilde, ettersom Staten (Politikerne) har ett til "våpen" som de bruker flittig for å bestemme hva vi skal eller ikke skal gjøre, nemlig lover og regler. (...), ’Anonymous’, ’i går, 00:54’)”*

Of course it is natural that Vigridd talks much about itself, with a GTF of 1.7 times the corresponding value in Diskusjon.no (NGTF ratio 65.1), which means that their writers talk more about Vigridd than Diskusjon.no writers do. For an example of how “vigridd” can be mentioned in Diskusjon.no, see the example for “jødene” earlier in this section. On Vigridd’s own website they of course use that word to tell what they as a group has a common opinion about something. In the following quote from line 3643, Vigridd boasts of itself the one to give real facts, not politically correct svada. In this excerpt it is also mentioned that Fridtjof Nansen supposedly was a right-extreme before that expression was invented:

*“Mens altså media gir deg politisk korrekt svada gir vi i Vigridd deg fakta som vil gi deg et mere realistisk syn på historien. Nansen var forøvrig høyreekstrem og medlem av Fedrelandslaget som ville vært kalt naziorganisasjon i dag.”*

“boken” (GTF ratio 1.1, NGTF ratio 41.2) implies that Vigrid has many more references to books than Diskusjon.no has.

The following is from line 2370 in the text file vigridtvedtnet.txt:

*“Da krigen endelig var over fikk nok ikke tyske soldater samme generøse tilbudet. De ble drevet til ulovlig minerydding og misbrukt på det groveste i forhold til all lov og rett som gjelder krigsfanger etter kampslett. Denne **boken** gir innblikk i en lite kjent ulykke og kaster lys over Norges ulovlige atferd.”*

The name Breivik, the last name of Norwegian neo-Nazi terrorist Anders Breivik Breivik, has a f2/f1 GTF ratio of 0.4, but a f2/f1 NGTF ratio of 14.4 This means that Breivik is really mentioned more frequent in f1, Diskusjon.no, but if we compare the normalized frequencies in the two websites, then that name is in a sense mentioned oftener in f2, Vigrid. Example from line 4124 in the Vigrid text collection file:

*“Da jeg leste Dagbladets artikkel om Breivik som er linket over kom jeg til å tenke på at dette er jo en litt mild utgave av noen historiske personer jeg har truffet ofte i litteraturen om den verste katastrofen som har rammet Europa siden kristendommen ble innført; den 2. verdenskrig.”*

The word “fanger” (prisoners) has a 26.5 times higher normalized term frequency in Vigrid (f2/f1 GTF ratio 0.7).

“militære” (military) has f2/f1 GTF ratio 0.7, and NGTF ratio 28.2). “militære” refers to any army during the Secound World War and today.

“løgnene” (the lies) has the f2/f1 GTF ratio 0.7, but the f2/f1 NGTF ratio 28.4. “løgnene” refers to lies that Vigrid and other Neo-Nazis claims that the allied nations told during the war and after it.

Example of “løgnene” in line 4780 in vigridtvedtnet.txt: *“(…) Fra de HoloCa\$htroende hører vi mye mas om moral og etikk, ved siden av alle **løgnene** de spyr ut av seg. La oss fronte dem også på det etiske området. Min påstand er at SS-Oberführer Oscar Dirlwanger i sine råeste handlinger og kamper i terrorkrigen sto på et vesentlig høyere etisk og moralsk nivå enn de som planla og gjennomførte bombingene av de tyske byene med sivilbefolkningen som prioritert mål.”*

One can see an example of how holocaust is used in the example for “jødene” above. Line 4056 in forum\_entries.txt shows an example of how it is used in Vigrid:

*“De anglozionistiske regimene under Roosevelt og Churchill gikk sammen om verdens største **Holocaust** og konsentrerte folkemord med sin teppebombing av de tyske byene for å drepe flest mulig sivile tyskere. Når bombene hadde gjort sitt og tyske kvinner og barn flyktet fra flammehavet gikk de angloamerikanske jagerflyene over til lavangrep med maskingevær mot sivilbefolkningen og ellers alt som rørte seg. (...)”*

Here the author refers to another Holocaust, the bombing of civilians in German cities, which like the genocide of the Jews really, but the Vigrid sympathisers think this is the really holocaust, where some Jews only accidentally were killed by the attacks of the USA and UK.

“palestina” has a normalized mentioning of 25.4 times higher than in the Diskusjon.no forum. The reason is that Vigrid supports the independence of Palestina because Vigrid is against the existence of the Israeli state.

Line 5304: “Hele historien er kokt opp av jødiske og allierte interesseorganisasjoner som krigs- og etterkrigs-propaganda for blant annet å sikre dem penger, medynk og støtte til å invadere **Palestina** for å opprette Israel. Hvor de straks gikk i gang med masseslakt og fordrivelse av den arabiske befolkningen som i Deir Yassim.”

“sannheten” has a 38.2 times higher NGTF. Vigrid wants the truth to be known, but their truth seems often to be very different from other people’s truth. Vigrid thinks that the Jews cooperated with the allied nations to fight down and kill all Germans, which is never proved to be true. Here is an text from line 4832 in vigridtvedtnet.txt that claims that Francisco Franco, the fascist leader of Spain, was not a dictator. He is praised from preventing communists (“den røde pesten” = “the red pest/plague”) from gaining the power in Spain.

“Francisco Franco (4. desember 1892 – 20. november 1975) blir av historieskriverne omtalt som en diktator. To europeere til fra denne perioden blir også omtalt som diktatorer og fiender av fred, frihet og menneskelighet, nemlig Hitler og Mussolini – de to statslederne som hjalp Franco og de spanske nasjonalistene å bekjempe den røde pesten som kjempet for å overta Spania under det historien kaller den spanske borgerkrigen. Våre historiebøker forteller **sannheten** for den som har lært å lese: noter deg det som står der og dra den logiske slutningen at **sannheten** er det stikk motsatte. I denne artikkelen forteller jeg den virkelige historien om den spanske borgerkrigen.”

Line 140970 refer to a decision the Norwegian Defence had to take some years ago about from where they should buy fighter planes. It was finally decided to buy the American F-35, not the Swedish Gripen. The following commentary claims that the truth is that the Swedish plane simply was not good enough for the Norwegian requirements:

“(’Diskusjon’, ’101\_t74’, ’Norges kjøp av kampfly - Debatt og synspunkter på kjøp!’), ’(...)Det er ikke tvil om at det blir svinaktig dyrt å kjøpe nye fly, men dessverre for Norge så eksisterer det pr. idag intet annet reellt fly-alternativ. Hadde Gripen vært bedre enn F-35 hadde jeg selvsagt støttet Gripen, men sannheten, som tydeligvis er svært vanskelig for noen å svelge, er at Gripen som flymaskin ikke nådde opp til Norges minimumskrav. Dette innlegget har blitt redigert av Vice: 10. mai 2009 - 16:54’, ’Anonymous’, ’10. mai 2009 - 16:50’)”

This shows that the two different websites discuss completely different things when they discuss the truth.

**Table 29: Top 30 words in Vigrid (f2) and Diskusjon.no (f1) sorted by f1/f2 NGTF**

Word	Translation	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
tråd	thread	861.6	22.1	12.0
sitat	quote	791.6	20.3	15.3
korea	Korea	669.4	17.1	10.0
sigøynere	Gipsy	646.8	16.6	10.0
mexico	Mexico	613.8	15.7	9.4
http	http	603.2	15.5	9.0
debatt	debate	515.7	13.2	7.2
samlings	collection’s	486.0	12.5	6.6
sveits	Switzerland	462.0	11.8	7.8

brorskap	brotherhood	450.0	11.5	7.1
skrev	wrote	429.6	11.0	3.6
diskuterer	discuss	427.0	10.9	6.7
frp	a Norw. populist party	417.2	10.7	5.8
oppfordring	request	415.8	10.7	6.6
seriøst	serious(ly)	383.8	9.8	6.6
bord	table	382.7	9.8	6.3
synspunkter	points of view	335.8	8.6	5.5
våpen	weapon	312.7	8.0	6.3
mars	March	312.2	8.0	2.8
www	www	312.1	8.0	5.9
forby	forbid	310.1	7.9	6.4
argumenter	arguments	305.6	7.8	6.9
negativt	negative	297.0	7.6	5.0
tvinge	to force	290.3	7.4	5.2
foreldrene	parents	268.9	6.9	6.4
enig	agreed	261.3	6.7	5.8
rask	fast	258.2	6.6	4.5
regime	regime	253.5	6.5	4.4
ideologi	ideology	252.7	6.5	6.2
overbevise	convince	231.8	5.9	4.1

Following words are used more in Diskusjon.no (f1) and the f1/f2 NGTF (f1/f2) are shown in parentheses. The people discussing at Diskusjon.no seem to be preoccupied by “sitat” (20.3), “korea” (17.1), “sigøynere” (16.6), “offisielt” (5.1), “sveits” (11.8), “debatt” (13.2). The first word does not seem very special. “korea” likely refers to the still ongoing cold war (propaganda war) between North Korea and South Korea. Lately, North Korea has namely threatened to attack South Korea and the USA.

Line 139332 in forum\_entries.sql:

*“(’Diskusjon’, ’9\_t79’, ’Nord Korea: Er fyren idiot?’, ’Skjønner ikke engang at de tør å true USA med atomkrig. Nord Korea har I kanskje et fåtall atombomber mens USA har, ja la oss bare si masse og de har midlene for å få de levert. Ønsker de en utholdelses atmokrig med USA har de tapt eller rettere sagt de har tapt uansett hvilken krig de starter men de burde se litt på mulighetene for å kunne vinne en slik krig før de starter. Og tomme trusler er nettop det, tomme trusler.’, ’Anonymous’, ’10. mars 2013 - 17:47’)”*

“sigøynere” is likely used much more here because of the many Gipsys or Romani coming from Eastern Europe to Norway to beg for money. It must be specified that not only Romani come to Norway to beg. In Norway there has therefore lately been a discussion whether begging should be forbidden again. It is surprising though, that ordinary people talk more about Romani than the Vigrid website does. The reason is maybe that Vigrid has not so many authors, and they hate the Jews more than the Romani. Begging became legal in Norway some years ago to reduce crime among drug addicts, but then the beggar immigration problem appeared instead. In the following example, it is mentioned a sign in a subway train that demands Gipsys to leave the train because they allegedly beg and play music.



Line 139936:

*“('Diskusjon', '100\_t66', '«Jeg vil ikke ha noen sigøynere om bord. Vennligst gå av»', '«Jeg vil ikke ha noen sigøynere om bord. Vennligst gå av» Hva syns dere om meldingen fra T-banesjåføren? Dette er selvfølgelig hårreisende, men det er faktisk et problem her i Oslo med tiggere & folk som spiller musikk på T-banen. Har opplevd det flere ganger og nekter å gi noe som helst.', 'Anonymous', '18. februar 2013 - 12:44')”*

The authors in Vigrid may be a little confusing in what they think about the Jews and Gipsys. In the following quote, German doctors are honouring for helping such people (line 8115 in vigridtvedtnet.txt):

*“Igjen, midt under krigen driver altså tyskerne ved hjelp av europeiske toppleger grunnleggende forskning for å hjelpe speiselt jøder og sigøynere. Hvordan kan dette rime med legenden om at tyskerne for enhver pris prioriterte utryddelse av de samme gruppene? Komplott idioti.”*

In this quote it is neither mentioned whether Vigrid likes Jews and Gipsys or not, but it is mentioned that the Germans did not prioritize extinction of them at any price. Of course not all Germans wanted these people gone, but it is difficult to see what wants to convey with this message.

“sveits” (Switzerland) is likely mentioned because of hiding of taxable money there, because Switzerland has a liberal, economical policy. That country is also praised and criticized at Diskusjon.no for its direct democracy, at least in line 140564:

*“('Diskusjon', '9\_t60', 'Partibarometer for januar, februar og mars 2013', 'Interessant at hele 10 av 15 som har stemt så langt ser på Sveits som et politisk "forbilde". Er det forholdet til EU som fascinerer, eller den liberale økonomiske politikken? Dette innlegget har blitt redigert av LucarioX: 13. januar 2013 - 17:03', 'Anonymous', '13. januar 2013 - 17:03')”*

Line 140572:

*“('Diskusjon', '9\_t60', 'Partibarometer for januar, februar og mars 2013', 'Jeg stemmer kanskje på DLF, jeg får se om det blir DLF eller ingenting. Ellers synes jeg Sveits er et bra forbildet. Liberale på omtrent alle aspekter. Renere demokrati(bedre enn Norsk iallefall), liberal økonomi, sikkerhet(fra staten), liberal på sosiale spørsmål som narkotika og våpen. For eksempel, den føderale regjeringen i Sveits må spørre innbyggerne hvert X år om de skal få lov til å hente inn skatt de neste årene. Det aller beste ville selvsagt vært et fritt samfunn.', 'Anonymous', '24. januar 2013 - 23:33')”*

A surprise is that “våpen” (weapon) has an 8 times higher NGTF in Diskusjon.no than in Vigrid. Weapons seem to be more discussed by ordinary people than by Nazis. This can of course be criticism of use of weapons. There is much writing about whether the police should also carry weapons or not.

If we examine the list of words that are in Vigrid, but not in Diskusjon.no, we find different conjugations for the words “tysk” (German), “jøde” (Jew) and “jødisk” (Jewish). Some of the conjugations of these words are in both forums, so we would not have seen these results if we had implemented stemming. The word “holoca” may not be easy to understand, but it is most likely the word “holoca\$h” which is cut because of the dollar sign, which our word filtering code thinks is a punctuation character, and therefore reads the words “holoca” and “a”

instead. Vigrid uses the words “holoca\$h” and “Jew\$A” to tell that the Jews and the USA used holocaust (which Vigrid does not believe really happened) to somehow earn money.

“auschwitz” and “birkenau” are concentration camps in Poland which were used during the war to imprison and execute enemies of the German nazi state. “gasskammer” (gas chamber) is also only mentioned in the Vigrid webpages. One big surprise is maybe that the word “kjærliighet” (love) is mentioned in Vigrid, but not in Diskusjon.no. The reason can be that Vigrid is biased by emotions, while the ordinary forum manages to keep its argumentations on-topic. Lines 14274 and 16443 in the file vigridtvedtnet.txt refer to love to the people and to the Nordic gods, respectively:

*Line 14274:*

*“I åresvis brukte han sin unge energi til å bekjempe nedrige regimer. Flere ganger unnslopp han såvidt døden. Han utviste beslutsomhet, forsakelse, kjærliighet til folket sitt og en dyp avsky for undertrykkende lederskap. Denne mannen har helt klart ingen mangler på gode kvaliteter.”*

*Line 16443:*

*“Har du en skog så plant et nytt solsymbol. Hjukors, hakekors, Thorskors eller livsrune er alle en hyllest til liv og vekst, til **kjærliighet** og lys så har du en skog eller et blomstefelt så plant et tegn til glede for deg selv og oss andre Åsatroere og Odinister.”*

“waffen” which comes from Waffen-SS, “internerte “ (internated), “churchill” and “hat” (hatred) also makes the neo-nazi forum differ from the ordinary forum in this case, but the conclusion here is more that the ordinary forum does not discuss war so much. A little interesting is maybe that Hitler is mentioned in both, but not his first name, Adolf, which only occurs in Vigrid.

“myrde” (to murder) also makes Vigrid distinguishable. “lyve” (to lie) is also such a word, but it is more likely Vigrid which claims other people lie than that they admit that they have lied. The word “forrhådt” (betrayed), “forrhæderne” (the betrayers), and “forrhæderi” (betrayal) also makes Vigrid differ from Diskusjon.no. Fidelity seems therefore to be important for the Neo-Nazis.

When looking for words that are in the political from on the website Diskusjon.no, but not in Vigrid, one can find the names of different Norwegian parties on the top of the list. These parties are Sosialistisk Venstreparti, DLF (Det Liberale Folkeparti), Fremskrittspartiet, Kristelig Folkeparti, (Rød) Valgallianse, Senterpartiet, Demokratene, Kystpartiet, Pensjonistpartiet, NKP, and KrF. The only party that Vigrid also mentions is Arbeiderpartiet. This implies that the political forum discusses the politics of different parties, while Vigrid does not. From what we can see Vigrid mainly write about its own opinions and stands in the political debate.

For some strange reasons, Diskusjon.no mentions “kampfly” (fight planes) and “verneplikt” (military duty), but Vigrid not.

Diskusjon.no has much more text included than Vigrid, and that is most probably the reason for many more unique words. Diskusjon includes expressions associated with politics in Norway: “valgflask” (promises done during an election campaign), “velferdsstaten”, “likestilling”, “politiet” (the police), “helsevesenet” (the health care system), and “mandatfordeling” (mandate distribution).

Expressions from politics in general also make the forum different from Vigrid: “partiprogrammet” (the party’s political program), “klassemotsetninger” (class oppositions), “anarkisme” (anarchism), “totalitært” (totalitary). There are also expressions from everyday political issues: “narkotika” (drugs), “pornografi” (pornography), and international conflicts “syriakrigen” (Syrian war), “nordkorea” (North Korea), “kartellene” (the (drug) cartels). All in all, we see more varied opinions at diskusjon.no about what happens in the world, and we therefore get a more complete and nuanced version of the news than what Vigrid writes about in its little world filled with hatred to different ethnic groups.

## 4.16 Word Frequency Analysis of Deutsche Stimme

The right-extreme National Democratic Party of Germany, NPD, publishes a newspaper called “Deutsche Stimme” (“German Voice”), which also has a website. The articles on this website more often than not have reader comments attached to them. This is almost like a forum with a very long first comment (which really is a news article). The news articles try to be matter-of-factly, but they fail because they are tremendously affected by the authors’ extreme political opinions. Not surprisingly, the most used non-stopword is “npd”, the abbreviation of the party name. It is used 2432 times. The German stopword list is not complete, so one must critically use German language knowledge to find the words that are interesting for the analysis. The most used interesting words are “deutschland” (Germany, used 1405 times), “deutschen”/“deutsche” (Germans, used 2332 times combined), “menschen” (humans or people, 900 times), “land” (738 times), “leben” (“life” or “to live”, 646), “euro” (635), “brd” (635), “partei” (618), and “volk” (“people”, 618). Many of these words can be associated with patriotism and even nationalism, which is confirmed in many cases when one reads the text in the forum. We present more of the characteristic words in Deutsche Stimme when comparing that forum with Turn to Islam later.

The word “Volk” (people) is not nationalistic alone, but it can be used in a nationalistic context. The following forum post sounds at least slightly nationalistic, since it encourages the German people to protect itself against criminal foreigners. Gangs from Eastern Europe seem to be dangerous murderers if we should believe the author “Odin” who praises the weapon law in the USA. He (or maybe she) uses some kind of gallows humour to argument for why this law works: “There are crazy people everywhere. If all Americans were crazy, there would be no one left of them. The Germans (Deutsche) are also mentioned in the text as defenceless between against the so-called criminal, foreign scum. Germans are presented as victims in the following example from line 108158:

*“(‘Deutsche Stimme’, ‘0\_t16’, ‘Comment’, ‘Über diesen Vorfall hört man wie immer nicht viel. Wieviele **Deutsche** wurden von Ausländern im eigenen Land totgeprügelt, totprügeln zählt nicht als Mord. Aber als der NSU auftauchte und angeblich die Ausländer erschossen hat, da war es Mord. In letzter zeit hört man wieder viel über marodierende Mörderbanden aus Osteuropa. Das ist der Dank für ein freizügiges Europa. Ich habe eigentlich mit Waffen nichts am Hut. Manchmal schwirren einem Gedanken im Kopf herum. In Amerika ist das Waffengesetz in Ordnung. Durchgeknallte gibt es überall. Wenn alle Amerikaner durchgeknallt wären dürfte es keine mehr geben. Die Meisten setzen die Waffe als letztes Mittel zur Selbstverteidigung ein. Der ganze kriminelle, ausländische Abschaum weiß die **Deutschen** sind wehrlos. Die polizei ist überfordert. Die Bonzen sitzen in Ihren gepanzerten*

*Limosinen und können gut reden. Deutsches Volk steh endlich auf und wehrt Euch.', 'Odin', '2013-03-13T18:18:12+00:00')*”

The text is, like in many other places on the forum, written in a way that one can easily get sympathies with the author. The writer of the above-mentioned message plays on the fear of violent criminals. That they are supposed to be foreigners seems to be even scarier than if they were Germans. If one does not read the forum critically, it is easy to get carried away. Why should foreign criminals be more dangerous than domestic criminals? A murder seems to be a murder, no matter who committed it. The politicians are referred to as “fat cats” (“Bonzen”) in shielded Limousins, and these people can just talk. By that it is implied that they cannot take any actions. The police are claimed to have too much to do when people cannot protect themselves. This statement seems to be exaggerated in comparison to how ordinary Germans experience their safety in their daily life, and even compared to how violence is reported in German TV news. That just a few people get robbed and killed does not have to mean that everyone must be afraid of these things to happen to oneself.

“menschen” is just a general word like “people” in English forum. It is not nationalistic at all. The Euro is the international currency in most of the countries in the European Union. There are some references to the financial Euro crisis in Deutsche Stimme, but quite often “euro” is simply mentioned when telling how many euros something costs. In the following example from line 108532, an estimate of fines is given:

*“(Deutsche Stimme', '0\_t11', 'Insel: Der Terror geht weiter', 'Von Michael Grunzel Der Behördenterror gegen die Einwohnerinnen und Einwohner der Altmarkgemeinde Insel geht weiter und nimmt immer groteskere Züge an. Inzwischen sollen mehrere Dorfbewohner wegen angeblicher Ordnungswidrigkeiten mehr als zehntausend Euro an Bußgeldern entrichten. Das in der Nähe von Stendal liegende und inzwischen auch zur Stadt gehörende Dorf Insel beherbergt seit fast zwei Jahren zwei gefährliche Sexualstraftäter, welche auf Grund eines Urteils des Europäischen Gerichtshofes für Menschenrechte 2011 aus einer baden-württembergischen Justizvollzugsanstalt entlassen wurden. (...) ', 'admin', '1 Tag zuvor')*”

The NPD members/sympathisers are worried about serial rapists living in an ordinary neighbourhood, so in that sense, extremists are like normal people. There is maybe something about this case we do not understand, but to us it seems odd to complain about fines for minor breaches of the law in the same context. That someone is acquitted of charges for rapists does not seem like an excuse for not paying fines for something quite different. This is a good example of an argument that just is based on the fears and feelings of people.

The articles about the euro in “Deutsche Stimme” are occasionally about the international financial crisis. The following article excerpt from line 108667 gives credits to the Norwegian pension fund for neither trusting the euro nor the dollar in its investments. It boasts somehow of Norway for staying independent and not joining the European Union.

*“(Deutsche Stimme', '0\_t20', 'Euro-Währung – Verantwortungsbewußte verlassen das â??sinkende Schiffâ??', 'von Udo Pastörs M. Großmann / pixelio.de Die Nachricht schlug wie eine Bombe ein. “Norweger trauen Weltwährung nicht mehr!”Â titelt das Handelsblatt vom heutigen Tage (12.03.2013) Worum geht's? Daß der norwegische Staat sein Volk mit in der Welt einmaligen Sozialleistungen versorgt, dürfte weitgehend bekannt sein. Weniger bekannt ist jedoch, daß das nordische Land einen Ölfonds mit Liquidität in Höhe von 512*

*Milliarden Euro unter Staatsaufsicht verwaltet. Begnügte man sich in den vergangenen 1 ½ Jahren damit, französische und britische Staatsanleihen zu verkaufen und diese durch russische, mexikanische und vor allen Dingen durch türkische Bonds zu ersetzen, so machte man nun “Nägel mit Köpfen”. Der sogenannte “Statens pensjonsfond utland” traut weder dem Dollar noch dem Euro. (...) Man kann die Norweger nur beglückwünschen. Gründet doch ihre wirtschaftliche Unabhängigkeit nicht nur auf reiche Öl- und Gasvorkommen, sondern auch auf eine Nichtmitgliedschaft in der EU. Dies bewahrte seine Staatsbürger vor der Einführung des Euro und der Verpflichtung, milliarden schwere Bankenrettungsaktionen mitfinanzieren zu müssen. (...), 'admin', 'vor 4 Tagen')”*

The word “Leben” (life) in is in the following example actually related to the end of life. The text speculates about whether it would be reported in the TV news if five Turks killed a German. The author “Sozialer Heimatfreund” claims that it would not be reported at all. However, if the murdered person were a Turk and the murderers Germans, the author thinks it would certainly appear in the news. This is demonstrated with a fictional news tip below which the author says could be sent to a news station for testing out if the news stations would actually send news about a German killed by foreigners.

Line 109560:

*“(Deutsche Stimme', '0\_t16', 'Comment', 'Ich habe hier noch einen Vorschlag: Folgende Text könnte man verwenden, um per Mail bei verschiedenen Redaktionen, zum Beispiel der Tagesschau-Redaktion, nachzuhaken, warum sie denn nicht ¼ber diesen schrecklichen Vorfall berichtet haben. “Tagesschau” ist in diesem Fall zu ersetzen mit der jeweiligen Redaktion, die man anschreiben möchte: “Sehr geehrte Damen und Herren, mich hat heute eine Nachricht mehr erschüttert als alles andere: Ein 25-jähriger Deutscher, wurde in Kirchweyhe von sechs Türken brutal und hinterhältig zusammengeschlagen. Heute starb er im Krankenhaus, weil sein Gehirn irreparabel war – seine Mutter hatte an seinem Totenbett gewacht (auf der Intensivstation), bis ihr Sohn aus dem **Leben** geschieden ist!” Sicherlich hat die “Tagesschau-Redaktion” diesen Vorfall mitbekommen. Meine Frage: Wieso wird darüber nicht berichtet??? Wenn das Opfer ein Türke gewesen wäre, dass von sechs Deutschen zusammengeschlagen worden wäre, hätte es von Ihrer Seite mit Sicherheit einen Aufschrei gegeben!!! Wie können Sie das mit Ihrem Gewissen vereinen, dass Sie tagtäglich Meinungspropaganda betreiben und nicht sachlich, neutral und objektiv berichten??? (...), 'Sozialer Heimatfreund', '2013-03-14T00:32:22+00:00')”*

The author also claims that journalism in the German press is not really critical and that Germany is not really a democracy, but is ruled by an information dictatorship. The news media are just propaganda machines for the government, the author above claims. Such kind of opinion seems to be normal among extremists, since they feel that their point of view cannot be expressed with much resistance in the public. The NPD is against democracy as we know it, but this party and its sympathizers claim that people shall still be free to talk if they gain power to rule in Germany. They claim to know about how democracy should really work, but as we have experienced from the 1930’s and 1940’s, their predecessors, the Nazis, seemed to even have a worse censorship than the Neo-Nazis claim that there is in Germany today. The NPD has not admitted that anything that the Nazis did was wrong, so they have a problem explaining themselves.

The word “Partei” (party) refers to one or another political party, and in many cases the NPD itself, like in line 108524. In this message the author “nationaler” writes that a party is a

good tool for winning the minds of people, because a party is a legal organisation, and it is practically impossible to forbid, in opposition to associations and fellowships:  
 “(‘Deutsche Stimme’, ‘0\_t25’, ‘Comment’, ‘Hallo! Eine **Partei** ist immer noch das beste Werkzeug im Kampf um die Köpfe unserer Volksgenossen, da eine partei eine legale Organisation ist, welche zu 100% auf dem Boden des GG steht und im Gegensatz zu Vereinen und Kameradschaften praktisch “unverbietbar” ist.’, ‘nationaler’, ‘2013-03-12T14:51:20+00:00’)”

It is not impossible to forbid a party in Germany, but one has to prove that a party is association is serious criminality or a militia to accomplish that. One has tried to declare the NPD illegal before, but without success. Two other political parties were prohibited in the 1950’s, the Sozialistische Reichspartei (SRP, Socialist Reich Party) in 1952 and the Kommunistische Partei Deutschlands (KPD, Communist Party of Germany) in 1956 [51].

## 4.17 Word Frequency Analysis of Turn to Islam (German)

“Turn to Islam” is an Islamist forum with sections in a lot of different languages. In this project we have chosen only to download and analyse the German part. The English forum seems (at least on the frontpage) to be excessively overflowed with questions about how to use the forum. In the German forum there are also some commentaries in English, which we can see from the list of words from that forum section. Some German commentaries have quotes from English texts in external articles and English-writing authors in Stormfront. This forum actually uses many of the same words as the English Islamist forums analysed in this project:

**Table 30: GTFs and examples for Turn to Islam**

Word	Translation	GTF	Example
Allah	Allah	2106	In the following excerpt, it seems like a non-German is writing, because the syntax is not quite correct and there are no punctuation characters. The author “Sabahudin” mentions his/her nationality as Albanian. He/she mentions that two German Muslim friends which he/she often meets, and they have the belief in Allah in common. Line 714: <i>“(‘Turn to Islam’, ‘p12-t10’, ‘Elhamudilah’, ‘Ich bin hiere auch neu hier im forum und habe gedacht ich muss hier auch was schreiben ich habe auch paar freunde deutsche muslime wo ich die oft treffe mir freuts wenn ich so was lese denn nur <b>allah</b> steht zwischen uns muslimsche bruder und schwester, ich rede von gläubigen muslimen bin albaische muslim <b>Allah</b> subhane ve te ala sagt denne die ich das herz vesiegelt habe wird nie was geben die rechtzuleiten bruder due bist eine von dene wo <b>Allah</b> dich</i>

			<i>rechtleitet hat Gruss Sabahudin', 'Sabahudin', '06-30-2007, 07:32 PM')"</i>
Islam	Islam	1309	<p>The author of the following excerpt writes that he/she feared that this forum was heavily influenced by radicals. He/she realized that those people had nothing to do with Islam, and that Allah would punish them. In the German section of Turn to Islam, "TekbirAllah" realized that the most entries had traces of the love of Islam. Muslims should not feel hatred, because it gives nothing but headache. This extract gives a clue that the German part of Turn to Islam is may not only be dominated by extremists, but also by ordinary Muslims. We have also not found any traces of hatred in this forum.</p> <p>Line 739:  <i>"('Turn to Islam', 'p12-t13', 'Untitled', '(...) Normalerweise halte ich nichts von solchen seite da ich immer wieder die befürchtung habe das ignredwelche radikalen sich hier rumtreiben, aber mir wurde klar das diese personen keine moslems sind , die haben nichts mit dem <b>islam</b> zutun.Allah wird denen schon die gerecht strfae erteilen.Als ich viele beiträe las die von deutschland und aus aller welt dachte ich masallah allahuelber bei den meisten beiträgen kann man die liebe des <b>islam</b> richtig spüren. Wir sind moslems , wir sollten kein hass empfinden da dies am ende sowieso nichts brignt ausser kopfschmerzen. Aleyküü selam meine brüder sowie schwestern.', 'TekbirAllah', '06-21-2007, 08:59 AM')"</i></p>
Allahs	Allah's	930	Genitive of "Allah". It is used to signal that something belongs to Allah.
Qur('an)	Qur('an)	566	"Qur'an" is a way of writing "Quran" to better show the pronunciation of the word.
Menschen	humans/people	829	It is likely to believe that a man has written this message, since we find "ibn" (son) in the nickname.
Propheten	prophet (object)	531	The prophet Muhammed is mentioned many times in the text. "Propheten" is the singular object form of "Prophet".
ibn	ibn = son	523	"ibn" is very common as a part of a last name like in the name of the author in the example for
Gott	God	517	Line 1624: <i>"('Turn to Islam', 'p8-t19', 'Untitled', 'selam</i>

		<p>schwester. also erstmal tut es mir leid zu lesen das es dir im moment so schlecht geht. inshallah wird es wieder besser werden. verliere nie die hoffnung und das vertrauen in ALLAH und bete zu ihm auch außerhalb der gebetszeiten. das gibt dir innere ruhe und etwas mehr gelassenheit. und natürlich kannst du jederzeit ein du´a auf deutsch beten. deinem <b>mann</b> solltest du geduld entgegenbringen. glaub mir, ich weiß wie du dich fühlst. kannst du mit ihm auch nicht mehr in ruhe reden? ich meine das ihr euch abends mal hinsetzt und darüber redet? ALLAH hat uns den islam angeboten und das wir ihn annehmen durften ist eine ehre. wir sind in einem nichtislamischen land groß geworden, das den glauben an <b>gott</b> mehr oder weniger verdrängt oder verloren hat. du hast den islam angenommen weil es der richtige weg und gut für dich und deine familie ist. frag ihn warum es ihn so stört das du muslima bist. nur weil er und seine familie nicht beten? wegen dem arabisch: der koran ist eben das originale wort gottes und auf arabisch herabgesandt worden. deswegen sollte man arabisch beten. mal davon abgesehen das die korantexte im original viel schöner sind als in deutsch. wer einmal die FATIHA auf deutsch gelesen/gebetet hat weiß was ich meine sicher, alle sprachen kommen von gott und natürlich versteht er alle. deswegen sagte ich ja das man die du´a in seiner muttersprache machen kann. aber für das rituelle gebet sollte man arabisch beten. das hat mit terrorismus und fanatismus oder dergleichen überhaupt nichts zu tun und ich finde es sehr schade das dein <b>mann</b> zu solchen argumenten greift. wahrscheinlich sind das die letzten die ihm einfallen, deswegen greift er jetzt in die klischeekiste mal davon abgesehen: sprich mal die gebete die du kannst ganz langsam und betont. das ist wie meditation und wirklich balsam für die seele. schwester, ich wünsche dir viel kraft für deinen weg, bleib standhaft und geduldig. inshallah wird sich alles wieder bessern. verlier nie dein vertrauen und deine hoffnung in ALLAH, er weiß was bei dir los ist. und was noch sehr wichtig ist : laß dich</p>
--	--	---



			<i>nicht vom islam abbringen. es ist der richtige weg auch wenn viele das nicht sehen wollen. erinnere dich daran was unser prophet und manch andere propheten durchmachen mußten. teilweise haben die ihre familien verloren wegen des glaubens. also bleib stark ich wünsche dir alles gute, schwester. falls ich was falsches geschrieben habe dann korrigiert mich bitte. und ALLAH weiß es am besten wa salam', 'Izz-Al-Din', '06-20-2007, 12:01 PM')"</i>
Prophet	prophet (subject)	464	Subject form of "Prophet".
leben/Leben	live/life	456	<p>This excerpt is written by a German man who has converted to Islam and at the moment lives in Belgium. He hopes to some time live in Germany again, and maybe also for a period in a Muslim country.</p> <p>Line 614:  <i>"('Turn to Islam', 'p1-t16', 'Untitled', '[QUOTE=Fluchs,537145]Salam, Ich bin seit 7 Monate konvertiert habe mich aber das erste mal vor 6-7 Jahre mit dem Islam befasst. Ich lebe derzeit in Belgien aber ich bin immer unterwegs demnächst werde ich wieder in Deutschland <b>leben</b>. Vorher habe ich in der Schweiz, Spanien gelebt. Ich danke ALLAH das er mir den richtigen Weg gezeigt hat ! ALLAH OU AKBAR Hast du dich noch in Deutschland konvertiert oder erst in dem Nahen Osten? Hier denken alle ich bin verrückt, kein Alkohol, kein Schwein, usw. kein Weihnachten, usw. ich hoffe das ich bald auch in einem Muslimischen Land <b>leben</b> kann oder eine muslimische Frau heiraten kann. Damit ich nicht immer zwischen zwei Kulturen bin. salam ja ich verstehe das istschwierig,,im westen zu <b>leben</b>..und du bist nur seit monaten konvertiert ich bin in england auf die univeritat konvertiert aber nach einenem jahr heriratete ih und bin nach saudi arabien gekommen das macht es leichter fur mich,,,,', 'esperanza', '12-04-2011, 12:01 PM')"</i></p>
Muslim	Muslim	427	A person who confesses to Islam.
Frau	woman	414	The following exerpt tells that a woman is free when she wears a hijab (Hidschaab), a kerchief on her head. Then men on the street will understand that she is married and they

		<p>will respect her and let her walk in peace. A woman shall protect her beauty, because she is too valuable to let all men see her. The only man who is permitted to see the hair is the husband. This will then lead to a reinforced love for the married couple, and the husband will take better care of the woman.</p> <p>From a western point of view, this commentary is a little difficult to understand, because ordinary men normally do not crave wildly for getting hold of an unknown woman. It does not matter whether she wears a kerchief or not. A woman is respected anyway. If a man wants to check if a woman is married, he can just see if she wears a wedding ring. The author does not seem to be quite informed, because nothing is mentioned in the Quran that women must wear a veil on the head. The tradition of women covering the head is older than Islam and exists also in the Christian and Jewish world [52].</p> <p>Line 990:  <i>“(Turn to Islam’, ‘p16-t3’, ‘Warum soll ich den Hidschaab tragen?’, ‘Bismillaahir-Rahmaanir-Rahiim Dies ist eine gute Frage und darauf gibt es eine wunderbare Antwort. Allah hat uns jede gute Tat befohlen und uns jede, die schlecht für uns ist, untersagt. Allah ordnet den muslimischen Frauen an, den hidschaab zu tragen, wenn sie die Sicherheit ihrer Häuser verlassen oder sich in der Anwesenheit fremder Männer befinden. Also ist für dich - eine muslimische <b>Frau</b> - das Tragen des hidschaab eine Quelle guter Dinge, und das aus verschiedenen Gründen. Darunter sind: Du gefällst Allah. Du folgst den Befehlen deines Herrn, wenn du den hidschaab trägst und kannst dafür große Belohnungen erwarten. Er ist Allahs Schutz für deine natürliche Schönheit. Du bist zu wertvoll, um dich für jeden <b>Mann</b> zur Schau zu stellen. (...) Allah möchte dich von anderen respektiert wissen, und das du dich selbst annimmst. Durch den hidschaab erhöht Allah deine Stellung. Wenn ein fremder <b>Mann</b> dich ansieht, wird</i></p>
--	--	--

			<i>er dich respektieren, da er sieht, dass du dich selbst respektierst. (...) Allah möchte, dass deine Schönheit nur für einen einzigen <b>Mann</b> bewahrt und aufgehoben wird - deinen Ehemann. Allah hilft dir durch den <b>hidschaab</b>, eine erfolgreiche Ehe zu genießen. Denn du sparst deine Schönheit für einen einzigen <b>Mann</b> auf, die Liebe deines Ehemannes wird zunehmen, er wird dich liebevoller umsorgen, dich mehr achten, mehr schätzen. (...)', 'ibn azem', '02-08-2007, 11:42 PM')"</i>
Mann	man	366	See example for "Frau". "Mann" can mean both "man" and "husband".

## 4.18 NGTF Comparison of Deutsche Stimme and Turn to Islam (German)

In the following table, we find words that are used more in Turn to Islam than in Deutsche Stimme.

**Table 31: 25 non-stop words with highest f2/f1 NGTF ratios. Turn to Islam is f2, and Deutsche Stimme is f1.**

Word	Translation	GTF ratio (f2/f1)	NGTF ratio (f2/f1)	NGTF-IDF ratio (f2/f1)
allah	Allah	175.500	202.667	20.000
jesus	Jesus	16.077	18.566	9.959
koran	Quran	11.080	12.795	4.525
erde	earth	10.400	12.010	5.635
himmel	Heaven/sky	10.083	11.644	6.247
gottes	God's	9.036	10.434	5.040
wahrlich	truly	8.923	10.304	5.192
paradies	Paradise	8.786	10.146	5.176
gott	God	8.763	10.119	4.688
muslimen	Muslims	8.500	9.816	4.892
botschaft	Message	7.545	8.713	4.666
kopftuch	head cloth	6.611	7.634	4.096
islam	Islam	6.578	7.596	1.360
muslime	Muslims	6.567	7.583	3.271
gemeinschaft	community	6.000	6.929	4.006
moschee	Mosque	5.864	6.771	3.403
muslimischen	Muslim (adj)	5.400	6.236	3.299
erlaubt	allowed	5.333	6.159	3.180

islamische	Islamic	4.913	5.674	2.629
taten	deeds/did	4.667	5.389	2.868
alkohol	alcohol	4.636	5.354	3.221
islamischen	Islamic	4.405	5.087	2.638
heilige	holy	4.400	5.081	3.005
religion	religion	4.388	5.067	2.599
offenbart	reveals	4.273	4.934	2.964
herz	heart	4.235	4.891	2.762

Mentioning of “religion” by the German Islamists is 2.6 times than by the German Neo-Nazis in addition to all the other religious words mentioned in Table 31. It looks therefore like there is more discussion about religion amongst the Islamists.

The right-extreme newspaper “Deutsche Stimme” seems to have more texts about politics than the Islamist forum.

**Table 32: 30 words with highest f1/f2 NGTF ratios. Deutsche Stimme is f1, and Turn to Islam is f2.**

Word	Translation	GTF ratio (f1/f2)	NGTF ratio (f1/f2)	NGTF-IDF ratio (f1/f2)
deutschen	Germans	30.667	26.556	13.856
parteien	parties	30.200	26.152	15.708
wählen	elect	28.727	24.876	14.942
groß	big/large/great	27.000	23.381	12.897
staat	state	26.563	23.002	13.425
demokratie	democracy	25.900	22.428	12.666
politik	politics	23.571	20.412	11.720
deutsche	German(s)	22.275	19.289	11.352
opfer	victim/sacrifice	20.417	17.680	12.626
politische	political	17.909	15.508	10.622
politischen	political	17.727	15.351	10.313
millionen	millions	16.500	14.288	10.211
land	country	16.400	14.202	8.642
gewählt	elected	16.000	13.855	9.280
deutschland	Germany	14.947	12.943	7.075
medien	media	13.636	11.808	8.219
durchaus	absolutely	13.273	11.494	8.351
system	system	12.862	11.138	7.527
arbeiten	work	12.800	11.084	8.628
ten	?	12.700	10.998	7.608
bevölkerung	population	11.929	10.330	7.134
feind	enemy	11.929	10.330	9.120
schrieb	wrote	11.571	10.020	8.579
zahlen	pay	11.500	9.958	7.261
mittlerweile	by now	11.500	9.958	7.287

übrigens	by the way	11.400	9.872	7.732
kosten	cost	11.091	9.604	7.553
maß	measure/extent	10.769	9.326	7.502
länder	countries	10.615	9.192	6.998
rechts	right	10.071	8.721	7.016

The NPD, the political party behind the newspaper Deutsche Stimme, is critical to democracy and elections, although their full name is Nationaldemokratische Partei Deutschlands, National democratic party of Germany. Deutsche Stimme discusses mass media (papers, websites, TV channels) and is quite critical to them. NPD thinks that all other political groups and news media only lies. It mentions all the political opponents as enemies.

The word “bevölkerung” with has an NGTF that is 10 times higher than in the Islamic forum. We think it is because the NPD is worried about the German population. The party thinks that foreigners will take over the country by becoming the majority.

Words that are used in Deutsche Stimme, but not in Turn to Islam, seem to have much to do with politics in Germany today. “npd” is “Nationaldemokratische Partei Deutschlands”, the National Democrat Party of Germany, which is a right-extreme nationalist party, “brd” is a German abbreviation for “Bundesrepublik Deutschland”, Federal Republic of Germany, “cdu” is “Christlich-Demokratische Union”, the Christian Democratic Union, a conservative, Christian party, “spd” is “Sozialdemokratische Partei Deutschlands”, the Social Democratic Party of Germany, “merkel” or Angela Merkel is the German federal chancellor, “ddr” is “Deutsche Demokratische Republik”, the German Democratic Republic, which was a communist dictatorship and today is the eastern part of the German federal republic. We also see words like “politiker” (politician), “regierung” (government), “bundestag” (parliament), “politisch” (political), “demokratisch” (democratic) and “rechtsstaat” (constitutional state). These words all refer to descriptions of how politics and government work in Germany today, and they are all written with explanations of why the NPD does not like that.

Here is a good example of how a commenter of a news article in Deutsche Stimme thinks democracy does not work in Germany today. NPD members and sympathizers even claim they know how democracy really should work. The party claims that Germany is governed indirectly from Washington D.C. like the DDR was governed from Moscow. The author then concludes that the DDR and the BRD were never democracies, and “Heinrich Müller” claims that the NPD is threatened by prohibition only because they tell that harsh truth (as they see it):

*“(‘Deutsche Stimme’, ‘0\_t10’, ‘Comment’, ‘Solche Zustände sind in einer (Gesinnungs-) Diktatur doch ganz normal. Die sog. “BRD” ist keine **Demokratie!** Gemäß Art. 20 (1) ist die “BRD” ein **demokratischer** (Bundes-) Staat, und auch die “DDR” bezeichnete sich selbst als **demokratisch**. (...) Über Wohl und Wehe, Sein oder Nichtsein des deutschen Volkes wurde im Falle der “DDR” in Moskau und wird im Falle der “BRD” in Washington entschieden. Daraus ergibt sich zwangsläufig, dass sowohl die “DDR” als auch die “BRD” zu keinem Zeitpunkt ihrer Existenz **Demokratien** waren. Eine Partei, die dies nicht nur vorgibt, sondern in Deutschland -wirklich- **demokratische** Verhältnisse schaffen will, muss daher verboten werden. Auf eine Lüge mehr oder weniger kommt es ja schon lange nicht mehr an.’, ‘Heinrich Müller’, ‘2013-03-17T11:27:06+00:00’)”*

While the words in Deutsche Stimme are related to politics, the words in Turn to Islam (the German part) are heavily influenced by religion, and particularly by Islam. The word “allah” is actually used in both forums. “allahs”, the genitive of same same word, is the most popular word in Turn to Islam that is not in Deutsche Stimme. “qur” (really qur’an), “quran”, “prophet”, “ramadan” (month of fasting in Islam), “Imam” (priest in Islam), “sunni” (a branch of Islam), “anbetung” (worship), “lehre” (teaching), “mohammed” (the most important prophet in Islam). Many of the words that make Turn to Islam unique from Deutsche Stimme are of the religious kind, mainly from Islam. There are of course exceptions like “Christus” (Christ) “Auferstehung” (resurrection) which is connected to Jesus in Christianity.

## 4.19 Author Vocabulary Analysis

We have examined which words are dominant in the vocabulary in five of the ten authors that have written most text in each of the forums Stormfront and Ummah. The words that we in the analysis mention as unique for an author are used by that given author and not the other nine authors in the top ten list of most productive authors. The so-called unique words may therefore be used by some of the less productive authors.

### 4.19.1 Ummah

The author “Anna” seems to be writing much about the religion Islam and English language characteristics. From the nickname, it looks like this author is a woman. She seems to write mainly in English and it looks like she writes sometimes in Arabic also. The author seems to have an interest in writing about ordinary life events, such as how to feed children, and how to maintain physical and mental health, but also some Islamic rules and rituals. For instance, boys and children are not allowed to wear necklaces. Therefore, how to put the young children on the right path without straightaway banning them from things that are forbidden according to Islam, and tips for how to do that are discussed. Her writings also look like helpful guidance for pregnant women, and for those who have children. There are in addition other general topics like love and marriage. There is some discussion about violence such as smacking children. The author seems to be against or have displeasure over smacking of children and female genital mutilation.

Some of the most used words can be seen in Table 33, and they are all from the file Anna.txt.

**Table 33: GTFs and examples for "Anna"**

Word	GTF	Example/Explanation
allah	508	“Allah” is of course used many times by an author in an Islamist forum. Therefore no example is needed.
arabic	173	“Arabic” is the language spoken on the Arabic peninsula and in Northern Africa, which comprise a large part of the Muslim world.
people	141	From line 289: <i>“re: muslim scholars speak out against female genital mutilation in rare conference bro irhab... many ppl are doing proper mutilating things in the name of female circumcision. because of that, and because it is not something compulsory for women, then it is better if we support it when ppl speak out against it rather than talking that much in favour of circumcision. the majority of <b>people</b> doing this are not actually doing the actual female circumcision of removing that piece of skin or whatever bt they are like sewing up certain parts, and cutting off complete parts n all sorts of crazy n sickening things, astaghfirullah. bc of the harmful effects of that, we shud speak in favour of efforts 2 stop these and educate the ppl against it inshaallah rather than saying each time "female circumcision is from islam" because when they mention it, they do not hav in mind the same thing which u have in mind... and it will jst reinstate their views that their mutilation practises hav some basis in islam, which they do not”.</i> From here we can understand how the author tries to convey the

		message to the society that female genital mutilation is bad and should not be continued anymore, and she specifically says that it is not related to Islam. The text states that different people may have different opinions of what female genital mutilation is.
word	135	The author is very interested in writing about English language grammar and syntax. For instance in line 409: <i>“if the verb was “choose” then the passive participle is “chosen” verbal noun - this is a naming <b>word</b> that had its origin in a verb”.</i>
time	99	Line 406: <i>“write a small paragraph about yourself and your free <b>time</b> (8 marks)”.</i> Here “time” as a duration/period is used in an English language learning exercise.
prophet	83	Line 83: <i>“in the context of the ayah, allah is telling the <b>prophet</b> and the believers that you will discover that most of the jews will betray you. despite all of this, allah is asking the <b>prophet</b> to practice forgiveness of the best kind”</i>
quran	73	Line 67: <i>“the people of the <b>quran</b> are the people of allaah, the ones who are closest to him”</i> The Quran is considered as the holy book for the Muslims, and it consists of religious texts of Islam.
lesson	68	The author gives some lessons in English grammer, as shown in the example for “tense”. The word “lesson” is therefore mentioned several times of natural reasons.
english	67	The author writes about English grammer, as shown in the example for “tense”. The word “English” is therefore mentioned several times of natural reasons.
fear	66	The author tries to explain the word “fear” many times. The author examines that word deeply according to the Quran, and the “fear of God” seems to be something slightly different from its general meaning found in an ordinary English dictionary, which he/she has already checked. For instance, in line 87: <i>“the believers are only those who, when allah is mentioned, feel a fear (wajl) in their hearts.? (8:2) rahb ???rahb is that kind of a <b>fear</b> that makes you very alert, very careful and does not let you relax. this is the type of <b>fear</b> that students feel during exam time: restless, careful, and concern for passing. it is also <b>fear</b> that you are afraid that you will disappoint the one you love. allahazzawajal commands us to have this <b>fear</b> of him: (...) verily, he is (the) only one deity. then, <b>fear</b> me much. (16:51) the believers call upon allah while fearing (rahbah) him, and hoping (raghbah) in him”</i> This message gives an analogy of how a Muslim should fear Allah, namely like one fears an exam, so it is a kind of fear of failure.
muslim	62	The author welcomes each forum members who has become Muslim to participate in a new private forum area. Example from line 4: <i>“new section for support for new muslims salams everyone we have created a new private area for new muslims, to help them with any questions they might have whether emotional, practical issues, general</i>



		<i>support or whatever. if you are somebody who is thinking about becoming <b>muslim</b>, or has become <b>muslim</b> whether recently or even not so recently and you would like access please make a help desk ticket, or else pm me and i will give you access to that new area insha allah. (...)"</i>
surah	57	<p>A surah is a chapter of the Quran. There are totally 114 of them and they vary in length from several words to several pages [53]. Each surah is divided into ayahs, which can be called verses in English [53]. The following message sample discusses ayahs that should be recited by the fellow forum members. In this religious sence, recitation means to formally read verses from a holy text as a prayer [54]. Ayahs from the Quran seem to be mentioned in this message, but due to an error with encoding, we only see question marks.</p> <p>Line 22:  <i>"re: uqp previous surahs revision revision of the red highlighted arabic words: surat al waqiah. these are the words which were highlighted daily by abu muslim when he gave our daily ayahs, i will collect them here for revision purposes. we can check down the list and see how many we have remembered insha allah. if you want, you can print the list out and learn the words. you can look at each word and recite an ayah from the <b>surah</b> which contains the word. you can use it any way you want [you don't need to post and say how many you remembered, its for personal use] *note: some words already covered in previous revision lists in the thread, i won't repeat inshaallah just to make it easier ?????? (...)"</i></p>
tense	56	<p>In line 57: <i>"the respectct and care that the prophet's of allahswt pay towards allahswt. ibraheem as does not say "when allah makes me ill" rather he says "when i fall ill", therefore not directly attributing his illness to allahswt. though he knows well that both health and illness is from allahswt. however because we all look at illness as a test, a trial, something difficult, so from the amount of respect he has for allahswt, he does not say when allah makes me sick, but rather he acknowledges his illness as his fault. then he continues, to say that allahswt is the one who restores his health. givingallahswt the prestige, and position of goodness, azzawajal.????????? this is in a <b>tense</b> that demonstrate that it will happen more than once "</i></p> <p>The sample above teaches about the simple present tense, a verb form used for describing actions done more than once.</p> <p>In line 374:  <i>"present <b>tense</b> verbs look back to lesson nine to see how we made verbs in the past. in this lesson we are going to talk about present <b>tense</b> verbs"</i></p> <p>Here she teaches about historic present tense. The author seems to show some enthusiasm in giving English lessons and similar concepts in Arabic to help others/new ones to understand and learn about the Quran.</p>
language	56	The author writes about English language, as shown in the example for "tense". The word "language" is therefore mentioned several times of natural reasons.
life	44	From line 231: <i>"u feel happy to do it cos u already love them even like</i>

		<p><i>with delivering the child that is probs 4 most people the most or one of the most painful things ull experience in <b>life</b>, but ur focus isnt really so much just all on the pain but same time about ur baby and really it is soooo worth it u wd just feel like going thru it all again str8 away because of hw much of a high u get wen u see ur baby (well some people dont they never wana do it agian, but some do feel like that ) seriously its like the best days of ur<b>life</b> wen u give birth even tho its most painful time of ur life too i feel all emo n mushy thinkinabt it”</i></p> <p>The author seems very passionate and optimistic about having babies even how painful it is.</p>
muslims	37	<p>In line 165:  <i>“<b>muslims</b> should be very careful about their motives for doing stuff like this. there is a verse in the qur'an about the one whose face turns dark when he is told he has a baby girl..... <b>muslims</b> should be happy with whatever allah decrees. wanting a boy cause you already have several girls and no boys is one thing, but wanting a boy because you think boys are better than girls is not right yeah because wanting only specifically one gender and then if you don't get it, may lead to ingratitude? so surely itll be better that if u do prefer to have one or the other, you dont keep telling yourself that soo much that u feel u onlyyyy want a boy/girl and will be disappointed with the opposite. also it wouldnt be nice for that child if this was showed in any way they wil feel unwanted”.</i></p> <p>From this long extract, we can understand the good attitude that the author has towards birth of a child. The author mentions the fact that whether a newly born infant is a boy or a girl, it should be considered as a valuable treasure, since not every woman is blessed with a child. The text says that loving a child, whether it is a boy or a girl, is important for all Muslims.</p>

The author “AbuMubarak” seems to write a lot about the religion Islam and society. The author seems to have a very negative and bad attitude to western countries. The author is highly criticizing western countries, their culture and their influence on Muslim countries. From many places in his/her writing, we can see signs of that. Politics also seems to be discussed. It looks like family values are praised by this author. For instance, for a married person, how he/she should behave towards his/her in-laws is considered to be very important as it is a key element for peace amongst the family members and also for the whole society.

Examples for some of the most frequent words are in Table 34, and they are from the file AbuMubarak.txt.

**Table 34: GTFs and examples for "AbuMubarak"**

Word	GTF	Example/Explanation
allah	989	<p>“allah” is the Arabic word for God who is worshipped by Muslims in Islam. Therefore it is likely to be written many times by an author in an Islamic forum like Ummah. For instance, in line 255 we read:  <i>“to the degree that you remember <b>allah</b>, your heart will be calm and cheerful (...) we should not be surprised when we hear that people who remember <b>allah</b> are at peace”.</i></p>

		From this extract we can understand how the author wants to tell how divine Allah is. We also find the word “allaah” used 534 times.
people	753	“people” is used as the general mankind. For instance, in line 255, “we should not be surprised when we hear that <b>people</b> who remember allah are at peace” is referred to the followers/devotees of Islam religion.
muslim	626	From line 136: <i>“west is shedding crocodile tears coming to <b>muslim</b> lands and staging themselves as the protectors of the <b>muslim</b> women, as they are the messiah for the enslaved <b>muslim</b> woman. first thing they want to do is liberate <b>muslim</b> woman from her religion i.e. deen!, it is their first priority and for their vested interests they are ready to do, spend anything”</i> From these examples we can clearly see how the author is very resistant to western countries.
prophet	574	From line 255: <i>“such is the belief of muslims, the followers of the <b>prophet</b> muhammad. and nothing happens in the universe except through allah?s knowledge, permission, and divine plan”</i> In Islam, prophet is a person who has been assigned by Allah to accomplish a mission of spreading his teachings worldwide.
islam	535	From line 136: <i>“the status of women in <b>islam</b> is often the target of attacks in the secular media. the ?hijab? or the islamic dress is cited by many as an example of the ?subjugation? of women under islamic law. it really seems immature and childish when the western media and govt. ignore big issues like hunger, cold, dislocation, murders, genocide and rather concentrate on portraying a negative picture of islamic principles and code of life especially when it comes to muslim woman. media has played its huge role in maligning islaam and it didn?t cease any opportunity to attack islam, it portrayed muslim women as an enslaved victim of an inhumane religious codes. muslim women are needed to be liberated from such mishap. and the west is playing its main role behind directing such unprecedented propaganda (or lies) againstislamic values and principles”</i> From this extract also, we can understand how the author is opposing and highly criticises western countries. He/she claims that these countries lie about Muslim women being subdued by covering themselves with clothes.The author criticises the same countries for not caring about what really matters: namely hunger, cold, and other humanitarian crises.
women	529	In 116 we can read: <i>“one of your jobs is to educate your wife in matters of sex especially in your likes and dislikes and do not compare her to other <b>women</b>”</i> . From this we can understand that the author wants to express his good attitude towards women and tries to convey his message that to get pleased yourself by your wife only. Line 137 says <i>“for many <b>women</b>, their new ?freedom? has brought the dismal experiences of exploitation, abandonment by men, abortions, financial hardships, single parenthood and isolation”</i> . From here, we can understand the author’s opinion about so-called freedom for women and its negative side. The author claims that this freedom only has

		brought unhappiness to women.
god	494	In line 100: <i>“islam is based upon reality, not illusion mankind is a creation of <b>god</b>, and mankind succeeds in submitting to its creator and fails when it tries to devise other methodologies that do not include its creator nothing about <b>god</b> is private, not the food you eat, the rain, the sun, your heartbeat only a stone cold kafir would even want to hide god”</i> Here the author is trying to give a definition to God. In Islam, “afir” is an Arabic word which means a person who does not believe in God and who tries to cover the truth.
peace	472	Line 87: <i>“filthy kaffir pig says she isnt sorry for abughuraib originally posted by uncle umar ironic these these white christians talk about <b>peace</b> and love all the time while at the same time being the ones that start all the wars, decieve, manipulate, pillage rob, rape, blackmail and extort people who are weak. and then they complain when the subjects of their violence lash out, like asif they (the christians) are the only ones allowed to kill. over a million iraqis dead and the excuse is "oops" o my god, you said white that makes you racist”.</i> The author discusses the irony that Christians want peace to the whole world, but are not able to achieve that. He/she seems to generalize the evil actions to all (white) Christians and white people. The torture that some soldiers and officers in the American army, but not all of them, did to prisoners in the Abu Ghuraib prison in Iraq, is regarded as gruesome by ordinary people. Although there is some belief that also other American military officers supported the torture, there is no reason to believe that everyone did.
Muhammad	411	The prophet Muhammad is of natural reasons mentioned several times by an author in an Islamist forum.
time	329	“time” is very often used in the sense of the duration of an event. In line 301 we as the example: <i>“this is because lusts and desires may be satisfied at some <b>time</b> yet the person still knows them to be wrong and he may control himself and repent and give up that evil deed”.</i> We can see how the author finds one’s spiritual life very important, and that it must not be based on evils or sins. The author speaks a lot about moral attitudes and good behaviour of the mankind.
bush	194	From line 101: <i>“when <b>bush</b> speaks of freedom as god's gift to humanity, does he mean the first amendment freedom of larryflynt to produce pornography and of salmanrushdie to publish the satanic verses, a book considered blasphemous to the islamic faith? if the islamic world rejects this notion of freedom, why is it our duty to change their thinking? why are they wrong? when the president speaks of freedom, does he mean the first amendment prohibition against our children reading the bible and being taught the ten commandments in school? if the president wishes to fight a moral crusade, he should know the enemy is inside the gates. the great moral and cultural threats to our civilization come not from outside america, but from within. we have met the enemy, and he is us”</i>

		The author does not seem to like former US president George W. Bush. He/she thinks that the freedom of speech from the first amendment of the American constitution is not as great as one could believe. The author seems to hate that everyone can write disrespectful commentaries about religion (blasphemy). The author seems to reveal some background from Christianity, since he/she writes about “prohibition against our children reading the bible”.
American	150	From line 273: “(...)america's defiance of a number of un resolutions, and list the fact that america has wmd's and the willingness (and track record) to use them america's undying support for terrorism (israel) then list <b>american</b> crimes about potential, and actual, danger to the land of muslims show how america has gone into muslim countries and raped, tortured, unlawfully imprisoned thousands of muslims show how america has a long record of oppression and racism show how warped <b>american</b> thinking is with its legalization of homosexuality and fornication maybe show how women are mistreated by using them as sex symbols teen pregnancy, abortions, etc would it be ok for you then?” From here again we can vividly see how the writer is very negative about the USA. The message also shows that the author does like homosexuality which he/she considers as wrong and sinful.
democracy	115	In line 95, “ <b>democracy</b> in the west tends to move away from moral and ethical values. this secularist and materialistic orientation of western democracy is totally rejected by iqbal” The author does not seem to believe in democracy as it is known in the west. He/she thinks that it is immoral.
president	109	Line 223: “ <b>president</b> barack obama's re-election, after nearly four years of failing to fulfill his vow to shutter the controversial facility, has compounded the despondency felt by many detainees. in one of his first acts in office, obama declared that he would close the doors of the george w. bush-era "war on terror" prison for good. but it remains open, housing 166 detainees on the eve of his inauguration next month to a second term” This is another example which shows the displeasure on USA. The author is displeased with the fact that the terrorist prison in the American military camp in Guantanamo Bay in Cuba is not closed yet. This is not necessarily an extremist opinion. Many ordinary people think that it is wrong to keep prisoners without judging them.
America	98	See the example above for the “American”.
western	100	In line 105 we can read the following: “my dad has a secret wife, is my family not good enough ? :( sisters better stop acting like polygyny is this evil thing it is not it is something all of the companions of the prophet and the prophet himself engaged in we have been so indoctrinated by " <b>western</b> " values, we have now learned to hate parts of islam”. This is another instance to show that the author has strong disapproval for the western countries and their cultures different from this author's culture which is enriched by the religion Islam. The author defends having several wives as a part of Islamic culture, where up to four

		wives are allowed. The author thinks that it is correct for a man to have many wives, in opposition in only one as it is in most of the western culture. A further discussion about this topic can be read in the section for the forum Turn to Islam in later in this chapter.
British	49	In line 312: <i>“look on helplessly when muslims are being persecuted in their homeland and then watch them being punished by the <b>british</b> asylum system. stuff your veil into your handbag because you'll never get that job if you cover your head. sacrifice prayer times and fasting to keep up with the crowd and stay in with the boss”</i> The author’s bitter feelings about western countries are again portrayed. This time the British asylum system is criticized for treating refugees coming to the United Kingdom badly. There is also some criticism about Muslims not getting permission or occasion to say prayers in the working time.

The author “Abdul Haq” seems to write about politics, war, terror and crisis situations. It seems like the author is interested in discussing the conflict between Israel and Palestine and their old and current dispute issues. The author seems also to protest against America, Great Britain, and other western countries. Examples of most used words in Table 35 are from the file “Abdul Haq.txt”.

**Table 35: GTFs and examples for "Abdul Haq"**

Word	GTF	Example/Explanation
intelligence	188	From line 2: <i>“in other words, the u.s. <b>intelligence</b> community had sufficient information of an impending al-qaeda hijacking attack, powers argues reviewing the available evidence, but was probably blocked from undertaking preventive action from above”</i> Here the word “intelligence” refers to domestic and foreign intelligence services of the USA.
camp	166	From line 15: <i>“palestinians have accused israeli forces of massacring civilians and committing warcrimes between april 2 and 12 in the jenin <b>camp</b>”</i> Here, Jenin is a city in Palestine in the northern West Bank where the Palestinian Jenin refugee camp is situated.
military	122	From line 2: <i>“fbi and <b>military</b> intelligence officials in washington say they were prevented for political reasons from carrying out full investigations into members of the bin laden family in the u.s. before the terrorist attacks of september 11”</i> Here the intelligence service of the American military is mentioned.
attacks	113	From line 2: <i>“the document also suggests <b>attacks</b> on cuban refugees in the united states ?even to the extent of wounding.? so if this document is supposed to show us what u.s. officials are morally capable of, it seems to suggest that they are capable of lying, deceit, conspiring to wage a war of aggression - but not killing u.s. citizens”</i>

		From this we can understand the author’s view about USA. He/she does not seem to have a positive feeling about that western country. American officials or bureaucrats are asserted to be capable of planning different cruel acts, except murder of American citizens.
international	68	From line 16: <i>“those palestinian groups that have carried out attacks against civilians have also violated the basic <b>international</b> legal principle of the inviolability of civilian life and property. acts of terror that take life violate the right to life set forth in the <b>international</b> covenant on civil and political rights”</i> The word “international” is here used for referring to international agreements about how to treat civilians.
war	66	See the comment for “bush”.
refugee	64	From line 16: <i>“israeli defence force incursion into jenin city and <b>refugee</b> camp, 3-18 april 2002 (...)”</i> A refugee is a person that has fled from war. If such persons cannot find other places to stay, they settle down in camps in areas which are relatively safe. In the example that was unfortunately not the case, when the enemy’s soldiers came to the camp. We see from context that it was Israeli military which attacked Palestinian refugees.
bush	63	In line 2: <i>“there must be something to make the supporters of <b>war</b> feel pleased with themselves because as far as their strategy for making the world a safer place goes, we now have every single middle eastern nation except israel lined up against america and britain. which is a remarkable achievement ? iran, iraq, saudiarabia, the lot. and do they think war against iraq will make the populations of these countries a) less resentful or b) more resentful? and when rumsfeld, <b>bush</b> and blair stood in silence yesterday, did they a) give a moment's thought to any of the 3,125 or more dead afghan civilians, or b) give no thought at all? and if they answered b) to both, do they possibly think that there might be a connection? the current strategy must make it more likely that more terrorist atrocities will follow, and so it will go on. maybe that's the plan. eventually there'll be so many one-minute silences, we'll all be stood around remembering all day every day, and the world will be in peace at last”</i> In this long extract the writer’s view of the USA (through former president Bush) and the UK (through former prime minister Blair) is revealed. He/she thinks America and Great Britain are cold-blooded and commit injustice to other countries, especially the ones in the Middle East. The author uses some gallows humour as a hope that there will be continuous peace when so many persons are killed that one has keep silent for many minutes in respect for the deceased.
terrorist	48	Line 17: <i>“the report highlights some key aspects of the events in jenin and other palestinian cities: <b>terrorist</b> attacks from palestinian cities the report describes the concerns of the government of israel that a number of the cities served as bases for armed palestinian groups engaged in terrorist acts against israel. for example, israel has charged that, from</i>

		<i>october 2000 to april 2002, 28 suicide attacks were planned and launched from the jenin camp”</i> The excerpt tells about the Israeli government’s worries about planning of terrorist attacks by people in the Palestinian refugee camp Jenin. The author seems to quote this statement from an external news source.
political	47	As shown in the example for “Bush”, politics is discussed in the forum. It is therefore natural that the word “political” is used some times.
killed	40	Line 2: <i>“and it is a novel approach to grieving, that war helps you through a trauma. perhaps this will catch on, and grief counsellors will advise: "you might consider easing the pain by bombing a random country. (...)the one that comes in with the minimum death toll of civilians in afghanistan, <b>killed</b> by american bombs, is 3,125. which is more than those killed in the world trade centre. maybe they are aware and enjoy the close race. perhaps the news programmes could have the score in the corner throughout their broadcast, the way they do now with football and cricket”</i> The author sharply criticizes the USA for bombing Afghanistan as a kind of remedy for the grieving after the terrorist attack against the World Trade Center in 2001. He/she compares this killing fight sarcastically with a sports event with a score table, where points are given for number of murders.
hijackers	16	In line 2: “alleged hijackers may have trained at u.s. bases”.

The author “hamza81” seems to write about the religion Islam and mainly about Islamic rituals such as fasting during the month Ramadan, and other Islamic religious acts and their benefits for Muslims. It looks like the author wants to convey others that fasting during Ramadan is very important for Muslims and they should never miss it. The author gives some tips and advices to devotees for a proper fasting and how to make it fruitful for the everyday in Ramadan. The author criticizes masturbation and provides tips for how to overcome it. Moreover, the author gives some tips for how to spend the youth in a proper way. He/she encourages the reader to read through the Quran and learn for a better life. Here follows hamza81’s most used words and examples for how they are used in Table 36 (we refer to lines in the file hamza81.txt):

**Table 36: GTFs and examples for "hamza81"**

Word	GTF	Example/Explanation
allah	724	Allah is the name of God in Islam and is therefore of course mentioned by an author in an Islamist forum.
prayer	284	Prayers are an important part of Islam.
ramadan	264	Ramadan is the ninth month of Islamic calendar when Muslims fast. The author seems to be very interested in writing about Ramadan and other religious rituals performed by Muslims. For instance in line 73: <i>“the month of <b>ramadan</b> is the one in which the quran was sent down, a guidance for mankind, clear proofs for the guidance, the criterion; so whoever amongst you witnesses this month, let him fast it”</i>
time	221	In line 47: <i>“this is the best <b>time</b> to gain god habits. this is the best <b>time</b> to go</i>



		<p><i>closer to allah! there is no <b>time</b> like the present and there may not be anytime left for us to see tomorrow. so now is the <b>time</b> we have to make the changes before its too late! this life is just a test and we are only on a short hourney here to our final and ultimate destination which is the hereafter. hereafter is for eternity and this life is but a few short hours so what should we work for? ”.</i></p> <p>From here, we can see how the author wants to tell that everyone should spend time only for good deeds and not for the evils and sins, since very short time is left and who knows that is the end of the journey of life. The text says that one must spend the time for things that really matters before it is too late.</p>
prophet	208	The prophet (Muhummad) is of course mentioned by a writer in an Islamist forum.
night	202	During Ramadan one can eat only at night.
pray	179	Equally important as the word “prayers”.
worship	140	<p>From line 50:  <i>“surely one of the best way we can make the best of each and every precious second in the blessed month of ramadan is to have a <b>worship</b> checklist that we can refer to”</i></p> <p>Here a worship checklist is something that includes instructions about performing religious acts for perfect worshipping of Allah.</p>
muslim	90	A Muslim is a believer of Islam.
rewards	83	<p>In line 58:  <i>“these <b>rewards</b> should not be missed at any cost! scholars have stressed that these special and blessed days and nights of the first ten days in particular should be spent doing intensive worship and virtuous acts as well as contemplation. allaah swears an oath by them, and swearing an oath by something is indicative of its importance and great benefit”</i></p> <p>Here, the rewards means the spiritual gains or merits by doing good deeds and behaving well as according to Islamic religious teachings taught by Allah.</p>
sins	74	<p>In line 75:  <i>“this is the month of charity in which believer's sustenance is increased. whosoever feeds a fasting person or gives a fasting person a single date or a sip of water, to break the fast, for him there shall be forgiveness of his <b>sins</b> and he will be saved from the fire of hell, and for him shall be the same reward as for him (whom he fed)”</i></p> <p>According to Islam, Muslims believe that when they engage in charitable activities during Ramadan, their sins would be washed away or vanish. The text says that a person a person who feeds a fasting person shall be forgiven for that and all his sins.</p>
fasting	49	<p>Fasting in Islam is where Muslims avoid from food, drink, gambling and all other desires from sunrise to sunset during the month of Ramadan. An example is shown in line 81:  <i>“Ramadan <b>fasting</b> is not just refraining from food and drink! jaabir said, “if you fast, then let your eyesight, your hearing, and your tongue fast from lying and the prohibited things, and leave [talking about] what has happened so that there may be peace and tranquillity upon you on the day of your fast, and do not make the day you fast equal to</i></p>

		<i>the day you do not fast”</i>
forgiveness	44	In line 83: <i>“this is an opportunity to gain a lifetimes of <b>forgiveness</b>, immense rewards for even the smallest of deeds and allah has made it easier for us to do good deeds in this blessed month”</i> Here forgiveness means when Muslims perform fasting, other observances and good deeds like charitable works during the month of Ramadan, then he or she get forgiveness from Allah for his/her sins.
eat	33	In line 17: “that the individual should not <b>eat</b> or drink too much. then sleep will overpower him, and waking up will be difficult”. This is a tip given to perform the night prayer in an easy way for Muslims during in the month of Ramadan.
equal	29	In line 81: <i>“if you fast, then let your eyesight, your hearing, and your tongue fast from lying and the prohibited things, and leave [talking about] what has happened so that there may be peace and tranquillity upon you on the day of your fast, and do not make the day you fast <b>equal</b> to the day you do not fast”</i> “equal” is in this example applied for comparing a normal day and a fasting day. A fasting day should not be equal to an ordinary day.
family	28	In line 47: <i>“(...) don't forget to spend time with your family and close relatives as islam emphasises a lot about maintaining ties of kinship. if there is a dispute then be the first to try and resolve it or get an elder of the community or <b>family</b> involved to mediate matters and get advice from a learned scholar. the disputes should not be allowed to continue becuse of petty worldly matters and should be resolved straight away”</i> This is another useful advice given for young people, but could actually in principle be important to anyone. One should spend time with family, and quarrels in a family should be solved.
married	12	In line 104: <i>“finally the best solution of all- get <b>married</b>. give some urgency to your family about the need for you to get <b>married</b> sooner rather than later. tell them of the benefits of getting <b>married</b> early and that it is a sunnah as the prophet (sallallahualaihiwasallam) recommended it as it will help avoid haraam(forbidden) relations with non mahrams and getting <b>married</b> early will help eradicate this problem and mean that one can indulge in their sexual urges and desires legitamatley as well as recieve reward for it”</i> The writer thinks that it is always good to get married early. Then the couple can enjoy sexual pleasures without feeling guilty for that. There is then also no need for affairs outside the marriages.

The author “Z-Blade” seems to write mainly about the religion Islam and mentions how to pay gratitude to fellow Muslims who have helped you and been kind to you. His/her most applied words are in Table 37, and the examples are taken from the file Z-blade.txt.

**Table 37: GTFs and examples for "Z-blade"**

Word	GTF	Example/Explanation
lol	115	<p>The author seems to frequently quote another author who uses the word “lol” (laugh out loud), and sometimes he/she seems to use it him-/herself.</p> <p>Here is an example from line 131, where one also can find a particular kind of smiley ( :) ):</p> <p><i>“re: 1...2...3...4...allahu akbar! originally posted by small sis jazakallahukhairan!!!! is it possible to ask something worse than i just did!!!! lol!!! i learnt something new - its what matters - and now i can listen to the reicitation!! alhamdullilah!! and jazakallahukhairan 2 u!! jazakallahukhairan for the hadith!! 3836 barakallahu feek! lol, yep it is, the person said something like: "hmm ok bro this going to sound kind of stupid but what are we suppose to do with the link? it just opens to a megaupload page " lol! but mashaallah it's not a stupid question at all :'). barakallahu feek to you! barakallahu feek! (lol ). 3837</i></p>
peace	91	Often used in the context: “Allah’s peace”.
lord	83	Another word for Allah/God.
evil	41	<p>In line 745:</p> <p><i>“allah burdens not a person beyond his scope. he gets reward for that (good) which he has earned, and he is punished for that (evil) which he has earned”</i></p> <p>The example states that Allah does not demand too much from humans, but every person gets rewards for the good and punishment for the evil he/she does.</p>
funny	27	<p>In line 705:</p> <p><i>“haha, you guys are <b>funny</b>, nice joke”</i></p> <p>We see that “Z-Blade” has some kind of humour.</p>
forgive	25	<p>In line 870:</p> <p><i>“remember, allah has said he is merciful to those who show mercy to others. so <b>forgive</b> and forget, and receive great reward inshaallah!”</i></p> <p>The author mentions that one should forgive people who have done wrong deeds and that one should not take revenge. Therefore it seems like the author gives some importance for values.</p>
earth	25	<p>In line 50:</p> <p><i>“those who break allah's covenant after ratifying it, and sever what allah has ordered to be joined and do mischief on <b>earth</b>, it is they who are the losers”</i></p> <p>In Islam, it is said, if the agreement (about doing good deeds, etc.) between the God and the devotees are broken by someone, then he/she would not be protected by God. That goes against Allah’s words and does evils.</p>
ramadan	22	<p>In line 365:</p> <p><i>“whoever observes night prayer in <b>ramadan</b> as an expression of his faith and to seek reward from allah, his previous sins will be blotted out.”</i></p> <p>Muslims believe that if they worship and do observances such as fasting, performing prayers during the month of Ramadan, they are set</p>

		free from sins by Allah.
wealth	21	In line 783: <i>“i thought about all types of <b>wealth</b>, but did not find a better wealth than contentment in little”.</i> From this we can understand the author’s view about wealth. He/she sees wealth as being contented or satisfied with what you already have, which is not something all people are satisfied with.
hearts	20	In line 640: <i>“i have seen wrong actions killing hearts, and their degradation may lead to their becoming addicted to them. turning away from wrong actions gives life to the <b>hearts</b>, and opposing yourself is best for it”</i> This extract which is originally posted by someone else, gives the idea about how peaceful it is for everyone when all people try to get rid of hatred and filth.
excellent	17	In line 104: <i>“thus islaam taught muslims to say assalamualaykum (peace be upon you), and replaced all other formulas of greeting. in this small greeting is hidden an <b>excellent</b> and comprehensive prayer, kindness and affection for the young, and respect and attention for the elders”</i> A very common greeting phrase used among Muslims is mentioned in the quote, and the author considers it as brilliant.
family	17	In line 135, <i>“therefore be joyous only for an increase in knowledge or in good works, for they are your two companions who will accompany you in your grave when your <b>family</b>, wealth, children and friends stay behind”</i> It is clearly seen that the author is more positive towards moral life and good deeds even though this is not originally posted by him.
patience	12	In line 783: <i>“i looked at all types of sustenance but did not find a better sustenance than <b>patience</b>”</i> The writer considers patience as a kind of intellectual nutrition (sustenance) for humans, and by that he/she means that patience is as important for the human mind as food is important to the human body.

## 4.19.2 Stormfront

The most productive author in the part of the Stormfront forum we have downloaded is kazan188. He/she writes both in English and French, dependent on the discussion thread we are looking at. We therefore obviously see some French words in the vocabulary. Among the French words that are not stopwords, we find “guerre” (war, used 352 times), and “allemagne” (Germany, 203).

For the most frequent English and multilingual words, we find the examples shown in Table 38 from the file kazan188.txt.

**Table 38: GTFs and examples for "kazan188"**

Word	GTF	Example/Explanation
German	1153	An Indo-European language or a citizen of the country Germany
Hitler	892	Surname of the Nazi dictator Adolf Hitler. See example in “people” in this table.
jews	816	See “people” in this table.
camp	608	A camp in the text is mostly referring to an internment camp in Germany or Eastern Europe during the Second World War. The following example is from an English quote in a French forum message. Line 488: “(...) apparently most were transported to auschwitz, although some were sent to the lodz ghetto, the theresienstadt ghetto-camp, the bergen-belsen <b>camp</b> , and elsewhere.(...) camps in germany, poland and” occupied russia." source: the new york times, june 15, 1943, p. 8. quoted in: arthur butz, the hoax of the twentieth century (ihr, 1997), p. 82; (..)
Polish	564	“polish” is clearly people from Poland, not some cream to wipe silver blank with: Example from line 1: “(...) the polish resistance had infiltrated the whole camp complex very early and regularly transmitted reports outside. moreover the allied authorities made their own investigations to know if the rumors were true or not. the reports of their investigations have naturally disappeared but we all know taht they finally chose not to bomb auschwitz-birkenau (...)”
Jewish	532	See “people” in this table.
Auschwitz	513	Name of a concentration camp during the Second World War.
Roosevelt	399	Surname of two American presidents with first names: Theodore and Franklin
people	398	“people” can refer to practically all kinds of humans. Here is an example where the Jewish <b>people</b> is mentioned in line 506: “(...) new arguments which should prove the jews to be a peculiar <b>people</b> , never amalgamating with their fellow-citizens ; on the other side science was bringing to light infallible marks by which the physical peculiarity of the jewish race should be made clearly manifest." david kaufmann (1878) "hitler was right in one thing. he calls the jewish <b>people</b> a race and we are a race." rabbi stephen wise (1938) (...)”
British	368	People from Great Britain or an adjective that something is from there. One of the winners of the Second World War.

France	333	Name of a European country. Word can be both in English and French. One of the winners of the Second World War after partly being occupied by Germany.
documents	332	Following message excerpt mentions documents that have disappeared. The message is French with quotes in English. It mentions the cruelties the Sovjet Union army did with Polish prisoners of war in the Katyn forest in 1940. Line 509: <i>“re: les usa connaissaient l'implication de l'urss dans le massacre de katyn l'internaute dont parle le professeur faurisson, soucieux de préserver la crédibilité de la parodie de justice de nuremberg, (...) shoot every member of the orchestra grind the bones of millions of people in portable bone-grinding machines grind the bones of 200 bodies [¾ ton] at one time as described in photographs and <b>documents</b> which have disappeared (...)”</i>
government	326	Governments from many countries from both during and after the Second World War, are mentioned by kazan188. Among those are governments in Germany, Belgium, the UK and the USA. From line 469: <i>“(...) and before september 1939 churchill wasn't powerless either. while it is true that he had little following in the house of commons during much of the 1930s, he was given considerable privileges by the <b>government</b>. (...) churchill continued to be consulted on many matters by the <b>government</b> or seen as an alternative leader(...)”</i>

In general, by just looking at the most frequent words presented in the table above, it is clear that kazan188 discusses topics from the Second World War.

Mjodr only writes in English. Examples of frequent words in Table 39 are from the file Mjodr.txt.

**Table 39: GTFs and examples for "Mjodr"**

Word	GTF	Explanation/Example
white	395	Used for referring to white people, especially those that think white people are superior compared to other people. Example from line 713 <i>“(...) there are 13 active hate groups in san diego county, four of which are <b>white</b> supremacist groups, according to a new report from the southern poverty law center. (...). as reported in san diego citybeat, american third position, a <b>white</b>-nationalist political group, has spread to san diego county. (...) the august shooting in a sikh temple in wisconsin was branded a hate crime because the shooter, wade michael page, was revealed to have ties to <b>white</b> supremacist groups and to be a leader in a skinhead band called end apathy. in july, a concert for <b>white</b> supremacist bands was held in el cajon. (...)”</i>
police	266	In the example from line 703, the author quotes the news tv channel Al-Jazeera, which talks about how the police in the USA have prioritized extremist monitoring the last decade: <i>“oh, the irony - aljazeera - white supremacy, the new terrorist threat. (...) during the war on terror, white supremacists were given less scrutiny by both local and federal <b>police</b>,</i>

		<i>while their victims - muslims and non-muslims - endured state-sponsored profiling and vilification. (...)</i>
national	265	It may seem obvious for a nationalist to use the word “national”, but “national” in the USA simply is another for federal, which means that something is associated with all of the USA, not only some single states. The word “national” is therefore actually not nationalistic in itself. Here is an example from line 688: “ <b>montana</b> governor <b>schweitzer</b> speaks at dnc. ass. i agree with him about romney (and disagree with his support for <b>obama</b> ), but he is still an ass. <b>schweitzer</b> , no one likes you. you have no friends. quote: charlotte, n.c. - montana governor brian <b>schweitzer</b> took aim at mitt romney in his speech thursday night at the democratic <b>national</b> convention. (...) <b>schweitzer's</b> appearance thursday marked the second time he's had a chance to speak at the <b>national</b> democratic convention. (...)”
Obama	253	The author does not seem to like Barack Obama. See example for “national”.
Montana	241	See example for “national”.
chick	225	“Mjodr” seems to also post at least this innocent message without anything politically related in line 526: “animal videos - critters in motion a bronx zoo caribbean flamingo <b>chick</b> takes its first steps - youtube a bronx zoo caribbean flamingo <b>chick</b> takes its first steps - youtube
park	209	The word “park” refers to a national park or a green area in a city.
youtube	207	Word used when referring to a video on the website YouTube. See example for “chick”
news	189	News is simply information that is new, and preferably it should be matter-of-factly. Line 164: “re: sub-forum for credible <b>news</b> /information sources i don't disagree. but i think it should be all new sources, not just making a new forum.”
family	167	Family seems to be important for “Mjodr”. (...) "we are very much supportive of the <b>family</b> —the biblical definition of the <b>family</b> unit," mr. cathy explained recently in an interview with the baptist press. "we are a <b>family</b> -owned business, a <b>family</b> -led business, and we are married to our first wives. we give god thanks for that." in virtually every culture, <b>marriage</b> is an institution involving a public commitment between a man and a woman. the complementary nature of men and women points to the unique purpose of <b>marriage</b> : to bear and raise children. one can recognize this fact and so conclude that "same-sex <b>marriage</b> " is an oxymoron—without being "anti-gay." (...) mr. richards, a roman catholic, and mr. robison, an evangelical christian, are co-authors of "indivisible: restoring, faith, <b>family</b> , and freedom before it's too late" (ignatius press/faithwords, 2012). jay richards and james robison: the chicken inquisition - wsj.com
gay	157	One can see from the example for “family” that “Mjodr” is against homosexual (gay) marriage, although just a quote of someone else is used there.
law	157	May refer in different laws, among others the American law.
black	146	In the most cases, the word “black” is used to refer to dark-skinned people, like African Americans, but funny enough it can also refer to one

		specific fair-skinned man called Don Black. From the excerpt from line 330, he is the founder of the white-supremacist website stormfront.org: (...) <i>he created volksfront while serving time in prison for assaulting a <b>black</b> man, according to the splc. (...) "if there's any threat to the white, we'll kill the ****ing <b>blacks</b>," he said in the interview. he added that he doesn't like killing but just cares deeply about his own race. don <b>black</b> was a former kkk leader before splitting with the klan and creating stormfront. (...) <b>black</b> and nine other white supremacists were arrested in 1981 for allegedly planning to invade dominica, overthrow the <b>black</b> government, and turn it into a "white state," according to the splc. he spent three years in federal prison. <b>black</b> ultimately split with the klan in the late 1980s and went on to create white supremacist website stormfront.org. (...)</i>
marriage	146	According to "Mjodr", a marriage is always between a man and a woman. He seems to think that same-sex marriage is a bad thing.
government	146	Government in the USA or other states
president	145	The president mentioned in the messages of "Mjodr" is usually Barack Obama. Line 321: "(...) <b>president</b> obama, today joining the chorus for gun control inside the united states in the wake of the batman massacre, has previously indicated that he would sign the treaty, which would then have to be ratified by the senate. (...)

From the words in the table above, it looks like mjodr discusses issues in the political situation in the USA today, which right-extremists, but not only they, seem to be concerned about. They think that family values must have a stronger position in society. A marriage according to "Mjodr" or whoever he quotes shall always be between a man and a woman, not two of the same sex.

In the list of words used by Mjodr, we also find "romney" (117) and "shooting" (121), so the previous American election and weapons are discussed by this author. It is a little unclear exactly what "Mjodr" thinks about weapon availability for the people, but he/she mentions a host of shooting episodes:

Line 554: "(...) *the brady campaign to prevent gun violence reported last month that the united states experiences an average of around 20 mass shootings each year as of late, and at the time of the early july aurora, colorado **shooting**, the study called the theater massacre the sixth mass **shooting** in the month of july alone. (...)*"

The author named WhiteRights seems to write about politics and society today, like Mjodr does. Amongst the most frequent words the examples in Table 40 stick out:

**Table 40: GTFs and examples for "WhiteRights"**

Word	GTF	Explanation/Example
world	198	Used in "third world", "world war" and other phenomenons in this world.
California	189	Line 43 seems to mention some kind of economic crisis in California: <i>"drdirtbag quote: originally posted by drdirtbag this stuff scares the heck outta me. when will these cities that are barrely hanging on start</i>



		<i>to fall and when they do we are all gonna end up with the short stick while the non whites take all the resources. right now, the cities that over-extended themselves financially, who started building new stadiums or new airports in 2007, are starting to go bankrupt. with each bankruptcy, that means more city workers unemployed and more pressure on the state of <b>california</b>, which is already in bad shape.”</i>
gun	158	“WhiteRights” claims in line 212 that the federal government will use harsh methods to confiscate guns: Troops are allowed to kill people who refuse to hand in their guns voluntarily. <i>“obama and un <b>gun</b> confiscation: us troops have orders to shoot americans who resist one more reason to vote against obama. he can't wait to start pushing <b>gun</b> control. <b>gun</b> confiscation: us troops have orders to shoot americans who resist troops ordered to kill all americans who do not turn in guns - youtube troops ordered to kill all americans who do not turn in guns – youtube”</i>
control	142	“gun control” is a frequent expression used by “WhiteRights” like in the example for “gun” above. It is the American federal government that wants to control who owns guns, so that as few people as possible get killed by accident or murder.
illegal	138	“WhiteRights” is very critical to immigration from poor countries to the USA, either directly from Mexico, or from other Latin American countries through Mexico. The author even wants the American government to spend lots of money on deporting the illegal aliens (foreigners). Line 283: “(...) if republicans and their conservative fiscal policies want to stand a chance in the future, they need to reverse the demographic policies of the democrats, cut off all <b>immigration</b> from needy third world nations, deport the 20 million <b>illegal</b> aliens before they all have anchor babies and open up immigration to white europeans so that it’s much easier for white people to come here.”
power	136	Both electric power and political power are mentioned by the author. He seems to be interested in energy politics, but it is unclear whether he/she is for or against renewable energy. <i>Line 337: “leucocyte quote: (...) the biggest problem with wind farms is that we can't build enough to <b>power</b> the us. (...) between 1978 and 1996, the aquatic species program (asp) focused on the production of biodiesel from high lipid-content algae growing in outdoor ponds and using co2 from coal-fired <b>power</b> plants to increase the rate of algae growth and reduce carbon emissions. (...)”</i>
crime	135	Many types of crime are mentioned, but it seems to be extremely bad if the committer is black in line 381: “re: <b>black</b> woman gets 12 years for snatching ny baby in 1987 what a horrible crime! apparently it was a <b>black-on-black</b> crime. (...)”
media	130	
permission	88	“permission” is most likely to be found in the expression “(article reprinted with permission.)”, which can be found several places in the texts by “WhiteRights”.
immigration	83	See example for “illegal”
Einstein	82	The author does not seem to be a fan of Albert Einstein. He/she even

		claims that Einstein ruined the physics science and even claims him for plagiating the equation $E = mc^2$ . It is no there Einstein was a Jew, so it is likely to believe that is the reason the author wants to disparage him. Example is from line 128: <i>“re: how <b>einstein</b> ruined physics (article reprinted with permission.) albert einstein: plagiarist and fraud by ian mosley albert einstein is today revered as “the father of modern science”. (...) the most glaring evidence against einstein concerns “his” most famous equation. one website notes “the equation <math>e=mc^2</math>, which has been forever linked to einstein &amp; his theory of relativity was not originally published by einstein. according to umberto bartocci, a professor at the university of perugia and a historian of mathematics, this famous equation was first published by olinto de pretto ...two years prior to einstein’s publishing of the equation. in 1903 de pretto published his equation in the scientific magazine atte and in 1904 it was republished by the royal science institute of veneto. einstein’s research was not published until 1905... (...)”</i>
public	80	public comment, public hearing, public Christmas displays, and so on. This word is used in many different contexts in the author’s writing.
birth	80	“WhiteRights” comes with a doubtless racist comment about putting people in poor countries on birth control, which means that would be just allowed to get a limited number of children. The author seems to think there are too many non-White people. Example from line 667: <i>“re: liberal cnn features wyclef jean: give hatti money. majority comments say “no way” how much is jean wyclef making? she might be able to feed everyone in _aiti herself. why rob more money from poor, over-taxed americans? the truth about most third world nations is that they need <b>birth</b> control more than anything else. if we had begun reducing third world populations beginning in 1900, we wouldn’t have one billion useless negroes with their average 70 point iqs. but the jews got the white race to fight two fratricidal world wars, instead of focusing on more important things like controlling the third world population, and keeping it out of white nations!!”</i>
liberals	80	In line 451, the “liberals” is the Liberal Party in France, who wants to introduce high taxes for rich people. It is a little unclear what the opinion of the author is. He/she seems to mostly quote news from other sources. <i>“re: france plans 75% tax rate on the rich these socialist tax grabs tend to harm the economy. the rich people will move to monaco and take their money with them. the <b>liberals</b> may “feel good” about passing high tax rates like that, but they won’t increase revenue. (...)”</i>

WhiteRights seems to live up to his/her nickname. He/she is worried about illegal immigration, and mentions cases related to that, which are reported in mass media. It looks like he/she is worried about illegal immigrants mainly from Mexico. There are some racist comments in his/her text, like in Mjodr’s texts. (examples)

Another writer called “revision” appears to write both in French and English. It is difficult to figure out exactly what his/her opinions are, since much of text he has posted in the Stormfront forum, is quotes from different news sources, but we get the gist of his/her opinions through them, that he/she seems to have sympathy with Nazis. It is likely to

believe that he/she at least He/she uses among others the French words “juifs” (Jews, used 258 times), “guerre” (war, 133), “allemagne” (Germany, 79), “monde” (world, 76), and “racisme” (racism, 77). Else, the English words shown in Table 41 are mentioned in the forum (examples from the file revision.txt):

**Table 41: GTFs and examples for "revision"**

Word	GTF	Explanation/Example
holocaust	290	<p>In line 2, “holocaust” is applied when mentioning the bombing of Dresden in 1945 which killed lots of German civilians:</p> <p><i>“dresden <b>holocaust</b> 1945: an apology to germany is due dresden <b>holocaust</b> 1945: an apology to germany is due as queen elizabeth ii opens a memorial to the 55,000 members of the royal air force bomber command who died during the second world war, increasing numbers of britons are questioning the history and legacy of that conflict. the political leaders (principally prime minister winston churchill) who sent those men of bomber command to their deaths – as well as condemning 500,000 german civilians to be burned alive across sixty towns and cities that were devastated in a deliberate bombing strategy – are now seen by some as war criminals. (...)”</i></p> <p>Holocaust is a word for the mass killing of Jews in Europe during the Second World War. Lots of Neo-Nazis and white supremacists claim that holocaust never happened. The following example from line 48 is about a court trial where:</p> <p><i>“white supremacist can seek help online in attempt to prove racist claims: (...) the court order gives terry tremaine permission to ask for information from denizens of online forums, some of which are notorious for virulent messages of intolerance, suggesting his eventual criminal trial could become a controversial marathon like ernst zundel?s and jim keegstra?s attempts to prove the <b>holocaust</b> was a fraud. white supremacist terry tremaine can seek help online in attempt to prove racist claims: court   news   national post”</i></p>
Nazi	257	<p>The word “Nazi” seems to be mostly used in quotes from news articles. Whether the author finds that word insulting or not, is not quite clear. In the following example from line 197, a former nazi guard seems to be soon deported from the USA, where he has lived for many years:</p> <p><i>“re: la traque des criminels nazis / «la vieillesse ne doit pas protéger les criminels de guerre» ex-nazi guard in mercer county to appeal deportation order pittsburgh (ap) - a former <b>nazi</b> guard who has lived for decades in mercer county is fighting the government's effort to deport him. geiser came to the u.s. in 1956 and became a citizen six years later. he's now an 87-year-old retired steelworker living in sharon. ex-<b>nazi</b> guard in mercer county to appeal deportation order - wcn - westminster cable network”</i></p>
Jewish	216	<p>Jews are often mentioned in a right-extremist forums, because Jews are hated by these people.</p>
News	172	<p>The word “news” is mostly used for when quoting a news source. “breaking news” is a frequent expression”. Sometimes “news” is a part of the name of the source, like in “bbc news”. From line 379: “ny</p>

		<i>comptroller to help locate holocaust victims' assets ny comptroller to help locate holocaust victims' assets new york state comptroller thomas dinapoli has agreed to help track down property that belonged to jewish victims of the holocaust, a jewish group said monday. "the issue of restoring jewish property to their owners is still key even though 70 years have passed since the war's end," said lea nass, deputy minister of pensioners affairs. jerusalem post - breaking <b>news</b> "</i>
anti	135	An ideology opposite to another ideology, like anti-racism
camp	111	"camp" is of course referring to concentration camp, since this is a right-extreme forum. The following excerpt from line 288 is from an Australian news site and is about Australian economical contribution to restoration of Auschwitz and Birkenau: <i>"australia gives \$500k for auschwitz-birkenau aust gives \$500k for auschwitz-birkenau australia will contribute \$500,000 to help preserve the world war ii concentration <b>camp</b> sites, auschwitz and birkenau, in southern poland. prime minister julia gillard says the <b>camp</b>s continue to serve as a solemn reminder of the terrible crimes perpetrated against millions of people during the war. (...)   news.com.au"</i>

An person nicknamed "Revilo" seems to only write in English. The main topic in his writing is the Second World War. Frequently used words with examples can be seen in Table 42: GTFs and examples for "Revilo".

**Table 42: GTFs and examples for "Revilo"**

<b>Word</b>	<b>GTF</b>	<b>Explanation/Example</b>
Jews	495	"Revilo" seems to be eager to find proofs that Jews are a race in line 44: <i>"re: dna links prove <b>jews</b> are a ?race,' says genetics expert righteous racism by professor revilo p. oliver righteous racism by professor revilo p. oliver (liberty bell, july 1987) i occasionally receive inquiries from persons who have come across some writing of mine and ask why i invariably refer to the <b>jews</b> as constituting a race, although americans have always been taught that the chosen people are only a religion, venerable because they discovered monotheism, which is a superior form of religion because they discovered it. (...)"</i>
revolution	78	"revolution" is a sudden change in politics according to line 279, often violence is used in this process like in the Russian October revolution: <i>"re: who financed lenin and trotsky? the jewish role in the bolshevik <b>revolution</b> and russia's early soviet regime the jewish role in the bolshevik <b>revolution</b> and russia's early soviet regime assessing the grim legacy of soviet communism by mark weber in the night of july 16-17, 1918, a squad of bolshevik secret police murdered russia's last emperor, tsar nicholas ii, along with his wife, tsaritsa alexandra, their 14-year-old son, tsarevich alexis, and their four daughters. (...)"</i>
family	78	Family values are not praised particularly by this author, but is just mentioned in a neutral way. In line 406, a fact that Karl Marx' family was rabbis (Jewish priests), is used in an attempt to convince readers that communism is based on Judaism, which is not true, since Marx,

		who admittedly was a Jew, did not believe in that religion: <i>“coldstar quote: originally posted by coldstar this guy looks jewish to me ... communism is jewish. communism is directly linked to the talmud, karl marx was from a rabbinical family.”</i>
power	75	power is political power in the messages by this author, like in line 408, where he/she claims that the Jews have the real power in the United States through lobbyism: <i>“re: democrats? ?jewish cash money team? israel lobby in the united states - wikipedia, the free encyclopedia j.j. goldberg wrote in his 1994 book jewish <b>power</b> that 45% of the democratic party?s fundraising and 25% of that for the republican party came from jewish-funded political action committees.[46] richard cohen, a columnist for the washington post, updated those figures in 2006 citing figures of 60% and 35% respectively for the democratic and republican parties. according to the washington post, democratic presidential candidates depend on jewish sources for 60% of money from private sources.[47]”</i>
god	68	“god” can be the God in Christianity and Judaism in the posts written by this author, for instance in line 336: <i>“re: jewish-born clerics helped push vatican ii reforms martin luther - on the jews and their lies: quote: (...) the sun has never shone on a more bloodthirsty and vengeful people than they are who imagine that they are <b>god</b>'s people who have been commissioned and commanded to murder and to slay the gentiles. in fact, the most important thing that they expect of their messiah is that he will murder and kill the entire world with their sword. (...)”</i> Jews regard themselves as God’s chosen people, but this author, who seems to hate Jews, stretches this too far. He/she claims that Jews really want to kill other people if they had power to do that.
church	68	A church is a God’s house in Christianity or a national/worldwide Christian community. From line 298: <i>“re: jewish converts who changed the <b>church</b> giles fraser - wikipedia, the free encyclopedia giles anthony fraser[1] (born 27 november 1964[3]) is a priest of the <b>church</b> of england. (...) he has been involved in social and political advocacy and according to the daily telegraph "would be the first to admit that he is fond of the sound of his own voice".[7] fraser has contributed to bbc radio 4's thought for the day. since 2004, he has had a weekly column in the <b>church</b> times. (...)”</i>
free	66	“free” can mean the ability of someone to do what one wants to do or that something does not cost anything. A little surprising is that the latter meaning is dominant in the text by “Revilo”, like in line 309: <i>“re: got scammed by jewish corporation on the jews and their lies - wikipedia, the <b>free</b> encyclopedia”</i> As commonly known, one cannot trust absolutely everything written on Wikipedia. However, we cannot find out in which Wikipedia article this quote is retrieved from, or whether it is from Wikipedia at all.

## Analysis of 10 most active authors in Ummah and Stormfront

In this section, we will look at the words that all the ten most active authors in Ummah and Stormfront use, respectively. We can then actually assume that many stop words are used by all of the authors in a complete forum. We do not need to prove this, since many of these words (like “and”, “is”, and “has”) are necessary for writing meaningful English sentences. In this analysis, like in the rest of the project, it is therefore no point including stop words. We have also excluded the words that are used less than ten times in the complete forum from the analysis. Some of the words could be a little important, because they are likely unique for the less than ten authors who use them. These words could have been useful features for recognizing exactly the authors who use them. However, if we included these infrequent words, we would also get some noise with many uninteresting words, because there are so extremely many words with a GTF less than 10. This would make the analysis difficult, so we have to make a trade-off, and analyse as best as we can without these rare words included.

As a result of this, there are no words in the analysis that absolutely all authors in a forum use. If we just analyse the ten most active authors in Stormfront, all ten use the words “president”, “party”, “free”, “time”, “israel”, “immigration”, “society”, and so on. Briefly, the ten most active Stormfront authors write about politics, society, family, war and Jews. Hence, they seem to discuss some of the same topics.

The top 10 productive authors in Ummah all use words “time”, “times”, “day”, “help”, “family”, “life”, “world”, “hand”, “people”, “words”, “little”, “instead”, “book”. The words used by the Ummah authors are more generic and less specific than the Stormfront authors’ words. “family” and “book” seem to be the most specific. Interesting is that both of the top ten author groups are preoccupied by family and some kind of literature. Both groups somehow discuss life, which is of course important to all human beings. Else most of the words are very general.

A vocabulary intersection of a set of authors is the set of words that all these authors apply. There are too many words in the intersection of the vocabulary of Stormfront authors to mention all the words that are different from the vocabulary intersection of Ummah, so we just mention some of the most important of these words here: “president”, “government”, “history”, “country”, “immigration”, “population”, “law”, “society”, “rights”, “social”, “nations”, “europe”, “war”, “political”, “act”, “news”, “jews”, “white”, “nation”, “children”, “national”, “america”, “money”, “attack”, “police”, “power”.

From this, we confirm that the ten most-writing Stormfront authors seem to write about politics, war, Jews, and the society in America (here: the USA) and Europe.

“help”, “able”, “hand”, “words”, “little”, “instead”, and “book” are words that all the ten most active authors in Ummah use, but the ten most active contributors in Stormfront do not use. These words from the Ummah authors are too generic to get any good clue of they write about. “book” may imply some writing about one or another book.

The words “time”, “times”, “day”, “family”, “life”, and “world” are words all 10 most active authors in each of the forums use. They need the three first words to express points in time, which is very normal for all human beings. The word “family” should imply that family is important for all these 20 active forum authors. The two latter words hint that they are interested in what happens in their lives and in the world.

## Unique words for authors in Ummah

The author “Anna” writes in English and sometimes also in Arabic. Among his/her unique words we can find “literally”, “root”, “marks”, “revision”, “depends”, “antecedent”, “lesson”, “patterns”, “pronoun”, “definite”, “allah”, “describing”, “radhiallahu”, “masculine”, “passive”, “kasra”, “sentences”, “plural”, “verb”, “participle”, “appreciation”, “vocabulary”, “adjective”, “grammar”, “translate”. The author seems to be interested in writing about English grammar lessons and exercises. He/she also writes about religion Islam.

The author “AbuMubarak” writes in English. Among his/her unique words we can find, “catholic”, “democratic”, “revolution”, “industry”, “syndrome”, “legislation”, “superiority”, “feminists”, “pork”, “russell”, “protest”, “prophethood”, “equality”, “saddam”, “leadership”, “politics”, “hypocrisy”, “liberated”, “abortions”, “sexuality”, “relationships”, “halal”, “converted”, “ordained”, “civilization”, “balance”, “suspicion”, “shaytaan”, “deity”, “constitution”, “enforcement”, “statistics”, “kafir”, “obeying”, “slaughter”, “forgiving”, “taliban”, “abused”, “evils”, “sexes”, “happiness”, “loving”, “powerful”, “gospel”, “harmony”, “radical”, “custody”, “raped”, “republic”, “arabia”, “clinton”, “colored”, “pentagon”, “osama”, “senate”, “ideologies”, “racist”, “racism”, “ferguson”, “mecca”, “spirit”, “hate”, “race”, etc. This author’s writing seems therefore to be much oriented about the religion Islam and western countries like USA and their cultural influence. The author highly criticizes western countries and his/her negative opinions about them are vividly captured in what he/she has written in the forum.

The author “Abdul Haq” seems to write in English as well. His/her set of unique words includes “camp”, “bethlehem”, “deaths”, “assault”, “helicopter”, “ambulances”, “jerusalem”, “humanitarian”, “territory”, “explosives”, “atrocities”, “warnings”, “incursion”, “curfews”, “palestinians”, “jenin”, “martyr”, “fired”, “wounded”, “palestine”, “israelis”, “albert”, “refugee”, “massacre”, “pearl”, “militants”, “ramallah”, “defensive”, “anthrax”, “northwoods”, “bombing”, etc. This author seems to be interested in writing about politics, war, terror and crisis situations such as the conflict between Israel and Palestine and their dispute issues.

The author “hamza81” writes in English and sometimes seems to write in Arabic as well. Among his/her unique words we find “congregation”, “tip”, “imdad”, “mukhtar”, “dhuhaa”, “file”, “nights”, “voluntary”, “allahumma”, “fatiha”, “ritual”, “offers”, “rewards”, “steps”, “performer”, “fasts”, “habit”, “lifetime”, “prays”, “checklist”, “breaks”, “virtuous”, “sunrise”, “imaam”, “qaidah”, etc. This author is very interested in writing about how Islamic devotees perform rituals and observances such as fasting and prayers, and engage in charitable work during the month of Ramadan. Muslims believe these rituals are very invaluable, and they also think they are set free from their sins when they perform the rituals with a honest heart during the month of Ramadan.

The author “Z-Blade” writes in English and it seems like he/she writes some in Arabic too. His/her unique words are “haha”, “assalamualaykum”, “kubs”, “jazakallah”, “joke”, “cool”, “prostrate”, “intercede”, “wassalamualaykum”, “hadithhadith”, “duaa”, “feek”, “haya”, “funny”, “sea”, “nasheed”, “yep”, “hehe”, “hmmm”, “blade”, “barakallahu”. This author seems to use more of greetings in his/her messages and writings such as, “assalamualaykum” (Thank you), “barakallahu”, “jazakallah”, etc. “jazakallah” is an Arabic word which is used as an expression of gratitude. In Islamic terminology it means “May Allah reward you with goodness”. It is also very popular among Muslims when it comes to thank someone.

## Unique words for authors in Stormfront

We have chosen to use only the writings of the ten most active authors in a forum to find the words unique to each single author. This means that an author-unique word, as we present it here, really is not necessarily unique to only that author in the forum. The forums Stormfront, Ummah, and Swiss English Forum have thousands of different authors and it will therefore take a lot of time to calculate which words are unique for each single author. Nonetheless, the unique words, as we define them, can be used to differentiate one author at least from the nine other most productive authors.

The author “kazan188” frequently write in French, so many of the unique words are therefore French. Many of the unique words, both in English and French, show that many of the commentaries are about the Second World War. We also see the French words “islamique” and “syrie”, which should mean that issues from today, like extreme islamists and the war in Syria are mentioned. Unique English words are “propagandists”, “hysteria”, “farm”, “cellar”, “thesis”, “pictures”, “inmates”, “publisher”, “novel”, “partisan”, “bombers”. He/she mentions “franklin”, the first name of the American president, Franklin D. Roosevelt.

The author “Mjodr” only writes in English. Among his/her unique words, we can find “veteran”, “camping”, “drink”, “hungry”, “transgender”, “illegally”, “urge”, “divorced”, “wildlife”, “write”, “girlfriend”, “divide”, “wolves”, “web”, “enjoy”, “bull”, “sorry”, “firearm”, “sentenced”, “restaurants”, “confiscated”, “pain”, “apartment”, “batman”, “dinner”, “whitehead”, “arrest”, “moose”, “abortion”, “cinema”, “thief”, “bags”, “eating”, “daughters”. This author seems to write about everyday topics. Not so many of the unique words seem right-extreme, but we can glimpses of some frustration of criminality: “illegally”, “arrest”, “thief”.

“WhiteRights” only writes in English. some of the unique words are “customer”, “electoral”, “celebrity”, “verdict”, “worldwide”, “suicide”, “screaming”, “breath”, “ordinary”, “lying”, “loan”, “defenseless”, “smoke”, “louisiana”, “neighborhoods”, “hits”, “judges”, “cop”, “infrastructure”, “corporations”, “bedbugs”, “birthday”, “katrina”, “camera”, “funny”, “mortgage”, “arafat”, “bankrupt”, “identical”, “homosexuality”, “santa”, “forgery”, “lovelock”, “furious”. Here, we find some words from economy (bankrupt, mortgage, loan). Else, the words look pretty random. “katrina” may be the 2005 hurricane which swept over Louisiana in the USA and caused a lot of destruction.

“revision” writes some messages in English, but he/she also write much in French. We especially notice the words from the latter language in the list. Unique French words are “résistance”, “problèmes”, “cérémonie”, “antisémite”, “internés” (interned), “condamnés” (sentenced persons), “légion”, “attentats” (assassinations). These words seem to be connected to racism (“antisémite”) and prisoners of crime or war (“internés” and “condamnés”). “résistance” may refer to the resistance against the occupation of most of France during the Second World War. “légion” is a division in the an army. It looks like “revision” has written very little in English, but some of his/her unique English words are “sugar”, “investment”, “hero”, “unearthed”, “insect”, “prehistoric”, “basketball”, “rocks”, “astronomers”, “cats”, “exploitation”. The English words seem random and they are few, so he/she has not used many unique English words in the forum discussion.



We even find the German word “Glückwunsh” (congratulation). Merkel, the German chancellor is mentioned, and also the right-extreme political party NPD. The use of some German expressions may imply that “revision” has some knowledge of what happens in German politics, or at least has some interest in that.

“Revilo” writes everything in English. He/she is interested in war history, politics, and racial ideology, but writes also about religion and family values. These are some of the words used in his/her writing: “Jewish”, “Israel”, “war”, “world”, “American”, “Germany”, “Russia”, “government”, “president”, “white”, “revolution”, “family”, “power”, “church”, “France”, “Christian”.

## 4.20 Word Colocation Analysis

Analysis of co-occurrences of all possible two-word combinations takes a lot of time to find with a computer program if all possible word pair combinations are to be examined. The Java program we wrote for finding such collocated word pairs and their accompanying odds ratios, could spend hours, or even days to test if thousands or even millions of possible word combinations really were in some forum messages. Therefore, we choose to only examine word pairs where each of the words were used at least a given amount of times. In the forums Steam, Islamic Web-Community, Turn to Islam and Deutsche Stimme, we excluded words that are used less than times from the research, so that the program was finished in less than an hour in each case. In Ummah we had to ignore words used less than 100 times, so that it could finish in just a few hours (about five hours). We cut away the words used less than 200 times from the Stormfront forum, but it still used days to finish. We still could not remove more words, because we were afraid that would impact the results too much. We remove words that are not so much used in the forum in comparison to the most popular words, because there are hundreds or even thousands of distinct words in a single forum, so computing the odds ratio of all word pairs is simply not feasible in a reasonable amount of time, as we just have described.

### 4.20.1 Steam

In the following analysis we have included the words that are mentioned 10 times or more in the forum Islamic Web-Community.

The words “originally” and “posted” have a very strong co-location relation to each other, an odds ratio of 4,530,225. They are used together when mentioning a quote after the expression “originally posted by” and the name of the quoted forum user. “bomb” and “level” have an odds ratio of 1,886,973. In a game level, one team (the terrorists) plant a bomb, while another team (the counter-terrorists) tries to prevent that from happening or they defuse it. In the following extract from the forum, as shown in line 66252, it is mentioned that bots (game characters not controlled by real persons) should never carry the bomb and they should have the same skill level as the players and follow the commands of the players. It is also said that players away from the keyboard should never get the opportunity to carry the bomb (with their playable characters):

*“('Steam', '10\_t24\_p1', 'Comment', 'Bots: All great, they should not get the **bomb**, they should follow commands and their skill **level** needs to be adjusted cause they are right now simply useless. 1. Great, Afkers should not get the **bomb** either. (...)', 'Kiccpe', '4 hours ago ')”*

“hacks” and “lose” have the odds ratio 1,270,833. That someone hacks the game system leads to the event that someone else loses. The words have no clear relation to each other in the following example from line 66383, but they are at least used in the same message:

*“('Steam', '10\_t9\_p1', 'Comment', 'point 1, 2: should be nice. point3: nonsense. THats the way the game works. point4: agreed point5: nonsense. Dont play if you dont wanna **lose**. (...)  
point9: see some post of a hacker finding guy: it is not so easy te be sure that someones **hacks**... if it is very clear, this should be oke. Better strategy is not to kick a hacker, but to follow him, record and sent the recording...', 'Snuijfe', '9 hours ago ')”*

The words “undesirable” and “abuse” have together an odds ratio of 1,261,960. These words fit well together because “abuse” is of course undesirable by the game producer and fair players. The following message seems to anyhow originate from Valve/Steam itself and it states clearly that abuse of the system in the multiplayer game “CS:GO” leads to undesirable experience for players of that game.

From line 66412:

*“('Steam', '11\_t16\_p1', 'About Global Offensive Bans', 'Please do not create topics complaining about in-game bans, they will be deleted or closed. A note from Support about in-game bans Player abandons, AFKing, and teamkilling result in an **undesirable** experience for all of the other players in the game. Our system is designed to limit the negative impact of users who repeatedly **abuse** these systems. (...)', 'ianskate', 'Feb 11, 2013 @ 1:57pm')”*

“smooth” and “drivers” (1,226,448) fit together because the game runs smooth when correct drivers for display, sound, etc. are installed on the computer and of course when the hardware is good enough for running the game. In line 66687 a user seems to struggle with hardware, drivers, and game update issues:

*“('Steam', '12\_t23\_p1', 'CS:GO running slow', 'CS:GO is running slower than usual on my Macbook Pro Retina + Windows 7. Unplayable. I usually run the game **smooth** on 2880x1800, but since the last update, the game runs terribly slow. At 1900x1200 it still has too low FPS. I completely reinstalled Windows 7, updated all **drivers**, but the problems isn't solved. Specs: i7 2.3GHz 16GB RAM NVIDIA GeForce GT 650M (1GB) 251GB SSD harddisk Anyone else experiencing problems?', 'KeesKachel', 'Dec 22, 2012 @ 4:21am')”*

“global” and “offensive” (752,893) of course belong together because of the name of the discussed game, “Counter-Strike: Global Offensive”.

The pair “weapon” and “firebomb” (644,893) makes sense because a firebomb is a weapon. The two words are mentioned in two different events right after each other in a video recording of some gameplay (at 4 min 49 sec and 5 min 55 sec from the start of the video).

Line 66829:

*“('Steam', '12\_t7\_p1', 'Comment', '(...)4:49 people hoping around and you are trying to kill a person with basic **weapon** when he has smg 5:55 you are running through a **firebomb** and you lose 40 health.(...)', 'zPx', '2 hours ago ')”*

Other examples of co-occurring word pairs can be seen in Table 43.

**Table 43: Co-occurring word pairs in Steam**

Word 1	Word 2	Odds ratio	Explanation
kill	enemies	780.1	The enemies that are killed in the Steam forum are playable characters in the game “CS:GO”, not real people. Line 69821: “('Steam', '24_t47_p1', 'Comment', 'ok, for example: <b>kill 100 enemies</b> with Deagle ... I had aprox. 70 and now is 0 :( this hapens in all weapons except finished archvivismts next... win 100 round in Italy/Dust ... I had 75 / 50 and now is 0 ...', 'vsn', 'Just now ')”
play	score	788.6	When one plays a game it is normal to get a score. Sometimes people become jealous someone gets a higher score than themselves. Line 69824: “('Steam', '24_t49_p1', 'Troll Clan of Idiots: XB_64', 'I recommend you folks do NOT <b>play</b> with the clan (if it even is a clan) called XB_64. (...) They decided to kick me because I was playing better than all of them were, whilst having the best <b>score</b> out of both teams on the session. This is one reason why MP gaming is very hit and miss nowadays, because of idiots.', 'Phazar', '14 hours ago')”
cheaters	appear		Fair players of course hate cheaters, and in “CS:GO”, the author “Toraqi” thinks that cheaters mostly appear among the most professional players in line 66254: “('Steam', '24_t9_p1', 'Comment', 'Some of the issues are not really that easy to solve. But the ban for competetive should be from 2 hours to 12h to 24h to 3 days to 7 days or something. Cheaters seem to appear only in the higher ranks... they will need to fix that though by being a lot less forgiving for ruining the experience of a lot of players.', 'Toraqi', '3 hours ago ')”

## 4.20.2 Islamic Web-Community

In the following analysis we have included the words that are mentioned 10 times or more in the forum Islamic Web-Community.

Among the pairs with highest odds ratios, we find a lot of words that are not related to the discussion in the forum like “gpl”, “license”, “copyright”. They seem to come from pieces of text on the website that are not really forum messages written by humans. One can see some examples below, on the top of the table.

**Table 44: Co-occurring words in Islamic Web-Community**

Word 1	Word 2	Odds ratio	Explanation
gpl	license	48884	“gpl” means “GNU Public License” Line 49: “ <i>PHP-Nuke is Free Software released under the GNU/GPL license.</i> ”
reserved	copyright	47736	Line 52: “ <i>Web site Engine's code is <b>Copyright</b> © 2002 by PHP-Nuke. All Rights <b>Reserved.</b></i> ”
reserved	released	47736	Same example as for “gpl” and “license”
deeds	praise	11546	Line 86: “(‘Islamic Web-Community’, ‘f1-t18’, ‘(...)’, ‘Ten Days of Righteous <b>Deeds</b> Even for those not performing the pilgrimage, Hajj, the first ten days of this month are considered very sacred and a time for increased reflection, seeking Allah's forgiveness, doing good and various other forms of <b>worship</b> . Prophet Mohammed (peace and blessings be upon him) has said about the first ten days of Dhul Hijja: "There are no days in which righteous <b>deeds</b> are more beloved to Allah than these ten days." The people asked, "Not even <b>Jihad</b> for the sake of Allah?" He said, "Not even <b>Jihad</b> for the sake of Allah, except in the case of a man who went out to fight giving himself and his wealth up for the cause, and came back with nothing (Bukhari). While any good deed done for the sake of Allah according to the way He approves will be rewarded immensely during the first ten days, Insha Allah, some of the more specific <b>actions</b> mentioned in the Traditions of the Prophet are as follows: 1. Fasting In terms of fasting, it is particularly encouraged to fast on the ninth day of Dhul-Hijja, known in Arabic as Yawm Arafa. The Prophet used to fast on this day (al Nisai and Abu Dawud). Fasting on this day will expiate a Muslim's sins for two years. 2. Dhikr The <b>verbal</b> remembrance of Allah is another meritorious act during these first ten days of Zul Hijjah. The Prophet upon him) encouraged Muslims to <b>recite</b> a lot of Tasbeeh ("Subhan-Allaah"), Tahmeed ("Al-hamdu Lillaah") and Takbeer ("Allahu akbar") during this time. The Takbeer may include the words "Allahu akbar, Allahu akbar, la ilaha ill-Allah, wa <b>Allahu akbar</b> wa Lillahi?l-hamd (Allah is Most Great, Allah is Most Great, there is no god but Allah, Allah is Most Great and to Allah be <b>praise</b> )," as well as other phrases. Men are encouraged to <b>recite</b> these phrases out <b>loud</b> and women <b>quietly</b> . 3. <b>Sacrifice</b> One of the good <b>deeds</b> that will bring a person closer to Allah during these ten days is

			<p>offering a <b>sacrifice</b>, by choosing a high-quality animal and fattening it, spending money for the sake of Allah. 4. Sincere repentance One of the most important things to do during these ten days is to repent sincerely to Allah and to give up all kinds of disobedience and sin. This means more than just a verbal expression of sorrow for past <b>misdeeds</b>. It also requires a firm resolution to avoid making the same <b>mistakes</b> in the future by giving up bad habits and behavior while sincerely turning to Allah. Generally, all good <b>deeds</b> are <b>rewarded</b> highly at this blessed time. These actions include praying, reading Quran, making Dua (supplication), giving in charity and being good to our families. Compiled From: "The first 10 days of Zul Hijjah: why they matter" - SoundVision.com "Virtues of the Ten Days of Dhul Hijjah" - Salih Al Munajjid' 'lubna', '21-11-2008, 23:11')"</p>
rewarded	deeds	11044	See example for “deeds” and “praise”.
rewarded	praise	11044	See example for “deeds” and “praise”.
akbar	allahu	10542	“Allahu akbar” is Arabic and means “God is the greatest”. See example for “deeds” and “praise”.
loud	sacrifice	9126	See example for “deeds” and “praise”.
recite	worship	8128	See example for “deeds” and “praise”.
produce	generate	7154	<p>Both “produce” and “generate” means to yield electrical power in of a power plant in the following example from line 455:</p> <p>“(‘Islamic Web-Community’, 'f2-t46', '(...) The Solar Plants will be the largest in the World and will <b>produce</b> 800MW equivalent to a medium coal-burning plant- A must read for those who are interested in a more ecological world. (...) Though the California installations will <b>generate</b> 800 megawatts at times when the sun is shining brightly, they will operate for fewer hours of the year than a coal or nuclear plant would and so will <b>produce</b> a third or less as much total electricity. (...) A recently built plant that uses mirrors to concentrate sunlight, called Nevada Solar One, can <b>produce</b> 64 megawatts of power. (...)”’, 'Tayeb', '19-08-2008, 13:06')”</p> <p>Synonyms are used for varying the language.</p>
mistakes	misdeeds	7140	See example for “deeds” and “praise”.
actions	jihad	6656	See example for “deeds” and “praise”.
verbal	quietly	6156	See example for “deeds” and “praise”.

### 4.20.3 Deutsche Stimme

In the word co-occurrence analysis in Table 45 we have included the words that are mentioned 200 times or more in the news/commentary website Deutsche Stimme. One fact that we can establish for certain is that all texts there are in German, which is no surprise, since this website belongs to the German nationalist party NPD.

**Table 45: Co-occurring words in Deutsche Stimme**

Word 1	Word 2	Odds ratio	Explanation
http	www	486.2	Both occur in URL addresses like in this example from line 108190: <i>“(‘Deutsche Stimme’, ‘0_t16’, ‘Comment’, ‘Komisch das bislang nur die BILD-Zeitung darüber schrieb: <b>http://www.bild.de/regional/bremen/totschlag/das-ist-der-killer-von-daniel-29496848.bild.html</b>, ‘Antagonist’, ‘2013-03-14T13:17:23+00:00’)”</i>
cdu	spd	76.2	Both are political parties in Germany. CDU is the Christian-Democratic Union (a conservative party), while SPD is the Social-Democratic Party of Germany (a liberal party). Both parties are listed together with other German parties in line 108330, where the author is unsatisfied with the most of the German political parties, because they do not seem to react as he/she thinks they should. The author thinks that foreigners should be expelled when they commit any crime, like he/she claims they do in Saudi-Arabia: <i>“(‘Deutsche Stimme’, ‘0_t18’, ‘Comment’, ‘(...)Die Leute wollen einfach nichts dagegen machen nur meckern hilft nicht <b>CDU,SPD</b>,Grüne,Linke ,FDP sollte man einfach nicht wählen. Zb. In Saudi Arabien leben fast nur Ausländer und passen sich vernünftig an, weil man dort sofort abgeschoben wird wenn man was kriminelles Verbrochen hat, dass könnte man hier auch einführen. (...) ’, ‘Anonymous’, ‘2013-03-15T14:04:24+00:00’)”</i>
verbot	npd	15.8	A lot of times, German authorities have tried to forbid the extreme nationalist party NPD. The following example from line 108353, explains a misunderstanding from the discussion. Somebody thought that the NPD had been forbidden once before, which is not true. Another right-extreme party called SRP (Socialist Reich Party) was banned in 1952. Many of its members were actually former members of NSDAP, the original Nazi party [51]. It was officially banned because of affiliation with paramilitary organisations, but the author of the

			<p>following example claims it was because of the success in Bundestag elections like in Schleswig-Holstein:</p> <p>“(‘Deutsche Stimme’, ‘0_t19’, ‘Comment’, ‘(..) Nat¼rlich meinte “VS-Agent 0815? nicht das <b>Verbot</b> der NPD – sondern das <b>Verbot</b> der Sozialistischen Reichspartei ( SRP ), wo genau aus diesem Grund ( Wahlerfolg in Schleswig – Holstein ) diese Partei verboten wurde – wenn auch OFFIZIELL mit anderer Begr¼ndung !!! (...)’, ‘M.’, ‘2013-03-14T16:07:46+00:00’)”</p>
opfer	fall	14.1	<p>In line 108659, the word “Fall” means a case, a situation. The “Opfer” is a victim of violence. The author claims that violence does not help stopping the foreigners. He/she thinks that elections must be boycotted and that the laws must be changed:</p> <p>“(‘Deutsche Stimme’, ‘0_t30’, ‘Comment’, ‘Das bringt gar nichts Hass ist Ballast Ja, aber im konkreten <b>Fall</b> steht die Frage aus, wer hier wen zu hassen scheint. Der, der dem <b>Opfer</b> den Kopf eingetreten hat (oder DIE?) oder die trauernden. Kommen Sie mal klar! Gewalt hilft nicht die Ausl¼nder zustoßen, sondern man muss die Wahlen boykottieren und ja die Gesetze ¼ndern h¼rtere Strafen f¼r das vergehen(...)’, ‘Papst LeoII.’, ‘2013-03-15T18:02:01+00:00’)”</p>
medien	opfer	10.4	<p>The following excerpt from a commentary written by Holger Apfel, the NPD party leader himself, claims that German women are victims of rapes done by Turkish and African men. He claims that media seem to care when a Turkish family’s home is burnt down, and that the media then claim implicitly that right-extremists are involved. From line 109540:</p> <p>“(‘Deutsche Stimme’, ‘0_t59’, ‘(...) Gerade junge Frauen werden h¼ufig <b>Opfer</b> triebgesteuerter T¼ter aus Kleinasien oder Afrika. In Dresden erw¼rgte ein Pakistani vor zwei Jahren seine deutsche Freundin, weil sie sich von ihm trennen wollte. Frauen, insbesondere deutsche Frauen, haben bei den Orientalen ohnehin einen geringen Stellenwert. In epischer Breite er¼rtern die <b>Medien</b> aktuell ein Brandungl¼ck in Backnang, bei dem eine t¼rkische Familien zu Schaden kam. Zwischen den Zeilen liest man das Bedauern der Meinungsmacher heraus, den Vorgang nicht ¼Rechtsradikalen¼ bzw. gleich der NPD in die Schuhe schieben zu k¼nnen. Soviel Heuchelei ist nicht zu ¼berbieten. (...), NPD-Parteivorsitzender ’, ‘admin’, ‘vor 4 Tagen’)”</p>
welt	menschen	8.3	<p>“katholischen Menschen”, Catholic people, are in line 110048 compared with “der muslimischen</p>

			<p>Welt”, the Muslim world. People and world are in this example both synonyms for community, people with common culture and belief, who comprise a part of the world. The author blames both Catholics and Muslims for double moral. Double moral is to be against something immoral, while supporting it in secret. The Catholic Church is formally against brothels, although there are lots of them in some Catholic regions. In Islam pork meat and alcohol is forbidden to consume, although there according to the author “Herbert” are some Muslims who consume these things:</p> <p><i>“(Deutsche Stimme', '0_t76', 'Comment', '(...)Denn sieht man bei uns gewisse religiöse Dinge und gerade die Doppelmoral der katholischen Kirche insbesondere im Süden, wird einem Übel. Dort verweigert man nicht katholischen <b>Menschen</b> Arbeit. Dort herrschen teilweise Zustände von der Denkweise wie im Mittelalter. Aber in gewissen nicht von Touristen durchreisten Regionen herrschte eine grosse Bordelldichte, die darauf schliessen lässt, dass eigentlich nur die Bewohner der Umgebung diese Etablissements aufsuchen können. Doch gibt es diese Moral auch in der muslimischen <b>Welt</b>. Schweinefleisch essen auch welche und Alkohol ist eigentlich tabu und wird von Teilen derer trotzdem konsumiert. (...).', 'Herbert', '2013-02-18T15:14:26+00:00')”</i></p>
geld	bekommen	8.0	<p>“Geld bekommen” means “get money”, and in the following excerpt, the author “Flipper der kluge Delphin” claims that an asylum seeker gets 40 euro less than an Hartz 4 receiver (unemployment benefit receiver). It is not mentioned for which kind of period these groups get this money, but maybe for a month. It is not mentioned how much money they actually get.</p> <p>Line 111506:</p> <p><i>“(Deutsche Stimme', '1_t145', 'Comment', '(...) Das Einkommen eines Wirtschaftsflüchtlings, Asylant genannt, liegt um 40 <b>Euro</b> unter dem eines Deutschen Hartz 4 – Empfänger. Glückwunsch, ich bin sehr gespannt auf die Argumentation der Richter, warum künftig der Asylant mehr <b>Geld bekommen</b> soo als der Deutsche Hartz 4 – Empfänger. Ceterum Censeo: Das “Grundrecht” auf Asyl muss zwingend abgeschafft werden!’, 'Flipper der kluge Delphin', '2012-07-18T16:17:14+00:00')”</i></p>
wählen	npd	8.0	<p>The NPD of course wants as many Germans as possible to vote for them in the elections. From line 108314, we see that the writer “M.” intend</p>



			<p>to elect the NPD:  <i>“(‘Deutsche Stimme’, ‘0_t18’, ‘Comment’, ‘(...) Ich werde dennoch wählen – und natürlich die NPD wählen . Wer einmal innerlich die brd hinter sich gelassen hat – der kann freier nicht sein . Â Beste Grüsse M.’, ‘M.’, ‘2013-03-17T13:50:40+00:00’)”</i></p> <p>“wählen” (elect) often occur together with other party names, but then the website against against electing these other parties and scares with how bad they think the society will be like if the other parties may rule also in the future.</p>
medien	politik	7.4	<p>In line 111772, “there is a claim that the media (Medien) and different political parties work together on a policy (politik) which do not support families getting children good enough:  <i>“(‘Deutsche Stimme’, ‘1_t159’, ‘Comment’, ‘So eine Forderung halte ich für nicht so sinnvoll. Auch in der Wählerschaft der NPD gibt es viele Kinderlose. (...) Das beste Mittel gegen Kinderlosigkeit ist Wohlstand u. Gemeinschaftsdenken. Auch die Medien und vor allem CDU, SPD, FDP, GRÜNE u. SED-Linke tragen hier mit ihrer <b>Politik</b> eine deutliche Mitschuld, auf die ich auch nicht näher eingehen möchte, da man hier dann Seitenweise schreiben müsste. (...), ‘Deutsche vs. BRD-Bonzen’, ‘2012-02-16T11:41:58+00:00’)”</i></p>
euro	geld	6.9	<p>Euro is money (Geld), more precisely the common currency for several members of the European Union. See example of usage in the same message in the example for “geld” and “bekommen”</p>
frau	kinder	6.9	<p>A woman (Frau) and a man are supposed to raise children (Kinder) together according to “Josef Klemens”. Same-gender relationships do not work for that purpose, he claims. Of course it is correct that a couple of the same gender cannot get children together in the natural way. Therefore the author claims that same-gender relationships are not the future, because children are the future.  Line 109889:  <i>“(‘Deutsche Stimme’, ‘0_t74’, ‘Comment’, ‘(...) Kinder gehen immer aus Beziehungen zwischen Mann und Frau hervor. Und Kinder sind die Zukunft, das sollte man nie vergessen. Aus diesem Grund sollte man eher die BeziehungenÂ zwischen Mann und Frau fördern, als gleichgeschlechtliche Beziehungen.(...), ‘Josef Klemens’, ‘2013-02-21T09:33:33+00:00’)”</i></p>
eigenen	land	6.9	<p>The NPD dicusses often what happens in their own country, “im eigenen Land”, which is Germany. That German expression is used to awake feelings of the</p>

			<p>reader, like in 111928, where readers are scared by a statement that Germans are getting in their own country by foreigners:</p> <p>“(‘Deutsche Stimme’, ‘I_t117’, ‘Comment’, ‘Ä?ber diesen Vorfall hÄ?rt man wie immer nicht viel. Wieviele Deutsche wurden von AuslÄ?ndern im <b>eigenen Land</b> totgeprÄ?gelt, totprÄ?geln zÄ?hlt nicht als Mord.(...)’, ‘Odin’, ‘2013-03-13T18:18:12+00:00’)”</p>
volk	politik	6.6	<p>People and politics and interconnected somehow like shown in line 111293. The excerpt tells that politics has a task to serve the people and the interests of the people. So do economics:</p> <p>“(‘Deutsche Stimme’, ‘I_t1138’, ‘Comment’, ‘Ich kann mich Fr. SchÄ?ssler, dem Kritiker und Herbert oben nur anschlieÄ?en. Ä Herbert schrieb: “WÄ?rden die Gewinne nicht an das Kapital, sondern an das <b>Volk</b> fließen, gÄ?be es auch annehmbare Renten.” Ä Daran allein sieht man schon die Pervertierung von <b>Politik</b> und Wirtschaft. Beide haben die Aufgabe bzw. die Pflicht!, dem <b>Volk</b> (...) zu dienen bzw. seine Interessen zu wahren! (...)’, ‘Nils’, ‘2012-12-11T05:33:53+00:00’)”</p>
demokratie	bürger	6.4	<p>In a democracy (Demokratie), citizens (Bürger) are important because they are the ones who vote for politicians. The following commentary from line 111261 claims that there is no real democracy anymore because of the debts in the financial crisis. The citizens have allegedly lost their faith in the politicians, which will be difficult to restore:</p> <p>“(‘Deutsche Stimme’, ‘I_t1136’, ‘Comment’, ‘(...) Die machen doch was sie wollen. <b>Demokratie</b> gibt es nicht mehr. Und dennoch sind wir die 80 millionen die es eigentlich in der hand haben und nicht diese verlogenen intriganten die nur nach geld stinken und denen nur noch das geld wichtig ist. Es muß irgendwann sich wirklich was ändern, denn so geht das nicht weiter. Es muß das vertrauen der <b>Bürger</b> wieder hergestellt werden, doch das wird nicht leicht bei den Schuldenberg. Liebe Grüße an alle die noch an Ihr Land glauben und die Tradition unserer Großeltern und dem Zusammenhalt.(...)’, ‘Müller’, ‘2012-12-26T14:11:19+00:00’)”</p>
volk	politiker	6.3	<p>The politicians (Politiker) represent the people. The following comment says that Germans are unemployed and without hope. The only hope is the NPD, according to the quoted author, because the political system and politicians (Politiker) are bad. He/she claims that the German people (Volk) did not deserve this. The comment seems to be written by</p>

			<p>Udo Voigt, former leader of the NPD.</p> <p><i>“(‘Deutsche Stimme’, ‘1_t133’, ‘(...) Als letzte Hoffnung bleibt in diesem Land nur die NPD. Millionen Deutsche sind in diesem Land nicht nur arbeits-, sondern hoffnungslos. Diese Millionen könnten ja mal aufwachen, darum will man jetzt die NPD verbieten. Unser Volk hat ein solches System und solche Politiker nicht verdient. Denken wir auch in diesen Tagen daran, daß der Kampf für ein besseres Deutschland noch lange nicht zu Ende ist! Weihnachten 2012 Dipl.sc.pol. Udo Voigt Bild: Gabriela Neumeier / pixelio.de’, ‘admin’, ‘25. Dezember 2012’)”</i></p>
deutscher	ausländer	6.1	<p>A German (Deutscher) and a foreigner can both live in Germany if they work and do not abuse the social benefits. Foreigners can come to Germany according to “Steinar” if they are not too many. They should have education, like medicine as the author mentions. Line 110222:</p> <p><i>“(‘Deutsche Stimme’, ‘0_t77’, ‘Comment’, ‘(...)1.sofort stop des bildungsfernen sozialkssenzuzuges aus c-staaten wie libanon,albanien,rumänien...ggf ein doktor ist o.k. 2.sofortiges einstellen aller sozialleistungen der ausländern die hier nur herummarodieren.du als <b>deutscher</b> bekommst dort auch kein geld geschenkt. 3.alle kriminellen wie dieser in den medienauftauchende libanesenclan rausschmeissen.dummfreche jungausländer in diversen schulklassen gleich mit rausschmeissen. was will ich? was will ich wie jeder andere der klar denken kann sagen? es darf <b>ausländer</b> in deutschland geben aber nur in geringer zahl und nur bestimmte. (...), ‘steinar’, ‘2013-02-22T14:05:38+00:00’)”</i></p>
politik	land	5.9	<p>From 108340 we a statement it is claimed that what is most important is that correct politics (Politik) is put into action in the author’s country (Land, which is Germany). He thinks that (foreign) criminals should not be treated with silk gloves, which means that they should not be treated in a polite and naïve manner, but get the punishment they deserve:</p> <p><i>“(‘Deutsche Stimme’, ‘0_t19’, ‘Comment’, ‘(...) Unterm Strich geht es doch einfach nur darum richtige <b>Politik</b> in diesem <b>Land</b> zu machen. Und die wahren Verbrecher, die wahren Verantwortlichen all unserer Probleme sind doch diejenigen, die zum Beispiel zugelassen haben, dass Deutschland in großen Teilen überfremdet, die zugelassen haben,</i></p>

			<i>dass Kriminelle mit Samthandschuhen angefasst werden usw. Das sind unsere wahren Feinde und das sind nun mal keine Muslime. Übrigens wurde ich vorhin anonym angerufen. (...)', 'UnerkannterGeheimdienstler0815', '2013-03-13T14:48:10+00:00')"</i>
letzten	jahren	5.7	“In den letzten Jahren” = “In the last years”. There are lots of word pairs in the list that are neighbours and represent words that often belong together in the German language. Other examples: “ganze” + “land” (odds ratio 5.8) from “das ganze Land”. “gibt” + “viele” from “es gibt viele...” (5.8)

#### 4.20.4 Turn to Islam

In word co-occurrence analysis in Table 46 we have included the words that are mentioned 10 times or more in the forum Turn to Islam.

**Table 46: Co-occurring words in Turn to Islam**

Word 1	Word 2	Odds ratio	Explanation
originally	posted	84460	“originally posted by [author]” is automatic text which is generated when quoting another message in the forum. The forum is really English, so therefore finds this pair also in the German forum section we analyzed.
answer	question	4388	There are some texts both in English in the German forum section. Sometimes complete messages are in English, and other times they include a quote from an external English source or from another forum member who writes in English. In the following excerpt, a German author refers to an external article in English. The article includes a question about masturbation and an answer to that question. That article writes that masturbation is forbidden in Islam according to the Quran. A man can only have sexual intercourse with his wives and slaves. Today this means in practice only with his wife, since monogamy is common practice amongst most muslims today, although polygamy is technically allowed in Islam. Line 1692: “(‘Turn to Islam’, ‘p9-t6’, ‘Untitled’, ‘Selbstbefriedigung ist nicht erlaubt, weder in Ramadan, nachts, morgens oder abends, und hier sind ein Paar Artikel darüber, die sind zwar auf

			<p>Englisch, aber man braucht nicht viele Kenntnisse um den Inhalt zu verstehen: Ruling on masturbation and how to cure the problem</p> <p><b>Question:</b> I have a <b>question</b> which I am shy to ask but another sister who has come to Islam recently wants an <b>answer</b> to and I do not have an <b>answer</b> (with dilals from the Qur''an and Sunnah). I hope you can help and I hope Allah will for give me if it is inappropriate but as Muslims we should never be shy in seeking knowledge. Her <b>question</b> was "Is it permissible in Islam to masturbate?". May Allah increase us all in knowledge. <b>Answer:</b> Praise be to Allaah. Masturbation (for both men and women) is <u>aram</u> (forbidden) in Islam based on the following evidence: First from the Qur'aan: Imam Shafi'i stated that masturbation is forbidden based on the following verses from the Qur'aan (interpretation of the meaning): "And those who guard their chastity (i.e. private parts, from illegal sexual acts). Except from their wives or (the captives and slaves) that their right hands possess, - for them, they are free from blame. But whoever seeks beyond that, then those are the transgressors." 23.5-7 Here the verses are clear in forbidding all illegal sexual acts (including masturbation) except for the wives or that their right hand possess. And whoever seeks beyond that is the transgressor. (...) Sheikh Muhammed Salih Al-Munajjid  <a href="http://www.islamqa.com/index.php?ref...t=masturbation">http://www.islamqa.com/index.php?ref...t=masturbation</a>, 'ayman', '09-25-2007, 10:04 PM'"</p>
marry	wives	3327	<p>Looks like polygamy, marriage of a man and more than one woman is discussed. According to line 1534, where quote from an English article is given, a man can marry up to four women if he is able to treat all of them fair and equally, else only one wife is allowed:</p> <p>"('Turn to Islam', 'p7-t2', 'Untitled', '(...) According to Islamic sharee'ah, a man is permitted to <b>marry</b> one, two, three or four wives, in the sense that he may have this number of wives at one time. It is not permissible for him to have more than four. This was stated by the mufasssireen (commentators on the Qur'aan) and fuqaha' (jurists), and there is consensus among the Muslims on this point, with no differing opinions. It should be noted that there are conditions attached to plural marriage: 1 – Justice or fairness. Allaah says (interpretation of the meaning): "but if you fear that you shall not be able to deal justly (with them), then only one" [al-Nisa' 4:3] This aayah is indicates that just</p>

			<i>treatment is a condition for plural marriage to be permitted. If a man is afraid that he will not be able to treat his wives justly if he marries more than one, then it is forbidden for him to <b>marry</b> more than one. What is meant by the justice that is required in order for a man to be permitted to have more than one wife is that he should treat his wives equally in terms of spending, clothing, spending the night with them and other material things that are under his control. (...) Al-Mufasssal fi Ahkaam al-Mar'ah, part 6, p. 290 Islam Q&amp;A Sheikh Muhammed Salih Al-Munajjid', 'ibn azem', '04-12-2008, 04:05 PM')"</i>
youtube	watch	3273	On Youtube, one can watch videos
erstens	zweitens	2745	Used in German forums when mentioning argument in a given order. They mean "first" and "second"
polygamy	society	2220	<p>Following excerpt in a quote from an external English article. The German introduction is removed. The answerer of the question claims that there are more women than men in the society. This must then lead to some women not getting married if only monogamy is allowed, and this will again lead to some women prostituting themselves. This will be harmful for the society.</p> <p>Line 1534:  <i>"('Turn to Islam', 'p7-t2', 'Untitled', '(...) The ruling on plural marriage and the wisdom behind it Question: I was really into becoming a Muslim. I came to this site to find out how to become Muslim, on doing so I found out alot about the religion I never knew before, and it's kind of disturbing and almost a let down. I'm sorry I feel like that but it's true. One of the things that bother me is the <b>polygamy</b> thing, I would like to know where it addresses that in the Holy Qu"ran, please try to give me tips on how to live like that and remain sane? Answer: Praise be to Allaah. Allaah concluded His Message to mankind with the religion of Islam, and He tells us that He will not accept any religion other than that. Allaah says (interpretation of the meaning): "And whoever seeks a religion other than Islam, it will never be accepted of him, and in the Hereafter he will be one of the losers" [Aal 'Imraan 3:85] Your backing away from the religion of Islam is considered to be a loss for you, and a loss of the happiness that awaited you, had you entered Islam. You should hasten to enter Islam, and beware of delaying, for that delay may lead to regrettable consequences. With regard to what you mention about the reason</i></p>

			<p><i>for your backing off being the idea of plural marriage [polygamy or polygyny], we will present to you the ruling on plural marriage in Islam, and then the wisdom and noble purposes behind it. 1 – The ruling on plural marriage in Islam: The shar’i text which permits plural marriage is: Allaah says in His Holy Book (interpretation of the meaning): “And if you fear that you shall not be able to deal justly with the orphan girls then marry (other) women of your choice, two or three, or four, but if you fear that you shall not be able to deal justly (with them), then only one or (slaves) that your right hands possess. That is nearer to prevent you from doing injustice” [al-Nisa’ 4:3] This is a Qur’aanic text which shows that plural marriage is allowed. (...) 2 – Statistics show that the number of women is greater than the number of men, if each man were to marry just one woman, this would mean that some women would be left without a husband, which would have a harmful effect on her and on <b>society</b>: The harmful effect is that she would never find a husband to take care of her interests, to give her a place to live, to spend on her, to protect her from haraam desires, and to give her children to bring her joy. This may lead to deviance and going astray, except for those on whom Allaah has mercy. With regard to the harmful effects on <b>society</b>, it is well known that this woman who is left without a husband may deviate from the straight path and follow the ways of promiscuity, so she may fall into the swamp of adultery and prostitution – may Allaah keep us safe and sound – which leads to the spread of immorality and the emergence of fatal diseases such as AIDS and other contagious diseases for which there is no cure. (...) Islam Q&amp;A Sheikh Muhammed Salih Al-Munajjid, 'ibn azem', '04-12-2008, 04:05 PM'”</i></p>
instrumente	musikinstrumente	2218	<p>These words occur in the same message because it takes too much time to write the word “music instruments” (Musikinstrumente) each time. From the second time they are just referred to as “instruments” (Instrumente)</p>
ehevertrages	eheschluss	2218	<p>“marriage contract” (Ehevertrag) and “wedding” (Eheschluss) obviously belong together. A couple has to sign a contract when marrying each other, which legally confirms that they are married. “Ehevertrages” is the genitive form of the noun “Ehevertrag”. The following excerpt from line 1693 describes how marriage works in Islam. This forum message (which is really very long), states</p>

			<p>that forced marriage is not a valid marriage in Islam. The woman is regarded as the central person in a marriage and can only marry if she wants to. The groom does not even have to be on site during the wedding ceremony, but the bride has to. A representant of the groom must be there in such a case:</p> <p><i>“(‘Turn to Islam’, ‘p9-t7’, ‘Die islamische Position zur <b>Zwangsheirat</b>’”, ‘(...) In diesem Beitrag nun wird anhand einer Frage-Antwort-Form dargelegt werden, dass jedwede Art von Zwangsausübung, Drohung etc. vom klassischen islamischen Eherecht abgelehnt wird und es also eine <b>islamische Zwangsheirat</b> per se nicht gibt. (...) Bezeichnenderweise können sich der Wali oder auch der Ehemann durch einen Sachverwalter (Wakil) vertreten lassen (etwa wenn der Vater der Braut in einem anderen Land lebt und einen Wakil ernennt, der zusammen mit der Braut, dem Bräutigam und den Zeugen den <b>Eheschluss</b> vornimmt). Die Frau aber (sowie praktischerseits auch die Zeugen) kann niemals während des <b>Ehevertrages</b> durch einen Vertreter oder eine Vertreterin ersetzt werden - sie ist die zentrale Person, und von ihrer Zustimmung ist die Gültigkeit der Ehe abhängig. (...) Abdurrahman Reidegeld 04.12.2004 Islamische Zeitung’, ‘ibn azem’, ‘09-04-2007, 02:06 PM’)”</i></p>
eheschluss	zwangsheirat	2218	<p>“wedding” is also discussed together with “forced/arranged marriage”. See example for “ehevertrages” and “eheschluss”.</p>
arabien	saudi	1825	<p>The country “Saudi-Arabien” or “Saudi Arabia” is read as two words by the analysis program.</p>
masturbation	fasting	1659	<p>Very disturbing colocation of words, but one must notice that these words are only three times in the same message. The surprising combination of “masturbation” and “fasting” is originally from a forum message about things one cannot do during Ramadan, the fasting month. One cannot eat at daytime (only at night-time), but it seems to be even worse to masturbate in exactly that month: Line 1693:</p> <p><i>“(‘Turn to Islam’, ‘p9-t6’, ‘Untitled’, (..) Praise be to Allaah. No doubt <b>masturbation</b> is haraam according to most of the scholars, as Shaykh al-Islam Ibn Taymiyah (may Allaah have mercy on him) said. If this happened in Ramadaan, this is worse, and if there was ejaculation of sperm because of the <b>masturbation</b>, the violation of the</i></p>



			<p><i>sanctity of the fast is even more severe in sin. From the time the maniy (semen) is ejaculated, the fast is broken, but the person should still fast for the rest of that day. (...), 'ayman', '09-25-2007, 10:05 PM')"</i></p> <p>There is also an example for the word "masturbation" in the second row of this table, where the usage of "question" and "answer" is exemplified.</p>
--	--	--	---

## 4.20.5 Ummah

In the following analysis we have included the words that are mentioned 100 times or more in the forum Ummah.

**Table 47: Word co-occurrences in Ummah**

Word 1	Word 2	Odds ratio	Explanation
wahhabism	terrorists	1809.2	<p>Wahhabism is a direction of Islam that exists mainly in Saudi Arabia [43]. The following excerpt tells that Wahhabism is in crisis, because the majority of the terrorists in the terrorist attack on 11 September 2011 were from Saudi Arabia. We understand that Wahhabism does <b>not</b> want to be associated to terrorism. Wahhabism seems to be the state religion in Saudi Arabia, although just a large minority of 40 % seem to belong to that confession.</p> <p>Line 75710:  <i>"In other such groups, like the American Muslim Council (AMC) and the Muslim Students Association (MSA) <b>Wahhabism</b> is in crisis, because of the devastating effect of 9/11. In addition, the Wahhabis are deeply compromised by the exposure of individuals like John Walker Lindh, Richard Reid, José Padilla, and John Muhammad. Lopez: Why were 15 of the 19 9/11 <b>terrorists</b> from Saudi Arabia? Schwartz: For three reasons. First, although no more than 40 percent of Saudi subjects, at the most, consider themselves Wahhabis, the Wahhabi clergy has controlled education in the kingdom, so that all subjects have been raised in an atmosphere of violent hatred for other Islamic traditions and for the other faiths."</i></p>
wahhabism	traditional	625.5	Traditional Muslims most likely confess to

			<p>Sunni and Shia, directions of Islam much older than Wahhabism.</p> <p>Line 75722:  <i>“The fact that the U.S. political and media elite have done almost nothing to enable <b>traditional</b> Muslims in this country to oppose <b>Wahhabism</b> makes the situation that much worse”.</i></p>
president	washington	421.5	<p>The president of the United States resides in the city Washington, D.C, and so do the Department of State, a federal office for foreign affairs.</p> <p>Line 106464:  <i>“The Global 2000 Report to the <b>President</b>: Global Future: Time to Act, prepared by the Council on Environmental Quality and the U.S Department of State, <b>Washington</b>”.</i></p>
operations	bomb	404.3	<p>In some military operations, there can be bombing of strategic targets. In the following excerpt the Japanese attack on the American military base Pearl Harbor in Honolulu, Hawaii is discussed. It is interesting to see that American war history is told in an Islamist forum.</p> <p>Line 75912 :  <i>“In late October, Haan finally convinced US Senator Guy Gillette that the Japanese were planning to attack in December or January. Gillette alerted the State Department, Army and Navy Intelligence and FDR personally. 24 September 1941, the "<b>bomb</b> plot" message in J-19 code from Japan Naval Intelligence to Japan" s consul general in Honolulu requesting grid of exact locations of ships pinpointed for the benefit of bombardiers and torpedo pilots was deciphered. There was no reason to know the EXACT location of ships in harbor, unless to attack them - it was a dead giveaway. Chief of War Plans Turner and Chief of Naval <b>Operations</b> Stark repeatedly kept it and warnings based on it prepared by Safford and others from being passed to Hawaii.”</i></p>
secondly	firstly	403.1	<p>The following example is from an argument which lists two statements sequentially.</p> <p>Line 91945:  <i>“<b>Firstly</b>, that ?general will? may contradict the ?Divine will? because those who are making the laws may or may not refer to the Divine laws.</i></p>

			<i>Secondly, the "general will" may not be the real "general will" as characterized by Rousseau because those who are making the laws may not be necessarily good people".</i>
president	director	388.2	<p>"president" and "director" are both leader titles, and it makes sense that these words co-occur in a report of a meeting where different leaders meet.</p> <p>Line 96912:  <i>"The <b>president</b> of NOCIRC, the <b>president</b> of NOHARMM 200, two other members of NOCIRC, a Jewish nurse, who is a conscientious objector who refuses to participate in circumcision, and myself (a Christian of Palestinian origin) were selected. The discussion lasted about 45 minutes. The <b>director</b> was very nice and allowed each a chance to speak."</i></p>
wahhabism	ideology	382.5	<p>Wahhabism is a form of Islam which seems to be very conservative and may by somebody seem intolerant for people standing outside that group, whereas ideology is a body of ideas, beliefs that reflects some group or social movement. The following mentions where the ideology of wahhabism exists.</p> <p>Line 75710:  <i>"Outside the Peninsula, <b>Wahhabism</b> is generally unpopular. But where trouble is found, Wahhabism may thrive. Hamas in Israel represents pure Wahhabism. Forms of neo-Wahhabi or Wahhabized <b>ideology</b> have been powerful in Egypt (the Muslim Brotherhood) and in Pakistan ? in both countries neo-Wahhabis lead attacks on other Muslims and other faiths."</i></p>
wahhabism	arabia	376.9	<p>As mentioned earlier, Wahhabism is a direction of Islam that exists mainly in Saudi Arabia [43].</p> <p>Line 75711:  <i>"When the Saudis needed to clear the Grand Mosque in Mecca of protestors in 1979, they employed French paratroops to kill Muslims within the walls of the mosque. Lopez: How widespread is it? Schwartz: <b>Wahhabism</b> is official in Saudi <b>Arabia</b>."</i></p>
administration	Washington	371.7	<p>It looks like American history is discussed in this Islamist forum. Here is some discussion about the attack on Pearl Harbor, Hawaii. It criticizes what the American administration in</p>

			Washington did wrong that time. Line 75912 : <i>“It was lack of information like this that lead to the exoneration of the Hawaii commanders and the blaming of <b>Washington</b> for unpreparedness for the attack by the Army Board and Navy Court. At no time did the Japanese ever ask for a similar bomb plot for any other American military installation. Why the Roosevelt <b>administration</b> allowed flagrant Japanese spying on PH has never been explained, but they blocked 2 Congressional investigations in the fall of 1941 to allow it to continue.”</i>
administration	agencies	315.0	Both words refer to parts of a government. This section discusses missing and bad communication between the administration and the federal agencies in the American government, which resulted that the attacks on 11 September, 2011 were not prevented. Line 91972: <i>“Within a year of the terrorist attack, the nation"s leader determined that the various local police and federal <b>agencies</b> around the nation were lacking the clear communication and overall coordinated <b>administration</b> necessary to deal with the terrorist threat facing the nation, particularly those citizens who were of Middle Eastern ancestry and thus probably terrorist and communist sympathizers, and various troublesome "intellectuals" and "liberals.”</i>
court	fbi	307.9	Both the FBI and the FISA Court (United States Foreign Intelligence Surveillance Court) [55] seem to be parts of the American Government system. The former is responsible for investigation domestic crimes and the latter is a kind of court.  Line 106472: <i>“In March 2001, an internal-debate ignited at the Justice Department and the <b>FBI</b> over wiretap surveillance of certain terrorist groups. Prompted by questions raised by Royce C. Lamberth, the Chief Judge of the <b>FISA Court</b>, the Justice Department opened an inquiry into Michael Resnick, an FBI official who coordinated the Act"s applications.”</i>
operations	forces	302.9	Military forces may put operations into action in some kind of war situation. Line 75812 : <i>“The investment by Western/American Oil</i>

			<i>companies is leading to killing of many in Afghanistan by Turkmen and Tajiks in current crisis in 2001. 36. UGUNDA: From 1971 to 1979, several thousand Muslim lost their lives when British American <b>forces</b> launched covert <b>operations</b> to oust Idi Amin.”</i>
infidel	mockery	288.5	<p>The word “infidel” means a person who does not believe in a religion. It is especially used in Christianity and Islam. Here the word “infidel” has been used in the name of a movie, and “mockery” seems to be used as a commentary for that movie, because the author thinks the movie makes fun of Islam in some kind of scornful way.</p> <p>Line 90097 :</p> <p><i>“(Ummah', '256494-The-Infidel-Movie-a-mockery-or-just-comedy%2Fpage2', 'Re: The <b>Infidel</b> Movie - a <b>mockery</b> or just comedy?', 'Originally Posted by Umm Zubayr Asalam alaykum, a friend sent me the trailer of "the <b>infidel</b>", a comedy film about a muslim man who discovers he is actually jewish. i have a few muslim friends who think it looks like harmless fun but it looks like a <b>mockery</b>. i know ummah forum is diverse, and i wanted to know what you think..</i></p> <p><i><a href="http://www.youtube.com/watch?v=BMudFOMQgC0">http://www.youtube.com/watch?v=BMudFOMQgC0</a> This film has been banned in Israel because it mocks the Jewish religion. Thought id give my two pence worth', 'Hijabistani', '17-04-10, 10:59 PM'),”</i></p>
bottle	breastfeeding	281.0	<p>These two words could be often observed in the general information on mother’s breastfeeding in early days, such as how to get rid of the problems during that period. Both words refer to ways of feeding a baby.</p> <p>Line 83109 :</p> <p><i>“If they know you're very committed to <b>breastfeeding</b>, and want their support, they're more likely to encourage you, rather than tempting you to try <b>bottle</b> feeding.”</i></p>
administration	federal	276.7	<p>These two words are part of a name of an authority called Federal Aviation Administration, which is the national aviation authority of USA.</p> <p>Line 104963:</p> <p><i>“Seven to eight weeks prior to the 11th September attacks, all internal U.S. security agencies were warned of an impending Al-</i></p>

			<p><i>Qaeda attack against the Untied States that would likely occur in several weeks time: ?Meanwhile, intelligence had been streaming in concerning a likely Al Qaeda attack. ?It all came together in the third week in June,? Clarke said. ?The C.I.A.?s view was that a major terrorist attack was coming in the next several weeks??.? On July 5th, Clarke ?summoned all the domestic security agencies?the <b>Federal Aviation Administration</b>, the Coast Guard, Customs, the Immigration and Naturalization Service, and the F.B.I.? and informed them ?of an impending attack.?”.</i></p>
operations	planned	240.8	
Iraq	troops	240.6	<p>We see from the following sample that the USA has military troops in Iraq, which former president Bush. The Americans do not want to watch to Iraq war on TV anymore, but they wear ribbons encouraging people they meet to support the troops in Iraq, likely economically. The quote below says that these supports do not know about the destructions the war has done in Iraq.</p> <p>Line 91234 :</p> <p><i>“Of course you may be suprised in that you might actually agree with Hitler on some things! flippantly toss around hitlers name? hitler was a hero to the germans, they placed him in power, they went on with their lives and left it to hitler to impose his methodology on the world it wasnt until they fell flat on their faces did the germans regret hitler i see a lot of similarities to hitler &amp; the third reich and bush &amp; neo-cons the americans dont even want to watch the <b>iraq</b> war on the telly any longer, but they paste these yellow ribbons speaking of how they "support their <b>troops</b>" oblivious to the path of destruction bush has embarked, the dollar is dropping, american prestige is dropping, many are seriously discussing war crimes for bush, and the americans watch the super bowl too many correlations to ignore’,”</i></p>
polygamy	divorce	22.2	<p>Following post is really a repost/quote of another writer’s message. The quoter author thinks that polygamy should be allowed, since the society, as it is now, forces to man to take just one choice, which are staying with the wife he has or divorce and marry a new one. The author asks why a such man could not stay in a monogamous marriage with his first wife and</p>

		<p>just be happy with that. Line 103293:  <i>“(Ummah', '352930-Give-quot-ugly-quot-girls-a-chance', 'Re: Give "ugly" girls a chance', 'Originally Posted by ??? They would be deluded if they believed that and were against polygamy, they would not be deluded however if society's pressures have forced them into that in the first place. For example, a man is already married. He wishes to take on another wife, but his first wife with whom he has children threatens to divorce him if he does. She is a sinner, and more than that he should not marry to avoid the fitnah on the children. That is a man who was pressured. Also when a man marries another woman, why must you think "because he thinks she's prettier". Is he not going to go through the same process anyway and determine he's attracted to the girl? If so, he would think her to be pretty, be it prettier or less so, the point is he found her acceptable and thus married. Nothing else needs to be noted. And again, when society has forbidden this practice, well we see the fruits now. Marriage as a whole is more difficult, and indeed society pressures one into making one choice and one choice only, and that means less of the compromising you are asking for in the opening post. I agree that polygamy is halal. However, some make it sound as if a man has to either choose between committing adultery, marrying another wife or staying monogamous and suffering. Can't he just be monogamous and be happy, in love and grateful for his wife? How hard is that?', 'DaughterOfAdam', '30-01-13, 04:52 PM)'”</i></p>
--	--	--

#### 4.20.6 Stormfront

For the following analysis in we have included words that are used 400 times or more in the forum Stormfront.

**Table 48: Word co-occurrences in Stormfront**

Word 1	Word 2	Odds ratio	Explanation
fil	chick	27929.6	<p>Line 25494:  <i>“(Stormfront', '902741', 'Re: JEW SOB Rahm Emanuel Vows To Fight "Chik-Fil-A Over Anti-Fruitcake Stance!', 'i am about to walk out the door ,I will read the article when i get back</i></p>

			<p><i>home. I wanted to say though . I LOVE ME SOME <b>CHICK FIL A</b> ! When you live in a 63 percent black city , it is very hard to get great service wherever you go but at <b>CHICK FIL A</b> ? WHite white white ! QUICK service , POLITE SERVICE , NO WRONG orders , and good food .', 'Popper504', '07-26-2012, 05:29 AM')”</i></p> <p>Chick-Fil-A is a fast food restaurant chain in the USA with headquarters in Atlanta, Georgia. The word is split up at the hyphens by our analysis program. The author “Popper504” seems to be a white person from a city where the majority is dark-skinned. He/she blames restaurants run by black people for giving him bad service. Therefore, he/she prefers to eat at a Chick-Fil-A restaurant, where he/she finds the service good and quick. “Popper504” seems to hate black people because he/she experiences them as impolite and sloppy.</p>
encyclopedia	Wikipedia	2726.8	<p>That Wikipedia is an online encyclopedia must be widely known. In several places in the forum there are quotes with “Wikipedia, the free encyclopedia” as source.</p>
dawn	golden	2416.4	<p>Golden Dawn is a Greek right-extreme political party which has got 18 seats in the Greek parliamentary and had a clearly racist, political action where Greeks were urged to donate blood only for Greeks like shown in line 903515: “<i>Stormfront</i>’, ‘903515’, ‘<b>Golden Dawn Wants Blood Only For Greeks</b>’, ‘<i>This article is from the beginning of the month but I just found it and don't recall seeing it posted on here. Quote: ATHENS ? Sitting in Parliament for only a couple of weeks, the neo-Nazi <b>Golden Dawn</b> party has expanded its anti-immigrant campaign to begin a drive to get Greeks to donate blood only for Greeks, outraging doctors and health care officials who said it was racist and would not be allowed. Reminiscent of Nazi Germany practices to keep ?pure? blood lines, <b>Golden Dawn</b> has put up posters around the city calling for volunteers to donate blood ?only for Greeks who need our help.? (...) One of the biggest doctors? unions in Athens said the <b>Golden Dawn</b> initiative was an ?insane, unscientific, illegal and racist action? and promised to do everything necessary to protect the ?sacred procedure of blood donation.? (...)</i>”</p>



			<i>Golden Dawn got 18 seats in the Parliament based on a virulent campaign to rid Greece of all immigrants, and some of its members have been tied to vicious gang assaults on immigrants, a favorite tactic. (...) Golden Dawn' "</i>
Mitt	Romney	1390.7	Mitt Romney was the presidential candidate for the Republican Party in the election in 2012. The writer "Oldsalt97" does not seem to like him, since Romney declares Jerusalem the capital of Israel. Ron Paul, another republican candidate is also not an option. Example from line 29208: "('Stormfront', '903624', 'Re: Romney declares Jerusalem capital of Israel', 'Clearly he is running for the Israeli presidency and not POTUS. We have only a few other choices this November and the two parties aren't it. Ron Paul is not the answer either, but neither is not voting. (...)', 'Oldsalt97', '07-30-2012, 05:58 AM)'"
Ron	Paul	827.8	Ron Paul was one of the many possible presidential candidates for the Republican Party before Mitt Romney finally became the candidate. See example for the pair "Mitt" and "Romney"
Barack	Obama	819.7	Barack Obama is the president of the United States while this thesis is written. The Stormfront members do not seem to like him. In line 29842 he is accused of telling public schools to ignore bad behaviour by black students. This can be a misunderstanding. The context may be that black students are often punished for bad behaviour when white students are not, and Obama therefore meant to say that black students must be treated better. This is of course speculation from our side, but the following is anyway odd: "('Stormfront', '903763', 'Obama asks public schools to ignore bad behavior by black students', 'Obama asks public schools to ignore bad behavior by black students <b>Barack Obama</b> 's "African American Education Initiative" creates a new Federal bureaucracy. One of its goals will be to stop disciplinary action against black public school students who misbehave Obama asks public schools to ignore bad behavior by black students - Charleston Charleston Conservative   Examiner.com', 'Patriot114', '07-30-2012,

			06:05 PM)''
Hitler	Adolf	481.6	Adolf Hitler is former dictator of Germany. Line 30249 he is praised by "FABIAN4" for being a saviour for Germany after an economic crisis: <i>"('Stormfront', '903862', 'Re: A new Confederacy', 'I think maybe the only way to start a new government National Socialist is when people get fed up with the current one like what's happening now in the U.S. and the economy crumbles by the Jewry Zionist then people see the truth more clearly, because they start asking questions. Would have to be like what happened to Germany before <b>Adolf Hitler</b> saved the people and Nation! History repeats and Jewry will come to a close as it always has!', 'FABIAN4', '07-31-2012, 01:08 PM')"</i>
minister	prime	463	The high odds of these two words must be due to the compound word "prime minister".
encyclopedia	free	434.3	These two words must be closely related because of the many quotes with "Wikipedia, the free encyclopedia" as source.
camps	concentration	353.0	Concentration camps were camps where prisoners of war and enemies of the state were held captive by the German army during the second world war.
Holmes	James	347.6	We often see examples of a first name associated with a last name in the list of associative word pairs. James Holmes shot and murdered a lot of innocent people in the premiere of the Batman movie "The Dark Knight Rises" in Colorado. That massacre is in the quote from line 35129 compared to a later one: <i>"('Stormfront', '904985', 'Re: Here we go again: Mass Shooting At Sikh Temple In Oak Creek, Wisconsin', 'Let's hope it wasn't a White perp although it appears it may be. This is not good following the James Holmes massacre. (...)', 'Proud White Chap', '08-05-2012, 02:48 PM')"</i>
Roosevelt	Churchill	276.5	Roosevelt and Churchill were the president of the United States and the prime minister of the United Kingdom during the Second World, respectively. They are therefore often mentioned together.
camp	concentration	264.6	Very similar to the pair of "camps" and "concentration".
negri	bianchi	174.0	"black" and "white" in Italian. These

			<p>oppositions are natural to find together in a right-extremist forums were racist topics often are discussed.</p>
Japanese	Japan	170.0	<p>A person from Japan is of course called a Japanese. Therefore these words are often used together. The author “AngryGoy” does not seem to hate Japanese, and as far as we know, right-extremists do not hate Asian people (except from Jews if one counts the Middle East as Asia). Japan was an ally of Germany and Italy during the Second World War. The message author seems to cheer on Japan rather England since Japan only have indigeneous players in its national football team, while there are ethnic Africans in the English team. It does not seem to help for him/her whether coloured people are grown up in England and are British citizens. Line 37993:</p> <p><i>“(‘Stormfront’, ‘904938’, ‘Re: Great Britain Loses to South Korea after Black Player’s Missed Penalty Kick’, ‘I was pulling for South Korea over England that game, the same way I was pulling for <b>Japan</b> over France in the women’s semi final. I respect the fact that these Asian countries only allow their true genuine <b>Japanese</b>, Koreans, Chinese to represent their respective countries. I just HATE seeing Negroes and other non-White supposedly represent countries like England and France. It is my hope and dream to see ONLY White Europeans represent European countries at the next Olympics in Rio. BTW, Seeing 3 Asians and 3 Whites make all their penalty kicks and having the Negro interloper miss was AWESOME!!!’, ‘AngryGoy’, ‘08-12-2012, 02:58 AM’)”</i></p>
Colorado	Holmes	135.8	<p>An author with a nickname of question mark has a strange comparison of a Jew called James Holmes who killed 14 random persons in a cinema, while a white racist called Wade Page killed “only” 6 in a sikh temple. A Jew is therefore allegedly “worse” because managed to kill more people during the assault. That kind of ranking seems odd to ordinary people like we who write this thesis. To murder people should be considered bad, no matter how many persons one kills, we think.</p> <p><i>Line 37521:</i></p> <p><i>“(‘Stormfront’, ‘905620’, ‘Re: Temple killings</i></p>

			<i>put spotlight on "hate rock", 'This guy(Wade Page) was a gift to the left.He took the heat off the Jewish guy in Colorado(James Holmes) and they all really couldn't be happier.Racist kills 6,trumps nut job who kills 14.Numbers vs ideology-there is no contest.One or two more like Page and Holmes could become a hero,that's how twisted and corrupt they are.', '?????? ??????????', '08-08-2012, 11:23 AM)'”</i>
Soviet	union	135.1	The Soviet Union was a communist state which consisted of Russia and other countries in Eastern Europe and Central Asia.
democrats	republicans	129.3	Democrats and Republicans are the two biggest political parties in the USA. They are therefore often mentioned together.
candidate	presidential	119.2	A presidential candidate is a person who makes himself/herself eligible for the position as federal president in a republic. Here is an example from the USA in line 39902: <i>“(‘Stormfront’, ‘906289’, ‘(...) Congress should have demanded Obama show his birth credentials when he declared he was a <b>presidential candidate</b> in Springfield, Ill., on Feb. 10, 2007. ‘(...), ‘Mjodr’, ‘08-11-2012, 06:59 AM)’”</i>
church	Catholic	103.8	The Catholic Church is the most widespread kind of Christianity in the world. Authors in Stormfront never stop surprising us about where the Jews appear, like when they allegedly infiltrate the Catholic church in line 40459: <i>“(‘Stormfront’, ‘906447’, ‘Re: Jewish-Born Clerics Helped Push Vatican II Reforms’, ‘The jews have been infiltrating the <b>Catholic church</b> for a long time now.’, ‘LusoAmerican’, ‘08-12-2012, 11:55 AM)’”</i>
camps	prisoners	101.3	In the last wars, it has been usual to keep prisoners in concentration camps. Normally prisoners of war are supposed of war are supposed to be treated with dignities in such camps, called concentration camps. That was not always the case for the prisoners the Nazis took during the Second World War. A Jewish prisoner who is quoted by “JamesMiller”, claims that. The Stormfront author does not believe the Jew in line 45524: <i>“(‘Stormfront’, ‘907687’, ‘Holocaust? Survivor Henry Friedman Says He Survived Seven (7) "Extermination Camps"', ‘Quote: (...) By</i>

			<p>definition, he said a concentration camp is a place where the captors are to treat <b>prisoners</b> “with dignity.” “They had extermination <b>camp</b>s,” he said. Article Link: Holocaust Survivor Speaks At Bethany Assembly - The Daily Telegram I thought Henry Friedman had died. Obviously, he's still on the payroll to peddle his Jewish written script to naive White school kids. (...), 'JamesMiller', '08-18-2012, 01:07 AM'”</p>
southern	poverty	87.4	<p>The two words are both part of the name of a non-profit organization, Southern Poverty Law Center, which helps victims of hate groups (for example victims of white supremacists). It works for tolerance among people in the USA. Line 46095:  “(‘Stormfront’, ‘907814’, ‘WASHINGTON POST: “White Supremacy Is The Greatest Danger We As Americans Face As A Source Of Domestic Terrorism”’, ‘Quote: (...) The <b>Southern Poverty</b> Law Center had tracked the Wisconsin Sikh temple shooting suspect Wade Michael Page, a 40-year-old Army veteran with links to a white supremacist movement. (...) - The Real Homegrown Terrorism .’, ‘JamesMiller’, ‘08-18-2012, 07:26 PM’)”</p>
Stalin	Soviet	87.1	<p>Josef Stalin was the leader of the Soviet Union from 1922 to 1953.</p>
Spain	Italy	86.8	<p>Spain and Italy are both countries in Southern Europe.</p>
Russian	Putin	84.6	<p>Vladimir Putin is the president of Russia.</p>
mixing	race	83.5	<p>The Stormfront forum members like “Bana phrionnsa do not seem to like race mixing, which is another word for a black and a white person to get a child together. The author claims that television shows try to deliberately convince people with different ethnicities to get children together. Example from line 50333:  “(‘Stormfront’, ‘908839’, ‘re: JC Penny ad promotes miscegenation’, ‘Is the woman white? The example is always the same however. Black guy and non-black woman. It is like the Kardashians. Those women certainly are not white but they are not black. (...) There are several race mixing propaganda shows on television. (...)The media seems to be doubling down their efforts to push <b>race mixing</b>.’, ‘Bana phrionnsa’, ‘08-23-2012, 02:05 PM’)”</p>

## Chapter 5: Discussion

From the numerous examples we have found for the different forums, we clearly see that the choice of words in an extremist forum differs from a so-called ordinary forum with topics like daily life and politics. There is also a difference between the word contents of a right-extremist forum and an Islamist forum. The most frequent words and the most frequent pairs of co-located words in a forum seem to work properly as characteristics for each of the different forums. When examining a list of the most frequent words in a forum, one can find out which topics are discussed.

When we observe the most popular words in Ummah and Islamic Web-Community forums, we can understand that there is much writing about the religion Islam. Amongst the most frequently used words are “Allah”, “Prophet”, “muslims”, “messenger”, “Quran”, “Hadith”, “fasting”, “women”, “God”, “Hajj”, and “peace”, which we have already explained in chapter 4. Sometimes, we see that these forums mention ordinary life events, both ones that are related to religion, and to family life. For instance, the forum authors discuss Islamic rituals such as fasting during the month Ramadan. It is shown that they also discuss women and their rights in different view angles. Women are claimed to have good rights in society, and wearing a kerchief is claimed by a German forum author to be a means for signalling unknown men to let a married woman walk around in peace.

When we examine the Stormfront forum, we understand it as a forum about racism, and right-extreme politics. For instance we see that the words “white”, “Jews”, “black”, “race”, “Hitler”, “Negro”, “Obama”, “world”, and “war” are popular. Islam is there sometimes mentioned as something bad, while Christianity is mentioned as something good. From many examples in Stormfront we see that so-called race theory is discussed. It looks like the authors think that white people are superior compared to other races. They also claim that Jews are bad people and that they all have an agenda of conquering the world. Jews and dark-skinned African Americans (by the Stormfront authors called Negroes) are regarded by these right-extremists as inferior compared to the so-called white master race. Barack Obama is the president of the USA. From examples in the Stormfront forum we see that the right-extremists hate him both for having coloured skin and for representing liberal politics. American politics is namely discussed in Stormfront. Republicans are not really better liked by the White Supremacists than the Democrats are. Stormfront writers seem to be even more conservative in politics than the Republicans. The words “world” and “war” are in the forum because texts about the Second World War, which is also discussed.

In the Swiss English Forum, which think of as an ordinary forum, extremist topics are practically not discussed. Instead we can read about day-to-day topics like food, drinks, school, and parties. From the contents of the forum, it seems like many of the forum writers are immigrants in Switzerland, that want to know about the society and social life in that county. Amongst the most frequently used words found we find “Switzerland”, “Swiss”, “Zurich”, “German”, “English”, “thanks”, “car”, and “live”. The two first words describe the nationality of the country this forum is about. Zürich is a city in that country, and German is the most popular mother tongue in that country. English is the language foreigners use before they learn German, French, or Italian, dependent on where they live. “Thanks” is used when the writers that ask questions are grateful for getting answers in the forum. “car” and “live”

are mentioned because people tend to need a car and a place to live.

The Steam forum about the game “Counter-Strike: Global Offensive” discusses mainly technical difficulties and cheating in the game. The extremist and ordinary forums differ clearly from each other in the extent extremism-loaded words are used. Some of the most popular words we find are “game”, “play”, “players”, “competitive”, “cheaters”, “ban”, “matchmaking”, and “team”. All these words are related to competitive online computer gaming. It is obvious that a forum about a computer game discusses how one plays that game. What is particular about this forum is that cheating and banning of people doing that are discussed.

We have found word pairs in the extremist forums Stormfront, Ummah, Turn to Islam, Deutsche Stimme, and Islamic Web-Community, and in the non-extremist forum Steam, where both words in each pair are from the same forum message, but not necessarily from the same sentence. Nevertheless, many of the most frequent and characteristic pairs actually happen to consist of word neighbours (contiguous words) in the original text. Some of these word neighbours comprise human names or compound words like “Adolf Hitler”, “Barack Obama”, “presidential candidate”, “race mixing”. There are also words that are not even in the same sentence, but still have some kind of semantic relation: “Democrats” – “Republicans”, “Stalin” – “Soviet”, “camps” – “prisoners”. The first pair consists of the two big, political parties in the USA. The second pair occurs due to the fact that Josef Stalin was state leader of the Soviet Union. The last pair originates like from some discussion of prisoners in concentration camps. The word pairs we have presented till now are from Stormfront. In Ummah we have “president” – “Washington”, because the American president has his seat in the city Washington, D.C., “wahhabism” – “Arabia”, because the that kind of Islam mainly exists in Saudi Arabia, “administration” – “agencies”, because both words describe parts of a national administration. For more examples, also from other forums, see section 4.20. We must admit that also a lot of word pairs found are just random and show no clear common meaning. This is often because these words are only in the same forum message, and not necessarily in the same sentence. We choose not to present the “false-positive” pairs, pairs which do not make sense, in this thesis.

A problem with our Java program for finding word pairs in a message runs very slow, because it had to test for a huge amount of combinations of two words. For the forums Ummah and Stormfront, there could be millions of such combinations, even after removing stop words. It seemed to run even slower for a program finding words in the same sentence, so to save time, we had to just analyse words that are in the same message. We had to drop analysing words used less than 10 times for Steam, Islamic Web-Community, and Turn to Islam, 100 words for Ummah, 200 words for Deutsche Stimme, and 400 words for Stormfront. If we had not performed these time-saving actions, the word pair finding would have needed several days to complete.

From the above-mentioned frequent words and many other words shown in chapter 4 we see they are good features for characterizing the text contents of a forum. The same we can say about the pairs of co-occurring words. Nevertheless, we need to check how a word is actually used, because a word can be used in quite different contexts. It can be namely be applied quite differently from how we immediately guess. Let us imagine that we did not know the history of Nazis and right-extremists at all. When seeing the word “Jews” used 9044 times in the Stormfront forum, we could guess that Jews are right-extremists, or that right-extremists like Jews. Of course we know that Jews are not like that, and from examples we have figured

out that many of the Stormfront writers actually hate Jews a lot. We can read about this for example in line 1864 in forum\_entries.sql:

*“(Stormfront', '897112', 'flyer33716', '(...) We can never afford to be as kindly as this gentle giant. He wished no harm to the **Jews** only to get them out of our lands. Had Britain allowed them to go to Israel or Madagascar, had the allies not deliberately bombed food supplies to the camps, we would not have seen the piles of bodies. International Jewry wanted as many dead **Jews** as they could get lined up for the world to see at the end of the war. Next time the **Jews** will be forceably removed, no camps, no waiting, just ordered out quick as you like. For those who refuse to go load them on ships and set them adrift on the seas with Jewish captains on board. That is if they do not genocide us first, and given the state of the British Nationalist parties it looks as though they will succeed.’, 'blenkarni', '07-01-2012, 02:22 PM’)*”

The excerpt does not show any clear sign that the author wants Jews killed, but he/she wants all Jews to go to a faraway land like Israel or Madagascar. From this and other examples in the report we can deduce that Jews are most likely **not** on the side of the White Supremacists or right-extremists on the Stormfront forum. This example shows that frequent words can often give good clues about what the text is about, but we need to look at how they are used in context with other words to really understand what the extremists want to express.

Hypothesis 1 is therefore proved to be true on the condition that we support the forum characterization done by means of frequent words (and words used together in the same message) with some examples from the messages for how these words are actually applied in the forum.

Hypothesis 2 is basically very similar to hypothesis 1. The difference is that with hypothesis 2 we wanted to find out whether single authors in a forum can be profiled and characterized by means of their vocabularies. We found the vocabulary of five of the ten most productive authors in Ummah, Stormfront, and Steam, respectively. It was clear that each of these authors had some unique words if we compared with all of the nine other most productive authors. If we try to find unique words for an author in the context of tens of thousands other authors in a forum, we find nothing. First reason is that stop words are not included in the analysis. The second reason is that it is very unlikely that even so few as hundred authors use some of the same words, except for stop words, of course. Not all of them write about exactly the same topic. We figured that out by some trying and failing with different numbers of authors in the set for analysis.

We can characterize each single of the five selected authors from each forum by means of non-unique words as well. We then see how words are used together in combination. In the Stormfront forum, the author “kazan188” frequently uses the words “Hitler”, “jews”, “camp”, and “Auschwitz”, which together should apply that he/she writes much about what he thinks happened with Jews in concentration camps during the Second World War. The author “Mjodr” in the same forum often uses the words “white”, “police”, “national”, and “Obama”, which imply that he/she discusses the situations for white people, and how government and politics work in the USA. One can see more examples of that in the in section 4.19.

This demonstrates that it is plausible that one can profile and characterize authors by means of the words they use frequently and how often these words are used. Words that are unique for an author are not well suited for profiling that author in a complete forum. We can namely not find author-unique words if we examine and compare too many authors at a time. If we first divide all authors into groups of 10 (or maybe 15) before profiling, it should be possible to use unique words for differentiating authors from each other in a given group. The



conclusion is that authors can be profiled on words and word frequencies, because they give good implications for what the author discusses, like in the example written shown with “kazan188” above. Unique words do not work alone. It may work in future research by first finding small groups of authors with words used in common, and nearly the same frequencies.

Hypothesis 3 seems to be likely because when we compare an extremist forum (Stormfront, Ummah, or Islamic Web-Community) with an ordinary forum (Steam or Swiss English Forum), we can in many cases easily see which words are typical and dominating for the extremist forum, and at the same time are not so prevalent in the other forum, and vice versa. We do this by means of NGTF comparison. The reader can see these results in the actual sections in chapter 4. In this chapter we show just a little of the comparison of Stormfront with the Ummah in section 4.12. Following words have been much more used in Stormfront than Ummah:

“whites” (NGTF ratio 417.7), “negro” (190.2), “blacks” (160.6), “Obama” (152.8), “nationalist” (87.5). These words are somehow related to racism. Barack Obama is president of the USA and has coloured, and is therefore hated by White Supremacists. “nationalist” is in some cases connected to racism, if a nationalist is also classified as “Nazi” or “Neo-Nazi”. Following words are much more used in Ummah than in Stormfront, and they are all without doubt connected to the religion Islam: “Allah” (NGTF ratio 228), “Quran” (216.2), “prophet” (134.6), and “Muhammad” (51.2).

By comparing word counts in the vocabularies in the Islamist forums with Stormfront, and each of these with the vocabularies in Steam and Swiss English Forum, respectively, we found the main topics discussed in the two different forums. We can tell that the members of the Islamist forums mainly discuss religion, and particularly Islam, while the right-extremists tend to discuss politics, and they spend much text on explaining why their politics is the best for the society. The Islamists use many words from Islam, which an average Muslim also could have written. Much of the contents in Ummah, Islamic Web-Community, and Turn to Islam are simply information and discussion about the religion Islam. According to authors in the Islamist forums, Islam seems to be the solution for many of the problems in the world, mainly the moral problems. In the Islamist forums, western culture is criticized for being immoral and bad in different kinds of ways. The American warfare in Iraq is as mentioned as one example for that.

The right-extremists generally claim that some European people actually support a slow extermination of themselves by genetic mixing with other races, and that they are not proud and sometimes even ashamed of their own culture. Mixing of races (children of biracial parents) and cultures is an extremely bad thing for the white supremacists, which they sometimes call themselves. According to right-extremists there is no room for different cultures in the same society. The predominating religion (for the authors who discuss religion) in the forum Stormfront seems to be Christianity, while the authors in Vigrid believe in Norse gods. In the right-extreme political party NPD in Germany, religion is not mentioned so much, except for some criticism of Islam. Some authors in the Stormfront forum actually think a little like some authors in Ummah, that the moral in the western society today is corrupted. Some Stormfront authors think that old, Christian values like marriage and family must return to the minds of people today, so that the society will work properly again in the way they want it. These are not only Christian values. We can see that Muslims in the Islamist forums are also preoccupied by marriage and family. The Stormfront authors do not like biracial marriages, and they want married couples to get many children, so

that the foreigners will not dominate the society. For Islamists it does not seem to matter what kind of race or ethnic group one belongs to, as long as one confesses to the religion Islam.

Both of the two types of extremists, which are studied in this thesis, seem to have problems understanding the opinions of other people. In most cases, they just sweep all other ways of looking at politics and society away, and just state that they are right, no matter what other people think. Typical claims are that other people are immoral and/or ignorant. According to the extremists, other people just lie, or are influenced by others to spread lies.

We also compared the Norwegian right-extremist website Vigrid with the political forum at the Norwegian diskusjon.no. In the comparison between these two websites we clearly see that many of the words in Vigrid are related to racism and to the Second World War. Vigrid uses for instead more often the word “holocaust” than diskusjon.no (with an NGTF ratio of 32.7). Although the ordinary forums also use these words, they have quite different opinions than the extremists about the same topics. Authors in the right-extreme websites claim that holocaust, the genocide of Jews by the German Nazi regime, did not happen, while ordinary are ashamed of the fact that it happened.

An example from line 6844 in vigriddtdnet.txt mentions a document which allegedly has been hidden or ignored by people wish to the story about holocaust remain accepted as a truth (the author thinks that holocaust never happened):

*“Dette utvetydige dokumentet og flere liknende er systematisk skjult eller ignorert av de som opprettholder utryddelseshistorien om holocaust.”*

Moreover, we compared the German online extreme nationalist news site Deutsche Stimme and the German section of the international Islamist forum Turn to Islam. Deutsche Stimme writes more about amongst others “Deutschen” (Germans), “Parteien” (parties), “Staat” (state), “Demokratie” (democracy), and “Politik” (politics), which are related to politics. The word “Deutschen” give a good clue that the German politics is discussed. The German Islamist forum in Turn to Islam discusses the religion Islam: “Allah”, “Koran”, “Paradies” (paradise), “Moschee” (mosque).

Analysis of Norwegian and German forums worked well, and as recently shown, the words were quite similar to the ones in the results for the English forums. We had some issues with the text encoding, the Norwegian letter æ, ø, and å, and the German letters ä, ö, ü, and ß were not written correctly in the CSV/TSV files we got as output from our Java text mining programs. In the tables and examples in chapter 4 these errors are corrected manually back to the original letters.

From all examples for word count comparisons above and the others presented in sections 4.6-4.12, 4.15, and 4.18, we see that it should fully possible to find characteristic words of an extremist forum by comparing it with the vocabulary of another kind of forum. We can actually also go the other way and find words that are typical for an ordinary forum and not for an extremist forum. Hypothesis 3 is therefore shown to be plausible.

## Chapter 6: Conclusion

In this thesis we wanted to figure out which words extremists tend to use, and which words they use that other people do not use to the same extent. We wanted to figure out how extremists express themselves to spread their ideas to fellow extremists and/or other people visiting the forums. What we did was downloading webpage (HTML) files from online forums, extracting forum message text data from these files and storing them in a database, generating word count statistics, and manually analysing them. Word statistics included frequent and important words, and word pairs, and statistics from forums were pairwise compared to each other.

We have shown in this thesis that words used frequently in an extremist forum, or used much more in that extremist forum than in an ordinary forum, can to some extent be used to find important topics the extremists write about. It also seems useful to find pairs of words that occur together in the same message when looking discussion topics in a forum. Finding words that tend to occur together seems to give at least a slightly better insight in the contents of forum messages than counts of single words alone. Nevertheless, one should read through text samples, where the frequent words or word pairs occur, to be able to somehow conclude what is really discussed in the forum, because the meaning of a word highly depends on how it is used in context. We have also demonstrated that it seems like authors can be characterized by which words they use and how frequently these words are. We then see which topics they tend to write about. Words that make an author unique in a group of ten seem to somehow work as author characteristics as well.

By means of the word frequency analysis for each forum, and pairwise word frequency comparisons of forums, we figured out that right-extremists discuss politics, and argument for why their political point of view is correct. We also saw that Islamists discuss the religion Islam, what is important in that religion, and why it is the correct religion according to them. Both right-extremists and Islamists claim that the western culture is immoral, and both groups are preoccupied by family values.

Analysis of websites/forums in Norwegian and German worked well, and we got results resembling the ones for English forums. This means that many of the most frequent words had the same meaning as the words from the corresponding English forums.

### Future Work

In future work of extremist forum research, it would be useful to find words that occur together in the same sentences in a complete, not only the same messages. It could also be useful to find sets of three or four words that tend to be together, not only just sets of two words. It would be useful to have own functionality in the text mining program for finding compound English words which without hyphen, and full names of persons consisting of one first name and last name, and maybe one or more middle names.

Furthermore, comparing frequencies of word pair occurrences in two different forums could be interesting to find out if more than one author actually uses the same word combinations and how they differ in usage frequency. For comparing pairs of co-occurring words in two different forums, we could divide the odds ratio of the word pair in the first forum by the

odds ratio of the same word pair in the other forum. In other words, we would compute ratios of odds ratios for word pairs that are in the two different forums.

When it comes to author text mining, one could divide all authors in a forum into groups, where all authors in a group had some important topic-related words in common, and at the same time each of these authors in one group would use some words that are unique for the group. Inside each group we then could create a unique profile for each single author by finding and using his/her unique words. For profiling each author, it would be useful to find words that tend to occur together in the same message or sentence. We then of course still need to include single words and their frequencies in the analysis as well.

In our thesis, we have only downloaded each forum at a specific point in time. What could be interesting in future research, would be creating a system that regularly looks for updates in the forums, and stores the downloaded text data in a kind of version control system. Then we would be able to follow and analyse forum changes over time. For author text mining, we could use publish times for the analysis to see when and how often an author tends to write.

## References

1. Stormfront (website). *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 17 May 2013. [Cited: 18 May 2013.] [http://en.wikipedia.org/wiki/Stormfront\\_\(website\)](http://en.wikipedia.org/wiki/Stormfront_(website)).
2. Extremism. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 15 May 2013. [Cited: 20 May 2013.] <http://en.wikipedia.org/wiki/Extremism>.
3. Browse Forums: Forums in English. *Dark Web Forum Portal*. [Online] The University of Arizona, 2013. [Cited: 16 May 2013.] <http://cri-portal.dyndns.org/portal/Home.action#BrowseForums>.
4. **Mitchell, T. M.** *Machine Learning*. s.l. : McGraw-Hill Science/Engineering/Math, 1997. ISBN: 9780070428072.
5. **Bradford, R. B.** Relationship discovery in large text collections using latent semantic indexing. [Online] 2006. [Cited: 5 February 2013.] <http://lsa-svd-application-for-analysis.googlecode.com/svn-history/r113/trunk/LSA/Other/LsaToRead/1-15.pdf>.
6. **Patman, F. and Thompson, P.** Names: A New Frontier in Text Mining. *Springer Link*. [Online] 2003. [Cited: 5 February 2013.] [http://link.springer.com/chapter/10.1007/3-540-44853-5\\_3](http://link.springer.com/chapter/10.1007/3-540-44853-5_3). ISBN 978-3-540-44853-2.
7. **Kramer, S.** Anomaly detection in extremist web forums using a dynamical systems approach. *ACM Digital Library*. [Online] 2010. [Cited: 5 February 2013.] <http://dl.acm.org/citation.cfm?id=1938614>. ISBN 978-1-4503-0223-4.
8. **Yilu, Z., et al.** US domestic extremist groups on the Web: link and content analysis. *IEEE Xplore Digital Library*. [Online] 26 September 2005. [Cited: 5 February 2013.] [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1511999&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1511999&tag=1). ISSN 1541-1672.
9. **Abbasi, A. and Hsinchun, C.** Applying authorship analysis to extremist-group Web forum messages. *IEEE Xplore Digital Library*. [Online] 26 September 2005. [Cited: 5 February 2013.] [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1512002&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1512002&tag=1). ISSN 1541-1672.
10. **Zhou, Y., et al.** Exploring the dark side of the web: Collection and analysis of US extremist online forums. *Springer Link*. [Online] 2006. [Cited: 5 February 2013.] [http://link.springer.com/chapter/10.1007/11760146\\_67](http://link.springer.com/chapter/10.1007/11760146_67). ISBN 978-3-540-34479-7.
11. **Pillay, S. R. and Solorio, T.** Authorship Attribution of Web Forum Posts. *IEEE Xplore Digital Library*. [Online] 18-20 October 2010. [Cited: 5 February 2013.] [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5706693](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5706693). ISBN 978-1-4244-7760-9.
12. **Dahlin, J., et al.** Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Boards. [Online] 26-29 August 2012. [Cited: 4 December 2012.] <http://users.isy.liu.se/rt/johda87/publications/Dahlin2012-EntityMatching.pdf>. ISBN 978-1-4673-2497-7.
13. Swiss English Forum. [Online] The Local Europe AB. [Cited: 18 May 2013.] [englishforum.ch](http://englishforum.ch).
14. The Local. *Switzerland news in English*. [Online] The Local Europe AB. [Cited: 18 May 2013.] [englishforum.ch/daily-life/](http://englishforum.ch/daily-life/).

15. Vigrid – norsk politisk organisasjon. *Store norske leksikon*. [Online] Norsk nettleksikon, 2 May 2012. [Cited: 18 May 2013.] <http://snl.no/Vigrid/norsk%20politisk%20organisasjon>.
16. Vigrid (organisasjon). *Norsk Wikipedia*. [Online] Wikimedia Foundation, Inc., 5 April 2013. [Cited: 16 May 2013.] [http://no.wikipedia.org/wiki/Vigrid\\_\(organisasjon\)](http://no.wikipedia.org/wiki/Vigrid_(organisasjon)).
17. **Schrenk, M.** *Webbots, Spiders, and Screen Scrapers: A Guide to Developing Internet Agents with PHP/CURL (2nd edition)*. San Francisco, CA, USA : No Starch Press, 2012. ISBN-13: 978-1593273972.
18. CNET download.com. [Online] 4 October 2006. [Cited: 13 February 2013.] [http://download.cnet.com/WinWSD-WebSite-Downloader/3000-2377\\_4-10562531.html](http://download.cnet.com/WinWSD-WebSite-Downloader/3000-2377_4-10562531.html).
19. **Hedley, J.** jsoup: Java HTML Parser. [Online] 2013. [Cited: 20 May 2013.] <http://jsoup.org/>.
20. **Odvarco, J.** Firebug. *Web Development Evolved*. [Online] [Cited: 20 May 2013.] <http://getfirebug.com/>.
21. **W3Schools.** HTML DOM Introduction. [Online] W3Schools. [Cited: 19 May 2013.] [http://www.w3schools.com/html/dom/dom\\_intro.asp](http://www.w3schools.com/html/dom/dom_intro.asp).
22. —. HTML DOM Nodes. [Online] W3Schools. [Cited: 19 May 2013.] [http://www.w3schools.com/html/dom/dom\\_nodes.asp](http://www.w3schools.com/html/dom/dom_nodes.asp).
23. tf-idf. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 22 May 2013. [Cited: 28 May 2013.] <http://en.wikipedia.org/wiki/Tf-idf>.
24. Odds ratio. *English Wikipedia*. [Online] [Cited: 12 May 2013.] [http://en.wikipedia.org/wiki/Odds\\_ratio](http://en.wikipedia.org/wiki/Odds_ratio).
25. Counter-Strike: Global Offensive. [Online] Valve Corporation, 21 August 2012. [Cited: 22 May 2013.] <http://store.steampowered.com/app/730/>.
26. myIWC.com Forums. *Islamic Web-Community*. [Online] 2013. [Cited: 27 May 2013.] <http://www.myiwc.com/forums/>.
27. java.util.regex Class Pattern. *POSIX character classes (US-ASCII only)*. [Online] Oracle. [Cited: 31 May 2013.] <http://docs.oracle.com/javase/1.4.2/docs/api/java/util/regex/Pattern.html>.
28. Allahu Akbar. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 19 May 2013. [Cited: 20 May 2013.] [http://en.wikipedia.org/wiki/Allahu\\_Akbar](http://en.wikipedia.org/wiki/Allahu_Akbar).
29. New Matchmaking & Maps. *Counter-Strike: Global Offensive*. [Online] Valve Corporation, 1 October 2012. [Cited: 26 May 2013.] <http://blog.counter-strike.net/index.php/2012/10/5256/#FAQ>.
30. What is E-Sports Entertainment? [Online] E-Sports Entertainment LLC, 2013. [Cited: 26 May 2013.] <http://play.esea.net/welcome.php?referer=http%3A%2F%2Fplay.esea.net%2F>.
31. First-person shooter. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 22 May 2013. [Cited: 26 May 2013.] [https://en.wikipedia.org/wiki/First-person\\_shooter](https://en.wikipedia.org/wiki/First-person_shooter).
32. Frame rate. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 11 May 2013. [Cited: 26 May 2013.] [http://en.wikipedia.org/wiki/Frame\\_rate](http://en.wikipedia.org/wiki/Frame_rate).
33. Counter-Strike: Global Offensive. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 22 May 2013. [Cited: 26 May 2013.] [http://en.wikipedia.org/wiki/Counter-Strike:\\_Global\\_Offensive](http://en.wikipedia.org/wiki/Counter-Strike:_Global_Offensive).
34. Haram. *English Wikipedia*. [Online] Wikimedia Foundation, 6 May 2013. [Cited: 26 May 2013.] <http://en.wikipedia.org/wiki/Haram>.

35. Mosque. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 21 May 2013. [Cited: 26 May 2013.] <http://en.wikipedia.org/wiki/Mosque>.
36. Dome of the Rock. *Encyclopædia Britannica*. [Online] Encyclopædia Britannica, Inc. [Cited: 28 May 2013.] <http://global.britannica.com/EBchecked/topic/168491/Dome-of-the-Rock>.
37. Subhan'Allah. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 27 February 2013. [Cited: 20 May 2013.] [http://en.wikipedia.org/wiki/Subhan\\_Allah](http://en.wikipedia.org/wiki/Subhan_Allah).
38. Migros. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 16 May 2013. [Cited: 20 May 2013.] <http://en.wikipedia.org/wiki/Migros>.
39. expat. [Online] Farlex, Inc. [Cited: 2 June 2013.] <http://www.thefreedictionary.com/Expat>.
40. Zug. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 4 April 2013. [Cited: 20 May 2013.] <http://en.wikipedia.org/wiki/Zug>.
41. St. Gallen. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 5 May 2013. [Cited: 21 May 2013.] [http://en.wikipedia.org/wiki/St.\\_gallen](http://en.wikipedia.org/wiki/St._gallen).
42. **Kjøll, G.** ibn. *Store norske leksikon*. [Online] Norsk nettleksikon, 28 February 2013. [Cited: 14 May 2013.] <http://snl.no/ibn>.
43. arabere - historie: Den osmanske tid. *Store norske leksikon*. [Online] Norsk nettleksikon, 14 February 2009. [Cited: 16 May 2013.] <http://snl.no/arabere/historie>.
44. **Vogt, K.** hadith. *Store norske leksikon*. [Online] Norsk nettleksikon, 28 February 2013. [Cited: 14 May 2013.] <http://snl.no/hadith>.
45. Ummah. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 17 April 2013. [Cited: 16 May 2013.] <http://en.wikipedia.org/wiki/Ummah>.
46. **Vogt, K.** sunna. *Store norske leksikon*. [Online] Norsk nettleksikon, 15 February 2009. [Cited: 16 May 2013.] <http://snl.no/sunna>.
47. Subhanahu wa ta'ala. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 28 February 2013. [Cited: 16 May 2013.] [http://en.wikipedia.org/wiki/Subhanahu\\_wa\\_ta%27ala](http://en.wikipedia.org/wiki/Subhanahu_wa_ta%27ala).
48. **Bratberg, T.** Ibn Saud. *Store norske leksikon*. [Online] Norsk nettleksikon, 28 February 2013. [Cited: 14 May 2013.] [http://snl.no/Ibn\\_Saud](http://snl.no/Ibn_Saud).
49. Ummah. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 17 April 2013. [Cited: 14 May 2013.] <http://en.wikipedia.org/wiki/Ummah>.
50. **Bryhn, R.** Jesse Owens. *Store norske leksikon*. [Online] Norsk nettleksikon, 14 February 2009. [Cited: 14 May 2013.] [http://snl.no/.versions/list/Jesse\\_Owens](http://snl.no/.versions/list/Jesse_Owens).
51. Sozialistische Reichspartei Deutschlands. *Deutsche Wikipedia*. [Online] Wikimedia Foundation, Inc., 17 April 2013. [Cited: 20 May 2013.] [http://de.wikipedia.org/wiki/Sozialistische\\_Reichspartei](http://de.wikipedia.org/wiki/Sozialistische_Reichspartei).
52. **Vogt, K.** hijab. *Store norske leksikon*. [Online] Norsk nettleksikon, 10 August 2012. [Cited: 24 May 2013.] <http://snl.no/hijab>.
53. surah. *Encyclopædia Britannica*. [Online] Encyclopædia Britannica, Inc. [Cited: 22 May 2013.] <http://global.britannica.com/EBchecked/topic/574938/surah>.
54. Recitation. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 5 March 2013. [Cited: 22 May 2013.] <http://en.wikipedia.org/wiki/Recitation>.
55. United States Foreign Intelligence Surveillance Court. *English Wikipedia*. [Online] Wikimedia Foundation, Inc., 18 May 2013. [Cited: 31 May 2013.]

[http://en.wikipedia.org/wiki/FISA\\_Court](http://en.wikipedia.org/wiki/FISA_Court).

56. **Kendall, D. G. J. and Sanders, A.** A Mobile Search & Data Mining Application for Terrorism Data. [Online] 2012. [Cited: 12 May 2013.]

[http://scholar.googleusercontent.com/scholar?q=cache:fuAuQahsHyEJ:scholar.google.com/+Kendall+kronos&hl=no&as\\_sdt=0,5](http://scholar.googleusercontent.com/scholar?q=cache:fuAuQahsHyEJ:scholar.google.com/+Kendall+kronos&hl=no&as_sdt=0,5).

57. Counter-Strike: Global Offensive. *Steam Community Discussions*. [Online] Valve Corporation. [Cited: 21 April 2013.] [steamcommunity.com/app/730/discussions/](http://steamcommunity.com/app/730/discussions/).



## Appendix A: Code, Text and Word Statistics

Program code used for the research of this thesis, as well as text downloaded and word statistics files generating by that code, are put in a zipped folder called “mefutm.zip”, which again is uploaded to a public Dropbox folder. The zip file is available at <https://dl.dropboxusercontent.com/u/23581429/mefutm.zip> at least throughout June 2013. It has a size of 208 MB.

In the zipped folder “mefutm.zip”, two files called “forum\_entries.sql” and “vigriddtvedt.txt” include all text we have used for the text mining and analysis. The SQL file is meant for recreating a table in a database, where all the forum text data (except for text from Vigrid) were stored before the text was analysed. The data in the database really come from HTML documents downloaded from online forums.

The folder “Code” contains code for HTML document download from forums in the subfolders “Eclipse\Vigrid”, and “NetBeans\Engchspider - PHP.zip”. The former is Java code for downloading all text in pages available from the front-page at vigriddt.net. The latter is a PHP webbot for downloading HTML document are forum threads available from the front-page. This code is for downloading from Swiss English Forum. We used similar or resembling code for downloading from the other forums. Be aware that most of the code is original code written by the author of the book “Webbots, Spiders, and Screen Scrapers” [17]. Only the file “engchspider.php” and some few other code lines are written by us.

Code for extracting text from HTML files to a database table can be found in the subfolder “NetBeans\HtmlToDbExtractor – Java.zip”. Code for mining/analysing the text found in the database and in the file “vigriddtvedt.txt”, can be found in the subfolder “Netbeans\GtfIdf.txt”.

The folder “WordCooccurrences” includes files that list words that occur together in the same forum messages, and their odds ratios.

In the folder “Word statistics”, one can find the statistical data generated from the forum text data extracted from the webpages. The subfolder “WordCount” includes the word count statistics GTF, IDF, NGTF, and NGTF-IDF. The subfolder “WordCountAuthor\AuthorWordCount” has similar data for some of the most-writing authors. The subfolder “WordCountAuthor\AuthorWordsInForum” includes statistics with intersection of words used by most active authors, and each of these authors’ unique words in the top 10 activity group.

The folder “WordInForumsComparison” includes files that show NGTF ratios of words that exist in two forums. There are also files in this folder with lists of words that are only used in one of two forums in a comparison.