

Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit

Thomas Schmidt, Florian Kaindl and Christian Wolff

Media Informatics Group, University of Regensburg, Germany

thomas.schmidt@ur.de

florian.kaindl@stud.uni-regensburg.de

christian.wolff@ur.de

Abstract. We present results of a project examining the application of computational text analysis and *distant reading* in the context of comparative religious studies, sociology, and online communication. As a source for our corpus, we use the popular platform *Reddit* and three of the largest religious subreddits: the subreddit *Christianity*, *Islam* and *Occult*. We have acquired all posts along with metadata for an entire year resulting in over 700,000 comments and around 50 million tokens. We explore the corpus and compare the different online communities via measures like word frequencies, bigrams, collocations and sentiment and emotion analysis to analyze if there are differences in the language used, the topics that are talked about and the sentiments and emotions expressed. Furthermore, we explore approaches to diachronic analysis and visualization. We conclude with a discussion about the limitations but also the benefits of distant reading methods in religious studies.

Keywords: Religious Studies, Distant Reading, Reddit, Sentiment Analysis, Computational Social Science, Collocation.

Introduction

With the concept of *distant reading*, Moretti [17] has argued for the application of statistical and computational methods, primarily in literary studies and linguistics. The general idea of distant reading is to explore large quantities of text via methods of computational text analysis and text visualization, thus enabling findings that would not be possible by qualitative or hermeneutical work alone. Some of the most popular methods in this field are stylometry, topic modeling and sentiment and emotion analysis [10]. The application of distant reading is also explored outside of literary studies. Similar concepts for other media types can be found in film studies ([1]: distant *viewing*) or digital musicology [5]. In addition, in the context of textual analysis, distant reading is also explored outside of literary studies in other text-oriented domains (e.g. [21]). In a similar way, we want to explore the application of distant reading in the

context of religious studies and sociology by analyzing the communication of different religious groups on the online platform *Reddit*¹.

With the rise of social media the research area of *computational social science* is gaining popularity [12]. However, most text analysis and visualization approaches are focused on areas like politics (e.g. [11]). In the context of religious content and religious studies one can find research about extremist groups like ISIS [2] or the application of distant reading techniques for famous religious texts (e.g. the Bible) [14, 25, 26].

However, the exploration of the online communication of “ordinary” or “moderate” religious and spiritual groups on social media channels is rather rare although one can assume that, considering the importance of social media for young adults, a lot of religious discussions find their place on those platforms. Pfahler et al. [21] show the benefit of applying distant reading on a Muslim forum by exploring the topics discussed via *topic modeling*.

To gather more insights about the subjects and language of religious discussions on social media across diverse religious creeds, we present the results of a project examining and comparing three subforums on Reddit: a Christian, a Muslim and an occult forum. We explore different techniques of distant reading and computational text analysis. As methods, we employ the analysis of most frequent words and bigrams, collocation analysis and visualization as well as sentiment and emotion analysis. Our research goals are (1) to identify differences and specific features concerning language usage as well as content discussed among those groups and (2) reflect upon the benefits and limitations of the different computational techniques used.

Corpus

In the following, we describe how we gathered and constructed the corpus. If not mentioned otherwise, we made use of *Python* and the popular library *NLTK* for all methods.

Corpus Acquisition

As a source for our corpus, we have chosen the platform *Reddit*². *Reddit* is a news aggregation website founded in 2005 and is ranked among the top 20 most visited websites in the world³. In recent years, the platform has evolved from its primary use, which is to share links and images. Nowadays, it is a collection of subforums for various topics. Users can subscribe to a subforum and via a voting system more popular posts are placed more prominently on the platform. A subforum, also called *subreddit* on *Reddit* consists of *submissions* (which are equivalent to a thread for general forums) and corresponding *comments*. Usually, the majority of entries consists of com-

¹ <https://www.reddit.com/>

² <https://www.reddit.com/>

³ <https://www.alexa.com/topsites>

ments. Due to the huge popularity, Reddit has been used for various research using text mining methods [6, 7, 8]. Furthermore, Reddit’s open source roots and various open source libraries for gathering data adds to its popularity in research.

As subreddits we have chosen three of the most popular religious subreddits. The first subreddit, */r/Christianity*⁴ is focused on discussions about Christian belief and practice. It is the biggest of the subreddits with 202,242 subscribers. The Muslim subreddit */r/Islam*⁵ is smaller (82,404 subscribers), possibly due to the fact that Reddit is more popular in English speaking countries. Next to those monotheistic religions, we also look at an esoteric forum: */r/Occult*⁶. The subreddit describes itself as “centered around discussion of the occult, mysticism, esoterica, metaphysics, and other related topics” (149,379 subscribers; all subscription counts are of September 2, 2019). Spiritual directions as discussed in this forum have become more popular, especially in the Western world and among the youth. Therefore, it is not surprising that this subreddit is larger than those of many world religions (e.g. Islam, Judaism, Buddhism). Please note that we do not want to explore the religious convictions in these online communities (which might be an interesting topic for religious studies) but rather want to explore the possibilities of corpus analysis and distant reading. For this purpose, the chosen subreddits are (1) large enough and (2) varied enough to investigate the subreddits on their own but also to compare them with each other. However, one limitation to keep in mind is the difference in size with */r/Christianity* being much greater. We will focus on normalized results to avoid problems because of these size differences.

To gather submissions, comments and metadata for a specific subreddit we use the *Python Reddit API Wrapper (PRAW)*⁷ library and save the data in the JSON format in a *MongoDB* database. All submissions and comments have been collected for a timeframe of one year (from the 1st of July 2018 to the 1st of July 2019). It is important to regard at least one year since religious communities and their communication behavior might be influenced due to specific holidays in the circle of the year.

Corpus Description

We have collected 115,556 submissions for all three subreddits. Nevertheless, we filtered out 74,162 submissions consisting of links only or lacking author information. Posts with no author information (e.g. deleted authors) are not visible anymore on the platform. Thus, 41,394 submissions remain. After extracting the comments of these submissions, 759,992 comments remain comprising more than 50 million tokens and over 3.5 million sentences. Table 1 summarizes some of the general metrics of the overall corpus and the specific subreddits after filtering noise.

⁴ <https://www.reddit.com/r/Christianity/>

⁵ <https://www.reddit.com/r/islam/>

⁶ <https://www.reddit.com/r/occult/>

⁷ <https://github.com/praw-dev/praw>

Table 1. Corpus metrics.

Metric/Forum	/r/Christianity	/r/Islam	/r/Occult	Sum
Submissions	28,896	4,123	8,275	41,394
Comments	618,719	64,886	76,387	759,992
Tokens	43,996,066	4,754,301	5,702,675	54,453,042
Sentences	2,897,575	300,854	365,962	3,564,391

Table 2 illustrates some statistics about the lengths of submissions and comments.

Table 2. Comparison of post lengths.

Metric/Forum	/r/Christianity	/r/Islam	/r/Occult
Sentences per submission	10.6	9.9	9.3
Tokens per sentence in submission	16.2	16.5	16.4
Sentences per comment	4.2	4.0	3.8
Tokens per sentence in comments	15.1	15.7	15.4
Comments per submission	21.4	15.7	9.1

Considering the post lengths, there are no specific differences. The only striking difference can be found concerning the Christian Forum since it is much larger than the other ones. Submission also have a higher number of comments. However, the overall sentence length does not differ since sentences are only slightly longer in the Christian forum.

Analysis

In the following, we present results for various statistical parameters, starting with word frequencies, followed by bigram frequencies, results for significant collocations and sentiment and emotion analysis.

Word Frequencies

To gain insights about the subjects discussed and the overall language we analyze the most frequent words used in the subreddits. For the preprocessing we have eliminated stop words and lemmatized the tokens using the *WordNet-Lemmatizer* which is a general purpose solution for lemmatization often used for social media content [19, 20]. The following figures illustrate the top 10 most frequent words (MFWs) per subreddit (Figure 1 to 3).

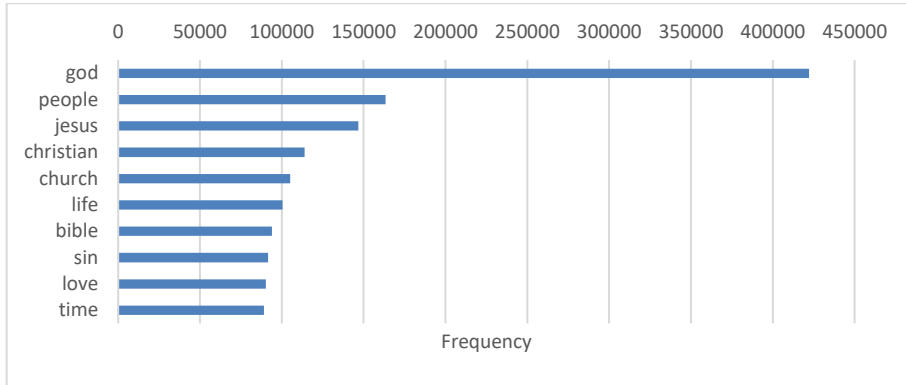


Fig. 1. MFWs in /r/Christianity.

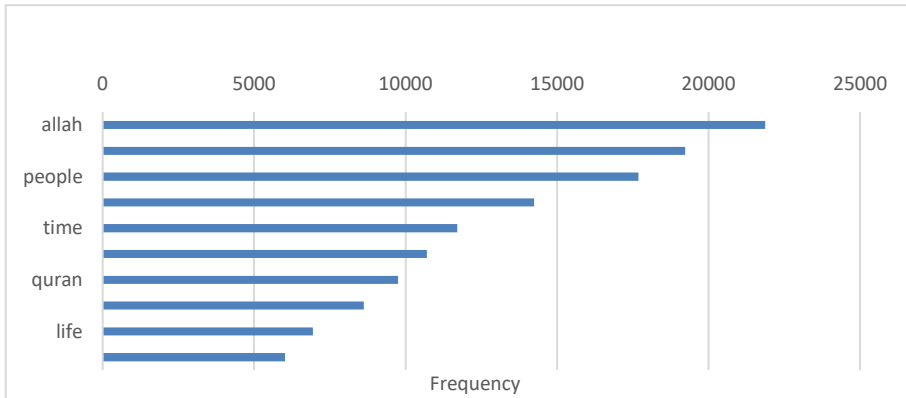


Fig. 2. MFWs in /r/Islam.

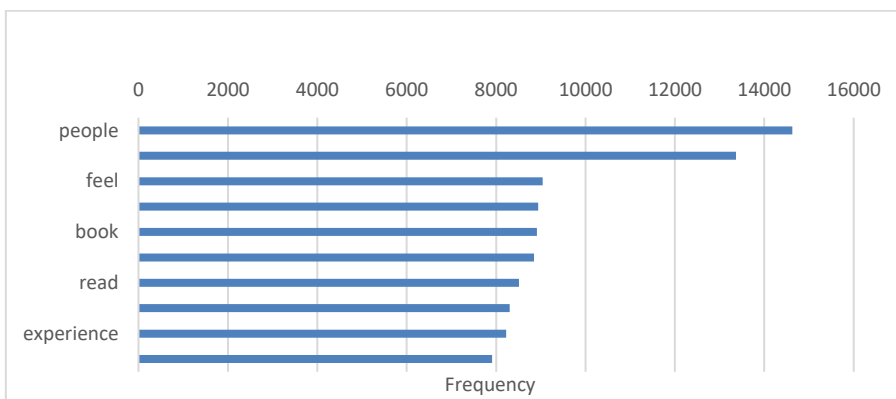


Fig. 3. MFWs in /r/Occult.

The results show that the word “god” is an important term in all three sub-reddits. It is also notable that the word is used with a higher relative frequency in /r/Christianity compared to the other forums. In /r/Islam it is notable that the words “muslim” and “islam” are much more common than their equivalents “christian” and “christianity” in /r/Christianity, suggesting these words are used differently in their respective domain, or that more meta discussion takes place in /r/Islam. The lack of the word “Mohammed” as one of the most frequent words in the Muslim forum is due to the numerous different spellings of this name (which have not been unified in this study). As a last observation on the top words, /r/Christianity is the only subreddit with a word for an emotion, “love”, in the top ten words, while /r/Occult’s top ten words uniquely feature two words relating to the senses, namely “feel” and “experience”.

Bigram Frequencies

A bigram is defined as two tokens appearing next to each other. More than unigrams, bigrams can give insights in the usage of more complex concepts. Figure 4 to 6 illustrate the 10 most frequent bigrams for each subreddit.

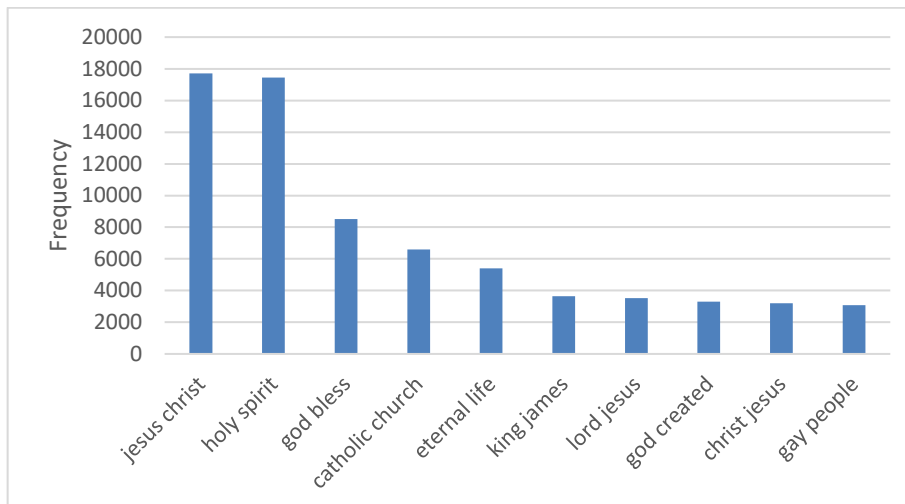


Fig. 4. Most Frequent Bigrams in /r/Christianity.

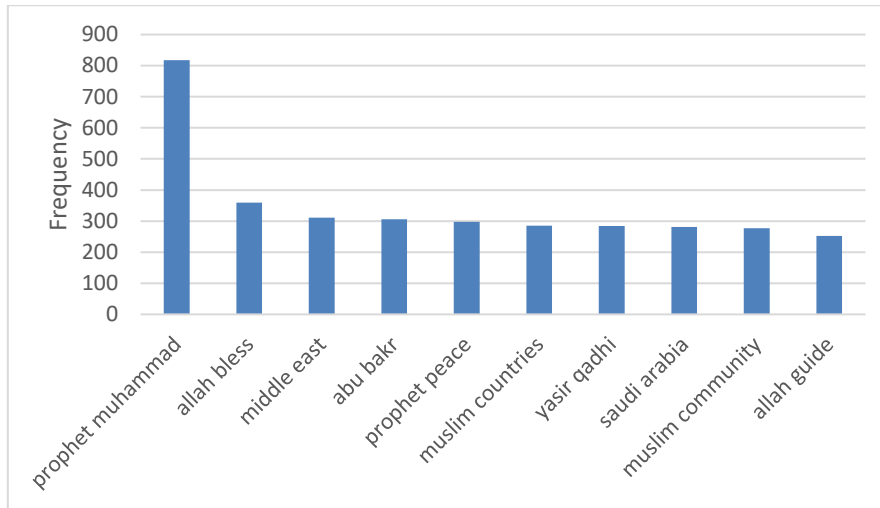


Fig. 5. Most Frequent Bigrams in /r/Islam.

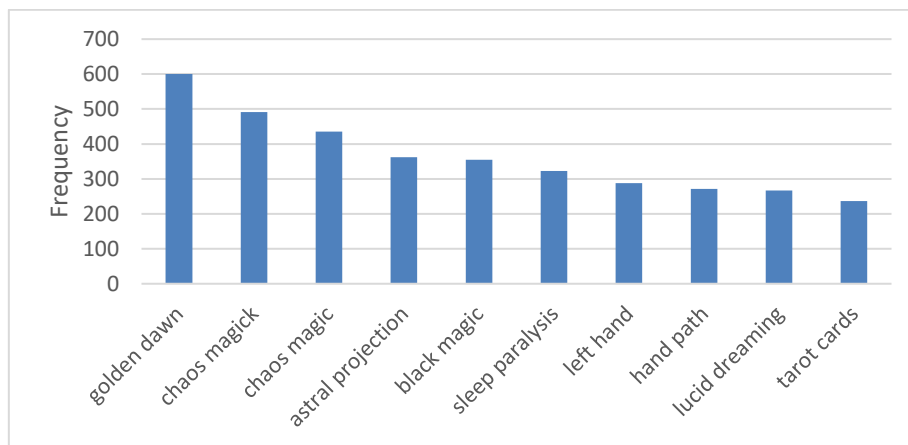


Fig. 6. Most Frequent Bigrams in /r/Occult.

Both in the Cristian as well as in the Muslim forum specific named entities are the most common bigrams e.g. “Jesus Christ”, “holy spirit”, “lord Jesus” and “prophet Muhammad”. One of the most frequent bigrams in the Christian subreddit is “gay people”. In comparison, bigrams consisting of the word gay are rather rare in the other forums. “Gay people” is ranked 44 for /r/Islam and there are no similar bigrams found for /r/Occult showing that this topic is not of interest for this specific community. In the Christian subreddit, a specific edition of the bible is often referred to (the King James Bible, the most important bible edition in the English speaking world). For the Muslim forum, geographical and political concepts are dominant e.g. “middle east”, “muslim countries”, “saudi arabia” as well as spiritual authorities (“Yasir

Quadhi”, “Abu Bakr”, “Ibn Taymiyyah”). Those findings are indeed in line with results of topic modeling on a similar corpus [21]. /r/Occult’s top bigrams refer mostly to esoteric concepts and practices which is interesting since religious practices are rarely discussed in the other forums

Collocations

To gain a better understanding about some of the religious key concepts we look at the collocations for words representing those concepts. As a text window for collocation analysis we choose five, meaning words can be a maximum of five positions away to be regarded as collocations. The collocation strength was measured as *Pointwise Mutual Information* (PMI) which scores the collocations based on their actual co-occurrence in the corpus in proportion to their expected co-occurrence if they were independent [4]. Because this can lead to high values for very low-frequency collocations, a minimum threshold was set for each measurement. We visualize the collocations similar to [3]. The key word is centered in the middle while the surrounding words are those that are frequent enough in the surroundings of the word according to the threshold. The lengths of the edges decrease with higher PMI-values, thus words that occur more frequent are closer to the centered word. We focus our analysis on various important religious words like god, death, life, love, experience or religion. In the following, we show the collocation usage for the words “god” and “death”.



Fig. 7. Collocation visualizations for the word “god” in r/Christianity/.



Fig. 8. Collocation visualizations for the word “god” in r/Islam/.

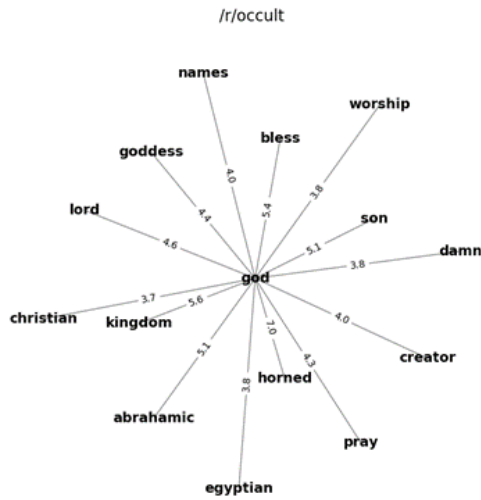


Fig. 9. Collocation visualizations for the word “god” in r/Occult/.

The collocations for the word god in Christianity (see figure 7–9) show some outdated verb forms pointing to bible quotes (“giveth”, “commendeth”). In line with the upcoming results about sentiment analysis, positive characterizations are more frequent (“forgives”, “loves”) than negative ones (“hates”, “punishing”). Similar holds true for the Muslim forum with words like “forgive”. Those positive collocations become

even more apparent when analyzing the word Allah instead of god (which is not shown here). It is striking that the existence of god seems to be discussed much more in the Muslim forum (“existence”, “exists”). Furthermore, the word “god” is probably (also) used in the Muslim forum to refer to a specific Christian or Jewish god (“son”, “Abraham”). For the occult forum, the multiple perspectives on God become very clear. The word god is mostly surrounded by other words clarifying which god is being discussed (“horned”, “Abrahamic”, “Egyptian”, “Christian”, “sun”). It is also the only forum showing some rather negative perspective on god via the collocation with “damn”. This might point to atheist or agnostic views.

The collocations for the concept death highlight the differences between the groups even more clearly (see figure 10–12).

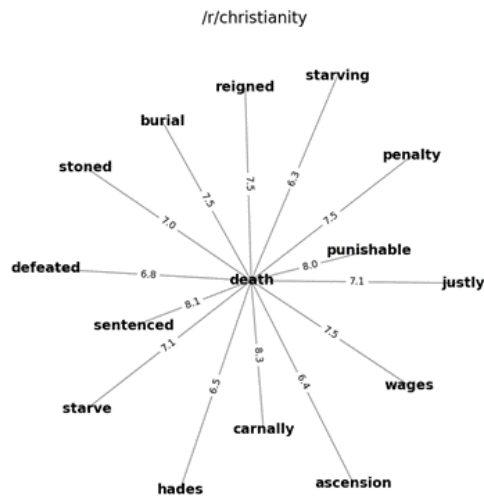


Fig. 10. Collocation visualizations for the word “death” in r/Christianity/.



Fig. 11. Collocation visualizations for the word “death” in r/Islam/.

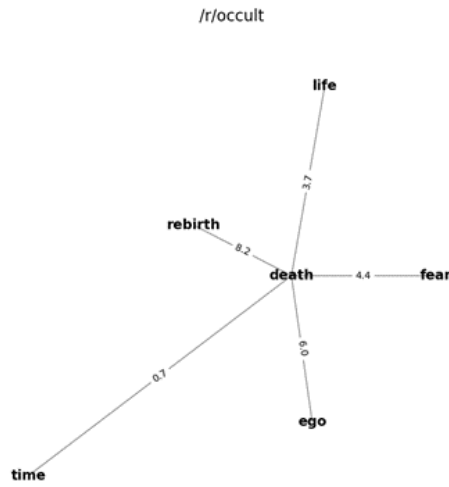


Fig. 12. Collocation visualizations for the word “death” in r/Islam/.

In /r/Islam as well as /r/Christianity strong correlations with the term “penalty” are found. Death is much more frequently discussed in the Christian forum, thus more collocations are identified. However, the collocations also point to the fact that death plays a much more important role in the life and narration of Jesus since we find a lot of collocations in this context (“ascension”, “resurrection”, “jesus”). The collocations

with “angel” and “taste” in the Muslim subreddit refer to specific Quran passages. For the occult forum, the esoteric and spiritual content becomes clear since death is strongly connected to words like “rebirth” and “ego” pointing also to spiritual concepts well-known in Buddhism.

Sentiment Analysis

Sentiment analysis means using computational methods for the analysis and prediction of sentiments, mostly in written text [13]. Most of the times, the prediction goal is whether the overall connotation of a text is negative, positive or neutral. This concept is also often referred to as polarity. Typical areas for sentiment analysis are product reviews but also social media [27]. In recent years, sentiment analysis has also gained a lot of interest in Digital Humanities [15, 18, 22, 23, 24].

To explore sentiment analysis in our specific corpus, we use *Vader*, an open source sentiment analysis library for Python⁸. *Vader* outputs a polarity score for each sentence, which allows for the classification of each sentence as positive, neutral or negative. Although *Vader* employs lexicon-based methods for sentiment analysis, it has been specifically developed for social media and shows very good evaluation results on this type of content [9]. Table 3 shows the percentage of sentences classified with a specific polarity class per subreddit.

Table 3. Ratio of Sentences Classified with a Polarity Class.

	Positive	Neutral	Negative
/r/Christianity	43.60%	31.85%	24.52%
/r/Islam	41.11%	36.20%	22.70%
/r/Occult	42.89%	37.59%	19.52%

While the sentiments expressed are rather similar, it is noticeable that /r/Christianity has the lowest ratio of neutrality and is thus more polarized than the other forums. /r/Occult has the lowest ratio of negative sentences, which might be because there are fewer negative topics like “sin” and “hell” discussed in this subreddit. Overall, it is rather striking that positivity dominates all subreddits. Please note however, that our findings are purely descriptive at the moment and we apply no significance tests for comparisons.

Emotion Analysis

The computational method of emotion analysis is closely related to sentiment analysis. The goal, however, is to analyze and predict more complex emotions instead of the simple polarity of a text. For our analysis, we use the *NRC Emotion Lexicon* [16], a general purpose sentiment and emotion lexicon. It consists of around 14,000 words

⁸ <https://github.com/cjhutto/vaderSentiment>

and their associations with a set of emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) but also with a polarity category (positive, negative). Words can be associated with one or more of those emotions and polarity categories. By counting the number of words associated with emotions one can investigate the emotionalization of the language used. However, please note that this lexicon, unlike *Vader*, is not optimized for social media language and is also not as sophisticated, as *Vader* also accounts for negations and valence shifters. The following graph illustrates the percentages of every category for each subreddit (see figure 13):

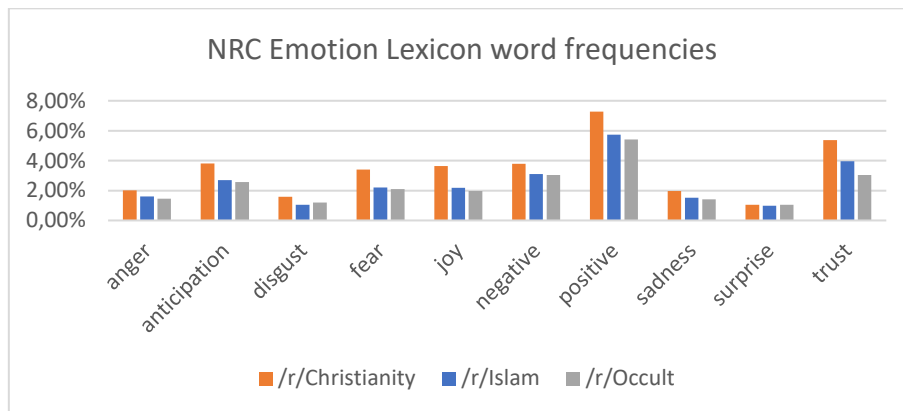


Fig. 13. Percentage of words associated with emotions per subreddit.

Like the results of the sentiment analysis, most emotions are much more frequent in the Christian forum than in the others. This especially accounts for the categories *anticipation*, *fear*, *joy* and *trust*. Similar to the results concerning *Vader*, we could also identify this effect for the two polarity categories *positive* and *negative*. This suggests that the discussions in */r/Christianity* are more emotionally charged. We also investigated what specific words of the NRC emotion lexicon lead to these results: Top words in the */r/Christianity* subreddit with a *trust* connotation include “god”, “church”, “faith” and “pray”, words that were found to be especially frequent in this sub-corpus. All of these, as well as “love”, are furthermore associated with *joy*. However, vocabulary from Abrahamic religions is also often associated with negative emotions. “God”, for example, is also associated with *fear* (its polarity is positive, however), as are “sin”, “pray” and “worship”, words most frequently found in */r/Islam* and */r/Christianity*. Concerning negativity and negative emotions */r/Occult* is much closer to the other forums. Reason for this are the negatively connoted and frequent words “occult”, “demon”, “black” and “chaos” (as commonly appearing in “black magic” and “chaos magic”).

Discussion

Via various methods of computational text analysis, we were able to gather some interesting insights concerning the topics that are talked about and the sentiments and emotions expressed. However, in the following we want to reflect upon the benefits and the limitations of the methods chosen:

Ngram-frequencies give a compact and easily to understand overview of the key concepts and topics that are discussed in the forums. The bigrams were more insightful than the unigrams showing some more general differences like the focus on politics and authorities in the Muslim forum and the focus on practices for the occult forum. The analysis of word frequencies also proved to be very helpful for the interpretation of more advanced methods like the collocation and sentiment/emotion analysis. However, comparisons are limited with this method, since similar concepts are often referred to differently (e.g. “God” vs “Allah”) and dependent of the specific vocabulary of a group. We also want to pursue methods to identify keywords that are specific for a sub-corpus e.g. using *tf-idf* weighting or comparative ranked lists.

The collocation analysis and visualizations did prove to be of the most interest for us. By focusing on specific words that represent important concepts, we were able to find interesting differences about the contexts of those words. Furthermore, to correctly interpret the data, in-depth knowledge about the religions is necessary e.g. to identify quotes of the scriptures. Comparisons are easier, since different words for the same concepts can be easily identified in the surroundings of a centered word. We recommend investigating collocation analysis for similar future work. We also plan to explore the possibility to construct a word embeddings model using our corpus to analyze word associations.

The sentiment and emotion analysis illustrates some interesting results concerning higher levels of emotional language for the Christian and Muslim forum. Although these findings are of interest, they should be validated by more in-depth analysis since now we can only speculate about the reasons for this result. We plan to analyze the most extreme manifestations of comments concerning the emotional values to gain more insights. Furthermore, we also want to precisely evaluate the performance of the sentiment analysis approaches since they have been proven rather problematic in other areas of Digital Humanities [22]. One obvious problem is the lack of an emotion lexicon which is specifically designed for the language used on Reddit or other social media platforms.

Finally, there are several limitations of our study one should keep in mind when interpreting the data. As already mentioned, the size of the subreddits was not equally distributed. We focused on the analysis of normalized data to avoid skewness because of the length. The reason for this disproportion might very well be the English language. */r/Islam* is very likely primarily used by Muslims living in Europe and America which are of course a minority compared to Christians in those countries. Furthermore, research has shown that Reddit is predominantly used by American male young adults⁹. Therefore, we want to point out that we cannot make any statements about the

⁹ <https://www.techjunkie.com/demographics-reddit/>

religious communities in general but only about this limited user group of Reddit and also just for the specific year we regarded. Nevertheless, we plan to explore distant reading methods to analyze religious groups on social media and improve our research by increasing the corpora and investigating other social media channels. We also want to examine other methods like stylometry, topic modeling, and named entity recognition to evaluate how religious studies and sociology can benefit of those methods.

References

1. Arnold, T., Tilton, L.: Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities* 34(Supplement 1), pp. i3–i16 (2019).
2. Badawy, A., Ferrara, E.: The rise of jihadist propaganda on social networks. *Journal of Computational Social Science* 1(2), pp. 453–470 (2018).
3. Brezina, V., McEnery, T., Wattam, S.: Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics* 20(2), 139–173 (2015)
4. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1), pp. 22–29 (1990).
5. Cook, N.: *Beyond the score: Music as performance*. Oxford University Press (2013).
6. De Choudhury, M., De, S.: Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In: *Eighth international AAAI conference on weblogs and social media* (2014).
7. Grover, T., Mark, G.: Detecting potential warning behaviors of ideological radicalization in an alt-right subreddit. In: *Proceedings of the International AAAI Conference on Web and Social Media*. vol. 13, pp. 193–204 (2019).
8. Guimaraes, A., Balalau, O., Terolli, E., Weikum, G.: Analyzing the traits and anomalies of political discussions on reddit. In: *Proceedings of the International AAAI Conference on Web and Social Media*. vol. 13, pp. 205–213 (2019).
9. Hutto, C.J., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth international AAAI conference on weblogs and social media* (2014).
10. Jänicke, S., Franzini, G., Cheema, M.F., Scheuermann, G.: On close and distant reading in digital humanities: A survey and future challenges. In: *EuroVis (STARs)*. pp. 83–103 (2015).
11. Karami, A., Bennett, L.S., He, X.: Mining public opinion about economic issues: Twitter and the us presidential election. *International Journal of Strategic Decision Sciences (IJSDS)* 9(1), pp. 18–28 (2018).
12. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al.: *Computational social science*. *Science* 323(5915), pp. 721–723 (2009).
13. Liu, B.: *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press (2016).
14. McDonald, D.: A text mining analysis of religious texts. *The Journal of Business Inquiry* 13(1), pp. 27–47 (2014).
15. Mohammad, S.: From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In: *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. pp. 105–114. Association for Computational Linguistics (2011).

16. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29(3), pp. 436–465 (2013).
17. Moretti, F.: Conjectures on world literature. *New left review*, pp. 54–68 (2000).
18. Nalisnick, E.T., Baird, H.S.: Character-to-character sentiment analysis in Shakespeare's plays. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. vol. 2, pp. 479–483 (2013).
19. Oyeboode, O., Orji, R.: Social media and sentiment analysis: The Nigeria presidential election 2019. In: *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. pp. 0140–0146. IEEE (2019).
20. Pandarachalil, R., Sendhilkumar, S., Mahalakshmi, G.: Twitter sentiment analysis for large-scale data: an unsupervised approach. *Cognitive computation* 7(2), pp. 254–262 (2015).
21. Pfahler, L., Elwert, F., Tabti, S., Morik, K., Krech, V.: What do you do with 5 million posts? versuche zum distant reading religioser online-foren. In: *Vogeler, G. (ed.) Book of Abstracts, DHd 2018*. pp. 335–338. Cologne, Germany (2018).
22. Schmidt, T., Burghardt, M.: An evaluation of lexicon-based sentiment analysis techniques for the plays of Gotthold Ephraim Lessing. In: *Proceedings of the Second Joint SIGHUMWorkshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. pp. 139–149. Association for Computational Linguistics (2018), <http://aclweb.org/anthology/W18-4516>
23. Schmidt, T., Burghardt, M., Dennerlein, K., Wolff, C.: Sentiment annotation in Lessing's plays: Towards a language resource for sentiment analysis on German literary texts. In: *2nd Conference on Language, Data and Knowledge (LDK 2019)* (2019).
24. Schmidt, T., Burghardt, M., Wolff, C.: Toward multimodal sentiment analysis of historic plays: A case study with text and audio for Lessing's Emilia Galotti. In: *Proceedings of the Digital Humanities in the Nordic Countries Conference 2019 (DHN 2019)*. pp. 405–414 (2019).
25. Slingerland, E., Nichols, R., Neilbo, K., Logan, C.: The distant reading of religious texts: A “big data” approach to mind-body concepts in early china. *Journal of the American Academy of Religion* 85(4), pp. 985–1016 (2017).
26. Verma, M.: Lexical analysis of religious texts using text mining and machine learning tools. *International Journal of Computer Applications* 168.
27. Vinodhini, G., Chandrasekaran, R.: Sentiment analysis and opinion mining: A survey. *International Journal* 2(6), pp. 282–292 (2012).