

Visualizing Collocations in Religious Online Forums

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

15th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2020
Ottawa, Canada
July 20-25, 2020
Conference Abstracts

Keywords: Collocations, Distant Reading, Religious Studies, Reddit, Visualization

Abstract.

We present results of a project examining the application of text visualization in the context of religious studies and sociology. Our goal is to analyze and compare the online communication of various religious directions. For this contribution we focus on the visualization of collocations for specific religious and spiritual key concepts. As a corpus, we acquired the content of the three religious subreddits /r/Islam, /r/Christianity and /r/Occult for a one-year time span. The overall corpus consists of 700,000 comments and around 50 million tokens. We explore and visualize collocations for the concepts “life”, “religion” and “love”. We discuss the results and to what extent we were able to gather new insights.

Link to version in the conference abstracts: https://dh2020.adho.org/wp-content/uploads/2020/07/510_VisualizingCollocationsinReligiousOnlineForums.html

Link to the poster on Humanities Commons: <http://dx.doi.org/10.17613/daq1q-1t69>

Please cite as:

Schmidt, T., Kaindl, F. & Wolff, C. (2020). Visualizing Collocations in Religious Online Forums. In *15th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2020, Conference Abstracts*. Ottawa, Canada.

1 Introduction

One of the most influential concepts in Digital Humanities (DH) in recent years is Moretti’s (2000) idea of Distant Reading, more precisely the application of computational methods to analyze and visualize large amounts of text to gather new insights. Distant Reading has led to various successful projects especially in literary studies and linguistics (cf. Jänicke et al., 2015) but also religious studies, e.g. to analyze famous religious texts (McDonald, 2014; Slingerland et al.; 2017; Verma, 2017). We want to build primarily upon the work of (Pfahler et al., 2018) who applied topic modeling on Muslim online forums to investigate what this community is predominantly talking about. They identified several main topic clusters about eating, family and politics which are talked about the most.

We want to further explore the application and potential benefit of Distant Reading-methods for the use case of religious online forums. Our research goal is to examine the content, language, topics and sentiments in religious online forums of different religious subgroups to identify differences and similarities and learn more about the way of life and beliefs of these communities.

While we explore multiple methods like named entity recognition, topic modeling and sentiment analysis, in the following contribution we report upon our results for the method of collocation analysis. Via collocations, we want to analyze differences in the way several religious key concepts are discussed in online forums of different religious subgroups.

2 Methods

We have chosen *Reddit* for data collection since it is rather easy to scrape and one of the largest platforms on the internet¹. Furthermore, various religious subgroups are represented enabling us to compare content more easily.

We have acquired all submissions (threads) for the time span of July 1, 2018 to July 1, 2019 for the three subreddits */r/Christianity*², */r/Islam*³ and */r/Occult*⁴. We chose the first two since they represent the two largest monotheistic religions and included the third one to also examine a rather esoteric religious direction.

We have acquired over 700,000 comments and around 50 million tokens⁵ (Table 1).

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 1. Corpus statistics

We have chosen five as maximum length for a collocation and measure the strength of collocations via *Pointwise Mutual Information (PIM)* which scores the collocations based on their actual co-occurrence in the corpus in proportion to their expected co-occurrence if they were independent (Church & Hanks, 1989). To visualize collocations, we place the key concept in the middle and the collocations around them. The higher the PMI-value, the closer the concept. We also put the exact PMI-score on the edges.

3 Results

In the following we showcase the use case for the spiritual key terms “love”, “religion” and “life” and highlight some insights we gained.

¹ <https://www.reddit.com/>

² <https://www.reddit.com/r/Christianity/>

³ <https://www.reddit.com/r/islam/>

⁴ <https://www.reddit.com/r/occult/>

⁵ The corpus is available upon request via mail: thomas.schmidt@ur.de

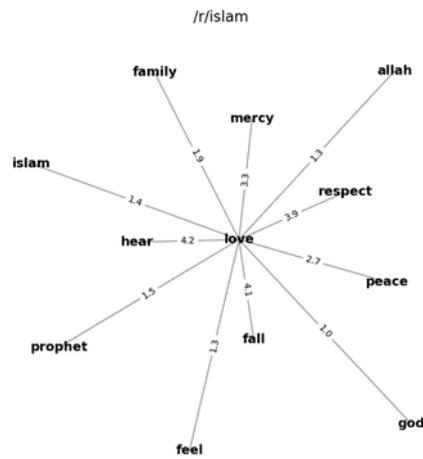


Figure 1. Collocations for “love” in /r/Islam

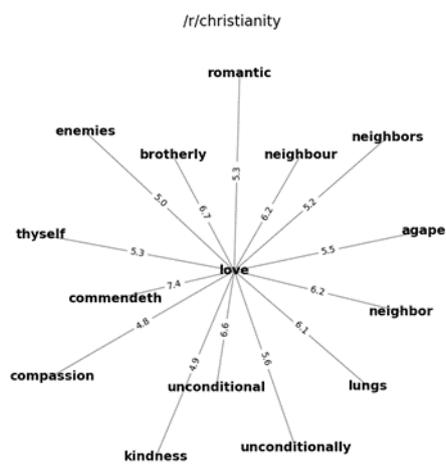


Figure 2. Collocations for “love” in /r/Christianity

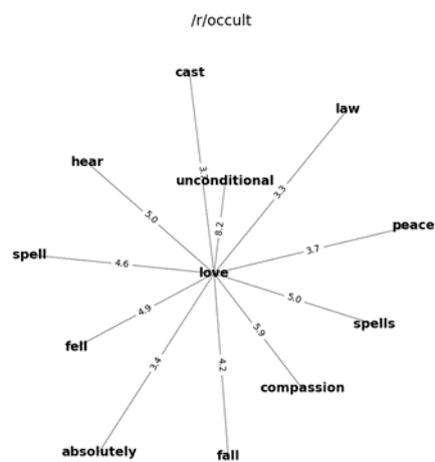


Figure 3. Collocations for “love” in /r/Occult

In the Christian subreddit, we find that love shows most connections with idioms/quotes from the bible (“unconditionally”, “enemies”, “agape”). In contrast, we find strong associations with positive terms, words for god and the prophet as well as for “family” in the Muslim forum which is in line with Pfahler et al. (2018) showing a strong focus on family-related topics in Muslim forums. For /r/occult we find rather fitting associations with the notion of magic, thus showing the rather esoteric content of this forum.

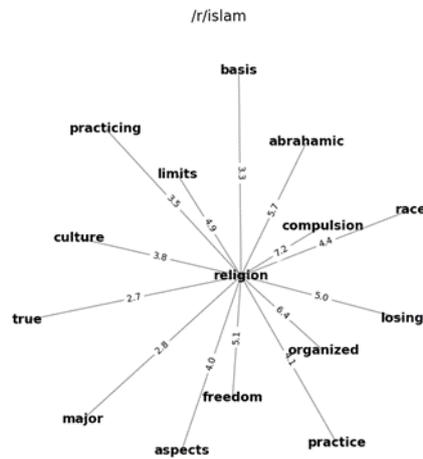


Figure 4. Collocations for “religion” in /r/Islam

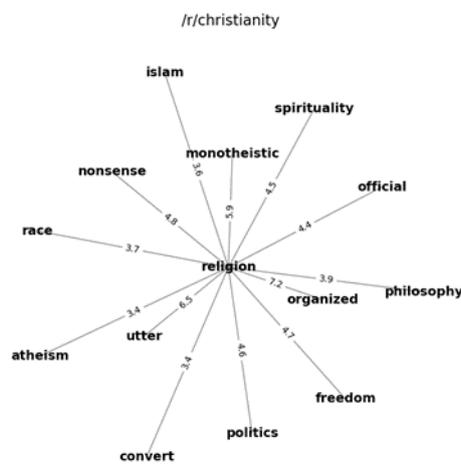


Figure 5. Collocations for “religion” in /r/Christianity

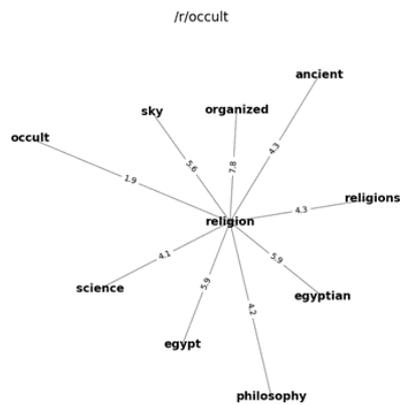


Figure 6. Collocations for “religion” in /r/Occult

Many terms in /r/Islam and the concept of religion point to discussions about religious directions e.g. “organized”, “abrahamic”, “culture”, “major”. The connection with race might be connected to the racism Muslims face in western countries. Quite similarly, /r/Christianity also shows collocations describing the discussion about other religions (“organized”, “islam”, “false”) also pointing to rather heated discussions (“utter”, “nonsense”). /r/Occult shows collocations specifying the religion and other world views (“Egypt”, “ancient”, “philosophy”, “science”).

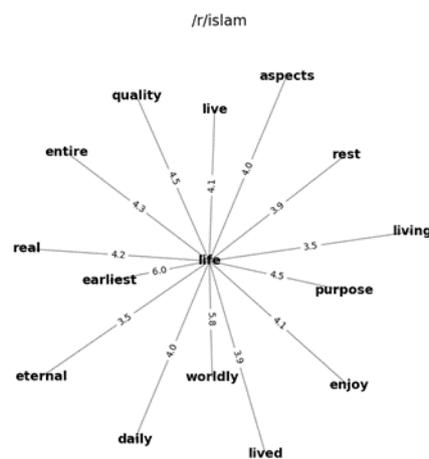


Figure 7. Collocations for “life” in /r/Islam

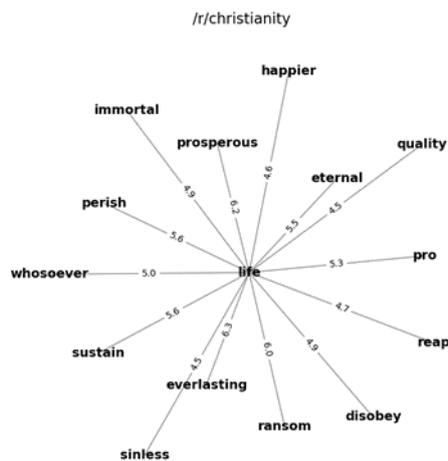


Figure 8. Collocations for “life” in /r/Christianity



Figure 9. Collocations for “life” in /r/Occult

In /r/Christianity, “life” is associated with words pointing to the afterlife (“everlasting”, “eternal”, “immortal”) while in /r/Islam, it is rather tied to terms describing a direction in life (“purpose”, “meaning”). However, both subreddits show connections with rather positive words except for death concepts. Those collocations are indeed stronger for /r/Islam (“rest”, “death”, “short”). The collocations are quite varied for /r/occult.

Overall, we were able to gather some first insights like the strong difference of /r/occult, connections to family and politics for some key concepts in the Muslim forum or the focus on discussions about religious directions for the concept of religion in all forums.

We plan to investigate other methods of computational text analysis but also want to apply more in-depth qualitative analysis of parts of our corpus via content analysis to confirm and evaluate some of our assumptions we derived via the collocation visualizations.

References

- Church, K. W., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*, 76–83. <https://doi.org/10.3115/981623.981633>
- Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015, May). On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In *EuroVis (STARs)* (pp. 83-103).
- McDonald, D. (2014). A text mining analysis of religious texts. *The Journal of Business Inquiry*, 13(1), 27-47.
- Moretti, F. (2000). Conjectures on world literature. *New left review*, 54-68.
- Pfahler, L., Elwert, F., Tabti, S., Morik, K., & Krech, V. (2018). What do you do with 5 million posts ? Versuche zum distant reading religiöser Online-Foren. *DHd Konferenz 2018*, 335-338.
- Slingerland, E., Nichols, R., Neilbo, K., & Logan, C. (2017). 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), 985-1016.
- Verma, M. (2017). Lexical analysis of religious texts using text mining and machine learning tools. *International Journal of Computer Applications*, 168(8), 39-45.