Semantic Stability in Wikipedia

Darko Stanisavljevic *, Ilire Hasani-Mavriqi *, Elisabeth Lex, Markus Strohmaier, and Denis Helic

Abstract In this paper we assess the semantic stability of Wikipedia by investigating the dynamics of Wikipedia articles' revisions over time. In a semantically stable system, articles are infrequently edited, whereas in unstable systems, article content changes more frequently. In other words, in a stable system, the Wikipedia community has reached consensus on the majority of articles. In our work, we measure semantic stability using the Rank Biased Overlap method. To that end, we preprocess Wikipedia dumps to obtain a sequence of plain-text article revisions, whereas each revision is represented as a TF-IDF vector. To measure the similarity between consequent article revisions, we calculate Rank Biased Overlap on subsequent term vectors. We evaluate our approach on 10 Wikipedia language editions including the five largest language editions as well as five randomly selected small language editions. Our experimental results reveal that even in policy driven collaboration networks such as Wikipedia, semantic stability can be achieved. However, there are differences on the velocity of the semantic stability process between small and large Wikipedia editions. Small editions exhibit faster and higher semantic stability than

Darko Stanisavljevic *
VIRTUAL VEHICLE Research Center, Inffeldgasse 21a Graz Austria, e-mail: darko.stanisavljevic@v2c2.at

Ilire Hasani-Mavriqi *
Graz University of Technology and Know-Center GmbH, Inffeldgasse 13, Graz, Austria, e-mail: ihasani@know-center.at

Elisabeth Lex
Graz University of Technology, Inffeldgasse 13, Graz, Austria, e-mail: elisabeth.lex@tugraz.at

Markus Strohmaier
GESIS and University of Koblenz-Landau, Unter Sachsenhausen 6-8, Cologne, Germany, e-mail: markus.strohmaier@gesis.org

Denis Helic
Graz University of Technology, Inffeldgasse 13, Graz, Austria, e-mail: dhelic@tugraz.at

* Both authors contributed equally to this work.

large ones. In particular, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level, whereas, in small Wikipedia editions, the number of needed successive revisions is much lower for the same level of semantic stability.

Key words: semantic stability, semantic similarity, TF-IDF, RBO, Wikipedia

1 Introduction

Wikipedia is one of the largest, freely accessible web-based encyclopedias and its content is open for editing by users. Wikipedia articles are mainly a contribution of volunteer editors who collaboratively create and manage the largest repository of human knowledge. This way, different editors can contribute with their expertise, ideas and opinions. Wikipedia contributors, however, may have different motivations and opinions, for example, it may take some time for them to agree if sufficient and correct information is provided within an article. If editors have different point of views on a particular topic, especially on controversial topics, they might end up overwriting each others content such that articles cannot become semantically stable. These are also known as edit wars [3, 5, 13, 17]. On the contrary, if Wikipedia editors achieve consensus on the content, implicitly, articles become semantically stable.

Problem & objectives. The goal of this paper is to investigate the semantic stability process in collaboration networks, such as Wikipedia, that are driven based on policies, guidelines and community standards. Based on these policies, both editors' behavior and the process of article production is managed [7].

Approach & methodology. In order to assess the semantic stability of Wikipedia, we turn to semantic similarity of consecutive revisions of Wikipedia articles. Semantic similarity of two textual documents expresses the extent to which two documents deal with semantically similar topics or content. This concept is key to understanding the comparison of documents written in natural language. Typically, semantic similarity is calculated by means of document statistics. An advantage of statistical approach is that it does not require predefined models, which describe the meaning of particular words (terms). The method applied in this work, i.e., Rank Biased Overlap, is also a statistical method and it is first introduced in [16]. The basic procedure carried out during the calculation of the semantic similarity is the modeling of the semantic space in accordance with the term distribution in a corpus of documents. In such a space, each document is represented by a vector and semantic similarity is calculated by performing vector operations on those vectors. This approach is based on the distributional hypothesis, according to which the terms with similar meanings show tendency to appear in similar contexts [8].

The concept of semantic stability applied in our paper is based on the work presented in [15], which studies the semantic stability of social tagging systems. In our work, we are interested in the semantic stability of Wikipedia. Thus, we take a Wikipedia corpus of documents that contains the complete edit history for each article and which includes all existing article revisions. The following Wikipedia language editions are used: English, German, French, Spanish, Italian, Czech, Finnish (Suomi), Danish, Greek and Swedish. The intention behind the choice of these particular languages is to have five Wikipedia editions with a large number of articles and five smaller editions. This enables us to study the relation between semantic stability and corpus size. Our long term goal is to investigate the consensus building process in Wikipedia based on the semantic stability. Authors of [15] state that semantic stability implies implicit consensus on the description of a resource in a social tagging system.

Findings & contributions. One of the contributions of our work is the software solution that we provide as an open source project[2], which is highly modular, configurable and flexible and can be applied by anyone looking for an efficient way to analyze the semantics of natural language documents contained, for example, in the Wikipedia XML dump files. From the empirical point of view, we conduct experiments in 10 different Wikipedia language editions and discuss the experimental results and their implications. Our experimental results reveal that the mean semantic stability of large Wikipedia editions is significantly lower compared to the mean semantic stability of small Wikipedia editions. In particular, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level, whereas, in small Wikipedia editions for the same level of semantic stability, the number of successive revisions needed, is much lower.

2 Technical Approach

2.1 Preliminaries

Particularly important for this paper is the theory describing: (i) evaluation of importance of terms in a single document or in a corpus of documents and their representation in a form of matrix - TF-IDF (Term Frequency - Inverse Document Frequency), (ii) calculation of semantic similarity measure and (iii) calculation of semantic stability over time.

We represent each revision of the parsed Wikipedia articles as a TF-IDF vector. *Term Frequency - Inverse Document Frequency* is one of the methods in the theory of Information Search and Retrieval used to represent the relevance of terms in a document belonging to a collection of documents - *corpus* [2, 9, 14, 18].

The comparison of the TF-IDF vectors is performed using a modified version of RBO (Rank Biased Overlap) method as in [15]. However, our approach is flexible and can be extended to include additional similarity measures. The RBO method is

[2] https://doi.org/10.5281/zenodo.153891

used to calculate the similarity measure of two given vectors, each of them representing the rankings of terms contained in a single Wikipedia article. Its main characteristic is that it takes the cumulative overlap of the given rankings as a measure for similarity. It is represented with the following mathematical equation:

$$RBO(\sigma_1, \sigma_2, p) = (1-p) \sum_{d=1}^{\infty} \frac{2 * \sigma 1_{l:d} \cap \sigma 2_{l:d}}{|\sigma 1_{l:d} + \sigma 2_{l:d}|} p^{(d-1)} \tag{1}$$

where $\sigma 1$ and $\sigma 2$ are not necessarily conjoint lists of ranking and $\sigma 1_{l:d}$ and $\sigma 2_{l:d}$ are ranked lists at depth d. RBO evaluates to a value in the range $[0,1]$, where 0 means disjoint and 1 means identical. The parameter p defines the steepness of the weights and takes a value in interval ($0 \leq p < 1$). When $p = 0$, RBO considers only the top ranked item of the lists and its value is either 0 or 1. When p is arbitrarily close to 1 the weights are almost the same for all depths and the analysis is arbitrarily deep.

The similarity measure described in Equation 1 is used as basis for determining the semantic stability over time. Based on [15], for a given value of RBO threshold k, an article is semantically stable if its RBO value at the point of time t is equal or higher than the threshold k. A rather simple mathematical formulation of this method for inspection of stabilization process in a given data set is as following:

$$f(t,k) = \frac{1}{n} \sum_{t=1}^{n} \begin{cases} 1, & \text{if } RBO(\sigma_{t-1}, \sigma_t, p) \geq k \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Based on the Equation 2, for each article in a Wikipedia corpus, the rank-biased overlap similarity measure is calculated. Inputs are the revisions before and after the time point t as well as the parameter p. If the calculated similarity is equal or greater than the threshold k, 1 is added to the sum, otherwise 0 is added. With no more articles in corpus to iterate, the sum is divided by the total number of iterated articles from the Wikipedia corpus. Thus, the result will be the percentage of the stable articles at time-point t for a predefined threshold value k.

For our experiments, the rank-biased overlap similarity measure algorithm is parametrized with the $p = 0.9$ which means that the first ten ranks of the ranking list have 86% of the weight of the evaluation as stated in [15]. Empirically, we also find that $p = 0.9$ is appropriate because of the value of parameter d (depth of evaluation) chosen for rank-biased overlap. This means that the TF-IDF vectors will be checked for similarity only up to the depth of 20. Of course, one can take a much higher depth, but that will increase the computation time as well as the storage space. Namely, the TF-IDF vector representing a single revision of an arbitrary article can have several thousands of values, but not all of those values are stored. Only the values up to the depth needed for rank-biased overlap calculation are stored. So, if 20 elements are used for rank-biased overlap measure, the first 10 elements of the ranking weight 86% of the evaluation and the other 10 elements weight only 14%. It is exactly because of this fact that there is no need to do the similarity calculation for much higher depths as those are not regarded as very important. In every case, the top 20 (most-weighted) elements of the TF-IDF vector are more than enough to precisely describe the semantics of the article revision they represent.

2.2 Experimental Setup

We study two different aspects of the stabilization process: (i) semantic stabilization of the Wikipedia corpus over a predefined period of time and (ii) semantic stabilization of the Wikipedia corpus after a number of successive revisions. The idea behind the examination of the Wikipedia corpus stabilization over the time is to choose a point in time t and count the number of articles existing at that point in time and the number of articles existing at that point in time that are also semantically stable. This is possible because of the fact that every article revision is uniquely identified in the database by the compound key consisting of the article ID and the revision timestamp.

Another way to inspect the stabilization process of the document corpus is to find out how many successive revisions are required before a percentage of the available articles becomes stable (in reference to the stability threshold). The idea is very similar to the previously discussed one, but now it is assumed that all articles have the first revisions starting at the same date and time. The timestamp information is now completely neglected and only the number of revisions per article is important. So, at the beginning, the first value of the similarity vectors of all articles is examined. The stability threshold takes the maximal value at the beginning of the calculation, 1. If the desired percentage of the articles is stable, the next value of the similarity vector is inspected. If not, the threshold is decreased and the calculation is repeated until the value of the stability threshold, for which the desired percentage of articles is stable, is found. Analysing the semantic stability from two different point of views, provides more useful insights about the examined corpus.

Dataset Preprocessing. The Wikimedia[3] provides XML dumps of all active Wikipedia projects. The basic building block of all Wikipedia editions is a page. Every page represents an article and every article has at least one, but usually more than one, revision. There are articles in bigger Wikipedia editions which have tens of thousands of revisions.

We analyze 10 Wikipedia language editions, five of which are (randomly selected) small language editions and the remaining five are the largest language editions. Our goal is not to analyze the full Wikipedia corpus of the large editions, thus, the sampled data of 10 thousand randomly selected articles with their complete revision history is used for 8 out of 10 Wikipedia editions. Only Czech and Finnish Wikipedia corpus is fully analyzed.

[3] https://dumps.wikimedia.org/

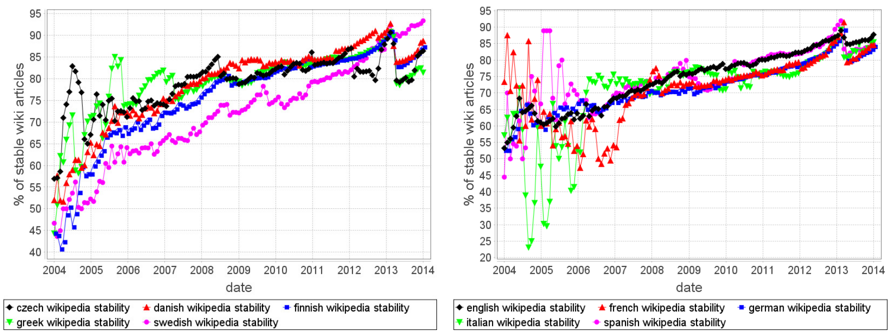
3 Results and Discussion

Figure 1 compares the stabilization process between small and large Wikipedia language editions over a period of time. A portion of the stable articles (in percentages) is shown for a chosen point in time t, in order to spot periods of increased stability or instability of an article corpus. The plots in Figure 1 correspond to the RBO threshold $k = 0.8$. We run experiments with two other values: $k = 0.4$ and $k = 0.6$, to investigate the role of the threshold parameter k in the stability calculation method proposed in [15]. Once the similarities of all revisions of a single Wikipedia article are calculated, the value representing the similarity in a given moment of time t is taken and compared to the value of the parameter k. Our intuitive assumption is that, for a low value of RBO threshold k, there are a lot of articles in the examined corpus, whose stability value in a given instant of time is higher than the chosen threshold. Our results are consistent with our initial assumptions. Thus, as the value of the RBO threshold increases, the number of stable articles decreases. The document corpus stability is inversely proportional to the value of parameter k. However, the steepness of the stabilization curves remains the same over different parameters k, thus, we include plots for only $k = 0.8$ to show the least stability.

From the plot in Figure 1a, it is noticeable that all small Wikipedia editions exhibit semantic stability variations in almost the same range (with a deviation $\pm 2\%$ from the average). The only exception to this is the case of Swedish Wikipedia that has the semantic stability well below the average semantic stability of the other four small Wikipedia editions.

Figure 1b shows that in large Wikipedia editions, semantic stabilization curves oscillate more at the beginning of the editorial process compared to small editions. Thus, they are, on average, more unstable than the small Wikipedia editions. Our explanation for this is that the small Wikipedia editions consist mainly of articles which are the translated versions of the articles from the main Wikipedia editions (for example from the English Wikipedia). Once translated and created, such articles are rarely edited a lot. Whereas, in large editions such as in the English one, a higher number of new articles that are authored from scratch is present. Of course, the editorial process of such articles is more dynamic.

We observe a very interesting phenomenon in both plots in Figure1, namely, in both small and large Wikipedia editions, a sudden increase of the semantic stability is noted, with a peak around year 2013. Right after this point of time, the stability decreases for all Wikipedia editions and than continues to increase again. We wanted to find an explanation for this observation by contacting the Wikipedia community by writing several posts in the *Wikimedia.org*[4]mailing list, but we did not receive any plausible answer. Some of the assumptions are that: some of the Wikipedia servers were down for a short maintenance, or some of the Wikipedia maintenance bots were active and editing Wikipedia contents was shortly blocked or malfunctioning of Wikipedia servers was induced by malicious software or hacker attacks. But, the temporary peak in semantic stability in year 2013 could also be seen as a consequence of a change in Wikipedia policies of how to handle edit wars (e.g, the



(a) Small Wikipedia editions (b) Five largest Wikipedia editions

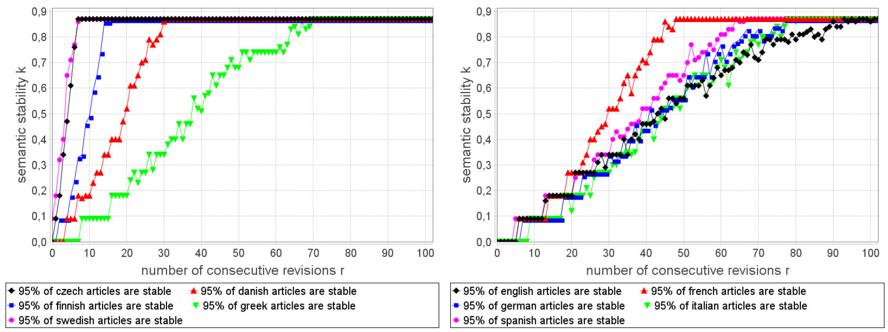
Fig. 1: **Semantic stabilization of the Wikipedia corpus over a period of time.** Percentages of stable articles (y-axis) are shown in relation to a predefined period of time (x-axis) for **(a)** small and **(b)** large Wikipedia editions. Semantic stability curves shown, correspond to the RBO threshold $k = 0.8$ and steepness parameter $p = 0.9$. For illustration, consider the plot in (a), for a chosen point in time, (e.g.,) year 2008, in (e.g.,) Czech edition, is indicated that 70% of articles have reached a semantic stability equal or higher than 0.8. The steepness of the stabilization curves remains the same over different parameters k, however, the percentage of stable articles decreases with increasing k. Comparing plots in (a) and (b), one can see that the mean semantic stability of small Wikipedia editions is significantly higher in contrast to large ones. This is in line with the fact that small Wikipedia editions contain large portions of articles simply translated from the English Wikipedia, for example. Such articles are usually rarely changed substantially and they increase the overall stability of small editions. In contrary, the editorial process in large editions is much more dynamic.

introduction of a new rule such as the three-revert rule). Still, no hard evidence was brought into light.

Figure 2 visualizes the number of consecutive revisions per article needed to achieve the stability of 95% in both small and large Wikipedia editions. This means that 95% of articles in a corpus become semantically stable, evaluated based on different RBO (for $p = 0.9$) thresholds k (y-axis in Figure 2), after r consecutive revisions (x-axis).

In Figure 2a, 95% of stable articles is reached after, for example, 70 revisions for the Greek Wikipedia and 30 or less revisions for all other small Wikipedia editions. It can be seen that for the Greek Wikipedia, 95% of the articles has the stability of 0.5 or higher after almost 35 revisions, where $k = 0.5$ is considered as a medium stability [15]. From this fact one can conclude that the Greek Wikipedia edition is the most frequently edited one amongst the analyzed small editions. The Czech

[4] https://lists.wikimedia.org/mailman/listinfo/wiki-research-l



(a) Small Wikipedia editions

(b) Five largest Wikipedia editions

Fig. 2: **Semantic stabilization of the Wikipedia corpus after a number of successive revisions.** 95% of articles in a corpus become semantically stable, evaluated based on different RBO thresholds k (y-axis), after r consecutive revisions (x-axis). The plot in (a) illustrates that almost all small editions exhibit, at the beginning, a fast increase of the stabilization curves, which remain relatively stable after few successive revisions. An exception presents the Greek edition, which is the most frequently edited among the small ones. The plot in (b) depicts that the stabilization process in large editions is delayed. This indicates that in large editions a higher number of successive revisions is needed in order to reach the same semantic stability level as in small Wikipedia editions. These results are consistent with the fact that the size of the community contributing to the large editions, such as English, can not be compared to the small ones. Large communities are characterized with heterogeneous contributors' expertise, motivation and opinions, which implicates that it takes time until contributors agree if sufficient and correct information is provided within an article.

and Swedish editions are showing much more semantic stability. 95% of the article corpus of this two editions has the semantic stability of 0.5 or higher after only about 5 revisions.

Figure 2b shows the stabilization process of large Wikipedia editions where the achieved stability is 95%. This time, as expected, the English Wikipedia is the most unstable one. Almost the complete corpus of analyzed articles becomes stable after almost 95 revisions of each article. The medium semantic stability of the corpus that is defined by the value of parameter $k = 0.5$ is, in the case of English Wikipedia, reached after about 45 revisions, and in the case of the French one (the most stable one) after about 30 revisions.

These results are in line with the fact that larger communities contribute to the largest Wikipedia editions (e.g., English, German or French), in comparison to the communities editing the small Wikipedia editions, written in languages, which are only used by a very small percent of the world population. Large authoring community indicates a heterogeneous community based on authors' expertise, ideas and opinions, which in turn implies that the contributed content is more colorful. If content contributors have different point of views on a particular topic, especially on controversial topics, they might end up overwriting each others content such that articles cannot become semantically stable. Thus, in large Wikipedia editions a higher number of revisions is needed until contributors agree if sufficient and correct information is provided within an article.

Key findings. Our findings can be summarized as follows: even in policy driven collaboration networks such as Wikipedia, semantic stability can be achieved. However, there are differences on the velocity of the semantic stability process between small and large Wikipedia editions. In large Wikipedia editions, semantic stability curves oscillate more at the beginning of the editorial process compared to small editions. Thus, the mean semantic stability of large Wikipedia editions is significantly lower in contrast to small Wikipedia editions. In other words, small Wikipedia editions stabilize faster and achieve higher levels of semantic stability.

4 Related work

The process of consensus reaching among Wikipedia editors has been on the focus of many recent studies [1, 3, 5, 6, 7, 13, 17]. Authors in [5] study the problem of edit wars in Wikipedia and model this phenomenon using agent-based systems, based on theories of group stability and reinforcement learning. Authors show that consensus is reached faster if the number of credible or trustworthy agents and agents with a neutral point of view is increased. In the contrary, consensus is hindered when agents with opposing views are in equal proportion. Similarly, authors in [13] apply also an agent-based model to emulate conflict scenarios in edit wars and validate their model by empirical Wikipedia data. Recently published work [3] uses hidden Markov models to approximate and characterize the computational structure of conflicts in Wikipedia.

The work presented in [7] investigates the role of conflict in the editorial process in Wikipedia by studying talk pages. Experimental results reveal that conflict is central to the editorial processes of Wikipedia; it is a generative friction that is used by Wikipedia editors as part of a coordinated effort within the community to improve the quality of articles.

There are several research approaches published in the field of semantic similarity measurements [4, 10, 11, 12]. Hajian et. al. [4] propose a multi-tree similarity algorithm as a non-linear technique for measuring similarity based on hierarchical relations which exist between attributes of entities in an ontology. This method compensates for the lack of semantic relatedness among features using taxonomic relations that exist among the features of two entities. In [10] authors implement a probabilistic method of measuring semantic similarity for real-world noisy short texts like microblog posts. Their method adds related Wikipedia entities to a short text as its semantic representation and uses the vector of entities for computing semantic similarity. The work presented in [11] shows that the combination of knowledge and corpus-based word-to-word similarity measures can produce higher agreement with human judgment than any of the individual measures. Authors in [12] present an approach for measuring semantic similarity between words using the snippets returned by Wikipedia and the five different similarity measures of association. Their results demonstrate that the snippets in Wikipedia have a significant influence on the accuracy of semantic similarity measure between words.

The Rank Biased Overlap or shortly RBO method is introduced in [16]. Our study is based on the scientific work [15], in which a modified version of RBO is applied to investigate the semantic stability of social tagging systems. However, in our work we assess the semantic stability of Wikipedia articles.

5 Conclusion and Future Work

In this work, we study the semantic stabilization of Wikipedia with a focus on the dynamics of Wikipedia articles' revisions over time. Our experimental results reveal that: (i) the analyzed Wikipedia language editions show medium semantic stability and (ii) large Wikipedia editions exhibit a significantly lower mean semantic stability value compared to the small Wikipedia editions.

Our first findings are in line with the research results of the work presented in [15], in which authors state that natural languages are semantically stable in their nature. In our case, all the analyzed datasets have at least medium semantic stability.

Our second experimental results indicate that the large Wikipedia editions, which were utilized for the purpose of this paper are semantically less stable than the small ones. This observation can be logically explained by the fact that large Wikipedia editions have much more contributors than the small ones. The sheer size of the community supporting and developing the English Wikipedia edition cannot be compared to e.g., the size of community working on the Czech Wikipedia edition. Having many more users contributing to the content means that higher semantic instability is brought to the system. The users of English Wikipedia are changing the content of the articles much more than the users of small Wikipedia editions. Additionally, many articles available in small Wikipedia editions are simply translations of the articles found in the English Wikipedia. Once translated, such articles are rarely changed significantly, which contributes to a higher semantic stability of the small Wikipedia editions.

One of the limitations of our work is that we evaluated only sampled data for the large Wikipedia editions. However, our software solution is flexible and could be easily extended to analyze the full Wikipedia corpus of the large editions.

For future work, we plan to investigate the consensus building among editors in different Wikipedia categories, in order to find out if there are categories that are unstable. We also want to specifically study the semantic stability of articles marked as controversial. One of our future plans is to combine the content based approach introduced in this work with a network based approach. Vandalism detection is also a topic that could benefit from our work.

Acknowledgments. This work is supported by the Know-Center, the VIRTUAL VEHICLE Research Center and the AFEL project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 687916. The Know-Center and the VIRTUAL VEHICLE Research Center are funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

References

1. Biancani, S.: Measuring the Quality of Edits to Wikipedia. In: Proceedings of The International Symposium on Open Collaboration, OpenSym '14. ACM, New York, NY, USA (2014)
2. Debole, F., Sebastiani, F.: Supervised Term Weighting for Automated Text Categorization. In: Proceedings of the 2003 ACM Symposium on Applied Computing, SAC '03. ACM, New York, NY, USA (2003)
3. DeDeo, S.: Conflict and computation on wikipedia: A finite-state machine analysis of editor interactions. Future Internet **8**(3) (2016)
4. Hajian, B., White, T.: Measuring Semantic Similarity using a Multi-tree Model. In: Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, ITWP 2011. CEUR Workshop Proceedings (2011)
5. Kalyanasundaram, A., Wei, W., Carley, K.M., Herbsleb, J.D.: An Agent-based Model of Edit Wars in Wikipedia: How and when is Consensus Reached. In: Proceedings of the 2015 Winter Simulation Conference, WSC '15. IEEE Press, Piscataway, NJ, USA (2015)
6. Müller-Birn, C., Dobusch, L., Herbsleb, J.D.: Work-to-rule: The Emergence of Algorithmic Governance in Wikipedia. In: Proceedings of the 6th International Conference on Communities and Technologies, C&T '13. ACM, New York, NY, USA (2013)
7. Osman, K.: The Role of Conflict in Determining Consensus on Quality in Wikipedia Articles. In: Proceedings of the 9th International Symposium on Open Collaboration, WikiSym '13. ACM, New York, NY, USA (2013)
8. Sahlgren, M.: An Introduction to Random Indexing. In: Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE '05. Copenhagen, Denmark (2005)
9. Salton, G., Buckley, C.: Term-Weighting Approaches in Automatic Text Retrieval. Information Processing and Management: an International Journal **24**(5) (1988)
10. Shirakawa, M., Nakayama, K., Hara, T., Nishio, S.: Probabilistic semantic similarity measurements for noisy short texts using Wikipedia entities. CIKM '13. ACM, New York, NY, USA (2013)
11. Stefanescu, D., Rus, V., Niraula, N.B., Banjade, R.: Combining Knowledge and Corpus-based Measures for Word-to-Word Similarity. In: Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference, FLAIRS '14. AAAI Press, Palo Alto, California (2014)
12. Takale, S.A., Nandgaonkar, S.S.: Measuring Semantic Similarity between Words Using Web Documents. International Journal of Advanced Computer Science and Applications, IJACSA **1**(4) (2010)
13. Török, J., Iñiguez, G., Yasseri, T., San Miguel, M., Kaski, K., Kertész, J.: Opinions, conflicts, and consensus: Modeling social dynamics in a collaborative environment. Phys. Rev. Lett. **110** (2013)

14. Turney, P.D., Pantel, P.: From Frequency to Meaning Vector Space Models of Semantics. Journal of Artificial Intelligence Research **37**(1) (2010)
15. Wagner, C., Singer, P., Strohmaier, M., Huberman, B.A.: Semantic Stability in Social Tagging Streams. In: Proceedings of the 23rd International Conference on World Wide Web, WWW '14. ACM, New York, NY, USA (2014)
16. Webber, W., Moffat, A., Zobel, J.: A similarity Measure for Indefinite Rankings. ACM Transactions on Information Systems, TOIS **28**(4) (2010)
17. Yasseri, T., Kertész, J.: Value production in a collaborative environment. Journal of Statistical Physics **151**(3) (2013)
18. Zaman, A.: Stop Word Lists in Document Retrieval Using Latent Semantic Indexing: an Evaluation. Journal of E-Technology **3**(1) (2012)