Can Plausibility Help to Support High Quality Content in Digital Libraries?

José María González Pinto(✉) and Wolf-Tilo Balke

pinto@ifis.cs.tu-bs.de balke@ifis.cs.tu-bs.de
Institut für Informationssysteme, Mühlenpfordstrasse 23
38106 Braunschweig, Germany
+49(531) 391 3271

Abstract. Presented herein is a novel approach to support high quality content in Digital Libraries by introducing the notion of *Plausibility* of new scientific papers when contrasted with prior knowledge. In particular, our work proposes a novel assessment of scientific papers to support the workload of reviewers. The proposed approach focus on a core component of a scientific paper: its claim. Our methodology exploits state of the art neural embedding representation of text and topic modeling on a Digital Library of scientific papers crawled from PubMed. As a proof of concept of the potential usefulness of the notion of Plausibility, we study and report experiments on documents with claims expressed as statistical associations. This type of claims is very often found in medicine, chemistry, biology, nutrition, etc. where the consumption of a drug, substance, product, etc., has an effect on some other type of entity such as a disease, another drug, substance, etc.

Keywords: plausibility, information discovery, quality assessment

1 Introduction

For years, digital libraries have been a valuable and trustworthy source of information due to the carefully *curated quality* of their content. Since collections are continuously growing with increasing publication numbers, the main challenge to preserve content quality lies in the inclusion of new articles in a collection. Today, peer review is the key to assess new articles and thus help digital libraries preserve high quality content. However, with increasing numbers of publications reviewers are facing the problem of workload scalability: there is less and less time to do this valuable and necessary task. This has also been recognized by the community[1] and while nobody has a perfect solution there are many approaches to at least aid the process, such as expertise profiling, matching submissions with possible reviewers, or resolving paper biddings. In this work, we aim at supporting peer review not at the process level, but with a clear focus on document level. We aim at assessing a new scientific paper's *Plausibility* in the light of prior knowledge represented by some digital library collection. With this novel assessment, the question of how many reviewers a new paper needs can be adjusted by its respective degree of Plausibility: the less plausible it is (i.e. the more its inclusion

adfa, p. 1, 2011.

would hurt the collections consistency), the more reviewers might be needed to come to a clear decision.

The notion of Plausibility in our work is based on the *knowledge-fit theory* from cognitive sciences [1]. Basically, it states that human plausibility judgements consist of two steps: firstly, a mental representation of current knowledge is built and secondly, an assessment examines how a new piece of information fits all prior knowledge. Of course, this is very hard to decide in general settings. Thus, we will focus our work on a particular type of documents to provide first insights on the general feasibility of the idea: in particular, we focus on documents containing *empirical claims* in the sense of *statistical associations between entities*. Empirical claims thus are given by sentences that express some kind of association between two entities and in what way one affects the other. Actually, our research shows that this simple type of claims can be found in many scientific papers: consider for instance medicine, chemistry, biology, nutrition, etc. where the consumption of a drug, substance, product, etc., has an effect on some other type of entity such as a disease, another drug, substance, etc.

What makes exactly this type of claims so interesting are findings like those reported by nutritional researchers in [2]. Basically, for 50 common basic foods the researchers performed literature searches using PubMed to obtain articles investigating the association between each ingredient and the respective cancer risk. To their surprise, 80% of the ingredients were indeed related to cancer risk. But what was even more surprising: out of 264 single-study assessments 191 (72%) concluded that the tested food was associated with an increased (n=103) or a decreased (n=88) risk *at the same time* [2]. What does that say about the concept of Plausibility? How can we account for this type of situations and still provide a consistent instantiation of Plausibility over digital libraries? And how many of this type of claims are there anyway?

As opposed to the first two questions, the last one is easy to answer. To estimate this number, we used a similar linguistic query pattern as in [3]: *(help AND prevent) OR (lower AND risk) OR (increase OR increment AND risk) OR (decrease OR diminish AND risk) OR (factor AND risk) OR (associated AND risk)*. Even with this simple filter there are currently almost 1 million articles in PubMed with empirical claims in the form of statistical associations. Figure 1 provides the cumulative number of articles per year. We can even observe a clear increase of the number of articles dealing with empirical claims every year.

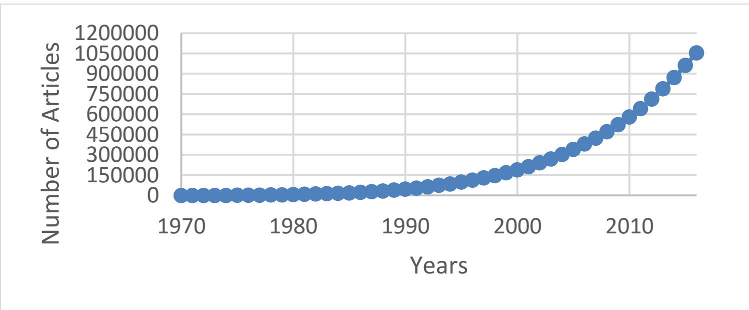

Fig. 1. Accumulated number of articles containing empirical claims in PubMed per year.

To tackle the challenge of the first two questions, in this paper we develop a data driven approach relying on a novel integration of state of the art neural embedding representation of text and generative topic models to operationalize the concept of Plausibility. Our goal is to provide a way to assess the consistency of each new document with respect to the current knowledge (i.e. the state of the art) so that we can answer questions such as: is a new document consistent with current knowledge? Do we have documents in our collection supporting or contradicting a new document? And can we represent our collection in a way such that we can derive a decision to reflect the consistency of new knowledge in the light of current knowledge?
To accomplish this, we first need to operationalize the concept of Plausibility. As a proof of concept, we then implement a new architecture integrating these ideas and providing first insights by analyzing empirical claims. In summary, our contributions are:

1. Firstly, a representation of document collections that combines topic modeling with a neural embedding to exploit two relevant metadata elements: conclusions and abstracts.
2. Secondly, a query facility to find semantically similar claims that may support or contradict a new document's claim.
3. And thirdly, a mechanism to finally assess the total Plausibility of a new document, e.g. to verify its consistency with respect to a collection's representation.

Our paper is organized as follows: in Section 2 we provide relevant related work. We then propose a general architecture with the formalization of Plausibility in Section 3. In Section 4 we present the experimental setting to evaluate our proposed solution with a discussion of our findings. Afterwards, we present concluding remarks and outline future work in Section 5.

2 Related Work

Many attempts to model arguments for different purposes exist in literature. Particularly relevant for our work is the body of research dealing with the semantic annotation of claims of scientific articles in the biomedical domain. For instance, in [3], a model is developed for the annotation of scientific hypotheses and claims in natural language using as a case study Alzheimer Disease. Nanopublications [4–6], promoted by the Concept Web Alliance, models core scientific statements with associated context and it is used for data integration across chemical and biological databases[7]. A more detailed model of scientific papers in the biomedical domain is Micropublications [8]. The model specified as an OWL 2 Vocabulary (the ontology language for the Semantic Web1) is developed around the idea that scientific claims are defeasible arguments [9, 10]. Thus, they support natural language statements, data, methods, materials specifications, discussion, challenge, and disagreement. In our work, we built on these ideas and represent one core component of scientific papers: claims. Moreover, we attempt

[1] https://www.w3.org/TR/owl-overview/

to operationalize the notion of Plausibility from Cognitive Sciences. In particular, a Plausibility theory that has been empirically proven to be strongly correlated with human judgements[11].

3 Methodology

In this section, we formalize the concept of plausibility and the problem we aim to solve. Plausibility in this work builds on the knowledge-fit theory from Cognitive Sciences. The theory states that human judgements consist of two steps: firstly, a mental representation of prior knowledge that allow us to comprehend and make sense of the world; secondly, assess how new knowledge fits this prior knowledge. Therefore, to operationalize Plausibility we need to formally define a) how to represent our current knowledge of a Digital Library and b) how to determine Plausibility of a new document given a).

Let's revisit the findings of [2] to better explain the rationale behind our proposed methodology. In [2] it was found that for some substances, there were papers that concluded that a given substance was a factor that increases the risk of cancer, while some other papers studying the same substance concluded the opposite. Our proposed decision process that accounts for this type of situations is as follows: if a new document agrees with our current view it is considered "plausible", otherwise "not plausible". Let's explain the difficult situation: if in our Digital Library, we have documents with claims that at the same time contradict as well as support the new claim, we decided to label the case as "controversial". At this point, we can stop and deliver a weighted claim measure based on the semantic similarity of the claims that support versus those that contradict the new claim. However, our hypothesis is that we can still do better: try to identify if the context of the documents in our Digital Library exhibit some characteristics that makes them belong to different groups e.g. "possible worlds". If we can find a possible world where the new document will fit and in this world the claims agree e.g. the world is consistent, then we can proceed to assess the plausibility of the new document as before. Otherwise, if the possible world is inconsistent, then we again have the "controversial" situation.

To operationalize this process, we hypothesized that we need to provide our Plausibility data driven approach with a representation capable of capturing two relevant and related components: a) the semantics of the relationship of the entities in the collection and b) how each claim context exhibit certain characteristics that makes it part of a possible world within the collection. Thus, towards this goal, we turn our attention to two aspects to assess the Plausibility of a new article:

1. *Claim of the paper*: the statement(s) of the contribution(s) of the paper. In a curated Digital Library, this is usually found in the conclusion metadata.
2. *Context of the claim of the paper:* the surrounding text of the claim that provides the "explanation" of how the authors of the paper reached the claim(s) stated in the paper.

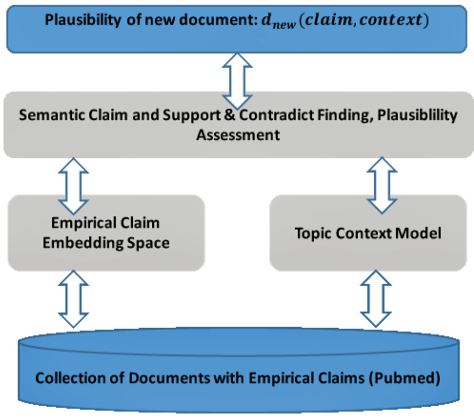

Fig. 2 Architecture of Plausibility.

Formally, we can define our Plausibility problem as:

Definition 1. Document Plausibility Problem: given a document collection with empirical claims $D = \{d_1 \dots d_n\}$, and a new claim in document d_{new}, we aim at finding how *consistent* is the claim in d_{new} with respect to D.
We approach this problem by breaking it down into three tasks:

1. Representation of the collection of documents D.
2. Finding documents in the representation of D with semantic similar claims that support or contradict a new claim in document d_{new}.
3. Calculate the Plausibility of the claim of d_{new}.

In Figure 2, we show the components of our proposed methodology that we further explain in this section.

3.1 Representation

Let's begin with how we model a document collection in our work. We consider a collection of documents $D = \{d_1 \dots d_n\}$, where each document d_i in D is a tuple $(Claim, BagOfContext)$.
Where:

- $Claim$ is a sentence that represents an empirical claim. In other words, a sentence that contains an association between two entities and in particular, how one of them affects the other. In this work, we use the conclusion metadata of each paper to find such sentences.
- $BagOfContext$ is a vector space model representation of the context of the claim. The context in this work is the abstract of the paper.

Empirical Claim Embedding Space. Because of the relevance in our work of the empirical claims, we use Neural Network language models to compute an embedding representation of them [12–15]. Embedding language models have shown interesting semantic properties to be able to find related concepts, related paragraphs, analogies, etc. [14, 16]. In this work, we rely on such representations to capture claim specific semantics. Moreover, we use this representation to find not only semantically similar claims but also to distinguish between claims that express supporting or contradicting positions with respect to the claim of the document we want to assess. In our experiments, we use the embedding implementation of [14]. It is relevant to mention that we decided to use the embedding space because it benefits our approach to find highly related claims. The idea is that entities used in similar contexts with respect to the effect on another entity are related and might help in the absence of explicit knowledge. In our experiments, we first train the word embedding in the entire collection of documents over the abstracts. Then, every claim is represented as a weighted point of embedded words. It is this representation that is used to query the embedding space over other claims to assess the Plausibility of a new document.

Topic Context Model. We use a generative probabilistic model to represent the $BagOfContext$ of each $d_i \in D$. In particular, the Latent Dirichlet Allocation (LDA). This model is an instance of a general family of mixed membership models for decomposing a collection into multiple latent components (topics). In LDA it is assumed that words of each document arise from a mixture of topics, where each topic is a multinomial over a fixed vocabulary. The topics are shared by all documents in the collection, but the topic proportions vary stochastically across documents, as they are randomly drawn from a Dirichlet distribution[17, 18]. In this work, we employ this powerful representation to operationalize the idea of possible worlds to account for cases where we find claims that support and contradict each other in our collection. In particular, our hypothesis is that by the instantiation of the possible world idea, we can provide additional insights to understand what we have called "controversies". Thus, we operationalize the idea of possible worlds, with a representation of the context of a claim given by its latent mixture of topics.

3.2 Finding semantic similar claims

Finding similar claims in our work is a crucial step given the embedding space representation. Moreover, given the claim of a document we would like to assess, we also need to distinguish between claims that support or contradict it. For this task, we proceed as follows: given the embedding space of claims, we first find similar claims by computing distance similarities as in [19, 20]. Because this distance is highly efficient in using the embedding space against some other alternatives[19], we rely on it in our first step. In Figure 3 we show an illustrative example of the semantics captured in the claim embedding space as given by the distance computation named Word Mover's Distance (WMD) of [19]. Please observe how entities such as "tomato sauce" and "lycopene" end up close to each other in the embedding space because of the semantics captured by the WMD. Thus, instead of endless list of synonyms, we rely on this type of representation to find similar entities used in similar contexts. Next, we distinguish between supporting or contradicting claims. As a proof of concept, we focus on claims expressing “increase” vs “decrease” associations that express clearly contradictory positions. Thus, for the claims that we investigate in our experiments we distinguish these two positions. A simple textual pattern mechanism with synonyms to the words “increase” and “decrease” were used to distinguish between the two. Synonyms in this work are words related to “increase” and “decrease” as captured in the embedding space.

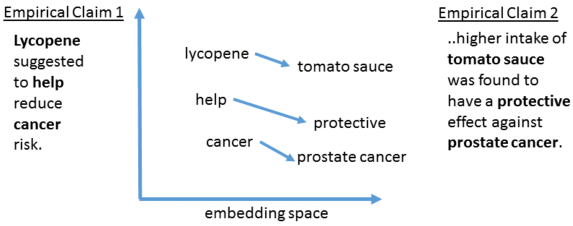

Fig. 3. Illustrative example of the power of the embedding space using WMD distance.

3.3 Computing Plausibility

In this section, we formally define how we determine the Plausibility of a new document. Plausibility in our work is the consistency of a new claim with current body of knowledge in terms of its agreement at the claim level or at the context level when a possible world consistency holds. Let d_{new} be a document that is currently not in our collection and we would like to know how plausible it is, given our current knowledge. Let $ClaimOf(d_{new})$ be the claim of document d_{new}. Let $DocSimClaim(d_{new})$ be the set of documents dealing with semantic similar claims to $ClaimOf(d_{new})$. Moreover, let $DocsContradict$ and $DocsSuppport$ be the documents that contradict and support respectively the $ClaimOf(d_{new})$. The distinction between these two sets of documents is given by finding first if the claim is in an “increase” or “decrease” association. Afterwards, we just map every claim found to one of the two groups. To be able to determine the Plausibility of the new document, we proceed as follows:

1. If $DocsContradict$ is empty and $DocsSuppport$ is not, then $ClaimOf(d_{new})$ is plausible.
2. If $DocsContradict$ is not empty and $DocsSuppport$ is empty, then $ClaimOf(d_{new})$ is not plausible.
3. If both $DocsContradict$ and $DocsSuppport$ are not empty, we initiate our quest for a possible world consistency:
   a. First, we need to find how document d_{new} would fit in our Topic Model representation of the collection. Let $topicOf(d_{new})$ be that topic. This is found by posterior inference and selecting the topic with the highest probability value.
   b. Second, if in $topicOf(d_{new})$ the claims agree, we call the world consistent and proceed to verify the consistency of $ClaimOf(d_{new})$. If the

possible world is not consistent, we declare a "controversial" situation regarding the claim of the new document.

4. If $DocsContradict$ and $DocsSuppport$ are both empty, then $ClaimOf(d_{new})$ calls for a special assignment of resources to manually assess its value.

4 Experiments and Findings

To demonstrate and evaluate our proposed Plausibility measure, we performed experiments with two primary goals. Firstly, we wanted to gather valuable insight into the notion of finding similar semantic claims in our corpus that either support or contradict a new document's claim. Because we do not have a ground truth, we manually observed the claims and set a threshold that to the best of our understanding can lead to highly related claims that may or may not support each other with respect to a disease. After this experimentation, we set a threshold of 0.50 for the experiments that we report here. Secondly, as a proof of concept, we needed to compare our approach to experts work. In particular, we chose the results reported in [2] that we mentioned in the Introduction. Thus, we first retrieved all the documents related to two of the ingredients reported in [2] salt and lycopene. To retrieve the documents, we used the query pattern mentioned in our Introduction. We chose these two cases to acknowledge the scope of our tool since they represent different situations: salt was found to be one of the few exceptions of the analysis regarding its risk effect that was not subject of "controversy" due to contradicting findings. On the other hand, lycopene represent a situation that cannot be plausible from our perspective: the increase/decrease effect at the same time. Thus, the goal of studying these two cases was to see if we could find a suitable explanation. For every document, we extracted the empirical claims contained in the conclusions section when available. Unfortunately, not all the documents contained this valuable metadata. We just ignored them and put it aside the task of metadata generation that is necessary to improve the knowledge representation of our work.

Thus, the collection of documents used for these experiments consist of 87k documents. We used this collection to train our embedding representation using default parameters as given by open source project Gensim[21]. For every experiment, we first selected one document at random and considered it as the new document and proceeded to assess its Plausibility.

The first case that we report here is the association of "salt" and "cancer". The document to evaluate was "*Salt intake and gastric cancer risk according to helicobacter pylori infection, smoking, tumour site and histological type*"[22]. The claim of the paper that we used to query our semantic embedding space is a very simple one: "*our results support the view that salt intake is an important dietary risk factor for gastric cancer, and confirms the evidence of no differences in risk according to h. pylori infection and virulence, smoking, tumour site and histological type*." After querying our semantic embedding space, we retrieved some related claims in other papers. Some examples are:

- "Dietary salt intake was directly associated with risk of gastric cancer in prospective population studies, with progressively increasing risk across consumption levels."
- "Improved dietary habits, reducing salt consumption and eradication of h. pylori infection may provide protection against gastric cancer in Turkey."
- "These data suggest that high intake of salt and smoked and pickled food may be associated with a high risk of gastric cancer, and this association could be due to intragastric formation of nitrosamines"

In this particular case, the "new" claim finds support in our current knowledge and our approach states a "plausible" situation. This is an example of how our approach could help a reviewer to assess if the new document is consistent with current body of knowledge. Basically, it could allow the reviewer to find similar studies dealing with the specified entities probably in similar ways.

Next, we report on a second experiment. In this second experiment we take "lycopene" as one of the ingredients where there was evidence of being in a situation that we call "controversial". Remember that "controversial" means that "lycopene" was found in an increased and decreased risk association in different research papers. In this particular case, we found 197 documents related to the association within our collection. We selected a document with the following claim: *"this study does not support a role for lycopene in prostate cancer prevention"*[23].

We found in our collection claims that both support and contradict the new document's claims. This leads to a controversial situation as defined in our methodology section. However, we can take our second step to make a final decision. In this case, the new document fits better in a possible world that is not consistent. Thus, we conclude that this is a controversial situation in need of human experts to carefully look into the new document. One should notice the level of complexity of this case. For instance, in the community curated archive Wikipedia entrance of lycopene[2], one can find the following: The FDA (US Food and Drug Administration), in rejecting manufacturers' requests in 2005 to allow "qualified labeling" for lycopene and the reduction of various cancer risks, stated:

"...no studies provided information about whether lycopene intake may reduce the risk of *any of the specific forms* of cancer. Based on the above, FDA concludes that there is no credible evidence supporting a relationship between lycopene consumption, either as a food ingredient, a component of food, or as a dietary supplement, and any of these cancers." Furthermore, two more experts of medical panels cited in the entrance of the Wikipedia page also confirmed this situation.

To get a better assessment of the potential of our approach, we performed simulated experiments with a selection of the 80 most recent meta-analyses found in our collection with respect to three other diseases: hypertension, diabetes and asthma in addition to cancer. A "meta-analysis" is a systematic review that uses statistics analysis to be able to combine several research papers on a particular topic. One characteristic of meta-analyses is that it may never be possible to include all the papers that deal with a particular phenomenon. Usually, researchers query a digital library using keywords to

[2] https://en.wikipedia.org/wiki/Lycopene

get a candidate set of papers and after that, they manually decide which candidates can be included in the analysis. Depending on the methodology chosen by the researchers, the final number of articles vary. In this set of experiments, we proceeded as follows: we took out of our collection the meta-analysis, and then we queried our representation using the claim of the meta-analysis. If we could agree with the claim of the meta-analysis in at least one possible world, then we consider that as a positive outcome. After our experimentations, our best result was a kappa of 0.7746 with a 95% confidence interval (0.7875, 0.9549). Notice that because the criteria that the experts use to include and/or exclude some papers in a meta-analysis are beyond our current text mining processing, we included all papers as given by our query pattern. However, one caveat of this type of experiments is the training time of the embedding and the LDA hyper parameters. In this particular setting, we trained the word embeddings with 100 dimensions and LDA with 8,000 iterations with a fixed 300 topics in a collection of 315k documents.

To provide insights of our results, let's look at one of the cases where our approach failed. Consider the findings of [24], where the study of alcohol regarding prostate cancer was analyzed. As stated in the paper, a total of 340 studies were found in the exploratory search, but only 27 satisfied the inclusion criteria of the researchers (manual assessment). For this case, we found a "controversial" situation. In other words, our proposed approach did not agree with the meta-analysis in any possible world. More specifically, all the possible worlds were inconsistent and our tool stated a "controversial" situation. Moreover, the researchers reported "Our study finds, for the first time, a significant dose–response relationship between level of alcohol intake and risk of prostate cancer starting with low volume consumption"[24]. Of course, this is an expert assessment and our tool is not aiming at replacing a decision but instead helping to detect situations that may require a better administration of reviewers, especially in cases of "controversy" where clearly major care should be taken.

Discussion. Our results look promising and there are some issues that we noticed during our experiments. Firstly, the assessment of the degree of association between the claims is something that only domain experts can properly adjust. For instance, the idea that "tomato sauce" and "lycopene" can be considered similar enough to retrieve claims that associated both of them with "cancer" depends on what the experts would consider "related". Moreover, the idea of considering or not considering related types of a disease, such as "prostate cancer", "lung cancer", "gastric cancer", etc., in the retrieval of "related" claims is again questionable. In our experiments, we did notice a difference when we filtered the results to restrict the retrieval to the specified entities. Nevertheless, we envision an application where the reviewer can actually experiment with this feature of our approach. Secondly, the final decision of "controversial" with the idea of the possible world explanation did help to some extend but stayed below expectations in the experiments. One possible explanation is the criteria of inclusion/exclusion of articles in a meta-analysis and the methodology used to assess its conclusion. These two aspects are of course beyond our approach capacities and not in the scope of what we want to achieve. And third, our approach could accurately find controversial situations as confirmed by the meta-analysis experiments. However, this was only possible when we did not restrict the entities to exact matches but instead expand them to the most related ones as motivated in the work of [19].

5 Conclusions

We introduced a novel approach to assess the Plausibility of a new document to support peer review not at the process level, but with a clear focus at the document level. Our results look promising towards the goal of novel management of resources in peer review. In particular, the question of how many reviewers a new paper needs can be adjusted by its respective degree of Plausibility. Of course, our experiments also reveal future work that is needed to crystalize our vision. For instance, assuming that "tomato sauce" and "lycopene" can be considered similar enough to retrieve papers that associate both of them with "cancer" depends on the goal of the analysis. And this is something that domain experts can properly adjust. Thus, in future experiments we will provide an online application to allow users of our system to personalize the degree of associations between the claims. Hopefully, we will learn some patterns to adapt to new users and new domains. The idea of "possible worlds" proved itself to be useful but in some domains one might consider a more restricted view and instead of proportions of topics as latent descriptors of documents, one might be interested in a hard clustering approach. We will also incorporate this notion into our system in the future.

Finally, the model of claims that we currently have must be extended to cope with other domains. To do that, we will need to account for more advanced model's representation of arguments in scientific papers. We are aware that the incipient field of Argumentation Mining in the last few years has shown tremendous potential to envision more powerful applications. We will also explore that line of research in future work.

6 References

1. Price, B.Y.S., Flach, P.A.: Computational Support for Academic Peer Review: A Perspective from Artificial Intelligence. Commun. ACM. 60 No. 3, 70–79 (2017).
2. Schoenfeld, J.D., Ioannidis, J.P.A.: Is everything we eat associated with cancer? A systematic cookbook review. Am. J. Clin. Nutr. 97, 127–134 (2013).
3. Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., Clark, T.: The SWAN biomedical discourse ontology. J. Biomed. Inform. 41, 739–751 (2008).
4. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. Inf. Serv. Use. 30, 51–56 (2010).
5. Velterop, J.: Nanopublications: the future of coping with information overload. LOGOS J. World B. Community. 21, 3–4 (2010).
6. Kuhn, T., Barbano, P.E., Nagy, M.L., Krauthammer, M.: Broadening the scope of nanopublications. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 487–501 (2013).

7. Groth, P., Loizou, A., Gray, A.J.G., Goble, C., Harland, L., Pettifer, S.: API-centric Linked Data integration: The Open PHACTS Discovery Platform case study. J. Web Semant. 29, 12–18 (2014).
8. Clark, T., Ciccarese, P.N., Goble, C. a: Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. J. Biomed. Semantics. 5, 28 (2014).
9. Toulmin, S.: The Uses of Argument. Ethics. 70, vi, 264 (1958).
10. Verheij, B.: The Toulmin Argument Model in Artificial Intelligence. In: Argumentation in Artificial Intelligence. pp. 219–238 (2009).
11. Connell, L., Keane, M.T.: A model of plausibility. Cogn. Sci. 30, 95–120 (2006).
12. Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A Neural Probabilistic Language Model. J. Mach. Learn. Res. 3, 1137–1155 (2003).
13. Turian, J., Ratinov, L., Bengio, Y.: Word Representations: A Simple and General Method for Semi-supervised Learning. Proc. 48th Annu. Meet. Assoc. Comput. Linguist. 384–394 (2010).
14. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Distributed Representations of Words and Phrases and their Compositionality. Nips. 1–9 (2013).
15. Mikolov, T., Yih, W., Zweig, G.: Linguistic regularities in continuous space word representations. Proc. NAACL-HLT. 746–751 (2013).
16. Le, Q., Mikolov, T.: Distributed Representations of Sentences and Documents. Int. Conf. Mach. Learn. - ICML 2014. 32, 1188–1196 (2014).
17. Blei, D.M.: Probabilistic topic models. Commun. ACM. 55, 77 (2012).
18. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. J. Mach. Learn. Res. 3, 993–1022 (2003).
19. Kusner, M.J., Sun, Y., Kolkin, N.I., Weinberger, K.Q.: From Word Embeddings To Document Distances. Proc. 32nd Int. Conf. Mach. Learn. 37, 957–966 (2015).
20. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 460–467 (2009).
21. Sojka, R. {\v R. u}{\v r}ek and P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. pp. 45–50. ELRA, Valletta, Malta (2010).
22. Peleteiro, B., Lopes, C., Figueiredo, C., Lunet, N.: Salt intake and gastric cancer risk according to Helicobacter pylori infection, smoking, tumour site and histological type. Br. J. Cancer. 104, 198–207 (2011).
23. Kristal, A.R., Till, C., Platz, E.A., Song, X., King, I.B., Neuhouser, M.L., Ambrosone, C.B., Thompson, I.M.: Serum lycopene concentration and prostate cancer risk: Results from the prostate cancer prevention trial. Cancer Epidemiol. Biomarkers Prev. 20, 638–646 (2011).
24. Zhao, J., Stockwell, T., Roemer, A., Chikritzhs, T., Bostwick, D. et. al.: Is alcohol consumption a risk factor for prostate cancer? A systematic review and meta–analysis. BMC Cancer. 16, 845 (2016).