The Journey is the Reward - Towards New Paradigms in Web Search

Harald Sack(✉)

Hasso Plattner-Institute for IT Systems Engineering,
University of Potsdam, Potsdam, Germany
harald.sack@hpi.de

Abstract. Without search engines the information content of the World Wide Web would remain largely closed for the ordinary user. Current web search engines work well as long as the user knows what she is looking for. The situation becomes problematic, if the user has insufficient expertise or prior knowledge to formulate the search query. Often a sequence of search requests is necessary to answer the user's information needs, whenever knowledge has to be accumulated first to determine the next search query. On the other hand, retrieval systems for traditional archives face the problem that there is possibly not always a result for an arbitrary search query, simply because of the limited number of documents available. Semantic search systems (try to) determine the meaning of the content of the archived documents first and thus in principle are able to overcome problems of traditional keyword-based search engines concerning the processing of natural language. Moreover, content-based relationships among the documents can be used to filter, navigate, and explore the archive. Content-based 'intelligent' recommendations help to open up the archive and to discover new paths across the search space.

Keywords: Semantic search · Exploratory search · Semantic annotation · Linked open data · Recommender systems

1 Introduction

The immense number of available documents in the World Wide Web today would remain locked for the users without search engine technology. Web search engines index the publicly accessible Web and facilitate to find the information the user is looking for. Web authors and search engine providers usually follow a common objective: while the authors want their documents to be found, the search engine providers want to deliver these documents to the users who are looking for the provided information. Without web search engines the users would have to move hand over hand from one document to the next by following the hyperlinks contained in the document. The search engine relieves the user from this tedious task by crawling the web beforehand and preparing a suitable index data structure for targeted search. Moreover, contemporary search

W. Abramowicz (Ed.): BIS 2015 Workshops, LNBIP 228, pp. 15–26, 2015.
DOI: 10.1007/978-3-319-26762-3_2

engines such as Google[1], Bing[2], or Yahoo![3] provide additional services, as e.g. auto-completion of search queries, search as you type [15], multimedia search, or query by example, to augment their usefulness as well as their convenience. Furthermore, the introduction of knowledge graphs has enabled the search engines to answer the user's questions rather than simply to return document search results. Knowledge graphs provide structured and detailed information about the entities recognized in the user's search query, sometimes also supported by a list of links to other search recommendations [9].

Due to their very nature current web search engines work well as long as the user knows what she is looking for, i.e. as long as the user knows how to phrase the search query. The situation becomes problematic, if the user has insufficient expertise or prior knowledge to formulate the search query. Often a sequence of search requests is necessary to answer the user's information needs, whenever knowledge has to be accumulated first to determine the next search query. Besides in web search, the same search engine technology is also applied within archives with a limited number of documents. Although the number of documents in the web is strictly speaking limited, their immense number make it much more likely to find results for almost any kind of query. However, smaller archives face the problem that there is possibly not always a result for an arbitrary search query, simply because of the manageable number of documents available.

One way to overcome this problem is to make use of the semantics of the information provided in these documents. Semantic analysis enables to determine the meaning of the content of the archived documents and thus in principle to overcome problems of traditional keyword-based search engines concerning the processing of natural language, such as synonymy and polysemy [4]. Moreover, semantic relationships can be identified among the archived documents, which can be used to filter and navigate the archive, although the original search term does not have to be present in the result documents. In this way, also documents closely related to the search query can be identified and recommended to the user. These content-based 'intelligent' recommendations help to open up the archive and to discover new paths across the search space [14]. In this way the user is able to explore the content of an archive even without having a specific information need beforehand enabling serependitious discoveries.

The paper is structured as follows. Section 2 explores situations, where search engines fail to fulfill the user's search interest and explain how these disadvantages at a second glance even might get useful. In Sect. 3, semantic analysis of multimedia documents is outlined and discussed, while in Sect. 4, building on that the possibilities of semantic and exploratory search including intelligent (search) recommendations are further investigated. The paper concludes in Sect. 5 with a short summary and a brief outlook.

[1] http://google.com/.
[2] http://bing.com/.
[3] http://yahoo.com/.

2 Drawbacks of Traditional Search Engines

Most users don't complain about search results of web search engines, because up to now they simply provide the best and most convenient way to get to the desired information. But, this is also because the user has learned about their basic function principles as well as the user has adapted her expectations about the obtained search results. Since the early days of web search engines users and at least with the arrival of Google in the late 1990s, the user knows that to search for information about a specific concept means to search for documents that contain the name of that concept. This refers to text documents and obviously the search engines provide basic natural language processing techniques such as lemmatization of search queries and index terms alike.

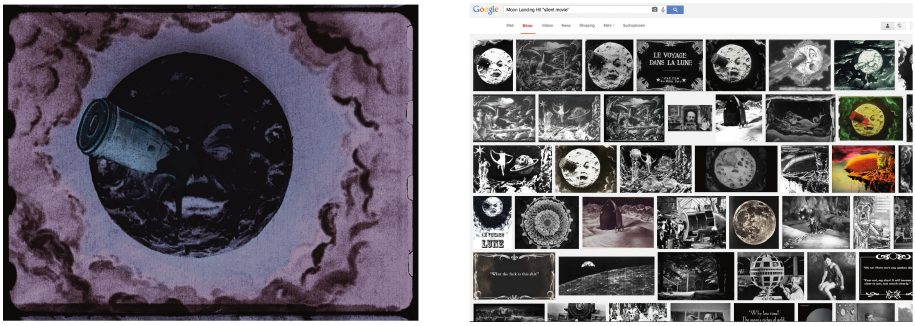
Besides, many additional helper techniques facilitate the ease of use of web search engines: autocompletion suggests the most probable search terms already while typing the search query. In the meantime the search is already computed in the background and instantly delivered to the browser in real-time. Moreover, the user might select one of several suggested search term completions. To achieve this, web search engines analyze their usage log files for information about co-occurrences of search terms and other more sophisticated statistical measures.

Modern web search engines also enable search on multimodal documents such as text, images, videos, or audio documents. But, only for a small fraction of multimedia documents search engines actually analyze the media content for indexing. In the web, multimedia documents most times are embedded within an HTML document via hyperlinks. Likewise, HTML documents that contain links to multimedia documents also provide a title or a short description of the linked content marked up by HTML anchor tags. In this way web search engines can make use of these descriptive texts to index the media documents by their content without the need for complex analysis. Further search engine extensions are the possibility to perform queries by example[4] or the support by powerful knowledge bases that are applied for disambiguation, question answering, or recommendation [3].

Although semantic technologies are already applied to support search engines, their basic search paradigm dates back to the early days of information retrieval, when among an index of (text) documents a similarity based mapping to given search query terms was computed [8]. For text documents similarity is often interpreted as string similarity, which does not necessarily mean similarity of content. Likewise the fundamental vector space model of information retrieval assumes index terms to represent orthogonal base vectors, i.e. the index terms are considered not to be related or similar to each other [7]. Thus, the interpretation of search results based on the vector space model does not take into account the similarity among different terms in the document. Extensions of a generalized vector space model also take into account the similarity of index terms by incorporating additional information for index entities such as e.g. class membership and class hierarchy information, but still are a topic of current research.

[4] as e.g., Google reverse image search https://images.google.com/.

But, search engines help the user to find a solution for her information needs only as long as the user knows how to phrase the search query. If the user lacks the knowledge to name the document or entity she is looking for, maybe because she is not familiar with the pertaining subject, web search engines are only of limited use. Imagine the following scenario: You are looking for a distinct movie, but you don't know the title or neither the director or any of the actors. But, you have an iconic picture in your mind which is part of that movie and is depicted in Fig. 1(a)[5].



(a) (b)

Fig. 1. (a) Search example for a movie, for which the user has only a visual memory and does not know title, director, or actor information. (b) Result of an image search experiment with descriptive terms "Moon, Landing, Hit, Silent Movie".

The only way to find information about the movie in question is to search for descriptive features that describe the media content. The retrieval success depends on which terms the user can think of and if they match the terms that have been used to describe the media on the web. The reader is advised to repeat the experiment with the image search feature of the search engine trying to retrieve this specific picture. Each single try will result in numerous pictures and only if the movie is sufficiently described by the search terms, the searched picture will be found among the search results (cf. Fig. 1(b)). However, by looking closer at the achieved search results, the user might be able to identify also pictures from similar movies, which might also be of interest. In this way, even traditional search enables the user to find results that originally were not intended, but might nevertheless be relevant and of interest.

In the following section, some fundamental techniques of semantic analysis and semantic annotation are presented, which are intended to support the search engine to fulfill the user's information need although the user might not know the best suited search terms.

[5] George Melies, Le Voyage dans la lune (A Trip to the Moon), 1902.

3 Semantic Analysis and Semantic Annotation

To understand the content of a document means that the content can be correctly interpreted. Often the content of a document is additionally annotated with metadata. This metadata can help to determine the correct interpretation of the document's content. But, metadata can originate from various sources of different reliability and thus, also influences document interpretation in different ways. For the correct interpretation of the content, always the information context has to be taken into account. Depending on the context the same information can be interpreted in different ways. Thus, semantic analysis starts with the specification of the context under which the information should be interpreted.

For our purpose, we define semantic analysis as the process required to understand information content in the sense that the content can be interpreted correctly. For natural language text, semantic analysis comprises the determination of named entities represented in the text and the correct mapping of the text terms representing named entities to unique entities of a knowledge base (Named Entity Disambiguation, NED). Within the knowledge base entities are mapped to ontologies that specify properties, relations, and constraints to further define the meaning of an entity.

In Sect. 3.1 the different aspects and types of metadata are discussed. Section 3.2 illustrates the process of semantic analysis and context establishment, while Sect. 3.3 describes our approach to NED based on Linked Data. Finally, Sect. 3.4 concludes this section by showing how to apply semantic analysis in semantic document annotation.

3.1 Heterogeneous Metadata and Reliability

Metadata can occur with different degrees of structure. Unstructured metadata comprises binary data as well as textual data in natural language. To derive the meaning of unstructured metadata manual interaction or additional automated analysis is required. Semi-structured metadata contains tags or other markup symbols, as e.g. in HTML or XML, to separate semantic elements, i.e. elements with a dedicated meaning or interpretation, and enforce hierarchies of records and fields within the data. Structured data on the other hand does conform with a given formal structure of a data model usually associated with a database or data tables. Formal structure and interpretation of structured data is given by the underlying data model. The less structured the data, the more possibilities of interpretation and more sources of errors might occur. Thus, metadata can be considered more reliable as higher the degree of structure and the more detailed and expressive the underlying data model is specified.

Considering natural language text, further degrees of structure can be distinguished. While plain text represents information in terms of sentences, metadata can also be provided on the basis of keyterms, i.e. single words. The interpretation of single words can be more difficult than the interpretation of an entire sentence due to missing context information. However, keyterms also might originate

from a predefined (restricted) vocabulary only, allowing for a unique interpretation. In that sense this different granularity of structure for natural language text results in a different degree of ambiguity and thus also possibilities to make mistakes in the interpretation. In general, the less mistakes can be made in the interpretation the more reliable the metadata might be considered.

Also the source of the metadata has to be regarded. While metadata can originate from the author of the data or an expert describing the data with high accuracy and reliability, it might also be metadata provided by an ordinary user of the data, who is no expert. Thus, authoritative metadata from experts are more likely to be more reliable or trustworthy than metadata from an arbitrary user without a certified expertise. In the same way, metadata originating from automated analysis processes might obtain a different degree of reliability and correctness depending on the quality of the analysis process and the quality of the original data. Furthermore, metadata from automated analysis might occur as low level feature data, as e.g., direct measurement results. Otherwise, high level feature metadata comprise metadata that originates from an interpretation, aggregation, and categorization of low level feature metadata, and thus are subject to an additional source of error. Provenance is a valuable source to determine reliability, trustworthiness, and correctness of metadata.

To fully understand and to interpret metadata correctly the process of semantic analysis has to consider different levels of abstraction within metadata as well as different degrees of reliability, trustworthiness, and correctness. To enable correct interpretation of metadata, semantic analysis must integrate information on context, pragmatics, as well as constraints and axioms which determine their validity.

3.2 Semantic Analysis

In conformity with the definitions of Carnap [2] and Russel [6], we define semantic analysis as the process of determining the meaning of data (information) in the sense of their correct interpretation. For a correct interpretation, semantic analysis must take into account all available metadata while considering its reliability, trustworthiness, and correctness. Nevertheless, available metadata might not be sufficient to achieve a unique and unambiguous interpretation. Moreover, the interpretation also depends on the context of the information and possibly also on its pragmatics, i.e. the intention of its originator. While pragmatics influences context, context might be considered as additional available data that has to be taken into account for the subsequent disambiguation. In [11] we have defined a formal context for the disambiguation of ambiguous interpretations. According to the contextual description the confidence of the context item is calculated depending on the reliability of the metadata source, the level of agreement among metadata sources, the structural degree of metadata, as well as the level of potential ambiguity of metadata interpretations. Based on this context model, each potential interpretation achieves a confidence score, which is further refined by metadata correctness, metadata relevance, and metadata ambiguity.

Semantic analysis determines a mapping among the original data and a knowledge base of formal semantic descriptions of unique entities. In the Linked Data environment and the semantic analysis of text-based data, semantic analysis as e.g. uniquely maps text tokens to DBpedia[6] entities. DBpedia entities are linked to ontologies, which define the entities' meaning by relating it to other entities or data values.

3.3 Named Entity Disambiguation

Named Entity Disambiguation (NED) as the process of identifying the correct meaning of an ambiguous information object is one of the core technologies of semantic analysis. Ambiguity is resolved with the help of context information. Applied on textual input, NED determines the correct meaning of text tokens that stand for named entities by taking into account the surrounding sentence, paragraph, or larger fraction of the text. For our approach, we distinguish four phases [13]:

1. **Detect named entities in text**
   Named entities in text usually are nouns, which can be identified via a Part-of-Speech tagger. A Named Entity Recognitizer (NER) cannot only determine nouns that represent named entities, but also categorize named entities into predefined classes such as persons, locations, organizations, or time expressions. For our approach we have applied the Stanford NER with three classes: persons, locations, organization [10]. N-gram analysis considers the number of consecutive text tokens that denote a named entity, as e.g. for compound names. Each single token of a compound name can denote an individual entity. Thus, all 1-grams, 2-grams, ..., n-grams containing a noun as the last term are potential named entities.
2. **Determine possible candidate entities**
   For all detected potential named entities in the text, candidate mappings from a knowledge base–here DBpedia–are generated. For this process, possible alternative names of the entities under consideration have to be determined. In DBpedia, there exist various properties that denote alternative denominations. According to the design of Wikipedia, which is reflected in DBpedia, so-called redirect pages also denote alternative names and have to be resolved. Moreover, so-called disambiguation pages provide possible referrals for homonyms and might also contribute alternative names. Redirects and disambiguation pages often contain chains or even cycles, which have to be resolved by aggregating all labels from redirect and disambiguation paths within the leafs of these paths. To speed up this process, a gazeteer is computed beforehand that connects a named entity with all its possible names. For all detected named entities in the text all potential entity candidates are collected via a gazeteer lookup.
3. **Filter entity candidates**
   To simplify the following tasks the number of potential entity candidates are

[6] http://dbpedia.org/.

reduced by a plausibility filter. Here, the results of the NER class from step 1 for a named entity under consideration are compared with the `rdf:types` of the entity candidates returned from the gazeteer. In case of conflicting types the entity candidates concerned will be deleted.

4. **Disambiguate entity candidates according to context**
For all remaining entity candidates of the context under consideration the induced link graph derived from DBpedia is created. This graph serves as the basis for the disambiguation process. The disambiguation relies on the assumption that the correct entities for a given context are most likely related with each other. In terms of a graph this means that there might be paths and even connected components found between the candidate entities, which help to identify the correct entities. The longer a connected component in the induced link graph, the higher the likelyhood that the connected nodes denote the right interpretation. The link graph can be considered as partitioned graph into sets of entity nodes which belong to the same text term. Thus, connected components have to be identified that cover the most term partitions. Links inside a term partition have to be neglected. If strongly connected components can be found, they further consolidate the prior selection. Figure 2 illustrates the concept of a link graph with term partitions and connected components according to a given example text context.
In addition to link graph analysis also co-occurrence analysis based on the texts of Wikipedia articles of the entities under consideration is performed. Here, for all the labels of all the entity candidates of a given context it is verified whether these labels co-occur in the entities' article texts [11]. If neither link graph analysis nor co-occurrence analysis is able to disambiguate an entity, the decision is made according to the most popular entity, which is assumed to be the correct entity with a higher probability than the remaining. As a measure for entity popularity the in-degree of an entity node in the link graph or also the pagerank algorithm can be applied [1]. If the popularity delivers only inconclusive results, as e.g. if the differences among entity popularity are too small, the concept of so-called negative context can be applied. Here, all entity candidates are excluded from the candidate list for which a connection to the already disambiguated entities is rather unlikely or even contradictory [12].
In general, a hierarchical approach has been chosen for NED, which always starts the disambiguation with the most reliable algorithm on the most accurate and reliable data. The remaining ambiguity is resolved with the less reliable algorithms on less reliable metadata. The following algorithms are applied in the given sequence:
(a) connected component analysis on the link graph
(b) co-occurrence on wikipedia text corpus
(c) popularity based link graph analysis (e.g. with indegree or pagerank)
(d) negative context analysis.

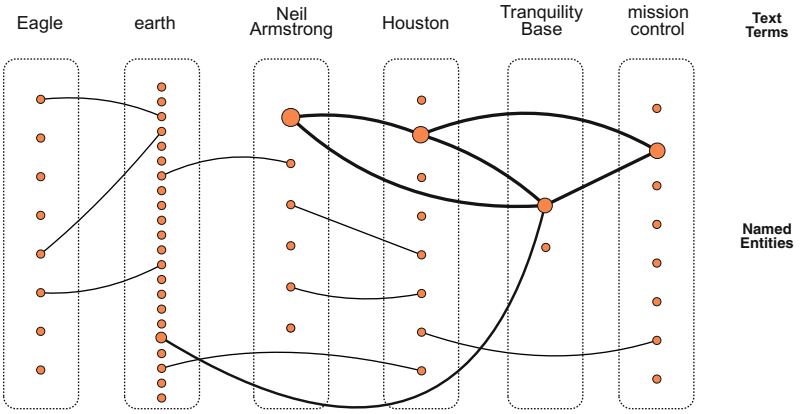


Fig. 2. Example text with highlighted named entities and the related term partition graph, where all edges have been eliminated except edges between term partitions. Strongly connected components and longest connected components are emphasized.

3.4 Semantic Document Annotation and How to Make Use of It

The process of semantic analysis results in a mapping of information objects to entities in a given knowledge base, as e.g. DBpedia. For natural language text, text tokens are mapped to DBpedia entities via URIs (Uniform Resource Identifier). The text annotation can be achieved via the NLP Interchange Format (NIF)[7] and RDFa[8]. For non-textual documents such as images, videos, or audio files, semantic annotation can be achieved by addressing temporal and spatial fragments of the media via URI media fragments[9]. Figure 3 shows a sample annotation for a video fragment with RDFa.

In this way, semantically anotated documents easily can be published on the web. Since the annotations can be dereferenced, the information content of the documents can be correctly interpreted, as long as the annotations are correct.

4 Exploratory Search and Intelligent Recommendations

One prominent application that benefits from explicit semantic annotations on the web are search engines. With semantic annotations, natural language texts can be interpreted correctly and ambiguities or errors induced by natural language can be avoided. By switching from keyword-based search to entity-centered search, the usual problems with synonyms, metaphoric language as well as ambiguities can be avoided and more precise as well as more complete search results can be achieved. But, as already pointed out in Sect. 2, there are more relevant search scenarios, where semantic annotations can be of benefit.

[7] http://persistence.uni-leipzig.org/nlp2rdf/.
[8] http://www.w3.org/TR/rdfa-syntax/.
[9] http://www.w3.org/TR/media-frags/.

```
<div vocab="http://www.w3.org/ns/oa#"
     prefix="dctypes: http://purl.org/dc/dcmitype/
             foaf: http://xmlns.com/foaf/0.1/"
     typeof="Annotation"
     resource="#contentAnnotation-001">
  <div property="hasTarget"
       resource="http://test.org/test.ogv#t=20,30&xywh=480,150,140,330"
       typeof="dctypes:video">
  </div>
  <div property="hasBody" typeof="SemanticTag">
    <a property="foaf:page"
          href="http://dbpedia.org/resource/Neil_Armstrong">
        Neil Armstrong
    </a>
  </div>
</div>
```

Fig. 3. Sample annotation of a temporal and spatial video fragment with URI media fragments and various annotation vocabularies as RDFa embedded in HTML.

4.1 Pinpoint Search vs. Exploratory Search

The usual web search scenario can be considered as so-called pinpoint search. The user knows what she is looking for and is able to provide the right query terms. In a traditional library, this is similar to the situation, when you are looking for a specific book that can be looked up in the library index. But, the situation changes, when the user is searching for the next book to read, which should be somehow similar or related to the first one. Likewise, the user possibly has first to gather more information before being able to put forward the right search query. Searches for complex answers, where the user is not familiar with the domain, or where in general the knowledge to pose the right search query is not available, are referred to as exploratory search [5].

Interestingly, in traditional libraries means for exploratory search are provided by the library classification system according to which the books in the library shelves are organized. To find related books, the user simply has to browse the shelf where he has found the original volume. If this procedure is not satisfactory, the user can ask the librarian for intelligent recommendations.

For an exploratory search scenario with semantically annotated documents, the search process also must consider the relations among the information content of the documents. By dereferencing the semantic entities within the document annotations, explicit or also inferred relations between entities can easily be exploited to compute measures of similarity and relatedness among the documents [14].

4.2 From Exploratory Search to Intelligent Recommendation

By taking into account similarity and relatedness among documents for exploratory search, the user has to decide which direction she wants to follow.

In general, relations derived from entities of the document's content can be of different relevance, which depends on the intention and the background of the user. If no information about the user is available, the situation is similar to the cold start problem in recommender systems, when no usage history is available to generate a recommendation via statistics. Content-based recommender systems derive recommendations in a similar way as exploratory search systems generate search results. They take into account similarity and relatedness, while deciding what aspect in the considered relations is of general relevance.

To generate more interesting recommendations, it is important not only to take into account similarity, because otherwise the user soon will be bored. The user wants to be positively surprised by a recommendation by not suggesting the obvious, but by finding unexpected while nevertheless relevant suggestions. Here, serendipity has become a decisive factor for the quality of recommendations as well as for search result suggestions of exploratory search systems. Thus, to fulfill the user's information needs, a search system should return as well pinpoint search results of high precision and recall, as well as additional results or search suggestions that have been generated from content-based relationships. If the user decides to follow the suggestions, she will be able to follow her personal interest to discover new and maybe previously unknown paths through the search space.

5 Conclusion and Future Work

In this paper some scenarios have been developed that go beyond the current (web) search paradigm of simple pinpoint search results, especially when the user lacks information to be able to phrase the right search query. One way to cope with this challenge is to apply semantic analysis and annotation to be exploited by semantic and exploratory search engines. Exploratory search engines provide additional search results and search suggestions for the user to discover new and maybe previously unknown paths through the search space. The question remains, whether the traditional presentation of search results as a linear list also holds for results in this extended scenario. Besides the improvement and extension of the described technology, future work therefore will also focus on well suited user interfaces for exploratory search.

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. In: Proceedings of the 7th International Conference on World Wide Web 7 (WWW7), Elsevier Science Publishers B.V., Amsterdam, The Netherlands, pp. 107–117 (1998)
2. Carnap, R.: Testability and meaning I. Philos. Sci. **3**, 419–471 (1936)
3. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2014), pp. 601–610. ACM, New York (2014)

4. Guha, R., McCool, R., Miller, E.: Semantic search. In: Proceedings of the 12th International Conference on World Wide Web, WWW 2003, pp. 700–709. ACM Press, New York (2003)
5. Marchionini, G.: Exploratory search: from finding to understanding. Commun. ACM **49**(4), 41–46 (2006)
6. Russell, B.: An Inquiry into Meaning and Truth. W.W. Norton & Co, New York (1940)
7. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11), 613–620 (1975)
8. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill Inc, New York (1986)
9. Singhal, A.: Introducing the Knowledge Graph: things, not strings, Official Google Blog (May 2012). http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html
10. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363–370 (2005). http://nlp.stanford.edu/manning/papers/gibbscrf3.pdf
11. Steinmetz, N., Sack, H.: Semantic multimedia information retrieval based on contextual descriptions. In: Cimiano, P., Corcho, O., Presutti, V., Hollink, L., Rudolph, S. (eds.) ESWC 2013. LNCS, vol. 7882, pp. 382–396. Springer, Heidelberg (2013)
12. Steinmetz, N., Sack, H.: About the influence of negative context. In: Proceedings of 6th IEEE International Conference on Semantic Computing (ICSC 2013), pp. 134–141 (2013)
13. Usbeck, R., Rder, M., Ngomo, A.N., Baron, C., Both, A., Brmmer, M., Ceccarelli, D., Cornolti, M., Cherix, D., Eickmann, B., Ferragina, P., Lemke, C., Moro, A., Navigli, R., Piccinno, F., Rizzo, G., Sack, H., Speck, R., Troncy, R., Waitelonis, J., Wesemann, L.: GERBIL - general entity annotator benchmark, in WWW 2015. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1133–1143. ACM (2015)
14. Waitelonis, J., Sack, H.: Towards exploratory video search using linked data. Multimed. Tools Appl. **59**(2), 645–672 (2012). doi:10.1007/s11042-011-0733-1
15. Zuccarino, S.: Updates to Google News US Edition: Larger Images, Realtime Coverage and Discussions, Google News Blog (May 2012). http://googlenewsblog.blogspot.com/2012/05/updates-to-google-news-us-edition.html