A short survey of linked data ranking

Conference Paper · March 2014

DOI: 10.1145/2638404.2638523

3 authors:

Semih Yumusak
KTO Karatay University

Erdogan Dogdu
Angelo State University

Halife Kodaz
Konya Technical University

A Short Survey of Linked Data Ranking

Semih Yumusak
KTO Karatay University
Konya, Turkey
semih.yumusak@karatay.edu.tr

Erdogan Dogdu
Georgia State University (adjunct)
TOBB University of Economics and Technology
Ankara, Turkey
edogdu@etu.edu.tr

Halife Kodaz
Selcuk University
Konya, Turkey
hkodaz@selcuk.edu.tr

ABSTRACT
Linked data systems are still far from maturity. Hence, the basic principles are still open for discussion. In our study on building a novel linked data search engine, we have surveyed fundamental methods of internet search technologies in the context of linked data, such as: crawling, indexing, ranking, and monitoring. The scope of the survey covers statistical ranking, database ranking, document level ranking, Web ranking, and linked data ranking techniques. In order to classify the linked data ranking methods, we identified a number of categories. These categories are ontology ranking, RDF ranking, graph ranking, entity ranking, document/domain Ranking. At the end of the survey, we have listed the ranking techniques based on the well-known PageRank algorithm.

Categories and Subject Descriptors
H.3.3 [**Information Storage And Retrieval**]: Information Search and Retrieval – *Information filtering, Relevance feedback, Selection process, Search process.*

General Terms
Algorithms, Documentation, Standardization, Theory.

Keywords
Ranking, Information Retrieval, Data Processing

1. INTRODUCTION
Semantic Web is referred to as the next generation of the Web and it aims to transform the Web into a machine-readable Web unlike the current Web that started as a Web of linked documents and still mostly is. Linked data is a term that refers to large volumes of data repositories that are created using semantic Web and linked data principles.

In order to store and classify the knowledge, semantic Web uses the idea of ontology over the Web of data. The ontologies are mostly identified by OWL Web Ontology Language[1] and entities are identified by RDF (Resource Description Framework)[2].

The Linking Open Data (LOD)[3] project [11] was started to implement the idea of Linked Data[4], which was designed to create connections between separate semantic Web knowledge bases. This leads to a connected Web of data, which is named Linked Data Cloud[5].

As the data grows in those domains, data search becomes a more complicated issue. In order to return meaningful results, search engines use ranking techniques to sort results.

Ranking is defined as;

"reorganizing search results so that the most relevant information appears higher in the list." [46]

"an important mechanism in the search process with the function of prioritizing data elements." [39]

"a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second."[6]

In the Web search domain, text based search engines rank the documents and domains according to their popularity and relevance [13, 44]. However, in Semantic Web and Linked Data domains, ranking is more complicated in search processes due to semantic relationships. Most of the ranking methods designed for Semantic Web and Linked Data are inspired by the common Web search ranking algorithms such as PageRank [13]. Moreover, there are several studies on ranking different properties of Semantic Web documents, entities, ontologies, or graphs.

In this survey study, we identify and classify the common patterns of ranking methods on linked data. By classifying those methods, we aim to provide an understanding on previous works and draw a roadmap for future applications. The ranking techniques are categorized according to their usage and functionalities ranging from local databases to big data, and from a single semantically annotated file to linked data. We identified the ranking methods as; Ontology Ranking, RDF Ranking, Graph Ranking, Entity Ranking and Document/Domain Ranking.

In section 2 we present the ranking categories we identified and the related studies. In section 3 semantic Web search engines are compared based on their ranking techniques briefly.

2. RANKING METHODS
In this section, semantic web and linked data related ranking studies are surveyed. We categorize the linked data ranking studies based on data sources they use. These are ontology ranking, RDF document ranking, graph ranking and entity ranking. We have also surveyed the studies which are not identified as linked data, however they are related to linked data.

[1] http://www.w3.org/TR/owl-features/

[2] http://semanticweb.org/wiki/RDF

[3] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[4] http://www.w3.org/DesignIssues/LinkedData.html

[5] http://lod-cloud.net/

[6] http://en.wikipedia.org/wiki/Ranking

2.1 Ontology Ranking

The ontology ranking techniques are of two types; ranking data by using ontologies and ranking ontologies themselves. One of the first ontology-based ranking method is applied on Web services. [53] presents a study on ranking "semantic Web service advertisements with respect to a service request". The ranking is based on semantic similarity between request parameters and the service advertisement by using a domain ontology.

Ontology ranking is studied in several works. One of the earliest studies on ontology ranking is OntoKhoj [49]. Table 1. Weights of Hyperlinks [49] is retrieved from [49], in which the semantic relations are given weights according to the importance level of the relationship types.

Table 1. Weights of Hyperlinks [49]

Priority (Weight)	Relationship	Language Specific
1	instantiation	rdf:type
2	subClass	rdfs:subClassOf, daml:subClass
3	domain/range	rdfs:domain, daml:range

In [54], an "ontology-based inferencing" is used to rank the query results and similar to this study [52] used a numerical weight which is assigned to every relation instance of an ontology. [52] named their method as "weight mapping" and [54] named their method as "ontology-based ranking".

Swoogle [24] is the first ontology search engine that uses an ontology ranking system, which they call "ontology rank". According to this ranking, the ranking level of an ontology increases as the usage of the ontology increases in the community. They state that, "Swoogle uses a Rational Surfer Model" [25] rather than a "random surfing model" [15]. [25] discusses their detailed approach on ranking in the following popularity based ranking categories; Semantic Web Documents (OntoRank based on Rational Surfer Model) and Ontology Dictionary (TermRank, Class-Property Bonds).

AKTiveRank study [1] ranks ontologies according to some structural metrics they define. They state that Swoogle [24] and OntoKhoj [49] style ranking "will not work for large numbers of existing ontologies because of their poor connectivity and lack of referrals from other ontologies" [1]. A user controlled multi-dimensional ranking approach is introduced and the ranking measures are defined as Class Match Measure, Density Measure, Semantic Similarity Measure, Betweenness Measure.

Watson [4] computes a simple structural or topic related quality measures for the data and store those values with the ontologies. These scores are used as a ranking parameter while the data is queried.

2.2 RDF Document Ranking

Resource Description Framework(RDF) is a common standard model for Web based data interchange.[7] In the semantic Web, RDF documents are used for data publication and exchange. By the growth of the RDF publishing on the Web, ranking of RDF documents became an important task.

[7] http://www.w3.org/RDF/

One of the early studies on ranking RDF resources, named QuizRDF [20], was designed as an RDF search engine. The ranking of the RDF resources is based on TF-IDF scoring[8].

[5] focused on ranking the contents of the RDF documents and significance of the ranking is its domain independence. RDF statement ranking algorithms are defined under two categories, ranking of "Topic-Related Sentences" and "Query-Related Sentences". The notion of the topic-related sentence ranking is summarized as "the more important the property is, the higher its corresponding priority is" [5]. Similarly, the notion of query-related sentence ranking is summarized as "if the node denotes a literal (e.g., string) and it contains the words the user entered, the node is a query-related node." After this relation, the ranking levels of statement types are listed in Table 2 [5].

Table 2. Statement sequence after ranking [5]

Rank	Statements
1	Query-Related Node + Predicate + Topic Node
2	TopicNode+ Predicate+ Query-Related Node
3	Query-Related Node + Predicate + Un-Topic Node
4	Un-Topic Node + Predicate + Query-Related Node
5	Topic Node + Predicate + Un-Query-Related Node
6	Un-Query-Related Node + Predicate + Topic Node

Sig.ma[9] is a data aggregator and visualizer for the semantic Web, which currently aggregates data from Sindice[10], OKKAM[11], YBoss[12], Lod Sparql Endpoint[13]. Sig.ma, in its data acquisition process uses two types of ranking methods. First, the resource descriptions extracted from the RDF graphs and retrieved from different sources are ranked. It is stated that the descriptions are "matched and scored against the keyword phrase, considering both RDF literals and (with a lower score) words in URIs". The details of the ranking method are not given in [59]. After the ranking, the highest-ranking resource description from each source is selected. Second, the properties are ranked in the consolidation phase, in which "all selected resource descriptions are merged into a single entity profile" [59]. The details of this ranking method are given in section 2.4.

2.3 Graph Ranking

In this section, we review mainly the RDF graph-based ranking. However, there are several studies on data graphs like Dirichlet PageRank [17], which introduces a different approach on ranking the nodes of a data graph.

[37] defines a novel graph based system on ranking, which is called the "naming authority". It is described that the algorithm uses PageRank as a basis and it assigns "authority values to data sources based on a naming authority graph". By using the "naming authority matrix", it selects the results of the Web search.

[8]http://nlp.stanford.edu/IR-book/html/htmledition/tf-idf1-weighting-1.html

[9] http://sig.ma/

[10] http://sindice.com/

[11] http://api.okkam.org/

[12] http://boss.yahoo.com/

[13] http://lod.openlinksw.com/sparql

As an RDF graph ranking method, [28] introduces a ranking based on a "language modelling approach" which uses the counts of witness triples as the feature of ranking. In order to compute the relevancy of the nodes "Maximal-Marginal Relevance" approach [14] is used to re-rank the top ranked results.

Hermes project [58] has a scoring method named EF-IDF to combine popularity with distinctiveness. The EF-IDF score is computed and then "associated with elements of keywords and the structure indices." The EF-IDF basically calculates the popularity of an element by using its number of occurrence in the dataset[57]. As for the user interface, it is stated that "computed query graphs are sorted and finally, presented to the user for selection."

Similarly, Semplore [62] focuses on "Relation-based ranking" and ranking on the nodes according to their relation with other nodes based on the scores coming from the underlying IR engine rankings and TF-IDF scoring.

A novel popularity based approach is introduced by [48]. The ranking algorithm was applied on DBpedia[14]. Unlike several other PageRank [13] studies listed in Table 3, they created a "*relative* ranking system" [48] in which the rank of a node is changeable among queries and the nodes are ranked with regard to their neighbors. The popularity of a node is defined as a combination of two perspectives: "the one related to the words occurring within web documents, and the other one exploiting the social nature of the current web" [48]. [41] is a novel study on the distributed indexing of Linked Data and the data graph ranking method, which uses "top-N ranking and skylines" techniques.

Triplerank [34] method uses a 3-dimensional tensor to rank the RDF graph by calculating subject, object and property scores. The importance of semantic Web data model for Triplerank is descried as; it "enables the seamless representation of arbitrary semantic links"[34].

Scalability of the Web of data is an important concern, which was argued in [22] under the cases of K-Search [10] and Semplore [62] studies according to their inverted index structures and ranking over RDF graphs. It is stated that whereas RDF graph ranking allows "keyword-based- tree-shaped queries" [22], the performance and the scalability of such system becomes very limited.

RDF Xpress [30] is a project on searching RDF data. The ranking method is based on the previous study of the author on RDF Graph search [29], in which the ranking is based on IR-style ranking for RDF data. Other studies [27, 28, 42] explains how they rank the subgraphs under the basis of statistical language-models (LMs) [50].

SWSE [36] is a semantic Web search engine, which uses a ranking mechanism based on PageRank[13] which computes the on the data sources graph and the RDF graphs which is called the ReConRank [38].

2.4 Entity Ranking

The entity ranking techniques are of two types; ranking of an individual entity and ranking of an entity type. One of the first entity-based ranking is applied on xml documents. XRANK [35] is an xml element ranking method developed similar to PageRank[13], which is named as ElemRank.

The Touchgraph[15] project focuses on ranking the semantic associations. The semantic ranking metrics rankings are listed as context, subsumption, and trust. The statistical metrics are listed as rarity, popularity and association length.[2] Similarly, SemRank[3] also focuses on semantic associations.
SemSearch[16] is a semantic Web search engine. In this study, the entities are ranked according to how well they match the user query but do not consider the source of data. [45]

In Sig.ma[17], the properties of an entity are ranked. The ranking metric is a type of popularity approach for the property, which is described in the article as "the number of sources that have values for the property"[59].

Falcons search engine[18] uses a ranking mechanism described as the objects are ranked according to their popularity and relevance to the query [16]. In the study of another ObjectRank[31] mechanism, which uses "tractable description logic DL-Lite" as an underlying ontology. The algorithm is introduced as a generalization of PageRank [13] on Web pages.

The semantic search engine NAGA[19] ranks the structured data by using different factors like the "extraction confidence" and the "query length". [42]

ObjectRank[6] method ranks the entity nodes in a graph which uses different ranking techniques including PageRank[13] in calculating a global rank.

EntityAuthority [55] introduces a method to search for entities embedded in Web pages. The method is explained as similar to HITS[44] and ObjectRank[6] methods, and it is introduced that the system presents a richer approach for different kinds of nodes and underlying mathematical definitions. In the article, "query-independent PageRank (PR)", "Page-Inherited Authority (PIA)", "Un-Typed Authority with PageRank (UTA- PR)", "HITS (UTA-HITS)", "Entity-Derived Authority (EVA)" methods are evaluated.

WebOWL [8] is a semantic Web search engine which ranks the introduced "OWL Objects" as entities. The ranking is applied with a PageRank-like algorithm and some heuristic extensions.

The TRank[56] study ranks the entity types instead of ranking individual entities. A novel technique is developed on ranking the entity types based on the relevance of a given context.

Another study on entity type ranking is named as resource description ranking and property ranking in ECSSE[19]. The resource description ranking uses basic word matching techniques. The property ranking is based on a simple metric as "the number of sources that have values for the property".

In a different study[40], the category structure of Wikipedia[20] is used to rank entities. Since the Wikipedia is a self-categorized

[14] http://dbpedia.org/

[15] www.touchgraph.com

[16] https://code.google.com/p/semsearch/

[17] http://sig.ma/

[18] http://ws.nju.edu.cn/falcons/objectsearch/index.jsp

[19] https://www.mpi-inf.mpg.de/yago-naga/naga/

[20] www.wikipedia.com

project, the human defined structures are a good way to rank entities.

[47] is a recent survey written on Web of data searching, which surveys the current search engine trends in common terms. The entity ranking trials from Wikipedia-like databases are discussed. The entity ranking is mentioned under the name of the INEX Entity Ranking Track and TREC Entity Search Tracks.
The INEX Entity Ranking Track[23, 61], TREC Entity Search Tracks[7, 18] and Semantic Search Challenge[21] are three evaluation campaigns over Web of data. As it is inferred in [12]; Semantic Search Challenge in general searches over structured data in RDF, whereas TREC 2010 Entity Track [7] searches free text in Web pages. Furthermore, Semantic Search Challenge uses xml based RDF data and RDF descriptions, whereas INEX Entity Track [23] extracts text from xml files.

2.5 Document/Domain Ranking

Document/Domain ranking methods are mostly used as a domain independent technique to rank the documents or its domains. The document ranking emerges in different stages of online and offline data processing applications. Even a crawler engine can use a ranking mechanism as it is seen on the selection of a candidate URL to crawl. For instance, [26] is a crawler ranking method to rank candidate Web pages according to their semantic similarity of a concept.

In [46], the fundamental concepts of document ranking are outlined as; Frequency ranking (occurrence ranking), Horvath ranking (facet ranking), Facet ranking and frequency ranking in combination, Cosine relevance ranking.

Document ranking methods using structured data fields are discussed in [47] with the examples of BM25F and PRM-S [43, 51]. As [47] implies, these techniques allows the users to integrate text and metadata while ranking.

Some question answering systems are also introduces a document ranking mechanism. The PowerAqua [32] question answering system developed in KMI also uses a document ranking mechanism. Although the details of the ranking method are not given in the article, it is stated as the documents are ranked according to their semantic relationships with the query.[32]

Sindice semantic Web search engine[22] uses ad-hoc rules such as “prefer data sources whose hostnames correspond to the resource host- names” [60] to rank the domain of the data.

The most prominent study on document ranking is the PageRank method used by Google [13]. There are several studies on document ranking, which uses PageRank idea as the basis of their study.

As the most influential algorithm in Web technologies, PageRank [13] inspired several studies we have discussed in this survey. We have listed PageRank inspired studies in Table 3 for a cross reference categorization for the rest of the paper.

Table 3

Project/Article Name	Category	Ref.
Global PageRank of web communities	Document/Domain Ranking	[21]
ObjectRank	Graph Ranking	[6]
ReConRank	Graph Ranking	[38]
Hermes	Graph Ranking	[58]
Swoogle	Ontology Ranking	[33]
OntoKhoj	Ontology Ranking	[49]
Using naming authority..	Graph Ranking	[37]
Dirichlet PageRank..	Graph Ranking	[17]
Triplerank	Graph Ranking	[34]
Ranking complex relationships	Entity Ranking	[2]
XRANK	Entity Ranking	[35]
Semplore	Graph Ranking	[62]
OWLRank	Entity Ranking	[8]
SWSE	Graph Ranking	[39]

[21] http://challenge.semanticweb.org

[22] http://sindice.com/

3. DISCUSSION

In this brief survey we identified and classified the studies on ranking methods over the Linked Data. Although the ranking methods such as PageRank [13] and HITS [44] highly influenced the Web, there is no significantly new algorithm to be accepted in semantic Web. Ranking can be applied in almost in every layer[23] of the semantic Web. At the lowest levels, the triples or XML related data are indexed and ranked, RDF descriptions and document files are ranked, ontology descriptions and properties are ranked; and at a higher level the results are ranked. Semantic Web search engines use basic level ranking approaches yet, and they are not focused on a novel ranking method yet. We expect there will be more work in this area as semantic Web and linked data are getting more attention in the Web search community.

4. REFERENCES

[1] Alani, H., Brewster, C. and Shadbolt, N. 2006. Ranking ontologies with AKTiveRank. *The Semantic Web-ISWC 2006* (2006), 1–15.

[2] Aleman-Meza, B., Halaschek-Weiner, C., Arpinar, I.B., Ramakrishnan, C. and Sheth, A.P. 2005. Ranking complex relationships on the semantic web. *IEEE Internet Computing*. 9, 3 (2005), 37–44.

[3] Anyanwu, K., Maduko, A. and Sheth, A. 2005. SemRank: ranking complex relationship search results on the semantic web. *14th International Conference on World Wide Web* (2005), 117–127.

[4] Aquin, M., Baldassarre, C., Gridinoc, L., Sabou, M., Angeletou, S. and Motta, E. 2007. *Watson: Supporting next generation semantic web applications*.

[5] Bai, X., Delbru, R. and Tummarello, G. 2008. RDF snippets for Semantic Web search engines. *On the Move to Meaningful Internet Systems: OTM 2008* (2008), 1304–1318.

[6] Balmin, A., Hristidis, V. and Papakonstantinou, Y. 2004. Objectrank: Authority-based keyword search in databases. *Proceedings of the Thirtieth international*

[23] http://www.w3.org/2001/12/semweb-fin/swlevels.png

conference on Very large data bases-Volume 30 (2004), 564–575.

[7] Balog, K., Serdyukov, P. and Vries, A. 2010. Overview of the TREC 2010 entity track. (2010).

[8] Batzios, A. and Mitkas, P. a. 2012. WebOWL: A Semantic Web search engine development experiment. *Expert Systems with Applications*. 39, 5 (Apr. 2012), 5052–5060.

[9] Berners-lee, B.T. and Hendler, J. 2001. The Semantic Web. *Scientific American*. May 2001 (2001).

[10] Bhagdev, R., Chapman, S., Ciravegna, F., Lanfranchi, V. and Petrelli, D. 2008. Hybrid search: Effectively combining keywords and semantic searches. *The Semantic Web: Research and Applications* (2008), 554–568.

[11] Bizer, C., Heath, T. and Berners-Lee, T. 2009. Linked data-the story so far. *International Journal on Semantic ….* (2009).

[12] Blanco, R., Halpin, H., Herzig, D.M., Mika, P., Pound, J., Thompson, H.S. and Tran, T. 2013. Repeatable and reliable semantic search evaluation. *Web Semantics: Science, Services and Agents on the World Wide Web*. 21, (Aug. 2013), 14–29.

[13] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 30, 1 (Apr. 1998), 107–117.

[14] Carbonell, J. and Goldstein, J. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (New York, New York, USA, 1998), 335–336.

[15] Chebolu, P. and Melsted, P. 2008. PageRank and the random surfer model. *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms* (2008), 1010–1018.

[16] Cheng, G. and Qu, Y. 2009. Searching Linked Objects with Falcons. *International Journal on Semantic Web and Information Systems*. 5, 3 (2009), 49–70.

[17] Chung, F., Tsiatas, A. and Xu, W. 2011. Dirichlet PageRank and trust-based ranking algorithms. *Algorithms and Models for the Web Graph* (2011), 103–114.

[18] Craswell, N. and Soboroff, I. 2005. Overview of the TREC-2005 Enterprise Track Email search task. (2005), 1–7.

[19] Cyganiak, R., Catasta, M. and Tummarello, G. 2009. Towards ECSSE : live Web of Data search and integration. *Proceedings of the Semantic Search 2009 Workshop* (2009).

[20] Davies, J. and Weeks, R. 2004. QuizRDF: Search technology for the semantic web. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (2004), 1–8.

[21] Davis, J. V. and Dhillon, I.S. 2006. Estimating the global pagerank of web communities. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, New York, USA, 2006), 116–125.

[22] Delbru, R., Campinas, S. and Tummarello, G. 2012. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*. 10, (Jan. 2012), 33–58.

[23] Demartini, G., Iofciu, T. and Vries, A.P. De 2010. Overview of the INEX 2009 Entity Ranking. *Focused Retrieval and Evaluation* (2010), 254–264.

[24] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V. and Sachs, J. 2004. Swoogle : A Search and Metadata Engine for the Semantic Web. *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (2004), 652–659.

[25] Ding, L., Pan, R., Finin, T. and Joshi, A. 2005. Finding and ranking knowledge on the semantic web. *The Semantic Web–ISWC 2005* (2005), 156–170.

[26] Du, Y. and Hai, Y. 2013. Semantic ranking of web pages based on formal concept analysis. *Journal of Systems and Software*. 86, 1 (Jan. 2013), 187–197.

[27] Elbassuoni, S. and Blanco, R. 2011. Keyword search over RDF graphs. *Proceedings of the 20th ACM international ….* (2011), 237–242.

[28] Elbassuoni, S., Ramanath, M., Schenkel, R., Marcin, S. and Weikum, G. 2009. Language-model-based ranking for queries on RDF-graphs. *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), 977–986.

[29] Elbassuoni, S., Ramanath, M., Schenkel, R. and Weikum, G. 2010. Searching RDF Graphs with SPARQL and Keywords. *IEEE Data Eng. Bull*. 33, 1 (2010), 16–24.

[30] Elbassuoni, S., Ramanath, M. and Weikum, G. 2012. RDF Xpress: a flexible expressive RDF search engine. *Proceedings of the 35th ….* (2012), 2011.

[31] Fazzinga, B., Gianforme, G., Gottlob, G. and Lukasiewicz, T. 2011. Semantic Web search based on ontological conjunctive queries. *Web Semantics: Science, Services and Agents on the World Wide Web*. 9, 4 (Dec. 2011), 453–473.

[32] Fernandez, M., Lopez, V., Sabou, M., Uren, V., Vallet, D., Motta, E. and Castells, P. 2008. Semantic Search Meets the Web. *2008 IEEE International Conference on Semantic Computing* (Aug. 2008), 253–260.

[33] Finin, T. Swoogle : A Semantic Web Search and Metadata Engine ∗.

[34] Franz, T., Schultz, A., Sizov, S. and Staab, S. 2009. Triplerank: Ranking semantic web data by tensor decomposition. *The Semantic Web-ISWC 2009* (2009), 213–228.

[35] Guo, L. and Shao, F. 2003. XRANK: ranked keyword search over XML documents. *Proceedings of the 2003*

ACM SIGMOD international conference on Management of data (New York, NY, USA, 2003), 16–27.

[36] Harth, A., Hogan, A., Delbru, R., Riain, S.O. and Decker, S. 2007. SWSE : Answers Before Links ! *Semantic Web Challenge* (2007).

[37] Harth, A., Kinsella, S. and Decker, S. 2009. Using naming authority to rank data and ontologies for web search. *The Semantic Web-ISWC 2009* (2009), 277–292.

[38] Hogan, A., Harth, A. and Decker, S. 2006. Reconrank: A scalable ranking method for semantic web data with context. *Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006)* (Athens, GA, USA, 2006).

[39] Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A. and Decker, S. 2011. Semantic Search- Reading - Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine. *Web Semantics: Science, Services and Agents on the World Wide Web*. 9, 4 (Dec. 2011), 365–401.

[40] Kaptein, R. and Kamps, J. 2013. Exploiting the category structure of Wikipedia for entity ranking. *Artificial Intelligence*. 194, (Jan. 2013), 111–129.

[41] Karnstedt, M., Sattler, K.-U. and Hauswirth, M. 2012. Scalable distributed indexing and query processing over Linked Data. *Web Semantics: Science, Services and Agents on the World Wide Web*. 10, (Jan. 2012), 3–32.

[42] Kasneci, G., Suchanek, F.M., Ifrim, G., Ramanath, M. and Weikum, G. 2008. NAGA : Searching and Ranking Knowledge. *IEEE 24th International Conference on Data Engineering, 2008. ICDE 2008* (2008), 953–962.

[43] Kim, J.Y. and Croft, W.B. 2012. A field relevance model for structured document retrieval. *ECIR'12 Proceedings of the 34th European conference on Advances in Information Retrieval* (2012), 97–108.

[44] Kleinberg, J. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*. 46, 5 (1999), 604–632.

[45] Lei, Y., Uren, V. and Motta, E. 2006. Semsearch: A search engine for the semantic web. *Managing Knowledge in a World of Networks* (2006), 238–245.

[46] Materne, A. and Sleightholme, G. 2013. Methods of ranking search results for searches based on multiple search concepts carried out in multiple databases. *World Patent Information*. (Oct. 2013), 1–12.

[47] De Melo, G. and Hose, K. 2013. Searching the web of data. *Advances in Information Retrieval* (2013), 1–4.

[48] Mirizzi, R., Ragone, A., Noia, T. Di and Sciascio, E. Di 2010. Ranking the linked data: the case of dbpedia. *Web Engineering* (2010), 337–354.

[49] Patel, C., Supekar, K., Lee, Y. and Park, E. 2003. OntoKhoj: a semantic web portal for ontology searching, ranking and classification. *Proceedings of the 5th ACM international workshop on Web information and data management* (2003), 58–61.

[50] Ponte, J.M. and Croft, W.B. 1998. A language modeling approach to information retrieval. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (New York, New York, USA, 1998), 275–281.

[51] Robertson, S., Zaragoza, H. and Taylor, M. 2004. Simple BM25 extension to multiple weighted fields. *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (2004), 42–49.

[52] Rocha, C., Schwabe, D. and Aragão, M.P. 2004. A Hybrid Approach for Searching in the Semantic Web. *Proceedings of the 13th international conference on World Wide Web* (2004), 374–383.

[53] Skoutas, D., Simitsis, A. and Sellis, T. 2007. A ranking mechanism for semanticweb service discovery. *2007 IEEE Congress on Services* (2007), 41–48.

[54] Stojanovic, N., Studer, R. and Stojanovic, L. 2003. An Approach for the Ranking of Query Results in the Semantic Web. *The Semantic Web-ISWC 2003* (2003), 500–516.

[55] Stoyanovich, J., Bedathur, S., Berberich, K. and Weikum, G. 2007. EntityAuthority: Semantically Enriched Graph-Based Authority Propagation. *WebDB* (2007).

[56] Tonon, A. and Catasta, M. 2013. TRank: Ranking Entity Types Using the Web of Data. *The Semantic Web–ISWC 2013* (2013), 640–656.

[57] Tran, T., Herzig, D.M. and Ladwig, G. 2011. SemSearchPro–using semantics throughout the search process. *Web Semantics: Science, Services and Agents on the World Wide Web*. 9, 4 (2011), 349–364.

[58] Tran, T., Wang, H. and Haase, P. 2009. Hermes: Data Web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*. 7, 3 (Sep. 2009), 189–203.

[59] Tummarello, G., Cyganiak, R., Catasta, M., Danielczyk, S., Delbru, R. and Decker, S. 2010. Sig.ma: Live views on the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*. 8, 4 (Nov. 2010), 355–364.

[60] Tummarello, G., Delbru, R. and Oren, E. 2007. Sindice.com: Weaving the open linked data. *The Semantic Web* (2007), 552–565.

[61] De Vries, A.P., Vercoustre, A.-M., A. Thom, J., Craswell, N. and Lalmas, M. 2008. Overview of the INEX 2007 entity ranking track. *Focused Access to XML Documents* (2008), 245–251.

[62] Wang, H., Liu, Q., Penin, T., Fu, L., Zhang, L., Tran, T., Yu, Y. and Pan, Y. 2009. Semplore: A scalable IR approach to search the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*. 7, 3 (Sep. 2009), 177–188.