Noname manuscript No.
(will be inserted by the editor)

A Survey of Approaches for Ranking on Structured Data

First Author · Second Author

Received: date / Accepted: date

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords First keyword · Second keyword · More

1 Style

Text with citations [**?**] and [**?**].

1.1 Subsection title

as required. Don't forget to give each section and subsection a unique label (see Sect. 1).

Paragraph headings Use paragraph headings as needed.

$$a^2 + b^2 = c^2 \tag{1}$$

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

Fig. 1: Please write your figure caption here

Table 1: Please write your table caption here

first	second	third
number	number	number
number	number	number

2 Introduction

Outline

3 The Ranking Problem

3.1 Ranking on Structured Data

Structured data: can be conceived as graphs
Ranking: traditionally, core IR problem, many techniques have been adopted to deal with structured data
Ranking over structured data: ranked list of results
Types of ranked results: entities, relations, subgraphs, entire datasets (graphs)
Evaluating ranked results: different metrics

3.2 Ranking on Structured Data vs. Ranking on Unstructured Data

Discuss differences
Say we focus on structured data

3.3 Ranking vs. Matching

Discuss differences
Say we focus on ranking

3.4 Applications

Optional / can be short but should provide good coverage of different application domains

Fig. 2: Please write your figure caption here

4 Ranking Approaches

4.1 Historical Development

4.2 Generic Architecture

4.3 Main Dimensions

Features: content, structure, context, two meanings of context: context of user; and context of the data such as trust values of source, truth value / confidence degree of individual records; make clear this survey does not discuss user-context based methods in details
Results: entity, relationships, subgraphs, graphs (entire dataset)
Techniques:
- may vary in terms of methods: used *NLP* for understanding keywords, used *ML* for classifying keywords (I know some ML-based query classification approaches for document retrieval, are there also examples for structured data ranking?) and for learning to rank, (other) statistics-based methods: Vector Space Model (VSM), Language Modeling (LM), Information Theory
- may vary in terms of heuristics: two main ones are query-relevance and popularity; other heuristics are proximity, informativeness (based on Information Theory, e.g. entropy) and context-based: trust, truth value, locality etc.

4.4 Foundation

Here we discuss all basics needed to understand the approaches presented in the following sections.
Vector-Space Model: also discuss pivoted normalization and point out problems of short document length etc. in the context of structured data
Language Modeling: also discuss smoothing strategies
Link Analysis
Learning to Rank
...

4.5 Taxonomy of Ranking Approaches

Classify approaches mainly based on the type of *heuristics* they used. Those that use same heuristics are distinguished in terms of *methods*. E.g. Query-relevance based solutions can be further distinguished in terms of VSM-based and LM-based approaches.

5 Query-Relevance

6 Popularity

7 Other Heuristics

8 Combining Heuristics through Learning to Rank

9 Sample Approaches from the Literature

Before, we discuss the concepts/techniques behind existing approaches. Here we provide a sample

10 Open Research Challenges

11 Conclusion

For consistency, we use bibtex entries from dblp, e.g. `http://www.dblp.org/rec/bibtex/conf/sigmod/BalminKT12`. Many bibtex entries are already in the paper.tex file so just searched for these entries there first and add only new entries when needed to avoid redundancy.

References

1. R. A. Baeza-Yates, A. Gionis, F. Junqueira, V. Murdock, V. Plachouras, and F. Silvestri, "The impact of caching on search engines," in *SIGIR*, 2007, pp. 183–190.
2. P. Mika, E. Meij, and H. Zaragoza, "Investigating the semantic gap through query log analysis," in *International Semantic Web Conference*, 2009, pp. 441–455.
3. K. Haas, P. Mika, P. Tarjan, and R. Blanco, "Enhanced results for web search," in *SIGIR*, W.-Y. Ma, J.-Y. Nie, R. A. Baeza-Yates, T.-S. Chua, and W. B. Croft, Eds. ACM, 2011, pp. 725–734.
4. E. Kaufmann and A. Bernstein, "How useful are natural language interfaces to the semantic web for casual end-users?" in *ISWC/ASWC*, 2007, pp. 281–294.
5. T. Tran, T. Mathäß, and P. Haase, "Usability of keyword-driven schema-agnostic search," in *ESWC (2)*, 2010, pp. 349–364.
6. C. Mangold, "A survey and classification of semantic search approaches," *IJMSO*, vol. 2, no. 1, pp. 23–34, 2007.
7. T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR*, 1999, pp. 50–57.
8. X. Wei and W. B. Croft, "Lda-based document models for ad-hoc retrieval," in *SIGIR*, 2006, pp. 178–185.
9. O. Egozi, S. Markovitch, and E. Gabrilovich, "Concept-based information retrieval using explicit semantic analysis," *ACM Trans. Inf. Syst.*, vol. 29, no. 2, p. 8, 2011.
10. G. Cheng and Y. Qu, "Searching linked objects with falcons: Approach, implementation and evaluation," *Int. J. Semantic Web Inf. Syst.*, vol. 5, no. 3, pp. 49–70, 2009.
11. G. Tummarello, R. Cyganiak, M. Catasta, S. Danielczyk, R. Delbru, and S. Decker, "Sig.ma: Live views on the web of data," *J. Web Sem.*, vol. 8, no. 4, pp. 355–364, 2010.

12. T. Tran, H. Wang, and P. Haase, "Hermes: Data web search on a pay-as-you-go integration infrastructure," *J. Web Sem.*, vol. 7, no. 3, pp. 189–203, 2009.
13. T. Tran, D. M. Herzig, and G. Ladwig, "Semsearchpro - using semantics throughout the search process," *J. Web Sem.*, vol. 9, no. 4, pp. 349–364, 2011.
14. V. Lopez, V. S. Uren, E. Motta, and M. Pasin, "Aqualog: An ontology-driven question answering system for organizational semantic intranets," *J. Web Sem.*, vol. 5, no. 2, pp. 72–105, 2007.
15. G. Kasneci, F. M. Suchanek, G. Ifrim, M. Ramanath, and G. Weikum, "Naga: Searching and ranking knowledge," in *ICDE*, 2008, pp. 953–962.
16. K. Anyanwu and A. P. Sheth, "P-queries: enabling querying for semantic associations on the semantic web," in *WWW*, 2003, pp. 690–699.
17. G. Li, S. Ji, C. Li, and J. Feng, "Efficient type-ahead search on relational data: a tastier approach," in *SIGMOD Conference*, 2009, pp. 695–706.
18. E. Kandogan, R. Krishnamurthy, S. Raghavan, S. Vaithyanathan, and H. Zhu, "Avatar semantic search: a database approach to information retrieval," in *SIGMOD Conference*, 2006, pp. 790–792.
19. A. Wagner, G. Ladwig, and T. Tran, "Browsing-oriented semantic faceted search," in *DEXA (1)*, 2011, pp. 303–319.
20. P. Heim, T. Ertl, and J. Ziegler, "Facet graphs: Complex semantic querying made easy," in *ESWC (1)*, 2010, pp. 288–302.
21. A. Harth, "Visinav: A system for visual search and navigation on web data," *J. Web Sem.*, vol. 8, no. 4, pp. 348–354, 2010.
22. R. Kaptein, P. Serdyukov, A. P. de Vries, and J. Kamps, "Entity ranking using wikipedia as a pivot," in *CIKM*, 2010, pp. 69–78.
23. M. Bron, K. Balog, and M. de Rijke, "Ranking related entities: components and analyses," in *CIKM*, 2010, pp. 1079–1088.
24. B. Fazzinga, G. Gianforme, G. Gottlob, and T. Lukasiewicz, "Semantic web search based on ontological conjunctive queries," *J. Web Sem.*, vol. 9, no. 4, pp. 453–473, 2011.
25. P. Castells, M. Fernández, and D. Vallet, "An adaptation of the vector-space model for ontology-based information retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 261–272, 2007.
26. M. Fernández, I. Cantador, V. Lopez, D. Vallet, P. Castells, and E. Motta, "Semantically enhanced information retrieval: An ontology-based approach," *J. Web Sem.*, vol. 9, no. 4, pp. 434–452, 2011.
27. O. Corby, R. Dieng-Kuntz, and C. Faron-Zucker, "Querying the semantic web with corese search engine," in *ECAI*, 2004, pp. 705–709.
28. T. H. Cao, T. D. Cao, and T. L. Tran, "A robust ontology-based method for translating natural language queries to conceptual graphs," in *ASWC*, 2008, pp. 479–492.
29. F. Giunchiglia, U. Kharkevich, and I. Zaihrayeu, "Concept search," in *ESWC*, 2009, pp. 429–444.
30. H. P. Giger, "Concept based retrieval in classical ir systems," in *SIGIR*, 1988, pp. 275–289.
31. F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *WWW*, 2007, pp. 697–706.
32. C. J. van Rijsbergen, "Towards an information logic," in *SIGIR*, 1989, pp. 77–86.
33. S. Vassiliadis, G. Triantafyllos, and W. Kobrosly, "A fuzzy reasoning database question answering system," *IEEE Trans. Knowl. Data Eng.*, vol. 6, no. 6, pp. 868–882, 1994.
34. C.-H. Chang, M. Kayed, M. R. Girgis, and K. F. Shaalan, "A survey of web information extraction systems," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1411–1428, 2006.
35. D. Trieschnigg, W. Kraaij, and M. J. Schuemie, "Concept based document retrieval for genomics literature," in *TREC*, 2006.
36. W. Zhou, C. T. Yu, V. I. Torvik, and N. R. Smalheiser, "A concept-based framework for passage retrieval at genomics," in *TREC*, 2006.
37. A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff, "Semantic annotation, indexing, and retrieval," *J. Web Sem.*, vol. 2, no. 1, pp. 49–79, 2004.
38. J. Chu-Carroll, J. M. Prager, K. Czuba, D. A. Ferrucci, and P. A. Duboué, "Semantic search via xml fragments: a high-precision approach to ir," in *SIGIR*, 2006, pp. 445–452.
39. J. Chu-Carroll and J. M. Prager, "An experimental study of the impact of information extraction accuracy on semantic search performance," in *CIKM*, 2007, pp. 505–514.
40. J. Kleb and A. Abecker, "Entity reference resolution via spreading activation on rdf-graphs," in *ESWC (1)*, 2010, pp. 152–166.
41. J. Zhu, Z. Nie, J.-R. Wen, B. Zhang, and W.-Y. Ma, "2d conditional random fields for web information extraction," in *ICML*, 2005, pp. 1044–1051.
42. O. Etzioni, A. Fader, J. Christensen, S. Soderland, and Mausam, "Open information extraction: The second generation," in *IJCAI*, 2011, pp. 3–10.
43. F. M. Suchanek, M. Sozio, and G. Weikum, "Sofie: a self-organizing framework for information extraction," in *WWW*, 2009, pp. 631–640.
44. J. Zhu, Z. Nie, X. Liu, B. Zhang, and J.-R. Wen, "Statsnowball: a statistical approach to extracting entity relationships," in *WWW*, 2009, pp. 101–110.
45. M. Zhong and X. Huang, "Concept-based biomedical text retrieval," in *SIGIR*, 2006, pp. 723–724.
46. Y. Qiu and H.-P. Frei, "Concept based query expansion," in *SIGIR*, 1993, pp. 160–169.
47. E. M. Voorhees, "Using wordnet to disambiguate word senses for text retrieval," in *SIGIR*, 1993, pp. 171–180.
48. ——, "Query expansion using lexical-semantic relations," in *SIGIR*, 1994, pp. 61–69.
49. E. Meij, D. Trieschnigg, M. de Rijke, and W. Kraaij, "Conceptual language models for domain-specific retrieval," *Inf. Process. Manage.*, vol. 46, no. 4, pp. 448–469, 2010.
50. X. Jiang and A.-H. Tan, "Ontosearch: A full-text search engine for the semantic web," in *AAAI*, 2006.
51. D. Damljanovic, M. Agatonovic, and H. Cunningham, "Natural language interfaces to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction," in *ESWC (1)*, 2010, pp. 106–120.
52. P. Cimiano, P. Haase, J. Heizmann, M. Mantel, and R. Studer, "Towards portable natural language interfaces to knowledge bases - the case of the orakel system," *Data Knowl. Eng.*, vol. 65, no. 2, pp. 325–354, 2008.
53. T. Tran, H. Wang, S. Rudolph, and P. Cimiano, "Top-k exploration of query candidates for efficient keyword search on graph-shaped (rdf) data," in *ICDE*, 2009, pp. 405–416.
54. V. M. Ngo and T. H. Cao, "Ontology-based query expansion with latently related named entities for semantic text search," in *Advances in Intelligent Information and Database Systems*, 2010, pp. 41–52.
55. G. Ladwig and T. Tran, "Index structures and top-k join algorithms for native keyword search databases," in *CIKM*, 2011, pp. 1505–1514.
56. C. Rocha, D. Schwabe, and M. P. de Aragão, "A hybrid approach for searching in the semantic web," in *WWW*, 2004, pp. 374–383.
57. K. Schumacher, M. Sintek, and L. Sauermann, "Combining fact and document retrieval with spreading activation for semantic desktop search," in *ESWC*, 2008, pp. 569–583.
58. J. Heflin, J. A. Hendler, and S. Luke, "Shoe: A blueprint for the semantic web," in *Spinning the Semantic Web*, 2003, pp. 29–63.
59. B. Aleman-Meza, C. Halaschek-Wiener, I. B. Arpinar, C. Ramakrishnan, and A. P. Sheth, "Ranking complex relationships on the semantic web," *IEEE Internet Computing*, vol. 9, no. 3, pp. 37–44, 2005.
60. K. Anyanwu, A. Maduko, and A. P. Sheth, "Semrank: ranking complex relationship search results on the semantic web," in *WWW*, 2005, pp. 117–127.

61. Z. Nie, Y. Zhang, J.-R. Wen, and W.-Y. Ma, "Object-level ranking: bringing order to web objects," in *WWW*, 2005, pp. 567–574.
62. T. Cheng, X. Yan, and K. C.-C. Chang, "Entityrank: Searching entities directly and holistically," in *VLDB*, 2007, pp. 387–398.
63. L. Ding, R. Pan, T. W. Finin, A. Joshi, Y. Peng, and P. Kolari, "Finding and ranking knowledge on the semantic web," in *International Semantic Web Conference*, 2005, pp. 156–170.
64. R. Delbru, N. Toupikov, M. Catasta, G. Tummarello, and S. Decker, "Hierarchical link analysis for ranking web data," in *ESWC (2)*, 2010, pp. 225–239.
65. A. Harth, S. Kinsella, and S. Decker, "Using naming authority to rank data and ontologies for web search," in *International Semantic Web Conference*, 2009, pp. 277–292.
66. A. Balmin, V. Hristidis, and Y. Papakonstantinou, "Objectrank: Authority-based keyword search in databases," in *VLDB*, 2004, pp. 564–575.
67. J. Stoyanovich, S. J. Bedathur, K. Berberich, and G. Weikum, "Entityauthority: Semantically enriched graph-based authority propagation," in *WebDB*, 2007.
68. T. Tao and C. Zhai, "An exploration of proximity measures in information retrieval," in *SIGIR*, 2007, pp. 295–302.
69. L. Guo, F. Shao, C. Botev, and J. Shanmugasundaram, "Xrank: Ranked keyword search over xml documents," in *SIGMOD Conference*, 2003, pp. 16–27.
70. T. M. V. Le, T. H. Cao, S. M. Hoang, and J. Cho, "Ontology-based proximity search," in *iiWAS*, 2011, pp. 288–291.
71. F. Liu, C. T. Yu, W. Meng, and A. Chowdhury, "Effective keyword search in relational databases," in *SIGMOD Conference*, 2006, pp. 563–574.
72. R. Blanco, P. Mika, and S. Vigna, "Effective and efficient entity search in rdf data," in *International Semantic Web Conference (1)*, 2011, pp. 83–97.
73. Z. Nie, Y. Ma, S. Shi, J.-R. Wen, and W.-Y. Ma, "Web object retrieval," in *WWW*, 2007, pp. 81–90.
74. V. Bicer, T. Tran, and R. Nedkov, "Ranking support for keyword search on structured data using relevance models," in *CIKM*, 2011, pp. 1669–1678.
75. S. Elbassuoni, M. Ramanath, R. Schenkel, and G. Weikum, "Searching rdf graphs with sparql and keywords," *IEEE Data Eng. Bull.*, vol. 33, no. 1, pp. 16–24, 2010.
76. L. Dali, B. Fortuna, T. Tran, and D. Mladenic, "Query-independent learning to rank for rdf entity search," in *ESWC*, 2012.
77. J. Coffman and A. C. Weaver, "Learning to rank results in relational keyword search," in *CIKM*, 2011, pp. 1689–1698.
78. C. Jonquet, P. LePendu, S. M. Falconer, A. Coulet, N. F. Noy, M. A. Musen, and N. H. Shah, "Ncbo resource index: Ontology-based search and mining of biomedical resources," *J. Web Sem.*, vol. 9, no. 3, pp. 316–324, 2011.
79. K. Balog, L. Azzopardi, and M. de Rijke, "A language modeling framework for expert finding," *Inf. Process. Manage.*, vol. 45, no. 1, pp. 1–19, 2009.
80. K. Balog, M. Bron, and M. de Rijke, "Query modeling for entity search based on terms, categories, and examples," *ACM Trans. Inf. Syst.*, vol. 29, no. 4, p. 22, 2011.
81. J. Pehcevski, A.-M. Vercoustre, and J. A. Thom, "Exploiting locality of wikipedia links in entity ranking," in *ECIR*, 2008, pp. 258–269.
82. H. Bast, A. Chitea, F. M. Suchanek, and I. Weber, "Ester: efficient search on text, entities, and relations," in *SIGIR*, 2007, pp. 671–678.
83. C. Wang, M. Xiong, Q. Zhou, and Y. Yu, "Panto: A portable natural language interface to ontologies," in *ESWC*, 2007, pp. 473–487.
84. Y. Li, H. Yang, and H. V. Jagadish, "Nalix: A generic natural language search environment for xml data," *ACM Trans. Database Syst.*, vol. 32, no. 4, 2007.
85. Y.-J. Han, T.-G. Noh, S.-B. Park, S.-Y. Park, and S.-J. Lee, "A natural language interface of thorough coverage by concordance with knowledge bases," in *IUI*, 2010, pp. 325–328.
86. D. A. Ferrucci, E. W. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. M. Prager, N. Schlaefer, and C. A. Welty, "Building watson: An overview of the deepqa project," *AI Magazine*, vol. 31, no. 3, pp. 59–79, 2010.
87. J. Chu-Carroll, K. Czuba, J. M. Prager, and A. Ittycheriah, "In question answering, two heads are better than one," in *HLT-NAACL*, 2003.
88. J. M. Prager, E. W. Brown, A. Coden, and D. R. Radev, "Question-answering by predictive annotation," in *SIGIR*, 2000, pp. 184–191.
89. R. van Zwol, B. Sigurbjörnsson, R. Adapala, L. G. Pueyo, A. Katiyar, K. Kurapati, M. Muralidharan, S. Muthu, V. Murdock, P. Ng, A. Ramani, A. Sahai, S. T. Sathish, H. Vasudev, and U. Vuyyuru, "Faceted exploration of image search results," in *WWW*, 2010, pp. 961–970.
90. J. Coffman and A. C. Weaver, "A framework for evaluating database keyword search strategies," in *CIKM*, 2010, pp. 729–738.
91. Y. Li, I. Chaudhuri, H. Yang, S. Singh, and H. V. Jagadish, "Enabling domain-awareness for a generic natural language interface," in *AAAI*, 2007, pp. 833–838.