Graph search and beyond

SIGIR 2015 workshop summary

Alonso, O.; Hearst, M.A.; Kamps, J.

DOI
10.1145/2766462.2767855

Publication date
2015

Document Version
Final published version

Published in
SIGIR 2015

Graph Search and Beyond

SIGIR 2015 Workshop Summary

Omar Alonso
Microsoft
Mountain View, CA
USA

Marti A. Hearst
UC Berkeley
Berkeley, CA
USA

Jaap Kamps
University of Amsterdam
Amsterdam
The Netherlands

ABSTRACT

Modern Web data is highly structured in terms of entities and relations from large knowledge resources, geo-temporal references and social network structure, resulting in a massive multidimensional graph. This graph essentially unifies both the searcher and the information resources that played a fundamentally different role in traditional IR, and "Graph Search" offers major new ways to access relevant information. Graph search affects both query formulation (complex queries about entities and relations building on the searcher's context) as well as result exploration and discovery (slicing and dicing the information using the graph structure) in a completely personalized way. This new graph based approach introduces great opportunities, but also great challenges, in terms of data quality and data integration, user interface design, and privacy.

We view the notion of "graph search" as searching information from your personal point of view (you are the query) over a highly structured and curated information space. This goes beyond the traditional two-term queries and ten blue links results that users are familiar with, requiring a highly interactive session covering both query formulation and result exploration. The workshop attracted a range of researchers working on this and related topics, and made concrete progress working together on one of the greatest challenges in the years to come.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Query formulation, Search process, Selection process*

Keywords: Graph search; Semantic search; Personalization; Exploration; Query suggest

1. INTRODUCTION

Information on the Web is increasingly structured in terms of entities and relations from large knowledge resources, geo-temporal references and social network structure, resulting in a massive multidimensional graph. This graph essentially unifies both the searcher and the information resources that played a fundamentally different role in traditional IR, and offers major new ways to access relevant information. In services that rely on personalized information like social networks, the graph plays an even more important role, in other words: *you are the query*.

Graph search affects both query formulation as well as result exploration and discovery. On the one hand, it allows for incrementally expressing complex information needs that triangulate information about multiple entities or entity types, relations between those entities, with various filters on geo-temporal constraints or the sources of information used (or ignored), and taking into account the rich profile and context information of the searcher (and his/her peers, and peers of peers, etc). On the other hand, it allows for more powerful ways to explore the results from various aspects and viewpoints, by slicing and dicing the information using the graph structure, and using the same structure for explaining why results are retrieved or recommended, and by whom.

This new graph based information seeking approach introduces great opportunities, but also great challenges, both technical ranging from data quality and data integration to user interface design, as well as ethical challenges in terms of privacy; transparency, bias and control; and avoiding the so-called filter bubbles. Graph search is already available today in many flavors with different levels of interactivity. Social network-based services like Facebook and LinkedIn provide flexibility to search their personal network form many diverse angles. Web search engines like Google and Bing rely more on using graphs to show related content as a mechanism to include other possible contexts for a given query. Clearly, it is not limited to web, and can be applied to other highly structured data. Just to give an example, the hansards or parliamentary proceedings are fully public data with a clear graph structure linking every speech to the respective speaker, their role in parliament and their political party. Graph search allows to explore politics from the viewpoint of individual members of parliament or government.

At a high level, graph search seems limited to familiar entity types (e.g., Facebook entities) and templates. How far can this scale? Will this work on truly open domains? There is a huge potential to use the graph to go beyond recommendations for new friends and contacts or semantically related content. Unlocking the potential of richer knowledge sources for new search strategies requires us to think outside

SIGIR'15, August 09–13, 2015, Santiago, Chile.
ACM 978-1-4503-3621-5/15/08.
DOI: http://dx.doi.org/10.1145/2766462.2767855.

the box, by combining different insights from IR, semantic search, data integration, query expansion and user interfaces to name a few.

2. MANY OPEN RESEARCH QUESTIONS

We view the notion of "graph search" as searching information from your personal point of view (you are the query) over a highly structured and curated information space. This goes beyond the traditional two-term queries and ten blue links results that users are familiar with, requiring a highly interactive session covering both query formulation and result exploration.

This raises many open questions:

IR Theory What happens if search gets personal? Does this break the classic dichotomy between users and documents, as users are nodes in the social network data themselves? What is the consequence of ultimate personalization, as the local graph differs for all users? As the local graph structure is key, does this obviate the need for large central indexes? Do these types of requests fit in the classic paradigm (e.g., Broder's taxonomy)? How does this shift the balance between the control of the searcher and the ranker over the result set?

Data Integration Building a knowledge graph requires massive data integration at many levels: are there trade-offs in simplicity and level of detail (such as the classic knowledge representation trade-off)? What levels of granularity and comprehensiveness are needed for effective deployment? What quality is needed: is any noise acceptable? How to deal with near duplicate detection, conflation, or entity disambiguation?

Use Cases and Applications Rather than a universal solution, graph search is particularly useful for specific types of information needs and queries. What are the data and tasks that make graph search works? What kind of scenarios that would benefit from a graph model? In what context can switching perspectives by showing results from the vista of other persons useful?

Query formulation How to move from singular queries to highly interactive sessions with multiple variant queries? What new tools are needed to help a searcher construct the appropriate graph search query using refinements or filters to better articulate their needs, or explore further aspects? How can we augment query autocompletion to actively prompt user to interactively construct longer queries exploring different aspects?

Result Exploration There is a radical shift towards the control of the searcher—small changes in the query can lead to radically different result sets—how can we support active exploration of slices of the data to explore further aspects? Unlike traditional facetted search options, the result space is highly dynamic, how can we provide adaptive exploration options tailored to the context and searcher, at every stage of the process?

Evaluation How do we know the system is any good? How to evaluate the overall process, given its personalized and interactive nature? Can we rely on the direct evaluation of query suggestions and query recommendations? Are there suitable behavioral criteria for in the wild testing, such as longer queries, multiple filters, longer dwell-time, more active engagement, more structured-query templates? Can we use are standard experimental evaluation methods from HCI and UI/UX design?

Privacy Access to personal data is fraught ethical and privacy concerns, is there is similarly structured public data for scientific research? As an extreme form of personalization, how to avoid the uncanny cave, filter bubbles and echo chambers? How ethical is it to privilege a particular query refinement suggestion over the many other possible candidates?

3. CONCLUSIONS

The workshop brought together researchers from a range of areas in information access, who worked together on searching information from your personal point of view over a highly structured and curated information space. One of the main lines of discussion was the considerable industrial activity around social graphs. The most famous example is Facebook Graph Search, a feature that allows users to perform more sophisticated searches on their social network [1]. Bing has been integrating Facebook into their web search results for the last couple of years. Similarly, Google has been annotating search results with Google+ profiles. And all the rest of the search industry is moving in the same direction.

There are also crucial links with work on searching structured data, and work on the appropriate query languages, in particular as part of semantic search. These branches of research in particular focus on complex querying of structured text or data, whereas the graph search addresses also, and perhaps primarily, the process of constructing series of complex queries interactively. This is directly related to exploratory search and sense making. The graph structure provides natural facets for exploring the data, from a local point of view, allowing for a more dynamic structure than traditional faceted search using rigid, global, hierarchical structure. This challenges our understanding of search user interfaces design and evaluation, with search results moving from the found links, to the HIT page as snippets, and now to query suggestion as previews of possible query extensions.

Graph Search has fundamental consequences for information access and offers tremendous opportunities for building new systems and tools that allow users to explore information from many different angles, shifting control back to the user. This is a radical departure from current systems where the machine learning dominate the interaction: the entire information space is determined by the user, and the user is in the driver's seat when expressing her needs and exploring the space of options interactive.

Acknowledgments This research is funded in part by the Netherlands Organization for Scientific Research (ExPoSe project, NWO CI # 314.99.108; DiLiPaD project, NWO Digging into Data # 600.006.014).

References

[1] N. V. Spirin, J. He, M. Develin, K. G. Karahalios, and M. Boucher. People search within an online social network: Large scale analysis of facebook graph search query logs. In *CIKM'14*, pages 1009–1018. ACM, 2014. URL `http://doi.acm.org/10.1145/2661829.2661967`.