



UvA-DARE (Digital Academic Repository)

Graph search and beyond

Alonso, O.; Kamps, J.

Published in:
Dagstuhl Reports

DOI:
[10.4230/DagRep.3.10.92](https://doi.org/10.4230/DagRep.3.10.92)

[Link to publication](#)

License
CC BY

Citation for published version (APA):
Alonso, O., & Kamps, J. (2013). Graph search and beyond. *Dagstuhl Reports*, 3(10), 108-111.
<https://doi.org/10.4230/DagRep.3.10.92>

General rights


It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

4.3 Graph Search and Beyond

Omar Alonso (Microsoft Research – Mountain View, US), Jaap Kamps (University of Amsterdam, NL)

License  Creative Commons BY 3.0 Unported license
© Omar Alonso and Jaap Kamps

4.3.1 Motivation

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. 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 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. The best examples at the time of writing are Facebook Graph Search and related efforts at Bing, Google and other commercial search engines. Similar approaches 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.

4.3.2 Issues

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, requiring a highly interactive session covering both query formulation and result exploration.

Two Step Interaction

Incremental Structured Query Input: Creating a graph query requires incremental construction of a complex query using a variety of building blocks. Current search engines treat this as a form of query suggestion or query completion, which offers tailored suggestions trying to promote longer queries that cover multiple entity types and relations and various filters. Suggestions and entity types may be based on the user's own activity. This goes beyond prevailing autocompletion techniques, with previews and surrogates from traditional result pages or SERPs (Search Engine Results Page) moving to a more dynamic query suggestion.

Dynamic Structured Result Set Exploration: Results are highly personalized: they are unique for the searcher at a given point in time. The result set is highly structured: rather than just showing the top-10 results from an almost infinite list, a faceted exploration based on your interests is needed. The structure is dynamically derived from the graph structure and the user's point of view, rather than a rigid facet and facet value hierarchy.

When to Use Graph Search?

Rather than a universal solution, the graph search is particularly useful for specific types of information needs and queries. This is also depending on the character of the data available. E.g., Facebook Graph Search emphasizes the social network structure, friends and other persons, locations and location-tagged objects. Social network data is abundantly available (although getting access presents a major barrier) but also notoriously skewed. Rather the searcher personal point of view, it can also be used to show results from the viewpoint of any person in the network. There are many interesting sets of data – both historically or modern – that capture both the persons and related information: think of parliamentary data in public government, or intranet data in organizations.

Query Classification

Graph search also requires a new query classification, beyond the traditional division into navigational, informational, and transactional queries. Is there a new way to characterize queries in this new model? Does the notion of information need change? It is the ultimate form of personalization, with the searcher not only responsible for the query but also determining the (slice of) the data being considered. What shifts in control and transparency are needed to accomplish this?

Graph Search Evaluation

This also presents a range of new evaluation problems. How to evaluate the overall process, given its personalized and interactive nature? How to evaluate the first stage as essentially a form of query autocomplete? And how to evaluate the second stage as to explore and exploit the result set?

4.3.3 Methods

Graph search requires a highly interactive session covering both query formulation and result exploration.

Query Exploration

There is a radical shift towards the control of the searcher, necessitating new tools that help a searcher construct the appropriate graph search query, and actively suggest refinements or filters to better articulate their needs, or explore further aspects. This leads to a far more dynamic interaction than with traditional result lists, or modern hit lists showing summaries of a static set of results.

This suggests a new form of “query autocomplete” that invites and allows users to issue longer queries constructed based on entities, relationships, and templates. In contrast to SERP, we define IQEP as the Incremental Query Exploration Page. IQEP allows the user to explore more the result set as part of the input query. We can think of IQEP as an interactive mechanism that promotes relevant results selected by the user from the traditional SERP to the input box. Figure 1 shows IQEP as a bi-directional channel that moves results from the search list to the input box or viceversa.

There are a range of suitable evaluation methods. The obvious way is by direct evaluation of query suggestion, query recommendations (are they any good?). There is also a range of criteria useful for behavioral observation for in the wild testing: users should issue longer queries, multiple filters, dwell-time, active engagement, structured-query templates. There are query segments where this type of querying is expected to be most useful: torso and tail queries; exploratory scenarios. Traditional head or navigational queries seem less interesting, although these could be part of a more complex underlying information need.

This goes beyond Broder’s taxonomy: queries are all navigational, informational, and transactional but they are entity-focused. Queries may aim to return a single or a small set (not unlike traditional Boolean querying over structured data), or there is a need for data analytics on the whole set of results.

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 – necessitating active exploration of slices of the data to explore further aspects.

This suggest a new form of search results unique for every user. Similarly to the query exploration mechanism, this interaction encourages users to explore over entities, relationships, and filters. Unlike traditional faceted search options, the result space is highly dynamic, and requires adaptive exploration options tailored to the context and searcher, at every stage of the process.

This is a radical departure from the traditional “ten blue links”. The IQEP moves from links to answers, and from answers to suggesting (expressions of) needs. This is a complete shift from the traditional dichotomy between query (the searcher’s responsibility) and results (the system’s responsibility). Traditional search results have moved to a hit list of result summaries (still a fix set of results, but the shown summaries are tailored to the searcher and her query). These summaries in terms of entities are now answers rather than links to answers. Now these results, or previews of them, are moving into the search box, in the form of structured query suggestions with some sort of preview indicating of the consequences on the result set (often in terms of numbers of results, or entity previews).

There are many options for the evaluation of components: (adaptive) captioning, (adaptive) filters, graph query templates. E.g., captioning should describe (relative to the entity), explain (relative to the user), and be contrastive (relative to the IQEP). There are standard experimental evaluation methods from HCI and UI/UX design. With a running

service, evaluation in the wild is very suitable. There are various implicit and explicit criteria: users should explore the result set, usage of multiple filters, dwell-time, active engagement, structured-query templates. Torso and tail queries, and exploratory scenarios are the most suitable query segments.

4.3.4 Conclusions

Graph search gives amazing power, and unleashes the potential of semantically annotated information with many entities, and relations between entities. It brings the control back to the searcher. Graph-based search systems also have the potential to solve part of the old IR problem of conceptual search.

In terms of IR research and required evaluation methods, as discussed in the sections above, there are some open problems. What we need is to work on sharable research data, that exemplifies most of the characteristics we want to study. There is no need to be on Facebook or Twitter, or hand over your personal data. Similar small data sets and systems are available (e.g., so.cl, NYT, Parliamentary data, etc.) It will be hard to share a realistic subset of social network data (unless there are enough volunteers?) but we could work on a simulated set. What would be a concrete task to study on this data? Instead of implementing all features, it would be useful to select a few components like query suggestion box, filters as facets, and captions to show the potential.

Search engine user interfaces has been very stable in the last 15 years. The input box and the 10 blue links are still the most optimal way to show search results. Can we do better in terms of user experience? This would be give users a lot of flexibility and options. However, remains to be seen if users would adopt such dynamic interface.

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? Finally, there are a number of ethical issues such as privacy, transparency, bias and control, and filter bubbles.