INVITED PAPER

The Future of Enterprise Search

Dagobert Soergel
University of Buffalo, NY, USA
Email: dsoergel@buffalo.edu

Denisa Popescu
The World Bank, USA
Email: dpopescu@worldbankgroup.org

This paper provides some pointers on how to improve enterprise search going beyond one-size-fits-all ranking algorithms by adding knowledge-based decision making into all stages of search.

Enterprise search has two main goals:

1. Make sure that staff have available the information that supports best decisions, solutions, and task execution in the time available. This means just-in-time, just-what-is-needed information.
2. Significantly reduce the time staff spend on searching for (and often not finding) information, both by system-initiated queries that bring results without any staff effort and by more system help for submitting user-initiated queries coupled with assistance in processing search results.

Search in an organization must be tailored to support specific tasks of the organization's staff. It must integrate seamlessly with tasks and workflows. This makes it quite different from general search engines such as Google or Bing. Even so, most enterprise search systems use general search engines using one-size-fits-all ranking algorithms and result displays. But many types of questions come up in the work of an organization, each requiring special handling. Furthermore, enterprise search must deal with many types of objects and the links among them–documents, people, projects, organizational units, places, and more–and provide integrated, one-stop access to all of them, not splintered access through many different specialized systems, as is often the case. This editorial provides pointers on how to improve enterprise search by adding knowledge-based decision making into all stages of search.

Synopsis

Processing a query must go through the following steps (see Fig. 1)

1. **Query analysis**. This is the most important step, but often does not get sufficient attention. The search system must interpret queries as entered by the user or as initiated by the system at predetermined steps in a workflow, to determine the user's need / intent. This determination must also consider the user's context, which is often a complex web of specific circumstances. The determination of the user's need / intent forms the basis for formulating a search plan that specifies the collection(s) to be searched, the search algorithm to use, and the query transformed in many ways: marking off phrases, disambiguating query terms, query expansion, and more).
2. **Search plan execution**, which may involve (a) just finding a quick answer, or (b) finding the five best Web pages for a short query (for example, just of a country name, or a person name), or (3) relevancy scoring search, or a linked data search with inference.
3. **Display of search** results and helping the user with selection/ navigation exploration, and with making sense of the results

It is also necessary to address new modes of interaction that will make search easier and more effective.

Search integrated with workflow and daily work

In the daily work of staff, search should not be a separate activity but completely integrated with the tasks staffs perform. A workflow may initiate a query to search at a point where the task requires information. For example, during the review of a project funded by the World Bank, the reviewer must check that the project meets the requirements of the *World Bank Environmental and Social Framework*, The project review system should send a number of queries to enterprise search, all using the place of the project as a geographic filter and each addressing a different topic, such as

- What people live at or close to the project site. Find some documents about their way of living, their agriculture and other land use, their culture
- Are there any endangered species living close to the project? What is their habitat?
- Are there any cultural heritage sites near the project site?

With each topic include documentation provided by the Bank client applying for project funding. The project reviewer wood then gets an overview of the results

for each topic. Section 3 makes some suggestions for supporting users in processing such results.

To facilitate staff-initiated queries, every screen should have a search box. The user may or may not type something into the box; the system would suggest queries based on the content of the search box and an analysis of the user's context (location, user in a meeting, content of present and previous screens, search history, and more). The user selects a query and perhaps modifies it.

This integration allows for much better evaluation of search. The system has access to what the user does after he gets the search results. If the user is successful in his task, the search results were useful. The fewer clicks were needed to get the results, the better.

1. Query analysis. Interpret the query to determine users' need/intent and formulate a search plan

Determining user needs accurately is the most important step in getting to a good answer. For any enterprise one can discern a number of query types related to different tasks. Since query analysis is often neglected, we made it the focus of this editorial in order to drive home an important message.

Most readers when hearing the term *query* think of the ubiquitous search box, and some may think about more ways for users to interact with the system, such as selecting further search criteria through filters or using an interface for fielded search. But an enterprise search system worth its salt must do more. A user working on a task, such as assessing the budget for a project or evaluating bids for software acquisition should not have to enter a query for information that is obviously needed at a given point in the workflow; the workflow system should issue the query to enterprise search and present the search results to the user. Likewise, if a staff member is assigned a new task, the workflow system should submit a query to enterprise search so that the staff member gets the task along with a packet of information needed to complete the task.

When enterprise search receives a user-entered query, it must first determine the user's need and intent, which often can be seen from the query type. The examples given in Table 1 demonstrate the range of query types and the wide variety of knowledge needed to determine the query type and transform the query into a form required for a successful search. Knowledge required includes linguistic and taxonomic knowledge and knowledge of the context in which the query was submitted. Much of the knowledge about user context (such as whether a staff

member is scheduled to travel to Ghana, or has worked on Ghana, or is preparing to attend or is attending a meeting on whether to fund a project in Ghana) is available in the organization's data systems; it just needs to be looked up. Enterprise search may also maintain user profiles that give more permanent elements of the user context (but not elements that are highly time or situation dependent). Again, good understanding of the query is key to good search results.

The determination of the user's need / intent forms the basis for formulating a search plan that specifies

- The collection(s) to be searched (both internal and external)
- For each collection
    - the query transformed in many ways: marking off phrases, disambiguating query terms, query expansion, specifying that some terms occur in certain data field(s), and more;
    - the search algorithm to use.

Table 1 indicates for each example what to do – the search plan. Enterprise search needs an intelligent component that can figure out the search plan based on query and context.

2. Search plan execution

Search the collection(s) specified with the collection-specific transformed query using the collection-specific search algorithm specified. Search plans vary in complexity:

- Simple: Search the collection of quick answers, return the one or two that are indexed with a search term.
- Simple: For a query that consists just of a country name: In the collection of best bets find the five most important for the query.
- Complex: For a query such as *Factors influencing the price of bread in Syria* search all document collections with an expanded query using a relevance scoring algorithm.
- Complex: For a search such as *Find all staff who are or were team leads for projects with a budget of $10M or more* do a linked data search with inference spanning several collections.
- Complex: In a search for *candidates to fill a job* search people collections using a scoring algorithm borrowed from job matching sites, such as Monster.com or Indeed.com.

Analyzing queries and then choosing the most efficient way to find the answer(s) makes better use of computing resources. In most organizations there are many administrative queries and many one-word queries; such queries can be executed very quickly with simple queries run against small collections, leaving the heavy lifting of computing expected relevance scores over large collections or linked data search for more complex queries.

3. Display of search results and helping the user with selection and making sense of the results

For some searches a simple ranked list of results might be most useful, but there are many other possibilities that provide more support to users, depending on the situation.

One simple improvement is to put 25 results on a "page" since people are more likely to scroll than to go to a second page, and looking at just 10 results is often not optimal. This advice is especially important if each result has next to it a check box so that the user can run through the results quickly, check those she wants to look at more closely, and then request a list of just the checked results.

In dealing with complex problems or issues, search is just the beginning of a larger process of exploration, sensemaking, building an understanding, and further search arising from these efforts. Result displays need to support this process. We give just some examples.

- Arrange results in the order best read or visited.
- Group results by type (documents, people, projects, organizations, …), showing the top two in each group, with an *expand* button for each group.
- Show an overview of results, such as a network diagram showing concepts, documents, people, events, . . ., linked with relationships, with the possibility to navigate, zoom in and out, drill down. This allows for exploration. The diagram may show just concepts – a concept map.
- For results dealing with places, show a geographical map with circles of different sizes indicating the number of results dealing with that place. Map result displays can be integrated with GIS (Geographic Information Systems) that can provide additional data and many ways of analysis.
- Provide a sensemaking assistant that helps with organizing results, extracting information from results, summarizing (multi-document summaries) and building structures or schemas and ultimately a deep

understanding of the problem or situation at hand. This may include machine learning for discovering patterns in large data sets.

All of this requires creative approaches to the design of intuitive interfaces that show information succinctly for quick comprehension, give easy-to-grasp visualizations, and provide affordances that support the complex task of exploration.

4. New modes of interaction

There are many technologies on the horizon that can be used with search now or in a few years: Touch screens, eye tracking (to select an option or a piece of text, to let the system see what the user is interested in and look for more of that), graphical interfaces that become more powerful through use of these technologies (for a simple example, searching for place by selecting it on a map through tapping on the screen, searching for information on a department by selecting it on an organization chart), voice input and output. And not to forget mobile devices for simple searches. For complex searches a very large screen can be very helpful. (Main points from White 2016.)

CONCLUSION

Enterprise search done right can make an enormous contribution to the success of the organization. Done right means

- Enterprise search integrated in daily work.
- Queries initiated by systems on the user's behalf at specified points of a work flow.
- Careful analysis of tasks and the many types of queries arising from the tasks.
- Intelligent processing of queries:
    - Query analysis to determine the type of query and specify a search plan.
    - Search plan execution.
    - Display of search results and helping the user with selection/ navigation exploration, and with making sense of the results.

Getting there presents a challenge to organizations in general and to IT departments in particular:

- Enterprise search done right requires a new way of thinking about search, putting more emphasis on query analysis.

- The degree of integration across all applications presents technical difficulties and, perhaps harder to overcome, difficulties with stakeholders in the organization who want to control the ways their content is searched.
- Improving search requires high-quality data and adherence to data standards in all data systems in the organization, something that, like herding cats, is notoriously difficult to achieve.
- Machine learning will play a major role in increasing the role of IT systems to help reach an organizations goal, but machine learning is not a silver bullet on its own; it must be deployed as part of a an overall vision of an integrated intelligent system.
- As should have become clear from the amount of data and intelligent processing described in the examples, to achieve the below-one-second response time users have become accustomed to, enterprise search done right requires computing resources far beyond what is normally allocated for search.
- The organization must strive to understand and, to the extent possible, quantify the benefits of enterprise search done right both in terms of staff time saved and in terms better information received, resulting in better solution and improved collaboration across the organization.
- For enterprise search done right to happen the organizational leadership must fully engage and turn the mantra that *information and knowledge are among the most important assets of an organization* into action, fostering a climate of integration and allocating sufficient resources.

REFERENCES

White, R. W. (2016). *Interactions with search systems*. Cambridge University Press.

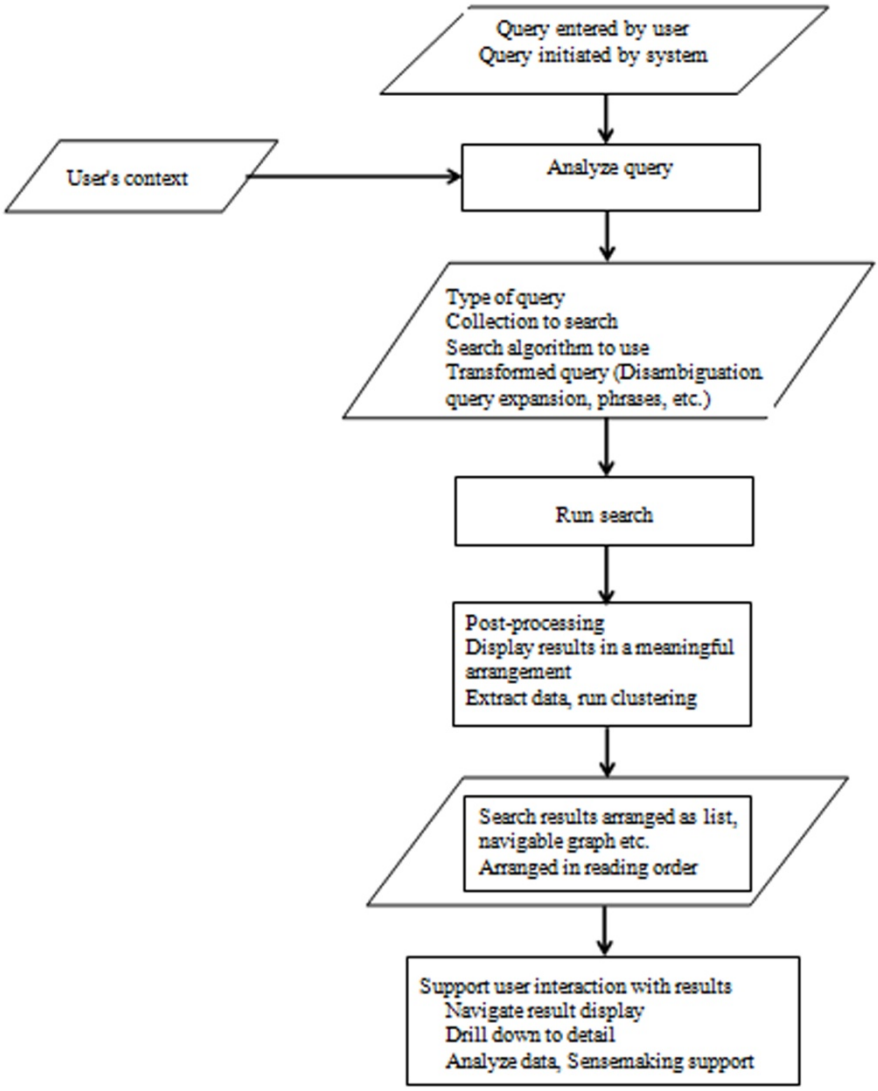

Figure 1. Processing a query in enterprise search

Table 1

Query interpretation and search plan generation

The table indicates what to do (the search plan) Enterprise search needs an intelligent component that can figure out the search plan based on query and context

User or workflow query	Query type and search plan
Travel phone number	The user needs the phone number of the organization's travel agency (This is a *simple administrative query*) 1. Search the collection of quick answers (a curated collection) Find the one quick answer by matching the query with one of the assigned keywords.
How do I create a Project/ Information Document? User context: Never did a PID before	Another *simple administrative query* Search the collection of policies, rules, and regulations (a curated collection that also includes forms and document templates with completed examples Map the query to the reference vocabulary used in this collection, then conduct a simple Boolean search.
How do I apply for early retirement? User context from data in the Human Resources system	A more *complex administrative query* Conduct a dialog with the user to determine through a series of questions the precise situation of the user (in addition of what is already known from HR (Human Resources Department) data) to find just the right answers.
Ghana User context: Upcoming travel to Ghana	*Administrative query* 1. Search the organization's *travel regulations and guidelines collection* with just *Ghana*. 2. Also search the Web for *Ghana travel warnings* 3. Also search the Web with the query *Ghana travel* and limit to the last 6 months , omit results from 2 Arrange results in three groups, one for each query

User or workflow query	Query type and search plan
Ghana User context: Has been assigned the task of evaluating a project located in Ghana. Has not worked on Ghana before.	One-word query. Needs general information about Ghana (This is an *knowledge query*) 1. Search the organization's *best bets collection* with just *Ghana*. The result is a list of internal and external URLs for Web pages such as A page that contains basic information on Ghana plus pointers to more A page that gives summary statistics on Ghana plus pointers to sources of statistical data. The home page of the World Bank Country Office. Ghana The home page for the World Bank Headquarters Unit. Ghana
Agriculture	One-word query. Needs general information about agriculture (This is an *knowledge query*) 1. Search the organization's *best bets collection* with just *agriculture*
PID Ghana	Query analysis finds that PID is an acronym with the most frequent expansion *Project Information document.* The user needs project information documents for projects in Ghana (This is a *knowledge query*) Search internal document collection Reformatted query: Document type filter: *Project Information document* Location filter: *Ghana* (documents must have *Ghana* in the metadata) Display message Searched for *Ghana Project Information document* Did you mean *Ghana "Public Improvement District" (PID)* [Query analysis selected *"Public Improvement District"* as the next most likely expansion of *PID* in the organization context.]

User or workflow query	Query type and search plan
Food supply Ghana	The user needs documents, statistical data, projects, experts on the topic (This is an *knowledge query*) Search the entire enterprise search database using the ranking algorithm. Could use query as is, but query analysis can do better using information from the organizations taxonomy: ("food supply" OR "food availability" OR "food price" OR "agricultur* productivity") Ghana Recall could be enhanced by hierarchic query expansion: food OR meat OR milk OR sorghum OR cassava … Ghana OR [list of provinces and major cities]
Person search General note applicable to all user contexts	Through autocomplete, search can detect from as few as three characters in the search box that the user is looking for a given person, giving preference to persons that are on the user's mind at the moment (upcoming meeting, person is in the user's circle, person appeared in recent email messages to or from the user, person is in the news, etc.) For all known person searches, search should show: • basic information, mostly contact information, link to a blog, on travel in from - to, thumbnail picture (this is like a quick answer, but compiled on the spot with most current data) • best bets for the person: Links to a profile page, the public HR page, the persons personal website, curriculum vitae • a box where the user can select categories of further information to be shown, such as documents the person authored, projects the person worked on, organizational units the person works with
Person search Need to call the person	Person search context several, including in charge of filling open position.

User or workflow query	Query type and search plan
Person search User context: Needs a profile of new boss or coworker just hired or of a person in another organization with whom the user will collaborate.	Search can determine this context from various sources. After basic information, show first a link to a profile page for the person. The profile page may need to be compiled or updated dynamically
Person search User context: User is looking to fill an open position The person searched for would be a reasonable candidate	After the basic information, show a box including the score indicating how well the person fits the position and the skills and prior experience that justify the score
Job opening search User context: A World Bank staff member wants to find a different job in the World Bank or the International Finance Corporation.	Collection to search: Job openings in the World Bank Group As query use the data about the searcher available in the HR system possibly amended by description of the position the user currently fills and other data about the user scraped from various sources on the Web plus any elements about future job aspirations formulated y the user Use algorithm for matching a people profile with position descriptions; such an algorithm returns a score, say 0 – 10. The higher the user's match score with a position the more likely it is that an application from the user for the job would be successful. Such an algorithm matches the applicant's skills with the skills required for the job and also assesses how well previous experience has prepared the applicant for the current position.
Search for courses, learning events, and learning materials	Assume the organization has a systematic approach for the continuing education of its staff. For each staff member HR maintains a training plan that specifies a time line of types and subjects of training events to be taken and/or learning goals to be achieved. The HR system uses a controlled vocabulary of concepts and skills important in the organization.

User or workflow query	Query type and search plan
User context: User wants to keep up to date with evolving job requirements and prepare herself for career advancement	With sufficient lead time, the HR system sends to enterprise search a query that identifies the staff member and specifies the training events or goals from the training plan. Enterprise search then selects several internal and external collections of training events and materials and for each of these collections transforms the query into a collection-specific query