Breakthrough Analysis: Two + Nine Types of Semantic Search

eScience Meeting Discussion
Eric Rozell
Search and Semantics

• “Both apply data-structuring techniques to make information more findable and usable. Join the two and you get semantic search, in essence, search made smarter, search that seeks to boost accuracy by taming ambiguity via an understanding of context.”
1. Related Searches / Queries

- Query expansion
  - “vegetables” => “carrots”

- Query correction
  - i.e., “Did you mean...”

- Query Recommendation
  - i.e., “You can also search for...”
2. Reference Results

• Google Knowledge Graph
• Bing Answers
• Search engine attempts at question answering
  – Bing calls it “intent”...
  – User intends to search for practical information (e.g., showtimes) instead of document hits
3. Semantically annotated results

- Not in the sense of rich results (e.g., RDFa, Schema.org, microdata tags)...
- Rather, it’s highlighting of terms semantically related to search query
4. Full Text Similarity Search

• Using full text documents to find similar documents using statistical / vector space model techniques

• Not sure where I heard it, but some tech support systems use this (transcribe your description of the problem to find similar problems and solutions)…
5. Search on Semantic / Syntactic Annotations

• Syntactic annotations...
  – E.g., “noun”, “verb”, etc.
  – For search engines, query modifiers:
    • E.g., site:wikipedia.org

• Semantic annotations
  – “<organization>center</organization>”
  – Only provide results for center where it is part of a phrase recognized as an organization with named entity recognition techniques
6. Concept Search

- “Ford films” => produce results for the term “movies”
- Taxonomy-based search
- Similar to technique 1.
7. Ontology-based Search

• Not only query expansion, but also expansion based on relationships...
  – E.g., “things that dogs chase” => results about “cats”, “cars”, and “tails”
8. Semantic Web Search

• Not clear what they were trying to capture here...

• Basically just discussing that the intent is for queryability of the “Web of data”
9. Faceted Search

• Exploratory interfaces based on (often) predefined categories for search
• Most of you are familiar, so I won’t dig into this too much...
10. Clustered Search

• “Like faceted search without predefined categories”
• Topics are extracted from the document corpus based on statistical similarities between documents (or, in some cases, semantic similarities)
11. Natural Language Search

• Think, IBM Watson

• Article asks:
  – “Noting that we're now habituated to two-to-three word searches, I wonder if [natural language search] ever will [catch on]”...
Impressions of the Article

• Overall, seems fairly dated
  – Many of the search engine examples have changed (e.g., Google no longer says, “Did you mean…”, they just fix the spelling error for you)

• It has good coverage of interesting applications you can build with semantics

• Has no regard to how you capture the semantics to begin with

• Also does not touch the limitations, or potential for noise in these kinds of applications
Why is this article useful?

• Being a Semantic Web shop, we’ve all encountered these kinds of use cases/applications at one time or another...

• It’s a handy reference to pull out for use case generation or idea generation for building applications on top of data