

When is Semantic Search Really Semantic Search?

Jake Spiegler & Thomas Burdick
Information Architecture and
Knowledge Management
Kent State University

The Research Goal

- Evaluating the effectiveness of a search system which claims to behave semantically in the context of a large organization's research division
- Challenge was that the search system was behaving in a way that *made no sense* to the searchers or to the search system engineers
- Goal was to identify improve the search experience for users by improving the behavior of the search system
- This meant leveraging more of its capabilities but also identifying capabilities which were not yet ready for prime time and should perhaps be turned off

Research Methodology

- Two-phased research methodology
- Phase 1: Evaluation of the Current System
 - Reviewed search logs to review general behavior patterns
 - Drew testing terms from recorded search logs to try to learn what kind of semantics were being used – most of the semantics targeted query
 - Selected set of search terms and watched the behavior in targeted search experiences
 - Collected document data for each term
 - Conducted usability study on what users thought about the search
- Phase 2: Implementing Improvements and Re-Evaluate
 - Throughout Phase 1 potential improvements were identified
 - Team learned a lot about what was semantic and what was not, what was working well and what was not

What Was Semantic

- Search engine exhibited some semantic aspects
 - Full text similarity search with multiple matching points of text as query and indexed data points
 - Leveraged no metadata
 - Did not index keywords or tags
 - Relevance calculation not understandable at any level

Observations

- Users were approaching search with a less than semantic mindset. Not used to approaching search in this way.
- Search results were not organized in the way a user would rank them.
- Search system couldn't derive context from excel documents
- Team learned a lot about what the vendor meant by semantics and how they had been integrated into the design
- A lot of the semantic design, though, is built deep into the design – appears not to be configurable - this was an important observation because it may influence the organization's decision to invest in semantics in the future

On-Going Research

- Should we expect a semantic search engine to leverage semantics such as metadata, surrogates, user generated tags or can a semantic search system still be focused largely on full text?
- Should we expect a semantic search system to understand and adapt to different formats? Does it approach indexing an excel spreadsheet the same way it approaches indexing a powerpoint or text documents?
- How do we train users to approaching the search with a semantic mindset? Or, is the issue that users have an old mindset about search that they need to unlearn?