Eleven+ Views of Semantic Search

Denise A. D. Bedford, Ph.d.
Goodyear Professor of Knowledge Management
Information Architecture and Knowledge Management
Kent State University

Presentation Focus

- Long-Term Vision of Semantic Search
- Semantic Search as a Radical Transformation
- Semantic Search as Incremental Improvement
- Status of Semantic Search R&D

1. Long-Term Vision of Semantic Search

Semantic Search?

- Two very big concepts – a term we find in the published literature, on the web, in conference presentations, in product descriptions and in general conversations.

- What does that mean? What does it look like? What does it feel like? How does it work? And what is the value proposition for semantic search?

- Trying to define semantic search assumes that someone knows what it is – I don't think that is the case. I think instead we need to build a vision of semantic search

- The vision also needs to be long-term and transformative, to incorporate all the work that is currently underway and a lot of work that isn't any where yet

Two Ways to Characterize Semantic Search

- Two fundamental ways in which we view semantic search today
- One is a radical transformation of search brought about by disruptive technologies and a radically different future semantic environment (Mills Davis and the Semantic Technology Folks)
- The second view is an incremental transformation of our current search environment with the addition of targeted semantics. (Seth Grimes and the Information Science Folks)
- These two views are not contradictory – rather they represent a continuum of work on semantic search

2. Semantic Search as Radical Transformation

Search in a Semantically Rich World

- The radical transformation perspective is promoted by researchers and visionaries in the field of semantic technologies
- In this context, search is an integrated and embedded part of a larger semantic world where it is supported by a wealth of other semantic applications.
- Mills Davis has identified 16 areas which are essential to the development of a future semantic landscape.
- Only one of these areas is focused on search – but it is not explicitly search. Rather it is focused on the human activity that is enabled by search – Knowing

Davis' Characterization of Semantic Landscape

- Semantic User Experience
- Semantic Content Tools
- Semantic Social Computing
- Semantic Collaboration
- Semantic Software Development
- Semantic Architecture
- Semantic Ecosystem
- Semantic Enterprise Processes
- Semantic Transport
- Semantic Processors
- Semantic Storage
- Automated Reasoning
- Intelligent Systems
- Knowledge-Based Applications
- Knowledge Representation
- **From Search to Knowing**

Each of these areas in the landscape has several subdomains each of which represents a distinct and rich focus of research and development.

Radical Transformation of Search

- We can place Davis semantic search vision on the transformative end of the scale – a very different vision of search than what we have today – because they integrate multiple semantic components into a semantic architecture

- Discovery, Analysis (Watson and Wolfram Alpha), Intelligent Search, Personalized Semantic Search, Semantic System Dialogs

- From Davis' perspective adding semantics changes the basic nature of search – it redefines the interactions between the computer and people, and it makes search an enabler than an end goal.

- Semantic search may ultimately be invisible to us as we take on more complex interactions with the "system."

Future Vision of Semantic Search

- Davis' landscape provides a context for bringing together the work that is currently underway in many disciplines

- Some of this work is underway in knowledge sciences, some in information science, some in artificial intelligence, user experience design, in computer science and computational linguistics.

- The literature – like the work – is currently scattered. This future vision of a semantic landscape provides an opportunity to bring all of this work on semantics into a common view.

3. Incremental Improvement of Current Search Using Semantics

Improving Upon a Common Foundation

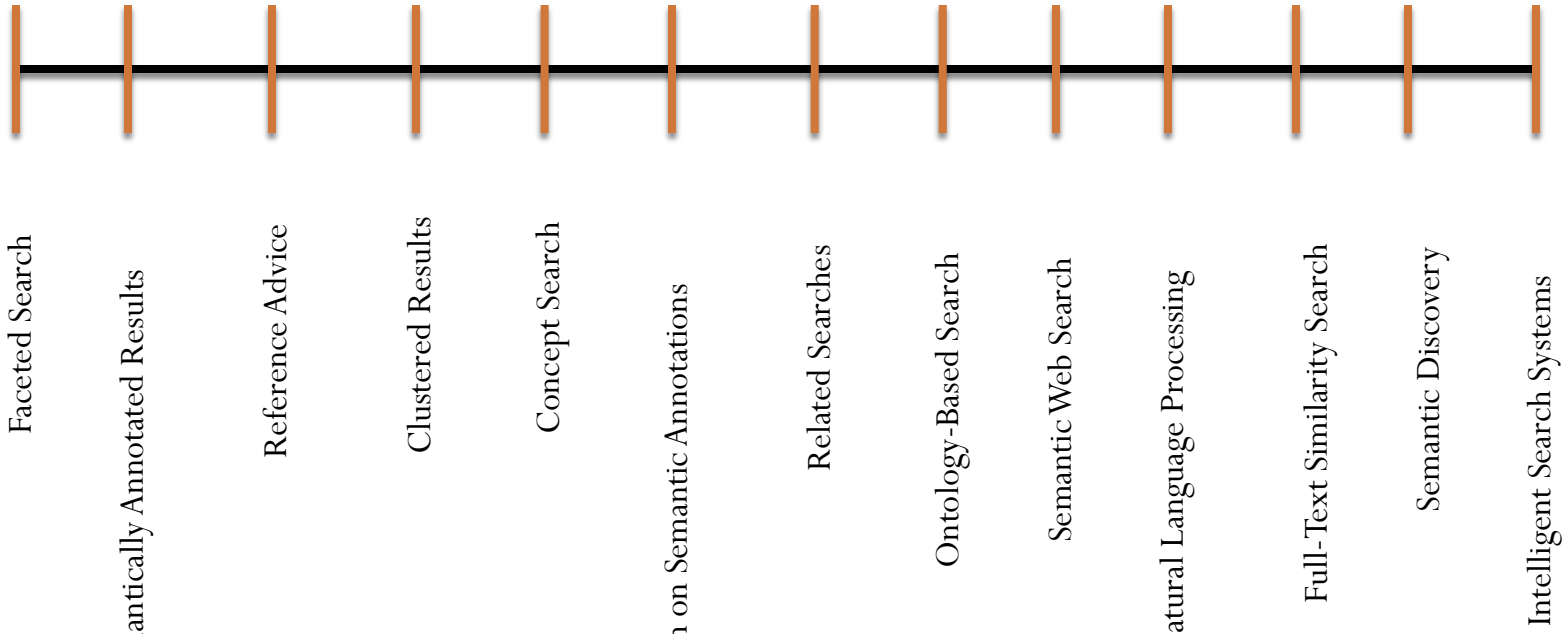
Seth Grimes Characterization of Semantic Search

- Seth Grimes has given us a wonderful characterization of the current work in semantic search http://www.informationweek.com/software/business-intelligence/breakthrough-analysis-two-nine-types-of/222400100
- Grimes has identified eleven areas in which R&D is underway
- I suggest that these eleven areas represent the incremental improvement to semantic search rather than the transformational view
- From the incremental perspective, search remains conceptually the same, but parts of search are semantically augmented – we use the old architecture, not the radically different architecture envisioned by Davis.

Enhancing Search With Simple to Complex Semantics

Current Search Architectures

Search in a Semantic Future



R&D Focused on Query Management

Type of Semantic Search Application	Description
Reference Advice	Results include materials that may provide further information about the terms used in the query.
Search on Semantic or Syntactic Annotations	Query processing approach which either implicitly or explicitly semantically tags query terms to improve the context in which the query terms are searched.
Concept Search	Search which expands the query terms to include semantically related concepts. The expansion can be done explicitly or implicitly.
Ontology-Based Search	Search system assigns semantic meaning to the query terms and associates the terms with other terms that may be related in other contexts.
Semantic Web Search	Searching complex relationships that will be available in the future web of data.
Faceted Search	Otherwise known as parametric or fielded search. Search is enabled for more than one dimension – multiple facets or parameters that enable a search to more precisely define their search needs.
Natural Language Search	Search query is semantically processed and tagged for more effective matching against other objects in the search system index.

R&D Focused on Results Management

Type of Semantic Search Application	Description
Full-Text Similarity Search	Search uses a submitted block of text or a full document to identify other results which may be similar. Similarity is generally determined based on statistical or vector-space similarity measures. There is typically no —semantic meaning associated with the similarity ratings,
Related Searches and Queries	Search suggestions that highlight objects that are similar in some way to the query terms. This can either be explicit or implicit suggestion.
Semantically Annotated Results	Search results have highlights for terms in the documents that are semantically-related to the search query.
Clustered Search	Search results are statistically clustered into categories to help the searcher more effectively navigate the results sets.

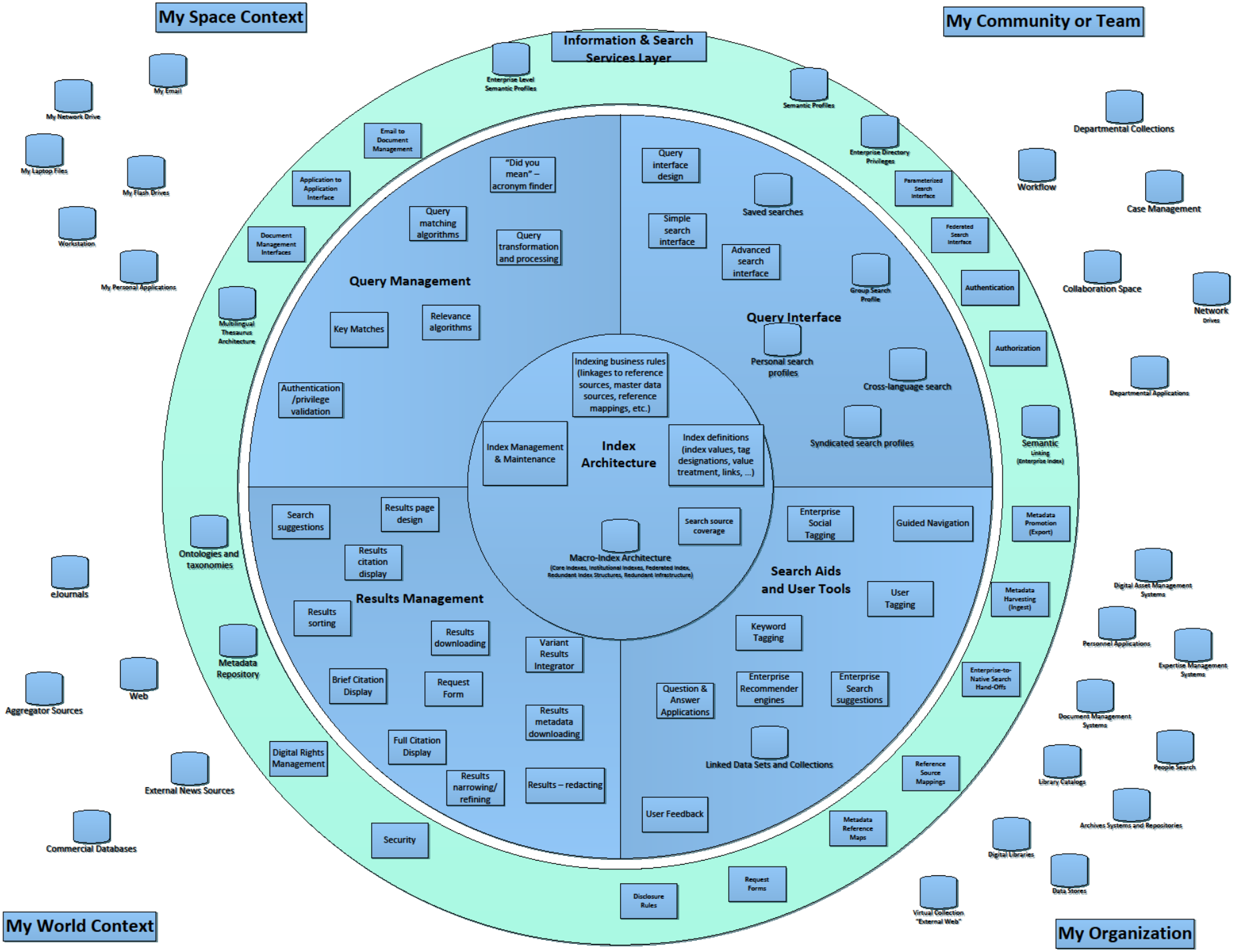
Incremental Improvement of Search Using Semantics

- When we think of the radical transformation we see search in a whole new context. When we think of incremental improvement to search we understand semantics being applied to every aspect of a search system

- By system here we are not referring to a tool or technology but to inputs, processes, outputs – the macro- and the micro-views of search as a system

- Let's overlay what Grimes describes as the current work in semantic search with a conceptual model of a search system

- This may help us to see how the current work aligns or needs to be aligned, where we have gaps and where there are opportunities for future work


My Space Context
My Community or Team
Information & Search Services Layer
My Email
My Network Drive
My Laptop Files
My Flash Drives
Workstation
My Personal Applications
Enterprise Level Semantic Profiles
Semantic Profiles
Enterprise Directory Privileges
Departmental Collections
Workflow
Case Management
Collaboration Space
Network Drives
Authentication
Authorization
Departmental Applications
Semantic Linking (Enterprise Index)
Metadata Promotion (Export)
Digital Asset Management Systems
Metadata Harvesting (Ingest)
Personnel Applications
Expertise Management Systems
Enterprise-to-Native Search Hand-Offs
Document Management Systems
People Search
Library Catalogs
Archives Systems and Repositories
Digital Libraries
Data Stores
Virtual Collection "External Web"
My Organization
My World Context
Commercial Databases
External News Sources
Aggregator Sources
Web
Metadata Repository
eJournals
Ontologies and taxonomies
Multilingual Thesaurus Architecture
Email to Document Management
Application to Application Interface
Document Management Interfaces
"Did you mean" – acronym finder
Query matching algorithms
Query transformation and processing
Query Management
Key Matches
Relevance algorithms
Authentication /privilege validation
Query interface design
Simple search interface
Saved searches
Advanced search interface
Parameterized Search Interface
Federated Search Interface
Group Search Profile
Query Interface
Personal search profiles
Cross-language search
Syndicated search profiles
Indexing business rules (linkages to reference sources, master data sources, reference mappings, etc.)
Index Management & Maintenance
Index Architecture
Index definitions (index values, tag designations, value treatment, links, ...)
Macro-Index Architecture
(Core Indexes, Institutional Indexes, Federated Index, Redundant Index Structures, Redundant Infrastructure)
Search source coverage
Search suggestions
Results page design
Results citation display
Results Management
Results sorting
Results downloading
Variant Results Integrator
Brief Citation Display
Request Form
Results metadata downloading
Full Citation Display
Digital Rights Management
Results narrowing/ refining
Results – redacting
Security
Enterprise Social Tagging
Guided Navigation
Search Aids and User Tools
User Tagging
Keyword Tagging
Question & Answer Applications
Enterprise Recommender engines
Enterprise Search suggestions
Linked Data Sets and Collections
Reference Source Mappings
User Feedback
Metadata Reference Maps
Request Forms
Disclosure Rules

Micro-Level View of Search System

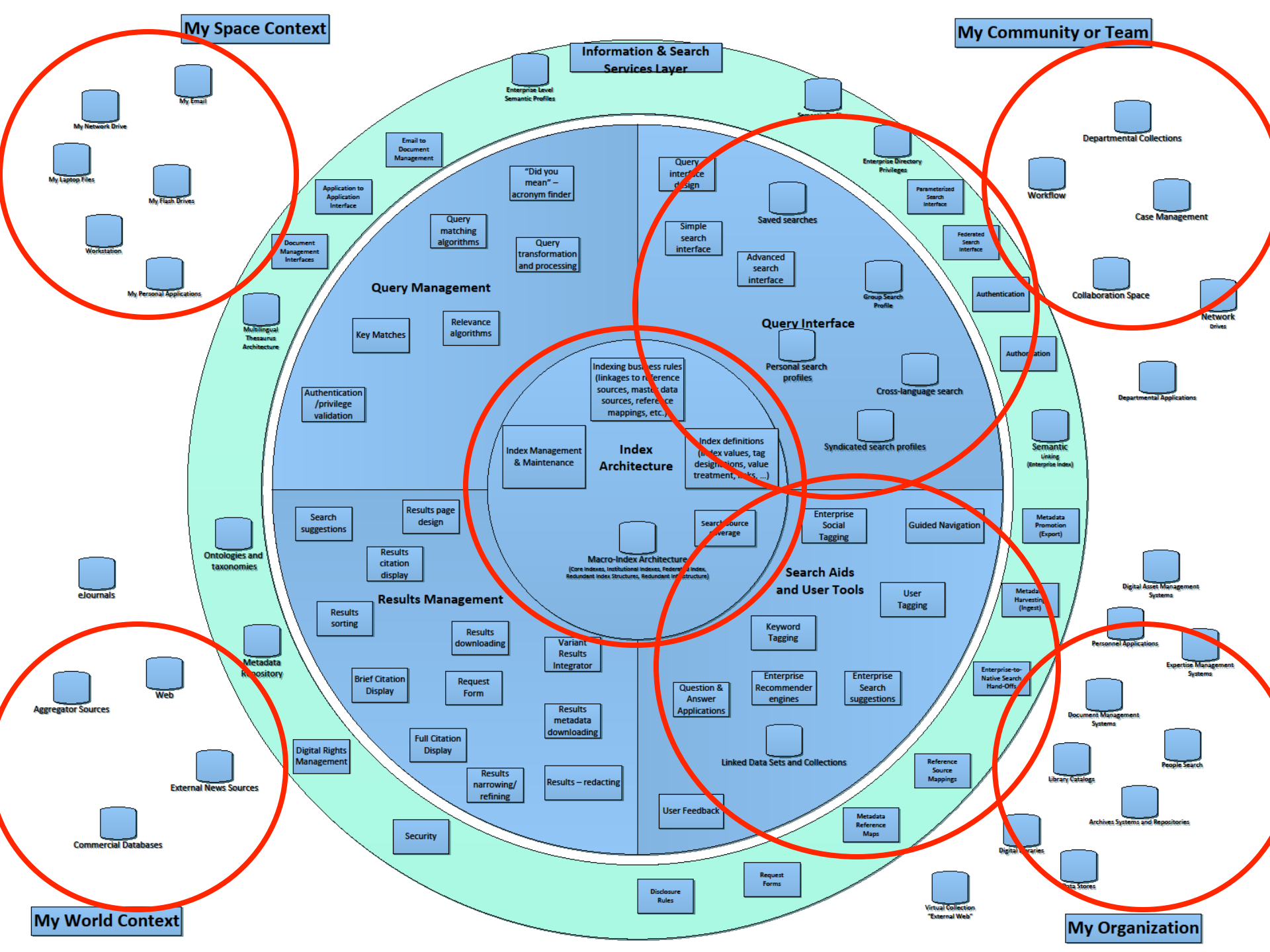
Index	The index is the heart of any search system. Index architecture includes indexing business rules, index definitions, search source coverage, and index management and maintenance.
Query Management	Query Management includes query transformation and enhancement, query matching algorithms, relevance calculations, key matches, dictionary look ups, misspelling management, and so on.
Results Management	Results management includes search suggestions, results scoring, results citation displays, results refining, sorting, results download authorization and authentication, and variant results management.
Query Interface	Query interface includes general design and presentation of functionality, personalized interfaces, personalized search profiles, cross-language search, and saved searches.
Search Aids and User Tools	This space includes social tagging, guided navigation structures, end user tagging, keyword tagging, recommender engines, search suggestions, question and answer applications, results linking and presentation (versions, editions, languages) and “others like this.” Also accommodates linked data that is created by end users.

Macro-Level View of Search System

- The macro-level of search focuses on the context in which it search takes place - search in a personal context, in a community or team space, in an organizational space, and in the larger universe

- The macro-level begins to describe Mills Davis' semantic landscape but on a much narrower scale - we would fold all other 15 levels into each of those four quadrants

- Making the macro-level semantic means ensuring both the content and the context are semantically enabled and semantically aware.
  - Being semantically aware means understanding something about the user not just what the user communicates explicitly in search.
  - Being semantically enabled means the content can tell the search system what it needs to know to make a good decision


My Space Context
My Email
My Network Drive
My Laptop Files
My Flash Drives
Workstation
My Personal Applications
My Community or Team
Departmental Collections
Workflow
Case Management
Collaboration Space
Network Drives
Information & Search Services Layer
Enterprise Level Semantic Profiles
Email to Document Management
Application to Application Interface
Document Management Interfaces
"Did you mean" – acronym finder
Query matching algorithms
Query transformation and processing
Query Management
Key Matches
Relevance algorithms
Multilingual Thesaurus Architecture
Authentication /privilege validation
Query interface design
Enterprise Directory Privileges
Parameterized Search Interface
Saved searches
Simple search interface
Federated Search Interface
Advanced search interface
Group Search Profile
Authentication
Query Interface
Personal search profiles
Cross-language search
Syndicated search profiles
Authorization
Departmental Applications
Semantic Linking (Enterprise Index)
Indexing business rules (linkages to reference sources, master data sources, reference mappings, etc.)
Index Management & Maintenance
Index Architecture
Index definitions (index values, tag designations, value treatment, ...)
Macro-Index Architecture
(Core Indexes, Institutional Indexes, Federated Index, Redundant Index Structures, Redundant Infrastructure)
Search suggestions
Results page design
Results citation display
Search Source Coverage
Enterprise Social Tagging
Guided Navigation
Metadata Promotion (Export)
Ontologies and taxonomies
eJournals
Results Management
Results sorting
Results downloading
Variant Results Integrator
Brief Citation Display
Request Form
Results metadata downloading
Full Citation Display
Digital Rights Management
Results narrowing/ refining
Results – redacting
Security
Search Aids and User Tools
User Tagging
Keyword Tagging
Question & Answer Applications
Enterprise Recommender engines
Enterprise Search suggestions
Linked Data Sets and Collections
Reference Source Mappings
User Feedback
Metadata Reference Maps
Disclosure Rules
Request Forms
Virtual Collection "External Web"
Metadata Harvesting (Ingest)
Enterprise-to-Native Search Hand-Offs
Digital Asset Management Systems
Personnel Applications
Expertise Management Systems
Document Management Systems
People Search
Library Catalogs
Archives Systems and Repositories
Digital Libraries
Data Stores
Metadata Repository
Aggregator Sources
Web
External News Sources
Commercial Databases
My World Context
My Organization

4. Current State of R&D in Semantic Search

Apparent and Real Gaps

- Grimes characterization from the perspective of search and the information science domain is a good one in my opinion
- If we accept this characterization, it looks like we have a lot of gaps and many opportunities for future development
- I checked my opinion, though, against the literature and found that we find a much richer R&D environment when we broaden our view to include:
  - (1) the whole search system -- micro- and macro- levels
  - (2) Mills Davis' characterization of the semantic landscape

Volume and Nature of Semantic Search Work

- I found the literature scattered across many domains, many different publications and conferences – I even experienced something of this sort on a personal level when I attended a Wolfram Alpha conference
- After four weeks, I am still going through all of the articles and reports I gathered - I will include a summary in the paper that will be openly published to the NKOS Workshop website
- While some of the work is taking place in the United States, an equal amount is underway in other regions of the world
- The work is taking place in universities, in research institutes, in vendor and venture capital contexts - not all of the work may be openly available

Open Questions and Discussion

- In my humble opinion, we didn't really do a good job understanding search architectures and system designs when search was simple -- we left this to the vendors to design and we trusted what they told us

- Only a handful of universities and information science programs teach search system design and architectures to information science students

- We definitely need a good grounding in search if we have any expectation of constructing a full-function seamlessly integrated semantic search

- Can you envision a semantic search system which includes all those features that Grimes' referenced? Are we still going to expect the user to do the integration for us?

Open Questions for Discussion

- Semantic search is going to be human knowledge intensive – it will involve major investments of people and brainpower to build and to sustain

- It may also be intensive for users to use if it is not designed well – instead of giving users more power, we may be adding to their burden

- What kinds of skills do users need to "search semantically"?

- What kinds of skills do organizations need to have to support semantically enhanced search?

Open Questions for Discussion

- Is the path of incremental development to semantic search likely to lead us to that transformational view that Mills Davis describes?

- Will we have a patchwork semantic search rather than a well designed and integrated capability?

- Having a vision of what semantic search will be is critical to ensuring that we achieve that transformational view.

Thank you!

dbedfor3@kent.edu