Seven Arguments for Semantic Technologies

by Mike Bergman - Monday, February 25, 2013

http://www.mkbergman.com/1626/seven-arguments-for-semantic-technologies/

A popular post on this blog has been the Seven Pillars of the Open Semantic Enterprise. That article described the building blocks – or foundations – to a semantic enterprise, at least from my own perspective. But it has always felt that the reason why anyone should even be interested in this semantic enterprise business deserved its own discussion.

This current article riffs off of that earlier blog to provide the seven rationales or arguments for why pursuing a semantic enterprise makes sense, especially in contrast to conventional or traditional approaches. This riff extends to even re-presenting the seven-spoke wheel from that original article:
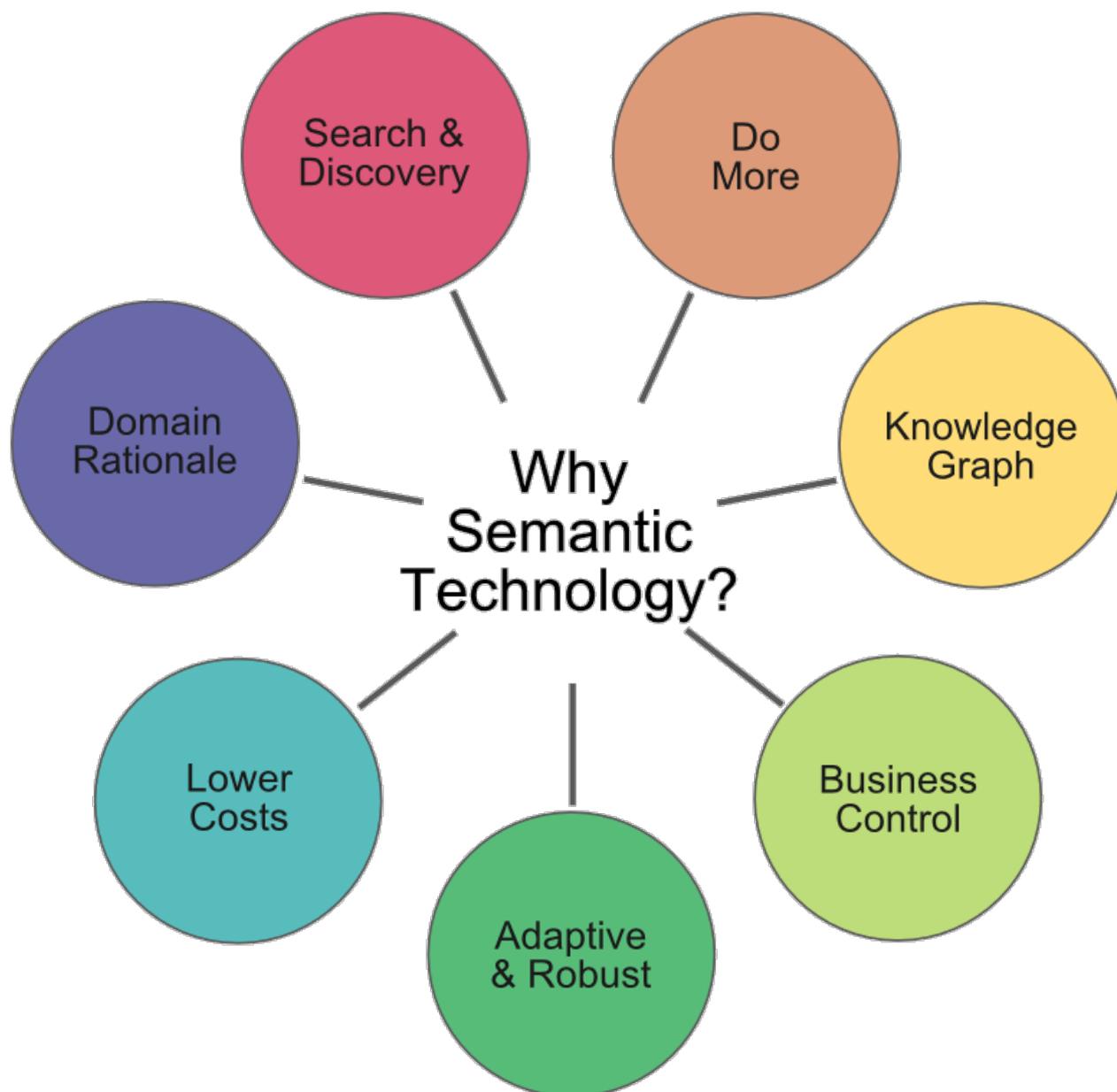
Each of these bubbles deserves some discussion.

Search & Discover

The first advantage with semantic technologies is that all kinds of information are unified. No matter what information you consider, any content type may become a 'first-class citizen'. For really the first time, all kinds of information ranging from traditional databases and spreadsheets ("structured"), to markup, Web pages, XML and data messages ("semi-structured"), and then on to documents and text ("unstructured") or multimedia (via metadata) can be put on a level playing field [1].

These data, now all treated on an equal footing, can be searched and retrieved by a variety of techniques.

These range from SQL, standard text search, or SPARQL, depending on content type. This unique combination enables all of the aspects of findability – find, discover, navigate – to be fulfilled. Because of the diversity of search options available, search results can be varied and "dialed in" depending upon circumstance and needs.
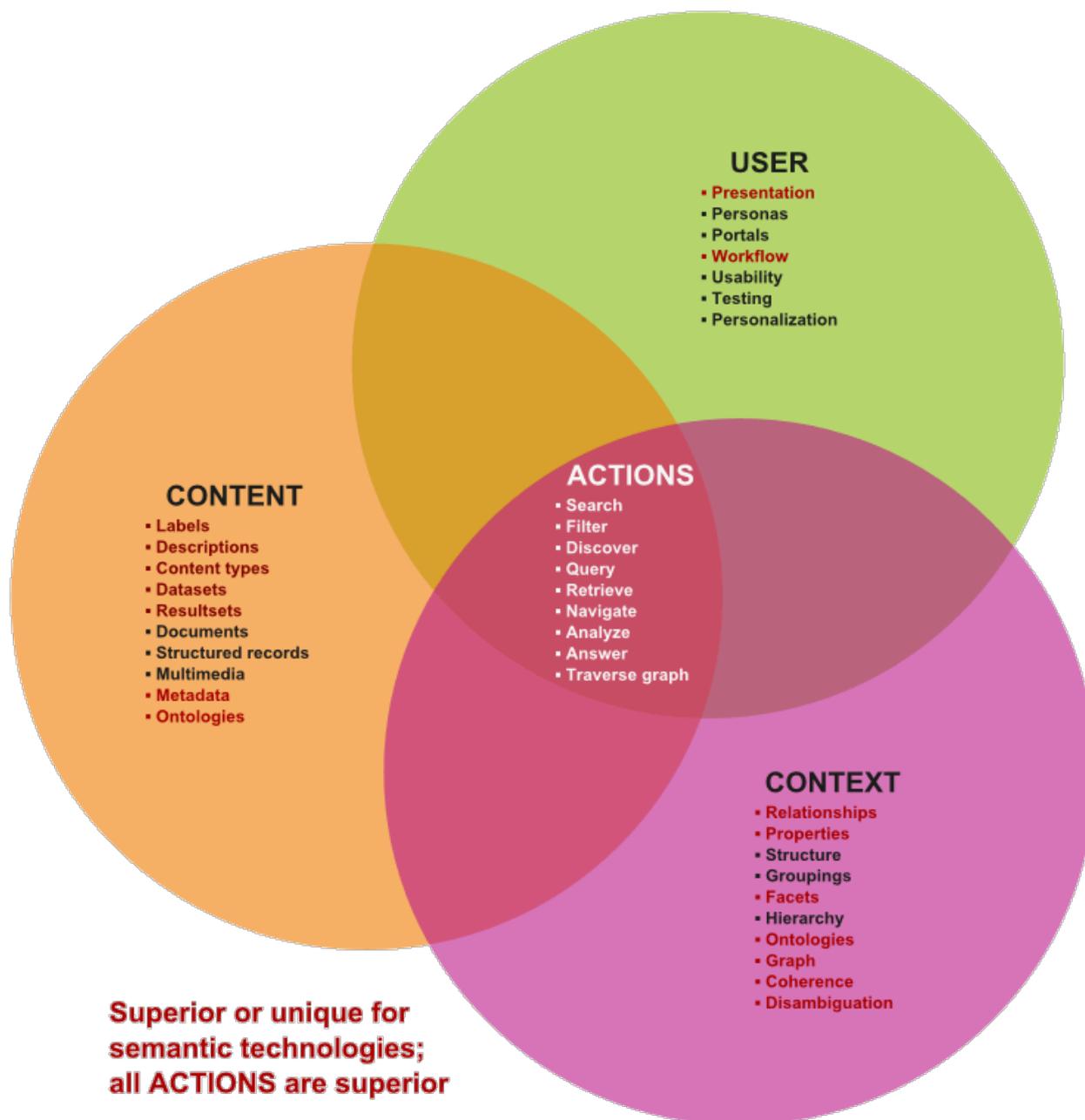
Because all content is represented either as a type of thing ("node" or noun) or the relationships between those things ("predicate", "property", "attribute" or characteristic or verb), any and all of those classifiers may be used for faceting or grouping. Further, the relationships put all things in context, useful to ensure results are relevant and disambiguated.

In all cases, these ways of describing things and their relationships to one another are based on the "idea of the thing" and are not bound or restricted to keywords. This means that all the various ways that things can be described – alternative terms, synonyms, acronyms or jargon – including in multiple languages, can be used to find or match these ideas or concepts.

When combined with the ability to infer relationships between things – even if not explicitly asserted – search and discovery for semantic technologies literally blows away any and all alternative approaches to search.

Do More

The classic information architecture (IA) diagram relates *users* to *content* and *context*. It is at the nexus of these ideas that actions and actionable information occurs:

USER
- Presentation
- Personas
- Portals
- Workflow
- Usability
- Testing
- Personalization

CONTENT
- Labels
- Descriptions
- Content types
- Datasets
- Resultsets
- Documents
- Structured records
- Multimedia
- Metadata
- Ontologies

ACTIONS
- Search
- Filter
- Discover
- Query
- Retrieve
- Navigate
- Analyze
- Answer
- Traverse graph

CONTEXT
- Relationships
- Properties
- Structure
- Groupings
- Facets
- Hierarchy
- Ontologies
- Graph
- Coherence
- Disambiguation

Superior or unique for semantic technologies; all ACTIONS are superior

Semantic technologies are superior in terms of the ability to capture all forms of *content* (structured, semi-structured and unstructured) as first-class citizens and to represent it through knowledge graphs (ontologies). Further, the ability to describe this content with multiple labels, languages and descriptors means the "idea of things" is much better captured than via keywords alone.

The explicit accounting of relationships between things with semantic technologies means the ability to capture better *context*, important for navigation and the reduction of ambiguity. The richness of relationships also means that the way things relate to one another can also be used for grouping, classifying, filtering or finding things.

Users can be better characterized and related to this content and its contexts because of this ability to match metadata to things and relationships. This leads to richer user experiences and the separation of content from presentation, giving the content more power.

In all cases, because of these basic information architecture advantages, the ***actions*** that can be taken upon content are far superior in comparison to any alternatives. Unique actions brought by semantic technologies include analysis, graph traversals and answering systems.

But, the ability to do more extends beyond content and context.

The ability of semantic technologies – specifically the RDF data model – to represent all content forms and any possible schema means that any existing content can be represented through a single representation. This makes RDF a form of "universal solvent" in which any content form can be distilled. This has huge implications and advantages.

One advantage is in providing data interoperability. The RDF data model enables any content form or schema to be represented, and the fact that the meanings of things can be mapped to agreed concepts and relationships also means that disparate information sources can be adequately related to one another. The fact that all data has a unique URI identifier means that any information accessible via HTTP can be included. This model with its marriage to ontology graphs leads to an excellent framework for interoperating data.

This same robustness for mapping different data leads to a second advantage, namely, semantic annotation. Concepts may be matched through so-called ontology-based information extraction (OBIE) and entities or things may be matched with named entity recognition (NER). The tags that result from these recognitions can then be placed back into documents via what is known as semantic injection. The result of all of these activities is that the very same content can now be equivalently understood by machines or humans.

The graph orientation of the system and its logic next means that the information structure is computable with unique analysis capabilities. The relationships between things can be understood and inferred, and the graph structures themselves may lead to unique traversal mechanisms and network analysis such as influence, clusters, neighborhoods, connectedness and so forth. No other information structure provides these unique advantages.

The graph structure also means that finding and relating stuff only need access a single index, after which relations can be traced and computed. Conventional relational database systems require joins and multiple index lookups to even approximate a portion of this ability, which can quickly run out of steam with complex requests or queries. (Also, recall that RDBMs can not accommodate the content or schema flexibilities that semantic technologies can, either.)

These advantages – combined with some of the other advantages discussed in next sections – also enable semantic publishing using these technologies. Semantic publishing offers new ways to let data and its characteristics drive how information gets presented.

The Knowledge Graph

Besides the RDF data model, the other pivotal aspect of semantic technologies is the knowledge graph, the ontology that captures the logical schema of the problem domain at hand. The knowledge graph is based on logic and is built from simple statements, or assertions.

The so-called "triples" that are these basic assertion statements in the semantic model are like a sentence of *subject – predicate – object*. The object of one statement can be the subject of another. In this manner, these "triple" barbells get connected together, growing in a graph-like structure as more statements are added. These basic building blocks are easy to understand and easy to correct if problems are found. Because each node (a subject or object) is the "idea of a thing" and not limited by individual labels or language, each of these things can be described in multiple ways with multiple terms or synonyms. Different people using different language to describe the same thing can thus communicate. Further, how these things relate to one another can be as diverse as how things relate to one another in the real world. The knowledge graph is phenomenally capable of describing the relevant world at hand.

All of these components themselves are based on basic first-order logic, which means these graph structures can be reasoned over, including being able to infer what is not strictly asserted, and to test that the assertions that are made make logical sense. This logical sense is what we term *coherence*. Because of this logical structure, and because of its graph nature, semantic technologies offer unique ways to find things and to analyze them. Increasingly over time we will see graph analysis become a more important aspect of how we analyze and solve problems.

Business in Control

How all of this affects the business is fundamental. Because the characterization of data and the structure of how it interacts together – and because the basic nature of these structures is relatively easy to understand – semantic technologies bring a tectonic shift to the enterprise. Control of how it works now shifts to those who need to consume and manage that information; that is, knowledge workers, managers, and subject-matter experts.

This content is separated from the presentation and the applications that use it. Since so much information is contained in the structure and relationships of the content, these patterns can inform how to present and use the information. For example, the fact that some information contains geo-locational attributes means that it can be mapped; or, we can know that cameras are a kind of device or product. This embedded knowledge can be used to inform how generic applications need to respond and display the info.

Thus, we see that the nexus of control around knowledge management can now shift to those who need and consume that information. The role of information technology moves to the background to provide the infrastructures and tools that can be driven from these information structures. We term these types of applications, "ontology-driven applications", or ODapps.

We are only now at the very beginnings of this transition. ODapp tools are still few and not mature, and few organizations have even made the cultural transition to shift this locus of control. But, embracing semantic technologies and its innate power to bring information management directly into the hands of those who need it will definitely disrupt the enterprise.

An Adaptive and Robust Fit

Because these structures and the data model behind them are a natural fit with the nature of information, semantic technologies prove to be both adaptive and robust. The data model is easily extended and modified without affecting the schemas already in place, a circumstance of having an open world logic. An abiding constant of relational technology – the basis for enterprise IT systems over the past few decades – has been its rigidity and difficulty in changing its structure or organization. Such a framework is perhaps the best for transactional systems, but is a poor choice for knowledge systems where the amount of content and its relationships are constantly changing.

A huge lever arising from these underlying semantic technologies is the ability to integrate across the different characteristics of information – its syntax, its structure, and its meaning. Units of measure, or different languages, or different ways to describe the same thing can all be boiled down to a common representation.

As the attention shifts to how we describe our domains and its concepts and instances, we can segregate off the questions of how we present and interact with that information. That means our human-computer interfaces can become more effective. It means that HCI itself can focus more on the channel or device. We are seeing it now with widgets and mobile apps, but our information will increasingly be presented through known interactions no matter what device we use. Semantic technologies are a natural and superior means for this adaptivity.

Much Reduced Costs

All of these benefits in productivity, responsiveness, and adaptivity translate into much reduced costs. These reductions come both in lower set-up and deployment costs and in lower maintenance and scope costs. Experience is that these functions can be undertaken on average at lower costs one or two orders of magnitude less with semantic technologies than with traditional approaches.

These reductions come about because we can leverage our existing information stores and schema into a single, "canonical" representation against which we can tool and present. The fact that new sources can be integrated into the system without re-architecting what already exists is another huge win, and a matter that almost always overruns budgets with conventional approaches.

An area little documented is the high cost of errors. It is ubiquitous in our current information systems. But, it is a hidden and huge cost. Because semantic technologies can help in putting information into context, can help resolve ambiguities, and can be tested for integrity and coherence, the chance of identifying errors before they are introduced into the system is great. These benefits are in addition to the measurable deployment and maintenance advantages.

A Domain Rationale

Every domain has its own rationale and arguments for why semantic technologies make sense. In this case, we use a biomedical example. It is particularly suitable because health care and biomedical knowledge, as indeed for all of biology, is a rich domain for semantics.

As the very aspects of life get scrutinized and dissected with our modern technologies and approaches, we are seeing a veritable explosion in the both the amount and nature of biomedical information. Many ideas and concepts not known five to ten or twenty years ago now define this dynamic domain. New and technical terminology keeps arising, but also because the relations are to life and health, these need to be expressed in human terms and at varying levels of sophistication.

This is a perfect example of the relevance of semantic technologies. It is no wonder that more than 250 ontologies now characterize this space, with growth in semantic solutions rapidly occurring because leading institutions and funding agencies are aggressively exploiting and promoting semantic technologies.

Slide Re-cap

For a slide presentation of some of these points, you may see:

[Seven Arguments for Semantic Technologies](#)

View more [presentations](#) from [mkbergman](#).

Thanks.

_____
[1] http://www.mkbergman.com/1618/the-primacy-of-search-in-the-semantic-enterprise

_____

PDF generated by *AI3:::Adaptive Information* blog