

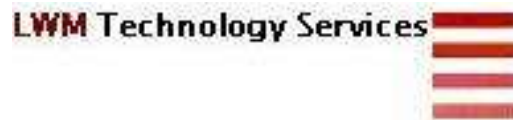
Applying Semantic Technology to Knowledge Assets

Presented by *Lynda Moulton*

- Consultant, LWM Technology Services
- Senior Analyst for Outsell's Gilbane Group

Boston KM Forum, January 20, 2011

lmoulton@lwmtechnology.com



Topics

- ▶ Knowledge Assets (Assumptions)
- ▶ Software Infrastructure (Assumptions)
- ▶ Semantic Software Classifications
- ▶ Linguistic Knowledge Structures
- ▶ Fitting the Pieces into a Semantic Framework
- ▶ Applications for Solving Problems

Knowledge Assets

PEOPLE



Leaders



Experts

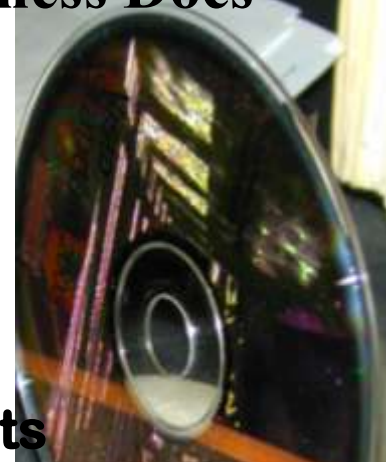
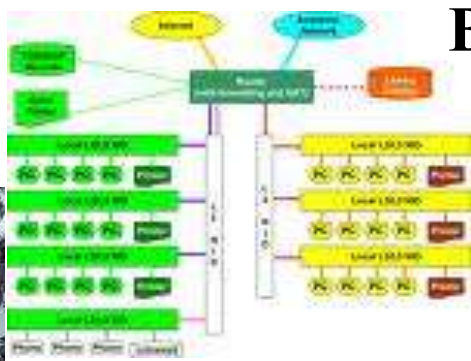
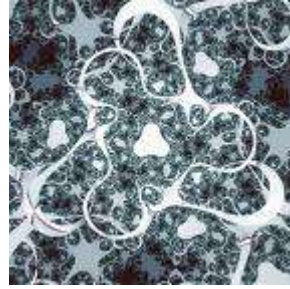
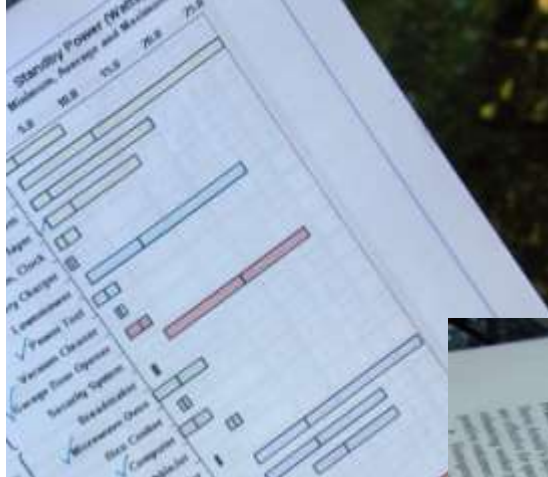


Teams/Communities

Knowledge Assets

Business Docs

Business Docs



Schematics/Drawings
Stored Mixed Formats

Slides/Images



Audio/Text/Artwork



Printed Collections

Data/Spreadsheet

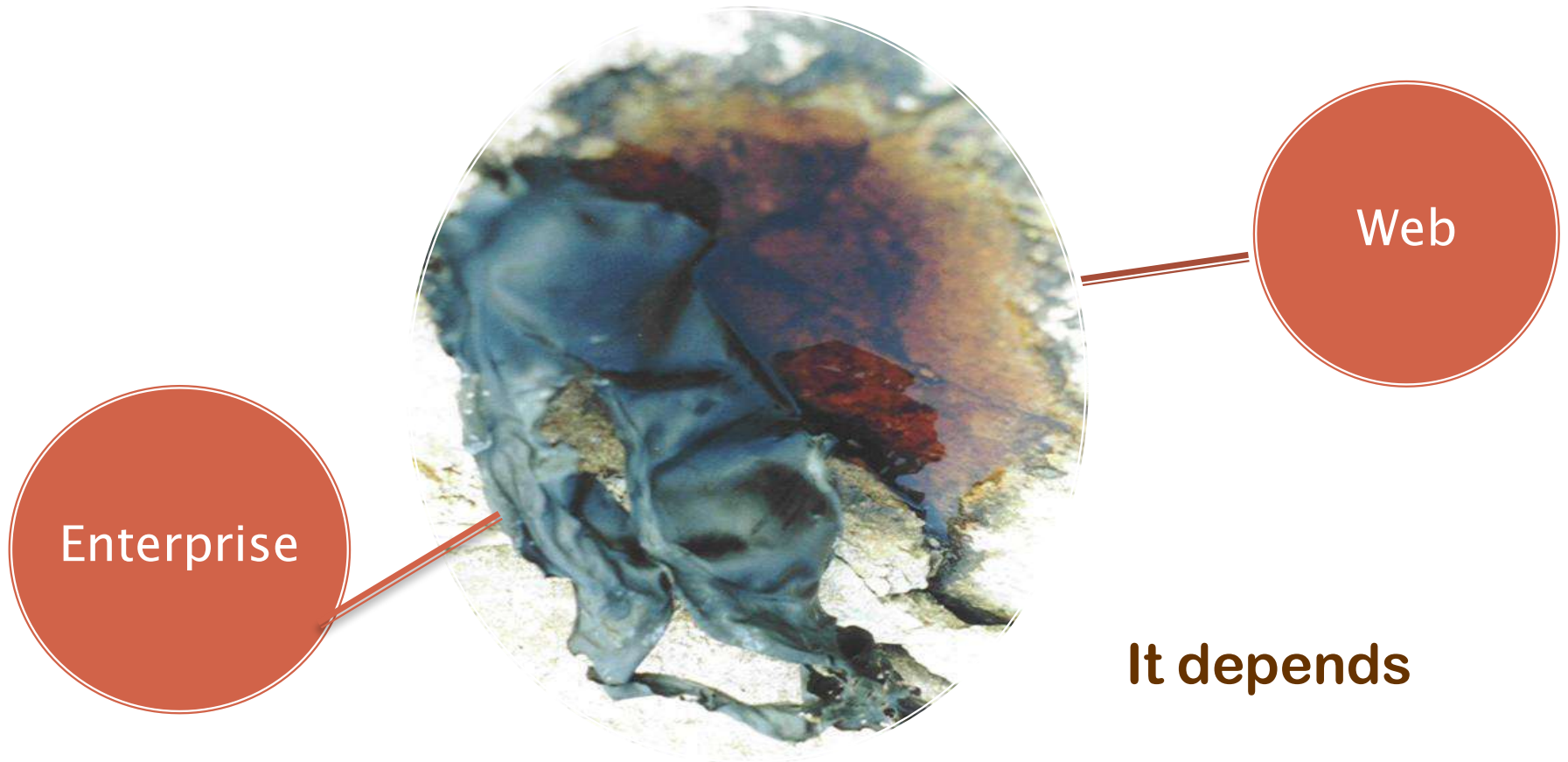
	A	B	C
1	Boards on mc.com as of 9/28/2006		
2			
3	PRODUCT	LAUNCH DATE	A/D and D/A Converter
4	1U Dual Cell-Based	6/11/2006	
5	Cell Accelerator	7/31/2006	
6	Dual Cell-Based	2/15/2006	
7	Dual-Port Interlink	1/1/1900	
8	Echotek Series ECAD-2-12210-PMC 2-Channel 12-Bit	1/1/1900	X
9	Echotek Series ECAD-DA-41-PMC Multi-Channel A/D	1/1/1900	X
10	Echotek Series ECAD-X-081500 2- Channel 8-Bit 1.5	1/1/1900	X
11	Echotek Series ECDR-2414/4814 and 2414-DL1 24/48 Channel Digital	1/1/1900	
	Echotek Series ECDR-3214/6414-		

Software Infrastructure



- ▶ [Real Story Group Networked Representation:](#)
2011 Content Technology Vendor Map
 - ▶ Semantic technology might be embedded in these vendors products
- OR
- ▶ Semantic technologies have relationships to these vendors products
- BUT
- ▶ There are dozens of vendors with semantic technologies not represented here

What problems are we trying to solve?



Arriving at Semantic Search Requires Significant Human and Technological Pre-Processing

State-of-the-Art:

To leverage:

- Managed vocabulary
- Content management, text mining and text analytics tools

to *improve* analysis & findability of
information when we use automated
search

When Does it Make Sense?

- ▶ Voluminous corpus (potentially millions of documents)
- ▶ Content is technically complex in scope and depth
- ▶ Content high-value to seekers who may only require limited portions
- ▶ Required by experts for use in their areas of expertise
- ▶ Poorly differentiated or uncategorized (for purposes of research or e-discovery interest)
- ▶ Significant bottom-line impact, directly or indirectly, when discovered

Semantic Content Tool Categories

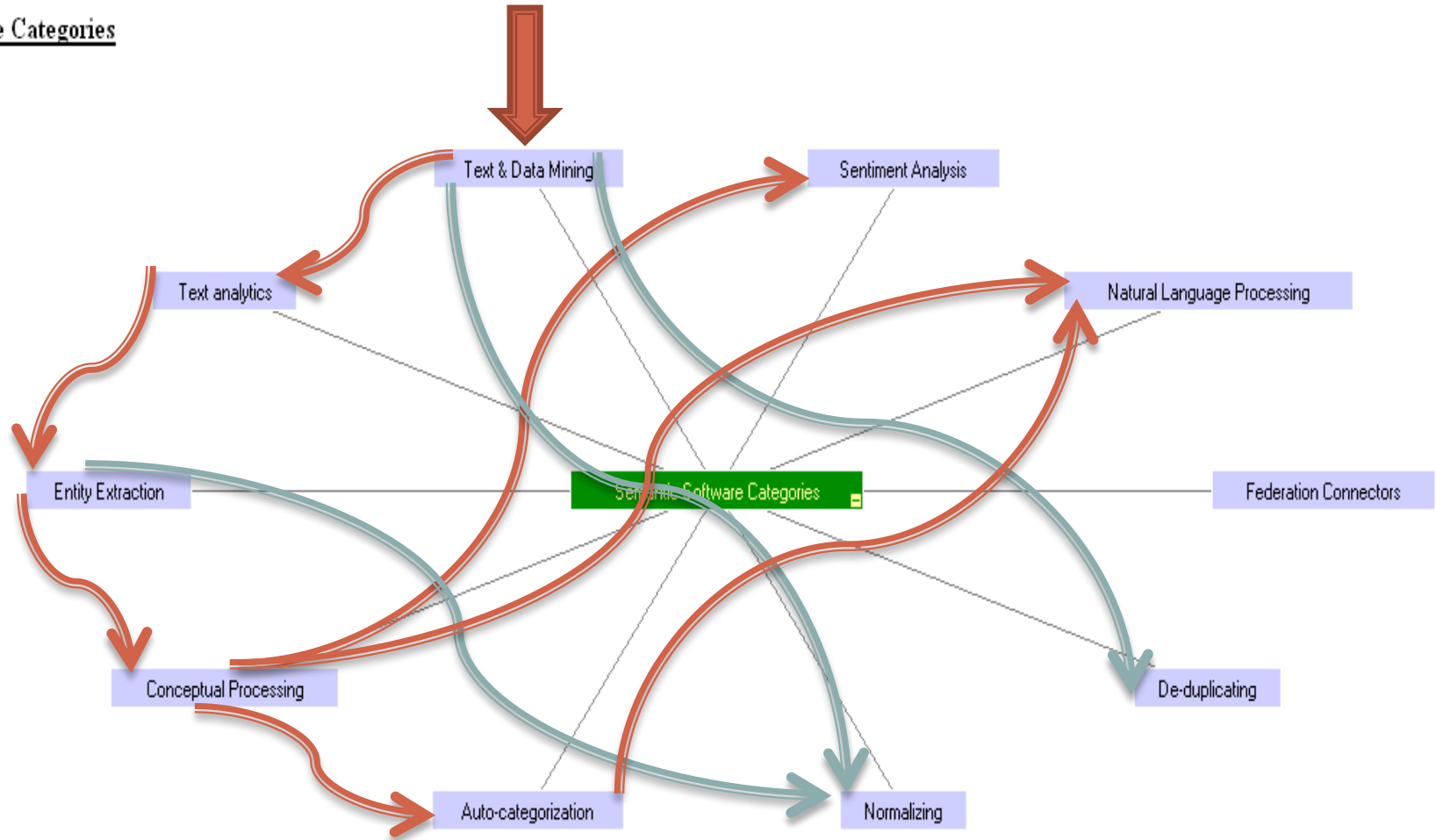
- ▶ Text mining and text analytics
- ▶ Concept and entity extraction
(embedded)
- ▶ Concept analysis (embedded)
- ▶ Natural language processing (NLP)
(embedded)
- ▶ Content data normalizing (embedded)
- ▶ Federating and de-duplicating
- ▶ Sentiment analysis
- ▶ Auto-categorization

[The embedded tools are sometimes called human language technologies HLTs]

Semantic Software Classifications



Semantic Software Categories

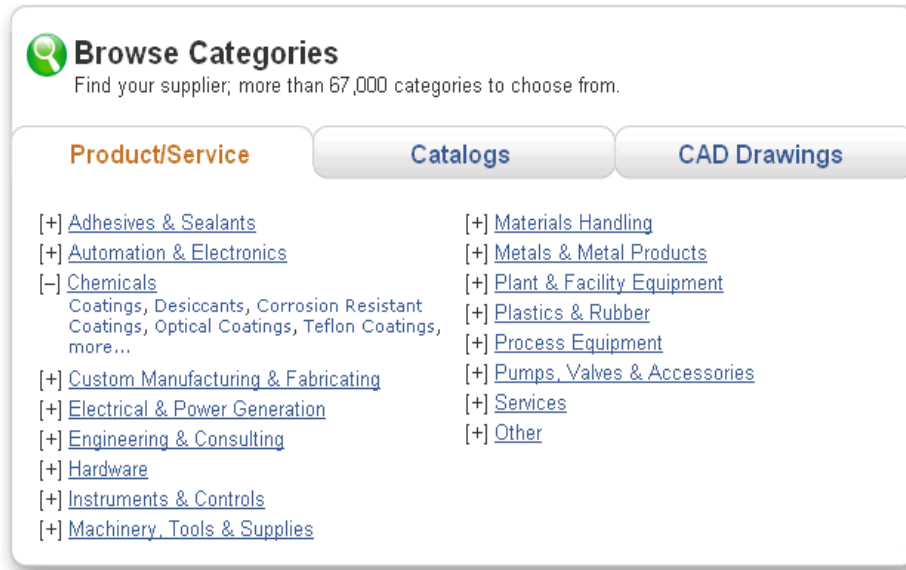


Semantic tools that are in use today are used for:

- ▶ Supporting Semantic Content enhancement
- ▶ Creating Supporting Assets (e.g. ontologies, semantic nets)
- ▶ Enhancing Semantic Search
- ▶ Building Semantic Platforms where all three layers are integrated

Linguistic Knowledge Structures

- ▶ Taxonomies ([e.g. Thomasnet.com](#))



The screenshot shows the 'Browse Categories' section of Thomasnet.com. It features a search icon and the text 'Find your supplier; more than 67,000 categories to choose from.' Below this are three tabs: 'Product/Service' (selected), 'Catalogs', and 'CAD Drawings'. The 'Product/Service' tab displays a list of categories, each with a plus or minus sign and a link. The categories are arranged in two columns.

Product/Service	Catalogs	CAD Drawings
[+] Adhesives & Sealants	[+] Materials Handling	
[+] Automation & Electronics	[+] Metals & Metal Products	
[-] Chemicals Coatings, Desiccants, Corrosion Resistant Coatings, Optical Coatings, Teflon Coatings, more...	[+] Plant & Facility Equipment	
[+] Custom Manufacturing & Fabricating	[+] Plastics & Rubber	
[+] Electrical & Power Generation	[+] Process Equipment	
[+] Engineering & Consulting	[+] Pumps, Valves & Accessories	
[+] Hardware	[+] Services	
[+] Instruments & Controls	[+] Other	
[+] Machinery, Tools & Supplies		

- ▶ Ontologies

http://www.slideshare.net/sveino/semantics-and-search?src=related_normal&rel=3592470
brought to our attention by Steve Arnold.

(Slides 6, 9, 14)

- ▶ Semantic nets

(Discussion:

<http://www.semtalk.com/pub/KimFillies.htm>)

- ▶ Thesaurus ([e.g. MeSH](#))

Meeting Semantic Challenges Requires Knowledge about Assets

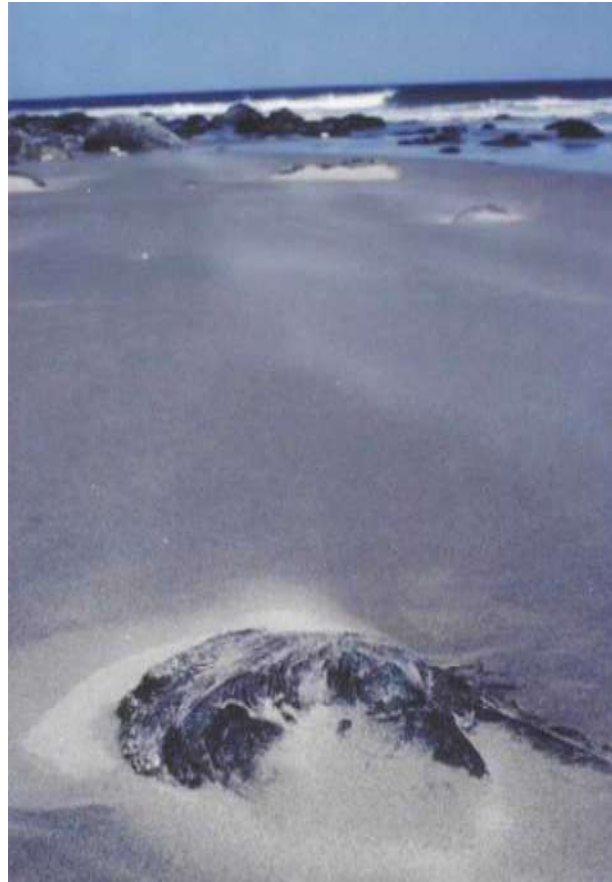
Language, language, language

- Terms & their Meanings
 - Context
 - Disambiguation
 - Curation
 - Perspective
 - Viewpoint
 - Concept extraction
 - Tone
- Grammar – Usage Rules
- Contextual interpretation

Automating language development & applying terminology rules to automatic (tagging) indexing for retrieval

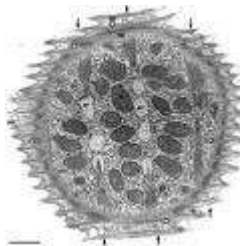
Semantic Problem 1

Context
required



Semantic Problem 2

One word many meanings



F	G	H
<i>Title</i>	<i>Company</i>	<i>Address</i>
Analyst	Partners HealthCare	
	You-know, Cognimetri	153 Dutton Rd.
Founder & Head Surgeon	BrainGrab Studios	

How does this cell get activated?

Sentence context helps disambiguate to discover what is intended but it is not the whole answer.

Semantic Problem 3



Beaver Surgical Instruments, Inc.



- Evidence of animal – image
- Entities named Beaver – image of person and company name

Beaver is attempting to correct some cutting defects

Again context (absence of ‘the’) provides some help but does not entirely resolve whether Beaver the man or the company is doing correcting. Curation needed.

Semantic Problem 4



More Perspective Helps Clarify



Semantic Problem 5

"I want to help clean up the state that is so sorry today of journalism. And I have a communications degree. I studied journalism, who, what, where, when, and why of reporting," xxx continued. "I will speak to reporters who still understand that cornerstone of our democracy, that expectation that the public has for truth to be reported. And then we get to decide our own opinion based on the facts reported to us.

Who said it? Why is it being said? What does it mean? What is the context? Does having the answers move you any closer to understanding what was meant? (viewpoint questions)



Extracting concepts when language is clearly foggy needs more contextual data to reveal the viewpoint.

Semantic Problem 7

Measuring tone or sentiment



Problem



Begin with These Realities

- ▶ Teasing concepts from any unstructured documents is among the most difficult of computational problems.
- ▶ There are relatively few experts in the field who have reached “rock star” status, yet.
- ▶ There are no companies or products in this field that have eclipsed all others offering universal semantic processing or semantic search, yet.
- ▶ Success with any semantic software option will depend on how it is implemented, supported by a team of people who understand the tool, and how much computing resource is devoted to the software.
- ▶ If you begin now, you will be in an early adopter category.
- ▶ Better results may well be achieved by using a number of software tools that complement each other, either in an integrated fashion or in tandem.
- ▶ Begin with a discrete problem you are trying to solve.

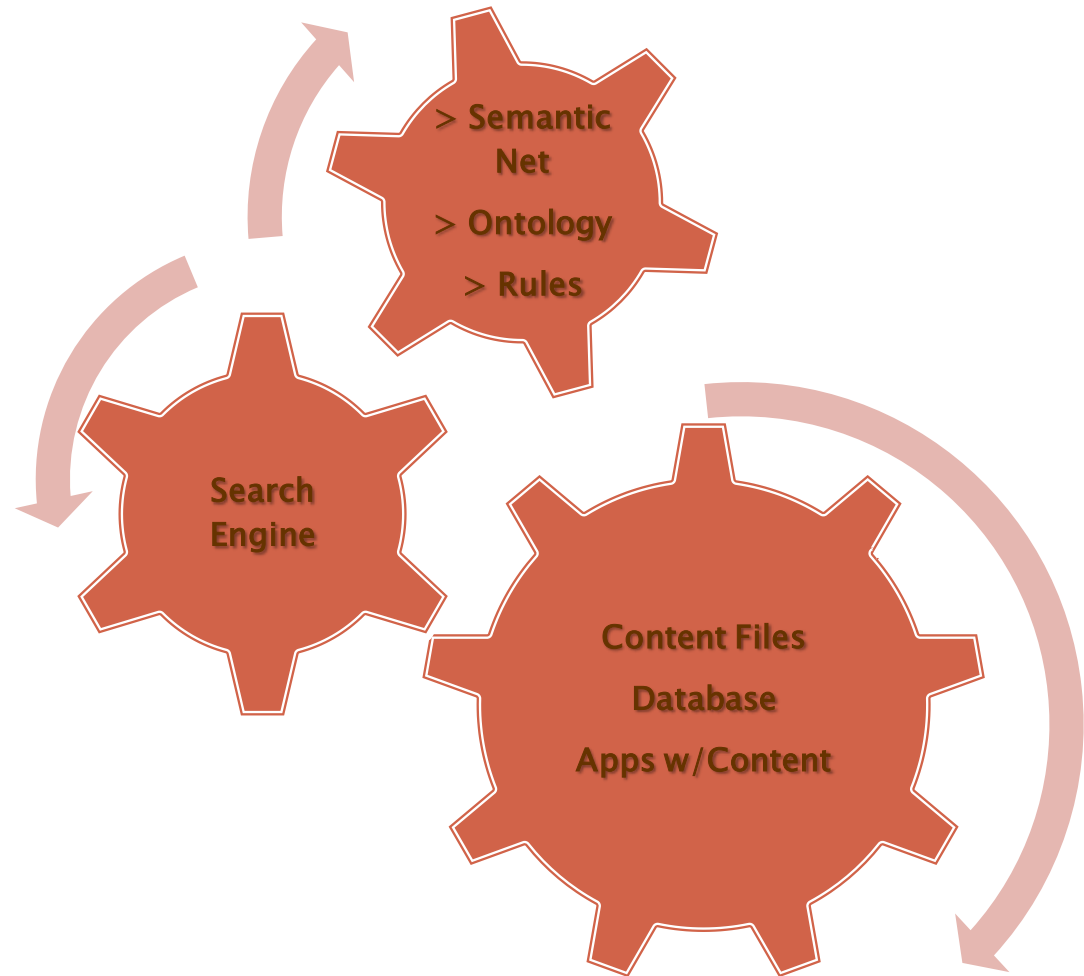


Primary Challenge: improving access to structured & unstructured Siloed content

Enrich the context, elicit concepts, improve tagging

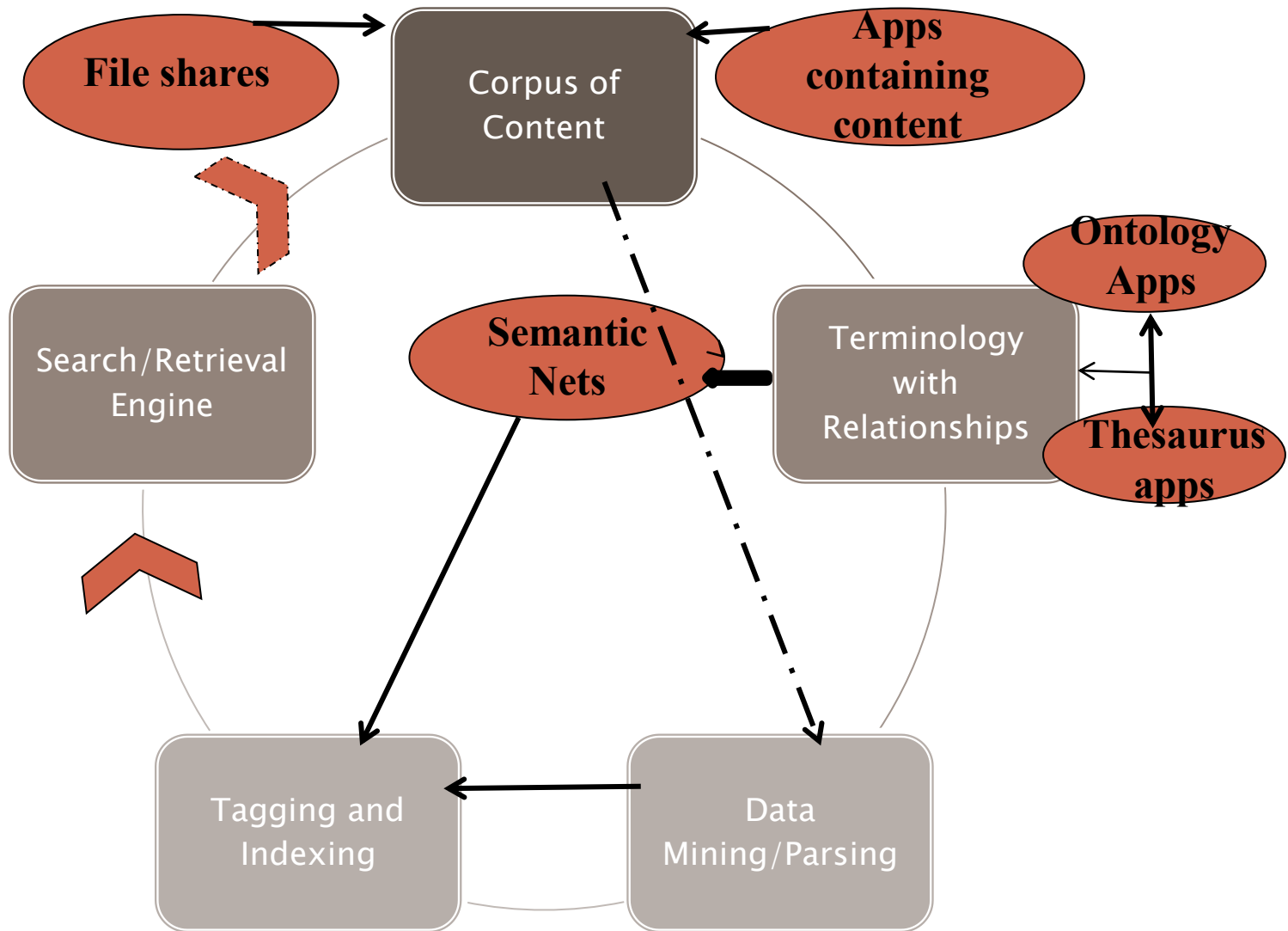
Leverage content for added enterprise value

- Meanings are defined in one location
- Data & content are elsewhere
- Search engine applies meanings to find appropriate content



Simple Platform Model

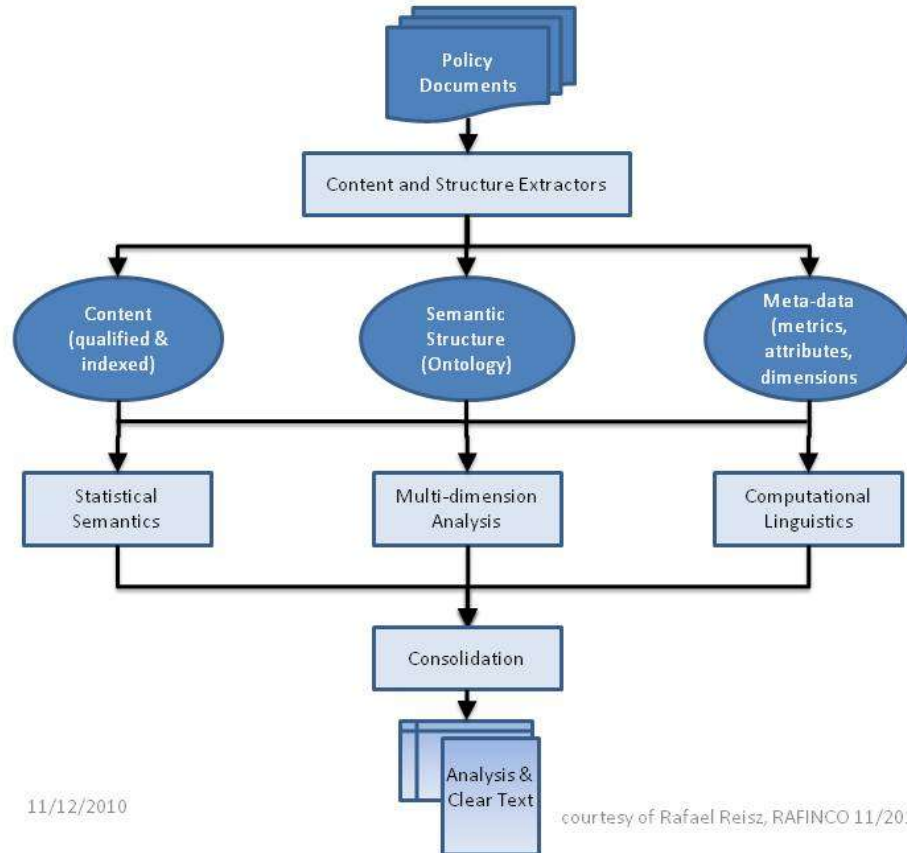
Potential Framework Model



IDEA: Model for Analyzing Public Policy Content



High Level Schematic of Policy Processing Flow



BIG Picture Components: Technology and People



- **No Technology Offers Solutions for ALL the Semantic Challenges an Enterprise Faces**
 - **No one Product does it ALL**
- **Quality People to Take Care of Language Management and Ongoing Curation are REQUIRED and Hard to Find**
 - **People are Needed to Interact with tools: NLP, text mining, text analytics, etc.**

Applications for Solving Problems

- ▶ 90+ commercially viable or installed in multiple institutions
- ▶ Embedded and standalone
- ▶ Integrated with search engines
- ▶ Representative products by application and industry

Vertical Industry Use Cases

- ▶ Corporate Governance – Cambridge Semantics (Federating and Normalizing Spreadsheets and Business Databases across disparate global entities)
- ▶ Energy exploration & production – Expert System (Semantic mapping and auto-categorization for signals on industry changes for competitive intelligence)
- ▶ Life sciences – Linguamatics (text mining for fact and concept extraction; tagging)

Vertical Industry Use Cases

- ▶ [Defense](#) – [Basis Technology](#) (multi-lingual entity extraction)
- ▶ [Data Warehouse Analytics](#) – [Attivio](#) (Platform for Building Content Analysis and Search Applications)
- ▶ [Financial](#) – [Lexalytics](#) (Industry news surveillance – sentiment analysis)

Partial List of Companies with Semantic Applications or Components

- [Attensity](#)
- [Attivio](#)
- [Basis Technology](#)
- [BiText](#)
- [Cambridge Semantics](#)
- [Clearwell](#)
- [Cognition](#)
- [Collexis](#)
- [Concept Searching](#)
- [Expert System](#)
- [JD Powers](#)
- [Lexalytics](#)
- [Linguamatics](#)
- [Metatomix](#)
- [MuseGlobal](#)
- [Nstein \(OpenText\)](#)
- [Patterns & Predictions](#)
- [Recommind](#)
- [SAS: Teragram](#)
- [Seal Software](#)
- [Sinequa](#)
- [Smartlogic](#)
- [TEMIS](#)
- [TopQuadrant](#)
- [Zylab](#)

Question and Answer



Thank you for your
interest

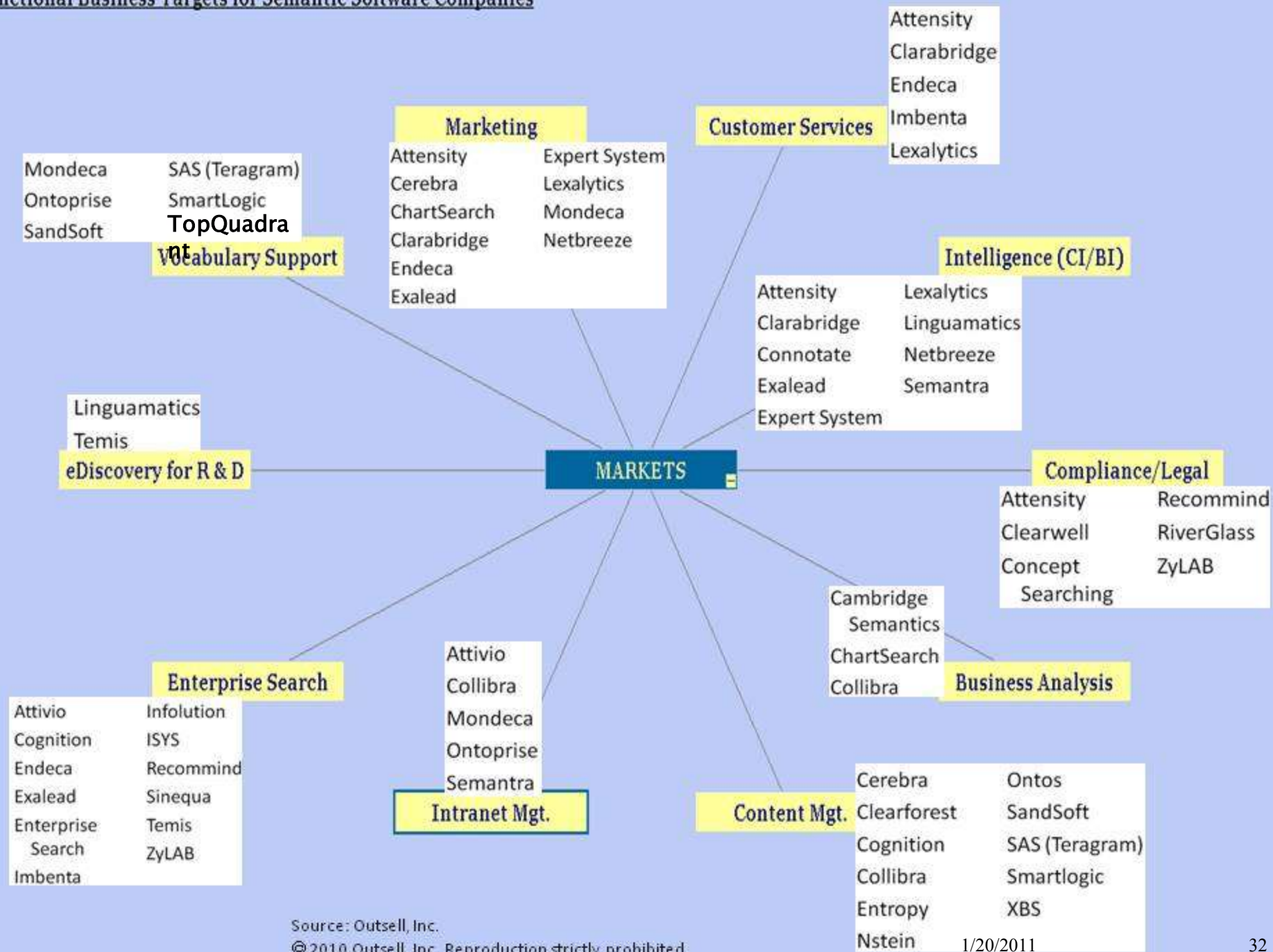
Lynda Moulton,
Consulting on strategic
planning and product
and process analysis

lmoulton@lwmtechnology.com

978-660-1845

Supplemental information >>>>

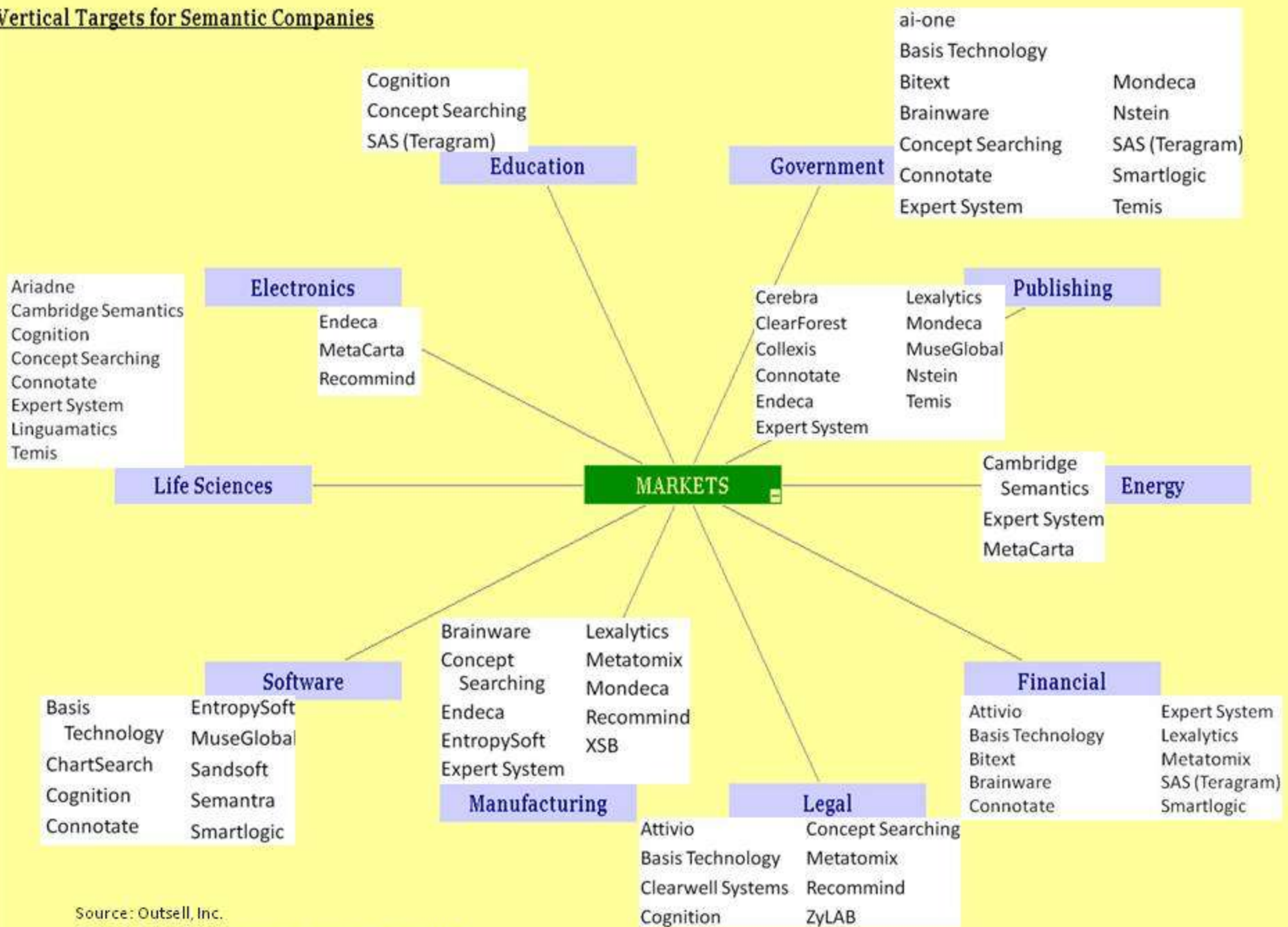
Functional Business Targets for Semantic Software Companies



Source: Outsell, Inc.
 © 2010 Outsell, Inc. Reproduction strictly prohibited.

1/20/2011

Vertical Targets for Semantic Companies



Source: Outsell, Inc.

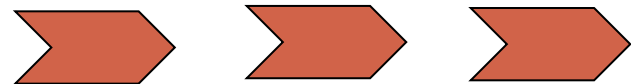
© 2010 Outsell, Inc. Reproduction strictly prohibited.

Cases in which semantic tools contribute business and retrieval benefits:

- ▶ [Knovel](#) – Publishing enhancing content for engineering & scientific problem solving
- ▶ [National Institute for Clinical Excellence](#); [UK National HS Information Centre](#)
- ▶ e-Discovery – legal and regulatory
- ▶ e-Discovery – expertise and relationships
- ▶ Fact relationships – life sciences
- ▶ Correlating relevant content across repositories – energy ([call to action for energy industry](#))
- ▶ Sentiment analysis – e-commerce and customer relationship management
- ▶ [Audio](#) and Image – tough (future)

Where to Begin

- ▶ What are you trying to achieve?
- ▶ What is the content domain?
- ▶ Who will be seeking content?
- ▶ What are the types of questions they will be asking?
- ▶ How is content curated & metadata added
- ▶ What is the current infrastructure?
- ▶ Will you be replacing existing components or enhancing tools?
- ▶ Where is the leadership and commitment?



Getting to Work on the Solution

- ▶ Team:
 - ✓ language (list management, vocabulary management (SMEs))
 - ✓ content management
 - ✓ metadata management
 - ✓ IT support
- ▶ Language: Ontology/Thesaurus
- ▶ Tool selection: Language management, Content management, Search

Links for Background Research & Study

- ▶ ***Semantic Software Technologies: Landscape of High-Value Applications for the Enterprise***, by Lynda Moulton, Outsell's Gilbane Group <http://gilbane.com/Research-Reports.html#semantic>
- ▶ Glossary for Search: <http://www.microsoft.com/enterprisesearch/en/us/search-glossary.aspx>
- ▶ Relating *Ontologies* to topics discussed: http://en.wikipedia.org/wiki/Ontology_%28information_science%29
- ▶ Semantic Net: http://en.wikipedia.org/wiki/Semantic_net (lexicon)
- ▶ Case Development & Rule Building: Ceusters W. *Medical natural language understanding as a supporting technology for data mining in healthcare*. In: KJ Cios (ed.) Medical Data Mining and Knowledge Discovery, Physica-verlag Heidelberg, New York, 2001;:41-67. <http://org.buffalo.edu/RTU/papers/MNLPminingpdf.pdf>
- ▶ Parser Illustration of One: <http://www.nyu.edu/pages/linguistics/parsers.html>
- ▶ ***Introduction to Chomsky's Grammar and Parsers***: <http://www.nyu.edu/pages/linguistics/parsers.html#TWO>
- ▶ **Historical Instruction: *The state of the art in corpus linguistics*** by Geoffrey Leech
<http://ccl.pku.edu.cn/doubtfire/CorpusLinguistics/Introduction/The%20state%20of%20the%20art%20in%20corpus%20linguistics.htm>) <http://www.nytimes.com/2010/10/05/science/05compute.html>
- ▶ NELL (Never-Ending Learning Language system: *Aiming to Learn as We Do, a Machine Teaches Itself*, by Steve Lohr, NY Times, 10/04/2010 <http://www.nytimes.com/2010/10/05/science/05compute.html>

Influencing Bodies & Organization

- ▶ [W3C](#)
- ▶ [SemTech](#)
- ▶ [SemanticWeb](#)
- ▶ Open Source Community (Committers)
- ▶ System Integrators
- ▶ Professional Associations (vocabulary control operations)