On The Cusp: A Global Review of the Semantic Web Industry

David Provost
A Global Review of the Industry and Leading Vendors

Sep 30 2008
Table of Contents

Executive Summary ............................................................................................................. 3
Who Should Read This Report .......................................................................................... 4
Method ................................................................................................................................ 5
Highlights and Findings ...................................................................................................... 6
Conclusion ............................................................................................................................. 9

Industry Participants:
- Aduna ................................................................................................................................. 10
- The Calais Initiative ........................................................................................................... 12
- Cambridge Semantics ......................................................................................................... 14
- Dow Jones Client Solutions Profile .................................................................................. 16
- Expert System ................................................................................................................... 18
- Franz .................................................................................................................................. 19
- Mondex ................................................................................................................................ 21
- ontoprise ............................................................................................................................. 23
- Ontos ................................................................................................................................. 25
- OpenLink Software ........................................................................................................... 27
- Primal Fusion .................................................................................................................... 29
- Saltlux ............................................................................................................................... 30
- Sindice ............................................................................................................................... 32
- Thetus ............................................................................................................................... 33
- TopQuadrant .................................................................................................................... 35
- Twine.com ....................................................................................................................... 37
- Yahoo! SearchMonkey ...................................................................................................... 38

© David Provost 2008, Some Rights Reserved

You are free:
- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work

Under the following conditions:
- Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page:
  http://creativecommons.org/licenses/by/3.0/
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

David Provost is a business development professional and a strategist. His involvement with the World Wide Web began in 1996 as one of the original team members at Gomez.com. Recognizing the fundamental role of technology in creating end user experiences, he turned his attention to emerging Web technologies and their effect on business results. David earned his BS and MBA from Boston University and an SM from MIT, where he wrote his thesis titled "Hurdles in the Business Case for the Semantic Web".
Executive Summary

The Semantic Web is proving itself as a commercially competitive technology.

With every passing day, the vendors profiled in this report gather more evidence that drives this point home. For several years the Semantic Web has been portrayed as full of promise, yet real-world evidence articulated in plain, non-technical language has been in short supply. This report is a step toward rectifying this situation by profiling the leading vendors in this space along with several “special cases” such as Twine, Primal Fusion, The Calais Initiative and others. The focus is very much on what these players “do” (their products and services) and not so much on the technical underpinnings of how they do it.

Several key events have occurred in the industry during the last six to twelve months, including the entry of major corporations like Thomson Reuters, Dow Jones, and Yahoo! Other developments include the launch of Twine’s beta site, Primal Fusion’s upcoming beta, and the continued growth and vitality of the vendor community. At the most basic level, both early stage companies and seasoned, global companies are investing in the Semantic Web because of the clear role they see for the technology in the achievement of their overarching strategic objectives.

Intriguing possibilities are emerging, such as the role of “linked data”, Social Network Analysis and how the Semantic Web may aid this practice, and how the convergence of Natural Language Processing, Semantically enabled search, and the traditional publishing industry will play out. Time will tell, but the potential effects could be substantial.

In the meantime, the dialog among industry players is shifting away from being intensely technical with a strong emphasis on low level concepts, toward being more focused on applications, solutions, and the problems the Semantic Web can solve. This is an important

---

The Semantic Web is a concept obscured by difficult acronyms and dense material. Some basic definitions:

**The Semantic Web is** an extension of the World Wide Web designed in part to standardize the integration of data and the interoperation of applications.

**RDF (Resource Description Framework)** is a language and paradigm for data integration.

**Ontologies should be thought of as thesauruses** – by looking up one word, related words are presented. In turn, looking up each of the related words causes the web of related terms to expand. Limiting terms to a specific concept, activity, or field would be akin to creating a model.

**A model** is a set of interconnected concepts, rules, and properties. A simple non-Semantic model might be a spreadsheet that models mortgages by using interest rates, principal amounts, payments, and the duration of the loan. Ontological models range from things as simple as spreadsheets to logical models that are more complex.

**SPARQL makes federated queries possible**, and federated queries can be thought of as queries made simultaneously to multiple, disparate databases that could be located anywhere on a network. This is a key concept in making disparate and distributed data stores seem like a single repository.
transition in cementing the Semantic Web’s place among commercial technologies and while it’s occasionally been painful, this transition is well underway and a growing market is receptive to the messages that are emerging.

**Who Should Read This Report**

**Business managers and strategists** seeking market opportunities and competitive advantages from innovations in Information Technology (IT). The Semantic Web is an emerging technology and its full range of uses has not yet been discovered. Since the Semantic Web is an extension of the World Wide Web (WWW), it will be ubiquitous – any location and any device that can reach the WWW will be able to reach the Semantic Web. Any rapidly evolving technology that can reach a global market almost instantaneously deserves close attention.

**Customers who are evaluating the Semantic Web** and the vendors in the industry. This report sharpens the distinctions between vendors and uses plain language to describe what their products actually “do”. Historically, the industry has struggled with articulating the value of the Semantic Web. This is partly due to the emerging status of this technology, and partly because the industry’s vocabulary has been heavily oriented toward developers and software engineers. Business managers who are seeking an edge in their own competitive spheres require different frames of reference and terminology and this report is a first step toward providing those tools.

**Vendors should read this report** to observe how other companies articulate their value and direction. Each profile in this report has been reviewed by the respective vendor for accuracy. This step was taken as a basic fact checking measure and to ensure close adherence to how each of these companies “tells their story”. With this knowledge, vendors can take another step in the evolution of their positioning, messaging, and competitive tactics and strategies.

**Developers, software engineers, and technical personnel** who are interested in communicating with the growing business audience for the Semantic Web. Business managers are always happy to learn about technologies that will improve their company’s performance. However, they face constraints of time and training when delving deeply into the workings of the Semantic Web. With this report developers may be better equipped with a slightly different vocabulary and/or perspective that allows the two groups to meet somewhere in the middle.
Method

The primary method for researching this report was to interview company representatives either by phone or in person. All interviews followed the same basic structure defined by a set of questions. There were minor variations in the question sets used for vendors vs. deployers, but within each of these groups the questions were the same and any variations that occurred happened strictly in the course of conversation.

The questions themselves were developed both independently based on personal experience and in concert with some of the companies reviewed. Industry participation was welcome in this process to ensure fairness, coverage of key issues, and care was taken to temper any potential for bias in the final questions.

Selecting the companies for inclusion in this report posed challenges due to the emerging nature of some ventures, the global dimension of the industry, and the generally imperfect nature of information in a fast-paced, rapidly evolving industry. Nonetheless, the basic guidelines were:

- Vendors had to be product based, not professional services based.
- Deployers like Dow Jones, The Calais Initiative, Twine, and Primal Fusion had to provide evidence of their efforts to create an online service or product. This requirement was easily satisfied in every case.
- Each company had to establish credibility as a going concern, meaning there had to be reason to believe these entrants were more than an idea, demo, or the proverbial “two guys in a garage.” In cases where the company was not familiar, factors like conference sponsorships and the like were taken into account. The rationale for assessing sponsorship activity was that most companies will invest in their products, company infrastructure, personnel, and sales and marketing before using precious cash to sponsor a conference.

Prepared with both questions and a list of companies, invitations were sent soliciting participation (25 companies were invited; 17 companies participated). Phone calls or meetings were scheduled, the interviews performed, and the company profiles were written based on those discussions. Every company received a draft copy of their profile to review for accuracy and correctness. Most profiles were returned with minimal, if any edits, and in all cases edits were considered and generally applied.

The finalized profiles were reviewed, patterns and themes were identified and the findings were written in the form of this report.
Highlights and Findings

The Semantic Web (SW) industry is global and the companies covered in this report reflect this span of geography, with ten based in North America, six in Europe and one in South Korea. European representation is significant among these players and some assert that long term research policies among European governments and on the part of the European Union (EU) itself have played a direct role in this outcome.

Whatever the case, the vendors this report have all the appearances of thriving, emerging technology companies and they have shown their readiness to cross borders, continents, and oceans to reach customers. Just as these companies are geographically diverse, this group is equally diverse in addressing different aspects of the SW stack.

![Diagram of SW stack](image)
The result is a range of capabilities and includes databases specific to the requirements of storing Semantic data, ontology development and management tools, application development platforms, Natural Language Processing (NLP – a means of extracting the structure and concepts contained in text and making the results ready for search and integration purposes) and more.

Importantly, some companies are beginning to focus on specific uses of Semantic technology to create solutions in areas like knowledge management, risk management, content management, and more. This is a key development in the Semantic Web industry because until fairly recently, most vendors simply sold development tools. In turn, customer organizations would use these tools to create solutions that suited their own internal needs. Obviously, this strategy limits the addressable market to those customers that have the knowledge, time, money, and motivation to invest in such a bottom-up approach.

<table>
<thead>
<tr>
<th>Company</th>
<th>Solution</th>
<th>Middleware</th>
<th>NLP</th>
<th>Database Platform</th>
<th>Ontology</th>
<th>Search</th>
<th>Consumer Web Service</th>
<th>Developer Web Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aduna</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Calais Initiative</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambridge Semantics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Dow Jones Client Solutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Expert System</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mondeca</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Ontoprise</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ontos</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenLink Software</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primal Fusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Saitlux</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sindice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Thetus</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TopQuadrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Twine/Radar Networks</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yahoo!/SearchMonkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

By emphasizing solutions, vendors can more sharply focus their development and sales efforts. This leads to the first point of this report:

**Semantic Web technology and solutions based on this technology are competitive with “traditional” Information Technology (IT).** Ontoprise’s SemanticGuide is making genuine inroads among service organizations tasked with troubleshooting complex systems. This product can be integrated with trouble ticket systems, related databases, and equipped with an integrated knowledge base. Once implemented, service personnel can take advantage of how an ontology-based solution can work, obviating the need for a predetermined, specific starting point and allowing any set of facts and observations to serve as the start to the problem resolution process.

Franz’s Allegrograph is an excellent example of a Semantic database that’s been very positively received by the market. Increasingly, the value of a schema-less database is being recognized as a way to deal with data storage requirements that include being able to quickly and easily add and store new and unplanned-for data types. Allegrograph’s continued success can easily be envisioned because there’s reason to believe this technology can make Social Network Analysis...
more effective and more easily scaled to Web proportions. A win here may open the door better targeting of advertising, recognition of (and consequent action on) social networking patterns and habits, and superior integration of metadata which would lend itself extremely well to better and broader search results.

Although it’s yet to be proven, Cambridge Semantics’ Anzo for Excel may prove to be a highly effective, economical way to integrate the vast quantities of data contained in these files across an enterprise. Of course the beauty of this starting point is that once the Excel “problem” has been addressed, the Anzo infrastructure can easily adapt to include other file formats and data types. As a result, customers won’t need to bring in a new solution, go through a new learning curve, or end up with multiple solutions all working toward the common goal of integration and interoperation.

The above examples are just a brief sampling of the commercial success that the Semantic Web has been experiencing. In broad terms, it’s easy to point out the longevity of many companies in this industry and use that as a proxy for commercial success. With more time (and space in this report), additional examples could be described but the most interesting prospect pertains to what the industry landscape will look like in twelve months. At that point, The Calais Initiative, Twine, SearchMonkey, and other brand new entrants will have had time to prove their worth. At the same time, the previously existing vendors in this report will have had even more time to refine their solutions and rack up customer wins. It seems quite reasonable to expect many more success stories as the Semantic Web continues its evolution and adoption.

Natural Language Processing (NLP) is emerging as a force for taming World Wild Content. Borrowing from Thomson Reuters (and it’s a simplification for discussion purposes), there are two kinds of content: High quality, professionally produced content and all the rest, or “wild” content. From this perspective, the professional content is taken as a known quantity. Comparatively, the wild content is unknown – it could be good, even great, or it could be bad and just plain wrong.

Bear in mind that some in the field feel that NLP is about as good as speech recognition (which this same constituency feels is barely good enough). This assessment may be overly pessimistic and regardless, NLP will continue to evolve and improve. As it does so, by extracting key concepts and document structures and integrating these data with other metadata, NLP serves as an important first step in preparing the global span of content for Semantic consumption and use, whether by people or applications. Since it has never been done before, it’s hard to estimate the value of NLP applied to the Web’s content, but current levels of investment, interest, and usage suggest the value could be quite high.

Linked data will have important and extremely valuable uses, and these are slowly gaining recognition. It doesn’t help matters that the concept of linked data is so murky, but it’s coming soon to an application near you. Semantic Web technology allows for linking to data directly down to the record level, not just the file level. Now, imagine linking data in several different databases that could be distributed anywhere on a network. In a bank, scenario might look like this:

A certain population of the bank’s credit card holders has been identified as being extremely lucrative. Data for these customers is stored in a number of different databases, and it’s even been discovered that certain publicly available leading economic indicators serve as significant factors in this group’s behavior. Linking data would enable the assembly of a composite data set drawn from all relevant data sources (and selecting the data only) and made ready for processing by a model developed to understand and predict this group’s behavior. This approach could position the bank to anticipate spending patterns over a timeframe during which the
organization could act to provide additional services, products, or work to reinforce the overall customer relationship.

The example above is a data intensive application of Semantic technology designed to support business decisions related to a valuable customer segment. It’s not the kind of thing that’s likely to be generally accessible on the Web, which highlights a related point: Some of the most meaningful uses of linked data may be difficult or even impossible to appreciate visually. Unlike the intense visual experiences found on the World Wide Web, many aspects of the Semantic Web are likely to be hidden from view, even though they may play a direct role in what appears on a computer screen.

**Marketing, technical, and solution partners will play a much greater role in selling Semantic Web-based solutions.** Every vendor in this report knows Semantic Web technology extremely well but most of them lack domain expertise, whether it’s in financial services, life sciences, manufacturing, etc. Some companies, like Ontoprise, Ontos, OpenLink Software, and others have well-developed partner relationships (Ontoprise may be the industry leader in this respect) that serve as productive channels for selling their products into a wide range of accounts. Whether these partners are called OEMs, ISVs, or VARs (Original Equipment Manufacturers, Independent Software Vendors, Value Added Resellers, respectively), they can play a valuable role in providing domain expertise, sales relationships, not to mention actual sales forces that are ready, willing, and able to sell new products and services into their customer base.

These partners generally focus on a limited number of industries or vertical markets. The relevant candidates for partnership deals have proven businesses in their own right, they focus on one or two vertical markets, and they have the technical ability to learn Semantic Web technology and how to use it in the markets they serve. More than one Semantic vendor has pointed out that it’s difficult to find the right partners, and it’s difficult to learn how to choose good partners.

**Conclusion**

As the Semantic Web industry continues to mature and targeted applications and solutions emerge, the nature of industry discourse will change. Instead of focusing on extremely low level concepts and terminology used by researchers, a much higher level of discussion will emerge and it will be increasingly driven by business managers. Ironically, as the transition from theory to practice occurs, many researchers are likely to re-focus on the fundamental aspects of their work. For this group, the appeal of exotic, cutting edge research and discovery is much greater than the “routine” of building businesses and expanding a market.

Interestingly, Semantic vendors are playing a direct role in changing industry dialogue, beginning with early steps in the process of differentiation. Referring once again to the Semantic technology stack, each layer is designed to support fairly specific capabilities. In turn, products that deliver any of these capabilities could be enough to base a business on. The current crop of vendors shows competitive overlap in some cases, but there are as many or more opportunities for collaboration, and this latter point may be only just starting to crystallize in the minds of these players and their customers. As a result, the dialogue, application of the technology, and the industry dynamics are poised to evolve to another level.

The Semantic Web industry is alive, well, and it’s increasingly competitive as a commercial technology. At this point, there are too many success stories and too much money being invested to dismiss the technology as non-viable. The Semantic Web is presently building a track record, which means the big wins and unanticipated uses are yet to come. In the meantime, adoption is occurring, and the early news is very good indeed.
Aduna

Company: Aduna
URL: http://www.aduna-software.com/ (Sesame: http://www.openrdf.org/)
HQ: Amersfoort, The Netherlands
Products (Primary): Aperture, AutoFocus, Sesame, AutoFocus Server, Spectacle
Survey Respondents: Jeroen Wester
Vendor Category: Solution/Middleware/Database

Employees: 10
Revenue: --
Installed base: (Sesame and AutoFocus) 25,000-30,000

Primary Offering:

Aduna is best known for Sesame, the widely adopted open source Semantic database. Without getting into the technical details, Sesame is a specialized database designed to handle the unique storage requirements associated with how Semantic Web technology reads and writes data. Semantic databases are also referred to as triple stores, RDF databases, or some variation on these terms.

Since it’s open source, Sesame is available at no cost, although implementing and maintaining the software will require professional support. Aduna is well positioned to provide these services since its own development team plays a lead role in contributing to this project. Furthermore, since Sesame is an essential part of the product suite offered by Aduna, it’s highly likely the company will remain deeply involved in Sesame’s development and use.

AutoFocus is the other key product offered by Aduna. This product is a highly customizable navigation and visualization tool that’s designed to connect with Sesame (or any other Semantic database). This product provides a number of intuitive visual cues, or Cluster Maps (screenshots available on aduna-software.com) of what’s being presented, along with user-selectable panes containing navigation elements leading to different views of the data. By the way, these different views are referred to as “facets”, and AutoFocus could be considered a “faceted browser”, which is a pretty common term for Semantic visualization tools.

“Hairballs” are a common problem when viewing Semantic data or models. This term refers to the sheer density and complexity of what’s being viewed. It’s a visual experience so it’s best seen first hand but nonetheless, AutoFocus presents search and exploration results in Cluster Maps that clarify the relation between data, files and Web pages, and data attributes, like file type, size and author. The user benefits from this way of presenting information because relations in the data set are clearly shown.

Key Differentiators:

Use of open source technology is normal among Semantic Web vendors but very few actually maintain their own open source communities. By making Sesame an open source project, Aduna gets the benefit of a committed group of developers who continually improve the product. It also ensures that there’s a larger group of professionals who are qualified to work with Sesame and use it in production environments.

In Aduna’s case, Sesame “opens the door” to customer accounts that download it and put it to use. Once these customers gain experience they may call upon Aduna to provide a larger and more tailored implementation, at which point the full suite of Aduna products may come into play.

No matter how the sales/implementation process begins, Aduna customers benefit from the many contributions, perspectives, and support associated with open source technology. Once in place, the company then stands ready to support Sesame with best practices learned from its community. As implementations grow, Aduna can follow on smoothly with AutoFocus and the rest of its product suite to support research, discovery, and the comprehension of complex models or environments.
**Six/Twelve Month Plans:**

Aduna has significant plans for improving Sesame, which include improving the database’s federated query (e.g., queries executed across multiple databases) capabilities and some form of predicting where “the answer” may most likely exist within a network of data stores. Other plans include continued enhancements to AutoFocus and AutoFocus Server, including focus and filtering facilities for refinement of query results and improved usability of the tools.

In response to growing interest from the US market, Aduna may open a subsidiary office there.

**Analysis:**

Sesame is well known and well regarded and it’s a valuable calling card for Aduna. The company has developed some very interesting visualization tools that work well to reduce the complexity often associated with actually “seeing” what an end user might be working with. These two factors, combined with Sesame’s open source status (eliminating price as a barrier to adoption) position the company well to penetrate accounts. Note that winning accounts is likely to expand Aduna’s open source community even further, which is almost certainly a good thing.

The company’s plans to address federated query capabilities is timely and in keeping with the direction of the industry. If successful, this will be an important step toward the concept of “linked data”, a powerful and complex topic worthy of its own report. Like many Semantic ventures, Aduna has strong ties to government funded projects, but the continued development and adoption of Sesame and the company’s other products are likely to pull it further into the commercial mainstream. Aduna relies on development and marketing partners to reach its broader market but it’s very possible the broader market may increasingly seek out solutions with “Aduna inside”.

The Calais Initiative

Company: Thomson Reuters, Inc.
URL: http://www.opencalais.com/
HQ: New York NY, USA
Products (Primary): The Calais Initiative
Vendor Category: NLP

Employees: 50,000
Revenue: US$13.94 Billion
Installed base: 5,000 developers, 1.2 million pieces of content processed per day

Primary Offering:

The Calais Initiative (Calais) comprises several tools for processing text, but the core product is a Natural Language Processing (NLP) engine. When presented with a body of text, the Calais Web service returns the “named entities” (the categories to which the document’s key terms and concepts are assigned), facts, and events it discovers within the document. The relationships between these items are also identified and embedded in the results. Essentially, the results are the Semantic metadata of the document and can be thought of as the document’s “knowledge content,” which can be published and made available for searching and navigation.

On its own, and applied to one or two small, short documents, this might not seem terribly valuable. But deployed on the Web and made available as a free service, Calais is in a position to process massive amounts of data (text, quantitative, graphic, etc.) and extract their knowledge content. Once this task is complete, this content can be searched individually or combined with other content and searched in a larger context. This larger context can be based on other Web content, proprietary Thomson Reuters content, a combination of the two or the context of select data sources that may address a specific area of interest.

Ultimately, Calais’s goal is to be the world’s best tool for extracting the structure of any kind of content, recognizing its type, the concepts that are contained, their relationships, and doing so not just within a single file, but across a span of files that could be as large as the Web itself.

Key Differentiators:

Thomson Reuters, for one – the fact that Calais is sponsored by a global information giant suggests that this entrant will be with us for a long time. Furthermore, at this time Calais is in the final stages of testing its “infinite scalability” initiative, (e.g., cloud computing) designed to address growth in demand and/or spikes in utilization.

Another distinguishing characteristic is the rate at which the service has been adopted (the fact that it’s free is worth repeating). The net effect has been to discard the original projections for usage because demand has so vastly exceeded expectations. Note that until very recently, demand for Calais has existed almost entirely outside of any Thomson Reuters media property. This state of affairs is changing rapidly, with internal inquiries arriving with greater frequency.

Deploying Calais over the vast, professionally developed and controlled content in the Thomson Reuters empire would be a remarkable step in the company’s evolution. After 150 years as a traditional news wire service and publisher, Thomson Reuters’ content could quickly become something not yet fully defined, but possibly far more powerful and useful than what traditional publishers have offered before.

Six/Twelve Month Plans:

In addition to the continued internal roll-out of Calais, outside demand is moving beyond “experimenters” and creative small companies exploring this new service. Demand from large organizations, including well established publishers, is growing at an unexpectedly high rate. As a result, larger organizations or ventures built around Calais can expect to see availability backed by Service Level Agreements (SLAs). Special situations are also being anticipated where Calais is deployed on an enterprise scale behind a corporate firewall. Presently,
management doesn’t expect SLA services or direct licenses to reflect the bulk of the value that Calais will create. The key to this value may be contained in the Analysis section, below.

Analysis:

Let’s start with the premise that Thomson Reuters has 150 years of experience creating, managing, and presenting content that people want. Over this period, the company has amassed a body of high quality content that’s possibly the largest in the world. This content will continue to grow, but the advent of the Web has unleashed a torrent of content on a genuinely planetary scale. Since this content is outside Thomson Reuters’ editorial and/or production controls, the company considers it to be “wild” content. This doesn’t mean it’s bad – some of it’s exceedingly good.

Based on the environmental factors below, Calais puts Thomson Reuters in a position to extend its core competencies to include content it controls as well as wild content because:

- The fundamental nature of publishing and using content is changing.
- “World Wild Content” will dwarf the content Thomson Reuters controls.
- Professionally produced content will continue to command a premium.
- The Open Access movement and similar efforts by academics, researchers, and other content authors seeking to retain control of their work will continue and grow.
- Thomson Reuters has extensive experience in every aspect of the content industry.
- Flexible integration/interoperation of different types of content may provide powerful added value.

Calais is a free service that stands to significantly benefit people and organizations around the world. The terms of use may vary to allow users to retain ownership of their content’s metadata or not, but unless you’re a major publisher, this won’t be much of an issue. What matters, at least to Thomson Reuters, is that Calais is a very concrete step toward organizing and integrating the vast span of wild content with its own high quality content. Offering customers your own content combined with the very best of free, Web-based content in an easily searched, highly flexible and exceptionally expansive product is a strong competitive advantage that may ensure another 150 years of operation. This is the strategic thrust of The Calais Initiative.
Cambridge Semantics

Company: Cambridge Semantics, Inc.
URL: http://www.cambridgesemantics.com/
HQ: Boston MA, USA
Products (Primary): Anzo for Excel, Anzo Semantic Application Server, Anzo on the Web
Survey Respondents: Lee Feigenbaum
Vendor Category: Solution

Employees: 10
Revenue: --
Installed base: --

Primary Offering:

By using Anzo for Excel, anyone familiar with spreadsheets can immediately begin working with semantically linked and organized data. Cambridge Semantics’ approach differs from most of the industry by its use of Microsoft Excel as the primary user interface (UI). Anzo for Excel is an add-in that people can use to build and edit ontologies, link to other spreadsheets elsewhere on a network or out on the Web, and work with data at the source level, meaning that changes and updates are immediately available – all while observing appropriate permission levels. In this mode, Excel can also be used to query for and bring together legacy, non-semantic databases.

Providing record level access is a key point – this is an intrinsic quality of Semantic Web technology. Until now, the most likely way this would occur would be if a semantic application did the reading and writing. Alternatively, a highly qualified IT professional could use an ontology editing tool to manually make these changes. Anzo for Excel broadens the potential user base substantially and places the power of semantic technology in the hands of anyone who’s even just moderately competent with Excel.

For it to work, Anzo for Excel requires Cambridge Semantics’ Anzo Semantic Application Server. This product plays the role of establishing “semantic awareness” for the Excel add-in and then coordinates the interactions with corporate IT infrastructure, including spreadsheets found across the enterprise.

Key Differentiators:

By focusing on Excel, the global standard for spreadsheets, Cambridge Semantics has clearly separated itself from the pack by taking advantage of an extremely familiar user interface and readily tapping into the vast quantities of data stored in spreadsheets. Most Semantic Web vendors offer editing tools, many of which are based on Eclipse, the open source Java application development platform. Eclipse is powerful in its own right, but hardly the kind of application or environment that’s right for non-technical end users. Even after taking advantage of Eclipse’s UI flexibility, most editing/application tools offered today are meant for experienced, highly trained users who typically undergo further training expressly for the tool in question.

Cambridge Semantics eliminates this barrier to adoption by taking advantage of an interface that everyone knows and can work with. Learning the full capabilities offered by Anzo for Excel still takes time, so there’s still a meaningful learning curve before “power users” will be able to take full advantage of semantic technology. Nonetheless, there’s no other product available today that lets people get started so quickly.

Equally distinguishing is the company’s targeting of access to and integration of Excel data. Spreadsheet use varies from one enterprise to the other and by professional category as well. Even so, enterprises of all sizes have very valuable information stored in Excel files and it’s safe to say that from a corporate perspective, this information is loosely managed, if at all. Anzo for Excel’s ability to integrate Excel files and export results in standardized formats is a very intriguing rifle shot at a very common and potentially high value problem.

Six/Twelve Month Plans:

Incorporated in 2007, Cambridge Semantics is a young company compared to many vendors in
the Semantic industry. Its near term plans reflect a new level of maturity with the planned introduction of consistently packaged support and training offerings, along with consistent licensing terms. The company will also continue exploring partner relationships that include Independent Software Vendors (ISVs), system integrators, and technical partners. Scheduled for late 2008 is the general release of Anzo for Excel and finally, Cambridge Semantics will continue refining its sales and marketing messages which are gaining greater traction now that it has the Anzo for Excel story.

Analysis:

This is possibly the first platform-oriented Semantic Web platform company to escape from the tyranny of being a tools vendor and focus on a specific, high value application. Excel's ubiquity has been discussed, but the concept of "shadow IT" has not. Shadow IT refers to the parallel, uncontrolled and uncoordinated data stores that exist in every company that's implemented an enterprise solution for Enterprise Resource Planning (ERP), Sales Force Automation (SFA), Supply Chain Management (SCM), or just about anything else that requires enterprise-wide adoption. These unsanctioned data stores are often in the form of Excel files.

Company personnel may regularly perform their jobs using this "shadow" structure and the common currency of information may be contained in these files. If the formal IT structure is lucky, these users will update the "official" applications with what they create in Excel. In the real world, it's unlikely that end users will take the time to do so. In addition to providing a powerful form of Excel integration, Anzo for Excel is a real step forward in deploying a standards-based solution to coordinate and integrate these "unsanctioned" files with large-scale corporate solutions.
Dow Jones Client Solutions

Company: Dow Jones Client Solutions
URL: http://solutions.dowjones.com/djcs/
HQ: New York NY, USA
Products (Primary): Synaptica
Survey Respondents: Christine Connors
Vendor Category: Solution / Ontology / Search

Employees: --
Revenue: --
Installed base: --

Primary Offering:

Dow Jones Client Solutions (DJCS) offers a host of products and services and Synaptica marks the company’s entry to the Semantic space. Synaptica can be used to build and manage vocabularies, taxonomies, thesauruses and the inherent metadata of these structures. Environments that deploy Synaptica are usually enterprise-oriented and behind a corporate firewall. In these settings, customer goals might be to improve enterprise search results, standardize corporate libraries for compliance purposes, scope out the information that exists within the enterprise, or support the creation of the elusive “single source of truth”.

Synaptica has actually been in general release for 12 years (and acquired by Dow Jones less than three years ago) and during that time increasingly sophisticated and Semantic capabilities have been added, such as support for RDF, OWL, and SKOS, the first two of these being crucial W3C recommendations. Note that Dow Jones’ use of the term “taxonomy” may be an expediency to ease the introduction of concepts like ontologies, inferencing, and other “more Semantic” terminology.

Unlike most of the vendors in this space, Dow Jones has a well established sales team and obviously, the company’s success demonstrates they’re good at what they do. But in the meantime, these personnel have been trained to listen for hints from customers rather than taking a more formal, pitch-oriented approach to rolling out Synaptica’s capabilities.

Key Differentiators:

1) The luxury of a global installed base of active and loyal users compounded by 2) Dow Jones’ (and now News Corp.’s) ownership of high quality, professionally produced content. Unlike nearly every other entrant in the Semantic industry, Dow Jones could remain happily busy by steadily introducing Synaptica to each of its existing customers and likewise, gradually integrating these capabilities with the vast span content and various media channels owned by News Corp. The opportunities within News Corp. alone would make most Semantic vendors ecstatic if they occupied a similar almost-preferred-vendor status. If the DJCS team is industrious and inventive (and they certainly appear to be), they may well introduce innovative uses of Semantic technology within their corporate bounds and also, among the company’s extensive installed base. This final point is worth bearing in mind – managing the company’s vast span of content has given Dow Jones a very clear understanding of metadata and Semantics and these lessons could be quite valuable to the company’s customers.

Six/Twelve Month Plans:

DJCS plans to deal with basic issues related to integration with key technologies. While on its own this may seem modest, given the range of customers, content, and channels this group serves, this goal alone is a substantial IT task. So for now, there’s nothing glamorous on the horizon – but potentially, the groundwork will be set for widespread deployment of Synaptica and derivative products and services.

Analysis:

There’s no doubt that News Corp.’s acquisition of Dow Jones was a distraction. It’s possible that plans for Synaptica were disrupted or delayed, but now that the dust generated by the merger has settled, it looks like these plans will start
moving full speed ahead. The market for behind-the-firewall corporate applications supporting information identification, organization, and retrieval, is certainly ripe and Dow Jones’ position as an incumbent vendor will help its success greatly. Likewise, it’s easy to believe that opportunities abound within the overall News Corp. empire.

What’s very clear is that this is yet another media giant moving forward with an investment in the Semantic Web. Given Dow Jones/News Corp.’s track record of success, it’s likely the company will discover interesting and productive uses for Synaptica and everything this product spawns. If your Dow Jones account rep isn’t already talking about Synaptica, ask them about it – the results could be very interesting.
**Expert System**

Company: Expert System S.p.A.
URL: [http://www.expertsysten.net/](http://www.expertsysten.net/)
HQ: Modena, Italy
Products (Primary): Cogito
Survey Respondents: J. Brooke Aker
Vendor Category: NLP

Employees: 140+
Revenue: US$12M FY’07
Installed base: --

**Primary Offering:**

Cogito is a family of modular products oriented toward a range of Natural Language Processing (NLP) activities such as semantic search (e.g., richer than keyword-driven results), content categorization, sentiment monitoring, an intelligence platform that processes information from open (publicly available) sources and more. Essentially, a Cogito module exists for most, if not all, primary NLP requirements for information that might be contained in documents, blogs, email, Web sites, or just about any other electronic format that contains language based information.

With such a full arsenal of capabilities, Cogito could be used to process text-based information and knit the results together in an integrated, searchable “knowledge space”. If desired, the results and underlying documents could all be pulled together with an analyst-oriented module designed to maximize the interaction of machine-driven results with the judgment of an informed human expert. Setting aside the obvious applications within the intelligence community, a team of human experts equipped with the full range of Cogito capabilities could be a very potent force in research, discovery, and analysis in virtually any industry.

**Key Differentiators:**

Given the company’s sole focus on NLP and the fact that its been in business since 1989, there are a fair number of differentiators, starting with the fact that anyone who’s ever used Microsoft Office in Italian has likely come into contact with Cogito’s spelling and grammar checker. The company has an extensive customer list that cuts across many industries with a significant presences in telecom, media & advertising, and others, including branches of the Italian government.

What’s clear is that these solutions are well established in a range of environments that are behind firewalls, outside firewalls, as well as combinations of the two. The years of experience this company demonstrates, along with what it must have learned from its dealings with Microsoft suggest that these products have been sharpened and are ready for deployment in production environments.

**Six/Twelve Month Plans:**

Expert System in an NLP company through and through, so the natural progression dictates the coverage of more languages. Presently, Cogito works with English and Italian, but starting in 2008 this will expand to include French, German, and Arabic, with even more languages scheduled to be added later.

It also seems that the company’s interest in the US market is growing, which would add to the relatively small installed base in this market.

**Analysis:**

The continued adoption of the Semantic Web is presenting new opportunities to Expert System. The results provided by NLP in the form of themes, concepts, and the structures of documents and text-based files is exceedingly well suited to deployment in Semantic Web solutions. Cogito has been developed and refined for nearly twenty years and its installed base is proof that the technology works well. This level of success means that Expert System can focus on how its already proven semantic (as in linguistics) technology can play well on the Semantic (as in integration and interoperability) Web. The company is certainly making strides in this respect and we should all expect to see more of Expert System and Cogito.
Franz
Company: Franz, Inc.
URL: http://www.franz.com/
HQ: Oakland CA, USA
Products (Primary): Allegrograph, Allegro CL, AllegroCache
Survey Respondents: Jans Aasman, Craig Norvell
Vendor Category: Semantic Database

Employees: ~35
Revenue: <$10 Million
Installed base: n/a

Primary Offering:
Franz’s primary offering is Allegrograph, a Semantic database. Semantic databases are specialized to store the specific data structure (RDF, or Resource Description Framework) that much of the Semantic Web is all about. These databases are distinct from traditional relational databases (RDBs) and are also known as “Semantic stores”, “graph databases”, “RDF databases”, “triple stores”, or even “quad stores”.

Semantic databases are necessary and differ from RDBs for a variety of reasons, such as: RDBs require a clearly defined structure (schema) when they’re designed; Semantic DBs do not require a similar structure. This means that if new data types are encountered, adding them to a Semantic DB is relatively easy, while doing so with an RDBMS can be difficult, time consuming, and error prone. Secondly, in an RDB it’s difficult to create “webs” of relationships within the data being stored and it’s much easier to do so in a Semantic DB (in this context the term “web” is synonymous with the term “graph”). This second point is important because the webs in question are a key element of what makes the Semantic Web work. (Microsoft provides a good and very brief summary: http://msdn.microsoft.com/en-us/library/aa303446.aspx.) Also note that Semantic DBs are well suited to Social Network Analysis (SNA), which is of increasing interest in the corporate world as well as the social networking world.

Databases in general and Semantic databases in particular can become highly complex and technical subjects very quickly, which is not the point of this profile. Nonetheless, AllegroGraph has been tuned for the purpose of storing very large quantities of Semantic data and supporting complex queries. Franz’s unique “Activity Recognition” package combines Semantic data with Geospatial data, Temporal Logic, Social Networking Analytics to allow users to easily combine this information and perform queries that would be impossible with traditional RDBs. For example (supplied by Franz): “Find all meetings that happened yesterday within 5 miles of Berkeley that were attended by the most important person in Jans’ friends and friends of friends.”

One look at Franz’s customer list for Allegrograph shows that major commercial players such as BAE Systems, Boeing, Sun Microsystems, Raytheon, Merrill Lynch, GlaxoSmithKline, as well as their equivalents in government and academia confirm the value of this technology and the purpose it serves.

Key Differentiators:
As a database company, Franz is under constant pressure to improve its technology and maintain a competitive edge over alternatives that might be open source or offerings from traditional RDB vendors. Other key differentiators include AllegroGraph’s Activity Recognition package (mentioned above) along with support for “range queries” (queries executed over a specified range of data). Support for range queries has the benefit of maintaining the speed associated with an RDB along with all the benefits of a Semantic database. Obviously, these benefits will continue lowering barriers to adopting of Semantic technologies and serve as an important step toward the use of “linked data”.
Six/Twelve Month Plans:

Not surprisingly, Franz is tightly focused on AllegroGraph and has some very specific plans that include automatic indexing, concurrency, introduction of full transactional capability and more.

Analysis:

For the past several years, Franz has been succeeding with its exclusive focus on Allegrograph. In the process, the company has racked up quite a few impressive account wins, positioning Franz as the incumbent vendor in these accounts and leaving the company well-equipped with an armload of prestigious references. As a leading Semantic database vendor, Franz has delivered features and functionality that its customers need, want, and adopt. This track record and its plans going forward leave Franz well positioned to continue in its leadership position.

The traditional database industry giants have been slowly moving to adapt their offerings to support the requirements of Semantic solutions, which poses a major competitive threat to Franz. But for now, Franz offers one of the best solutions available for customers who recognize the value of a schema-less database which also supports the definition of specific relationships within the data being stored.
Mondeca

Company: Mondeca
URL: http://www.mondeca.com/
HQ: Paris, France
Products (Primary): ITM T3, ITM e-Catalog, ITM e-Knowledge, Semantic Portal
Survey Respondent: Jean Delahousse
Vendor Category: Platform

Employees: 18
Revenue: --
Installed base: 23

Primary Offering:

ITM T3 is a platform that integrates the development and management of controlled vocabularies, terminologies, taxonomies, metadata dictionaries, and thesauruses (ontologies.) This approach allows authoring and collaboration when working on large scale reference systems such as catalogs, knowledge bases that might be used for technical support purposes, and conventional thesauruses focused on specific topics such as tourism, medicine, or law. The solution supports multiple terminologies and terminology mapping.

ITM T3 is designed for organizations that have the expertise to work with well-established knowledge management practices and then output the results in XML, RDF, and SKOS (Simple Knowledge Organization System, a W3C pre-recommendation model for sharing and linking knowledge organization systems.) These output capabilities allow disparate approaches to information organization to be combined and in turn create a result that could not be achieved by any single approach. The input flexibility ITM T3 offers and the technology’s process of synthesizing output could be extremely valuable in the right environments.

Mondeca’s ITM e-Catalog and ITM e-Knowledge are extensions of ITM T3 tailored to suit specific applications and provide customers with a head start in creating their own solution. In the company’s words, Semantic Portal “…is an Internet site that offers a single point of access to a broad range of resources and services centered on a knowledge base.” In other words, a “semantically-enabled” mashup platform.

Key Differentiators:

Not surprisingly, European companies are well positioned to excel in multi-lingual support. In Mondeca’s case this capability is well proven by recently winning the contract to manage the European Union’s official thesauruses. The company has a historic focus on multi-lingual support which seems to be paying off handsomely with this prestigious win.

Other differentiators include pre-built ontology models, or templates, for specific industries such as tourism, medicine, and law. As with the company’s extensions of ITM T3 and Semantic Portal “widgets”, customers can get off to a fast start and produce results quickly (compared to starting from scratch.) These templates likely provide Mondeca with a competitive advantage in sales situations as well.

Six/Twelve Month Plans:

Mondeca plans to act on the growing number of large accounts that are aware of Semantic Web technology, reflected by the RFP’s (Requests for Proposal) the company receives. Additionally, the company’s recent win of the EU’s thesaurus project is likely to keep this team very busy. Other plans include the evaluation of Software as a Service (SaaS) to serve coalitions of smaller vertical industry players and as the need arises, international customers as well.

Analysis:

Mondeca was founded on the basis of first-hand experience in the financial services industry. The immense database relationship tables in this industry are difficult to build and their maintenance is extremely time consuming. Semantic Web technology was recognized as a promising approach, something which has been proven out by Mondeca and other companies.
Like many vendors, Mondeca finds that one of its primary competitors is market confusion. It’s a cliché, but customers care about the results that solutions deliver first and foremost. At the same time, customers want to be well positioned for future innovations. Solutions like Mondeca’s which deliver clear and demonstrable results that are grounded by W3C recommendations should reduce this confusion and the related unease.

Interestingly, Mondeca observes that W3C standards are distinctly more durable and effective than prior standards governing data and application integration. By drawing on the comprehensive body of standards-related experience in Natural Language Processing (NLP), Artificial Intelligence (AI), documentation and linguistics, Semantic Web standards benefit from existing rigor and the diversity that results from drawing upon such a wide range of experience.

Mondeca’s origins in real-world, complex data integration problems, its observance of rigorous forward-looking standards, and its success with major accounts like the European Union make this a vendor that has earned serious consideration.
ontoprise

Company: ontoprise GmbH
URL: http://www.ontoprise.de/
HQ: Karlsruhe, Germany
Survey Respondents: Andreas Nierlich, Joachim Redmer
Vendor Category: NLP / App Platform / Solution / Ontology

Employees: ~50
Revenue: --
Installed base: --

Primary Offering:

ontoprise relies on a combination of products and solutions for its success. On the product side, OntoBroker is the company’s primary offering and serves as the underlying platform upon which Semantic applications can be built and operated. OntoStudio is the company’s tool for actually building ontologies and creating models from these ontologies.

However, by offering solutions such as SemanticGuide, SemanticMiner, and Services for SemanticMediaWiki+, ontoprise takes a very significant step beyond selling a proven set of development tools for IT professionals. For example, SemanticGuide is the company’s solution for customer service organizations in need of a knowledge base. There are many advantages to an ontological approach to troubleshooting, a major one being the inherent flexibility in where and how the problem resolution process begins.

By taking advantage of how ontologies work, SemanticGuide doesn’t need to adhere to a strict, sequential process for problem diagnosis and remedy. Instead, each “fact” about a given incident can be presented and then expanded upon as further facts are either added or discovered. Steadily, as the model of a problem situation is built and refined, potential remedies begin to surface.

When supporting complex systems such as what might exist in a refinery, an integrated manufacturing facility in the automobile industry, or a paper mill, being able to rely on any set of facts as a starting point can lead to substantial time savings when trying to get operations back online. With this approach, support and service personnel may be able to avoid consuming conventional troubleshooting procedures that require a specific starting point followed by a rigid discovery process.

ontoprise’s other products and solutions are each tailored to emphasize different implementations, but SemanticGuide may be their best and most successful example of giving customers and end users a head start toward a practical, valuable Semantic deployment. (And the point is to solve a specific business problem, not deploy Semantic technology for its own sake.)

Key Differentiators:

ontoprise benefits from several distinguishing characteristics, such as the solutions mentioned above, an effective partnership strategy including a partnership with Silicon Graphics, Inc. (SGI), blue chip account wins at companies like Alstom, Audi AG, KUKA Robot Group, Vulcan, Inc., and a well developed consultative sales approach that uses plain language to discuss the business benefits of Semantic Web technology. These differentiators are a result of a fundamentally solid family of products and solutions that has grown and improved since the company’s founding in 1999. Note that the SGI partnership provides ready access to high-performance hardware, if a customer has the need.

Perhaps a simpler description of ontoprise’s key differentiators would be that the company has a track record of success in major accounts. In the Semantic Web industry, that’s saying quite a lot.

Six/Twelve Month Plans:

ontoprise has been distilling its formula for success for awhile, so radical changes are
unlikely. Instead, the company’s plans focus on accelerating its business performance in its core markets of Germany and more broadly, Europe. The measure of success the company has set for itself is to “grow faster than the market”. A more specific definition wasn’t requested, but regardless, ontoprise clearly has every intention of expanding its business, which is further supported by the next point: ontoprise is already getting “substantial” revenue from the US and plans are in place to open a subsidiary in this market. The company’s reliance on marketing and development partners probably means that the technologists will remain in Karlsruhe, but a US business presence would make market penetration and expansion much easier to control and manage. Interestingly, ontoprise reports that US-based customer inquiries now reflect a better understanding of the Semantic Web than typical European inquiries.

**Analysis:**

ontoprise has the markings of a successful company because among other things, the management team has learned how to build good technology, shape solutions that address well defined business problems, identify and qualify productive partners, and then articulate the value of its products and solutions in clear business terms.

Addressing each of these points, the evidence for the quality and reliability of the ontoprise's products can be found in its customer list. Consistently winning major accounts is a clear indicator of a company’s product set and related, its ability to sell. By developing solutions geared toward specific business practices (like customer service or knowledge bases), ontoprise again sets itself up for success by solving a specific, well defined problem. Furthermore, its partnership strategy is a key element in providing clear cut domain expertise and account relationships, although the company readily admits that its lessons in developing these relationships were hard-earned. Finally, in sales conversations, ontoprise has clearly succeeded in moving away from using technical acronyms and language. For decision makers at customer organizations, this must be a huge relief.

On their own, each of the points above probably wouldn’t be enough to make any company successful. But ontoprise is successfully demonstrating that by executing on many fronts better than its competitors, the account wins will start to add up.
Ontos

Company: Ontos AG (Research & Development and Intellectual Property) and Ontos International AG (Sales and Marketing, Operations)

URL: http://www.ontos.com/
HQ: Nidau, Switzerland

Products (Primary): OntosMiner, LightOntos, Ontos SOA, TAIS Ontos

Survey Respondents: Daniel Hladky
Vendor Category: NLP and Ontology

Employees: 35
Revenue: --
Installed base: >30

Primary Offering:

OntosMiner reflects the company’s focus on Natural Language Processing (NLP) enhanced by ontological and multilingual processing. Essentially, OntosMiner will analyze an article, a book, or some equivalent body of text and through well-established NLP procedures, the key subjects, concepts, and relationships will be extracted and summarized in list form. These results can be further processed by LightOntos, an ontology-based analysis tool. The results can be further edited and expanded by LightOntos to produce even more accurate representations of a document’s central concepts.

Ontos SOA rounds out the company’s offerings by providing wide-ranging retrieval of data and the subsequent storage and distribution of results. It consists of several components for crawling, aggregating, and merging. The company presently has a patent pending on a method for semantic navigation and web services as well as a new method for storing large quantities of Semantically formatted data.

Key Differentiators:

Ontos distinguishes itself by offering “tuned” ontologies for specific areas like business (economics) and politics, particularly for the support of Web-based news portals. As a matter of fact, since late 2007 Ontos has been powering a news portal presently in development. The publisher’s intention is to present content that’s automatically linked by Ontos technology and increase the likelihood that viewers will read more articles and remain on the site longer. This in turn stands to increase advertising revenue. Calculating the return on investment (ROI) is relatively simple – the real question becomes how optimistic the projected “lift” (increase in viewership, response, and/or revenue) should be.

Regardless, this approach takes any required learning curve away from the general viewing audience and gives it to the portal managers. Ontos can tailor its implementation to the needs of these managers and provide an interface suitable for use by non-linguists. Furthermore, the company can provide “Web widgets” to ease information extraction and embedding within applications.

Another distinction is the company’s experience in law enforcement. TAIS Ontos was created in 2001 and is used to power Criminal Intelligence Systems (CIS). This solution can span numerous disparate data sources and support the exploration of relationships, all within the context of criminals, their activity, and their movement within the judicial system. The company’s visualization component helps analysts explore previously unknown links in a variety of formats for viewing the data.

Six/Twelve Month Plans:

Ontos has ambitious plans to expand its business, starting with the launch of an API platform for named entities and widgets, which would serve as an enhancement to the company’s NLP and ontology capabilities. Other planned applications of NLP technology include the creation of a “Competitive Dashboard” which would assess sentiment within specified message forums and separately, the company may launch an NLP plug-in for Microsoft Outlook. The company is also working on an idea it calls “semantic bookmarking.” Finally, like other European companies in the Semantic Web space,
Ontos may open a US-based office to better respond to North American demand for its products.

**Analysis:**

Ontos is using NLP/ontology based solutions to target specific opportunities such as news portals and publishers, law enforcement, and interestingly, the company is emphasizing the Russian market as well. Ontos has been thriving by providing closely tailored solutions, “tuned” ontologies, not to mention its continued support of TAIS Ontos, which is geared toward law enforcement – a perennially reliable customer. By making thoughtful choices in pursuing opportunities and nurturing strengths in its traditional markets, Ontos has crafted competitive, tailored solutions and is steadily gaining experience and valuable domain expertise.
OpenLink Software

Company: OpenLink Software
URL: http://www.openlinksw.com/
HQ: Burlington MA, USA
Products (Primary): Virtuoso, Universal Data Access Drivers, Data Spaces
Survey Respondent: Kingsley Idehen
Vendor Category: Platform

Employees: --
Revenue: <$10 Million
Installed base: --

Primary Offering:

Virtuoso is OpenLink’s primary Semantic Web product and it provides a “virtual” layer above actual data stores and sources. On one side, disparate data stores, data formats, as well as internal and external sources (from an enterprise perspective) connect to the virtual layer. On the other side, consistently and predictably formatted access to the data is provided for consumption by applications, other data stores, and/or end users. The effective elimination of barriers to working with combined sets of data allows processing to occur across integrated bodies of data that might otherwise only exist after time consuming and error prone procedures found in traditional Information Technology (IT).

Consumers of these data retain control over the range of data employed for a particular purpose (the entire body of data doesn’t have to be called) and this control allows for retrieval down to the record level. As a result, unique bodies of data can be assembled and processed, which in turn have the potential to produce equally unique and valuable results. These uniquely assembled (or called) bodies of data are identified by an equally unique URI (Uniform Resource Identifier) and this is where the concept of “linked data” begins to take shape (the subject of “linked data” is large, complex, and outside the scope of this report.) Note, data can be written back to the original stores in addition to being read.

Virtuoso is W3C standards compliant and ready for deployment in production environments. OpenLink has also created Data Explorer 0.19, a Firefox add-on that enables exploration of linked data on the Web sites that provide this access.

Key Differentiators:

Virtuoso is distinguished by its public availability through high profile (Semantic Web-wise) sites that include DBpedia, where it supports the project’s RDF data set and provides Semantic query access. Other sites powered by Virtuoso include Science Commons and Bio2RDF.org. On this basis, potential customers can explore the value Virtuoso offers, verify the technology’s reliability and readiness for production environments, and network with other users to learn from their experience and the value they’ve gained in the process.

Six/Twelve Month Plans:

With Virtuoso extending its track record each day it continues to support the growing projects mentioned above, OpenLink plans to expand its efforts to spread the word in the enterprise market. In particular, the company’s messaging is likely to evolve from discussing a virtual layer to that of a “conceptual” layer, which will play to Virtuoso’s strength.

For example, managers within an enterprise think of a “customer” as having many different characteristics, and these may be described by the data stored in applications for CRM, ERP, general ledger or accounting systems, etc. At the human level, a customer is thought of as a concept and an enterprise’s body of data provides a description, or model of a customer. This model doesn’t capture every imaginable facet of a customer, but the intention is to capture the important things like what products have been purchased, key contacts, payment history, shipping, and perhaps inventory requirements.

By presenting such a conceptual view and allowing managers to drill down into the source data (since it’s intrinsically linked to the view)
managers stand to have access to information that cannot be easily matched (or at least not on economic terms) by traditional IT solutions.

Analysis:

OpenLink Software's history lies in traditional approaches to data integration based on JDBC, ODBC, and other well established connection schemes, so it shouldn’t be a surprise that the company is pursuing the “next level” of data integration and access. From a decision perspective, this background provides OpenLink with a sound fundamental business but at least as importantly, it provides solid evidence that the company has extensive experience solving real-world problems.

Virtuoso doesn’t have a readily identifiable competitor – only substitutes, which might come in the form of Business Intelligence products or other, proprietary mashup technologies. But by adopting Virtuoso as their platform, DBpedia, Science Commons, and others are providing a very credible endorsement that potential customers would do well to take note of.
Primal Fusion

Company: Primal Fusion, Inc.
URL: http://www.primalfusion.com/
HQ: Waterloo Ontario, Canada
Products (Primary): primalfusion.com
Survey Respondents: Peter Sweeney
Vendor Category: Thought Networking

Primal Fusion, presently in closed alpha testing, is building a “thought networking” service for consumers. This service is envisioned as a mechanism to help people make their thinking tangible and take it online.

Starting with a simple word or phrase to describe a thought, a person might enter “itchy eyes, runny nose”, and then Primal Fusion would return related concepts or ideas. These ideas could be reviewed and added to the original “thought” at the user’s discretion. Iterating this process would yield an increasingly sharper and clearer thought based on elements drawn from the interaction with Primal Fusion. (Note that the thoughts generated by Primal Fusion are synthesized from existing “knowledge stores” like Wikipedia.)

Let’s say that the original thought has grown and is now something like: “red, itchy eyes, runny nose, headache, low energy.” Presumably, this person has been reviewing their related results and adding relevant terms, concepts, or new thoughts to their original thought. As a result, they now have a much more accurate description of their condition. With Primal Fusion this description could be saved and re-deployed against other knowledge bases. A “traditional” Semantic Web professional would recognize the thought that’s been developed as being equivalent to a “model” that might be more formally developed using purpose-built tools.

Now imagine deploying this thought as a query over a specific knowledge store. Let’s assume Wikipedia was used to expand and refine the original thought. After learning this thought might be about allergies, our knowledge seeker might want to query other open domains (such as the National Library of Medicine, if it was available) to learn about treatments for their condition.

At this point, the refined thought would produce related thoughts from the National Library of Medicine (NLM) pertaining to the individual’s unique perspective, experience, and condition. Going further, the NLM results could be added to this thought to create an evermore personalized description of the original thought that began the whole process.

Now imagine this kind of process being applied to activities ranging from “how do I frame my basement” to “what are the limits to using Bose-Einstein condensates to turn light into matter and then back again?” If successful, Primal Fusion stands to change the traditional research process into one that’s more interactive, richer, productive, and efficient. This scenario could eventually turn computers and online knowledge stores into collaborators in the learning and discovery process instead of simply serving as enabling mechanisms or support devices.

Primal Fusion’s plans to operate as a business include a consumer-facing Web service supported by advertising and a more developer-oriented fee based service for third-party applications. The latter business may or may not include a “freemium” component, where fees are assessed for activity above a specified level. The company’s service is meant to be general-purpose and segmented by activity, not necessarily a traditional model of “this product for that market”.

Behind the scenes, Primal Fusion’s venture is based on proprietary technologies for Semantic analysis and synthesis. According to the company, great care has been taken to address the user experience, so it will be interesting to see their approach to something abstract like information discovery and knowledge synthesis. Initial steps are likely to be small, but Primal Fusion certainly seems like a venture worth watching.
Saltlux

Company: Saltlux, Inc.
URL: http://www.saltlux.com/
HQ: Seoul, South Korea
Products (Primary): IN2 Product Suite
Survey Respondents: Jun Hyung Ahn, Lucy Cho, Tony Lee
Vendor Category: Search/NLP/Ontology

Employees: 77
Revenue: --
Installed base: --

Primary Offering:

IN2 (pronounced “in two”) is the company’s lead product and it serves as a development platform and operating environment for Saltlux’s capabilities in NLP, ontology integration, and semantic search. A possible scenario might be a document- and data-rich environment with some amount of metadata available, most likely found in the more data intensive stores.

IN2 could be applied to the document stores and through its text mining/NLP function, extract initial metadata for integration in the larger context of metadata. This integrated metadata could be maintained in a suitable data format for processing by an ontological model for exploration of the “information space” and potentially, discovery of previously unknown relationships.

At any point, Saltlux’s semantic search capability could be applied to assemble and present concepts and findings that are related to a given term. These results would likely become richer as the information space expands through further integration of metadata, NLP results, and further ontological modeling development. In complex, discovery-oriented environments found in research activities, this process of metadata extraction, integration, model assembly and exploration could be quite valuable. Complex, process oriented environments could also benefit from a structured discovery process and the clearer understanding that this approach may provide.

In short, Saltlux provides a sound starting point for developing and delivering a knowledge-oriented Semantic Web solution. As is common in the industry, the company relies on components from solid partners to fill in some key functional areas and as an integrated whole, the result stands to be a very powerful IT integration, modeling, and discovery capability.

Key Differentiators:

Saltlux is in Korea, for one thing. In the larger sense, Asia has all the same IT integration and legacy problems found anywhere else in the world. Saltlux is one of the very few companies in this region that has emerged as a focused, viable vendor of Semantic Web solutions. As a result, this company has ready access to opportunities throughout the region, including countries with known interest in this technology such as China, Japan, and Singapore. Aside from any unique cultural considerations, Saltlux’s strength in NLP for Asian languages likely gives the company easier access and greater credibility as well.

Additionally, Saltlux has demonstrated an aggressive approach to partnering with leading Semantic vendors like Franz, Inc. and ontoprise GmbH to round out its offerings. These factors combined with the backing of Saltlux’s established IT business in other non-Semantic disciplines means this company may emerge as a solid contender for the long term.

Six/Twelve Month Plans:

Not happy simply to pursue opportunities in and around some of the world’s most dynamic economies, Saltlux has three specific product initiatives, which are the development of a customizable mobile and ubiquitous service system, a B2C Semantic search system with topic clustering, and the release of a search appliance for the SMB (Small to Medium sized Business) market. Each one of these poses an interesting expansion on delivering Semantic technology to a broader audience possessing a variety of needs.
**Analysis:**

Saltlux benefits from the foundation provided by its business in search solutions, mining, and consulting while it pursues its Semantic Web venture. As a result, it’s likely the company can allow itself more time to get established and build a reputation. As a matter of fact, this approach seems to be working very nicely for Saltlux given account wins that include LG Electronics, SK Telecom, Korea Telecom, and different branches of the Korean government.

As a region, Asia may be less penetrated by vendors although customer awareness of the Semantic Web appears significant. The experience Saltlux is gaining through its success to date, combined with what it learns from its aggressive approach to partnering may place the company in a leadership position in its part of the world for quite some time to come.
Sindice

Company: Sindice
URL: http://sindice.com/
HQ: Galway, Ireland
Products (Primary): Sindice.com
Vendor Category: Microformat data aggregation service

Sindice offers a set of Application Programming Interfaces (APIs) which make it easy to locate and access data published to the Web. Increasingly, “microformats” and RDF (lightweight formats for publishing small amounts of data) are being used by Web sites to publish snippets of relevant data. These formats are an evolving framework identifying and presenting relatively simple concepts like contact information, calendar and/or event information in a way that both humans and computers can recognize and use. Similarly, as a result of being embedded in Web pages using the RDFa specification, RDF in general is becoming increasingly popular as a format for expressing metadata.

Structured information collected this way is made available for reuse by application developers, including Web site developers and content managers. As a result, data from thousands of locations on the Web are standardized by Sindice. In turn, Sindice can be queried for events, people, locations, products, and more in a uniform fashion, regardless of the point of origin or the original Semantic format used.

The data made available via Sindice are exposed directly by their publishers and can usually be integrated directly with the results end users receive (of a query, for example). Or, the functionality of a Web application can be enhanced to provide additional information about an event, a person, and eventually, other similarly tightly defined concepts.

Sindice is targeting developers as its audience and as an early stage beta, the venture is seeking to balance offering powerful APIs with APIs that are easy to use by non-Web developers. Five months after a soft launch, Sindice is experiencing roughly 10,000 queries per day via its API. Current plans call for a more formal launch in late 2008, when a more comprehensive microformat search service API will be offered.

Sindice is a research project out of DERI (Digital Enterprise Research Institute) in Galway, Ireland and funding for the next three years is expected from the European Union’s Okkam and Romulus projects and from Science Foundation Ireland.
Thetus

Company: Thetus Corp.
URL: http://www.thetus.com/
HQ: Portland OR, USA
Products (Primary): Thetus Publisher, Tupolo
Survey Respondents: Lisa Epstein, Philip Pridmore-Brown
Vendor Category: Application Platform / Modeling Solution

Employees: 25
Revenue: --
Installed base: --

Primary Offering:

Publisher is Thetus Corp.’s lead product and its typical customer can be found at the center of the information storm that seems to grow larger with every passing day. Personnel charged with finding their way through this turbulent environment may have to tap a range of data sources, some of which may be unstructured or in disparate formats. Publisher shows its strengths by connecting to these sources, supporting the development of models (a set of interconnected concepts, rules, and properties), and aiding the process of pattern discovery. Supporting this “discovery process” is important – in the course of their research or analysis, end users might uncover useful (and unexpected) information which relates to the original task. Being able to capture these relationships and place them accurately within the larger context is a very valuable capability in an information/knowledge intensive environment.

Publisher is also a high performance server capable of managing multiple ontologies for the purpose of building and testing models. Thetus finds its products supporting goal modeling, risk modeling, and other forms of data modeling which may lend themselves to discovering patterns in data. The company’s modeling experience fits well with its competency in developing end user-oriented views. These views can be tailored by end users, a feature which can aid the process of comprehension and discovery.

Note that Publisher also offers workflow components and a search capability. The workflows can be set to take action when certain pre-specified conditions exist and the search function supports the creation of query models that can be deployed across a model to seek out inferences, or relationships not explicitly defined when the data were originally loaded.

Key Differentiators:

It’s common knowledge in the industry that intelligence agencies around the world have found the Semantic Web to be a very valuable technology. In-Q-Tel, the Central Intelligence Agency’s (CIA) venture capital group, has added emphasis to the value this technology represents by entering into a funding agreement with Thetus. Ordinarily, a funding source wouldn’t garner much attention, but this situation varies from the norm – not only does Thetus get funding from the deepest imaginable pockets, the company is uncommonly well positioned to deploy its technology in high stakes, leading edge environments. As expected, Thetus can’t discuss these initiatives, but the experience the company gains in the course of intelligence oriented work would likely be quite valuable in many other environments as well.

In December, 2006, the company closed a $3.5 million round of funding with Nth Power as lead investor. This foray into the commercial sphere of VC funding may have been a contributing factor to Thetus’s new customer wins in the construction engineering and enterprise program management industries during the first half of 2008.

Six/Twelve Month Plans:

Thetus has established a clear strength in finding patterns contained across a range of disparate data stores and then building models that describe these patterns and support further investigation. The company’s plans call for branching out to include resource modeling
(there’s a difference), while integration with enterprise IT infrastructures and collaboration tools such as SharePoint and Clearspace will be another focus of the company. In the latter respect, Thetus will be working to expose the Semantic capabilities of its products to make their integration easier and more comprehensive.

**Analysis:**

On the face of it, Thetus may strike the casual observer as being like many Semantic technology vendors, that is, a company with a good platform, good modeling tools, and possibly very good visualization tools. But there are questions that can’t be avoided, such as what the company’s experience in the intelligence community has taught it about integration, reliability, and scalability – key considerations in any major production environment and scarcely unique to the intelligence business. Furthermore, given the global span of data that intelligence analysts deal with, what are the important lessons that Thetus has under its belt, and how can these benefit commercial customers?

Thetus certainly seems to be a successful classified government contractor. Success in that market isn’t always easy to transfer to commercial markets, but that doesn’t necessarily correlate to the quality or capability of the technology. The company has a clear interest in winning commercial customers and their success suggests that Thetus’s message is one that deserves close attention.
TopQuadrant

Company: TopQuadrant, Inc.
URL: http://www.topquadrant.com/
HQ: Alexandria VA, USA
Products (Primary): TopBraid Suite
Survey Respondents: Irene Polikoff, Robert Coyne
Vendor Category: Application Platform/Ontology

Employees: --
Revenue: --
Installed base: 500+

Primary Offering:

TopQuadrant leads with TopBraid Suite, an integrated platform comprising TopBraid Composer, TopBraid Live, and TopBraid Ensemble. Using TopBraid, customers can integrate data, develop and deploy Semantic applications and infrastructure, as well as create applications that process data that have been linked or “Semantically combined”.

Looking closer, Composer is a full strength ontology development tool and supports modeling, application development, data source configuration and more. As a server platform, Live is used to deploy Semantic applications, mashups, and in general, any of the solutions developed with Composer. Ensemble is a collection of out-of-the-box, configurable user interface components. With these components, developers can quickly build Semantic applications that end users can use to view and interact with rich, connected collections of information. The net effect is that TopBraid is flexible enough to be used as a content management system, wiki, and with add-ins it can support faceted search, calendaring, maps, timelines, and charts and reports created with BIRT (Business Intelligence and Reporting Tools, a suite of open source Business Intelligence tools).

TopBraid Suite has an open architecture and integrates with best of breed third party reasoners and Semantic databases such as Oracle 11g and AllegroGraph. This suite is well suited to companies investigating the practicality and value of deploying enterprise applications of Semantic Web technology. The three components of TopBraid target corporations with any level of sophistication – whether there’s an interest in starting small with a proof of concept, moving to a pilot, or beyond into the range of enterprise applications.

Training on the fundamentals of Semantic technology and its range of products is a key element of TopQuadrant’s global reputation. TopQuadrant almost certainly holds more publicly accessible training courses than any other Semantic Web vendor. Additionally, customer-tailored training programs are regularly designed for any level of expertise in the current state of the art, a clear indication of one of the company’s central values: To be the thought leader in the Semantic Web industry.

Key Differentiators:

In addition to training, TopQuadrant offers unique capabilities in its toolset, such as SPARQLMotion, which is geared toward script developers as a higher level graphic scripting language. With SPARQLMotion, script writers can connect a series of automated, pre-defined routines (which can also be user-written) in a way that resembles Yahoo! Pipes or OSX’s “Automator” function. As a higher level language, SPARQLMotion allows a larger team to participate in the development and maintenance of solutions created by the lower level tools found in TopBraid Suite.

SPARQL, which supports federated query (querying multiple, disparate databases) is becoming a key area of focus for the company. This W3C recommendation plays a central role in integrating legacy data stores as well as contemporary Semantic databases. Aside from integration benefits, customers can use this capability to develop a solution which taps a given database and then re-use the solution to process a different data store with little or no modification to the code.
Six/ Twelve Month Plans:

TopQuadrant is in execution mode – they’ve made their plans and now they’re focused on “linked data exploration” (uniquely enabled by SPARQL), Semantically-enabled content management, and enterprise architecture in a few different industries. The company’s solution areas of focus are based on customer demand – TopQuadrant is one of the very few Semantic Web vendors with an extensive installed base. Composer’s low price point and a self-serve purchase process on topquadrant.com has attracted a range of buyers, many of whom are now ready to take the next step in the deployment of their Semantic efforts.

Analysis:

TopQuadrant is well known in the Semantic Web industry for the quality of its products, its large customer base, the global stature of its personnel, and the remarkable goodwill the company has generated through its numerous annual training events and conference presentations. The time is right to focus on specific applications of Semantic technology and the company is committed to moving forward on the basis of its expertise in SPARQL, ontologies, and Semantic application development.

For customers who want to get started but need help in scoping out a development roadmap, TopQuadrant’s experience provides a very clear understanding of and support for the life cycle of Semantic projects. The company is well positioned to work at any scale within an enterprise ranging from an initial demo to a full blown production environment application deployed across the enterprise. In turn, enterprise customers can be confident they’ll be working with proven tools, backed by perhaps the industry’s strongest track record for training and support. Global IT vendors of all kinds thrive on a combination of products and services, and there’s every reason to believe this formula for success will work for TopQuadrant as well.
Twine.com

Company: Twine
URL: http://www.twine.com
HQ: San Francisco CA, USA
Products (Primary): Twine.com
Survey Respondents: Nova Spivack
Vendor Category: Interest Network

Twine is an interest network designed to let people share links, comments, files, and more about topics they’re interested in. When Twine launched, it appeared to be an interest site for people involved with the Semantic Web. Since then the range of interests has grown rapidly and now a quick look at the Top 100 Twines show interests as diverse as Green Business and Investing, Science Discoveries, Geopolitics, Sustainable Living and there are thousands more not on the list, all started by individuals who have an interest in these topics.

While the company is easily one of the best known Semantic ventures today, what’s refreshing about Twine is that it emphasizes what it does (its business mission) and not how it does it (Semantic Web technology). Longstanding members of the Semantic Web community may be left wondering “where’s the beef”, because there aren’t any ontology editing screens, model visualizers, or application development environments. On the other hand, the general public may come to believe they’ve finally “found the beef” and it’s on Twine.

The reason for this disparity in perception are simple – Twine was made to support people who have interests they’d like to share. If you’re a Semantic developer, you’ll probably find other like minded developers that you can share ideas with, but if Twine is an evolutionary step beyond Facebook or MySpace, hands-on development features won’t be part of the experience.

On the other hand, if you’ve started a twine on cooking, this may be perfectly OK with you. In fact, you may not care about the technology at all. Instead, you may be far more interested in the bookmark someone just posted to your cooking twine that leads to a recipe you’ve never thought of before.

Visible or not, there are still a lot of Semantics at work under the hood in the form of auto-tagging, Natural Language Processing (NLP), and linked data. But the technology and the messages associated with it have been subsumed by Twine’s business goals and its point for existing in the first place: To be a money making venture.

On the topic of money, Twine’s audience demographics and their behavior may position the company front and center as a viable media property. For instance, Twine’s target demographic is young professionals. To date, the audience is older (which usually means more affluent) than what’s found on Facebook or MySpace, where advertisers have been frustrated in crafting effective campaigns.

Compared to “discovery” sites like Delicious, Digg, or Technorati, where visitors may linger for two minutes, Twiners remain on the site for fifteen minutes. In the world of advertising that’s a mind-boggling jump that represents a highly motivating business opportunity, particularly when these visitors are deeply engaged in an interest they hold dear (and that interest that can be identified, quantified, targeted, and served).

Whether in the form of cash profits or equity appreciation, Twine is a business that sees an opportunity to use Semantic technology in a way that other technologies can’t easily replicate, if at all. This means that Semantic technology is providing Twine with a competitive advantage in two critical processes: 1) Developing a valuable audience; and 2) providing advertisers with a highly targeted, systematic way of reaching this audience. The fact that Twine launched its invitation-only beta in October ‘07 means it will have a one year lead (or more) before another genuine competitor comes along. That’s a good cushion.

Vendors in this report should be paying attention – if there’s an upswell in Semantically-based social networks, e-commerce sites, or other viable ventures, it’s a market that they should be selling into – hard.
Yahoo! SearchMonkey

Company: Yahoo!
URL: http://developer.yahoo.com/searchmonkey/
HQ: Sunnyvale CA, USA
Products (Primary): SearchMonkey
Survey Respondents: Amit Kumar
Vendor Category: Search

The search industry is serious business and anything that can provide a competitive edge is fair game. SearchMonkey is Yahoo!’s opening shot at using Semantic technology (RDFa, eRDF, and microformats) to produce a search experience that sways market share in their favor. The use of microformats is an evolving practice, but a simple example would be annotating a Web page so that computers can recognize that a string of numbers is actually a phone number or a date.

A scenario might be the publisher of a Web site that sells concert tickets embedding microformats that deliver telephone numbers and a running count of tickets remaining to a concert. Someone searching for tickets to this concert would see the site’s phone number and the remaining ticket count in their search results, eliminating the need to navigate to the ticket seller’s site to find the same information.

In this case, the site owner gets a jump on any competing ticket sellers and Yahoo! can claim they’re offering a competitive tool. Consumers of search results win because they stand to get essential information presented on a single page and not distributed across several sites. Even if all ticket sites in the example above used microformats to deliver information, all parties would still win because of the quick comparisons this would make possible.

SearchMonkey is a key element in what the Yahoo! calls its “Yahoo! Open Strategy” (Y!OS), which is an effort to build a community of developers and publishers for their search platform. In turn, the company hopes that search consumers will find the experience compelling enough to start submitting more and more searches through Yahoo!

SearchMonkey presently delivers enhanced results for Yelp, LinkedIn, StumbleUpon, and a range of other sites (found here: http://gallery.search.yahoo.com/). It will be extremely interesting to see what effect this enhancement to search results has – not just in terms of Yahoo!’s share of the search market, but how directly the company’s business results can be attributed to this variation on Semantic technology.