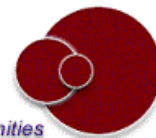


Newsletter

Nº. 7, January - March 2004

KTweb

Connecting Knowledge Technology Communities



Welcome to the seventh issue of the **KTweb** Newsletter, providing an overview of selected news and events from the world of knowledge technologies over the preceding quarter, together with information and commentary on specific issues and major events.

PERSPECTIVES

The battle for Web search

Looking back over earlier Newsletters, a good number of these lead articles focus on the world of Web search. And the simple reason is that this is where the action is happening. Many other knowledge technologies that we cover are essentially still in development, however long they may have been around.

Agent technologies and Semantic Web are having a fairly long gestation - research goes on, the product market develops slowly, standards develop even more slowly. But like a chrysalis, these technologies are still developing and have yet to emerge fully grown onto the market.

Two areas notable for their much greater visibility are Web Services and Web search. And it is fairly clear why - corporate interest. No implied criticism here of the business world; simply that where potential for business benefit or profit is seen, things tend to move faster and get done faster. Businesses take an interest in such technologies, and software publishers and service providers are quick to enter the market.

There is, however, a difference between Web Services and Web search. Web Services depends on a degree of co-operation between vendors. Their products must interoperate, and so standards and protocols must be developed and agreed, and this takes time. Even so, not everyone is always in agreement with everyone else - read about the perhaps non-complementary specifications proposed in this last quarter by two different groups of manufacturers - and this again does not help progress.

Web search is different. A Web search product does not really depend on anyone else's. Build a better search, and the world will beat a path to your web site. Take a look at Google. Unknown just a few years ago, and now it's almost a part of the English language. If you can deliver a great search service, people will use it, and they will read your sponsored results too, and profits will start to arrive.

Web search is one of the most lucrative services on the Web, despite being essentially free to its users. Google, Overture and Yahoo have shown this, and now others want a slice of the action. Perhaps not always motivated purely by profit, but the market is getting more crowded nevertheless.

A number of the newer arrivals are discussed on the next page. These companies are typically offering new ways of navigating your search, by automatic classification or visualisation of results. Other new facilities are appearing, such as personalisation, and its cousin localisation; now searches would be able to understand better your own personal preferences and interests, or would limit your search for "pizza delivery" to the area where you live. All of these techniques have been around for a while, and in Newsletter Nº. 4 we noted the growing level of interest in them.

IN THIS ISSUE

PERSPECTIVES	1
IN THIS ISSUE	1
The battle for Web search	1
NEWS	2
OVERVIEW	2
Information Retrieval	2
Web Services	3
Grid computing	3
Semantic Web	3
XML	3
FP6 and the research domain	4
STANDARDS UPDATE	6
Web services	6
Semantic Web	6
XML	6
EVENTS	6
KTweb UPDATE	6

But the moves by the big players are getting the attention. The Web search leaders are the likes of Google, Yahoo and AskJeeves. Old favourites such as Altavista, Excite, HotBot, and Lycos have been bought, merged, changed into portals, or otherwise have lost their place in the Web search world.

We also reported in Newsletter Nº. 4 on Microsoft's moves into search. More news is now emerging - about search technology to be embedded in Longhorn (the next major version of Windows), and about overall strategy. Search is the one key area in which Microsoft not only does not dominate, but also hardly has a toe in the water. This will change. In March, the company took the unusual step of admitting that they missed out. CEO Steve Ballmer said: "People say that Microsoft does it all, but this is the case where we didn't do it all." First-generation technology from Microsoft will likely appear in the next 12 months. Exactly what will appear is hard to predict, but history shows that when Microsoft puts its mind to anything, it usually delivers something significant.

And meanwhile, a surprising new player has emerged. It's still in test mode, but Amazon's A9 search engine has the industry talking. Although powered by Google, A9 is heavy on personalisation, helping users to tailor their searches, and it has a number of other features such as annotation of pages from search results and a full search history.

Amazon is unlikely to want to face up to the giants like Google and - maybe soon - Microsoft; its interest is e-commerce and A9 is designed to support that. However, Chris Winfield, president of search-engine marketing firm 10E20 notes: "Search is the biggest thing online besides shopping and e-mail" and we will see some battles in the coming months. ●

OVERVIEW

Looking over the News items that we reported in the first quarter of 2004, information retrieval seemed to predominate, and the bulk of that concerned Web search, as reviewed in Perspectives in this issue.

There were also a good number of announcements of Calls for Papers - an average of six per month, so perhaps we are seeing here the kick-off of this year's conference season.

We saw a mixture of miscellaneous items - Bill Gates receives an honorary knighthood from Britain, Tim Berners-Lee receives a real one (the benefit, if you like, of being British). On a more sombre note, George Pake, who founded Xerox's famous Palo Alto Research Centre (PARC), died in March.

And there were a few items which might signal bigger news in the future. Voice technology appears to be gaining momentum. The W3C's VoiceXML standard and its Speech Recognition Grammar made progress, becoming Recommendations; Opera launched a voice-enabled browser; and still in the commercial world we saw product announcements from specialist companies like Nuance, plus what many in the industry will undoubtedly see as the Big Event in this area - the launch, by Bill Gates (Knight Commander of the British Empire) himself, of Microsoft's Speech Server product (<http://www.eweek.com/article2/0.1759.1553679.00.asp>).

Another area where there seems to be some movement is RSS (the most common expansions of which look to be "Rich Site Summary" or "RDF Site Summary" but you may see variants such as "Really Simple Syndication" - what's in a name?). The number of RSS newsfeeds available on the Web is growing, and we are seeing some big IT names taking an interest - Sun announced that it will develop RSS for internal communications within its company, as well as delivering external information among developers, customers and partners (<http://www.computerweekly.com/articles/article.asp?liArticleID=129396>).

Information Retrieval

Corporate activity in the domain of information retrieval continued, most in Web search, but more of that later. In purely business terms, there were a number of product releases and acquisitions.

In January, iPhrase acquired Banter. iPhrase is a "search and online customer interaction solutions provider", while Banter develops classification software for unstructured content and informal communications. iPhrase claimed that the acquisition allows it to extend its natural language processing support from Web-based interactions to e-mail, chat and secure messaging. Fast Search and Transfer, now focussing on enterprise systems after selling its web search to Overture (now bought by Yahoo), introduced FAST ESP, stating this to be an enterprise search platform that for the first time enables organizations to intelligently and dynamically retrieve and analyse the complete collection of information existing across their entire enterprise and from the Web.

Verity completed its acquisition of "the intellectual property, certain customer agreements, and other strategic assets from NativeMinds Inc., a privately held software company and provider of integrated self-service solutions" and also announced the availability of K2 Enterprise version 5.5, and Stanford Linear Accelerator Center's selection of Ultraseek.

Another of the big boys in enterprise search, Convera, announced its new Taxonomy Workbench and five new industry-specific taxonomies. The Taxonomy Workbench is designed to help deliver more accurate search results within corporate intranets, providing geographically scattered employees access to diverse information, or on an external Web site for customer service and product information.

And a third major player, Hummingbird, formed a partnership with Recommend to allow its customers to add Recommend's MindServer search and categorization capabilities into their own suite of information management solutions, enabling automated metadata enhancement and broad search across a range of information sources and locations.

Basis Technology, which provides software for multilingual text mining and information retrieval, announced that InQuira will incorporate its Rosette Language Analyzers into the InQuira 6.5 platform. The move is intended to help InQuira to expand its multinational customer base by providing search capabilities in Japanese, French, Italian, German, and Spanish.

The number of new search product providers continues to grow, most seeking to add smarter functionality to your search experience. Some, such as Australia's Mooter (<http://www.mooter.com/> - possibly one to watch), Groxis with its Grokker product (<http://www.grokker.com/>) and TouchGraph (<http://www.touchgraph.com/>), provide visualisation facilities to search results - either their own or those returned by your favourite search engine - offering graphical classification of the results with a view to help you navigate them more easily. Results classification may of course be text-based. Teoma has been offering a form of this for a while now, and Northern Light (<http://www.northernlight.com/>) was a pioneer in this field although they have withdrawn their publicly available search site until July 2004. Vivisimo (<http://www.vivisimo.com/>), which started in 2000, does an impressive job. And lastly, keep an eye open for Dipsie (<http://www.dipsie.com/>), less than a year old and planning to launch, in summer 2004, their Deep Web search service. The Deep Web - all of that information, which cannot be reached by web crawlers, hidden in the databases feeding web pages - has been estimated as containing the vast majority of the information potentially available on the Web. Dipsie aims to index all of that and will thus offer an index of 10 billion documents - more than twice the size of Google's.

Amongst the big players in Web search, Google continues to move slowly towards an IPO, but the idea of a stock market listing apparently continues to be not a top priority for its management.

Research is becoming serious - Yahoo established its own Research Labs as a reincarnation of Overture Research, and Google announced that it will open an R&D centre in Zurich, hoping to capitalise on European expertise.

Web Services

One area which attracted plenty of coverage was the potential convergence of Web Services and Grid computing. The concept is not new, being mooted in 2002 when Paul Messina of Caltech, chairman of the Grid Forum Advisory Committee, said in opening remarks to Global Grid Forum 4 that Grid computing and Web services may eventually converge until there is no distinction between the two. Now things seems to be moving more in that direction.

In January Akamai, The Globus Alliance, HP, IBM, Sonic Software and TIBCO proposed new Web services specifications "that will integrate Grid and Web services standards" to provide a "foundation for the Open Grid Services Architecture". These specifications are the WS-Resource framework and WS-Notifications, which enable grid resource management with standard Web services protocols.

Later that month at the Finexpo City Technology Strategies conference, JP Morgan and UBS discussed how they are using grid computing and web services together, reporting positive returns from early small-scale implementations of Web Services and Grid computing, and raising expectations for wider uptake in new application areas across the enterprise.

February brought a report of the University of Newcastle in the UK building the Grid Application Framework, called WS-GAF, a new kind of framework to join these two worlds, aiming to bring commercial Web services software to grid networkers. But work remains to be done - Ian Foster, one of the pioneers of Grid computing, said in an interview "Ultimately, interoperability requires the standardization of a wide range of WS and Grid standards. The full value of adopting Web services as a basis for the Grid will be achieved only when such standards are in place and in a manner that meets Grid requirements." And standardization takes time, but the January announcement of the specifications noted above is clearly a major step.

Other Web Services news saw greater alignment with the Web Services model from Oracle, previously a lukewarm player in this field, and a flurry of other specifications and proposals. The commercial world is becoming rather full of companies offering WS products and services, and some analysts are predicting a shakeout soon. A British report in January did however find a surprisingly high number of large UK companies - more than half - that claimed to have adopted Web Services, but also found that fears over return on investment (ROI) and security are holding some firms back.

Grid computing

The big news in Grid computing was the happy marriage with Web Services (see above).

And Europe should capitalise on its potential in Grid computing, according to a European Commission workshop "Delivery of industrial-strength Grid Middleware" in January, whose message was that Europe is well positioned for building up its lead in this area but needs to ensure a continued dynamic environment that stimulates innovation coupled with more efficient approaches for commercial exploitation.

Semantic Web

Not much news on the Semantic Web front in this quarter, apart from the launch of a raft of new research projects supported by the European Commission - there is more on that below.

The World Wide Web Consortium published recommendations for two key SW specifications (see Standards Update below), and in January a new Special Interest Group entitled Semantic Web and Information Systems (SIGSEMIS) was established (<http://www.sigsemis.org>).

The Web Services community also tested the water in February, considering adding some semantic capabilities in the next version of UDDI. Two draft specifications were published, and comments are invited. If you want to have your say, take a look at the documents:

<http://www.oasis-open.org/committees/download.php/5337/uddi-spec-tc-req11-14-semantics-20040205.doc>
<http://www.oasis-open.org/committees/download.php/5322/uddi-spec-tc-taxonomies-discussion-20040205.doc>

XML

XML-specific news, like that for the Semantic Web, is a little thin on the ground, but XML pervades so much these days that much of the other news that you are reading concerns XML in some way. XML - and its extended family of offspring: RDF, OWL, VoiceXML, ebXML, and so on - underpins much of the activity in the world of knowledge technologies. "The world has gone crazy with XML ...," remarked Sun's senior IT architect Victoria Livschitz in February.

And Microsoft likes XML as well. Built into Longhorn, the Next Big Thing in Windows, will be an XML-based markup language for developers, codenamed XAML (eXtensible Application Markup Language).

However, Microsoft's more recent excursions into the XML domain have been more contentious.

In 2003 it publicly released, and made freely downloadable, its XML schemas for Word 2003, Excel 2003 and InfoPath 2003. As a gesture of goodwill, it said, but now it has filed patent applications, in New Zealand and the European Union, that cover word processing documents stored in XML format. The proposed patent covers methods for an application other than the original word processor to access data in the document.

And although this embrace of XML seems at first sight to be a move towards openness, some analysts felt that the patents could create a barrier to competing software. "This is a direct challenge to software vendors who want to interoperate with Word through XML," said Rob Helm of Directions on Microsoft. "For example, if Corel wanted to improve WordPerfect's support of Word by adopting its XML format...for import/export, they'd probably have to license this patent."

FP6 and the research domain

Four KT projects from the European Commission Fifth Framework Programme finished in this first quarter of 2004.



<http://swws.semanticweb.org/>

SWWS

SWWS sought to enable Semantic Web Enabled Web Services which can be defined using a formal description language, allowing web services to be discovered using a semantics-driven approach, aiming towards a scalable semantics-based Web Service environment where business partners can seamlessly share services across dynamic business communities.



<http://www.cogproject.org>

COG

Grid technologies are one of the most promising of the recent developments in computing, and COG's objective was to demonstrate the applicability of Grid technologies to industry, addressing the problem of accessing and communicating data which is held in heterogeneous data formats and scattered across disparate systems by developing an ontological system able to interact with disparate data sources.



<http://www.invisip.de/invisip/index.html>

INVISIP

A basic problem in the site planning process is the search for actual and expressive data and their analysis. In particular spatial data are needed to analyse and realise planning objectives. Planners need spatial and context awareness to formulate evaluation criteria in the specific discipline (e.g. traffic and environmental aspects). INVISIP aimed to create a framework to support all involved parties in the site planning process: municipal authorities and departments, planning offices, data suppliers and citizens.

Information Visualisation techniques were used to improve search and analysis tasks, and to facilitate the decision-making process based on an existing metadata information system (MIS) for geographic data.

<http://monet.nag.co.uk>

MONET

Mathematics and its applications have become increasingly pervasive in many areas of modern life, making it important that companies, researchers and individuals use the most appropriate mathematical algorithms for their purpose.

The aim of the MONET project was to demonstrate the applicability of the latest ideas for creating a semantic web to the world of mathematical software, developing a framework for mathematical web services, with advanced prototypes of key technologies used for deploying and interacting with mathematical services on the Web, together with the development of ontologies and tools.

In the last newsletter, we listed the new projects selected for negotiation following the response to Call 1 of the EC's Sixth Framework Programme (FP6). Some of these projects are now underway, and the list below shows you the ones about which we have found something, such as their URL where this is available (or, at least, the best URL that we can find for the project).

	ACEMEDIA	Integrating knowledge, semantics and content for user-centred intelligent media services (http://www.acemedia.org/)
	AGENTLINK III	AgentLink III: A Co-ordination Network for Agent-Based Computing (http://www.agentlink.org/)
	AIM@SHAPE	Advanced and Innovative Models And Tools for the development of Semantic-based systems for Handling, Acquiring, and Processing Knowledge Embedded in multidimensional digital objects (http://dlforum.external.forth.gr:8080/AIM@SHAPE/)
	ALVIS	Superpeer Semantic Search Engine (http://cosco.hiit.fi/search/alvis.html)
	DIP	Data, Information, and Process Integration with Semantic Web Services (http://dip.semanticweb.org)
	DIRECT-INFO	Media monitoring and multimodal analysis for time critical decisions (http://www.direct-info.net)
	KNOWLEDGE WEB	Realizing the semantic web (http://knowledgeweb.semanticweb.org/)
	METOKIS	Methodology and tools infrastructure for the creation of knowledge units (http://metokis.salzburgresearch.at/)
	MUSCLE	Multimedia Understanding through Semantics, Computation and Learning (http://www.cwi.nl/projects/muscle/)
	REVERSE	Reasoning on the Web with Rules and Semantics (http://www.reverse.net/)
	SEKT	Semantically-Enable Knowledge Technologies (http://sekt.semanticweb.org/)
	SIMAC	Semantic Interaction with Music Audio Contents (http://www.semanticaudio.org/htmls/simac.html)
	VIKEF	Open software framework for future Semantic Web applications (http://www.unin.it/en/internazionale/area_ric/projects/vikef.htm)

STANDARDS UPDATE

Web services

In January Microsoft, BEA Systems and Tibco Software published a specification designed to communicate events between Web services. The specification, called Web Services Eventing, is intended to simplify the development of applications that rely on events to trigger an action. However, they did not indicate to which standards organisation they would submit the specification, and IBM - another key figure in Web Services - did not show any interest.

IBM's reasoning perhaps became more apparent later that month, since it was, however, fully involved in the specifications, noted above, designed to integrate Grid and Web services standards, also announced in January. In certain ways these specifications bypass the earlier specification. Other companies concerned were Akamai, The Globus Alliance, HP, Sonic Software and Tibco. This time, of course, it was Microsoft that was notable by its absence, but Tibco seemed to be playing on both sides.

But Microsoft were once again involved - along with Canon, BEA Systems, and Intel - in the release of a new Web services specification, to help provide Web services interoperability and support for occasionally-connected devices and systems. This specification, WS-Discovery, acts like the Universal Description, Discovery and Integration (UDDI) standard, but it will focus on dealing with devices and systems that are not always connected to the network.

Semantic Web

The World Wide Web Consortium published recommendations for Resource Definition Framework (RDF) and the Web Ontology Language (OWL).

W3C also launched Phase 2 of the Semantic Web Activity, with two new Working Groups; the Best Practices and Deployment WG (charter) and the RDF Data Access Working Group (charter). These join the RDF Core and Web Ontology WGs, the Semantic Web Interest Group, and the Semantic Web Coordination Group.

XML

Little happened in standards for XML - a sign of maturity of the standard? The only event of note was the release by the W3C of its Recommendation for version 2.0 of VoiceXML, "designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations. Its major goal is to bring the advantages of Web-based development and content delivery to interactive voice response applications."

EVENTS

The early part of the year is typically a quiet time on the event circuit. KTweb listed just 23 events in the first quarter of 2004 - 15 in the Europe Union, 1 in other European countries, 5 in the USA and 2 elsewhere. Even then, seven of these events belonged to the same 'travelling' series of workshops. That said, one event in that quarter was CeBIT, probably the world's biggest IT event.

KTweb UPDATE

As you can see by virtue of reading this newsletter, KTweb is soldiering on despite the end of EU funding. The past three months have been a period of consolidation and planning, and we will be looking to implement some of our ideas in the months to come.

One area that we will need to explore seriously is advertising and sponsorship, to provide us with the funds to keep Katie healthy. If your organisation believes that it could help in this way, please get in touch with us.

For answers to your questions and enquiries about **KTweb** and its portal and services, please send an email to contact@ktweb.org.