Session VIII: Search as an architectural component

Chair: Loretta Anania (EC– DG INFSO),
Th. Zahariadis (Synelixis), J. Domingue (Open University), P. Daras (ITI), A. Galis (UCL)

Objectives

As the FIA activities evolve, it turns out that there are going to be different stratums that support the Future Internet. All these stratums have different types of resources and also virtual representations. At the network stratum, we may consider as resources the nodes themselves (e.g. routers, servers, printers, terminals, etc.) and their physical characteristics (e.g. processing power, network interfaces, routing capabilities, memory, caching capabilities, traffic, etc.). At the Service stratum, we may consider the services, the service components and their APIs and interfaces as resources. At the Content Stratum, the different types of content (e.g. text, pictures, voice, video of different types and formats, slices and sub) can be considered as resources. Finally, any piece of information or any application (including social networks and link open data links) may also be considered as a resource.

The initial objective of this session was to highlight the search and discovery issues of different resources at the different stratums. This is a very broad topic. Quite different mechanisms (even philosophy and language) are currently used to discover/mine, describe, index and retrieve resources at each stratum. Due to time restrictions, the session was focused only in the area of content searching and retrieval, which may be the most challenging ones as the content is and will grow much faster than any other resource and the search engines need in most cases to extract and combine information rather than just discover it.

Presentations

• Introduction and context	Loretta Anania	EC – DG INFSO
• Search as an architectural component: searching for a new paradigm	Francesco Nucci	Engineering
• Next Generation Search	Ricardo Baeza-Yates	Yahoo!
• Visual Search as an architectural component	Alexis Joly	INRIA
• Discussion	Moderator: L. Anania (EC) Panel: R. Baeza-Yates (Yahoo!) F. Nucci (Engineering) A. Joly (INRIA) J. Domingue (Open University)	
• Conclusions- Discussion	Loretta Anania	EC– DG INFSO

*The presentations are available on the FIA Ghent website (*http://fi-ghent.fi-week.eu/slides/*)*

Summary of Presentations

Search as an architectural component: searching for a new paradigm
(Francesco S. Nucci – Engineering R&D Lab, Director European Projects)

"Search" is not limited to everyday's experience of searching the web but it is a *must* feature of business intelligence applications for the enterprise. The size of digital archives has grown enormously in many application domains (e.g. enterprise, news agencies, TV broadcasters, advertising agencies,...) the characterization of their interactions (i.e. messages, events, etc.) and their distribution. Moreover, content is increasingly going to be Multimedia based: combinations of text, audio, still images, animation, video and interactivity into a single form. As a result and based on various analysis 50% of the applications will include a search facility as a primary interface for end users.

This need for search engines poses a number of challenges, namely:

- Lack of a unified framework for multimodal content search and retrieval
- Lack of Dynamic and Scalable Indexing techniques for searching over the network
- Lack of user-centric and context-aware S&R framework
- Lack of efficient tools for searching Intelligent Content

Dr. Nucci concluded that in the current Internet we are using some core services – such as search engines. Others, e.g., to provide geo-information, people search or social networking, have seen tremendous growth in recent years. Mostly we are using these services in isolation from each other, e.g., via independent websites and user interfaces. Some services – like search – are starting to become integrated but here a significant increase can be expected.

Next Generation Search
(Dr. Ricardo Baeza-Yates, VP of Yahoo! Research for EMEA & Latin America)
Content is increasing at very high rates. Even in 2007, the published content was more than 4GB per day, however the private data is even much higher (user generated content more than 80GB per day, private text ~3TB, upper bound of typed text ~700TB). This creates specific trends for search engines: User Generated Content (quality vs. quantity), Social Networks, Real time content creation (people + physical sensors). On the other hand also the search engines are evolving, moving from mining into text into identifying what are the user's task and context. This affects not only the retrieved information but also the way to present that to the users.
Dr. Baeza-Yates concluded that we are moving from a web of pages to a Web of Objects. Objects are people, places, businesses, restaurants … (named entities), which have attributes (missing, noisy, etc). Search engines should be intelligent enough to find and retrieve content objects based on the context and the user need. The key feature in search engines will be the user experience (information + presentation).

Visual Search as an architectural component
(*Alexis Joly, INRIA*)
Visual Search supports already very important applications such as TV channels/archives searching, copy detection (beyond DRM issues like contents links), large scale object retrieval. The next steps in visual search will be visual objects mining, linking visual contents at the object level, visual query suggestion and visual based events tracking. The visual object's model in the Future Internet it is expected to have neither concepts nor low level features, but support visual objects represented as independent entities.

Panel Presentations and Discussions
Some conclusions based on the discussions are that the term "searching" has quite different usage and meaning in each stratum (networks, content, services). Moreover, there are quite different mechanisms to discover information, so we should be very careful as the term is evolving. On the other hand, it is obvious that discovery of resources should be inherently supported by Future Internet Architecture.

Another important topic was the conflicting issues between privacy, data security, personalization and user aggregation. Everyone agrees that privacy and security are very important issues. Yet, in order to increase the personalized behavior of the Search Engine more info has to be extracted so privacy is challenged.

Finally the issue of open source search was discussed and proposed as one of the lessons learned by Engineering in Pharos project.