

# IRFS2008

5<sup>th</sup> - 8<sup>th</sup> NOVEMBER 2008  
Austria Trend Hotel Savoyen  
Vienna, Austria

“RESEARCH MEETS PRACTICE”

## PRELIMINARY PROGRAMME

### Programme Committee

C.J. van Rijsbergen, Yves Chiamella, John Tait,  
Stephen Adams, Henk Tomas, Barrou Diallo

[www.irfs2008.org](http://www.irfs2008.org)

## 5<sup>th</sup> NOVEMBER 2008

---

10:00 - 18:00

### Pre-Symposium Tutorials “IP for IR” and “IR for IP”

The “IP for IR” and “IR for IP” tutorials aim at giving Information Retrieval scientists and Intellectual Property professionals an overview and understanding of the IR world and the IP world: what are the main concepts, fields of expertise, methods, tools, challenges and more. The “IP for IR” will be run by an IP professional and the “IR for IP” will be run by an IR scientist.

The morning session will be repeated in the afternoon and will be followed by a joint Practice Class, which all participants are invited to join. IR scientists will be given a search question and will endeavour to find relevant results with a given set of patent search tools. IP professionals are invited to guide and coach the IR scientists through this process.

We expect this session to be highly interactive and result in an increased understanding of patent searching challenges. It is a great opportunity for IR and IP professionals to share their knowledge over their respective techniques.

18:00 - 20:00

### Welcome Reception & Registration

---

09:00 - 18:00

**Main Symposium Part I**

19:30 - 23:00

**Banquet Dinner**

09:00 - 17:00

**Main Symposium Part II**

17:00 - 19:00

**Closing Reception**

.....

**Multilinguality (Asian and Cyrillic Languages)**

Patent corpuses cover a wide variety of natural languages, and search processes about a given patent may be issued by users whose languages are extremely different. Moreover, a given set of documents may contain words or sentences from different languages and in any case the relevance of a foreign language retrieved patent must be properly evaluated.

This linguistic gap between documents and searchers often results in poor retrieval performances in multilingual search contexts, due to ill-formed queries, lack of recall, or incorrect relevance evalua-

tion of retrieved documents. Automatic translation of natural languages is an essential tool for addressing these problems: techniques from this domain may be used for translating documents, or passages of documents to help understanding their content, while automatic translation of user queries is the basis of multilingual Information Retrieval and cross language Information Retrieval, two techniques aimed at processing retrieval queries issued in a language different from the one of the searched documents.

**Annotation & Ontologies in the Context of Patent Retrieval**

Annotations and indexes are common tools used in patent retrieval. While annotation techniques mainly deal with XML-based environments used for creation of entirely new searchable parameters – and eventually large retrospective databases – ontologies deal with the management of terminologies used in such contexts.

While there is still little agreement on how to standardise production of electronically filed patent applications in a way that captures information elements valuable to the searcher, ontologies aim to

offer partial solutions to this problem by allowing ways to define and use semantic relations among terminologies coming from various sources.

When included in information retrieval systems, they may be viewed as advanced thesauri allowing a better control of indexing and query languages. This topic considers the value of markup languages for improving document retrieval, what the future may hold beyond XML and the foreseeable improvements of such environments based on ontology management.

**Non-Text Retrieval**

With the exception of specialised chemical structure databases, most retrieval of patent information still relies upon plain text. However, non-text items such as drawings, photographs, flow-charts, electrical circuit diagrams can all form important parts of the information disclosure. In trademarks, additional parameters such as colour, sound or smell can all

be applied for, yet we have no search engines capable of systematically retrieving such information. This topic will explore the need for inclusion of non-text material in applications for intellectual property rights, and some of the possibilities in developing search tools capable of storing and retrieving this type of material directly.

## Interfacing (Collaborative Assessment of Relevance)

Many existing IR interfaces have been designed to support a process which only involves one single person. In patent matters, however, a team of people – patent searcher, librarian, patent attorney, R&D, manager etc – is assigned to complete tasks that require a high level of efficiency and exhaustiveness. This raises several challenges that have not been

fully exploited in the IR community. What is the information seeking behaviour in patent retrieval? What contextual factors influence their behaviour? How can a search interface support collaboration in patent retrieval? How can we divide labour, share knowledge, and resolve disagreement on relevance in patent retrieval tasks?

## Tuning IP-IR Channels for Improved Communication and Collaboration

The purpose of this session is to get a common understanding over ways to proceed for taking common initiatives between IP and IR communities, but also between IP professionals in the future. A first goal aims at identifying communication tracks

and workflow predicates, allowing delegates to practically exploit the spirit of the IRF Symposium. A second goal consists in identifying means for avoiding communication bottlenecks and promoting new sources for project initiatives.

## Project Presentations

The support and promotion of scientific projects is one of the main tasks of the IRF.

Taking into account the learnings from the IRFS 2007, special attention is brought to the fact that the projects deliver answers to important issues of patent professionals.

Experts from top universities present their research projects and latest findings in corpus enrichment and patent retrieval.

The applied techniques are the outcome of the latest insights in information extraction, linguistics, probabilistic retrieval and language modelling.

## Matrixware Contribution

### Alexandria

The Alexandria Archive is designed as an extensible, global storage facility for high quality scientific, technical, and business information. The archive provides an authoritative information source providing data services to global R&D communities. Beginning with the world patent corpus, Alexandria will deliver an infrastructure capable of supporting the diversity of data services required by professional researchers, information specialists, and software engineers.

### Leonardo

Is the application development layer that allows access to data repositories. It allows information retrieval professionals and researchers to take advantage of an ever expanding suite of components or tools for searching and retrieving data from an increasing number of enhanced data repositories such as Alexandria. The Leonardo development platform is an open, transparent and flexible framework based on Eclipse with tools that encourage software engineers and developers to create their own components. In this way it acts as a technology transfer platform for professionals wanting to access cutting edge search tools and for researchers to access and further enhance large, relevant data archives.

## IRFS2008 Exhibition 6<sup>th</sup> - 8<sup>th</sup> November 2008

The IRFS2008 Exhibition – showcasing latest IR research projects and innovative commercial products – will form an integral part of the IRFS2008. There will be dedicated time slots within the symposium programme for delegates to visit the exhibition area, learn about new approaches and technologies for patent searching, and engage in lively discussions with exhibitors and co-delegates.

## 8<sup>th</sup> NOVEMBER 2008

---

09:00 - 15:00

### Outreach Exhibition Day

The Outreach Exhibition Day is an open day primarily aimed at schools, universities and any other groups with an interest in IR and IP. IRFS2008 participants are most welcome to stay and use this extra day to visit the exhibition and have in-depth discussions with IR research project teams and selected commercial companies showcasing innovative search technologies for patent searching.



#### IRF Society Office

Palais Eschenbach, Eschenbachgasse 11, 1010 Vienna, Austria, Europe

Phone +43-1-585 73 61-0, Fax +43-1-585 01 41

E-mail: [office@ir-facility.org](mailto:office@ir-facility.org)

[www.ir-facility.org](http://www.ir-facility.org), [www.irfs2008.org](http://www.irfs2008.org)

The IRFS2008 is endorsed by



Association for  
Computing Machinery

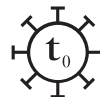
---

**IRFS2008 delegates may also attend the following accompanying event, organised independently by the World-Information Institute:**

8<sup>th</sup> November, 11:00 - 20:00

**World-Information Institute Symposium**

**"Deep Search" – The digital future of finding out**



With the explosion of information in all shades and languages, issues of orientation and navigation in the oceans of knowledge pose themselves with renewed urgency. How is computer readable significance produced, how is meaning involved in machine communication? These questions of culture, context und classification in information systems should not be ignored since what is at stake is nothing less than how we, as individuals and institutions, come to find out about the world. "Deep Search" addresses the social and cultural dimension as well as the information politics and societal implications of search.

<http://world-information.org/wii>

<http://www.t0.or.at/>