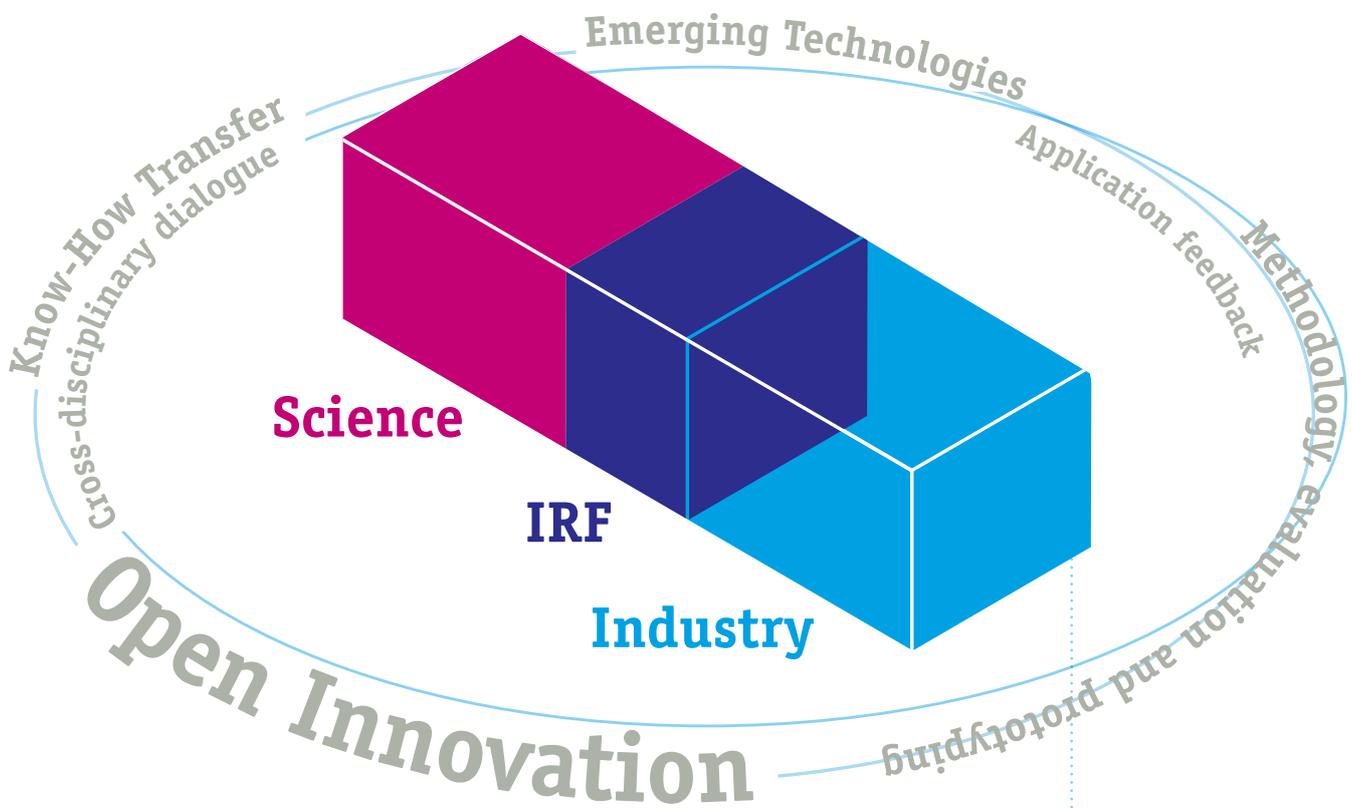




Information Retrieval Facility

The Science and Industry Platform for a Sustainable Innovation Cycle



Industrial Researchers & Analysts

Information Professionals

Information Publishers

Professional Services Firms

**Governmental &
Non-Governmental Organizations**



The Bridge between Science and Industry

The Information Retrieval Facility (IRF) is an independent, non-profit research institute founded by internationally renowned information retrieval scientists with the mission to facilitate research in large-scale Information Retrieval (IR) and foster knowledge transfer between academia and industry. The IRF is a member-based organization with over 250 scientific members around the globe and is managed by a Scientific Board composed of international IR experts. One of the objectives of the IRF is to make the latest academic research developments available to industry and thereby develop solutions to search challenges that are not yet answered by current tools. To do this, the IRF operates on two levels:

- As an **Open Science Institution**, the IRF maintains a semantic supercomputer infrastructure that facilitates large scale IR experiments and provides access to a large and high-quality corpus of scientific and technical data, enabling scientists to study methods beyond limited laboratory conditions.
- As a **Technology Transfer Platform**, the IRF brings IR scientists and end-users from the industry together, ensuring that both communities engage in a fruitful dialogue. This interaction between experts from research and industry ensures a high scientific standard and eventually a better relevance of the research work to business.

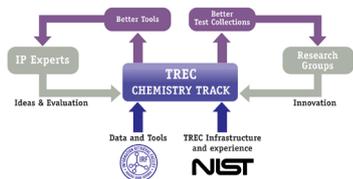
The IRF is engaged in several research projects on a national and European level, for example:



PLUTO: Patent Language Translations Online: EU-funded project started in 2010 with a duration of 3 years that will develop a collaborative online translation platform for patents.



KHRESMOI: Medical Information Analysis and Retrieval: EU-funded project started in 2010 with a duration of 4 years that will develop a multimodal, multilingual search and access system for biomedical information and documents.



TREC-CHEM: Annual evaluation campaign launched in 2009 by the IRF with the Text Retrieval Conference in the US to identify how current IR methods adapt to text containing chemical names and formulae.



CLEF-IP: Launched in 2009 by the IRF with the Cross Language Evaluation Forum, this annual evaluation track assesses IR techniques in the intellectual property domain.

About Information Retrieval

Information Retrieval is the science of searching for documents and for information within documents, as well as searching relational databases and the World Wide Web. It has many practical implications for the industry: Beyond the development of search engines and optimization of digital libraries, IR applications encompass automatic classification and annotation, information discovery, alerting and routing systems. They reach out to automatic translation engines, providing companies with business development tools of immediate strategic relevance.



“ The IRF provides a unique opportunity to evaluate approaches to information retrieval in realistic environments with real data and users. ”

C.J. “Keith” van Rijsbergen, *Chairman of the IRF Scientific Board*

The IRF Symposium – Where Research meets Practice



The IRF offers several communication platforms to foster a fruitful dialogue between the science and industry communities in the field of professional information retrieval. One of them is the Information Retrieval Symposium (IRFS), held on a regular basis in Vienna. The IRFS is an international conference which brings together high-ranking delegates from the Intellectual Property and Information Retrieval communities. It gives these experts a forum to exchange ideas on specific problems and related solutions, with the objective of making the latest IR technologies finally available to the industry.

Since 2007, the IRF Symposium has established itself as a unique conference where academia and industry meet to discuss and solve patent retrieval challenges. After a discovery phase, during which the patent information professionals explained their needs and showed the limits of the systems they work with, the last IRF Symposium in June 2010 marked the start of the integration phase: Information retrieval scientists have started to integrate these parameters in their research. In 2011, the IRF Symposium will demonstrate further examples of retrieval techniques developed to better solve the information professionals' needs in 3 specific fields: **Image Retrieval, Machine Translation and Interfacing.**



The IRF Symposium 2011 is placed under the sign of **collaboration and results**: Information retrieval researchers will apply their systems to a standardized patent collection and patent searchers will evaluate the results. The past Symposium has shown the need for the involvement of a third

party, the technology implementers, who can take on the results to develop practical tools. The IRF Symposium 2011 intends to give them this opportunity.

Highlights of the IRF Symposium 2011 programme will include: **Open discussions**, presentations involving **demonstrations and specific use cases**, an exhibition showcasing **new technologies and prototypes**, as well as an enhanced version of the **PatOlympics**. Dedicated tracks showing best retrieval practices from other technical fields like **biomedicine and trademarks** will be introduced, thus opening up new perspectives on how to improve patent search and analysis.



“ The IRF is doing a fantastic job of bringing together two communities that were previously unaware of each other's existence. This combination of unique domain expertise will help fashion novel solutions to real, complex problems for patent information professionals. ”

Anthony Trippe, Director, IP Analytics - 3LP Advisors

IRF Key Research Areas

Information Retrieval applications provide concrete solutions to handle huge amounts of business, legal, scientific or web data, and offer techniques to extract the most relevant information from sometimes very complex documents. The IRF is particularly active in research areas of immediate relevance to industry:

Multiple Indexing

The purpose of creating an index is to optimize the speed and the performance of finding relevant documents. The creation of multiple indices improves the accessibility of different kinds of data within an organization which increases the productivity of the employees.

Automated Text Annotation

Annotation, or tagging, is about attaching names, attributes, comments, descriptions etc. to a document or to a selected part in a text. It provides additional information (metadata) about an existing piece of data. The implementation of automatic and semi-automatic annotation techniques will raise the data quality for the whole organization, give more meaning to existing documents and allow generating new connections between them.

Automated Information Extraction

Automated information extraction enables search without language barriers: The relevant information is extracted from unstructured text and made available in the form of a customized database.

Information Visualization

Information visualization is the use of graphical representations to provide intuitive views of large amount of information, thereby simplifying interpretation and analysis.

Image Retrieval

Image Retrieval is the search for images based on their visual characteristics. As with text retrieval, image retrieval performs better in narrower application domains. The IRF is developing a technique that enables efficient image searches for a variety of technical drawings, such as flow charts, block diagrams, time charts and graph plots.

Automated Document Categorization

Document classification (or categorization) consists in assigning an electronic document to one or more categories, based on its contents. The implementation of classification techniques will not only raise the accessibility of data within your organization, it will also reduce the overhead costs of information management by automating content access.

Statistical Machine Translation

Statistical Machine Translation (SMT) engines generate translations on the basis of statistical models derived from the analysis of bilingual text corpora. The IRF Chinese-to-English translation engine has been trained with more than 4 million bilingual aligned sentences obtained from human-translated patents. It can be open to other domains and trained with other languages.

Complex Professional Search Strategies

One of our projects aims at translating extensive information search tasks that rely partly on complex search algorithms in an easy-to-use graphical interface based on the workflow paradigm. One of the benefits is the reusability of complicated search patterns by less experienced professionals.



Why cooperate with the IRF?

- **Scientific Excellence**
Access the know-how of leading IR scientists
- **Independence**
Benefit from our neutral guidance and evaluation know-how
- **Light Weight Projects**
Get quicker results and lower your development costs through the integration of existing academic tools and open source software
- **Result Ownership**
Decide for proprietary, shared or open research projects
- **Confidentiality**
Early NDA agreements protect your know-how and data
- **Costs Savings**
Reduce the project costs by joining the IRF as Project or Strategic Member

Contact

membership@ir-facility.org
for more information.

Industry Cooperation

Whether for proprietary, shared or open research projects, you can work with some of the best IR brains in the world under the guidance of the IRF as vendor-neutral, independent research institute.

The most obvious beneficiaries of efficient information retrieval are companies that generate large amounts of data or work with complex data like chemical process descriptions or mechanical drawings, and industries with high innovation pressure and high R&D activity like the manufacturing and pharmaceutical industries.

Frequently, the implementation of existing academic tools combined with well-established open source components will deliver a quick, but sustainable increase in the efficiency of the search processes within an organization. More specific retrieval challenges will require the set up of a dedicated R&D project that will deliver a customized tool. In both cases, the IRF provides the necessary IR skills and resources to maximize the impact for your organization while minimizing the development costs.

Our services include:

- **Discovery Workshop:** Discover cost-efficient approaches to improve search and analysis processes in your organization
- **Training:** Improve the knowledge of your own experts in order to optimize existing search processes
- **Expert Finding:** Find IR experts with appropriate skills for your specific project requirements
- **R&D Projects:** Work with leading IR scientists under the guidance of an independent project coordinator to develop a solution customized to your needs

We offer to:

- Conduct and manage research projects
- Write scientific reports
- Build prototypes
- Evaluate tools
- Produce metadata
- Provide coaching for public funding
- Build and coordinate research consortia

Through our global network of recognized IR specialists, we give you access to the best experts to develop the search solutions that will meet your specific needs.

Industry Membership

Industry Membership is open to companies interested in accessing the IRF research infrastructure and/or conducting R&D projects with the IRF.

Access Membership

Use the IRF infrastructure (large data collections, semantic supercomputer, evaluation framework) for experimentation and testing

Project Membership

Save money on your different projects with the IRF: Training, evaluation, research, prototyping etc...

Strategic Membership

Influence the scientific direction of the IRF and benefit from a research fellowship at the IRF

Visit www.ir-facility.org or contact membership@ir-facility.org to find out more about collaboration opportunities with the Information Retrieval Facility Society.

The IRF at a Glance

Year of foundation:	2007
Founding members:	9 of the top IR scientists* worldwide and Matrixware GmbH
Scientific Board:	C.J. "Keith" van Rijsbergen* (Chairman), <i>University of Glasgow, UK</i> Yves Chiaramella* (Vice-Chairman), <i>Université Joseph Fourier, FR</i> Maristella Agosti, <i>University of Padova, IT</i> Jamie Callan*, <i>Carnegie Mellon University, US</i> Kilnam Chon*, <i>Korea Advanced Institute of Science and Technology, KR</i> W. Bruce Croft*, <i>University of Massachusetts Amherst, US</i> Hamish Cunningham*, <i>University of Sheffield, UK</i> Norbert Fuhr*, <i>Universität Duisburg-Essen, DE</i> David Hawking*, <i>Funnelback Internet & Enterprise Search, AU</i> Noriko Kando, <i>National Institute of Informatics, JP</i> Arcot Desai Narasimhalu*, <i>Singapore Management University, SG</i> John Tait, <i>Information Retrieval Facility, AT</i> Benjamin T'sou, <i>City University of Hong Kong, HK</i>
Executive Board:	Francisco Webber (Chairman) Daniel Schreiber (Treasurer) Sylvia Thal (Secretary)
IP-Expert Committee:	Stephen Adams, <i>Magister Ltd, UK</i> Teresa Loughbrough, <i>Unilever R&D, UK</i> Anthony Trippe, <i>IP Analytics - 3LP Advisors, US</i>
Data collections:	<ul style="list-style-type: none">• 19 million patent documents in a standardized ST 36 XML format (MAREC)• 128.9 million Linked Life Data (medical entities)• Web crawl: Over 1 billion web pages in 10 languages
Infrastructure:	<ul style="list-style-type: none">• SGI supercomputer for very large in memory experiments (LDC - Large Data Collider)• SGI FPGA RASC modules• 240 TB primary high speed storage• IBM x86 supercomputer (MDC - Medium Data Collider)• IBM x86 computing cloud



Information Retrieval Facility

Tech Gate | Donau City Straße 1 | A-1220 Vienna | Austria | Phone: +43-1-236 94 74 | ZVR-Number: 589671732
office@ir-facility.org | www.ir-facility.org | <http://ir-facility.net>