IRF Industrial Cooperation Opportunities

ICIC 2010

Marie-Pierre Garnier, Corporate Liaison Manager
mp.garnier@ir-facility.org
www.ir-facility.org
http://ir-facility.net
The Information Retrieval Facility (IRF)

- Founded in 2007 in Vienna, Austria
- Independent non-profit research institute
- Member-based organization with over 250 scientific members from around the globe
- Managed by a Scientific Board composed of leading IR experts
- Chairman of this Board is Keith van Rijsbergen, one of the founders of modern Information Retrieval

OBJECTIVES:
- Promotes and facilitates research in large scale information retrieval
- Fosters knowledge transfer between academia and industry
What is Information Retrieval (IR)?

IR is the science of searching:
- For documents, for information within documents, and for metadata about documents
- Relational databases and the World Wide Web

IR is interdisciplinary:
- Computer Science
- Mathematics
- Library Science
- Information Science
- Information Architecture
- Cognitive Psychology
- Linguistics
- Statistics

IR applications include:
- Search Engines
- Digital Libraries
- Information Discovery
- Alert & Route
- Automatic Classification & Annotation

Domain-specific applications of IR include:
- Retrieval of:
  - Patents
  - Legal documents
  - Scientific & technical documents
  - Images
  - Chemical structures
  - Market information
IRF Key Research Areas

- Multiple Indexing
- Automated Text Annotation
- Automated Information Extraction
- Automated Document Categorization
- Complex Professional Search Strategies
- Image Retrieval
- Interfacing Professional Users
- Statistical Machine Translation
Current Research Projects

PLuTO: Patent Language Translations Online
- Funded by the EU (ICT Policy Support Programme). Duration: 2010-2013
- Consortium Partners: CNGL/Dublin University, ESTeam, Cross Language, WON
- Goal: Develop a collaborative online translation platform for patents

KHRESMOI: Medical Information Analysis and Retrieval
- Funded by the EU (ICT FP 7 Programme). Duration: 2010-2014
- Consortium Partners: HES-SO, Atos, Dublin University, Ontotext, University of Sheffield + 7 additional European partners
- Goal: Develop a multimodal, multilingual search and access system for biomedical information and documents

Key Technologies:
- Machine Translation, Translation Memory, Information Retrieval
- Automated Information Extraction, Automated Image Analysis and Indexing, Linking of Unstructured Documents with Structured Knowledge Base, Cross-Language Search, Adaptive User Interfaces
Current Research Projects

Service Detective
- Funded by the Austrian FFG (FIT-IT Programme). Duration: 2009-2011
- Consortium Partners: University of Innsbruck, Seekda
- Goal: Automatic discovery of web services

LarKC: Large Knowledge Collider
- Funded by the EU (ICT FP 7 Programme). Duration: 2008-2011
- Consortium Partners: STI Innsbruck, AstraZeneca R&D, CEFRIEL, Max Planck Institute + 8 additional European partners
- Goal: Build an integrated platform for large scale semantic computing

Key Technologies:
- Web Crawling, Semantic Annotation
- Large Scale Information Retrieval, Semantic Computing
Other Research Activities

- Development of a technique to enable efficient image searches for a variety of technical drawings, such as flow charts, block diagrams, time charts and graph plots.

- Phrase-based SMT technology (MOSES)
- Chinese to English automatic translation, customized for patents
- 4 million bilingual aligned sentences obtained from human translated patents

- Development of a framework for professional search based on the workflow paradigm
- Goal: Translate extensive information search tasks in an easy-to-use graphical interface

CLEF-IP (cross-lingual search of patents) and TREC-CHEM (retrieval of chemical documents)

Statistical Machine Translation

Image Retrieval

Evaluation Tracks

Interfacing Professional Users
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 adaptation 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
IRF Industrial
Cooperation Opportunities

Discovery Workshop
- Identification of the retrieval challenges in your company

IRF Services
- IR Training & Knowledge Transfer
- Expert Finding
- R&D Projects
- Prototyping

IRF Industry Membership
- Access Membership
  Use the IRF infrastructure for experimentation and testing
- Project Membership
  Buy IRF services at discounted rates
- Strategic Membership
  Influence the scientific direction of the IRF and benefit from a research fellowship at the IRF
The IRF manages a pool of experts in the fields of information retrieval, natural language processing and semantic technologies, all of which have been selected for their proven track record in their area. We find the experts best suited to your needs!

We offer to:

- Conduct and manage research projects
- Write scientific and technical reports
- Build prototypes
- Evaluate tools and systems
- Produce metadata
- Provide coaching for public funding (EU & national)
- Build and coordinate research consortia

The IRF organizes in-house trainings dedicated to the optimization of search strategies and processes to achieve business relevant search results from the Internet and other information sources.
Access Membership

**Unique data collections:** Access to large scale data of technical and scientific documents (patents, scientific data)

**Powerful processing environment:** Level 1-access to the semantic supercomputing infrastructure

**Reference laboratory QM:** Access to evaluation framework

**Global expert pool:** Discounted access to an international, multidisciplinary network of leading scientists

**EU projects interface:** Access to an experienced project initiator and coordinator for publicly funded projects.

**Discounted rates for IRF events:** 50% discount on IRF Conference and IRF Symposium tickets

**For IT & Professional service providers:**
- Presentation of prototypes on the IRF Prototype Portal
- Posting of service profiles on the IRF Marketplace
Project Membership

Includes all benefits of the Access Membership with an upgrade to level 2 - access to the semantic supercomputing infrastructure.

- **Discovery workshop** to determine the project scope
- **Cost savings for R&D projects and trainings:** 20% discount on IRF services fees
- **Student platform:** Possibility to post an internship need
- **Discounted rates for IRF events:**
  - 2 complementary tickets for the IRF Conference and Symposium;
  - 25% discount on IRF Strategic Seminar tickets

For **IT & Professional service providers**:
- Free exhibitor booth to expose prototypes at the IRF Symposium
- Free sponsor package “Friend of the IRFS”
Strategic Membership

Includes all benefits of the Project Membership

**Industrial Advisory Board:**
Possibility to send one representative to:
- influence the scientific and strategic direction of the IRF
- Assess the relevance of the IR research activities in the light of industry needs and make recommendations to the IRF Scientific Board

**Research fellowship at the IRF:**
1 dedicated research position (excl. staff costs)

**Discounted rates for IRF events:**
3 additional complementary tickets for the IRF Conference and Symposium; one complimentary IRF Strategic Seminar ticket

**For IT & Professional service providers:**
- Free exhibitor booth to expose prototypes at the IRF Symposium
- Free sponsor package “Friend of the IRFS”
- Possibility to run a workshop at the IRF Symposium
Data Collections - Example

MAREC

Consisting of 19 million patent documents in different languages, that are normalized to a highly specific XML format, MAREC is the first standardized patent data corpus for research purposes.

MAREC serves as a global storage facility for high quality scientific, technical and business information.

**MAREC at-a-glance**

- 19 million XML documents
- From 4 patent organizations:
  - European Patent Office (EPO)
  - World Intellectual Property Organization (WIPO)
  - United States Patent and Trademark Office (USPTO)
  - Japan Patent Office (JPO)
- Patent applications and granted patents from 1976 to 2008
- Unified fields, numbering scheme and citation format
- Comparable corpus

Industry Membership provides access and support to the IRF document collections such as MAREC.
Leonardo is a premium toolkit framework for retrieval and analysis tasks:
- Designed to enable the elimination of data and technical barriers so that your experts can focus intelligently on the task at hand and share their results with others at all steps of the process.

Leonardo sets a new standard for the human intelligence enterprise:
- It can plug into virtually any data source, enabling you to connect different kinds of information and integrate different tools.

Leonardo has been custom-engineered:
- It is built on a IR Workflow model, which allows retrieval tasks to be solved similarly in a scalable and extensible environment.

Leonardo has an easy-to-use graphical interface:
- It allows to use complicated search processes without IR knowledge.
• 3 plenary sessions: Image Retrieval, Machine Translation and Interfacing

• Panel discussions at the end of each session

• Presentations involving demonstrations and specific use cases

• Exhibition showcasing new technologies and prototypes

• PatOlympics - Second edition of the interactive prototype evaluation competition

• Biomedicine and Trademark tracks for new perspectives on how to improve patent search and analysis

• Second edition of the IRF Scientific Conference, on 6 June 2011