

**New challenges in Information Retrieval and Text Mining
in an open source initiative**
4-6 Febbraio 2009
CNIPA- Via Isonzo 21/b, Roma

Il programma del Workshop si suddivide in tre giornate a tema:

- Il 4 febbraio, con tema **Information Retrieval for Public Administrations**, sarà una giornata dedicata alla presentazione delle esperienze italiane ed europee nel settore della Pubblica Amministrazione.
- Il 5 febbraio, con tema **Existing Terrier extensions and experiences**, è una giornata di riunione, di confronto e sintesi riservata alla comunità open-source operante nello sviluppo della piattaforma Terrier.
- Il 6 febbraio, con tema **Implementation of the Collaboration**, è una giornata di confronto sulle possibili forme di collaborazione e di finanziamento alla ricerca.

4 febbraio 2009: Information Retrieval for Public Administrations

Morning:

Welcome and Introduction (CNIPA)

Part I (Sala Conferenze): **Case studies.**

- Applications in Justice (CNIPA).
- Pronouncements by Antitrust Authority (*Italian Antitrust Authority*).
- Case studies at CNIPA (*Laboratorio di Sperimentazione Tecnologie Innovative - CNIPA*).
- Evaluation and selection of competitive Information Retrieval solutions for PA (*CONSIP*).

Coffee break

Part II (Sala Conferenze): **Experience in Applications**

- AlmavivA projects and perspective
- Patent retrieval at Matrixware
- Document Management for National Archives (*Nestor*)
- Information Retrieval issues in CMS platforms (*Sistemi Informativi CNR*).
- Enterprise search (University of Glasgow) (to be confirmed).
- Tagging Services and Public Administration - A case study using BibSonomy (University of Kassel)

Afternoon:

Round table with participants (Chairman *Mauro Draoli*):

- Application Areas & Advanced projects
- Building common and shared resources for the development and evaluation of IR solutions in European PA
- Joint participation to International IR competition (e.g. CLEF ResPubliQA2009, Legal track, ect.).

Existing Terrier extensions and experiences

5th February 2009

Topics:

- Modelling
 - Query language (Boolean retrieval, conditional operators, contextual search, p-norms, proximity search etc.)
 - Integration with DB and XML retrieval, passage retrieval, hypertext modelling
 - Combining indexes and/or fusion of outputs
 - Dictionaries, ontologies and use of external resources
 - Integrating 3rd parties tools/models
 - Building up native applications
 - Automatic tagging & information extraction
 - Beyond unigram approach and curse of dimensionality:
 - K-grams, fingerprints, NDD, clustering, similarity tools, LSA, etc
 - SVM and ML techniques
 - Query intent classification
 - Statistical tools and data mining
 - Clickthrough and query logs analysis,
 - Modelling & identifying burstiness
 - Time-dependent retrieval
 - Technologies
 - Effective & focused crawling
 - Concurrent Distributed Models
 - Map Reduce indexing and retrieval
 - Compression and index architectures consistent with concurrent solutions
 - Applications
 - Enterprise & Expert Search
 - Social search
 - Local and geographic search
 - News search
 - FAQ search

- OA
- Blog and Opinion Search
- Legal search

È prevista la partecipazione dei ricercatori della

- Fondazione Ugo Bordoni
- Tor Vergata-Nestor
- CNIPA
- IASI-CNR
- Università di Padova
- Università dell'Aquila
- Matrixware, Austria
- Università della Coruna, Spagna
- Università di Glasgow, GB
- Università di Kassel, Germania