

Sunday July 25, 2004

09:00 – 17:00 **Doctoral Consortium** (Halifax Hall of Residence)

08.00 - 19.30 **Registration** (Arts Tower)

08.30 - 12.15 **Tutorials** (Arts Tower, Coffee break 10.00 - 10.30)

- XML Information Retrieval
 - Ricardo Baeza-Yates, *University of Chile*
 - Norbert Fuhr, *University of Duisburg Essen*
- Machine Learning for Text Classification Applications
 - David Lewis
- Multilingual Information Retrieval
 - Fred Gey, *UC Berkeley*
- User Interfaces and Visualisation for Information Access
 - Marti Hearst, *UC Berkeley*

12.15 - 13.15 Lunch (Arts Tower)

13.15 - 17.00 **Tutorials** (Arts Tower, Coffee break 15.00 - 15.30)

- Internet search
 - Jan Pedersen
 - Knut Magne Risvik,
- High Performance Indexing and Query Evaluation for Information Retrieval
 - Justin Zobel, *Royal Melbourne Institute of Technology*
- Bioinformatics and Genomics for Information Retrieval
 - Bill Hersh, *Oregon Health and Science University*
- Text summarization
 - Dragomir Radev, *University of Michigan*

See <http://sigir.org/sigir2004/tutorials.htm> for more information about tutorials.

14.00 - 17.00 **Coach Trip** (Chatsworth House)

19.30 - 21.00 **Reception** (Millennium Galleries)

Monday July 26, 2004 (Octagon Centre & Union of Students Auditorium)

07.30 - 09.00 Newcomer's Breakfast (University House)

08.00 - 18.00 Registration open (Octagon Centre)

09.30 - 10.30 **Opening & Keynote Address** (Octagon Centre)

- Challenges in Using Lifetime Personal Information Stores
 - Gordon Bell, *Microsoft Bay Area Research Center*

10.30 - 11.00 Coffee Break (University House)

11.00 - 12.30 **Session I: Opening session** (Octagon Centre)

- Evaluating High Accuracy Retrieval Techniques
- Scaling IR-System Evaluation using Term Relevance Sets
- Temporal Profiles of Queries

12.30 - 14.00 Lunch (University House)

14.00 - 15.30 **Session IIa: Test collections** (Octagon Centre)

- Retrieval Evaluation with Incomplete Information
- Test collections with no system pooling
- Building an Information Retrieval Test Collection for Spontaneous Conversational Speech

14.00 - 15.30 **Session IIb: Formal models – 1** (Union of Students Auditorium)

- A Formal Study of Information Retrieval Heuristics
- Probabilistic Model for Contextual Retrieval
- Discriminative Models for Information Retrieval

15.30 - 16.00 Coffee (University House)

16.00 - 17.30 **Session IIIa: XML retrieval** (Octagon Centre)

- The overlap problem in content-oriented XML retrieval evaluation
- Length Normalization in XML Retrieval
- Configurable Indexing and Ranking for XML Information Retrieval

16.00 - 17.30 **Session IIIb: Dimensionality reduction** (Union of Students Auditorium)

- Locality Preserving Indexing for Document Representation
- Polynomial Filtering in Latent Semantic Indexing for Information Retrieval
- On Scaling Latent Semantic Indexing for Large Peer-to-Peer Systems

18.00 - 21.00 **Poster and Demo Session with buffet** (Octagon Centre)

Tuesday, July 27, 2004 (Octagon Centre & Union of Students Auditorium)

08.00 - 18.00 Registration opens

08.30 - 10.00 **Session IVa: Formal models – 2** (Octagon Centre)

- GaP: A Factor Model for Discrete Data
- Belief Revision for Adaptive Information Retrieval
- Tuning Before Feedback: Combining Ranking Discovery and Blind feedback for Robust Retrieval

08.30 - 10.00 **Session IVb: Cross-language information retrieval** (Union of Students Auditorium)

- Translating Unknown Queries with Web Corpora for Cross-Language Information Retrieval
- Resource Selection for Domain-Specific Cross-Lingual IR
- Using the Web for Automated Translation Extraction in Cross-Language Information Retrieval

10.00 - 10.30 Coffee Break (University House)

10.30 - 12.30 **Session Va: Language models** (Octagon Centre)

- Dependence Language Model for Information Retrieval
- Parsimonious Language Models for Information Retrieval
- Cluster-based retrieval using language models
- Corpus structure, language models, and ad-hoc information retrieval

10.30 - 12.30 **Session Vb: Clustering** (Union of Students Auditorium)

- Document Clustering by Concept Factorization
- Learning To Cluster Search Results
- Document Clustering via Adaptive Subspace Iteration
- Restrictive Clustering and Metaclustering for self-organizing Document Collections

12.30 - 14.00 Lunch (University House)

14.00 - 15.30 **Session VIa: Text classification** (Octagon Centre)

- Feature Selection using Linear Classifier Weights: Interaction with Classification Models
- Web-page Classification through Summarization
- Parameterized generation of labeled datasets for text categorization based on hierarchical directory

14.00 - 15.30 **Session VIb: Disambiguation** (Union of Students Auditorium)

- Information Retrieval using Word Sense: Root Sense Tagging Approach
- An Effective Approach to Document Retrieval via Utilizing WordNet and Recognizing Phrases
- Web-a-where: Geotagging Web Content

15.30 - 16.00 Coffee Break (University House)

16.00 - 17.30 **Session VIIa: Recognising and using named entities** (Octagon Centre)

- Focused Named Entity Recognition Using Machine Learning
- Learning Phonetic Similarity for Matching Named Entity Translations and Mining New Translations
- Text Classification and Named Entities for New Event Detection

16.00 - 17.30 **Session VIIb: Efficiency and scaling** (Union of Students Auditorium)

- Assigning Identifiers to Documents to Enhance the Clustering Property of Fulltext Indexes
- Filtering Algorithms for Information Retrieval Models with Named Attributes and Proximity Operators
- Hourly Analysis of a Very Large Topically Categorized Web Query Log

19.00 - 23.00 **Banquet** (Cutler's Hall)

Wednesday, July 28, 2004 (Octagon Centre & Union of Students Auditorium)

08.30 - 17.30 Registration open

09.00 - 11.00 **Session VIIa: Content-based filtering & collaborative filtering** (Octagon Centre)

- A Collaborative Filtering Algorithm and Evaluation Metric that Accurately Model the User Experience
- An Automatic Weighting Scheme for Collaborative Filtering
- Using Bayesian Priors to Combine Classifiers for Adaptive Filtering
- A Nonparametric Hierarchical Bayesian Framework for Information Filtering

09.00 - 11.00 **Session VIIb: Image retrieval, users and usability** (Union of Students Auditorium)

- Automatic Image Annotation by Using Concept-Sensitive Salient Objects for Image Content Represent
- A Search Engine for Historical Manuscript Images
- Display Time as Implicit Feedback: Understanding Task Effects
- Human versus Machine in the Topic Distillation Task

11.00 - 11.30 **Coffee Break** (University House)

11.30 - 12.30 **Keynote Address** (Octagon Centre)

- Chemoinformatics: an Application Domain for Information Retrieval Techniques
 - Professor Peter Willett, *University of Sheffield*

12.30 – 14.00 Lunch / **ACM Business Meeting**

14.00 - 15.30 **Session IXa: Machine learning for IR** (Octagon Centre)

- Learning Effective Ranking Functions for Newsgroup Search
- Language-specific Models in Multilingual Topic Tracking
- Web Taxonomy Integration through Co-Bootstrapping

14.00 - 15.30 **Session IXb: Natural language processing** (Union of Students Auditorium)

- Evaluation of an Extraction-based Approach to Answering Definitional Questions
- Query Based Event Extraction along a Timeline
- Sentence Completion

15.30 - 16.00 **Coffee Break**

16.00 - 17.30 **Session X: Web structure** (Octagon Centre)

- Block-Level Link Analysis
- Usefulness of Hyperlink Structure for Query-Biased Topic Distillation
- Block-based Web Search

17.30 - 17.40 **Conference Close**

Thursday, July 29, 2004 (Arts Tower)

09.00 - 16.00 **Workshops**

- New Directions For IR Evaluation
- RIA and "Where can IR go from here?"
- Peer-to-Peer IR
- Information Retrieval in Context
- Search and Discovery in Bioinformatics
- Integration of Information Retrieval and Databases (IR + DB)
- XML and Information Retrieval
- Geographical IR
- Mathematical Formal Methods
- Semantic Web
- Information Retrieval For Question Answering

See <http://sigir.org/sigir2004/workshops.htm> for more information about workshops.