

ECIR 2012. Tutorial programme.

Title: Quantum Information Access and Retrieval.

Brief description

Quantum Theory, or more precisely Quantum Probabilities, offers a formalism based on probabilities, logic and geometry - three ingredients of most Information Access and Retrieval (IAR) models - based on the mathematics of Hilbert spaces. As a generalization of standard probability theory, quantum probabilities offer more expressiveness, which has begun to be leveraged by research in this area to tackle diverse challenges in IAR. In practice, it has been already used for ad-hoc information Retrieval, the diversification of search results, contextual Information Retrieval, and extractive summarization.

The intended audience is researchers in IR who want a detailed understanding of both what the potential is and how to work with such a formalism, from a theoretical and practical point of view. No knowledge of Physics in general, and of Quantum Physics in particular, is required to participate to the tutorial as it will focus on the mathematical formalism that comes from Quantum Physics. Examples coming from Physics will be used to give some background information about the origins of such a formalism. A basic knowledge of probabilities and IR is expected from participants.

During the tutorial, we will detail what Quantum Probabilities are and how researchers in IAR attempt to use them. We will give an overview of the field of research and of its potential. To this end, this tutorial will pay a great deal of attention to how the different concepts of Quantum Probabilities (superposition, subspaces, linear operators, entanglement, interference) that distinguish it from standard probability theory have been used in IAR. Further, the two "quantum" frameworks that have been developed by the presenters will be presented.

The overall goal is to provide enough information and motivation so that participants, by the end of the tutorial, can start to design new Quantum Probability based models and experiment with them. To help experimenting with this formalism, examples of code using the KQP (Kernel Quantum Probability) API will be given.

Learning objectives

1. Provide an overview of the mathematics of Quantum Probabilities;
2. Describe the main uses of this formalism in the current works in the fields of Information Retrieval and Access;
3. Provide computational details and give pointers to the code so that participants can experiment with their own ideas based on Quantum Probabilities.

Duration

A half-day (3 hours)

Presenters

B. Piwowarski (University Pierre et Marie Curie, Paris, France) - <http://www.bpiwowar.net/>

M. Melucci (University of Padua, Padova, Italy) - <http://ww.dei.unipd.it/~melo/>

Title: Designing the Search Experience

Brief description

This tutorial explores the fundamental concepts and principles of User-Centred Design for information search and discovery and demonstrates how to apply them in a range of practical contexts. Participants will learn how to differentiate between various types of search behaviour, develop an understanding of the key dimensions within the search user experience, and discover how to apply UI design patterns to commercial search applications. The session concludes with a group exercise applying these skills to a range of practical design challenges.

Learning objectives

The aim of this tutorial is to deliver a learning experience grounded in good scholarship, integrating the latest research findings with insights derived from the practical experience of designing and optimizing an extensive range of commercial search applications. It focuses on the development of transferable, practical skills that can be learnt and practised within a half-day session.

Participants in this tutorial will:

- Explore the fundamental concepts and principles of Human-Centred Design for information search and discovery
- Study models of human information-seeking behavior (e.g. Bates, Belkin, Jarvelin & Ingwersen, Marchionini, Norman, Sutcliffe & Ennis, etc.), and how to apply interaction design principles based on those models
- Learn how to differentiate between various types of search behaviour: known-item, exploratory, lookup, learning, investigation, etc. and understand how they may be combined to form composite search strategies and patterns
- Develop an understanding of the key dimensions of user type, goal and mode of interaction, and how to apply these dimensions when designing for different user contexts
- Understand the role of design patterns, and how to apply UI design principles and patterns from various libraries in designing search user interfaces
- Gain an awareness of the key design resources available within the HCIR community and how to apply these to practical design challenges

Duration

Half-day (3 hours)

Presenters

Dr. Tony Russell-Rose MBCS CITP, UX Labs and City University London

- personal: <http://isquared.wordpress.com/>

- corporate: <http://uxlabs.co.uk>

Title: Music Information Retrieval 2.0

http://www.cp.jku.at/tutorials/ecir2012_mir20.html

Brief description

Music is an omnipresent topic on the Web, and everyone enjoys listening to his or her favorite tunes. Music information retrieval (MIR) is a research field that aims, among other things, at automatically extracting semantically meaningful information from various representations of music entities, such as a digital audio file, a band's Web page, a song's lyrics, or a tweet about a microblogger's current listening activity. A key approach in MIR is to describe music via computational features, which can be broadly categorized into three classes: music content, music context, and user context.

In this tutorial, we first summarize the ideas behind the three categories of computational features and discuss advantages and disadvantages of each. We subsequently review briefly some standard content-based feature extraction techniques, before focusing on the contextual aspects of music which are accessible through Web technology. To this end, we give an introduction to the field of Web-based MIR and an overview of popular data sources (e.g., Web pages, (micro-)blogs, social networks, user tags, lyrics). Then we present approaches to exploit these sources to (a) mine descriptive and relational metadata (e.g., band members and instrumentation, country, album covers, genres, related artists), to (b) construct similarity measures for music artists and songs based on collaborative and cultural knowledge, and to (c) automatically index and retrieve music.

All presented concepts are illustrated and discussed using exemplary applications and case studies. After this tutorial, the participants will have a solid knowledge of current research in MIR with respect to content-based and Web-based methods, its potential and limitations, and future directions.

Learning Objectives

The half-day tutorial reports on the state-of-the-art in mining music-related information from the Web and further giving the interested audience an introduction to content-based feature extraction. The main goal is to give a sound and comprehensive, nevertheless easy-to-understand, introduction to exploiting Web- and community-based media in the music domain. The presented approaches are highly valuable for tasks and applications such as automated music playlist generation, personalized Web radio, music recommender systems, and intelligent user interfaces to music. Participants will leave the tutorial with a solid knowledge on current research in MIR with respect to content-based and Web-based methods, its potential and limitations, and future directions.

Duration

Half-day (3 hours)

Speakers

Markus Schedl, Johannes Kepler University Linz, Austria - <http://www.cp.jku.at/people/schedl/>

Peter Knees, Johannes Kepler University Linz, Austria - <http://www.cp.jku.at/people/knees/>

Title: Question Answering Systems: History and Architecture

Brief description

The most prevalent way for computer users to find the required information is to surf the Web and search through Internet pages. To this aim, the user expresses his/her information need by a small set of keywords as a query and receives a list of documents which he must then sort through to find the desired answer. Question answering systems, on the other hand, have a different approach. Unlike search engines which receive a set of keywords as an input query and return a list of retrieved documents, the goal of question answering systems is to return a short answer string which exactly addresses the user's question. In the recent past, open-domain question answering has become one of the most actively investigated topics in natural language processing. Its popularity stems from the fact that a user expresses his/her query as a natural language question and receives an exact answer to his/her question rather than being overwhelmed with a large number of retrieved documents.

To build a question answering system, a pipeline of different components is required which constructs the whole architecture of the system. This goal is usually achieved using a combination of three important techniques: (1) natural language processing methods which analyze the input question, (2) information retrieval methods which search for a set of relevant documents from the input corpus and find relevant sentences from the retrieved documents, and (3) information extraction methods which extract candidate answers from the relevant sentences and select the best answer to the input question.

In addition to these techniques which are used in any factoid question answering system, text mining methods are also required for adapting question answering systems to deal with opinion questions such that the system can find opinionated sentences from the text and recognize whether they have a positive or negative polarity.

Learning Objectives

The goal of this tutorial is to provide a comprehensive introduction on question answering systems by describing their importance and usages. We will review the history of question answering systems including the main changes that have been made to improve them from the early stages. In addition, we will describe the question answering tracks at TREC and CLEF and present their evaluation criteria. After this comprehensive introduction, the tutorial will cover the most important components which are required for building a question answering system. We will provide a clear and detailed overview of the state-of-the-art techniques and models that are applied on different components.

Duration

Half-day (3 hours).

Speakers

Dr. Saeedeh Momtazi

Information Systems, Hasso-Plattner-Institut, Potsdam University, Potsdam, Germany

saeedeh.momtazi@hpi.uni-potsdam.de

<http://www.hpi.uni-potsdam.de/naumann/people/smomtazi.html>

Prof. Dietrich Klakow

Spoken Language Systems, Saarland University, Saarbrücken, Germany

dietrich.klakow@lsv.uni-saarland.de

<http://www.lsv.uni-saarland.de/klakow.htm>

Title: From Expert Finding to Entity Search on the Web

<http://diuf.unifr.ch/xi/EntitySearchTutorial>

Brief description

Instead of merely exploiting the syntactic structure of the Web and its documents, it is now possible to leverage semantic information about Web resources. Today, Web Search goes beyond finding documents, and evolves into an interface for finding Web-mediated solutions for user goals involving any type of entity. Therefore, current research challenges include the extraction of information buried within Web pages that can be aggregated and presented to the end user. One example is the query "New York restaurants" which is not aiming at a ranked list of Websites any of which could provide information about a restaurant in New York; rather, the user would be best satisfied by a list of entities with additional information such as the average price and, possibly, a map displaying the results.

To go beyond current Web search, the next step is to rank, rather than documents, information units of varying type, complexity, and structure, that is, entities. Being able to retrieve entities rather than documents would also allow current Search Engines to answer more complex user queries such as "Countries where I can pay in Euro" or "Italian Nobel Prize winners". In the commercial setting, new prototypes going in this direction are being developed.

In this full-day tutorial (6 hours) we present an overview of techniques and algorithms for different entity search tasks: Expert Finding, Entity Ranking in Wikipedia, and ranking structured data extracted from web pages or published as Linked Data.

Learning objectives

The objective is to show how core IR and Semantic Web techniques are making a difference in tasks such as expert finding, entity ranking, semantic search, etc. Additionally, we report results from evaluation studies performed at standard evaluation initiatives such as TREC and INEX. We use the evolution of the Entity Search field through time as a guideline for organizing the tutorial content.

By attending the tutorial, attendants will:

- * Acquire knowledge about the different search tasks involved in the Entity Search area
- * Acquire an understanding of current Expert Finding, Entity Search, and Semantic Search techniques
- * Understand how to properly evaluate Entity Search systems
- * Understand how novel search systems based on semantic data operate

Duration

Full-day (6 hours plus breaks)

Speakers

Dr. Gianluca Demartini (eXascale Infolab, University of Fribourg, Switzerland)

<http://diuf.unifr.ch/main/xi/demartini>

Dr. Peter Mika (Yahoo! Research, Barcelona, Spain)

http://research.yahoo.com/Peter_Mika

Dr. Thanh Tran (Institut AIFB, Universitat Karlsruhe, Germany)

http://www.aifb.kit.edu/web/Duc_Thanh_Tran/en

Prof. Arjen P. de Vries (Centre for Mathematics and Computer Science, CWI, The Netherlands)

<http://homepages.cwi.nl/~arjen/>