Knowledge Orders and Science: Workshop Program

PROGRAMME

08:45-9:00	**Registration, Coffee/Tea**
09:00-9:15	**Welcome: KNOWeSCAPE Project, Its Aims, Scope and Knowledge Orders/Visualization**
	Andrea Scharnhorst, *DANS, Royal Netherlands Academy of Arts and Sciences (Netherlands)*
	SESSION 1: TEXT CLASSIFICATION & VISUALIZATION
09:15-09:45	**Linking Knowledge Spaces**
	Christophe Guéret, DANS, *Royal Netherlands Academy of Arts and Science (Netherlands)*
09:45-10:15	**Disciplines: Implicit and Opposing Structures Overlaying Interdisciplinarity Research**
	Scott Weingart, *Cyberinfrastructure for Network Science Center, Indiana University, Bloomington (USA)*
10:45-11:15	**Using Big Data to Quantify the Evolution of Written Corpora at the Micro and Macro Scale**
	Alexander M. Petersen, *IMT Lucca Institute for Advanced Studies (Italy)*
11:15-11:45	**Coffee/Tea**
	SESSION 2: KNOWLEDGE ORGANIZATION AND CULTURAL HERITAGE
11:45-12:15	**Digital Cultures and Universality in Knowledge Organization (DIGIKO)**
	Widad Mustafa ElHadi & Laurence Favier, *Geriico, University of Lille 3 (France)*
12:15-12:45	**Vision, Understanding and Power: Creating Maps for Cultural Heritage Materials**
	Kathryn La Barre, *Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign (USA)*
12:45-13:30	**Lunch**
	SESSION 3: SCIENCE MAPS

SESSION 3: SCIENCE MAPS

13:30-14:00	Visualizing the Flow of Ideas through Science
	Martin Rosvall, *Department of Physics, Umeå University (Sweden)*
14:00-14:30	World Citation and Collaboration Networks: Uncovering the Role of Geography in Science
	Raj Kumar Pan, *Department of Biomedical Engineering and Computational Science, Aalto University, (Finland)*
14:30-15:00	**Coffee/Tea**
	SESSION 4: VISUALIZATION OF VISUAL ARTS & MUSIC GENRES
15:00-15:30	**A World Without Categories?**
	Lev Manovich, *Graduate Center, CUNY (USA)*
15:30-16:00	**Music Visualization by Audio Content Description**
	Emilia Gomez Guiterrez, *Music Technology Group, Universitat Pompeu Fabra, Barcelona (Spain)*
16:00-17:00	**DISCUSSION SESSION**

ABOUT THE WORKSHOP

Chairs	Almila Akdag Salah (Netherlands) KNOWeSCAPE WG3
	Aida Slavic (Croatia) KNOWeSCAPE WG1
Date	Wednesday, 23 October 2013

Objectives and scope of the workshop

Knowledge Orders and Science is the first joined WG (work group) meeting of the KNOWeSCAPE Cost Action. The aim of this workshop was to set the scene for the overall goals of this Cost Action, which are to bridge new information spaces (such as online data sources like Wikipedia) and traditional institutions, applying new methods of data representation and data analysis, and to explore new ways and interfaces to navigate in complex information spaces.

Organized by the action's two workgroups "Phenomenology of knowledge spaces"(WG1) and "Visual analytics of knowledge spaces" (WG3), one objective of the workshop is to promote cross-community collaboration between these intertwined groups of researchers. Apart from the general issues of knowledge orders and classification, a key issue addressed by the project is the understanding of how expertise in data mining fits with information and knowledge discovery. A special emphasis of this workshop was on the exploration and visualization of different knowledge ordering systems and their behavioural patterns, evolutions and co-evolutions, mappings between these systems and possible application areas in data representation, visualization and interfaces.

The workshop is opened by the principal investigator and the leader of the KNOWeSCAPE action Andrea Scharnhorst, who introduced the COST Action TD1210 and summarised its main goals. The workshop is concluded with a discussion session chaired by Andrea Scharnhorst.

TALKS OVERVIEW

Christophe Guéret

Department of Artificial Intelligence, Vrije Universiteit Amsterdam (Netherlands)

Linking Knowledge Spaces

The movement toward Open Data and the increasing adoption of computational techniques by many scientific fields is leading to the creation of several on-line knowledge spaces. One can now explore the concepts and publications of a given scientific community, browse through the research activities and staff of a university or look at social networks with various centres of interests. While most of these spaces are now isolated, there is a huge interest in interconnecting them into one global knowledge space to be explored, and visualised, as a single entity. This talk will describe how the Semantic Web, and in particular Linked Data technologies, are key to that goal. An overview of relevant concepts will be given before moving to presenting some concrete use-cases and examples.
About the speaker: http://www.few.vu.nl/~cgueret/

Scott Weingart & Cassidy Sugimoto

Disciplines: implicit and opposing structures overlaying interdisciplinarity research

Cyberinfrastructure for Network Science Center, Indiana University, Bloomington (USA)

The recent popularity of maps of science and bottom-up classification schemes have inspired a wave of research into the flow of knowledge within and across disciplines, as well as into the nature of interdisciplinarity. Often implicit in the attempts to measure and define interdisciplinarity are categorizations and classifications of disciplines that themselves remain uncritiqued and at odds with one another. This presentation brings to light the many implicit definitions of discipline and disciplinarity that science mapping research glosses over, and in-so-doing attempts to inspire a more nuanced understanding of the term for future research.

About the speaker: http://www.scottbot.net/HIAL/?page_id=22226

Alexander Petersen

Using big data to quantify the evolution of written corpora at the micro and macro scale

IMT Lucca Institute for Advanced Studies (Italy)

Using the Google Inc. n-gram dataset spanning 200+ years, we show patterns consistent with competitive dynamics at the level of individual words (tokens) as well as at the level of entire corpora. At the micro scale, we demonstrate tipping points in the life-cycle of new words, growth patterns consistent with competition for limited "market opportunities", and evolutionary selection induced by modern editing software (Petersen et al, Sci. Reports 2012). At the macro scale we show that languages "cool as they expand", a dynamic property that highlights periods of political conflict which are characterized by heightened levels of language fluctuations (Petersen et al, Sci. Reports 2013). We will show that these general methods can be extended to other evolving categorical systems such as the MeSH (Medical Subject Headings) vocabulary used by the United States National Library of Medicine.

About the speaker: http://physics.bu.edu/~amp17/

Widad Mustafa El Hadi & Laurence Favier

Digital Cultures and Universality in Knowledge Organization (DIGIKO)

Geriico, University of Lille 3 (France)

The pervasive power of digitization causes scientific, educational, economic and cultural communities to change modes of accessing, sharing and disseminating knowledge and leads to a convergence between our cultural heritage, classic culture and technical culture. It is no surprise that some researchers have called for a "digital humanism", pointing out at the force by which new technologies are becoming a sort of «culture» since they drive us into a new global cultural destiny/context". This presentation derives from the DIGIKO project, submitted by Geriico, University of Lille 3, with our four partners (DFG – Germany, ESRC – UK, NOW – Netherlands). The project aims to analyze, assess and provide a state-of the-art overview of recent vocabulary standards and underlying technologies and the most advanced developments and tools in KOS management and implementation; the most advanced theory and methodology underlying the construction and use of KOS; resource discovery in a digital landscape: digital libraries, repositories, portals, communities of practice, social networks, bibliographic services.

About the speaker: http://www.wmustafa.com/

Kathryn La Barre

Vision, understanding and power: creating maps for cultural heritage materials

Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign (USA)

Abstract: At a time of intense interest in enhanced access to digital cultural heritage resources the obstacles to optimal access seem intractable. This presentation will discuss ongoing work to address the difficulties in creating effective access for two types of cultural heritage materials – films and folktales. This discussion will center upon the role that facets can play in enhancing access to these kinds of complex materials through evaluating the lessons learned from two projects: *Films and Facets* and *Folktales and Facets*. Chief among the lessons that will be discussed is the abiding positive valence of frames of reference, viewpoints and context as guideposts for search.

About the speaker: http://www.lis.illinois.edu/people/faculty/klabarre

Martin Rosvall

Department of Physics,
Umeå University (Sweden)

Visualizing the flow of ideas through science

To comprehend the hierarchical organization of large integrated systems, we have introduced an information-theoretic approach that exploits the duality between compression and pattern detection. By compressing a description of a random walker as a proxy for real flow on a citation network, we find regularities in the network that induce this system-wide flow of ideas. From the pattern of scientific communication, we reveal scientific fields organized in major disciplines and visualize this organization in a multilevel map of science.

About the speaker: http://www.tp.umu.se/~rosvall/

Raj Kumar Pan

Department of Biomedical
Engineering and
Computational Science,
Aalto University (Finland)

World citation and collaboration networks: uncovering the role of geography in science

Modern information and communication technologies, especially the Internet, have diminished the role of spatial distances and territorial boundaries on the access and transmissibility of information. This has enabled scientists for closer collaboration and internationalization. Nevertheless, geography remains an important factor affecting the dynamics of science. Here we present a systematic analysis of citation and collaboration networks between cities and countries, by assigning papers to the geographic locations of their authors' affiliations. The citation flows as well as the collaboration strengths between cities decrease with the distance between them and follow gravity laws. In addition, the total research impact of a country grows linearly with the amount of national funding for research & development. However, the average impact reveals a peculiar threshold effect: the scientific output of a country may reach an impact larger than the world average only if the country invests more than about 100,000 USD per researcher annually.

About the speaker: http://becs.aalto.fi/~rajkp/

Lev Manovich

Graduate Center, The City
University of New York
(USA)

A world without categories?

Computational data analysis produces often more subtle descriptions of the data than traditional classification systems which use language labels to divide the world into discrete classes. Instead, computers can extract multiple "features" from the data, and then position each object in a multi-dimensional "feature space." In this representation (which now used throughout modern societies in every area of business and science), every object occupies a unique position in a multi-dimensional space. The distance between objects in this space represents a "difference" between these objects. I will illustrate this "post-categorical" model by using examples from our work with various cultural data sets including all paintings of van Gogh, 1 million manga pages, and 1 million user-generated artworks.

About the speaker: http://manovich.net/

Emilia Gomez Guiterrez

Music Technology Group,
Universitat Pompeu Fabra,
Barcelona (Spain)

Music visualization by audio content description

There has been a large amount of research within the Music Information Retrieval (MIR) field intended to extract meaningful descriptions from music in audio format, to compute similarity between music pieces and to classify them according to semantic concepts such as mood, style or preference. However, less effort has been devoted to investigate which are the best strategies to present, in a visual way, this information to users with different profiles (e.g. expert musicians and people with no theoretical musical knowledge) and in different contexts (e.g. music listening or education). The main challenges are to provide intuitive visualizations of large music collections, to present information related to different temporal scales (from real-time to global descriptors), and to combine descriptions related to different musical facets such as score, rhythm, tonality or instrumentation. In this talk I will review some relevant approaches to music visualization in terms of tonality, dynamics, tempo, structure, mood and music preference. I will also present how these approaches are being considered in the PHENICX project (http://phenicx.upf.edu) to enrich live music concert performances in classical music. I will finally discuss about the need of multi-scale, personalized and adaptive representations of music collections.

personalized and adaptive representations of music collections.

POSTERS

Ashkan Ashkpour
Erasmus University, Rotterdam (Netherlands)
Albert Meroño-Peñuela
Vrije University, Amsterdam (Netherlands)

The CEDAR Project: Classifying the Dutch Historical Censuses

The censuses are a rich source of historical information for researchers providing demographic, social and economic structures, yielding a wealth of data on many issues in the course of time. The Dutch historical censuses are currently digitized, but notoriously difficult to compare, aggregate and query in a uniform fashion: meaningful historical information is currently hidden in thousands of disconnected Exel Files and over 2,300 tables of aggregated data. The CEDAR project (eHumanities group) aims at enabling greater access and use of this dataset by applying a specific datamodel (exploiting the Resource Description Framework RDF technology), to make census data interlinkable with other hubs of historical socioeconomic and demographic data; and various harmonization practices. A large part of census data harmonization depends on the classification of the data. Querying these RDF data, we create visualizations in order to explore the thousands of variables in our data set and create bottom up classifications for housing variables, occupations, religious denominations, and so on. These visualizations correspond to different moments in history. We leverage animation techniques to display the conceptual changes that modified the social landscape in fundamental centuries of Europe's history.

Junte Zhang
Matthijs BrouwerHennie BrugmanMatthijs DroesMarc Kemps-SnijdersJan Pieter Kunst
Nicoline van der Sijs
René van Stipriaan
Erik Tjong Kim Sang
Rob Zeeman
Meertens Institute, Royal Netherlands Academy of Arts and Sciences (Netherlands)

Nederlab: visual analytics in a virtual research environment for humanities

Nederlab (www.nederlab.nl) is a virtual research environment or laboratory for research on the patterns of change in the Dutch language and culture. Linguists and historians could use Nederlab to research Dutch language and cultural heritage by searching for and having interactive access to large amounts of historical texts and rich and structured metadata describing these resources. The text collections covered by Nederlab include literature i.e. fiction and non-fiction resources, massive amounts of newspaper articles, and the list of collections is set to increase. We demonstrate as example a concrete scenario for literary scholars, and show when, how and which visual analytics on metadata are powerful tools for exploring, finding, collecting and analyzing these texts for (historical and language) research. This includes visualizing the temporal and spatial dimensions for interactive search, and other contextual information such as the names and gender of authors, and comparative analytics of selected results.