

Scientific report on the ELIAS science meeting 4425

EU Summer School Information Foraging (27-8-2012/7-9-2012)

Meeting convenors: Prof. Wessel Kraaij, Prof. Theo van der Weide

Radboud University Nijmegen

Website: <http://www.ru.nl/is/ifl/@842933/pagina/>

Summary

The EU summer school Information Foraging has been organized for the second year thanks to an Intensive Programme (IP) grant of the Erasmus lifelong learning programme. This year, the additional ELIAS grant enabled us to support the participation of additional students and teachers in order to substantially increase the impact and sustainability of the activity.

The participation of students (32) more than doubled with respect to the first year (2011). In addition, the number of lecturers (13) doubled as well, which resulted in a significant increase in focus and quality of the summer school program.

The scientific theme of the summer school Information Foraging is interactive search in all its aspects. The name of the summer school is inspired by the theory of Information Foraging (Pirulli and Card, 1999). Typical characteristics of an information foraging scenario are:

- *The searcher does not have a clear idea what exactly he is looking for or does not know the correct terminology.*
- *Search session is longer than typical lookup search (could span over days / weeks).*
- *The searcher combines information elements from different web pages.*

The IP presents a wide range of courses that i) help students to master the different scientific disciplines related to interactive search/browsing behavior, task based search, ii) help to understand the interaction process, iii) introduce the technology required to support effective interactive search (including user interfaces and scalable search algorithms) iv) provide an overview of classic and more recent evaluation methods and metrics. In addition the school has stimulated the students to combine elements from the different subthemes of the IP in a group assignment.



IP Partners



Sponsors



Scientific content and discussion at event

Wessel Kraaij opened the event and explained the five building blocks of the summer school. From user to system, the following chain of disciplines is covered by the programme: 1) Information seeking behaviour: Understanding the cognitive process and search patterns of users; 2) Interaction: Information foraging is a process where multiple steps are involved, thus understanding interaction (HCI) is important; 3) Interfaces: the user operates with the system through the user interface. A good interface is a key element for a successful system; 4) Systems/ algorithms: Information retrieval models and enhancement techniques such as query expansion, result clustering, spelling correction are important assets for the user interface but critically depend on powerful algorithms.; 5) Evaluation: How can interactive and information foraging systems be evaluated? How can we use crowd sourcing for evaluation?

The first course was given by **Norbert Fuhr** (University of Duisburg-Essen). He gave a tutorial on „Information Retrieval Models“. He first introduced some basic concepts like representations used in IR, probabilistic event spaces and the Probability Ranking Principle, after which he presented some well-known probabilistic models like the binary independence retrieval model, BM25 and language models. The evaluation lecture introduced the concepts of relevance, distributions and preferences, which served for the definition of classic measures like recall, precision and fallout as well as for different variants of cumulated gain. The next part addressed interactive retrieval, discussing the Interactive Probability Ranking Principle as well as the Quantum Probability Ranking Principle along with application examples. The final lecture on Web retrieval presented Page Rank and the Learning to Rank framework.

Ian Ruthven (University of Strathclyde) presented three lectures on information behaviour, information seeking and information interfaces. The first two lectures covered concepts such as relevance, context and ways to study information behaviour before extending into theories and methodologies for studying information seeking and creating information seeking models. The lecture on interfaces started by sketching the history of IIR systems from catalogs to internet systems and presented ways to think about interface design and how creative interface design could change user search behaviour.

Pia Borlund (Royal School of Library and Information Science) gave a course on Interactive, task-based IR evaluation. The first lecture started out by establishing the focus of interactive (I)IR by relating it to the related research areas of information seeking and information behavior. Hereafter the concept and test instrument of a ‘simulated work task situation’ was introduced, the guidelines and requirements for its design and use were presented and discussed, and illustrative examples of successful and less successful designed and tailored simulated work task situations were shown and discussed in order to gain an understanding of what makes a good simulated work task situation. After this the students were given assignments on how to design and tailor simulated work task situations according to specific user groups. The second lecture concerned the planning and design of IIR evaluations studies. In this course the ‘tool box’ of IIR evaluation studies was presented, and considerations concerning the planning of test design (such as rotation and counterbalancing of search tasks, purpose of protocols, the function and types of tutorials, and the importance of pilot

testing) were discussed with strong emphasis on how IIR evaluation studies can vary in focus and hence must be designed according to the research focus of the study in question.

Sascha Kriewel Fuhr (University of Duisburg-Essen) gave two lectures, the first of which concentrated on social media in information retrieval. The focus of the lecture was on tagging, folksonomies, and the use of social tagging for search and information foraging. Additionally, the concept of searching within and with the help of social online communities was introduced. The second lecture introduced the technology of eye tracking for information foraging research. The lecture provided an overview on the history and technology of eye tracking, before talking about the use of eye tracking in information retrieval evaluations and for gathering implicit feedback data. Example data from eye tracking experiments was analyzed live.

Wessel Kraaij (Radboud University) gave a lecture on the potential use of querylog and clickthrough data for search. Querylog analysis was compared to other sources of relevance and the potential issues of releasing query logs for research were discussed in the context of the AOL scandal. Finally, some examples of possible uses of clickthrough data were illustrated by presenting recent work from the Radboud University research group.

Elaine Toms (University of Sheffield) concluded the first week with three lectures. The first one concerned 'information foraging methods' i.e. browsing strategies and tasks. The second lecture addressed the problem of how to evaluate interactive browsing systems. Her final lecture was devoted to an interactive session where students presented their intermediate results on the group tasks.

Theo van der Weide (Radboud University) gave a lecture on searching in large data collections. Modern Information Research more and more involves processing of (very) large data sets. Therefore distributed storage and distributed processing more and more becomes a standard tool for researchers. In his lecture he discussed Hadoop as a distributed file management system and MapReduce as a effective way to process the distributed files. He presented the typical algorithms that are part of many programs from Information Retrieval researchers. He also presented the numerical stability in the context of very large computations. As a case he discussed the computation of cosine measure, and showed that the obvious implementation is not numerically stable.

Leif Azzopardi (University of Glasgow) combined lectures, group work, and practical session with theory during his time at the summer school. He first presented work on children and their information seeking needs where he focused on introducing Piaget's Stages of Development to provide students with an understanding of the difference between children and adults. Then, presented a series of applications specifically designed for children, which try to account for these differences. Most of the examples presented were from the EU PuppyIR project. Students were then split into break groups to design an application for children and present how their application fits with the child users' stage of development. Students were then given to option of undertaking a systems based tutorial to build a search engine mash up or to undertake an interactive information retrieval experiment. Most students elected to undertaken the experiment in order to get an appreciation for what it is like to participate in such an experiment. During the following lectures, Dr Azzopardi then introduced the theory on the economics of search interaction, which underpinned the experiment. He also, presented how simulation of interaction could be used to investigate a number of interesting problems regarding how users interact with systems. At the end of the day, Dr

Azzopardi returned to present the findings from the experiments that the students undertook earlier in the day. This provided students with unique experience where they saw a walk through of how theory informs experimentation, how the experiment was designed (as well as trying it out) and how to present the main findings of an experiment.

Mohand Boughanem (Université Paul Sabatier) presented a tutorial on “context and personalization”. This course was divided into two parts. The first one concerns context definition and context modelling. It first provides some definitions of context and context-aware IR and then focuses on the different dimensions (factors) that compose the context and the way they are exploited in IR process. To better understand the range and the diversity of the factors that might be useful for context-aware IR various context taxonomies are presented. The second part focused on one of the most exploited dimension of context in information retrieval, namely the user context referred to as user profile. It first presents profile modelling, it includes the information/data sources used to build the profile, the collection strategies (the way the sources are collected, implicitly or explicitly), the internal representation of the profile (keywords or topics/concepts), and then describes how profile is exploited during the search. In practice, the user profile is integrated in one of the three main retrieval steps of the IR process: (1) query reformulation, (2) document re-ranking, (3) document retrieval. Several approaches exploiting the profile in these three steps are presented.

In the course titled “Uncertainty issues in Query Formulation”, **Gabriella Pasi** (University Milano Bicocca) has addressed the problem of dealing with uncertainty in the process of query formulation. To this aim, uncertainty has been considered from both the system side and the user side. In the former case the two problems of spelling corrections and query ambiguity have been considered. In the latter case the problem of how to cope with the user uncertainty in query formulation has been considered; to address this problem the issue of defining flexible query languages (i.e. languages tolerant to vagueness and imprecision in query formulation) has been analysed, and possible solutions have been presented.

Keith Van Rijsbergen (University of Glasgow), presented a special invited keynote to identify open research issues in relation to the topic “Uncertainty in IR”. This lecture was a public lecture with some 60 attendees. The talk started with an introduction to the nature of uncertainty in general: how it might arise and how to measure it. Keith continued with a description of how uncertainty arises in the various components of an IR system, that is, in relevance, indexing, retrieval, and evaluation. He explained the different ways uncertainty can be captured and how it is incorporated in a model for separate components. Handling uncertainty for IR requires the use of probability, logic, and geometry, in information spaces. He finally demonstrated how these three concepts can be integrated when the information space is represented by a vector space, making use of non-classical logic, and non-Kolmogorov probability theory.

Jaap Kamps (University of Amsterdam) presented a course on “Searching Structured Data”. Evaluation is key in information access, yet scientifically much evaluation is focusing on the narrow question of what system is best. All research is guided by research questions, and the choice of method (including evaluation methods) should be determined by the research question at hand. Standard evaluation benchmarks in the Cranfield/TREC paradigm have served our field well by quantifying system effectiveness in a meaningful way and by aligning the research agendas of many groups around the world. But this is currently under challenge by rapid recent developments,

requiring us to rethink our methods. Our content is changing, and structured content -- both in terms of document structure and annotations, as well as the overall collection structure -- is prevalent. Also any action on the Web leaves its trails and rich contextual information is available, especially in a mobile setting. So also the search context is changing -- with rich information about the task, the searcher, and prior interactions becoming available. How do such changes factor into the old search problem?

The final tutorial entitled Information Extraction and Topic Modeling in a Retrieval Context and given by **Marie-Francine Moens** (KU Leuven) was composed of four parts spread over two half days of lecturing. The tutorial focused on content descriptors for text documents that go beyond bag-of-words representations and on their use in probabilistic retrieval models. After an introductory part in which the course themes were extensively motivated, part two of the tutorial especially focused on opinion mining and retrieval, named entity recognition, expert retrieval, entity linking, and relation extraction and retrieval. A demo was given that showed the potential of cross-modal video annotation based on the jointly mining of image and text data, in which the video frames were automatically labeled with the names of the faces appearing in them, the actions of these persons and the scene locations. Part three introduced the concept of probabilistic topic modeling, focusing on the most prominent topic models such as probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA) and studied their integration in retrieval models. Finally in a fourth part cross-lingual topic models based on a bilingual Latent Dirichlet Allocation model were discussed. The value of the topic models was illustrated in cross-lingual retrieval, clustering and categorization of documents. In addition, it was shown how semantically similar words obtained by the cross-lingual topic models can be integrated as useful additional evidences in cross-lingual retrieval models. The lectures were interwoven with suitable small exercises.

The summer school was concluded by student presentations of the group work on the design task of an information foraging system. Six groups presented their work with topics ranging from a system to find experts, a course recommender system and a media firm scenario. All presentations included a scenario, requirements, an interface design, a back end composed of analysis components and a proposal for an evaluation methodology.

Assessment of the results and impact of the event on the future direction of the field

In 2012, the summer school Information Foraging was co-located with the Information Interaction in context 2012 conference, which is thematically very related. Several students stayed in Nijmegen for the full three weeks, in addition many key people from the IliX community gave a course at the summer school.

The summer school consisted of a series of 6hr length tutorials balanced across two days, starting in the afternoon and finishing the next day. This structure promoted interaction between lecturers and facilitated one to one sessions between lecturers and students. In addition the cohesion of the overall school was promoted by a transversal design exercise, where student groups were stimulated to incorporate and apply the material of the complete summer school onto a practical scenario.

The course was awarded with 4ECTS and students were handed out certificates to recognize their attendance.

Each individual lecturer and the school as a whole have been evaluated by each students and the evaluation has been very positive without exceptions. We think that the program of the summer school is complementary to the existing ESSIR summer school, since the focus is on interaction, behavior and context. We expect that the school will help to grow the field of information interaction in context, by recruiting master students for the field and providing an international specialized course for PhD students.

All in all, the summer school has been received by all participants as a very successful event. There will be a third summer school again taking place in 2013 in Nijmegen.

Final programme of the meeting

					Topic	Lecturer
27 August 2012		09:00	-	10:00	Registration	
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Introduction to IP	Kraaij / vd Weide
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Group introduction activity	Kraaij / vd Weide
	lunch	12:30	-	14:00		
		14:00	-	15:00	Information Retrieval Models	Fuhr
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Information Retrieval Models	Fuhr
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Information Retrieval Models	Fuhr
28 August 2012		09:00	-	10:00	Information Retrieval Models	Fuhr
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Information Retrieval Models	Fuhr
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Information Retrieval Models	Fuhr
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	information behaviour	Ruthven
	coffee break	15:00	-	15:15		
		15:15	-	16:15	information behaviour	Ruthven
	coffee break	16:15	-	16:30		
		16:30	-	17:30	information behaviour	Ruthven
29 August		09:00	-	10:00	information seeking and	Ruthven

2012					information interfaces	
	coffee break	10:00	-	10:15		
		10:15	-	11:15	information seeking and information interfaces	Ruthven
	coffee break	11:15	-	11:30		
		11:30	-	12:30	information seeking and information interfaces	Ruthven
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Evaluation of Interactive and Task based IR	Borlund
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Evaluation of Interactive and Task based IR	Borlund
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Evaluation of Interactive and Task based IR	Borlund
30 August 2012		09:00	-	10:00	Evaluation of Interactive and Task based IR	Borlund
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Evaluation of Interactive and Task based IR	Borlund
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Evaluation of Interactive and Task based IR	Borlund
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Social Media and IR	Kriewel
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Social Media and IR	Kriewel
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Social Media and IR	Kriewel
	dinner	??	-	??	Social Dinner	
31 August 2012		09:00	-	10:00	Implicit tagging: Eye Tracking	Kriewel
	coffee	10:00	-	10:15		

	break					
		10:15	-	11:15	Implicit tagging: Eye Tracking	Kriewel
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Implicit Tagging: Query Log Analysis	Kraaij
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Information Foraging	Toms
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Evaluation of non Goal-based Applications	Toms
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Prototyping a Browsing Application	Toms
					Topic	Lecturer
3 September 2012		09:00	-	10:00	Handling Big Data	van der Weide
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Handling Big Data	van der Weide
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Handling Big Data	van der Weide
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Adaptive Interfaces	Azzopardi
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Adaptive Interfaces	Azzopardi
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Adaptive Interfaces	Azzopardi
4 September 2012		09:00	-	10:00	Adaptive Interfaces	Azzopardi
	coffee break	10:00	-	10:15		

		10:15	-	11:15	Adaptive Interfaces	Azzopardi
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Adaptive Interfaces	Azzopardi
	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Context and Personalization	Boughanem
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Context and Personalization	Boughanem
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Context and Personalization	Boughanem
5 September 2012		09:00	-	10:00	Context and Personalization	Boughanem
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Context and Personalization	Boughanem
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Context and Personalization	Boughanem
	lunch	12:30	-	13:30		
		13:30	-	14:30	Uncertainty issues in query formulation	Pasi
	coffee break	14:30	-	14:45		
		14:45	-	15:45	Uncertainty issues in query formulation	Pasi
	coffee break	15:45	-	16:00		
		16:00	-	17:15	Modeling Uncertainty for IR (keynote)	van Rijsbergen
		17:15	-	18:30:00	Keynote reception (sponsored by TNO)	
6 September 2012		09:00	-	10:00	Searching Structured Data	Kamps
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Searching Structured Data	Kamps
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Searching Structured Data	Kamps

	lunch	12:30	-	13:15		
		13:15		14:00	lecturer consultation	
		14:00	-	15:00	Information extraction and linking	Moens
	coffee break	15:00	-	15:15		
		15:15	-	16:15	Information extraction and linking	Moens
	coffee break	16:15	-	16:30		
		16:30	-	17:30	Information extraction and linking	Moens
7 September 2012		09:00	-	10:00	Information extraction and linking	Moens
	coffee break	10:00	-	10:15		
		10:15	-	11:15	Information extraction and linking	Moens
	coffee break	11:15	-	11:30		
		11:30	-	12:30	Information extraction and linking	Moens
	lunch	12:30	-	13:00		
		13:00		14:45	Group presentations	Moens/Kraaij/vd Weide
		14:00	-		Closing, certificates	Moens/Kraaij/vd Weide