Scientific report on the ELIAS science meeting 4846

EU Summer School Information Foraging (19-8-2012/30-8-2013)

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

Radboud University Nijmegen

Website: http://ifl.cs.ru.nl/ip

Summary

The EU summer school Information Foraging has been organized for the third year thanks to an Intensive Programme (IP) grant of the Erasmus life-long learning programme. This year, the additional ELIAS grant enabled us to support the participation of additional students and teachers, increasing the impact and sustainability of the activity.

The participation of students (22) was lower than expected, which was probably due to the fact that there were related summer schools in the same period, which was not the case in 2012 when 32 students participated. We managed to maintain a diverse and strong line-up of twelve lecturers, refining the successful 2012 format.

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 (Pirolli 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 behaviour, 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.

Scientific content and discussion at event

Theo van der Weide 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 **Birger Larsen** (Royal School of Library and Information Science). Search is a major enabler in almost all information access and use. Most current search applications, including web search engines, work because of their root in a long line of research in the field Information Retrieval (IR). The aim of IR systems is to retrieve many or all relevant documents for a user's information need - and at the same time to retrieve as few irrelevant documents as possible. This course introduced the main components of IR systems: how to build search engine indexes, how to match information needs to documents, and how to rank documents effectively. The main matching and ranking models were introduced, including Boolean, Vector Space, language modelling and probabilistic models. Incorporation of user feedback and link popularity in web search was also discussed.

Jaap Kamps (University of Amsterdam) continued with the basics of evaluation. 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?

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 catalogues 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.

Elaine Toms (University of Sheffield**)** continued the interactive IR theme 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.

Sascha Kriewel Fuhr (University of Duisburg-Essen) closed the first week with a lecture on eye tracking. 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 analysed live.

Theo van der Weide (Radboud University) started the second week with a lecture on searching in large data collections. Modern Information Research involves more and more processing of (very) large data sets. Therefore distributed storage and distributed processing increasingly 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) presented an overview of the main theories and formal models of information retrieval and information seeking. This started with an introduction about what is information and what is the value of information from an economic viewpoint. Then proceeded to explain a number of different economic models that had been proposed since the 1960's and how they have shaped Information Retrieval. The presentation focused specifically on describing Information Foraging Theory, Search Economics and the Interactive Probability Ranking Principle, where the emphasis was on applying such theories to Information Seeking scenarios and how to build new models from these theories. The series of lectures concluded with discussions about how the theories could be applied to the student's projects.

Mounia Lalmas (Yahoo Labs) gave a lecture on user engagement, in particular its measurement. User engagement refers to the quality of the user experience that emphasises the phenomena associated with wanting to use an online application longer and frequently. User engagement is a multifaceted, complex phenomenon, giving rise to a number of approaches for its measurement: self-reporting (e.g., questionnaires); observational methods (e.g., desktop actions); and web analytics using online behavior metrics (click-through rates). These methods represent various trade-offs between the scale of the data analyzed and the depth of understanding. She presented various efforts aiming at measuring user engagement, and discussed a number of studies carried out in various domains and scenarios. She also related those aspects of user engagement and its measurement impacting the development and deployment of information foraging approaches and their evaluation.

In the online world, user engagement refers to the quality of the user experience that emphasizes the phenomena associated with wanting to use a web application longer and frequently. User engagement is a multifaceted, complex phenomenon, giving rise to a number of approaches for its measurement: self-reporting (e.g., questionnaires); observational methods (e.g., facial expression analysis, desktop actions); and web analytics using online behaviour metrics. These methods represent various trade-offs between the scale of the data analysed and the depth of understanding. For instance, surveys are hardly scalable but offer rich, qualitative insights, whereas click data can be collected on a large-scale but are more difficult to analyse. Still, the core research questions each type of measurement is able to answer are unclear. This lecture will present various efforts aiming at combining approaches to measure engagement and seeking to provide insights into what questions to ask when measuring engagement. The lecture will emphasise those aspects impacting the development and deployment of information foraging approaches and their evaluation.

The 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 lecture of **Gabriella Pasi** (University Milano Bicocca) has addressed the issue of contextual search. To overcome the “one size fits all” behavior of most search engines, in recent years a great deal of research has addressed the problem of defining techniques aimed at tailoring the search outcome to the user context to the aim of improving the quality of search. The main idea is to

produce context-dependent and user-tailored search results. Search tasks are subjective, and often complex; the user-system interaction based on keyword-based querying and on the presentation of search results as a list of web pages ordered according to their estimated relevance is often unsatisfactory. In this lecture a short overview of the main issues related to contextual search have been outlined.

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.

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

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. Each group had to write a short paper discussing the objective of the system, the interaction and interface aspects, information processing components and its evaluation. This group work was scheduled as a regular activity, allowing for a balanced teaching model with alternated classroom and groupwork sessions. Groups presented their design on the closing Friday afternoon.. Lecturers reported that the group of students had a more participatory attitude than in 2012. We conjecture that the variation between tutorial lectures and group work was beneficial to maintain a good level of attention.

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 student 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.

It is fair to say that like in 2012, the summer school has been received by all participants as a very successful event. The plan is to continue the school as a biannual event alternating with ESSIR years. A steering committee (consisting of Norbert Fuhr, Leif Azzopardi, Pia Borlund and Wessel Kraaij) has been installed to ensure quality and continutity.

Final programme of the meeting

EU Erasmus IP Information Foraging 2013 programme : detailed schedule IP IF 2013

WEEK 1									
							Topic	**Lecturer**	**Room**
	Mon	19 August 2012		09:00	-	10:00	Registration		HG00.071
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Introduction to IP	Theo van der Weide	HG00.071
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Group introduction and assignment	Theo van der Weide	HG00.071
			lunch	12:30	-	13:30			
				13:30	-	14:30	Information Retrieval Systems and Models	Birger Larsen	HG00.071
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Information Retrieval Systems and Models	Birger Larsen	HG00.071
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Birger Larsen	HG00.071
	Tue	20 August 2012		09:00	-	10:00	Evaluation 101: Is it any good and why?	Jaap Kamps	HG00.071
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Evaluation 101: Is it any good and why?	Jaap Kamps	HG00.071
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Jaap Kamps	HG00.071
			lunch	12:30	-	13:30			
				13:30	-	14:30	Information Seeking and Information Behaviour	Ian Ruthven	HG00.071
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Assignment	Ian Ruthven	HG00.071
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Information Seeking and Information Behaviour	Ian Ruthven	HG00.071
	Wed	21 August 2012		09:00	-	10:00	Assignment	Ian Ruthven	HG00.308
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Information Seeking and Information Behaviour	Ian Ruthven	HG00.308
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Ian Ruthven	HG00.308
			lunch	12:30	-	13:30			
				13:30	-	14:30	Evaluation of Interactive and Task based IR	Pia Borlund	HG00.308

			coffee break	14:30	-	14:45			
				14:45	-	15:45	Evaluation of Interactive and Task based IR	Pia Borlund	HG00.308
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Pia Borlund	HG00.308
	Thu	22 August 2012		09:00	-	10:00	Evaluation of Interactive and Task based IR	Pia Borlund	HG00.304
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Evaluation of Interactive and Task based IR	Pia Borlund	HG00.304
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Pia Borlund	HG00.304
			lunch	12:30	-	13:30			
				13:30	-	14:30	Information Foraging	Elaine Toms	HG00.304
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Information Foraging	Elaine Toms	HG00.304
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Elaine Toms	HG00.304
	Fri	23 August 2012		09:00	-	10:00	Evaluation of non global based apps	Elaine Toms	HG00.310
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Evaluation of non global based apps	Elaine Toms	HG00.310
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Elaine Toms	HG00.310
			lunch	12:30	-	13:30			
				13:30	-	14:30	Implicit tagging (eye tracking) / clickdata)	Sascha Kriewel	HG00.310
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Implicit tagging (eye tracking) / clickdata)	Sascha Kriewel	HG00.310
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Sascha Kriewel	HG00.310
WEEK 2									
							Topic	**Lecturer**	**Room**
	Mon	26 August 2012		09:00	-	10:00	Big Data	Theo van der Weide	HG00.071
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Big Data	Theo van der Weide	HG00.071
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Theo van der Weide	HG00.071
			lunch	12:30	-	13:30			
				13:30	-	14:30	Information Foraging Theory	Leif Azzopardi	HG00.071
			coffee break	14:30	-	14:45			

				14:45	-	15:45	Information Foraging Theory	Leif Azzopardi	HG00.071
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Leif Azzopardi	HG00.071
	Tue	27 August 2012		09:00	-	10:00	Search Economic and Interactive PRP	Leif Azzopardi	HG00.071
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Search Economic and Interactive PRP	Leif Azzopardi	HG00.071
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Leif Azzopardi	HG00.071
			lunch	12:30	-	13:30			
				13:30	-	14:30	User Engagement	Mounia Lalmas	HG00.071
			coffee break	14:30	-	14:45			
				14:45	-	15:45	User Engagement	Mounia Lalmas	HG00.071
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Mounia Lalmas	HG00.071
	Wed	28 August 2012		09:00	-	10:00	User Engagement	Mounia Lalmas	HG00.071
			coffee break	10:00	-	10:15			
				10:15	-	11:15	User Engagement	Mounia Lalmas	HG00.071
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Mounia Lalmas	HG00.071
			lunch	12:30	-	13:30			
				13:30	-	14:30	Information Extraction and Linking	Marie Francine Moens	HG00.071
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Information Extraction and Linking	Marie Francine Moens	HG00.071
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Marie Francine Moens	HG00.071
	Thu	29 August 2012		09:00	-	10:00	Information Extraction and Linking	Marie Francine Moens	HG00.307
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Information Extraction and Linking	Marie Francine Moens	HG00.307
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Marie Francine Moens	HG00.307
			lunch	12:30	-	13:30			
				13:30	-	14:30	Retrieval in Context	Gabriella Pasi	HG00.307
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Retrieval in Context	Gabriella Pasi	HG00.307
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Assignment	Gabriella Pasi	HG00.307

	Fri	30 August 2012		09:00	-	10:00	Context and Personalization	Gabriella Pasi / Wessel Kraaij	HG00.304
			coffee break	10:00	-	10:15			
				10:15	-	11:15	Context and Personalization	Gabriella Pasi / Wessel Kraaij	HG00.304
			coffee break	11:15	-	11:30			
				11:30	-	12:30	Assignment	Gabriella Pasi / Wessel Kraaij	HG00.304
			lunch	12:30	-	13:30			
				13:30	-	14:30	Assignment Presentations		HG00.304
			coffee break	14:30	-	14:45			
				14:45	-	15:45	Assignment Presentations		HG00.304
			coffee break	15:45	-	16:00			
				16:00	-	17:00	Certificates, Closing		HG00.304