

Advances in Formal Models of Search and Search Behaviour

Full Day Tutorial

Leif Azzopardi
University of Strathclyde
Glasgow, United Kingdom
leifos@acm.org

Guido Zuccon
Queensland University of Technology
Queensland, Australia
g.zuccon@qut.edu.au

ABSTRACT

Searching is performed in the context of a task and as such the value of the information found is with respect to the task. Recently, there has been a drive to developing formal models of information seeking and retrieval that consider the costs and benefits arising through the interaction with the interface/system and the information surfaced during that interaction. In this full day tutorial we will focus on describing and explaining some of the more recent and latest formal models of Information Seeking and Retrieval. The tutorial is structured into two parts. In the first part we will present a series of models that have been developed based on: (i) economic theory, (ii) decision theory (iii) game theory and (iv) optimal foraging theory. The second part of the day will be dedicated to building models where we will discuss different techniques to build and develop models from which we can draw testable hypotheses from. During the tutorial participants will be challenged to develop various formal models, applying the techniques learnt during the day. We will then conclude with presentations on solutions followed by a summary and overview of challenges and future directions. This tutorial is aimed at participants wanting to know more about the various formal models of information seeking, search and retrieval, that have been proposed. The tutorial will be presented at an intermediate level, and is designed to support participants who want to be able to understand and build such models.

Keywords

Search Behaviour, User Models, Retrieval Strategies, Evaluation

1. INTRODUCTION

In the field of Information Retrieval we have developed lots of models - mainly retrieval models for ranking documents. In this tutorial, however, the focus will be on theories that model the interaction between the user and the system,

thus examining models of interactive information seeking and retrieval. These models not only provide insights into how documents should or could be ranked but also shed light on search behaviours and how the interface/system affects such behaviours. Less attention has been paid to such models but there has been a growing interest in their development. However, building such models introduces numerous complexities and greater appreciation of how people interact with systems. Consequently, developing such models has been hailed as a (somewhat) grand challenge for Interactive Information Retrieval [14]. This is because such models are becoming increasingly important as they form the basis of many performance measures, provide a better understanding of how people interact with such systems, and enable more holistic session/task based evaluation.

The tutorial will be composed of two parts: (1) Theory and (2) Practice. In the theory sessions we will cover how: (i) economic theory, (ii) decision theory (iii) game theory and (iv) optimal foraging theory have been applied to Information Retrieval and Seeking. In the practical sessions we will focus on building various models throughout a series of activities. The day will conclude with a discussion of challenges and future directions.

1.1 Introduction and the Economics of Search

In the first session we will briefly provide the context and motivation for developing formal (by which we mean mathematical) models of Information Seeking and Retrieval (ISR). To begin, we will discuss and describe various conceptual and descriptive models of ISR, including Bates' Berry Picking Model [13] and the ISR framework proposed by Ingwersen and Kalvero [22], along with other such models (e.g. [39, 23, 18]). This will provide the background for the tutorial where we point out the limitations of such models and motivate the need to develop models that are not only descriptive in nature, but predictive and crucially explanatory. To set the context, we will discuss the qualities of information, the poverty of attention, information as a good, and how information can be valued. This is important because information seeking is intrinsically embedded within a task. Following on from this we will review various models based on economics. In this session we will cover:

1. Information as Good and the Value of Information [37].
2. Optimal Search Behaviour and the Pandora's Box Problem [35, 38, 37]

3. Optimal Amount of Information (when to stop and deciding at the margin) [15]
4. Azzopardi's Model of Searching (Assessing vs Querying) [3, 5, 9]
5. Cooper's Model of Searching (User vs System or User vs User) [17].
6. Cost Models for Querying [10]
7. Cost Model for Relevance Feedback/Find-Similar [10]
8. Cost Model for Result Browsing [11]

1.2 Decisions, Games and Foraging

In this session we will consider two contemporary and emerging models (based on Decision Theory and Game Theory) before discussing models based on Information Foraging Theory and then showing how all the models relate together. First we shall show how the PRP can be extended to consider interaction by presenting the Interactive Probability Ranking Principle [20]. Then we will cover the Card Model and the POMPD model. Both of these models are derived from the idea that the user and the system are playing a game - this leads to game theoretic models. Then, we will delve into Information Foraging Theory (IFT), which has become very popular and is often used to motivate experiments, but yet infrequently used in terms of modelling and prediction. Here, we will focus on the patch model, and explain the different implications of the model/theory, and how user behaviour is expected to change under varying circumstances. We will explain how these implications can be generated through a graphical analysis (and for the more math savvy, how they can be analytically derived).

In summary, the topics we shall cover in this session are:

1. Probability Ranking Principle [28].
2. Interactive Probability Ranking Principle [20]
3. Card Model [41, 40]
4. POMPD model [24]
5. Optimal Foraging Theory [34].
6. Information Foraging Theory [26, 27, 30, 31]:
 - Information Scent Model
 - Information Diet Model
 - Information Patch Model
 - Charnov's Marginal Value Theorem [16]
7. IFT Example: Exploration / Exploitation [29, 1]

1.3 A Guide to Modelling

In the third session, we will focus on optimisation models, in general, how to build such models, and how to use them to generate hypotheses using various methods (analytical, graphical and computational) [36, 25]. The essence of such models is to develop a cost function and a gain function given the different interactions that users can perform. Given these functions, it is then possible to establish under what conditions cost is minimized and/or gain is maximized. This section will explore the following topics:

- Decision Making and Optimization Models [21, 25].
- Optimality and Rationality [33, 32, 35].
- What choices are available? What are the limitations?
- What are the interactions, costs and benefits?
- What will vary? Parameters?
- Define the problem and the goal
- Construct a cost and gain function
- What method to choose: analytical, graphically, computational
- Solve, plot, compute
- Draw inferences and generating hypotheses

To help illustrate how to undertake the model building process, we will use an example based on the simple scenario of finding the first highly relevant document in a result list. We will describe how we can characterize the problem, enumerate the different variables, and show how the different variables impact the overall cost and the design of the interface. If time permits, then we will also include some notes on developing simulations to explore such models (i.e., computational approaches).

1.4 Practical Session

The final session of the tutorial will be dedicated to building models. For example, participants will be given a context such as a user trying to find an app on a mobile phone or tablet. The goal here is to find the app that the user wants to use. So the focus will be on building a cost function to model the different ways in which the user can find the app, e.g. search or browse. To add realism to the scenario, we will consider different types of users, ones that can remember where the app is, and those who cant (i.e., best and average/worse cases) as well as consider how screen size and app icon size can be modelled to arrive at an optimal size and layout for such interaction. During the practical session, participants will be encouraged to abstract away the details to form a representation of the problem, and identify the main variables that are likely to influence the interaction (i.e., the number of apps on the phone, the number of apps per screen page, the cost of moving between screens, etc.). These will be used to formulate the costs, then we will be able to reason about when it is better to search for an app and when it is better to browse. We will also consider alternative designs, such as presenting the most used apps first, or a hierarchical browsing structure, and whether they are likely to be more efficient or not, or under what circumstances.

2. INTENDED LEARNING OUTCOMES

By the end of the tutorial, participants should be able to:

- Define and describe the different types of models;
- Explain the rationale for formal models of ISR;
- Define an optimisation model;
- Describe the main contemporary models;

- Explain and infer the predicted user behaviour given the contemporary models;
- Design a formal model and generate hypotheses regarding user behavior.

3. BIOGRAPHY

Leif Azzopardi has been recently awarded a Chancellor's Fellowship at the University of Strathclyde within the Department of Computer and Information Sciences. Prior to this appointment, he was a Senior Lecturer in the School of Computing Science at the University of Glasgow. His research focuses on building formal models for Information Retrieval - usually drawing upon different disciplines for inspiration, such as Quantum Mechanics, Operations Research, Microeconomics, Transportation Planning and Gamification. Central to his research is the theoretical development of statistical language models for Information Retrieval, where his research interests include:

- Models for the retrieval of documents, sentences, experts and other information objects [12, 19];
- Probabilistic models of user interaction and the simulation of users for evaluation [2, 6, 7];
- Microeconomic models of information interaction, specifically how cost and effort affect interaction and performance with search systems [3];
- Methods which assess the impact of search technology on society in application areas such as, search engine bias and the accessibility of e-Government information [8], and;
- Search for fun (i.e. the SINS of users) [4].

He received his Ph.D. in Computing Science from the University of Paisley in 2006, and he received a First Class Honours Degree in Information Science from the University of Newcastle, Australia, 2001. In 2010, he received a Post-Graduate Certificate in Academic Practice and has been lecturing at the University of Glasgow since then. He has given numerous invited talks on Formal Models of Information Seeking and Retrieval throughout the world and lectured at the Information Foraging Summer School (2011, 2012 and 2013) and Symposium of Future Directions in Information Access (2007-2013).

Guido Zuccon is a lecturer within the School of Information Systems at the Queensland University of Technology. His research interests include formal models of search, ranking principles for information retrieval, and retrieval models for health search. Guido has actively contributed to the area of document ranking and search result diversification. During his Ph.D. he performed an extensive analysis of document ranking principles and introduced the quantum probability ranking principle [45, 42] and was the first to empirically evaluate the interactive PRP [43]. His work on formal models of search result diversification based on facility location analysis [44] received the best paper award at ECIR 2013 and then in 2014 he received a best reviewer award at ECIR.

He received a Ph.D. in Computing Science from the University of Glasgow in 2012, and he received a Master in Computer Engineering with *summa cum laude* from the University of Padua, Italy, in 2007. Before joining the Queensland

University of Technology as a lecturer in 2014, he was a postdoctoral research fellow at the CSIRO, Australia.

4. REFERENCES

- [1] K. Athukorala, A. Oulasvirta, D. Glowacka, J. Vreeken, and G. Jacucci. Narrow or broad?: Estimating subjective specificity in exploratory search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, pages 819–828, 2014.
- [2] L. Azzopardi. Query side evaluation: an empirical analysis of effectiveness and effort. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 556–563. ACM, 2009.
- [3] L. Azzopardi. The economics in interactive information retrieval. In *Proc. of the 34th international ACM SIGIR conference*, pages 15–24. ACM, 2011.
- [4] L. Azzopardi. Searching for unlawful carnal knowledge. In *Proceedings of the SIGIR Workshop: Search for Fun*, volume 11, pages 17–18, 2011.
- [5] L. Azzopardi. Modelling interaction with economic models of search. In *Proc. of the 37th ACM SIGIR Conference*, pages 3–12, 2014.
- [6] L. Azzopardi and M. de Rijke. Automatic construction of known-item finding test beds. In *Proceedings of SIGIR '06*, pages 603–604, 2006.
- [7] L. Azzopardi, M. de Rijke, and K. Balog. Building simulated queries for known-item topics: an analysis using six european languages. In *Proc. of the 30th annual international ACM SIGIR conference*, pages 455–462. ACM, 2007.
- [8] L. Azzopardi and V. Vinay. Retrievability: An evaluation measure for higher order information access tasks. In *Proc. of the 17th ACM CIKM*, pages 561–570, 2008.
- [9] L. Azzopardi and G. Zuccon. An analysis of theories of search and search behavior. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR '15*, pages 81–90, 2015.
- [10] L. Azzopardi and G. Zuccon. An analysis of the cost and benefit of search interactions. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval, ICTIR '16*, pages 59–68, 2016.
- [11] L. Azzopardi and G. Zuccon. *Two Scrolls or One Click: A Cost Model for Browsing Search Results*. 2016.
- [12] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '06*, pages 43–50, 2006.
- [13] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.
- [14] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42:47–54, 2008.
- [15] U. Birchler and M. Butler. *Information economics*. Routledge, 2007.
- [16] E. L. Charnov. Optimal foraging: attack strategy of a

- mantid. *The American Naturalist*, 110(971):141–151, 1976.
- [17] M. D. Cooper. A cost model for evaluating information retrieval systems. *Journal of the American Society for Information Science*, pages 306–312, 1972.
- [18] S. Erdelez. Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context*, pages 412–421. Taylor Graham Publishing, 1997.
- [19] R. T. Fernández, D. E. Losada, and L. A. Azzopardi. Extending the language modeling framework for sentence retrieval to include local context. *Information Retrieval*, 14(4):355–389, 2011.
- [20] N. Fuhr. A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11(3):251–265, 2008.
- [21] F. S. Hillier and G. J. Lieberman. Introduction to operations research. NY, US, 2001.
- [22] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-Verlag New York, Inc., 2005.
- [23] C. C. Kuhlthau. Developing a model of the library search process: Cognitive and affective aspects. *RQ*, pages 232–242, 1988.
- [24] J. Luo, S. Zhang, X. Dong, and H. Yang. *Advances in Information Retrieval: 37th European Conference on IR Research, ECIR 2015*, chapter Designing States, Actions, and Rewards for Using POMDP in Session Search, pages 526–537. 2015.
- [25] K. G. Murty. Optimization models for decision making: Volume. *University of Michigan, Ann Arbor*, 2003.
- [26] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.
- [27] H. L. Resnikoff, H. Resnikoff, and H. Resnikoff. *The illusion of reality*. Springer-Verlag New York, 1989.
- [28] S. E. Robertson. The probability ranking principle in ir. *Journal of documentation*, 33(4):294–304, 1977.
- [29] T. Ruotsalo, K. Athukorala, D. Glowacka, K. Konyushkova, A. Oulasvirta, S. Kaipainen, S. Kaski, and G. Jacucci. Supporting exploratory search tasks with interactive user modeling. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*, ASIST '13, pages 39:1–39:10, 2013.
- [30] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT/SIGCHI*, pages 269–276, 1993.
- [31] P. E. Sandstrom. An optimal foraging approach to information seeking and use. *The library quarterly*, pages 414–449, 1994.
- [32] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118, 1955.
- [33] H. A. Simon. Theories of bounded rationality. *Decision and organization*, 1:161–176, 1972.
- [34] D. Stephens and J. Krebs. Foraging theory. *Princeton: Princeton University Press*, 1(10):100, 1986.
- [35] G. J. Stigler. The economics of information. *The journal of political economy*, 69(3):213–225, 1961.
- [36] H. R. Varian. How to build an economic model in your spare time. *American Economist*, 41:3–10, 1997.
- [37] H. R. Varian. Economics and search. *SIGIR Forum*, 33(1):1–5, 1999.
- [38] M. L. Weitzman. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654, 1979.
- [39] T. D. Wilson. Human information behavior. *Informing science*, 3(2):49–56, 2000.
- [40] C. Zhai. Towards a game-theoretic framework for information retrieval. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, pages 543–543, New York, NY, USA, 2015. ACM.
- [41] Y. Zhang and C. Zhai. Information retrieval as card playing: A formal model for optimizing interactive retrieval interface. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, pages 685–694, New York, NY, USA, 2015. ACM.
- [42] G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *Advances in Information Retrieval*, pages 357–369. Springer, 2010.
- [43] G. Zuccon, L. Azzopardi, and C. van Rijsbergen. The interactive prp for diversifying document rankings. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1227–1228. ACM, 2011.
- [44] G. Zuccon, L. Azzopardi, D. Zhang, and J. Wang. Top-k retrieval using facility location analysis. In *Proc. of ECIR*, pages 305–316. 2012.
- [45] G. Zuccon, L. A. Azzopardi, and K. Rijsbergen. The quantum probability ranking principle for information retrieval. In *Proc. of the 2nd ICTIR*, pages 232–240, 2009.