Interactive Information Retrieval: Models, Algorithms, and Evaluation

ChengXiang Zhai
czhai@illinois.edu
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA

ABSTRACT

Since Information Retrieval (IR) is an interactive process in general, it is important to study Interactive Information Retrieval (IIR), where we would attempt to model and optimize an entire interactive retrieval process (rather than a single query) with consideration of many different ways a user can potentially interact with a search engine. This tutorial systematically reviews the progress of research in IIR with an emphasis on the most recent progress in the development of models, algorithms, and evaluation strategies for IIR, ending with a brief discussion of the major open challenges in IIR and some of the most promising future research directions.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**; **Retrieval models and ranking**; *Evaluation of retrieval results.*

KEYWORDS

Interactive information retrieval; mathematical models of retrieval; search engines; user interaction

ACM Reference Format:

ChengXiang Zhai. 2021. Interactive Information Retrieval: Models, Algorithms, and Evaluation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3404835.3462811

1 MOTIVATION

Information Retrieval (IR) is, in general, an iterative process with users interacting with a search engine in various ways to complete an information seeking task. As such, it is highly important to study Interactive Information Retrieval (IIR), where we would attempt to model and optimize an entire interactive retrieval process (rather than a single query) with consideration of many different ways a user can potentially interact with a search engine. Indeed, recent years have seen a rapid increase in research work on a wide range of related topics to IIR such as conversational search and recommendation, online learning to rank, search result diversification, and reinforcement learning for IIR.

SIGIR ’21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462811

The interactive nature of the IR process and the need for query reformulation have been recognized by IR researchers since as early as in 1960s when MEDLARS, the first large-scale IR system was developed, and over years, much research work has been done on this topic. However, today’s search engine applications are still based on algorithms designed mostly for optimizing the results for a single query. A main reason for this is due to the lack of understanding of how to mathematically model the problem of IIR and how to generalize a principle such as Probability Ranking Principle, which has been the foundation of existing ranking algorithms, to deal with interactive retrieval. How to generalize the Cranfield evaluation methodology to quantify performance of an interactive retrieval system is also a challenge that has hindered the progress in developing algorithms for interactive retrieval. Recently, there has been some promising progress in tackling these challenges, including e.g., the Game-Theoretic Framework for IR [20], the Probability Ranking Principle for Interactive IR [6], the Interface Card Model [22, 23], Economics IR models [2], conversational search model [15], and interactive retrieval algorithms such as dynamic retrieval models [19], and conversational question answering [14, 16]. Progress has also been made in evaluation of interactive retrieval systems [3, 11, 12, 21]. While there have been some focused tutorials on a few specific lines of research work (see a detailed discussion in the related tutorials section), there is a lack of a general tutorial that attempts to synthesize all these scattered lines of work, systematically review all the major frameworks, models, and algorithms for IIR developed so far, and identify important future research directions. This tutorial is meant to fill in this gap so as to facilitate further research in this increasingly important topic area.

2 OBJECTIVES

The proposed tutorial will systematically review the progress of research in IIR with an emphasis on the most recent progress in the development of frameworks, models, algorithms, and evaluation strategies for IIR. The tutorial will provide (1) a broad overview of research in IIR including both important early milestones and most recent progress, (2) a systematic review of multiple formal models for IIR using a very general cooperative game framework, including decision-theoretic models such as the Interface Card Model and Probability Ranking Principle for IIR, economics models, conversational search model, and dynamic retrieval models, (3) a review of some representative specific techniques and algorithms for IIR, such as interactive query formulation, various forms of feedback techniques and diversification of search results, (4) a review of multiple strategies proposed recently for evaluating IIR, including particularly user simulation, evaluation as a service, and living lab evaluation, and (5) a discussion of the major open challenges in IIR and some of the most promising future research directions.

The generality of the framework and ideas covered in this tutorial makes it broadly appealing to any graduate students, researchers, and practitioners who are interested in obtaining a complete review of all the major ideas, models, algorithms, and techniques for interactive information retrieval (IIR) developed in the past decades and learning about the major research challenges in IIR. The attendants would be exposed to the rich history of research in IIR done by both information scientists and computer scientists, as well as important milestones in the development of both conceptual and formal frameworks for IIR. They will learn in detail how the IIR problem, broadly including special instances such as conversational search, conversational recommender systems, question answering systems, and intelligent personal assistants, can all be framed mathematically as a sequential Bayesian decision problem involving formal models of users and their behavior as well as optimization of human-system collaboration. They will also learn in detail about the major state-of-the-art formal models, techniques, and algorithms for IIR and how simulation-based evaluation of IIR works. Finally, they will learn about the current trends of technology and important open research challenges in IIR. In general, researchers would benefit from learning about the state of the art in IIR research and interesting new research problems, while practitioners would benefit from being exposed to a wide range of ideas and techniques for improving a search engine application system by enhancing its support for interactions with users.

3 RELEVANCE TO SIGIR

A majority of topics of research papers published in ACM SIGIR conferences are more or less related to IIR, including e.g., query intent understanding and modeling, user task inference, query formulation and refinement, all kinds of feedback techniques, learning to rank from a user’s interaction history, as well as various modes of interaction in IR such as question answering, conversational search, conversational recommendation, mobile IR, multi-media IR, integration of browsing and search, and many variants of search user interfaces. Many papers on user studies are also highly relevant since they contribute to the modeling of users and evaluation of IIR systems.

4 RELATED TUTORIALS

There are many related tutorials to this tutorial, including e.g., a SIGIR’19 tutorial on Building Economic Models and Measures of Search [1], a SIGIR’19 tutorial on Deep Chit-Chat: Deep Learning for Chatbots [18], an IJCNLP’17 tutorial on Neural Approaches to Conversational AI later summarized in a survey article in the Foundations and Trends in IR [7] and further evolved into a SIGIR’20 tutorial on Recent Advances in Conversational IR [8], an ICTIR’17 tutorial on Bandit Algorithms in IIR [9], a SIGIR 2016 tutorial on Online Learning to Rank [10], and tutorials on Conversational Recommender Systems given at SIGIR’20 [13], RecSys’20 [4], and WSDM’21 [5]. Another SIGIR’21 tutorial on Bandit Feedback for IIR [17] provides an in-depth treatment of machine learning algorithms for IIR with a focus on bandit feedback. The proposed tutorial is complementary with all these related tutorials by providing a more complete review of the topic with a focus on general frameworks and models that can often tie those existing tutorials together in a general formal or conceptual framework (e.g., the cooperative game framework can cover both conversational search systems and conversational recommender systems), whereas the existing tutorials have values in treating specific lines of research in more detail, especially in covering many machine learning algorithms in more depth. A tutorial with the same title was offered at SIGIR’20 as a half-day tutorial. This full-day tutorial is a significant expansion of that SIGIR’20 tutorial with elaborated explanation of many topics and some new content from two relevant SIGIR’20 workshops on “Deep Reinforcement Learning for IR" and “Applied Interactive Information Systems," respectively.

5 FORMAT AND DETAILED OUTLINE

The tutorial will be a 6-hour full-day tutorial with breaks scheduled by the SIGIR organizers. The anticipated detailed outline of the tutorial is as follows.

5.1 Background

This part will provide any necessary background to the audience, including especially a high-level historical review of research in IIR, to both give the attendants a high-level overview of the major milestones of research in IIR and prepare them for understanding the main content which starts from the next part.

5.1.1 What is Interactive Information Retrieval? This part will provide a general definition of IIR, characterized in 3 dimensions, i.e., Information, Retrieval, and Interaction, all interpreted very broadly to enable examination of all the variants of IIR in the same general conceptual framework. It also includes a detailed discussion of the different notions of IIR, including Cognitive IR view, Human-Computer Interaction view, and Search Engine application view.

5.1.2 Historical Overview of Research in IIR. This part will give a high-level historical view of all the major research milestones in IIR, divided into the following periods: (1) From 1964 to Early 1970s, covering MEDLARS, MEDLINE, Rocchio’s algorithm, and Bennett’s Design Challenges; (2) From Middle 1970s to Middle 1980s, covering cognitive views of IIR developed at Royal School of Librarianship, Copenhagen and University College, London, especially the ASK Theory developed by Belkin and the THOMAS system; (3) From Middle 1980s to Early 1990s, covering various IIR models such as Berry Picking model, Ellis’ behavioral model, Scatter/Gather, and Okapi system; and (4) From Middle 1990s to Present, covering the newest development in the Web era including TREC tasks, the use of Machine Learning for learning from user interaction data, A/B test, formal models of IIR, and many new applications such as mobile search and conversational search.

5.2 Formal Models for IIR

This part will introduce the Cooperative Game Framework for modeling IIR and use the framework to explain all the major formal models developed for IIR with a focus on explaining the Interface Card Model and the PRP for IIR in detail.

5.2.1 *A cooperative game framework for IIR.*

This part will describe the general cooperative game framework of IIR where the IIR problem is framed as a cooperative game played by a search engine and its users [20]. The framework allows us to formally frame the problem as a Bayesian decision problem, where we formally integrate research in user studies, evaluation, retrieval models, and efficient implementation of IR systems in a single unified framework. The formal framework will be discussed in detail with a simplified instantiation leading naturally to a Partially Observable Markov Decision Process (POMDP) view of IIR, where the state consists generally of the retrieval situation and a formal model of the user. The duality of a user's decision making process and the search engine's decision process will be discussed where we see that the user can also be modeled as following a POMDP with the retrieval environment (collection and search engine) and the retrieval situation as the state. Multiple ways to instantiate all the components of the formal framework so as to make the framework operational will also be discussed.

5.2.2 *Interface card model .*

This part will present the interface card model as a specific instantiation of the cooperative game framework and describe the major ideas and formal mechanisms of interface card model in detail. Multiple special cases of the general interface card model will be discussed in detail including both Markov Decision Process (MDP) and POMDP formulations, plain card models and navigation card models, models of a user's stop action. The PRP for IIR will be shown as a special case covered by the general interface card model under certain assumptions. An algorithm for optimizing the interaction navigation interface using the interface card model will be discussed along with some user study experimental results that show the interface card model can be used to derive an algorithm that can automatically generate an adaptive interface for navigation that is adaptive to both the screen sizes of a display and the system's confidence in modeling a user's information need and that works better than a statically designed adaptive interface.

5.2.3 *Probability Ranking Principle for IIR.*

This part will discuss PRP for IIR in more detail including its derivation, a discussion of the generality of PRP for IIR in modeling various interactive IR tasks and how it can be used to optimize a ranked list.

5.2.4 *Economics IR Models.*

This part will introduce the basic idea of applying Economics to IR models, briefly review the major progress in this area with an emphasis on its modeling of user decisions, which is important for optimizing user interactions. Multiple economics retrieval models will be briefly discussed.

5.2.5 *Conversational Search Models.*

This part will briefly discuss a general framework for conversational search models and a high-level overview of some recent work in this direction.

5.2.6 *Dynamic Retrieval Models.*

This part will introduce the major ideas of the line of work on dynamic retrieval models with an emphasis on the Dual-Agent Stochastic Game (DASG) model and its major contributions in modeling the decision processes of both the user and search engine jointly and modeling the communications between a search engine and its users.

5.3 Techniques and Algorithms for IIR

This part will systematically review the major techniques and algorithms developed for supporting and optimizing an IIR system. Due to the large space of work in this area, the review will focus on the major ideas and the most influential techniques. Whenever it is appropriate, pointers will be given for additional refined tutorials or references for specific lines of work.

5.3.1 *Conceptual models for information seeking and retrieval .*

This part will review a number of important conceptual models for information seeking and retrieval, which can be regarded as informal instantiation of the cooperative game framework and is useful to serve as roadmap for designing and improving practical IIR systems. A number of important conceptual models will be discussed, including Marchionini's Typology of Search Tasks, Taxonomies of Web search and E-com search, Ellis' behavioral model and its extension by Meho and Tibbo, the Berry picking model, Belkin's ASK and Ingwersen's cognitive IR model. This will be followed by a discussion of multiple lines of specific techniques and algorithms.

5.3.2 *Interactive query formulation and refinement.*

This part discusses various strategies for supporting interactive query formulation and refinement.

5.3.3 *Feedback .*

This part will provide a high-level review of various forms of feedback techniques and algorithms, characterized by different ways to obtain user interaction data, different ways to learn from user interaction data, and different objectives in feedback (e.g., active feedback where the goal is to actively learn from the user about his/her preferences). This area has a large number of algorithms proposed, including particularly the online learning to rank algorithms that have been proposed recently. In this tutorial, the coverage will be mainly to show the connection of feedback techniques and the general cooperative game framework and how the framework can naturally suggest many novel ways of doing feedback (e.g., explanatory feedback). Pointers will be given to major references including related tutorials that provide more in-depth coverage of some of the algorithms.

5.3.4 *Diversification of search results.*

This part will systematically discuss the important issue of search result diversification using the cooperative game framework. In particular, four different reasons for diversifying results will be discussed, including (1) redundancy reduction (to reduce user effort), (2) modeling uncertainty of user's preferences (to improve robustness), (3) provide an overview of results (to increase immediate utility), and (4) learn efficiently from user interactions (to increase future utility).Depending on the reason, diversification algorithms need to be designed in different ways. How to derive such algorithms for different scenarios of diversification using the cooperative game framework will be discussed.

5.3.5 *Whole session/page optimization.*

This part will briefly review the most recent progress in optimizing the whole page of search results. The work can be regarded as a special case of the interface card model where features are used to allow supervised machine learning to be used for optimizing a search result page. Some work attempting to optimize results over a whole search session will also be discussed.

5.4 Evaluation of IIR

This part will discuss the challenges in evaluating IIR and cover various strategies for evaluating IIR quantitatively with a focus on the use of user simulation for evaluating IIR.

5.4.1 Challenges in IIR Evaluation . This part briefly discuss the major challenges in IIR evaluation, especially those in generalizing the Cranfield evaluation methodology to evaluate an arbitrary interactive retrieval system.

5.4.2 Simulation-based Evaluation . This part will cover the general idea of evaluating an IIR system using user simulators, which can be shown as a natural generalization of Cranfield evaluation methodology for doing component evaluation of IIR systems. General evaluation metrics based on IR simulation as well as how they can cover traditional IR measures will be discussed. A case study of using this kind of evaluation methodology to evaluate a simple IIR system will be presented.

5.4.3 Formal models for user simulation. This part will briefly review the recent work on building simulators of retrieval system users including how to simulate user's queries, query reformulation, and user's clicking behavior.

5.4.4 Other strategies of IIR evaluation. This part will review some additional strategies of IIR evaluation, including Belkin's three-level evaluation, evaluation as a service, and living lab evaluation.

5.5 Summary

This part will summarize the major points of the tutorial with the main take-away messages and a discussion of remaining challenges and future research directions.

6 PRESENTER BIOGRAPHY

ChengXiang Zhai is a Donald Biggar Willett Professor in Engineering of Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC), where he also holds a joint appointment at Carl R. Woese Institute for Genomic Biology, Department of Statistics, and School of Information Sciences. His research interests include intelligent IR, text mining, natural language processing, machine learning, and their applications. He has published over 300 papers in these areas. He served as Associate Editors for major journals in multiple areas including information retrieval (ACM TOIS, IPM), data mining (ACM TKDD), and medical informatics (BMC MIDM), Program Co-Chairs of NAACL HLT'07, SIGIR'09, and WWW'15, and Conference Co-Chairs of CIKM'16, WSDM'18, and IEEE BigData'20. He is an ACM Fellow and a member of ACM SIGIR Academy. He received numerous awards, including ACM SIGIR Test of Time Paper Award (three times), the 2004 Presidential Early Career Award for Scientists and Engineers (PECASE), an Alfred P. Sloan Research Fellowship, UIUC Rose Award for Teaching Excellence, and UIUC Campus Award for Excellence in Graduate Student Mentoring. He has graduated 36 PhD students and over 50 master students. He has two MOOCs on Coursera on Information Retrieval and Text Mining, respectively. He has given many tutorials, including a tutorial on Statistical Language Models for IR at HLT-NAACL'04, SIGIR'05, SIGIR'06, and HLT-NAACL'07, a tutorial on Axiomatic Analysis and Optimization of IR Models at ICTIR'13 and SIGIR'14, a handson tutorial on MeTA Toolkit for Text Retrieval and Mining at KDD'17, a tutorial on Probabilistic Topic Models for Text Data Retrieval and Analysis at SIGIR'17 and SIGIR'18, and a tutorial on IIR: Models, Algorithms, and Evaluation at SIGIR'20.

REFERENCES

[1] Leif Azzopardi, Alistair Moffat, Paul Thomas, and Guido Zuccon. 2019. Building Economic Models and Measures of Search. In *Proceedings of ACM SIGIR 2019*. 1401–1402.

[2] Leif Azzopardi and Guido Zuccon. 2018. Economics models of interaction: a tutorial on modeling interaction using economics. (2018).

[3] Ben Carterette, Ashraf Bah, and Mustafa Zengin. 2015. Dynamic test collections for retrieval evaluation. In *Proceedings of the 2015 international conference on the theory of information retrieval*. 91–100.

[4] Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2020. Tutorial on Conversational Recommendation Systems. In *Fourteenth ACM Conference on Recommender Systems*. 751–753.

[5] Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2021. WSDM 2021 Tutorial on Conversational Recommendation Systems. In *Proceedings ACM WSDM 2021*. 1134–1136.

[6] Norbert Fuhr. 2008. A probability ranking principle for interactive information retrieval. *Information Retrieval* 11, 3 (2008), 251–265.

[7] Jianfeng Gao, Michel Galley, Lihong Li, et al. 2019. Neural approaches to conversational ai. *Foundations and Trends® in Information Retrieval* 13, 2-3 (2019), 127–298.

[8] Jianfeng Gao, Chenyan Xiong, and Paul Bennett. 2020. Recent Advances in Conversational Information Retrieval. In *Proceedings of ACM SIGIR 2020*. 2421–2424.

[9] Dorota Glowacka. 2017. Bandit algorithms in interactive information retrieval. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. 327–328.

[10] Artem Grotov and Maarten de Rijke. 2016. Online learning to rank for information retrieval: Sigir 2016 tutorial. In *Proceedings of ACM SIGIR 2016*. 1215–1218.

[11] F. Hopfgartner, A. Hanbury, H. Müller, I. Eggel, K. Balog, T. Brodt, G. V Cormack, J. Lin, J. Kalpathy-Cramer, N. Kando, et al. 2018. Evaluation-as-a-service for the computational sciences: overview and outlook. *Journal of Data and Information Quality (JDIQ)* 10, 4 (2018), 1–32.

[12] Rolf Jagerman, Krisztian Balog, and Maarten De Rijke. 2018. Opensearch: lessons learned from an online evaluation campaign. *Journal of Data and Information Quality (JDIQ)* 10, 3 (2018), 1–15.

[13] Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2020. Conversational Recommendation: Formulation, Methods, and Evaluation. In *Proceedings of ACM SIGIR 2020*. 2425–2428.

[14] Chen Qu, Liu Yang, Cen Chen, Minghui Qiu, W Bruce Croft, and Mohit Iyyer. 2020. Open-retrieval conversational question answering. In *Proceedings of ACM SIGIR 2020*. 539–548.

[15] Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. 117–126.

[16] Siva Reddy, Danqi Chen, and Christopher D Manning. 2019. Coqa: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics* 7 (2019), 249–266.

[17] Huazheng Wang, Yiling Jia, and Hongning Wang. 2021. Interactive Information Retrieval with Bandit Feedback. In *Proceedings of ACM SIGIR 2021*. Tutorial Description.

[18] Wei Wu and Rui Yan. 2019. Deep Chit-Chat: Deep Learning for Chatbots. In *Proceedings of ACM SIGIR 2019*. 1413–1414.

[19] Grace Hui Yang, Marc Sloan, and Jun Wang. 2016. Dynamic information retrieval modeling. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 8, 3 (2016), 1–144.

[20] ChengXiang Zhai. 2016. Towards a game-theoretic framework for text data retrieval. *IEEE Data Eng. Bull.* 39, 3 (2016), 51–62.

[21] Yinan Zhang, Xueqing Liu, and ChengXiang Zhai. 2017. Information retrieval evaluation as search simulation: A general formal framework for ir evaluation. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. 193–200.

[22] Yinan Zhang and Chengxiang Zhai. 2015. Information retrieval as card playing: A formal model for optimizing interactive retrieval interface. In *Proceedings of ACM SIGIR 2015*. 685–694.

[23] Yinan Zhang and Chengxiang Zhai. 2016. A sequential decision formulation of the interface card model for interactive IR. In *Proceedings of ACM SIGIR 2016*. 85–94.