What Should We Teach in Information Retrieval?

Ilya Markov
University of Amsterdam
i.markov@uva.nl

Maarten de Rijke
University of Amsterdam
derijke@uva.nl

Abstract

Modern Information Retrieval (IR) systems, such as search engines, recommender systems, and conversational agents, are best thought of as interactive systems. And their development is best thought of as a two-stage development process: offline development followed by continued online adaptation and development based on interactions with users. In this opinion paper, we take a closer look at existing IR textbooks and teaching materials, and examine to which degree they cover the offline and online stages of the IR system development process. We notice that current teaching materials in IR focus mostly on search and on the offline development phase. Other scenarios of interacting with information are largely absent from current IR teaching materials, as is the (interactive) online development phase. We identify a list of scenarios and a list of topics that we believe are essential to any modern set of IR teaching materials that claims to fully cover IR system development. In particular, we argue for more attention, in basic IR teaching materials, to scenarios such as recommender systems, and to topics such as query and interaction mining and understanding, online evaluation, and online learning to rank.

1 Introduction

While there are several definitions of IR [81], many agree that IR is about technology to connect people to information. In our view, this includes search engines, recommender systems, and task-oriented dialogue systems. Technology changes rapidly. Naturally, this impacts disciplines such as ours that have a strong technology component. The ways in which people interact with information continue to change, with different devices, different search engines, different recommender systems, and different conversational agents catering for different niches and needs. Our understanding of people's information interaction behavior changes, as people interact through increasingly diverse interfaces, leaving behind rich sets of digital trails that go beyond the traditional query, click, timestamp format. New types of data require new analysis and mining methods. And with new information interaction formats, the algorithms underlying IR solutions change, enabling increased levels of personalization and task support.

The changes noted above are sufficiently profound and long-term to encourage us to rethink what material we should be presenting in our IR textbooks and teaching. But there is also a more fundamental perspective that should be integrated into our IR teaching materials. IR systems are best thought of as *interactive systems* – systems that are first developed offline and then put online (in the sense of being exposed to users), where they continue to be developed based on interactions with their users. Thus, the process of developing a modern IR system can be described as a two-step process: the initial development of the system can be called an *offline* phase, while the evolution of the system after its initial deployment can be called an *online* phase.

For those of us who are tasked with teaching future generations of IR practitioners, researchers, and entrepreneurs in a technology-oriented program (CS, AI, IS), this perspective of an IR system as an interactive system forces us to ask whether we have the right teaching materials in IR. Are our IR textbooks up-to-date? Do they cover the materials that we should be teaching in a typical technology-oriented IR course in a CS, AI, or IS curriculum given the rapid changes in technology and the reality of an IR system as an interactive system? If not, what is missing?

Below, we begin by outlining an extensive list of topics that we believe should be included in a modern IR curriculum, divided into offline and online phases, application areas, and the broader impact of IR technology. We then identify which of those topics are covered in today's textbooks. We find that discussions of emerging scenarios are (necessarily) limited, and that treatments of the online development phase and of the impact of IR are largely lacking. Thus, we follow with a section in which we identify topics related to online development of IR systems that future IR textbooks should cover.

It is important to note that in this paper we focus on a list of scenarios and a list of topics in IR rather than on developing specific IR courses or curricula. To create a standalone IR course or a set of connected IR courses, the topics presented in this paper have to be filtered, (possibly) re-grouped, and the presentation level has to be adjusted according to a specific program, whether it is CS, AI, or IS, and to a specific student level, whether it is undergraduate, graduate, or post-graduate, a job that we leave as future work.

2 Scenarios and Topics in IR

We start by listing the scenarios and topics that we believe should be included in the IR teaching curriculum. First, we point out that today's IR scenarios are very diverse and go way beyond search. Second, we focus on the search scenario as the most studied and best understood one, and give an extensive list of topics that we believe should shape current and future IR teaching.

2.1 Scenarios

In Table 1, we list IR *scenarios* that we believe should be a part of IR teaching. The scenarios that we consider are different instantiations of a context of a user interacting with technology so as to obtain information. Parameters that can be varied include the type of input provided by the user (query, profile, ...), the type of device used (desktop, mobile, screenless, ...), the type of results returned (documents, answers, snippets, ...), and the type of interaction supported (clicks, swipes, ...).

We believe that today's IR teaching materials should cover a number of scenarios beyond traditional search, where a user has access to a keyboard, enters a keyword query, and is offered a ranked list of documents as a response. Additional scenarios that deserve attention in our basic IR teaching include recommender systems (where the user is the query), mobile search (where query entry, result presentation, and interaction signals differ from the traditional search scenario), question answering (with a question instead of a query and answer(s) instead of documents containing the answer), and conversational search (with stateful interaction with result(s) possibly generated instead of retrieved, possibly screenless).

There is a set of teaching materials that cover IR scenarios beyond search. For instance, recommender systems and question answering are covered extensively in, e.g., [1] and [29, Chapter 23], respectively, mobile search has a dedicated tutorial [45] and survey [26], and even a young topic such as conversational search is also covered in a tutorial [30]. By and large, however, extensive teaching materials for these scenarios are yet to be developed.

In the rest of the paper, we focus on one scenario, namely traditional search, and use it as an example to demonstrate the variety of topics that we believe should be included in the IR teaching curriculum. But the discussion and arguments presented in the rest of the paper are also applicable to other scenarios mentioned in Table 1.

Table 1: Scenarios to be covered in IR teaching.

Scenario	Description
Conversational search	Taxonomy of conversational agents (task-oriented, open-dialogue, etc.), text- and voice-based agents, approaches to conversational search, retrieval vs. generation, evaluation of conversational search
Mobile search	Interaction signals, result presentation, contextual awareness, evaluation of mobile search
Question answering	Answer passage retrieval, information extraction, answer extraction, answer ranking, reading comprehension
Recommender systems	Collaborative filtering, content-based recommendation, user profiles, matrix factorization, cold-start problem
Search	Any aspect of the traditional desktop-based search engine pipeline, its development, usage and evaluation

2.2 Topics

In Table 2, we list teaching *topics* for the search scenario. This list is generated by considering existing IR textbooks, online courses and video lectures, surveys, conference and summer school tutorials and discussions with colleagues from both academia and industry. The textbooks considered are listed in Appendix A, while surveys and tutorials are listed in Appendix B.

Topics in Table 2 are categorized as follows. First, we list topics that discuss the offline phase of the IR system development, i.e., the phase of building the system from scratch till the moment of its deployment. We categorize a topic as offline if it covers algorithms and methods that do not require access to a live IR system. Offline topics cover data acquisition, processing and storage, offline evaluation, scoring and matching methods that can be computed and/or learned offline, etc. Second, we list topics that are concerned with the online phase of the IR system development, i.e., the phase where an already deployed system is constantly improved based on the way users interact with it. Such topics cover algorithms and methods that rely on a live IR system and cannot be applied otherwise. Online topics include queries and interactions, online learning and evaluation, personalization, etc. Third, we list IR applications, such as entity search, multimedia IR, sponsored search, etc. Each of these applications usually contains its own offline and online phases. Finally, we list topics related to the impact of IR technology.

The distinction between offline and online topics is, of course, limited. First, some topics include algorithms from both offline and online categories. For example, query expansion can be done offline using controlled vocabularies and thesauri, but can also be performed online based on query logs. In such cases, we categorize topics according to our personal preferences. Second, the offline phase of the IR system development does not stop after the system deployment. Clearly, indexing and offline evaluation must be used throughout the entire system life-cycle, offline scoring and ranking methods must be continuously improved to keep up with new types of information, etc.

In Table 2, we further group topics by their function to simplify reading, such as data acquisition and storage, retrieval and ranking, evaluation, etc. The table also shows how each topic is covered in the most recent and widely used IR textbooks [6, 15, 27, 56] and the online course [79]. The following four-point scale is used to denote coverage: not covered ($-$), partly covered ($\pm$, one or two subsections of a textbook), covered ($+$, from three subsections to one chapter), extensively covered ($*$, more than one chapter). The information on exact coverage of each topic is given in Table 3, Appendix A.

Table 2: Topics to be covered in IR teaching. The topics are grouped by category: offline phase, online phase, IR applications, and impact of IR. Within each category the topics are further grouped by their function, e.g., data acquisition and storage, retrieval and ranking, evaluation, etc. The latter grouping is used to simplify reading the table. The five columns on the right show the coverage of topics in the most recent and widely used IR textbooks, namely, Baeza-Yates and Ribeiro-Neto [6], Büttcher et al. [15], Croft et al. [27], Manning et al. [56], and in the online course by Zhai [79]: $-$ means that a topic is not covered, $\pm$ means that a topic is mentioned in a textbook and the textbook devotes one or two subsections to the topic, $+$ means that a topic is covered in more than two subsections and up to one chapter, $*$ means that a topic is extensively covered in more than one chapter. The exact parts of textbooks devoted to each topic are presented in Table 3, Appendix A.

Category	Topic	Content	[6]	[15]	[27]	[56]	[79]
	Crawling	Crawler architecture, storing the crawled repository, extending the crawled repository, updating the crawled repository, distributed crawling, duplicate detection, spam and quality control	+	+	+	+	±
	Indexing	Inverted index, index construction, index compression, index optimization, term vocabulary/dictionary, query processing on indices, distributed indexing and query processing	+	*	+	*	+

	Link analysis	Structure of the Web, Web graph, PageRank, HITS	+	+	+	∗	+
	Text analysis	Statistical properties of text, Zipf's and Heaps' laws, tokenization, stop-words, normalization, stemming, lemmatization, spelling correction	+	+	+	+	±
	Learning to Rank (LTR)	Ranking signals/features, pointwise/pairwise/listwise approaches	±	+	+	+	+
	Semantic scoring	Latent semantic indexing/analysis (LSA/LSI), probabilisitic LSA, topic models, word embeddings, neural IR, learning to match	+	−	±	+	−
	Term-based scoring	Boolean retrieval, vector space model, TF-IDF, probabilistic retrieval, BM25, language modeling	+	∗	+	∗	∗
	Term dependency	Term-based scoring that considers dependencies between terms	−	−	+	±	−
	Efficiency evaluation	Efficiency criteria, efficiency metrics, stress testing, modeling efficiency criteria	−	+	+	±	−
	Hypothesis testing	Statistical background on hypothesis testing, significance tests in IR	−	+	±	−	±
	Offline evaluation	Test collections, obtaining relevance judgements, offline effectiveness metrics, experimental design	+	+	+	+	+
Offline phase	User studies	Evaluation based on controlled user studies	−	−	−	−	−
	Parallel computing	Parallel and distributed processing of IR algorithms	+	−	−	−	−
	Search optimization	Query scheduling, caching, inexact top-k retrieval, index elimination and ordering, champion lists, cluster pruning, etc.	±	+	−	+	±
	Distributed IR	Resource representation, resource selection, data fusion, score normalization, peer-to-peer IR	∗	−	+	−	−
	Diversification	Diversifying search results in terms of content and according to possible user intents	−	−	−	−	−
	Search interfaces	Query interfaces, results presentation (blended, tabbed, etc.), snippets, presentation of heterogeneous content (verticals, direct answers, multimedia and structured information)	∗	−	+	+	−
	Text classification	Probabilistic classifiers, linear classifiers, unsupervised and supervised classification, feature selection, applications of text classification	+	∗	+	∗	−
	Text clustering	Flat clustering, hierarchical clustering	−	−	+	∗	−
	Intent	Query taxonomy, vertical intents, other intents (freshness, entity, direct answer, etc.)	±	±	−	±	−
	Query analysis	Query understanding, query suggestion, query auto-completion	−	−	−	−	−
	Query expansion	Controlled vocabularies, manual and automatic thesauri, query-log-based expansion	+	−	±	+	−

	Sessions and tasks	Short- and long-term sessions, tasks and missions, methods for satisfying corresponding information needs	±	−	−	−	−
	User profiling	Long-term interests, short term interests, behavioral patterns	−	−	−	−	
	Biases	Results examination biases (position bias, attraction/attention bias, etc.), people's beliefs and unconscious biases	−	−	−	−	−
	Interactions	Taxonomy of interactions, modeling and predicting interactions, applications of interaction data	±	±	−	−	−
Online phase	Logging	What to log, logging techniques, storing and accessing logged data, data anonymization	−	−	+	−	−
	Counterfactual LTR	Learning to rank based on historical/logged user interactions	−	−	−	−	−
	Online LTR	Just-in-time learning to rank based on an incoming stream of user interactions	−	−	−	−	−
	Personalization	Personalizing search results for individual users and groups of users	−	−	−	−	−
	Relevance feedback	Relevance feedback for term-based scoring methods (vector space model, probabilistic retrieval, language models), pseudo-relevance feedback, indirect/implicit relevance feedback	+	+	±	+	+
	Counterfactual evaluation	Offline evaluation based on historical/logged user interactions	−	−	−	−	−
	Online evaluation	A/B testing, interleaving, online effectiveness metrics	+	±	±	±	−
	Entity search	Entities, knowledge base, knowledge graph, entity linking	−	−	±	−	−
	Multimedia IR	Content-based image retrieval (color, texture, etc.), audio and music retrieval, video retrieval	+	−	+	−	−
	Product search and recommendation	Collaborative filtering, content-based recommendation, cold-start problem, review analysis	−	−	±	−	+
Applications	Retrieval of semi-structured documents	Query languages for semi-structured documents, indexing and query processing, scoring and ranking, evaluation	+	+	±	+	−
	Sponsored search	Business models of sponsored search, placing ads, user behavior in sponsored search	−	−	−	−	−
	Vertical search	Verticals, vertical selection, results presentations, complex SERPs, whole page optimization, evaluation	−	−	−	−	−
	Web search	Scalability, matching, ranking, quality control, evaluation	*	+	+	*	+
Impact of IR	Cognition	Search effects, search as learning, cognitive effects, filter bubble	−	−	−	−	−

Privacy	Differential privacy, anonymization, sampling	–	–	–	–	–
Responsible IR	Fairness, accountability, transparency	–	–	–	–	–

3 Today's IR Textbooks

In this section, we discuss how the topics presented in Table 2 are covered in today's IR teaching materials. Specifically, we focus on the most recent and widely used textbooks, namely, those by Baeza-Yates and Ribeiro-Neto [6], Büttcher et al. [15], Croft et al. [27], Manning et al. [56], and on the online course by Zhai [79]. We split the discussion according to the categories in Table 2: offline phase, online phase, IR applications, and impact of IR.

3.1 Offline phase

The offline phase is concerned with building an IR system from scratch till the moment of its deployment. The corresponding topics, thus, cover algorithms and methods that do not require access to a live IR system. These topics constitute the largest part of Table 2 and are grouped based on their function as follows: (i) data acquisition, processing and storage; (ii) scoring and ranking; (iii) evaluation; (iv) efficiency; and (v) other fundamental topics in IR, such as diversification, query expansion, classification, etc.

The topics describing data acquisition, processing and storage, namely crawling, indexing, link analysis, and text analysis, are fully covered in all considered textbooks and are mostly covered in the online course. Moreover, indexing is extensively covered in [15, 56, 76]. In our own teaching, we also use an excellent recent survey of data acquisition and storage techniques in web search by Cambazoglu and Baeza-Yates [17].

The core part of the scoring and ranking group, namely term-based scoring (i.e., TF-IDF, BM25, language modeling, etc.) is covered in all considered teaching materials and extensively in [15, 56, 79]. Another topic of this group that is mostly covered is Learning to Rank (LTR). However, the considered textbooks only cover a very early LTR method, namely ranking SVM [43], while the discussion of more modern LTR techniques, such as RankNet and LambdaRank [14], and other methods surveyed in [55], is missing. Other topics related to scoring and ranking are only partly covered, e.g., term dependency in [27, 56] and semantic scoring in [15, 27, 56]. This is natural as some topics, such as word embeddings for ranking [2] and neural IR [48, 58, 60] have only been introduced recently. Given the above observations, we believe that today's IR textbooks have to be updated to include recent developments in scoring and ranking, especially in the fields of LTR and semantic learning to match (especially, neural IR).

The topics related to evaluation during the offline phase of the IR system development are mostly covered in today's teaching materials. One important aspect that is relatively underrepresented is experimental methodology and study design; Sakai [67] fills this gap. Another important aspect of evaluation that deserves more attention is user studies; an excellent survey by Kelly [47] covers this aspect. Overall, today's IR teaching materials, with the addition of [47, 67], fully cover offline evaluation.

The efficiency aspect of IR systems is also present in the textbooks considered and in the online course. Specifically, search optimization techniques are either mentioned or fully covered in most teaching materials, while distributed processing of IR algorithms is extensively discussed in Büttcher et al. [15]. We believe this material is enough to represent the topic within the IR teaching curriculum.

Finally, other fundamental topics, such as distributed IR, search interfaces, and text clustering, are fully (and sometimes extensively) covered in at least two textbooks each, while text classification is covered in all textbooks and extensively in [15, 56]. We would also like to note that there is an excellent book on search interfaces by Hearst [36]. The only topic missing in today's IR teaching materials seems to be diversification. This gap is filled with a recent survey by Santos et al. [68]. We also have to note that some topics in this group could also be placed in other categories and groups, e.g., a part of distributed IR can be attributed to search optimization, while diversification and search interfaces are tightly connected to users and could be a part of the online phase. However, we believe that the algorithms behind diversification and search interfaces do not require access to a live IR system and, thus, we categorize these topics as offline instead of online. We also consider distributed IR as a separate search paradigm rather than an optimization technique and, thus, place this topic under other fundamental topics.

Overall, the offline phase of the IR system development is well covered by today's teaching materials. Some topics, such as semantic scoring and LTR, have been advanced considerably since the publication of the textbooks. We believe that recent developments in those areas have to be included in the next editions of the textbooks. Additionally, we argue for including such topics as user studies and diversification, which are now completely missing in the IR teaching materials.

3.2 Online phase

The second largest part of Table 2 covers the online phase of the IR system development. Online topics include algorithms and methods that rely on and can only be used within a live IR system. These topics are grouped as follows: (i) querying, (ii) interacting with results, (iii) scoring and ranking, and (iv) evaluation.

Topics related to querying, such as query analysis, user intents, sessions and tasks, user profiling (used, e.g., in the case of proactive or zero-query retrieval), have from little to no coverage in the textbooks and online course considered. Some textbooks [6, 15, 56] mention a basic taxonomy of query intents, i.e., informational, navigational and transactional queries [11]. However, a wider notion of user intents, such as vertical intent, freshness, entity intent, etc., is not discussed in the textbooks.

Topics related to user interactions are partly covered in the textbooks. Baeza-Yates and Ribeiro-Neto [6] and Büttcher et al. [15] mention click interactions and their potential for improving IR systems, while Croft et al. [27] discuss problems related to logging. However, there is way more to logging and interactions that has been recently developed in both academia and industry [74].

In terms of scoring and ranking within the online phase, the only covered topic is relevance feedback, because it is tightly connected to offline scoring and ranking methods and is usually discussed in connection with term-based scoring. However, relevance feedback is online in nature, because it may only happen when users interact with a functioning IR system.[1] Other online scoring and ranking topics, such as counterfactual and online LTR and personalization, are not discussed in the considered teaching materials.

[1]Here, we refer to true relevance feedback as opposed to pseudo-relevance feedback, which, of course, can be performed during the offline phase.

Finally, in terms of evaluation, most books mention A/B testing, while Baeza-Yates and Ribeiro-Neto [6] cover it in detail. At the same time, counterfactual evaluation [44] and other types of online evaluation, such as interleaving [19], are not covered in the considered teaching materials.

In summary, the online phase of the IR system development is covered in much less detail than the offline phase. In fact, most topics are either not covered or covered only marginally. At the same time, for interactive systems (which IR is an instance of) their evolution after deployment is as important as the offline development of an initial working system. In Section 4, we fill this gap by bringing together the existing teaching efforts related to the online phase and outline specific topics to include into future versions of IR textbooks and courses.

3.3 Applications

IR is applied in many areas and no course in IR is able to cover all the field's many applications. Having said so, there is a considerable amount of attention, both in textbooks and in surveys and tutorials for entity search [7, 9], multimedia IR [65], product search [66], sponsored search [72], vertical search [5], and, obviously, web search, which is fully covered in all textbooks considered and extensively by Baeza-Yates and Ribeiro-Neto [6] and Manning et al. [56]. Retrieval of semi-structured documents, e.g., XML retrieval, supported by the community's longstanding interest in document structure and by a range of community-based evaluation efforts related to semistructured documents [51], is covered in a relatively large number of today's textbooks.

As mentioned above, no single textbook or course can cover the variety of IR applications. Still, we believe these applications have to be at least mentioned in today's IR teaching materials, devoting some attention to largely missing topics of entity search, recommender systems, sponsored search, and vertical search.

3.4 Impact of IR

IR technology affects our lives in many ways. Our algorithmic innovations give rise to many"small-scale" interactions through clicks, swipes, etc., in search, recommendation, and conversational environments, that become informative because of their scale. IR technology underpins social media, where slightly richer interactions such as "likes" can easily be mined to uncover highly personal characteristics [50]. And highly specialized IR technology with fewer but higher-stakes interactions is being deployed in domains as diverse as criminal justice, finance, human resources, medicine, and security. At all levels of impact, IR algorithms can have unintended consequences.

The impact of IR technology, and of algorithms more generally, is gaining attention in the popular press. And in our community we are seeing growing awareness of the issue too [4]. However, as Garcia-Gathright et al. [31] point out, beyond these calls to action, standard processes and tools for researchers and practitioners do not readily exist to assess and address the potential negative impact of IR technology. The underlying topics are diverse, ranging from fairness and accountability to transparency, the literature is relatively scattered, and interdisciplinary approaches seem to be required – all of this means that very different communities are working on the topic and that for many pertinent topics tutorials and surveys are scarce if not absent.

There are (partially) relevant tutorials that we have found. Olteanu et al. [59] survey recent warnings against the naive usage of social data; they highlight that there are *biases* and inaccuracies occurring at the source of the data, but also introduced during data processing pipeline; there are methodological limitations and pitfalls, as well as ethical boundaries and unexpected consequences that are often overlooked. Garcia-Gathright et al. [31] describe an early approach to assessing and addressing *algorithmic and data bias* in practice that attempts to translate the literature into processes for (production) teams wanting to assess both intended data and algorithm characteristics and unintended, unfair biases. Hajian et al. [35] present a general tutorial, not limited to IR, on *algorithmic fairness*, that discusses discrimination discovery as well as fairness-aware data mining. Barocas and Hardt [8] also focus on the general topic of *fairness* in machine learning, but do touch on fairness in recommender systems at several points. Zhang and Chen [80] provide a comprehensive overview of *explainability* in the context of recommender sytems. Finally, Knijnenburg and Berkovsky [49] present a tutorial on *privacy* in recommender systems, Hui Yang and Zhang [41] present a tutorial on *differential privacy* in IR, and White [74] discusses *privacy* in the context of interactive systems. Unfortunately, we have not been able to identify tutorials or surveys that consider topics such as surveillance, accountability, or transparency in the context of IR technology.

4 Online Phase Revisited

In Section 3, we have shown that today's IR teaching materials cover the offline phase pretty extensively, while only touching on the online phase. In this section, we take a step towards filling this gap by bringing together surveys and tutorials related to the online phase of the IR system development and by giving an overview of the current teaching efforts in this direction. This section covers online topics in the same order as they appear in Table 2 and Section 3: (i) querying, (ii) interacting with results, (iii) scoring and ranking, and (iv) evaluation.

4.1 Querying

Queries encapsulate evidence about user needs and intents, sessions and tasks, user profiles and context. Understanding and analyzing queries is a vital part of a live IR system. Several surveys and tutorials are available on the topic. White [74] gives a broad overview of user querying behavior and the taxonomy of query intents, while Silvestri [69] zooms in on query log analysis, White and Roth [75] discuss exploratory search, and Mehrotra et al. [57] focus on user tasks.

Silvestri [69] discusses statistics and applications of query logs. The survey presents distributions and trends behind user querying behavior and shows how to use query logs to solve various query-related IR problems, such as query expansion, query suggestion, personalization, spelling correction, etc. The list of these applications can be expanded with a survey on query auto-completion [16]. White and Roth [75] discuss exploratory search, which goes beyond the query–response paradigm. The survey formally defines exploratory search and discusses its features, such as rapid query refinement, collaboration between users, support for facets, visualization, and search trails, etc. Mehrotra et al. [57] focus on sequences of queries, namely sessions and tasks. The tutorial discusses in detail task understanding, task extraction algorithms, task-based evaluation, and applications of tasks, such as personalization, recommendation, and dialogue systems.

Still, we believe that there is much more about querying out there, concerning query understanding beyond sessions and tasks, query intents beyond Broder's taxonomy [11], zero-query and proactive retrieval based on a user profile and context, etc. The teaching materials on these topics are yet to be developed.

4.2 Interacting with results

User interactions with an IR system, such as clicks, mouse movements, scrolling, time, etc., are a valuable source for understanding users of the system and for improving and evaluating its quality. White [74] gives a comprehensive overview of the area, discussing the literature on a wide range of user interactions and interaction models, on logging and privacy issues, and on applications of user interactions in terms of system support and evaluation. This book can be considered as an encyclopedia of user interactions with an IR system and as a basis for creating corresponding teaching materials.

However, these materials are not yet readily available and still have to be developed. An exception is the survey and tutorials on clicks models [21–24].[2] This set of materials presents a unified framework for existing click models, discusses the most widely used models, such as UBM [28], DBN [18], and DCM [34], presents the corresponding estimation and evaluation techniques, available datasets and tools, discusses advanced models of user interactions and their applications in ranking, evaluation, and simulation. These teaching materials are also accompanied by the PyClick software package that implements the discussed click models.[3]

We believe that particular attention should also be paid to various biases in user interactions, from biases in the way users examine search results, such as position bias [25] and attraction/attention bias [20, 71], to biases due to people's beliefs [73]. This topic also provides a basis for studying unbiased ranking and evaluation (e.g., [3, 44]).

4.3 Scoring and ranking

A stream of user interactions, such as clicks, is used to continuously improve IR systems through online LTR. This problem is usually formalized as a multi-armed bandit problem [63] or a contextual bandit problem [53]. Both views are extensively covered in recent tutorials [32, 33, 61], which discuss such problems as dueling bandit gradient descent [77] and the exploration vs. exploitation trade-off [38]. Lattimore and Szepesvári [52, Chapter 32] present a theoretical framework for using bandit algorithms for IR, and highlight unique challenges and ways to address them in the online setting. The book also cites and discusses other literature on bandit algorithms for IR.

The problem of LTR from historical data [39] is also well-covered by the above materials. The related problem of learning from logged bandit feedback is discussed in a tutorial on counterfactual learning [44] with a focus on such approaches as policy optimizer for exponential models (POEM) [70] and in a tutorial on unbiased LTR [3] with a focus on removing bias.

[2] `https://clickmodels.weebly.com`
[3] `https://github.com/markovi/PyClick`

4.4 Evaluation

Continuous evaluation of a deployed IR system is crucial for maintaining and improving its quality. The survey on online evaluation by Hofmann et al. [40] gives an overview of this area. It covers A/B testing and interleaving, online metrics, counterfactual evaluation and practical considerations. Each of these topics is covered in further detail in at least one of several recent tutorials [12, 13, 44, 61]. A/B testing, online metrics and practical aspects of online evaluation are presented by real-world practitioners, from Yandex,[4] in [12, 13]. Interleaving and multileaving are excellently covered by Oosterhuis [61]. Counterfactual evaluation is extensively discussed by Joachims and Swaminathan [44].

Although today's textbooks present the general procedure of A/B testing, they need to be enhanced with a discussion of practical aspects of A/B testing, such as conflicting experiments, user sampling, duration and seasonality of user interests, appropriate logging of user interactions, hypothesis formulation and success criteria, efficiency of experiments, etc. [12, 13]. In terms of interleaving, at least basic algorithms have to be included into IR teaching, such as team draft interleaving [64], probabilistic interleaving [37], and optimized interleaving [62]. As outlined by Hofmann et al. [40], online metrics have to be discussed along two dimensions: granularity (document-, ranking-, and session-level) and absolute vs. relative. Finally, counterfactual evaluation has to cover the inverse propensity scoring estimator [54] and the bias vs. variance trade-off [44].

5 Conclusion

In this opinion paper we have brought together various IR scenarios and topics that we believe should be a part of the IR teaching curriculum. We argued that although a number of teaching materials exists that cover IR scenarios beyond search, these scenarios should be more prominent in today's IR teaching. Then, we zoomed in on the search scenario and divided a list of corresponding teaching topics into an *offline* phase and an *online* phase, depending on whether they are concerned with the initial development of an IR system or with its continuous improvement during operation, and into IR applications and the broader impact of IR. We showed that topics concerned with the offline phase are covered well by existing IR teaching materials, such as textbooks and online courses. We also showed that topics concerned with the online phase, IR applications, and the impact of IR are covered in little to no detail in these teaching materials. We then focused on the online phase of IR and brought together an extensive set of existing surveys and tutorials on such aspects as querying, interacting with results, online scoring and ranking, and online evaluation. We summarized these additional teaching materials and discussed how to incorporate them into the IR curriculum.

Our point of view on IR, unfolded in this paper, has its limitations. Particularly, we still focus on a *system* aspect of IR, i.e., we treat IR as a set of algorithms which provide certain functionality, although here we focus on *interactive* or *interaction-based* algorithms. We acknowledge that IR is not only about that and that there are more user-oriented aspects out there: user studies [47], a cognitive perspective of information seeking [10, 42], etc. These aspects should also be reflected in IR teaching and, particularly, in textbooks. However, the discussion of the corresponding topics goes beyond the scope of this paper.

[4] `https://yandex.com`

With this paper we made a first step towards a new generation of IR teaching materials. The next step is to build a coherent set of slides, textbooks, and online courses, that cover the topics that are missing in the current teaching materials and, particularly, focus on scenarios beyond search and topics beyond the offline phase of IR. These new teaching materials will, in turn, be the basis for specific IR courses and whole IR curricula, targeting particular study programs (e.g., CS, AI, IS) and student levels (undergraduate, graduate, post-graduate). We invite everyone in the community to take the next step towards a new generation of IR teaching materials and design – and share – those courses and curricula.

Acknowledgements

We would like to thank the following people who contributed to this opinion paper by participating in discussions or by providing comments on an initial draft (in alphabetical order of the last name): Ricardo Baeza-Yates, Aleksandr Chuklin, W. Bruce Croft, Alexey Drutsa, Dorota Glowacka, Vladimir Gulin, Gleb Gusev, Katja Hofmann, Thorsten Joachims, Evangelos Kanoulas, Yiqun Liu, Maarten Marx, Daan Odijk, Harrie Oosterhuis, Anne Schuth, Adith Swaminathan, Csaba Szepesvári, Ryen White, Min Zhang.

This research was partially supported by Ahold Delhaize, the Netherlands Institute for Sound and Vision, and the Netherlands Organisation for Scientific Research (NWO) under project nr CI-14-25. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

A Textbooks considered

For each of the topics listed in Table 2, we indicate whether and, if so, where they are covered in today's most widely used textbooks and online course in IR [6, 15, 27, 56, 79].

Table 3: Coverage of topics from Table 2 in IR textbooks and online course. Materials are ordered by the last name of the first author. The order and grouping of topics is the same as in Table 2.

Topic	Textbooks and online course				
	Baeza-Yates and Ribeiro-Neto [6]	Büttcher et al. [15]	Croft et al. [27]	Manning et al. [56]	Zhai [79]
Crawling	12	15.6	3.2, 3.6–3.8	20.1–20.2, 20.4, 19.6	5.4
Indexing	9.2–9.3, 10.3.3–10.3.4	2.1, 4–7, 14	5	1.2, 2.3–2.4, 3.1–3.2, 4, 5.2–5.3, 6.1, partly in 7.1, 7.2.1, 20.3	2.4–2.5, 5.5
Link analysis	11.3, 11.5.2	15.1, 15.3.1–15.3.4	4.5, 10.3.2	19.2, 21	5.6–5.8

Table 3: Coverage of topics from Table 2 in IR textbooks and online course. Materials are ordered by the last name of the first author. The order and grouping of topics is the same as in Table 2.

Topic	Textbooks and online course				
	Baeza-Yates and Ribeiro-Neto [6]	Büttcher et al. [15]	Croft et al. [27]	Manning et al. [56]	Zhai [79]
Text analysis	6.5–6.6	3	4.1–4.3, 6.2.1–6.2.2	2.1–2.2, 3.3–3.4, 5.1	Some in 2.4
Learning to Rank (LTR)	11.5.4–11.5.5	11.7	7.6.1, 11.2	15.4	6.1–6.3
Semantic scoring	3.4.1–3.4.3	–	7.6.2	18	–
Term-based scoring	3.2–3.3, 3.5	2.2.1–2.2.3, 8, 9	7.1.1–7.1.2, 7.2–7.4	1.3–1.4, 6.2–6.4, 11, 12	1.5–1.6, 2.1–2.3, 4
Term dependency	–	–	11.3	11.4.2	–
Efficiency evaluation	–	13.1	8.5	8.6.1	–
Hypothesis testing	–	12.3	8.6.1	–	Some in 3.6
Offline evaluation	4.2–4.4, 4.5.1–4.5.2, 4.5.4	12.1–12.2, 12.4–12.5	8.2, 8.4	8.1–8.5	3
User studies	–	–	–	–	–
Parallel computing	10.4–10.6	–	–	–	–
Search optimization	11.4.3	13.3–13.4	–	7.1	2.6
Distributed IR	10.3.1–10.3.2, 10.7–10.8, 11.4.2, 11.4.4–11.4.5	–	10.5	–	–
Diversification	–	–	–	–	–
Search interfaces	2, 11.7	–	6.3	8.7	–
Text classification	8	10, 11.1–11.6	9.1	13, 14, 15.1–15.3	–
Text clustering	–	–	9.2	16, 17	–
Intent	7.2.3	15.2.1	–	19.4.1	–
Query analysis	–	–	–	–	–
Query expansion	5.5–5.6	–	6.2.3	9.2	–
Sessions and tasks	7.2.5	–	–	–	–
User profiling	–	–	–	–	–
Biases	–	–	–	–	–
Interactions	7.2.1–7.2.2	15.2.2, 15.5.2	–	–	–

Table 3: Coverage of topics from Table 2 in IR textbooks and online course. Materials are ordered by the last name of the first author. The order and grouping of topics is the same as in Table 2.

Topic	Textbooks and online course				
	Baeza-Yates and Ribeiro-Neto [6]	Büttcher et al. [15]	Croft et al. [27]	Manning et al. [56]	Zhai [79]
Logging	–	–	8.3	–	–
Counterfactual LTR	–	–	–	–	–
Online LTR	–	–	–	–	–
Personalization	–	–	–	–	–
Relevance feedback	5.2–5.3	8.6	6.2.4, 7.3.2	9.1	5.1–5.3
Counterfactual evaluation	–	–	–	–	–
Online evaluation	4.5.3, 4.5.5, 5.4	15.5.2	8.6.3	8.6.3	–
Entity search	–	–	11.4.2	–	–
Multimedia IR	14	–	11.6	–	–
Product search and recommendation	–	–	10.4.2	–	6.5–6.9
Retrieval of semi-structured documents	13.4–13.6	16	11.4.1	10	–
Sponsored search	–	–	–	–	–
Vertical search	–	–	–	–	–
Web search	11, 12	15	3.2, 4.5, 7.5	19, 20	5.4–5.8
Cognition	–	–	–	–	–
Privacy	–	–	–	–	–
Responsible IR	–	–	–	–	–

B Tutorials and surveys considered

Table 4: Tutorials and surveys referenced in this paper.

Author(s)	Topic
Ai et al. [3]	Unbiased learning to rank
Arguello [5]	Aggregated search
Barocas and Hardt [8]	Fairness in machine learning
Bast et al. [9]	Semantic search, knowledge graphs

Budylin et al. [12, 13]	Online evaluation
Burges [14]	Learning to rank
Cai and de Rijke [16]	Query auto-completion
Cambazoglu and Baeza-Yates [17]	Infrastructure
Chuklin et al. [21, 22, 23, 24]	Click models
Crestani et al. [26]	Mobile information retrieval
Gao et al. [30]	Conversational search
Glowacka [32]	Bandit algorithms
Grotov and de Rijke [33]	Online learning to rank
Hajian et al. [35]	Algorithmic bias
Hofmann et al. [40]	Online evaluation
Hui Yang and Zhang [41]	Differential privacy in information retrieval
Joachims and Swaminathan [44]	Counterfactual evaluation and learning
Jones [45]	Mobile search
Kanoulas [46]	Online and offline evaluation
Kelly [47]	User studies
Kenter et al. [48]	Neural methods in information retrieval
Knijnenburg and Berkovsky [49]	Privacy in recommender systems
Lalmas [51]	XML retrieval
Lattimore and Szepesvári [52]	Bandit algorithms
Liu [55]	Offline learning to rank
Mehrotra et al. [57]	Task understanding
Mitra and Craswell [58]	Neural methods in information retrieval
Onal et al. [60]	Neural methods in information retrieval
Oosterhuis [61]	Online evaluation and ranking
Ren et al. [66]	E-commerce
Sakai [67]	Experimental design and methodology
Santos et al. [68]	Diversification
Silvestri [69]	Mining query logs
Wang et al. [72]	Online display advertising
White and Roth [75]	Exploratory search
Zhai [78]	Language modeling
Zhang and Chen [80]	Explainable recommendation

References

[1] C. C. Aggarwal. *Recommender Systems – The Textbook*. Springer, 2016.

[2] Q. Ai, L. Yang, J. Guo, and W. B. Croft. Analysis of the paragraph vector model for information retrieval. In *Proceedings of ICTIR*, pages 133–142. ACM, 2016.

[3] Q. Ai, J. Mao, Y. Liu, and W. B. Croft. Unbiased learning to rank: Theory and practice. In *Proceedings of CIKM*, pages 2305–2306. ACM, 2018. URL `https://sites.google.com/site/tutorialofultrforwebsearch/home`.

[4] J. Allan, J. Arguello, L. Azzopardi, P. Bailey, T. Baldwin, K. Balog, H. Bast, N. Belkin, K. Berberich, B. von Billerbeck, J. Callan, R. Capra, M. Carman, B. Carterette, C. L. A.

Clarke, K. Collins-Thompson, N. Craswell, W. B. Croft, J. S. Culpepper, J. Dalton, G. Demartini, F. Diaz, L. Dietz, S. Dumais, C. Eickhoff, N. Ferro, N. Fuhr, S. Geva, C. Hauff, D. Hawking, H. Joho, G. Jones, J. Kamps, N. Kando, D. Kelly, J. Kim, J. Kiseleva, Y. Liu, X. Lu, S. Mizzaro, A. Moffat, J.-Y. Nie, A. Olteanu, I. Ounis, F. Radlinski, M. de Rijke, M. Sanderson, F. Scholer, L. Sitbon, M. Smucker, I. Soboroff, D. Spina, T. Suel, J. Thom, P. Thomas, A. Trotman, E. Voorhees, A. P. de Vries, E. Yilmaz, and G. Zuccon. Report from the third strategic workshop on information retrieval in lorne (SWIRL 2018). *SIGIR Forum*, 52:34–90, 2018.

[5] J. Arguello. Aggregated search. *Foundations and Trends in Information Retrieval in Information Retrieval*, 10(5):365–502, 2017.

[6] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Publishing Company, 2nd edition, 2011.

[7] K. Balog. *Entity-Oriented Search*. Springer, 2018.

[8] S. Barocas and M. Hardt. Fairness in machine learning. In *Proceeding of NIPS*, 2017. URL `https://fairmlbook.org/`.

[9] H. Bast, B. Buchhold, and E. Haussmann. Semantic search on text and knowledge bases. *Foundations and Trends in Information Retrieval*, 10(2–3):119–271, 2016.

[10] N. J. Belkin. People, interacting with information. *SIGIR Forum*, 49(2):13–27, 2016.

[11] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3, 2002.

[12] R. Budylin, A. Drutsa, G. Gusev, E. Kharitonov, P. Serdyukov, and I. Yashkov. Online evaluation for effective web service development. In *Proceedings of TheWebConf*, 2018. URL `https://research.yandex.com/tutorials/online-evaluation/www-2018`.

[13] R. Budylin, A. Drutsa, G. Gusev, P. Serdyukov, and I. Yashkov. Online evaluation for effective web service development. In *Proceedings of KDD*. ACM, 2018. URL `https://research.yandex.com/tutorials/online-evaluation/kdd-2018`.

[14] C. J. C. Burges. From RankNet to LambdaRank to LambdaMART: An overview. Technical report, Microsoft Research, 2010.

[15] S. Büttcher, C. L. Clarke, and G. V. Cormack. *Information Retrieval: Implementing and Evaluating Search Engines*. MIT Press, 2010.

[16] F. Cai and M. de Rijke. A survey of query auto completion in information retrieval. *Foundations and Trends in Information Retrieval*, 10(4):273–363, 2016.

[17] B. B. Cambazoglu and R. Baeza-Yates. Scalability challenges in web search engines. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(6):1–138, 2015.

[18] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of WWW*, pages 1–10. ACM, 2009.

[19] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue. Large-scale validation and analysis of interleaved search evaluation. *ACM Transactions on Information Systems*, 30(1): 1–41, 2012.

[20] D. Chen, W. Chen, H. Wang, Z. Chen, and Q. Yang. Beyond ten blue links: enabling user click modeling in federated web search. In *Proceedings of WSDM*, pages 463–472. ACM, 2012.

[21] A. Chuklin, I. Markov, and M. de Rijke. An introduction to click models for web search: SIGIR 2015 tutorial. In *Proceedings of SIGIR*, pages 1113–1115. ACM, 2015. URL `https://clickmodels.weebly.com/sigir-2015-tutorials.html`.

[22] A. Chuklin, I. Markov, and M. de Rijke. Advanced click models and their applications to IR: SIGIR 2015 tutorial. In *Proceedings of SIGIR*, pages 1111–1112. ACM, 2015. URL `https://clickmodels.weebly.com/sigir-2015-tutorials.html`.

[23] A. Chuklin, I. Markov, and M. de Rijke. *Click Models for Web Search*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2015.

[24] A. Chuklin, I. Markov, and M. de Rijke. Click models for web search and their applications to IR. In *Proceedings of WSDM*, pages 689–690. ACM, 2016. URL `https://clickmodels.weebly.com/wsdm-2016-tutorial.html`.

[25] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of WSDM*, pages 87–94. ACM, 2008.

[26] F. Crestani, S. Mizzaro, and I. Scagnetto. *Mobile Information Retrieval*. Springer, 2017.

[27] W. B. Croft, D. Metzler, and T. Strohman. *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2010.

[28] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of SIGIR*, pages 331–338. ACM, 2008.

[29] P. Education. *Speech and Language Processing*. Jurafsky, Daniel and Martin, James H., 2008.

[30] J. Gao, M. Galley, and L. Li. Neural approaches to conversational AI. *arXiv preprint arXiv:1809.08267*, 2018.

[31] J. Garcia-Gathright, A. Springer, and H. Cramer. Assessing and addressing algorithmic bias – but before we get there. In *AAAI Spring Symposium Series*, 2018.

[32] D. Glowacka. Bandit algorithms in interactive information retrieval. In *Proceedings of ICTIR*, pages 327–328. ACM, 2017.

[33] A. Grotov and M. de Rijke. Online learning to rank for information retrieval. In *Proceedings of SIGIR*, pages 1215–1218. ACM, 2016. URL `https://staff.fnwi.uva.nl/m.derijke/talks-etc/online-learning-to-rank-tutorial/`.

[34] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *Proceedings of WSDM*, pages 124–131. ACM, 2009.

[35] S. Hajian, F. Bonchi, and C. Castillo. Algorithmic bias: from discrimination discovery to fairness-aware data mining. In *Proceedings of KDD*, pages 2125–2126. ACM, 2016. URL `http://francescobonchi.com/algorithmic_bias_tutorial.html`.

[36] M. A. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.

[37] K. Hofmann, S. Whiteson, and M. de Rijke. A probabilistic method for inferring preferences from clicks. In *Proceedings of CIKM*, pages 249–258. ACM, 2011.

[38] K. Hofmann, S. Whiteson, and M. de Rijke. Balancing exploration and exploitation in listwise and pairwise online learning to rank for information retrieval. *Information Retrieval Journal*, 16(1):63–90, 2012.

[39] K. Hofmann, A. Schuth, S. Whiteson, and M. de Rijke. Reusing historical interaction data for faster online learning to rank for IR. In *Proceedings of WSDM*, pages 183–192. ACM, 2013.

[40] K. Hofmann, L. Li, and F. Radlinski. Online evaluation for information retrieval. *Foundations and Trends in Information Retrieval*, 10(1):1–117, 2016.

[41] G. Hui Yang and S. Zhang. Differential privacy for information retrieval. In *Proceedings of WSDM*, pages 777–778. ACM, 2018. URL `https://privacypreservingir.org/`.

[42] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005.

[43] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of KDD*, pages 133–142. ACM, 2002.

[44] T. Joachims and A. Swaminathan. Counterfactual evaluation and learning for search, recommendation and ad placement. In *Proceedings of SIGIR*, pages 1199–1201. ACM, 2016. URL `http://www.cs.cornell.edu/~adith/CfactSIGIR2016/`.

[45] M. Jones. Mobile search tutorial. In *Proceedings of MobileHCI*. ACM, 2009.

[46] E. Kanoulas. A short survey on online and offline methods for search quality evaluation. In *Proceedings of RuSSIR*, pages 38–87. Springer, 2016.

[47] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2):1–224, 2009.

[48] T. Kenter, A. Borisov, C. Van Gysel, M. Dehghani, M. de Rijke, and B. Mitra. Neural networks for information retrieval. In *Proceedings of WSDM*, pages 779–780. ACM, 2018. URL `http://nn4ir.com`.

[49] B. P. Knijnenburg and S. Berkovsky. Privacy for recommender systems. In *Proceedings of RecSys*, pages 394–395. ACM, 2017.

[50] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15):5802–5805, 2013.

[51] M. Lalmas. *XML Retrieval*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.

[52] T. Lattimore and C. Szepesvári. *Bandit Algorithms*. 2018.

[53] L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of WWW*, pages 661–670. ACM, 2010.

[54] L. Li, W. Chu, J. Langford, and X. Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of WSDM*, pages 297–306. ACM, 2011.

[55] T.-Y. Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

[56] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[57] R. Mehrotra, E. Yilmaz, and A. H. Awadallah. Understanding & inferring user tasks and needs. In *Proceedings of TheWebConf*, 2018. URL `https://task-ir.github.io/Task-based-Search/`.

[58] B. Mitra and N. Craswell. An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 2018. To appear.

[59] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman. Social data: Biases, methodological pitfalls, and ethical boundaries. *SSRN*, 2016. `https://ssrn.com/abstract=2886526`.

[60] K. D. Onal, Y. Zhang, I. S. Altingovde, M. M. Rahman, P. Karagoz, A. Braylan, B. Dang, H.-L. Chang, H. Kim, Q. Mcnamara, A. Angert, E. Banner, V. Khetan, T. Mcdonnell, A. T. Nguyen, D. Xu, B. C. Wallace, M. Rijke, and M. Lease. Neural information retrieval: At the end of the early years. *Information Retrieval Journal*, 21 (2-3):111–182, 2018.

[61] H. Oosterhuis. Learning to rank and evaluation in the online setting. In *Proceedings of RuSSIR*, 2018. URL `https://sites.google.com/view/harrieoosterhuis/russir-2018`.

[62] F. Radlinski and N. Craswell. Optimized interleaving for online retrieval evaluation. In *Proceedings of WSDM*, pages 245–254. ACM, 2013.

[63] F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. In *Proceedings of ICML*, pages 784–791. ACM, 2008.

[64] F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality? In *Proceeding of CIKM*, pages 43–52. ACM, 2008.

[65] R. Raieli. *Multimedia Information Retrieval*. Chandos Publishing, 2013.

[66] Z. Ren, X. He, D. Yin, and M. de Rijke. Information discovery in e-commerce. In *Proceedings of SIGIR*, pages 1379–1382. ACM, 2018. URL `https://sites.google.com/view/sigir2018-info-ec/home`.

[67] T. Sakai. *Laboratory Experiments in Information Retrieval - Sample Sizes, Effect Sizes, and Statistical Power*, volume 40 of *The Information Retrieval Series*. Springer, 2018.

[68] R. L. T. Santos, C. Macdonald, and I. Ounis. Search result diversification. *Foundations and Trends in Information Retrieval*, 9(1):1–90, 2015.

[69] F. Silvestri. Mining query logs: Turning search usage data into knowledge. *Foundations and Trends in Information Retrieval*, 4(1–2):1–174, 2009.

[70] A. Swaminathan and T. Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *Proceedings of ICML*, pages 814–823. JMLR.org, 2015.

[71] C. Wang, Y. Liu, M. Zhang, S. Ma, M. Zheng, J. Qian, and K. Zhang. Incorporating vertical results into search click models. In *Proceedings of SIGIR*, pages 503–512. ACM, 2013.

[72] J. Wang, W. Zhang, and S. Yuan. Display advertising with real-time bidding (RTB) and behavioural targeting. *Foundations and Trends in Information Retrieval*, 11(4-5): 297–435, 2017.

[73] R. White. Beliefs and biases in web search. In *Proceedings of SIGIR*, pages 3–12. ACM, 2013.

[74] R. W. White. *Interactions with Search Systems*. Cambridge University Press, 2016.

[75] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.

[76] I. Witten, A. Moffat, and T. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann Publishers, 1999.

[77] Y. Yue and T. Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of ICML*, pages 1201–1208. ACM, 2009.

[78] C. Zhai. Statistical language models for information retrieval. *Synthesis Lectures on Human Language Technologies*, 1(1):1–141, 2008.

[79] C. Zhai. Text retrieval and search engines. https://www.coursera.org/learn/text-retrieval, 2018.

[80] Y. Zhang and X. Chen. Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*, 2018.

[81] J. Zobel. What we talk about when we talk about information retrieval. *SIGIR Forum*, 51(3):18–26, 2018.