Report on the International Conference on Design of Experimental Search & Information REtrieval Systems (DESIRES 2018)

Omar Alonso
Microsoft
omalonso@microsoft.com

Gianmaria Silvello
University of Padua
silvello@dei.unipd.it

Abstract

This is a report on the first edition of the International Conference on Design of Experimental Search & Information REtrieval Systems (*DESIRES* 2018) held in Bertinoro, Italy, from August 28 to August 31, 2018.

1 The DESIRES conference series manifesto

DESIRES is a systems-oriented conference, complementary in its mission to the mainstream Information Access and Retrieval conferences like SIGIR, ECIR, and other symposiums focusing on specific aspects of IR such as ICTIR or CHIIR, emphasizing the innovative technological elements of search and retrieval systems. This conference is inspired by the *Conference on Innovative Data Systems Research* (CIDR) [1] conference series of the database community.

DESIRES is thought as a *biennial* retreat-like event that favors discussion and brainstorming.

DESIRES gathers researchers and practitioners from both academia and industry to discuss the latest innovative and visionary ideas in the field. The goal of this conference is to provide the IR community with a venue for presenting innovative search systems architectures, as well as a publication opportunity. DESIRES does not compete with the mainstream conferences presenting rigorous treatises in established areas; instead its goal is to air radically new ideas.

DESIRES mainly encourages papers about innovative and risky information access and retrieval system ideas, systems-building experience and insight, resourceful experimental studies, provocative position statements, multidisciplinary takes on IR, and any new application domains. DESIRES especially welcomes contributions focusing on implementation details, successful or failed reproducibility attempts, technological breakthroughs, and new uses of old ideas.

Traditionally, program committees in our IR field reward scholarship on narrow ideas, operate by consensus, discard "loosey-goosey" papers including half-baked ideas and submissions that resemble war stories from the field. The major IR conferences usually reject

[1] http://cidrdb.org/

such submissions because they are not scientific. However, these are often the very papers that offer long-term value to the field and should be widely disseminated. DESIRES values innovation, experience-based insight, and vision.

DESIRES looks for original work that could not be submitted in parallel to another venue. To encourage authors to submit only their best work, each person could be an author or co-author of just a single full paper. DESIRES also accepts abstracts and demos presenting ideas that are still in very early stages or challenge the current trends in IR. Any author of a full paper could additionally submit one abstract or one demo. Abstracts and demos are inserted in the conference proceedings as "short papers" and must have a single author.

2 DESIRES 2018

DESIRE 2018 was organized under the Bertinoro international Center for informatics (BiCi) [2] umbrella. BiCi is an association whose mission is to foster cutting-edge research and advanced education in Computer Science. The conference was held from August 28 to August 31, 2018, at the University Residential Centre of Bertinoro (Ce.U.B.) [3] in Bertinoro, Italy.

DESIRES received a total of 33 submissions in two categories: 20 full papers and 13 abstracts.

All full papers were reviewed by at least three members of an international Program Committee formed by experts from industry and academia. Of the full papers submitted to the conference, 13 were accepted for oral presentation. All the abstracts were reviewed by at least two reviewers, and they were all accepted to be presented to the conference. The accepted contributions represent state of the art in information retrieval, cover a diverse range of topics, propose new uses for IR techniques, querying, personal search, expert search, visual systems, IR research tools, teaching in IR, evaluation, NLP and collaborative search.

There were 49 unique authors with papers or abstracts accepted at *DESIRES 2018* with the following geographical distribution (affiliation): 12 authors from the USA (24.5%), 9 authors from Germany (18%), 7 authors from the UK (14%), 7 authors from The Netherlands (14%), 5 authors from Australia (10%), 4 authors from Italy (8%), 2 authors from Qatar (4%) and one author from France, Sweden and Switzerland (2%).

There were three keynotes:

- *Computing without Servers, V8, Rocket Ships, and Other Batsh*t Crazy Ideas in Data Systems*, Jimmy Lin, University of Waterloo, Canada.
- *Learning to Rank at Bloomberg - From Theory to Production*, Diego Ceccarelli, Bloomberg, UK.
- *Balancing Efficiency and Effectiveness Trade-Offs in Large Scale Multi-Stage Search Engines*, Shane Culpepper, RMIT University, Australia.

There was one industry panel. The goal of the panel is to discuss topics that are of importance in industrial settings and may not be of significance in academia. Moreover, another goal is to examine problems, rising trends, and their implications for the future of the field.

The speakers of the panel were:

[2] `http://www.bici.eu/index.html`

[3] `http://www.ceub.it/?lang=en`

- Emre Kiciman, Microsoft
- Mark Najork, Google
- Jeremy Pickens, Catalyst Repository Systems
- Tony Russel Rose, UXLabs

DESIRES 2018 has been held under the patronage of the Department of Information Engineering of the University of Padua. *DESIRES 2018* would not have been possible without the financial support from Bloomberg (silver sponsor) and Google (bronze sponsor).

DESIRES 2020. The next edition of DESIRES will take place in September 2020. The submission deadline for full papers, abstracts and demos will be in Spring 2020. Bookmark DESIRES website `http://desires.dei.unipd.it/` and follow `DESIRES_IR` on Twitter to keep posted.

3 Keynotes

Computing without Servers, V8, Rocket Ships, and Other Batsh*t Crazy Ideas in Data Systems. Jimmy Lin [20] discussed two trends in the context of Information Retrieval Systems: centralized computing based on large data centers and distributed computing based on personal devices. In the first part of the talk, Jimmy discussed "serverless architecture" showing how it is possible to enable computing without servers by raising the level of abstraction to computations, isolating the developer from mundane details of execution on physical machines. In the second part, Jimmy moved to the other end of the spectrum by presenting IR systems, a relational database system, and a neural network architecture that runs on the JavaScript engine of a Web browser. Finally, Jimmy elaborated on some futuristic scenarios where we may need to use a Web search engine on Mars by presenting some solutions that enable document search in high latency/small bandwidth situations.

Learning to Rank at Bloomberg - From Theory to Production. Diego Ceccarelli [5] reported on the internal Bloomberg search engine by describing the significant challenges that engineers have to tackle to guarantee efficient news search in almost real-time to their user base. In particular, Diego presented compelling war stories about the steps that brought the Bloomberg team to efficiently incorporate Learning to Rank advances in Apache Solr at a production level.

Balancing Efficiency and Effectiveness Trade-Offs in Large Scale Multi-Stage Search Engines. Shane Culpepper [6] discussed recent work on managing trade-offs between efficiency and effectiveness in modern multi-stage ranking architectures which are comprised of a candidate generation stage followed by one or more reranking stages. Moreover, Shane reported about the discussion that took place at SWIRL III (Lorne, Australia, 2018) [8]. Shane engaged the participants by asking them to propose their long-range issues of the Information Retrieval field; this request triggered a lively discussion that brought up some themes as: progressive visualization for IR models to aid the comprehension of search results in almost real time, understandability and explainability of IR models especially in the

context of Neural IR and in Health IR, conversational search, how to open up expert search to non-experts (e.g., how to explain documents to non-experts) and narrative construction.

4 Scientific program

4.1 Systems

Answering What If, Should I, and Other Expectation Exploration Queries Using Causal Inference over Longitudinal Data. This paper presents a system for running ad-hoc online causal inference analyses. The idea is to leverage social media and other micro-blogging platforms to generate pros/cons lists for decision support and timeline representations to show how situations evolved. The goal of this system is to aid people to take decisions and to build on the "experiences of the crowd". Why should we trust a search engine to find the best product to buy or to form an opinion about an event and not to "evaluate options and take action with one-click" [14]?

WASP: Web Archiving and Search Personalized. This paper targets "lifelogging" and presents a fully functional prototype of a personal web archive and search system available open source. The idea of this system is to log user searches to create a personal Web archive. The authors, on the one hand, present the design challenges they tackled and on the other hand discuss how a personal Web archive has to be tailored on specific user profiles – e.g., which pages to archive? Which pages to index? How to deal with privacy issues [15]?

4.2 Models

Hyperdimensional Utterance Spaces. This paper presents a hyperdimensional processing model for language data as an extension of models previously used for words to handling text level information. The idea is to represent a broad range of linguistic and extra-linguistic features in a common framework to be used as a bridge between symbolic and continuous representations, as an encoding scheme for symbolic information and as a basis for feature space exploration. The paper discusses hyperdimensional computing and provides empirical validation of the framework based on "authorship profiling", a task specifically requiring an understanding of linguistic content [13].

Search Agent Model: A Conceptual Framework for Search by Algorithms and Agent Systems. This paper presents a Search Agent Model that addresses some of the current challenges faced by search systems: confidence estimation, task state, and expressing complex long-term retrieval models. The focus of this paper is on search systems used by bots rather than by humans. The authors discuss two use-cases based on a travel agent system and entity-aware retrieval models. The model presented in this work represents the first step towards enabling algorithmic and agent system users to develop new information agent applications [9].

Joint Modeling and Optimization of Search and Recommendation. This paper presents a joint model for search where search engines and recommender systems

models can be used together to improve the performances of both. The authors propose a general framework that simultaneously learns a retrieval model and a recommendation model by optimizing a joint loss function. Preliminary tests on product search show the viability of this solution. Furthermore, the paper discusses how the presented joint framework can be used in contexts where limited data is available [31].

4.3 Experimentation

Mix and Match: Collaborative Expert-Crowd Judging for Building Test Collections Accurately and Affordably. This paper discusses how relevance judgments can be created by wisely mixing crowdsourcing and traditional techniques – i.e., how some “”document-topic pairs are assigned to in-house assessors for relevance judging while crowd workers assess the rest”. The idea is to define a method to distinguish between documents which are better suited to be judged by crowd judges and those that should be assigned to expensive trusted judges. The proposed method based on “collaborative judging” shows that leveraging on the crowd in combination with trusted judges for an accurate and affordable building of IR test collections is a promising research direction to pursue [17].

APONE: Academic Platform for ONline Experiments. This paper presents an open source platform to set up controlled experiments on the Web. APONE offers a ready to use solution to common problems researchers have to face when conducting online experiments – e.g., how to define/start/stop an experiment, how to randomly assign users to the different system variants, and how to analyze the impact of one system/interface change. The paper also presents the outcomes of using APONE in the classroom with a large group of students showing the potential of this tool also for teaching IR [23].

Agile Information Retrieval Experimentation with Terrier Notebooks. This paper presents the last release of the widely-used Terrier IR system (ver. 5.0). In particular, it focuses on Terrier notebooks that enable fast prototyping and debugging of IR systems. Terrier notebooks work in a Scala environment which allows for code that is more concise than the equivalent Java. Moreover, the paper describes how Terrier notebook can be employed to teach a modern undergraduate- and a postgraduate-level elective course on IR [21].

4.4 Querying and Retrieval

Towards Efficient and Effective Query Variant Generation. This paper presents a new take on query expansion and data fusion approaches to boost retrieval performances. The primary goal is to mimic the performance achievable through fusion over human query variations by combining relevance models induced from multiple external corpora. The authors show that this approach achieves top effectiveness performances. Moreover, their method is also efficient when it relies on many short synthetic queries generated by a stochastic random process [4].

Retrieval and Richness when Querying by Document. This paper examines Query-By-Document (QBD) techniques with a focus on active learning of text classifiers in a legal context (e.g., electronic discovery and corporate investigations). The authors discuss the idea of “richness” of a document collection – i.e., the prevalence of relevant documents;

they use 658 categories from the RCV1-v2 collection to study the impact of richness on QBD variants supported by Elasticsearch. They also port BM25 to the machine learning package scikit-learn and discuss lessons learned for reproducibility [30].

4.5 Novel Applications

On the Development of a Collaborative Search System. This paper presents an open-source collaborative search system called "SearchX". This is an on-going effort with the aim of providing a tunable system able to handle tens or hundreds of users. SearchX can be accessed from multiple platforms without the need for user-side installations; its main features comprises: division of labor (e.g., group chat), sharing of knowledge (e.g., bookmarking, document rating) and awareness (e.g., query history, group summary) [26].

Pilot Experiments of Hypothesis Validation Through Evidence Detection for Historians. This paper discusses evidence detection for hypothesis validation in the humanities domain, in particular focusing on historians. The authors conduct a user study that highlights how current state-of-the-art argument mining methods are well-suited for historical validation of hypotheses. They conclude that one size does not fit all; hence, an evidence detection method must be trained interactively to adapt to the user's needs [29].

2dSearch: A Visual Approach to Search Strategy Formulation. This paper describes a visual approach to search strategy formulation. The use of a visual approach enables the elimination of many sources of syntactic error, makes the query semantics more transparent and offers opportunities for query refinement. The authors provide architectural and implementation details of their system and conduct an evaluation on PubMed Word2vec model showing better results than the best publicly available embedding model [27].

4.6 Short papers

DESIRES short papers are all single-authored and are presented in a gong-show fashion: 10 minutes for the presentation plus up to 5 minutes for questions. Short papers touched upon several hot and new research topics in IR at large comprising fact checking [18], keyword search over structured data [11], moving beyond single query optimization [7], off-line web search [3], user profiling for searching digital libraries [2], interactive document summarization [12], medical case-based retrieval [22], privacy-safe on-device personal search [24], cost-aware total recall e-discovery [10], topic modeling [25], machine-assisted annotation [16], IR and NLP advances from under an industrial lens [19] and searching for arguments [28].

5 Participants Feedback

The participants to DESIRES 2018 were asked to fill in a questionnaire to gather feedback about several aspects of the conference ranging from the scientific program to the venue, social events, and the overall organization. All answers are in a $[0, 4]$ range where 0 means "really bad',' and 4 means "very good". An overview of the feedbacks is reported in Figure 2. We can see that DESIRES 2018 had a pretty good turn out even though some aspects must be improved in the next edition. Most notably, all the scientific aspects have been considered

Figure 1: Bertinoro, 29 August 2018. *DESIRES 2018* group picture taken on the Ce.U.B. terrace.

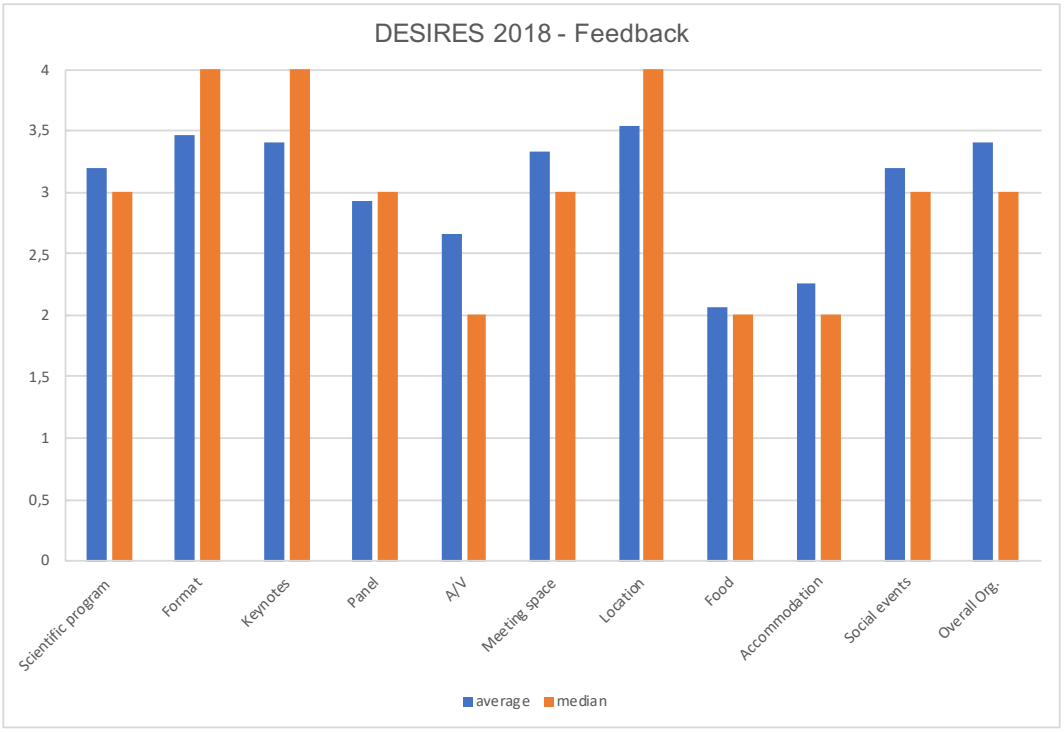

Figure 2: participants feedback about different aspects of *DESIRES 2018*.

a success. The feedback questionnaire contained an open part where the participants could write what they appreciated the most about the conference. The format of the conference was

particularly appreciated by the participants who encouraged the organizers to keep the event "small" and "retreat-like" to favor discussion and ideas exchange. Another important point is that many participants highlighted how DESIRES is different than other conferences since in this context authors were "more willing to point out the weakness of their research than maintain the illusion of perfection inherent in [another conference name] papers". Others appreciated the "academia-industry collaboration focus of the conference". Another critical aspect that many appreciated is that senior researchers/professors gave many insights and suggestions to students and young researchers dedicating much time to discussion. Finally, the vast majority enjoyed the fact that paper reviews focused on the ideas and not on performance gains over some baseline.

Acknowledgments

We desire to thanks all the people that encouraged us and helped us to make DESIRES happen. In particular, we would like to thank the top-tier program committee members (see `http://desires.dei.unipd.it/2018/#organization`) for the timely and accurate reviews and the advisory board in the person of Maristella Agosti, Ricardo Baeza-Yates, Bruce Croft, Susan Dumais, Norbert Fuhr, Donna Harman, Kalervo Järvelin, Marc Najork, Jan Pedersen and Gerhard Weikum for all the work, support and suggestions they provided.

References

[1] O. Alonso and G. Silvello, editors. *Proc. of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems*, volume 2167 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[2] M. Barifah. Building Interaction Profiles for Better Search Tools in DLs. In *DESIRES 2018* [1], page 110.

[3] R. Benham. Topic-specific off-line web search. In *DESIRES 2018* [1], page 101.

[4] R. Benham, J. S. Culpepper, L. Gallagher, X. Lu, and J. Mackenzie. Towards Efficient and Effective Query Variant Generation. In *DESIRES 2018* [1], pages 62–67.

[5] D. Ceccarelli. Learning to Rank at Bloomberg - From Theory to Production. In *DESIRES 2018* [1], page 7.

[6] J. S. Culpepper. Balancing Efficiency and Effectiveness Trade-offs in Large Scale Multi-Stage Search Engines. In *DESIRES 2018* [1], page 8.

[7] J. S. Culpepper. Single query optimisation is the root of all evil. In *DESIRES 2018* [1], page 100.

[8] J. S. Culpepper, F. Diaz, and M. D. Smucker. Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *SIGIR Forum*, 52(1):34–90, 2018.

[9] J. Dalton and J. Foley. Search agent model: A conceptual framework for search by algorithms and agent systems. In *DESIRES 2018* [1], pages 22–28.

[10] G. M. Di Nunzio. Finding all the Needles in the Haystack. A System to Estimate the Costs of e-Discovery and Systematic Reviews. In *DESIRES 2018* [1], page 106.

[11] D. Dosso. Keyword search on RDF graphs. In *DESIRES 2018* [1], page 99.

[12] B. Hättasch. Towards Interactive Summarization of Large Document Collections. In *DESIRES 2018* [1], page 103.

[13] J. Karlgren and P. Kanerva. Hyperdimensional Utterance Spaces. In *DESIRES 2018* [1], pages 29–35.

[14] E. Kiciman and J. Thelin. Answering What If, Should I, and Other Expectation Exploration Queries Using Causal Inference over Longitudinal Data. In *DESIRES 2018* [1], pages 9–15.

[15] J. Kiesel, A. P. de Vries, M. Hagen, B. Stein, and M. Potthast. WASP: web archiving and search personalized. In *DESIRES 2018* [1], pages 16–21.

[16] J.-C. Klie. INCEpTION: Interactive machine-assisted annotation. In *DESIRES 2018* [1], page 105.

[17] M. Kutlu, T. McDonnell, A. Sheshadri, T. Elsayed, and M. Lease. Mix and Match: Collaborative Expert-Crowd Judging for Building Test Collections Accurately and Affordably. In *DESIRES 2018* [1], pages 42–46.

[18] M. Lease. Fact checking and information retrieval. In *DESIRES 2018* [1], pages 97–98.

[19] D. Lewis. We Don't Know S***** (Search). In *DESIRES 2018* [1], page 107.

[20] J. Lin. Computing without Servers, V8, Rocket Ships, and Other Batsh*t Crazy Ideas in Data Systems. In *DESIRES 2018* [1], pages 3–6.

[21] C. Macdonald, R. McCreadie, and I. Ounis. Agile Information Retrieval Experimentation with Terrier Notebooks. In *DESIRES 2018* [1], pages 54–61.

[22] S. Marchesin. Implicit-Explicit Representations for Case-Based Retrieval. In *DESIRES 2018* [1], page 109.

[23] M. Marrero. APONE: Academic Platform for ONline Experiments. In *DESIRES 2018* [1], pages 47–53.

[24] M. Najork. Training On-Device Ranking Models from Cross-User Interactions in a Privacy-Preserving Fashion. In *DESIRES 2018* [1], page 108.

[25] A. Purpura. Non-negative Matrix Factorization for Topic Modeling. In *DESIRES 2018* [1], page 102.

[26] S. Rikarno Putra, K. Grashoff, F. Moraes, and C. Hauff. On the Development of a Collaborative Search System. In *DESIRES 2018* [1], pages 76–82.

[27] T. Russell-Rose and P. Gooch. 2dSearch: A Visual Approach to Search Strategy Formulation. In *DESIRES 2018* [1], pages 90–96.

[28] C. Stahlhut. Searching Arguments in German with ArgumenText. In *DESIRES 2018* [1], page 104.

[29] C. Stahlhut, C. Stab, and I. Gurevych. Pilot Experiments of Hypothesis Validation Through Evidence Detection for Historians. In *DESIRES 2018* [1], pages 83–89.

[30] E. Yang, D. D. Lewis, O. Frieder, D. A. Grossman, and R. Yurchak. Retrieval and Richness when Querying by Document. In *DESIRES 2018* [1], pages 68–75.

[31] H. Zamani and W. B. Croft. Joint Modeling and Optimization of Search and Recommendation. In *DESIRES 2018* [1], pages 36–41.