Sim4IR: The SIGIR 2021 Workshop on Simulation for Information Retrieval Evaluation

Krisztian Balog  
University of Stavanger  
Stavanger, Norway  
krisztian.balog@uis.no

Paul Thomas  
Microsoft  
Canberra, Australia  
pathom@microsoft.com

David Maxwell  
Delft University of Technology  
Delft, The Netherlands  
d.m.maxwell@tudelft.nl

Shuo Zhang  
Bloomberg  
London, United Kingdom  
szhang611@bloomberg.net

ABSTRACT
The use of simulation techniques is not foreign to information retrieval. In the past, simulation has been employed, for example, for constructing test collections and for model performance prediction and analysis in a broad array of information access scenarios. Nevertheless, a standardized methodology for performance evaluation via simulation has not yet been developed. The goal of this workshop is to create a forum for researchers and practitioners to promote methodology development and more widespread use of simulation for evaluation by: (1) identifying problem settings and application scenarios; (2) sharing tools, techniques, and experiences; (3) characterizing potentials and limitations; and (4) developing a research agenda.

ACM Reference Format:

1 THEME AND PURPOSE
Sim4IR is one-day workshop dedicated to the topic of using simulation techniques for information retrieval (IR) evaluation. The goal of this workshop is to create a forum for researchers and practitioners to present and discuss methods, tools, techniques, and experiences related to the use of simulation as a means to evaluate IR systems, and to develop a research agenda that drives methodology development and allows to unlock the potential of simulation techniques.

1.1 Motivation and Relevance
Simulation is not foreign to IR. In the past it has been employed, among others, to facilitate automatic construction of known-item test collections [1], to generate synthetic test collections [7], to analyze search behavior [12], and to evaluate interactive tasks such as search sessions [3, 5], typing search queries [4], filling values in a table [15], or conversational item recommendation [16] and search refinement [13]. Still, the use of simulation is not widespread. The potential for using simulation has only been recognized by relatively few IR researchers so far.

Recently, the need for simulation has become ever more apparent, with the emergence of areas where other types of evaluation are infeasible. One specific area is conversational information access scenarios, such as conversational item recommendations [6, 9, 10, 16], where human evaluation is both very time and resource intensive at scale. Another example is the case of test collections, which cannot be shared, e.g., because of privacy constraints [7].

We therefore argue that it is time to move more fully embrace simulation as a means of evaluation, and to start working towards establishing a standard methodology around it—as it has been done for offline (test collection based) evaluation [14], online evaluation [8], or user studies [11]. The IR community is uniquely suited to drive research and development in this area, given its rigorous focus on evaluation methodology that dates back to the inception of the field.

1.2 Topics
Topics for the workshop include, but are not limited to:
- Problem settings and application scenarios that lend themselves to evaluation via simulation, for example:
  - Simulation of IR test collections
  - Simulation of users and user interactions
- Characterizing the capabilities and limitations of simulation approaches for various IR problems
- Simulation methods, tools, techniques, and toolkits
- Evaluation of simulation

2 FORMAT AND PLANNED ACTIVITIES
Our aim is to create a dynamic, interactive, energetic workshop structured to encourage group discussion and active collaboration among attendees. The workshop will feature two keynote talks,
paper presentations, multiple (parallel) breakout sessions, and a final discussion session to wrap up the event.

As this field is at an early stage of development, there is still a lot of uncertainty about which of the approaches will lead to successful deployment; we aim to attract and discuss a wide range of ideas and perspectives by soliciting multiple types of contributions (regular, position and demonstration papers, as well as talks featuring already published works).

We invited submissions of:

- **regular papers** (4-6 pages) that present original technical, theoretical, or experimental contributions;
- **position papers** (2-4 pages) that explore controversial, risk-taking or nascent ideas that have the potential to spark debate and discussion at the workshop;
- **demonstrator papers** (max 4 pages) that present first-hand experience with research prototypes or operational systems;
- **featured talks**, to present work that has already been published in a leading conference or journal, but is relevant to the topics this workshop.

3 **OBJECTIVES AND OUTCOMES**

The expected outcomes of the workshop are:

- concrete insight into the potential of simulation techniques, the barriers to success, and concrete steps to take this research forward;
- synchronize related research happening in IR, AI, and NLP in ways that combine the strengths of each discipline; and
- have a lively, interactive workshop were everyone contributes and that inspires attendees to think "outside the box."

The results will be disseminated in various ways:

- A high quality, peer reviewed workshop proceedings.
- A report on the results of the workshop will be submitted to SIGIR Forum.
- We consider co-authoring a comprehensive white paper on the deliberations of the workshop for publication in a suitable journal.
- If the outcome lives up to our high expectations, we will consider a special issue in an appropriate journal.

4 **RELATED WORKSHOPS**

The SIGINT workshop at SIGIR 2010 looked at the simulation of interactions [2]—although a decade old now, this report is useful for framing the benefits of employing simulation within IR/IIR contexts. "The main conclusion and general consensus was that simulation offers great potential for the field of IR; and that simulations of user interaction can make explicit the user and the user interface while maintaining the advantages of the Cranfield paradigm." [2] Nevertheless, it is about time to revisit the potential of simulation and to establish a research agenda for its broader use.

REFERENCES


