

Online Search Evaluation with Interleaving

Filip Radlinski
Microsoft
Cambridge, UK

filiprad@microsoft.com

Abstract

Online evaluation allows information retrieval systems to be assessed based on how real users respond to search results presented. Compared with traditional offline evaluation based on manual relevance assessments, online evaluation is particularly attractive in settings where reliable assessments are difficult or too expensive to obtain.

However, the successful use of online evaluation requires the right metrics to be used, as real user behaviour is often difficult to interpret. I will present interleaving, a sensitive online evaluation approach that creates paired comparisons for every user query, and compare it with alternative A/B online evaluation approaches. I will also show how interleaving can be parameterized to create a family of evaluation metrics that can be chosen to best match the goals of an evaluation.

ACM Classification

H.3.3 Information Search and Retrieval

Keywords

Interleaving; Web search; A/B evaluation

Short Bio

Filip Radlinski is an applied researcher at Microsoft in Cambridge, UK, and works for Bing. He is also an honorary lecturer in the Department of Computer Science at University College London (UCL). His current research focuses on developing machine learning and online evaluation techniques for learning from, evaluating with, and optimizing to implicitly collected feedback from web users. He is particularly interested in applications to search both in the typical web setting and on new form factors, with a current focus on personalization, contextualization and in-situ evaluation. Before joining Microsoft, he completed his PhD in Computer Science at Cornell University.