Invited Talk: Understanding Faceted Search from Information Retrieval, Information Science, and Data Science Perspectives

Xi Niu

University of North Carolina at Charlotte, USA

1 Abstract

I am honored to be invited to give a talk on faceted search at the BIRDS Workshop (Bridging the Gap between Information Science, Information Retrieval, and Data Science) at SIGIR 2020. This talk reflects well my research area, which uniquely connects Information Retrieval (IR), Information Science (IS), and Data Science (DS). My research area is attributed to my educational and working background: I obtained my Master's degree in Software Engineering and my Ph.D. degree in Information Science. Now I am a faculty member at a computing department.

Historically, there has been a divide between systems and users in the IR research communities. I believe this divide reflects how researchers weigh the relative importance between developing new retrieval algorithms compared to understanding how people search for information. However, there is increasing agreement that the interaction between the users and the search engine is a fundamental part of the IR process. Due to my unique background, I am the researcher who builds models, develops algorithms, as well as runs user studies in order to understand how people seek information. My research has provided evidence that more effective information access can be achieved by a system that actively supports user interaction.

IR researchers make modern search engines or recommender systems rank better, retrieve better, or just function better, and most of today's IR papers are very well-defined or in other words, narrowly focused. I believe this BIRDS workshop is eye-widening, and will provide thoughts from a bigger, richer, or more comprehensive view of how information retrieval systems can better serve people, not only for their short-term needs but also for their long-term benefits, not only for the immediate relevance, but also for broader discovery.

Our research project on faceted search is an example of how IR, IS, and DS complement each other, offering a holistic approach that goes beyond each discipline alone. Faceted search has become a common feature on most search interfaces in e-commerce websites, digital libraries, government's open information portals, and so on. Beyond the existing studies on developing algorithms for

faceted search and empirical studies on facet usage, this talk investigated user real-time interactions with facets over the course of a search from information retrieval, data science and human factor perspectives. It adopted a Random Forest (RF) model to successfully predict facet use using search dynamic variables. In addition, the RF model provided a ranking of variables by their predictive power, which suggests that the search process follows rhythmic flow of a sequence within which facet addition is mostly influenced by its immediately preceding action. In the follow-up user study, it is found that participants used facets at critical points from the beginning to end of search sessions. Participants used facets for distinctive reasons at different stages. They also used facets implicitly without applying the facets to their search. Most participants liked the faceted search, although a few participants were concerned about the choice overload introduced by facets. The results of this research could be used to understand information seekers and propose or refine a set of practical design guidelines for faceted search. This research demonstrates how we could intersect Information Retrieval, Information Science, and Data Science to better tackle the traditional research questions in IR.

Keywords: Information Retrieval · Information Science · Data Science · faceted search · machine learning · predictive analytics · user studies.