# Measuring Job Search Effectiveness

Alfan Farizki Wicaksono
The University of Melbourne
wicaksonoa@student.unimelb.edu.au

## CCS CONCEPTS

• **Information systems → Web searching and information discovery**; **Retrieval effectiveness**; *Task models*;

## KEYWORDS

User model; evaluation; job search

Users of online job search websites interact with ranked lists of job summaries generated in response to queries, hoping to identify one or more jobs of interest. Hence, the quality of job search rankings becomes a primary factor that affects user satisfaction. In this work, we propose methodologies and measures for evaluating the quality of job search rankings from a user modeling perspective. We start by investigating job seekers' behavior when they are interacting with the generated rankings, leveraging job search interaction logs from `Seek.com`, a well-known Australasian job search website. The output of this investigation will be an accurate model of job seekers that will be incorporated into an effectiveness metric.

Recent proposals for job search ranking models used using two types of metrics to evaluate the quality of the ranking generated by the models: (1) offline metrics, such as NDCG@k ($k$ is set to the number of job summaries shown in the first page), Prec@1, or Mean Reciprocal Rank (MRR); and (2) online metrics, such as click-through rate and job application rate [3, 6].

For job search, we believe that the users are more recall-oriented, meaning that they are willing to invest more time in perusing job rankings into deeper positions in order to uncover options. Wicaksono and Moffat [8] have computed best-fit parameters for RBP and INSQ on job search data and suggest that RBP provides the closest fit when $\phi = 0.86$ for SERPs with infinite scrolling ($\phi = 0.77$ for SERPs with pagination); and that INSQ gives the best approximation when $T = 3.7$ for SERPs with infinite scrolling ($T = 1.8$ for SERPs with pagination). Hence, the behavior of job search users is different from that of Web search users that issue head queries, at least with regard to user persistence. A new model that is capable of explaining job seeker behavior is then needed.

User behaviors are variable [2, 7]. However, it has been shown that some factors, such as the user goal and the task complexity, govern the user behavior, especially the *persistence* of a user [5]. In our work, we aim at understanding the variability that occurs within a user, a query, or between all users in conjunction with several characteristics, such as *persistence* [4], user goal, and task complexity. We employ the C/W/L framework [5], in which any user model can be characterized by three properties: the continuation probability at rank $i$, denoted by $C(i)$; the weights associated across a ranking, denoted by $W(i)$; and the stopping likelihood at rank $i$, denoted by $L(i)$. This framework allows us to decide which one among the existing user models best explains each variation.

To get a better understanding of behavioral variability in search activities, we analyze interaction logs containing a large sample of users and queries. One particular interest is to study variability from the perspective of user persistence and user stopping behavior. We can employ various approaches, such as a sequential pattern mining approach [7] and a probabilistic approach [2], to explore variability of behavior within each user, within each query, and in the whole interaction log; or we might propose a new method to explore behavioral variability within the interaction data.

Finally, we investigate whether existing user models such as INST [5], and the IFT-model [1] are sufficient to explain the nature of job seekers' behavior; or otherwise we need to build a new user model that is more precise than existing user models. One important step is to identify factors that contribute to $C(i)$. We start from several factors that have been identified by Moffat et al. [5]: (1) the anticipated volumes of relevance (user goal), $T$; (2) the ranking position of current inspection, $i$; and (3) the extent to which relevance had been accumulated after the user examined the results at ranking positions 1 to $i$, $T_i = T - \sum_{j=1}^{i} r_i$. Note that Moffat et al. [5] verified these factors in the context of Web search, which is different from job search. It would be useful to again verify these factors leveraging interaction logs from the domain of job search, before planning fresh experimentation to discover new factors specific to job search.

**Acknowledgment.** This work was funded by the Australian Research Council (LP150100252), and by Seek.com; and the author was additionally supported by SIGIR and University of Melbourne Google-funded travel grants.

## REFERENCES

[1] L. Azzopardi, P. Thomas, and N. Craswell. Measuring the utility of search engine result pages. In *Proc. SIGIR*, pages 605–614, 2018.
[2] B. Carterette, E. Kanoulas, and E. Yilmaz. Incorporating variability in user behavior into systems based evaluation. In *Proc. CIKM*, pages 135–144, 2012.
[3] J. Li, D. Arya, V. Ha-Thuc, and S. Sinha. How to get them a dream job? Entity-aware features for personalized job search ranking. In *Proc. KDD*, pages 501–510, 2016.
[4] A. Moffat and J. Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Sys.*, 27(1):2.1–2.27, 2008.
[5] A. Moffat, P. Bailey, F. Scholer, and P. Thomas. Incorporating user expectations and behavior into the measurement of search effectiveness. *ACM Trans. Inf. Sys.*, 35(3):24:1–24:38, 2017.
[6] A. Saha and D. Arya. Generalized mixed effect models for personalizing job search. In *Proc. SIGIR*, pages 1129–1132, 2017.
[7] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proc. WWW*, pages 21–30, 2007.
[8] A. F. Wicaksono and A. Moffat. Empirical evidence for search effectiveness models. In *Proc. CIKM*, pages 1571–1574, 2018.


*SIGIR '19, July 21–25, 2019, Paris, France*

ACM 978-1-4503-6172-9/19/07.
https://doi.org/10.1145/3331184.3331421