

Proactive Information Retrieval

Procheta Sen
Dublin City University, Ireland
procheta.sen2@mail.dcu.ie

Abstract

Users interact with digital systems with some underlying tasks in their minds. In our research scope, a task can be either single or multi-staged. A single-staged task is associated with a single information need, whereas a multi-staged one is associated with more than one information needs. An example of a single-staged task is suggesting related papers to an author while they are writing a section of the research papers. An example of a multi-staged task is planning for a vacation, where the different underlying information needs could be ‘places to visit’, ‘booking accommodation’ etc. In the process of accomplishing their task objectives, a user often needs to interact with an information retrieval (IR) system to address one or more information needs. For instance, for writing a research paper on a chosen topic, a user needs to look for existing research work related to the topic.

Traditional IR systems do not take into account a user’s task intent while showing search results to the user. In our work, we propose a methodology towards developing next generation IR systems (i.e. *proactive* IR systems) which seek to anticipate the task intent of a user from their interactions with digital systems in order to proactively suggest potentially relevant information sources to assist them to complete their tasks. Specifically speaking, in this PhD, we proposed an embedding approach that captures the task semantics from the interactions of a user with a digital system (e.g. laptop, desktop, smartphone etc.). The proposed embedding approach is then applied for the downstream tasks of providing proactive suggestions in both single and multi staged scenarios. For the single-staged task, we propose a simulation setup to simulate a user’s reading and writing interactions in a desktop environment. For the multi-staged task, we focus on web search sessions where a user can have multiple information needs corresponding to a search task. We also proposed a reproducible evaluation framework to compare between different proactive suggestion models.

Awarded by: Dublin City University, Ireland on 10 August 2021.

Supervised by: Gareth Jones.

Available at: <https://procheta.github.io/sprocheta/Thesis.pdf>.

Selected Publications

- Procheta Sen, Debasis Ganguly, and Gareth Jones. Procrastination is the thief of time: Evaluating the effectiveness of proactive search systems. In *Proceedings of the 41st International ACM Conference on Research & Development in Information Retrieval (SIGIR)*, pages 1157–1160, 2018a.
- Procheta Sen, Debasis Ganguly, and Gareth J.F. Jones. Tempo-lexical context driven word embedding for cross-session search task extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), NAACL-HLT*, pages 283–292, 2018b.
- Procheta Sen, Debasis Ganguly, and Gareth J. F. Jones. Word-node2vec: Improving word embedding with document-level non-local word co-occurrences. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), NAACL-HLT*, pages 1041–1051, 2019.
- Procheta Sen, Debasis Ganguly, Manisha Verma, and Gareth J.F. Jones. The curious case of ir explainability: Explaining document scores within and across ranking models. In *Proceedings of the 43rd International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 2069–2072, 2020.
- Procheta Sen, Debasis Ganguly, and Gareth J. F. Jones. I know what you need: Investigating document retrieval effectiveness with partial session contexts. *ACM Transactions on Information Systems (TOIS)*, 40(3), 2021.