

Keynote: Table Retrieval and Generation

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

Abstract

Tables are powerful and versatile tools for organizing and presenting data. Tables may be viewed as complex information objects, which summarize existing information in a structured form. Therefore, for many information needs, returning tables as search results may be more helpful to the user than serving a ranked list of items (documents or entities). This talk is centered around utilizing (relational) tables as units of retrieval.

We introduce and address the problem of *ad hoc table retrieval*: answering a keyword query with a ranked list of tables. In another variant of this task, referred to as *query-by-table*, the input is not a keyword query, but an incomplete table. Tables can be ranked much like documents, by considering the words contained in them. Our main research objective is to move beyond lexical matching and improve table retrieval performance by incorporating semantic matching. We achieve that by representing tables and queries in multiple semantic spaces (employing both discrete sparse and continuous dense vector representations).

It may happen that the exact table the user is looking for does not exist. Our third task, termed *on-the-fly table generation*, addresses this very scenario: given a query, generate a relational table that contains relevant entities (as rows) along with their key properties (as columns). This problem is decomposed into three specific subtasks: (i) core column entity ranking, (ii) schema determination, and (iii) value lookup. We show that the first two subtasks are not independent of each other and can assist each other in an iterative manner.

Biography

Krisztian Balog is a full professor at the University of Stavanger where he leads the Information Access & Interaction research group. He received his PhD from the University of Amsterdam, and worked as a postdoc at the Norwegian University of Science and Technology (NTNU), before joining the University of Stavanger. His general research interests lie in the use and development of information retrieval, information extraction, and machine learning techniques for intelligent information access tasks. His current research concerns entity-oriented and semantic search, and novel evaluation methodologies. He serves as a senior programme committee member at SIGIR, CIKM and ECIR, as an Associate Editor of the ACM Transactions on Information Systems, and as a current and former coordinator of information retrieval benchmarking efforts at TREC and CLEF.

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In: Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH18). Co-located with SIGIR 2018, Ann Arbor, Michigan, USA – 12 July 2018, published at <http://ceur-ws.org>