SmartTable: Equipping Spreadsheets with Intelligent Assistance Functionalities

Shuo Zhang
University of Stavanger
shuo.zhang@uis.no

ABSTRACT

Tables are one of those "universal tools" that are practical and useful in many application scenarios. Tables can be used to collect and organize information from multiple sources and then turn that information into knowledge (and ultimately to support decision-making) by performing various operations, like sorting, filtering, and joins. Because of this, a large number of tables exist already out there on the Web, which represent a vast and rich source of structured information and could be utilized as resources. Recently, a growing body of work has begun to tap into utilizing the knowledge contained in tables. A wide and diverse range of tasks have been undertaken, including but not limited to (i) searching for tables [4], (ii) extracting knowledge from tables, and (iii) augmenting tables (e.g., with new columns and rows [1, 3]).

The objective of this research is to develop a set of components for a tool called SmartTable, which is aimed at assisting the user in completing a complex task by providing intelligent assistance for working with tables. Imagine the scenario that a user is working with a table, and has already entered some data in the table. We can provide recommendations for the empty table cells, search for similar tables that can serve as a blueprint, or even generate automatically the entire table that the user needs. The table-making task can thus be simplified into just a few button clicks. Motivated by the above scenario, we propose a set of novel tasks such as row and column heading population, table search, and table generation. The following specific research questions are addressed: (**RQ1**) How to populate table rows and column heading labels? (**RQ2**) How to find relevant tables given a keyword query? (**RQ3**) How to find tables relevant to the table the user is currently working on? (**RQ4**) How to generate an output table as response to a free text query?

For **RQ1**, the task of *row population* [1, 3] relates to the task of *entity set expansion*, where a given set of entities is to be completed with additional entities. Row population focuses on populating entities in the "core column" of a relational table. We develop a two-step pipeline for this task utilizing a table corpus and a knowledge base. In the first step, candidate entities sharing the same categories with seed entities or co-occurring in similar tables are selected. In the second step, they are ranked by a probabilistic model. *Column population* shares similarities with the problem of *schema complement*, where a seed table is to be extended with additional columns. For column population, we regard column headings from similar tables as candidates and rank them using a probabilistic model.

For **RQ2** and **RQ3**, we address the problem of *table search*. This task is not only interesting on its own but is also being used as a fundamental building block in many other table-based information access scenarios, such as table completion or table mining. To search related tables, the query could be some keywords [2, 4] or it can also be an existing (incomplete) table. Based on the query type, this task is divided into two sub-tasks, which are *table retrieval for keyword query* and *query-by-table* respectively.

For **RQ4**, we introduce and address the task of the on-the-fly table generation: given a query, generate a relational table that contains relevant entities (as rows) along with their key properties (as columns) [5]. In terms of the table elements in a relational table, this task boils downing to core column entity ranking, schema determination and value look-up. We propose a feature-based approach for entity ranking and schema determination, combing deep semantic features with task-specific signals. For value lookup, we combine information from existing tables and a knowledge base.

So far, we have proposed methods and evaluation resources for addressing the tasks of row/column population, table search, and table generation. Future research directions for this project include looking up table values, interacting with tables using natural language, and generating table embeddings.

CCS CONCEPTS

• **Information systems → Environment-specific retrieval**; *Users and interactive retrieval*; *Recommender systems*; Probabilistic retrieval models;

KEYWORDS

Intelligent table assistance; structured data search; table completion; table search; table generation

ACM Reference Format:
Shuo Zhang. 2018. SmartTable: Equipping Spreadsheets with Intelligent Assistance Functionalities. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3209978.3210219

REFERENCES

[1] Shuo Zhang, Vugar Abdulzada, and Krisztian Balog. 2018. SmartTable: A Spreadsheet Program with Intelligent Assistance. In *Proc. of SIGIR '18.*
[2] Shuo Zhang and Krisztian Balog. 2017. Design Patterns for Fusion-Based Object Retrieval. In *Proc. of ECIR '17.* Springer, 684–690.
[3] Shuo Zhang and Krisztian Balog. 2017. EntiTables: Smart Assistance for Entity-Focused Tables. In *Proc. of SIGIR '17.* 255–264.
[4] Shuo Zhang and Krisztian Balog. 2018. Ad Hoc Table Retrieval using Semantic Similarity. In *Proceedings of The Web Conference 2018 (WWW '18).*
[5] Shuo Zhang and Krisztian Balog. 2018. On-the-fly Table Generation. In *Proc. of SIGIR '18.*

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07.
https://doi.org/10.1145/3209978.3210219