University of Fribourg / Faculty of Science and Medicine / Department of Informatics

Automatic Knowledge Graph Entity Refinement
Based on Social Networks

Hussein Hazimeh

Knowledge Graphs (KGs) like Wikidata or DBpedia allow to store, process and visualize knowledge facts about real-world entities (nodes) and the interrelations between them (edges). These allow users to visualize consolidated knowledge from heterogeneous data resources in a unified and clustered graph. In this work, we propose several contributions to the fields of Knowledge Graph construction and enrichment. The overarching contribution of this thesis is to introduce new methods to increase the coverage of certain entities in a KG. The four main technical contributions of this thesis are as follows: (1) Comparative study that compares the state-of-the-art profile matching methods on Online Social Network (OSNs). (2) We propose new techniques to refine specific entities such as academic entities (e.g., authors) and social event entities (e.g., festivals). For academic entities, we introduce new methods to identify their corresponding Online Social Network (OSN) links such as Facebook links. To address this objective, we investigate methods for matching OSN user profiles based on novel features. For social events, we introduce in addition a new approach to evaluate the overall public sentiment related to these events over time collected from OSNs. We evaluate the performance of our methods on several real-world datasets and show that they outperform the state of the art and produce high-quality results. (3) We introduce a number of measures to explore the user profile scope on multiple OSNs. Through these measures, we analyze three axes: (a) the user profile attributes, (b) user profile content, and (c) user social network. (4) We introduce a novel user profile matching method to interlink users across multiple OSNs that leverages two fundamental matching features: (a) life events and (b) profile biographies. Life events (e.g., graduation) are used to improve the content matching process, and biographies (a short description that OSN users write about themselves) are used to improve the attribute matching process. In conclusion, we show how leveraging data from multiple OSNs is important to complete missing information of many entities inside KGs.

Jury:
Prof. Dr. Philippe Cudré-Mauroux (thesis supervisor)
Prof. Dr. Elena Mugellini (thesis co-supervisor)
Prof. Dr. Mohamed Dbouk (external examiner)
Prof. Dr. Omar Abou Khaled (internal examiner)
Prof. Dr. Laurent Donzé (president of the jury)