| | |
|---|---|
| **Reference:** | **Truth Inference Methods (KG&HC:TI)** |
| **Topic:** | **Truth Inference Methods for Extending Knowledge Graphs with Human Computation** |
| LVA-Type: | Diplomarbeit |
| Start: | Ab sofort |
| End: | Nach Vereinbarung |
| Contact: | Marta Sabou (marta.sabou@ifs.tuwien.ac.at) |
| | Stefan Biffl (Stefan.Biffl@tuwien.ac.at) |
| | Artem Revenko (artem.revenko@semantic-web.com) |

## Background

Knowledge graphs (KG) are novel kinds of graph-based data structures that enable the creation of intelligent applications such as advanced search engines, recommender systems and question answering systems. A critical problem in the life-cycle of a KG is extending and keeping it up-to-date. This is a costly and time-consuming task that is hard to achieve within the boundaries of one organization, but could be achieved by involving large groups of contributors through Human Computation & Crowdsourcing (HC&C) techniques. HC&C techniques rely on splitting large and complex problems into multiple, small and easy tasks solvable by an average contributor in a suitable population and then coordinating the collection and aggregation of individual micro-contributions into a larger result (see Fig.1).
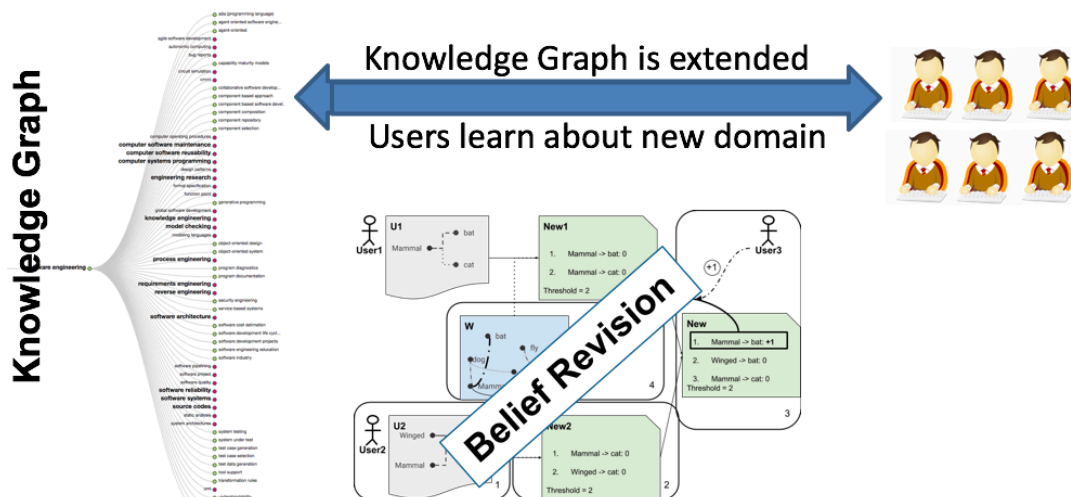


*Figure 1: Knowledge Graph extension with Human Computation.*

One important aspect in HC&C projects is **truth inference**, that is the method used to aggregate, verify and make sense of diverse contributions provided the HC@C participants, keeping those that are of the highest quality. While a range of such truth inference algorithms were proposed for generic HC&C approaches [2], HC&C methods on knowledge graphs require techniques that also take into account the knowledge stored in the KG, for example, by adopting belief revision style approaches, as we have proposed in [1].

## Tasks

This master thesis will investigate, implement and evaluate truth inference algorithms suitable for the task of extending knowledge graphs by leveraging HC&C approaches. The thesis will focus on the following topics:

- **Truth Inference Methods**: the student will extend existing or propose new truth inference methods (a minimum of two) for the process of aggregating extensions provided to a knowledge graph as part of human computation approaches. A starting point will be an approach for knowledge graph extension based on belief revision theory, see [1].
- **Knowledge Collection Interfaces**: the truth inference methods will be implemented and used to instrument two concrete human computation interfaces, namely: (1) an open-source web based interface created as part of the PROFIT project and (2) an interface based on the FigureEight crowdsourcing platform.
- **Evaluation**: Evaluation will contrast the quality of the KG extension process (both in terms of the number and correctness of the provided extensions) depending on the truth inference methods and knowledge collection interfaces used. Evaluation could be performed with diverse populations, e.g., students enrolled in advanced knowledge acquisition courses at TU Wien and/or layman crowds. Final decisions will be made within the student project.

## Experience and skills needed

The following skills are recommended, but some of them could also be learned during the thesis itself.

- Semantic Web/Linked Data technology basics
- Knowledge representation basics
- Python programming
- Java programming
- Web application development
- Software engineering
- Crowdsourcing basics and platforms (FigureEight, CML)

## References

[1] Artem Revenko, Marta Sabou, Albin Ahmeti, Martin Schauer: Crowd-Sourced Knowledge Graph Extension: A Belief Revision Based Approach. HCOMP (WIP&Demo) 2018.

[2] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. 2017. Truth inference in crowdsourcing: is the problem solved?. Proc. VLDB Endow. 10, 5 (January 2017), 541-552.