Editorial: Special Issue on Data Linking

In this special issue of the Journal of Web Semantics, we present two papers dealing both with one of the most important problem in the field of web data management: data interlinking. This field has gained significant interest over the last years, with the evolution of web technologies enabling the emergence of a web of data. The exponentially increasing number of data sources published as linked data or embedded in web pages through the use of dedicated schemas require techniques able to efficiently identify common entities appearing across these sources.

Over the last years many systems were developed involving a wide range of techniques taking into account various information about the data sets involved in order to find the most accurate links between them. Vocabularies, existing links, data ranges, ontology alignments, and user input are combined for the best results. Most efficient systems are semi-automated as they require the user to input a linkage specification, indicating what to link with what and thus guiding the tool in the process. However, for web scale data interlinking, the amount of user input in a link specification is still too high. Most recent research thus focus on minimizing the user input. The two papers in this special issue are presenting research results going in this direction, each of them following a specific path to achieve a similar goal.

In the first paper Active Learning of Expressive Linkage Rules using Genetic Programming, the authors of the interlinking tool Silk present a technique to automate the construction of linkage specifications through active learning and genetic algorithms. The resulting system only requires the user to validate a few links until an acceptable specification is reached.

In the second paper An Automatic Key Discovery Approach for Data Linking, Fatiha SAIS, Nathalie Pernelle, and Danai Symeonidou propose a technique to automate the selection of predicates to be compared during the interlinking process. The method discovers sets of properties allowing to identify data resources uniquely in a given data set, similarly to the notion of keys in relational databases.

Both articles have gone through a very rigorous selection process and were both improved since their first submission. It was an editorial choice to only retain articles meeting a very high standard, resulting in only two articles published. We believe this will ensure a stronger field of research. Enjoy reading!

The Editors:
Alfio Ferrara
Andriy Nikolov
François Scharffe