

Knowledge Linking for Online Statistics

Marc Spaniol*

Max-Planck-Institut für Informatik, Saarbrücken, Germany
mspaniol@mpi-inf.mpg.de

Natalia Prytkova

Max-Planck-Institut für Informatik, Saarbrücken, Germany
natalia@mpi-inf.mpg.de

Gerhard Weikum

Max-Planck-Institut für Informatik, Saarbrücken, Germany
weikum@mpi-inf.mpg.de

The LAWA project investigates large-scale Web (archive) data along the temporal dimension. As a use case, we are studying Knowledge Linking for Online Statistics.

Statistic portals such as eurostat's "Statistics Explained" (http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Main_Page) provide a wealth of articles constituting an encyclopedia of European statistics. Together with its statistical glossary, the huge amount of numerical data comes with a well-defined thesaurus. However, this data is not directly at hands, when browsing Web data covering the topic. For instance, when reading news articles about the debate on renewable energy across Europe after the earthquake in Japan and the Fukushima accident, one would ideally be able to understand these discussions based on statistical evidence.

We believe that Internet contents, captured in Web archives and reflected and aggregated in the Wikipedia history, can be better understood when linked with online statistics. To this end, we aim at semantically enriching and analyzing Web (archive) data to narrow and ultimately bridge the gap between numerical statistics and textual media like news or online forums. The missing link and key to this goal is the discovery and analysis of entities and events in Web (archive) contents. This way, we can enrich Web pages, e.g. by a browser plug-in, with links to relevant statistics (e.g. eurostat pages). Raising data analytics to the entity-level also enables understanding the impact of societal events and their perception in different cultures and economies.

Key Words: Entity Disambiguation, Knowledge Base Alignment, eurostat Linking Experiments, Entity-level Web (Archive) Analytics