

Mining Future Internet Workshop Abstract

Olga Streibel¹ and Achim Rettinger² and Jose Quesada³

¹ Institute for Computer Science, CSW Group, Free University Berlin,
`streibel@inf.fu-berlin.de`,

² Karlsruhe Institute of Technology,
`rettinger@kit.edu`

³ Max Planck Institute
`quesada@gmail.com`

Knowledge representations used for the next generation Internet are expected to provide machine understandable semantics based on structured representations (i.e. RDF). The main research focus so far has been on how deductive reasoning can be utilized in those representations to provide more advanced applications. However, having this more expressive structured representations opens up new opportunities for more intelligent data processing: While the current Web is mainly about key word search, the future Internet and future Web offers data mining and machine learning techniques an extensive, ready to use source of structured information for sophisticated inductive inference tasks, like uncertain query answering. The First International Workshop on Mining the Future Internet offers a platform for discussing algorithms and applications for probabilistic inductive inference and statistical analysis of future internet knowledge bases like the large amounts of Linked Data sets already being available. From efficient representation of feature vectors based on RDF data, through intelligent extraction of complex features under the use of semantic technologies, up to learning complex relations and analyzing data sets build on expressive formal logics like OWL - MIFI brings together practical and theoretical research focused on statistical learning approaches that aim to contribute to the vision of the Future Internet. MIFI should also open up the discussion concerning the relevance of factors like time, context and user influencing mining approaches for the Future Internet.

Data representations on the Internet are constantly evolving and a trend towards more structured, more semantic-based representation of data can be observed on the Web. How easy will it be to use the semantically annotated data for automated processing? Are we ready to take benefits from the semantics added to the data? How are inconsistency and dynamics of data and its structure, quantity of data, influencing the evolvement of future internet knowledge bases? We believe that statistical learning methods will play a key role in extracting information, processing, predicting and analyzing the data from FIKBs. How will this affect the statistical machine learning? Will a new generation of machine learning techniques arise within the Future Internet?