ANNUAL REPORT 2012

Mission

The Word Wide Web (WWW) has undergone significant changes, since it was initially launched, evolving from a set of interlinked static pages to a global platform for accessing resources, dynamic content and a multitude of different media. One of the main driving factors behind the evolution and growth of the Web is undoubtedly the social aspect, which is visible through the plenitude of social network and collaborative pages. However, up to date, the fact that the Web represents not only an indispensible source of data but also a collection of a wide range of opinions, viewpoints, mindsets and backgrounds, has remained commonly ignored. In particular, the design foundations and core technological components of the Web have led to an unprecedented growth and mass collaboration. However, the existence of many different perspectives and ways of reflecting on an event, or describing the characteristics of a particular product, person, company, country, etc., is still unexplored in terms of browsing tools or Web applications. Currently there is no widely used approach towards aggregating, exploring, analysing and harvesting the multi-faceted view that the Web provides on a single topic.

RENDER addresses precisely these challenges by developing methods, techniques, software and data sets that leverage **diversity as a crucial source of innovation and creativity**, whilst providing enhanced support for feasibly managing data at very large scale, and for designing novel algorithms that reflect diversity in the ways information is selected, ranked, aggregated, presented and used. In particular this is realized through RENDER's information management solution, which can handle very large amounts of data and hundreds of thousands of users, as well as a plurality of viewpoints and opinions. This is showcased through the usage of realistic data sources with billions of items; through open source extensions to popular communication and collaboration platforms such as MediaWiki, Drupal, and Twitter, and through three high-profile case studies.

Visit us at www.render-project.eu

Supported by SEVENTH FRAMEWORK PROGRAMME

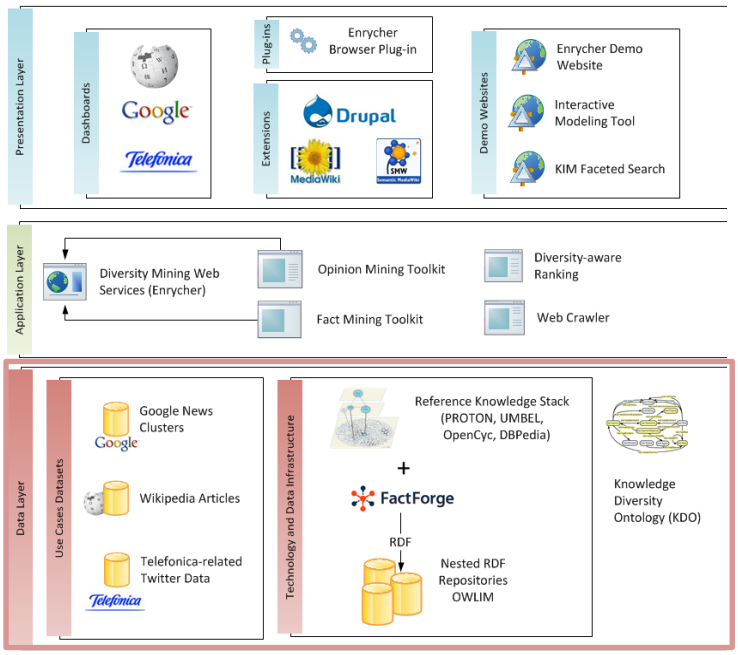
RENDER

Reflecting Knowledge Diversity

STI · INNSBRUCK KIT Google WIKIMEDIA Telefónica ontotext Jožef Stefan Institute

ANNUAL REPORT 2012

Summary of the second year of the project



RENDER has completed a very successful second year, building on the foundation laid during the first year and providing the necessary data and infrastructure for realizing the case studies, which are in the focus of the final year of the project. In particular, the conducted scientific and technical work is visible through the enriched pallet of diversity tools and visualization support. Dissemination, exploitation and community building have yield promising results: we were very active in the Wikimedia community, participating through different partners at a variety of events addressing specific stakeholders of this community; our main achievements have been published at workshops and conferences in all relevant scientific communities, and we maintain a lively Web site and social media communication ecosystem. Through initiatives such as the EU project networking session, during the European Semantic Web Conference ESWC2012, we had the chance to exchange ideas and identify avenues of collaboration with other ongoing projects in the area of intelligent information management.

In a nutshell, the second year of the project has fully met its targets both from scientific and technology perspectives. The project has built a robust, mature set of research components, dealing with knowledge representation, opinion mining, fact checking, search, ranking and summarization, which have been integrated into various application scenarios, including the three use cases. Our large-scale data management technology and associated knowledge sources (OWLIM, FactForge, news and social media crawler) have been further developed and are now an integral part of the case study prototypes. In addition to the showcases realized by Wikimedia, Telefónica and Google, the project has created diversity-minded extensions of well-known communication and collaboration platforms. In particular, we created a news-oriented demo of Drupal, which brings together services such as Enrycher, OWLIM, KDO and our diversity-aware ranking methods, and contributed to the Wikidata project, which relies on MediaWiki.

RENDER

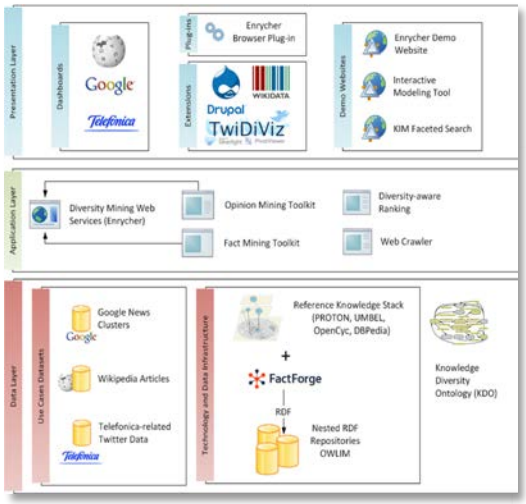
Reflecting Knowledge Diversity

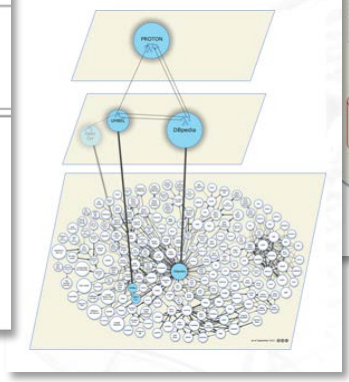
STI · INNSBRUCK · KIT · Google · WIKIMEDIA · Telefónica · ontotext · Jožef Stefan Institute

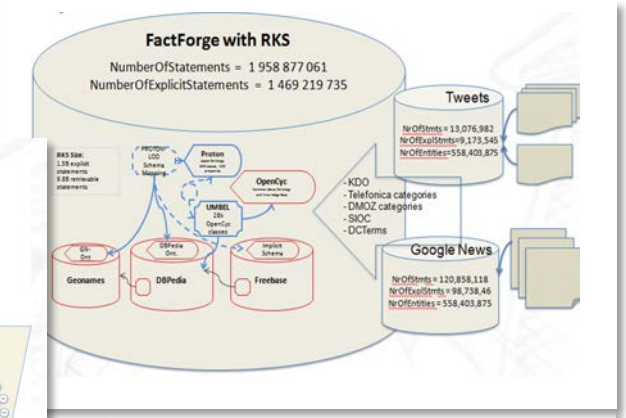
ANNUAL REPORT 2012

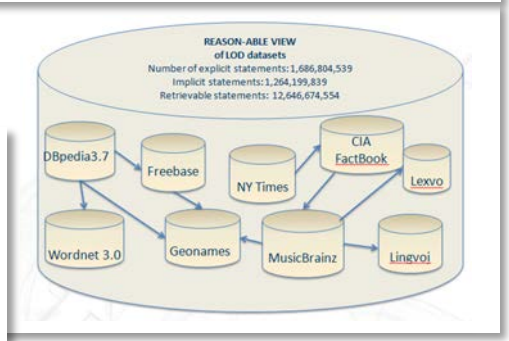
Data management infrastructure

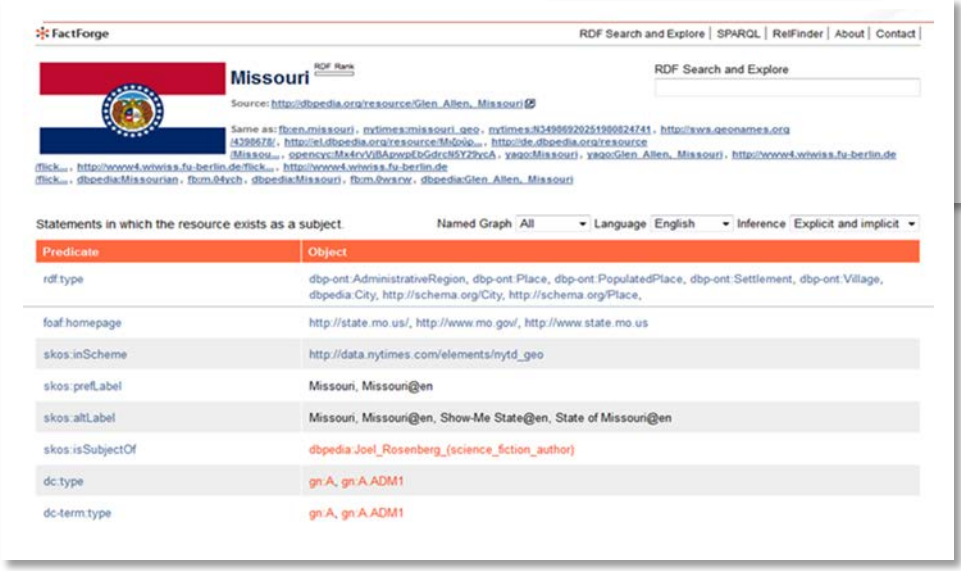
Stable, scalable and fully functional version of the infrastructure











Practical application of nested repositories

The implementation of nested repositories is completed and is now available to the project. Test adjusting and loading of news and tweets in two nested repositories was achieved successfully.

Visit us at www.render-project.eu

Supported by SEVENTH FRAMEWORK PROGRAMME

RENDER
Reflecting Knowledge Diversity

ANNUAL REPORT 2012

Advanced data collection

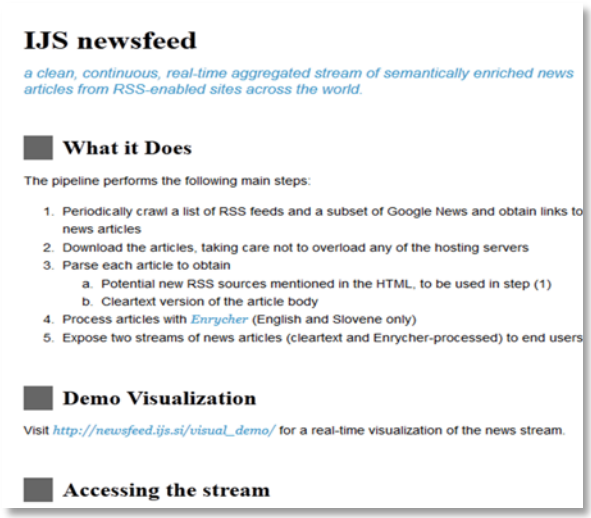
Stable, scalable and fully functional version of the infrastructure

Data sources are RSS-enabled news sites worldwide and Subset of Google News.

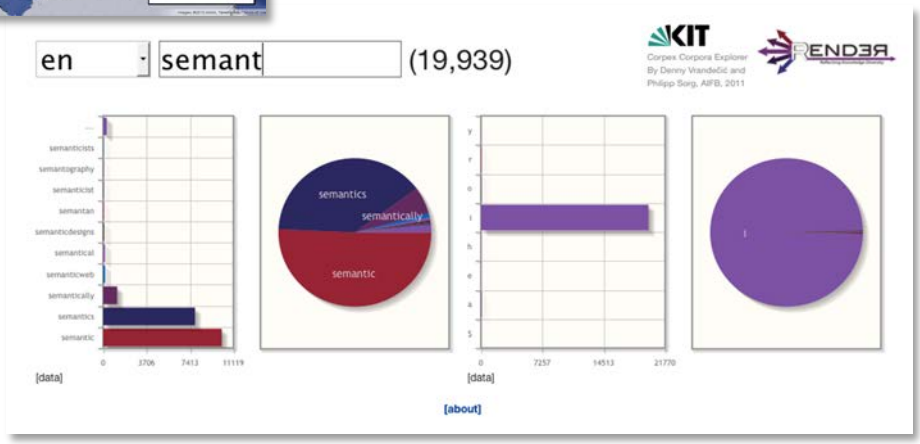
Statistics

- about 110.000 unique websites
- about 120.000 articles / day
- Languages: 50% English, 9% German, 7% Spanish, 5% French, 3% Italian
- current archive of about 60.000.000 articles

Output: Cleartext, Metadata (URL, time, title, publisher, geotags and Linked to a subset of news.google.com clusters), Language detection and Semantic annotations (Enrycher; English articles only) e. g. named entities, DMOZ categorization, sentiment and deep parsing.



Corpex is a natural language corpus extracted from Wikipedia, for several dozen languages, where some of them did not have any such language corpus before. Corpex is provided via a web service API, a web site (Screenshot), and as full downloads of the complete dataset.



Visit us at www.render-project.eu

Supported by SEVENTH FRAMEWORK PROGRAMME

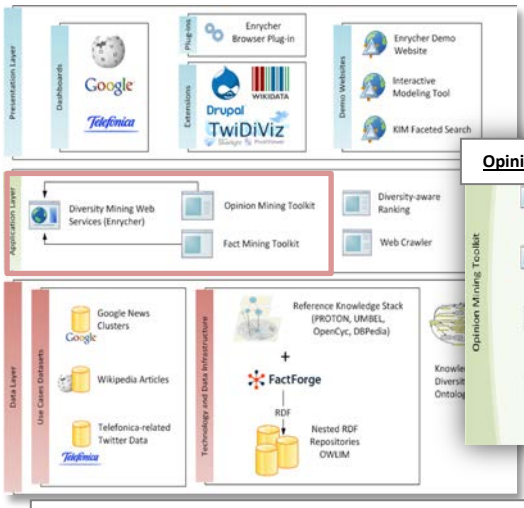
RENDER

Reflecting Knowledge Diversity

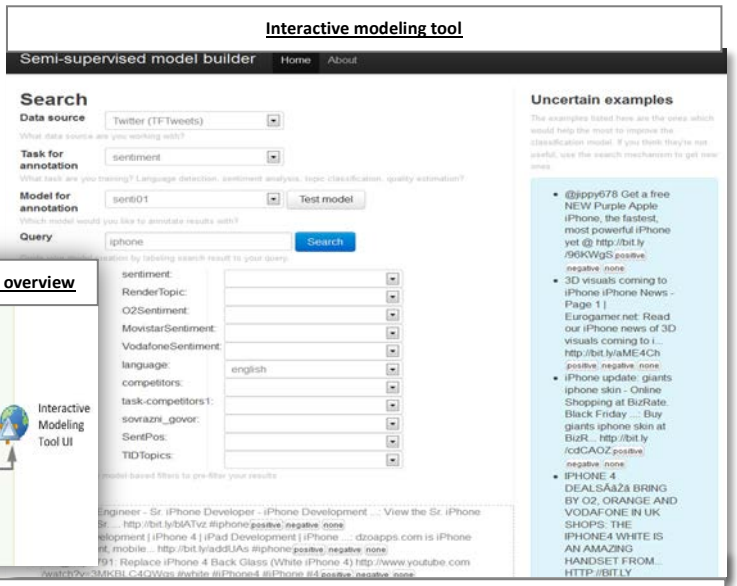
ANNUAL REPORT 2012

Diversity mining services

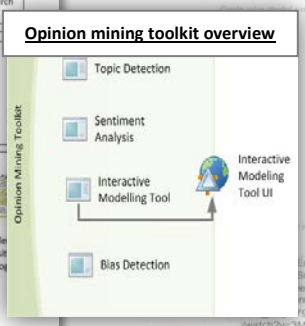
Integrated with the data management infrastructure, applied at Wikipedia and Telefónica



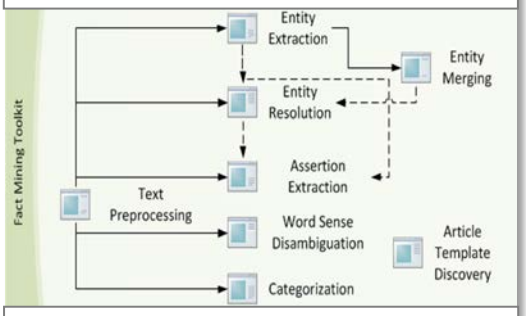
Interactive modeling tool



Opinion mining toolkit overview



Fact mining toolkit



- **Maintenance of components (tweaks, bug fixes)**
- **Evaluation of assertion extraction**
- **Reworked the article template discovery component**

Enrycher Integration

The **diversity mining services** developed within RENDER are applied when analysing Wikipedia articles or data related to Telefonica. In the case of Wikipedia articles, we use various diversity services such as topic and entity identification together with Wikipedia articles specific features, like article reference changes in order to learn which Wikipedia articles express neutral points of view. Furthermore, we identify topics, sub-topics and sentiments expressed towards Telefonica products and services using a combined approach based on active learning and classification.

RENDER

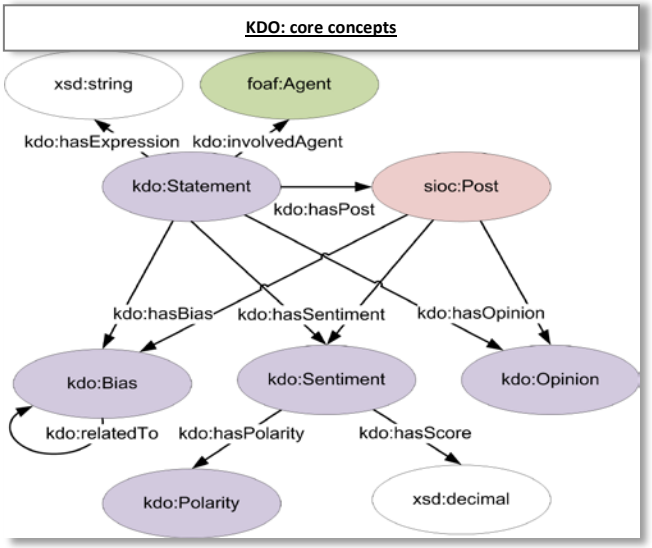
Reflecting Knowledge Diversity

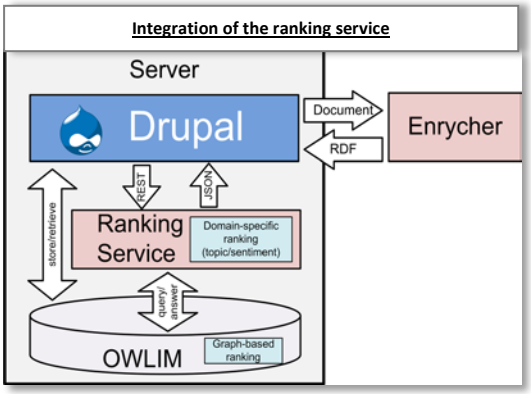
STI · INNSBRUCK KIT Google WIKIMEDIA Telefónica ontotext Jožef Stefan Institute

ANNUAL REPORT 2012

Diversity ranking service

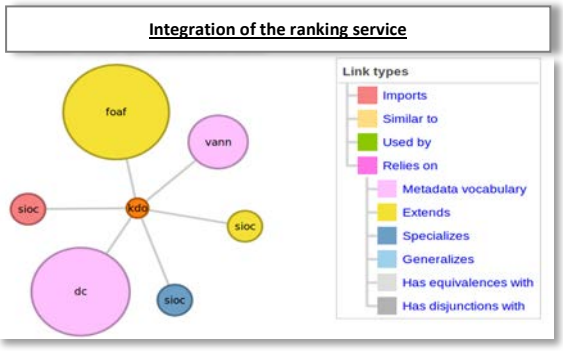
Knowledge diversity ontology and prototype of diversity ranking service





The focus of **diversity-aware ranking** on designing and implementing is a ranking algorithm that takes into account the output of the opinion and fact mining toolkits, i.e. structured data in KDO format. The dimensions of topic coverage and sentiment are considered. The prototype is available as a RESTful service that enables to rank data on any RDF store, which contains data in KDO compliant format. The diversity-aware ranking also based on spreading activation, RDFrank and clustering. The work also includes the creation of highly efficient library for spreading activation implemented for NVIDIA CUDA GPU.

The **Knowledge Diversity Ontology (KDO)** is updated, extended and made dereferencable. In addition KDO is also registered with LOV (Linked Open Vocabularies). An evaluation framework was defined and used to evaluate the KDO ontology. KDO is used in describing the diversity data used by the other components of the RENDER infrastructure and valuable feedback was received from parties using the ontology. We also performed an extensive evaluation of the Knowledge Diversity Ontology.

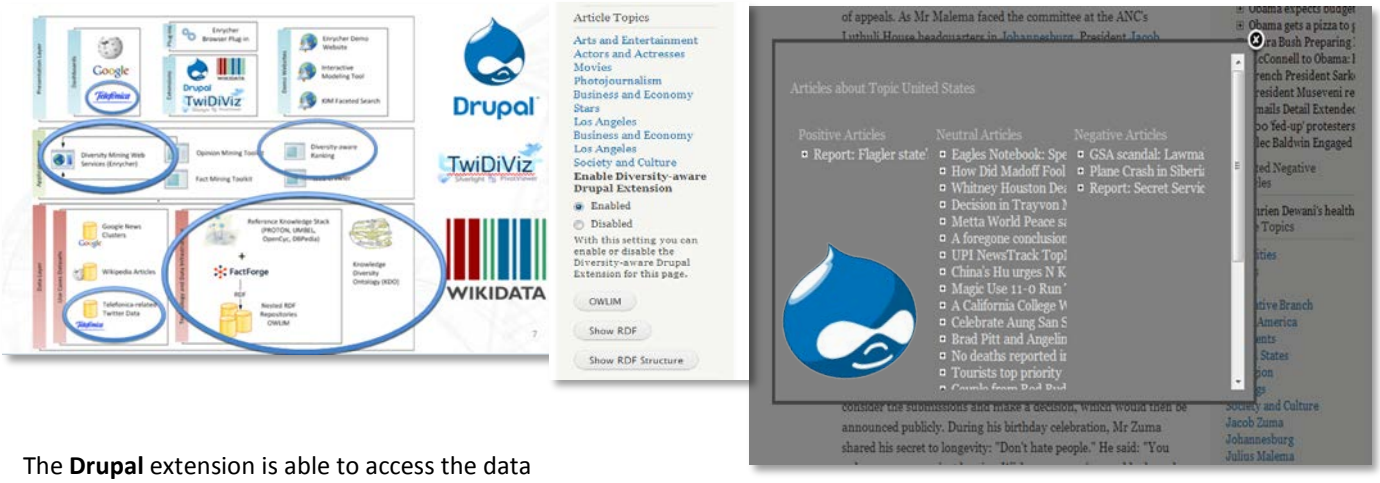


RENDER

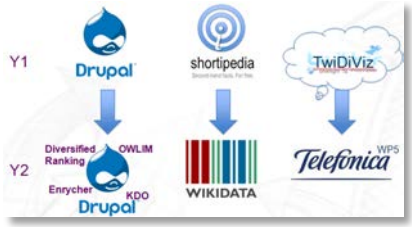
Reflecting Knowledge Diversity

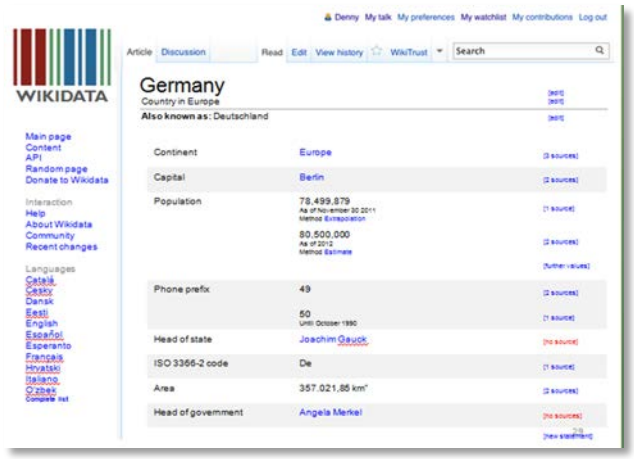
ANNUAL REPORT 2012

Diversity toolkit



The **Drupal** extension is able to access the data layer and integrate the diversity aware ranking service. We further worked on an even closer integration of the Drupal extension with the OWLIM repository.

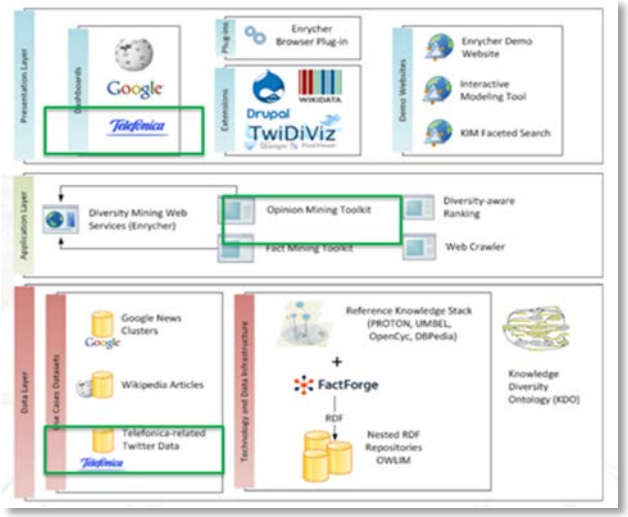


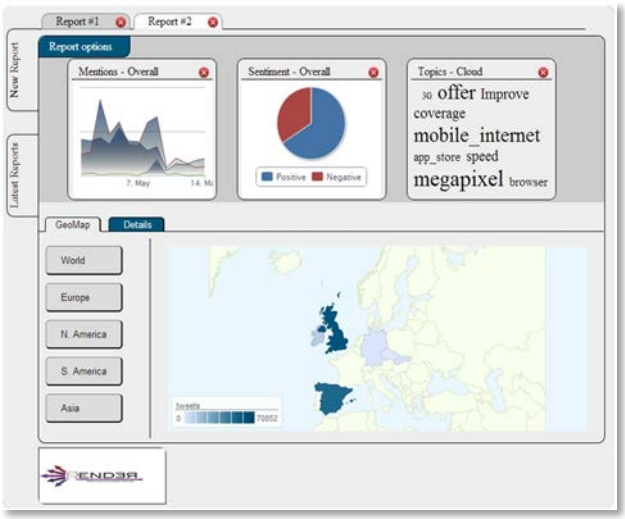


During the second project year our efforts were focused on further developing the already built prototypes of diversity-aware extensions to collaborative systems (as well as developing a new Twitter visualization tool) for testing and putting them to work in the best-suited use case scenarios.

ANNUAL REPORT 2012

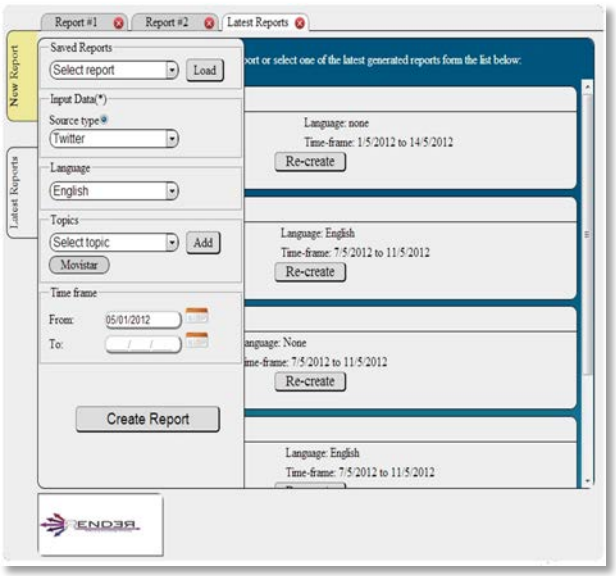
RENDER Case Studies I





Work on the architecture definition is done, according to the requirements of the case study. We provide a user interface (UI) for annotating topics and sentiments and creating learning models. Furthermore we tested the UI based on topic detection and sentiment analysis and checked the first results. The tests are conducted on the Twitter data, but can be easily extended to other data available in the case study: e-mails, surveys.

The Interactive Modelling Tool, which supports the annotation of Twitter data is extended. The tool now operates on the entire TID dataset; moreover, additional metadata is extracted from the tweets, in order to support the Telefónica dashboard. The tool's appearance and the ways of interacting with it were improved.



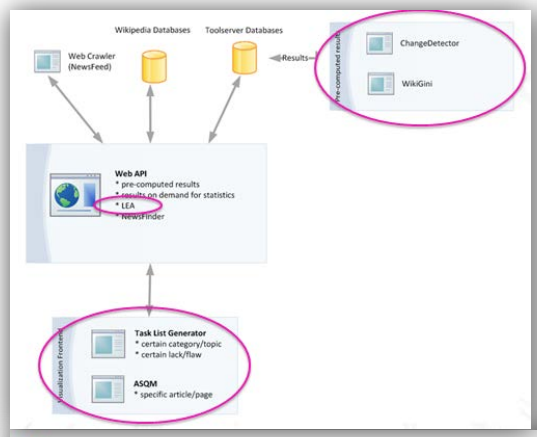
RENDER

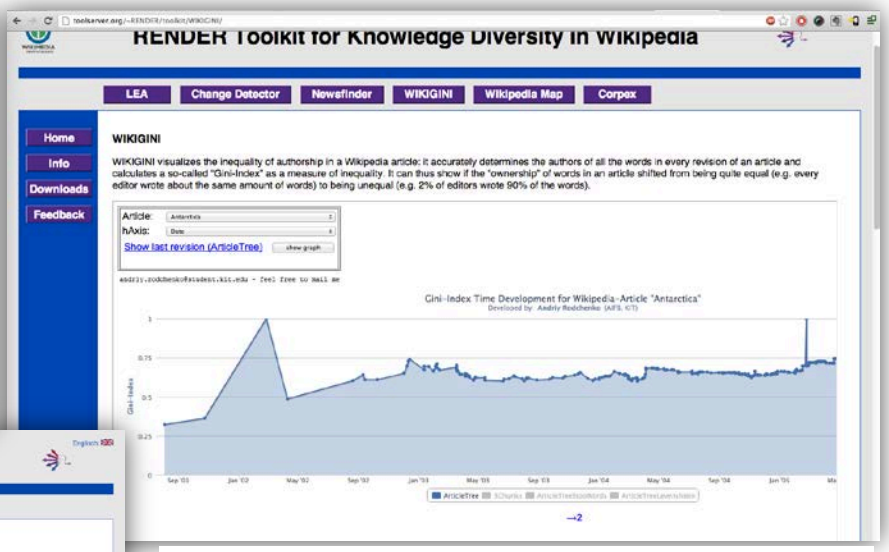
Reflecting Knowledge Diversity

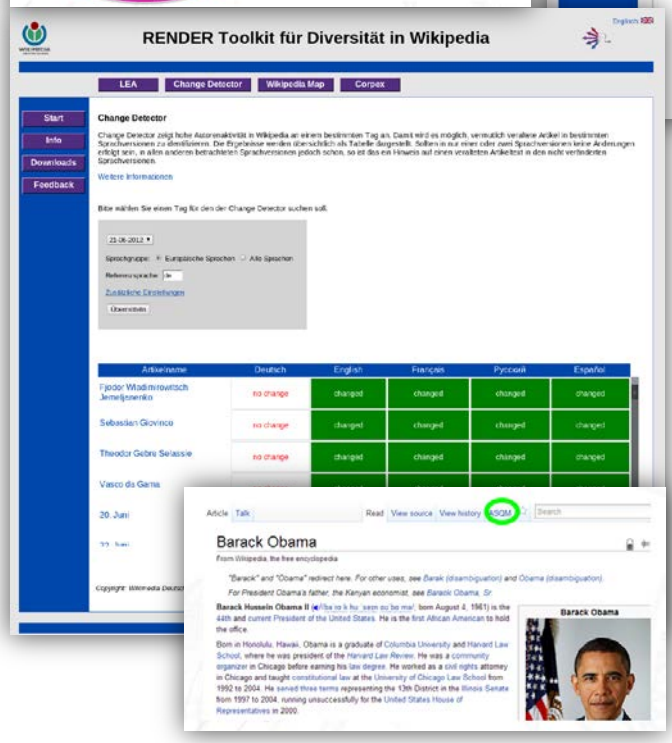
STI · INNSBRUCK · KIT · Google · Wikimedia · Telefónica · ontotext · Jožef Stefan Institute

ANNUAL REPORT 2012

RENDER Case Studies II







Wikimedia designed and visualized the overall architecture for the case study. Additionally, we started the development of two analysis tools to calculate different aspects of diversity in Wikipedia - the timeliness and the fact coverage - by using different language versions of an existing Wikipedia article. The Article Statistics and Quality Monitor (ASQM) displays the diversity analysis results for every article. This tool will include the possibility to check for different metrics and provide a quick overview of the quality and the state of an article. The Task List Generator (TLG) enables a Wikipedia author to generate lists of articles related to a specific topic or preferred category, which need to be improved. The ASQM tool has been installed as a gadget in the Wikipedia user page and the TLG tool can be accessed on the Toolserver.

RENDER
Reflecting Knowledge Diversity

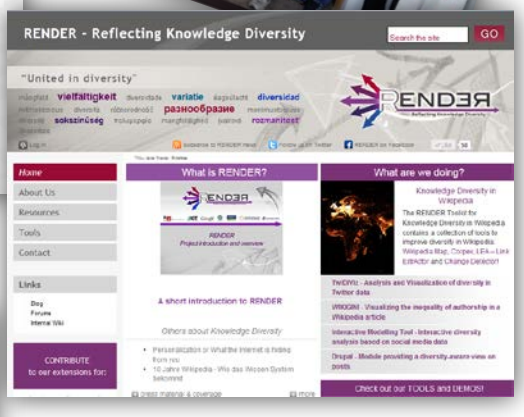
STI · INNSBRUCK KIT Google WIKIMEDIA Telefónica ontotext Jožef Stefan Institute

ANNUAL REPORT 2012

Impact and community building

Community building activities in Y2 include press releases, blog entries, meet-ups with the community, sessions organized at conferences and other community gatherings.

- **WikiData Summit 2012**
  Dissemination: RENDER presentation by Wikimedia
  Venue: Berlin, Germany
  Date: May 31 – June 1, 2012
  Web site: http://meta.wikimedia.org/wiki/RENDER/Berlin_summit_2012/
- **Berlin Hackathon 2012**
  Dissemination: Discussions with Wikimedia community developers about REDNER use cases
  Venue: Berlin, Germany
  Date: June 1-3, 2012
  Web site: http://www.mediawiki.org/wiki/Berlin_Hackathon_2012
- **ACM WebScience**
  Dissemination: Poster "A Knowledge Diversity Dashboard for Wikipedia" by Wikimedia
  Venue: Evanston, IL, USA
  Date: June 22-24, 2012
  Web site: http://www.websci12.org/



The RENDER website, Twitter and Facebook

The project website offers a wide variety of information for all interested parties. Besides general information on the project, its goals and case studies, and its partners, the website offers research results and publications as well as different press materials like the project factsheet and flyer, and related information. Visit us at www.render-project.eu