



ANNUAL REPORT 2012

Mission

The World 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 indispensable 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

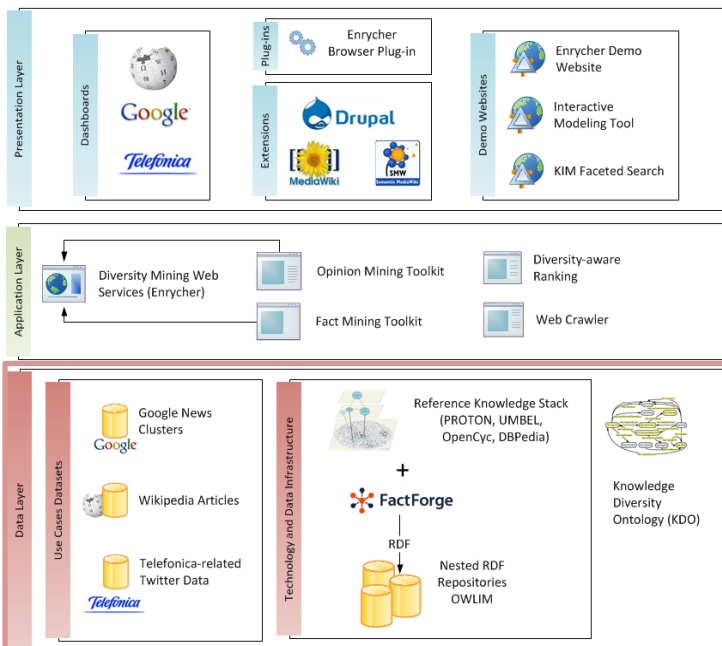




ANNUAL REPORT 2012

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.

Visit us at www.render-project.eu

Supported by





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.



IJS newsfeed

a clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS-enabled sites across the world.

What it Does

The pipeline performs the following main steps:

1. Periodically crawl a list of RSS feeds and a subset of Google News and obtain links to news articles
2. Download the articles, taking care not to overload any of the hosting servers
3. Parse each article to obtain
 - a. Potential new RSS sources mentioned in the HTML, to be used in step (1)
 - b. Cleartext version of the article body
4. Process articles with *Enrycher* (English and Slovene only)
5. Expose two streams of news articles (cleartext and Enrycher-processed) to end users

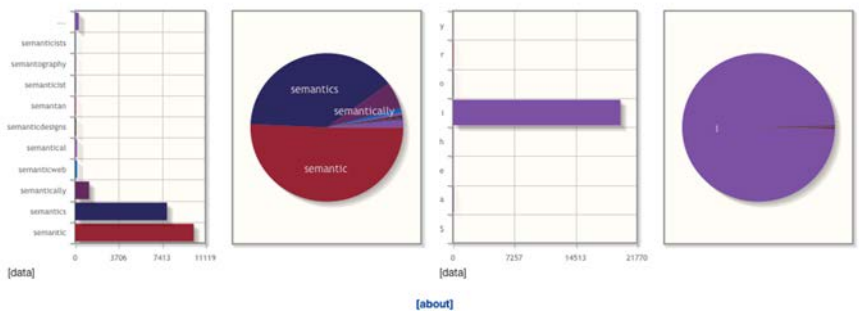
Demo Visualization

Visit http://newsfeed.ijs.si/visual_demo/ for a real-time visualization of the news stream.

Accessing the stream

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.

en - semant (19,939)



Visit us at www.render-project.eu

Supported by

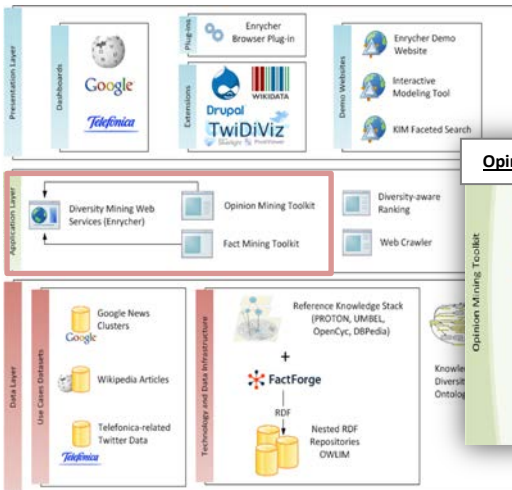




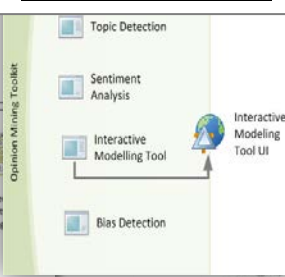
ANNUAL REPORT 2012

Diversity mining services

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



Opinion mining toolkit overview



Interactive modeling tool

Semi-supervised model builder Home About

Search

Data source:

What data source are you working with?

Task for annotation:

Model for annotation:

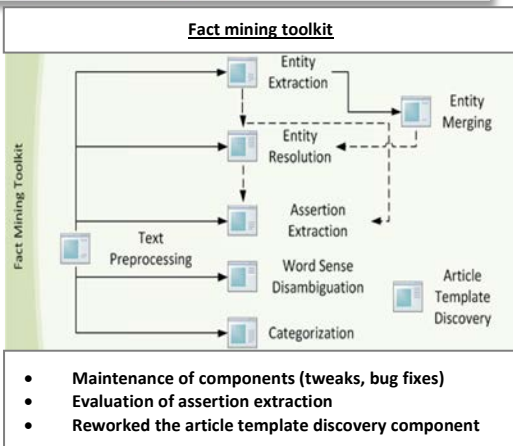
Which model would you like to annotate results with?

Query:

Uncertain examples

The examples listed here are the ones which would help the most to improve the classification model. If you think they're not useful, use the search mechanism to get new ones.

- @bippy678 Get a free NEW Purple Apple iPhone, the fastest, most powerful iPhone yet @ http://bit.ly/i96KWgS positive negative none
- 3D visuals coming to iPhone iPhone News - Page 1 | Eurogamer.net. Read our iPhone news of 3D visuals coming to i... http://bit.ly/iAME4Ch positive negative none
- iPhone update grants iPhone skin - Online Shopping at BuRate Black Friday ... Buy grants iPhone skin at BuR... http://bit.ly/iC8CAOZ positive negative none
- IPHONE 4 DEALS! A2a2a BRING BY O2, ORANGE AND VODAFONE IN UK SHOPS: THE IPHONE 4 WHITE IS AN AMAZING HANDSET FROM... HTTP://BIT.LY/...



Enrycher Integration

text	entities	keywords
<p>Dutch police arrest seven suspected of planning attack</p> <p>AMSTERDAM (AP) -- Police in the Netherlands say they've broken up what may have been a terrorist plot to attack stores in Amsterdam, including an Ikea. Among the seven people arrested in the case is a relative of one of the terrorists who died in the 2004 bombings in Madrid. Amsterdam's mayor says an anonymous tip to police warned that terrorists planned to set off explosives in shops to cause casualties in busy places. The tip last night came through an unregistered cell phone and identified a suspect and locations for police to search. Earlier today, authorities in Amsterdam shut down a major shopping street near a soccer stadium, sealed off an Ikea and canceled a concert by the American group, "The Killers." The suspects are identified as six men and one woman. All are Dutch nationals of Moroccan ancestry.</p>	<ul style="list-style-type: none"> • AMSTERDAM • AP • Police • the Netherlands • Madrid 	<p>Society, Regional, North_Holland, Netherlands, Europe, Law_Enforcement</p> <p>categories</p> <ul style="list-style-type: none"> • Regional/.../North_Holland/Amsterdam/ • Regional/.../Netherlands/North_Holland/ • Society/.../Law_Enforcement/Police • Society/.../Police_Associations/ • Regional/.../Amsterdam

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.

Visit us at www.render-project.eu

Supported by

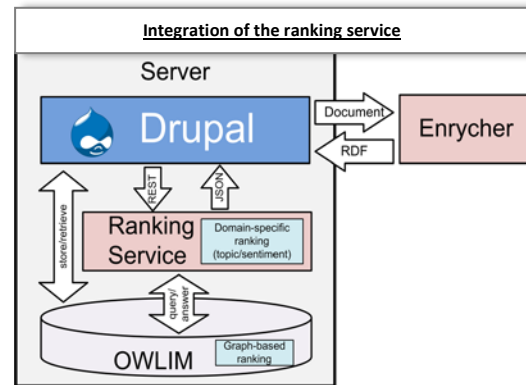
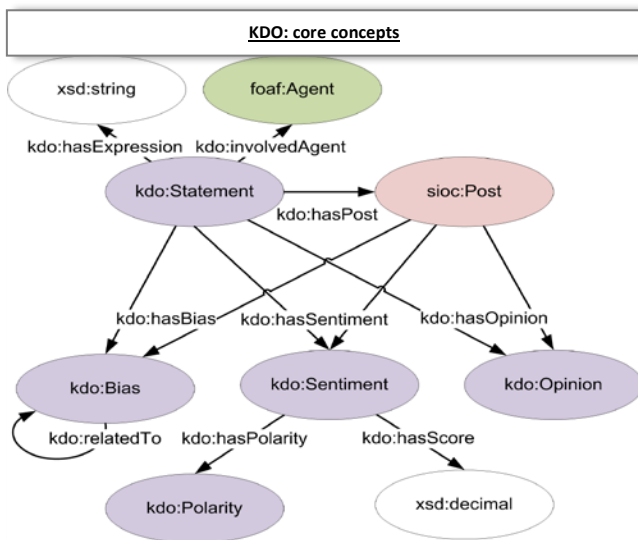




ANNUAL REPORT 2012

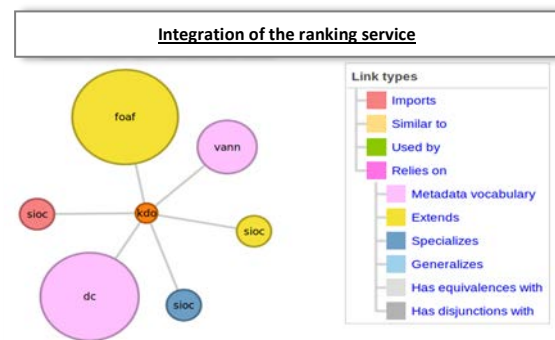
Diversity ranking service

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.



Visit us at www.render-project.eu

Supported by





ANNUAL REPORT 2012

Diversity toolkit

The screenshot shows a dashboard with several tool icons and settings. On the left, there are icons for Google, Drupal, TwiDiViz, and Wikidata. In the center, there are settings for 'Diversity-aware Ranking' and 'Enable Diversity-aware Drupal Extension'. On the right, there are settings for 'Article Topics' and 'Diversity-aware Ranking'.

The screenshot shows a web page with a search bar and a list of articles. The articles are categorized into 'Positive Articles', 'Neutral Articles', and 'Negative Articles'. The 'Positive Articles' list includes 'Report: Flagler state', 'Eagles Notebook: Spe', 'How Did Madoff Fool', 'Whitney Houston Des', 'Decision in Trayvon?', 'Metta World Peace s', 'A foregone conclusion', 'UPI NewsTrack Topi', 'China's Hu urges N K', 'Magic Use 11-0 Run', 'A California College V', 'Celebrate Aung San S', 'Brad Pitt and Angelin', 'No deaths reported i', and 'Tourists top priority'. The 'Neutral Articles' list includes 'GSA scandal: Lawma' and 'Plane Crash in Siberi'. The 'Negative Articles' list includes 'Report: Secret Serv'.

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.



The screenshot shows the Wikidata page for Germany. It includes a search bar, a list of languages, and a table of facts. The table includes:

Continent	Europe	(2 sources)
Capital	Berlin	(2 sources)
Population	78,499,879	(2 sources)
Phone prefix	49	(2 sources)
Head of state	Joachim Gauck	(10 sources)
ISO 3166-2 code	De	(1 source)
Area	357,021,85 km²	(2 sources)
Head of government	Angela Merkel	(10 sources)

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.

Visit us at www.render-project.eu

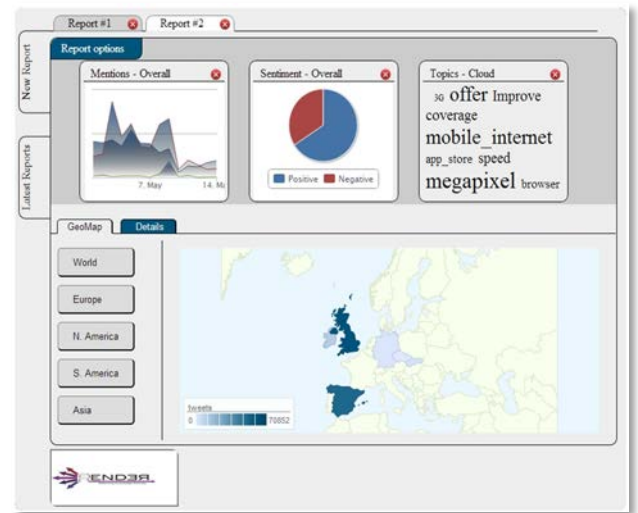
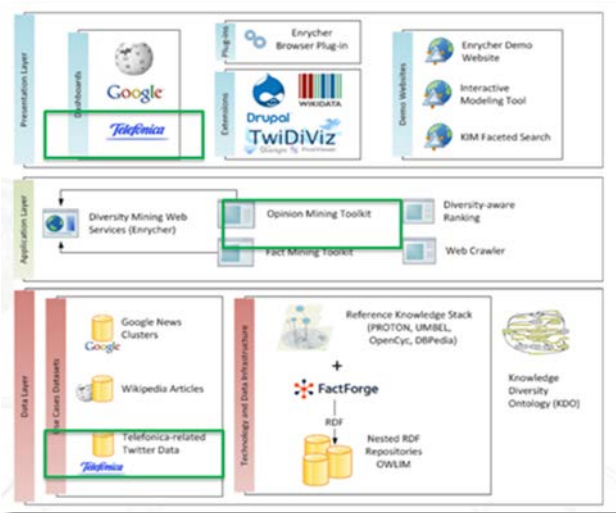
Supported by





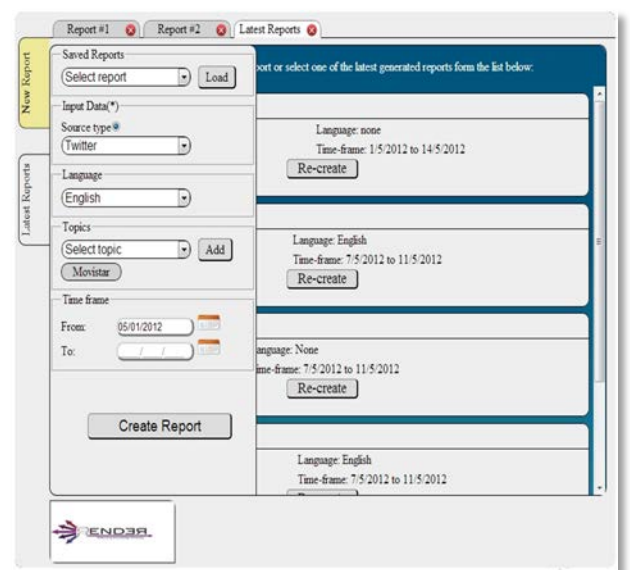
ANNUAL REPORT 2012

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.



Visit us at www.render-project.eu

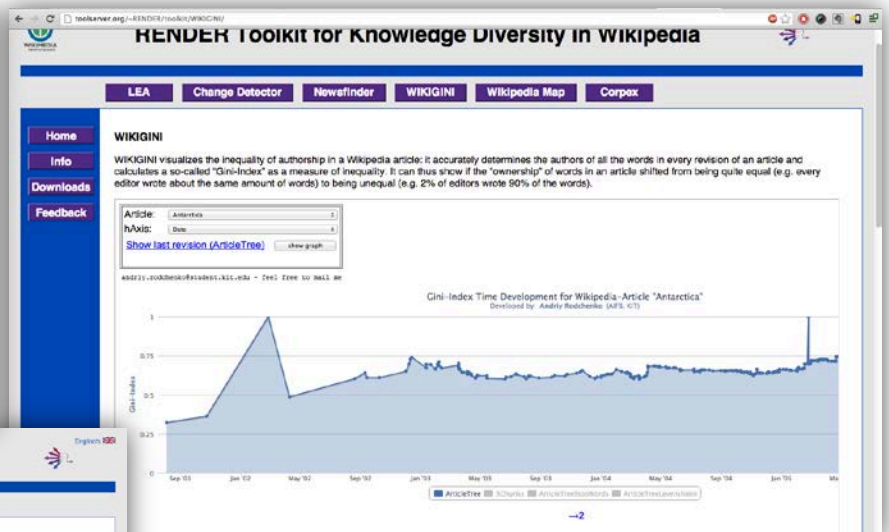
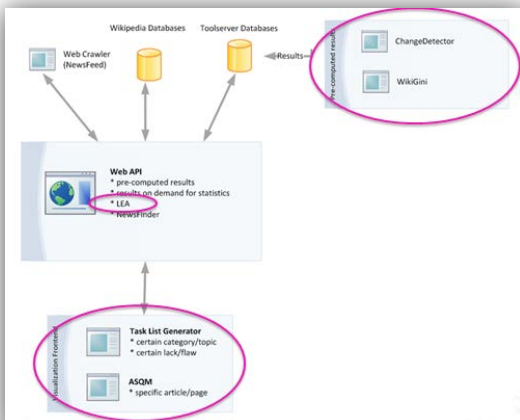
Supported by





ANNUAL REPORT 2012

RENDER Case Studies II



Artikelname	Deutsch	English	François	Pycckий	Español
Fodor Wladimirovitsch Jemeljanenko	no change	changed	changed	changed	changed
Sebastian Giovinco	no change	changed	changed	changed	changed
Theodor Gebre Selassie	no change	changed	changed	changed	changed
Vasco da Gama					

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.

Visit us at www.render-project.eu

Supported by



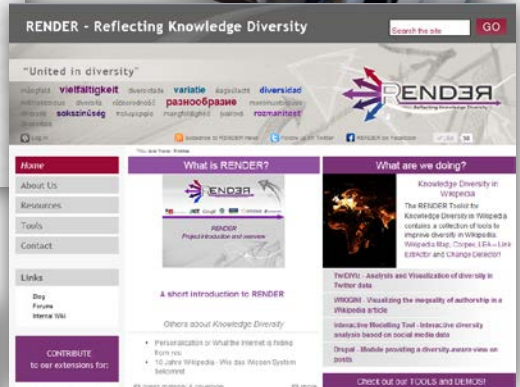


ANNUAL REPORT 2012

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

Visit us at www.render-project.eu

Supported by

