RENDER
FP7-ICT-2009-5
Contract no.: 257790
www.render-project.eu

RENDER

Deliverable D4.1.3

Impact and Refinement of Diversity-Aware Extensions to Collaborative Systems

Editor:	Fabian Flöck, KIT
Authors:	Andreas Thalhammer, UIBK; Fabian Flöck, KIT; Lydia Pintscher, Wikimedia
Deliverable nature:	Prototype (P)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	30 September 2013
Actual delivery date:	30 September 2013
Suggested readers:	Diversity aware tool developers
Version:	1.0
Keywords:	Deliverable, diversity, Wikipedia, fact coverage, collaborative, web 2.0, extensions, RDF, semantic, community

Disclaimer

This document contains material, which is the copyright of certain RENDER consortium parties, and may not be reproduced or copied without permission.

In case of Public (PU):

All RENDER consortium parties have agreed to full publication of this document.

In case of Restricted to Programme (PP):

All RENDER consortium parties have agreed to make this document available on request to other framework programme participants.

In case of Restricted to Group (RE):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement. However, all RENDER consortium parties have agreed to make this document available to <group> / <purpose>.

In case of Consortium confidential (CO):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the RENDER consortium as a whole, nor a certain party of the RENDER consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Full Project Title:	RENDER – Reflecting Knowledge Diversity
Short Project Title:	RENDER
Number and Title of Work package:	WP4 – Diversity Toolkit
Document Title:	D4.1.3 – Impact and Refinement of Diversity-Aware Extensions to Collaborative Systems
Editor (Name, Affiliation)	Fabian Flöck, KIT
Work package Leader (Name, affiliation)	Fabian Flöck, KIT

Copyright notice

© 2010-2013 Participants in project RENDER

Executive Summary

This deliverable describes the advances made in regard to the diversifying extensions to collaborative platforms in the third project year as well as their final state. These are: Wikidata, the Drupal “Diversity Enricher” module and TwiDiViz. In the case of the Drupal extension, it also includes a short evaluation, as there is no other deliverable dedicated to this content. To avoid redundancy, we excluded a description of the extensions to Wikipedia, as they are described in detail in WP5, although they have been partly developed in WP4 as well.

The first outcome is the final version of the “Diversity Enricher” module developed by RENDER, extending the widely used content-management and publishing platform Drupal. During the project, we constantly developed this extension further, learning in the process from the interaction with the community and iterating persistently over the software. No disruptive changes were implemented in the third year. Instead, a great quantity of fine-tuning was carried out to adapt the module to the conditions of a realistic usage scenario on the web. This strategy of small steps while listening to the Drupal crowd proved to be very valuable, as the results of the evaluation verify as well as the reactions of the community. These changes included application of roles, e.g., administrators, interface remodeling and unification. In addition, several code cleaning and documentation tasks were performed. We deployed the final version of the extension and enhanced the showcase with a real Blog export of http://crossingthebaltic.com. Our extension is implemented as a Drupal module that can be obtained online under an open source license.

We finalized TwiDiViz, our MS Pivot Viewer extension, which we had adapted to be a visualization tool to the Telefónica T-OMT service during the second project year. This decision served us very well as an appropriate community for this single Silverlight-based tool was not in sight and would have been difficult to construct. We got valuable feedback during the Telefónica use case evaluation, which was incorporated in the third year iterations. The updates included new and revised filters to suit the users’ needs as well as the integration with Ontotext’s RENDER Tweet repository and an update of the data interface with the T-OMT. The integration of TwiDiViz with the T-OMT was finalized while still preserving its adaptability to potential future use cases and data sources via the defined data interface. We also finalized a use case with Twitter data incorporated in the Telefónica evaluations.

We describe the Wikidata project after its completed one-year development. While Wikidata was split out in an early stage from RENDER as a standalone project, RENDER was a crucial incubator and supported its development during the initial phases. The project went to meet all expectations and has already surpassed Wikipedia in editor activity. It allows providing diverging and differently sourced data for the same facts, and as such makes the diversity in Wikipedia and its sister projects explicit. It connects the different language version via a unified semantic repository and thus ensures easier, complete and verified connections between diverse language versions of an article.

In summary it can be said that although the initial planning of certain extensions to platforms at the project start in WP4 were transformed or dropped, we adapted quite successfully and shifted our efforts and expertise in the project in a very goal-oriented manner. This turned out to be good, adaptive strategic planning: by push-starting Wikidata, by switching one tool use case from Wordpress to the larger and more developer-friendly (and –active) Drupal platform and by integrating the Twitter visualization tool with the Telefónica business use case, we achieved a higher impact for the available person months than could have been created by strictly adhering to the original project outline for WP4. Additionally, some tools for Wikipedia were initially developed in this work package. All in all it can be said that the technologies developed and deployed in WP4 have achieved most of the goals set at the beginning of RENDER; they made many platforms more diversity-friendly.

© RENDER consortium 2010 - 2013

Table of Contents

Executive Summary 3
Table of Contents 4
Abbreviations 5
1 Introduction 6
2 Wikidata 7
2.1 Technical Development 7
2.2 Deployment 8
2.3 Impact 9
2.3.1 Community Engagement and Feedback 9
2.4 Further Support 11
3 Diversity Enricher Drupal Module 12
3.1 Updated Description 12
3.1.1 Use of Roles 12
3.1.2 Revision of the Interface 13
3.1.3 Finalization and Publication of the Source Code 14
3.2 Showcase 15
3.3 Evaluation 15
3.3.1 Results 16
3.3.2 Discussion 18
3.4 Impact 19
3.5 Further Support 19
4 TwiDiViz: Analysis and Visualization of Diversity in Twitter Data 20
4.1 Updated Description 20
4.1.1 Filtering by Country 20
4.1.2 Strings for Topic Filtering 20
4.2 Showcase 20
4.3 Further Support 24
5 Conclusions 25

 © RENDER consortium 2010 – 2013

Abbreviations

CMF	Content management framework
CMS	Content management system
CXML	Collection eXtensible Markup Language
KDO	Knowledge Discovery Ontology
KIT	Karlsruhe Institute of Technology
LOD	Linked Open Data
MS	Microsoft
RDF	Resource Description Framework
SDK	Software Development Kit
SMS	Short Message Service
TID	Telefónica Investigación y Desarrollo
T-OMT	Telefónica Opinion Mining Tool
TwiDiViz	Twitter Diversity Vizualization Tool
UI	User Interface
URI	Unique Resource Identifier
WP	Work Package
SMW	SemanticMediaWiki

© RENDER consortium 2010 - 2013

1 Introduction

This deliverable describes the advances made in regard to the diversifying extensions to collaborative platforms in the third project year as well as their final state. This includes Wikidata, TwiDiViz and the Drupal extension, for which we also include a short evaluation, as there is no other deliverable dedicated to this content. To avoid redundancy, we excluded a description of the extensions to Wikipedia, as they are described in detail in WP5, although they have been partly developed in WP4 as well. We describe the Wikidata project after its completed one-year development, as it was incubated to a significant degree in the RENDER project and added substantially to reaching the goal of heightened diversity in collaborative platforms. The remainder of this deliverable is organized as follows. The Wikidata project and outcomes are described in Section 2. Section 3 describes the Drupal extension and Section 4 the TwiDiViz tool for analysis and visualization of Twitter data. Finally, Section 4 concludes the deliverable.

© RENDER consortium 2010 – 2013

2 Wikidata

The Wikidata[1] project was triggered by the idea to create a representation of diverse data sources and diverging versions for specific data facts in a Wiki, which was in fact one of the initial challenges approached within the RENDER project. Therefore, the RENDER inspired and strongly influenced the initiation of the Wikidata, which developed into a self-funded, stand-alone project alongside RENDER. Wikidata as a platform is now an essential part of the Wikimedia Foundation's strategy for structured fact data storage, linking and inclusion in Wikipedia and other Wikimedia projects. In this last deliverable on diversifying collaborative systems we will therefore report on the advances Wikidata has made during RENDER's project runtime and how it helped to achieve goals targeted by RENDER.

Wikidata has the goal to create a free knowledge base that can be read and edited by humans and machines alike. According to its main objective it aims to establish a "New open source software enabling dynamic and collaborative analysis, sharing and editing of heterogeneous datasets across multiple disciplines." It provides data in more than 300 languages, and allows for the central access to the data in a similar vein as Wikimedia Commons does for multimedia files. With the end of the initial one year development that is now finished, Wikidata is providing a powerful, wiki-based platform for the creation and maintenance of structured data, used in Wikipedia and beyond.

Wikimedia Deutschland started working on Wikidata in April 2012, with some preparatory work in March 2012. The initial development of the project was funded with donations by the Allen Institute for Artificial Intelligence [ai]2, the Gordon and Betty Moore Foundation, and Google, Inc. The initial development of Wikidata ended as planned in March 2013.

In the following we report on the 1-year (and, for now, final) results of the Wikidata project in brief. We describe the technical development, deployment, impact and future maintenance.

2.1 Technical Development

The Wikidata project implemented several MediaWiki extensions to be used in the different Wikipedia projects and in the new stand-alone platform called Wikidata (see also the following Section "Deployment"). The following extensions have been developed:

- Wikibase Repository – enabling to use a MediaWiki instance for the creation and maintenance of structured data
- Wikibase Client – enabling to use the structured data from a Wikibase repository
- Wikibase Lib – common functions used by both the repository and the client
- Diff – a PHP library for generating the differences between structured data
- DataValues – a set of libraries for handling data values, including parsers, serialization, and JavaScript interfaces for displaying and editing them
- Ask – a PHP library to represent structured queries to wiki-based knowledge bases

Wikibase Client is used in the Wikipedias, and Wikibase Repository in Wikidata.

Besides these extensions, some changes have been done to the core MediaWiki engine itself. Among them is the biggest single change to MediaWiki so far, the ContentHandler, which allows MediaWiki to deal with different types of content, and still reuse its powerful history and user features. This feature is currently used to represent structured data in Wikidata. But it is also usable for other features like an SVG editor based on MediaWiki.

In general, Wikidata is being developed fully in the open, which is why many of the design documents and drafts are publicly available. Changes and suggestions can be done directly on the documents or their related talk pages, or on the Wikidata mailing list.[2]

[1] http://www.wikidata.org/
[2] The full Technical Proposal of Wikidata is available here: http://meta.wikimedia.org/wiki/Wikidata/Technical_proposal

Wikidata itself builds on a specially designed data backend (storing triples in JSON and RDF), based on a data model that's has been created and extensively discussed with several communities. Every item in Wikidata can have a set of statements. Every statement consists of a main part (e.g., the key-value pair “Population: 3500000”). Every statement can further have qualifiers (e.g., giving a date to the main part and explicating the method it was derived from). Finally, statements can (and should) have references to corroborate their truthfulness. An API enables retrieval and editing of the content.

The user interface of the Wikidata repository provides a page for every item described by an article in one of the Wikipedia language editions. An example is given in Figure 1.

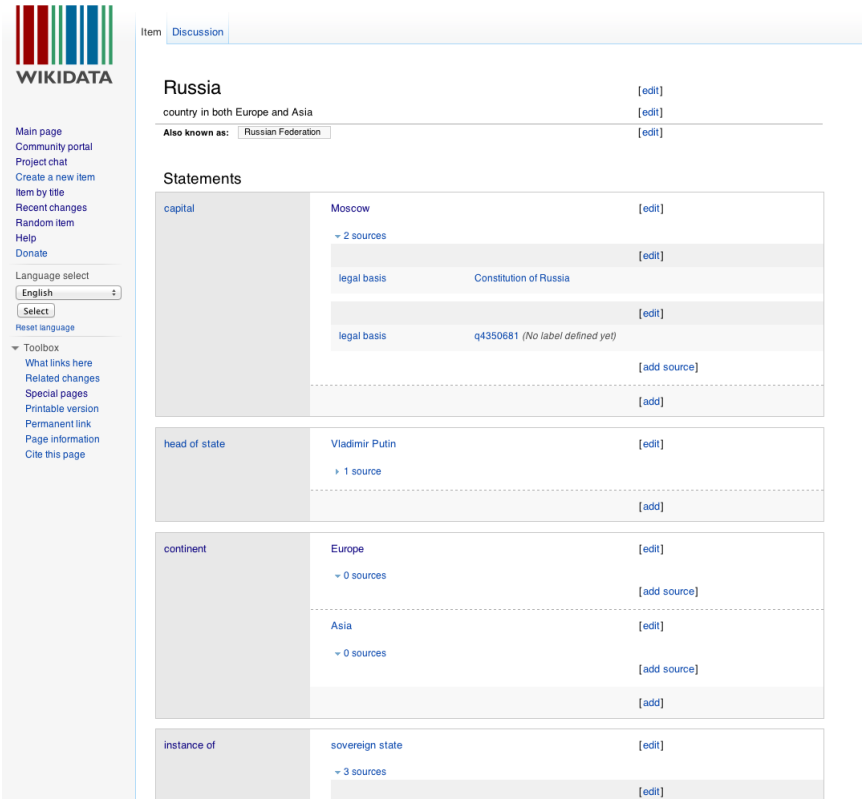


Figure 1: An Item in Wikidata

The user interface of a Wikidata repository allows direct editing of the stored items. Statements and references can be created in the UI and collaboratively maintained. Language links can be added with the help of an auto-completion suggestion of what was typed. These are offered from the Wikipedia that was chosen in the language field. The language links in the Wikidata repository are pointing to the actual Wikipedia language editions.

The technical implementation of the Wikidata platform comprises a wide range of other backend systems and extensions (for data storage, curation and entry).

2.2 Deployment

The major goal of the Wikidata project was not only to develop extensions to MediaWiki as described in the previous section, but also to deploy Wikidata itself. This section details the current status of this work. Deployment and maintenance of the Website will continue beyond the initial Wikidata development project. Since May 2012, the state of development can be followed on a public demo system that also

 © RENDER consortium 2010 – 2013

provides feedback functionality. The demo is available at http://wikidata-test.wikimedia.de. In October 2012, Wikidata itself has gone online. The site is available at http://www.wikidata.org.

The infrastructure for replacing local language links with global links has been deployed to all Wikipedias. The first Wikipedia to deploy Wikidata was Hungarian on January 2013, followed by Hebrew and Italian in January. English Wikipedia was switched in February, and the other 282 language editions followed in March. As of time of writing, more than 27 Million language links are available from Wikidata, which means that it covers basically all Wikipedia articles in all languages.

Infobox data can be added to Wikidata since February, limited to two data types (other items and multimedia files from Wikimedia Commons). On March, the data type for strings was enabled. The ability to add references to statements has been deployed in February and extended in March. Usage of infobox data on the Wikipedias has been switched on for eleven Wikipedias at the time of writing (Italian, Hungarian, Hebrew, Russian, Ukrainian, Turkish, Uzbek, Croatian, Serbian, Bosnian, and Serbo-Croatian), representing approximately 10% of the page views to the Wikipedia projects. The Wikidata team is carefully monitoring the performance impact of the rollout.

2.3 Impact

Two months after the launch of the infobox feature, more than 3.8 Million statements had been collected, of which 57% are sourced with references. Almost 300 properties have been created and are currently in use. The community has set up a process for the creation and maintenance of properties. More than 10,000 human editors had been reported mid-April, who have thus far made 3.4 Million edits. Including the three dozen or so bots running on Wikidata, there have been more than 24 Million edits, making this the currently most active Wikimedia project. Other Wikimedia projects have expressed interest in being supported by Wikidata, and the Wikidata team will explore possible use cases and their implementation costs.

As has been said, 27 Million language links are available from Wikidata, meaning that it is now the close-to-universal language interlinking method on Wikipedia. *Wikidata has about 2-3 times the edit activity of the English Wikipedia, and is thus possibly the most edited wiki in the world at the moment.*[3] In general, the Wikidata project has had more than 6,300 commits by overall 40 contributors until the end of March 2013.[4]

2.3.1 Community Engagement and Feedback

Various activities took place to involve and build the community around Wikidata. The community is engaged and following the project, for example through the weekly status updates.[5] The community was approached at numerous events and through numerous channels, which we will not list here in detail, as it is out of scope of this deliverable.

[3] http://stats.wikimedia.org/wikispecial/EN/draft/TablesWikipediaWIKIDATA.htm
[4] http://www.ohloh.net/p/wikidata/
[5] http://meta.wikimedia.org/wiki/Wikidata/Status_updates

© RENDER consortium 2010 - 2013

The following figures 2-4 show the progress of the community engagement since the beginning of the Wikidata project until end of March 2013.

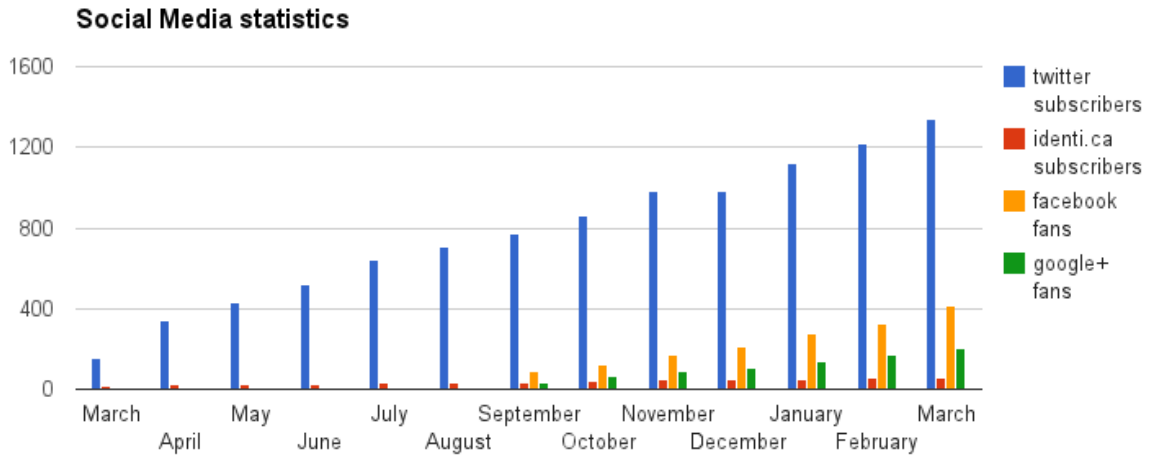


Figure 2: Social Media Statistics

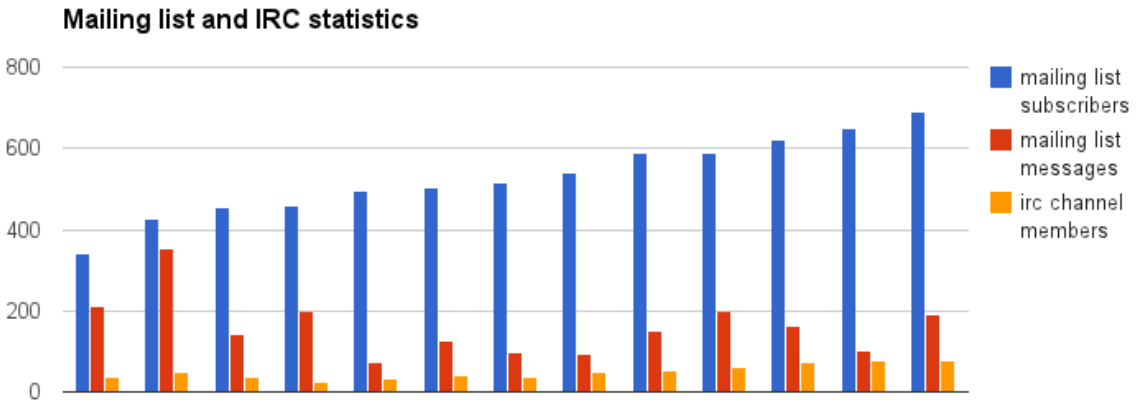


Figure 3: Mailing List and IRC Statistics

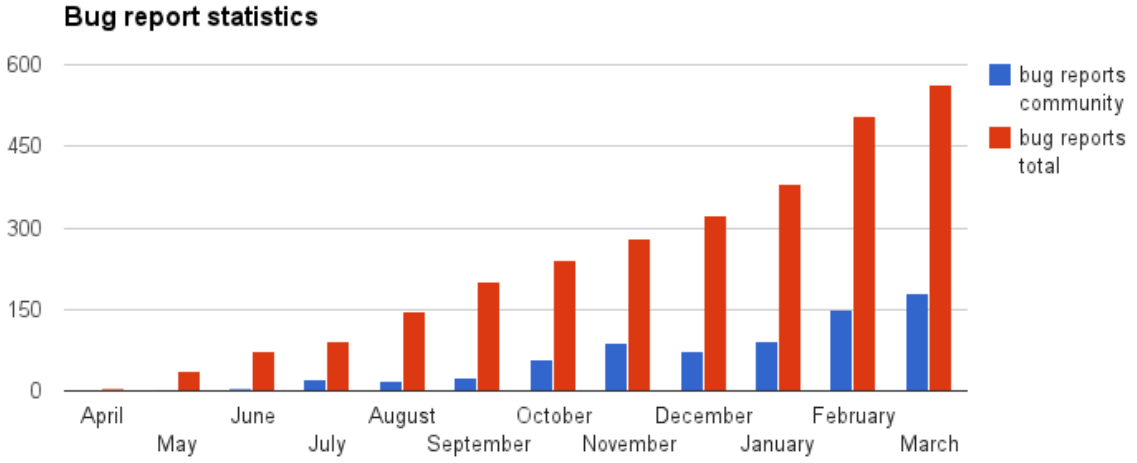


Figure 4: Bug Report Statistics

© RENDER consortium 2010 – 2013

2.4 Further Support

Wikimedia Deutschland e.V. has committed to support the development of Wikidata throughout 2013 with a reduced team. With the growing usage of Wikidata comes an increasing request for new features from the community. Other Wikimedia projects have shown interest in an integration with Wikidata. Promising candidates are Wikisource, Wikimedia Commons, Wiktionary, and Wikivoyage.

A major functionality of Wikidata could be to enhance the growing Wikimedia Commons project with language independent tags and metadata. Currently, the over 16 Million multimedia items of Commons are very hard to search and browse through, especially for users in another language than English. Wikidata is expected to be an enabler for a new language-independent search and categorization infrastructure. Wikivoyage and Wikisource both could use structured data from Wikidata in order to enrich their pages with more information coming from a common database. Also, Wikivoyage has the same interlanguage linking issue that the Wikipedias have, and thus could benefit from the same solution. Wiktionary is a project that could increase extremely in value based on a more supported, structured approach.

Besides the sister projects, Wikidata also will cover a number of new functionalities which have arisen from the actual usage of Wikidata and development of the Wikidata community, which include to deal with the huge amount of activity on Wikidata; extended content maintenance functionality like different ranks for statements; or improved support for inline editing in the client wikis.

The continued development will also allow the Wikidata team to fulfill the requirements that have not been fully implemented yet, including further data types, a query answering module, and the usage of query result formatters and their integration in Wikibase clients. Based on the external code review, the team will also improve the code base.

It is planned to continue the Wikidata development for the foreseeable future. Wikidata is on a good trajectory to become an integral part for the Wikimedia projects. But Wikidata aims for much more: the mission of Wikimedia is to provide everyone with access to the sum of all knowledge. A small number of external websites have already started using data from Wikidata, and on Wikidata we are collecting a quickly growing number of external identifiers for items like books, authors, movies, geographical features, etc. We expect external Websites to start using or linking to Wikidata identifiers for their own tags and structured data.

The Wikimedia movement is in the unique position to become a widely accepted, central authority for maintaining and curating identifiers. The goal is for Wikidata to become the core repository of a Web of knowledge, a rich and widely trusted free knowledge base for the whole world.

© RENDER consortium 2010 - 2013

3 Diversity Enricher Drupal Module

This section covers the final improvements on the Drupal module coined “Diversity Enricher”.

In the 3rd year of the project, the experience that has been gathered trough the showcase, the technical work with the project partners and connections to the community in the first two years was used to enhance the Diversity Enricher Drupal module in several specific ways to prepare it for real use on the web. This included the application of roles, e.g., administrators can configure the module while users have no further rights. Furthermore, the interface was polished and unified. In addition, several code cleaning and documentation tasks were performed. We deployed the final version of the extension on http://render-project.eu/drupal and enhanced the showcase with a real Blog export of http://crossingthebaltic.com for which the authors generously granted their permission. Our extension is implemented as a Drupal module that can be obtained from http://drupal.org/sandbox/sti-innsbruck/1991696.

3.1 Updated Description

This section provides details on the third-year changes to the Drupal extension. In particular, it will explain details of the updates of the administration functionality (i.e., the use of roles), the interface and the finalization and publication of the source code at http://drupal.org.

3.1.1 Use of Roles

By default, there are three roles defined in each Drupal deployment: anonymous user, authenticated user, and administrator. After the second year of the project, several parameters like the endpoint, the URL of the Enrycher service, etc. had to be configured in a configuration file. This functionality is now available to administrators through the “Configuration” tab. The click on “Diversity Enricher: Admin” now offers to configure the following parameters:

- Sesame Server URL (text field)
- Sesame Repository Name (text field)
- Use RENDER Ranking Service (radio buttons “enabled” and “disabled”)
- Ranking Service URL (text field)
- Enrycher Service URL (text field)

It is important to note that if the RENDER Diversity-Aware Ranking Service (3rd field) is enabled, the Sesame repository has to be available to the specified ranking service deployment. This is particularly important if the Sesame server is available only on localhost. In this case, a localhost deployment of the ranking service is recommended.

The “import” functionality that was previously available to all (even anonymous) users has also been moved to the administration interface. A strict requirement here is the availability of articles in the SIOC and KDO format like the Enrycher service is producing. Figure 5 depicts the administration options for the Diversity Enricher module.

Anonymous users and authenticated users can still use the interface options of the Diversity Enricher module. The updates on the interface described in the following section.

 © RENDER consortium 2010 – 2013

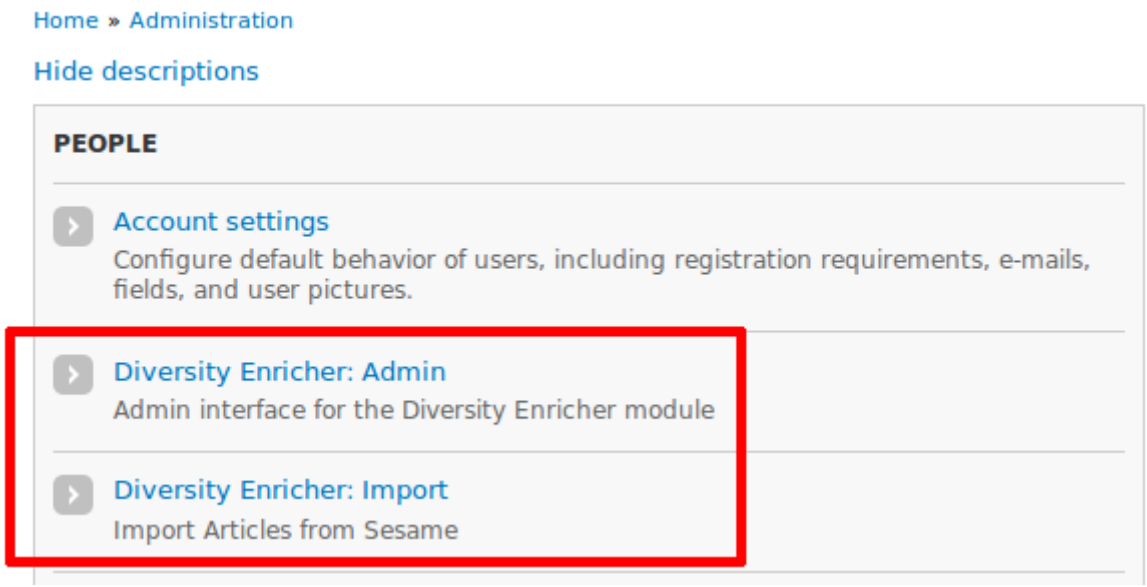


Figure 5: Administration Options for the Diversity Enricher Module

3.1.2 Revision of the Interface

The interface has been restructured in order to separate the functionalities of "diversity" and "export". Named entities and topics that are covered in the article text are highlighted. There exist topics that are extracted but do not occur as such in the article text: these topics are listed as "Further Topics" at the bottom of each article. Moreover, at the right hand side, below the "Related Articles" feature, we introduced a tag cloud that addresses the covered tags automatically attached to the article and weights them by their total occurrence in the dataset. The export functionalities have been extended to support RDF+XML, RDF/JSON and Turtle. Moreover, the sentiment score of each article is now shown at the top of each article.

Figure 6 shows an overview on the new interface of the extension. On the left hand side the original article is presented, whereas on the right hand side the main functionality of the extension is located. There, related articles within the Drupal database are listed, split up according to their extracted overall sentiment. An article is considered to be related if it has at least one topic in common with the currently shown one and is located in the same cluster of the Diversity-Aware Ranking Service (c.f. D3.3.2). The topics of the related articles can be shown by clicking on the + button in front of the article titles. In addition, tags extracted from the currently shown article are presented in a tag cloud below the related article's tree. The size of the respective tag is determined by its number of occurrences in the triple store. Named entities are recognized within the text and get marked. It is possible to click on all tags, named entities, and topics in order to get articles with the same tag/topic (see Figure 7).

© RENDER consortium 2010 - 2013

Figure 6: Screenshot of the Diversity Enricher Drupal Module.

Figure 7: Related Articles with Topic "United States".

3.1.3 Finalization and Publication of the Source Code

The Drupal developer community follows strict coding standards.[6] For the final version of the Diversity Enricher Drupal module, we went through a long process of code cleaning and commenting before even submitting the code for the official code reviews. After that, the tool http://pareview.sh/ automatically checks for syntax requirements (e.g. all functions should be prefixed with the modules name) on each initially submitted tool. Depending on how well the coding standards were applied, this may result in a number of issues that have to be resolved initially. For the Diversity Enricher module, this check already

[6] https://drupal.org/coding-standards

© RENDER consortium 2010 – 2013

resulted in more than 100 issues. After that followed a long process with several code reviewers of the community involved. In the current state, the module has already undergone several review rounds with community members. The source code of the Diversity Enricher module is published at https://drupal.org/sandbox/sti-innsbruck/1991696.

3.2 Showcase

As for a real world showcase, we were searching for a blog that has a high impact and discusses controversial topics. In the course of our search, we came across http://crossingthebaltic.com, a blog maintained by a group of European academics with focus on the Baltic Sea region. The covered topics range from politics and history to arts and culture.[7] We were given permission to extract the articles from the blog and use it for our showcase at http://render-project.eu/drupal. For the extraction and integration, we created a small Drupal module that helps automating this task.

Figure 8 depicts an article from crossingthebaltic.com and Figure 9 shows the diversity-aware version with named entities marked, related articles at the right hand side (split by sentiment) and the automatically generated tag cloud in the bottom right corner. This showcase was used for the evaluation in the following section.

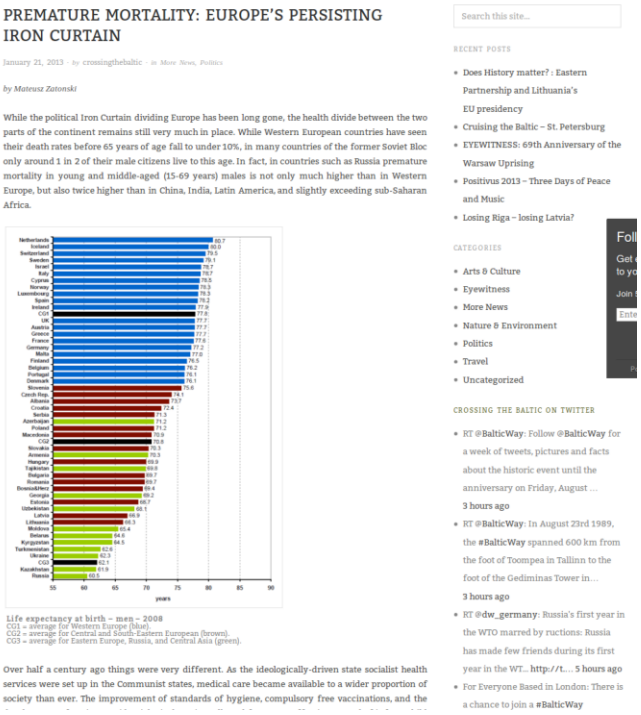


Figure 8: Original Crossingthebaltic.com.

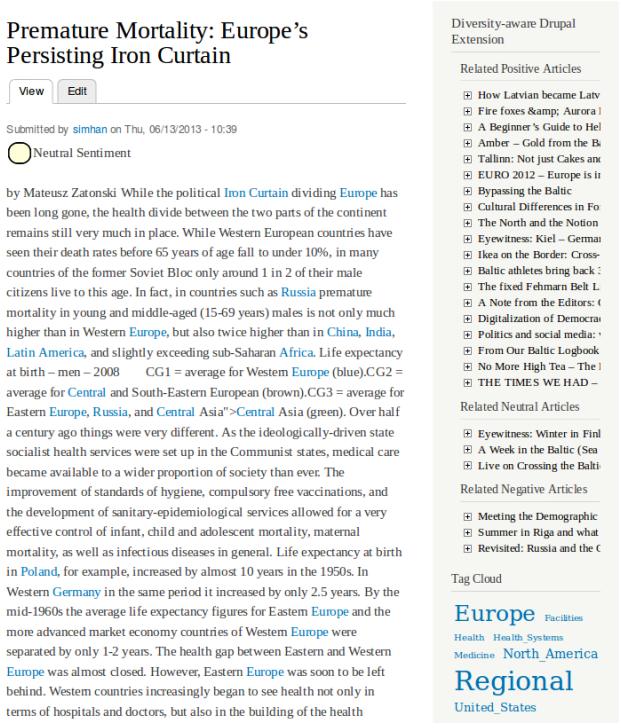


Figure 9: Diversity-Aware Crossingthebaltic.com.

3.3 Evaluation

The evaluation was carried out in the months of July and August 2013. We conducted a survey with 14 participants, mostly with an IT background. The participants were asked to browse the http://render-project.eu/drupal Web site, select an article and answer the following questions (help text in grey):

- Do you find displaying the sentiment at the top of the article useful? *
  positive, neutral, negative
- Choose one of the links in the article text that interests you and click on it. Is the information you get presented useful? *

[7] http://crossingthebaltic.com/about/

You should see a separate window with positive, neutral, and negative articles about the topic you just clicked on.

- Do you find the "Related Positive/Neutral/Negative Articles" fields at the right hand side useful? * Sometimes there are no entries for one or two of the categories.
- Do you find the automatically generated tag cloud at the right hand side useful? *
- Do you find the option to expand each article by its topics useful? *
  [+] markers

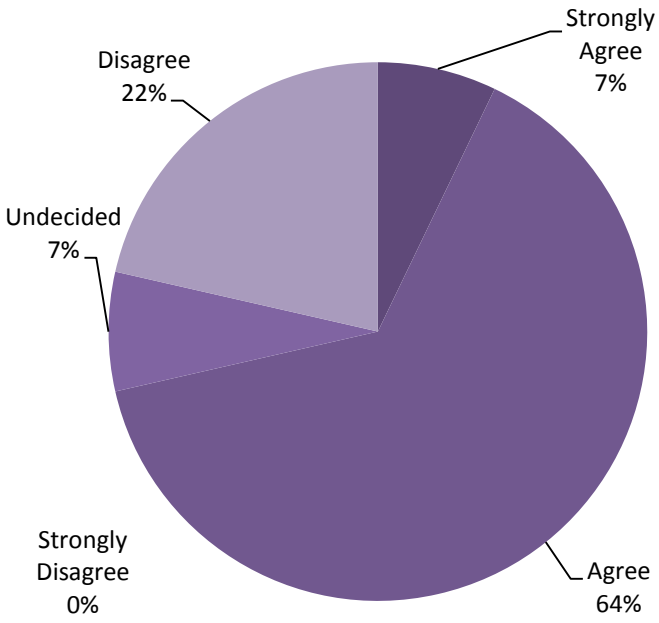
* mandatory

In the remainder of this section, we first present the results of the evaluation, followed by a discussion.

3.3.1 Results

- **Do you find displaying the sentiment at the top of the article useful?**

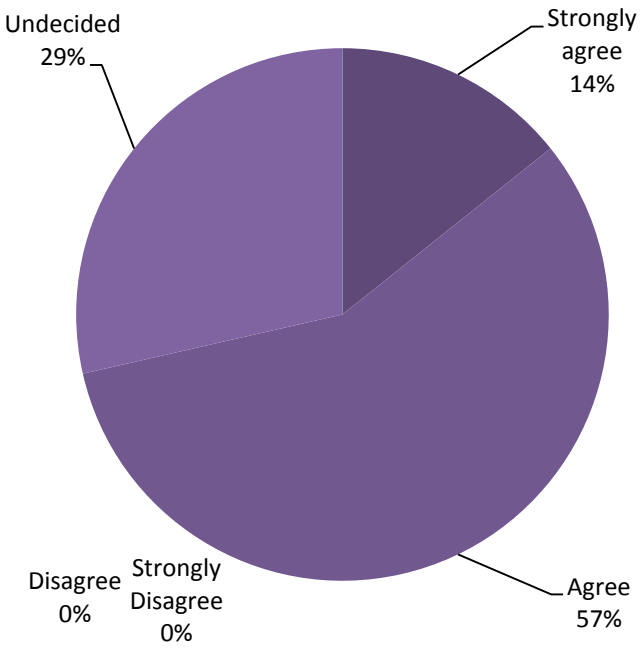
Strongly agree	1
Agree	9
Undecided	1
Disagree	3
Strongly Disagree	0



© RENDER consortium 2010 – 2013

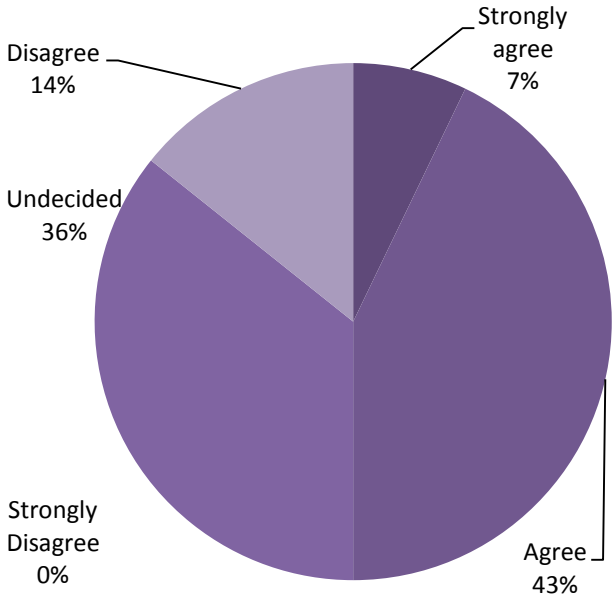
- **Choose one of the links in the article text that interests you and click on it. Is the information you get presented useful?**

Strongly agree	2
Agree	8
Undecided	4
Disagree	0
Strongly Disagree	0



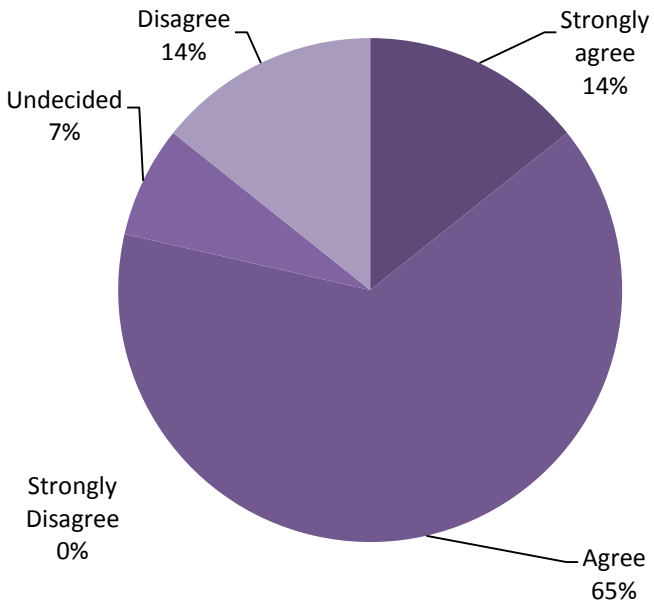
- **Do you find the "Related Positive/Neutral/Negative Articles" fields at the right hand side useful?**

Strongly agree	1
Agree	6
Undecided	5
Disagree	2
Strongly Disagree	0



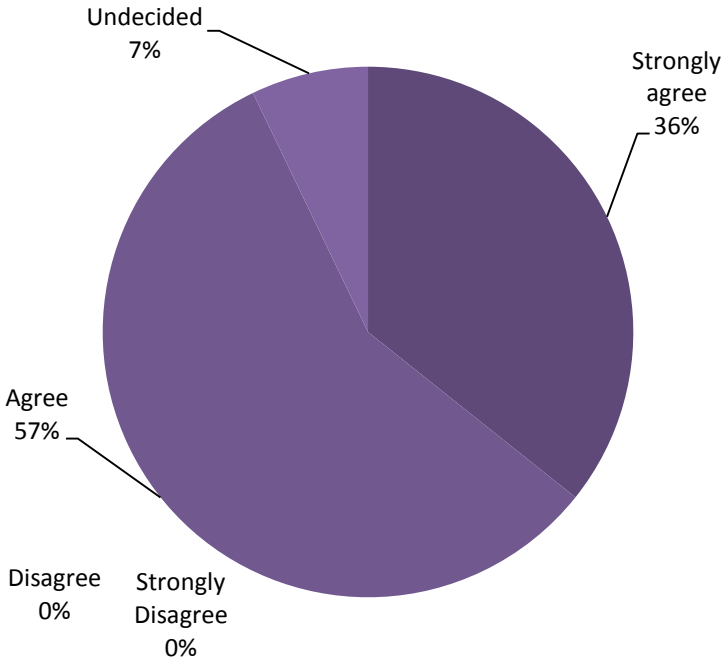
- **Do you find the automatically generated tag cloud at the right hand side useful?**

Strongly agree	2
Agree	9
Undecided	1
Disagree	2
Strongly Disagree	0



- **Do you find the option to expand each article by its topics useful?**

Strongly agree	5
Agree	8
Undecided	1
Disagree	0
Strongly Disagree	0



3.3.2 Discussion

The overall results of the evaluation are positive. The evaluators found the option to expand an article by its topics very useful as well as the option to explore further topics and tags (e.g., see Figure 7), with no disagreement on usefulness. The tag cloud as such is a well-known tool in blogging and was mostly found useful. Displaying the sentiment score of an article as also found useful. Both, the tag cloud and the

 © RENDER consortium 2010 – 2013

sentiment score received mixed feedback with the majority on the positive side. Most mixed feedback received the recommendation of "positive, neutral, and negative articles" with half of the participants either agreeing or strongly agreeing, but with the other half being mostly undecided or even disagreeing.

3.4 Impact

Due to the unexpectedly long period of the code review, we were not able to achieve "full project" status[8] at the drupal.org web site after the second year of the project but effort is going on in order to accomplish this status after the third year. Thus, a more long time study with the module being a fully maintained Drupal project will show insights about the number of downloads and installations. In the code reviews there have been about six developers from the community involved so until the time of writing.

3.5 Further Support

At UIBK's side, we plan to further develop Diversity Enricher in the context of the SESA project[9]. This project focuses on scalable dissemination of content into various channels, including Web sites. As, nowadays, many Web sites are managed with a Content Management System, Drupal is one of the target platforms of this project. As a very interesting step, we plan to support content creators in to see their output analyzed in accordance to diversity aspects such as sentiment and covered topics before publishing the content. As soon as the module achieved full project status we plan to hand the project over to selected members of the Drupal community. This step is planned as a two-phase scenario: in the first phase we plan to get external developers involved and still maintain the project. In the second phase, we plan to reduce our efforts and slowly turn the project into a full community-maintained effort.

[8] https://drupal.org/node/1011698
[9] http://sesa-project.sti2.at/

4 TwiDiViz: Analysis and Visualization of Diversity in Twitter Data

TwiDiViz enables faceted browsing of Tweets. Originally developed at UIBK, in the second year of the project Telefónica and UIBK initiated the integration of TwiDiViz as a visualization component in T-OMT. This section reports about the final amendments.

In the third year of the project, requirements of the end users of the business units at the Telefónica side demanded for several amendments that are detailed in this section. One of the changes was that filtering by language is rather superfluous while filtering by country has a much higher impact. Thus, TwiDiViz doesn't support language filtering any longer but rather filtering by country. For the country codes we use the two-letter notation of ISO 3166-1. Moreover, for easier filter exchange between TwiDiViz and T-OMT we now use the string representation of the topics rather than their URIs for topic identification. Finally, for the showcase, both systems now use the RENDER Tweet repository at http://rendertweets.ontotext.com as a common backend.

4.1 Updated Description

This section provides details on the third-year changes to TwiDiViz. In particular, it will address the changes that have been demanded by the business units at Telefónica in order to establish seamless integration of TwiDiViz and T-OMT.

4.1.1 Filtering by Country

At the Telefónica business units, the demand was raised to support Country rather than Language filtering. In order to comply with this request, the exchange format had to be changed to the following JSON notion:

```
{
“source” : “...”,
“countries” : [“...”,”...”,...],
 “topics” : [“...”,”...”,...],
“from_date” : “...”,
“to_date” : ”...”,
“query_date” : “...”
}
```

Listing 1: T-OMT Filter Exchange Format

In order to support country filtering properly, we use the two-letter notation of ISO 3166-1 for countries and other regional specifications (e.g. EU for Europe). The language filtering option was dropped.

4.1.2 Strings for Topic Filtering

In the previous version of TwiDiViz, the JSON format for T-OMT integration considered URIs rather than the string representation of specific topics. As the demand was raised by Telefónica, this has been changed to a pure string representation (i.e., the rdfs:label).

4.2 Showcase

This section covers the update of the showcase of the TwiDiViz faceted browsing tool. In the new showcase and therefore in the newly provided screenshots several of the previously discussed changes (cf. Section

 © RENDER consortium 2010 – 2013

4.1) are also visible. Before explaining Figure 10, we want to emphasize on the most significant change the showcase has undergone: at the beginning of year three, the Enrycher processed Twitter data was made available in an OWLIM repository and therefore, we were able to hook the showcase on to a real dataset.[10] Figure 10 exemplifies the position of the functions:

1. **Filter summary** – all selected filter options are shown. Note the difference of "Topics" compared to Figure 11 where URIs are used to describe topics.
2. **Document type** – Different document types can be selected. In this case, only Tweets are available.
3. **Filters** – Further filters include Country (cf. Section 4.1), Topic, Tag, and Sentiment
4. **Graph vs. Grid view and sorting** – Here, either graph or grid view may be selected. The dataset is sorted in accordance to the selected attribute (in this case Sentiment)
5. **Vertical division** – In case graph view is selected the dataset is divided by the values of the attribute in whose accordance is sorted (in this case "negative" and "positive").

Figure 11 depicts the previous version of TwiDiViz without Twitter data.

[10] The data is available at http://rendertweets.ontotext.com

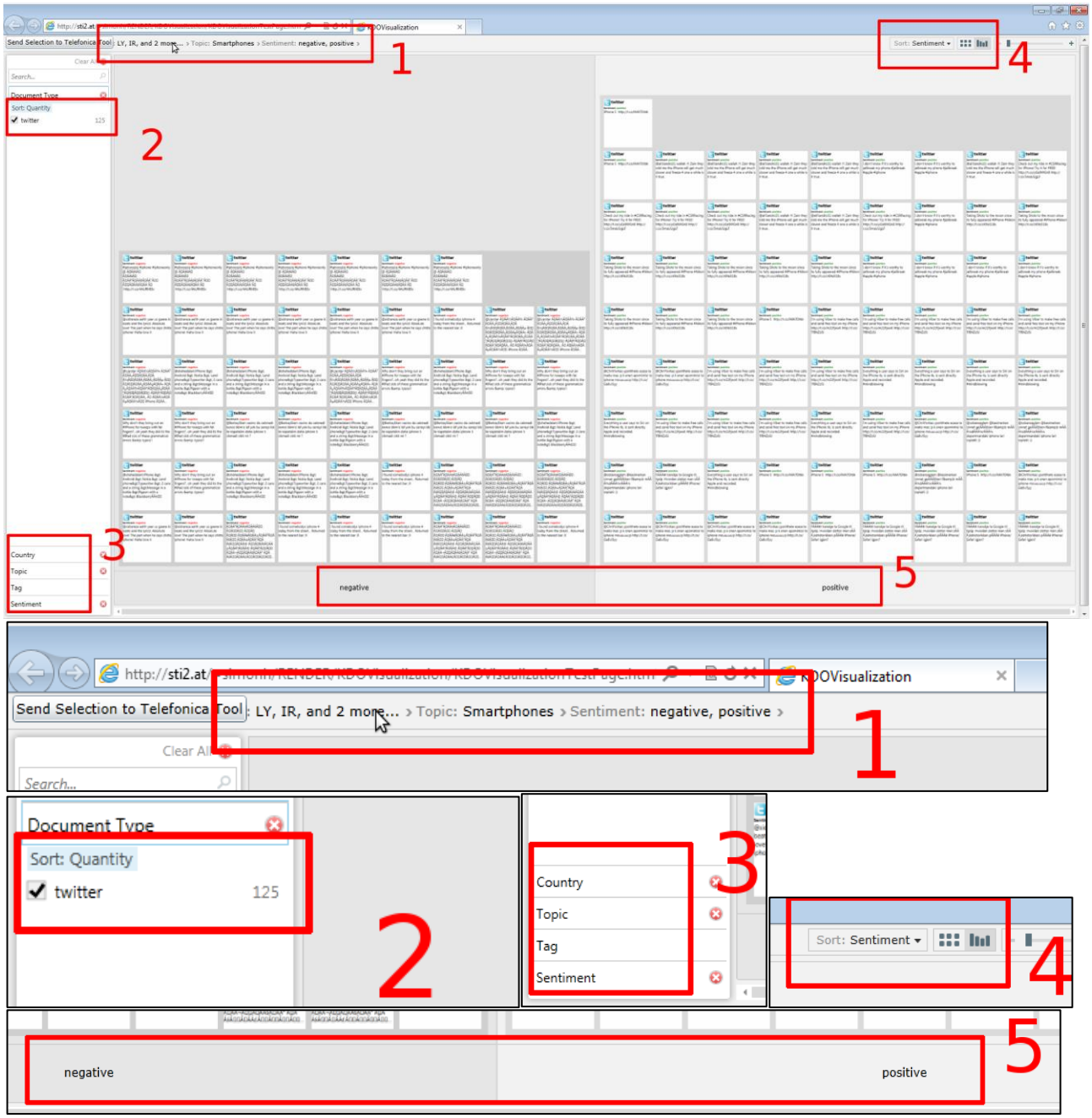


Figure 10: Current Showcase with Tweets from the RENDER Twitter Repository with Overview (above) and Clippings of Individual Functions 1-5 (below)

© RENDER consortium 2010 – 2013

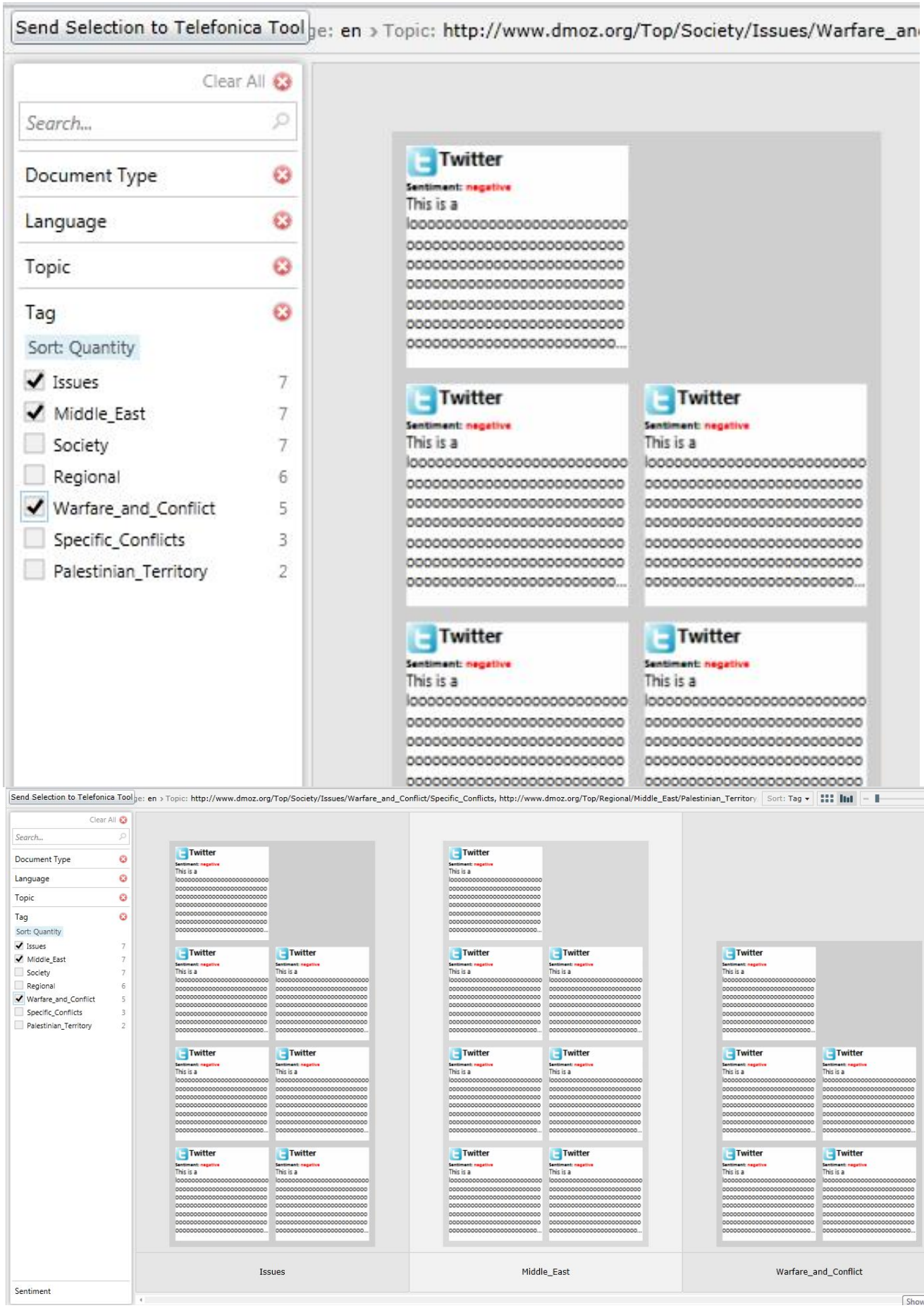


Figure 11: Previous Showcase (without Twitter Data). Above: Clipping, below: Whole Tool View.

4.3 Further Support

As for the UIBK part, the source code of the TwiDiViz tool is available under the MIT license and therefore open for commercial reuse. This licensing option is intended to make further development at the Telefónica side as easy as possible. The source code is publically available at https://github.com/athalhammer/TwiDiViz. As TwiDiViz is based on a version of MS PivotViewer that simply communicates with the T-OMT via an API, it's functionality can be used in a simple fashion by other project and tools that aim to visualize their diverse Twitter (or similar) data in a versatile manner.

© RENDER consortium 2010 – 2013

5 Conclusions

In this deliverable, we reported on the final status on three extensions to collaborative systems, namely the Wikimedia projects, Drupal and the T-OMT by Telefónica via an Open Source modification of the MS Pivot Viewer available for reuse. We describe the Wikidata project after its completed one-year development. While Wikidata was split out in an early stage from RENDER as a standalone project, RENDER was an important incubator and supported its development during the initial phases. As reported, the project went to meet all expectations and has already surpassed Wikipedia in editor activity. It allows providing diverging and differently sourced data for the same facts, and as such makes the diversity in Wikipedia and its sister projects explicit. It connects the different language version via a unified semantic repository and thus ensures easier, complete and verified connections between diverse language versions of an article.

We did not cover the extensions to Wikipedia itself in this deliverable – as developed by Wikimedia in RENDER – as they are described in detail in the deliverables of WP5.

Extending the widely used content-management and publishing platform Drupal, we presented the "Diversity Enricher" module developed by RENDER. During the project, we constantly developed this extension further, learning in the process from the interaction with the community and iterating persistently over the software. No disruptive changes were implemented in the third year. Instead, a great quantity of fine-tuning was carried out to adapt the module to the conditions of a realistic usage scenario. This strategy of small steps while listening to the Drupal crowd proved to be very valuable, as the results of the evaluation verify as well as the reaction of the community. The evaluation was reported in this deliverable as no dedicated deliverable was foreseen for this content.

We finalized TwiDiViz, our MS Pivot Viewer extension, which we adapted to be a visualization tool to the Telefónica T-OMT service. This decision was taken during the second project year and served us very well as an appropriate community for this single Silverlight-based tool was not in sight and would have been difficult to construct. We got valuable feedback during the Telefónica use case evaluation, which was incorporated in the third year iterations. During this last project year the integration of the tool with the T-OMT was finalized while still keeping its a adaptability to potential future use cases and data sources via the defined data interface. Therefore TwiDiViz can be considered an extension to any platform that wants to visualize and explore diverse twitter data. We finalized a use case with data to be incorporated in the Telefónica evaluations.

In summary, it can be said that although the initial planning of certain extensions to platforms at the project start in WP4 were transformed or dropped, we adapted quite successfully and shifted our efforts and expertise in the project in a very goal-oriented manner. This turned out to be good, adaptive strategic planning: by push-starting Wikidata, by switching one tool use case from Wordpress to the larger and more developer-friendly (and –active) Drupal platform and by integrating the Twitter visualization tool with the Telefónica business use case, we achieved a higher impact for the available person months than could have been created by strictly adhering to the original project outline for WP4. Additionally, some tools for Wikipedia were initially developed in this work package. All in all it can be said that the technologies developed and deployed in WP4 have achieved most of the goals set at the beginning of RENDER; they made many platforms more diversity-friendly

© RENDER consortium 2010 - 2013