RENDER

Deliverable D6.3.1

Report on Clustering and Liaison Activities – Year 1

Editor:	Delia Rusu, JSI
Author(s):	Delia Rusu, JSI
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)	Public (PU)
Contractual Delivery Date:	September 2011
Actual Delivery Date:	September 2011
Suggested Readers:	Everybody interested in the clustering and liaison activities carried out during the first project year
Version:	1.0
Keywords:	Collaboration activities, research projects

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:	WP6 Dissemination, exploitation, and community building
Document Title:	D6.3.1 - Report on Clustering and Liaison Activities Year 1
Editor (Name, Affiliation)	Delia Rusu, JSI
Work package Leader (Name, affiliation)	Ioan Toma, UIBK

Copyright notice

Executive Summary

During the first year of the RENDER project, we aimed at establishing contacts to several networks of excellence as well as to other projects which share topics of interest with RENDER.

The envisaged networks of excellence are **Pascal**, **MetaNet** and **PlanetData**. Pascal is a European Union funded network which aims at supporting collaboration between experts in Machine Learning. MetaNet's goal is building the technological foundations of a multilingual European information society. PlanetData is focused on fostering a common ground for sharing different approaches to large-scale data provisioning and management.

The collaboration with the Pascal network of excellence is important mainly from the point of view of Work Package 2 – Diversity Mining, which heavily employs machine learning techniques for identifying and analyzing diversified information. Secondly, the collaboration with MetaNet enables the language data (written corpora) and language processing tools developed within RENDER to be made available to a broader community. Thirdly, PlanetData offers support for large-scale data provisioning and management, which is an important topic for RENDER.

In the collaboration with other projects, three paths have been pursued. We jointly organized a workshop co-located with the World Wide Web Conference together with the **LivingKnowledge** project. In the case of the **ALERT**, **CoSyne** and **Wikidata** projects we started defining common research tasks. Finally, RENDER was invited to join a virtual centre of excellence established by the ROBUST project.

LivingKnowledge aims at moving forward the emerging area of diversity-, bias- and evolution-aware retrieval of information. ALERT is an Open Source project whose goal is to improve the overall bug resolution process in Open Source developers' collaborative environments. The aim of the CoSyne project is to automate the dynamic multilingual synchronization process of Wikis, which support collaborative content creation. ROBUST is dedicated to the management of risks and opportunities in large scale online business communities. Wikidata is a planned instance of a diversified Semantic MediaWiki (which is an extension of Wikipedia's wiki engine with Semantic Web features).

In the next year of the project, we are going to clearly define the joint small research tasks supported by the networks of excellence, continue the collaboration established thus far with the aforementioned projects, and foster collaborations with other initiatives, be they coming from other research projects or companies.

Table of Contents

Abbreviations

NoE   Network of Excellence

FET   Future and Emerging Technologies

1   Introduction

During the first year of the RENDER project, we aimed at establishing contacts to several networks of excellence as well as to other projects which share topics of interest with RENDER.

The envisaged networks of excellence are **Pascal**, **MetaNet** and **PlanetData**. Pascal is a European Union funded network which aims at supporting collaboration between experts in Machine Learning, Statistics and Optimization, and at the same time sustains the usage of machine learning in various application domains. MetaNet aims at building the technological foundations of a multilingual European information society, via a dynamic and influential community, offering at the same time an open distributed facility for sharing and exchanging resources and collaborating with neighbouring technologies. PlanetData is focused on fostering a common ground for sharing multiple approaches to large-scale data management from different disciplines in order to create holistic solutions, while offering solutions to support large-scale data provisioning and management.

The collaboration with the Pascal NoE is important mainly from the point of view of Work Package 2 – Diversity Mining, which heavily employs machine learning techniques for identifying and analyzing diversified information. Secondly, the collaboration with MetaNet enables the language data (written corpora) and language processing tools developed within RENDER to be made available to a broader community. Thirdly, PlanetData offers support for large-scale data provisioning and management, which is an important topic for RENDER.

As far as the collaboration with other projects is concerned, we contacted the **LivingKnowledge**, **ALERT CoSyne and ROBUST** projects, as well as the **Wikidata** initiative. LivingKnowledge aims at moving forward the emerging area of diversity-, bias- and evolution-aware retrieval of information. ALERT is an Open Source project whose goal is to improve the overall bug resolution process in Open Source developers' collaborative environments. The aim of the CoSyne project is to automate the dynamic multilingual synchronization process of Wikis, which support collaborative content creation. ROBUST is dedicated to the management of risks and opportunities in large scale online business communities. Wikidata is a planned instance of a diversified Semantic MediaWiki, which is an extension of Wikipedia's wiki engine with Semantic Web features.

Three paths have been pursued: either jointly organizing an event together with the LivingKnowledge project– more precisely a workshop co-located with the World Wide Web Conference, defining common research tasks, as was the case of ALERT, CoSyne and Wikidata projects or joining a virtual centre of excellence established by the ROBUST project.

The main section of this deliverable provides details regarding the abovementioned NoEs and projects RENDER is collaborating with.

2   Collaboration Activities

In this section we are going to describe the collaboration activities carried out during the first year of the project. Firstly, we'll describe the relevant networks of excellence that we established contacts with, followed by the collaboration with other projects.

2.1   Networks of Excellence

JSI established contacts with the following networks of excellence (NoEs): Pascal, MetaNet and PlanetData. The collaboration will take the form of partly shared technology development and joint small research tasks, supported by NoEs.

Pascal [1] is a European Union funded NoE. It has established a distributed institute that brings together researchers and students across Europe, and is now looking to expand to other countries all over the world. The network supports collaboration between experts in Machine Learning, Statistics and Optimization, while promoting the use of machine learning in many relevant application domains such as: Machine Vision, Speech, *Natural Language Processing*, *Information Retrieval*, *Textual Information Access*.

We chose the Pascal NoE as the last three application domains are highly relevant for the RENDER project, more exactly the diversity mining related tasks of fact extraction and opinion mining.

MetaNet [2] is another NoE comprising 47 research centres from 31 countries, and having as purpose to build the technological foundations of a multilingual European information society. The three goals supported by META-NET are in the lines of:

- META-VISION, by fostering a dynamic and influential community around a shared vision and strategic research agenda,

- META-SHARE, by creating an open distributed facility for the sharing and exchange of resources,

- META-RESEARCH, by building bridges to relevant neighbouring technology fields.

Language data (written corpora) and language processing tools developed within RENDER can thus be available to a broader community.

PlanetData [3] is the third NoE that we contacted. The network defines three main objectives:

- *Research*: bringing together approaches to large-scale data management from different disciplines in order to create holistic solutions,

- *Data provisioning and management*: To provide software to support large-scale data provisioning, made available via the PlanetData Lab, supporting various formats, for researchers to test and validate their techniques. To create definitive vocabularies for the description of data sets, to build a catalogue of data, to publicize guidelines and best practices for provisioning,

- *Impact*: To provide a medium through which the research results and empirical findings of the PlanetData network can be used to improve the education level related to large-scale data management in both academia and industry.

The fact that RENDER deals with large –scale data motivates the collaboration with PlanetData. The solutions to large-scale data management developed within RENDER can thus be disseminated, while, in the same time, the RENDER project can benefit from the data provisioning and management approaches provided by PlanetData.

2.2   The ALERT Project

ALERT [4] is an Open Source project with a focus on the Open Source developers' collaborative environments, where it aims to improve the overall bug resolution process. The project will develop methods and tools that improve **free and open-source software** coordination through real-time, personalized, context-aware notification – therefore maintaining awareness of community activities.

To achieve this, **ALERT** will create an active collaboration platform. In this platform, there would be a virtual actor with the following tasks which would enable developers to work better together:

- to interact with other developers, process and recognize various kinds of interactions,

- to suggest actions on the basis of the interactions

- to remember and bring past interactions into the developers' attention

The collaboration with the ALERT project will be in the form of applying technology developed by RENDER for a specific scenario defined by ALERT. More precisely, within RENDER JSI will develop an automatic text annotation tool for linking text to external Linked Open Data resources, thus providing an additional structured layer to the textual layer. For now the tool links text to WordNet resources [10], but in a future deliverable we are going to release a version of the tool for linking to OpenCyc resources as well. We are going to test the feasibility of using the above mentioned tool by the ALERT project, to link bug resolution specific terms to OpenCyc.

2.3   The LivingKnowledge Project

LivingKnowledge [5] is an FET project aimed at moving forward the emerging and increasingly important area of diversity-, bias- and evolution-aware retrieval of information. The main outcomes of the project are twofold. On the theoretical side, the project will provide radically new insights for dealing with diversity, bias and evolution in information access and search, with innovative methods and algorithms for exploiting available diversity on the Web in all main dimensions. On the technology side, LivingKnowledge will provide a testbed incorporating all infrastructures necessary to manage, retrieve and exploit facts and opinions, traceable along several dimensions: content, time and bias. Thus, the goal of the project is to improve search and knowledge management technology in order to obtain more concise, complete and contextualised search results.

Together with the LivingKnowledge project, RENDER organized the First International Workshop on Knowledge Diversity on the Web – DiversiWeb 2011, a workshop collocated with the World Wide Web 2011 Conference held in Hyderabad, India. The workshop is described as a dissemination activity in [11].

2.4   The CoSyne Project

CoSyne [6] is a Research and Technological Development project co-funded by the European Union. The overall aim of the CoSyne project is to automate the dynamic multilingual synchronization process of Wikis, which contain a combination of dynamic user-generated content combined with multilingual aspects.

The outcome of the CoSyne project is:

- robust translation of noisier user-generated content between 6 core languages – 4 major languages and 2 languages with limited resources,

- to improve machine translation quality by segment-specific adaptive modelling,

- to identify textual content overlap between segments of Wiki pages across languages to avoid redundant machine translation,

- to identify the optimal insertion points for translated content to preserve coherence,

- to analyse user edits in order to distinguish between factual content changes and corrections of machine translation output; the latter idea will be exploited to improve machine translation performance using a self-learning approach.

Wikimedia contacted the CoSyne Project via Wikimedia Netherlands, one of the CoSyne use case partners. One of the CoSyne lines of research is multilingual content synchronization and the collaboration with RENDER will mainly focus on fact coverage from a multilingual perspective.

2.5   The ROBUST Project

The EU project **ROBUST** [7] is dedicated to the management of risks and opportunities in large scale online business communities. This involves the analysis and discovery of communities as well as the reaction to situations and developments in online communities that might pose a danger to the health of a community or a chance for prospering.

RENDER will join a virtual centre of excellence in the field of Online Communities which is currently being established in the context of the ROBUST project. The aim of this centre is to bring together leading experts and projects in the field of research on Online Communities and to exchange experiences, latest results as well as to find common grounds for joined work or future projects.

As first step we will join the kick-off of the centre of excellence in a concerted event in Paris mid-October 2011 and will consider next steps for cooperation.

2.6   The Wikidata Proposal

One of Wikimedia's project proposals, which they host and maintain is Wikidata [8]. The project contains a three-phased approach, as follows:

- **Interwiki links** consists of creating an entity base for the Wikimedia projects

- **Infoboxes** gathers first infobox-related data for a subset of the entities, having as goal to augment infoboxes with data from data.wikimeda.org

- **Lists** expands the set of properties beyond those related to infoboxes, and provides ways of exploiting this data within and outside of the Wikimedia projects

Wikidata is a planned instance of a diversified Semantic MediaWiki, which is an extension of Wikipedia's wiki engine with Semantic Web features [9]. Semantic MediaWiki today cannot represent diverse data -- i.e. it cannot display different, conflicting values for one property. The Wikidata extension to Semantic MediaWiki allows adding metadata to facts -- most importantly the source -- so that it can be further refined or selected. Shortipedia provides a prototype implementation of this functionality, but the connection of sources to facts is currently merely mocked-up in the prototype. The Wikidata extensions will implement the functionalities demonstrated in Shortipedia on top of Semantic MediaWiki, thus creating a truly diversifiable Semantic MediaWiki.

3   Future Work

In the next year of the project, we are going to clearly define the joint small research tasks supported by the networks of excellence, as well as give account for the first outcomes of these tasks.

We are going to continue the collaboration established thus far with the aforementioned projects, either in the form of jointly organized events (as was the case with the LivingKnowledge project), by defining common research tasks (e.g. ALERT, CoSyne, Wikidata) or by joining a virtual centre of excellence established by the ROBUST project.

Furthermore, we are going to foster collaborations with other initiatives, be they coming from other research projects or companies.

References

[1]   http://www.pascal-network.org/

[2]   http://www.meta-net.eu/

[3]   http://www.planet-data.eu/

[4]   http://www.alert-project.eu/

[5]   http://livingknowledge-project.eu/

[6]   http://www.cosyne.eu

[7]   http://robust-project.eu/

[8]   http://meta.wikimedia.org/wiki/New_Wikidata

[9]   http://semantic-mediawiki.org/wiki/Semantic_MediaWiki

[10]   B. Fortuna, D. Rusu, M. Trampus, L. Dali, T. Stajner, M. Grobelnik. Deliverable 2.2.1 - Prototype of the Fact Mining Toolkit. 2011.

[11]   I. Toma. Deliverable 6.1.3 - Report on dissemination activities Y1. 2011.