FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)
Large-scale, Cross-lingual Trend Mining and Summarisation of Real-time Media Streams

D8.1.2 Dissemination and Exploitation Report –v2

France Lasfargues, IM – Thierry Decleck, DFKI,
With contribution by all partners

Abstract
FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)
Deliverable D8.1.2 – v2 (WP 8)

This deliverable outlines the results of dissemination and exploitation activities of the TrendMiner project, co-funded by the European Commission, within the 7th Framework. This final deliverable is based on the non-public deliverable D8.1.1 Dissemination and exploitation Plan – v1, which has been released in August 2012. This initial dissemination plan introduced the methods and tools used for the dissemination and exploitation of TrendMiner activities. This final deliverable highlights the scientific and societal impact achieved by the project.

Keyword list: Dissemination, Exploitation

Nature:Report
Dissemination: PU
Contractual date of delivery: M36
Actual date of delivery: 31 Oct 2014
Reviewed By: All
Web links: http://www.trendminer-project.eu/index.php/publications;
http://www.trendminer-project.eu/index.php/downloads

TrendMiner Consortium

This document is part of the TrendMiner research project (No. 287863), partially funded by the FP7-ICT Programme.

DFKI GmbH
Language Technology Lab
Stuhlsatzenhausweg 3
D-66123 Saarbrücken,Germany
Contact person: Thierry Declerck
E-mail: declerck@dfki.de

University of Southampton
Southampton SO17 1BJ, UK
Contact person: Mahensan Niranjan
E-mail: mn@ecs.soton.ac.uk

Internet Memory Research
45 ter rue de la Rvolution
F-93100 Montreuil, France
Contact person: France Lafarges
E-mail: contact@internetmemory.org

Eurokleis S.R.L.
Via Giorgio Baglivi, 3
Roma RM 0016, Italia
Contact person: Francesco Bellini
E-mail: info@eurokleis.com

University of Sheffield
Department of Computer Science
Regent Court, 211 Portobello St.
Sheffield S1 4DP, UK
Contact person: Kalina Bontcheva
E-mail: K.Bontcheva@dcs.shef.ac.uk

Ontotext AD
Polygraphia Office Center fl.4,
47A Tsarigradsko Shosse,
Sofia 1504, Bulgaria
Contact person: Atanas Kiryakov
E-mail: naso@sirma.bg

Sora Ogris and Hofinger GmbH
Bennogasse 8/2/16
A-1080 Wien, Austria
Contact person: Christoph Hofinger
E-mail: ch@sora.at

Hardik Fintrade Pvt Ltd.
227, Shree Ram Cloth Market,
Opposite Manilal Mansion,
Revdi Bazar, Ahmedabad 380002, India
Contact person: Suresh Aswani
E-mail: m.aswani@hardikgroup.com

DAEDALUS - DATA, DECISIONS AND LANGUAGE, S. A.
C/ López de Hoyos 15, 3º, 28006 Madrid, Spain
Contact person: José Luis Martínez Fernández
Email: jmartinez@daedalus.es

Institute of Computer Science Polish Academy of Sciences
5 Jana Kazimierza Str.,Warsaw, Poland
Contact person: Maciej Ogrodniczuk
E-mail: Maciej.Ogrodniczuk@ipipan.waw.pl

Universidad Carlos III de Madrid
Av. Universidad, 30, 28911, Madrid, Spain
Contact person: Paloma Martínez Fernández
E-Mail: pmf@inf.uc3m.es

Research Institute for Linguistics of the Hungarian Academy of Sciences
Benczúr u. 33., H-1068 Budapest, Hungary
Contact person: Tamás Váradi
Email: varadi.tamas@nytud.mta.hu

Executive Summary

This deliverable outlines the results of dissemination and exploitation activities of the TrendMiner project, co-funded by the European Commission, within the 7th Framework. This final deliverable is based on the non-public deliverable D8.1.1 Dissemination and exploitation Plan – v1, which has been released in August 2012. This initial dissemination plan introduced the methods and tools used for the dissemination and exploitation of TrendMiner activities. This final deliverable highlights the scientific and societal impact achieved by the project, and presents in a separate section the individual exploitation plans of the partners of the consortium.

Contents

Executive Summary **3**
Contents **4**
1 Introduction **5**
2 Dissemination **5**
2.1 TrendMiner: Dissemination Results 5
2.2 List of Public deliverables (for the whole project period) 7
2.3 List of Scientific Publications (year 3) 8
2.4 Presentations (year 3) 9
2.5 Fairs (year 3) 10
2.6 Open Source Software Releases (Year 1 -2 -3) 10
2.7 Online demos 12
2.8 Cooperations 12
2.9 Online presence 13
2.9.1 Website Final results 13
2.9.2 Twitter account 13
3 Exploitation **14**
3.1 DFKI 14
3.2 USFD 15
3.3 Ontotex 15
3.4 University of Southampton 16
3.5 Internet Memory Research 17
..... 19
..... 19
3.6 Eurokleis 20
3.7 Sora Ogris & Hofinger 21
3.8 Hardik Fintrade Pvt Ltd 23
3.9 Daedalus – UC3M Exploitation for the Health Use Case 23
3.10 Daedalus exploitation strategy 29
3.11 RILMTA - Department of Corpus Linguistic of the Hungarian Academy of sciences 29
3.12 IPIPAN Institut of Computer Science Polish Academy of Sciences 30

1 Introduction

This deliverable outlines the results of dissemination and exploitation activities of the Trendminer project, co-funded by the European Commission, within the 7th Framework
Dissemination and exploitation were continuous tasks in the project. This deliverable is therefore not only a report on the last year of the project, but it also describes activities that took place during the whole lifetime of the project.
Section 2 concentrates on the dissemination activities and section 3 presents the planned exploitation by the partners of the results of the project.

2 Dissemination

2.1 TrendMiner: Dissemination Results

In order to promote TrendMiner and its results, the consortium undertook dissemination activities addressing various communities, since the project dealt with various technologies, such as:

- Computational Linguistics / Natural Language processing
- Information Retrieval
- Machine Learning
- Semantic Web
- Knowledge Management
- Data/Text Mining
- …

In total, TrendMiner has been presented at 82 conferences and fair trade events. The active collaboration with other projects allowed increasing impact and visibility of TrendMiner.

	Year 1	Year 2	Year 3	Total
Publications	11	12	15	38
Presentations	9	15	8	32
Events	6	2	4	12
Collaborations	5	9	11	25
Open source code	0	3	9	12
Book / Press release	1	2	2	5

Table 1: Dissemination Results

The consortium also submitted papers to journals and conferences that will be published (if accepted) in the year 2015.

Furthermore, the consortium organised satellite events at high-level conferences, and stand-alone events, such as:

- TrendMiner sponsored the 1st Workshop on ***P***ractice ***a***nd ***Th***eory of ***O***pinion Mining and ***S***entiment Analysis at Konvens 2012
- Diana Maynard (USFD) gave an opinion mining tutorial at LREC 2012, jointly organised by TrendMiner and Arcomem projects
- Course on "Mining social media with GATE", co-organised and co-delivered by TrendMiner staff at Sheffield. Part of the Annual GATE Summer school. June 18-22nd 2012.
- Kalina Bontcheva and Dominic Rout (USFD) delivered a 1 day course on NLP for social media, delivered at the GATE summer school, June 7th 2013.
- Kalina Bontcheva (USFD) teaches a 4 day course on NLP for social media to PhD students in Szeged, Hungary (December 9 - 12, 2013)
- USFD delivered a tutorial at EACL 2014 on Natural Language Processing for Social Media
- USFD delivered a tutorial at ISWC 2014 on Semantic annotation of social media
- USFD delivered a week-long training course on NLP for social media, June 9 - 13, 2014
- DFKI co-organizer of the ES3LOD Workshop, (LREC 2014, Reykjavik, May 2014.
- TrendMiner had a booth at the HLT Village at both LREC 2012 and LREC 2014

TrendMiner was also presented at non-academic events (see 2.4)

TrendMiner has also been trying to be as open as possible regarding the publication of results. Almost three-quarter of the deliverables (28 out of 37) have been be made publicly available and they have been all published on TrendMiner website (http://www.trendminer-project.eu/index.php/publications/85-public-deliverable).

2.2 List of Public deliverables (for the whole project period)

Deliverable	Title	Lead	Date
D2.1.1	Knowledge & Provenance Modeling and Stream Reasoning - v1	SOTON	6
D2.2.1	Multilingual, ontology-based IE from stream Media-v1	DFKI	12
D3.1.1	Regression models of trends in streaming data - v1	SOTON	6
D4.1.1	Multilingual summarisation of stream media software	USFD	12
D5.1.1	Real-time Stream Media Collection - V1	IM	12
D5.1.2	Architecture for distributed text annotation and trend mining over media stream - v1	OTON	12
D8.2.1	Market Watch - v1	SORA	12
D2.2.2	Multilingual, ontology-based IE from stream Media - v2	DFKI	24
D2.3.1	Multilingual resources and evaluation of knowledge modeling - V1	DFKI	18
D3.1.2	Regression models of trends - v2	SOTON	18
D3.2.1	Clustering models for discovery of regional and demographic variation - v1	USFD	24
D4.1.2	Multi-lingual sumarisation of stream media software- v2	USFD	24
D5.2.1	Real- Time Stream Media Collection - V2	IM	24
D5.3.1	Architecture for distributed text annotation - v2, Real-Stream Media Processing Platform - v1	ONTO	24
D2.3.1	Multilingual resources and evaluation of knowledge modeling - V1	DFKI	14/30
D2.3.2	Multilingual resources and evaluation of knowledge modeling - V2	DFK	30
D3.3.1	Tools for Mining Non stationary Data - v2 Clustering models regional and demographic variation - v2	SOTON	30
D4.2.1	Multi- paradigm search software - v2 Croww lingual web UI for trends / sentiments in streaming media	ONTO	34
D5.3.2	Real-Time Stream Media Processing Platfrom and Cloud Based Deployment - V2	ONTO	36
D5.4	Experimental results on sequential computation	SOTON	36
D6.3	Application final results	EK	36
D7.3	Application final results	SORA	36
D8.1.2	Dissemination and exploitation Report - V2	IMR	36
D8.2.2	Market Watch v2	EK	30
D2.4	Integration of lexical and terminological data	IPIPAN	36
D5.5	Deployment of web services for new use cases	DAEDALUS	36
D9.1	Integration of annotation generated by annotation tools	RILMTA	36
D10.1	Harmonization of the new language data and tools	UC3M	36

2.3 List of Scientific Publications (year 3)

Institution	Conférence	Article
USFD	ACL 2014	Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly - Vasileios Lampos, Daniel Preotiuc-Pietro, Sina Samangooei, Douwe Gelling, Trevor Cohn. ACL 2014, Workshop on Language Technologies and Computational Social Sc
UC3M	EACL 2014	Detecting drugs and adverse events from Spanish social media streams. Isabel Segura-Bedmar, Ricardo Revert, Paloma Martínez. 5th International Workshop on Health Document Text Mining and Information Analysis (Louhi), EACL 2014, Gothenburg, April 27.
UC3M	ACL 2014	Extracting drug indications and adverse drug reactions from Spanish health social media. Isabel Segura-Bedmar, Santiago de la Peña González and Paloma Martínez ACL 2014, Workshop BioNLP 2014. Baltimore, MD, June 26-27, 2014
UC3M	SEPLN 2014	ADRSpanishTool: a tool for extracting adverse drug reactions and indications. Santiago de la Peña, Isabel Segura-Bedmar, Paloma Martínez, José Luis Martínez.. Demo at SEPLN 2014, 17-19 September 2014, Girona (Spain), http://taln.upf.edu/pages/sepln2014/es/
UC3M	SEPLN 2014	TrendMiner: Large-scale Cross-lingual Trend Mining Summarization of Realtime Media Streams Paloma Martínez, Isabel Segura, Thierry Declerck, José L. Martínez. SEPLN 2014, 17-19 September 2014, Girona (Spain)
DFKI	CCL	Thierry Declerck. Harmonizing Lexical Data for their Linking to Knowledge Objects in the Linked Data Framework. Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing, Dublin, Ireland, The COLING 2014 Organizing Committe, International Committee on Computational Linguistics (ICCL), Sheffield, Dublin, 8/2014 (PDF)
DFKI	CCL	Guy Emerson, Thierry Declerck. SentiMerge: Combining Sentiment Lexicons in a Bayesian Framework. Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing, Dublin, Ireland, The COLING 2014 Organizing Committe, Dublin, 8/2014 (PDF)
DFKI	LREC-2014	Thierry Declerck, Hans-Ulrich Krieger. Harmonization of German Lexical Resources for Opinion Mining. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Ireland, ELRA, Paris, 5/2014 (PDF)
DFKI	LREC-2014	Hans-Ulrich Krieger, Thierry Declerck, TMO - The Federated Ontology of the TrendMiner Project. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014), European Language Resources Association, 2014 (PDF)
USFD		L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak, K. Bontcheva. Analysis of Named Entity Recognition and Linking for Tweets. To appear in Information Processing and Management.
RILMTA	MSZNY 2015	Márton Miháltz, Tamás Váradi: Trendminer: politikai témájú közösségimédia-üzenetek feldolgozása és szociálpszichológiai elemzése ("Trendminer: Processing and Socio-psychological Analysis of Political Social Media Messages"). Paper submitted to the Eleventh Conference on Hungarian Computational Linguistics (MSZNY 2015), to be held in Szeged, January 15-16, 2015.

Institution	Article
UC3M	Lessons learnt from the DDIExtraction-2013 shared task. Journal of Biomedical Informatics. Isabel Segura-Bedmar, Paloma Martínez, María Herrero-Zazo. Accepted manuscript, available online: 21-MAY-2014 / http://www.sciencedirect.com/science/article/pii/S1532046414001245
USFD	L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak, K. Bontcheva. Analysis of Named Entity Recognition and Linking for Tweets. To appear in Information Processing and Management.
USFD	G. Gorrell and K. Bontcheva. Classifying Twitter Favorites: Like, Bookmark or Thanks? To appear in Journal of the Association for Information Science and Technology.
EK	DATA MANAGER ONLINE, Grandi dati, grandi informazioni?, Francesco Bellini, available on http://www.datamanager.it/2014/07/grandi-dati-grandi-informazioni/
UC3M	Improvements in detecting drug-drug interactions, Isabel Segura and Paloma Martínez, (English: http://portal.uc3m.es/portal/page/portal/actualidad_cientifica/noticias/drug_interactions, Spanish: http://portal.uc3m.es/portal/page/portal/actualidad_cientifica/noticias/interaccion_farmacos)
DFKI	Thierry Declerck (DFKI) cheif-editor of a special issue of the JLCL Journal on "Practice and Theory of Opinion Mining and Sentiment Analysis"
UC3M	Isabel Segura-Bedmar and Paloma Martínez Guest editors of the Journal of Biomedical informatics - Special Issue on MIning the Pharmacovigilance Literature.

2.4 Presentations (year 3)

- Thierry Declerck is invited to present lexical encoding in SKOS at the "SKOS applied - Tools and Strategies for the Design, Application and Sharing of Controlled Vocabularies", DARIAH-DE expert workshop. 18. Februar 2014. *Cologne, Germany.*

- Daniel Preotiuc-Pietro - Discovering the multifaceted information hidden within large user-generated text streams. 23 April, University of Pennsylvania, Positive Psychology Center Colloquium.

- Daniel Preotiuc-Pietro - Discovering the multifaceted information hidden within large user-generated text streams. 8 May, University of Bristol.

- Márton Miháltz - TrendMiner Projekt. 25 September 2014. Presentation at the Hungarian Natural Language Processing MeetUp, Colabs, Budapest.

- Thierry Decleck presents the DFKI work on harmonizing German Sentiment lexicons at the IGGSA Worskhop collocated with Konvens 2014, October 07, Hildenshein, Germany.

- Maciej Ogrodniczuk, Mateusz Kopeć, Aleksander Wawer. Large Scale Processing of Polish Political Tweets. To be presented at the 9th International Conference on Practical Applications of Language Corpora (PALC 2014), Łódź, Poland, 20–22 November 2014.

- Maciej Ogrodniczuk, Mateusz Kopeć, Aleksander Wawer. TrendMiner: Large-scale Cross-lingual Trend Mining Summarization of Real-time Media Streams. To be presented at the Natural Language Processing

Seminar organised by the Linguistic Engineering Group at the Institute of Computer Science, Polish Academy of Sciences (ICS PAS). Warsaw, Poland, 24 November 2014.

- Daniel Preotiuc-Pietro - Bilinear Models for Text Regression. 24 September, University of Pennsylvania, Computational Linguistics seminar.
- Andrea Varga will present TrendMiner and the WP3 research therein to Signal, London (http://signal.uk.com/), UK

2.5 Fairs (year 3)

Partner	Fair	Presentation
IMR	Data Tuesday	Une plate-forme B2B de traitement de l'information à grande échelle 04/2014, Paris, France
DFKI	HLT Village	HTL Village, 05/2014/Reykjavik, Iceland
IMR	IIPC	From Web Archiving services to Web scale data processing Plug & Play Platform 05/2014, Paris France
IMR	Time2Marketing	Renforcer son brand marketing avec la richesse du Web 06,2014 Paris, France

2.6 Open Source Software Releases (Year 1 -2 -3)

https://github.com/danielpreotiuc/twitter-collection-utils - data collection utilities for streaming collection, user-based collection, etc. This is a set of Python scripts that performs some usual data collection tasks from Twitter. This includes continuous crawling of tweets of a list of users or getting all the tweets from a user's timeline.

https://github.com/danielpreotiuc/trendminer-clustering - clustering code from D3.2.1 with pre-trained topic for 1 year of TMSORA data with cluster coherence and labels provided by SORA; released as a webservice

https://github.com/sinjax/trendminer - preprocessing pipeline working as a standalone tool, Hadoop tool, webservice

Code to transform the data using RBF kernels for use with the DMR topic model from Mallet. Method is introduced in D3.3.1: https://github.com/andreavarga/trendminer-sptempclustering

EU Summaries dataset, associated with Deliverable D3.1.1

Spatio-temporal clustering (WP3 - D3.3.1)
https://github.com/andreavarga/trendminer-sptempclustering

BGGR text regression model (WP3 - D3.3.1) https://github.com/sinjax/trendminer-python/tree/master/bivariate

Spectral clustering - removed Matlab dependencies (WP3 - D3.3.1). This algorithm builds hard clusters of words that co-occur in the same tweets.
https://github.com/sinjax/trendminer-python/tree/master/clustering

Merging of (German) Polarity lexicons (WP2: Task 2.1):
This repository contains the source code and data for SentiMerge, a Bayesian combination of four German sentiment lexicons. Details can be found in this paper, which was published at LG-LP 2014, a workshop at COLING 2014.https://github.com/guyemerson/SentiMerge:

Hungarian NLP pipeline for social media text analysis:
Trendminer Hungarian Processing Pipeline (trendminer-hunlp): a suite of scripts that perform Hungarian NLP processing steps (tokenization, pos-tagging, morphological analysis, lemmatization) by extending existing tools (huntoken, hunmorph, hunpos) to be able to deal with some of the challenges presented by the special language of social media messages, which differs from the domain of standard language (generally newswire) texts that were used to develop and train the exising tools
https://github.com/mmihaltz/trendminer-hunlp

Facebook data download, Java NooJ import/export format conversion
This package contains tools that were used to collect and process Hungarian data in the Trendminer project: 1) tools to periodically download new and updated public posts and comments published on specific Facebook pages, 2) a tool to convert NLP output files (from trendminer-hunlp) to NooJ input XML files, 3) a tool to extract annotations from NooJ output files.
https://github.com/mmihaltz/trendminer-hutools

Java NooJ command-line interface, sources
This repository contains configuration files to build the project with maven (maven.apache.org) and uses bintray.com/maven/grand-hifi/nooj-releases as an online release repository.: https://github.com/tkb-/nooj-cmd

Java NooJ command-line interface, binaries
https://bitbucket.org/tkb-/nooj-cmd/downloads

The TrendMiner Ontologies, associated with Deliverables D2.2.2 (Multilingual, ontology-based IE from stream media -v2) and D2.1.2 (Knowledge & Provenance Modelling and Stream Reasoning - v2) : http://www.dfki.de/lt/onto/

See also http://www.trendminer-project.eu/index.php/downloads

2.7 Online demos

- Sentipejd – a sentiment analysis tool for Polish made available as a Web service in Multiservice framework for chaining NLP Web services
- Prototype for monitoring drugs and medical events in Spanish social media (login: daedalus, passwd: trendminer2014)
- Ontology-based IE - online demo in English, German, Hindi, and Bulgarian
- Ontology-based IE - REST service for integration
- Prototype to annotate Spanish health user comments (GATE pipeline)
- Webservice for the LODIE disambiguation pipeline (http://demos.gate.ac.uk/trendminer/lodie/) . See for use D.2.2.2
- Demo of the Summarisation prototype for TrendMiner: http://demos.gate.ac.uk/trendminer/summarization/ Related to D.4.1.2 (Multi-Lingual Summarisation of Stream Media Software - v2: D4.1.2.pdf)

2.8 Cooperations

- TrendMiner (DFKI) and the project Pheme collaborating on the development of a multilingual ATC (*Anatomical Therapeutic Chemical* Classification System) ontology (http://www.pheme.eu/). DFKI
- TrendMiner (DFKI) and the project Eurosentiment (http://eurosentiment.eu/) collaborating on the LOD compliant representation of polarity lexicons. DFKI.
- There is a collaboration between WP2 work on evaluation datasets and the uComp project on human computation (https://www.ucomp.eu/), where uComp is helping us with expertise in acquiring gold standard corpora via CrowdFlower (and later - games with a purpose).
- USFD is also in ongoing discussions with http://www.cosmosproject.net/, on potential integration of the LODIE tools into their platform.
- The LODIE service by USFD will be used in the ForgetIT project (http://www.forgetit-project.eu/) for analysing social media and other web content
- USFD have also adapted and evaluated LODIE with the British Library, on environmental science content (http://gate.ac.uk/projects/envilod/EnviLOD-WP5-Quantitative-Evaluation-Report.pdf). That was a short 6 month collaboration, further details on which can be found here: http://gate.ac.uk/projects/envilod/index.html
- Cooperation with the Pheme project (USFD, ONTO and DFKI)
- Cooperation with the Gesis insitute on social science on the TheSoz taxonomy (bi-lateral meeting at DFKI on the 30.1.2014)
- DFKI cooperating with the Interest Group on German Sentiment Analysis (IGGSA) on developing and sharing polaritiy and sentiment lexicons for German
- DFKI Cooperating witht the GFAI (Saarbrücken) partner of the European EUMSSI project http://www.eumssi.eu/home/ on issues related to German lexicons and grammars for sentiment analysis
- IPIPAN collaborated with University of Social Sciences and Humanities in preparation of lexicons of several emotion-related and cognitive

dimensions, for instance: proactive and defensive orientation, also Polish version of WordNetAffect lexicon.

2.9 Online presence

2.9.1 Website Final results

We have installed the google analytics tools in order to monitor the visits of our webpage.
Here we give some few results delivered by the google analytics.

	Year 1	Year 2	Year 3
Sessions	NA	2,446	3,031
Users	NA	1,592	2,117
Pageviews	NA	9,881	12,332
Pages/Session	NA	4.04	4.07
Avg. Session Duration	NA	00:02:40	00:02:59
Bounce rate	NA	42.35%	44.18%
% New Sessions	NA	65.04%	68.90%

Table 2: Website statistics final results

While the used Content Management System used as backend for the TrendMiner Webpage gives us the following information on the most popular pages:

Popular Items	Created	Hits
Home	2012-03-12	30667
Objectives	2012-03-11	5065
About	2012-03-11	4969
TM-News	2012-03-15	4955
Data Sets, Web Services, Demos	2012-05-03	4860

Table 3: Most popular items

2.9.2 Twitter account

Tweets	81
Following	89
Followers	178

Table 4 : Tiwtter statistics account

Example of Tweets

e

Figure 1: Tweet

3 Exploitation

3.1 DFKI

As a private but non-profit research institute, DFKI has no direct commercial exploitation plans. But indirectly, DFKI is promoting results of R&D projects via the creation of spin-off companies[1].

It is intended to establish closer technology transfer agreements with the DFKI spin-offs that are potentially interested in the outcome of TrendMiner. A good candidate for this is the company "Attensity" http://www.attensity.com/home/), which emerged from a former spin-off, Xtramind. This kind of transfer agreement is not restricted to spin-offs.

Technology transfer and future research directly funded from industry is the main focus for our group in order to establish a longer term development in the fields covered by TrendMiner. We are in contact with 3 different companies, which are offering opinion or sentiment analysis, but this on the basis of methods not using linguistic or semantic features.

We will be careful here in not entering in conflict with strategies to be implemented by the use case partners involved in TrendMiner.

As the scientific level, DFKI established contacts and exchanges of data and results with other groups and initiatives, like the IGGSA[2] (Interest Group on German

1 See http://www.dfki.de/web/about/spin-offs/dfki-spin-offs?set_language=en&cl=en for a list of such Spin-Offs,.

Sentiment Analysis) society. DFKI delivered to IGGSA merged German lexical data sets for supporting advances in the state of the art in NLP-based opinion and sentiment analysis..
Last but not least, we are in contact with two academic institutions dealing with the domains of the two use cases of TrendMiner, and a deeper exchange of methodologies and data sets has already started. An extension to other academic and industrial institutions will ensure a longer term research goal.
DFKI published all their publicly available results either on its web page (for example the TrendMiner Ontologies: www.dfki.de/lt/onto) or in the github framework (for example the SentiMerge data set: https://github.com/guyemerson/SentiMerge). We hope that this data will be used by a large community, and that it will help establishing new strategic parternships.

3.2 USFD

USFD is an academic partner so exploitation activities are not its core aim. Nevertheless, the USFD work arising from TrendMiner offers a low barrier to entry for generating consultancy and commercial income from the know-how already available within the organization.

Additionally, USFD has an established history of exploiting the results of previous projects via the GATE open-source HLT tools and infrastructure. We are committed to maintaining the TrendMiner results beyond the bounds of the project and actively promoting their exploitation both by the NLP research community and to industry through consultancy services. The central role played by the USFD tools in TrendMiner will help reaffirm USFD's reputation as one of the world-leading centres for this type of R&D.

Some of the TrendMiner results will be provided also as web services from the AnnoMarket marketplace which will increase the awareness within the text mining industry regarding USFD's work in the area. This can lead to an increased number and size of consultancy contracts that USFD is approached for. The University development strategy actively encourages good contacts with industry and collaboration through consultancy and commercially-funded research projects. For instance, based on LODIE and the TrendMiner political ontology, we are already advising a UK charity on successful monitoring of the social media presence of UK political parties and politicians in the run up to the next general election (which is expected to be in May 2015, so beyond the bounds of this project).

3.3 Ontotex

Ontotext is an enterprise technology & solutions provider for semantic metadata management and analytics. The technology portfolio of Ontotext includes a semantic data integration platform, text mining & semantic annotation platforms for various

2 See http://iggsa.sentimental.li/

domains – such as Media & Publishing[3], Life Sciences & Healthcare[4], and Digital Libraries[5] – as well as a a high-performance semantic database (GraphDB[6]).
Within TrendMiner Ontotext has already successfully applied subsets of its technology portfolio, in particular:

- Semantic data integration for RDF-izing various data sources and integrating with open Linked Data sources
- the semantic database was used as the central semantic data warehouse where all integrated information and metadata is stored and queried.

The exploitable Trend Miner results for Ontotext can be summarised as:

- valuable experience with applying the of semantic database of Ontotext (GraphDB) in a scenario with real-time high velocity and high volume data streams. The results achieved by the University of Southampton within TrendMiner provided valuable insight to Ontotext regarding scalability limitations of the GraphDB database as well as practical approaches to extend the database with real-time reasoning capabilities.
- Important hands-on experience with high-performance computing and GPU technology. Even though the HPC research within TrendMiner was applied in a limited context, Ontotext had a great opportunity to test the technology potential and limitations in scenarios which are relevant to its commercial offerings. Adding high-performance computing / GPU related extensions is already on the long-term roadmaps of Ontotext's text mining & semantic database product lines.
- Insight and experience with sentiment mining technology developed within TrendMiner. Even though Ontotext is not involved in the actual development of the sentiment mining related components, as the main integration partner Ontotext got a hands-on experience with these (open source) components developed by various Trend Miner partners. One of the open source technologies developed within TrendMiner by the University of Sheffield was already integrated in a commercial product line of Ontotrext – the Self Service Semantic Suite (S4)[7] – in order to provide social media (Twitter) analytics capabilities
- Opportunities to deliver text analytics and metadata management solutions in the domain of social media analytics for political and PR campaigns with SORA. The two companies have a non-formal agreement to pursue together monetization opportunities combining the domain expertise of SORA and the enterprise technology of Ontotext.

3.4 University of Southampton

As an academic partner in the consortium, SOTON's exploitation strategy for the TrendMiner project is centred on the use of technologies and approaches developed

[3] http://www.ontotext.com/semantic-solutions/media-publishing/
[4] http://www.ontotext.com/semantic-solutions/life-sciences/
[5] http://www.ontotext.com/semantic-solutions/galleries-libraries-archives-museums/
[6] http://www.ontotext.com/products/ontotext-graphdb/
[7] http://www.ontotext.com/products/ontotext-s4/

within the project in our future research, rather than on explicit commercialisation. SOTON's main contributions are in two areas (stream reasoning and data mining of streaming media), and our exploitation plans for TrendMiner results are therefore concentrated here.

The Squall stream reasoner (from WP2) has formed the basis of a general framework for stream processing that has contributed to SOTON's existing portfolio of Semantic Web capabilities, and has been used as a foundation both for ongoing research within SOTON and for a number of grant proposals.

These tools have been made available to the community under an open source license, and we intend to feed the results from this work into the W3C RDF Stream Processing Community Group as a prelude to possible future standardisation.

Similarly, the work on non-stationary data mining (from WP3) contributes to SOTON's long-standing strengths in data mining and machine learning.

3.5 Internet Memory Research

After 36 months, IMR has widely benefited from TrendMiner Project to extend and improve the scope of its Web-scale data processing platform, mignify (http://mignify.com). According to Trendminer' topic, this led to major developments on active resources for political and financial sources and on resources from conversational Web.

Active sources

In the framework of Trendminer project, IMR provides fresh and qualitative sourcing for economic and financial trends. It includes the capacity to detect and to update on a relevant rate active sources such as RSS. From TrendMiner corpus based on specific domains and on European news, IMR industrialize the process of calculating refreshment rate and processing large amount of active sources to deliver a real added value to the platform and to its customers.
This output will benefit to the large scale data processing platform, but could also be implemented in IMR web archiving platform (http://archivethenet.com, dedicated to heritage institutions). Indeed, until now, web archiving customers needed to defined on their own the crawl frequency rate. With these new developments, IMR platforms will handle it automatically.

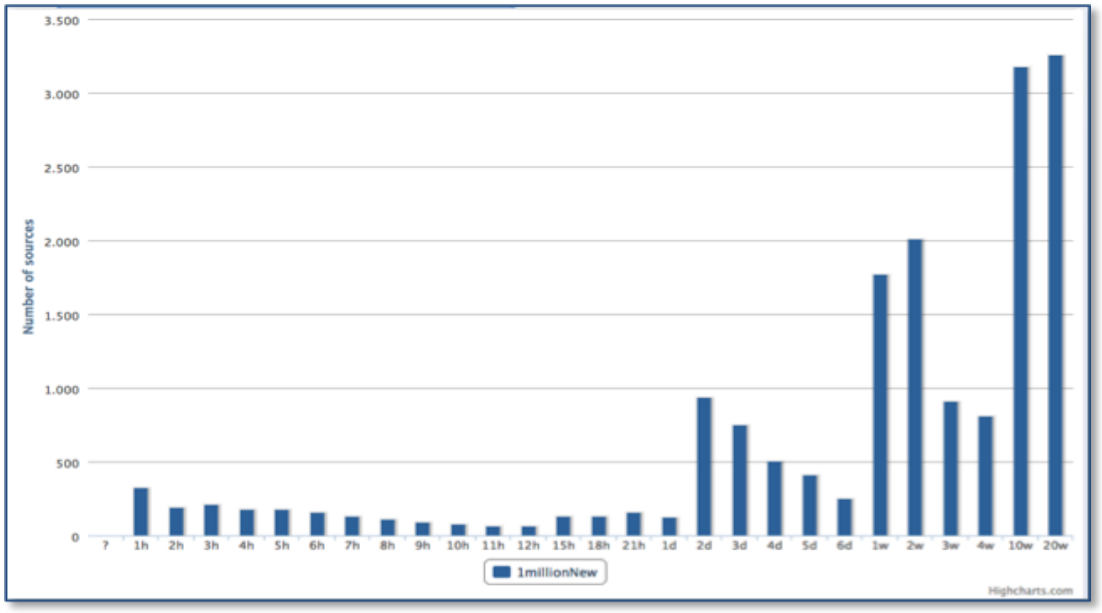

Figure 2 Example of RSS refreshment rate (Sample)

Conversational Web

One of the other important exploitable impact of TrendMiner project concerns discussions displayed on forums. Indeed, forums are rich in information, they contain not only personal stories but also discussions about products, brands, services and the way to consume them. But collecting discussions on the web and processing them at large scale is not a small issue and quite challenging.

This issue has been addressed during TrendMiner project and process has been industrialized in order to annotate, and extract valuable information such as the Title of the thread, the Post, and the Date. A annoator tool has been developped included wrapper extraction technology (Figure 3-4) And this data processing chain has been designed to be adaptable to several framework, such as market places, online catalogs,...

At the end, this web scale data processing output will benefit to many customer use-cases in many domains (marketing insight, profiling, price optimization, counterfeiting, piracy etc.)

Figure 3: Example of Page with its annotation

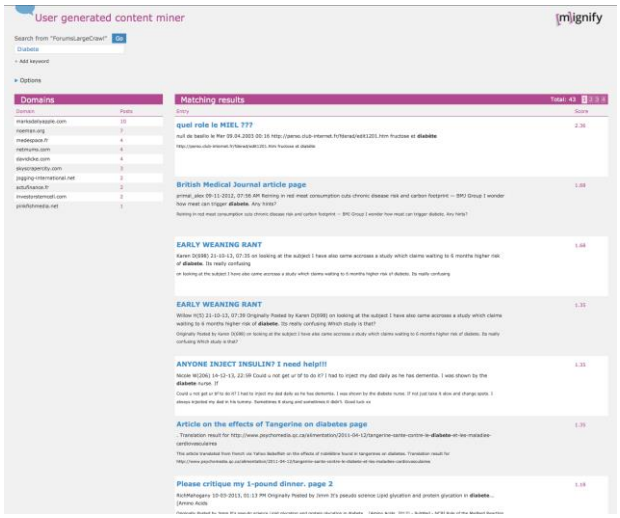

Figure 4 Research UI on Forums extraction

3.6 Eurokleis

Eurokleis has identified an exploitation strategy relying on the decision making capacity improvement of investors, analysts, governmental authorities and general public, in order to fully benefit from the potential of TrendMiner project to outperform innovative trading strategies.

Eurokleis will achieve this objective by creating three main products and several services, such as:

- **Pay per use - APIs (e.g. financial sentiment API, trading indicator API):** users would be able to post raw messages into our API and will receive back a response that gives the financial sentiment contained in the message.

- **Freemium model - Customized monitoring terminals/Online Tool:** Eurokleis will develop a real-time financial decision tool primarily aimed to support investors, analysts, advisors, stock brokers, but also to financial regulators, journalists and citizens which will display in real time the summary of the opinions regarding the selected financial instrument. This web application will be distributed free of charge, while the use of the underlying model can be sold to institutions interested in operating within financial markets.

- **Freemium model/subscription based - MOBILE APPs: to** optimize the information delivery to the costumers of the online tool, by for example alert systems. Other (free or premium) APPs will be developed for educational, recreational and professional purposes.

- **Subscription based - Basic account:** to provide basic information (e.g. like sentiment index and a rough expected price movement of selected financial instruments) towards interested persons.

- **Fact-based consultation - Customized proprietary portfolio management models:** to provide customized implementations addressing banks and financial professionals who need to use effective proprietary portfolio management models based on sentiment analysis.§ Additional services: e.g. financial consulting services, traditional and online training services about the use of the Eurokleis financial product and support services.

- **Potential future use case: anti-fraud detection**

Based on the knowledge achieved thanks to the TrendMiner project, Eurokleis will provide services to better support the identification of the financial frauds from the discovery of potential threats to the identification of the Beneficial Owner and the actual Ownership Structure.
The following functions will be provided for the Identification of stylized facts representing a signal of possible fraud, allowing zooming in/out for regions and time windows:

For market abuse:

- a trader places both buy and sell orders with the same price
- savvy online message board posters (a.k.a. "Bashers") who make up false and/or misleading information about the target company in an attempt to get shares for a cheaper price
- a group of traders create activity or rumours in order to drive the price of a security up
- selling and repurchasing the same or substantially the same security for the purpose of generating activity and increasing the price.

For VAT and money laundering:

- Business has changed commodities and sectors quickly.
- Turnover of business grows substantially in short time period
- Unsolicited approaches from organisations with little history in the market offering a guaranteed profit on high-value deals.
- Repeat deals at the same or a lower prices and small or consistent profit, e.g. €1 per item.
- Instructions to pay less than the full price to the supplier.
- Using goods that are of high value and low volume and attract a high tax rate such as computer parts or mobile phones.
- The total amount of money channeled through an account is considerable, although the balance is usually very low.
- Not publishing the companies' annual records, pursuing activities that are not part of the corporate goals.
- Obligatory elements of invoices, such as VAT number, date etc. are missing.
- Foreign nationals in charge of companies, who have often never been a director of a company in the jurisdiction and may not have an address in the jurisdiction.
- Invoices for services not usually associated with the business.

3.7 Sora Ogris & Hofinger

SORA is seeking agreements for cooperation with ONTO and DFKI for continued testing and development of the TrendMiner prototype for the Political Use Case after the end of the project.

The goal would be to explore the possibility of developing a marketable service, along the lines of a client-oriented, easy to use, smartphone-ready front-end platform together with a human curated back-end platform to offer personalised Twitter analyses and reports with the working title "STAT" (SORA Twitter Analysis Tool). Development and marketing would ideally follow an "early-access" logic, coupled with viral marketing methods that rely on word-of-mouth and engaging influential members for optimum effect.
The end-stage of STAT would be a real-time service for Social Media Monitoring that offers personalised analysis and reporting for journalists, political professionals and

activists, along the lines envisioned in the Business Plan (D8.3.2). Current coverage range (Twitter only) would be extended to Blogs and possibly even Newspaper Articles.

Short-term outlook (End of Project until mid 2015)

To this end, the approach chosen during the external testing phase in Y3 would initially be continued. The "closed-beta" phase begun during Y3 would be extended for another 6 months. The TrendMiner prototype would be disseminated during this time within an even wider circle of selected heavy users of Twitter in Austria, who would provide further feedback. These heavy users should be mainly journalists, PR and political professionals as well as other political activists, who represent the intended customer base of STAT.
Based on this feedback development of the prototype´s UI functions, data sources, object- and sentiment-detection capabilities and analytical components can continue. Additionally the back-end platform for curated Social Media Monitoring will be built, to create the first true version of STAT for external testing.
The "closed-beta" phase of the TrendMiner prototype should be followed by an official launch of STAT, timed to coincide with an important election campaign e.g. such as the 2015 Vienna elections for maximum publicity among future users and clients.
Select members of the initial circle of closed-beta testers can be offered free "early-access" trials of the personalised service and would be encouraged to discuss their use of the TrendMiner prototype on Twitter. Additional trial-accounts can be offered as and when useful for publicity and development. The early-access version of the personalised service would serve as both a vehicle for publicity and allows for the training of curators who would later manage the paid accounts.
Additionally, a freely accessible version of a non-curated version of STAT would be made available for the public in general and the intended client base of STAT in particular.

Mid-term outlook (Mid 2015 until ca. mid 2016)

Following a successful early-access phase, development would concentrate on optimising the interplay between back-end curation and front-end visualisation as well as usability features. Several releases, focused rather on optimising existing features than adding new ones, may be necessary until the early-access stage can be considered finished.
Furthermore interaction with the early-access users of personalised STAT, while curating their accounts would provide valuable expertise and support the development of a standardised and efficient model of curation for STAT.

Long-term outlook (mid 2016 onwards)
Having reached a point where the interplay between front-end and back-end softwares as well as curation techniques are considered adequate, existing users and beta-users will be offered a transition to the paid premium model. The freely-accessible version of STAT would remain online, however only in the form of a non-interactive demonstration and publicity platform.

3.8 Hardik Fintrade Pvt Ltd

Throughout the Trendminer project, HFPL's focus has been to help in developing products that offer a helping hand to sub-brokers and research analysts in their day to day research activities.
In the last few months, we have been validating two ideas of products with the help of our clients, sub-brokers, research analysts and other stock brokers in the market.
First, the summarisation application that summarises the content of a webpage and thus speeds up the process of analysing news articles. Second, a service that links entities (e.g. names of persons, organizations, locations, products) in news articles with the knowledge base to obtain more information about these entities.
Our findings suggest that:

- Most market research analysts find it very difficult to identify links among the entities mentioned in various articles and obtaining additional information about these entities is really a pain problem. It is also an unavoidable exercise without which it is difficult to come to an informed decision.

- Our survey of the B2B market suggests that use of mobile devices is rare as most researchers prefer to use laptops/PCs for their work usage. It suggests that investing into browser extension, at this stage, would be a better choice over getting a mobile application done in the B2C market. However, it is also clear that only mobile application should be the choice when entering into the B2C market.

- Until all the tools of Trendminer project have been evaluated (the exercise that is being carried out by various partners), it is risky to take them to the end-users (i.e. B2C). However most research analysts are eager to try out the beta versions of entity linking and summarzation services.

Thus, it was decided to go for a chrome extension that uses the Sheffield's LODIE and text summarising applications to annotate the content. For the integration of the summarisation application, we are waiting for it to be made available in the English language. So far the summarisation application is available in the German language, which, unfortunately, is not the language used in India.
As part of the dissemination activities, HFPL is planning to organise a workshop towards the end of the October 2014 where it would be a great opportunity for HFPL to showcase the demos and applications of the Trendminer technology. Members of the stock exchange and key influentials figures would also be invited to spread the word of achievements of Trendminer in the community.

3.9 Daedalus – UC3M Exploitation for the Health Use Case

Health use case business exploitation plan

Social Media Monitoring in the Health Sector

Social media can be defined as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation

and exchange of user-generated content"[8]. This broad definition covers user generated data, from Twitter to specialized blogs through Facebook. Nowadays, every company should be aware about opinions and mentions to them given by their customer in any of these social media and health companies are not an exception. The number of people with Internet access seeking for health information through the net ranges from 70 to 75% in the U.S. Besides, 42% of them used social media to get information about health issues. Patients talk with others about their feelings about a health problem, the way their bodies react to a given drug, how they mix different drugs to fight against some disease they have and many other issues related to their health situation. As an example of the importance of social media interactions in the health sector, according to a study developed by Price Waterhouse Coopers, 45% of consumers said information from social media would affect their decisions to seek a second opinion[9].

Age distribution of social media users must also be considered; there are surveys indicating that 89% of 18-29 year olds use social media in contrast with the 43% of people aged 65+. The difference of utilization by age groups will diminish over the next years and decades as digital natives increase their involvement and influence professionally and privately within their networks.

In order to analyze this market, the heavily regulated environment around health companies and prevention of direct-to-patient interactions must be taken into account, especially in Europe. This prevents pharmaceutical companies to get involved in social networks campaigns and only the half of the top 50 pharmaceutical companies in the world interact with patients through social networks. It is also worth mentioning that, outside the U.S., there are a lot of regulatory restrictions forcing pharmaceutical companies to behave in a conservative way. Nevertheless, the interest in listening patients' opinions through social networks as a first step through bidirectional communication with patients is increasing.

Current workflow

The process followed for social media monitoring in the healthcare domain does not differ from the ones applied in other domains. This process takes the following steps:

[8] Kaplan A, Haenlein M. *Users of the world, unite! The challenges and opportunities of Social Media*. Business horizons. 2010; 53 (1): 59-68.

[9] PriceWaterhouseCoopers, *Social media "likes" healthcare,* http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml

- *Listen*:

Select the conversations that matter. There are a lot of web sites and information sources to monitor, so there is a lot of information to process. This is one of the features to tag this problem as a Big Data problem: volume (a great amount of data), variety (data related to very different subjects) and velocity (interactions of patients with social media are frequent).

- *Analyze*:

The data gathered by the listening process must be deeply analyzed in order to get the subject of the conversation, the frequency of a given subject, the mood and sentiment expressed by user related to that subject and possible existing trends for those issues.

- *Engage*:

Once the relevant conversations and issues have been identified, it is time to engage with selected customers. For example, a compliant has been identified and a quick and planned answer must be given to the customer.

- *Evaluate*:

Finally, success of our social media strategy must be measured in some way. There are a lot of parameters to measure, all of them depending on the goals defined for the company and what your audience wants. There are a lot of possible dimensions to measure but it is recommended to go beyond superficial ones such as number of followers. For example, new referrals or the number of comments or ideas gathered from the audience could constitute interesting figures.

According to the regulatory issues previously mentioned, the engagement phase must be carefully defined but, in the U.S market, there are already examples of good practices like the Mayo Clinic[10].

This is a generic social media monitoring process but, if specific markets must be addressed, specificities of those markets must also be addressed. For this reason, there is still a high level of manual intervention to fulfill monitoring tasks. If natural language processing is going to be applied to reduce manual intervention, resources allowing the identification of drugs, diseases, conditions and so on, must be available, and it is something not at hand of most social media monitoring tools.

Deficits

There are several issues forcing companies to make big investments in manual resources to process social media contents.

- *Language is ambiguous*: that is the most relevant problem when dealing with natural language. It is needed to distinguish between Sonata as the name of a car model from Hiunday and Sonata as the name of an antidepressant.
- *Topic identification*: relevant elements for the target market or sector must be properly identified. It is very different recognizing smartphones models, brands and their features than identifying diseases, drugs, conditions, symptoms and pharmaceutical companies. Besides, language style used by patients to talk about their illnesses or health problems it is different from the one they use to talk about their mobile phone. At the end, they try to remember and imitate language expressions from their physicians. On the other hand, the vocabulary used by physicians is also different from the terms used by patients. For example, a physician talks about a cephalalgia or cephalea but patients say it has headache.
- *Sentiment analysis*: monitoring not only means knowing when somebody talks about a brand or topic but also identifying if they are doing it on a positive or negative way. From the natural language processing point of view, this requires to analyze sentences to relate sentiment expressions to a topic in the sentence. For example, in the sentence: 'The battery of the iPhone is bad but I like its screen', the usual thing is label it as neutral regarding the iPhone brand but a deeper analysis should relate 'battery' and 'screen' to iPhone and then 'bad' with 'battery' and 'like' with 'screen'. Current tools for sentiment analysis hardly reach this level of detail in content analysis.
- *Precision*: the most common dimension to measure recognition quality in natural language processing tasks is precision (along with recall, but precision is perceived as most relevant for users). The reachable level of precision depends on the given task. For example, it is common to reach a 90% precision in topic and entity recognition but when sentiment analysis is considered this percentage gets reduced to 70 or 80%. In any case, users of sentiment analysis tools usually say that they get better results with human intervention but, how much texts humans can process? And, do they apply common criteria to decide between positive and negative situations?

[10] Wipro, Transforming Healthcare through Social Media, http://www.wipro.com/Documents/resource-center/library/impact_of_social_media_in_healthcare.pdf

- *Trend identification*: the next step in social media monitoring functionalities goes through trend detection. News in social media spread quickly so means to recognize most commented topics with time are important for marketers. Trend recognition is a feature not provided by current social media monitoring tools.

TrendMiner Health use case

TrendMiner considers a healthcare business case covering the deficits mentioned in the previous section. The functionality covered by the health application in TrendMiner includes:

- *Listening*: to detect mentions to relevant health related entities (hospitals, pharmaceutical companies, drugs, diseases, physicians), filtering out those conversations.
- *Analysis and visualization*: extracting the relevance and trend related to identified entities, the concepts commonly accompanying those entities. This information is used to show a dashboard helping decision taking by managers and allowing a drill down to the source of the aggregated data shown in the dashboard.
- *Trend detection*: analyzing the relevance of given issues through time is relevant information for any domain, healthcare included. This trend analysis can result, for example, in a estimation of drug consumption or in the identification of relevant conversation themes around an entity.

For this purpose, specialized resources are available and adaptations for health language specificities have been developed:

- *Brand monitoring*: entities related to the health domain: pharmaceutical companies, most relevant hospitals, insurance companies. Concepts commonly related to these entities are identified. Besides, user contributions relating a brand can be classified according to a reputational model.
- *Drug monitoring*: common names for drugs, diseases, conditions and medical vocabulary has been built taking into account existing ontologies and approaches like UMLS[11], CIMA[12], MedDRA[13] (for drugs), the list of most relevant drugs from the European agency and so no. Besides recognition of mentions to different health topics, it is also possible to mark if those topics are seen as positive, negative or neutral by patients. This goes from the extraction of positive comments regarding a pharmaceutical company or an specific drug supplied by a pharmaceutical company or many other combinations.
- *Adverse Drug Reactions identification*: besides, once it has been possible to recognize drugs mentions, it is also possible to extract conversations where

[11] *Unified Medical Language System*, http://www.nlm.nih.gov/research/umls/
[12] Centro de Información online de Medicamentos de la AEMPS, http://www.aemps.gob.es/cima/pestanias.do?metodo=presentacionInicial
[13] Medical Dictionary for Regulatory Activities, http://www.meddra.org

reactions to drug consumption are expressed and the effects (positive or negative) of those reactions.

Main competitors in the market

There social media monitoring market is a crowded one. There are a lot of companies providing monitoring services, but not all of them have a clear offer in the health sector. Among the most important ones it is possible to find:

- **IMS Health:** IMS Health is an information, services and technology company dedicated to making healthcare perform better. By applying cutting-edge analytics and proprietary application suites hosted on the cloud, the company connects more than 10 petabytes of complex healthcare data on diseases, treatments, costs and outcomes to enable clients to run their operations more efficiently. Drawing on information from 100,000 suppliers, and on insights from more than 45+ billion healthcare transactions processed annually, IMS Health's 9,500+ professionals drive results for over 5,000 healthcare clients globally. Customers include pharmaceutical, medical device and consumer health manufacturers and distributors, providers, payers, government agencies, policymakers, researchers and the financial community.

- **3M Healthcare:** 3M is an innovation company serving customers in five different business groups, which increase speed and efficiency by sharing technological, manufacturing, marketing and other resources. In healthcare, 3M contributes supplying innovative and reliable products that help health care professionals improve the quality of care, providing medical and oral care products, drug delivery and health information systems.

- **Oracle:** The company has an specific offer in the healthcare sector. Oracle Healthcare solutions for payers and providers improve patient care and outcomes while maximizing efficiency and containing costs. Oracle delivers powerful solutions that address critical areas in the healthcare industry including integrated clinical data, and service line analysis features ensure lower-cost high-quality care, patient safety concerns and enable better patient outcomes.

- **Salesforce:** Salesforce.com provides social and mobile cloud technologies—including our sales and CRM applications— helping companies connect with customers, partners, and employees in entirely new ways. And thanks to the cloud, all of it comes with low cost, low risk, and fast results. 100,000+ successful customers tell the story best. The company has a specific offer in the health sector, although there is no defined product, the company provides specialized consulting for health applications.

On the other hand, all these competitors could also be seen as integrators of TrendMiner Health use case technology, which could be applied to the improvement of their offer. The main drawback of products of the competitors is that they do not cover deeply other languages than English, so it could be interesting for them to increase their coverage.

3.10 Daedalus exploitation strategy

Daedalus exploitation strategy is base on a Software as a Service model, through its product **Textalytics, meaning as a service**, www.textalytics.com. This product includes a set of API (Application Programming Interfaces) devoted to facilitate the integration of semantic processing in any application. The business model is based on pay-per-use scheme and it is driven to developers wanting to include text analytics technology into their applications. The TrendMiner approach goes a little bit further because it is driven to final users, not to developers, so dashboards to visualize data are needed. The experience of the rest of the partners and Daedalus in the development of these visualization strategies will be a key issue in TrendMiner product.

Target groups
Companies in different sectors can be interested in these services in the health domain. Among them it is possible to find:

- Pharmaceutical companies.
- Medical insurance companies.
- Medical centers.
- Marketing agencies working with pharmaceutical companies.

3.11 RILMTA - Department of Corpus Linguistic of the Hungarian Academy of sciences

Downloadable Resources

RILMTA publishes Hungarian resources produced during the Trendminer project in order to facilitate unrestricted future exploitation by third parties. The data explores reactions to posts by politicians by Hungarian social media users. It can serve as a basis for future analyses, surveys, studies etc. by political analysts, social scientists, journalists, politicians, NGOs etc.

On the RILMTA Trendminer project website (http://www.nytud.hu/depts/corpus/trendminer.html) the following datasets/resources will be available for unrestricted download:

Corpus of Public Facebook Political Comments

In the framework of the TrendMiner Project, RILMTA collected 1.93 million public Facebook comments addressed to public posts published on 500+ different Facebook pages belonging to Hungarian political organizations (political parties and their regional and other branches, political movements etc.) and politicians (members of Parliament, members of government etc.)
The corpus contains 1,939,356 Facebook comments (comprising 46,211,723 tokens) which are dated between 10.1.2013 and 9.22.2014. Meta-information associated to all comments includes date of publication, author user id (anononymized), Facebook comment id and ids of the Facebook post and the page where it was published. All

comments also have the following annotations: standard NLP (sentence and token boundaries, token surface form, lemma, part-of-speech tag and morphological analysis, named entity tag), socio-psychological content analysis tags (10+ tags for various categories) and scores (14 different counts and scores calculated from the annotations including sentiment polarity etc.)

Hungarian Political Ontology

The Hungarian Political Ontology represents the domain of Hungarian Politics during 2013-2014, focusing on the 2014 National Assembly elections and the 2014 European Parliament elections.
The ontology is available in standard OWL Turtle format and contains concept classes, properties and instances related to the domain. In addition to the general entity hierarchy created for all Trendminer political uses case parties, the Hungarian Political Ontology contains 18 new entities (classes, properties and individuals). It is populated with 1610 additional individuals that represent Hungarian political parties and persons (election candidates as well as office holders) involved in the 2010-2014 Hungarian Parliament, the 2014-2018 Hungarian Parliament (in 2014) and the 2014-2019 European Parliament (in 2014)..

Software Releases

Trendminer Hungarian Processing Pipeline (trendminer-hunlp): a suite of scripts that perform Hungarian NLP processing steps (tokenization, pos-tagging, morphological analysis, lemmatization) by extending existing tools (huntoken, hunmorph, hunpos) to be able to deal with some of the challenges presented by the special language of social media messages, which differs from the domain of standard language (generally newswire) texts that were used to develop and train the existing tools.
https://github.com/mmihaltz/trendminer-hunlp
Trendminer Hungarian Tools (trendminer-hutools): This package contains tools that were used to collect and process Hungarian data in the Trendminer project: 1) tools to periodically download new and updated public posts and comments published on specific Facebook pages, 2) a tool to convert NLP output files (from trendminer-hunlp) to NooJ input XML files, 3) a tool to extract annotations from NooJ output files.
https://github.com/mmihaltz/trendminer-hutools
Java NooJ Commandline (nooj-cmd): this tool, developed for the Trendminer project, extends the open source Java NooJ API (http://nooj4nlp.net/) to provide command line interface for most NooJ features (tagging of input documents in various formats using Java NooJ dictionaries and grammars).
Sources:https://github.com/tkb-/nooj-cmd
Binary: https://bitbucket.org/tkb-/nooj-cmd/downloads

3.12 IPIPAN Institut of Computer Science Polish Academy of Sciences

The plan aims at exploiting the results of the TrendMiner project in regard to ontology and social media resources and tools. Journalists, NGO's activists, individual citizens, political researchers, as well as political consultants and PR and

lobbying professionals, may directly benefit from the TrendMiner project. The project results will help to develop simple applications and custom services supporting the decision making of the individuals.

ONTOLOGY

Free of charge
It will be possible for watchdog NGO's and informal citizen networks to trace the careers and volatility of individual politicians since 1989 until present. This information will be especially important, as it will help optimize political decisions of millions of potential voters at the time of top electoral campaigns, as in the home country, as at the European Union level.
Basing on the existing and easy to develop database of Polish politicians, political sociologists will be able to identify and describe the political careers of Polish politicians, which will enhance systematic knowledge regarding the scope and nature of political change in the democratic transformation process. It will as well as help the evolving power clusters influencing different areas of cultural, social, economic, political life.

Proprietary
Lobbying and PR companies will be interested to identify the clusters of the decision makers, who could be approachable and potentially support the argument.

TWITTER

Free of charge
Journalists and media researchers will be able to trace the relationship between the discourse on Twitter between politicians and the discourse of the media. Media companies will be able to better monitor the public discourse.

Proprietary
The twitter sentiment trend identification tool will be developed for all those political parties, media and interest groups that search for the current knowledge on sentiment regarding public issues within different segments of public opinion, especially the young, new coming voters.

- Lobbying and PR companies will be interested to identify the clusters of the decision makers, who could be approachable and potentially support the argument.