
PROJECT FINAL REPORT



Grant Agreement number: ICT 316404

Project acronym: NewsReader

Project title: Building structured event indexes of large volumes of financial and economic data for decision making

Funding Scheme: FP7-ICT-2011-8

Period covered: from 1/1/2013 to 1/1/2016

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Prof. dr. Piek Vossen, Faculty of Arts, VU University Amsterdam

Tel. + 31 (0) 20 5986466

Fax: Fax. + 31 (0) 20 5986500

E-mail: piek.vossen@vu.nl

Project website: <http://www.newsreader-project.eu>

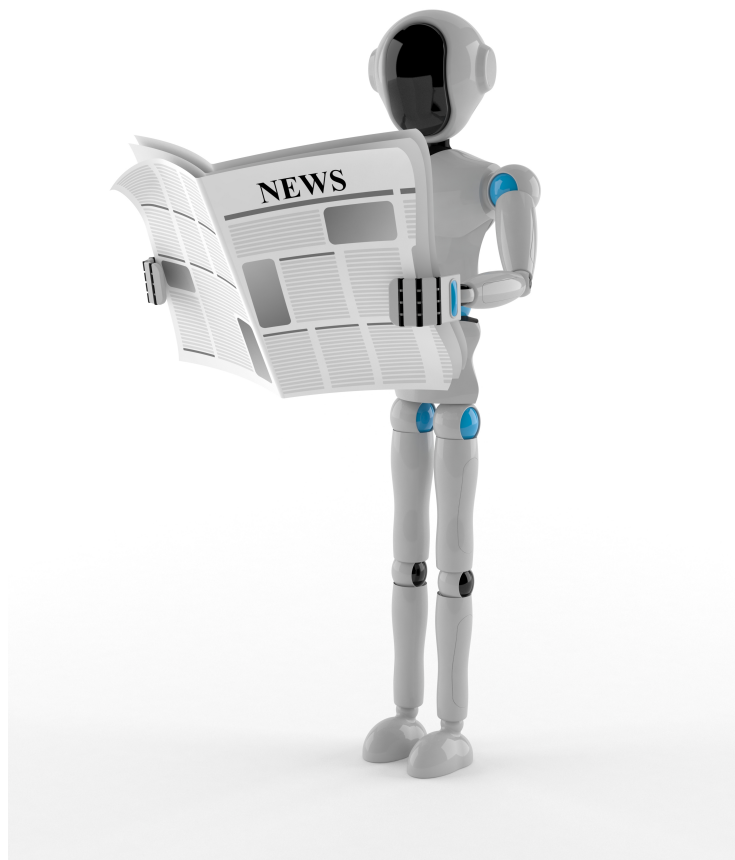
Contents

1	Final publishable summary report	3
1.1	Executive summary	3
1.2	Summary description	4
1.3	Main results and foreground	8
1.3.1	Architecture and design	8
1.3.2	Interoperable Natural Language Pipelines in four languages	13
1.3.3	Cross-document and cross-language event modelling	13
1.3.4	Manually annotated evaluation data	15
1.3.5	KnowledgeStore technology	17
1.3.6	Processed data and Event Centric Knowledge Graphs	17
1.3.7	Modelling events and their implications	18
1.3.8	NewsReader intelligence	20
1.3.9	Data visualisations and interactions	25
1.3.10	Hackathons and end-user-evaluations	27
1.4	Potential impact	29
1.5	Consortium details and contact	36
2	Use and dissemination of foreground	37
3	Report on Societal implications	108

1 Final publishable summary report

1.1 Executive summary

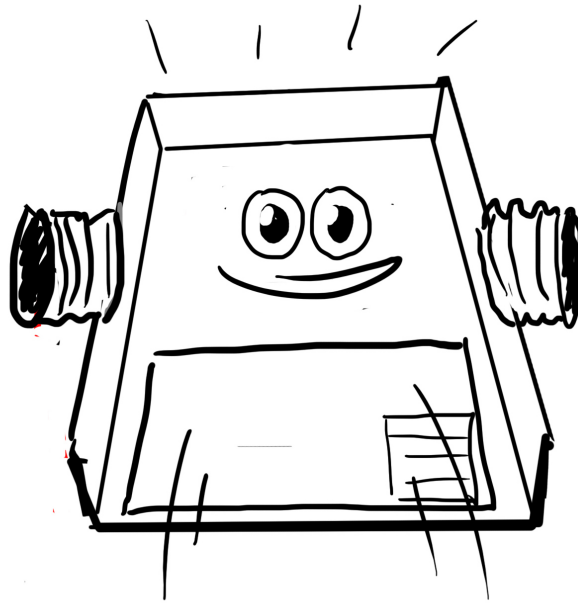
We developed a so-called Reading Machine that can read massive streams of news in 4 languages: English, Spanish, Italian and Dutch. The machine extracts *what* happened and *who* was involved, *where* and *when*. This information is represented in the form of billions of Semantic Web *triples* and stored in a KnowledgeStore that supports reasoning over the data. This allows us to detect trends, events with impact and social networks of people over time and regions. We can query long-term developments spanning decades for individuals or types of individuals to discover events that remained unnoticed. The project developed new and unique visualisations of the rich and complex data structures that provide efficient and intelligent access to the data. Currently, NewsReader technology is used in almost 40 follow-up projects.



1.2 Summary description

Sheila is a senior spokesperson of the Ministry of Internal Affairs. She monitors the daily stream of news: hundreds of documents per day! When Sheila reads an article that rumours about falsifications in CO2 emissions by a national automobile manufacturer she needs to advise her Minister on a response in the media within sixty minutes. To do this well, she needs to have an accurate picture of the full history, spanning decades, of the manufacturer, the management currently in charge and the connections with the government. Access to an overwhelming amount of data - millions of relevant documents - makes it almost impossible for Sheila to find the information she needs in the short time-span she has to give a well-informed advice.

Watch the story of Sheila and NewsReader's solution here on Youtube: <https://www.youtube.com/watch?v=rYLaVN3oqLI&feature=youtu.be>.



LexisNexis estimates the total volume of news that they archive per working day on 1.5 million articles. About 25% is about finance and economy: five-hundred-thousand documents per working day. A period of 10 years spanning the financial crisis will add up to an enormous volume of news and data. This accelerating growth of knowledge and information makes it nearly impossible to stay on top of developments. Making informed decisions and finding out about the consequences of your decisions becomes more difficult for businesses, governments and also citizens. Information and data not only grow exponentially because they become digitized. At the same time, our (online) activity and mobility accelerate, expanding our networks and intensifying the dynamics between them.

Professionals in any sector depend on access to accurate and complete knowledge to make well-informed decisions. Think about lawyers, politicians, heads of purchase in large firms, compliance directors and journalists. A missing piece of crucial information can be fatal to make the right decision. To find this crucial piece they need to search for a needle in a haystack, simply because there is more data than ever, it is highly interconnected through the Internet and quickly gets out of date in this rapidly changing world.

NewsReader developed the reading machine: a machine that can read millions of news items in four languages (English, Spanish, Italian and Dutch) day by day. This way, NewsReader helps these professionals allowing them to find that needle in the haystack by structuring information as stories: relating new events to past events. It stores all the details in the so-called KnowledgeStore and can intelligently reason over it by reconstructing histories over decades. It is thereby able to tell all the details about any person or organisation as told by thousands of different sources. NewsReader is also capable of measuring the impact of events on the people involved rather than the impact of news on the journalists and people that follow the news. Whereas the latter measures how much talk there is about topics or how trendy they are, NewsReader models the extent to which people are affected by the event: e.g. increase or decrease of ownership or sales, loss of jobs, etc.

How this works? In four simple steps: Identification, Deduplication, Aggregation and Perspectivation. IDAP.

Identification: NewsReader first identifies an event in text through similar components, by extracting *what* happened to *whom*, *when*, and *where*.

Deduplication: NewsReader makes sure similar information is represented only once, referencing every article across many sources in the haystack.

Aggregation: NewsReader aggregates complementary information across thousands of different sources in a single representation.

Perspectivation: NewsReader makes sure differences and different view points are traceable through their sources and mentions in text.

The result of this IDAP method is a complete, exact and rich record of the past, with access to original sources. The information from the news is stored as billions of 'factoid' statements, so-called **RDF triples**.

The overall architecture for this process is shown in Figure 1. In a first step, we use Natural Language Processing technology (NLP) to detect events, actors and time expressions in the news in 4 languages. The result is stored in XML files according to the Natural Language Processing Annotation Format (NAF), that we defined. NAF is interoperable across different languages. The result is stored in the KnowledgeStore, which is a scalable database platform for storing massive amounts of source data and interpretation layers of this data.

But this is not it! Typically, the same entities, dates and events are mentioned many times in news articles and especially across many different articles published around the

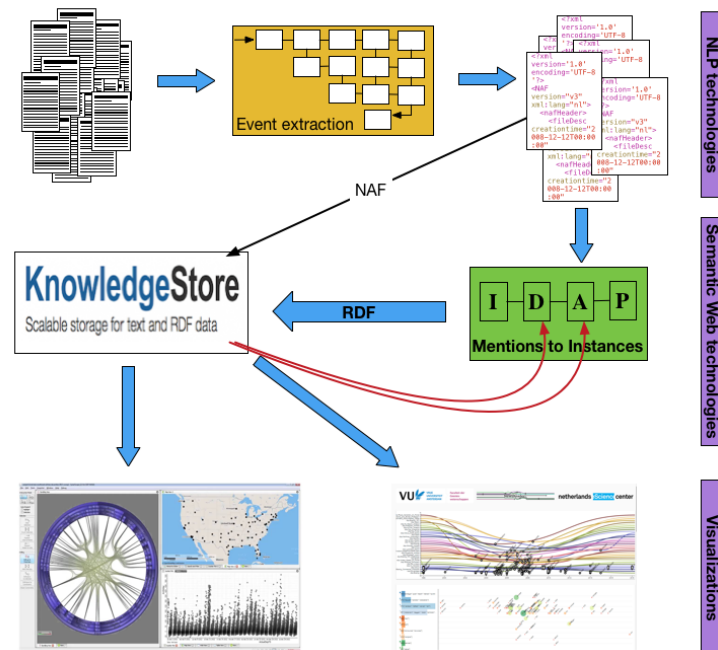


Figure 1: Global Architecture of the NewsReader reading machine

same day, which we can expect to report on similar events. We therefore make a distinction between mentions in (textual) sources and instances in the (assumed) world. We therefore *reinterpret* the Natural Language Processing output in NAF to an instance level, where each unique event and entity is represented only once using the IDAP method. The result is an RDF representation of the knowledge on the event, following the Simple Event Model (SEM) that is also stored in the KnowledgeStore.

This second processing step is illustrated in Figure 2, where show two descriptions of the same event from two different sources that use different words and expressions. The information is mapped to a unique representation of an event instance (Event₁₂, with labels *buy* and *sell*). The event has *buyer* and *seller* roles to entities that are identified through their DBpedia¹ identifiers. Since NewsReader interprets the events as instances of event types (represented here as *Commerce_money_transfer*), the system also understands what the transferred goods between these entities are. In this example, we see that similar information is deduplicated despite the different ways it was expressed. The event type predicts that there is also money involved but this is not expressed in the current source texts. Other (future) sources may tell us the amount of money paid and thus can provide this missing information to complete the picture. This will then lead to aggregation. We call these data structures Event-Centric Knowledge-Graphs or ECKGs, because all information is aggregated from different sources around the event rather than an entity.

¹DBpedia is a Semantic Web database with content derived from Wikipedia, containing millions of entities and properties of these entities

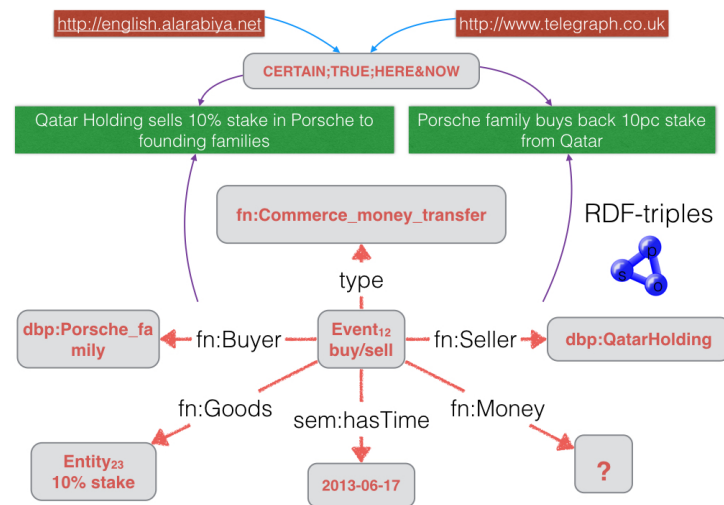


Figure 2: Representing instances of events and entities across sources

Aggregation and deduplication are important to make the correct inferences. If for example 10 articles report on a *sale of 10% stake*, we need to know if we can add up these sales to 100%, which implies they report on different sales, or 10%, in which case all articles report on the same event, or anything in between.

Since we keep the link between the RDF representations of event and entity instances and their mentions in the original sources, we can go back to the original sources at any point to show where the knowledge and information is derived from. This allows us to model the perspective of the sources of the news on the events (both the authors and publishers of the news as well as the cited sources in the text). Our *reading machine* builds a data structure that also makes explicit *who* said what, how *certain* they are, whether they *deny* or *confirm* it and what *emotion* they have towards the events. This data layer is also stored as RDF triples in the KnowledgeStore and can be used to model the perspective on events across many different sources.

Processing millions of news articles over decades, as has been done in NewsReader, results in a KnowledgeStore filled with billions of RDF triples which are little 'factoid' statements on events and perspectives linked to the source texts. Each data element is bound to time and sometimes also to place, and is semantically typed according to ontologies. This allows for reasoning over the data (what are the implications of events) and allows for deep semantic search (using SPARQL). Querying for types of people (e.g. *management*) and companies (e.g. *car manufacturers*) and also for types of events (e.g. *financial transactions* or *crimes*) in which they are involved makes it possible to visualise trends over time and/or in regions, show biographies and social networks and event storylines of sequences of events with causal connectivity. At the bottom of Figure 1, we show two high-end visualisations developed in the project that can be used to efficiently

access the data, detect correlations and trends and discover hidden events that remained unnoticed so far. The project used these interfaces and the KnowledgeStore in end-user experiments and hackathons that study the effectiveness of our data processing and modelling for professionals.

1.3 Main results and foreground

NewsReader can rapidly read texts in four different languages and creates a single Semantic Web representation (RDF triples) to represent so-called event-centric data across different text sources and different languages. The information is stored in a scalable KnowledgeStore that can hold background knowledge and supports reasoning. The reading machine as a complete system is a major achievement that integrates many different components that are also important achievements in themselves. We summarise these components briefly below. In Section 2, we provide a complete list of all the foreground results.

1.3.1 Architecture and design

We defined a unique system architecture that is open, flexible and extendable and that combines Natural Language and Semantic Web technology providing a technology bridge between unstructured and structured data. The data flow is shown in Figure 3. Textual sources are processed through pipelines of Natural Language Processing (NLP) modules that store the result in NAF-XML format. Next we interpret the mentions in NAF to instances in SEM and compare these across different articles. We store the final results as SEM-RDF triples.

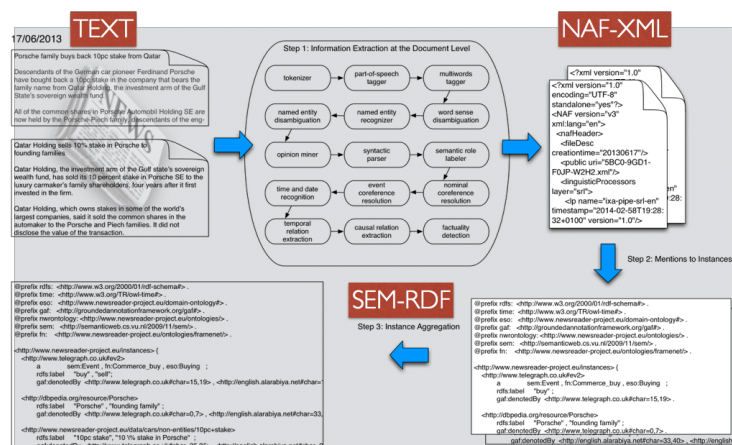


Figure 3: Overview of process and data flow

In Figure 4, we show the abstract SEM model that is used to capture the resulting RDF data. SEM allows modelling relations between event instances, actors, time and place.

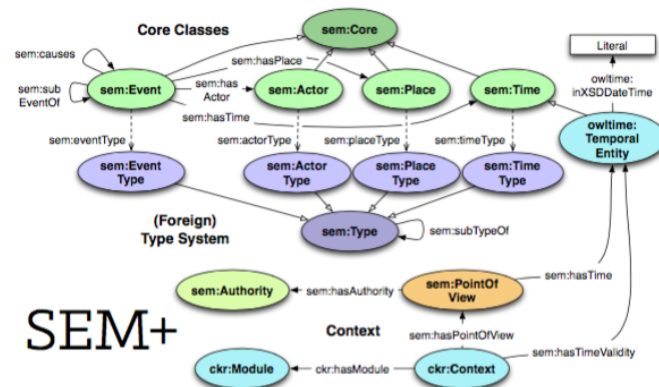


Figure 4: Simple Event Model

Whereas SEM is instance based, NAF representations allows for the annotation of mentions in text with interpretations. We defined the Grounded Annotation Format to link instance to mentions using `gaf:denotedBy` links. Each mention of information is also attributed to a source, which can either be the author or somebody cited. We model these attribution relations in our GRASP model (General Representations and Annotations of Sources and their Perspectives), which allows for the expression of perspectives of sources on events represented in SEM. Since each mention can represent a different perspective, we link the perspective to each mention. In Figure 5, we show a schematic overview how GAF combines all three models where different mentions in textual sources are mapped to the same SEM instance through `denotedBy` links and each mention is mapped to a perspective through `hasAttribution` and to a source through `wasAttributedTo` links.

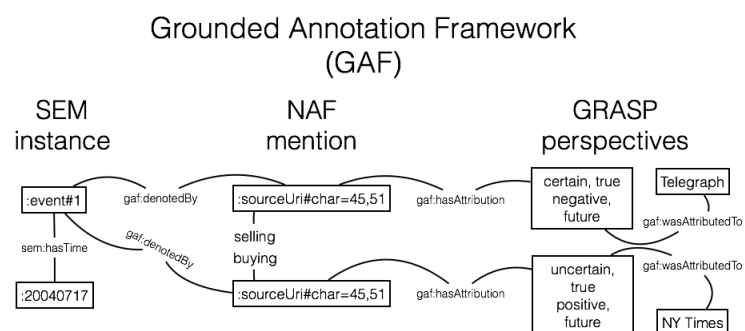


Figure 5: High level overview of the NewsReader models

Figure 6 then shows all formal triple relations for the semantics of the two example sentences presented before. We can see here that there is sharing of the SEM data, there are links to background ontologies indicating the type of event, there are links to their

mentions represented in NAF and to the perspective represented in GRASP.

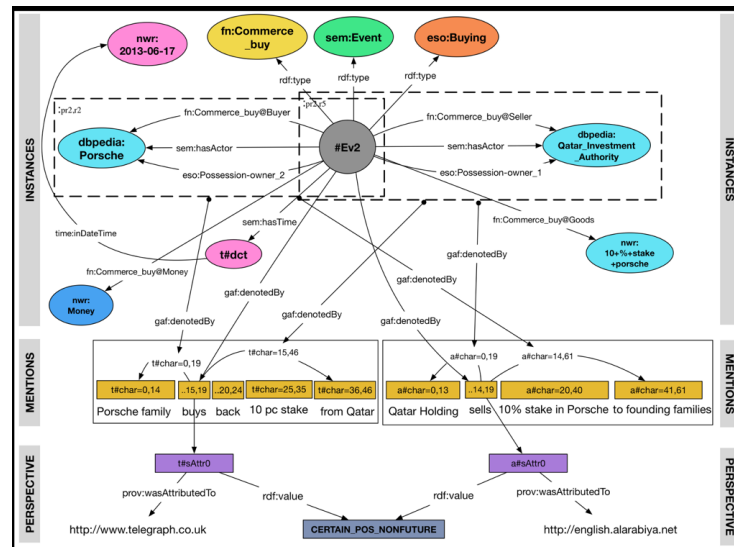


Figure 6: Example represented in NewsReader models: NAF, SEM, GRASP following GAF

Modelling the perspective allows us to find all statements of spokesmen over time and get an overview of their position and sentiment with respect to the events. In Figures 7 and 8, we show such lists for the Porsche CEO *Wiedeking* extracted from a large data set of news articles on the automotive industry processed by NewsReader.

The rich and complex modelling of data in NewsReader has a high potential for future research and technology development. Our models can deal with textual and non-textual sources and can be applied to any language in the world. We have been invited to participate in an ISO working group to investigate the standardisation of NAF. Our framework is already applied beyond the project's lifetime by external third parties.

The software design to implement the processing is shown in Figure 9. It shows a series of Natural Language modules that process incoming sources, such as news batches provided by LexisNexis, adding different interpretation layers. The central box represents the KnowledgeStore that contains all the source data, the interpretation layers and the final SEM-RDF triples with data. The KnowledgeStore can be queried by tools such as the visualisations developed by Synerscope.

We defined different architectures for parallel processing in batch mode (using Hadoop) and streaming mode (using Storm and a Mongo database). The Hadoop architecture performs best for large batches of data (millions), whereas the Storm architecture is optimal for continuous streams of data. We have demonstrated that our architecture can deal with massive amounts of news within operational limits: hundreds of thousands articles per day.

WENDELING WIEDEKING'S PERSPECTIVE

Event	Nr.	Perspective	Event	Nr.	Perspective
achieve	10	CERTAIN_u_POS	interest	8	CERTAIN_NON_FUTURE_POS
decision	6	CERTAIN_NON_FUTURE_NEG	interest	8	CERTAIN_u_POS
earnings	19	CERTAIN_FUTURE_POS	plans	9	CERTAIN_NON_FUTURE_POS
earnings	19	CERTAIN_NON_FUTURE_POS	plans	9	CERTAIN_u_POS
earnings	17	u_u_u	predict	17	CERTAIN_FUTURE_POS
focus	6	CERTAIN_FUTURE_POS	predict	17	CERTAIN_NON_FUTURE_POS
goal	7	CERTAIN_u_POS	predict	17	u_u_u
increase	18	u_u_u	stake	21	CERTAIN_u_POS
increase	15	CERTAIN_NON_FUTURE_POS	stake	19	CERTAIN_NON_FUTURE_POS
increase	11	CERTAIN_u_POS	stake	10	u_u_u
increase d	9	CERTAIN_NON_FUTURE_POS			
increase d	8	CERTAIN_u_POS			

Figure 7: Perspective on event expressed by Porsche CEO Wiedeking

WENDELING WIEDEKING'S SENTIMENT

Event	Nr.	Sentimen
interest	6	positive
voting	6	positive
location	3	negative
production	3	negative
waging	3	negative
war	3	negative
exercising	3	positive
gives	3	positive
have	3	positive
position	3	positive
restructuring	3	positive
build	2	positive
expertise	2	positive
hoping	2	positive
models	2	positive

Event	Nr.	Sentimen
retain	2	positive
wants	2	positive
avoid	1	negative
questioning	1	negative
team	1	negative
acquired	1	positive
influence	1	positive
maintain	1	positive
manufacturer	1	positive
order	1	positive
raise	1	positive
seeGermany	1	positive
stake	1	positive

Figure 8: Sentiment on event expressed by Porsche CEO Wiedeking

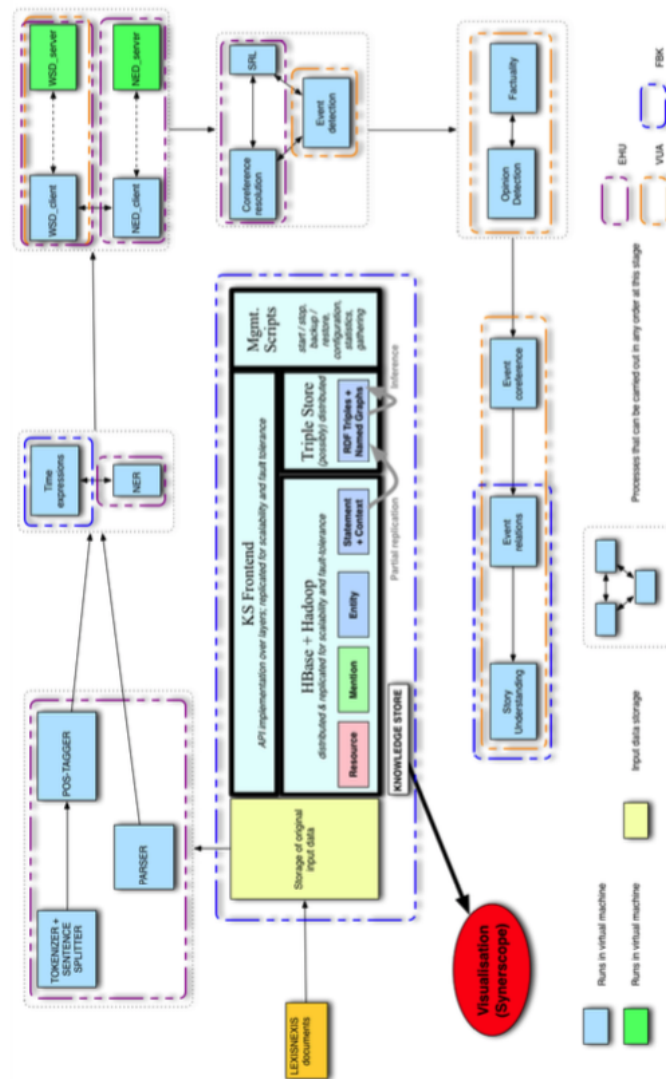


Figure 9: Overall software architecture

1.3.2 Interoperable Natural Language Pipelines in four languages

We created *reading machines* in four languages that can be downloaded as Virtual Machines and deployed for parallel processing. Each reading machine consists of a pipeline with many modules for processing textual sources, ranging from tokenisation and grammatical analysis up to detecting entities, linking these entities to databases, detecting time expressions and normalising them to ISO dates, detecting events and relations between actors, time and events.



Figure 10: Overview of the English pipeline

We tested the main modules on standard data sets comparing them to the state-of-the-art. For all four languages, our systems perform at or above state-of-the-art levels. This is remarkable since the NewsReader pipelines have not been trained specifically for these testing data sets. This also implies that NewsReader could perform relatively stable across general documents or news, while there is sufficient room for further improvement and adaptation to other domains. Even more remarkably, the performance for the Spanish, Dutch and Italian pipelines is similar to the English performance while having less appropriate linguistic resources and annotated datasets.

1.3.3 Cross-document and cross-language event modelling

Although each NewsReader pipeline is different, the output is interoperable across the different languages. Entities are mapped to English DBpedia entries (URIs), dates are normalised to ISO dates (e.g. *yesterday* will be interpreted as a date) and even events are

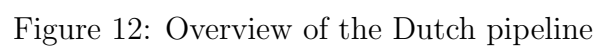
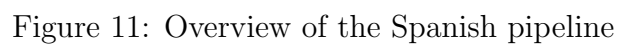




Figure 13: Overview of the Italian pipeline

mapped to a shared ontology across the languages (see Subsection 1.3.7 below). As a result of that, we are capable of representing the pipeline output of the four languages in the same SEM representation using the same language independent approach. We developed software that can translate the interoperable interpretations to these SEM representations following the IDAP method. The module first extracts event data from a single source document by detecting mentions of the same events and entities. Figure 14 illustrates how *coref* relations connect mentions referring to the same event in a single source thus gathering information on the event that is spread throughout the document. After aggregating the event data into a Composite Event representation, we then compare these representations across different sources. When comparing events, we abstract from the way the information is expressed, i.e. we do not base our comparison on the exact words that are used to describe the event, but rather on the kind of event they refer to as modelled by our language-independent ontology. Following this approach, we can merge information within and across sources and as well as across languages.

The granularity of identity can be adapted depending on the need of the users and the type of data that is processed.

1.3.4 Manually annotated evaluation data

We created two unique data sets (MEANTIME and ECB+) to test the semantic processing of text. Both are freely available and are already used by other researchers outside the consortium. In Figure 15, we show the MEANTIME corpus that was manually annotated for many of the semantic layers in NewsReader. We translated the English originals to other languages and annotated the translations in the same way as the English sources. The

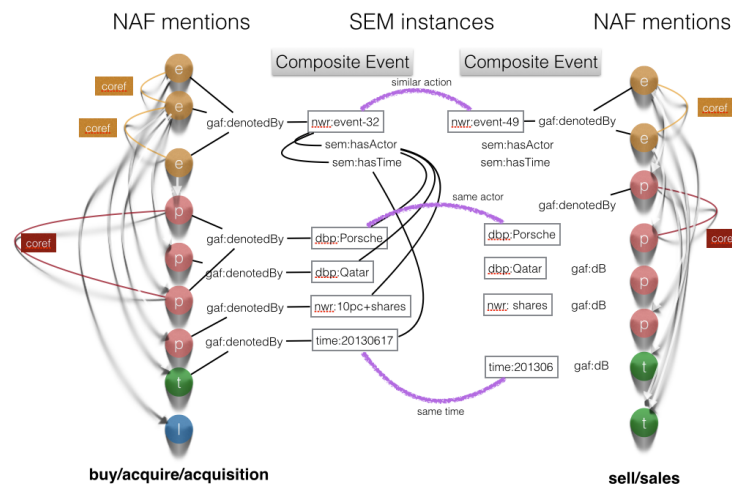


Figure 14: Interpreting mentions as instances, applying IDAP

data set is unique because it combines many semantic annotations, contains annotations across documents and annotations across languages.

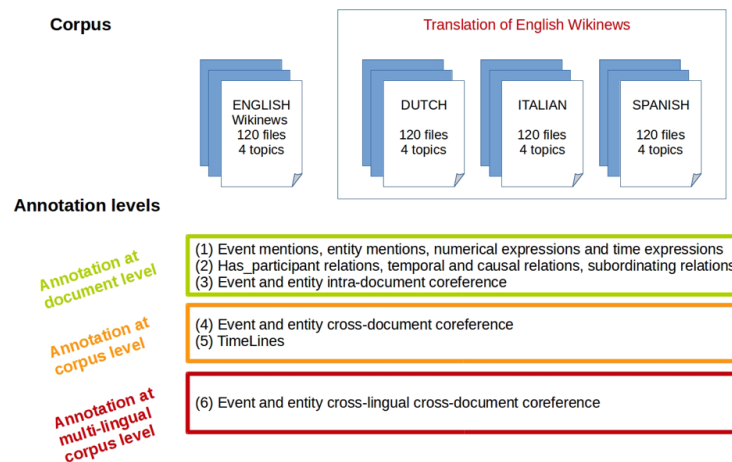


Figure 15: MEANTIME benchmark dataset with annotations across 4 languages

We also created the ECB+ data set. This is an extension of the original Event Coreference Bank (ECB) developed at Stanford University for the purpose of cross-document event-coreference. The cross-document coreference task consists of determining that two different documents make reference to the same event in the world. We added more seminal events to the data set to make the task less trivial. This has already led to publications by other researchers testing on our data.

1.3.5 KnowledgeStore technology

The KnowledgeStore is at the heart of the NewsReader system. It is a scalable database platform that can handle a variety of data streams and the relation to the interpretation of these data streams in the form of RDF triples. It allows for reasoning and inferencing on the data, possibly combined with background data. Figure 16 shows an overview of the architecture.

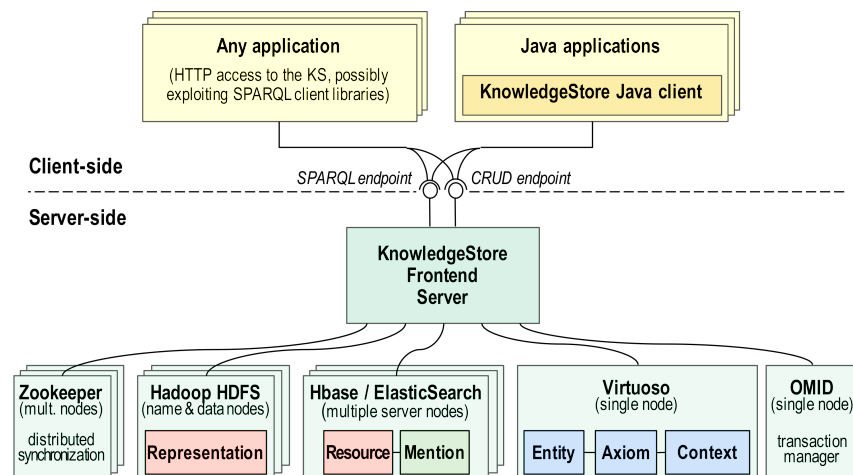


Figure 16: KnowledgeStore architecture

During the project, we populated the KnowledgeStore with massive data sets and its performance was thoroughly tested through a series of hackathons where hundreds of thousands of queries were fired by several teams of developers. We demonstrated that the KnowledgeStore performed well during these stress tests.

1.3.6 Processed data and Event Centric Knowledge Graphs

In addition to the software, NewsReader processed massive streams of news (millions of articles) thus generating large and rich data sets. In Table 1, we provide an overview of the data sets that were all loaded into the KnowledgeStore. We list the number of articles processed, the number of mentions of things (events and entities) and the actual number of instances, where the entities are subdivided into persons, organisations and locations. We also show how many have been mapped to DBpedia and how many have not. The latter make up a large proportion of the data making them an important object of research. We call *dark entities* as there is no background information about them available. The table also shows that we can apply the system to different domains without any adaptation. The final rows show the number of statements (Triples) for each data set. These statements are divided into background knowledge (from DBpedia) and those based on the mentions in the news (from Mentions). In the case of the largest data set, 2.3 million English articles on the Automotive Industry, we see that more than a billion statements are derived and

stored. This data set represents a massive history of the industry during the financial crisis over a period of more than 10 years.

1.3.7 Modelling events and their implications

NewsReader creates event-centric knowledge graphs or ECKGs. These ECKGs represent the changes in the world on which the news reports. It is very important to model these changes properly. We therefore developed the Event and Situation Ontology or ESO. The ontology captures the main types of events that occur in the automotive industry data set as shown in the hierarchy of Figure 17.

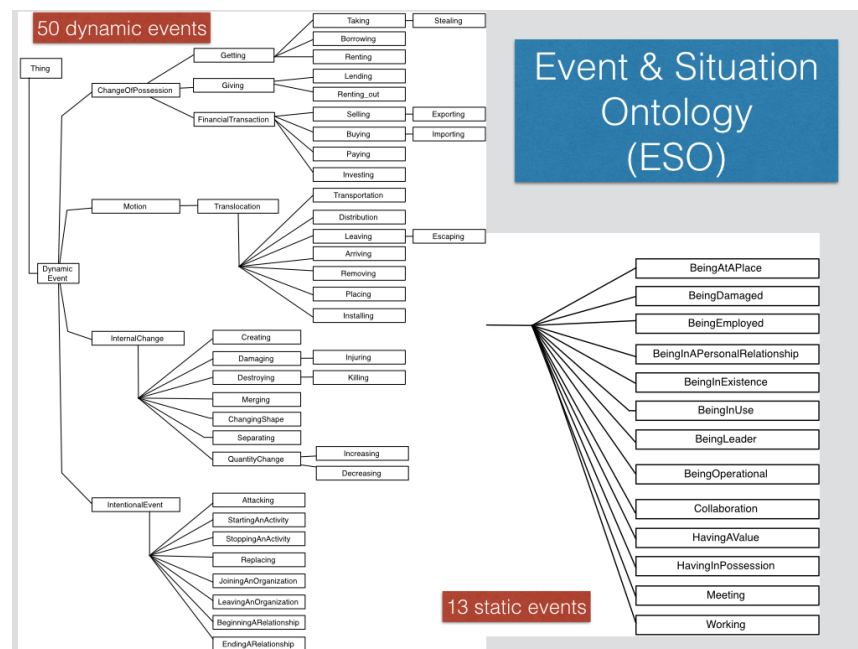


Figure 17: Event and Situation Ontology hierarchy

The hierarchy in ESO not only captures the most important events, it was specifically designed to formally model the implications of the events. An event that happens at a point of time, implies that something changed, e.g. *ownership* in case of *selling* or *buying*, or *employment* relations as a result of *hiring* or *firing*. In Figure 18, we show how ESO captures these implications for the involved actors. ESO makes explicit what entities are affected how and when by the events reported in the news. This allows us to reason over the changes and create timelines of changes for specific individuals.

The ESO hierarchy has been mapped to other well-known event ontologies as well as to wordnets in the different languages. Through the PredicateMatrix, another major resource developed in the project, we are capable of connecting ESO to the words in the 4 languages. The reading machines in NewsReader thus can detect what ESO event is

Table 1: Data processed during the NewsReader project. Numbers are obtained from the KnowledgeStore. Some data sets were processed several times.

	MEANTIME		WikiNews (Ver. 2)		FIFA WorldCup	Dutch Parliament	Cars (Ver. 3)
Topic	General News	General News	General News	Sport, Football	Financial crisis	Automotive Industry	
Period	2003-2015	2003-2015	2004-2014	2004-2014	around 2008/2009	2003-2015	
News Providers	wikinews.org	wikinews.org	LexisNexis	LexisNexis	Dutch House	LexisNexis	
Language	English	English	BBC, The Guardian,	BBC, The Guardian,	of Parliament	LexisNexis	
Populated in	October 2015	English	English	English	Dutch	English	
Pipeline Version	3.0	3.0	October 2015	May 2014	June 2015	October 2015	
			3.0	1.0	1.0-dutch	3.0	
News Articles	120		19,755	212,258	597,530	2,316,158	
Mentions	35,237		5,206,202	76,165,114	9,231,113	842,639,827	
Event instances	3,333		632,704	9,387,356	5,383,498	42,296,287	
Entity instances	339		40,314	858,982	111,579	2,263,156	
Persons	82		17,617	403,021	43,546	895,541	
in DBpedia	46		10,784	40,511	13,942	126,140	
Organizations	172		14,358	431,232	44,139	1,139,170	
in DBpedia	115		4,940	15,984	12,907	44,458	
Locations	85		8,339	24,729	23,894	228,445	
in DBpedia	81		7,369	16,372	11,167	76,341	
Triples	95,219,534		110,861,823	240,731,408	188,296,316	1,240,774,944	
from Mentions	1,046,544		16,688,833	136,135,841	65,631,222	1,146,601,954	
from DBpedia	94,172,990		94,172,990	104,595,567	122,665,094	94,172,990	
distilled from	DBpedia 2015		DBpedia 2015	DBpedia 3.9	DBpedia 2014	DBpedia 2015	

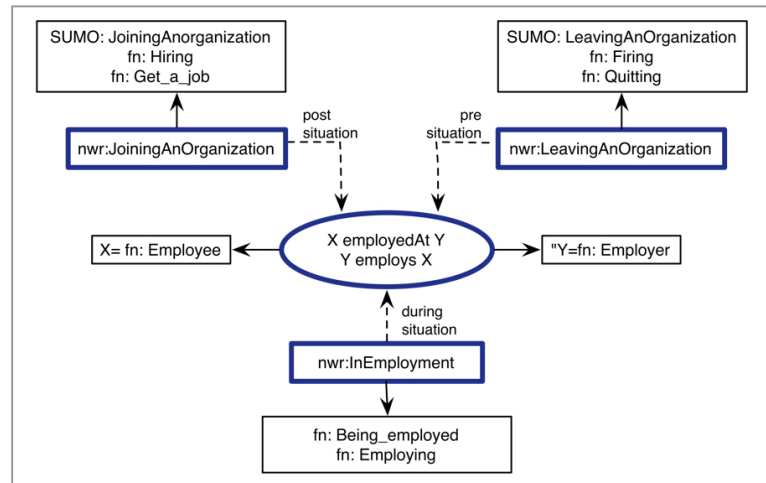


Figure 18: Pre and post-situations modeled in ESO

mentioned and what the ESO roles are of the actors in the event. Through the ontology, the KnowledgeStore can then infer the implications of the ESO situations for individuals.

1.3.8 NewsReader intelligence

The semantic data structures and ECKGs produced by NewsReader are very powerful if it comes to finding information, detecting trends, receiving notifications on unexpected developments or high-impact events and observing connections between people, organisations and events. The data sets contain millions of people and organisations as well as millions of events. Obviously, you can search for a specific person such as the Porsche CEO *Wiedeking* involved in a specific event such as being *sued*. You may find out that this indeed happened and was described in certain news articles at a specific point in time. But if you do not know what happened you also do not know what questions to ask.

However thanks to the ontologies in NewsReader, you can also ask more general questions such as all key persons or CEOs being involved in any court examination at any moment in time. This will give you the complete set of events and their reports in the news, including the case in which *Wiedeking* was *sued*. This is possible because NewsReader interprets the data using its background ontologies such as ESO for events and DBpedia for entities, as is shown in Figure 19 by the red dotted lines at different abstraction levels.²

²Since a large proportion of the entities is not in DBpedia, we also show a DBpedia version here with the dark entities and knowledge that NewsReader can derive for the entities from the news.

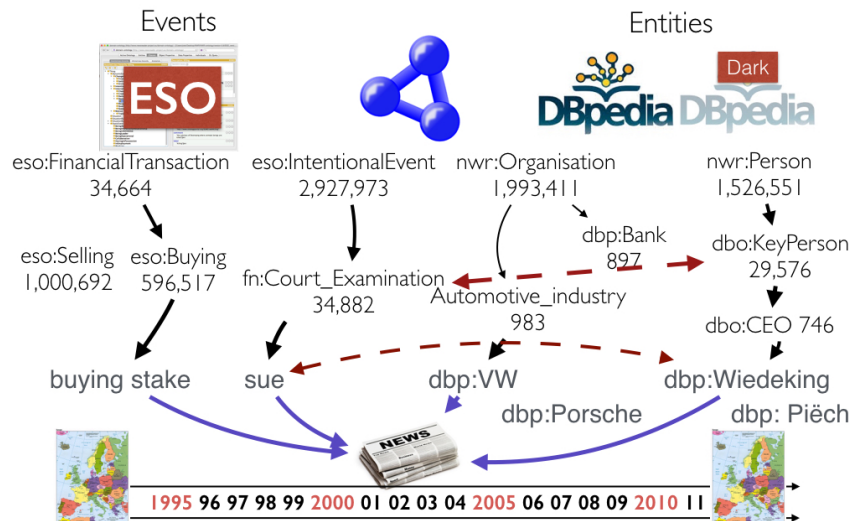


Figure 19: Semantic search in NewsReader using ontological classifications of event and entity instances

Although our data sets contains millions of events, people and organisations, the ontologies define intermediate levels to generalise over these individuals in many different ways. We can use this to observe more general trends (e.g. increase of lawsuits involving CEOs over time) and get complete overviews but we can also use this to find single events that are 'hidden' in the massive data. Hidden events are events that have once been reported in the news but that we are not or no longer aware of. These events can be still very critical for professional decision makers but are difficult to find in the data, e.g. a decision to acquire a company may depend on knowing with whom they did business in the past. On the basis of the semantic search, we designed the so-called *the triple haystack method* to discover such hidden events in the NewsReader data. The method is based on the assumption that the *need for news* can be roughly divided into three specific questions:

Relevant impact event of today: Everyday there is a massive stream of news but not all of it is interesting and there is too much to follow. The first problem people have is to monitor this stream of news and find the bits that matter to them.

Anybody involved in an impact event: Assume you happen to hear about an impact event, the second question could be to find everybody that is involved so that you can trace their history and role in relation to the impact event.

Hidden events that explain the impact event: Assume you know the names of the people and organisations that are involved you need to get an overview of all the

past events and see how they connect. This history may contain information that was published a long time before but nobody realised its value and importance for the impact event of today.

The answers to these three questions correspond to three needles that need to be found in three haystacks. Figure 20 represents this situation. We see a haystack to the right that represents 1 million events in today's news, one of which can be crucial. We see another haystack in the middle with 1 million people in the database of which one may be important and involved in today's news. Finally, the haystack at the left represents all the 42 million events from the past. One or some of these events may be crucial in connection to the person in the second haystack and today's event in the first haystack.

How to find these 3 needles? You can start the search process with the middle haystack and list all the people or organisations you care about, in which case you simply pick your first needle and then look for events with impact that they are involved in from the daily news: the right-most haystack. Alternatively, you may first look for impact events in general in the daily news regardless of who is involved and then look who is involved. In either case, it is important to know what are the events with impact.

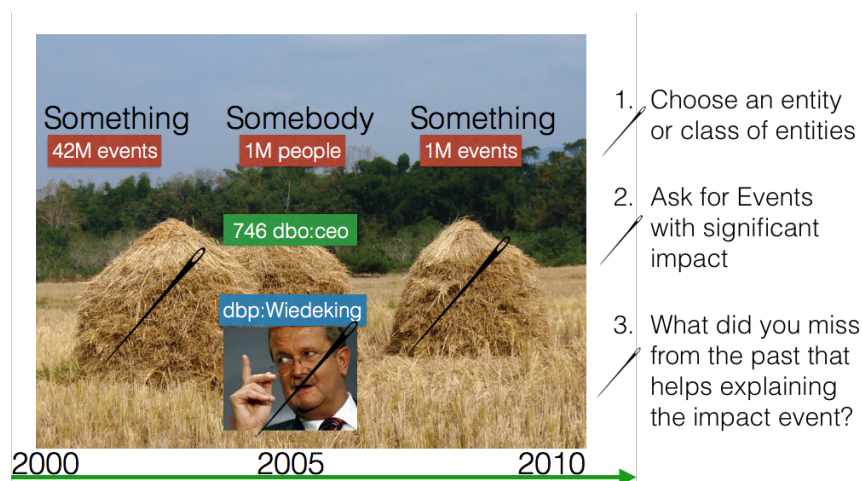


Figure 20: Triple hay stack method to find hidden events with impact

There are various 'classical' retrieval and alert solutions to help finding the first two needles. A traditional method is to trust the newspaper editor who decides to put certain news on the front page and other news not. A more modern method is to make a profile of your interest, e.g. *CEOs*, *Wiedeking*, *legal events*. Whenever there is impact news that matches your profile, you will get an alert.

The impact of an event can then be measured by considering the volume of news, microblogs or queries (compare Google trend) or the strength of the sentiment in social media.³ An example of a news tracking solution is the European Media Monitor. Figure

³We can also check structured data such as stock exchange values or financial business news to see

21 shows a screen dump of their Newsbrief that tracks trending topics over time, based on volume and clustering. The more trending they are, the more they will stand out.



Figure 21: European Media Monitor Newsbrief tracking trending topics in the news. Every coloured line is a separate topic for which the volume of news within a topic cluster is measured by hour. Topics go to sleep around midnight and tend to wake up in the morning.

These classical solutions can only work if the news is somehow spread and there is some measurable activity as a response (e.g. tweets). Typically, the topics in Newsbrief *go to sleep* with the people whose activity is measured (i.e. news providers) and topics *wake up* again with these people in the morning. This shows that these solutions measure the impact of the news on the wider crowd and not the impact of the event on the people directly involved. As such they measure the talk about the event and not the change implied by the event. NewsReader's ESO can also deal with the latter. In Figure 22, we show how pre- and post-situations of many events can be interpreted as positive and negative changes with respect to the condition of the participant of the event. In this way, we model for example that companies or markets get better or go down over time.

In Figure 23, we show that we can use the ESO model to trace the volume of events reported with negative and positive impact on Volkswagen and Porsche over time. A concentration of such events in time may point to a *needle*. This can then be used to find who else was involved (a *second needle*). Likewise, our software can trace all negative-impact events involving CEOs to find the fact that *Wiedeking* is being *sued*.

If we assume that the first and second needle are detected, NewsReader can reconstruct alarming changes that are not expected. Once observed, we can start digging into news for explanations

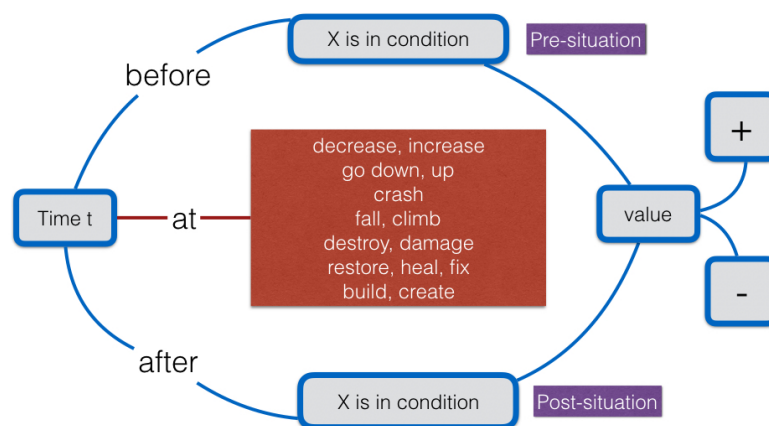


Figure 22: Positive and Negative impact of events modeled in ESO

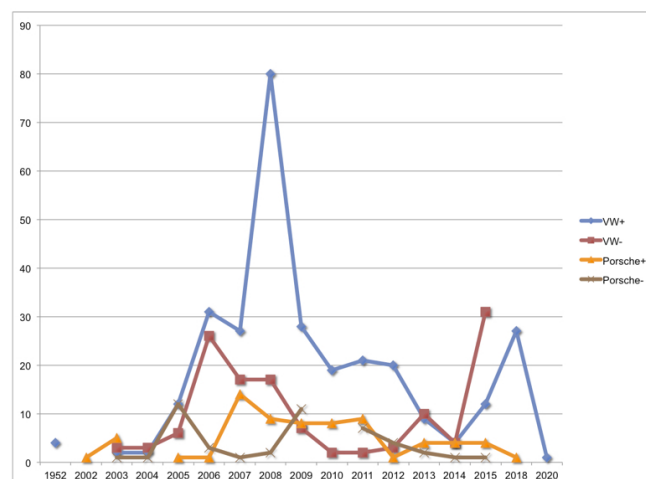


Figure 23: Positive and Negative impact events of VW and Porsche measured through in ESO

the past for the involved entity or entities leading up to the impact event, revealing what happened before (a *third needle*). For this we defined a computational storyline model, in which the impact event is considered the climax in the story and other events are connected to this event through so-called bridging relations. Bridging is achieved when events on a timeline share participants (e.g. all involve *Wiedeking*) and have some causal or coherence relation. Our model allows us to use thresholds and types of bridging relations to create different stories: short, long, tightly or loosely connected, big or small. In the next subsection, we show a visualisation of these storylines that helps finding 'hidden events' connected to the high-impact climax event. Note that NewsReader can also create these storylines on top of classical solutions of detecting relevant impact events in the news such as Newsbrief.

Concluding, we defined NewsReader's intelligence in terms of the semantic capabilities as follows:

- NewsReader can find *associations* between events and entities as needles in haystacks.
- NewsReader cannot always find *correlations*, but can support data scientists in finding them
- NewsReader cannot decide on *legal liability* nor provide any *scientific proof*
- NewsReader does not know what is *true or false* but can show what sources claim
- NewsReader is good at helping to *find a story* that may reveal so-called 'hidden events'
- NewsReader can *tell many different stories* from the data that is extracted

1.3.9 Data visualisations and interactions

Synerscope, member of the NewsReader consortium, is a start-up company that is specialised in visual tools for interacting with large and complex data sets. Their key asset is a series of views on data that are fully connected. Figure 24 shows an overview of different views supported. Any selection or filtering in a single view is projected on all the other views. This makes it possible to simultaneously analyse complex data represented in separated visualisations, each specialised in analysing a different aspect.

Synerscope adapted their tool to the event structures of NewsReader. Figures 25 and 26 show screen dumps of the tool showing social network graphs based on actors in thousands of events, event data plotted on maps and timelines and event data as word clouds. The Synerscope tool was used in a series of end-user-evaluations by professionals. The results of this are reported in user-studies but also raised a lot of interest from the participants for the use of NewsReader.

Whereas the primary data unit of NewsReader is the event, the ultimate goal is to derive narrative sequences of events that exhibit some storylines. Stories are explanatory structures that help us to understand the changes in the world. Whereas individual news

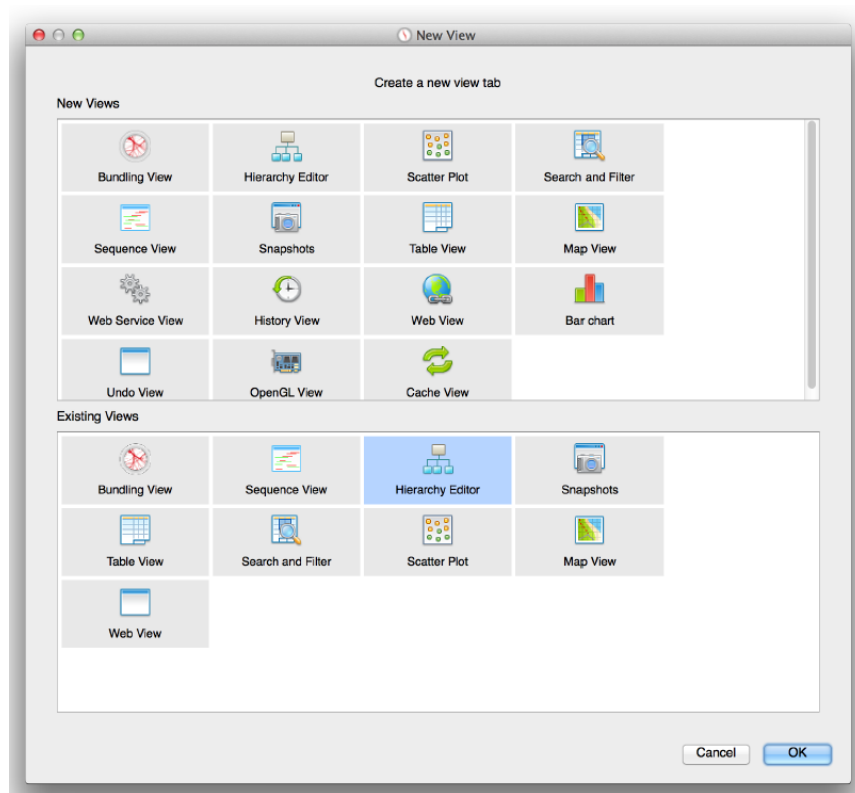


Figure 24: Multiple data views in Synerscope

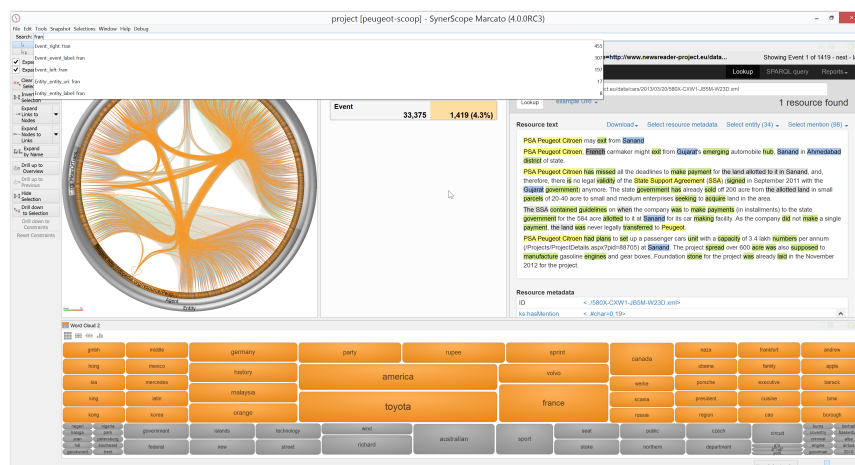


Figure 25: Social network of actors in the automotive industry visualised in Synerscope



Figure 26: Other views in Synerscope

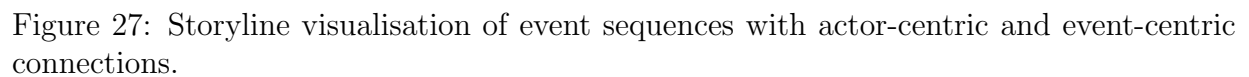
articles tell only part of the story, the NewsReader database may contain many stories that have not yet been discovered. We thus developed the Storyteller tool to visualise sequences of events around storylines, where we assume that a story contains at least one climax event with impact, preceded by events that lead up to it and followed by events that are the consequence.⁴ In Figure 27, we show a screen dump of the Storyteller. It loads ECKGs generated by NewsReader and structures sequences of events to approximate stories using actor and topic relations. The upper part of Figure 27 is actor-centric. It lists all the actors or entities that participate in events and for all each events. Each line represents the events in which the actor is involved and shared events result in intersecting lines. The most 'intersecting' actor is at the bottom of the figure. The middle part is event-centric: each row represents a group of events that approximate a story. The largest circle is the climax event and others are grouped in the same row as preceding and following events based on shared actors and topics. The bottom part allows you to select parts of the data, based on a period or the degree of climax.

1.3.10 Hackathons and end-user-evaluations

The project organised four hackathons in which many different types of users participated, ranging from individuals, to start-ups, to big firms and organisations. ScraperWiki developed the SimpleAPI using the REST protocol and JSON formats to lower the threshold for developers to use the technology.⁵ Participating groups could either directly query the KnowledgeStore or use the SimpleAPI to query the data. Most participants managed to understand the basics of the project, design an application and build a mock-up system within less than a day. We received many very enthusiastic responses from the participants.

⁴The tool can be tested online: <http://www.newsreader-project.eu/results/demos/json-timeline-structure/>

⁵www.newsreader-project.eu/tag/simple-api



All the project results are available as open source, with source code and resources downloadable from Github. Results are also available as Virtual Machines, either binary or installable from source code through scripts.

1.4 Potential impact

From a scientific point of view, the project has set some important steps in the area of semantic processing of text: making a fundamental distinction between mentions and instances in our models, creating an effective collaboration between natural-language processing and semantic web paradigms, deriving more complex event structures such as timelines and storylines. We developed many high-level semantic modules that perform at and beyond state-of-the-art level, not only for English but also for Italian, Spanish and Dutch. We created unique data sets for evaluating these tasks and defined the next level for cross-lingual semantic processing for the research field. We also modelled event data in RDF, including the formal representation of the implications of these events and a reasoning component to derive these from the text-based event representations. From a technical point of view, we have shown that we can arrive at interoperable semantic interpretations across languages, that we can efficiently process millions of news articles using parallel processing and store billions of triples in the KnowledgeStore in combination with background knowledge. We have shown that this massive amount of data can be used in several hackathons by different groups of developers, firing massive queries to the same data hub. We have also shown how to provide access to the large and complex data sets through visualisations.

The scientific and technical success is illustrated by the more than 120 peer reviewed publications but also by the many (20) follow up projects that are using NewsReader technology or are building on it and which are briefly summarised below:

Spin off projects

1. **Investigating Criminal Networks:** VUA is the project coordinator of this project that aims to develop understandable and useful visualisations of news story lines around illicit trade of humans, wildlife and drugs. The NewsReader pipeline is utilised to perform linguistic and semantic analysis on the data sources in this project. SynerScope is also involved as a partner in this project and they are further developing and testing their visualisation tool for this project (results of which are taken into account in the NewsReader project). Various LN clients have participated in a presentation/round table of the KIEM results on June 23rd. Participants mainly came from Government bodies, including various Ministries, as well as the Tax Department and National Police. (KIEM Trafficking: “Building structured event indexes of large volumes of financial and economic data for decision making”): NWO project number 314-98-030 (2014-2015).
2. **Medical Trust Networks:** In this project, the VUA computational linguistics research group collaborates with the social sciences faculty to investigate the detection

of belief systems in online medical forums. Online debates on for example vaccination are intense and it can be difficult to establish the trust relationships between the government and science institutions, and citizens. This project uses the NewsReader pipeline to process the news resources as well as the attribution model to analyse the content of the extracted beliefs. VU Network Institute funding (2014-2015).

3. **Mining Causal Graphs from Patient Records:** In this project, the VUA computational linguistics research group collaborates with the semantic web group in the computer science department to develop tools for automatically extracting symptoms, diagnoses and treatments from electronic patient records. This project serves as a preliminary investigation for comparing current medical practice with clinical guidelines in order to study the extent to which guidelines differ from current practice and perhaps to suggest updates to the guidelines based on observations from the field. The tools in this project are based on the NewsReader processing pipeline and attempts to adapt them to the medical domain. VU Network Institute funding (2014-2015).
4. **Political Discourse in the News:** In this project, the VUA computational linguistics research group collaborates with the department of communication science to explore the interaction between the news and parliamentary debates. The main purpose is to examine changes in how media talks about politics, whether they focus more on issues, political parties or individuals. This project focuses on the relation between media coverage and parliamentary discourse, using linguistic techniques such as topic modelling, grammatical analysis and event extraction. NewsReader tools, such as the opinion miner, are used to identify relevant information. VU Network Institute funding (2014-2015).
5. **BiographyNet:** The BiographyNet project aims to extract relations between historical people and events from over 125,641 biographies on 80,000 different people collected in the Dutch Biography Portal of the Netherlands. In the project, computational linguists, semantic web researchers and historians work together to develop automated tools for discovery of relations between people and events. The project uses the NewsReader Dutch pipeline as well as the grounded annotation framework to process text and organise its extracted information. Netherlands eSciencecenter (NLeSc) funding (2012 - 2016).
6. **Text Mining (undergraduate course):** VUA uses examples from the NewsReader project to teach 70 undergraduate students yearly about the state-of-the-art in linguistic processing. During the lab sessions of the course, modules from the NewsReader processing pipeline are used for the students to obtain hands-on experience with real tools.
7. **The borders of ambiguity:** This project aims to define the extent of the closed world of language problem and the optimal solution given the vast volumes of textual

data that are available by treating it as a system of language relations. The ULM-1 project builds upon the NewsReader processing pipeline. Currently, the project is using the tokeniser, pos-tagger, NERC and NED modules from NewsReader and further developing the word-sense disambiguation module. NWO Spinoza prize (2014 - 2019).

8. **Stories and world views as a key to understanding language:** The project aims to investigate how the interpretation of newswire can be used to model changes in the world. The assumption here is that the changes in the world are long-term stories concerning human intentions and goals that are expressed through texts about the world. The project builds upon the NewsReader processing pipeline, grounded annotation framework (GAF) and attribution model. NWO Spinoza prize (2014 - 2019).
9. **A quantum model of text understanding:** This project aims at finding a new model of natural language processing. Currently, most NLP architectures are set up as pipelines, meaning that early module can provides an erroneous result as input to a later module, effectively causing errors to propagate. This project investigates whether the pipeline model can be replaced by a model where later modules can influence earlier modules. The project uses the NewsReader processing pipeline as a starting point for their experiments. NWO Spinoza prize (2014 - 2019)
10. **INclusive INsight.** In this project, VUA is collaborating with an information management company and an entrepreneurial advisory company specialised in sustainable trade to improve the organisation of data resources around the production of coffee, tea and cocoa. The NewsReader pipeline is used to perform linguistic analyses on the resources to build search indexes. Understanding the sustainability network (KIEM project creative industry NWO Project number 314-98-031).
11. **Reading Between the Lines:** NewsReader team member Antske Fokkens received a personal research grant from the NWO. The project will develop methods to identify the more subtle perspectives and worldviews in text (e.g. do papers use different labels when they describe similar events involving different ethnicities?). Particular focus lies on how this methodology can be used by researchers from other domains (such as historians and communication scientists). The English and Dutch NewsReader pipelines provide the basis for the technologies used in this research. Naturally, GAF and the NewsReader attribution model will be used to represent alternative perspectives. NWO VENI grant (2016-2018).
12. **CLARIAH semantic annotation:** VUA is a consortium partner in the CLARIAH project, a large road map project funded by the NWO. This project aims to continue and enlarge the digital infrastructure for the Humanities. In this project, VUA is responsible for converting textual data sources and making them interoperable with structured data sources. Here the GAF model developed in NewsReader can

be applied, as well as various elements of the processing pipeline. NWO roadmap funding: (2015-2018).

13. **Visualising Uncertainties And Perspectives:** This project funded by the NLeSc and a collaboration between this institute and historians and computational linguists from the VU. The goal is to model and visualise uncertainty and perspectives based on automatic analysis of text. This includes both the uncertainty and various perspectives expressed in text as well as the uncertainty and different perspectives that result from (different) approaches used for automatic analyses. NewsReader forms a primary use cases and the project builds upon the work that has been done within NewsReader to represent perspectives, including GAF. Netherlands eScience Grant.
14. **Modelling Perspectives in Philosophy:** This project is a collaboration between the computational linguistics group at the VU and prof. dr. Arianna Betti from the University of Amsterdam's Institute for Language, Logic and Computation. It investigates methods for extracting and interpreting information about perspectives as expressed in philosophical texts in a computational way. The work done in NewsReader on modelling of and tooling for event factuality and sentiment and opinion mining serve as a starting point here. A Computational Experiment on Quine's Word & Object: VU Network Institute funding (2015-2016).
15. **Identifying implicit stereotypical views in natural language through automatic linguistic analyses:** This project is a collaboration between the computational linguistics group and dr. Camiel Beukeboom who is involved with the departments of communication science and social psychology at VUA. The project aims at determining which linguistic cues express stereotypes in text. The next step is to investigate how these cues can be identified using automatic analyses. The NewsReader pipeline will be used as a basis for this study. VU Network Institute funding (2015-2016).
16. **Ber2Tek, SKATER, LiMoSINE and QTLeap:** EHU is part of various research projects in which has collaborated in different aspects. As a consequence, the ixa-pipes, the NAF representation and the Predicate Matrix is now used in the following projects: Ber2Tek, SKATER, LiMoSINE and QTLeap. In addition, Freeling, an open source language analysis tool suite, offers the option of using the NAF representation and the Predicate Matrix. Finally, as a result of the collaboration in the SKATER project, the ixa-pipes includes now the Galician language.
17. **HAP-LAP and EMLCT Master programs (2014-):** EHU uses examples from the NewsReader project to teach master students about the state-of-the-art in linguistic processing. During the lab sessions of the course, modules from the NewsReader processing pipeline are used for the students to obtain hands-on experience with real tools.

18. **Digital Humanities:** Researchers at the LATTICE lab (CNRS, ENS, Paris 3) used the IXA Pipes Semantic Role Labeling and Coreference modules for Digital Humanities. They analyse a corpus on international climate negotiations (the Earth Negotiations Bulletin), in order to identify negotiation points supported or opposed by negotiating parties. The aim of the analysis is to help social science researchers understand participants' positions in the negotiation. The corpus was a subset of volume 12 of the Earth Negotiations Bulletin, and consisted of about 250 daily reports on the COP summits that have taken place yearly since 1995. The period analysed was 1995 to 2014.
19. **Understanding Multimedia Content:** This is an explorative FBK funded project that aims to define a unified conceptual framework for extracting, aligning and integrating knowledge extracted from different media, such as textual resources, images, audios and videos. Several NewsReader technologies (i.e. Italian and English pipelines, PIKES, KnowledgeStore) are used in the project and will be extended to cope with multimedia content. FBK funded (2015 - onward).
20. **DECIPHER:** FET proposal (involving VUA, FBK, EHU) submitted in September 2015 for multichannel knowledge-to-data processing. Under review.

With respect to the industrial and application objective, the project has received very positive feedback from the end-user sessions in which various professional participated. They had to carry out tasks to find complex and elaborate answers using the visualisation and interaction tool Synerscope in a data set spanning more than 10 years of news (more than 2.3 million articles). Our active dissemination (166 presentations given) and the success of the project already resulted in 18 concrete collaborations with commercial and semi-commercial organisations, shortly summarised below:

1. **Dasym** Dasym, a Dutch investment company, is the first company that takes the plunge with an installation of the complete NewsReader system at their company to monitor companies and industries, and to track opinions and events with potential impact. This project is privately funded and involves installation of the NewsReader system and training of Dasym staff members. (Jan - Sept 2016)
2. **OntoRadiology:** Through a small project funded through the VU Amsterdam Research Fellow programme, VUA is looking into adapting the NewsReader pipeline to the radiology domain. In this domain, many reports are available in the form of unstructured, transcribed text. Through the NewsReader pipeline, these will be converted to structured data. (Jan - Sept 2016)
3. **Spanish Ministry of Industry:** EHU has started a fruitful collaboration with the Spanish Ministry of Industry to set up a domain adapted pipeline for English and Spanish to help a technology watch system. For that, members from EHU and representatives and technical staff from the ministry have been working together and meeting in a regular basis under a funded project. The technical staff from

the ministry have downloaded and used the English and Spanish pipelines. As a result of this collaboration, a new prototype to test the NewsReader technology in the surveillance domain is running on the ministry's system. It contains a modified version of the NewsReader pipeline where not all the modules have been integrated. The collaboration has continued during the last year of the project.

4. **Dutch House of Representatives:** Members from the information department at the Dutch House of Representatives have met with the NewsReader consortium in various settings to investigate whether the NewsReader technology could aid them in their daily business. Two members of the department have participated in the NewsReader end-user evaluation in Y2 and Y3, and in June, the consortium presented a use case for which >600,000 Dutch documents were processed concerning the financial crisis. This document set was a combination of the House of Representatives internal documents, as well as 'external' news documents, provided by LN.
5. **Dutch National Bank:** VUA met with several members of the staff at the Dutch National Bank to discuss the use of NewsReader technology in their daily business where they analyse for example news on the financial markets.
6. **Commerzbank:** Since November 2015, the NewsReader consortium has had several conference calls and a face-to-face meeting in Amsterdam with representatives of the Risk Department in Commerzbank's Frankfurt HQ on potentially using the NewsReader technology. Interaction with the consortium was triggered by the consortium's presence at the Luxembourg-based European Data Forum that month. The next step the members of the NewsReader consortium are working on is a visit to the Commerzbank offices in Frankfurt to demonstrate the NewsReader toolkit to the bank's decision makers in the Risk department.
7. **Financial Times:** The Financial Times (London) is interested in integrating structured RDF data as outputted by NewsReader into their newsrooms.
8. **VVOJ (Dutch Association of research journalists):** The Dutch Association of Research Journalists invited Piek Vossen to present on the NewsReader project during their yearly conference. They are currently seeking funding for a NewsReader installation that can help research journalists to handle large archives of news.
9. **SURFsara:** SURFsara, a collaborative organisation for ICT in Dutch education and research are willing to host NewsReader pipeline for a long time to offer services to companies, researchers and journalists.
10. **Dutch Central Bureau of Statistics:** The Dutch Central Bureau of Statistics is interested in using NewsReader NLP tools in addition to their structured data analysis.
11. **The Netherlands Authority for the Financial Markets (AFM):** AFM is interested in using NewsReader technology to do market, company and industry analyses.

12. **Trivago:** Trivago is a travel meta search engine focusing on hotels. They are seeking funding for using the NewsReader technology to mine user experiences in the tourism domain.
13. **Underlined:** Underlined is a Dutch company that is looking for reputation monitoring technologies that can deal with fine-grained opinions.
14. **Olery:** Olery is a Dutch startup that would like to enhance their hotel review technologies, currently based on OpeNER, with the latest NewsReader updates. Seeking funding.
15. **Brandweer Amsterdam-Amstelland:** The Amsterdam Fire Department is interested in using natural language processing and semantic web technology to enrich their work maps and procedure descriptions.
16. **Almawave:** This is a project proposal submitted to the Trento province by FBK, in collaboration with Almawave, one of the leading Italian players globally operating in Customer Relations Management. One of the goal of the project is to investigate how NLP and Semantic technologies can improve customer experience, especially in call-center settings. The proposal, currently under evaluation, builds on several NewsReader technologies (i.e., Italian pipeline, PIKES, KnowledgeStore). Project proposal (2016 - onward).
17. **Treccani:** FBK is conducting initial experiments to apply some of the technologies developed within NewsReader (i.e., PIKES, KnowledgeStore) to build an advanced semantic search engine for a cultural web-portal, in conjunction with Treccani, the publisher of the Italian Encyclopaedia of Science, Letters, and Arts.
18. **Euregio:** FBK has an on-going collaboration with Euregio, a company doing business in Media Intelligence. In particular, research activities are being conducted to exploit semantic technologies to effectively monitor media (e.g., TV, radio, press), extracting relevant information to perform detailed socio-political analysis. The Italian pipeline is currently applied in the project, and will be used as basis to build an extraction pipeline also for German.

The potential of NewsReader as an infrastructure stretches beyond the current cases. We are currently considering pipelines for other languages that are interoperable with the current pipelines. We already collaborated with researchers from the Bulgarian Academy of Sciences, Sofia and managed to build a NewsReader pipeline for Bulgarian in a few weeks and started with German. This widens the scope for interoperability at a global scale. The technology can be applied to information extraction and modelling but also to new approaches for Machine-Translation with deep semantic awareness. NewsReader has already been used for historical and social research. We are also talking to scientist in the medical domain, in economy and finance, in the legal domain and religion studies. The main difference between NewsReader and other Natural Language Processing is that we

achieved to reach a semantic level of interpretation through robust and efficient processing of open and free text. Whereas the range of users for non-semantic processing of text (word level or syntactic analysis) is limited to specialists, almost anybody can benefit from technology that takes text processing to a semantic level.

1.5 Consortium details and contact

Project website: www.newsreader-project.eu

Project github: <https://github.com/newsreader>

Project coordinator: Prof. Dr. Piek Vossen, piek.vossen@vu.nl

Partner	Country	Contact	Email
Faculteit of Arts, Vrije University Amsterdam	Netherlands	Piek Vossen	piek.vossen@vu.nl
Euskal Herriko Unibertsitatea San Sebastian	Spain	German Rigau	german.rigau@ehu.es
Fondazione Bruno Kessler Trento	Italy	Luciano Serafini	serafini@fbk.eu
LexisNexis, Amsterdam	Netherlands	Pim Stouten	pim.stouten@lexisnexis.com
ScraperWiki, London	United Kingdom	Aidan McGuire	aidan@scraperwiki.com
SynerScope, Helvoirt	Netherlands	Thomas Ploeger	thomas.ploeger@synerscope.com

Table 2: Consortium members and contacts

2 Use and dissemination of foreground

Section A (public)

This section includes two templates

- Template A1: List of all scientific (peer reviewed) publications relating to the foreground of the project.
- Template A2: List of all dissemination activities (publications, conferences, workshops, web sites/applications, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters).

These tables are cumulative, which means that they should always show all publications and activities from the beginning until after the end of the project. Updates are possible at any time.

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO	Title	Main Author	Title of the periodical or the series	Number date or, frequency	Publisher	Place	Year	Pages	Permanent identifiers	Open Access
1	Building Event-Centric Knowledge Graphs from News	Rospoche, Marco	Journal of Web Semantics				2016		http://dx.doi.org/10.1016/j.websem.2015.12.004	Y
2	Big data for Natural Language Processing: A streaming approach	Agerri, Rodrigo	Knowledge-Based Systems	0			2015			Y
3	The KnowledgeStore: a Storage Framework for Interlinking Unstructured and Structured Knowledge	Corcoglioniti, Francesco	International Journal on Semantic Web and Information Systems	April-June			2015	1-35	http://www.igi-global.com/article/the-knowledgestore/136832	Y
4	The Predicate Matrix and the Event and Implied Situation Ontology: Making More of Events	Segers, Roxane	Proceedings of the Global WordNet Conference (GWC2016)			Bucharest, Romania	2016		http://gwc2016.racai.ro/proceedings.pdf	Y
5	Predicate Matrix. Automatically extending the semantic interoperability between predicate resources	de Lacalle, Maddalen López	Journal of Language Resources and Evaluation				2016			Y
6	Word vs. Class-Based Word Sense Disambiguation	Izquierdo, Ruben	Journal of Artificial Intelligence Research				2015	83-122		Y

7	EHU-ALM: Similarity-Feature Based Approach for Student Response Analysis	Aldabe, Itziar	Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)	Jun 14-15	Association for Computational Linguistics	Atlanta, GA, USA	2013		http://www.aclweb.org/anthology/S13-2097	Y
8	Interlinking Unstructured and Structured Knowledge in an Integrated Framework	Corcoglioniti, Francesco	7th IEEE International Conference on Semantic Computing (ICSC2013)	Sept 16-18		Irvine, CA, USA	2013		http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6693492&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6693492	Y

9	Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution	Cybulska, Agata	Proceedings of Recent Advances in Natural Language Processing (RANLP-2013)	Sept 7-14	INCOMA Ltd.	Hissar, Bulgaria	2013		http://aclweb.org/anthology//R/R13/R13-1021.pdf	Y
10	Denoting Data in the Grounded Annotation Framework	van Erp, Marieke	Proceedings of the 12th International Semantic Web Conference and the 1st Australasian Semantic Web Conference Posters & Demos Track (ISWC2013)	Oct 21-25		Sydney, Australia	2013		http://iswc2013.semanticweb.org/sites/default/files/iswc_poster_3.pdf	Y
11	Learning with the Web: Spotting Named Entities on the intersection of NERD and Machine Learning	van Erp, Marieke	Proceedings of the #MSM2013 Concept Extraction Challenge	May		Rio de Janeiro, Brazil	2013		http://nerd.eurecom.fr/ui/paper/vanErp_Rizzo_Trancy-msm2013.pdf	Y

12	Offspring from Reproduction Problems: What Replication Failure Teaches Us	Fokkens, Antske	Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)	August 4-7	Association for Computational Linguistics	Sofia, Bulgaria	2013		http://aclweb.org/anthology//P/P13/P13-1166.pdf	Y
13	GAF: A Grounded Annotation Framework for Events	Fokkens, Antske	Proceedings of the 1st workshop on Events: Definition, Detection, Coreference, and Representation at the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL2013)	Jun 9-15	Association for Computational Linguistics	Atlanta, GA, USA	2013		http://aclweb.org/anthology/W/W13/W13-1202.pdf	Y

14	Outsourcing FrameNet to the Crowd	Fossati, Marco	Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013).	August 4-7		Sofia, Bulgaria	2013		http://www.aclweb.org/anthology/P13-2130	Y
15	ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling	Laparra, Egoitz	Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)	August 4-7	Association for Computational Linguistics	Sofia, Bulgaria	2013		http://www.aclweb.org/anthology/P13-1116	Y
16	Sources of Evidence for Implicit Argument Resolution	Laparra, Egoitz	Proceedings of the 10th International Conference on Computational Semantics (IWCS2013)	March 19-22		Potsdam, Germany	2013		http://www.aclweb.org/anthology/W13-0114	
	Y									

17	Frame Semantics Annotation Made Easy with DBpedia	Fossati, Marco	Proceedings of the 1st International Workshop on Crowdsourcing the Semantic Web (Crowd-Sem2013)			Sydney, Australia	2013		http://ceur-ws.org/Vol-1030/paper-03.pdf	Y
18	BiographyNet: Managing Provenance at multiple levels and from different perspectives	Ockeloen, N.	Proceedings of the Linked Science Workshop at ISWC2013				2013		http://linkedscience.org/wp-content/uploads/2013/04/paper7.pdf	Y
19	Extractivism. Extracting activist events from news articles using existing NLP tools	Ploeger, Thomas	Proceedings of the 3rd Workshop on Detection, Representation and Exploitation of Events in the Semantic Web (DeRiVE 2013)			Sydney, Australia	2013		http://ceur-ws.org/Vol-1123/paper5.pdf	Y

20	WordNet: principles, developments and applications	Vossen, Piek					2013		http://www.degruyter.com/view/product/175228?rskey=PczMPQ&onlyResultQuery=International%20Encyclopedia%20of%20Lexicography,%20Supplementary%20volume:%20Recent%20developments%20with%20special%20focus%20on%20computational%20lexicography	Y
21	Wordnet-LMF: a standard representation for multilingual wordnets	Vossen, Piek			Hermes / Lavoisier / ISTE		2013		http://onlinelibrary.wiley.com/doi/10.1002/9781118712696.ch4/summary	Y
22	New Trends of Research in Ontologies and Lexical Resources	Oltramari, A.			Springer-Verlag		2013		http://www.springer.com/computer/database+management+%26+information+retrieval/book/978-3-642-31781-1	Y

23	Proceedings of the 3rd International Workshop on Linked Science 2013 - Supporting Reproducibility, Scientific Investigations and Experiments (LISC2013) In conjunction with the 12th International Semantic Web Conference 2013 (ISWC 2013)	Groth, Paul		10/21/2016	CEUR-WS	Sydney, Australia	2013		http://ceur-ws.org/Vol-1116/	Y
24	Proceedings of the Joint Workshop on NLP&LOD and SWAIE: Semantic Web, Linked Open Data and Information Extraction	Maynard, Diana		September	INCOMA Ltd. Shoumen, BULGARIA	Hissar, Bulgaria	2013		http://www.aclweb.org/anthology/W13-52	Y
25	Proceedings of the third Workshop on Detection, Representation and Exploitation of Events in the Semantic Web (DeRiVE 2013) co-located with The 12th International Semantic Web Conference (ISWC 2013)	van Erp, Marieke		10/21/2016	CEUR-WS	Sydney , Australia	2013		http://ceur-ws.org/Vol-1123/	Y
26	NewsReader project	Agerri, Rodrigo	30th Conference of the Spanish Society for Natural Language Processing ({SEPLN})		SEPLN		2014	215-218		Y

27	IXA pipeline: Efficient and Ready to Use Multilingual NLP tools	Agerri, Rodrigo	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/775_Paper.pdf	Y
28	A stream computing approach towards scalable NLP	Artola, Xabier	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/670_Paper.pdf	Y
29	EuroLoveMap: Confronting feelings from News	Atserias, Jordi	Proceedings of OpeNER Hackathon.			Reykjavik, Iceland	2014		http://adimen.si.ehu.es/~rigau/publications/opener14-aemrs.pdf	Y
30	RDFpro: an Extensible Tool for Building Stream-Oriented RDF Processing Pipelines	Corcoglioniti, Francesco	Proceedings of ISWC Developers Workshop colocated with 13th Int. Semantic Web Conference (ISWC'14), Riva del Garda, Italy	12(68)	CEUR-WS		2014	49-64	http://ceur-ws.org/Vol-1268/paper9.pdf	Y

31	Granularity for Event Coreference Resolution	Cybulska, Agata	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/1103_Paper.pdf	Y
32	Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution	Cybulska, Agata	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/840_Paper.pdf	Y
33	Finding Stories in 1,784,532 Events: Scaling up computational models of narrative	van Erp, Marieke	Workshop on Computational Models of Narrative (CMN'14)	July 31 – August 2		Quebec City, Canada	2014		http://drops.dagstuhl.de/opus/volltexte/2014/4660/pdf/27.pdf	Y
34	Georeferencing Animal Specimen Datasets	van Erp, Marieke	Transactions in GIS				2014		http://onlinelibrary.wiley.com/doi/10.1111/tgis.12110/abstract	N
35	Discovering and visualising stories in news	van Erp, Marieke	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/645_Paper.pdf	Y

36	NAF and GAF: Linking linguistic annotations	Fokkens, Antske	Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Inter-operable Semantic Annotation			Reykjavik, Iceland	2014	9		Y
37	CROMER: A Tool for Cross-Document Event and Entity Coreference	Girardi, Christian	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/726_Paper.pdf	Y
38	Number frequency on the web.	van Hage, Willem Robert	WWW (Companion Volume)				2014	571-572		Y
39	A Simple API to the KnowledgeStore	Hopkinson, Ian	Proc. of ISWC Developers Workshop colocated with 13th Int. Semantic Web Conference (ISWC'14), Riva del Garda, Italy		CEUR-WS		2014	7/12/2014	http://ceur-ws.org/Vol-1268/paper2.pdf	Y

40	From Text to Political Positions	Kaal, B.			John Ben-jamins Publishing Company		2014		http://www.academia.edu/6998238/From_text_to_Political_Positions_The_convergence_of_political_linguistic_and_discourse_analysis	N
41	First steps towards a Predicate Matrix	de Lacalle, Maddalen López	Proceedings of the 7th Global WordNet Conference (GWC2014)	Jan 25-29		Tartu, Estonia	2014		http://adimen.si.ehu.es/~rigau/publications/gwc14-llr.pdf	Y
42	Predicate Matrix: extending SemLink through WordNet mapping	de Lacalle, Maddalen López	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/589_Paper.pdf	Y
43	Generating Polarity Lexicons with WordNet propagation in 5 languages	Maks, E.	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/847_Paper.pdf	Y
44	OpeNER and the automatic generation of sentiment lexicons in five languages	Maks, E.	Proceedings of CLIN-2014				2014			Y

45	Event Factuality in Italian: Annotation of News Stories from the Ita-TimeBank	Minard, Anne-Lyse	Proceedings of CLiC-it 2014, First Italian Conference on Computational Linguistic				2014		http://clic.humnet.unipi.it/proceedings/vol1/CLICIT2014150.pdf	Y
46	FBK-HLT-time: a complete Italian Temporal Processing system for EVENTI-Evalita 2014	Mirza, Paramita	Proceedings of the First Italian Conference on	December 9-10		Pisa, Italy	2014		http://clic.humnet.unipi.it/proceedings/vol2/clicit201428.pdf	Y
47	Annotating Causality in the TempEval-3 Corpus	Mirza, Paramita	Proceedings of the eacl workshop on computational approaches to causality in language (catocl2014)	April 26-30		Gothenburg, Sweden	2014		http://www.aclweb.org/anthology/W14-0702	Y
48	An Analysis of Causality between Events and its Relation to Temporal Information	Mirza, Paramita	Proceedings of the 25th international conference on computational linguistics (coling2014)	August 23-29		Dublin, Ireland	2014		http://www.aclweb.org/anthology/C14-1198	Y

49	Classifying Temporal Relations with Simple Features	Mirza, Paramita	Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL2014)	April 26-30		Gothenburg, Sweden	2014	http://www.aclweb.org/anthology/E14-1033	Y
50	CAT: An Advanced Environment for manual annotation of text and corpora	Moretti, Giovanni	Proceedings of the 35th Conference of the International Computer Archive of Modern English (ICAME2014)	April 30 - May 4		Nottingham, UK	2014		Y
51	What implementation and translation teach us: the case of semantic similarity measures in wordnets	Postma, Marten	Proceedings of the 7th Global WordNet Conference (GWC2014)	Jan 25-29		Tartu, Estonia	2014	https://www.aclweb.org/anthology/W/W14/W14-0118.pdf	Y
52	Benchmarking the Extraction and Disambiguation of Named Entities on the Semantic Web	Rizzo, Giuseppe	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014	http://www.lrec-conf.org/proceedings/lrec2014/pdf/176_Paper.pdf	Y

53	Inductive Entity Typing Alignment	Rizzo, Giuseppe	Linked Data for Information Extraction 2014 (LD4IE 2014).	Oct		Riva del Garda, Italy	2014		http://ceur-ws.org/Vol-1267/LD4IE2014_Rizzo.pdf	Y
54	Integrating NLP and SW with the KnowledgeStore	Rospoche, Marco	ISWC 2014 Posters & Demonstrations Track, within the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 21, 2014		CEUR-WS.org		2014	69-72	http://ceur-ws.org/Vol-1272/paper_56.pdf	Y
55	Integrating Unstructured and Structured Knowledge with the Knowledge-Store	Rospoche, Marco	Proceedings of the Posters and Demos of the 19th International Conference on Knowledge Engineering and Knowledge Management (EKAW2014)				2014		http://ceur-ws.org/Vol-1272/paper_56.pdf	Y

56	Hope and Fear: Interpreting Perspectives by Integrating Sentiment and Event Factuality	van Son, Chantal	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/188_Paper.pdf	Y
57	Crowdsourcing for the identification of event nominals: An experiment	Sprugnoli, Rachele	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/1157_Paper.pdf	Y
58	Simple, Robust and (almost) Unsupervised Generation of Polarity Lexicons for Multiple Languages	San Vicente, Iñaki	Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL2014)	April 26-30		Gothenburg, Sweden	2014		http://www.aclweb.org/anthology/E14-1010	Y

59	Introduction to Computational Methods for the Analysis of Political Texts	Vossen, Piek	From Text to Political Positions Converging approaches to estimating party positions		John Benjamins Publishing Company		2014		http://www.academia.edu/6998238/From_text_to_Political_Positions_The_convergence_of_political_linguistic_and_discourse_analysis	N
60	Onomasiological lexicography: Wordnets and Thesauruses	Vossen, Piek	International Handbook of Modern Lexis and Lexicography		Springer-Verlag	Heidelberg, Germany	2014		http://tshwanedje.com/members/gmds/documents/Springer_Handbook_of_Modern_Lexis_and_Lexicography_ToC.pdf	N
61	NewsReader: recording history from daily news streams	Vossen, Piek	Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)	May 26-31		Reykjavik, Iceland	2014		http://www.lrec-conf.org/proceedings/lrec2014/pdf/436_Paper.pdf	Y
62	Proceedings of the 4th Workshop on Linked Science 2014 - Making Sense Out of Data (LISC2014), Riva del Garda, Italy, October 19, 2014.	Zhao, Jun					2014		http://ceur-ws.org/Vol-1282/	Y

63	Proceedings of the 7th Global WordNet Conference (GWC2014)	Orav, H.		Jan 25-29		Tartu, Estonia	2014		http://gwc2014.ut.ee/proceedings_of_GWC_2014.pdf	Y
64	Proceedings of the Third Workshop on Semantic Web and Information Extraction	Maynard, Diana		August	Association for Computational Linguistics and Dublin City University	Dublin, Ireland	2014		http://www.aclweb.org/anthology/W14-62	Y
65	Recognizing Biographical Sections in Wikipedia	Aprosio, Alessio Palmero	Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015		The Association for Computational Linguistics		2015	811–816	http://aclweb.org/anthology/D/D15/D15-1095.pdf	Y
66	A Contextual Framework for Reasoning on Events	Bozzato, Loris	DeRiVE2015 Workshop Proceedings				2015		http://ceur-ws.org/Vol-1363/paper_6.pdf	Y
67	SPINOZA VU: An NLP Pipeline for Cross Document TimeLines	Caselli, Tommaso	Proceedings of the 9th International Workshop on Semantic Evaluation				2015			Y

68	What happened to ...?" Entity-based Timeline Extraction	Caselli, Tommaso	Proceedings of the Computational Linguistics in the Netherlands (CLIN 2015)			Antwerp, Belgium	2015			Y
69	When it's all piling up: investigating error propagation in an NLP pipeline	Caselli, Tommaso	NLP Applications: completing the puzzle			Passau, Germany	2015			Y
70	Demonstrating the Power of Streaming and Sorting for Non-distributed RDF Processing: RDFpro	Corcoglioniti, Francesco	ISWC 2015 Posters {&} Demonstrations Track, within the 14th International Semantic Web Conference (ISWC 2015), Bethlehem, USA, October 11-15, 2015				2015		http://ceur-ws.org/Vol-1486/paper_52.pdf	Y

71	Extracting Knowledge from Text with PIKES	Corcoglioniti, Francesco	ISWC 2015 Posters {&} Demonstrations Track, within the 14th International Semantic Web Conference (ISWC 2015), Bethlehem, USA, October 11-15, 2015				2015		http://ceur-ws.org/Vol-1486/paper_66.pdf	Y
72	Processing Billions of RDF Triples on a Single Machine using Streaming and Sorting	Corcoglioniti, Francesco	ACM SAC 2015 Proceedings				2015		http://dl.acm.org/citation.cfm?doid=2695664.2695720	Y
73	"Bag of Events" Approach to Event Coreference Resolution. Supervised Classification of Event Templates	Cybulska, Agata	proceedings of the 16th Cicling 2015 (co-located: 1st International Arabic Computational Linguistics Conference)	April 14-20		Cairo, Egypt	2015			Y

74	Translating Granularity of Event Slots into Features for Event Coreference Resolution	Cybulska, Agata	Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation (co-located with NAACL-2015)			Denver, Colorado	2015			Y
75	Analysis of named entity recognition and linking for tweets	Derczynski, Leon	Information Processing & Management	March			2015	32-49	http://www.sciencedirect.com/science/article/pii/S0306457314001034	Y
76	Missing Mr. Brown and buying an Abraham Lincoln - Dark entities and DBpedia	van Erp, Marieke	Proceedings of the third NLP&DBpedia workshop				2015			Y
77	LOTUS: Linked Open Text Unleashed	Ilievski, Filip	Proceedings of COLD 2015				2015			Y
78	Named Entity Disambiguation with two-stage coherence optimization	Ilievski, Filip	Accepted for oral presentation at Computational Linguistics in the Netherlands (CLIN 2015)			Antwerp, Belgium	2015			Y

79	Topic Modelling and Word Sense Disambiguation on the Ancora corpus	Izquierdo, Ruben	Journal of the Spanish Society for Natural Language Processing (SEPLN2015)		SEPLN		2015			Y
80	Document Level Time-anchoring for TimeLine Extraction	Laparra, Egoitz	Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)			Beijing, China	2015			Y

81	From TimeLines to StoryLines: A preliminary proposal for evaluating narratives	Laparra, Egoitz	Proceedings of the 1st Workshop on Computing News StoryLines (CNewS 2015) at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)			Beijing, China	2015			Y
82	SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering	Minard, Anne-Lyse	Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)	June	Association for Computational Linguistics	Denver, Colorado	2015	778–786	http://www.aclweb.org/anthology/S15-2132	Y

83	FacTA: Evaluation of Event Factuality and Temporal Anchoring	Minard, Anne-Lyse	Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015	December 3-4		Trento, Italy	2015			Y
84	HLT-FBK: a Complete Temporal Processing System for QA TempEval	Mirza, Paramita	Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)	June	Association for Computational Linguistics	Denver, Colorado	2015	801–805	http://www.aclweb.org/anthology/S15-2135	Y
85	Semantic Technologies for Historical Research: A Survey	Peñuela, Albert Meroño	Semantic Web Journal				2015		http://www.semantic-web-journal.net/sites/default/files/swj301.pdf	Y
86	VUA-background: When to Use Background Information to Perform Word Sense Disambiguation	Postma, Marten	Proceedings of the 9th International Workshop on Semantic Evaluation				2015			Y
87	Open Source Dutch WordNet	Postma, Marten	Proceedings of the Computational Linguistics in the Netherlands (CLIN 2015)			Antwerp, Belgium	2015			Y

88	bRol: The Parser of Syntactic and Semantic Dependencies for Basque	Salaberri, Haritz	Proceedings of Recent Advances in Natural Language Processing (RANLP-2015)			Hissar, Bulgaria	2015			Y
89	IXAGroupEHUSpaceEval: (X-Space) A WordNet-based approach towards the Automatic Recognition of Spatial Information following the ISO-Space Annotation Scheme	Salaberri, Haritz	Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)	June	Association for Computational Linguistics	Denver, Colorado	2015	856–861	http://www.aclweb.org/anthology/S15-2145	Y
90	IXAGroupEHUDiac: A Multiple Approach System towards the Diachronic Evaluation of Texts	Salaberri, Haritz	Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)	June	Association for Computational Linguistics	Denver, Colorado	2015	840–845	http://www.aclweb.org/anthology/S15-2142	Y
91	EliXa: A Modular and Flexible ABSA Platform	San Vicente, Iñaki	Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)	June	Association for Computational Linguistics	Denver, Colorado	2015	748–752	http://www.aclweb.org/anthology/S15-2127	Y

92	ESO: A Frame based Ontology for Events and Implied Situations	Segers, Roxane	Accepted poster for Computational Linguistics in the Netherlands (CLIN 2015)			Antwerp, Belgium	2015			Y
93	The Event and Implied Situation Ontology	Segers, Roxane	Accepted abstract for Computational Linguistics in the Netherlands	12/1/2015		Amsterdam, the Netherlands	2015			Y
94	ESO: A Frame based Ontology for Events and Implied Situations	Segers, Roxane	Proceedings of MAPLEX 2015			Yamagata, Japan	2015			Y
95	Towards a Dutch Frame-Semantic Parser	van Son, Chantal					2015			Y
96	Details from a distance? A Dutch pipeline for event detection	van Son, Chantal					2015			Y
97	Cross-Language projection of multilayer semantic annotation in the NewsReader Wikinews Italian Corpus (WItaC)	Speranza, Manuela	Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015	December 3-4		Trento, Italy	2015			Y
98	EliXa: A Modular and Flexible ABSA Platform	San Vicente, Iñaki	Proceedings of SEMEVAL 2015			Denver, USA	2015			Y

99	Words in context: a reference perspective on the lexicon	Vossen, Piek	Proceedings of MAPLEX 2015	February 9-10		Yamagata, Japan	2015			Y
100	Storylines for structuring massive streams of news	Vossen, Piek	Proceedings of the 1st Workshop on Computing News StoryLines (CNewS 2015) at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)			Beijing, China	2015			Y

101	Interoperability for Cross-lingual and cross-document Event Detection	Vossen, Piek	Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation. EVENTS workshop at NAACL-HLT 2015			Denver, Colorado	2015			Y
102	Illuminating Dark Entities: a study on information discovery using Semantic web and Natural Language Processing	Vrijenhoek, Sanne					2015			Y
103	A Contextual Framework for Reasoning on Events	Bozzato, Loris	CILC2015 - 30th Convegno Italiano di Logica Computazionale, Genova, Italia	July, 1-3 2015			2015			Y
104	Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2015)	Zervanou, Kalliopi					2015			Y
105	Proceedings of the First Workshop on Computing News Storylines (CNewsStory 2015)	Caselli, Tommaso					2015			Y

106	A Sequence Labelling Approach to Attribution Relation Labelling	Caselli, Tommaso	Transactions of the Association for Computational Linguistics (TACL, ISSN: 2307-387X)			reviewed under revision	2016			Y
107	Exposing Predicate Models as Linked Data by Extending the Lemon Model	Corcoglioniti, Francesco	Proceedings of the 10th language resources and evaluation conference (LREC2016), Portoroz (Slovenia)			Portorož, Slovenia	2016			Y
108	A 2-phase Frame-based Knowledge Extraction Framework	Corcoglioniti, Francesco	Proc. of ACM Symposium on Applied Computing (SAC'16)				2016			Y
109	Evaluating Entity Linking: An Analysis of Current Benchmark Datasets and a Roadmap for Doing a Better Job	van Erp, Marieke	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
110	Crowdsourcing Salient Information from News and Tweets	Inel, O.	Proceedings of LREC 2016				2016			Y

111	Two architectures for parallel processing for huge amounts of text	Kattenberg, Mathijs	Proceedings of Language Resources and Evaluation Conference (LREC)			Portorož, Slovenia	2016			Y
112	A Multilingual Predicate Matrix	de Lacalle, Maddalen López	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
113	MEANTIME, the NewsReader Multilingual Event and Time Corpus	Minard, Anne-Lyse	Proceedings of LREC 2016			Portorož, Slovenia	2016			Y
114	A Comparison of Domain-based Word Polarity Estimation using different Word Embeddings	Pablos, Aitor Garcia	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
115	Addressing the MFS bias in WSD systems	Postma, Marten	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
116	Open Dutch WordNet	Postma, Marten	Proceedings of the Global WordNet Conference (GWC2016)			Bucharest, Romania	2016			Y

117	Context-enhanced Adaptive Entity Linking	Rizzo, Giuseppe	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
118	The Event and Implied Situation Ontology: Application and Evaluation	Segers, Roxane	Proceedings of Language Resources and Evaluation Conference (LREC 2016)			Portorož, Slovenia	2016			Y
119	A multi-layered annotation scheme for perspectives	van Son, Chantal	Proceedings of Language Resources and Evaluation Conference (LREC)				2016			Y
120	NewsReader: How Semantic Web helps Natural Language Processing helps Semantic Web	Vossen, Piek	Special Issue Knowledge Based Systems, Elsevier				2016			Y
121	Toward a truly multilingual Global Wordnet Grid	Vossen, Piek	Proceedings of the Global WordNet Conference (GWC2016)			Bucharest, Romania	2016			Y
122	Robust Multilingual Named Entity Recognition with Shallow Semi-supervised Features	Agerri, Rodrigo	Journal of Artificial Intelligence				Revised and Re-submitted			Y

TEMPLATE A2: LIST OF DISSEMINATION ACTIVITIES								
NO	Type of Activities	Main leader	Title	Date/Period	Place	Type of audience	Size of	Countries addressed
1	press release	Piek Vossen (VUA)	VU to Develop History Recorder	2/5/2012	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
2	interview	Piek Vossen (VUA)	Video interview "De Voorlopers"	2013	Amsterdam, The Netherlands	Civil Society	large	Netherlands
3	interview	Piek Vossen (VUA)	Video Interview VPRO	2013	Amsterdam, The Netherlands	Civil Society	large	Netherlands
4	presentation	Luciano Serafini (FBK)	Presentation on NewsReader at "Facts, Truths, Arguments"	2013-01-18	Fondazione Bruno Kessler, Trento, Italy,	Scientific Community & media	medium	Italy
5	presentation	Marieke van Erp (VUA)	NewsReader: Automating detective work	2013-01-23	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
6	presentation	Luciano Serafini (FBK)	CKR: a general framework for context in Semantic Web (Theory, prototype and extension to ASP)	2013-02-28	VU University Amsterdam	Scientific Community	medium	Netherlands
7	presentation	Egoitz Laparra (UPV/EHU)	Sources of Evidence for Implicit Argument Resolution	2013-03-20	Postman, Germany	Scientific Community	medium	International
8	presentation	Antske Fokkens (VUA)	Reproducing results in NLP	2013-04-07	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
9	presentation	Piek Vossen (VUA)	The news of today writes the history for the future. Recording history in the NewsReader project	2013-04-19	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
10	presentation	Piek Vossen (VUA)	Do big data hide or reveal stories? Recording History in the NewsReader project	2013-04-24	Wassenaar, The Netherlands	Scientific Community	medium	Netherlands

11	presentation	Piek Vossen (VUA)	Workshop on the Lexicon in Functional Discourse Grammar	2013-05-09	Vienna, Austria	Scientific Community	small	International
12	presentation	Piek Vossen (VUA)	NewsReader	2013-05-10	The Hague, The Netherlands	Childre	medium	Netherlands
13	presentation	Piek Vossen (VUA)	NewsReader	2013-05-24	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
14	presentation	Marieke van Erp (VUA)	From Events to Stories: Different ways of structuring the same bag of events over time	2013-06-13	Ravenstein, The Netherlands	Scientific Community	medium	Netherlands
15	presentation	Piek Vossen (VUA)	GAF: A Grounded Annotation Framework for Events	2013-06-14	Atlanta, GA, USA	Scientific Community	small	International
16	poster	Itziar Aldabe (UPV/EHU)	Similarity-Feature Based Approach for Student Response Analysis	2013-06-15	Atlanta, GA, USA	Scientific Community	medium	International
17	interview	Piek Vossen (VUA)	De Stemming, radio show	2013-06-30	The Netherlands	Civil Society	large	Netherlands
18	presentation	Antske Fokkens (VUA)	Offspring from Reproduction Problems: what replication failure teaches us	2013-07-08	Sofia, Bulgaria	Scientific Community	large	International
19	interview	Piek Vossen (VUA)	Weespernieuws	2013-07-08	The Netherlands	Civil Society	medium	Netherlands
20	presentation	Piek Vossen (VUA)	1st Annual Netherlands eScience Symposium	2013-07-11	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
21	presentation	Piek Vossen (VUA)	What e-connects Science and Humanities	2013-07-11	Amsterdam, The Netherlands	Scientific Community	large	Netherlands

22	presentation	Bernardo Magnini (FBK) and Tommaso Caselli (VUA)	“Events in Computational Linguistics” at the FBK workshop Perspectives on events	2013-07-18	Fondazione Bruno Kessler, Trento, Italy,	Scientific Community	large	Italy
23	interview	Piek Vossen (VUA)	Dit is de Dag, radio show	2013-07-23	The Netherlands	Civil Society	large	Netherlands
24	presentation	Antske Fokkens (VUA)	Linked Open Data & Delphi-In	2013-07-29	St. Wendel, Germany	Scientific Community	medium	International
25	interview	Piek Vossen (VUA)	Volkskrant	2013-08-01	The Netherlands	Civil Society	large	Netherlands
26	presentation	Egoitz Larra (UPV/EHU)	ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling	2013-08-06	Sofia, Bulgaria	Scientific Community	large	International
27	presentation	Bernardo Magnini (FBK)	The KNOWLEDGE-STORE: an Integrated Framework for Ontology Population	2013-09-06	INRIA, Nice, France	Scientific Community	medium	France
28	presentation	Piek Vossen (VUA)	the History Recorder: Today’s News is Tomorrow’s History	2013-09-12	Maastricht, The Netherlands	Scientific Community	medium	Netherlands
29	presentation	Piek Vossen (VUA)	De geschiedenisrecorder: het nieuws van vandaag is de geschiedenis van morgen	2013-09-12	Maastricht, The Netherlands	Civil Society	medium	Netherlands
30	presentation	Piek Vossen (VUA)	DE GESCHIEDENIS-RECORDER: HET NIEUWS VAN VANDAAG IS DE GESCHIEDENIS VAN MORGEN	2013-09-13	Oostende, Belgium	Scientific Community	small	Benelux

31	presentation	Marco Rospocher (FBK)	“Interlinking Unstructured and Structured Knowledge in an Integrated Framework” at the Seventh IEEE International Conference on Semantic Computing (IEEE-ICSC2013)	2013-09-16	Irvine, USA	Scientific Community	large	International
32	interview	Piek Vossen (VUA)	BNR Nieuwsradio	2013-09-27		Civil Society	large	Netherlands
33	presentation	Piek Vossen (VUA)	Ontsnappen aan de Gesloten Wereld van Taal	2013-09-27	The Hague, The Netherlands	Civil Society	large	Netherlands
34	presentation	Piek Vossen (VUA)	Help de computer taal begrijpen!	2013-10-05	The Hague, The Netherlands	Civil Society	small	Netherlands
35	interview	Piek Vossen (VUA)	Ad Valvas	2013-10-06	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
36	interview	Piek Vossen (VUA)	Hoe? Zo! Radio	2013-10-06	The Netherlands	Civil Society	large	Netherlands
37	interview	Piek Vossen (VUA)	Kennislink	2013-10-06	The Netherlands	Civil Society	large	Netherlands
38	interview	Piek Vossen (VUA)	New Scientist	2013-10-06	International	Civil Society	large	International
39	interview	Piek Vossen (VUA)	NOS Nieuws	2013-10-06	The Netherlands	Civil Society	large	Netherlands
40	other	Piek Vossen (VUA)	Nationale Denktank on Big Data	2013-10-10	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
41	interview	Piek Vossen (VUA)	OBA Live	2013-10-11	Amsterdam, The Netherlands	Civil Society	large	Netherlands
42	presentation	Piek Vossen (VUA)	Digital Humanities: Hype or Revolution – Part 1: “De digitale kaart van Landschap, Architectuur en Verstedelijking”	2013-10-12	Amsterdam, The Netherlands	Scientific Community	large	Netherlands

43	presentation	Piek Vossen (VUA)	Van feiten naar meningen, naar feiten over meningen	2013-10-16	Amsterdam, The Netherlands	Civil Society	large	Netherlands
44	presentation	Thomas Ploeger (SYN)	Extractivism. Extracting activist events from news articles using existing NLP tools	2013-10-21	Sydney, Australia	Scientific Community	medium	International
45	presentation	Marieke van Erp (VUA)	NewsReader: Making Large News Streams Manageable	2013-10-22	Sydney, Australia	Scientific Community	small	Australian
46	poster	Marieke van Erp (VUA)	Denoting Data in the Grounded Annotation Framework	2013-10-23	Sydney, Australia	Scientific Community	large	International
47	interview	Piek Vossen (VUA)	Algemeen Dagblad	2013-11-06	The Netherlands	Civil Society	large	Netherlands
48	interview	Piek Vossen (VUA)	NRC Handelsblad	2013-11-06	The Netherlands	Civil Society	large	Netherlands
49	interview	Piek Vossen (VUA)	NRC Next	2013-11-06	The Netherlands	Civil Society	large	Netherlands
50	interview	Piek Vossen (VUA)	Parool	2013-11-06	The Netherlands	Civil Society	large	Netherlands
51	interview	Piek Vossen (VUA)	Trouw	2013-11-06	The Netherlands	Civil Society	large	Netherlands
52	interview	Piek Vossen (VUA)	Volkskrant	2013-11-06	The Netherlands	Civil Society	large	Netherlands
53	presentation	Antske Fokkens (VUA)	Entities, Time and Event in BiographyNet and NewsReader	2013-11-13	Nijmegen, The Netherlands	Scientific Community	medium	Netherlands
54	presentation	Antske Fokkens (VUA)	What Replication Failure Teaches Us	2013-11-13	Nijmegen, The Netherlands	Scientific Community	medium	Netherlands
55	presentation	Piek Vossen (VUA)	Gala van de Amsterdamse Wetenschap	2013-11-26	Amsterdam, The Netherlands	Civil Society	small	Netherlands

56	presentation	Piek Vossen (VUA)	Help Computers understand language	2013-11-27	Eindhoven, The Netherlands	Scientific Community	large	Netherlands
57	presentation	Piek Vossen (VUA)	Language lessons for ICT	2013-11-27	Eindhoven, The Netherlands	Scientific Community	large	Netherlands
58	presentation	Piek Vossen (VUA)	Het internet als bron voor alle kwalen en ziektes	2013-11-29	Nijmegen, The Netherlands	Industry	medium	Netherlands
59	other	Piek Vossen (VUA)	Nationale Wetenschapsquiz	2013-12-29	Dutch National Television	Civil Society	large	Netherlands
60	interview	Piek Vossen (VUA)	The Why Projects, Bright New World	2014	the Netherlands	Scientific Community	large	Netherlands
61	interview	Piek Vossen (VUA)	Surf Magazine	2014	the Netherlands	Industry, Scientific Community, Policy Makers	large	Netherlands
62	interview	Piek Vossen (VUA)	Brabants Nieuwsblad	2014	the Netherlands	Civil Society	large	Netherlands
63	interview	Piek Vossen (VUA)	NWO Annual Report	2014	the Netherlands	Scientific Community	large	Netherlands
64	interview	Piek Vossen (VUA)	Hannover Messe	2014	Germany	Industry	large	International
65	interview	Piek Vossen (VUA)	Hypothese NWO	2014	the Netherlands	Scientific Community	large	Netherlands
66	interview	Piek Vossen (VUA)	NWO Annual Report	2014	the Netherlands	Scientific Community	large	Netherlands
67	presentation	Marieke van Erp (VUA)	Evaluating Named Entity Recognition and Disambiguation in News and Tweets	2014-01-17	Leiden, The Netherlands	Scientific Community	medium	Netherlands

68	presentation	Piek Vossen (VUA)	Digital Humanities: Hype or Revolution – Part 2: De Gouden Eeuw: de ontsluiting van een schatkamer aan informatie	2014-01-21	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
69	interview	Piek Vossen (VUA)	Kenniscafé Almere	2014-01-23	Almere, Flevoland	Civil Society	medium	Netherlands
70	presentation	López de Lacalle, Madalen (UP-V/EHU)	First steps towards a Predicate Matrix	2014-01-25	Tartu, Estonia	Scientific Community	medium	International
71	presentation	Itziar Aldabe (UP-V/EHU)	NewsReader project	2014-02-12	Donostia - San Sebastian, Spain	Scientific Community	small	Spain
72	presentation	Piek Vossen (VUA)	Invited lecture for Alumni Letteren	2014-02-14	Amsterdam, The Netherlands	Civil Society	small	Netherlands
73	presentation	Piek Vossen (VUA)	NewsReader: recording history by processing massive streams of daily news	2014-03-19	Athens, Greece	Industry, Scientific Community, Policy Makers	large	International
74	presentation	Piek Vossen (VUA)	TopToets lecture	2014-04-16	Amsterdam, The Netherlands	Children	medium	Netherlands
75	presentation	Piek Vossen (VUA)	In zeven stappen naar totaal onbegrip	2014-04-18	Amsterdam, The Netherlands	Civil Society	large	Netherlands
76	poster/demo	Rodrigo Agerri (UP-V/EHU)	Multilingual, Efficient and Easy NLP Processing with IXA Pipeline	2014-04-28	Gothenburg, Sweden	Scientific Community	medium	International

77	presentation	Iñaki San Vicente (UP-V/EHU)	Simple, Robust and (almost) Unsupervised Generation of Polarity Lexicons for Multiple Languages	2014-04-28	Gothenburg, Sweden	Scientific Community	large	International
78	poster/demo	German Rigau (UP-V/EHU)	EuroLoveMap: Confronting feelings from News	2014-05-26	Reykjavik, Iceland	Scientific Community	medium	International
79	presentation	Piek Vossen (VUA)	A Collaborative Interlingual Index for harmonizing word nets	2014-05-26	Reykjavik, Iceland	Scientific Community	large	International
80	presentation	Piek Vossen (VUA)	NewsReader: recording history from daily news streams	2014-05-26	Reykjavik, Iceland	Scientific Community	large	International
81	presentation	German Rigau (UP-V/EHU)	Predicate Matrix: extending SemLink through WordNet mappings	2014-05-28	Reykjavik, Iceland	Scientific Community	large	International
82	poster/demo	Rodrigo Agerri (UP-V/EHU)	IXA pipeline: Efficient and Ready to Use Multilingual NLP tools	2014-05-30	Reykjavik, Iceland	Scientific Community	medium	International
83	presentation	Piek Vossen (VUA)	Digital Humanities: Hype or Revolution – Part 3: Het gezicht van de sociale media	2014-06-02	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
84	other	Piek Vossen (VUA)	Tafelwetenschapper Avond van de Wetenschap & Maatschappij	2014-06-10	The Hague, The Netherlands	Scientific Community, Other, civil society	small	Netherlands
85	workshop	Ian Hopkinson (SCW)	World Cup articles hack day	2014-06-10	London, UK	Industry	medium	United Kingdom
86	presentation	Piek Vossen (VUA)	Kunnen computers ooit de taal van mensen begrijpen	2014-07-02	The Hague, The Netherlands	Civil Society	small	Netherlands
87	presentation	Piek Vossen (VUA)	Ontsnappen aan de gesloten wereld van Taal	2014-08-04	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands

88	presentation	Piek Vossen (VUA)	Kan een computer taal begrijpen?	2014-09-03	Amersfoort, The Netherlands	Civil Society	medium	Netherlands
89	presentation	German Rigau (UPV/EHU)	NewsReader project presentation (roundtable)	2014-09-14	Girona, Spain	Scientific Community	medium	International
90	presentation	Piek Vossen (VUA)	What if computers could read the news?	2014-09-19	Utrecht, The Netherlands	Scientific Community	medium	Netherlands
91	presentation	Piek Vossen (VUA)	NewsReader	2014-10-04	Hannover, Germany	Industry	medium	International
92	presentation	Piek Vossen (VUA)	tablespeaker Nationale Denktank	2014-10-10	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
93	presentation	Bernardo Magnini (FBK)	“The KNOWLEDGE STORE: an Integrated Framework for Ontology Population”	2014-10-16	Darmstadt University (TU), UKP group	Scientific Community	medium	Germany
94	presentation	Piek Vossen (VUA)	Understanding language by machines	2014-10-17	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
95	presentation	Marco Rospocher (FBK)	ISWC Developers Workshop talk : “A Simple API to the KnowledgeStore”	2014-10-19	Riva del Garda, Italy	Scientific Community	large	International
96	presentation	Bernardo Magnini (FBK)	“Al tempo nel linguaggio” at “Festival della Scienza”	2014-10-24	Genova, Italy	Civil Society	large	Italy
97	presentation	Piek Vossen (VUA)	Gollandskaja sreda: Dutch Wednesday 2014	2014-10-29	St. Petersburg, Russia	Scientific Community	medium	International
98	presentation	Marco Rospocher (FBK)	FBK Seminar on “KnowledgeStore”	2014-11-04	Fondazione Bruno Kessler, Trento, Italy,	Scientific Community	medium	Italy
99	presentation	Piek Vossen (VUA)	data-driven research and big-data	2014-11-04	Amsterdam, The Netherlands	Scientific Community	medium	International

100	poster	Marieke van Erp (VUA)	Nieuwe Namen Leren Linken	2014-11-14	The Hague, The Netherlands	Industry, Scientific Community, Policy Makers	large	Netherlands
101	presentation	Piek Vossen (VUA)	Studium Generale	2014-11-19	Eindhoven, The Netherlands	Scientific Community	medium	Netherlands
102	presentation	Piek Vossen (VUA)	NewsReader DNB	2014-12-18	Amsterdam, The Netherlands	Industry	small	Netherlands
103	workshop	Marieke van Erp (VUA)	Automotive articles hack day	2015-01-21	Amsterdam, NL	Industry	medium	International
104	presentation	Piek Vossen (VUA)	De taal van de verbeelding: is denken meer dan rekenen?	2015-01-28	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
105	workshop	Ian Hopkinson (SCW)	Automotive articles hack day	2015-01-30	London, UK	Industry	medium	International
106	presentation	Marco Rospocher (FBK)	“Integrating Unstructured and Structured Data in the Knowledge Store”: presentation at the “Deep and Large-Scale Semantic Processing Workshop”	2015-03-10	Trento, Italy	Scientific Community	medium	International
107	presentation	Piek Vossen (VUA)	NewsReader: the reading machine	2015-03-12	Amsterdam, The Netherlands	Industry	small	Netherlands
108	presentation	Marco Rospocher (FBK)	“Dati e Ricerca”: Presentation at the “Beni Comuni Digitali”	2015-03-18	Riva del Garda, Italy	Civil Society	medium	Italy
109	presentation	Antske Fokkens (VUA)	Political Discourse in the News	2015-03-24	The Hague, The Netherlands	Scientific community	large	Netherlands
110	presentation	Piek Vossen (VUA)	NewsReader	2015-03-24	Amsterdam, The Netherlands	Industry	medium	Netherlands

111	presentation	Piek Vossen (VUA)	What's in the news?	2015-03-26	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
112	presentation	Marco Rospocher (FBK)	Presentation at SAC2015 Conference: "Processing Billions of RDF Triples on a Single Machine using Streaming and Sorting"	2015-04-13	Salamanca, Spain	Scientific Community	medium	International
113	presentation	Piek Vossen (VUA)	Abraham Kuiperlezing	2015-05-16	Amsterdam, The Netherlands	Civil Society	large	Netherlands
114	presentation	Piek Vossen (VUA)	ehumanities KNAW	2015-05-21	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands
115	presentation	Loris Bozzato (FBK)	Presentation at DeRiVE2015 Workshop: "A Contextual Framework for Reasoning on Events"	2015-05-31	Portoroz, Slovenia	Scientific Community	medium	International
116	presentation	Piek Vossen (VUA)	ESO: a frame based Ontology for events and implied situations	2015-06-02	Antwerp, Belgium	Scientific Community	medium	Netherlands
117	presentation	Tommaso Caselli (VUA)	"What happened to ...?" Entity-based Timeline Extraction	2015-06-02	Antwerp, Belgium	Scientific Community	medium	Netherlands
118	presentation	Iñaki San Vicente (UPV/EHU)	EliXa: A Modular and Flexible ABSA Platform	2015-06-04	Denver, USA	Scientific Community	medium	International
119	presentation	Anne-Lyse Minard (FBK)	"SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering" and "HLT-FBK: a Complete Temporal Processing System for QA TempEval"	2015-06-05	Denver, Colorado	Scientific Community	large	International

120	presentation	Piek Vossen (VUA)	Modellling provenance and perspectives in Biographical Data	2015-06-10	Vienna, Austria	Scientific Community	large	International
121	presentation	Piek Vossen (VUA)	NewsReader ‘Linguistic Linked Open Data’	2015-06-18	Cercedilla, Spain	Scientific Community	medium	International
122	presentation	Piek Vossen (VUA)	The Global Wordnet Grid	2015-06-18	Madrid, Spain	Scientific Community	medium	International
123	presentation	Loris Bozzato (FBK)	Presentation at CILC2015 Conference: “A Contextual Framework for Reasoning on Events”	2015-07-01	Genova, Italy	Scientific Community	large	Italy
124	interview	Piek Vossen (VUA)	Quest interview	2015-07-05	The Netherlands	Civil Society	large	Netherlands
125	interview	Piek Vossen (VUA)	NTR Academie	2015-07-06	Hilversum, The Netherlands	Civil Society	large	Netherlands
126	interview	Piek Vossen (VUA)	MediaMix Radio	2015-07-06	Amsterdam, The Netherlands	Civil Society	large	Netherlands
127	interview	Piek Vossen (VUA)	A computer that really understands language	2015-07-07	the Netherlands	Scientific Community	large	Netherlands
128	poster	Egoitz Laparra (UPV/EHU)	Document Level Time-anchoring for TimeLine Extraction.	2015-07-28	Beijing, China	Scientific Community	medium	International
129	presentation	German Rigau (UPV/EHU)	From TimeLines to StoryLines: A preliminary proposal for evaluating narratives	2015-07-31	Beijing, China	Scientific Community	medium	International
130	presentation	Antske Fokkens (VUA)	What do you think? Unfolding perspectives in text	2015-08-05	Amsterdam, The Netherlands	Scientific Community	small	Netherlands
131	presentation	Tommaso Caselli (VUA)	Storylines for structuring big data?	2015-08-05	Amsterdam, The Netherlands	Scientific Community	small	Netherlands

132	presentation	Piek Vossen (VUA)	From mentions in text to instances in RDF: cross-lingual interpretation of unstructured news in the NewsReader project	2015-08-09	Hissar, Bulgaria	Scientific Community	small	International
133	presentation	Piek Vossen (VUA)	Modelling uncertainties and perspectives in the news	2015-08-10	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
134	presentation	Alessio Palmero Aprosio (FBK)	Poster presentation at EMNLP2015 Conference: "Recognizing Biographical Sections in Wikipedia"	2015-09-17	Lisbon, Portugal	Scientific Community	large	International
135	presentation	Piek Vossen (VUA)	NewsReader: extracting event-centric knowledge graphs from massive news streams	2015-09-25	Hilversum, The Netherlands	Scientific Community, Industry	large	Netherlands
136	presentation	Marieke van Erp (VUA)	Missing Mr. Brown and buying an Abraham Lincoln - Dark Entities and DBpedia	2015-10-11	Bethlehem, Pennsylvania, USA	Scientific Community	medium	International
137	presentation	Piek Vossen (VUA)	Elsevier Web Lecture	2015-10-11	Online	Industry	large	International
138	presentation	Marco Rospocher (FBK)	Demo presentation at ISWC2015 Conference: "Demonstrating the Power of Streaming and Sorting for Non-distributed RDF Processing: RDFpro"	2015-10-13	Bethlehem, Pennsylvania, USA	Scientific Community	large	International
139	presentation	Marco Rospocher (FBK)	Demo presentation at ISWC2015 Conference: "Extracting Knowledge from Text with PIKES"	2015-10-13	Bethlehem, Pennsylvania, USA	Scientific Community	large	International

140	presentation	German Rigau (UP-V/EHU)	NLP & Linked Data: OpeNER and NewsReader	2015-10-20	Madrid, Spain	Scientific Community	medium	Spain
141	presentation	Piek Vossen (VUA)	NewsReader	2015-10-20	The Hague, The Netherlands	Scientific Community	medium	Netherlands
142	presentation	German Rigau (UP-V/EHU)	Cross-lingual Event Detection in Discourse	2015-11-09	Hissar, Bulgaria	Scientific Community	medium	International
143	presentation	Piek Vossen (VUA)	Cross-lingual Event Detection in Discourse	2015-11-09	Hissar, Bulgaria	Scientific Community	medium	International
144	exhibition	Piek Vossen (VUA)	NewsReader	2015-11-17	Luxembourg, Luxembourg	Industry, Scientific Community, Policy Makers	medium	International
145	poster	Piek Vossen (VUA)	NewsReader poster	2015-11-17	Luxembourg, Luxembourg	Industry, Scientific Community, Policy Makers	medium	International
146	flyer	Piek Vossen (VUA)	NewsReader brochure	2015-11-17	Luxembourg, Luxembourg	Industry, Scientific Community, Policy Makers	medium	International
147	presentation	Piek Vossen (VUA)	The NewsReader vision	2015-11-17	Luxembourg, Luxembourg	Industry, Scientific Community, Policy Makers	medium	International
148	presentation	Piek Vossen (VUA)	Sentimenten, meningen en perspectieven in taal	2015-11-19	Amsterdam, The Netherlands	Scientific Community	medium	Netherlands

149	presentation	Piek Vossen (VUA)	Beter zoeken, beter graven	2015-11-21	The Hague, The Netherlands	Industry	medium	Netherlands
150	presentation	Piek Vossen (VUA)	Stories behind Data	2015-11-23	The Hague, The Netherlands	Industry	large	Netherlands
151	presentation	Bernardo Magnini (FBK)	Language and Knowledge: Toward Understanding Multimedia Content	2015-11-24	Public Library, Amsterdam	Civil Society	large	Netherlands
152	workshop	Marieke van Erp (VUA)	Automotive articles hack day	2015-11-24	Amsterdam, NL	Industry	medium	Amsterdam
153	presentation	Antske Fokkens (VUA)	NewsReader: unfolding stories and perspectives in large amounts of data	2015-11-27	Leuven, Belgium	Scientific Community	medium	Benelux
154	other	Antske Fokkens (VUA)	Panelist at the “Social Media: Incubators of a renewed news media landscape?” Symposium	2015-11-27	Leuven, Belgium	Scientific Community	medium	Benelux
155	other	Piek Vossen (VUA)	Panelist at the BIG DATA DEBAT	2015-11-29	Amsterdam, The Netherlands	Civil Society	medium	Netherlands
156	presentation	Anne-Lyse Minard (FBK)	Presentation at CLiC-it: “FacTA: Evaluation of Event Factuality and Temporal Anchoring”	2015-12-04	Trento, Italy	Scientific Community	large	Italy
157	presentation	Manuela Speranza (FBK)	Poster presentation at CLiC-it: “Cross-language projection of multilayer semantic annotation in the NewsReader Wikinews Italian Corpus (WItaC)”	2015-12-04	Trento, Italy	Scientific Community	large	Italy
158	presentation	Bernardo Magnini (FBK)	Reconstructing TimeLines in the KNOWLEDGESTORE	2015-12-10	IRIT, University of Toulouse	Scientific Community	medium	France

159	presentation	Filip Ilievski (VUA)	LOTUS: Linked Open Text UnLeashed	2015-12-10	Bethlehem, Pennsylvania, USA	Scientific Community	medium	International
160	presentation	Piek Vossen (VUA)	OpenMinTeD interoperability workshop	2015-12-11	The Hague, The Netherlands	Other	medium	International
161	press release	Piek Vossen (VUA)	Beter zoeken dan Google	2015-12-16	Amsterdam, The Netherlands	Scientific Community	large	Netherlands
162	presentation	German Rigau (UP-V/EHU)	NewsReader project	2016-01-22	Donostia - San Sebastian, Spain	Public Administration	small	Spain
163	presentation	Rodrigo Agerri (UP-V/EHU)	IXA pipes: Easy and ready use NLP tools for language communities	2016-01-31	Brussels, Belgium	Developers	large	International
164	video	Piek Vossen (VUA)	NewsReader video	2016-02-12	The Netherlands	Civil Society	large	International
165	presentation	Piek Vossen (VUA)	Netwerken in Perspectief VU Lustrumconferentie”	2016-03-18	The Netherlands	Scientific Community	large	Netherlands
166	presentation	Piek Vossen (VUA)	NewsReader	2016-04-03	Konya, Turkey	Scientific Community	large	International

Table 4: List of dissemination activities where audience size ‘small’ indicates 10-20people, ‘medium’ 20-50, and ‘large’ 50+ people

Section B (Confidential or public: confidential information to be marked clearly)

Part B1

APPLICATIONS FOR PATENTS, TRADEMARKS, REGISTERED DESIGNS: Not applicable

Part B2

Type of Exploitable Foreground	Description of Exploitable Foreground	Confidential YES/NO	Foreseen embargo date	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
software	The NewsReader KnowledgeStore is a scalable, fault-tolerant, and Semantic Web grounded storage system to jointly store, manage, retrieve, and semantically query, both structured and unstructured data. The KnowledgeStore plays a central role in the NewsReader EU project: it stores all contents that have to be processed and produced in order to extract knowledge from news, and it provides a shared data space through which NewsReader components cooperate.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK

software	RDFpro (RDF Processor) is a public domain, Java command line tool and library for RDF processing. RDFpro offers a suite of stream-oriented, highly optimized RDF processors for common tasks that can be assembled in complex pipelines to efficiently process RDF data in one or more passes. RDFpro originated from the need of a tool supporting typical Linked Data integration tasks, involving dataset sizes up to few billions triples.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	CAT is a general-purpose web-based tool for text annotation	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	Cromer is a Web-based tool to manually annotate event and entity coreference across clusters of documents.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	textpro 2.0 is a suite of modular Natural Language Processing (NLP) tools for analysis of Italian and English texts	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK

software	timepro identifies the tokens corresponding to temporal expressions in English or Italian, assigns them to one of the 4 TIMEX classes defined in ISO-TimeML and normalizes them following TIDES specifications. The English model has been trained on TempEval3 data and the Italian model on EVENTI corpus.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	temprelpro extracts and classifies temporal relations between two events or an event and a time expression in English and Italian texts. It also annotates time anchors of predicates (i.e. the time when an event occurred). English models have been trained on TempEval3 data and Italian models on EVENTI corpus.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	causalrelpro extracts explicit causal relations between two events in the same sentence in English.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	NAF populator is a homogeneous multi-threading queue processing which convert the NAF structure to the KS data-model.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK

data	The NewsReader MEAN-TIME corpus consists of English WikiNews articles annotated with temporal informations, has participant relations, entities, etc. at the document level and the corpus level following NewsReader guidelines.	NO	N/A	Research Data	J63 - Information service activities	Currently used	N/A	EHU, FBK and VUA
data	TimeLines built from a selection of seed entities. The corpus has been created for the TimeLine task at SemEval 2015.	NO	N/A	Research Data	J63 - Information service activities	Currently used	N/A	EHU, FBK and VUA
software	eventpro detects event extents and classifies them in one of the 7 TimeML classes. It is trained on the EVENTI corpus.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	factpro detects for each event annotated in a text its polarity, certainty and time. It is trained on the Fact Ita-Bank corpus.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
data	Fact-Ita Bank consists of 170 news stories (for a total of 65,455 tokens) selected from Ita-TimeBank annotated with factuality information on top of TimeML event annotation.	NO	N/A	Research Data	J63 - Information service activities	Currently used	N/A	FBK

software	RDFpro reasoner contains some additional modules for RDFpro. In particular: @esoreasoner apply the ESO ontology to the TriG files and extract the resulting triples (statements); @reformat-time convert the time formulas into an integer value; @filtertype assures that for each event there is only one type of link (eso ontology, framenet, propbank).	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	PIKES is a Java-based suite that extracts knowledge from textual resources. The tool implements a rule-based strategy that reinterprets the output of semantic role labelling (SRL) tools in light of other linguistic analyses, such as dependency parsing or co-reference resolution, thus properly capturing and formalizing in RDF important linguistic aspects such as argument nominalization, frame-frame relations, and group entities.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK

software	fbk-srl is a Semantic Role Labelling system for Italian. For each predicates annotated by fbk-eventpro it annotated its arguments and disambiguate its sense using MultiWordNet and an interlingual index (ili).	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	fbk-eventcoref creates intra-document event coreference relations. It uses lexical, semantic and morpho-syntactic criteria.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	nwr-eval is a suite of evaluation scorers for SRL, NERC, NED, nominal coreference, event coreference, event factuality, time expression, temporal relations, causal relations, timeline creation. The format of the gold standard corpus should be CAT and of the system output NAF. The package contains scripts to convert CAT and NAF files in the various evaluation format (CoNLL, NAF to CAT, CAT to NAF).	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK, EHU and VUA

software	The newsparser tool processes xml files and produces in output a document compressed archive with some meta information. The tool is the starting point of the current NewsReader event detection pipeline.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	FBK
software	Vua-Event-Coreference is a Java package for creating intra-document and cross-document event coreference relations. It includes various functions to establish these relations within a single document creating coreference layers in NAF. It also includes a function to convert NAF to RDF-TriG, establishing cross-document coreference relations that connect event instances to all mentions in the text and to create event-participant, event-place and event-time relations.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	Vua-CoreferenceEvaluation: Java package of functions for evaluating event coreference: conversion of NewsReader CAT annotations to CoNLL2011/2012 format for coreference; conversion of NewsReader NAF coreference to CoNLL2011/2012 format for coreference; function to reduce the key file to sentences with annotations; function to reduce the response file to the sentences of the key file; function to generate the scripts to compare key files with response files in CoNLL coreference format using CorScorer function to collect the results and calculate the macro and micro?averages. CorScorer can be downloaded from: https://code.google.com/p/reference?coreference?scorers/ The package include the WikiNews evaluation data for coreference.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
----------	--	----	-----	-------------------	--------------------------------------	----------------	-----	-----

software	vua-MultiWordTagger: This module reads a KAF or NAF file to detect multiword sequences of terms according the WordNet in LMF format. It replaces matched sequences of terms by a multiword term in WordNet with the elements as components in the terms. It creates a new identifier for the term by extending it with “mw” and fixes all further term references. to components in the chunk and dependency layers of NAF/KAF.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-KafSaxParser: Java package to read KAF/-NAF files and to access all the layers and data elements. It can also serialize KAF/NAF.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-WordnetTools: Java package to read any wordnet in LMF format and to apply the standard similarity and relatedness functions to it.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	vua-Ontotagger: Inserts semantic tags into the term layer of NAF/KAF as external references. Tags can be provided in table format or as OWL. Intended to run on top of the output of a WSD system.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
Resource	Open Dutch Wordnet: Dutch Wordnet	NO	N/A	Research resource	J63 - Information service activities	Currently used	N/A	VUA
Dataset	Extended EventCoreferenceBank: English Google news annotated with event-participant-place-time tags and with cross document and cross-topic coreference relations.	NO	N/A	Research data	J63 - Information service activities	Currently used	N/A	VUA
software	VUA-factuality: Module that aims to determine whether a statement is (a) confirmed or denied (b) certain, probably or possible and (c) future or non?future	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	Discourse Module: Module that splits a LexisNexis XML file into headers and body text. It writes the output to KAF. The LexisNexis data is structured using the NITF News Industry Text Format.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	vua-srl-nl: Semantic Role Labeller for Dutch trained on the Dutch SoNaR corpus. The module is a reimplementaion of the SRL package described in: Orphée De Clercq and Veronique Hoste and Paola Monachesi (2012) Evaluating automatic cross-domain semantic role annotation. In: Proceedings of the 8th International Conference on Language Resources and Evaluation Conference (LREC'2012). Istanbul, Turkey. pp 88-93 and includes a NAF wrapper for integration in the NewsReader pipeline.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-nedtype-reranker: Module that reranks the output of the Named Entity Disambiguation Module based on the most frequently occurring DBpedia types measured for the global automotive industry domain.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

resource	The Event and Situation Ontology (ESO) provides a populated model of events, the implied situations of events and the roles of the entities affected by an event. The ontology includes mappings to SUMO and Framenet on class level and to Framenet on role level. The ontology consists of 59 event classes.	NO	N/A	Research Resource	J63 - Information service activities	Currently used	N/A	VUA, FBK
software	VUA-opinion_miner: This is a module for extracting fine-grained opinions from KAF or NAF files. It can be trained given a set of KAF/NAF annotated files. Once trained, it detects opinion expressions, their targets and holders and create opinion triples that are stored in KAF/-NAF format.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	VUA-svm_wsd: System developed in Python for performing automatic Word Sense Disambiguation in Dutch text. It has been trained using the annotated data resulting from the DutchSemCor project, where it was firstly created.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	VUA-it_makes_sense: Wrapper around the it-makes-system WSD system, which is one of the best performing systems currently on the field. It allows to work with KAF/NAF files as input and output	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-morphosyntactic_parser_nl: This module implements a wrapper for the Alpino parser for Dutch text. It takes as input Dutch plain text, and it obtains the tokens, terms, part-of-speech and syntactic trees by calling to Alpino. The results are stored in KAF/NAF format.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	Vua-Heideltime is a KAF/NAF wrapper around HeidelTime. HeidelTime can identify and normalize temporal expressions in text and works on Dutch and English, among others. The wrapper allows users to use the standalone version of HeidelTime in a pipeline of modules using KAF or NAF as representation format.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	vua-srl-postprocess: This module is a Semantic Role Labeling post-processing script which uses the ESO ontology to decide which of the FrameNet frames and roles for a predicate mention are potentially valid. A frame/role is valid if it corresponds to an ESO class detected for a mention. The invalid predicates/roles are assigned a “?”, while the valid are assigned a “+”.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	DBpedia NER: A module that applies DBpedia spotlight on the token and term layer and creates entities based on DBpedia spotlight’s findings. Entity types come from DBpedia.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-nominal-event-detection: Identifies nominal elements that refer to an event and adds them as predicates in the semantic role layer.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	vua-srl-dutch-nominal-events: Checks if predicates without semantic roles are nominals and if so, creates relations with its PP modifiers	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA

software	KafNafParserPy: Parser in Python that can read, interpret, create, alternate and convert KAF and NAF	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	POCUS This module takes disambiguated entities as input and checks what the most logical interpretation is given the rest of the documents. In some cases, it will create new entities by (de)composing existing entities.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	VUA
software	ixa-pipe-tok is a multilingual rule-based tokenizer and sentence segmenter. ixa-pipe-tok is part of IXA pipes, a multilingual NLP pipeline developed by the IXA NLP Group. [http://ixa2.si.ehu.es/ixa-pipes]. This module provides Multilingual Sentence Segmentation and Tokenization for a number of languages, such as Basque, Dutch, German, English, French, Galician, Italian and Spanish.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU

software	ixa-pipe-pos provides POS tagging and lemmatization several languages. We provide Perceptron (Collins 2002) and Maximum Entropy (Ratnapharki 1999) POS tagging models: a) POS tagging models for English trained and evaluated using the WSJ treebank as explained in K. Toutanova, D. Klein, and C. D. Manning. Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of HLT-NAACL'03, 2003; b) POS tagging models for Spanish trained and evaluated using the Ancora corpus via 5-fold and 10-fold cross-validation.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
----------	---	----	-----	-------------------	--------------------------------------	----------------	-----	-----

software	ixa-pipe-nerc provides NERC for Basque, Dutch, English, German, Spanish and Italian. The named entity types are based on: a) CONLL: LOCATION, MISC, ORGANIZATION and PERSON; b) Ancora: LOCATION, MISC, ORGANIZATION, PERSON, NUMBER, DATE; c) SONAR-1: LOCATION, MISC, ORGANIZATION, PERSON, PRODUCT, EVENT	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	ixa-pipe-parse provides: a) Constituent parsing for English trained on the Penn Treebank and for Spanish trained on the Ancora corpus; b) HeadFinders based on Collins head rules (Michael Collins PhD thesis, 1999).	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU

software	CorefGraph is an independent module to perform coreference resolution. View as a NLP task which consists of determining the mentions that refer to the same entity in a text or discourse. CorefGraph is a python reimplementation of the Stanford Multi Sieve Pass system (Lee et al., 2013). The module provides resources for English and Spanish coreference resolution but it can be adapted to other languages.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	Ixa-pipe-srl provides a wrapper for English and Spanish dependency parser and semantic role labeller using mate-tools (https://code.google.com/p/mate-tools/). The module takes tokenized and POS-tagged text in NAF format as standard input and outputs syntactic and semantic analysis also in NAF. The semantic annotation provided by the module is enriched using the PredicateMatrix (http://adimen.si.ehu.es/web/PredicateMatrix).	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU

software	This repository contains the Named Entity Disambiguation tool based on DBpedia Spotlight. Providing that a DBpedia Spotlight Rest server for a given language is running, the ixa-pipe-ned module will take NAF or KAF as input (containing elements) and perform Named Entity Disambiguation for your language of choice.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	NAF_UKB is a tool to add sense information to a NAF input, producing a NAF output. This tool uses UKB, a collection of programs for performing graph-based Word Sense Disambiguation and lexical similarity/relatedness using a pre-existing knowledge base.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	NAF wrapper that detects time expressions based on HeidelTime in Spanish texts.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU

software	This repository contains the Named Entity Disambiguation tool based on DBpedia Spotlight. Providing that a DBpedia Spotlight Rest server for a given language is running, the ixa-pipe-ned module will take NAF or KAF as input (containing elements) and perform Named Entity Disambiguation for your language of choice.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	The module is based on the Multilingual Eurovoc thesaurus descriptors and it makes use of the JRC Eurovoc Indexer JEX to extract a set of topics given a document. The module takes NAF as input (containing terms) and writes the topic information of the document.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	Kaflib is a library which allows creating, editing and reading NAF (and KAF) documents.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
data	The Predicate Matrix is a new lexical resource resulting from the integration of multiple sources of predicate information including FrameNet, VerbNet, PropBank and WordNet.	NO	N/A	Research Data	J63 - Information service activities	Currently used	N/A	EHU

software	A set of scripts with the aim of automatically create a fully working cluster for distributed NLP processing.	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	A set of scripts with the aim of automatically create a fully NewsReader integrated system	NO	N/A	Research Software	J63 - Information service activities	Currently used	N/A	EHU
software	SynerScope is an explorative analytics platform that augments human reasoning. Can be run on-premises (local instance or server) or in the cloud (public or private)	NO	N/A	Commercial Software	J63 - Information service activities	Currently used	Yes	SYN
software	The Knowledgestore interface is a connector tool interfacing with the Knowledgestore by querying (using actor name(s)) resulting in a ready-made SynerScope project (querying the Knowledgestore and modeling the resulting data into a SynerScope investigation)	NO	N/A	Commercial Software	J63 - Information service activities	Currently used	No	SYN

The foreground consists of open source code and resources and can be exploited by anyone. We intend to shape exploitation through projects and consultancy, further details are described in Deliverable D9.2: Exploitation Plan.

3 Report on Societal implications

A General Information	
Grant Agreement Number:	316404
Title of Project:	NewsReader: Building structured event Indexes of large volumes of financial and economic Data for Decision Making
Name and Title of Coordinator:	Prof. dr. Piek Vossen
B Ethics	
1. Did your project undergo an Ethics Review (and/or Screening)? If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 'Work Progress and Achievements'	NO
2. Please indicate whether your project involved any of the following issues:	NO
Research on Humans	
Did the project involve children?	NO
Did the project involve patients?	NO
Did the project involve persons not able to give consent?	NO
Did the project involve adult healthy volunteers?	NO
Did the project involve Human genetic material?	NO
Did the project involve Human biological samples?	NO
Did the project involve Human data collection?	NO
Research on Human embryo/foetus	
Did the project involve Human Embryos?	NO
Did the project involve Human Foetal Tissue / Cells?	NO
Did the project involve Human Embryonic Stem Cells (hESCs)?	NO
Did the project on human Embryonic Stem Cells involve cells in culture?	NO
Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	NO
Privacy	
Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	NO
Did the project involve tracking the location or observation of people?	NO
Research on Animals	
Did the project involve research on animals?	NO
Were those animals transgenic small laboratory animals?	NO
Were those animals transgenic farm animals?	NO
Were those animals cloned farm animals?	NO
Were those animals non-human primates?	NO
Research Involving Developing Countries	
Did the project involve the use of local resources (genetic, animal, plant etc)?	NO
Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	NO
Dual Use	
Research having direct military use	NO
Research having the potential for terrorist abuse	NO

C Workforce Statistics			
3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).			
Type of Position	Number of Women		Number of Men
Scientific Coordinator	0		4
Work package leaders	1		5
Experienced researchers (i.e. PhD holders)	9		13
PhD Students	4		3
Other	10		19
4. How many additional researchers (in companies and universities) were recruited specifically for this project?			4
Of which, indicate the number of men:			1
D Gender Aspects			
5. Did you carry out specific Gender Equality Actions under the project?			NO
6. Which of the following actions did you carry out and how effective were they?			N/A
7. Was there a gender dimension associated with the research content-s i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?			NO
E Synergies with Science Education			
8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?			
	Text Mining course taught to 70 3rd Year BSc students (VUA)		
	Masterlanguage "Kan de computer concreetheid meten" (VUA)		
	Keynote Toptoets 2014 (VUA)		
	Several lectures at Studium Generale sessions in the Netherlands (VUA)		
	Lectures Alumni students (VUA)		
	Lecture Klokhuis (science education for children 8-12) (VUA)		
	Science Lecture Museon (VUA)		
	Two University of Trento students did an internship on KnowledgeStore activities (FBK)		
	Annotation by Humans and Machines: specialised course in Digital Humanities taught to 17 students (VUA)		
	HAP-LAP and EMLCT Master programs: hands-on lab sessions of various courses with NWR NLP pipeline 2014/2015 academic year: 12 students; 2015/2016 academic year: 11 students (EHU)		
	User evaluation of the SynerScope software with high-school students (SYN)		
9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?			
	A Massive open online course on the use of SynerScope software (SYN)		
F Interdisciplinarity			
10. Which disciplines (see list below) are involved in your project?			
Main discipline:			
Associated discipline:		Associated discipline:	
G Engaging with Civil society and policy makers			
11a Did your project engage with societal actors beyond the research community? (if 'No', go to Question 14)			YES

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?			
Yes - in implementing the research			
Yes, in communicating /disseminating / using the results of the project			
11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?			YES
12. Did you engage with government / public bodies or policy makers (including international organisations)			
Yes- in framing the research agenda			
Yes - in implementing the research agenda			
Yes, in communicating /disseminating / using the results of the project			
13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?			
Yes- as a primary objective (please indicate areas below- multiple answers possible)			
Yes- as a secondary objective (please indicate areas below - multiple answer possible)			
13b If Yes, in which fields?			
Information Society, Public Health, Research and Innovation			
13c If Yes, at which level?			
Local / regional levels			
National level			
European level			
International level			
14. How many Articles were published/accepted for publication in peer-reviewed journals?			11
To how many of these is open access ^[2] provided?			0
How many of these are published in open access journals?			7
How many of these are published in open repositories?			11
To how many of these is open access not provided?			4
Please check all applicable reasons for not providing open access:		lack of time and resources	
15. How many new patent applications ('priority filings') have been made? ("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).			N/A
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).			
	Trademark	0	
	Registered design	0	
	Other	0	
17. How many spin-off companies were created / are planned as a direct result of the project?			
Indicate the approximate number of additional jobs in these companies:			0
18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:			
	Increase in employment, or	In small & medium-sized enterprises	

	Safeguard employment, or	In large companies
19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (FTE = one person working fulltime for a year) jobs:		
Difficult to estimate / not possible to quantify		
I Media and Communication to the general public		
20. As part of the project, were any of the beneficiaries professionals in communication or media relations?		NO
21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?		NO
22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?		
	Press Release	Coverage in specialist press
	Media briefing	Coverage in general (non-specialist) press
	TV coverage / report	Coverage in national press
	Radio coverage / report	Coverage in international press
	Brochures /posters / flyers	Website for the general public / internet
	DVD /Film /Multimedia	Event targeting general public (festival, conference, exhibition, science café)
23 In which languages are the information products for the general public produced?		
	English	Other language(s)