NewsReader (project nr. 316404):

recording history from daily news streams

website: http://www.newsreader-project.eu

Coordinator: Prof. Dr. Piek Vossen, VU University Amsterdam
Start date: January 2013
End date: January 2016

CONSORTIUM
1. (Coordinator) Faculteit of Arts, Vrije University Amsterdam, the Netherlands.
2. Fondazione Bruno Kessler, Trento, Italy
3. Euskal Herriko Unibertsitatea, San Sebastian, Spain
4. LexisNexis, Amsterdam, the Netherlands
5. ScraperWiki, London, United Kingdom
6. Synerscope, Eindhoven, the Netherlands

Funded by the European Union FP7 Work Programme under Call FP7-ICT- 2011-8 – Objective 'Cooperation' – Research theme 'Information and Communication Technologies' - Challenge 4.4 - Area Intelligent Information Management. EU contribution is 2.8 million Euros (total cost: €3.77m).

Headline

A History Recorder is a computer program that "reads" daily streams of news and stores exactly what happened, where and when in the world, and who was involved. The program uses the same strategy as humans to build a story by merging new incoming information with previously stored. In contrast to humans, the History Recorder can do this for massive amounts of news streams for long periods of time and will not forget any details.

Abstract

The European NewsReader project[1] develops a '*history-recorder*' to process daily streams of news in 4 language (English, Spanish, Italian and Dutch) to determine **what** happened, **where** and **when** in the world, and **who** was involved. The history-recorder applies the same strategy as humans to merge news with previously stored information, creating a longer-term history rather than storing separate events. Like humans, the program removes duplicate information and complements incomplete information while 'reading'. The result is a single storyline for all the events. Unlike humans, NewsReader will not forget any detail, will be able to recall the complete story as it was told, know who told what part of the story, and identify what sources contradict each other. Since it keeps track of all the original sources of information, the history-recorder can also provide insights into how the story was told. This will tell us about the different perspectives from which different media sources present the news, both the news of today and the news of the past. NewsReader is a collaboration of three research groups and three companies (LexisNexis, ScraperWiki and Synerscope). NewsReader will be tested on ***economic-financial*** **news** relevant for ***political and financial*** decision-makers. About 25% of the news is about

finance and economy. LexisNexis estimates the total volume of daily news items for this domain on about five-hundred-thousand.

Problem addressed

The volume of news data is enormous and expanding, covering billions of archived documents and millions of documents as daily streams, while at the same time getting more and more interconnected with knowledge provided elsewhere. Professional decision-makers that need to respond quickly to new developments and knowledge or that need to explain these developments on the basis of the past are faced with the problem that current solutions for consulting these archives and streams no longer work, simply because there are too many possibly relevant and partially overlapping documents and they still need to distinguish the correct from the wrong, the new from the old, the actual from the out-of-date by reading the content and maintaining a record in memory. Consequently, it becomes almost impossible to make well-informed decisions and professionals risk to be held liable for not being able to handle the incoming information streams

An example

To explain the potential of NewsReader, we describe a more detailed example. We search events on *takeovers* and *buying stakes* involving *Volkswagen* and *Porsche*, as illustrated in Table-1 for a small selection of text snippets from online news. We also obtain information on the publication date. When ordered by date, a story unfolds in which, first, *Porsche* buys stake in *Volkswagen* with the assumed intention of taking over *Volkswagen*. The year 2009 turns out to be critical. *Porsche* first plans a takeover of *Volkswagen* but instead *Volkswagen* buys a large stake in *Porsche* and finally takes over *Porsche*. In that year, the economy collapses and luxury sports cars are less wanted. Other events spinn off from this event, such as hedge funds suing *Porsche* and its executives (among which the CEO *Wiedeking*) for not buying *Volkswagen* and *Volkswagen*, being one of the world biggest car producers, buying other companies. Many mentions of the takeover of *Volkswagen* by *Porsche* thus turn out to be speculative and never took place. This is indicated by various markers in the text, such as *failed to*, *could*, *plans to take full control*, *adjust or delay plans*. The true story is therefore more basic than the anticipated stories in the news. Eventually from all the mentions in the text, one event remains of a *takeover* of *Volkswagen* by *Porsche* that never took place and one event of a *takeover* of *Porsche* by *Volkswagen* that took place.

Table 1: Structured data extracted from textual sources revealing car industry take overs

Date	Who	Who	What	Textual Source
2005	VW	Porsche	Take(Porsche, 18.65% stake, VW)	In late 2005, Porsche took an 18.65% stake in the Volkswagen Group, further cementing their relationship, and preventing a takeover of Volkswagen Group
2007,Mar, 26	VW	Porsche	Take (Porsche, 30.9% stake, VW)	On 26 March 2007, Porsche took its holding of Volkswagen AG shares to 30.9%, triggering a takeover bid under German Law.
2007,Oct, 22	VW	Porsche	Takeover(Porsche, VW)	Porsche could launch a full takeover bid for Volkswagen, Europe's biggest car manufacturer, this week if the EU's highest court makes its widely expected decision to ban a post-war law giving the German state effective control over VW.
2008,Sep ,16	VW	Porsche	Take (Porsche, 35% stake, VW)	On 16 September 2008, Porsche increased its shares by another 4.89%, in effect taking control of the company, with more than 35% of the voting rights.
2009, Jan 6	VW	Porsche	Take (Porsche, 50.8% stake, VW)	Porsche AG took step closer to controlling the much larger Volkswagen AG by upping its share holdings to 50.8% in late Monday trading.
2009, Apr,7	VW	Porsche	Takeover(Porsche, VW)	With present economic conditions shrinking Porsche's available cash, the automaker may have to adjust or delay its plans to gain full control of Volkswagen. In January, Porsche raised its stake in Volkswagen to 50.76% gaining a majority stake.
2009,Jun, 29	VW	Porsche	Takeover(VW, Porsche)	29-June-2009 Porsche Rejects VW Takeover Offer. The power struggle between German automakers Porsche and Volkswagen escalated on Monday with Porsche rejecting VW's takeover bid as unfeasible.
2009,Jul, 23	VW	Porsche	Takeover(VW, Porsche) Step-aside (Wiedeking, Porsche)	23 Jul 2009 – Porsche Chief Executive Wendelin Wiedeking has stepped aside in a sign that Volkswagen takeover of its local rival is almost secured.
2009,Jul 23	VW	Porsche	Takeover(VW, Porsche)	VW takeover of Porsche edges closer. Resignations of executives at German sports car maker opens way for deal. A Volkswagen takeover of Porsche moved closer on Thursday with the resignation of the two executives running the German sports car maker.
2009,Jul, 31	VW	Porsche	Takeover(VW, Porsche)	31-July-2009 - The board of the giant German carmaker Volkswagen agreed Thursday to forge a new global automotive powerhouse by mounting a two-step 8-billion-euro ($11.4 billion) takeover of the legendary sports vehicle group Porsche.
2009	VW	Porsche	Take(VW, 49.9% stake, Porsche)	Volkswagen will initially take a 49.9 percent stake in Porsche AG by the end of 2009, and it will also see the family shareholders selling the automobile trading business of Porsche Holding Salzburg to Volkswagen AG

Our Approach

NewsReader processes news through a complex pipeline of modules that interact with a central repository of knowledge, called *KnowledgeStore* (Corcoglioniti et al. 2013), to store the results. The main tasks of the modules are: (1) entity/event mention recognition, to identify all the mentions of relevant entities and events in the input texts; (2) entity and event co-reference: to cluster the entity/event mentions that refer to the same real world entity/relation; (3) integrate all the information about entities/event contained in the co-referring mentions into a unique semantic structure to be stored in the *KnowledgeStore* and, when possible, associate it with a URI for the entity/event available in the *KnowledgeStore*. (4) detect factuality and context of events, which will detect if an event mentioned in the text really happened or is only a reported event (e.g., a prediction or an opinion about an event) and the context in which the event happened (5) find relations between events, like causal and temporal relations, which allow to structure events into consistent stories, called *storylines*. The result is an extremely compact structure shown in Figure-1.

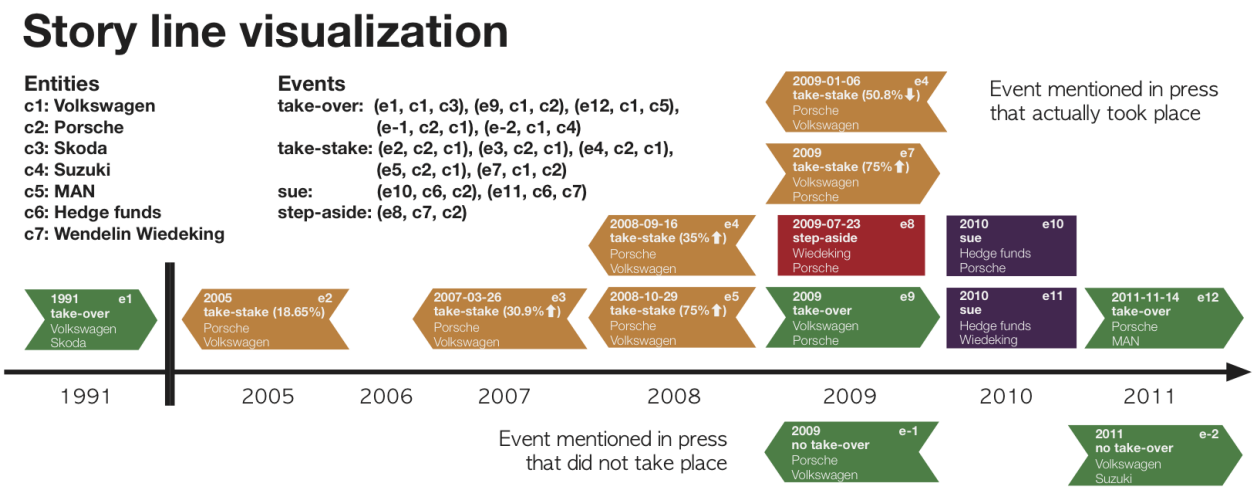

Figure 1: Participant-centred time-line view

Figure-1 shows the compact structure for our example that visualises the events as a sequence involving a selection of participants and an indication of trust (continuous line versus dotted line) and factuality (above or below the timeline). Such a representation can be used to quickly navigate storylines. Stories can be represented in various ways, depending on the narrative that the user selects: participant-centred, event-centred, location-centred, individual-centred or group/type-centred, something that is not possible in text-based summaries that are static. Synerscope will develop different visualisations in NewsReader, such as: timelines, maps, tables, graphs, plot-browsers, to show trends, correlations, generalizations, predictions and reasoning over the data.

Overall architecture

The overall architecture for the NewsReader platform is given in Figure-2. A range of NLP modules is deployed around the central *KnowledgeStore*. The modules use input-output textstream APIs according to the Natural-language-processing Annotation Format (NAF, for reference see paper also submitted to this conference). NAF, a successor of KAF (KYOTO Annotation Format, Bosma et al 2009), is a layered standoff-representation, following existing W3C, ISO and de-facto standards such as the Linguistic Annotation Format (Ide & Romary 2004). It enables to add new analyses to documents using, but not affecting the output of previous modules. Each of the modules can store its analysis as a NAF-resource in the *KnowledgeStore* and retrieve it from there. Alternatively, the modules can pass the output to each other when organized as a pipeline, in which case there is no intermediate storage. The modules are further grouped according to their independence on each others output. This allows for parallel distribution of the processing of texts. In the final paper, we will elaborate more on the different modules and architecture aspects of the system.

The *KnowledgeStore* has different components for the different type of data. In addition to the original sources, e.g. XML files provided by LexisNexis, pointers to resources, ***mentions*** of events and entities are stored in an Hbase and Hadoop platform with a specification of the context in which entities and events are mentioned in sources. Ultimately, mentions in sources are mapped to relations between ***instances*** of events and

entities, represented as RDF triples stored in a separate triple store. The triples are related to the sources through named-graphs, which allows us to formally model provenance of all information that is stored in the KnowledgeStore. The systematic separation of event/entity ***mentions*** and event/entity ***instances*** follows a formal model for the semantic interpretation defined in the Grounded Annotation Framework[2] (GAF, Fokkens et al 2013).

Whereas most modules for NLP interact with each other or with the Hbase/Hadoop part of the *KnowledgeStore*, the final modules in the chain access the RDF data to compare new event descriptions with past events. This involves intense reasoning over the data stored for the past. Basically it consists of two steps:

1. For all news published on a single day, we determine which sources discuss the same events in the 'real' world (even if using different wordings).
2. Once event descriptions are grouped through event-coreference relations, we relate these potential event-instances with all the instances stored in the past to determine whether they truly represent new events in the world. This process is defined as historical-event-coreference
3.
4. After determining coreference it is possible to aggregate information on the same events from the different sources and to create or unfold longer storylines from event sequences.

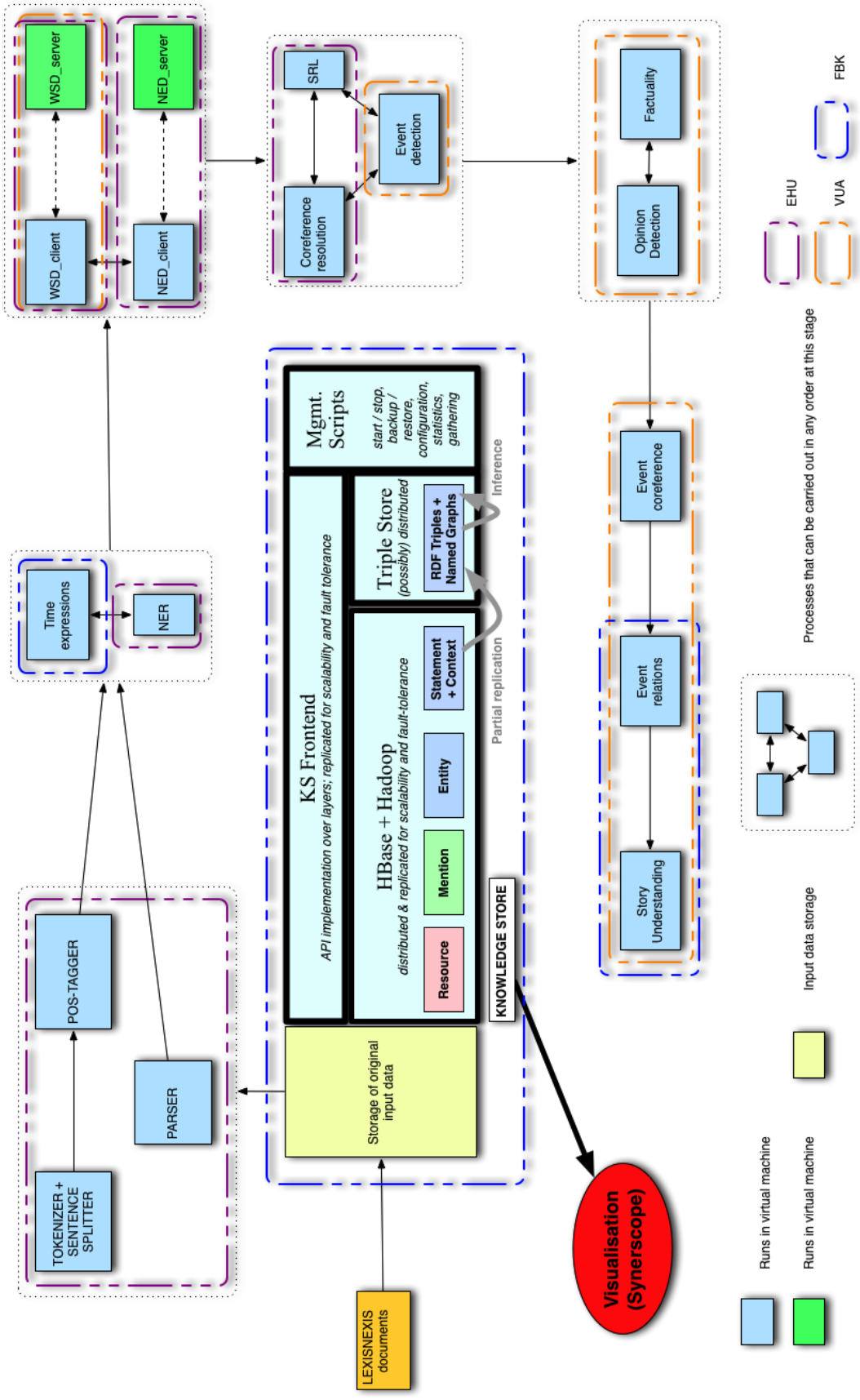

Figure 2: Overall architecture

Use cases and evaluation

In the first year of the project, we defined 4 use-cases to carry out user-evaluations in the first year of the project:

- TechCrunch/Crunchbase: a well-structured wiki-like database (Crunchbase) and news documents (TechCrunch) on the same topic: information technology. We anticipate that events in structured Crunchbase data will be reflected in non-structured TechCrunch articles;
- Dutch House of Representatives: focusing on information-intensive Parliamentary Inquiries, we identified several challenges:
    - coverage: 'understanding' an event, its key actors and entities
    - mapping the gaps: identifying areas with insufficient information coverage
    - creating networks of events, people and entities (companies, government bodies)
    - fact checking
- Global Automotive Industries: using a large, multilingual data set on the automotive industry, NewsReader will help (re)construct:
    - complex structures, ie the ownership structures of automotive conglomerate
    - complex events, ie mergers, acquisitions and corporate restructuring in this industry
- Business Intelligence: gathering information about companies to evaluate them as potential business partners. This evaluation may relate to a business' ability to repay a loan, to comply with anti-money laundering legislation or to carry out regular due diligence investigations.

The next image (Figure-3) is automatically generated from data produced for documents on Volkswagen and its factories in the Belgium city Vorst. Around 2006, Volkswagen decided to move production factories from Pamplona to Vorst and a few years later from Vorst to Germany and other countries. This had a big social-economic impact on Vorst. The image shows participants connected through events as story lines in time, as extracted from the news. This visualisation is based on about 50 documents with shallow processing. When large volumes of data are processed, more complex graph-visualisations are needed. Figure-4 shows the big-data visualisation developed by Synerscope and applied to the Crunchbase data. It shows a table view of investors in an IT company but also a graph view on all connections. Through filtering, selection and expansion, users can find correlations and patterns.

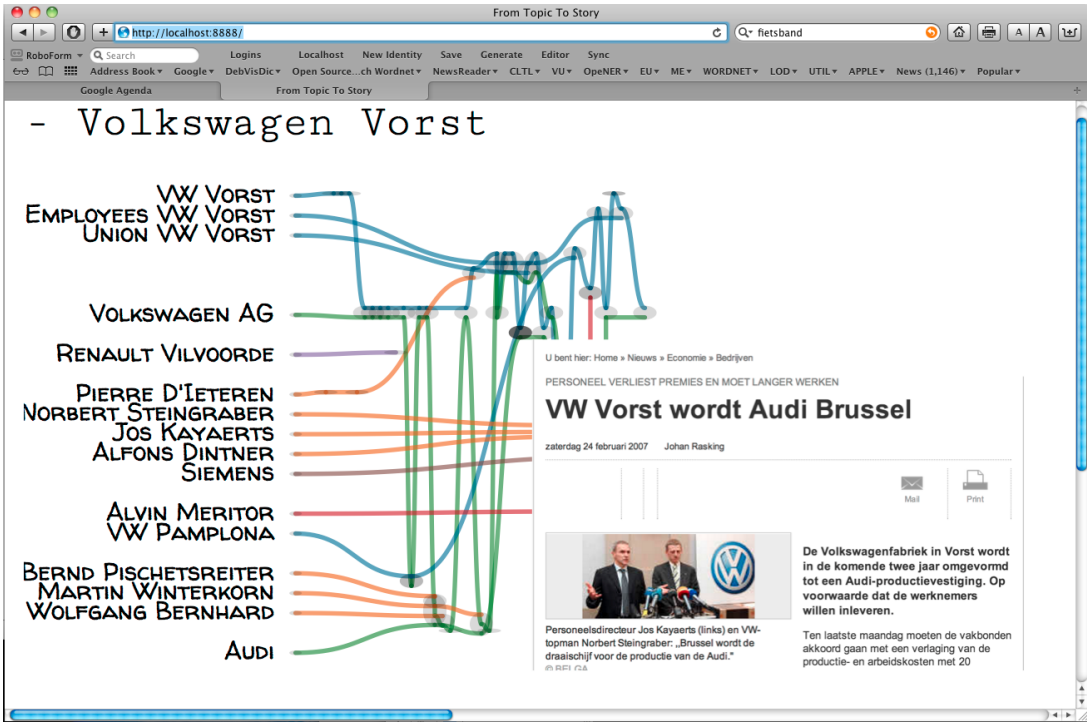

Figure 3: Automatic visualisation of story lines from data on Volkswagen and Vorst

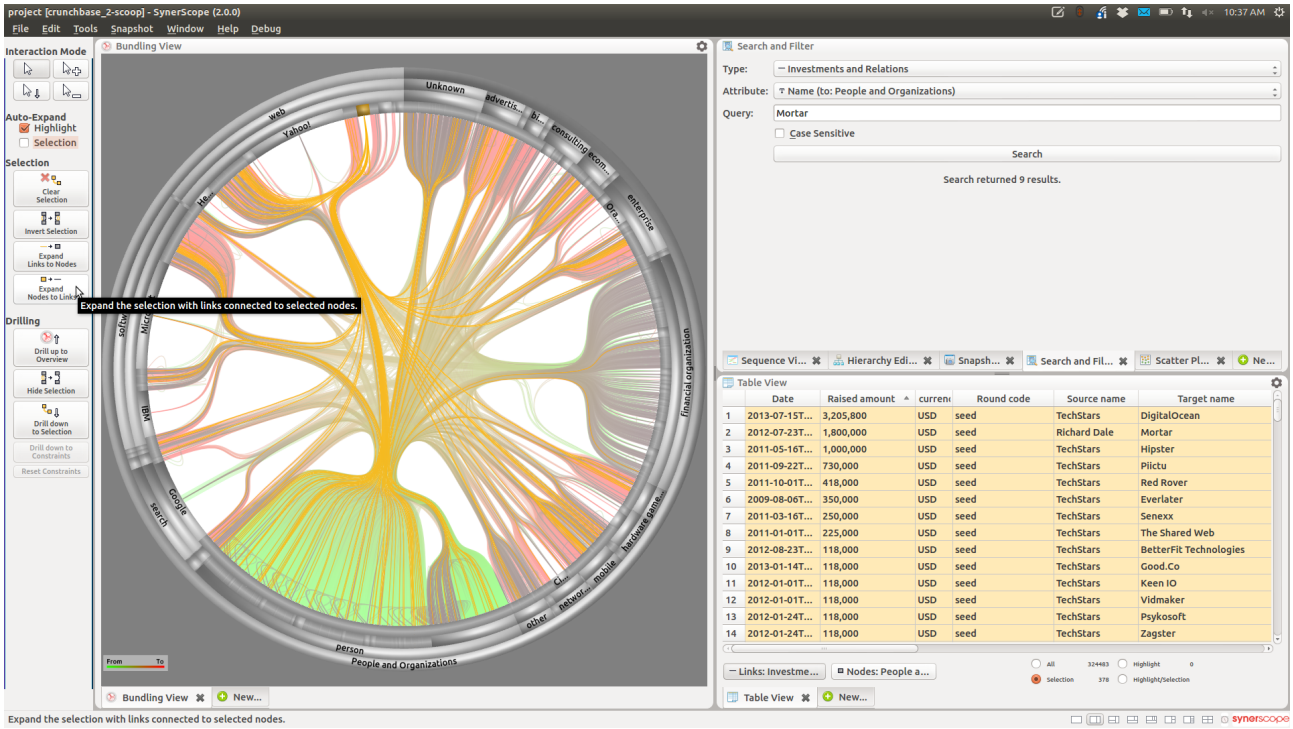

Figure 4: Marcoto tool developed by Synerscope for visualising big data graphs

Big data

Big data is one of the central aspects of the project. Estimates show that each working day, over 2 million news articles are produced. On the basis of 416 news articles that were manually annotated for their major events and participants, we derived estimates for processing a data volume of 1 million items, which is shown in Figure-5.

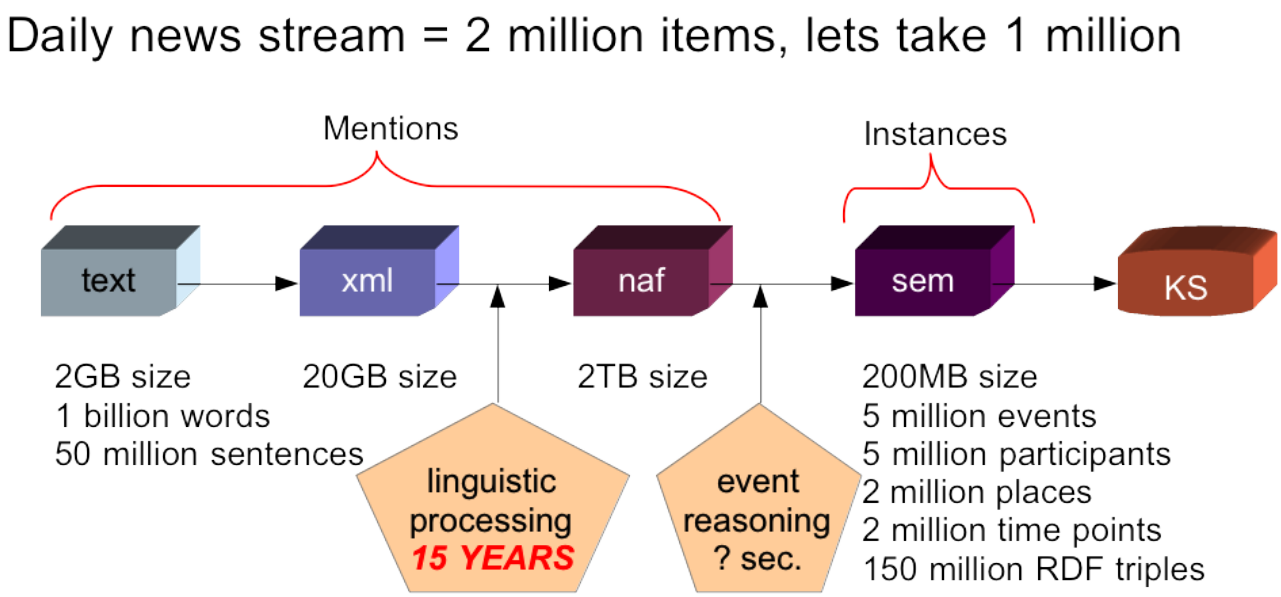

Figure 5: Estimated volumes for news on a regular working day

On a single day, 2GB of text will come in (1 billion words and 50 million sentences). In XML, this is 20GB and after processing the text and representing it in the Natural Language Processing Annotation Format (NAF) it becomes 2TB. On a single linux server with 8 cores and 12GB RAM it will take 15 years to process the data. This makes clear that the project needs to develop heavy parallelization of processing to be capable to process a single-day's batch of news before the next day.

After processing the text, we need to interpret the mentions of events and participants as instances, determining which mentions refer to the same instances. This generates another data layer which is more compact since many mentions refer to the same instances. Still this adds another 200MB of data, while the processing time is currently unknown since the software is still an experimental prototype. The ultimate result is stored in the KnowledgeStore (KS). We estimate roughly that a single day of news thus involves 5 million events and 5 million participants, related to 2 milion places and time points references. Locations and time-points have not been normalised to instances in this calculation.

The news of a singel day is processed as one batch, where we plan to apply clustering of news across sources for main topics. Within each topic-cluster, we next divide the news by the time and place indications. Different sources can only talk about the same events if both time and place match: a hostage in Bosnia is not the same event as a hostage in Somalia on the same day. With a cluster and time-place boudnaries, we next compare the linguistic expressions (the mentions) to find out what news talks about the same events. These events of the day are then compared to the events of the past to find out if they represent new events or old events already identified in the past. This process is schematically represented in Figure-6 below. Dealing with the news of the day, thus will involve large scale computing and high parallelization of software and data access.

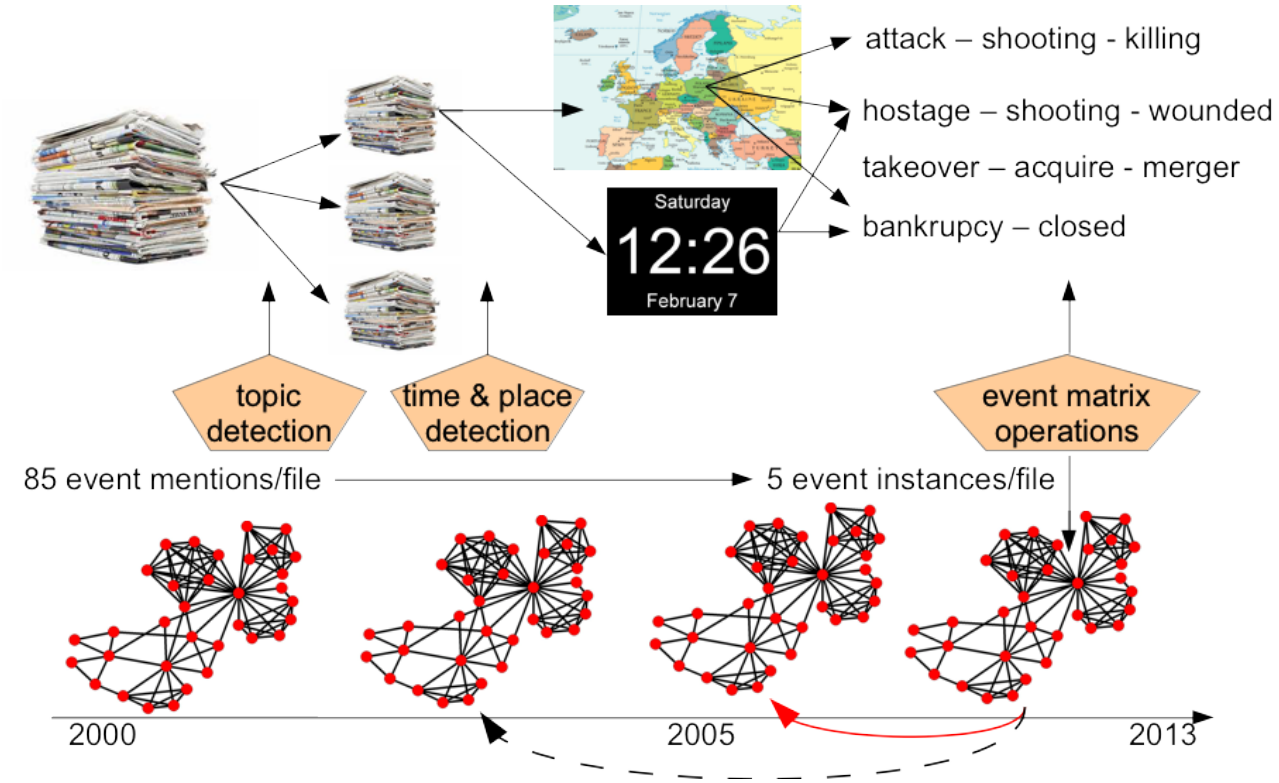

Figure 6: Schema for event processing on a daily basis, relating it to the past

Formal representation format

NewsReader defined a formal representation to make a fundamental distinction between mentions of events and entities in text and the instances in the 'real' world. This is achieved by combining two representation frameworks coming from the Natural Language community and the Semantic Web Community. Our approach is illustrated by the exampe in Figure-7. It shows two different earthquakes and tsunamies (2004 and 2009) depicted on a timeline that assumes events in the world. These are instances of events. The second timeline is used to mark mentions of events in sources at the time the source is 'published'. These sources can be anything from sensors, database to textual sources. The example shows that sources can talk about recent events, furture events and reflect on past events. Each time sources provide information on an event this will be related to the assumed instance. The information can be a repetition of old information, adding new information or contradicting other information. In this example, some sources report immediately after the 2004 tsunami but also later, in which case they start talking about possible future tsunamies and a tsunami warning system. In 2009, another tsunami takes place and the sources not only report on this new event but also reflect back on the event of 2004. Yet other sources speculate in 2013 that the tsunami in 2004 was not caused by the tectonic plates but by a US military vessel that tested a secret weapon in the region. The 2013 source thus introduces a new event before the tsunami of 2004.

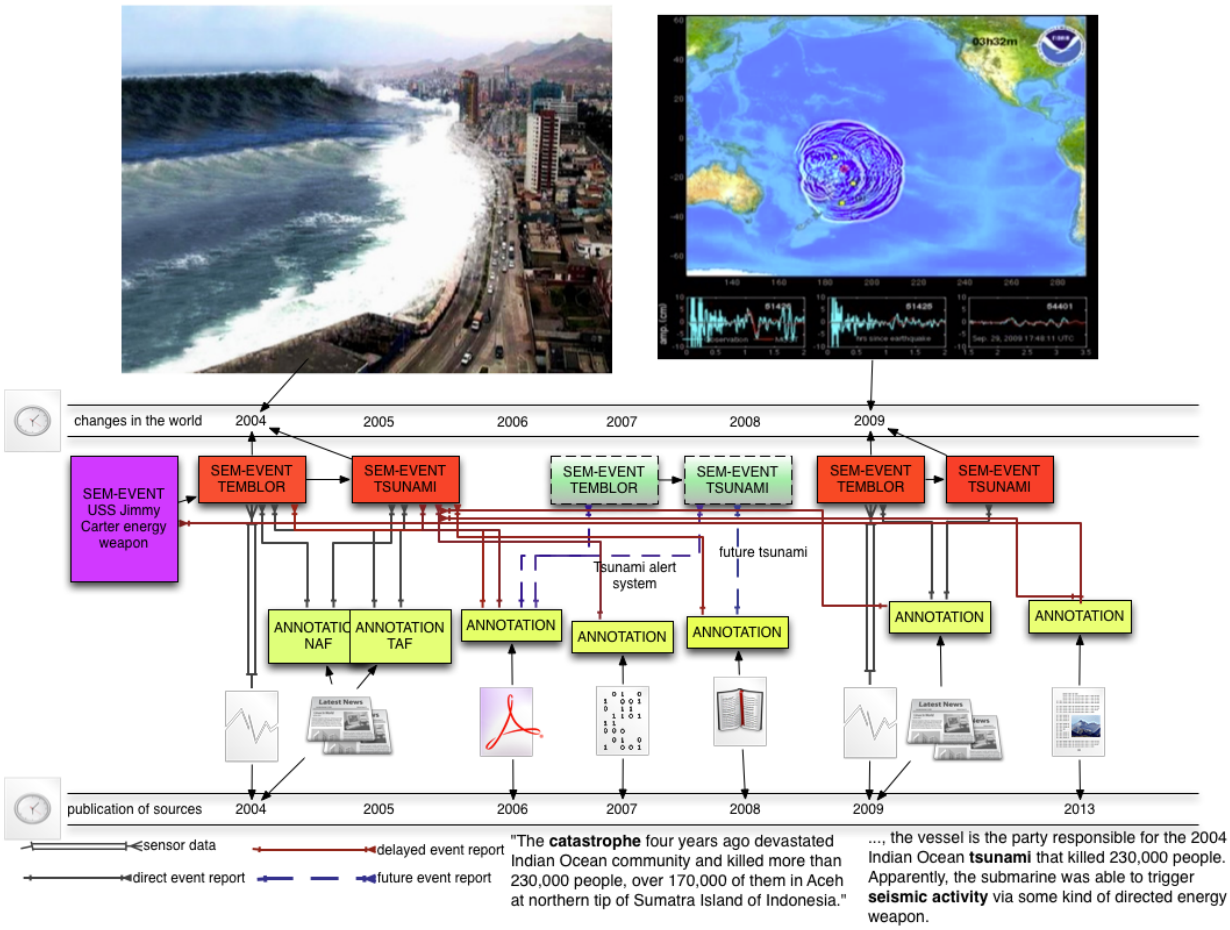

Figure 7: Example of NewsReader formal approach to mentions and instances

The formal model to handle this distinction between mentions and instance is the Grounded Annotation Framework (GAF, http://groundedannotationframework.org/) that was developed within NewsReader. Figure-8 shows a graphical represention of a single sentence referring to the first earthquake and tsunami. Mentions of information are represented as named-graphs bound to their source and eventually end up in a semantic representation of the assumed instance of the event. Different sources can contribute to the semantic represenation and their contribution is fully traceable. GAF thus provides a way to trace the provenance of any piece of information on events and entities to the sources. This adds another valuable dimension to NewsReader: users not only get an overview of long-term developments and trends but also insight in the variety of sources that tell the same story or different stories.

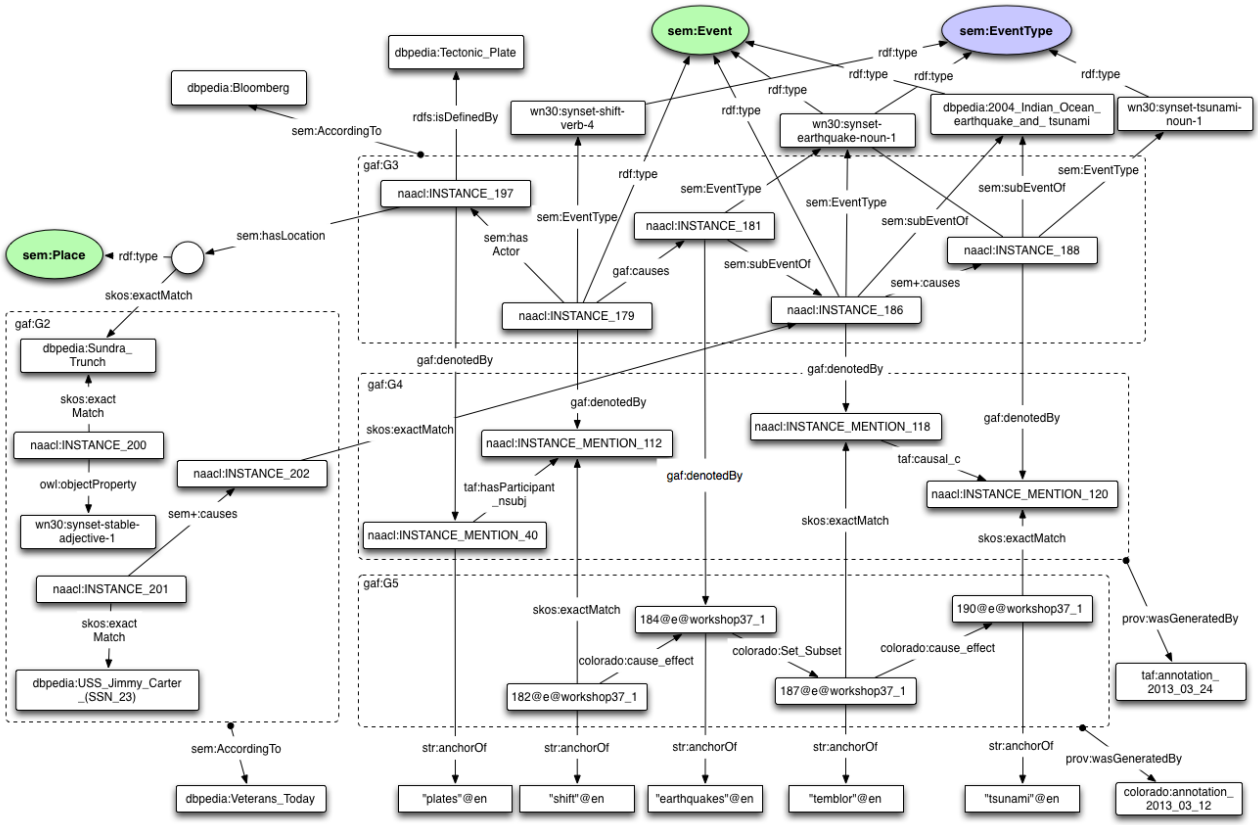

Figure 8: Example of the Grounded Annotation Framework modeling the information on the 2004 earthquake and tsunami

References

W. E. Bosma, P. Vossen, A. Soroa, G. Rigau, M. Tesconi, A. Marchetti, M. Monachini, and C. Aliprandi. 2009. Kaf: a generic semantic annotation format. In Proceedings of the GL2009 Workshop on Semantic Annotation.

Francesco Corcoglioniti, Marco Rospocher, Roldano Cattoni, Bernardo Magnini and Luciano Serafini. 2013. Interlinking Unstructured and Structured Knowledge in an Integrated Framework. In: Proceedings of the 7th IEEE International Conference on Semantic Computing (ICSC). Irvine, USA, September 16-18 2013.

Antske Fokkens, Marieke van Erp, Piek Vossen, Sara Tonelli, Willem Robert van Hage, Luciano Serafini, Rachele Sprugnoli and Jesper Hoeksema. 2013. GAF: A Grounded Annotation Framework for Events. In: Proceedings of the 1st workshop on Events: Definition, Detection, Coreference and Representation at NAACL2013. E. Hovy, T. Mitamura and M. Palmer (Eds). p. 11-20.

Nancy Ide and Laurent Romary, 2004. International Standard for a Linguistic Annotation Framework. Journal of Natural Language Engineering, 10:3-4, 211-225

Piek Vossen, Eneko Agirre, German Rigau and Aitor Soroa. KYOTO: a knowledge-rich approach to the interoperable mining of events from text. Bookchapter in: New Trends of Research in Onotologies and Lexical Resources. A. Oltramari, P. Voosen, L. Qin and E. Hovy (Eds.). Springer Verlag. Heidelberg, Germany. p. 59-83.

Publications

Vossen P., Ch. Fellbaum (Eds.), Proceedings of the 7th Global WordNet Conference (GWC2014), January 25-29, 2014, Tartu, Estonia.

Maddalen López de Lacalle, Egoitz Laparra and German Rigau, First steps towards a Predicate Matrix. Proceedings of the 7th Global WordNet Conference (GWC2014), January 25-29, 2014, Tartu, Estonia.

Thomas Ploeger, Maxine Kruijt, Lora Aroyo, Frank de Bakker, Iina Hellsten, Antske Fokkens, Jesper Hoeksema and Serge ter Braake. (2013) "Extractivism. Extracting activist events from news articles using existing NLP tools." In: Proceedings of the ISWC workshop DeRiVE 2013. Sydney, Australia, October 21, 2013

Marieke van Erp, Antske Fokkens, Piek Vossen, Sara Tonelli, Willem Robert van Hage, Luciano Serafini, Rachele Sprugnoli and Jesper Hoeksema (2013) "Denoting Data in the Grounded Annotation Framework". In: Proceedings of the ISWC 2013 Posters & Demos Track. Sydney, Australia, October 20-24, 2013.

Francesco Corcoglioniti, Marco Rospocher, Roldano Cattoni, Bernardo Magnini, and Luciano Serafini (2013) "Interlinking Unstructured and Structured Knowledge in an Integrated Framework", In 7th IEEE International Conference on Semantic Computing (ICSC), Irvine, CA, USA, September 16-18, 2013.

Cybulska, Agata and Piek Vossen (2013): Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution, Proceedings of RANLP 2013, Hissar, Bulgary, September 7-13, 2013.

Marco Fossati, Claudio Giuliano, Sara Tonelli (2013) Outsourcing FrameNet to the Crowd. (short paper), The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Sofia, Bulgaria, August 4-7, 2013.

Fokkens A., M. van Erp, P. Vossen, S. Tonelli, W. van Hage, L. Serafini, R. Sprugnoli, J.

Hoeksema (2013) "GAF: A Grounded Annotation Framework for Events", in: 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), E. Hovy, T. Mitamura, M. Palmer (Eds.), Association for Computational Linguistics (Publ.), ISBN 978-1-937284-47-3, p.11-20, Atlanta, USA, June 9-15, 2013.

Aldabe I., Maritxalar M., Lopez de Lacalle O. 2013, EHU-ALM: Similarity-Feature Based Approach for Student Response Analysis Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, USA, June 14-15, 2013.

Article on NewsReader to be published in the 'Kamerbode'; internal Newsletter of the Dutch House of Representatives.

Laparra E. and Rigau G. ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling. The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Sofia, Bulgaria, August 4-7, 2013.

Antske Fokkens, Marieke van Erp, Marten Postma, Ted Pedersen, Piek Vossen and Nuno Freire (2013) Offspring from Reproduction Problems: What Replication Failure Teaches Us. The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Sofia, Bulgaria, August 4-7, 2013. **Runner up best paper award**

Antske Fokkens, Marieke van Erp, Piek Vossen, Sara Tonelli, Willem Robert van Hage, Luciano Serafini, Rachele Sprugnoli and Jesper Hoeksema (2013) GAF: A Grounded Annotation Framework for Events. The 1st Workshop on Events: Definition, Detection, Coreference, and Representation. Workshop in conjunction with NAACL-HLT 2013, Atlanta, GA, USA, June 14, 2013.

Marieke van Erp, Giuseppe Rizzo and Raphaël Troncy (2013) Learning with the Web: Spotting Named Entities on the intersection of NERD and Machine Learning (preprint) #MSM2013 Concept Extraction Challenge. Rio de Janeiro, Brazil, May 2013. System ranked 2nd in challenge

Laparra E. and Rigau G. Sources of Evidence for Implicit Argument Resolution. 10th International Conference on Computational Semantics (IWCS2013). Postdam, Germany, March 19-22, 2013.

Media coverage

Piek Vossen: National radio interview on Spinoza Price and NewsReader @ "OBA Live" with Theodor Holman, Openbare Bibliotheek (OBA), Amsterdam, the Netherlands, October 11, 2013.

Piek Vossen: National radio interview @ BNR Newsradio, September 27, 2013.

Piek Vossen: interview I/O Magazine

Piek Vossen: national radio interview @ "Dit is de Dag", Hilversum, the Netherlands, July 23, 2013.

Piek Vossen: Invited guest @ regional radio L1 De Stemming, café Forum, Maastricht, the Netherlands, June 30, 2013.

Piek Vossen: Interview in regional newspaper Weespernieuws, August 7, 2013.

Piek Vossen: interview @national newspaper **Parool**, June 11, 2013.

Piek Vossen: interview @national newspaper **Volkskrant**, June 11, 2013.

Piek Vossen: interview @national newspaper **Trouw**, June 11, 2013.

Piek Vossen: interview @national newspaper **Algemeen Dagblad**, June 11, 2013.

Piek Vossen: interview @national newspaper **NRC Handelsblad**, June 11, 2013.

Piek Vossen: interview @ national newspaper **NRC Next**, June 11, 2013.

Piek Vossen: interview @**Kennislink**, June 10, 2013.

Piek Vossen: interview in **New Scientist**, June 10, 2013.

Piek Vossen:: interview @**Ad Valvas**, June 11, 2013.

Piek Vossen: interview @national radio **NOS Nieuws**, Radio 1, June 10, 2013.

Piek Vossen: interview @national radio NTR **Hoe?Zo! Radio**, June 10, 2013.

http://www.dutchdailynews.com/dutch-university-to-develop-history-recorder/

Piek Vossen: Invited speaker in Casa Luna, Dutch National radio interview on new EU-project **NewsReader**, Mediapark, Hilversum, October 8, 2012 (announcement video on Youtube by Casa Luna; @ Radio 1, NCRV Casa Luna, part 1 (on NewsReader); @ Radio 1, NCRV Casa Luna, part 2 (live interaction with listeners)

Blog from Marco Leeuwerink on the interview at Casa Luna on **Extend Limits**.Piek Vossen: Invited speaker in Dutch radio interview on new EU-project **NewsReader**, in Radio Station FunX, Amsterdam, May 10, 2012.

Workshops:

FBK workshop Perspectives on events (http://www.newsreader-project.eu/fbk-workshop-perspectives-on-events/), including two talks on Newsreader:

- Luciano Serafini: ”Events in Logic“
- Bernardo Magnini and Tommaso Caselli ”Events in Computational Linguistics“

Symposium on Events and Text Mining, Radboud University Nijmegen, organised by CLTL (VUA) and CLS (Nijmegen), November 13th 2013.

Invited talks and interviews

2014:

Piek Vossen: Invited keynote speaker on "Kunnen computers ooit de taal van mensen begrijpen" at the "Specialistenserie": theme: Imitation, Museon, Den Haag, the Netherlands, February 7, 2014.

2013:

http://www.youtube.com/watch?v=Wjbh1nu3I1Y

Luciano Serafini: Presentation at the Trento Workshop on Fact checking, 2013

Marco Rospocher: Presentation on "Interlinking Unstructured and Structured Knowledge in an Integrated Framework" at the Seventh IEEE International Conference on Semantic Computing (IEEE-ICSC2013), Irvine, USA, September 16-18, 2013.

Piek Vossen: interview I/O Magazine.

Piek Vossen in the "Nationale Wetenschapsquiz", national television, December 29, 2013.

Piek Vossen: video-interview VPRO.

Piek Vossen: video-interview on NewsReader in "De Voorlopers", SURFnet & Fast Moving Targets.

Piek Vossen: Invited keynote speaker on NewsReader at Studium Generale: one presentation for foreign students (knowledge engineering, european studies, university college) in English: "the History Recorder: Today's News is Tomorrow's History" and one presentation for broader public of region Maastricht (HOVO-publiek) in Dutch: "De geschiedenisrecorder: het nieuws van vandaag is de geschiedenis van morgen", Maastricht University, Maastricht, the Netherlands, December 9, 2013.

Piek Vossen: Invited keynote speaker at NOTaS-event Taal in Bedrijf 2013, Nijmegen, the Netherlands, November 29, 2013.

Piek Vossen: Invited keynote speaker at ICT.OPEN 2013, Eindhoven, the Netherlands, November 27-28, 2013.

Piek Vossen: Invited keynote speaker @ "Gala van de Amsterdamse Wetenschap", in cooperation with Parool and Folia, Stadsschouwburg, Amsterdam, the Netherlands, November 26, 2013 (buy tickets).

Piek Vossen: Invited keynote speaker at 1st Annual Netherlands eScience Symposium, by Netherlands eScience Center (NleSC), The Rosarium, Amstel Park, Amsterdam, the Netherlands, November 7, 2013.

Marieke van Erp: poster presentation on NewsReader at The 12th International Semantic Web Conference, Sydney, Australia, October 23, 2013 (video trailer)

Marieke van Erp: Invited talk at Schwa lab, The University of Sydney, School of Information Technologies, Sydney, Australia, October 22, 2013.

Thomas Ploeger: Presentation at the DeRiVE 2013 Workshop. "Extractivism. Extracting activist events from news articles using existing NLP tools". Sydney, Australia, October 21, 2013.

Piek Vossen: Invited keynote speaker at Nederlandse Vrouwenclub Amsterdam/Internationale Lyceumclub, Industrieele Groote Club, Amsterdam, October 16, 2013.

Piek Vossen: Invited guest on Spinoza Price and NewsReader @ "OBA Live" with Theodor Holman, Openbare Bibliotheek (OBA), Amsterdam, the Netherlands, October 11, 2013.

Piek Vossen: Invited keynote speaker on NewsReader at the Weekend van de Wetenschap in samenwerking met het Klokhuis – thema: Zoek het uit!, organized

by Ministerie van OC&W for children of 8-12 years, Weekend of Science, Museon, Den Haag, the Netherlands, October 5, 2013.

Presentation by Vossen on Spinoza-price, awarded by State Secretary for Education, Culture and Science Mr Sander Dekker and NWO Chairman Professor Jos Engelen, Nieuwe Kerk Den Haag, the Netherlands, September 27, 2013.

Piek Vossen: National radio interview @ BNR Newsradio, September 27, 2013.

Piek Vossen: Invited keynote speaker on NewsReader at 'Informatie aan Zee', a 2-day conference by the Flemish Association for Library, Archive and Documentation (VVBAD), Oostende, September 13, 2013.

Bernardo Magnini: a seminary at INRIA (Nice) on "The KNOWLEDGESTORE: an Integrated Framework for Ontology Population"

Piek Vossen: Invited panel member at the Workshop on the Lexicon in Functional Discourse Grammar, Vienna, Austria, September 5-6, 2013.

Antske Fokkens and Marieke van Erp: Presentation at the Association for Computational Linguistics (ACL 2013), August 4-9, 2013, Bulgaria Sofia (video trailer, slides, video of the presentation)

Piek Vossen: Invited guest at round table meeting with Stichting de Nationale DenkTank on conference on Big Data in 2014, October 10, 2013.Piek Vossen: Invited interview on Big Data @national newspaper Volkskrant, August, 2013.

Antske Fokkens: presentation on NewsReader at the Annual Network Institute Event, Amsterdam, the Netherlands, July 4, 2013.

Piek Vossen: presentation of GAF framework at the 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation, in conjunction with NAACL-HLT, Atlanta, GA, USA, June 14, 2013.

Marieke van Erp: Invited speaker on "From Events to Stories: Different ways of structuring the same bag of events over time", at Soeterbeeck eHumanities Workshop, a 2-day workshop bringing together researchers working in humanities and ehumanities. Ravenstein, the Netherlands, June 13-14, 2013.

Ian Hopkinson: speaker on Data Science London on June 12th, entitled "Scraping and parsing PDFs in Python"

Piek Vossen: Invited guest @ national radio "Dit is de Dag", Hilversum, the Netherlands, July 23, 2013.

Antske Fokkens: Invited speaker at the Delph-in Summit. Linked Open Data and DELPH-IN. St. Wendel, Germany. July 29, 2013.

Piek Vossen: Invited guest @ regional radio L1 De Stemming, café Forum, Maastricht, the Netherlands, June 30, 2013.

Piek Vossen: interview Quest on Spinoza Prize, special issue NWO / Experiment NL, 2013.

Itziar Aldabe: Presentation on "Similarity-Feature Based Approach for Student Response Analysis" at the Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, USA, June 14-15, 2013.

Piek Vossen: Interview in regional newspaper Weespernieuws, August 7, 2013.

Piek Vossen: interview @national newspaper Parool, June 11, 2013.

Piek Vossen: interview @national newspaper Volkskrant, June 11, 2013.

Piek Vossen: interview @national newspaper Trouw, June 11, 2013.

Piek Vossen: interview @national newspaper Algemeen Dagblad, June 11, 2013.

Piek Vossen: interview @national newspaper NRC Handelsblad, June 11, 2013.

Piek Vossen: interview @ national newspaper NRC Next, June 11, 2013.

Piek Vossen: interview @Kennislink, June 10, 2013.

Piek Vossen: interview in **New Scientist**, June 10, 2013.
Piek Vossen:: interview @**Ad Valvas**, June 11, 2013.
Piek Vossen: interview @national radio **NOS Nieuws**, Radio 1, June 10, 2013.
Piek Vossen: interview @national radio NTR **Hoe?Zo! Radio**, June 10, 2013.
Piek Vossen: invited talk on "**Processing large streams of news in the NewsReader project**" at the Center for Creation, Content and Technology (**CCCT**), University of Amsterdam, the Netherlands, May 24, 2013.
Piek Vossen: Invited keynote speaker on "**Do big data hide or reveal stories? Recording history in the NewsReader project**" at the workshop "Computational Humanities, reference cultures and identity formation", Netherlands Institute for Advanced Study (NIAS), Wassenaar, the Netherlands, April 24-26, 2013.
Piek Vossen: Invited keynote speaker on "**The news of today writes the history for the future Recording history in the NewsReader project**" at the Dijksterhuislezing 2013,**Stichting Academisch Erfgoed**(SAE), VU University Amsterdam, the Netherlands, April 19, 2013 (presentation in **Dutch**).
Egoitz Laparra: Presentation Presentation on "Sources of Evidence for Implicit Argument Resolution", at the 10th International Conference on Computational Semantics (**IWCS2013**). Postdam, Germany, March 19-22, 2013.
Luciano Serafini: Presentation on "**CKR: a general framework for context in Semantic Web (Theory, prototype and extension to ASP)**", VU University Amsterdam, February 28, 2013.
Marieke van Erp: Presentation on "**NewsReader: Automating detective work**" at **PoliMedia Symposium**, VU University, Amsterdam, the Netherlands, January 23, 2013.
Luciano Serafini: **Presentation** on NewsReader at "**Facts, Truths, Argumentations**", Fondazione Bruno Kessler, Trento, Italy, January 18, 2013.

2012:

Piek Vossen: Invited speaker in **Casa Luna**, Dutch radio interview on new EU-project **NewsReader**, Mediapark, Hilversum, October 8, 2012
announcement video on Youtube by Casa Luna;
@ Radio 1, NCRV Casa Luna, part 1 (on NewsReader)
@ Radio 1, NCRV Casa Luna, part 2 (live interaction with listeners)
Blog from Marco Leeuwerink on the interview at Casa Luna on **Extend Limits**.
Piek Vossen: Invited speaker in Dutch radio interview on new EU-project **NewsReader**, in Radio Station FunX, Amsterdam, May 10, 2012.
press release on Newsreader by VU University Amsterdam, May 2, 2012.