3XL News: A Cross-lingual News Aggregator and Reader

Evgenia Belyaeva[1,2], Jan Berčič[1], Katja Berčič[1], Flavio Fuart[1(✉)], Aljaž Košmerlj[1], Andrej Muhič[1], Aljoša Rehar[3], Jan Rupnik[1], and Mitja Trampuš[1]

[1] Jožef Stefan Institute, Jamova Cesta 39, 1000 Ljubljana, Slovenia
{evgenia.belyaeva,jan.bercic,katja.bercic,flavio.fuart,aljaz.kosmerlj, andrej.muhic,jan.rupnik,mitja.trampus}@ijs.si
[2] JSI International Postgraduate School, Jamova Cesta 39, 1000 Ljubljana, Slovenia
[3] Slovenian Press Agency, Tivolska Cesta 50, 1000 Ljubljana, Slovenia
aljosa.rehar@ijs.si

Abstract. We present *3XL News*, a multi-lingual news aggregation application for iPad that provides real-time, comprehensive, global and multilingual news coverage. Using methods, developed within the XLike project, for semantic data extraction from news articles and linking of news stories we are able to construct a concise, yet in-depth view of current news stories and their semantic relation. This enables users real-time monitoring of current global events and analysis of diverse reporting in different languages and navigation across related news stories.

1 Introduction and Motivation

Real-time access to the latest news any time and from any location has become possible with widespread adaptation of mobile networks and devices. Increasingly, users of such devices expect custom-made, native applications to access their information. In this article we describe **3XL News**: a system stemming from the **X**Like EU project, offering **X(cross)**Lingual analysis of e**X**tra Large News. *3XL News* is an iOS application targeting news professionals and the general public. It shows how semantic technologies can be used in a real-world scenario to provide real-time global news monitoring and analysis across several languages. The main novelty is the linking of stories across six languages using semantic data derived from multi-lingual entity detection and cross-language news linking.

Related Work: A multitude of news monitoring mobile applications is available on the market, broadly divided into two groups: publishers' own applications (like BBC, Al-Jazeera and RTV Slovenija) and news aggregators (like News Republic, Yahoo News, EMM mobile app [7] and iDiversiNews [8]). Those services reduce the overwhelming amount of information by summarizing news events and filtering them according to predefined user preferences or reading habits, however, they do not link news stories across languages. There are systems that perform

F. Gandon et al. (Eds.): ESWC 2015, LNCS 9341, pp. 3–8, 2015.
DOI: 10.1007/978-3-319-25639-9_1

cross-language linking such as NewsReader [10] and SPIGA [4] but support only four and two languages respectively and lack the capability to explore news along other dimensions (sentiment, location etc.).

2 Technological Background

XLike Project. XLike stands for Cross-LIngual Knowledge Extraction. The main goal of the project[1] was to develop technology to monitor and aggregate knowledge spread across mainstream and social media, as well as across different languages. This is achieved by applying computational linguistics and semantic technologies to extract formal knowledge from multilingual texts.

Developed methods were integrated into an extensive linguistic and semantic text analysis pipeline [2], which provides input data for *3XL News*. News is gathered from the Internet [9], annotated with semantic data [2], similar articles are linked across languages [5], news articles are clustered into stories and finally, news stories are linked across languages using content (i.e. text) and semantic data. We have performed manual evaluation of the obtained clusters [1] but we omit it here due to lack of space.

Cross-Lingual Similarity Function. To compute similarities between documents written in different languages, we model them in a latent, language-independent vector space [6]; projections into the latent space are inferred using techniques from linear algebra and statistics. The method [5] is related to Canonical Correlation Analysis (CCA) [3], which we apply on a multilingual corpus of documents obtained from Wikipedia[2].

Semantic Data Extraction. Semantic annotation consists of *named entity recognition* [2] and *Wikipedia Miner Wikifier* [11]. The former detects named entities (persons, organisations, locations), while the later relates them to Wikipedia entries. For each entity a list of identifiers across all supported languages is provided. Thus, semantic annotation is used not only to better present data to users, but also to identify the same entity across different languages.

Linking News Stories Across Languages. Automatic linking of news stories (clusters of closely related news articles describing about the same event) across languages is an important addition to existing aggregation approaches. The main goal is to interconnect many influential and often related news that are reported constantly by numerous news outlets in different languages. We use CCA and semantic data (entities) to link stories as described in [1].

[1] http://www.xlike.org.

[2] http://www.wikipedia.org.

3 *3XL News* Application

The application consists of the following views:

Languages Overview (Fig. 1, left). The graph represents stories grouped by language (colour-coded) with images of the top-mentioned entities for each language. Node size corresponds to the number of articles in each language, while the width of each edge is computed from the overall similarity of stories in the two corresponding languages. News sources are shown on the world map, in which the display of any language can be switched *on* or *off*.

Entity Overview (Fig. 1, right). Top-mentioned entities are aggregated across all languages, with possible filtering for major categories. Furthermore, users can choose a single entity to read related stories. Selected entities are shown on the world map below where news sources can be compared.

News Stories List. Contains a list of news stories filtered by language or entity and ordered by relevance with basic information for each story: title, summary, photo, publication date, number of news articles and related stories per language. The user can select a story and explore it.

News Story Exploration (Fig. 2). Components representing different aspects of a story are shown. The graph represents related stories across languages, with

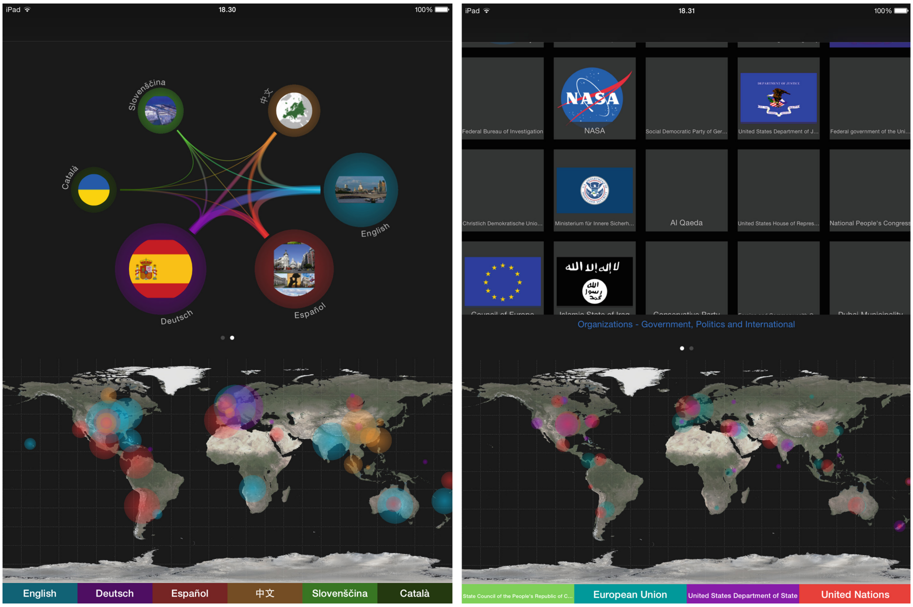
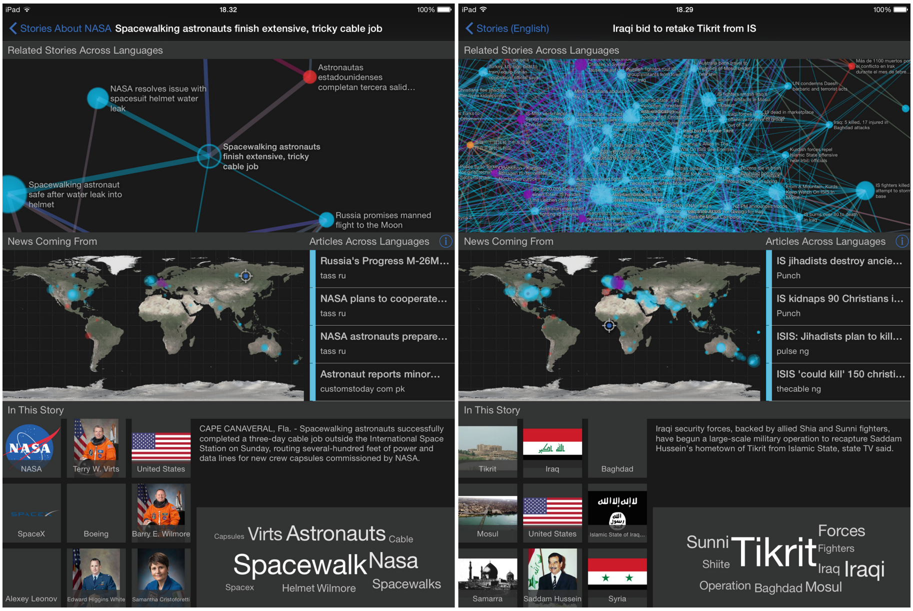

iPad
18.30
100%
Slovenščina
中文
Català
English
Deutsch
Español
English
Deutsch
Español
中文
Slovenščina
Català
iPad
18.31
100%
Federal Bureau of Investigation
NASA
Social Democratic Party of Ger...
United States Department of J...
Federal government of the Uni...
Christlich Demokratische Unio...
Ministerium für Innere Sicherh...
Al Qaeda
United States House of Repres...
National People's Congress
Organizations - Government, Politics and International
State Council of the People's Republic of C...
European Union
United States Department of State
United Nations

Fig. 1. Overview screens: by language and by entities (Color figure online)


iPad
18.32
100%
Stories About NASA
Related Stories Across Languages
Astronautas estadounidenses completan tercera salid…
NASA resolves issue with spacesuit helmet water leak
Spacewalking astronauts finish extensive, tricky cable job
Spacewalking astronaut safe after water leak into helmet
Russia promises manned flight to the Moon
News Coming From
Articles Across Languages
In This Story
NASA
Terry W. Virts
United States
SpaceX
Boeing
Barry E. Wilmore
Alexey Leonov
Edward Higgins White
Samantha Cristoforetti
Capsules
Virts
Astronauts
Cable
Spacewalk
Nasa
Spacex
Helmet
Wilmore
Spacewalks
iPad
18.29
100%
Stories (English)
Related Stories Across Languages
Turkey, US sign deal to train, equip Syrian opposition, official says
Kurdish fighters rout IS group militants from town near Iraq
Australia bans travel to Iraqi city of Mosul under new law
Más de 1100 muertos por el conflicto en Irak durante el mes de febre…
UN condemns Daesh barbaric and terrorist acts
Christians flee jihadists after Syria kidnappings
More Christians abducted by IS
IS fighters smash Iraq's ancient artifacts in Mosul museum
IS-Miliz verschleppt in Syrien mindestens 90 Christen
Iraqi forces launch offensive to drive IS group out of Tikrit
19 dead in marketplace bombings
Iraq: 5 killed, 17 injured in Baghdad attacks
Iraqi bid to retake Tikrit from IS
America's Allies In The War On ISIS See Enemies Everywhere
Kurdish forces repel Islamic State offensive near Irbil: officials
Police fly to Turkey for girls' recruit…
Dozens die in fresh Baghdad bombings
From A Mountain, Kurds Keep Watch On ISIS In Mosul
IS fighters killed attempt to storm base
Australian man killed fighting IS in Syria
NZ PM announces troops
IS burns over 40 to death in Iraq
Syria 3
News Coming From
Articles Across Languages
In This Story
Tikrit
Iraq
Baghdad
Mosul
United States
Islamic State of Iraq…
Samarra
Saddam Hussein
Syria
Sunni
Tikrit
Forces
Fighters
Shiite
Iraq
Iraqi
Operation
Baghdad
Mosul

Spacewalking astronauts finish extensive, tricky cable job

Russia's Progress M-26M…
tass ru

NASA plans to cooperate…
tass ru

NASA astronauts prepare…
tass ru

Astronaut reports minor…
customstoday com pk

CAPE CANAVERAL, Fla. - Spacewalking astronauts successfully completed a three-day cable job outside the International Space Station on Sunday, routing several-hundred feet of power and data lines for new crew capsules commissioned by NASA.

Iraqi bid to retake Tikrit from IS

IS jihadists destroy ancie…
Punch

IS kidnaps 90 Christians i…
Punch

ISIS: Jihadists plan to kill…
pulse ng

ISIS 'could kill' 150 christi…
thecable ng

Iraqi security forces, backed by allied Shia and Sunni fighters, have begun a large-scale military operation to recapture Saddam Hussein's hometown of Tikrit from Islamic State, state TV said.

Fig. 2. NASA story exploration (left) and Middle East story exploration (right)

node sizes corresponding to the number of articles in a story and edge widths corresponding to the similarity of given two stories. The geographic selector sorts articles according to sources, while the last section gives a quick overview of the story by giving its top-mentioned entities, summary and keywords.

Articles List. Contains a list of articles for a selected story, ordered by user selection, is shown on a separate screen. Each article can be selected and the original web page, from which the text was extracted from, is displayed.

4 Demonstration

We will demonstrate a typical user session by selecting and analysing few news stories as shown in the figures.

Katja checks the overview screen (Fig. 1, left) and top-mentioned entities (Fig. 1, right). She then compares several entities by reporting locations. Being interested in space-related news, she selects “NASA” to view a list of related news stories (screen-shot not provided).

Katja now explores the main story she selected, together with its semantically related stories. A thick connection shows strongly related stories, while thin connections represent weak relations. For example, stories related to NASA astronauts space-walk are strongly related, while stories about Russian plans to get to the moon are weakly connected (Fig. 2, left).

Switching to earthly events (Fig. 2, right), Katja finds an abundance of interconnected stories from all languages. Compared to space-related news, current events are more widely covered by the world media. Finally, she can list all the articles making up the current story and read them in the integrated browser (screen-shot not provided).

5 Conclusion

In this article we presented *3XL News*, an iPad news aggregation app that delivers a global view on the news media landscape by applying advanced computational linguistics and semantic approaches. The main advantage compared to similar systems is semantic linking of stories across six languages.

3XL News currently supports six languages, but the system is being extended to twelve languages with more planned. Through xLiMe[3], a research project dedicated to fusing the knowledge from different media content in different modalities, we plan to include annotated video and audio materials.

The demonstration video is available from the *3XL News* homepage[4]. The application will also be submitted to the *Apple App Store* and available free of charge.

Acknowledgements. This work was funded by the European Union through project XLike (FP7-ICT-2011-288342).

References

1. Belyaeva, E., Košmerlj, A., Muhič, A., Rupnik, J., Fuart, F.: Using semantic data to improve cross-lingual linking of article clusters. J. Web Seman. (submitted)
2. Carreras, X., Padró, L., Zhang, L., Rettinger, A., Li, Z., García-Cuesta, E., Agić, V., Bekavec, B., Fortuna, B., Štajner, T.: Xlike project language analysis services. In: Proceedings of EACL 2014, pp. 9–12. Gothenburg, Sweden, April 2014
3. Hardoon, D.R., Szedmak, S., Szedmak, O., Shawe-Taylor, J.: Canonical correlation analysis; an overview with application to learning methods. Technical report (2007)
4. Hennig, L., Ploch, D., Prawdzik, D., Armbruster, B., De Luca, E.W.: Spiga - a multilingual news aggregator. In: Proceedings of GSCL 2011 (2011)
5. Rupnik, J., Muhič, A., Škraba, P.: Cross-lingual document retrieval through hub languages. xLiTe: Cross-Lingual Technologies, NIPS 2012 Workshop (2012)
6. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. In: Information Processing and Management, pp. 513–523 (1988)
7. Steinberger, R.: Multilingual and cross-lingual news analysis in the europe media monitor (EMM) (Extended Abstract). In: Lupu, M., Kanoulas, E., Loizides, F. (eds.) IRFC 2013. LNCS, vol. 8201, pp. 1–4. Springer, Heidelberg (2013)
8. Trampuš, M., Fuart, F., Pighin, D., Tadej, Š., Berčič, J., Novak, B., Rusu, D., Stopar, L., Grobelnik, M.: Diversinews: surfacing diversity in online news. AI Magazine (2015, accepted for publishing)

[3] http://xlime.eu/.

[4] http://ailab.ijs.si/tools/3xl-news/.

9. Trampuš, M., Novak, B.: The internals of an aggregated web news feed. In: Proceedings of IS-2012. Ljubljana, Slovenia (2012)
10. Vossen, P., Rigau, G., Serafini, L., Stouten, P., Irving, F., Hage, W.V.: Newsreader: recording history from daily news streams. In: Proceedings of LREC 2014 (2014)
11. Zhang, L., Rettinger, A.: Semantic annotation, analysis and comparison: a multilingual and cross-lingual text analytics toolkit. In: Proceedings of EACL 2014, pp. 13–16 (2014)