User Controlled News Recommendations

Jon Espen Ingvaldsen
Norwegian University of Science and Technology, Department of Computer and Information Science, Trondheim
Norway
jonespi@idi.ntnu.no

Jon Atle Gulla
Norwegian University of Science and Technology, Department of Computer and Information Science, Trondheim
Norway
jag@idi.ntnu.no

Özlem Özgöbek
Department of Computer Engineering, Ege University,
Izmir,
Turkey
ozlem.ozgobek@ege.edu.tr

ABSTRACT
The adoption of mobile devices is pushing the Internet into a more personal and context aware space. A common challenge for online news services is to deliver contents that are interesting to read. In this paper, we describe the user interface design of the SmartMedia news recommender prototype. Through deep analysis of textual news contents it is able to deliver local, recent and personalized news experiences, and the user interface is designed to give the users control over the news stream compositions. We will present its innovative user interface and the approach taken to transform raw textual data into well defined and meaning bearing entities.

Categories and Subject Descriptors
H.4.7 [**Information Systems Applications**] Communications Applications – *Information browsers*

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
Recommender system, news, mobile, user interfaces, user control

.

1. INTRODUCTION
The Smartmedia project[1] at NTNU targets construction of context aware news experiences based on deep understanding of text in continuous news streams [4, 9]. The goal of the Smartmedia project is to deliver a mobile and context aware news experience based on deep understanding of textual contents, combining both geo spatial exploration and context aware recommendations. The system is designed with scalability in mind and ability to support multiple languages.

Privacy is an important aspect when engineering recommender systems and exploitation of user interaction and context data. When dealing with personal data and privacy, transparency tools are tools that can provide to the concerned individual clear visibility of aspects relevant to these data and the individual's privacy. The combination of transparency tools and user control yields viable functionality to empower users to protect their privacy [5].

In the Smartmedia project, we want to build transparent news recommender systems where the user can control gathered data and how their news streams are composed based on geo spatial location, personal interest profile and time. When designing user-friendly systems for mobile devices, we need to be careful about the amount of buttons and menu items introduced. In this paper we will describe the news recommender system prototype and its mobile user interface where the users can control their news stream recommendations from three toggleable buttons.

2. IMPLEMENTATION
The backend of the news recommender prototype developed is constructed as a pipeline of operations harvesting and transforming Rich Site Summary (RSS) entries and raw text data into a semantic and searchable representation. The pipeline and its operations are implemented with using Apache Storm[2]. This distributed computing framework enable scalability and ability to handle large amounts of news items from a magnitude of publishers continuously.

As shown in Figure 1, the news processing pipeline consists of five steps. The first step creates an input stream by continuously monitoring a large set of RSS feeds. Whenever a new news item occurs, properties such as the title, lead text and HTML sources are extracted. The HTML sources are parsed and cleaned to extract a representative body text. In the second step, natural language processing operations such as language identification, sentence detection and part-of-speech tagging is applied to extract entity mentions from the textual data. The third step uses supervised models to map entity mentions to referent entities in the WikiData[3] and Geonames[4] knowledge bases. These models combine textual similarities, graph relations and entity frequency and co-occurrence statistics to classify the relevance of multiple referent candidates. First Story Detection (FSD) is applied in the forth step to group news items describing the same news story. In the fifth step this semantic representation is indexed and made searchable. As this backend architecture is stream based, it is able to index and promote recent news items.

WikiData is the community-created knowledge base of Wikipedia [12]. Since its public launch in 2012, the knowledge base has gathered more than 15 millions entities, including more than 34 million statements and over 80 million labels and descriptions in more than 350 languages [3]. Most geographical entities in WikiData provide a reference to Geonames containing more detailed geographical properties. In the implementation of the Smartmedia prototype, the news and entity information including news text, titles, publication timestamps, entity labels and

[1] http://research.idi.ntnu.no/SmartMedia

[2] http://storm.apache.org/

[3] https://www.wikidata.org/

[4] https://www.geonames.org/

geospatial properties are indexed in a Lucene based search index. This index makes the news items and their related entities searchable and creates a foundation for detailed querying.

When a user is opening the news app on the mobile a request containing user id, location and preferences are sent to the backend. Here, a multi factor search query is formed to retrieve relevant news entries from the index.

3. USER INTERFACE

A web-based user interface is developed to make the news stream contents explorable on mobile devices. In this interface, the user is allowed to extract news items that are relevant to the geo special locality context, personal interests and given point of time. These three relevance factors are customizable and the user can select whether or not they should influence the retrieval and ranking of available news items.

To customize the geographical locality, the user specifies a circular relevance region on a map. Figure 2a shows an example of such a relevance region. By default, the relevance region is set to users current GPS location with a 50 km radius. By moving the region or modifying the radius, users can generate a local newspaper for any region of the world. If the location factor is disabled, it means that the system is recommending news from any location in the world and news that are not containing location information.

In the current Smartmedia prototype, we have predefined a handful of user interest profiles. Examples of such profiles are stock trader, soccer fan, technology geek, etc. Each profile consists of a weighted concept vector, where each entry is a WikiData entry associated with an interest score between 0 and 1. By selecting any of these interest profiles, the retrieved news will be influenced and biased towards the interest topics. When the personal interest factor is disabled, the user retrieve a news composition which is general and without such bias.

To customize the time-factor, the user is presented with a calendar where it is possible to move in time and retrieve either recent or historic news items. When, the time-factor is disabled the user will retrieve news solely based on the other relevance factors (location and personal interests).

Figure 2b shows an example of how news stories are presented. Here we see one news article *"Theresa May urges media restraint in coverage of terror suspects"* from the Guardian about politics and terror, followed by another news story from BBC. The three circular buttons on the bottom of the screen allow users to toggle whether their locality, personal interest profile and time setting such influence news story retrieval.

By clicking on a news story, the user gets the ingress of the news story and a list of the most salient entities for the selected news story. Figure 1c shows the ingress and relevant WikiData entities from the news article about Theresa May. As we can see, our news story about politics and terror related to Syria, Theresa May, ISIL and Sky News. By hovering these items, the user is presented with their textual WikiData description. On figure 2c, we can see that the WikiData entity for Theresa May contains the description *"British politician"*.

In general, the three buttons at the bottom of the screen for location, interest profile and time can at any time be activated and de-activated to provide very different recommendation strategies. For example, keeping all buttons active with default parameters means that the system will recommend news articles that have recently takes place in the vicinity of the reader and are consistent with her profile. Figure 3 describes different combinations of recommendation factors and summarizes how the user can control the retrieval and composition of news items.

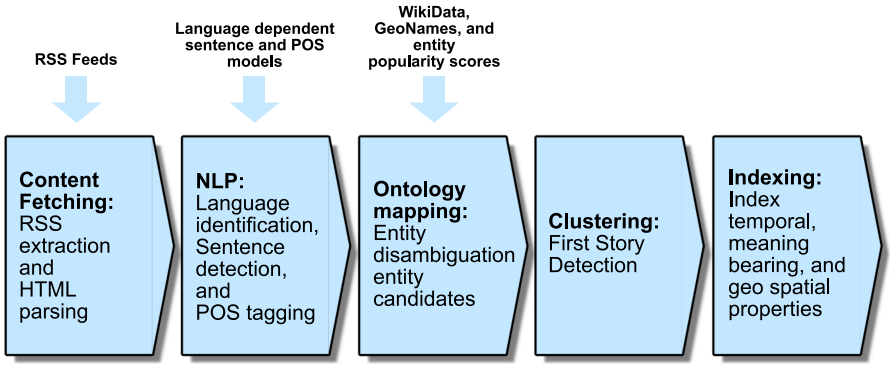
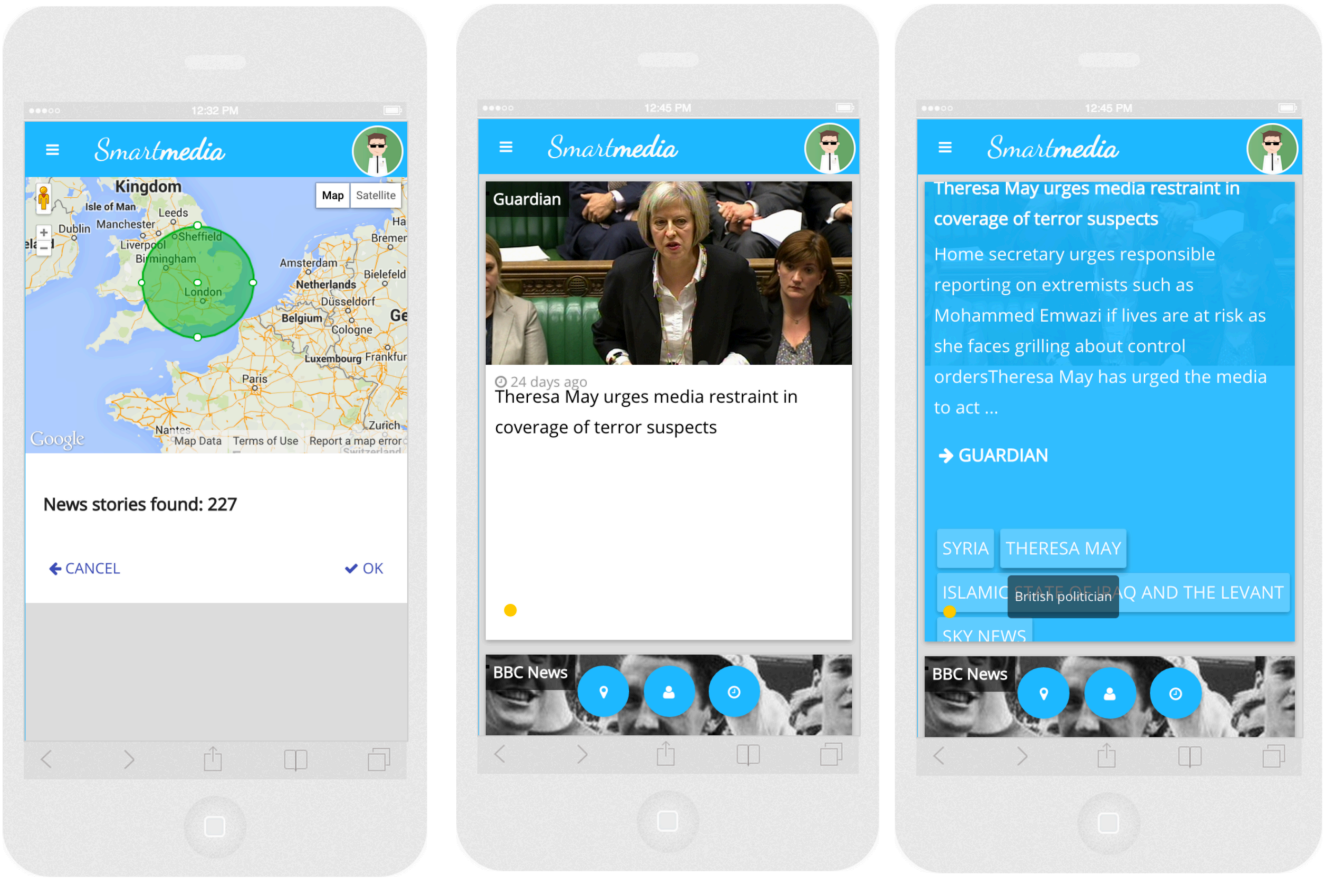


Figure 1. Steps of the news stream processing pipeline.



a) b) c)

Figure 2. Screenshots from the Smartmedia prototype. a) The map query interface. b) Presentation of news stories. c) Presentation of news details.

4. RELATED WORK

People nowadays have access to more worldwide news information than ever before. As Internet services get more information about their users and their context, they can deliver personal and customized contents and user experiences.

The prototype system, described in this paper, share similarities to other academic news applications such as NewsStand [8, 10] and News@Hand [1, 2]. Both these systems map textual news contents to entities defined in a knowledge base.

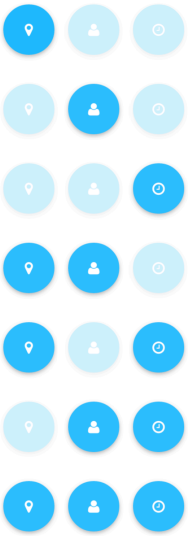
NewsStand targets geo spatial exploration of news. It is an example application of a general framework developed to enable people to search for information using a map query interface. It utilize maps both to explore and find news stories and to visualize and present single news events.

News@hand combines textual features and collaborative information to make news suggestions. It uses Semantic Web technologies to describe the news contents and user preferences. Both news items and user profiles are represented in terms of concepts appearing in domain ontologies, and semantic relations among those concepts are exploited to enrich the above representations, and enhance recommendations.

Both these NewsStand and News@Hand have user interfaces targeting desktops and larger device screens. They both provide user control over the retrieved set of news, either through a map or category based navigation or preferences settings.

Tran and Herder [11] have looked at the studied news event timelines and shown that manually constructed timelines are subjective and often missing important dates or other information. By complementing the timelines with elements extracted algorithmically from multiple sources, it is possible to create more objective and argumentative timelines. However, the manual processing and editing efforts are still needed to enhance the communicative qualities of the timelines, and to adapt it to the needs of the readers

Parra et al. [6, 7] presents SetFusion, a visual user-controllable interface for hybrid recommender system. Their approach enables users explore and control the importance of recommender strategies using an interactive Venn diagram visualization. Their evaluations indicate that this interface had a positive effect on the user experience and improved users engagement. Their idea of using the Venn diagram to explain intersections among recommendation approaches is transferable and valuable to the news domain.



Local
Retrieve news from a geospatial area

Personal
Retrieve news matching the interest profile of the user
Temporal
Retrieve news published after a given date

Local and personal
Retrieve news that both relate to a geospatial area and match users interest profile.
Local and temporal
Retrieve news related to a given geospatial area and published after a given date.
Personal and temporal
Retrieve news matching the interest profiles of the user and published after a given date.
Local, personal and temporal
Retrieve news with relevance for the selected geospatial area and interest profile, and published after a given date.

Figure 3. Combinations of selectable recommendation factors

5. CONCLUSIONS AND FUTURE WORK

The predefined user profiles can be replaced or used in combination with more personal profiles trained on traced interaction logs from the system. As users leave interaction data behind, we can gather knowledge about what the users interests are. However, for new users where no past interaction records exist, we have a cold-start problem where we still benefit on predefined stereotypes.

In future work we plan to use trained personal profiles with predefined stereotypes in combination. We will also gather user feedback and evaluate to which extent users want to control and customize their news presentations and study how their requirements can be met in a mobile user interface design.

Deep understanding of textual contents together with knowledge base structures provides a fundament for innovative and intelligent applications. This paper has described one such innovation from the news domain, and how its mobile user interface allow users to control the composition of news. A screencast video demonstrating the prototype and its user interface is available at: http://vimeo.com/121835936

6. REFERENCES

[1] Cantador, I. et al. 2008. News@ hand: A semantic web approach to recommending news. *Adaptive hypermedia and adaptive web-based systems*. (2008).

[2] Cantador, I. et al. 2008. Ontology-based personalised and context-aware recommendations of news items. *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 1, (2008).

[3] Erxleben, F. et al. 2014. Introducing Wikidata to the Linked Data Web. *The Semantic Web–ISWC 2014*. (2014).

[4] Gulla, J.A. et al. 2013. Learning User Profiles in Mobile News Recommendation. *Journal of Print and Media Technology Research*. II, 3 (2013), 183–194.

[5] Hansen, M. 2008. Marrying transparency tools with user-controlled identity management. *The Future of Identity in the Information Society*. (2008).

[6] Parra, D. et al. 2014. See what you want to see. *Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14* (New York, New York, USA, Feb. 2014), 235–240.

[7] Parra, D. and Brusilovsky, P. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies*. (2015).

[8] Samet, H. et al. 2014. Reading news with maps by exploiting spatial synonyms. *Communications of the ACM*. 57, 10 (Sep. 2014), 64–77.

[9] Tavakolifard, M. et al. 2013. Tailored news in the palm of your hand: a multi-perspective transparent approach to news recommendation. (May 2013), 305–308.

[10] Teitler, B. and Lieberman, M. 2008. NewsStand: A new view on news. *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. (2008).

[11] Tran, G. and Herder, E. 2015. Detecting Filter Bubbles in Ongoing News Stories. *Extended Proc. UMAP 2015*. (2015).

[12] Vrandečić, D. and Krötzsch, M. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*. (2014).