Using Linked Data to Reduce Learning Latency for e-Book Readers

Julien Robinson, Johann Stan, and Myriam Ribière

Alcatel-Lucent Bell Labs France, 91620 Nozay, France
`Julien.Robinson@alcatel-lucent.com`

Abstract. Web 2.0 technologies provide an opportunity to transform learning into a social experience: social learning can directly benefit from user-generated content analysis. The e-book device is an object central to the learning process; it becomes an ideal medium to promote innovative learning tools. In this paper, we describe how we analyse user annotations in e-books using Linked Data to reduce the latency between professor knowledge, book content and student assimilation.

Keywords: Social Learning, Semantic Web, Linked Data, e-Book.

1 Introduction

People learning with a given e-book form a de-facto community, which is even more relevant if they are reading at the same time, with the same purpose or during the same activity, e.g. a class. Currently, very few platforms allow social interactions between such people that are dedicated to their reading activity; readers can connect but at a very high level and generally outside the platform (e.g. forums in LibraryThing[1], chatting in Bookglutton[2] or more general social cataloging features GoodReads[3] and Shelfari[4]).

The main problem with these platforms is that interactions cannot be directly linked to the context of reading and thus do not allow fine-grained interactions about a specific part of the book. Bookglutton goes a step further by letting users chat while reading, but does not exploit the contents and context of annotations in the e-book. This issue was already explored [1] [2], but no solution was proposed based on the semantic analysis of the annotations.

An important observation for our work is the fact that in most cases, annotations in e-books are in the form of short, unstructured textual messages, associated to images, paragraphs or chapters. These can include transcriptions, reformulations, additional comments, questions or explanations (such as student annotations in BookGlutton[5]). Clearly, services that leverage these annotations

---

[1] `http://www.librarything.com/` - visited September 2010
[2] `http://www.bookglutton.com/` - visited January 2011
[3] `http://www.goodreads.com/` - visited October 2010
[4] http://www.shelfari.com/ - visited March 2011
[5] e.g. `http://www.bookglutton.com/glutton/melissthangx0x0/14459`

can increase the social dimension of reading. They can help readers determine a common environment to socially learn by seeing others' annotations or to generate interactions between students during a same activity.

In this paper we present our ongoing work on services that leverage semantic technologies and especially Linked Data for providing a social dimension for reading and implicitly, learning. First we summarize how we associate annotations, in natural language, to Linked Data concepts; then we describe the heatmap, the social navigation feature that leverages this work for efficient reading; finally we present the Tagging Beak, a research prototype that should pave the way for the next version of the heatmap.

2   Management of Shared Annotations Using Linked Data

Linked Data is an evolving component of the Semantic Web. It can be defined as a community effort to extract or publish data from various sources, interlink it and store it in a semantic web format, such as RDF. The content of these online linked datasets is diverse. It consists mostly of data about people, companies, books, geographic locations, films, music, television and radio programs, information about medicine (genes, proteins, clinical trials) and online communities (statistical data). The most used online dataset is certainly DBPedia[3].

An important source of social knowledge is the set of annotations produced by the user. In our framework, our primary hypothesis is to consider these annotations on a conceptual level. More concretely, this means that we associate each annotation to one or several concepts in a Linked Data source (e.g. DBPedia). There are several reasons to make this association: (i) provide a rich approximation of their meaning, (ii) the capability to retrieve related concepts and category terms, operation that we call semantic expansion[6] and (iii) the fact that annotations are generally short and that contextual cues are not always available. Therefore, Linked Data sources can enrich the description of such annotations with additional metadata and related concepts.

The association of annotations to concepts in Linked Data is performed by a disambiguation process that leverages the context of the annotation to match it with the most relevant concept in the knowledge base. Each annotation provided by the user is associated to a context, which is constructed from different sources, such as previous annotations shared by the user in the e-book and the set of annotations shared by others that are related to the given paragraph. Parallel to this, a set of candidate concepts are retrieved from DBPedia, that contain the given annotation in their label. The candidate concepts are ranked using cosine similarity between the vector constructed from their abstracts and the contextual cue vector. Once this disambiguation has been performed, the second use of Linked Data is the expansion of the concept by exploring both its semantic neighborhood and hierarchical links.

---

[6] As an example, here is the semantic expansion of the annotation "Facebook": [Global Internet Community, Student Culture, Social Information Processing, Social Networking Service, Social Media, Web 2.0, Internet Technologies].

3   Easy Social Navigation with Heatmaps

Key to the exploitation of Linked Data results in social learning is the way the user will perceive the analysis results and whether it will be useful or not. In [1], Jae K. Kim et al. describe how they let students place symbolic evaluations as annotations, and then graphically represent the synthesis of these annotations as colors in the e-book, thus enabling what is called **Social Navigation**. We follow the same investigation track, but instead of focusing on design choices and symbolic evaluations, we analyze free text annotations using Semantic Web techniques.

The user, a learning reader (our target is post-graduate students), opens an e-book and states their interest using freely chosen keywords; the keywords express their goal for the reading session. The keywords are semantically analyzed and a global similarity computation is run on the annotations of the book. Each annotation is attributed a heat value based on its relevance to the user's goal in reading the book. The appearance of the book is modified to display in a colored way the zones where the annotations are the most heated.

Figure 1(a) presents an example of the heatmap feature. This was produced by our prototype of the sBook [4] using data gathered from annotations from BookGlutton (229 annotations by 39 different users on the e-book "Franken-stein" by Mary Shelley).

The choice was made to use the paragraph as the finest precision of annotation heat. The heat of a paragraph is the total sum of the heat of each annotation on the paragraph. Note that the heatmap is a navigation indication for the user, it is not meant to be the final step for a learning reader; it is completed by the display of authors of annotations (on the right-hand panel) and the use of what we call annotation boards: these are separate windows that browse through all the annotations of a paragraph and the comments inspired by these annotations. The annotation board is where the user may find valuable information and participate in discussions.

The choice of the algorithm for heating an annotation is the focus of our current research. Figure 1(b) presents the administration interface that enables to choose different implementations for the user to experiment with. Annotations are heated based on their proximity to the resulting vector of words. We currently implement three expansion methods:

– no expansion, we simply search for the keyword
– a dictionary-based heatmap, that may take into account synonyms, antonyms and derived words (e.g. adjectives for a noun)
– a Linked-Data-based heatmap, that takes into account the expansion of the search keyword as described in the present paper.

Our first experiments show that different expansion techniques may be relevant for different types of words. For instance, the word "Paris" is expected to be linked to "France", which is the case using DBpedia but not using a dictionary; this hints that Linked Data would be appropriate for named entities. On the other hand, the word "character" expands to "personage" and "reputation" using
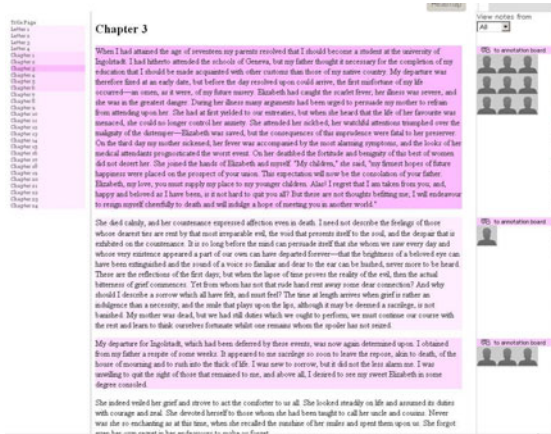
(a) sBook with Heatmap activated (overview)
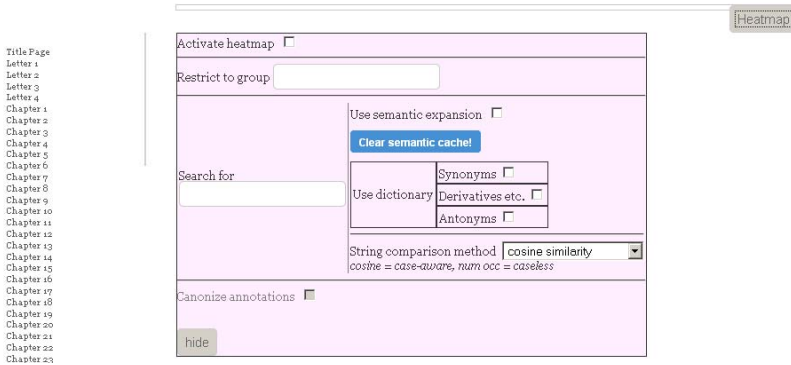
(b) Administration Heatmap Configuration

Fig. 1. Heatmap screenshots

a dictionary, while no such expansion is provided by DBpedia; this would indicate that dictionary expansion is more appropriate for literary theory terms. This is why we keep both options open and also separate. We intend to validate the utility of each expansion based on user experimentation by the end of 2011.

In a first step the annotations are processed as text; currently being implemented, we first process each annotation by associating it to a cluster of tags, obtained by semantic evaluation techniques. The similarity measure will be computed between the cluster of tags of the annotation and the vector of words expanded from the user keyword.

It may be noted that, while we explore keyword-based heatmaps, we also intend to explore other options: time-based heatmap, where the heat of an annotation is related to its creation date; location-based heatmap, taking into account the location of the annotation's author; or more generally context-based and activity-based heatmaps (which present intersections). These will be used for

further user experimentations, first to compare their effectiveness with keyword-based heatmaps, second because they are relevant to our general research on the link between the digital and physical worlds.

4   User Interaction Profiles

While the heatmap is a social navigation tool based on a request by the current reader, we intend to go further in our recommendation strategy. We aim to recommend not only annotations or parts of a book, but also other members of the reading community. This social recommendation process requires specific user profiles.

In this section, we focus on our global approach to building these user profiles based on interactions such as shared annotations or messages. Such a profile is called *"user interaction profile"*, it is based on individual content sharing activity or interactions with peers. Its aim is to identify concepts that either represent an expertise field of the user or a topic that motivates them to interact. A more detailed definition of interaction profiles and the corresponding algorithms can be found in [5].

In a first step, the algorithms to build these interactions profiles have been developed in the scope of a Twitter analysis tool. Twitter proposes a wealth of public social interactions, in a format close to annotations: short but meaning-ful messages, sometimes general comments and sometimes addressed to specific contacts. We use this as a means to tune and validate the algorithm before inte-grating it in the social book. We propose an online interactive tool, the Tagging Beak[7], to display these profiles for Twitter users (Figure 2(a)) and also to recom-mend people relevant to an information need expressed as a question in natural language (Figure 2(b)).

Users connect with their Twitter account; the system reads their tweets and extracts keywords and named entities. Profiles are built by contextualized seman-tic matching (*disambiguation*) and expansion of such items: extracted keywords are used to find concepts in DBpedia; in order to disambiguate homonyms, the concepts are then evaluated for their proximity to the previous tweets of the user and of their community[8] using semantic techniques. The most relevant concepts are then expanded along chosen dimensions in DBpedia, which yields neighbour-ing concepts, who are in turn filtered based on their proximity to the user profile. The resulting set of ponderated concepts is the user's interaction profile.

Sentiment analysis is also used, as well as declarative feedback from users. This helps determine not only the expertise of a user on a subject but also their willingness to share. They can for example say that they have high expertise on a subject but that they wish to share this expertise only with classmates and teachers and that they are ready to interact on this topic only when they are at the library or at school. With the help of DBPedia, we also retrieve the

---

[7] The Tagging Beak Collaborative Social Search System - `http://tbeak.com`
[8] In Twitter, communities are built upon the following / follower link between users.

(a) Interaction Profile of a user in the Tagging Beak (overview)

(b) People Recommendation Interface in the Tagging Beak (overview)

Fig. 2. Tagging Beak screenshots

hierarchical tree associated to the concept, which allows the user to share the same concept with other social spheres, but with less granularity. In this way, the user can share different levels of granularity about the same knowledge with multiple social spheres (e.g. detailed knowledge for the class, less detailed for friends).

Based on expertise and willingness to share, a user may be recommended to another user for answering a question. This recommendation strategy is key to social learning. It ranks people in the close and more distant (e.g. friends of friends) community of the user according to the semantic similarity between the topics of the question and their interaction profiles. The Tagging Beak then enables either to put the question to the recommended contact or to start following them in Twitter (if this is not already the case).

Linked Data allows to retrieve the hierarchical tree and semantic neighborhood associated to a given concept. In the case of collaborative learning, we leverage this feature by offering users a more granular way of sharing their interaction profile, constructed from the shared annotations. Such interaction profiles make it possible for the system to select the right person to ask a question on a specific topic. In order to overcome the problem of cold start in case of new members in such learning communities, we are currently investigating the possibility to connect this service to other Social Platforms and to build a user interaction profile from annotations shared in them. In this way, the user will be recommended people to interact with even if they did not share sufficient annotations. More specifically, each time the user opens the e-book, they will be presented with the most relevant readers to interact with, based on complementary expertise in specific topics of interest and declared sharing preferences related to that topic. This provides students with a seamless way of linking to co-learners and thus form spontaneous learning communities on a specific part of a lecture.

5   Conclusion

In this paper we have presented our research, which focuses on integrating Linked Data results into e-book navigation tools for students in order to enhance social learning. We base our work on user annotations, we process these annotations with Linked Data technologies, we use the produced synthetized results to modify our e-book display; we hope to maximize reading and learning efficiency, to create common environments and implicitly, communication opportunities. At the basis of our innovation is the hypothesis that the huge amount of shared content in different Social Platforms offers exciting opportunities to enrich the learning experience and reduce the knowledge latency in the class, which we define as the delay between the encoding of the knowledge of a professor on a specific subject into information (e.g. a lecture or slides) and the assimilation of this information as knowledge by students.

References

1. Kim, J.K., Farzan, R., Brusilovsky, P.: Social navigation and annotation for electronic books. In: BooksOnline 2008: Proceeding of the 2008 ACM Workshop on Research Advances in Large Digital Book Repositories, pp. 25–28. ACM, New York (2008)
2. Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., Coyle, M.: Collecting community wisdom: integrating social search & social navigation. In: IUI 2007: Proceedings of the 12th International Conference on Intelligent User Interfaces, pp. 52–61. ACM Press, New York (2007)
3. Lehmann, J., Bizer, C., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - a crystallization point for the web of data. Journal of Web Semantics 7(3), 154–165 (2009)
4. Ribière, M., Picault, J., Squedin, S.: The sbook: towards social and personalized learning experiences. In: Proceedings of the Third Workshop on Research Advances in Large Digital Book Repositories and Complementary Media, BooksOnline 2010, pp. 3–8. ACM, New York (2010)
5. Johann, S., Viet-Hung, D., Pierre, M.: Semantic user interaction profiles for better people recommendation. In: Proceedings of ASONAM 2011, Kaohsiung, Taiwan (to appear, 2011)