BooksOnline'12: 5th Workshop on Online Books, Complementary Social Media and their Impact

Gabriella Kazai[1]   Monica Landoni[2]   Carsten Eickhoff[3]   Peter Brusilovsky[4]

[1] Microsoft Research, Cambridge UK, v-gabkaz@microsoft.com
[2] USI, University della Svizzera Italiana, Lugano, Switzerland, monica.landoni@usi.chl
[3] Delft University of Technology, Delft, The Netherlands, c.eickhoff@tudelft.nl
[4] University of Pittsburgh, Pittsburgh, US, peterb@pitt.edu

ABSTRACT

BooksOnline'12, the fifth workshop in the series, aims to offer a forum for bringing together expertise from academia, industry and libraries to facilitate the exchange of research results and technology in the field of digital libraries with specific focus on online books and complementary media. The focus of this year's workshop is "engaging reading experiences", starting from the act of deciding what to read, through the exploration and interpretation of a book's content, to sharing the overall experience. Within this overall umbrella theme, the accepted papers naturally showed three salient themes: (1) Search and Discovery, (2) Personalization and Recommendation, and Reading Experiences beyond Text. The contributions demonstrate a range of technologies, including a collaborative tabletop visual approach to support the searching and discovery of books, co-citation methods to enhance document retrieval; exploring open issues in audio-book production to support non-text based reading and improving e-book accessibility; new approaches to recommendation that take into account writing style as well as looking specifically to young readers and their needs in order to develop recommendation tools that consider both content and reading level and match these against the readers' specific interests and reading ability. Following in the theme of the reader playing a central role in the future of our digital era, we are honored to welcome Maribeth Back from FX Palo Alto and Natasa Milic-Frayling from Microsoft Research as our keynote speakers.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries

General Terms

Algorithms, Design, Human Factors

1. INTRODUCTION

In the past decade, millions of books have been digitized in collaborations between libraries and companies like Amazon and Google, e.g., through the Million Book project, and made available online, e.g., through the Internet Archive. At the same time, e-books and eReaders have gained wide acceptance and popularity, their sales skyrocketing year after year. For example, in January 2012, American publishers sold 22.6 million children's and young adults e-books (up from 3.9 million in 2011 January)[1]. In parallel to these developments, during the past years, the Web culture has grown more and more enticing, centring many services around social media and collaboratively shared content. The vast range of possible exploitations of such community platforms includes viral marketing, collaborative tagging, recommendation and content creation. The harnessing of these considerable forces has the potential to revolutionize online reading both technically and in terms of interaction paradigms.

To match the great momentum in creating online book repositories and related social media platforms, the BooksOnline workshop series aims to foster research initiatives that are focused on innovation opportunities and challenges created by large collections of digital books and complementary media. This year, the workshop focuses on exploring the role that social media may hold in enhancing digital libraries and online repository services towards more engaging reading experiences. The accepted contributions naturally showed three salient themes: (1) Search and Discovery, a theme that encompasses work addressing serendipitous content discovery, customizable information filtering as well as faceted topic exploration; (2) Personalization and Recommendation, which introduces reader-centric means of content selection based on comprehensibility and entertainment; and (3) Reading Experiences beyond Text that proposes novel content presentation paradigms addressing accessibility concerns.

Following in the theme of the reader playing a central role and looking more to the future of digital content and reading, we are honored to welcome our two keynote speakers Maribeth Back from FX Palo Alto and Natasa Milic-Frayling from Microsoft Research. Maribeth Back shares some of the ideas behind the design and research process that led to creating eleven different experiences of new forms of reading for a museum exhibit called "XFR: eXperiments in the Future of Reading" [2]. Natasa Milic-Frayling highlights some of the challenges brought on by the unique nature of digital media and describes recent research efforts focused on devising methods and techniques to facilitate long term access to digital content [6].

The remainder of the paper summarizes the accepted contributions, grouped by the common themes that emerged.

---

[1] http://www.publishers.org/press/62/

2. ACCEPTED PAPERS

2.1 Search and Discovery

Among the contributions that focus on how to support readers in deciding what to read, the common theme is to apply advanced IR techniques to guide readers not only by looking at author names and bibliographic references but also by considering writing styles, reading abilities and personal interests.

Roman Rädle, Andreas Weiler, Stephan Huber, Hans-Christian Jetter, Svetlana Mansmann, Harald Reiterer and Marc Scholl propose a solution for the collaborative searching and serendipitous discovery of large digital book collections [8]. They present an interactive tabletop prototype for collaborative visual information seeking in a large online book repository. The tabletop interface and the use of tangible items stimulate collaboration among users. The main features of the prototype are the use of a visual representation of user queries, the facilitation of back-and-forth comparisons of search results and similarity search.

Exploiting co-citation relationships, Masaki Eto [4] explores ways to enhance co-citation retrieval and proposes a measure that uses a spread co-citation relationship, which is a linkage between pairs of documents that share co-cited documents. Two retrieval methods using the spread co-citation relationship are evaluated, including or excluding co-citation context. The initial results on a collection comprising about 152,000 academic articles show that this relationship tends to be able to detect relevant documents which are undetectable using a traditional co-citation relationship, and that using context has a positive effect to reduce the number of noise documents.

Marc-Allen Cartright, Jeff Dalton and James Allan present a novel interface called Proteus for interacting with multiple retrieval types extracted from scanned books of the Internet Archive [3]. Users can retrieve and interact with books, pages, people, locations, images, and topic clusters. The primary purpose of Proteus is to provide a rich interactive experience for users to explore collections with automatically extracted and linked entity data. The system supports seamlessly shifting perspectives between books, entities, and topics. Proteus provides a starting point for a variety of exploratory search tasks.

2.2 Personalization and Recommendation

Focusing on the needs of young readers, Maria Soledad Pera and Yiu-Kai Ng present an extended version of their book recommender system for K-12 students: eBReK12 [7]. In its original version BReK12 performs content and readability analysis to identify books that may be potentially appealing to its users. The extended version incorporates a multi-criteria analysis method which captures the complex interests of its users in order to enhance the precision of recommended books. The system is further enhanced by new ways of predicting the readability levels of books, for example, by considering the topical density of its content.

Paula Cristina Vaz, David Martins de Matos and Bruno Martins look at how to help readers choose a book in a reading for fun scenario [9], a very different situation then the traditional search by topic performed by content-based recommendation systems used to recommend scientific papers or news. They suggest that writing style has a strong influence on readers' choices and propose a hybrid set-up that combines a collaborative filtering algorithm with stylometric relevance feedback. Their findings are promising as they suggest that writing style indeed influences book selection, and that book content, characterized by writing style, can be used to improve collaborative filtering results.

2.3 Reading Experiences Beyond Text

The contributions in this category focus on studying or designing rich, interactive and social reading experiences and tools and concepts to support or to evaluate these. The concept of reading is stretched and expanded to go beyond text and embrace a more multimedia reading experience, particularly suitable to special need readers.

David A. Evans and John Reichenbach argue for the need to automatically generate narration for audio books to benefit the reading-disabled community [5]. They make their case describing the state of the art in audio books and in Text To Speech (TTS) technology. They suggest that the best current TTS engines are approaching the quality threshold that will create natural-sounding prosody and permit variable-speed TTS synthesis to be comprehensible and acceptable for automatically generated narration. This would then allow to make the long tail of book titles accessible as audio books.

Abbas Attarwala, Ronald Baecker and Cosmin Munteanu tackle various accessibility challenges and present a new iPad application, the Accessible, Large-print, Listening, Talking e-book (ALLT) [1]. ALLT extends the capabilities of standard e-book readers to enable individuals with a variety of sensory and motor impairments to experience reading. ALLT provides the capabilities of electronic large-print and talking books, as well as, the ability to synchronize an audio recording of a user reading aloud with the text. This, for example, enables visually impaired users to re-experience a reading session with a loved one through replaying a recording of the session.

3. ACKNOWLEDGMENTS

4. REFERENCES

[1] A. Attarwala, R. Baecker, and C. Munteanu. Accessible, Large-Print, Listening & Talking E-book (ALLT). In *CIKM BooksOnline*, 2012.

[2] M. Back. Revisiting the Future of Reading: The Research and Design Behind XFR. In *CIKM BooksOnline*, 2012.

[3] M.-A. Cartright, J. Dalton, and J. Allan. Search and Exploration of Scanned Books. In *CIKM BooksOnline*, 2012.

[4] M. Eto. Spread Co-citation Relationship as a Measure for Document Retrieval. In *CIKM BooksOnline*, 2012.

[5] D. A. Evans and J. Reichenbach. Need for Automatically Generated Narration. In *CIKM BooksOnline*, 2012.

[6] N. Milic-Frayling. The Future of Digital. In *CIKM BooksOnline*, 2012.

[7] M. S. Pera and Y.-K. Ng. Personalized Recommendations on Books for K-12 Readers. In *CIKM BooksOnline*, 2012.

[8] R. Rädle, A. Weiler, S. Huber, H.-C. Jetter, S. Mansmann, H. Reiterer, and M. Scholl. eBook meets Tabletop: Using Collaborative Visualization for Search and Serendipity in On-line Book Repositories. In *CIKM BooksOnline*, 2012.

[9] P. C. Vaz, D. M. de Matos, and B. Martins. Stylometric Relevance-feedback towards a Hybrid Book Recommendation Algorithm. In *CIKM BooksOnline*, 2012.