eBooks in the Cloud: Desirable Features and Current Challenges for a Cloud-based Academic eBook Infrastructure

Wei Shen

GESIS – Leibniz Institute for the Social Sciences, Lennéstr. 30, 53113 Bonn, Germany.
Email: wei.shen@gesis.org.

Ute Koch

GESIS – Leibniz Institute for the Social Sciences, Schiffbauerdamm 19, 10117 Berlin, Germany.
Email: ute.koch@gesis.org.

Abstract: *With the rapid development of the mobile technology, the use of multifarious mobile devices, such as tablet PCs, eReader devices and mobile phones for electronic reading, has become an important part of everyday life. In December 2010, Google launched its Google eBookstore with more than 3 million e-books. The most significant feature of the new Google application is that it stores the book content in the cloud, which enables the reader to access the book content at any time on any mobile device in a seamless manner. A cloud-based eBook infrastructure has the characteristics of providing a vast amount of publication information, storing and processing the data in the cloud and displaying the results on demand on diverse desktops or mobile services. Taking into account the characteristics of a cloud-based eBook infrastructure, the paper examines the current eBook distribution lifecycle consisting of eBook publishing, cloud computing and mobile reading technologies. Potential features necessary for a cloud-based academic eBook infrastructure which will better support publishing, searching and reading on the web, as well as promote communication in the scientific community, are outlined.*

Keywords: *Cloud-based eBook infrastructure; semantic publishing; mobile devices; mobile reading; electronic books; collaboration; information search.*

Introduction

Neither mobile technologies nor electronic book publishing is a newly emerging field, but with the success of Amazon's Kindle and the rapid growth of Amazon eBook store, they have received increasing attention in recent years and are becoming the craze fields of innovations.

There is a wide variety of dedicated eBook readers on the market today. EBooks can be read on various portable mobile devices, such as mobile phones, PDAs, tablet PCs. The rapid development of specialized software applications for mobile services (iBook, Stanza and Good Reader) provide new reading possibilities.

In December 2010, the search engine giant, Google, entered the eBook sales world and launched Google eBooks offering more than three million titles. One important feature of the new Google application is the ability to store the content of a book in the Google-Cloud. It enables the reader to access any eBook content "seamlessly at any time on any device" (Google books overview, 2010). The current world's largest eBook online store Amazon unveiled an expanded version of *Kindle for the Web* early in 2011 only one month after Google launched its eBook store.

The idea of a cloud-based model for an eBook system is not new. Safari Books (Version 6.0) released in October 2009 attempted to improve interactivity (e.g. inline notes) and collaboration (categorization sharing) with the basic concept "everything is always in sync because your library is in the cloud" (O'Reilly, 2009).

The Internet Archive's BookServer project (www.archive.org/bookserver) utilizes the cloud concept as well. It allows readers, booksellers, authors, libraries and publishers to access and distribute eBooks available in its distributed system through any devices they have.

The emergence of cloud-based eBook libraries is certain to provide new research opportunities, influence digital publishing and enhance the acceptance of eBooks and other reading devices by readers.

Although eBooks are noted as "the obvious next step to ring a full line-up of web-based basic library resources" (Dillon, 2001), many people remain reluctant to read digital documents from screens (O'Hara & Sellen, 1997). In particular, academic use of eBooks encounters great resistance. Michael Gorman, former president of the American Library Association, believed that "massive databases of digitized whole books, especially scholarly books, are expensive exercises in futility" (Gorman, 2004). Recently, eBooks are being presented in a much more user-friendly manner on computer screens, dedicated eReaders and mobile devices (e.g. free open eBook standard format by IDPF in the digital publishing world and the introduction of E Ink and LCD Display technology for mobile reading devices).

However, the "paper-like" reading experience or effect alone is not enough to convince the conservative book readers to accept eBooks. Instead of being the surrogate for print books, eBooks should take advantage of the dynamic nature of being digital text and provide innovative functionalities which supersede traditional paper books.

Definition and Purpose

In this paper, important and desirable features for cloud-based academic eBook infrastructure are proposed. Taking into account the perspectives of both the eBook publishers and the end users, we have examined current eBook publishing, mobile technologies on the one hand and the search and social reading behavior in scientific communities on the other. The proposed features are considered essential and necessary features in a cloud-based academic eBook infrastructure.

We adopt the definition of cloud in the white paper "A Berkeley View of Cloud Computing", namely cloud is used to define hardware and software in the datacenter, which provides applications as services referred to as Software as a Service (SaaS) over the Internet (Armbrust *et al*., 2009). With this definition of cloud, it follows that a cloud-based model of eBook infrastructure allows storing digital content in the cloud and providing end users with services for discovering, selecting and accessing the eBook they need. Furthermore, it presents end users with the possibility of reading eBooks from anywhere on any device.

Due to the vast number of eBooks in the cloud, end users require and demand that searching and discovering of eBooks in a cloud-based infrastructure be intelligent and precise. Given that nowadays, multiple devices with internet connection (PC, Laptop, Tablet PC, mobile/smart phone, etc.) are increasingly used by end users to access the eBooks on the web, it is a natural step to move eBook relevant services and applications to the cloud. Such a move facilitates communication and cooperation among end users, most notably among the users in the academic community. With eBook relevant services and applications located in the cloud environment, researchers are able to communicate and collaborate with one another at anytime and anywhere and can be benefited by using these services in the cloud during their whole research work process.

The emergence of cloud-based eBook infrastructure has altered the publishing process, from traditional printing to internet digital publishing. Recent studies by Outsell's Gilbane Group emphasize the following five disruptive technologies as crucial for publishing industries: cloud computing, mobile computing, business intelligence, semantic technology and Enterprise 2.0. They expect that these five technologies "will form the bedrock for the next generation of content selling, provisioning, and monetization" (Guenette, Trippe, & Golden, 2010). In addition to reducing cost in the process of adopting cloud-based technologies, publishers also face the challenge of understanding the shifts from the "digitized world" to a "community-centric web world".

In our view, improvement of the mechanisms for digital content discovery is the key to "eBooks in the cloud". Innovative technologies, such as collaborative recommendation and various semantic-based

approaches discussed below, provide both the end users and eBook publishers with possible means of achieving a desired level of efficiency in content discovery. However, the cost efficiency and implementation of such features will need to be evaluated by eBook publishers and eBook vendors.

Using Cloud-based Infrastructure to Select and Identify the eBook Needed

While cloud computing technology facilitates access to a vast number of electronic books on the web, the task of providing a seamless reading environment with good search experience has been of great interest to the researchers.

Content Searching Inside the Book

Full text search is the most used strategy for searching in a book in the electronic environment. But if the reader types a synonym or misspells the search term, the search result is not what s/he is looking for. A semantic search service in the cloud should be the appropriate tool to simplify the search process for great convenience. With the current semantic technology, such semantic search services can operate on the input search term in the eBook cloud environment, predict the term the user is searching for in the book with artificial intelligence and return the searching results to the reader's dedicated reading device properly.

Academic readers often search for footnotes and citations for their research goals. Sometimes a URL link is addressed by the author in the reference list, if the full text paper or electronic book is available on the web. Otherwise, the reader must search the literature via a web search engine, such as Google Scholar, or go to the library to find the literature s/he needs. On account of this, hyperlinking the reference lists of the book, which provides the reader with further information on the references, for example the functionality to indicate in the reference list of a book itself whether the full text of the referenced literature is available in the cloud library, is desirable.

Searching among the eBook Collections in the Cloud

The substantial numbers of eBook collections in the cloud and the capability of consistent access to the content across variable mobile devices, enable the user access to effortlessly and rapidly distributed eBooks on the web, conversely making search among the book collections in the cloud more challenging. The semantic web advocates a revolution that presents the publication information in a meaningful way and facilitates efficient data aggregation, and the capability of consistent access to the content across variable mobile devices accelerates automated information discovery. Therefore approaches of "semantic-aware search" are needed to bridge the gap between the heterogeneous publication data providers over the internet and to improve the readers' search experience.

Ontology is machine processable meta information which describes data semantics (Fensel, 2003). The task of developing a general eBook-cloud ontology model is to conceptualize the eBook-cloud domain knowledge using Web Ontology Language (WOL) (W3C, 2004), which is composed of a set of modular ontologies with relevant domain concepts, and relationship between the concepts. The goal of a general eBook-cloud ontology model is to "enable semantic interoperability between resources, services, databases, and devices via inter-related knowledge structures" (Cheung, Drennan, & Hunter, 2008).

Impact of Semantic Web Publishing

Shotton (2009) stated that semantic publishing aims to enrich the semantic tenor of a publication, enable the automatization of information discovery, compute the semantic relevance to other publications, aggregate the heterogeneous data sources, and promote the interactivity of data within the publication.

As with eJournals[1] STM publishers are vanguards in implementation of new ways in publishing eBooks and exploring new features that were mostly launched first in textbooks within eLearning environments.

If publishers in recent years thought about "writing once, publishing in multiple formats", they should now contemplate making the content readable not only for humans but also for machines. If this can be solved, content will not only be significantly improved but also easily found. In addition, it will enable a broader view on the context of research output, as well as provide substantial benefits to academic communication, particularly when social networks become an integrated part of the eBook cloud.

Utilizing semantic web technologies like RDF (W3C, 2004), ontologies, taxonomies, approaches are already being made to achieve the goal of semantic publishing. The PAUX Company has developed an innovative content management system in which the basic content component is semantic micro-content, linked to each other not by hyperlinks but by "PAUX-Links". These PAUX-Links can be rated and extended to enhance the semantic meaning of the contents, which enables semantic search and filtering.

The cloud-based concept of eBook systems has the potential and points the way: "we need servers which expose existing data systems using the RDF standards, documents 'marked up' and ontologies" so enhanced publications can interact in a semantic way (Vernooy-Gerritsen 2009).

Establishing Multiple Recommendation Services for eBooks in the Cloud

Recommendation tools are often seen as a useful supplement mechanism to information discovery. Wang (1998) defined a recommender system as "a system that has as its main task, choosing certain objects that meet the requirements of users, where each of these objects are stored in a computer system and characterized by a set of attributes".

Cloud-based eBook infrastructure provides readers with a distributed framework not only for sharing and synchronizing over mobile reading devices, but also for opening up an opportunity for deducing and recommending the valuable eBook information that meets the needs of the reader actively.

The following are some proposed recommendation services for an academic eBook cloud-based infrastructure.

A content-based recommendation service should be a useful tool for mobile readers to discover further eBook items that they would be interested in reading. According to content-based filtering algorithms, this service retrieves and analyzes the metadata of the item individuals buy or read and recommends other eBook items with similar content. Typical examples are functionalities such as the "see more related books" button or "new arrivals". The reader can also create a personal profile that instructs the cloud infrastructure to operate his preference inquiry and make recommendations to his end mobile device.

Collaborative recommendation service is "based on collecting and analyzing a large amount of information on users' behavior, activity or preferences and predicting what users will like based on their similarity to other users" (TVGENIUS, 2010). A cloud-based eBook infrastructure is capable of collecting the interactive user information implicitly from web-connected handheld mobile devices or multiple desktops the reader uses. With the help of these data, the collaborative recommendation service can compare and calculate the different interaction attributes such as watch list of the user, previous reading history, the frequency of an eBook access of the reader as well as reading time of an eBook, and finally recommend eBooks stored in the cloud to the user's device.

For academics, recommendations from colleagues or researchers with the same research interest play an important role in the eBook finding process. Social recommendation service is widely used in the social network portals such as Facebook. The eBook cloud-based infrastructure can incorporate such

[1] The "Article of the Future" project from Elsevier and Cell Press, the semantically enhanced version of an original PLoS NTD paper by Reis et al. (2008), undertaken by Shotton, Portwin, Klyne, & Miles (2009), or the Semantic Biochemical Journal are just some of the works worth mentioning here.

a social recommendation service that allows the reader to evaluate the eBooks s/he read and recommend them to the community.

Using Cloud-based eBook Infrastructure to Facilitate Collaboration and Communication among Academic Communities

Book reading in a group is a social activity and requires more "in-use features" (Vasileiou & Rowley, 2008) such as annotations, highlights and bookmarking. For academics who have the same research goals, social reading activity is common and has been proved to exchange research ideas, contributions and research results efficiently. As cloud computing technology offers institutional, corporate and individual users the ability to share storage space and application resources from a central, off-site location accessible from the internet (Carr, 2009), meanwhile the online activity of academic staff and students is increasingly distributed across multiple devices; this centralized facility can promote simultaneous collaboration and allow the researchers independence from their physical location, and the reading devices they use work on shared digital text together. Current applications of cloud-based eBook infrastructure by Google eBook do not support this kind of user collaboration activities in the Google environment; the user can neither bookmark the location in the Google eBook nor highlight the text .

One of the challenges is how to handle the shared references (Marshall, 2009) in eBooks on different reading devices. The traditional page number is apparently still not suitable for shared eBook referencing. Different eBook reading devices transfer the same eBook with different page numbering schemes, and the page number always changes depending on the zoom level. Thus, page numbers are not a consistent way of referencing in mobile devices. However, the eReader producers have already provided possible solutions. For instance, the Kindle eReader adopts document location to identify the specific location within the eBook. Google Reader Software offers another page referencing model, which intelligently displays the actual print page number of the physical book. In the case of a changed zoom level, a page number indicates the exact page number of the print page, such as "page 2-3". However, since such resolutions are all device-specific, interchanges with diverse mobile services are still obviated. In order to allow collaboration among the common reading practices mentioned above in the community, dynamic eBooks should be published with semantics encoded. The basic idea is similar to the "PAUX"-concept: fractionizing the eBook fragments to a fine granularity level. The fragment can be a paragraph, a sentence, or a word, linking these fragments semantically. Furthermore, hard/software of eReader devices or other mobile eBook reading devices should be improved and developed continuously for a better customization possibility in dynamic eBooks.

We have further observed that the process of actively sharing knowledge among academics and students in a social reading group is extremely goal-oriented. It often aims to allow team-work to achieve a specific goal, such as a research paper or a term paper. There are a number of collaboration applications developed for such a purpose. For instance, the application Google Docs enables users to share and work on a document in the Google-Cloud. However, handheld eBook reading devices are designed to enable the user to read and provide little support for the writing activity. This is also the case in Apple iPad, whose virtual touch keys continue to present a certain level of inconvenience, even in writing a short e-mail. An easy-to-use writing functionality is necessary in mobile reading devices.

Despite the hardware problem of writing, the design of collaboration tools, so-called "eBook publishing tools in the cloud", is also very challenging. Specific Content Management Services should be composed to organize and manage the consistency and composition of the content units with varied media type and data size, in order to make them "look and feel like a book", and another important feature for such a collaborative publishing application is certainly the automated apposition of semantic meaning for an enhanced eBook. Furthermore, a reviewing process for quality control is also a necessary application for the Cloud Content Management Service.

Personalization

The prerequisite to enjoying "seamless reading" of Google eBooks is that the reader own a Google user account. Personalization in the cloud environment is a general trend. It is a requisite factor for

many add-value features such as the personalized recommendation service mentioned above. Since storing the personal user profile in the cloud has the advantage of data persistence and consistence across the multiple devices, the personalization feature is highly desirable. Nevertheless personalization also brings challenges in user security. Taking an example of the famous eBook seller Amazon, Amazon remotely deleted eBooks from the customers' Kindles without any prior notification (Stone, 2009). It is imperative that any cloud service providers not only protect the sensitive personal data stored in the cloud, but also obey the privacy requirements of the customer.

Conclusions

Google CEO Eric Schmidt (2007) stated in an interview that "It's pretty clear that there's an architectural shift going on. These occur every 10 or 20 years. The previous architecture was a proprietary network with PC clients called client-server computing. With this new architecture you're always online; every device can see every application; and the applications are stored in the cloud." Cloud computing was one of the technology trends in 2010; it is also expected to be the case in 2011.

When Tom Mulally, founder of Numagic, a consulting company specializing in new media and emerging technology projects, presented a concept for the next generation of content production process, he stated that

> first the cloud would give us instant connectivity to whatever we need, when we need it, with transactions handled automatically. Then semantic search and related technologies would be combined to provide decision support, analytics, reasoning. . .essentially an omniscient intelligent-assistant running in the background. And the multi-sensor capabilities of mobile smart devices (iPhone, Android, iPad, etc.) connected via high speed wireless creates a sort of positive feedback loop, keeping all stakeholders, including the end user/client as active participants throughout the process. (Mulally, 2010)

Cloud computing and mobile technology continue to influence the development of eBook reading. They have the potential of revolutionalizing the traditional publishing industry. We consider it extremely important that approaches be taken to strengthen and improve the technologies in these two areas.

There are a variety of impacts on electronic publishing which we did not consider in detail here. For example James McQuivey from Forrester Research argued "the ultimate effect of Google eBooks. . . will be the creation of an ad-supported publishing model" (McQuivey, 2010). Another change discussed by Hugh McGuire on the O'Reilly radar points to the arbitrary distinction between "the internet" and "books" which he expects will disappear in the near future (McGuire, 2010). Nevertheless, the features and challenges summarized in the paper provide a step forward into an intelligent cloud-based academic eBook infrastructure with the capability of effective publishing, disseminating and sharing electronic resources on the web. These issues continue to be the focus of our work. With the initial findings, we hope to open up new avenues of research in this arena.

References

Armbrust, M., et al. (2009). Above the clouds: A Berkeley view of cloud computing. Technical Report EECS-2009-28, EECS Department, University of California, Berkeley. Retrieved May 2, 2011 from http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

Carr, J. (2009). Box up your files and move 'em to the cloud'. *Information Today*, *26*, 20.

Cheung, K., Drennan, J., & Hunter, J. (2008). Towards an ontology for data-driven discovery of new materials. In D.L. McGuinness, P. Fox & B. Brodaric (Eds.) *Papers from the AAAI Spring Symposium, March 26-28, 2008, Stanford University, CA.* (Technical Report SS-08-05). Menlo Park, CA: AAAI Press. Retrieved May 29, 2011 from http://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-05/SS08-05-003.pdf.

Dillon, D. (2001). E-books: the University of Texas experience, part 1. *Library Hi Tech*, *19*(2), 113-125.

Fensel, D. (2003). Ontologies: A silver bullet for knowledge management and electronic commerce. Retrieved May 2, 2011 from http://software.ucv.ro/~cbadica/didactic/sbc/documente/silverbullet.pdf

Google books overview. (2010). Retrieved May 2, 2011 from http://www.books.google.com/help/ebooks/overview.html

Gorman, M. (2004). Google and God's Mind: The problem is, information isn't knowledge. Retrieved May 2, 2011 from http://serials.infomotions.com/colldv-l/archive/2005/att-0054/M-Gorman___Google_and_God_s_Mind.pdf

Guenette, D.R., Trippe B., & Golden K. (2010). A blueprint for book publishing transformation: Seven essential processes to re-invent publishing. (Research report of Outsell's Gilbane Group). Retrieved May 2, 2011 from http://www.bisg.org/docs/gilbane_blueprint.pdf

Marshall, C.C. (2009). *Reading and writing the electronic book*. Morgan & Claypool Publishers.

McGuire, H. (2010). The line between book and Internet will disappear. The inevitability of truly connected books and why publishers need APIs. Posted on 10 September 2010. Retrieved January 23, 2011 from http://radar.oreilly.com/2010/09/beyond-ebooks-publisher-as-api.html

McQuivey, J. (2010). Google eBooks paves the way for ad-supported publishing. Posted on 06 September 2010, Retrieved January 23, 2011 from http://blogs.forrester.com/james_mcquivey/10-12-06-google_ebooks_paves_the_way_for_ad_supported_publishing

Mulally, T. (2010). Key emerging technology trends for media professionals. Retrieved January 22, 2011 from http://emerging-technologies-information08.blogspot.com

O'Hara, K. & Sellen, A. (1997). A comparison of reading paper and on-line documents. In *CHI '97: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 335-342). New York, NY. ACM. Retrieved May 29, 2011 from http://old.sigchi.org/chi97/proceedings/paper/koh.htm.

O'Reilly, T. (2009). Safari Books Online 6.0: a cloud library as an alternate model for ebooks. Retrieved December 20, 2010 from http://radar.oreilly.com/2009/10/safari-books-online-60-a-cloud.html

Reis, R.B., Ribeiro, G.S., Felzemburgh, R.D.M., Santana, F.S., Mohr, S., et al. (2008). Impact of environment and social gradient on Leptospira infection in urban slums. *PLoS Neglected Tropical Diseases*, 2: e228. Retrieved May 2, 2011 from http://dx.doi.org/10.1371/journal.pntd.0000228 .

Schmidt, E. (2007). Text of Wired's interview with Google CEO Eric Schmidt. Retrieved May 2, 2011 from http://www.wired.com/techbiz/people/news/2007/04/mag_schmidt_trans

Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing, 22,* 85-94.

Shotton D, Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in semantic publishing: Exemplar semantic enhancements of a research article. *PLoS Computational Biology*, 5(4): e1000361. doi:10.1371/journal.pcbi.1000361.

Stone, B. (2009). Amazon erases Orwell books from Kindle. *The New York Times*. Retrieved May 2, 2011 from http://www.nytimes.com/2009/07/18/technology/companies/18amazon.html?_r=1

TVGENIUS. (2010). TV recommendations using relevance maps. Retrieved May 2, 2011 from http://www.tvgenius.net/resources/white-papers/an-integrated-approach-to-tv-recommendations/

Vasileiou, M. & Rowley, J. (2008). Progressing the definition of "e-book". *Library Hi Tech*, 26(3), 355-368.

Vernooy-Gerritsen, M. (2009). *Enhanced publications*. Amsterdam: University Press.

W3C (2004). OWL Web Ontology Language overview. Retrieved May 2, 2011 from http://www.w3.org/TR/owl-features/

Wang, P. (1998). Why recommendation is special? In *Papers from the 1998 Workshop on Recommender Systems, part of the 15th National Conference on Artificial Intelligence (AAAI 1998)* (pp. 111-113), Madison, Wisconsin: EUA.