

ELPUB Digital Library v2.0

Application of semantic web technologies

Anand BHATT^a and Bob MARTENS^b

^a*ABA-NET/Architexturez Imprints, New Delhi, India*

^b*Vienna University of Technology, Vienna, Austria*

Abstract. This paper presents the ongoing efforts to further develop the ELPUB digital library, which highly supported the dissemination of published materials within the community in the past years. elpub.scix.net has been serving the ELPUB-community since a decade, predominantly aiming to archive the output of the annual conferences. Doubtless, there is still a need for a platform to maintain the “collective memory” of the association and so far over 700 entries from 17 conferences were recorded. The repository, which utilized the SciX-technology, delivered the user access to the individual papers and made the data available via an OAI-interface as well. However, in the course of time digital library technology evolved and an association dedicated to Electronic Publishing ought to be at the forefront of novel developments. For this reason a shift was performed towards the Architexturez platform (library.elpub.net) aiming to implement advanced semantic web features. Especially the display of evolving topics and their gradual development is appealing and moreover the aggregation of individual bibliographies. Many of the features were designed in consultation with research communities in, among others, architectural computing and real estate. While deploying features and capabilities are well established in the digital library domain, the system is designed to support further research by the ELPUB community and this paper will elaborate on the transition and deliver an overview on the current prospects along with the technical capabilities.

Keywords. Digital repository, Open Access, Semantic Web, Mining, Social interaction

1. Introduction

Shortly before the turn of the millennium, comprehensive web developments took place and solutions to create topical repositories popped up. The rationale targeted to work out digital libraries - on a shoe-string budget -, which would not disappear shortly after their launch. Furthermore, there was no business model in the background aiming to keep the digital libraries running. Contrariwise, a closed pocket model based on volunteering capacities in academic environments defined the starting point (likewise in ELPUB, which has no formal legal entity), which would then again donate an open access contribution to the community.

The initial work to setup a digital library can be characterized in these days as the extraction of metadata. Fortunately hardly any back digitization had to be handled for ELPUB, but even after a couple of elapsed years it took some efforts to gather the digital data. Indeed by its original launch a critical mass was already available and the ELPUB Digital Library has been extended after every annual crop [1].

Given these framework conditions the *SciX-technology* was a convenient working environment, however, trends were not covered, as further development did not take place. Nonetheless, the maintenance as such has been secured with insignificant disruptions in the course of time [2]. For this reason a search towards alternatives was started. In terms of quantities, we're moreover not talking about millions of records. However, even a couple of hundred recorded entries require a solution, which would go far beyond the basic idea of sustainable archiving.

2. Shift from SciX to the Architexturez Platform

The origin of the system goes back to 1999 when the creators of the system were required to develop experimental technology demonstrators, while writing technical standards for digital information management. The system has been developed over the years as the creators assisted in writing new standards, such as Unstructured Information Management Architecture (UIMA) and Extensible Resource Indicators (XRI) for the internet and provided the first statements of use.

The requirements for Architexturez were defined by way of a grant for Vienna University of Technology to initiate collaborative projects to initiate collaborative projects and proposed to the CumInCAD research community for review [3]. At the outset, it was determined that the novel platform should have added features such as named entity disambiguation, required to build lists of papers by entities such as authors and keywords, and semantic web features to better facilitate discovery and relationship mining activities in future.

2.1. New System Architecture

Figure 1 illustrates the reference architecture for the Architexturez platform. It is built with a *Drupal* web applications framework for the end-user facing tasks (items 5-7 in figure 1), from a heterogeneous set of modules, primarily based on *Apache Solr* and *Mahout libraries* for the applications part (layers 2-4), and a storage layer for data required by system. Proprietary data may be retrieved at runtime from external sources, such as *social network containers*, *citations databases* and *institutional repositories*. Unstructured Information Management Architecture (UIMA) and Mahout libraries provide, respectively, for search and advanced text analysis and document clustering capabilities.

Figure 1 illustrates a *reference architecture* for the Architexturez platform, reading from bottom-to-top, ① the data storage layer, including data connectors to external services and repositories, ② a database abstraction layer, ③ system core modules such

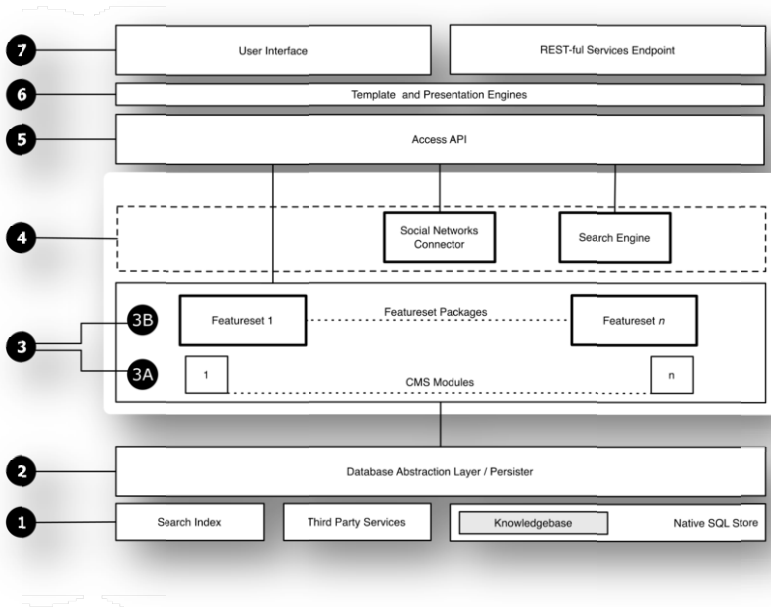


Figure 1. The Architexturez repository system architecture.

as (3a) rate calculation logic and user finder and (3b) other modules; ④ system services engines, such as faceted search engine, recommendation engine and people relevator engine, ⑤ an access API; ⑥ template engine for HTML 5 user interface and RDF/JSON, OAI-PMH and URN-NBN serializations for external customers, and ⑦ user interface and a REST services endpoint for data extraction.

If we take one instance, author names and keywords, as over time, there will occur instances where an author has used different names or abbreviations, and synonyms for the same keyword are provided. At other times there may occur multiple authors of the same name. SciX relied on human judgment in disambiguating named entities. Comparable systems, such as *CiteSeer* [4], provided machine-learning frameworks of named entities [5] and others use rank-based algorithms [6].

Architexturez preferred to combine automated methods and individual human judgment in disambiguating named entities. It unites automated and human-edited methods and clustering algorithms that build list of authors that may be related and then provides human editors with tools to make the final decision, shown in figure 2.

3. Novel Functionalities

The Architexturez-interface differs very much from the previous SciX-solution. When entering the upgraded ELPUB Digital Library environment, the home page displays the

Prefix

First Name

Initials

Last Name

Suffix

Peter

Linde

Complete Name *

Peter Linde

The value in this box will be constructed from the individual name parts fields above.

☐ Do not reformat

Selecting this will prevent the styles from trying to reformat the contributor name. The text in the "Complete Name" field will be used as is.

Affiliation

University, Company or Organization that the author is affiliated with.

Drupal User ID

(None)

If this author has a an account (Drupal User ID) on this web site, you may select it here. This will help to link the authors publications with other content.

▼ Author Link / Merge

Select other author names which will be linked or merged. Merging removes all the selected authors, then replaces any references to these authors with the author being edited above. You should only do this if you are sure the other authors represent the same author as the one being edited. **IF you do not select "Retain as alternate form" then THIS CANNOT BE UNDONE!**

Author name	Link	Merge	Retain as "alternate form"
C. Garcia Landa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Landau, Luis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Landau, Steven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Landy, S.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lenat, Douglas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ruslyanov Leonid	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hans Lind	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tay Linda	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Linde, B.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Linde, Peter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lund, Andreas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jens Lunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other authors which could be linked to this author

Add author

Save

Cancel

Figure 2. Author disambiguation screen – the system proposes clusters of authors with similar names, and human editors can (1) merge authors (2) retain author names as alternate forms or (3) do nothing.

previous set of conferences by way of the published proceedings. However, a user can also enter a search string at the top right. The bottom region of the page allows to embed aggregated information packages, such as news, calls for papers, or link collections, etc.

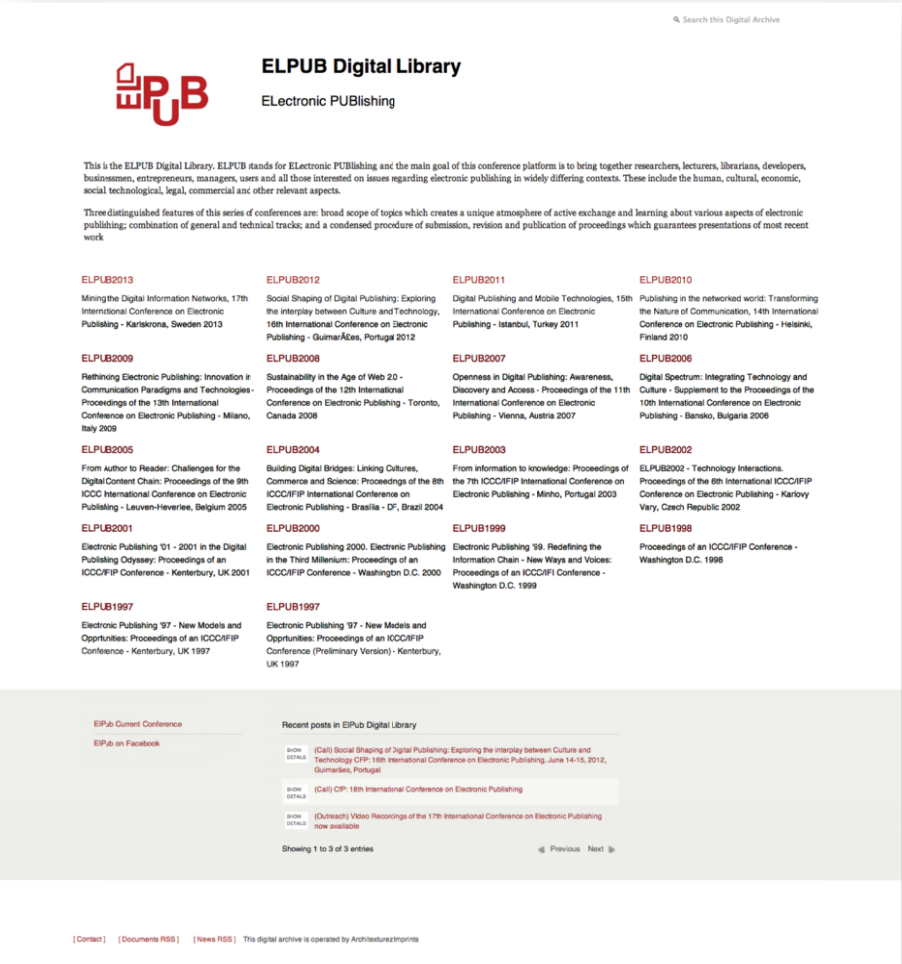


Figure 3. Architecturez: Screenshot of the front page.

Once a user has selected certain proceedings, a new screen will open, where in the central part the recorded entries are displayed. On the left side, there is a collation of keywords, and on the right a list of contributing authors in these proceedings appears. Paper abstracts and keywords are searchable on this screen as well (see figure 4).

The screenshot displays the ELPUB2013 website interface. At the top, the ELPUB logo is on the left, and the title 'ELPUB2013' is in the center. Below the title, a subtitle reads: 'Mining the Digital Information Networks, 17th International Conference on Electronic Publishing, Edited by Niklas Levisson, Peter Jönde and Panayiota Polydoratos. ELPUB, Karlskrona, Sweden: IOS Press BV, 2013.' A search bar is located in the top right corner with the text 'Search this Digital Archive'.

On the left side, there is a 'Keywords in ELPUB2013' section with a list of keywords including: 3-D Intersection, 3-D Visualization and Navigation, adaptive hypertext, Algorithms, Article-Level Metrics, Clustering, Cognitive Simulation, Conference Management System, Connectionism and Neural Nets, Controlled Vocabularies, Copyright, Cover Sheets, Crawling, Data Mining, Data Mining and Intelligent Computing, Deduplication, Design, Design Since, Design Since Research, Digital Humanities, and Digital Libraries.

In the center, there is a 'Show 100 entries' section. Below this, there are four entries, each with a 'SHOW DETAILS' link and a brief description of the paper. The entries are:

- de Antonio, Angélica, Cristian Moral, Daniel Kispel, and Martín J. Abente. "3D gesture-based exploration and search in document collections." In *Mining the Digital Information Networks, 17th International Conference on Electronic Publishing*. ELPUB, Karlskrona, Sweden: IOS Press BV, 2013.
- Caldera, Christian, René Berndt, and Dieter W. Fellner. "COMFY - A Conference Management Framework." In *Mining the Digital Information Networks, 17th International Conference on Electronic Publishing*. ELPUB, Karlskrona, Sweden: IOS Press BV, 2013.
- Tonkin, Emma I., Stephanie Taylor, and Gregory J. L. Tourte. "Cover sheets considered harmful." In *Mining the Digital Information Networks, 17th International Conference on Electronic Publishing*. ELPUB, Karlskrona, Sweden: IOS Press BV, 2013.
- Wendykier, Piotr. "Deduplication of metadata harvested from Open Archives Initiative repositories." In *Mining the Digital Information Networks, 17th International Conference on Electronic Publishing*. ELPUB, Karlskrona, Sweden: IOS Press BV, 2013.

On the right side, there is an 'Authors in ELPUB2013' section with a list of authors including: Abente, Martín J., Amado Alves, Mario, Andersson, Stefan, Bankin, Marianne, Baptista, Ana Alice, Berndt, René, Caldera, Christian, Croel, Stacy, Curado Malta, Mariana, de Antonio, Angélica, Di Cesare, Rosa, Erlandsson, Fredrik, Fellner, Dieter W., Fenner, Martin, Firmino, Helder Noel, Freire, Nuno, Giannini, Silvia, Hubbard, Bill, Hussain, Azhar, Jahn, Najko, Johnson, Henric, and Kispel, Daniel.

Figure 4. Display of proceedings with collated keywords and authors.

At this point the next step allows to follow a certain concept on the one hand to follow a keyword or, on the other hand, to trace back the publication output of a selected author. Especially novel members in the ELPUB community may take real advantage, in order to retrieve relationships between the evolving ELPUB topics and the involved authorship. At present predominantly PDF-files are serving as "multimedia" documents. However, other media types could be attached as well.

In term of intuitive searching/browsing, two features are being offered: (1) off-device navigation displaying citations and networks, (2) system-generated vector-space model to assist users in discovering new relationships. For the further development of the Digital Library as it stands, feedback from the side of the ELPUB-community is more than welcome.

The system embraces tools for ontology extraction for offline analysis of the documents and citations. It also provides a vector space model, for filtering and ranking relevant documents.

An important check such proposed and system-generated knowledge models must pass is the test of intuition, when authors and subject-experts review the related content. They should affirm that the related content contains documents that they would consider the most similar to the record being displayed.

ELPUB Digital Library - volume: ELPUB2013

Search this Digital Archive



3D gesture-based exploration and search in document collections

PDF Download



This paper describes an approach towards the interaction with 3D representations of large document collections. The goal was to provide the user with a highly dynamic environment in which even the very mapping strategy to position documents in space can be adjusted by the user depending on the specific task at hand, on his preferences, or on the context. A modification to the FDP algorithm is proposed, as well as a new gesture-based interaction paradigm in which the user can explore and search information in the collection just by simple hand movements. An experimental user evaluation was carried on to investigate the impact of the proposed approach on the precision of the mental model built by users through exploration, on the effectiveness in information search tasks, and on the general user satisfaction and perception of utility.

de Antonio, Angelica, Cristian Moral, Daniel Klepel, and Martin J. Abente. "3D gesture-based exploration and search in document collections." In *Mining the Digital Information Networks. 17th International Conference on Electronic Publishing. ELPUB. Karlskrona, Sweden: IOS Press BV, 2013.*

Keywords

K-Means, Information Retrieval, Force-Directed Placement, Clustering, 3-D Visualization and Navigation and 3-D Interaction

More like this

(1) *Digital Publishing and Mobile Technologies. 15th International Conference on Electronic Publishing*. Edited by Yasar Tonta, Umut Al, Phyllis Lepor Endogan and Ana Alice Baptista. ELPUB. Istanbul, Turkey, 2011..

(2) Krottmaier, Harald. "The need for sharing user profiles in digital libraries." In *Building Digital Bridges: Linking Cultures, Commerce and Science: Proceedings of the 8th ICCG/IFIP International Conference on Electronic Publishing*. ELPUB. Brasilia - DF, Brazil: Universidade de Brasilia, 2004.

(3) Gancarski, Alda Cristina R., and Pedro Manuel San Henriques. "IXDIRQL: An interactive XML Data and Information Retrieval Query Language." In *From information to knowledge: Proceedings of the 7th ICCG/IFIP International Conference on Electronic Publishing*. ELPUB. Minho, Portugal: Universidade do Minho, 2003.

(4) Rutledge, Lloyd, Lynda Hardman, Jacco van Ossenbruggen, and Dick C. A. Bulterman. "ADDRESSING PUBLISHING ISSUES WITH HYPERMEDIA DISTRIBUTED ON THE WEB." In *Proceedings of an ICCG/IFIP Conference*. ELPUB. Washington D.C.: ICC Press, 1998.

(5) Golovuz, Petr, Ilyashchuk, and Karel Izak. "A web-based user profile generator foundation

Citation Links

EndNote Tagged - RIS

Figure 5. Document record, with (1) a system generated preview of the full content and download link, (2) document title and abstract, (3) links to the author and keyword page, (4) related content provided by the vector space model, and (5) citation links.

4. Conclusion and Outlook

This paper described the switch of the ELPUB Digital Library towards a novel environment. The design parameters have been pointed out and a depiction of the system’s interface was presented. It is to be regarded as an open invitation to the ELPUB community to use the novel platform, i.e. to experiment and research beyond the annual conferences.

The contribution intends to think about possible options for further developmental tracks. Above all the ELPUB community should be on the edge of novelties as far as the own publication output is concerned. For example data aggregation can be regarded as an important issue, where “internal” and “external” data sets are combined in a structured way.

A system to harvest citations and references to ELPUB papers has been created, and it is currently processing references from external sources. It has been stated already many times, that the amount of information is exponentially growing and even so, the need for expedient navigation (sorting) is becoming more and more important.

References

- [1] Martens, B., Linde, B. and Turk, Z. (2003) A Digital Library for ELPUB Proceedings: The Use of a Web-Based Prototype ELPUB2003. From information to knowledge: Proceedings of the 7th ICC/IFIP International Conference on Electronic Publishing held at the Universidade do Minho, Portugal 25-28 June, 2003.
- [2] Martens, Bob; Linde, Peter; Kline, Robert; Holmberg, Per (2008) Enhancing the sustainability of electronic access to ELPUB proceedings: means for long-term dissemination ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing held in Toronto, Canada 25-27 June 2008.
- [3] Bhatt Anand; Martens, Bob (2006) ON-TO-CAAD: Investigating the Knowledge Contained within the Corpus of CAAD Research CAADRIA 2006 [Proceedings of the 11th International Conference on Computer Aided Architectural Design Research in Asia] Kumamoto (Japan) March 30th - April 2nd 2006.
- [4] http://web.mit.edu/seйда/www/Papers/IST-TR_DisambiguationCiteSeer.pdf
- [5] <http://www.kdd.org/kddcup2013/sites/default/files/papers/track2place2.pdf>
- [6] http://research.microsoft.com/pubs/154452/CIKM_CameraReady.pdf