



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2006

TeNDaX, a Collaborative Database-Based Real-Time Editor System

Leone, Stefania ; Hodel-Widmer, Thomas B ; Böhlen, Michael ; Dittrich, Klaus R

Abstract: TeNDaX is a collaborative database-based real-time editor system. TeNDaX is a new approach for word-processing in which documents (i.e. content and structure, tables, images etc.) are stored in a database in a semi-structured way. This supports the provision of collaborative editing and layout, undo- and redo operations, business process definition and execution within documents, security, and awareness. During document creation process and use meta data is gathered automatically. This meta data can then be used for the TeNDaX dynamic folders, data lineage, visual- and text mining and search. We present TeNDaX as a word-processing ‘LAN-Party’: collaborative editing and layout; business process definition and execution; local and global undo- and redo operations; all based on the use of multiple editors and different operating systems. In a second step we demonstrate how one can use the data and meta data to create dynamic folders, visualize data provenance, carry out visual- and text mining and support sophisticated search functionality.

DOI: https://doi.org/10.1007/11687238_76

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-56137>

Conference or Workshop Item

Originally published at:

Leone, Stefania; Hodel-Widmer, Thomas B; Böhlen, Michael; Dittrich, Klaus R (2006). TeNDaX, a Collaborative Database-Based Real-Time Editor System. In: EDBT 2006, Munich, Germany, 26 March 2006 - 30 March 2006, 1135-1138.

DOI: https://doi.org/10.1007/11687238_76

TeNDaX, a Collaborative Database-Based Real-Time Editor System A Word-Processing ‘LAN-Party’

Stefania Leone¹, Thomas B. Hodel-Widmer²,
Michael Boehlen³, and Klaus R. Dittrich¹

¹ University of Zurich, Department of Informatics,
Winterthurerstrasse 190, 8057 Zurich, Switzerland
{leone, dittrich}@ifi.unizh.ch

<http://www.tendax.net>

² Swiss Federal Institute of Technology (ETH Zurich),
Leonhardshalde 21, 8092 Zurich, Switzerland
hodel@sipo.gess.ethz.ch

³ Free University of Bolzano-Bozen, Piazza Domenicani 3,
39100 Bolzano, Italy
boehlen@inf.unibz.it

Abstract. TeNDaX is a collaborative database-based real-time editor system. TeNDaX is a new approach for word-processing in which documents (i.e. content and structure, tables, images etc.) are stored in a database in a semi-structured way. This supports the provision of collaborative editing and layout, undo- and redo operations, business process definition and execution within documents, security, and awareness. During document creation process and use meta data is gathered automatically. This meta data can then be used for the TeNDaX dynamic folders, data lineage, visual- and text mining and search.

We present TeNDaX as a word-processing ‘LAN-Party’: collaborative editing and layout; business process definition and execution; local and global undo- and redo operations; all based on the use of multiple editors and different operating systems. In a second step we demonstrate how one can use the data and meta data to create dynamic folders, visualize data provenance, carry out visual- and text mining and support sophisticated search functionality.

1 Introduction

Text documents are a valuable resource for virtually any enterprise and organization. Documents like papers, reports and general business documentations contain a large part of (business) knowledge. Documents are mostly stored in a hierarchical folder structure on file servers and it is difficult to organize them with regard to classification, versioning etc., although it is of utmost importance that users can find, retrieve and edit documents in a user-friendly way.

In most of the commonly used word-processing applications documents can be manipulated by only one user at a time and tools for collaborative document editing and management are rarely deployed. Documents should be seen as a valuable

business asset which requires an appropriate data management solution. The need to store, retrieve and edit these documents collaboratively with guaranteed mechanisms for security, consistency, availability and access control is obvious.

In the following, we present the database-based TeNDaX editor system which enables collaborative document editing and management, all within a interactive multi-user database environment.

2 The TeNDaX Editor System

TeNDaX stands for a Text Native Database eXtension. It enables the storage of text in databases in a native form so that text editing is finally represented as real-time transactions. Under the term ‘text editing’ we understand the following: writing and deleting text (characters), copying and pasting, defining layout and structure, inserting notes, setting access rights, defining business processes, inserting tables, pictures, and so on, i.e. all the actions regularly carried out by word processing users. By ‘real-time transactions’ we mean that editing text (e.g. writing a character/word) invokes one or several database transactions so that everything which is typed appears within the editor as soon as these objects are stored persistently. Instead of creating files and storing them in a file system, the content and all of the meta data belonging to the documents is stored in a special way in the database, which enables very fast transactions for all editing tasks [3]. The database schema and the transactions are designed to be used in a multi-user environment, as is customary in the database context. As a consequence, many of the database features (data organization and querying, recovery, integrity and security enforcement, multi-user operation, distribution management, uniform tool access, etc.) are now, by means of this approach, also available for word processing.

TeNDaX creates an extension of DBMS to manage text. This addition is carried out ‘cleanly’ and the responding data type represents a ‘first-class citizen’ [1] of a DBMS (e.g. integers, character strings, etc.).

Since the document data is stored in the database, we automatically gather meta data during the whole document creation process [6]. We gain meta data on document level (creator, roles, date and time, document object ID, document names, structure affiliation, note affiliation, security settings, size, authors, readers, state, places within static folders and user defined properties), on character level (author, roles, date and time, copy-paste references, local and global undo / redo, security settings, version and user defined properties) and from structure, template, layout, notes, security and business process definitions.

3 Demonstration: Word-Processing ‘LAN-Party’

In our word-processing ‘LAN-Party’ we focus on the TeNDaX editor system. Editors installed on different operating systems (Windows XP, Linux, Mac OSX) will support the demonstration of the following TeNDaX features:

- **Collaborative editing:** We will concurrently work with multiple users on the same document. Editing a document includes operations like writing and deleting characters [4], inserting pictures, creating tables, applying layout and structure [2], local and global undo- and redo operations, setting access rights etc. All these operations are carried out in a dynamic multi-user database environment.
- **Business process definitions and flow:** We will define and run a dynamic workflow within a document for ad-hoc cooperation on that document [5]. Tasks such as translation or verification of a certain document part can be assigned to specific users or roles. The workflow tasks can be created, changed and routed dynamically, i.e. at run-time.
- **Dynamic Folders:** On the base of the automatically gathered document creation process meta data we will build dynamic folders [6]. Dynamic folders are virtual folders that are based on meta data. A dynamic folder can contain all documents a certain user has read within the last week. Its content is fluent and may change within seconds (e.g. as soon as a document changes). This represents a novel method for document management and text retrieval.
- **Data Lineage:** We can display document content provenance. Meta data about all editing and all copy- and paste actions is stored with the document. This includes information about the source of the new document part, e.g. from which other document a text has been copied (either internal or external sources). We use this meta data to visualize data lineage as depicted in Figure 1.
- **Visual Mining:** The information visualization plug-in provides a graphical overview of all documents and offers a variety of interaction modalities. It is possible to navigate the document and meta data dimensions to gain an understanding of the entire document space. A visualization of a set of documents is shown in Figure 2.
- **Search:** The meta data based searching and ranking plug-in offers sophisticated search options. Documents and parts of documents can either be found based on the document content, or structure, or document creation process meta data. The search result can be ranked according to different ranking options, e.g. ‘most cited’, ‘newest’ etc.

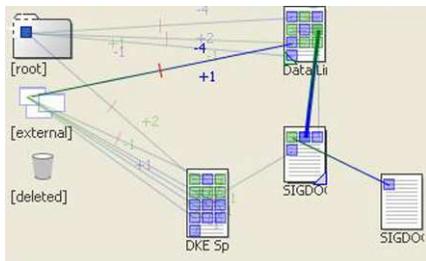


Fig. 1. Data Lineage

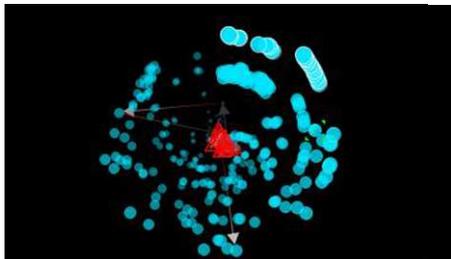


Fig. 2. Visual Mining

4 Conclusion

The collaborative database-based real-time TeNDaX editor system enables the storage and management of documents within a database. It natively represents text in fully-fledged databases, and incorporates all necessary collaboration support. It offers functions such as editing, awareness, fine-grained security, sophisticated document management, versioning, business processes, text structure, data lineage, text and visual mining - all within a collaborative multi-user database environment.

Within the above presented TeNDaX editor system we use database technology to provide full word-processing functionality and sophisticated document- and meta data visualization. TeNDaX extends database technology and offers database-based universal data management for text documents.

References

1. Abiteboul, S., Agrawal, R., Bernstein, P., Carey, M., Ceri, S., Croft, B., DeWitt, D., Franklin, M., Molina, H. G., Gawlick, D., Gray, J., Haas, L., Halevy, A., Hellerstein, J., Ioannidis, Y., Kersten, M., Pazzani, M., Lesk, M., Maier, D., Naughton, J., Schek, H., Sellis, T., Silberschatz, A., Stonebraker, M., Snodgrass, R., Ullman, J., Weikum, G., Widom, J., and Zdonik, S.: The Lowell Database Research Self-Assessment, CACM Vol. 48(5), 111-118 (2005)
2. Hodel, T. B., Businger, D., Dittrich, K. R.: Supporting Collaborative Layouting in Word Processing. IEEE CoopIS (2004) 355–372
3. Hodel, T. B., Dittrich, K. R.: Concept and Prototype of a Collaborative Business Process Environment for Document Processing. Data & Knowledge Engineering, 52, Special Issue: Collaborative Business. Process Technologies (2004) 61–120
4. Hodel, T.B., Dubacher, M., Dittrich, K. R.: Using Database Management Systems for Collaborative Text Editing. ECSCW CEW (2003)
5. Hodel, T.B., Gall, H., Dittrich, K. R.: Dynamic Collaborative Business Process within Documents. ACM SIGDOC (2004) 97–103
6. Hodel, T.B., Hacmac, R., H., Dittrich, K. R.: Using Creation Time Meta Data for Document Management. CAiSE (2005)