

# Docear: An Academic Literature Suite for Searching, Organizing and Creating Academic Literature

Joeran Beel  
Docear  
Magdeburg, Germany  
beel@docear.org

Bela Gipp  
Docear  
Berkeley, USA  
gipp@docear.org

Stefan Langer  
Docear  
Magdeburg, Germany  
langer@docear.org

Marcel Genzmehr  
Docear  
Magdeburg, Germany  
genzmehr@docear.org

## ABSTRACT

In this demonstration-paper we introduce *Docear*, an ‘academic literature suite’. *Docear* offers to scientists what an office suite like *Microsoft Office* offers to office workers. While an office suite bundles various applications for office workers (word processing, spreadsheets, presentation software, etc.), *Docear* bundles several applications for scientists: academic search engine, PDF reader, reference manager, word processor, mind mapping module, and recommender system. Besides *Docear*’s general concept, its special features are presented in this paper, namely a modular composition, free full-text access to literature, information management as mind map, automatic metadata extraction of PDFs and recommendations.

## Categories and Subject Descriptors

H.m [Information Systems]: Miscellaneous

## General Terms

Management, Documentation

## Keywords

Paper management, document management, literature management, pdf management, software suite

## 1. INTRODUCTION

Literature management is an important task for most researchers. It consists of searching for relevant literature (via keyword-based search or recommender systems), organizing the literature (reading papers, annotating them, etc.) and eventually creating own literature (drafting, writing, and citing) (see also Figure 1). Many software tools in the market try to facilitate the literature management process. For instance, digital libraries such as *ACM Digital Library* help finding relevant literature, tools like *JabRef* and *Endnote* help managing references, and PDF readers help reading and annotating documents.

However, full-text of academic literature is often costly or difficult to find, recommender for scientific articles such as TechLens [1] are not even close to the quality of music and movie recommender such as *Last.fm* and *Netflix*, and researchers having read and annotated hundreds of papers will easily lose track of what was written in which paper.

In this paper we introduce *Docear*<sup>1</sup>, the successor of *SciPlore MindMapping* [2]. *Docear* is what we call an ‘academic literature suite’, comparable to an office suite such as

*Microsoft Office*, but for researchers. While an office suite bundles various applications for office workers (word processing, spreadsheets, presentation software, etc.), *Docear* bundles several applications for scientists (see also Figure 2):

- A digital library containing some millions of research articles in full-text and their metadata (title, authors, journal, publishing year, etc.).
- A research module consisting of keyword search and a recommender system for the articles in the library.
- A PDF viewer for reading and annotating electronic literature (i.e. creating bookmarks and comments and highlighting text).
- A mind mapping module for drafting new literature and managing all information including files, document drafts, references and annotations.
- A word processing module for creating new literature.
- A reference manager for creating reference lists and bibliographies.
- Filters and converters and a RESTful Web Service for exchanging data with 3<sup>rd</sup> party applications.

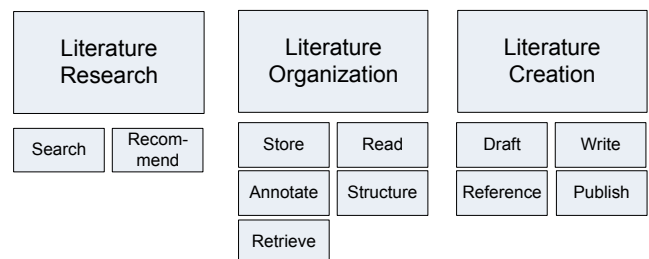
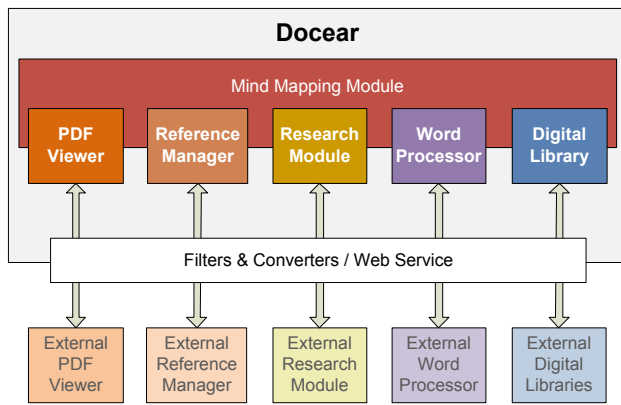


Figure 1: Literature Management Process

To reduce development efforts, some components are based on existing open source solutions. For instance, the mind mapping module is based on *Freeplane*, the successor of the popular *FreeMind* and the reference management is based on *JabRef*.

The first public Beta of *Docear*’s desktop version for *Windows*, *Linux* and *Mac OS* will be presented at the *ACM/IEEE Joint Conference on Digital Libraries 2011 (JCDL)* in Ottawa, Canada. However, as of now, not all components are completely developed yet (this paper just outlines *Docear*’s basic concept and ideas). Also, in the long run, a web application and smartphone app for *Android* and *iOS* are planned in addition to the Desktop version. All versions will be free and the desktop version is published as open source under the *GNU GPL*.

<sup>1</sup> <http://docear.org>



**Figure 2: The components of Docear**

In the following, the unique features of Docear are introduced.

## 2. A HOLISTIC CONCEPT

By integrating several software tools into a software suite, data exchange between the different tools is facilitated respectively made possible. All kind of items and information (documents, annotations, references, ideas, etc.) is available wherever the researcher might need it – in the reference manager, the PDF reader, while creating a paper draft and so on. For instance, when a user opens a PDF, the PDFs metadata (author, year, journal, etc.) is displayed<sup>2</sup>. Also, the user may drag & drop a PDF to a draft of a new article and the bibliographic data is inserted automatically as a citation<sup>2</sup>.

## 3. MODULAR COMPOSITION AND USE OF STANDARD FORMATS

Each of Docear's modules is exchangeable, except the mind mapping module. That means users not liking Docear's PDF viewer, reference manager or search engine may use another one instead or in addition to Docear's solution. Also, all data is stored in standard formats, e.g. BiTeX for references, PDF (ISO 32000) for annotations in documents, and as Open Document Standard (.odt) for text documents. Additional converters help importing and exporting data *from* respectively *to* other applications.

## 4. FREE FULL-TEXT ACCESS

Docear searches the Web for academic articles<sup>3</sup> (similar to *Google Scholar* and *CiteSeerX*). Currently, around 2 million articles including full-text are in Docear's database. Additionally to the standard search functionality, Docear automatically searches the database for literature when an article is mentioned in a document read by the user. For instance, Docear links entries in the reference list of a PDF with their full-text (compare Figure 3 and Figure 4)<sup>2</sup>.

## 5. INFORMATION STRUCTURING AS MIND MAP

Docear utilizes the power of mind maps for structuring information. Mind maps are well suited to structure document

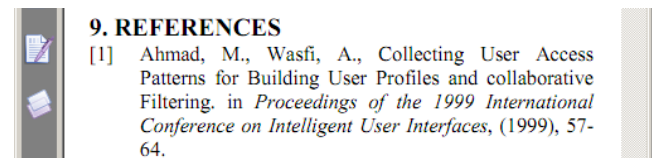
collections, the content of documents respectively annotations, and references. They may also be used to draft documents because the structure of a mind maps is similar to an outline. Docear provides a superior solution for structuring information in contrast to other solutions, using simple lists or social tags (which may be used in Docear in future versions additionally).

## 6. METADATA EXTRACTION

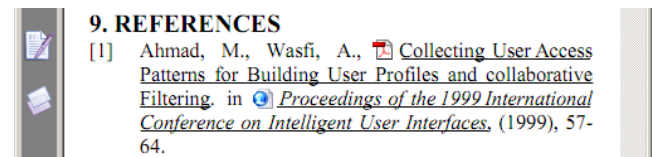
Docear extracts metadata such as title and author from PDF files. Additional metadata such as the year and journal is retrieved from Docear's bibliographic database<sup>3</sup>. With the extracted metadata users can structure their document collection and automatically insert references into their written articles.

## 7. RECOMMENDATIONS

Docear offers recommendations for scholarly literature and in future versions for conferences and journals the user could publish in, authors working on similar projects as the user, and for research grants the user could apply to. Potentially, these recommendations are of high relevance because Docear should be able to determine the interests of the users very well: Due to Docear's complete software suite, Docear knows what users are searching for, reading, which passages in a document interest them most, and what a user is currently working on.



**Figure 3: Original entry in a reference list**



**Figure 4: Modified entry by Docear linking to the full-text**

## 8. REFERENCES

- [1] R. Torres, S.M. McNee, M. Abel, J.A. Konstan, and J. Riedl. Enhancing digital libraries with TechLens. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 228–236. ACM New York, 2004.
- [2] Joeran Beel, Bela Gipp, and Christoph Mueller. 'SciPlore MindMapping' - A Tool for Creating Mind Maps Combined with PDF and Reference Management. *D-Lib Magazine*, 15 (11), November 2009, Brief Article.
- [3] Joeran Beel, Bela Gipp, Stefan Langer, Marcel Genzmehr, Erik Wilde, Andreas Nürnberger, and Jim Pitman. Introducing Mr. DLib, a Machine-readable Digital Library. In *Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL '11)*, 2011.

<sup>2</sup> Not yet implemented

<sup>3</sup> In cooperation with *Mr. DLib* [3]