

## The Project Setting

Docear ([www.docear.org](http://www.docear.org)) is what we call an “academic literature suite”. It integrates everything you need to search, organize and create academic literature in a single application: a digital library, reference manager, PDF and file manager, note taking and mind mapping. And the best: Docear works seamlessly with many existing tools like Mendeley, Microsoft Word, and Foxit Reader.

Docear is a part of the *Data and Knowledge Engineering Group* (DKE) from Prof. Dr. Andreas Nürnberger at the computer science department of OvGU. It’s free and open source, funded by the German Federal Ministry of Technology and developed by scientists from around the world, among others from OvGU, and the University of California, Berkeley.

## Your Research Project

We want to integrate a function into Docear that extracts header information (title, author, journal, ...) from PDF files, to automatically create new references. The reference management tool “Mendeley” has already a descent solution and also we have some basic functionality to do so. However, it could be better. Your research question to answer will be “How can header information be extracted from PDF files?”.

To answer the question, you will do literature research, evaluate existing solutions, create ideas for improvements and finally implement and evaluate your ideas. Of course, you don’t have to do all of this alone – you will be closely integrated into the Docear team and you can use the work we have already done in this field.

Finally, your work will be integrated into Docear and used by thousands of researchers around the world. If your work is outstanding, we will write a research paper with you about your results.

## Requirements

You should have a profound knowledge of the programming language JAVA. Knowledge in statistics, machine learning, other programming languages (especially C++ or Python) and/or MySQL, Hibernate, Jersey, REST Web Services is beneficial, but not required. Of course, we would appreciate if you spoke German but it would be no problem, if you don’t. We would prefer, if you apply for a long internship (12 weeks) but you can also apply for a shorter internship. If you are interested in combining your internship with writing a Bachelor thesis, please let us know in advance (this would be highly welcome).

## The University and Around

The Otto-von-Guericke University is a rather young university located in the center of Magdeburg. Especially the computer science department is well equipped and in a newly renovated building. For more information about the university read here: [http://en.wikipedia.org/wiki/Otto-von-Guericke-Universit%C3%A4t\\_Magdeburg](http://en.wikipedia.org/wiki/Otto-von-Guericke-Universit%C3%A4t_Magdeburg)

Magdeburg has a nice cultural life, beautiful parks, and a big shopping mall. Housing is rather cheap. You may get a room in a shared flat for as little as 150€ a month (however, slightly over 200€ is more realistic). Read more about Magdeburg on Wikitravel and keep in mind that Berlin (German’s vibrant capital) is just an hour away: <http://wikitravel.org/en/Magdeburg>

## Contact

Please do not hesitate to contact us if you have any questions: [info@docear.org](mailto:info@docear.org)



Docear is funded by

