Call for Bachelor Thesis

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Analyse the Impact of Text Representation Methods for Iconclass Label Recommendation

ICONCLASS[1] is the de facto global standard for the subject classification of cultural heritage content. It consists of alphanumeric "notations" which document the subjects in images in a language independent way. Each one of these notations has a set of textual labels to describe iconographic subjects in natural language.

In your thesis, you will have the chance to evaluate different representation methods for text, such as word embeddings[2], in the task of retrieving correct ICONCLASS codes from textual queries.

You will gain valuable experience in NLP and Information Retrieval techniques as well as provide fruitful insights on how to improve the current ICONCLASS text search system.

This thesis will be supervised by **Prof. Dr. Harald Sack, Information Service Engineering at Institute AIFB, KIT, in collaboration with FIZ Karlsruhe**.

[1] https://iconclass.org/
[2] https://en.wikipedia.org/wiki/Word_embedding

Which prerequisites should you have?
- Good programming skills in Python
- Knowledge of machine learning and NLP libraries (e.g., scikit-learn, NLTK)

It also helps if you have an interest or affinity with Art History or Cultural Heritage Contents

Contact person:

Etienne Posthumus
Etienne.Posthumus@partners.fiz-karlsruhe.de

Prof. Dr. Harald Sack
harald.sack@kit.edu

7 · Bible

71 · Old Testament

71B · Genesis from the descendants of Cain and Seth to Abraham

71B4 · story of the Tower of Babel (Genesis 11:1-9)

71B42 the Tower of Babel

aifb

Institute of Applied Informatics and Formal Description Methods
http://www.aifb.kit.edu/

KIT
Karlsruhe Institute of Technology