

EXCITE – A toolchain to extract, match and publish open literature references

Azam Hosseini, Behnam Ghavimi, Zeyd Boukhers and Philipp Mayr

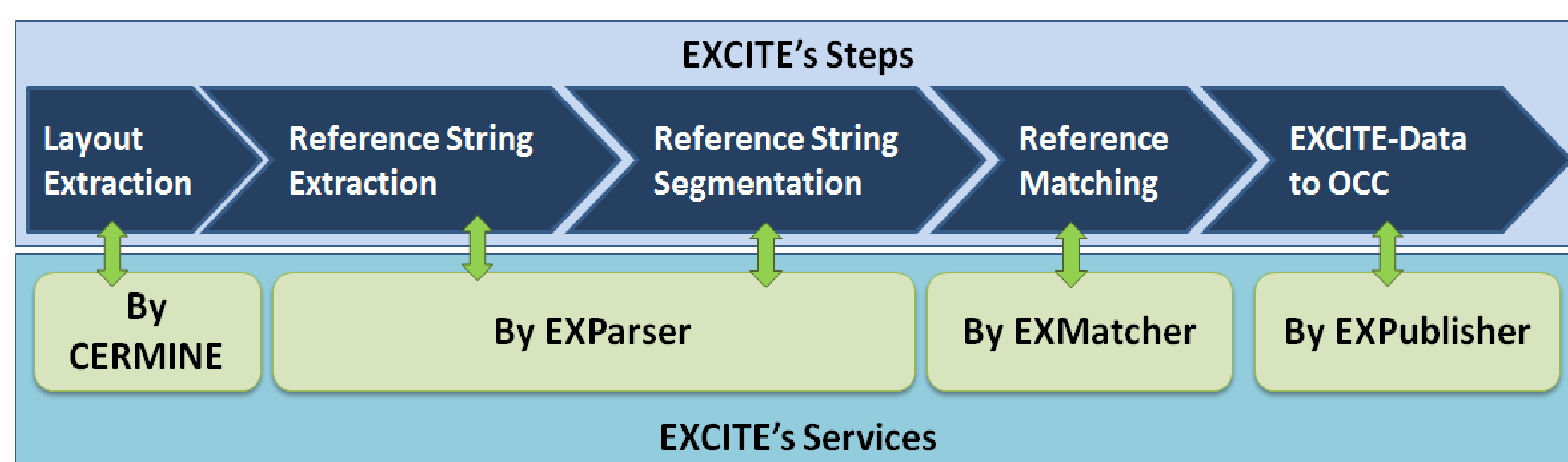
Background

- **Open access** to references is still **insufficient**
- Shortage of citation data for the international and **German social sciences**

Objectives

- Presenting a **generic and open toolchain** to extract, segment, match and publish literature references
- Making more citation data available to researchers
- **Opening up citation data** (esp. in the Social Sciences)

EXCITE toolchain



1. Text extraction by CERMINE
2. Reference Extraction by EXparser
3. References Matching by EXmatcher
4. Publication of references by EXpublisher

Demo system

EXCITE

Upload file (Choose File, Upload File)

Follow up code (Enter code to load data. 190121-BMRYN-EX-28254)

Check the result (Display Refs)

Extracted references in different formats

Reference String	Reference Segment(XML)	Reference BibTeX	Reference Matching
Huinink, J./Wagner, G. (1998): Individualisierung und Pluralisierung von Lebensformen. In: Friedrichs, J. (Hrsg.): Die Individualisierungsthese. Opladen	literaturportal27650e27f58e6314ab0448f07c7f1408	High	gesis-bib-83658
		High	Crossref doi: 10.1007/978-3-663-09724-2_7

Example of matched reference

EXparser

- Extracts references using a **probabilistic** approach
- Reference strings extraction and segmentation in one step to reduce error rate
- **Line based** algorithm

EXmatcher

- Uses **reference strings** and **segments** as input
- Uses **probabilities** of segments (generated by EXparser)
- Uses **Solr** for indexing and **SVM** for classification
- Matches against Sowiport, GESIS search, and **Crossref**

Technologies used

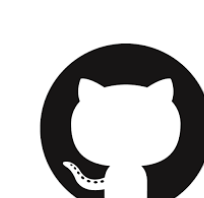


EXPublisher

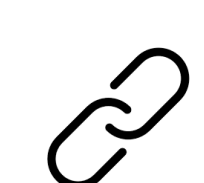
- **Enriches** data with **match information**
- Converts data into **OCC ontology**
- **First version** of data in **OCC ingested**
- All data in **OCC** with **0110** prefix. E.g.: <http://opencitations.net/corpus/be/01101.html>



Demo:
<http://excite.west.uni-koblenz.de/excite>



Git Repository: <https://github.com/exciteproject>



Website: <http://excite.west.uni-koblenz.de/website/>

**Begrenzung für Inhalt
bitte nicht überschreiten!**