



# Research Objective

## Base

### RT1: Corpus

- ① conversion and linking pipeline
- ② extensive corpus analysis

[**Scientometrics'20**]

## Research Gap

### RT2: Citations

- ① improved corpus pipeline
- ② blocking method for reference linking

[**ULITE'22**]  
[**JCDL'23**]

### RT3: Non-English

- ① extraction method for cross-lingual cit
- ② extensive citation analysis

[**ICADL'20**]  
[**IJDL'22**]

### RT4: Artifacts

- ① novel IE task and data set
- ② IE model development and eval.

[**ECIR'24**]



Data Quality

