

Using RecordLinkage and Parsing Methods to Create a Structured Dataset on Corporate Bonds

Project description

- Novel datasets often lack common firm identifiers used by commercial databases.
- A well-established approach is to use firm names and other firm characteristics and to apply linking methods exploiting machine learning and stochastic approaches.
- The goal of the IDP is twofold:
 - i) Link two datasets on Corporate Bonds for U.S Companies using the python package “RecordLinkage”, to generate an optimal crosswalk between the different datasets.
 - ii) Extract corporate filings for the matched company years and parse bond-specific information.

What we are looking for

- Advanced python skills
- Ideally experience with the RecordLinkage package
- Interest to learn something about finance, in particular about corporate datasets
- 1-2 persons

What we offer

- Knowledge in (corporate) finance
- Kick-off session including introduction to relevant finance and/or business topics
- Experience with IDPs
- Open dialogue and support
- Access to prime capital markets database (Orbis, Compustat)

Interested?

Please send an e-mail with CV, academic transcript and your preference for this project to lisa.knauer@tum.de.

Questions?

In case of any (e.g. topic related) questions, please contact Lisa Knauer (lisa.knauer@tum or call +49 89 289 25196).