Linking company data using RecordLinkage

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 to link two datasets on German companies using the python package “RecordLinkage”.
- In particular, you will link a large-scale dataset on firms’ job postings to a ready-made standard dataset on firm finances using this package.
- The outcome of the project is an optimal crosswalk between the two datasets.

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)

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 25482).