Interdisciplinary Project: Update of patent database PATSTAT

The Professorship for Economics of Innovation is looking for applications for an interdisciplinary project (IDP).

**Update of the patent database PATSTAT and the corresponding legal status database INPADOC, implementation of additional data fields, and upload to database**

**Evaluator:** Prof. Dr. Hanna Hottenrott

**Advisor:** M.Sc. Linda Bandelow (linda.bandelow@tum.de)

**Planned starting date:** as soon as possible

**Duration:** (6 month part-time or 3 month full-time)

**Required skills:** SQL, MySQL, data transfer to database

**Detailed background:**

Intellectual property rights are one of the most critical enablers of many fields of innovation and research which are key to our knowledge-based economy. In turn, one of the strongest forms of IP protection is through the application for and grant of patents. Due to the inherent design of the patent system, a vast amount of information is publicly available and therefore a great source of information for various research projects.

At the Economics of Innovation chair, we are using this patent data in various ways – be it to understand the proliferation of design rights, the invalidation of patents through law-suits, the necessity of patents to establish a market for technology or the application of ICT patents around the globe.

In order to support our research, we need to replace an old version of the linked patent databases PATSTAT and INPADOC (Oct 2015) with the latest available version from 2019. Subsequently, several additional queries will need to be performed in order to create customized data fields that will help facilitate our research.

**Work packages:**

1. **Preparation**
   a. **Familiarize yourself with the topic of patents:** Get up to speed on the working mechanisms of the patenting system and understand the differences between the approaches by different countries
   b. **Introduction to the databases:** Understand the data structure and relations between datasets within PATSTAT and INPADOC using a descriptive Entity-Relationship-Model

2. **Update of the PATSTAT and INPADOC database:** Aggregate the parts from the original data provision to a single SQL file and verify functionality.
3. Extension of database

The database needs to be expanded by the following variables (this is a preliminary list; exact requirements will need to be discussed with students):

a. Time of granting
b. Time of invalidation/revocation of patent right
c. Calculation of Grant Lags
d. Number of “children”
e. Number of “parents”
f. Calculation of “family size”
g. Forward and backward citations
h. IPC classification

Some of these variables have been implemented in previous, well-documented IDPs whereas some are new and will require you to develop new queries.

4. Upload to the database

The 2019 version of Patstat needs to be uploaded to the newly established database server.

5. Application of the above extensions for bespoke analysis

In the context of our various research projects, we will work in close cooperation with the IDP student to apply some of the extensions for customized analysis in order to both validate the correct programming but also to show to the student how his/her fundamental work feeds into our actual research. The scope of this part is also subject to discussion with the student and we would gladly take their interests into account in scoping this section.

Overview of project