Challenges and Solutions in the construction of Chinese Patent Database for innovation studies

Deyun YIN

Economics and Statistics Division, World Intellectual Property Organization (WIPO), Geneva, Switzerland

Chinese patent database has become the second largest patent database around the world. However, it is still underutilized as many key information, e.g., unique identifiers inventors, standardized applicant's names and citation, are unavailable yet. Facing these challenges, I wish introduce my work on the development of China's patent database (for short, CNIPA data). It includes:

1) Inventor name disambiguation of Chinese inventors' names: a brief introduction of global disambiguation, status, major algorithms, accuracies, and pitfalls in using existing database of disambiguated inventors, like USPTO's PatentsView, EPO's PATSTAT, JPO, CNIPA, as well as our algorithms and results.
2) Applicant's harmonization: challenges and solutions.
3) Geocoding of CNIPA and JPO patent database and integration with PATSTAT.
4) Extraction of citation data from raw text and construction of complex indicators for patent quality based on machine learning.

The estimated time for the presentation would be 10-30 minutes.