Research and innovation screening tool

When proposing new research or business ideas, one must cover substantial publication or patent material in order to ensure it is indeed new. Often, researchers look mainly at publications whereas industry tend to look at patents, though both sources contain such information, and increasingly so as industries tend to work in a more research based manner.

This project aims to look at how we can combine publication and patent information in order to answer questions like which authors/players are dominant for a given subject? What are the hubs (networks) for a subject? How are the publication and patent data interlinked (or not) for a given subject? Which university-company collaborations are visible through the patents and publications data?

A subset of data used for this study could e.g. be retrieved using scholar.py (on Github) for retrieval of Google Scholar data and patent data could e.g. be retrieved using the PatentsView API. Other data sources are also available and can be considered.

We will specify particular examples of interest, like e.g. 3d printers, additive manufacturing, or digital twins.

Text mining, bag of words, recurrent neural nets, transfer learning, network analysis or machine learning methods can be used to make visualizations and analyses to answer key questions and combine the data sources.

Prerequisites: The student should be familiar with machine learning and text mining methods/deep learning, and preferably also network analysis.

Supervisor: Line Clemmensen, lkhc@dtu.dk

The project is joint work with GE, US.

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads
Building 324
2800 Kgs. Lyngby
Denmark
Tel +45 45 25 30 31
www.compute.dtu.dk