Measuring the novelty and impact of inventions based on patent text

Dominique Guellec, Observatoire des Sciences et Techniques (OST), Paris

The OST is initiating a project that aims at measuring the novelty and impact of inventions by analysing the text and metadata of patent documents, using advanced NLP methods. The project will notably:

1. Propose a conceptual framework for novelty and impact in the context of patent data, elaborating on the notion of similarity and leading to measurement.
2. Identify and put together the relevant databases, including text data (patent documents, at least the title, abstract and claims) and metadata (citations, technical classes etc.); the current choice is to use EPO data, possibly extended to family data (after testing).
3. Test and select appropriate NLP methods among the various candidates already implemented in the literature (e.g. word embedding, LSTM, GRU, transformer etc.).

The presentation will report on the main options taken for the project regarding these three issues.