

From patents to technologies – A new level of analysis?

In innovation research, patents are used as an indicator for the output of application-oriented research and development of a country that allows assessments of its current and future technological competitiveness.

However, technologies generally are not only protected by a single patent but by a whole series or clusters of related patents, including one or more core patents. In addition to these core patents, companies often construct "patent thickets" (Shapiro 2001) in order to increase the protective effect of the patents or gain benefits from blocking competitors to file patents in the same or adjacent areas (Blind et al. 2006, Neuhäusler 2012). By grouping patents which are related to the same technology, the impact of the individual patents is reduced and analyses can directly be performed on a "technology level". This so to say offers the possibility to perform analyses on the level of technologies instead of patents. This also allows interesting insights about the average number of patents that protect a technology or to identify core patents within a patent cluster.

The objective of this feasibility study is to create an algorithm that aggregates similar patents representing a specific technology. This will be achieved via the information that is included in the patent specification or that emerges during the registration process. The aggregation of patents can be performed on company level or on a "global" level. The company specific view gives insights of the number of technologies a firm is active in. Leaving out this limitation provides us with a broader picture.

We propose a multi-step procedure as a method for the identification of technologies on the level of patents. The first step consists of grouping only those patents that fulfill some pre-defined essential basic conditions. The next step involves the application of more fine-grained selection criteria for the patents within the same technology.

More precisely speaking, we will analyze the patents which were registered during the last ten years by different patent applicants. In order to be chosen as candidates for the same technology group, the patents have to satisfy the following conditions. First, only patents of the same patent applicant may be grouped together. A second restriction refers to the technology field the patent was classified in. Patents may only be classified as protecting the same technology if they belong to the same of five technology fields which are defined based on the list of 35 WIPO classes (Schmoch 2008).

If two patents meet these basic requirements, more fine-grained selection criteria are applied in order to evaluate if both patents protect inventions related to the same technology. These criteria are combined in a normalized similarity measure. Thus, the similarity between two patents p_1 and p_2 is defined as

$$sim(p_1, p_2) = \frac{\sum_i w_i f_i(p_{1i}, p_{2i})}{\sum_i w_i}.$$

The function f_i computes the similarity between the two values p_{1i} and p_{2i} of a criterion i for the patents p_1 and p_2 . The weight w_i indicates how much impact the criterion i attributes to the similarity computation. For each criterion, an individual similarity measure as well as an individual weight is used. Especially, the value range may be quite distinct for the different criteria, thus these similarity values have to be mapped to normalized values.

Possible selection criteria are:

- Are patents technologically related by definition? This is determined through “technical relations” within the patent system. During the registration process of new patents, the patent applicant is provided the option to indicate relations between patents.
- Is the patent a continuation of an already existing patent?
- Are the patents associated with the same IPC class (four-digit code)? Besides being in the same technology field, this allows us to narrow down the relatedness between patents regarding a certain technology.
- How similar are the title and abstract of two patents? The similarity is determined by means of cosine similarity between two term vectors representing the title and abstract in the vector space model (Manning et al. 2008).
- Is patent p_1 cited by patent p_2 ? Is patent p_3 related to p_1 because it is also cited by p_2 ?
- How many previous patents are cited by each of the two patents and how many of those are identical (co-citation analysis)?
- Is the inventor of the patent the same for both patents?

Combining this criteria catalog, the similarity measure described above will be designed and calibrated on a training data set. This training data set consists of a collection of patents for certain technologies that are known in advance (e.g. MP3 technology). Especially, the weights of the different criteria have to be determined by means of known pairs of similar patents. Next, a clustering algorithm based on the similarity measure will be applied in order to identify different technologies. The results of the method will be evaluated on a test data set and analyzed and visualized using descriptive statistics.

References

Blind, K. & Edler, J. & Frietsch, R. & Schmoch, U. (2006): Motives to patent: Empirical evidence from Germany, *Research Policy*, 35, 655-672.

Manning, C. D. & Raghavan, P. & Schütze, H. (2008). Introduction to Information Retrieval, *Cambridge University Press*.

Neuhäusler, P. (2012): The use of patents and informal appropriation mechanisms - Differences between sectors and among companies, *Technovation*, 32, 681-693.

Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons. *Final Report to the World Intellectual Property Office (WIPO)*, Karlsruhe: Fraunhofer ISI.

Shapiro, C. (2001): Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting, Available at SSRN: <http://ssrn.com/abstract=273550> or <http://dx.doi.org/10.2139/ssrn.273550>