Global commercialization trends of microbial products and processes

Charles Parker\(^1\), Grace Rodriguez\(^1\), David P. Labe\(^2\) and George M. Garrity\(^1,3\)
\(^1\)NamesforLife, LLC, East Lansing, MI, \(^2\)USDA Agricultural Research Service, Peoria, IL, \(^3\)Michigan State University, East Lansing, MI

**Project Goals**

A fundamental premise in industrial microbiology and biotechnology is that microorganisms are a source of commercially useful products and processes. While there is always an element of serendipity in the discovery process, success is often defined by access to appropriate resources, including collections of well-documented strains and historical knowledge about their metabolic and genetic potential. Within communities, organizations still maintain private strain collections many of which have abandoned this strategy due to cost. There are, however, public resources that can serve as a substitute, chief of which are the international patent repositories. These collections hold the strains that are key to many patented microbiological inventions. Once a patent expires, these strains can be released without restriction. These strains are distinct from non-patent strains in public collections because of the amount of information that is publically available, although that information is difficult to obtain or use. Our objective is to make the connections between strains and the patent literature easy to navigate and to make the information about patented microbial products and processes more readily accessible.

We recently completed a pilot project through the USDA ARS Patent Collection (NRRL, Peoria, Ill.). Using proprietary text mining methods, we were able to identify global commercialization trends in 162 technology classes over a 70 year time span by following more than 4,000 district NRRL strains referenced by over 16,000 US and foreign patents drawn from a corpus of 180 M patent documents.

**Background**

Microorganisms and their metabolites and products are protectable intellectual property in many countries. Securing a patent on a microbial invention requires filing an application in each jurisdiction where protection is sought in which the invention(s) must establish the novelty, utility and non-obviousness of the claimed invention. Their detailed description must fully disclose the nature and workings of the invention so that it enables others who are “skilled in the art” to replicate the invention. Failure of enablement is grounds for revocation of the rights granted under a patent.

Microbiological inventions are distinct from chemical, mechanical and electrical inventions in that enabling of the invention typically requires the live organism to fulfill the legal requirements of full and complete disclosure. Since 1949, the USPTO has required that viable samples of the subject microorganisms be deposited in a public culture collection in conjunction with the filing of the patent application. Similar rules went into effect in most other industrialized nations soon after, leading to the establishment of a number of national repositories that hold collections of those preserved microorganisms that would become available when an issued patent was presented. In 1980, the process of depositing microorganisms in association with patent filings was simplified with the ratification of the Budapest Treaty, which allowed a single deposit in one of 12 international repositories. This resulted in the placement of the patents into 158 clusters, ranging in size from 2 - 865 documents.

**Meta-analysis**

The extracted metadata was loaded into the R statistical computing environment to explore some of the general trends in commercialization of microbial products and processes associated with the deposits in the USDA ARS Patent Collection. Semiotic fingerprinting analysis was also performed in which the subset of patents for which both strain identity and IPC codes were available (n=10,784). This resulted in the placement of the patents into 158 clusters, ranging in size from 2 - 865 documents.

**Future Work**

With this retrospective meta-analysis now completed, we are exploring the utility of automatically generating periodic updates to clients about emerging technology trends and the availability of strains following patent expiration or abandonment. We are also exploring the possibility of extending this approach to other international patent repositories as part of an alerting service. This meta-analysis also serves as a preliminary step in our terminology and ontology development efforts, extending this approach to other international patent repositories as part of an alerting service. This meta-analysis also serves as a preliminary step in our terminology and ontology development efforts. The metadata extracted from the subset of 10,784 patents that could be positively associated with NRRL patent strains and for which IPC classification codes were also available were subjected to a series of sequential cluster analyses. The results were then projected as heatmaps of the input similarity matrices that were reordered along both dimensions as a means of indexing the related patents. The similarity matrices are symmetric, with identity (patent match) all along the diagonal. Highly similar results are colored dark blue and should be found along the diagonal. Less similar results are less intense. Intermediate levels of similarity fall between the diagonal and off-diagonal. Top left, classification based on semiotic identity of patent strains; top right, classification based on IPC codes of patents; bottom left, classification based on the taxonomy identity; IPC code as semiotic fingerprint; bottom right, classification based on the taxonomy identity; IPC code as semiotic fingerprint. The top 10 most highly similar results are shown for each of the three analyses. Each of the three analyses performed associated a group of NRRL patent strains with a highly similar set of IPC codes as the most similar patent strain. Each group of IPC codes was placed in their respective technology boxes on the heatmaps. These heatmaps were then used as a preliminary step in our terminology and ontology development pipeline.

**Acknowledgments**

Funding for this project was provided through the DOE SBIR/STTR program (DE-SC0006191). Funding for the NamesforLife infrastructure was received from the DOE SBIR/STTR program (DE-FG02-07ER84488). The Michigan Small Business Technology Development Corporation, the Michigan Strategic Fund, the Michigan Economic Development Corporation, and the Michigan State University Crowdfunding initiative.