



Environmental Patents in Brazil: Proposed methodology for identification and compatibility with different databases

Francisco de Souza Marta*

IBGE, Rio de Janeiro, Brazil, francisco.marta@ibge.gov.br

Paulo Gonzaga Mibielli de Carvalho

IBGE, UNESA, Rio de Janeiro, Brazil, paulo.mibielli@ibge.gov.br

Denise Britz do Nascimento Silva

IBGE, ENCE, Rio de Janeiro, Brazil, denise.silva@ibge.gov.br

Frederico Cavadas Barcellos

IBGE, Rio de Janeiro, Brazil, fredcavadas@ibge.gov.br

Abstract

The concept applied to delimitation of new technologies related to environmental issues is constantly evolving, encompassing processes products or services to improve operational performance and increase efficiency and productivity while reducing costs, inputs, energy consumption and, with the primary purpose of generating lower environmental impact (WINTER, 2011). Patent environmental information obtained through different patent classification systems (EPO and WIPO) can present major technological limitations, since in some cases have content too generic or too comprehensive framework for the environmental concept. In some cases, depending on convenience, can even submit commercial or political interests. In this regard, an efficient tool to identify these new technologies related to environmental dimension may be the patent system by identifying environmental patents. In Brazil, the initiatives of companies in relation to mitigation of environmental problems caused by the productive activities is relatively recent. Among the possibilities to monitor these initiatives, we can highlight the Annual Industrial Survey - PIA, through variable Investment in Environmental Control, in the Survey of Technological Innovation - PINTEC, both from IBGE, and information about deposits of patent applications, National Institute of Industrial Property - INPI. However, the handling, processing and reconciling these different databases constituted a major challenge in view of the need for standardization and formatting of data, identification and elimination of duplicate records or missing data, harmonization of names of depositors, among others. Thus, the aim of this paper is to present a methodology for identifying environmental patents filed with the INPI, between 1998 and 2007, based on patents obtained by MIMOSA system developed by the European Patent Office, Japan Patent Office and the Office of Patent and Trademark United States. The results achieved with the application of the proposed method allowed the information obtained through patents could be compatible with the database of PIA and PINTEC.

Keywords: environmental patent; patents applications database; environmental technologies; database compatibility.

1. Introduction

The concept applied to delimit new technologies related to environmental issues¹ is in constant evolution, covering processes, products and either services to improve working performance and increase efficiency and productivity while reducing costs, inputs, and energy consumption, with the primary purpose of generating no or minor environmental impact (WINTER, 2011). In this context, an

¹ Also known as green technology and that according to definition in Agenda 21: "These are technologies that protect the environment, are less polluting, use all resources in a more sustainable way, recycle more waste and products, and handle residual wastes in a more palatable way."

efficient tool to find these new technologies related to environmental dimension may be the patent system by identifying environmental patents.

It is understood that the environmental patents content represents an important source of information about new environmentally sound technologies (EST), such as, those that deal with the combat against the causes of climate change, solve or mitigate environmental impacts or keep the environment and the natural resources; discuss about technologies related to environmental quality, energy conservation, development of renewable energy resources or reducing emissions of greenhouse gases; help recycling, improve resource management, etc.. (SANTOS, 2011). For this reason it is important the process of identifying patents for environmental purposes.

Currently there are two systems for identification of patents related to environmentally sound technologies (EST)². 1) The first was created by the World Intellectual Property Organization (WIPO) and is based on the International Patent Classification (IPC). This system allows patent identification related to (i) alternative energy production, (ii) transport, (iii) energy conservation, (iv) waste management, (v) agriculture, (vi) nuclear energy and (vii) administrative issues (WINTER, 2011). 2) The second is the European Classification System of Patent (ECLA), developed by the European Patent Office (EPO) that has specific sub classifications to better classify green technologies, such as: (i) clean energy technologies, (ii) combustion technologies with mitigation potential, (iii) specific technologies for climate change mitigation, (iv) technologies with direct or indirect potential contribution to mitigation of emissions, (v) reducing emissions and increasing fuel efficiency in transport; (vi) energy efficiency in buildings and lighting systems (PORT and KANNEBLEY JR., 2012).

However, it is worth noting that the environmental patent information obtained through different patent classification systems may have big limitations, because, sometimes, has generic content for the environmental concept. In some cases, depending on convenience, can even express commercial or political interests.

In order to overcome such limitations in data recovery of national patent applications, we chose to build a database with information on patents related to environmental issues and match these records with data from Annual Industrial Survey of Enterprise 2007 (PIA) and Survey of Technological Innovation 2005 (PINTEC). Thus, the aim of this article is to present a method for identification and treat this information based on the records of patent applications filed at the Brazilian Institute of Industrial Property (INPI) between 1998 and 2007, using the criteria for choice and classification the information contained in the titles and abstracts of patents applications as environmental patent, to enable compatibility with data from PIA and PINTEC surveys.

2. Method definition

To do the access and extraction of the database of national patent applications, or simply designated here as patent database it was used the MIMOSA system³ (Mixed Mode Application Software) that uses the data server of the European Patent Office (EPO), Global Patent Index (GPI). The EPO Global Patent Index server assemble all records of deposits patent applications in the world available through the patent search service online (esp@cenet) of the European Patent Office, and combines them with the search tools offered by MIMOSA recovery system. With this kind of tool it is possible to get access to all records of patents filed anywhere in the world, including Brazil, at any time.

The criterion used in the choice and extraction of patent database is that the deposits of patent applications were made in Brazil in the past ten⁴ years, with between 01/01/1998 to 31/12/2007. With that it was retrieved 215,124 patents registered in Brazil in this period containing the following

² The concept of Environmentally Sound Technologies (Environmentally Sound Technologies), is quoted by the UN Framework Convention on Climate Change (UNFCCC) (PORT and KANNEBLEY, 2012).

³ Provided by the European Patent Office at: Available at: <http://mimosa.european-patent-office.org/storagearea/MIMOSA_V6.0.7.5.exe> Accessed: October 10, 2010. The MIMOSA software is based on Microsoft Windows platform, now in version 6.0.7.5, and is designed by the Trilateral European Patent Office, Japan Patent Office and the Patent and Trademark Office of the United States. The software interface is available in ten different languages..

⁴ Brunnermeier and Cohen (2003) also used last ten years patent records (1983-1992) in their work on the determinants of environmental innovation in the United States.

information: order number, priority number, type of patent, name (s) of depositor (s), name (s) of inventor (s), date of deposit, Country, inventor Country, Portuguese title, English title, Portuguese abstract, English abstract, International Patents Classification, patent granted (Y or N).

Records of deposits of patent applications are extracted in TXT format, with fields separated by a specific identifier. Once retrieved the records is made the importing / conversion of the patent database for a database⁵ manager.

The next steps are: i) handling and treatment of patents database, and ii) identification of information related to the environmental dimension of the patent to consider such requests as "environmental patents."

The handling and processing of the database consist of standardization and formatting of data, identification and elimination of duplicate records or missing data, and matching the names of depositors.

2.1 Handling and treatment of the patent database

With the aid of a manager database, rules are performed to optimize the handling of patent database with more than 215,000 records filed between 1998 and 2007 containing 12 distinct fields.

Since a large volume of information requires a high performance of the hardware for data processing, which can affect negatively the increasing in response time, some steps have been taken to optimize the handling of basic patents. Initially the database was divided into two files. The first, called simplified database has only fields with basic information to find the record, composed by order number and the name of the depositor. The second file or complete database was set up with the same ID fields of the first record by adding other fields containing more detailed information. The bases integrity was kept throughout the relationship of field identification key, in this case through the application number. With this it was possible to work with a "light" database for the last process, that is, the compatibility of the environmental patent database with PIA-Enterprise 2007 and PINTEC 2005.

The filing of a patent can be done on behalf of one or more people or companies. Therefore, the field NAME OF DEPOSITOR in the simplified database presents features of be owned by one or more companies by an identifier as a "/" (slash). Thus, it was necessary to find all records that contained more than one depositor, and build more than one field depositor for the patent registration that was related to the names on the list of depositors. At the end, each patent number would have one or more depositors, with cases of duplication in these two fields.

Other procedures that were taken at the treatment step for the adequacy of the data: locating and exclusion of missing or duplicate data, standardization and formatting dates and numeric fields, the spelling errors corrections in text fields and harmonization of names of depositors. At this stage records were eliminated resulting in 19,083 from 196,041 patents that does not duplicated.

2.2 Identification of patent applications related to environmental protection

Phase 2 consisted of identifying information related to the environmental dimension of patent registration. Thus, we identified some keywords⁶ in the title and summary of the application for patent registration, which could result in a significant response to consider such requests as "environmental patents".

The process of "identifying" those environmental patents occurs through an extra field construction at the database, called environmental patent identifier. This identifier construction was conducted in three distinct stages based on the following criteria:

1st stage - Automatic review: based essentially on criteria for identification of environmental patents proposed by Brunnermeier and Cohen (2003) that consists in counting as environmental patents records that involves the destruction or disposal of hazardous or toxic wastes, recycling or reused waste, acid rain prevention, solid waste treatment, alternative energy sources, air and water pollution

⁵ In this article we used the Microsoft Office Access 2007 system.

⁶ Patents involving the destruction or disposal of hazardous or toxic wastes, recycling or reuse of waste, prevention of acid rain, solid waste treatment, alternative energy sources, prevention of air pollution and water pollution, are counted as "environmental patents" (Brunnermeier and Cohen, 2003, p.282).

prevention. In this first stage it was filtered 62,699 non-duplicate records from a total of 196,041 records filed between 01/01/1998 and 31/12/2007;

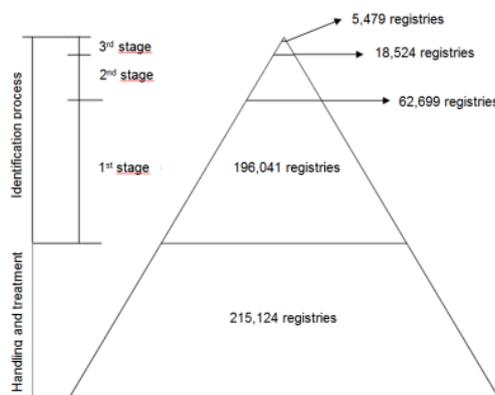
2nd stage – Semiautomatic review: it consists on the refinement of information previously identified in the 1st stage by using a list of incompatible⁷ terms in title or summary of the patent application, for the purpose of environmental improvement. Once the system identifies a record from this list, it is requested the researcher intervention to confirm or not the selected record and follow-up with the semiautomatic process. For example, trash⁸ recycling⁹, energy reduction for fluid¹⁰ device, etc. At this stage the remaining 18,524 records of 62,699 from the previous step;

3rd stage – Manual review: manual validation and analysis of the records identified in the two previous stages as patents related to environmental quality. This stage was most complex and slow, as was the analysis and manual validation of the records identified in the previous two steps as related environmental quality patents. Initially, the records were sorted alphabetically by titles to easily find the records started with the words ARTICLE, CASE, PRODUCTS, SYSTEM, METHOD, COMPOSITION, INSTALLATION, ARRANGEMENT, UNIT, SET, UNIT, DEVICE. Thus, it was possible to quickly qualify the patent content to decide whether to exclude them from the base. At the end of this stage 5479 records were identified and classified as "environmental patents."

In some cases more dubious, where there was a high content or complexity of a great uncertainty about the intended use of the patent (too generic) the criteria adopted for decision-making¹¹ or to seize such records were not particularly agreed on a case. Thus, in future studies, such rules can be reproduced to help with other results beyond those that have been achieved.

Figure 1 shows a scheme illustrating the results achieved in each stage of the analysis of the deposit database of patent applications until the last identification of the 5479 environment patent applications.

Figure 1: Results obtained in the stages of handling, processing and identification of environmental patents deposits on the databases of patent applications.



Source: MARTA, 2011.

At the end of the identifying process of environmental patents in the database of patent applications it was conducted pairing this database with data of PIA-2007 and PINTEC 2005 which identifies 560 CNPJs companies associated with 1997 records of environmental patents.

⁷ Arbitrated by author's own criteria of basic patents manipulation.

⁸ In this case, the word "trash" in the list of incompatible terms is detected by system and identifying this record as a possible patent without improvement of environmental purpose, even though title has the term "recycling". Therefore, the system stops and waits for an intervention by the researcher who can ratify or not the exclusion of this record on the database.

⁹ As the process of recycling materials, means the separation and processing of waste and scrap into secondary raw materials by compression, chemical treatments, physical, etc., Allowing further processing. The products obtained by recovering materials are reused by industry.

¹⁰ Incompatible term with the environmental aim.

¹¹ Via research sites of scientific information available through the internet (INPI, CAPES, GOOGLE Books, etc.).

2.3 Harmonization of patent database with micro data of PIA-Enterprise 2007 and PINTEC 2005 and its limitations

The compatibility between the patents database and micro-data of PIA 2007 and PINTEC 2005 consisted of the relationship of common identification key fields between these databases, allowing for each company, it was possible to any patent application filed on their behalf. However, as there was no common identification field in between the database of national patents achieved through the MIMOSA system, and the micro data of PIA and PINTEC, it was not possible to make a direct link between these databases. Thus, it was necessary to create a field that could relate the records of patents with companies.

Micro data of PIA-Enterprise 2007 and PINTEC 2005 have the COMPANY field characterized by the root number of CNPJ (National Register of Legal Entities), i.e., the first eight digits of the CNPJ and Company Name. Since the patent database has the field NAME DEPOSITOR, which can be owned by one or more person or entities, the only way to relate these two bases is through the association between the names of the depositor PJ and Company Name.

The main limitation to make compatible the information contained in the patent database achieved by MIMOSA software with the micro-data of PIA and PINTEC is the lack of a common key field to create the relationship between these two databases.

In order to overcome this serious limitation, the first step taken was the construction of a new field at the patents database to use it as a key relationship with the PIA and PINTEC. The new field chosen was the CNPJ number, identified by the root number of PIA and PINTEC enterprise, or the first eight digits.

Despite being the most proper solution, its implementation was extremely complex due to the need for at least one common field between these two databases so that they could create this new field CNPJ. For this we used the DEPOSITOR field on the database patents, and the NAME field, created into PIA-Enterprise and PINTEC.

The following describes the key steps for recovery CNPJ field on the patents database:

- ✓ Because patent applications can be filed on behalf of companies domiciled outside Brazil, there are many records of depositor on behalf of foreign companies, and although many of these companies also operating in Brazil, do not have the same designation in a foreign country, making difficult a direct comparison between the fields CORPORATE NAME from PIA or PINTEC with the DEPOSITOR NAME from the patent database. It was also observed that even companies who place national patent applications may have significant changes between the spellings of one record to another. Thus, a second problem is the difference in spellings, abbreviations / acronyms and misspellings, among others just to mention the most common cases. To minimize such problems, a new field was constructed analogous to the name of the depositor, called NEW_DEPOSITOR containing only numbers and words without any spelling signs (accents, hyphens, quotation marks, apostrophes, etc.), Spaces or punctuation characters and generic names were also abolished, such as INDUSTR*, COMERC*, LTD, LTD, LDA, SA, GmbH, CO, CIA, articles, prepositions, pronouns, linking words, etc.;
- ✓ Pairing databases of PIA and PINTEC and construction of a new field NEW_CORPORATE NAME compatible with the same rule of field NEW_DEPOSITOR from the patents database;
- ✓ Crossing and elaboration of proper reviews for compatibility between the textual fields of the patent database and new database PIA/PINTEC. For example, direct comparison, comparison of the N first and the M last characters of the sequence; comparing the first N terms of the sequence, etc.
- ✓ Planning and review of individual cases not associated (not found) between the two distinct bases for construction of possible criticism.

3. Final remarks

In Brazil, the statistics on industry initiatives in the environmental area are relatively recent and few. In IBGE, we have information on PIA-Enterprise and PINTEC, and the INPI, has a rich database of patent applications which, however, does not identify the environmental nature of patents.



The statistics raised by IBGE advances to create a good database of information to enable monitoring in industrial organizations by the implementation of initiatives that result in the reduction of emissions and use of resources. However, experience in the use of patent database and its consequent compatibility with the IBGE data shows the need of progress.

The existence of types of environmental patents created from the WIPO and ECLA studies, which are based just on the International Patent Classification does not solve the problem of systematization for identifying environment patents, since in some cases, have a too generic or too comprehensive framework for the environmental concept. Therefore, the developed method here was based on that proposed by Brunnermeier and Cohen in which patents are selected from the information contained in textual fields. It is expected therefore that the quality of information achieved through this process is better when compared to the technique using the information on the classification of the patent based on the IPC.

The challenge of match the environmental patents database of the INPI, with the micro-data of PIA Enterprise and PINTEC showed some difficulties that need to be solved.

We believe that the main contribution of this paper is to show how environmental patent information can be achieved with good quality and reliability and can be used for construction of environmental statistics from the data compatible with the PIA and PINTEC. Moreover, this article also gives visibility to the theme "environmental patents" that has been little studied in Brazil but that tends to gain focus in the discussions about "green economy" which was a theme of the RIO +20.

References

- BRUNNERMEIER, S. B., COHEN, M. A. Determinants of environmental innovation in US manufacturing industries. *Journal of Environmental Economics and Management* 45 (2003) 278–293
- MARTA, F.S Determinantes do Investimento em Controle Ambiental das empresas industriais Brasileiras em 2007. 231p. Dissertation (Masters in Population Studies and Social Research) ENCE, 2011; 220p. Available: http://www.ence.ibge.gov.br/c/document_library/get_file?uuid=ce946ace-3ca3-413f-a22f-c7bd1879b4e3&groupId=37690208. Accessed: October 5, 2012.
- NUNES, J. S; OLIVEIRA, L. G. Universidades Brasileiras - Utilização do Sistema de Patentes de 2000 a 2004. Instituto Nacional da Propriedade Industrial. DART/CEDIN/DIESPRO. Rio de Janeiro. 2007.
- PORTO, G.; KANNEBLEY JR., S. Subprojeto 3 - Rotas tecnológicas e sistemas de inovação produto 03 – Estrutura SASTec. Faculty of Economics, Administration and Accounting of Ribeirão Preto. University of São Paulo. FUNDACE – BNDES. Ribeirão Preto – SP. 2012
- SANTOS, D. A et al. Proteção Patentária Ante o Cenário das Mudanças Climáticas: Caso das patentes verdes. Instituto Nacional da Propriedade Industrial (INPI). PPT presentation. Available: http://lige.mgt.unm.edu/pdf/presentations/Santos_B.pdf. Accessed on: September 25, 2012. 2011.
- WINTER, E. INPI e patentes verdes. Academia da Propriedade Intelectual, Inovação e Desenvolvimento – ACAD. PPT presentation. Available: http://www.biocologia.ufscar.br/uploads/inpi_e_patentes_verdes.pdf. Accessed on: September 25, 2012.