Technological Forecasting & Social Change 81 (2014) 250–258

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Capturing waste recycling science

Gaizka Garechana [a,*], Rosa Rio-Belver [b], Ernesto Cilleruelo [c], Javier Gavilanes-Trapote [b]

[a] Technology Foresight Management (TFM) Group, Department of Industrial Engineering, University of the Basque Country UPV/EHU, calle Elcano 21, 48030 Bilbao, Spain
[b] Technology Foresight Management (TFM) Group, Department of Industrial Engineering, University of the Basque Country UPV/EHU, calle Nieves Cano 12, 01006 Vitoria, Spain
[c] Technology Foresight Management (TFM) Group, Department of Industrial Engineering, University of the Basque Country UPV/EHU, alameda Urquijo s/n, 48030 Bilbao, Spain

ARTICLE INFO

Article history:
Received 16 February 2012
Received in revised form 20 June 2012
Accepted 5 July 2012
Available online 11 August 2012

Keywords:
Tech mining
Search strategies
Waste recycling
Bibliometric analysis

ABSTRACT

Many institutions from the public and private sector are interested in the characterization of the research taking place in waste recycling (WR) science. Tech mining analysis can be applied to scientific databases with this purpose in mind, but difficulties do arise when designing the search strategy to effectively capture this multidisciplinary area. This paper introduces the process followed to build a query system that aims to solve this problem. This system has been applied to a selection of scientific databases, and the steps followed to download and clean the data are detailed. Initial results are explained, indicating the relevance of each database and quantifying the overlap among them. The main subjects behind the retrieved data have been identified, namely, chemistry, biology and environmental sciences. A precision test conducted by random sampling indicated that, with a confidence level of 95%, the proportion of WR articles is between 74.2 and 79.2% of the retrieved items, while recall is expected to be high, according to available classifications. These results are deemed to be satisfactory enough for basing forthcoming tech mining analyses on this query system.

1. Introduction

There are many institutions from the public and private sector interested in the characterization and forecasting of the research taking place in waste recycling (WR) science. This is a highly multidisciplinary field that comprises knowledge coming from various branches of science, including social sciences [1]. In addition to this, the waste management industry is formed by a wide range of activities including collection, transport, processing, recycling/disposal and the monitoring of waste materials [2]. This variety of activities and scientific research areas adds extra complexity to the decision making process in this field, enhancing the need for timely and accurate information about what is going on in it.

Tech mining analysis can play a key role in the assessment of R&D policies in WR, providing valuable information about the worldwide scientific research that is being conducted, the main actors in the field and many other innovation-related indicators. Tech mining can be defined as the application of text mining tools to science and technology information, informed by an understanding of technological innovation processes [3]. Scientific publication databases are a reliable global source of information that can be effectively exploited via text mining in order to extract information about research taking place in WR. However, the problem arises from the very multidisciplinary nature of this science, which invariably calls for the building of an ad-hoc search strategy to capture the nucleus of this science from the databases, and clearly define its subfields. The aim of this paper is to provide an overview of the method followed to build a set of queries oriented to capture the scientific production in WR, for the eleven year interval 2000–2010. The database choice, as well as the downloading and data cleaning processes are described, and some initial results are reported. This work is previous to the elaboration of other tech mining studies oriented to WR research activity, including social network studies and the building of a WR map of science.

* Corresponding author. Tel.: +34 686989172.
E-mail address: gaizka.garechana@ehu.es (G. Garechana).

0040-1625/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2012.07.005

2. Methodology

2.1. Precedents

The starting point of this paper is the work done by Garechana et al. [1] , tracking the cognitive changes experienced in WR science from the years 2005 to 2010, via tech mining tools. The criteria then followed to retrieve the items published in WR research were derived from the identification of the main underlying key cognitive areas in this field, represented by the Web of Science (WOS) Subject Categories (SC) .

A thorough study of the keywords found in the sample of Garechana et al. gave an overall impression of the major research topics in this field; however, some doubts remained as to what activities were to be included under the WR concept. These doubts led to the search of formal definitions, mainly on governmental and official public agency websites. An overall consensus was found, with some exceptions, about two facts of what WR can be considered as:

a) It does not only involve the remanufacturing or reprocessing of materials so that they can be used again, but also the collection, classification/separation of waste and the subsequent marketing of recycled products.
b) Recycling is not only about the reutilization of the spent materials in their original form. A discarded rubber tire might be reused as an industrial resource not in the form of recycled rubber, but in form of fuel for a cement kiln.

Some sources explicitly exclude the thermal combustion of wastes from the umbrella of WR [4], and obviously it would be as such in the case of mere incineration without energy recovery. However, the studies conducted by Garechana et al. (2012) and consulted experts point out that the Waste-to-Energy concept is an important area in waste management. Direct combustion methods are the least desirable option to recover energy from waste, and are subject to strict regulations [5] but other procedures aimed at obtaining fuel from waste are increasingly important. Taking these facts into account, the authors opted to choose the following inclusive definition of WR:

> "A method of recovering waste as resources which includes the collection, and often involving the treatment, of waste products for use as a replacement of all or part of the raw material in a manufacturing process." [6]

This definition is broad enough to include the recovery of many inputs like water or energy while considering the recycling process as a whole, including collection, characterization and classification of waste. Other available definitions are considerably broader but somehow ill defined: "Reusing materials and objects in original or changed forms rather than discarding them as waste" [6] while others put the energy recovery from waste under the definition of waste recovery, eluding the word "recycling" [7].

2.2. The query system

The work conducted by Porter et al. [8] to build a set of queries that define the boundaries of the science of nanotechnology has been a source of inspiration and methodology for this paper, as well as the process detailed in the work of Kostoff [9] to retrieve the literature corresponding to a particular area. The authors opted for a Boolean search term approach, given the fact that temporal and financial limitations of this study discouraged the alternative approach referred to as "bootstrapping" by Porter et al. [8], and considering on the other hand the advantages of modularity and flexibility derived from the Boolean approach.

2.2.1. Compilation of terms

An exploratory set of queries, based on the previously mentioned keyword study, was run on SCI and SSCI, and the results analyzed. The overview of WR activities and different types of waste sources given by Demirbas [2] was also a good starting point to define the subfields within WR. The items retrieved in this first attempt were examined by taking samples and reading the available fields, especially title and abstract, to determine if the results fitted in with the aforementioned WR definition. This process proved to be extremely effective to discover various synonyms and acronyms of WR related terms, as well as to identify some non-desired meanings of the initial keywords and "toxic terms". Important words in WR science were identified thanks to the reading of literally hundreds of titles and abstracts, and their suitability considered for including them in the queries. The correct interpretation of the retrieved items was considerably eased by the qualifications of the authors of this paper, the main author holding a university degree in chemical engineering and the second author a master's degree in environmental sciences. The analysis was further complemented by checking the glossary service of the European Environmental Agency [6] and US Environmental Protection Agency [10]. These two websites were intensively used as vocabulary source and qualified information point; both were selected due to their trustworthiness and the richness of their terminology services.

2.2.2. Adjusting retrieved items to WR

At this point, the researchers had collected a fairly lengthy set of keywords from various branches of WR, but when testing them via simple queries in various databases, the following problems arose:

– Some queries introduced significant noise in the sample, for example, retrieving publications about mere disposal or incineration of waste, did not deal with the aforementioned WR definition. A clear case can be found when retrieving publications about landfilling. Landfills produce various gases that can be used as an energy source [11], however, mere landfilling cannot be considered as WR, according to the chosen definition.

- Other important keywords have too broad a meaning. The term biomass is an example; this term refers to any biological matter, from a population of animal or vegetal organisms (fish biomass) to crops specifically grown for obtaining biofuel. It is necessary to modify the query to obtain publications oriented to extract energy from waste biomass resources.
- The same acronym can refer to very dissimilar things. Acronyms should be included in the queries after carefully evaluating the retrieved items.

The solution proposed by the authors was to develop some co-occurrence modules, to restrict the retrieval of the queries to the boundaries of WR. The choice of the proper module to co-occur with each keyword was made by trial and error, looking for at least roughly 70% of retrieved items directly related with WR. Modules are listed below:

- WASTE MODULE: (SCRAP* OR LITTER* OR GARBAGE* OR LEFTOVER* OR TRASH* OR RESIDUE* OR WAST* OR DEBRIS*).
- RECYCLE MODULE: (TREAT* OR RECYCL* OR RECOV* OR REUT* OR REUS*).
- ENERGY MODULE: (ENERG* OR FUEL*).

The keywords were usually searched in the "title" field, with the aim of minimizing noise in the sample, but good results were obtained when unmistakably WR-related terms were searched in "all fields". Co-occurrence with modules was applied in "all fields", when necessary (see Table 1). Some technical terms corresponding to specific chemical, metallurgical or biological treatments were detected in WR, but in many cases they had other applications apart from WR activities and introduced

Table 1
Set of queries, the syntax is that of WOS databases. Queries adapted for other databases are available upon request.

Query Number	Query
Exclusion queries:	
	TI = ("time wast*" OR "wasted time*" OR "waste of time*" OR "wast* time" OR "heat wast*" OR "wasted heat*")
	TS = ("product life cycl*" OR "radioactive wast*" OR "nuclear wast*" OR endosom*)
Optional query[a]	
	KEYWORD = recycle*
Waste-to-energy	
30	TS = ((biomass AND (ENERGY MODULE)) OR ((bioethan* OR biogas* OR biofuel* OR "energy recov*") AND (WASTE MODULE)))
29	TS = ("waste to energy" AND "energy from wast*")
28	TI = (SRF OR "solid refuse fuel*" OR "solid recovered fuel*" OR RDF OR "refuse derived fuel*") AND TS = ENERGY MODULE
27	TI = (methana* OR methani* OR methane*) AND TS = (WASTE MODULE)
26	TI = (incinerat*) AND TS = (ENERGY MODULE OR RECYCLE MODULE)
25	TI = (landfill*) AND TS = (ENERGY MODULE)
Paper industry	
24	TS = ("black liquor*" AND gasificat*)
23	TI = ("mill sludge*" OR deink*)
Various industries	
22	TS = ("product reu*" OR "reverse logistic*" OR "reverse suppl*" OR "closed loop suppl*" OR remanufact*)
21	TS = ("construction and demolit*")
20	TI = ("automotive shredder residue*")
19	TI = ("cement kiln") AND TS = (WASTE MODULE)
18	TI = (cullet*)
17	TI = (depolymer* NOT (depolymerase)) AND TS = (WASTE MODULE)
Agricultural residues	
16	TI = ("cereal straw*" OR "corn stover*" OR "rice husk*" OR "agricultural resid*" OR manure*)
Electric/Electronical residues	
15	TI = ("printed circuit board*") AND TS = (RECYCLE MODULE)
14	TS = (WEEE*)
Land and substratum recovery	
13	TS = (bioremediat*)
12	TI = (biochar* OR "terra preta" OR charcoal*) AND TS = (WASTE MODULE)
11	TI = (compost* OR vermicompost* OR brownfield* OR biosolid*)
Water recovery	
10	TI = (effluent*) AND TS = (RECYCLE MODULE)
9	TI = ((denitrif* AND water*) OR "reject water*" OR (sewage AND sludge))
General vocabulary	
8	TI = (anaerob* OR aerob*) AND TS = (WASTE MODULE)
7	TI = (feedstock*) AND TS = (WASTE MODULE)
6	TI = (scrap* OR leftover*) AND TS = (RECYCLE MODULE)
5	TI = (gasific*) AND TS = (WASTE MODULE)
4	TS = ("post consumer*")
3	TI = ("fly ash*" OR "bottom ash*") AND TS = (RECYCLE MODULE)
2	TI = ("mechanical biological treat*")
Waste recycling	
1	TI = (wast* OR recycl*) OR TS = ((wast* AND recycl*) OR (recycl* near/5 (scrap* OR garbage* OR leftover* OR trash* OR residue* OR debris*)))

[a] To be included when the database under study allows running queries restricted to keyword field.

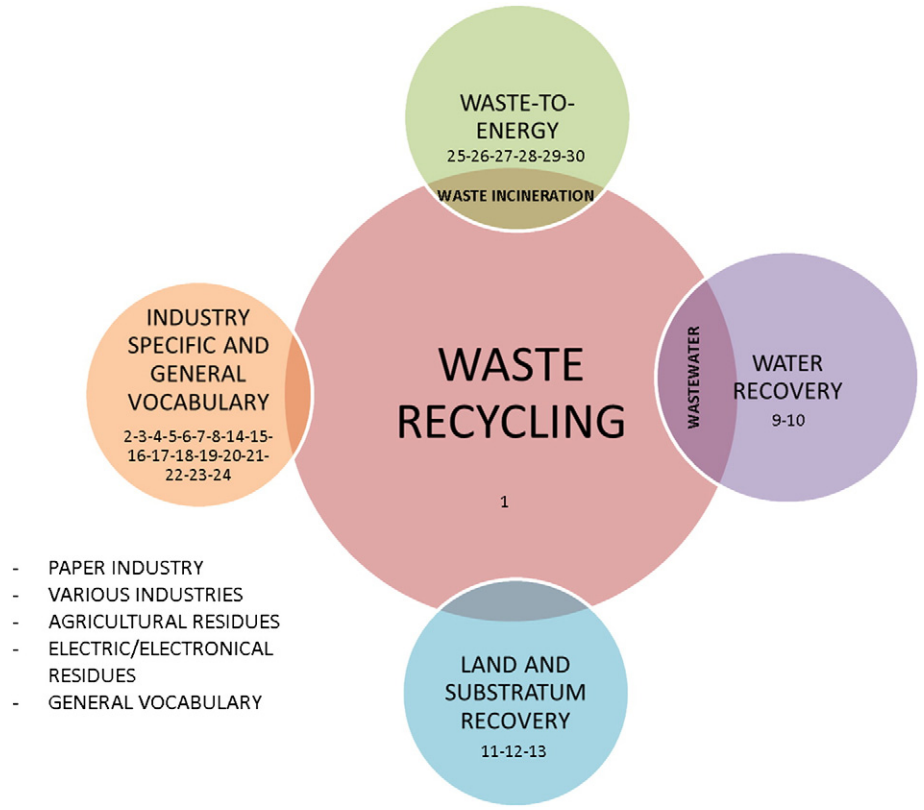

Fig. 1. Venn diagram showing the structure of the query system. The size does not exactly reflect the records retrieved by each node.

appreciable noise in the sample. The authors decided not to include them for two additional reasons: first, they added very few items that had not already been captured by the query system, and second, a low number of queries considerably eased the running of the set in multiple databases. The corrective effect exerted upon the ambiguous terms by the system of co-occurrence modules was deemed to be satisfactory, after analyzing some random samples of those retrieved in WOS databases.

2.2.3. Final structure and expert assessment

The final structure of the query system is shown in the Venn diagram (Fig. 1).

Each number in the diagram corresponds to a query as shown in Table 1. A great deal of scientific production regarding WR (60%) is captured by query 1 (central node), being surprisingly accurate, given the broad terms used. The nodes named "Industry specific and general vocabulary" and "Land and substratum recovery" both retrieve roughly 12% of items while "Waste-to-energy" node retrieves 10%. In spite of being a very important area of WR, the specific "Water Recovery" queries only retrieve 6% of items, this is due to the term "wastewater", the main keyword in water recovery activities, which gets captured by query 1 via the truncation of term "wast*". The percentages given here are orientative data obtained when running this query set in SCI and SSCI databases for year 2010.

The final set is listed in Table 1, containing 30 queries plus 2 exclusion queries according to WOS syntax, where field tag TI refers to field "Title", and TS refers to "All fields" (title, abstract and keywords). An optional query is given to be run restricted to "Keyword" field, where available. This query has not been used in this work, given that not all search engines under study allowed running a query restricted to "Keyword" field, and the comparative across databases required the query system to be homogeneous.

Exclusion queries were introduced in this system to avoid capturing non WR-related items. An important step in the building of exclusion queries is the detection of "toxic terms" by carefully examining the vocabulary suggested by the database search engine when running the searches, this service was originally intended to help finding related items, but it proved to be very useful for us to detect unwanted terms. This step can also be conducted by downloading the retrieved items and importing them to text mining software Vantage Point[1] for keyword analysis, this software offers many different options in this sense. The very laborious task of testing the different variations of queries against databases not only provides the analyst with many synonyms and new words to try, but also with the identification of "toxic terms" that, if properly identified, can be included in an exclusion query to increase the precision of the system. Roughly 4% of the retrieved items were identified as noise by these two exclusion

[1] See http://www.thevantagepoint.com/.

queries, and they were excluded prior to the information downloading phase. Please note that queries may have to be adapted to the syntax of each database.

A pilot set of queries and the Venn diagram were shown to experts working in metallurgic, e-waste and waste oil recycling sectors, who endorsed the overall approach, showing interest in future tech mining analysis carried out in WR field. Expertise from academic colleagues at the University of the Basque Country has been obtained from the early stages of this study, in areas like wastewater treatment or biology. In some cases, experts recommended candidate terms to be included in the set, mainly referring to particular technologies they applied in their respective industries or research, and these terms were tested against the scientific databases prior to their inclusion in the set. Some problems encountered when working with very specific, technical words have been explained in Section 2.2.2. It was also found that the contributions of actors in WR industries were limited to their particular activity or technology, and in spite of being valuable opinions, the authors found it necessary to look for the feedback of decision makers in a wider scope, given the breadth of the field object of study.

A meeting with IHOBE was arranged to obtain this prime endorsement. IHOBE is the public corporation in charge of supporting the Basque Government's Department for the Environment, Spatial Planning, Agriculture and Fisheries, in the development of its environmental policy [12]. Project documentation was made available prior to the meeting and feedback was asked for concerning methodology followed and the activity areas considered in WR. The answer given by the managers of IHOBE was positive, suggesting minor modifications that were subsequently implemented, and showing great interest in future tech mining analysis based on this search strategy. The WR map of science, currently in process, raised special interest because of its possibilities to support environmental decision making.

2.3. The databases

The aim of this study is to capture the research literature corresponding to WR research, so scientific publication databases are the target of our query set. The possibility of merging the results of several databases using Vantage Point made it possible for the authors to look for the maximum amount of information, but other factors had to be considered: On the one hand, the broadest possible coverage of scientific databases is desirable to capture the maximum scientific activity taking place, whereas on the other, there are computational limits for processing such amounts of information and database subscriptions are limited by available financial resources.

The final goal is to have a sufficiently large amount of items to support any kind of tech mining analysis. Further studies may require the adaptation of this query system to patent databases with the purpose of capturing the technological landscape of WR, but this is beyond the scope of this paper.

The criteria followed to choose the databases were based on the amount of information they contained and its availability. Together with pure scientific databases, social-sciences oriented databases were also to be included, taking into account the role they play in WR field [1]. An initial guide for choosing the best databases in WR science was found in the University of Connecticut database locator [13], and their proposal for environmental sciences databases was analyzed by the authors.

Two WOS databases were included, namely SCI and SSCI, looking at its broad coverage of several scientific disciplines and the availability of full citation data. The ease of its downloading process makes this database a very suitable one for tech mining purposes. SCOPUS was judged as an adequate complement to WOS databases, due to the large amount of information available, not only in the form of scientific or congress publications but also trade journals whose information the authors consider interesting for further analysis. This database only allows 2000 results of any query to be checked, so the downloading process had to be adapted to retrieve a maximum of 2000 items in each step, which caused a multiplication in the number of queries and a marked increase in the workload. Additionally, we were interested in including INSPEC in the database set with the purpose of introducing extra technical information, given its specialization in sciences and engineering. This database limits checking to 4000 results, so queries had to be adapted as previously explained. Finally, EBSCO Green file was included to further complement the dataset in environmental aspect, , this is a multidisciplinary database specialized in research about the human impact on the environment, including recycling issues [14].

Vantage Point software was used to create sub-datasets for particular analysis, as well as for merging the items retrieved from the different databases, subsequently eliminating the strong existing duplicities, as databases do considerably overlap in their coverage. Computational resources were insufficient to accomplish this task in a single step, so merging was done taking biannual intervals and limiting the fields only to those that were essential for the analysis taking place. The expression “WR database” is going to be used from this point onward to refer to the global database obtained with this work.

3. Results and discussion

3.1. Items and subjects across databases

Some characterization studies have been conducted on the downloaded WR database. For normalization purposes, only peer reviewed journal articles have been considered. Fig. 2 shows the amount of articles retrieved by the query system from each considered database, after eliminating duplicities. SCOPUS is leading the set, probably due to its wider coverage of peer-reviewed publications: 18,500 in July 2011 [15] vs. 11,471 publications covered by SCI + SSCI, as indicated in their master journal list [16]. Both INSPEC and Green File show a similar number of items. The overall trend shows an increasing number of publications over time for all databases, excepting Green File.

It should also be noted that a significant overlap exists among these four databases, estimated at roughly 38% for journal articles, meaning that almost 4 journal articles out of 10 are present in more than one database; this is clearly appreciated in the "FUSION" curve in Fig. 2. After eliminating the duplicities, a total of 102,301 journal articles have been captured in the indicated databases for the yearly interval 2000–2010.

Some of the studied databases add extra information by adding a "subject" or "classification" field to locate each item in a particular area of knowledge. These categories can be used to characterize the cognitive framework of the retrieved WR database: WOS subject categories have been widely used to characterize scientific fields [17] or even as a cognitive unit to build a global map of science [18]. Subject areas in SCOPUS have also been used for research characterization [19]. Table 2 contains the top 10 subjects as retrieved in WOS and SCOPUS with this query system. The authors considered it interesting to add a third column with INSPEC classification code. The exact contents of each subject category may differ between databases, but the authors consider that a good overall picture of the areas behind the WR database can be appreciated from the contents of Table 2.

Environmental sciences dominate the items from WOS, SCOPUS and INSPEC, followed by chemistry (including chemical engineering) and biology (considering a group formed by biotechnology, microbiology and biochemistry subjects). Energy and agricultural sciences also have their marked share of importance. INSPEC categories show a fairly different structure if compared with SCOPUS and WOS, this is partly due to the higher disaggregation level available in this field and the specialization of this database in engineering issues, but no strong discordances are observed. We could say that Table 2 points out the scientific areas a WR researcher or decision maker should pay attention to, especially when dealing with this field in the broad sense, aligned with the scope of this paper. Narrower scopes, whether it is focused on a particular technology or a particular type of waste, would require an ad hoc analysis far from the aims of this paper.

3.2. Precision and recall

An analysis about the precision and recall of this query system has been conducted on the data retrieved from WOS (61,635 journal articles), according to the definition cited in Section 2.1.

The precision has been determined by analyzing a sample of journal articles to calculate the proportion of noise in the sample, that is to say, items not directly related with WR science that have been unwittingly captured. A confidence level of 95 % and a margin of error of 5% have been pragmatically established to determine the size of the sample to be taken. The initial proportion of articles meeting the definition has been conservatively estimated at 50%, thus maximizing the size of the sample. This led to the random selection of 382 journal articles to be part of the sample. The title and abstract of every item in this sample were carefully read and analyzed by the authors to decide whether a particular item should be considered WR or not, using web glossaries [6,10] and other web resources to clear up the meaning of some specialized terms. The full paper was read when doubts raised, and academic colleagues from the University of the Basque Country were consulted when needed. This process was considerably eased by the qualification of the main authors (see Section 2.2.1) and the typical redaction of a WR paper, given the fact that these papers usually clearly state their purpose of studying the recycling/recovery of some kind of material or energy. 293 papers were found to meet the definition, so we are 95% confident that the proportion of WR articles present in the database lies between 74.2

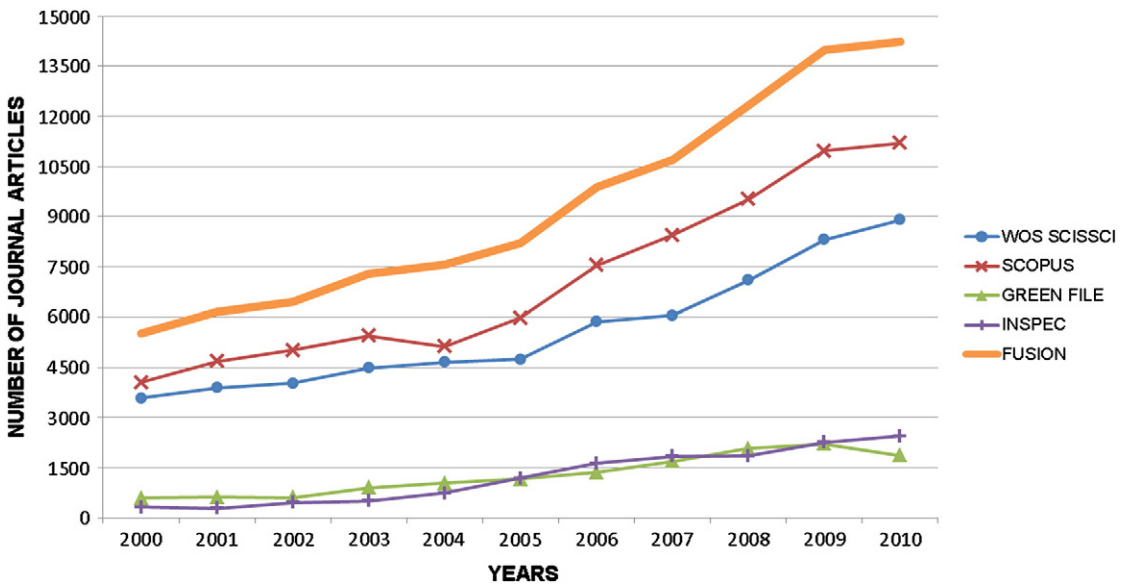

Fig. 2. Articles retrieved in peer reviewed journals, from years 2000 to 2010.

Table 2
Subject characterization of WR database, the %share of each category is indicated in brackets.

WOS (subject category)	SCOPUS (subject area)	INSPEC (classification code)
Environmental sciences (17.3%)	Environmental science (25.2%)	E0230 Environmental issues (16.5%)
Engineering, Environmental (10.9%)	Chemical engineering (11.4%)	E1840 Recycling (11.8%)
Engineering, Chemical (6.9%)	Agricultural and biological sciences (10.3%)	E1525 Industrial processes (10.4%)
Biotechnology and applied microbiology (6.6%)	Chemistry (7.8%)	E1710 Engineering materials (7.1%)
Water resources (5.9%)	Biochemistry, genetics and molecular biology (7.0%)	E3628 Biotechnology industry (2.4%)
Energy and fuels (4.4%)	Engineering (7.0%)	E1780 Products and commodities (2.3%)
Agricultural engineering (2.6%)	Immunology and microbiology (5.9%)	E3624 Fuel processing industry (1.7%)
Engineering, Civil (2.4%)	Materials science (5.7%)	E3030 Construction industry (1.6%)
Soil Science (2.3%)	Earth and planetary sciences (5.2%)	A3636 Metallurgical industries (1.5%)
Chemistry, Multidisciplinary (2.3%)	Energy (3.7%)	E3626 Chemical industry (1.5%)

and 79.2%. This is a proportion deemed adequate by similar studies in the field of nanotechnology [8] that have been successfully used to conduct tech mining studies [17].

The recall will quantify the number of truly relevant records about WR science that we are missing. This is somewhat complex, since it implies the prior determination of the total universe of WR publications, or at least, the essential core of WR science that should be captured by this strategy. In a large, highly multidisciplinary field like WR, with its fuzzy boundaries, this is a fairly difficult issue to deal with and, in fact, this paper would not be necessary were some definitive criteria available to detect the whole research taking place around WR. The many existing definitions of WR (see Section 2.1) add extra complexity to this task, making it difficult to find any subject categorization or other type of publicly available cognitive unit that fits in with the WR definition chosen for this work. The option of selecting a set of key journals and considering them the vessel of WR knowledge was considered, but there are many underlying problems regarding this option, such as determining the degree of alignment of journals with the aforementioned WR definition, as well as deciding the type of journals (academic journals, trade journals, JCR indexed or not…) to be considered. The issue of defining the WR research universe by taking a set of journals is a subject of research by itself.

Another option, widely used with research characterization purposes, consists of using the classification codes available in databases as cognitive categories that bring together the knowledge pertaining to a certain discipline. The accuracy of these classification codes has been brought into question by some papers [20], but successful works have been conducted using them as an analytical unit [17–19]. INSPEC database offers an interesting classification scheme containing a specific category for recycling research (E1840) that can be easily compared with the items retrieved by this query system by looking into the search field "Classification Code". Although far from being the definitive criteria, the authors believe that E1840 code could be used for giving a ballpark figure of the recall of this query system, taking into account the following shortcomings: First, the contents of this category are conditioned to the recall of the database itself, and second, our inclusive definition of WR is considerably wider than the scope of E1840 category. Despite this, capturing a significant part of the items under this category would be a plus for our approach. The recall test was conducted by restricting the items to journal articles written in English, and using the optional query considered for search engines allowing the search field "Keyword" (see Section 2.2.3). The query system captured 97% of the items under the E1840 category, predicting a high recall rate for this retrieval strategy.

4. Conclusions

"Waste recycling" is a concept comprising of a wide range of activities, including collection, classification/separation of waste and the subsequent marketing of recycled products. The wide range of recyclable waste materials and recycling processes make this a profoundly interdisciplinary field.

A set of 30 queries complemented by two exclusion queries and an optional query has been built with the aim of capturing the scientific production in this field. The methodology followed combines the areas and keywords detected by previous works [1] as well as keywords coming from qualified glossary services. A trial and error fine-tuning of the terms to be included in each query has been conducted, assisted by expert assessment in the more complex issues. This process proved to be a very effective method to delve deeper into the jargon of WR. An approach of co-occurrence modules has been adopted to avoid capturing articles beyond the target. The overall approach has been reviewed with IHOBE, a public corporation in charge of supporting government decision makers when developing the environmental policies of the Basque Country, obtaining positive feedback from them.

This query system has been run on four databases that were chosen either according to their coverage of worldwide scientific production (WOS, SCOPUS) or their specialization in environmental sciences and engineering (Green File, INSPEC). 102,301 journal articles were retrieved in this process for the yearly interval 2000–2010, and the overlap among the databases has been quantified (roughly 38%). Subject classifications have been analyzed where available, and the more relevant sciences behind WR have been identified, in broad terms they would be: environmental sciences, chemical sciences and biological sciences. Agricultural and energy studies also show a significant share of importance.

Finally, a random sample of articles has been analyzed to test the accuracy of the query system; the results indicate a proportion of 74.2–79.2% of WR articles with a confidence level of 95%, therefore validating the precision of the followed methodology. A recall test has been conducted using INSPEC "E1840 Recycling" classification code, obtaining highly positive but

no conclusive results, considering the particular nature of the field object of study. Some tech mining studies are currently being carried out using this WR database, the main goal being the building of a WR science map, a project which IHOBE has already expressed interest and with whom we expect to continue working, supporting our analysis with an expert-opinion source directly related with decision making in the field.

5. Lessons learned

This section is a compilation of facts and suggestions that may help tech miners to capture the research taking place around other complex, maybe ill-defined, large research areas with fuzzy boundaries like WR. Please feel free to contact with the corresponding author in case any additional information is needed.

This study considered it interesting to design a query system that could be run with minimum variation on multiple databases, looking to enlarge, as much as possible, the amount of information retrieved. This is a troublesome task, and the following facts should be taken into account:

- Databases often differ in the syntax of queries, and the "help" section of the search engine should be carefully read before anything else, especially when dealing with truncated terms and the like. Boolean operators and proximity operators can also be different across databases, and search fields may also differ in their description. These facts are usually well explained in the "help" section.
- The bias of some databases towards a particular branch of science strongly determines the precision of some terms, a high-precision query in INSPEC may be undesirable to be run in WOS. The proper use of co -occurrence modules helped to avoid this problem, but attention must be paid to the behavior of specialized databases.

"Toxic terms" to be included in exclusion queries can be detected by paying attention to the vocabulary suggested by the database search engine when running the searches. A similar analysis can be conducted importing the retrieved items to Vantage Point and analyzing the main keywords present in sample. This has been explained in Section 2.2.3.

Many databases limit the number of items that can be viewed or downloaded to a few thousand. This limitation can be successfully avoided by dividing the query in time intervals that meet the limit and subsequently merging the slices using Vantage Point.

The process of validating a query is labor-intensive and it can be quite boring, but it is a really good source of synonyms, acronyms and extra terms to be included in the strategy, as well as to detect "toxic terms". Acronyms should be carefully tested before being included; the same acronym can be used in various fields with very different meanings. The analyst should be sure of exhausting this trial-and-error method, which must be adequately complemented with expert assessment.

WR science, as defined in this study, is a highly multidisciplinary field involving a wide range of industries and activities that considerably differ in their raw materials, processes and products. The authors believe that tech miners dealing with similar circumstances should make sure they rely on enough experts with a wide scope of the field object of study. Specialized knowledge coming from experts working on a particular industry or branch of research is quite a valuable thing, but the authors strongly believe that it must be necessarily complemented by individuals with a global perspective on the field.

Acknowledgments

The authors would like to gratefully thank the peers and experts that have selflessly collaborated in the conception and correction of this paper. We are also grateful to the suggestions of a number of anonymous referees.

References

[1] G. Garechana, R. Rio, E. Cilleruelo, J. Gavilanes, Tracking the evolution of waste recycling research using overlay maps of science, Waste Manage. 32 (2012) 1069–1074.
[2] A. Demirbas, Waste management, waste resource facilities and waste conversion processes, Energy Convers. Manage. 52 (2) (2011) 1280–1287.
[3] A.L. Porter, S.W. Cunningham, Tech mining: exploiting new technologies for competitive advantage, Wiley-Interscience, United States of America, 2005.
[4] National Recycling Coalition, NRC Guiding Principles, http://nrcrecycles.org 2012.
[5] C. European Parliament, Directive 2000/76/EC of the European Parliament and of the Council of 4 December 2000 on the incineration of waste, 32000L0076, 2000.
[6] European Environment Agency, Environmental Terminology and Discovery Service (ETDS), http://glossary.en.eea.europa.eu/ 2012.
[7] European Environmental Agency, Eionet Gemet Thesaurus, http://www.eionet.europa.eu/gemet 2012.
[8] A.L. Porter, J. Youtie, P. Shapira, D.J. Schoeneck, Refining search terms for nanotechnology, J. Nanopart. Res. 10 (5) (2008) 715–728.
[9] R.N. Kostoff, Systematic acceleration of radical discovery and innovation in science and technology, Technol. Forecast. Soc. Chang. 73 (8) (2006) 923–936.
[10] US EPA, Environmental Issues Terms and Acronyms, http://iaspub.epa.gov/sor_internet/registry/termreg/searchandretrieve/termsandacronyms/search.do 2012.
[11] R. Bove, P. Lunghi, Electric power generation from landfill gas using traditional and innovative technologies, Energy Convers. Manage. 47 (11) (2006) 1391–1401.
[12] IHOBE, IHOBE.net, http://www.ihobe.net/ 2012.
[13] University of Connecticut, Research Database Locator: Find Articles & More, http://rdl.lib.uconn.edu/subjects/1898 2012.
[14] EBSCO Publishing, Green File, http://www.ebscohost.com/academic/greenfile 2012.
[15] Elsevier B.V., SciVerse Scopus. What does it cover? http://www.info.sciverse.com/scopus/scopus-in-detail/facts 2012.
[16] Thomson Reuters, Master Journal List, http://ip-science.thomsonreuters.com/mjl/ 2012.
[17] A.L. Porter, J. Youtie, How interdisciplinary is nanotechnology? J. Nanopart. Res. 11 (5) (2009) 1023–1041.
[18] L. Leydesdorff, I. Rafols, A global map of science based on the ISI subject categories, J. Am. Soc. Inf. Sci. Technol. 60 (2) (2009) 348–362.
[19] S. Kosecki, R. Shoemaker, C.K. Baer, Scope, characteristics, and use of the US Department of Agriculture's intramural research, Scientometrics 88 (3) (2011) 707–728.
[20] I. Rafols, L. Leydesdorff, Content-based and algorithmic classifications of journals: perspectives on the dynamics of scientific communication and indexer effects, J. Am. Soc. Inf. Sci. Technol. 60 (9) (2009) 1823–1835.

Gaizka Garechana holds a Bachelor's Degree in Chemical Engineering, and a degree in Industrial Engineering with MBA. He is a lecturer in Business Administration at the University of the Basque Country UPV/EHU. Current research is focused on his dissertation, a map of science of waste recycling. His research interests are competitive intelligence, management of innovation, knowledge management and the overall study of sources of competitiveness for nations and firms.

Rosa María Río-Belver is a PhD in Engineering, specializing on Industrial Engineering and Msc in Environmental Sciences and Technology. She is a University Professor at the University of The Basque Country; teaching on subjects such as Operations Research, Industrial and Technological Politics and Competitiveness–Innovation. Her main research interests are focused on innovation management, competitive intelligence, technology maps, emerging technologies, tech mining, knowledge engineering, and foresight. She is the author of numerous papers, while she is involved in national and European projects.

Ernesto Cilleruelo is a PhD in Industrial Engineering and Full Professor in the knowledge area of management. He has developed his academic and researcher career at the Faculty of Engineering at Bilbao, University of the Basque Country. His research interest is innovation in management, a research line that has been working with for twenty years.

Javier Gavilanes-Trapote holds a Bachelor's Degree in Mechanical Engineering, and a degree in Industrial Engineering. He is lecturer at the University of The Basque Country, teaching on subjects such as Business Economics, Business Strategy and Financial Management. His main research interests are focused in Patents, Innovation Management, Competitive Intelligence, Emerging Technologies, Tech mining, and Foresight.