G. Garechana, R. Rio-Belver, E. Zarrabeitia and I. Alvarez-Meaza, "From Research to Industry: A Quantitative and Qualitative Analysis of Science-Technology Transferences and Emergence Patterns in Bioremediation," in **IEEE Transactions on Engineering Management**, vol. 68, no. 5, pp. 1520-1531, Oct. 2021, https://doi.org/10.1109/TEM.2019.2936364.

Title

From research to industry: a quantitative and qualitative analysis of science-technology transferences in bioremediation.

Abstract

Bioremediation consists on utilizing living organisms for removing contaminants from various substrates. This work uses text mining techniques to determine the time lapse of knowledge transfer between research activity and technology development in bioremediation, using maps of science in order to extract patterns that could be of interest for decision making in this field. Our results show that technology developments in heavy metal bioremediation promptly follow scientific advances, as opposed to developments in bioremediation of organic chemical components. The science mapping reveals three distinct areas 1) Heavy metal remediation and phytoremediation 2) Aerobic and anaerobic remediation of chemical elements and 3) Bioremediation techniques for treating specific contamination sources such as oil. The emergence analysis points at activities involving energy recovery by bioremediation, and shows an increasing amount of technologies involving specific strains of microorganisms, which could gain significant traction in this field in an estimated time horizon of ten years.

Keywords: Bioremediation, Text mining, Science maps, Science-technology transfer

1. Introduction

The utilization of living microorganisms for changing the physical-chemical properties of substances is a millennia-old technique, fermentation being probably one of the first microbial processes that were put into practice by humans with this purpose. The natural biodegradation of substances more-or-less considered “waste” has also given a good service to mankind in key tasks such as improving the fertility of soils or simply by naturally wiping out lots of stuff that otherwise would form a huge - and probably toxic and dangerous - pile of material over the years. This work uses quantitative text mining techniques to determine the time lapse of knowledge transfer between research activity and technology development in bioremediation, and characterizes recent research activity using maps of science in order to extract patterns that could be of interest for decision making in this field.

1.1. Bioremediation: a brief introduction

The term “bioremediation” applies to the techniques aimed at using biological activity (mainly but not exclusively microorganism based) to remove contaminants in multiple sources. Initially focused on the treatment of pesticides that persist in soils long after being delivered, bioremediation has substantially expanded its range of applications to many different contaminants, such as those present in industrial wastewater and contaminated groundwater. This is a significantly multidisciplinary field that combines knowledge from various disciplines such as microbiology, chemistry, toxicology and environmental engineering (Alexander, 1999). In this section we aim at providing a bird’s eye view of bioremediation field, in order to familiarize the reader with the context where the science-technology transference analysis presented in this paper has been performed.

Some of the main contaminants that can successfully be removed by bioremediation are presented in Table 1 (Vidali, 2001). As new materials such as carbon nanotubes and its derivatives enter the economy, they become a new source of contaminants: some bioremediation techniques have also been tested in this field (Chen *et al.,* 2017). In any case, we can expect that the list presented in Table 1 will continue expanding in the future.

Class of contaminant	**Type of bioremediation**	**Potential industrial sources**
Chlorinated solvents	Anaerobic	Drycleaners Chemical manufacture
Polychlorinated biphenyls	Anaerobic	Electrical manufacturing Power station
Chlorinated phenol	Anaerobic	Timber treatment
BTEX	Aerobic and anaerobic	Oil industries Paint manufacture Chemical manufacture
Polyaromatic hydrocarbons	Aerobic	Oil industries Coke plants Engine works Power stations

Pesticides	Aerobic and anaerobic	Agriculture Pesticide manufacture

Table 1. Main contaminants and industrial sources that can be processed by bioremediation. Source: (Vidali, 2001)

Bioremediation is a particularly promising technology for treating petroleum hydrocarbon contamination in both soils and groundwater, especially considering the advantages it offers for *in situ* treatment of contaminants. Under certain circumstances, properly bio-stimulated indigenous microbial activity can be enough to offer an adequate contaminant-removal rate without further addition of exogenous bacteria (Baker and Herson, 1994). The most common methods for dealing with contaminated soils typically involve the mechanical removal of the contaminated substrate and its disposal via landfilling or its incineration, the latter creating secondary pollution by the emission of volatile organic compounds. Chemical treatment of contaminated soils poses a similar problem (Vineetha and Shibu, 2012). In spite of being significantly slower and sometimes having a lower contamination removal rate, bioremediation techniques often offer a cost-effective solution, particularly when *in situ* techniques are feasible (Prince, 2010). The standard toolkit for a composting treatment includes the use of surfactants, bulking agents, bioaugmentation techniques and both compost addition or direct composting, depending on the substrate and the contaminant type, among other factors, one or more of these techniques will be used (Chen et al., 2015).

An important factor to take into account in bioremediation is the bioavailability of the specific compound at which the treatment is aimed. Biological uptake mainly takes place in aqueous phase and under certain physicochemical conditions (Boopathy, 2000) that could be controlled in some sites, but that are difficult to control in others such as the groundwater or vadose soils, that require infiltration galleries, injection wells or other engineered means for introducing materials (Juwarkar et al., 2010). The availability of adequate amounts of oxygen, artificially inserted (bioventing) or not, determines the type of bioremediation to be used: Aerobic microbes include *Pseudomonas*, *Alcaligenes*, *Sphingomonas*, *Rhodococcus*, and *Mycobacterium*, and their efficiency to degrade pesticides and hydrocarbons in this environment has been proven. Anaerobic bioremediation is less frequently used, but there is an increasing interest in anaerobic bacteria for bioremediation of polychlorinated biphenyls in river sediments, dechlorination of the solvent trichloroethylene (TCE), and chloroform (Vidali, 2001). Underground aquifers are an example of anaerobic environment that gets polluted by BTEX (benzene, toluene and xylene isomers), perchlorates and halogenated solvents due to petrochemical activities. Pharma, solvent, dye, pesticide and ammunition industries also release mutagenic and carcinogenic nitroaromatic compounds that –in the case of polar compounds – can be fairly well treated by anaerobic bioremediation, and dangerous radionuclides can be captured by sulfate reducing organisms (Coates and Anderson, 2000). Bioavailability of contaminants is usually also reduced if the contamination persisted for long time in the substrate (Alexander, 2000), and the strength of the sorption of the element (Chen et al., 2015).

Mining activities also significantly degrade both soils and water sources in their surroundings and beyond. Acid Mine Drainage (AMD) is a hazardous phenomenon that takes place when sulfide minerals become exposed to atmospheric air and water due to mining activity, causing a double effect on the environment: toxic metals are released in high concentrations and both water and soil are significantly acidified. The conventional methods for treating this problem involve chemical precipitation reactions that need ample drying facilities for concentrating the toxic sludge. Solutions based on passive bioreactors-wetland treatment systems or active ones offer cost-efficient alternatives that under certain circumstances can lead to the recovery of metals, thus enhancing the efficiency of the process (Cohen, 2006; Martins et al., 2010).

It becomes necessary when pointing at some aspects that could be relevant for the future of this technology to mention the field of bioengineering, which has opened new ways to improve the microbial strains and their chemotactic activity (Juwarkar et al., 2010). Directed evolution techniques, for example, can dramatically change the rules of the game by allowing researchers to skip the shortcomings of natural evolution: random noise, historical accidents and ignorance of the selection pressures at work during adaptation are some of the factors that could be suppressed thanks to directed evolution (Arnold et al., 2001). The development of transgenic plants for phytoremediation (bioremediation by means of vegetal organisms) by transferring genes from microbes/other eukaryotes to plants is a real option nowadays (Eapen et al., 2007). Another shortcoming of bioremediation techniques lies in the inability to control the efficacy of the treatments without the use of expensive chemical analysis methods such as gas chromatography or mass spectrometry (Sayler and Ripp, 2000). The development of new biosensors that can be combined with conventional bioremediation methods can enhance the cost advantage of bioremediation as opposed to landfilling and/or incineration (Purohit, 2003).

1.2. Tech mining analysis of science and technology

Decision making in research and technology development can be supported by quantitative studies that complement the knowledge of managers with indicators and other empirical facts that increase the likelihood of making the right decision. Tech mining tools can be integrated in a decision support system to achieve this goal (Porter and Newman, 2011), enabling vital information present in patent and scientific publication databases to be incorporated into the decision making process. A large amount of the information contained in such databases is in textual format, making it impossible for a human analyst to read and interpret the text of thousands of documents in order to extract "activable information" ready to feed the decision making process. "Text mining" is the name given to the application of data mining tools to the analysis of unstructured, textual information (Tan, 1999), which enable the partial automation of data extraction and dimensionality reduction phases, inherent to every data analysis task. On the other side, "Tech mining" is the name coined by Porter and Cunningham (Porter and Cunningham, 2005) to describe the application of text mining tools to science and technology information databases, looking for answers to a wide set of key technology management questions. There are many examples of what can be achieved by means of tech mining analysis; in the following paragraphs we succinctly describe some of the works that better contextualize the study presented in this paper.

The evolution of scientific research in the waste recycling field has been thoroughly analyzed by Garechana et al. (Garechana et al., 2014; Garechana, Rio-Belver, Cilleruelo and Gavilanes, 2012) using tech mining tools for building science maps, among other indicators. The emergence of research areas such as waste to energy (WTE) and bioremediation from year 2002 to 2012 was detected in these maps, one of the most interesting applications of tech mining being the detection of emergent scientific areas (Upham and Small, 2009). Science maps are graphical characterizations of research fields (Garechana, Rio-Belver, Cilleruelo and Gavilanes-Trapote, 2012) that allow decision makers to have a visual grasp of the main scientific concepts that dominate an area and their evolution across time, as well as the changing relationships and potential knowledge transfers between disciplines. The application of tech mining based mapping techniques goes further than the study of science itself, enabling the study of interactions between research agents such as universities and geographical regions (Yoon et al., 2010). Text mining tools can be used to exploit information sources other than scientific articles or patents in order to answer relevant questions from a managerial or policy-making perspective, such as the effects that implementing an innovation management standard has on the innovative behavior of a firm. This can be studied by text mining the annual reports of the firms certified under such a standard (Garechana et al., 2017), and the same approach can be used to determine sustainability trends in the process industries, as shown by Liew *et al*. (Liew et al., 2014).

Similar tools can be used for technology analysis; in this case patent databases provide the information to feed the tech mining process and the methods for forecasting the evolution of technology are some of the most attractive works in this area, particularly those related to emerging technologies (Bildosola et al., 2015, 2017). An interesting application of tech mining to technological information is the elaboration of technology roadmaps, a tool combining quantitative and qualitative information in a visual representation of the evolution of technology across time (Petrick and Echols, 2004), that can improve decision making in technology, particularly in terms of building common understanding across internal and external organizational boundaries (Phaal et al., 2005).

2. Methodology

This work uses tech mining tools in order to study the time lapse of knowledge transfer between research activity and technology development in bioremediation, and characterizes recent research activity in that field. This section starts by explaining the criteria followed to collect and clean the data, after that the process to quantify the knowledge transfer time lapse and the method for building the science maps is described.

2.1. Data collection and cleaning

The evaluation of the sources available for characterizing the scientific activity in bioremediation led us to choose the Scopus database for the following reasons: First, considering its coverage, Scopus is the world's largest abstract and citation database of peer-reviewed literature. Second, considering the quality of the sources included in Scopus, every potential source is carefully curated and ultimately selected by an international group of scientists, researchers and librarians who represent the major scientific disciplines. Every year a review process of all new titles that are suggested to Scopus is conducted, in addition to reviewing and ensuring the quality of existing content (Elsevier, 2018). The database chosen for analyzing the bioremediation technologies was Patseer, a patent database comprising full INPADOC (the main database produced and maintained by the European Patent Office) records as well as contents from patent offices from several countries (Patseer, 2018).

The following query was run on the Scopus database to retrieve the scientific journal articles dealing with bioremediation techniques. Peer reviewed journal articles guarantee the quality of the indexed publications via a peer review process, and generally offer a great deal of well-structured indexing information, perfectly suitable for text mining analysis (Porter and Cunningham, 2005). The time interval of scientific production to be analyzed was set from year 1994 to 2017, since there are almost no bioremediation patents filed before year 1994, and scientific articles become scarce for years before 1994. The following query returned 31909 results on 18th may 2018:

TITLE-ABS-KEY (bioremediation) AND PUBYEAR > 1993 AND PUBYEAR < 2018 AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SRCTYPE , "j"))

Having downloaded the records corresponding to scientific activity, the next decision consists in choosing the textual field(s) that is/are going to be text mined in order to extract the science/technology indicators. At this point the authors considered the advantages of analyzing the information contained in the “author keyword” field: this field contains the keywords that, in the opinion of the author that originally submitted the journal article, best characterize the research done, creating a powerful tool with which to characterize the research being undertaken. In addition to this, journals often encourage authors to select a reduced number of terms that best describe the essence of the article, avoiding abbreviations and being as field-specific as possible. This considerably reduces the ambiguity and noise inherent to any text-mining study but at the same time it also reduces the amount of data available to conduct the study, since a certain amount of articles (particularly data from the nineties and before) do not have this field available. After removing the duplicates and taking only the articles with an “author keyword” field, our final dataset is reduced to 25035 journal articles. These records were downloaded in csv format and exported to Vantage Point text mining software (Search Technology, 2018) in order to extract the main concepts related to bioremediation science over time.

2.2. Science-technology analysis

We propose an approach for detecting the emergence and decline of “significant terms” in bioremediation, identified using a term frequency - inverse document frequency (TFIDF) algorithm on the data contained in the “author keyword” field. The detected trends have been compared, when feasible, with the trends shown by these same significant terms in technology, by building search queries adapted to Patseer patent database. According to the patterns described in linear innovation paths, advances in research are supposed to precede those developments in industry, this is particularly true in science-intensive fields (Balconi et al., 2010; Etzkowitz, 2008) such as bioremediation. We try to quantify the presumed time delay between the peak of a certain concept in research and its equivalent peak in technology developments by comparing the data we get from scientific articles and patents. To detect the patents associated with each research concept we built a search query adapted for each of the concepts, where a truncated form (if necessary) of the concept was forced to co-occur with either “remediation” or “bioremediation” terms in the title, abstract or claim (TAC) fields of the patent. The following is an example of the query corresponding to the concept “*Pseudomonas aeruginosa*” run on Patseer database to retrieve the patents related with the use of this technology for remediation purposes:

TAC:(pseudomonas aeruginosa) AND TAC:(remediation OR bioremediation)

The text mining procedure applied to the data is shown in the flowchart of Figure 1.

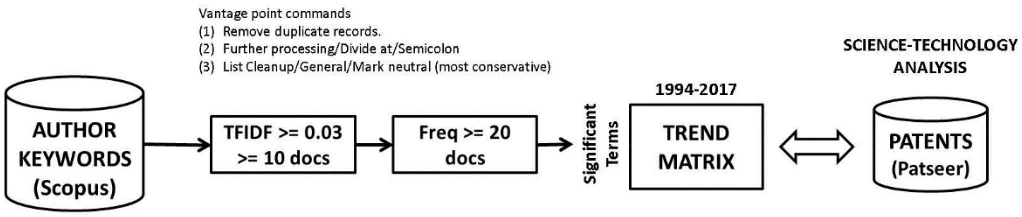

Figure 1. Text mining process for the science-technology analysis.

The "author keyword" field synthetizes the main topics of the scientific paper with a fairly decent amount of unambiguity and data "cleanliness" that are highly desirable in a text mining study. Despite this, there are several synonyms, plural/singular forms of the same term and other sources of noise that have to be cleaned in the first steps. The text mining software Vantage Point has been used to merge all equivalent forms of the same word in order to build a set of homogenized terms and track their occurrences across the years. The sequence of commands executed using Vantage Point is described in Figure 1.

This process generates a list containing the homogenized forms of the author keywords present in the dataset, the presence of terms such as "remediation" and "bioremediation" being overwhelmingly higher in absolute terms when compared to terms more specific to a certain technology or subfield of bioremediation. A trend analysis should be precisely focused on the latter type of terms, given the fact that the maturity and emergence of techniques often go unnoticed if attention is paid to the broad picture, which is dominated over the years by roughly the same broad concepts in the absence of dramatic, radical innovations that could change the overall landscape of a scientific area. In order to pick the more meaningful - for the purposes of this study - or "significant" terms, we set a term filter based on the TFIDF algorithm: this process weighs a set of terms present in a collection of documents (journal articles, in this case), penalizing the terms that occur in many documents (these are likely to be general terms and not particularly relevant for characterizing the contents of a document) and giving advantage to those terms that occur frequently but are not widespread in the document collection (Porter and Zhang, 2012; Robertson, 2004). We pragmatically set a minimum TFIDF threshold of 0.03 that should be achieved in at least 10 documents in order to select the significant terms for the trend analysis. An additional restriction was also considered by only including terms that occur at least 20 times in the dataset, in order to suppress overly infrequent terms from the trend analysis.

The last step of this analysis consisted in building a trend matrix where the emergence and decline of significant concepts could be analyzed. It should be noted that the number of bioremediation publications presents a somewhat steady yearly growth rate that is also present in many other scientific fields, as the number of scholars and institutions devoted to research increase globally. For a concept to qualify as “emerging” or “declining”, its relative growth/decline rate has been normalized by the overall trend of publications in the field of bioremediation. After carefully analyzing the trend matrix we set a guideline based on the average growth of each term in the last six target years (2011-2017) of the study. Those terms averaging a negative normalized growth in that interval will be considered declining or mature concepts in bioremediation, while terms averaging a normalized growth higher than 25% will be considered emerging concepts.

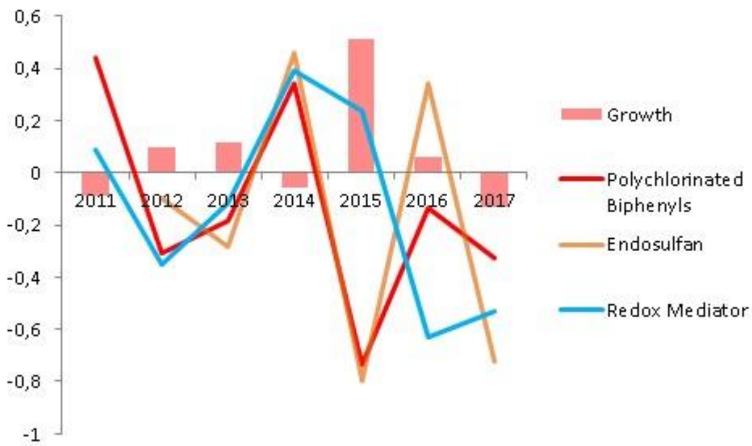

Figure 2. Normalized relative change of declining/mature concepts in bioremediation.

Figure 2. shows the normalized rate of relative change on three clearly declining/stagnant concepts in bioremediation, namely “Polychlorinated Byphenils”, “Endosulfan” and “Redox Mediator”. The bars show the yearly rate of growth (an average yearly growth of 7.3%) in bioremediation scientific publications.

2.3. Science map analysis

The science-technology analysis is complemented by the building of a set of science maps aimed at characterizing the broad changes that took place in the field of bioremediation during the time interval under study. With this purpose in mind, we build two separate maps for the intervals 1994-2006 and 2007-2017 (hereinafter referred to as PRE and POST, respectively), plus a map corresponding to the emergent significant terms detected in the trend analysis explained in the section 2.2, built using 2015-2017 (hereinafter we will refer to this interval as “emergent”) data. This latter map aims at providing a foresight of the scientific concepts that may be translated into technological developments in coming years, according to the results obtained in the trend analysis. No TFIDF filtering has been used to build the maps corresponding to PRE and POST intervals, merely mapping the 500 most frequent terms for the sake of interpretability and simplicity of the visualizations. Figure 3 shows the process followed to build the maps.

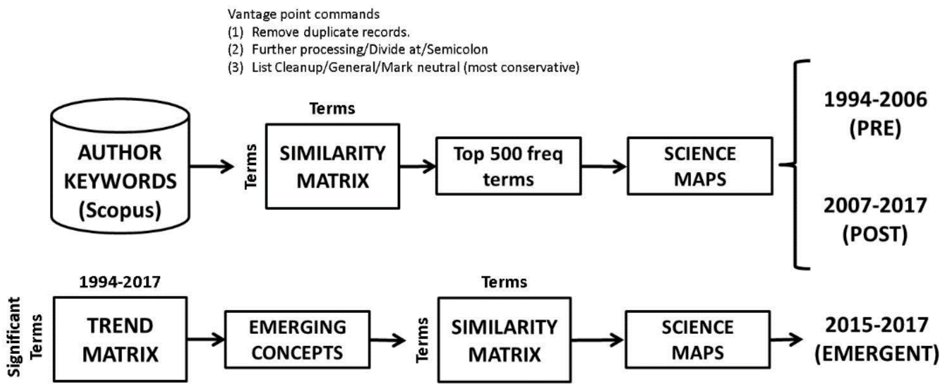

Figure 3. Text mining process for the science mapping analysis.

These similarity matrices contain the absolute number of times each of the terms co-occurs in the “author keyword” field with the rest of the terms of the dataset. We can expect that terms that co-occur frequently with each other maintain a close cognitive relationship, as pointed out by the amount of scientific work directly relating them to each other. However, this co-occurrence data has to be normalized, since very frequent terms will quite likely have an above the average co-occurrence rate with a higher number of words. There are many available methods for normalizing the raw co-occurrence data, Klavans and Boyack (Klavans and Boyack, 2006) make a thorough review of the more commonly used methods, although there are criticisms regarding the accuracy of some of the measures (Van Eck and Waltman, 2009). We opted for taking into account the explanations of Van Eck and Waltman (Van Eck and Waltman, 2009) on probability-based similarity measures and calculated the indirect Salton’s Cosine described by these authors (Van Eck and Waltman, 2008). There is at least one more similarity/distance indicator with the same name that should not be confused with our choice, since the calculation of “our” Salton’s Cosine considers the full vector of co-occurrences of each word with the rest in the term x term co-occurrence matrix, as shown in the following formula:

$$Cos_{i,j} = \frac{\sum_{q=1}^{N} C_{iq} C_{jq}}{\sqrt{\sum_{q=1}^{N} C_{iq}^{2} \sum_{q=1}^{N} C_{jq}^{2}}} \quad (1)$$

Where Ciq, Cjq are the co-occurrence values of i and j terms in their co-occurrences with the rest of the terms, in the N-term square co-occurrence matrix. The result of this approach is a term similarity matrix that quantifies the degree of similarity between the main concepts in bioremediation science. The maps have been generated using visualization and clustering software VOSViewer (Van Eck and Waltman, 2010).

3. Results and discussion

The analysis of declining vs emergent concepts in bioremediation described in section 2.2 generated a list of declining concepts that show a negative normalized growth rate according to the data corresponding to the last six years under study. These are concepts that experienced their peak in relative attention from the bioremediation scientific community in the past, and the aim of this study is to test the assumption under which advances in research are supposed to precede the developments in industry, this being particularly true in science-intensive fields (Balconi et al., 2010; Etzkowitz, 2008), and to quantify the time delay between the peak in attention from the scientific community and the peak in technological developments, as shown by the number of patent applications related to each concept. Table 2 shows the list of declining concepts in bioremediation we detected using our approach, indicating the year in which the relative peak of attention was achieved by such concept (measured by the relative presence of the concept in the data corresponding to that year) in research, and the year where the number of patent applications related to that concept had its peak. The concepts that did not show any appreciable patenting activity were discarded from the analysis.

Research concept	**Research/ patenting peaks**	**Time (years)**	**Research concept**	**Research/ patenting peaks**	**Time (years)**
16S rdna	2011/2010	-1	Rhizoremediation	2003/2014	11
Polycyclic aromatic hydrocarbons	2011/2010	-1	Endosulfan	2000/2012	12
Lead	2016/2016	0	*Pseudomonas putida*	2001/2013	12
Redox mediator	2008/2010	2	Chlorinated solvents	2000/2012	12
Zn	2014/2016	2	Hydrocarbons	1997/2010	13

Nitrite	2008/2010	2	Ryegrass	2001/2014	13
Reactive transport	2007/2010	3	Immobilized cells	1997/2010	13
Eichhornia crassipes	2011/2014	3	Polychlorinated biphenyls	1998/2012	14
Biostimulation	1999/2003	4	Organic matter	2003/2017	14
Pentachlorophenol	1995/2000	5	White-rot fungi	1998/2014	16
Copper	2009/2014	5	Chlorinated ethenes	1996/2013	17
Hexavalent chromium	2005/2011	6	Sediment	1996/2015	19
Geobacter	2002/2009	7	Benzene	1995/2015	20
Chelator	2002/2010	8	P-Nitrophenol	1995/2015	20
Toluene	1995/2003	8	*Pseudomonas aeruginosa*	2009/2015	6
Soil washing	2006/2014	8	Chromium	2007/2015	8
Biodiversity	2001/2010	9	Endocrine disruptors	2004/2006	2
Composting	2002/2011	9	Aeration	1999/2015	16

Table 2. Declining/stagnant concepts in bioremediation. Both peak years in research and patenting activity are shown, and the delay between these is given.

The classic linear innovation hypothesis of research preceding technical development seems to be confirmed by our data. Roughly 8% of our observations contradict this hypothesis, and do so in an amount that could easily be attributable to the measurement error inherent to any text-mining based approach. The average delay from peak to peak is some 9.6 years, however, the high standard deviation (5.4 years) of data is probably signaling the presence of subgroups in the declining concept set. Delving a bit deeper into the data, we detected a pattern after dividing the declining concepts into heavy metals, living organisms and chemicals, as shown in Table 3:

Category	**Concepts**	**Time (years)**
Heavy Metals	Lead Zinc Copper Chromium	3.75
Living Organisms	*Eichhornia crassipes* *Pseudomonas aeruginosa* *Geobacter* Rhizoremediation *Pseudomonas putida* Ryegrass White-rot fungi	9.71
Chemicals	Pentachlorophenol Toluene	13.5

	Endosulfan Chlorinated solvents Polychlorinated biphenyls Chlorinated ethenes Benzene p-Nitrophenol	

Table 3. Subgroups detected in the declining concepts.

There is evidence suggesting that the science-technology transferences may occur slightly faster in the case of research for bioremediation of heavy metals, than in the bioremediation of chemical contaminants, while the research directly dealing with particular applications, features or modification of living organisms also shows its own idiosyncrasy. The reasons for this differentiated behavior may be related with the effects of regulation and the proven effectiveness of the bioremediation treatments in each case, these are both factors - among many others - that could increase the propensity to patent and thereby reduce the time lapse between research advances and technological developments. Substantial increases in the presence of a certain type of contaminant can also lead to increased efforts both in research and patenting activity seeking to solve this problem.

Table 4 shows the top 10 research concepts that qualify as “emergent” under our approach: these are terms that show a 6-year average normalized growth higher or equal to 0.25. A total amount of 200 terms qualify as emergent under our approach.

Term	**Average growth coefficient**

Transport	1.87
Oil	1.76
Biogas	1.75
Bacterial diversity	1.65
Rhodococcus	1.31
Rhizobacteria	1.31
Aerobic granular sludge	1.31
Bioindicator	1.29
Leaching	1.25
Plasmid	1.19

Table 4. Top 10 emerging terms, according to the normalized average growth coefficient.

The PRE and POST science map analysis are shown in Figures. 4 and 5. The size of the circles indicates the relative frequency of each concept in the dataset, and the colors signal the thematic clusters detected by VOSviewer software based on the cosine similarities explained in section 2.3. The lines are drawn according to such similarities, but the number of lines mapped has been reduced for clarity in the display.

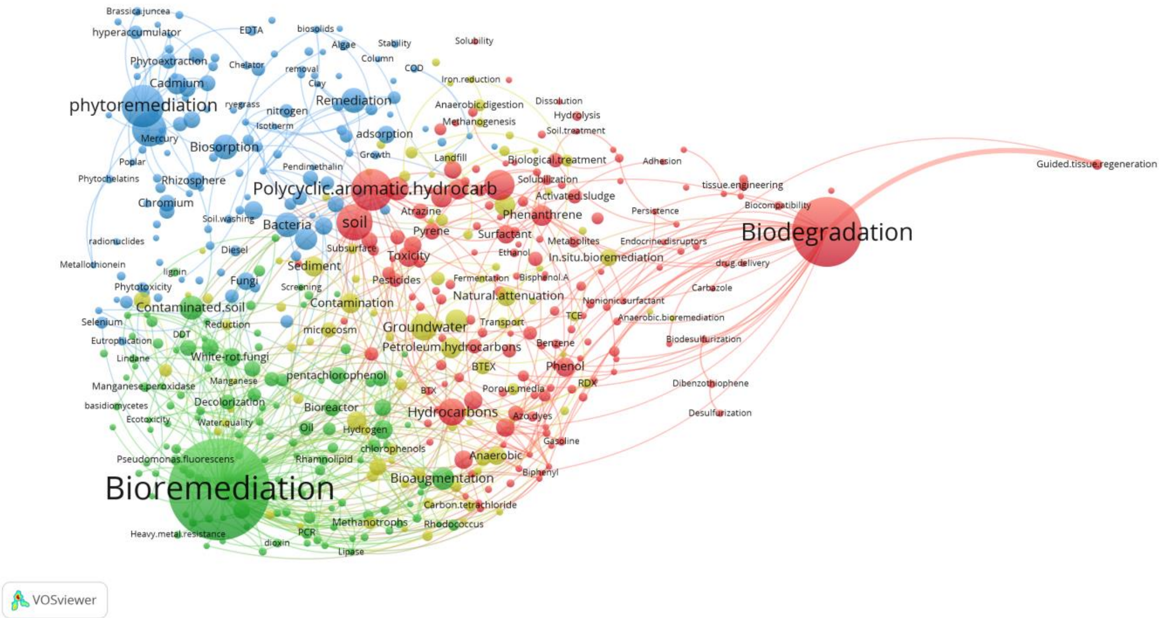

Fig 4. PRE bioremediation science map.

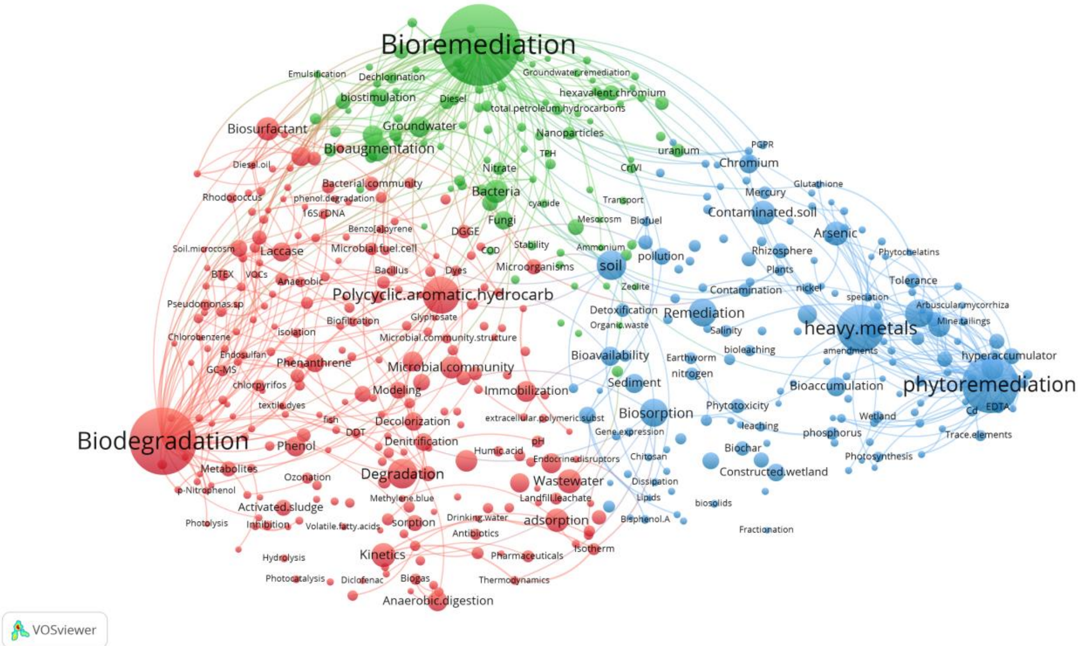

Figure 5. POST bioremediation science map.

The analysis of these maps shows the prevalence over time of three large, clearly distinct topics:

- Phytoremediation and heavy metals (blue): Our data shows that a substantial amount of research in phytoremediation techniques is aimed at the sorption of heavy metal contaminants.
- Remediation of polycyclic aromatic hydrocarbon, phenolic and polychlorinated compounds, both using aerobic and anaerobic organisms (red). Most of the concepts are involved with specific contaminant substances.
- Research involving specific contamination problems due to petroleum spills, groundwater contamination or textile effluents (green). Most of the concepts are involved with bioremediation techniques and/or industry specific contamination sources.

The term composition of these clusters is fairly robust across all the intervals analyzed, pointing at a stable, well developed scientific discipline where advances in research happen gradually, with no radical breakthroughs that could cause noticeable changes to the conceptual fingerprint of the field. We rely on our approach to detect emergent concepts explained in section 2.2 to identify the transformations that bioremediation is undergoing: Figure 6 shows the cluster structure formed by these emergent concepts. The size of the circles in this map does not correspond with the relative frequency of the terms but rather with the growth rate of each term, consequently infrequent terms that are dramatically increasing in their relative importance could become more conspicuous on the map.

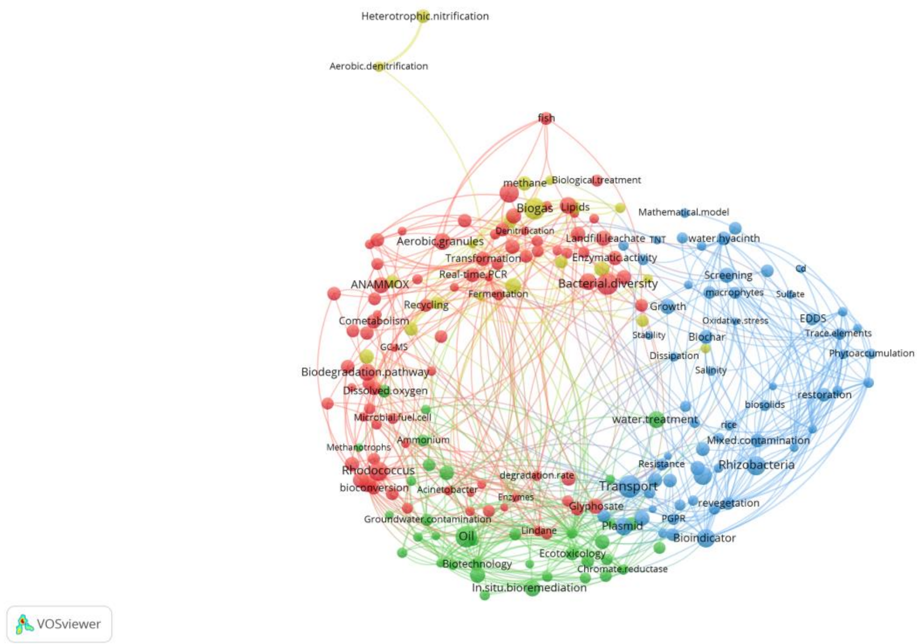

Figure 6. Emergent concepts in bioremediation.

The mapping of emergent terms coherently shows a cluster structure similar to PRE and POST science maps: the emerging concepts in each of the clusters detected in the previous science maps are strongly linked to each other, thus maintaining the cluster structure described above. It is remarkable that the emergent terms map shows a cluster closely related to energy recovery (yellow) overlapping with the terms associated with the red cluster (chemical contaminant remediation), and so pointing at a distinct specialization emerging therein. Microbiology research as a whole is growing considerably above average in this latter cluster, as opposed to the clearly defined specialization that is taking place in the emergent terms corresponding to the green cluster (bioremediation techniques and specific industrial contamination): Oil-related bioremediation techniques and in-situ bioremediation are among the topics that are experiencing faster growth in this area. Regarding emerging terms in the blue cluster (phytoremediation, heavy metal removal) our data shows that increasing attention is being paid to the phytoremediation potential of the water hyacinth, nonetheless, “hot topics” are more focused on transport problems - techniques aimed at providing adequate contact between microorganisms and reactants - the rhizobacteria ecosystems and techniques for checking the progress of the bioremediation treatments. It is remarkable that some of the fastest growing concepts correspond to this area, as shown in Table 4. According to these results and the data shown in Table 3, it seems reasonable to expect an increasing amount of technologies involving specific strains of - perhaps genetically engineered - microorganisms during the next ten years, this hypothesis being subject to the evolution of many variables related to the evolution of technology.

4. Conclusions

This research is a quantitative analysis conducted using data from Scopus, fundamentally a scientific publication database, and Patseer, a patent database, in order to study the transference of knowledge between research and technology development in bioremediation. In addition to this, the conceptual landscape of bioremediation has been mapped and its changes across the years characterized. The stability shown by this field led us to complement the mapping with an emergent term map that shed more light onto the change dynamics this field is undergoing.

A query system was designed to extract the information dealing with bioremediation from Scopus database and the “author keyword” field of scientific publications was text mined in order to extract the significant terms that contained the essence of subfields and research problems being tackled by the bioremediation research community. A trend matrix was built containing the evolution of the relative importance of these significant terms over time, allowing us to detect declining/stagnant vs emergent concepts in bioremediation. A data subset formed by the declining/stagnant terms was used to detect the peak in relative importance - as measured by the relative presence - of each of these concepts and to detect the time lapse between the apex of the attention paid by the scientific community and the peak in technology developments concerning these concepts.

It should be noted that we lack a systematic approach to study the dynamic reciprocal dependence between scientific and technological activity (Zhao and Guan, 2013), however, knowledge intensive fields are prone to experience a faster knowledge transference between science and technology (Finardi, 2011). Our results show considerable differences in the time lags between the predominance of concepts that science deals with, and the elements of technology that come later, depending on the scope of the research concepts. On one side, the technology developments in heavy metal bioremediation seem to follow shortly after the scientific advances (average time lag of 3.75 years), as opposed to developments in bioremediation of organic chemical components (average time lag of 13.5 years). The development of technology is a highly multivariate problem where factors such as the market demand, the rivalry between firms, the concentration of the market, the patentability of the technology or pressing needs caused by a sudden increase in a certain type of pollution (the hydrocarbon leakages in underground water due to shale gas extraction being a recent example), among others, play a key role. Looking for an answer to the question as to why such noticeable divergences exist would be a research project in itself, given the systemic complexity involved in the technology development process. However, our research suggests that the relationship between science and technology in these areas may be significantly different. The developments involving a specific type of organism take the middle ground, on average 9.7 years pass from the peak attention given to these organisms by the scientific community and the peak in the technical developments that make explicit use of them. According to our method, however, the linear innovation theory - of research preceding technology - seems to be proven, this being an expected outcome taking into account the knowledge-intensive features of the bioremediation field.

The science mapping of bioremediation research gives us enough evidence to describe this field as stable and subject to incremental improvements, rather than exposed to impacts from innovation shocks. Three distinct areas are neatly defined in the mappings, with few noticeable changes taking place before and after year 2007, this indicates certain maturity in the specialties inside bioremediation. The clusterization algorithm run using VOSviewer software detected homogeneous groupings formed by 1) Heavy metal remediation and phytoremediation 2) Aerobic and anaerobic remediation of chemical elements and 3) Bioremediation techniques for treating specific effluents and contamination sources such as oil. The persistence of this clusterization solution across the analyzed years is a signal of well-established boundaries between these activities.

With the aim of adding a foresight dimension to our work, we mapped the emergent terms (those with a 6-year normalized rate of growth higher or equal to 25%) in bioremediation, confirming that the cluster structure revealed in the field-broad maps of science was still present in the emergent terms. This reinforces our conclusions regarding the maturity of the detected specialties, since even the most dynamic concepts still remain inside the specialty boundaries. An interesting cluster related to energy recovery by bioremediation is detected, overlapping with the 2) cluster explained above. It should be noted that this cluster is formed by emergent concepts that show a certain amount of internal cohesion, so this trend could lead to the formation of a new discipline inside bioremediation science in the future. The remarkable emergence of energy-related issues in waste management has already been pointed out by other studies, thus supporting our conclusion (Garechana et al., 2015; Garechana, Rio-Belver, Cilleruelo and Gavilanes, 2012). The dynamism detected in the microbiology concepts coupled with the time delays presented in Table 3 led us to conclude that the next ten years may produce an increasing amount of technologies involving specific strains of microorganisms, which could be the result of genetic engineering. Whether these technologies will turn up as patents remains to be seen, given the multiple factors cited in this paper that have an important influence on the evolution of technology.

We believe that the method and results explained in this paper can offer valuable resources and a deeper insight about science and technology to managers involved in environmental management decision making processes, particularly those that include bioremediation in their quandaries.

References

Alexander, M. (1999), *Biodegradation and Bioremediation*, Academic Press.

Alexander, M. (2000), “Aging, Bioavailability, and Overestimation of Risk from Environmental Pollutants”, *Environmental Science & Technology*, American Chemical Society, available at:https://doi.org/10.1021/ES001069+.

Arnold, F.H., Wintrode, P.L., Miyazaki, K. and Gershenson, A. (2001), “How enzymes adapt: lessons from directed evolution”, *Trends in Biochemical Sciences*, Vol. 26 No. 2, pp. 100–106.

Baker, K.H. and Herson, D.S. (1994), *Bioremediation*, McGraw-Hill, Inc.

Balconi, M., Brusoni, S. and Orsenigo, L. (2010), “In defence of the linear model: An essay”, *Research Policy*, Vol. 39 No. 1, pp. 1–13.

Bildosola, I., Rio-Belver, R. and Cilleruelo, E. (2015), “Forecasting the big services era: novel approach combining statistical methods, expertise and Technology Roadmapping”, Springer, Cham, pp. 371–379.

Bildosola, I., Río-Belver, R.M., Garechana, G. and Cilleruelo, E. (2017), “TeknoRoadmap, an approach for depicting emerging technologies”, *Technological Forecasting and Social Change*, Vol. 117, pp. 25–37.

Boopathy, R. (2000), “Factors limiting bioremediation technologies”, *Bioresource Technology*, Elsevier, Vol. 74 No. 1, pp. 63–67.

Chen, M., Qin, X. and Zeng, G. (2017), “Biodegradation of Carbon Nanotubes, Graphene, and Their Derivatives.”, *Trends in Biotechnology*, Elsevier, Vol. 35 No. 9, pp. 836–846.

Chen, M., Xu, P., Zeng, G., Yang, C., Huang, D. and Zhang, J. (2015), “Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: Applications, microbes and future research needs”, *Biotechnology Advances*, Elsevier, Vol. 33 No. 6, pp. 745–755.

Coates, J.D. and Anderson, R.T. (2000), “Emerging techniques for anaerobic bioremediation of contaminated environments”, *Trends in Biotechnology*, Elsevier Current Trends, Vol. 18 No. 10, pp. 408–412.

Cohen, R.R.H. (2006), “Use of microbes for cost reduction of metal removal from metals and mining industry waste streams”, *Journal of Cleaner Production*, Elsevier, Vol. 14 No. 12–13, pp. 1146–1157.

Eapen, S., Singh, S. and D’Souza, S.F. (2007), “Advances in development of transgenic plants for remediation of xenobiotic pollutants”, *Biotechnology Advances*, Elsevier, Vol. 25 No. 5, pp. 442–451.

Elsevier. (2018), “Scopus content”, available at: https://www.elsevier.com/solutions/scopus/content (accessed 15 March 2018).

Etzkowitz, H. (2008), *The Triple Helix: University-Industry-Government Innovation in Action*, Routledge.

Finardi, U. (2011), “Time relations between scientific production and patenting of knowledge: the case of nanotechnologies”, *Scientometrics*, Springer Netherlands, Vol. 89 No. 1, pp. 37–50.

Garechana, G., Río-Belver, R., Bildosola, I. and Salvador, M.R. (2017), “Effects of innovation management system standardization on firms: evidence from text mining annual reports”, *Scientometrics*, Springer Netherlands, Vol. 111 No. 3, pp. 1987–1999.

Garechana, G., Río-Belver, R., Cilleruelo-Carrasco, E. and Larruscain-Sarasola, J. (2014), “Clusterization and Mapping of Waste Recycling Science. Evolution of Research From 2002 to 2012”, *Journal of the Association for Information Science and Technology*, Vol. In press.

Garechana, G., Rio-Belver, R., Cilleruelo, E. and Gavilanes-Trapote, J. (2012), “Visualizing the scientific landscape using maps of science”, *Industrial Engineering: Innovative Networks*, Springer, pp. 103–112.

Garechana, G., Rio-Belver, R., Cilleruelo, E. and Gavilanes, J. (2012), “Tracking the evolution of waste recycling research using overlay maps of science”, *Waste Management*, Vol. 32 No. 6, pp. 1069–1074.

Garechana, G., Rio-Belver, R., Cilleruelo, E. and Larruscain Sarasola, J. (2015), “Clusterization and mapping of waste recycling science. Evolution of research from 2002 to 2012”, *Journal of the Association for Information Science and Technology*, Vol. 66 No. 7, pp. 1431–1446.

Juwarkar, A.A., Singh, S.K. and Mudhoo, A. (2010), "A comprehensive overview of elements in bioremediation", *Reviews in Environmental Science and Bio/Technology*, Springer Netherlands, Vol. 9 No. 3, pp. 215–288.

Klavans, R. and Boyack, K.W. (2006), "Identifying a better measure of relatedness for mapping science", *Journal of the American Society for Information Science and Technology*, Wiley Online Library, Vol. 57 No. 2, pp. 251–263.

Liew, W. Te, Adhitya, A. and Srinivasan, R. (2014), "Sustainability trends in the process industries: A text mining-based analysis", *Computers in Industry*, Vol. 65 No. 3, pp. 393–400.

Martins, M., Santos, E.S., Pires, C., Barros, R.J. and Costa, M.C. (2010), "Production of irrigation water from bioremediation of acid mine drainage: comparing the performance of two representative systems", *Journal of Cleaner Production*, Elsevier, Vol. 18 No. 3, pp. 248–253.

Patseer. (2018), "Patseer coverage", available at: https://patseer.com/detailed-coverage/.

Petrick, I.J. and Echols, A.E. (2004), "Technology roadmapping in review: A tool for making sustainable new product development decisions", *Technological Forecasting and Social Change*, Vol. 71 No. 1–2, pp. 81–100.

Phaal, R., Farrukh, C.J.P. and Probert, D.R. (2005), "Developing a technology roadmapping system", *Technology Management: A Unifying Discipline for Melting the Boundaries*, Ieee, No. 1, pp. 99–111.

Porter, A.L. and Cunningham, S.W. (2005), *Tech Mining: Exploiting New Technologies for Competitive Advantage*, Wiley-Interscience, United States of America.

Porter, A.L. and Newman, N.C. (2011), “Mining external R&D”, *Technovation*, Elsevier, Vol. 31, pp. 171–176.

Porter, A.L. and Zhang, Y. (2012), “Text clumping for technical intelligence”, *Theory and Applications for Advanced Text Mining*, available at:https://doi.org/10.5772/50973.

Prince, R.C. (2010), “Can we improve bioremediation?”, *Handbook of Hydrocarbon and Lipid Microbiology*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3351–3355.

Purohit, H.J. (2003), “Biosensors as molecular tools for use in bioremediation”, *Journal of Cleaner Production*, Elsevier, Vol. 11 No. 3, pp. 293–301.

Robertson, S. (2004), “Understanding inverse document frequency: on theoretical arguments for IDF”, *Journal of Documentation*, Emerald Group Publishing Limited, Vol. 60 No. 5, pp. 503–520.

Sayler, G.S. and Ripp, S. (2000), “Field applications of genetically engineered microorganisms for bioremediation processes”, *Current Opinion in Biotechnology*, Elsevier Current Trends, Vol. 11 No. 3, pp. 286–289.

Search Technology, I. (2018), “The Vantage Point. Turn information into knowledge”, available at: https://www.thevantagepoint.com/.

Tan, A.H. (1999), “Text mining: the state of the art and the challenges”, *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, Citeseer, pp. 65–70.

Upham, S.P. and Small, H. (2009), “Emerging research fronts in science and technology: patterns of new knowledge development”, *Scientometrics*, Vol. 83 No. 1, pp. 15–38.

Van Eck, N.J. and Waltman, L. (2008), “Appropriate similarity measures for author co-citation analysis”, *Journal of the American Society for Information Science and Technology*, Wiley Online Library, Vol. 59 No. 10, pp. 1653–1661.

Van Eck, N.J. and Waltman, L. (2009), "How to normalize cooccurrence data? An analysis of some well-known similarity measures", *Journal of the American Society for Information Science and Technology*, Wiley Online Library, Vol. 60 No. 8, pp. 1635–1651.

Van Eck, N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Springer, Vol. 84 No. 2, pp. 523–538.

Vidali, M. (2001), "Bioremediation. An overview.", *Pure Appl. Chem*, Vol. 73 No. 7, pp. 1163–1172.

Vineetha, V. and Shibu, K. (2012), "Bioremediation of oil contaminated soil", *International Conference on Green Technologies (ICGT)*, Trivandrum, pp. 99–102.

Yoon, B., Lee, S. and Lee, G. (2010), "Development and application of a keyword-based knowledge map for effective R&D planning", *Scientometrics*, Springer, Vol. 85 No. 3, pp. 803–820.

Zhao, Q. and Guan, J. (2013), "Love dynamics between science and technology: some evidences in nanoscience and nanotechnology", *Scientometrics*, Springer Netherlands, Vol. 94 No. 1, pp. 113–132.