Journal of Cleaner Production 129 (2016) 305–314

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A pre-assessment of past research on the topic of environmental-friendly electronics

Alptekin Durmuşoğlu*

Gaziantep Univ, Dept Ind Engn, TR-27310 Gaziantep, Turkey

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form
13 April 2016
Accepted 16 April 2016
Available online 23 April 2016

Keywords:
Text mining
Environmental-friendly electronic
Technology forecasting
Bibliometrics
Singular value decomposition
K-means clustering

ABSTRACT

Environmental-friendly products, processes, tools, methods and etc… have been under the interest of both industry and academy. However, this interest has not been analyzed systematically for the academic studies. On the other hand, bibliometric analyses have been a widely used approach to measure and to analyze the interest of academic world on a certain topic. In this regard, this study intends to provide insight about the research on environmental-friendly electronic using the related literatures from the Thomson Reuters Web of Knowledge database during the period of 1980–2016. This study consists of two parts. In the first part, 7288 academic papers having the "environmental" and "electronic" phrases on "Title", "Abstract" or "Keywords" were retrieved and analyzed using bibliometric analysis methodology. These two adjectives has been selected on purpose since current studies on textual analysis indicate that phrase building most frequently starts after adjectives. In the second part of this work, Singular Value Decomposition (SVD) method, concept extraction and k-means clustering method was performed to gain more insight about the textual structure of the retrieved articles. Findings indicate that, approximately one third of publications were written by the authors addressing USA. It is also clarified that the topic was not in the agenda of researchers between 1980 and 1990. The number of publications on the area had significantly increased and had reached its peak in 2014. Text mining results showed that, the most important research focus was on "life-cycle" that was followed by "e-waste, sensor, recycling and solder" respectively.

1. Introduction

The growth of greenhouse gas emissions affects the entire world in a negative way. Therefore, developing novel green strategies and policies have been very significant factor of maintaining sustainability. Such strategies are expected to avoid companies from wasteful and environmentally harmful production and thereby canalize them to the resource efficient consumption alternatives.

In parallel to these sustainability issues, quality perspectives of consumers have significantly changed excessively. The concept of quality has evolved over the years. Formerly, the quality was approximated by the measures like: production volume, the cost-competitive advantages that make a firm superior to its competitors. In the early 2000s, quality of products, companies' quick response and adaptation of firms to the changes and innovativeness were the significant indicators showing the evolution in quality perception. However, this perspective has continued to evolve in the recent years. Nowadays, issues related to environmental considerations have become an increasingly important discussion topic and have taken the attention of policy-makers in the developed and developing countries. Since increasing the quality of product and process can enhance the organizations productivity, their market domination and profitability; environmental concerns have taken a considerable role in the strategies of the companies.

A number of policies regarding environmental aspects of countries had been announced by the European Union (EU) comprising directives on the limited use of harmful substances in electrical and electronic product and WEEE (waste electronics and electrical equipment) (Tseng et al., 2013). With the help of these directives which push manufacturers attention on minimizing the unwanted environmental impacts of electric and electronic products has been triggered. It has been recognized that, companies will be punished and either comprehensive or selective commercial sanctions will be employed when the companies launch new products to the market containing hazardous ingredients for the

* Tel.: +90 342 3172605; fax: +90 342 3604383.
E-mail address: durmusoglu@gantep.edu.tr.

http://dx.doi.org/10.1016/j.jclepro.2016.04.068
0959-6526/

public health. Besides, increase in conscious of consumers on environmentally-friendly products has positively affected the request of consumers (Kilbourne and Pickett, 2008). Therefore, producing environmentally-sensitive green products and using the environmental processes have been one of the strategically important targets of companies. In this regard, it has been necessary for companies to update their traditional production visions towards the suitable environmental and technology innovation policies. As a result of the fast depletion of the natural resources and increasing amount of various type of waste (industrial, municipal, agricultural, electronic and etc.) creating eco-friendly equipment and developing new techniques for recovering products and management of waste have drawn the researchers' attention to this topic (Gungor and Gupta, 1999).

Today, academic publications and patents are considered one of the most important resources of monitoring the research trends in academic studies (Murray, 2002; Tseng et al., 2007). By analyzing these documents, companies and researchers may get useful information about future technology without wasting time, money and effort and incorporate results with existing knowledge in further R&D activities (Yoon and Park, 2004). It is also important to use the correct methodology to save time, money and effort. In this regard, text mining has been an effective method to deal with large number of documents. Text mining is known to be a useful tool of processing unstructured text with machine support. Moreover, text mining can be helpful to detect the trends in large scale text documents (Miner et al., 2012). Text mining techniques have been widely used by different researchers for different purposes. For example (Daim et al., 2012) used text mining to create linkages between patenting patterns and the crucial events effecting wind energy technologies. Liew et al. (2014) have employed the methodology to identify the top priorities of chemical industries and their sustainability trends. Liu (2013) has performed "a visualization analysis" for terahertz technology. There are also some other studies identifying research trends using publication analysis (Daim et al., 2012, 2006; Tseng et al., 2007; Yoon and Park, 2004).

The aim of this study is to discover the recent trends of researchers and countries related to 'environmental' and 'electronic' between 1980 and January 2016. Bibliometric data were utilized in the characterization of the literature by journals, publication type, language, subject categories, countries, authors and citations. The detailed content analysis of the abstracts and titles of papers was performed by using several text mining methodologies. The documents related with environmental and electronic field were downloaded from Thomson Reuters Web of Knowledge (WoK) database (lists the all articles in the journals indexed by SCI-Soc-SCI Expanded-SCIE published by about 500 publishers). Topic term was selected as "environmental" and "electronic". It is known that the "topic term search" in WoK corresponds to the searches on: "Title", "Abstract", "Keywords" or "Keywords Plus®" fields within a record.

The keywords "environmental" and "electronic" are both adjectives. Previous studies have demonstrated that adjectives are good indicators of subjective, evaluative sentences (Hatzivassiloglou and Wiebe, 2000; Wiebe, 2000). It is remarkable to state that these two adjectives have been selected on purpose since phrases most frequently starts after adjectives and the most of English compound nouns are noun phrases that are described by the adjectives.

In the second part of this work, Singular Value Decomposition (SVD) method, concept extraction and k-means clustering method was performed to gain more insight about the textual structure of the retrieved articles. Frequencies of the words in retrieved articles were also analyzed. The remaining sections of this paper are organized as follows. Section 2 describes the bibliometric and text mining methodologies which is the basis of the analysis. The data employed in this work and the results of bibliometric and text mining analysis are presented in Section 3. Finally, Section 4 reveals the conclusion and discusses future research opportunities.

2. Data retrieval and methodologies

The documents related with environmental and electronic field were retrieved from Thomson Reuters Web of Knowledge (WoK) database (which lists the all articles in the journals indexed by SCI-Soc-SCI Expanded-SCIE published by about 500 publishers). Topic terms were selected as "environmental" and "electronic". They are both adjectives. Previous studies have demonstrated that adjectives are good indicators of subjective, evaluative sentences (Hatzivassiloglou and Wiebe, 2000; Wiebe, 2000). It is remarkable to state that these two adjectives have been selected on purpose, since phrases can be extracted during the process of consecutive word reading and they start most frequently after adjectives (Klinov and Mouromtsev, 2013). Some of the articles that are out of scope have been eliminated after a manual scan (such as the ones including "electronic data", "electronic customers"). Since the purpose of this study is to identify the research on environmental-friendly and electronic "products/devices/equipment/processes/tools/methods and etc", the words used after the given adjective "electronic" in the retrieved articles is illustrated in Table 1.

Subsequent to the query to detect the related articles, two fundamental sections of the papers (title and abstracts of the papers) were downloaded. Complete bodies of papers have not been included since adding them would dramatically increase the time for an automated analysis (due to the tables, images, references, etc.). The study was performed in two stages. The purpose of the first stage was to provide information regarding the characterization of the environmental electronic related studies by the use of bibliometric analysis. For this analysis, all extracted records (7288 papers) were used. In the second stage, text mining analysis was performed for the papers which have accessible abstracts. The number of records was decreased to 7145 after elimination of those papers without an abstract.

2.1. Bibliometric analysis

The term "bibliometrics" is the combination of words "biblio" that refers "books" and "metrics" that refers to "measurement" was first coined by Allan Pritchard in 1969 (Norton, 2000). It is a method that uses both qualitative and quantitative methods to organize, describe, analyze and explore the documents. Bibliometric analysis has been one of the most important methods that are used in measuring scientific progress in the field of information science (Raan, 2005). It helps to clarify the top most active authors, performance of countries, the most cited works, the most trend topic that studied by researchers, the frequent keywords that used by authors and etc. (Daim et al., 2006).

Some papers employing bibliometric analyses on environmental science have been published in the recent years. For example, 8244 publications were analyzed by Du et al. (2012) to discover the characteristics of energy efficiency literature for the years between 1991 and 2010. Similarly, the trends in carbon market from 1992 to 2011 were described by Du et al. (2015), according to their findings, the focus of carbon market related studies were "climate change" and "carbon emissions". Bibliometric analysis and topic modeling approach used by Jiang et al. (2016) to establish academic concerns of the Three Gorges Project (TGP) which is world's largest hydropower project. In the study 8240 abstracts of Chinese articles from 2001 to 2013 in aforementioned topic were used as inputs (Jiang et al., 2016). By using samples of 113,468 publications in a 20-year period the global environmental assessment literature trends were discovered by Li and Zhao (2015).

Table 1
The nouns that is following the adjectives: environmental and electronic.

Environmental + nouns	Counts	Environmental + nouns	Counts	Electronic + nouns	Counts	Electronic + nouns	Counts
Conditions	464	concerns	77	devices	349	industry	60
Factors	257	issues	77	equipment	339	control	56
Protection	207	changes	76	products	199	circuits	56
Impact	194	scanning	76	structure	175	medical	54
Effects	192	applications	65	properties	167	structures	49
Monitoring	174	friendly	56	waste	174	states	48
Pollution	139	regulations	56	components	142	sensors	41
Impacts	102	remediation	52	nose	103	system	40
Problems	101	control	50	packaging	86	conductivity	36
Stability	98	temperature	50	systems	81	applications	35

According to their predictions, topic of environmental assessment in the future will continue to grow and the studies will be doubled by the year 2019. In another study, Valipour (2014) performed a comprehensive study on previous works to arouse awareness of drainage, waterlogging, and salinity and to avoid trial and error policies in these fields. Academic output productivity and its influence factors in environmental sciences and ecology are measured by Dragos and Dragos (2013). 92 countries were ranked by Dragos and Dragos (2013) with the utilization of bibliometric analysis and Environmental Performance Index. One of the other focus areas of researchers on environmental science was biological invasions. Qiu and Chen (2009) employed bibliometric analysis to 3323 articles to reveal the trends of publications. According to results of Qiu and Chen (2009), the research collaborations of US between continents, universities, and countries help US to become the most productive country in biological invasions research.

Classic bibliometric studies generally take metadata into consideration (such as year of publication, language, citation, author(s), etc). Since the abstracts of papers are representative of the full papers, in addition to metadata this paper analyzed the abstracts of the papers with their titles. The analysis was enriched with the inclusion of abstracts, titles and the metadata and it was expected that it will be more explanatory and comprehensive than traditional bibliometric analysis. The keywords have not been included in the analysis while the abstracts could likely comprise the keywords.

2.2. Text mining

A huge number of data is available in web environment and can be collected in electronic form and it is known that 85–90% of these documents are not in a structured format (Delen and Crossland, 2008). Text mining plays a critical role for the analysis of unstructured data. Text mining involves various applications of different areas such as data mining, information retrieval, natural language processing and information extraction (Sumathy and Chidambaram, 2013).

Text mining approaches have been widely applied for environmental analysis. Altaweel and Bone (2012) examined the spatial patterns of public media on the topic of environmental reporting of water issues. The term frequency and principal component analysis methods were used in the evaluation. In another study, global map of science and text mining methods were used to extract valuable information from waste recycling related research between 2005 and 2010 (Garechana et al., 2012). There are also hybrid methodologies that are combining two or methods. As an example, solar lighting devices related patents were analyzed by integrating text mining, technology road mapping and quality function deployment methods (Jin et al., 2015). Climate change related literature of "conservative think-tanks" was analyzed by Boussalis and Coan (2016) to recognize the signal of climate change. They employed the latent Dirichlet allocation (LDA) method to determine the topic distribution. The analysis about TGP (Jiang et al., 2016) was also utilized LDA method while extracting topics. Topic modeling is important method used in text mining. There are different "topic modeling techniques" used in literature. While some researchers used LDA (Boussalis and Coan, 2016; Jiang et al., 2016) method, some of them employed "auto correlation maps" (Sunikka and Bragge, 2012). Brainstorming is another method used in topic modeling (Jourdan et al., 2008). On the other hand, clustering with expectation maximization algorithm (Delen and Crossland, 2008) has been another method that was preferred for topic extraction purposes.

The present study attempted to extract latent topics from environmental electronic related documents. For this purpose, SVD and k-means clustering methods were employed to identify trends and clusters of related research topics.

The text mining steps defined by Cerrito (2009) was also used in this study. These steps can be summarized as follows:

- **Determination purpose of study**: the aim of study and research area of interest is determined.
- **Collection of an initial set of data**: this data may contain web pages, textual documents, short notes, e-mails, patents and XML files.
- **Preparing the data**: raw data transformed into structured format so that each data, originally in natural language format, are defined as one token.
- **Converting organized documents into a term by document matrix**: the nominal data are concatenated into a text string so each identifier is represented by one text.
- **Simplify term by document matrix**: generating stop terms or including terms together with synonyms and specific phrases, stemming which aims reducing words and reaching root form of word.
- **Normalization of matrix**: raw frequency of input documents normalized by using different normalization methods to achieve best representation of a matrix and extraction of accurate pattern. The most common methods are log frequency, binary frequency and term frequency–inverse document frequency (TFIDF).
- **Reducing the dimensionality of matrix**: to reach maximum variability degree between terms and documents. The input matrix can be reduced by using different dimension reduction techniques such as, SVD, principal component analysis (PCA), etc.
- **Extracting knowledge**: using different text mining methods such as, prediction (regression, classification etc.), trend analysis, clustering (e.g. outlier analysis), association (link analysis and affinity analysis) novel patterns and various knowledge are obtained (Cerrito, 2009; Delen and Crossland, 2008; Miner et al., 2012).

For this study, subsequent to the query and data collection processes, papers having missing data were eliminated. Afterward, term-by-document matrix, indicating the numerical representation of the textual data, was generated. Terms were extracted after parsing the documents. Term-by-document matrix had a spreadsheet-like structure. The dimension of this matrix is equal to the number of documents multiplied by the number of terms. We have tried to simplify the matrix since the large term-by-document matrix (size increases with the size of the data) brings huge computation costs. Removal of the "stop words" is a widely used simplification method. Certain parts of English language, like conjunctions ("for", "or") or the word "the" are meaningless to a topic analysis. These terms are called stop words and can be removed from the text. In this regard, the stop words were also excluded in this study.

By using synonyms of the words, some replacement was also performed to decrease the variation of the words. Then, matrix was normalized using TFIDF (TF stands for term frequency of term i in document j and IDF stands for inverse document frequency of term i) method to represent words semantically and reflect relative frequency of words occurrence on the document rational. Thereby, instead of number of terms themselves, proportional importance of the words was used as the identifiers of documents. Weighting factor was assigned as "0" if that word occurred in all documents and assigned as "1" if a word occurred in only one document (Delen and Crossland, 2008). The calculation of TFIDF (Uğuz, 2011) was performed as in Eq. (1):

$$a_{(i,j)} = tf_{ij} idf_i = tf_{ij} \times \log_2\left(\frac{N}{df_i}\right) \quad (1)$$

where a_{ij} is the weight of term i in document j. According to Eq. (1), df_i is the document frequency for the ith word (the number of documents that include this word), tf_{ij} is the term frequency of term i in document j and N is the number of all of the documents.

As a consequence of this analysis, important words were extracted. Despite using stop-word-list the term by document matrix had still lots of words. SVD technique was used to decrease the dimension of matrix to a convenient size. The main aim was representing largest degree of variability with such an input (document × term) matrix having minimum dimensional space where each consecutive dimension was the representation of the possible largest degree of variability between documents and terms. SVD technique will rearrange term by document matrix. This matrix consists of m words (terms) and n documents (m × n matrix) where the values in the matrix represent relative frequencies such as TFIDF or word frequencies. In this method, singular value was computed by decomposing the original term by document matrix X in a product of three other matrices U, Σ, and V^t. Equation of SVD as follow:

$$X = U \times \Sigma \times V^t \quad (2)$$

where matrix X is m × n matrix and its entries are dataset components. According to Eq. (2) the SVD computes the m × r orthogonal matrix U, n × r orthogonal matrix V^t (where V^t stands for the transposed **V** matrix) and (r × r) diagonal matrix Σ (Song et al., 2013). The Σ matrix entries are non-negative values and sorted with descending order and they are singular values of X. Details of the mathematics underlying the SVD method can be found in Deerwester et al. (1990). Generally, it is accepted that two or three most obvious dimensions are used to reveal most of variability between terms and documents.

In this paper, SVD technique was utilized to have more compact matrix. In the last step, by using the extracted keywords and components from SVD analysis k-means clustering analysis was performed to identify trends of environmental electronic related studies. The distance matrix that was used in k-means clustering was Euclidian distance matrix. Euclidean distance is the distance between two documents. For any two n-feature samples, say $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$, their Euclidean distance is measured as Eq. (3) (Su, 2011):

$$\mathrm{dist}(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} \quad (3)$$

3. Results and discussion

3.1. Results of bibliometric analysis

A bibliometric analysis was performed to reveal the patterns in the articles which issued "environmental" and "electronic" as the topic. Analysis was constructed on the following dimensions: language, type, publication years, countries, subject categories, citations of distribution of publications. Details of the analysis for these dimensions were presented in the following sub-sections.

3.1.1. Type and language of publications

The published papers were written in sixteen different languages. The vast majority of the documents (97.77%) were written in English. It is followed by the publications which were written in Chinese, German, Portuguese, and French respectively. Among those 7288 articles; 4472 were journal articles (61.361% of the total), 2492 were meeting abstracts (34.193% of the total) and the remaining was: proceeding papers, reviews, editorial materials, book chapters, letters, notes, bibliographies, corrections and discussions.

3.1.2. Publication analysis by the years

The annual number of academic papers published on "environmental" and "electronic" topics are as shown in Fig. 1. The number of academic documents was between 1 and 7 from 1980 to 1990. It can be concluded that the topic was not in the agenda of researchers between 1980 and 1990. In the early 1990s, this number was rather modest-ranging from 42 to 87 papers per year. In 1997 (158 papers), the number of publications had almost doubled compared the previous year (87 papers). The reason of this increase can be linked with "Kyoto protocol" which was legally binding agreement aimed at reducing greenhouse gases emissions and preventing human-made global warming. The number of publications was fluctuated from 1996 to 2006, but general trend was upwards. From the year 2006 to 2010 the number of publications continued to fluctuate and the growth was accelerated. The number of publications had significantly increased and had reached its peak in 2014 (734 papers). In 2015 the number of publications (620 papers) decreased when compared to previous year. In 2016 which is not shown in Fig. 1; there was 17 papers published in January. If the year 2016 continues with the average of 17 papers/month, around 600 papers are expected to be published.

3.1.3. Countries performance

Authors address 115 different countries (58% of the all countries) in the retrieved articles. Table 2 presents top 30 countries (their corresponding percentage) that are addressed on the relevant topic. Over the 35 years, the United States, People's Republic of China (PRC), Germany, England and Italy constituted together the top 5 countries that the papers originated. Obviously, it was seen that nearly one third of publications belong to researchers from USA. These results correspondence to 'polluter pays principle'

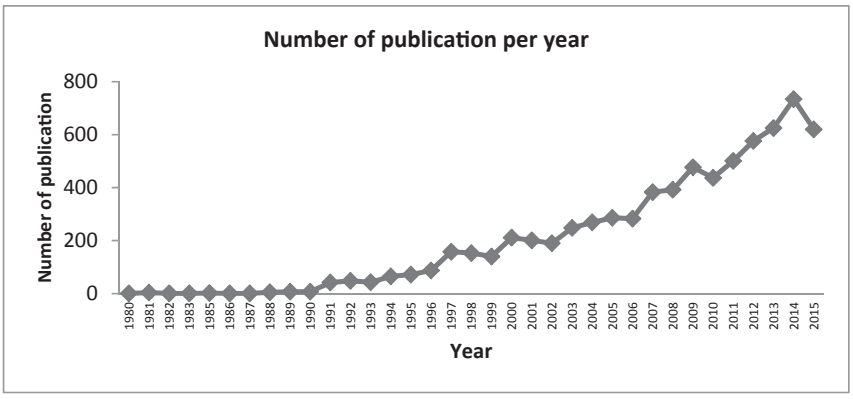

Fig. 1. Number of published academic papers on "environmental electronic" topic, 1980–2015.

which is an environmental policy principle that indicates "the payment of pollution should be made by those who cause it" (Tobey and Smets, 1996). Since developed countries are the largest producers of pollution (total U.S. greenhouse gas emissions have increased by 7.7 percent from 1990 to 2014 – US EPA) and they are expected to generate more significant solutions. In this regard, the relatively higher number of academic studies from those developed countries should not be something unexpected.

PRC had the second place as stated in Table 2. It could be stated that Chinese scientists had studied on the relatively poor environmental performance of PRC. It has been questionable whether these academic studies had a real impact on policy makers to put the findings into practice. According to Shanghai Academy of Social Sciences, Chinese cities, Guangzhou, Beijing and Shanghai were chosen "barely suitable" for living among 40 major cities around the world because of their intense environmental pollution (Kostka, 2014). According to forecast of "Solving the E-waste Problem (StEP)" the volume of e-waste in the world will be 72 million tones or grown by a third by 2017 (around 40 million tons per year). While PRC will be leading producer of it; will be followed by USA with the amount of 12.2 million and 11 million e-waste respectively.

Germany had been the third mostly addressed country among the relevant publications. Next to the Germany, there were 13 other EU (European Union) members in the list: England, Italy, France, Spain, Netherlands, Sweden, Denmark, Poland, Belgium, Greece, Romania, Finland and Austria. In total, 2631 papers addressed 14 EU members (more than one third (33.456%) of the total publications).

3.1.4. Distribution of publications by subjects

There were different research areas including engineering, physics, computer science, robotics, energy and etc. on "environmental" and "electronic" fields. The 7288 academic papers on environmental electronic field comprised more than 100 subject categories. Table 2 represents distribution of top fifteen research areas. More than one third of the publications were published in engineering field (34.399% of total records), followed by chemistry (19.018% of total records), and environmental sciences/ecology (14.476% of total records).

It is also remarkable to state that "Business Economy" subject appeared in top twenty research areas which showed that studies on environmental electronic field focused on the importance of environmental science on sustainable development for the economic growth of the countries and companies.

3.1.5. An analysis of citations

In bibliometric studies, citations have established themselves as an indirect indicator for a paper's quality and its usefulness in particular (Garfield, 1979). The provided graph in Fig. 2 shows the citations of the 7288 academic papers by the years. Citation analysis was also performed by using Thomson Reuters Web of Knowledge (WoK) database (which lists the all articles in the journals indexed by SCI-Soc-SCI Expanded-SCIE published by about 500 publishers).

According to the graph, during the 1980–1990 decade, the number of citations varied between 0 and 54. Numbers of citations in year 1997 had increased enormously compared to previous year. One of the reasons of this growth can be adaptation of Kyoto protocol which sets binding targets for industrialized countries and

Table 2
Top 30 countries ranked by numbers of papers published in (1980–2016) on "environmental electronic" topic.

Rank	Country	Records	% of records	Rank	Country	Records	% of records
1	USA	2204	30.241	16	Sweden	120	1.647
2	PRC	1048	14.380	17	Switzerland	119	1.633
3	Germany	540	7.409	18	Denmark	80	1.098
4	England	467	6.408	19	Poland	80	1.098
5	Italy	351	4.816	20	Belgium	78	1.070
6	France	300	4.116	21	Romania	67	0.919
7	Japan	288	3.952	22	Greece	66	0.906
8	Canada	255	3.499	23	Scotland	66	0.906
9	Australia	241	3.307	24	Iran	64	0.878
10	Spain	241	3.307	25	Finland	60	0.823
11	India	201	2.758	26	Malaysia	55	0.755
12	Netherlands	181	2.484	27	Russia	54	0.741
13	Taiwan	180	2.470	28	Austria	52	0.714
14	South Korea	174	2.387	29	Israel	52	0.714
15	Brazil	128	1.756	30	Turkey	52	0.714

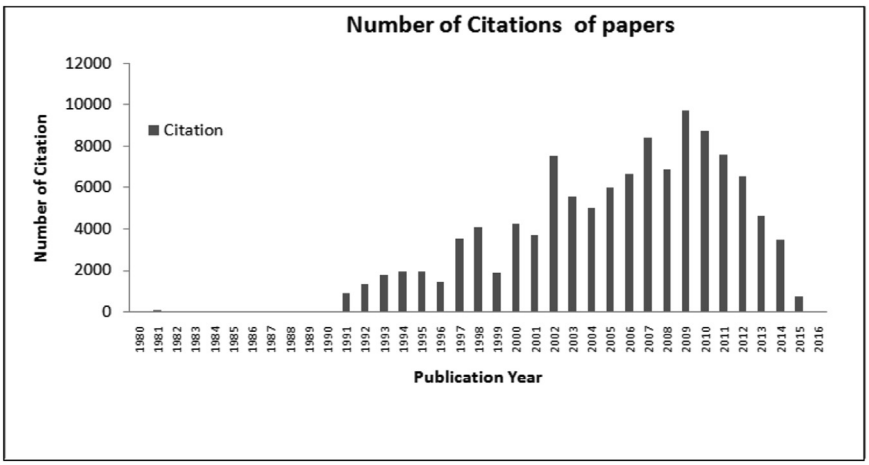

Fig. 2. Number of citations of published academic papers on "environmental electronic" topic, 1980–2016.

European community for reducing greenhouse gas (GHG) emissions (Hildén, 2011). The number of citations dropped from 4065 to 1855 between 1998 and 1999; subsequently increased to 4232 in 2000. Although there were fluctuations in the numbers of citations they had tended to increase until 2009. However, after 2009 there was continuous decrease in the number of citations.

3.2. Results of text mining

3.2.1. Word importance

Definitely, words that occur with greater frequency in a document are assumed to be better descriptors of the contents of those documents. However, since the ratio of frequencies of two concepts may not indicate the ratio of importance for those two words. Therefore using just the frequency counts of words may have misleading results in a textual analysis. In this regard, using "word importance scores" may help us to gain much more insight about the importance of concepts. For this study, the abstracts and title of 7145 documents yielded 47,646 words and 573 keywords. Top 20 of these keywords and their importance score were shown in Table 4. Subsequent to normalization of "term by document matrix" (described with details in section 2.2), the equation (1) was used to calculate the weights of the terms.

Findings indicate that, "life-cycle" was the most important word that was followed by "e-waste", "sensor", "recycling" and "solder" respectively. Life-cycle related subjects such as: life-cycle assessment (an analysis tool to evaluate the environmental impact that was caused by a product or process from the extraction of raw materials up to the waste disposal) was the most popular methodology used by the researchers. The top five words showed that, studies having topic "environmental" and "electronic" gave a special attention to the issues about environmental pollution caused by e-waste (electrical and electronic waste) disposal and recycling. The researchers were aware of the impact of e-waste which was accepted as the fastest growing waste type in the world (Lundgren, 2012). Energy harvesting which had the potential to reduce to need for battery power or wiring was also accepted an important subject by researchers. Dealing with sensor technology and solar power generation by using thin film solar cell were also attracting attention of researchers. The impacts of designing these systems on human health were not ignored in those studies. Moreover, researchers focused on the issue that developing sustainable methods, models or technologies in industrial waste processing. It is also understood from Table 3 that "solder" was important concept and researchers have worked on finding solution for soldering problem. In electronics assembly, lead usage is very popular because of the characteristics of lead, such as high-conductivity, low melting point, high-fatigue resistance and high-strength ductility which give extremely good solder connections (Johler, 2002). But the effect of lead on human health increase human awareness and generating new technology on electronic soldering is still searching by technology developers. Another research topic that was of interest to scientists in the world involved managing and processing hazardous chemical materials and recycling their wastes from environment.

Table 3
Distribution of publications on environmental and electronic by top 15 subjects (1980–2016).

Rank	Research areas	Records	% of records
1	Engineering	2507	34.399
2	Chemistry	1386	19.018
3	Environmental Sciences Ecology	1055	14.476
4	Materials Science	985	13.515
5	Physics	949	13.021
6	Computer Science	565	7.752
7	Science Technology Other Topics	450	6.175
8	Instruments Instrumentation	346	4.748
9	Optics	301	4.130
10	Energy Fuels	256	3.513
11	Public Environmental Occupational Health	249	3.417
12	Electrochemistry	210	2.881
13	Business Economics	190	2.607
14	Metallurgy Metallurgical Engineering	145	1.990
15	Agriculture	141	1.935

By using word importance table one can extract different information about countries situation. Fig. 3 illustrates the perspective of countries in terms of "life-cycle". As it can be seen from Fig. 3, the authors addressing Malaysia, Ireland and Thailand have relatively more focus on "life-cycle" when compared to others.

3.2.2. Dimension reduction

According to SVD method there are 23 components (concepts) for the widely used words in the relevant articles. First category explains the 13.6511% of the words. Second component explains 6.7663% and the third component is 5.7151%. The first two components, as illustrated in Fig. 4, have the highest percentage of singular value explanation. In clustering analysis section of this study, these 2 components were used to extract more valuable information from the text.

Table 4
Word importance score of phrases detected.

Rank	Word	% of importance	Freq.	Rank	Word	% of importance	Freq.
1	life-cycle	100.000	2508	11	oxygen	69.636	1408
2	e-waste	94.058	1021	12	molecule	68.598	1566
3	sensor	92.458	2371	13	product	68.559	4269
4	recycling	90.087	1671	14	temperature	68.545	2319
5	solder	89.686	943	15	thermal	66.002	1258
6	film	80.033	939	16	energy	64.226	2554
7	waste	78.193	1740	17	material	64.063	3605
8	power	75.939	1513	18	solar	63.648	1037
9	health	74.568	1874	19	review	63.569	1602
10	metal	71.7652	1692	20	management	61.7119	1584

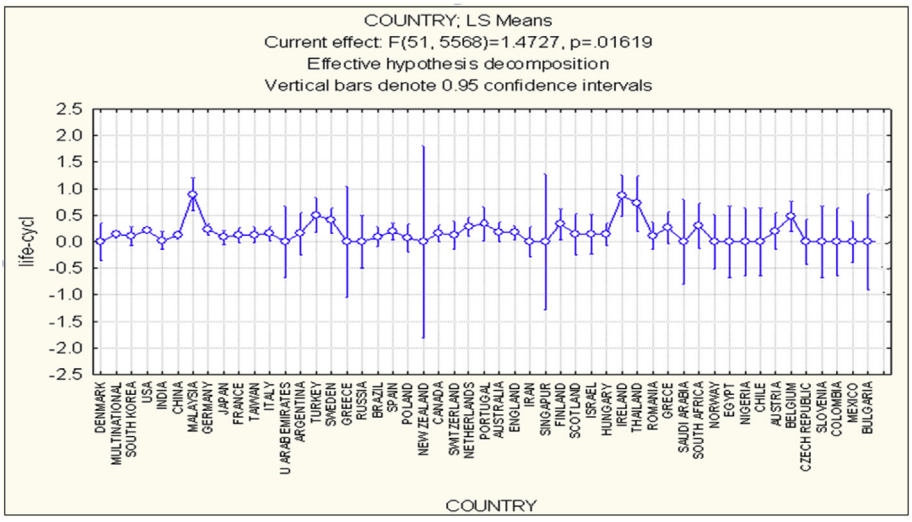

Fig. 3. Studies on "life-cycle" by the different countries.

3.2.3. Concept extraction

Concept extraction (putting similar documents into the same concept, by identifying common concepts) helps scientists to access the information in an easier way. Therefore, "concept extraction" has been also a preferable analysis for this study. The collected papers yielded 47,646 words after using stop-word list, 47,073 words were eliminated. Dimension of document-word matrix was reduced by SVD analysis and a list of most frequent 573 terms was divided into 23 components and 10 of them (having the highest explanation power) were shown in Table 4. As presented in Table 4, the first concept explains 13.651% of the research area. This concept simply, focused on the "systematic and effective improvement of

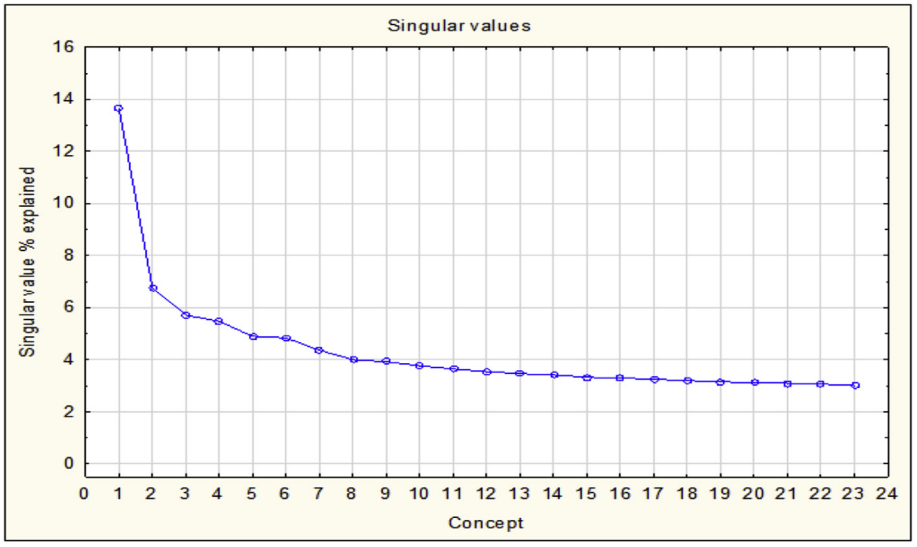

Fig. 4. Explanation power of concepts obtained via singular value decomposition method.

product and process in high-tech manufacturing industries". In Concept 2, "life-cycle assessment of waste/e-waste recycling technologies and their health risks" are the focus of the research. According to Concept 3, researchers study on the "identification of environmental risk factors of electronic products and their effects on children and patients". Thermal management of electronic device packaging and reliability tests were taken into consideration in Concept 4. The one noticeable result is that "China" was one of the important words of Concept 9. In the same concept, publications have also drawn attention on the chip, solder and packaging technologies (Table 5).

3.2.4. Cluster generation

Clusters are usually identified as the "topics," and learning these topics extends to newer sub-domain topic analyses in the publication analysis. With this purpose; subsequent to dimensionality reduction through SVD method, clustering analysis was performed in order to put the similar papers into similar clusters without having a pre-defined set of classes. K-means clustering algorithm was used in this study since it has been the most widely used methodology to understand the clusters of data sets (Dhillon et al., 2004; Eisa, 2012; Kim et al., 2015; Macqueen, 1967). In k-means clustering method, number of clusters k must be decided before the calculation of the distance metrics. In this study, Euclidean distance metric was used. The proper number of cluster is important to represent the whole dataset. Unfortunately, there is no optimal way in the determination of cluster numbers (Delen and Crossland, 2008). In this study, this decision was made by considering the results of the F-measures of the experiments. F-measure (harmonic mean of precision and recall) was used on generated clusters of different size (Rosell et al., 2004). Since K-Means clustering is like a multivariate analysis of variance in which the groups are not known, it focuses on reducing the within-groups sum of squares. Therefore larger F-ratios have been preferable while largest F-value indicates that clusters are different from each other (Bergman, 2014). In this study, after various trials and errors, the number of clusters varied between 2 and 21 and finally, the number of clusters was decided to be 8 which had the largest F value as shown in Table 6.

These eight clusters and their keywords with counts number and percent frequency of papers were illustrated in Table 7.

The descriptive labels were created by looking at the keywords as presented in Table 7. The most popular subject cluster includes the articles written on "product life cycle assessment" (cluster number 6). Almost 50% of the studies had been focused on this issue. Note that, this result was similar with word importance result in which the most important word was extracted as "life-cycle". It is understood from both concept extraction and clustering results that environmental electronic related studies were focused on the issues on life-cycle. Cluster number 1 was the second popular cluster with 16.916% which was included the documents having topic on "energy saving in manufacturing industry". Reducing the amounts of energy requirement of product and services by generating innovative technology was the main of energy efficiency studies. Moreover, usage of renewable energy resources in the design and production process resulting significant economic advantages, reducing the effects of global warming and increasing energy security. The paper in cluster 4 and cluster 5 constitutes approximately 7.577% and 7.737%, respectively. The result of cluster 4 (e-waste recycling) was also similar with the concept results. In this regard it can be concluded that "e-waste recycling problems" were found one of the most important topics studied by researchers. Moreover, cluster 5 indicates that "generating new models for effective investing methodology on environmental electronic issues" has also taken great attention. "Monitoring and control of power sensor systems" was another research subject (cluster number 7) that was issued in 195 studies.

Table 6
Number of clusters and corresponding F values.

Number of clusters	F value	Number of clusters	F value
2	665,7879	12	1802,105
3	892,6260	13	1800,293
4	976,2861	14	1668,030
5	1021,152	15	1633,425
6	1229,344	16	1577,576
7	1540,883	17	1687,742
8	1834,472	18	1673,883
9	1508,803	19	1428,109
10	1521,156	20	1389,578
11	1593,119	21	1380,715

4. Conclusions

This paper focused on the "environmental" and "electronic" related papers published in the indexed journals in the last thirty-six years (1980–2016) to provide an overall picture of the studies on the area. For document types, articles were the most common publication type (about 61% of all articles), followed by meeting abstracts (about 34% of all articles). The most active country with nearly one third of publications was USA. People's Republic of China (PRC) was the second mostly addressed country with 1048 publications. This fact may be linked to the relatively poor environmental performance of PRC. Germany, England and Italy made up the remainder of the

Table 5
The most important 10 words according to top 15 components.

Concept number	% of explanation	Most important 10 words
1	13.651	systematic, product, material, database, technical, process, high, develop, effective, methodology
2	6.766	recycling, management, health, waste, e-waste, life-cycle, risk, information, impact, assessment
3	5.715	database, health, children, evidence, conclusion, patients, review, identify, factor, risk
4	5.469	power, sensor, package, design, test, solder, thermal, systematic, device, reliability
5	4.898	sensor, molecule, energy, management, detect, monitor, systematic, network, sensitivity, information
6	4.837	sensor, detect, concentration, monitor, e-waste, sensitivity, nose, sample, measure, gas
7	4.364	film, review, material, polymer, oxygen, synthesis, carbon, application, properties, health
8	3.999	sensor, solder, detect, life-cycle, chemical, sensitivity, nose, gas, monitor, process
9	3.924	e-waste, solder, package, china, alloys, molecule, information, countries, health, chip
10	3.767	metal, review, circuit, board, printed, molecule, children, power, energy, chemical
11	3.647	film, recycling, waste, recovery, thin, optical, circuit, electric, patients, spectrum
12	3.545	life-cycle, e-waste, sensor, exposure, film, children, sensitivity, polymer, chain, molecule
13	3.471	life-cycle, optical, e-waste, spectrum, health, absorption, nm, laser, assessment, state
14	3.418	oxygen, sensor, life-cycle, network, surface, metal, site, mobile, film, temperature
15	3.329	film, health, thin, air, chain, supply, risk, emission, manufacturing, exposure

Table 7
Descriptive terms and clusters definitions.

#	Descriptive terms	Percent (%)	Clusters definitions
1	design + technology + package + increase + require + component + reliable + manufacture + industry + energy	16.916	Designing requirements and reliability for saving energy in manufacturing industry
2	integrate + provide + inform + potential + current + need + various + field + new + generate + recent + measurement + condition + time + include + detect + wide + ever + operation + data + change + temperature	4.464	Generating new information for measuring temperature change
3	surface + metal + factor + differ + order + research + relation + describe + object + determine + properties + perform + use + active + analysis + chemic + mechanic + technique + present + function + pollution + level + test + report + state + experiment + organ + health + approach + work + base	5.958	Experimental approach for testing the effects of chemical pollutants level in health organizations
4	e-waste + equipment + recycle + waste + electrical	7.577	Waste electrical equipment recycling
5	structure + study + investing + method + effect + model	7.737	Modeling effective investing methodology
6	product + life-cycle	48.932	Product life-cycle assessment
7	device + control + system + monitor + application + power + sensor	3.468	Monitoring and control of power sensor systems

top five most active countries on environmental electronic field. EU originated publications encompassed totally the 33.456% of the total records. It has been noticeable that the rank of Japan (9th) has been relatively lower when compared to the overall research performance of the country (between 4th and 6th).

The number of academic documents was between 1 and 7 from 1980 to 1990. It can be concluded that the topic was not in the agenda of researchers between 1980 and 1990. In the early 1990s, this number was rather modest-ranging from 42 to 87 papers per year. In 1997 (158 papers), the number of publications had almost doubled compared the previous year (87 papers). The reason of this increment was the "Kyoto protocol" which was legally binding agreement aimed at reducing industrialized countries' greenhouse gases emissions and preventing human-made global warming. The number of publications was fluctuated from 1996 to 2006, but general trend was upwards. From the year 2006 to 2010 the number of publications continued to fluctuate and the growth was accelerated. The number of publications had significantly increased and had reached its peak in 2014 (734 papers).

According to subject categories, engineering field was found hottest subject category followed by chemistry and environmental science ecology. In accordance with the citation results of bibliometric analysis, until year 2009 the number of citation that paper received was increased but between 2010 and 2016 there was continuous decrease in the number of citations. The reason of this situation was the papers published in after 2010 were relatively new those of papers that published before 2010.

In the second stage of the analysis, some text mining approaches were applied using SVD dimension reduction method. As a typical analysis of unstructured texts; concept extraction and k-means clustering methods were implemented. Due to elimination of unavailable abstracts of record, the number of records was decreased to 7145 papers in the relevant stage. 7145 environmental electronic studies' abstracts and titles yielded 47,646 words and 573 keywords. The most important word was found as "life-cycle". It was followed by "e-waste, sensor, recycling and solder" respectively. "It has been discovered that; life-cycle" related issues has taken more attention by the authors addressing Malaysia, Ireland and Thailand when compared to other countries.

The concept extraction results showed that "systematic and effective improvement of product and process in high-tech manufacturing industries" was the most focused topic. "waste/e-waste recycling" related topics were also found important. Environmental risk factors of electronic products had also been studied by researchers. It was detected that the studies about "China" has been a specific research area under the interest.

The present study can be also a valuable source for the people who have research interest on environmental and electronics. Delen and Crossland (2008) state that "*for the typical Ph.D. student, trying to find an appropriate dissertation topic can be a formidable task. Even if he or she has a good mentor as a dissertation chair, finding an appropriate topic, one that is both "new enough" and yet still considered "adding to an established body of knowledge" by being supported by the existing literature, can be difficult at best*". In this regard, findings of this article are expected to be assistive to check specific research areas.

Present paper expected to identify the determinant factors and widely focused subjects of academic world on environmental issues based on the electronics. Current paper used the publications from the start year of Thomson Reuters Web of Knowledge database to the extent that permitted.

A further analysis can be performed by the use of patent documents. In addition to that, environmental status (GHG emission, production and disposal of wastes, etc.) of the countries (countries with the most contribution in the topic of this study) can be analyzed to show whether academic investigations are effected from governmental policies or not.

References

Altaweel, M., Bone, C., 2012. Applying content analysis for investigating the reporting of water issues. Comput. Environ. Urban Syst. 36, 599–613. Special Issue Advances in Geocomputation.
Bergman, J., 2014. Practical Multivariate Analysis, fifth ed. PB-Chapman & Hall/CRC, Boca Raton, Florida.
Boussalis, C., Coan, T.G., 2016. Text-mining the signals of climate change doubt. Glob. Environ. Change 36, 89–100.
Cerrito, P., 2009. Text Mining Techniques for Healthcare Provider Quality Determination: Methods for Rank Comparisons, first ed. Medical Information Science Reference, Hershey, PA.

Daim, T., Iskin, I., Li, X., Zielsdorff, C., Bayraktaroglu, A.E., Dereli, T., Durmusoglu, A., 2012. Patent analysis of wind energy technology using the patent alert system. World Pat. Inf. 34, 37–47.
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Change 73, 981–1012.
Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R., 1990. Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. 41, 391–407.
Delen, D., Crossland, M.D., 2008. Seeding the survey and analysis of research literature with text mining. Expert Syst. Appl. 34, 1707–1720.
Dhillon, I., Kogan, J., Nicholas, C., 2004. Feature selection and document clustering. In: Berry, M.W. (Ed.), Survey of Text Mining. Springer, New York, pp. 73–100.
Dragos, C.M., Dragos, S.L., 2013. Bibliometric approach of factors affecting scientific productivity in environmental sciences and ecology. Sci. Total Environ. 449, 184–188.
Du, H., Li, B., Brown, M.A., Mao, G., Rameezdeen, R., Chen, H., 2015. Expanding and shifting trends in carbon market research: a quantitative bibliometric study. J. Clean. Prod. 103, 104–111. Carbon Emissions Reduction: Policies, Technologies, Monitoring, Assessment and Modeling.
Du, H., Wei, L., Brown, M.A., Wang, Y., Shi, Z., 2012. A bibliometric analysis of recent energy efficiency literatures: an expanding and shifting focus. Energy Effic. 6, 177–190.
Eisa, H., 2012. Text clustering on latent semantic indexing with particle swarm optimization (PSO) algorithm. Int. J. Phys. Sci. 7.
Garechana, G., Rio, R., Cilleruelo, E., Gavilanes, J., 2012. Tracking the evolution of waste recycling research using overlay maps of science. Waste Manag. 32, 1069–1074.
Garfield, E., 1979. Is citation analysis a legitimate evaluation tool? Scientometrics 1, 359–375.
Gungor, A., Gupta, S.M., 1999. Issues in environmentally conscious manufacturing and product recovery: a survey. Comput. Ind. Eng. 36, 811–853.
Hatzivassiloglou, V., Wiebe, J.M., 2000. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. Association for Computational Linguistics, pp. 299–305.
Hildén, M., 2011. The evolution of climate policies-the role of learning and evaluations. J. Clean. Prod. 19, 1798–1811. Promoting Transformation towards Sustainable Consumption and Production in a Resource and Energy Intensive Economy – the Case of Finland.
Jiang, H., Qiang, M., Lin, P., 2016. Finding academic concerns of the Three Gorges project based on a topic modeling approach. Ecol. Indic. 60, 693–701.
Jin, G., Jeong, Y., Yoon, B., 2015. Technology-driven roadmaps for identifying new product/market opportunities: use of text mining and quality function deployment. Adv. Eng. Inf. 29, 126–138.
Johler, W., 2002. Impact of lead-free soldering processes on the performance of signal relay contacts. In: Proceedings of the Forty-Eighth IEEE Holm Conference on Electrical Contacts, pp. 43–55.
Jourdan, Z., Rainer, R.K., Marshall, T.E., 2008. Business intelligence: an analysis of the literature. Inf. Syst. Manag. 25, 121–131.
Kilbourne, W., Pickett, G., 2008. How materialism affects environmental beliefs, concern, and environmentally responsible behavior. J. Bus. Res. 61, 885–893.
Kim, G.J., Park, Jang, D.S., 2015. Technology forecasting using topic-based patent analysis. J. Sci. Ind. Res. 265–270.
Klinov, P., Mouromtsev, D., 2013. Knowledge engineering and the semantic web. In: 4th Conference, KESW 2013, St. Petersburg, Russia, October 7–9, 2013. Proceedings. Springer.
Kostka, G., 2014. Barriers to the Implementation of Environmental Policies at the Local Level in China. SSRN Scholarly Paper No. ID 2487614. Social Science Research Network, Rochester, NY.
Li, W., Zhao, Y., 2015. Bibliometric analysis of global environmental assessment research in a 20 year period. Environ. Impact Assess. Rev. 50, 158–166.
Liew, W.T., Adhitya, A., Srinivasan, R., 2014. Sustainability trends in the process industries: a text mining-based analysis. Comput. Ind. 65, 393–400. ICT for Sustainability in Industry.
Liu, G., 2013. Visualization of patents and papers in terahertz technology: a comparative study. Scientometrics 94, 1037–1056.
Lundgren, M.K., 2012. The Global Impact of E-waste: Addressing the Challenge (Report).
Macqueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.
Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., Nisbet, R.A., 2012. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. ISSN 9780123870117, first ed. Academic Press/Elsevier.
Murray, F., 2002. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. Res. Policy 20 (31), 1389–1403.
Norton, M., 2000. Introductory Concepts in Information Science, second ed. In: Asist Monograph Series. Information Today, Inc., ISBN 978-1-57387-394-9
Qiu, H., Chen, Y.F., 2009. Bibliometric analysis of biological invasions research during the period of 1991 to 2007. Scientometrics 81, 601–610.
van Raan, A.F.J., 2005. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics 62, 133–143.
Rosell, M., Kann, V., Litton, J.E., 2004. Comparing Comparisons: Document Clustering Evaluation Using Two Manual Classifications. In: Presented at the ICON 2004, India.
Song, M., Yang, H., Siadat, S.H., Pechenizkiy, M., 2013. A comparative study of dimensionality reduction techniques to enhance trace clustering performances. Expert Syst. Appl. 40, 3722–3737.
Su, M.Y., 2011. Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification. J. Netw. Comput. Appl. 34, 722–730. Efficient and Robust Security and Services of Wireless Mesh Networks.
Sumathy, K.L., Chidambaram, M., 2013. Text mining: concepts, applications, tools and issues an overview. Int. J. Comput. Appl. 80, 29–32.
Sunikka, A., Bragge, J., 2012. Applying text-mining to personalization and customization research literature-who, what and where? Expert Syst. Appl. 39, 10049–10058.
Tobey, J.A., Smets, H., 1996. The polluter-pays principle in the context of agriculture and the environment. World Econ. 19, 63–87.
Tseng, M.L., Wang, R., Chiu, A.S.F., Geng, Y., Lin, Y.H., 2013. Improving performance of green innovation practices under uncertainty. J. Clean. Prod. 40, 71–82. Special Volume Sustainable Consumption and Production for Asia Sustainability Through Green Design and Practice.
Tseng, Y.H., Lin, C.J., Lin, Y.I., 2007. Text mining techniques for patent analysis. Inf. Process. Manag. 43, 1216–1247.
Uğuz, H., 2011. A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. Knowl. Based Syst. 24, 1024–1032.
US EPA, C.C.D., U.S., Greenhouse Gas Inventory Report [WWW]. http://www3.epa.gov/climatechange/ghgemissions/usinventoryreport.html (accessed 03.02.16.).
Valipour, M., 2014. Drainage, waterlogging, and salinity. Arch. Agron. Soil Sci. 1–16.
Wiebe, J., 2000. Learning subjective adjectives from corpora. In: Proceedings of the Seventeenth National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence. AAAI Press, pp. 735–740.
Yoon, B., Park, Y., 2004. A text-mining-based patent network: analytical tool for high-technology trend. J. High Technol. Manag. Res. 15, 37–50.