Technological Forecasting & Social Change, 2017, Volume 118C, Pages 270-280

Science foresight using life-cycle analysis, text mining and clustering: a case study on natural ventilation

M. Rezaeian[*,1], H. Montazeri[2,3], R.C.G.M. Loonen[2]

1 Faculty of Economics, Management & Accounting, Yazd University, Iran
2 Building Physics and Services, Department of the Built Environment, Eindhoven University of Technology, The Netherlands
3 Building Physics Section, Department of Civil Engineering, KU Leuven, Leuven, Belgium

Abstract

Science foresight comprises a range of methods to analyse past, present and expected research trends, and uses this information to predict the future status of different fields of science and technology. With the ability to identify high-potential development directions, science foresight can be a useful tool to support the management and planning of future research activities. Science foresight analysts can choose from a rather large variety of approaches. There is, however, relatively little information about how the various approaches can be applied in an effective way. This paper describes a three-step methodological framework for science foresight on the basis of published research papers, consisting of (i) life-cycle analysis, (ii) text mining and (iii) knowledge gap identification by means of automated clustering. The three steps are connected using the research methodology of the research papers, as identified by text mining. The potential of combining these three steps in one framework is illustrated by analysing scientific literature on wind catchers; a natural ventilation concept which has received considerable attention from academia, but with quite low application in practice. The knowledge gaps that are identified show that the automated foresight analysis is indeed able to find uncharted research areas. Results from a sensitivity analysis further show the importance of using full-texts for text mining instead of only title, keywords and abstract. The paper concludes with a reflection on the methodological framework, and gives directions for its intended use in future studies.

Keywords: Science foresight, Life-cycle analysis, knowledge gap identification, Sensitivity analysis, Text mining, Wind catcher.

1. Introduction

Effective management and planning of research and development activities requires strategic allocation of available resources (Arroyabe, Arranz, & de Arroyabe, 2015; Berloznik & Van Langenhove, 1998). This issue manifests itself at different scales and plays a role in private companies and public authorities as well as academia. For example, individual scientists and research departments have a keen interest in spending their time and money in areas with potential for high impact (Kajikawa, Yoshikawa, Takeda, & Matsushima, 2008; Ronald N. Kostoff,

* Corresponding author: Mina Rezaeian, Department of Economics, Management and Accounting, Yazd University, P.O. box 89195 - 741, Yazd, Iran. *E-mail address: mina.rezaiyan@ymail.com*

2008; R. N. Kostoff & Schaller, 2001; Leydesdorff, Cozzens, & Van den Besselaar, 1994; Ogawa & Kajikawa, 2015). Likewise, (inter)national governmental institutions seek to establish policy instruments (e.g. legislation and funding schemes) that give priority to development and application of innovative solutions with the highest positive contribution for society (Coccia, 2009; Kidwell, 2013).

Identification of such high-potential research and development areas is a challenging task. Making well-informed decisions requires detailed knowledge of past findings and current trends, and a deep understanding of emerging technology pathways (Leydesdorff et al., 1994). At the same time, it asks for a broad perspective to oversee future needs while identifying the opportunities that arise in neighboring research domains. The context in which such decisions are made is becoming increasingly complex because traditional science and engineering domains are getting more and more interconnected (Morillo, Bordons, & Gómez, 2003; Porter & Rafols, 2009). In addition, the information that is documented in patents, reports and research papers continues to grow in size at an exponential rate (Bengisu & Nekhili, 2006; Kajikawa et al., 2008; R. N. Kostoff & Schaller, 2001; Larsen & Von Ins, 2010). The availability of input for research and technology planning can therefore be perceived as overwhelming, especially for decision-makers who are new to the field. The inability to properly analyze and comprehend all this information may lead to wrong recommendations and suboptimal priorities in research and development agendas.

Science foresight refers to the collection of analysis and prediction methods that can assist the development of a science vision in order to prepare for future challenges or needs in science (Ben R Martin, 1995; Ben R. Martin, 2010). It has successfully been implemented in different fields, such as economy (Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014), environmental science (Dubarić, Giannoccaro, Bengtsson, & Ackermann, 2011; Iniyan & Sumathy, 2003), foresight (Saritas & Burmaoglu, 2015; H.-N. Su & Lee, 2010), health science (Abbott, Foster, Marin, & Dykes, 2014; Pereira & Escuder, 1999), politics (Coates, 1985), nano science and technology (de Miranda Santo, Coelho, dos Santos, & Fellows Filho, 2006; Huang, Notten, & Rasters, 2011; Robinson, Ruivenkamp, & Rip, 2007), and social science (Baloglu & Assante, 1999; Singh, Hu, & Roehl, 2007). The literature on science foresight covers a wide variety of qualitative and quantitative means for monitoring clues and indicators of evolving trends and developments (Coates, 1985). To facilitate successful science foresight analyses, it is clear that the methodology needs to be matched with e.g., the purpose of the study, the size and quality of the data base, and the type of output that is expected. However, the available information about the relative effectiveness of different science foresight methods is very limited, and it is therefore difficult to support such decisions. In addition, most methods perform well at some, but typically not all aspects of science foresight. The potential of combining the positive sides of different science foresight methods into one overall framework has so far remained relatively unexplored.

The main objective of this paper is to develop and evaluate a three-step methodological framework that can be used to identify knowledge gaps and provide new insights into development directions of a well-defined technological field. Although we aim at wider applicability, in this paper, the framework is developed and demonstrated with respect to wind catchers; a sustainable natural ventilation system for buildings. This topic was specifically chosen because it is manageable in scope and size (i.e. the veracity of the results can be checked), yet has experienced a complex development history, is an active field with mixed research methods, and has a non-trivial future outlook. This paper uses a combination of existing methods: life-cycle analysis, text mining and cluster analysis, but combines them in a novel way that has not been described before. Given the importance of

the impact of textual data on the accuracy of text mining, a sensitivity analysis is also carried out for three cases, when (i) title, (ii) title, abstract and keywords, and (iii) full-text of the papers are considered as the textual data. This evaluation is based on the methodology of the research papers.

This paper continues by describing the development of a methodological framework for science foresight on the basis of life-cycle analysis, text mining and cluster analysis (section 2). The sensitivity analysis is also presented in this section. Characteristics of wind catchers, the topic of the application study, are introduced in section 3. In section 4, this methodology is applied in the case of wind catchers, to describe the status of research in this field, predict future trends and identify knowledge gaps in order to identify possible opportunities for new research and development activities. In section 5, a reflection on the methodological framework and its potential in future studies is given.

2. Methodology

2.1. Life-cycle analysis

Life-cycle analysis is a widely-used data analysis technique that can be applied to describe the historical development of a technology or research domain, and, subsequently, to estimate the future trend or perspectives. Ernst (Ernst, 1997) suggests that the accumulation of patent applications is useful for measuring technology trends. The evolution over time can be plotted as S-shape curve to represent its technology life-cycle. There are four stages in a technology life cycle: introduction, growth, maturity and saturation (Ernst, 1997). During the introduction stage, there is a little growth in the number of patent applications. The growth stage, on the other hand, is characterized by exponential growth. As the patent application rate declines, the mature stage is entered. The saturation stage indicates limited growth with only few additional patent applications (Charles V. Trappey, Wu, Taghaboni-Dutta, & Trappey, 2011).

If the current stage of a science or technology is known, it would be possible to forecast the future trends and predict the saturation level and therefore, estimate the potential of the field for further and deeper studies. Knowledge about the maturity and future growth potential of science or technology innovations helps researchers, for example, to decide whether to continue investing resources or switch research directions (Campani & Vaglio, 2014; Charles V Trappey, Trappey, & Wu, 2010; Charles V. Trappey et al., 2011).

In this study, cumulative paper publications are used for predicting future development trends using Loglet analysis. The analysis is performed using "Loglet Lab" software. It refers to the decomposition of growth and diffusion into S-shaped logistic components, roughly analogous to wavelet analysis, popular for signal processing and compression (Meyer, Yung, & Ausubel, 1999).

Eq. (1) presents the equation to calculate the logistic growth:

$$N(t) = \frac{K}{1 + exp[-\frac{ln(81)}{\Delta t}\ (t - t_m)]} \quad (1)$$

where K is the asymptotic limit that the growth curve approaches and shows the saturation level of the growth, Δt is the characteristic duration that specifies the time required for a trajectory to grow from 10% to 90% of the limit K and t_m is the midpoint of the growth trajectory (Fig. 1).

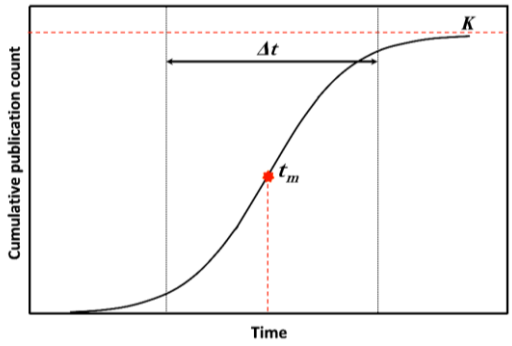

Fig. 1. A logistic curve and its parameters.

First, the logistic growth is visualized by simply plotting data on an absolute and linear scale. The Fisher-Pry transform is used to transform the logistic curve into a linear one. By doing so, Δt, t_m and K can be determined. Further information is presented by Mater et al. (Meyer et al., 1999).

Many growth and diffusion processes consist of several sub processes. Systems with two growth phases are called "bi-logistic". In such models, growth is the sum of two discrete wavelets, each of which is a three-parameter logistic, as presented in Eq. (2).

$$N(t) = N_1(t) + N_2(t) \quad (2)$$

2.3. Text mining

One of the most popular methods for science foresight is text mining. Text mining is used to identify valuable information such as relations, patterns or trends in textual data (Choudhary, Oluikpe, Harding, & Carrillo, 2009; Delen & Crossland, 2008; Ghazinoory, Ameri, & Farnoodi, 2013). For example, it has been widely adopted to explore the complex relationships among scientific documents (de Miranda Santo et al., 2006; Singh et al., 2007). A main theme supporting text mining is the transformation of text into numerical data. This transformation uses statistical methods to convert text mining into a classical data mining encoding. Despite the inability to explicitly understand linguistic concepts such as grammar or word meaning, statistical text mining has proven remarkably successful (Weiss, Indurkhya, & Zhang, 2010). Many projects have used different techniques of statistical text mining in various fields of science or technology. In these studies, the full-text or abstract of papers or patents are considered as the database. Table 1 provides an overview of previous studies in which different techniques of text mining were implemented in science and/or technology.

2.3.1. Impact of textual data

Text mining of research papers can be performed on different parts of the papers. However, while the full-text of papers is widely available in electronic versions, most applications of text mining are restricted to the abstract (Andrade & Bork, 2000; Daim, Rueda, Martin, & Gerdsri, 2006; de Miranda Santo et al., 2006; Dubarić et al., 2011; Charles V. Trappey et al., 2011; Trappey., Wu, Taghaboni-Dutta, & Trappey, 2011). Therefore, it would be worthwhile to investigate if full-text papers, instead of only abstract and meta-data would lead to different, more insightful results. Given the importance of the impact of textual data on the accuracy of text mining, a sensitivity analysis is carried out for three cases, where (i) title, (ii) title, abstract and keywords, and (iii) full-text of the research papers are considered as the textual data. The evaluation is based on the methodology of the papers. This choice is inspired by the fact that methodology is normally mentioned in different parts of the paper including the title.

Table 1. Overview of previous studies in which different techniques of text mining is used in science and/or technology

Year	Ref.	Field of study	Database	Text mining technique
2005	(Glenisson, Glänzel, Janssens, & De Moor, 2005)	Scientometrics	Paper	Bibliometric analysis, Categorization, Clustering
2005	(Yoon & Park, 2005)	TFT-LCD	Patent	Factor analysis
2006	(de Miranda Santo et al., 2006)	Nanotechnology	Paper	Bibliometric analysis, Text analysis, Visualization
2006	(Hsu, Trappey, Trappey, & Hou, 2006)	Hand tool industry	Patent	Information extraction, Clustering
2007	(Singh et al., 2007)	Human resource management	Paper	Clustering
2007	(Ronald N. Kostoff et al., 2007)	Technical articles	Paper	Clustering, Bibliometric analysis
2008	(Kim, Choe, Choi, & Park, 2008)	Mobile service	Patent	Keyword extraction, Text analysis
2009	(Lee, Yoon, & Park, 2009)	Personal digital assistant technology	Patent	Keyword extraction, PCA, Patent mapping
2008	(Delen & Crossland, 2008)	Management information systems	Paper	Clustering
2009	(Choudhary et al., 2009)	Construction	Post project reports	Information retrieval and extraction, Text analysis, Categorization, Summarization, Visualization
2010	(H.-N. Su & Lee, 2010)	Technology foresight	Paper	Visualization
2010	(Greenacre & Hastie, 2010)	Journal Vaccine	Paper	Clustering, Visualization
2011	(Ronald N. Kostoff, 2011)	Severe acute respiratory syndrome	Paper	Clustering, Auto-correlation mapping, Factor analysis
2011	(Charles V. Trappey et al., 2011)	Radio Frequency Identification (RFID)	Patent	Clustering
2012	(Choi, Park, Kang, Lee, & Kim, 2012)	Proton exchange fuel cell technology	Patent	Clustering
2012	(Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2012)	Fuzzy sets theory	Paper	Information retrieval, Visualization
2012	(Sunikka & Bragge, 2012)	Personalization & Customization	Paper	Bibliometric analysis, Visualization
2013	(Thorleuchter & Van den Poel, 2013)	German defence research program	R&D project	Classification
2014	(No, An, & Park, 2014)	Postage metering system	Patent	Clustering
2014	(Yoon, Park, & Coh, 2014)	LED technology	Patent	Text analysis, Keyword extraction
2014	(Liew, Adhitya, & Srinivasan, 2014)	Main sectors of the industry	Reports	Keyword extraction
2014	(Jun, Park, & Jang, 2014)	News/ Document & clustering	Patent-News	Clustering
2015	(Wang, Fang, & Chang, 2015)	Microalgal biofuel	Patent	Clustering
2015	(Kundu, Jain, Kumar, & Chandra, 2015)	Supply chain	Paper	Factor analysis
2015	(Moro, Cortez, & Rita, 2015)	Banking industry	Paper	Classification

The sensitivity analysis is carried out in three steps: First, a dictionary for keywords related to the research methodologies is developed based on the experts' opinion. Second, the frequency of the keywords in the papers is calculated and the research methodology of each paper is identified by comparing the frequency of the methodology keywords using Eq. (3).

$$T_{ij} = \sum_{m=1}^{T(KP_i).} KPF_m \tag{3}$$

where T_{ij} is the frequency of research methodology i in paper j, $T(KP_i)$ is the number of methodology keywords (related to research methodology i) in paper j, and KPF_m is the frequency of keyword m (related to research methodology i) in paper j.

Finally, the research methodology of each paper predicted by the text mining is compared to the "correct" methodology that was determined manually by the experts.

2.3.2. Content analysis

Text mining is normally performed with the use of advanced computer algorithms. However, using experts is of importance to interpret the results and to analyze the relevance of acquired information. Therefore, the reliability of a text mining activity is correlated to the skills and knowledge of the experts that are consulted. In this study, QDA MINER and WORDSTAT software are used for text mining process. In Fig. 2, the assistance of domain experts along with the text mining process is schematically shown.

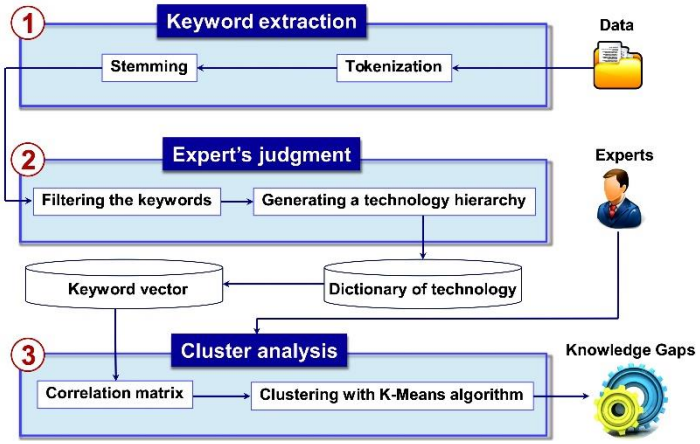

Fig. 2. Text mining process.

a. Keyword extraction

As shown in Fig. 2, the first step in handling text is to break the textual data into words or, more precisely, tokens. This is important for further analysis because without identifying the tokens, it is difficult to imagine extracting higher-level information from the document. Therefore, the textual data is analyzed to identify all the words and phrases in the papers. Once the papers have been segmented into a sequence of tokens, the next step is to convert each of the tokens to a standard form. In English, as in many other languages, words occur in text in more than one form. Often, but not always, it is advantageous to eliminate this kind of variation before further processing. When the normalization is confined to regularizing grammatical variants such as singular/plural and present/past, the process is called stemming (Weiss et al., 2010). After the process of stemming, the extracted words and phrases, are filtered by the experts to define the most relevant and meaningful keywords. These keywords are used to make a dictionary.

b. Expert's Judgment

- *Filtering the keywords:*

Expert's judgment is used in an iterative process to choose and agree on the number of manageable keywords for deriving the most reasonable clusters that can be interpreted. This choice could be critical because if the number of keywords is too small, the keyword-clusters will be too broad to reveal details. On the other hand, if the number is too large the results might not be manageable (Singh et al., 2007). Therefore, in this study, a technology hierarchy is applied to define the related keywords in this field.

- *Generating a technology-publication hierarchy:*

In some studies, only high frequency keywords are taken into account to build the dictionary keywords. This approach is useful to identify the existing themes or sub-fields. In order to find the knowledge gaps efficiently, experts' opinion is needed to consider all aspects of a scientific field, even those that are not explicitly mentioned in research papers. An ontology tree, or hierarchy, can help the experts specify a general overview of a scientific field. This hierarchy specifies the relationships between concepts and is used to extract meaningful words and phrases from the papers. Since generating a hierarchy is domain specific, experts must select the keywords and phrases related to this technology. In this study, the hierarchy is generated based on experts' opinion. Using the keywords that were automatically extracted from the papers, experts were able to generate this hierarchy easier. Note that the experts are allowed to add, remove and replace keywords during this step. To make the technology hierarchy, the method presented by Yoon and Park (Yoon & Park, 2005) is used. They used this method to generate a technology tree and transform it into a morphology box. This method helps to break a complex problem or technology into several parts (dimensions), so that it can be analyzed more easily. In the present study, three sections are considered: "Procedural", "Structural" and "Application". The "Procedural" section deals with the aspects related to the methodology of the papers. The "Structural" section addresses the different features of a technology and the factors affecting its performance. The "Application" section is related to the function of this technology in various sectors. Using this method, also the aspects that have not yet been taken into consideration in existing studies are included in the hierarchy. It can also be useful to find all relevant keywords that could help to find the knowledge gaps more efficiently.

c. Generating the dictionary and keyword vector

The next step is to develop a technology dictionary based on the filtered keywords and the hierarchy. Using this dictionary, the documents are transformed into key phrase vectors by analyzing the frequency (F) of each keyword and phrase (KP) in each paper. These vectors can be calculated by WORDSTAT software as shown in Table 2. For example, $F_{i,j}$ is the frequency of keyword *i*, in paper *j*.

Table2. Frequency of keywords in papers

	KP1	**KP2**	**KP3**	**…**	**KP$_N$**
Paper 1	$F_{1,1}$	$F_{1,2}$	$F_{1.3}$	…	$F_{1,N}$
Paper 2	$F_{2,1}$	$F_{2,2}$	$F_{2,3}$	…	$F_{2,N}$
Paper 3	$F_{3,1}$	$F_{3,2}$	$F_{3,3}$	…	$F_{3,N}$
…	…	…	…	…	…
Paper M	F_{M1}	F_{M2}	F_{M3}	…	F_{MN}

Assigning key phrase weights by the terms' frequency (TF) of appearance in a document is a popular method in text mining (Weiss et al., 2010). Regarding the fact that each paper can have a different number of pages and words, it is reasonable to conclude that the size of a paper and the average number of words, can affect the frequency of the keywords. As a remedy, this study used the method proposed by Trappy et al (Trappey. et al., 2011), which normalizes the weights for the frequency of key phrases by the number of words in each document. The function normalized TF-IDF or NTF approach is expressed by Eq. (4) where tf_{ik} is the number of key phrases *i* in document *k*, WN_k is the number of words in document *k*, *n* is the total number of documents in the document

set, and df_i is the number of documents of key phrase i in the document set. With this function, the effect of the size of a paper can be eliminated (Charles V Trappey et al., 2010).

$$NTF_{ik}-IDF = tf_{ik} * \frac{\sum_{s=1}^{n} WN_s}{n} * \frac{1}{WN_k} * \log_2\left(\frac{n}{df_i}\right) \quad (4)$$

d. Cluster analysis

- *Correlation measurement:*

A correlation between a set of data is a measure of how well and to which extent they are related. The correlation between keywords of documents has been used in many studies for clustering and discovering the relationship between the keywords. In these studies, the correlation, similarity or co-occurrence are calculated between all dictionary keywords (Andrade & Bork, 2000; Daim et al., 2006; de Miranda Santo et al., 2006; Dubarić et al., 2011; Charles V. Trappey et al., 2011). In the present study, however, the research methodology of each paper is used as the clustering variables. Using this method, the correlation of each research methodology with dictionary keywords can be determined, which is important to support effective research planning. Further information will be given in Section4.3 . In addition, the knowledge gaps can also be identified, by analyzing cases where there is no high co-occurrence or correlation between the research methodology and dictionary keywords. In the interpretation of the knowledge gaps, it is particularly interesting to link these findings back to the evolution of research methodologies, as identified through the combination of text mining and life cycle analysis. Therefore, after normalizing the frequency of the keywords, the correlation between these keywords and the keywords related to the methodology of the papers, are measured with the Pearson correlation coefficient as shown in Table 3. Note that the keywords related to each methodology are determined by the experts. For each methodology, the total frequency of all related keywords is used as an input for the correlation measurement. This measurement is based on the keywords co-occurrence with the methodology keywords in each paper. In Table 3 KP represents the keywords or key phrases and $R_{i,j}$ is the correlation between keyword i, and the group of keyword related to methodology j.

Table 3. Correlation matrix between keywords and research methods

	Methodology 1	**Methodology 2**	**…**	**Methodology N**
KP 1	$R_{1,1}$	$R_{1.2}$	…	$R_{1,N}$
KP 2	$R_{2,1}$	$R_{2,2}$	…	$R_{2,N}$
KP 3	$R_{3,1}$	$R_{3,2}$	…	$R_{3,N}$
…	…	…		…
KP N	R_{M1}	R_{M2}	…	R_{MN}

- *Clustering:*

Clustering is a statistical approach for classification of patterns into groups based on similarities of internal features or characteristics. K-means is a common clustering algorithm that has been used in a wide variety of applications to partition a data set into K groups (Chemchem & Drias, 2015; de Miranda Santo et al., 2006). To do so, the user must assign a number K as the expected number of clusters. Since the centroids of clusters are randomly chosen, the algorithm must repeat many times to adjust centroids and can only achieve locally optimal clustering results. The optimal number of clusters is defined by the experts after analyzing the result of clustering

with different numbers. So choosing the best clusters is an expert-based procedure. In addition, to verify the authenticity of the experts' opinion, the Davies-Bouldin index (Bezdek & Pal, 1998) is used.

3. Description of the application study

In this section, the framework presented in section 2 is demonstrated with respect to wind catchers; a sustainable natural ventilation system for buildings.

3.1 Wind catchers

Finding solutions that enable cost-effective operation of buildings with good comfort conditions and less negative impact on the environment is identified as one of the most compelling challenges of the 21st century (IEA, 2013; Kolokotsa, Rovas, Kosmatopoulos, & Kalaitzakis, 2011). Research and development of innovative building systems is expected to play an important role in facilitating this transition towards sustainable building design (Loonen, Singaravel, Trčka, Cóstola, & Hensen, 2014). Natural ventilation strategies have the potential to become a viable alternative for energy-intensive air-conditioning systems, but new approaches are needed to lead to wider adoption. Inspiration for new concepts can be found by looking at traditional design strategies such as the "wind catcher", which was used to provide natural ventilation and passive cooling in hot and arid regions of Iran and neighboring countries (Saadatian, Haw, Sopian, & Sulaiman, 2012). The cooling performance of both ancient and modern wind catchers has been analyzed and optimized by several researchers, using experimental, computational or analytical methods (M. Bahadori, Mazidi, & Dehghani, 2008; Calautit, Hughes, & Ghani, 2013; Calautit, O'Connor, & Hughes, 2014; Hassan & Lee, 2014; Montazeri, 2011; Montazeri & Azizian, 2008; Montazeri & Azizian, 2009; Montazeri, Montazeri, Azizian, & Mostafavi, 2010; Saadatian et al., 2012; S Soutullo, Sanchez, Olmedo, & Heras, 2011; Y. Su, Riffat, Lin, & Khan, 2008). In addition to academic contexts, industry has also paid attention to the development of these systems. These technology transfer activities have led to the invention of multiple new commercial wind catcher systems that meet the requirements of modern-day building design (Hughes, Calautit, & Ghani, 2012). However, considering the relatively slow uptake in practice, it can be argued that further research and development is still needed to fully utilize the potential of wind catchers.

3.2 Data base

Establishing a database is the first step in a science foresight analysis. In this study, peer-reviewed papers were collected based on the keywords "wind catcher", "windcatcher", "wind tower + building", "windvent + building" and "cool tower + building". The search was performed in the title, abstract and keywords of papers in the Scopus database. The peer-reviewed papers published until the end of 2014 were considered. The search led to 119 papers. However, not all of these papers focus on the application of natural ventilation or passive cooling in buildings. For example, the term "wind tower" can also be used to refer to a cooling system in power plants. To make sure that only relevant papers are included, each abstract was optically scanned be the experts. After filtering out 27 papers, the database consisted of 92 papers.

The experts were carefully selected based on their knowledge about this technology; consist of 2 professors, 2 post-doctoral researchers, 5 PhD students and an engineer in the fields of mechanical engineering and building physics from 3 countries: Iran, The Netherlands and Belgium.

4. Results

4.1. Life-cycle analysis

Fig. 3 presents the results of the life cycle analysis on research papers on wind catchers. The Fisher-Pry transform of the publication growth is shown in Fig. 3(a). The time in which the value is between 10^{-1} and 10^{1} is equal to Δt, and the time at 10^{0} is the point of inflection (t_m). The publications follow a bi-logistic curve (Fig. 3(b)). The first curve started in 1984 and its growth phase began in 1990. Then, after a relatively short growth and maturation phase, it reached its saturation level in 1997. Around the year 2000, another logistic started.

This pause in publications and the renewed interest after the year 2000 can be explained by the increasing application of Computational Fluid Dynamics (CFD) as a method for investigating the performance of wind catchers. This can be clearly observed in Fig. 4 in which the actual number of papers per methodology that were published in a given period is presented. Moreover, the pattern presented in Fig. 3(b) marks the transition from analysis of ancient wind catchers towards the development of innovative wind catcher systems in the 21st century. The second curve has currently entered the growth stage of the publications life-cycle. It has had a continuing growth since 2006. Using the growth model described in Section 2.1, the publication count is forecasted to increase by about 54% and reach its upper limit at 2020 with 142 papers. Therefore, this extrapolation indicates a clear potential for the development of this field in science and technology. Note that the growth period can still be extended, if there are new breakthrough innovations in this area. Therefore, scientists and inventors should analyze potential opportunities in this field.

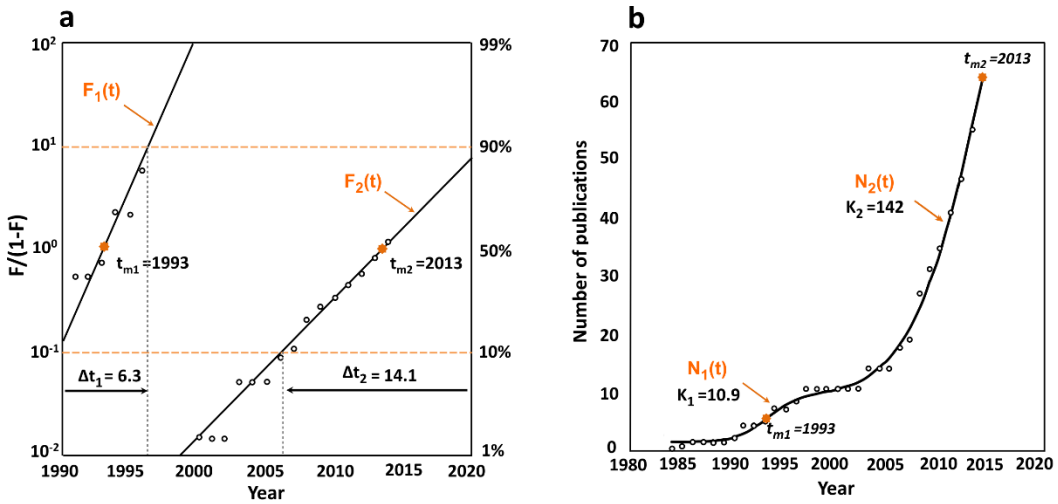

Fig. 3. (a) Determining Δt and t_m of the logistic growth using Fisher-Pry transform that renders the logistic linear. (b) Growth of wind catcher publications fitted to a bi-logistic curve.

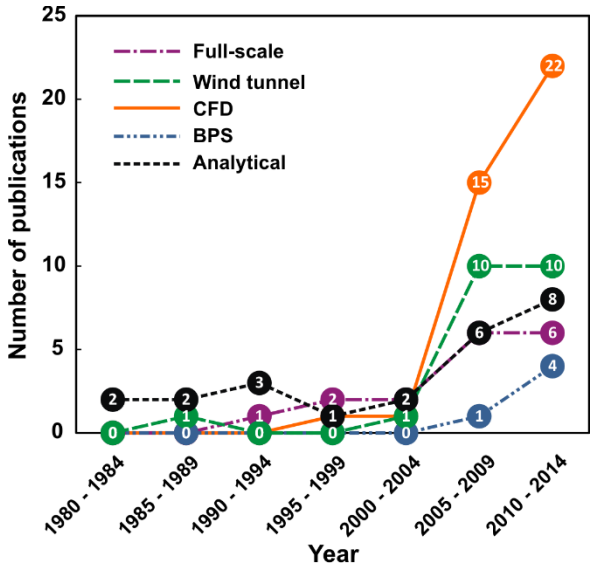

Fig. 4. Research methodologies applied in the wind catcher data base.

4.2. Text mining

4.2.1 Impact of textual data on text mining

The sensitivity analysis is carried out based on a dictionary for keywords related to the research methodologies. These keywords are presented in Table 4.

Table 4. Descriptive statistics of the sensitivity analysis

Method	Representative phrases	No of papers with this methodology	Top cited papers (Scopus)
Wind tunnel	Scale model, Scale models, Wind tunnel	20	(Karakatsanis, Bahadori, & Vickery, 1986), (Montazeri & Azizian, 2008), (Montazeri et al., 2010)
Full-scale	Full-scale measurement, Full-scale experiment, Field study, Field analysis, Prototype, Test cell, Questionnaire, Survey, Post occupancy evaluation, IN SITU, Occupied	21	(Pearlmutter, Erell, Etzion, Meir, & Di, 1996), (M. Bahadori et al., 2008), (Kalantar, 2009)
CFD	CFD, Computational fluid dynamics, Numerical analysis, Numerical simulation, CFX, Fluent, Ansys/Fluent, Openfoam, Turbulence model, Large eddy simulation, LES, RANS, Grid, Mesh, Reynolds-average Navier-stokes	39	(Li & Mak, 2007), (Montazeri et al., 2010), (Montazeri, 2011)
BPS & AFN	Building energy simulation, Enrgyplus, TRNSYS, Building simulation, ESP-r, Airflow network, Thermal model, Energy balance, CONTAM	6	(Nouanégué, Alandji, & Bilgen, 2008), (S Soutullo et al., 2011), (S. Soutullo, Sanjuan, & Heras, 2012)
Analytical	Analytical model, Analytically, Power law model, Simplified model, Mathematical model	24	(Bansal, Mathur, & Bhandari, 1994), (Mehdi N Bahadori, 1985), (Karakatsanis et al., 1986)
Review	Review, Literature review, Literature survey	14	(Khan, Su, & Riffat, 2008), (Hughes et al., 2012), (Mehdi N. Bahadori, 1994)

The overall capability of text mining to correctly identify the research methodology based on different parts of the paper is gained through the sensitivity analysis. Text mining using full-texts identifies the research methodology of the papers with an accuracy of 93.6%. The accuracy reduces to 67.7% for the combination of title, abstract and keywords and to 12.3% when only the titles of the papers are taken into account. In this study, therefore, full-texts of the papers are used for the text mining process.

4.2.2 Keywords and clusters

Using the text mining approach, all 92 papers are analyzed. Considering the extracted words and phrases from the papers, the experts established a hierarchy based on wind catcher publications (Fig. 5). Finally, the wind catcher dictionary is developed based on the filtered keywords and wind catcher publication hierarchy. Among 5783 words and phrases, 57 key phrases were found by the experts as the most related words to this technology.

The keywords are clustered with the K-means algorithm, based on their correlation and co-occurrence with internal characteristics of the papers. In this study, clustering has been performed based on the research methodology of the papers. Research on wind catchers is usually performed using (i) full-scale measurements, (ii) wind-tunnel measurements, (iii) CFD simulations and (iv) Building Performance Simulation (BPS), Air Flow Network (AFN) and analytical methods. In this study, therefore, the clusters are determined based on these four research methodologies as the clustering variables. After analyzing the clusters with different numbers of K, nine clusters were chosen to be the optimal number of clusters according to the experts' opinion. To verify the authenticity of the experts' opinion, it was also verified using the Davies-Bouldin (DB) index. This index is a function of the sum of within-cluster variance to between-cluster-center distances. The appropriate number of clusters is indicated by the minimum value of the DB index (Bezdek & Pal, 1998; Wu, Tang, Yang, Liu, & Guo, 2013). The results are presented in Table 5 for different numbers of clusters. It can be seen that K = 9 is superior with DBI = 2.97.

Table 5. Davies-Bouldin index for different numbers of clusters

Number of clusters (K)	**6**	**7**	**8**	**9**	**10**	**11**	**12**
DBI	3.79	3.83	4.03	2.97	3.18	3.01	3.07

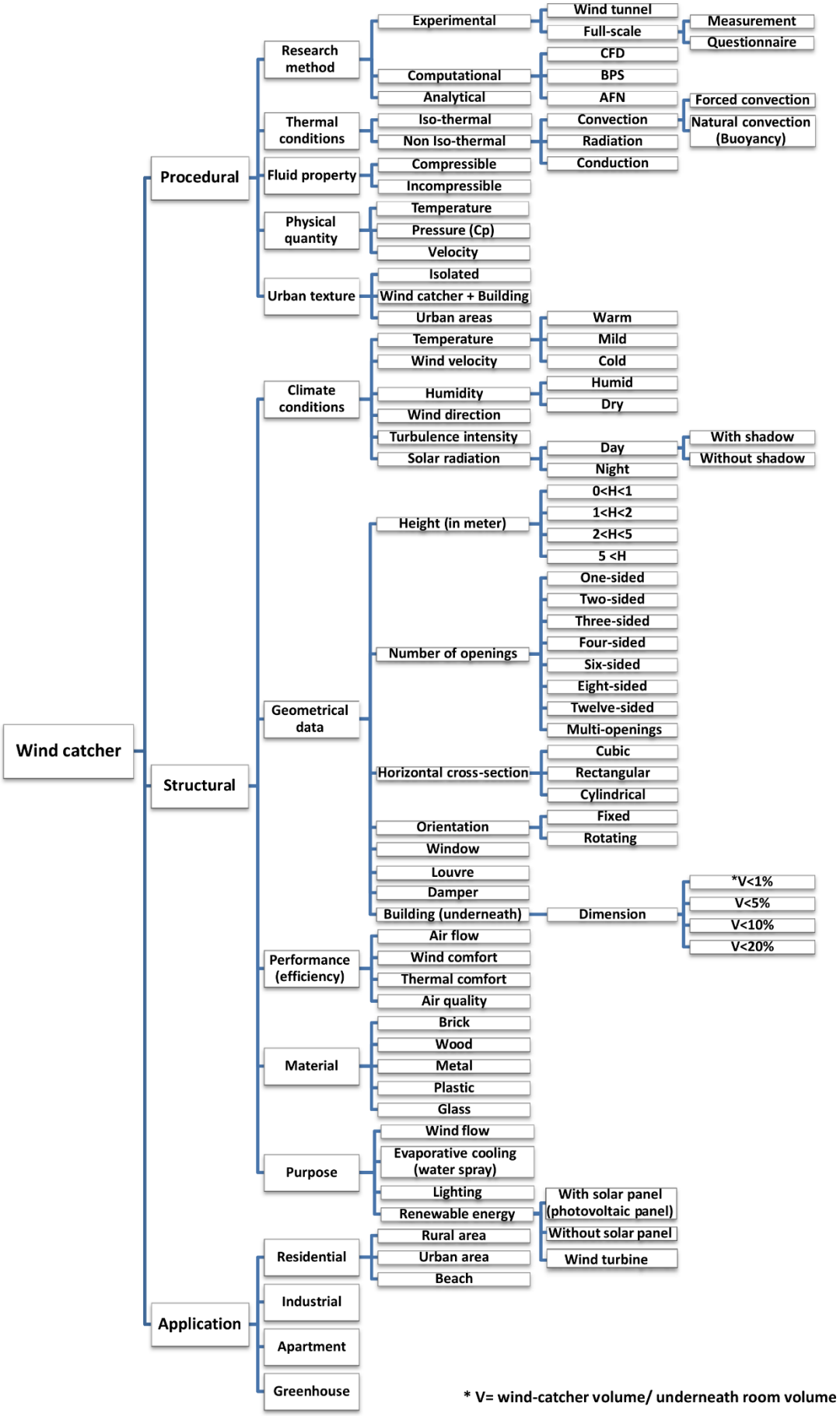

Fig. 5. Wind catcher-publication hierarchy.

By analyzing the relationship between the keywords in each cluster and the correlation between these keywords and the methodology of the papers, characteristics of each cluster are defined by the experts. The nine clusters including all keywords are presented in Table 6.

Table 6. Clusters along with keywords

Cluster	Methodology	Representative phrases
1	WT[1]	Cylindrical, Three-sided, Six-sided, Twelve-sided
2	FS[2]	Plastic, Residential, Day, Thermal comfort, Night, Mild, Evaporative cooling, Water spray, Air quality, Shadow, Warm, Glass, Lighting
3	CFD[3]	Apartment, Rectangular, Louvre, Turbulence intensity, Buoyancy, Natural convection
4	BPS, AFN & AN[4]	Humid, Conduction, Cold
5	WT + CFD	Cubic, Damper, Metal, Industrial, Four-sided, Wind speed, Wind velocity, Velocity, Incompressible
6	WT+ CFD + (BPS, AFN & AN)[5]	Rural, Wood, Two-sided, One-sided, Solar panel, PV panel
7	WT + CFD + (BPS, AFN & AN)[5]	Air flow, Pressure, C_P, Wind direction, Wind angle
8	FS+ (BPS, AFN & AN)	Temperature, Dry, Radiation, Beach, Urban
9	None	Forced convection, Brick, Greenhouse, Iso-thermal, Wind comfort, Wind turbine

[1]**WT:** Wind tunnel measurement; [2]**FS:** Full-scale measurement; [3]**CFD:** Computational fluid dynamics; [4] **BPS**: Building Performance Simulation, **AFN**: Air Flow Network, **AN**: Analytical method

[5] In these clusters, keywords have a high correlation with "CFD", "Wind Tunnel" and "BPS, AFN & Analytical" methods but in Cluster 6, the correlation between keywords and (BPS, AFN & AN) is lower compared to the correlation in Cluster 7

Note that the number of clusters greatly depends on the number of clustering variables. The total number of clusters reflects how the keywords in the dictionary keywords have been investigated from a methodological point of view.

Depending on the strength of co-occurrence and correlation between keywords and research methodologies, different scenarios can be distinguished. These scenarios are therefore an aggregation of possible cases. Given the clustering results, four possible scenarios can be considered for research on a specific keyword:

a) There is a high correlation between the keyword and a single research methodology. This scenario indicates that only one research methodology has been used to investigate the keyword. In this study, the total number of possible clusters for this scenario is 4 clusters (WT, FS, CFD, and BPS, AFN & AN). In total, 26 out of 57 keywords in the dictionary keywords have been investigated using only one methodology.

b) There is a high correlation between the keyword and multiple research methodologies. In this case, the keyword has been investigated using a combination of two or more research methodologies, in which the keyword has a high correlation with all those research methodologies, and no correlation with others. In this study, the total number of possible clusters for this scenario is 11 (WT+FS, WT+CFD, WT+BPS, AFN & AN, etc.). Clusters 5, 6 and 8 have been identified in this category. The results show that 20 keywords have been investigated using two or three research methodologies.

c) There is a low correlation between the keyword and one methodology but high correlation with multiple research methodologies. Cluster 7 represents this scenario. It comprises keywords that have been investigated using two or more research methodologies (three in this case), in which the correlation between the keywords and at least one of the research methodologies is significantly lower than the other methodologies in the cluster, but not zero. Out of many possible combinations, only one such cluster was identified in this study. Such a moderate correlation indicates that a trend is certainly discernible, but that the correlation is less pronounced than is the case for the two previous scenarios. In the present case, it is interesting to observe that keywords in cluster 6 (belonging to the second scenario) can all be characterized as 'structural', following the classification introduced in Section 2.3.2. Keywords in cluster 7 are connected to the same research methodologies as cluster 6, but can be connected to either 'procedure' or 'application'.

d) The correlation between the keyword and the four research methodologies is zero. It indicates that no significant research on this combination has taken place yet. Keywords in cluster 9 belong to this scenario.

The various scenarios indicate how the keywords have been investigated from a methodological point of view. Keywords in scenario 1 can be linked to mono-disciplinary research activities, as it concerns topics that have been studied using only one research methodology. This gives input for exploring potential new research directions by introducing research approaches that have not yet been employed in that context, as will be explained in Section 4.3. Keywords in Scenario 2, on the other hand, relate to phenomena or aspects that have already been investigated in a multi-disciplinary approach. The third scenario can provide additional information as the second scenario. Given the relatively low correlation of the keywords and one of the research methodologies, this information should be treated with caution, preferably in combination with expert's opinion. The last scenario provides significant information to identify research gaps, as will be discussed in Section 4.3.

4.2.3 Relationships between research methodologies and clusters

To better understand the qualities of the automated text mining and clustering methods, this section explores features of each research methodology and connects these to several keywords in the clusters.

Wind-tunnel measurement: Reduced-scale wind-tunnel measurements allow a strong degree of control over the boundary conditions, however at the expense of – sometimes incompatible – similarity requirements. Furthermore, wind-tunnel measurements are usually also only performed in a limited set of points in space (Montazeri & Blocken, 2013). The results show that wind-tunnel measurements mainly have been used to investigate the aerodynamic characteristics of wind catchers (Clusters 1 and 6). In recent years, wind-tunnel measurement and CFD studies are used in a complementary way, as the accuracy and reliability of CFD are of concern, and validation studies are normally performed using wind-tunnel measurement data [e.g. (Ghadiri, Lukman, Ibrahim, & Mohamed, 2013; Montazeri et al., 2010)]. The reason for the occurrence of the word "four-sided" is that the measurement data for this type of wind catcher is available in the literature and is widely used for CFD validation purposes.

Full-scale measurement: This method offers the advantage that the real situation is studied and the full complexity of the problem is taken into account. However, full-scale measurements are usually only performed in a limited number of points in space. In addition, there is no or only limited control over the boundary conditions

(Montazeri & Blocken, 2013). Therefore, keywords like "wind direction" do not occur in this cluster because of its limited control over the boundary conditions. The reason of occurrence of the phrases "evaporative cooing" and “water spray” with full-scale measurement is related to the complexity involved in the two-phase flow in water sprays as the evaporation process depends on several physical parameters that are not easily varied independently (H. Montazeri, B. Blocken, & J. Hensen, 2015; H. Montazeri, B. Blocken, & J. L. Hensen, 2015).

Computational fluid dynamics (CFD): This simulation-based method provides whole-flow field data, i.e. data on the relevant parameters in all points of the computational domain. Unlike wind-tunnel testing, CFD does not suffer from potentially incompatible similarity requirements because simulations can be conducted at full scale (Blocken, 2014). This could be reason for the occurrence of the word “apartment” in Cluster 3. In addition, CFD simulations are flexible and easily allow parametric studies. Therefore, the impact of “louvre” (Cluster 3) or “damper” (Cluster 5) is widely investigated using CFD.

BPS, AFN and analytical methods: Compared to the other methods, this group of computational approaches has a stronger emphasis on design support instead of product development and analysis (Loonen et al., 2014), as it is able to predict the dynamic performance of proposed buildings considering occupant behavior, indoor comfort conditions and meteorological boundary conditions (Clarke & Hensen, 2015). Another characteristic feature of this group of methods is the fact that thermal considerations, instead of aerodynamic effects, can easily be taken into account in the analysis. It is for this reason that, key words such as “conduction”, “temperature” and “radiation” are strongly correlated to BPS, AFN and analytical methods.

Fig. 4 shows a breakdown of research methodologies that were applied in the wind catcher database in relation to the publication life-cycle (Fig.3). The trends that can be observed in this figure are used to assist in the identification of knowledge gaps, as described in the next section.

4.3. Knowledge gaps and research direction discovery

In this study, a distinction is made between the research gap, as a research question or problem that has not yet been addressed, and information regarding specific characteristics of a research methodology compared to other methodologies that can be implemented to investigate a specific keyword. Clustering based on research methodology can lead to finding potential areas for future research, or support ongoing research in specific fields. The absence or presence of co-occurrence and correlation can play a role in multiple ways, for example:

a. There is one cluster including keywords that have no high co-occurrence and correlation with any of the research methodologies (Cluster 9), i.e. these keywords have not yet been investigated using any of the research methodologies. As all relevant research methodologies have been taken into account in clustering, these keywords and phrases can be assessed by the experts to find new fields in wind catcher studies and therefore, it can be used as a guide for researchers to decide about research directions to fill knowledge gaps, find innovative design solutions, and prolong the growth level in the publication life-cycle in this field. Some of the knowledge gaps in wind catcher studies are identified as follows:
    - The phrase “wind comfort” in Cluster 9 indicates that issues related to wind discomfort have not yet been considered in wind catcher studies.
    - Knowledge of convective heat transfer in building spaces is required to improve occupant thermal comfort and indoor air quality. CFD has been used to investigate the natural convection in wind catchers (Cluster 6). However, “forced convection” (Cluster 9) has not yet been investigated. Note

that research on building energy and building component durability is dependent on detailed information of the local and mean interior and exterior forced convective heat coefficient (CHTC) (Montazeri, Blocken, Derome, Carmeliet, & Hensen, 2015). Using inappropriate models to calculate CHTC can lead to considerable errors in Building Energy Simulation (BES).

- The potential of integrating (small) "wind turbines" in wind catchers has not yet been investigated. Therefore, it would be interesting to evaluate wind catchers as an energy-harvesting technology.
- The word "greenhouse" gives an indication that the use of wind catchers for buildings in different applications has not been studied and it can be another suggestion for future studies about this technology.

Most of the knowledge gaps that can be extracted from this study correspond with those identified by a recent conventional (i.e. man-made) literature review on wind catcher technologies as suggestions for further studies (Saadatian et al., 2012), however, our outcomes are more specific.

b. Each research methodology may have advantages and disadvantages compared to other methodologies. In this study, clusters with co-occurrence or correlation with the research methodology and dictionary keywords (Clusters 1 to 8) can be useful for researchers to identify how different research methodologies have been used. This can therefore result in choosing more appropriate research methodologies for a specific application. For example, research on the impact of evaporative cooling in enhancing the cooling performance of wind catchers has been performed using full-scale measurements. Given the advantages of CFD, the results of the current study suggests the possibility of using CFD as a powerful tool for the investigation of water spray systems in wind catchers. It is expected that the results by CFD will lead to new insights and information that could not have been obtained with full-scale measurements.

c. In some occasions, different research methodologies are used in a complementary way. For example, Clusters 1 and 2 provide useful information on whether experimental data (e.g. full-scale or wind tunnel) is available in the literature for validating the mathematical models developed for a specific application. For example, wind-tunnel data for three-sided, six-sided and twelve-sided wind catchers (Cluster 1), and full-scale data for water spray systems in wind catchers (Cluster 2).

5. Discussion and conclusions

5.1 Summary of findings

This paper has introduced a three-step methodological framework for science foresight analysis, consisting of (i) life-cycle analysis, (ii) text mining and (iii) knowledge gap identification by means of automated clustering. We have demonstrated that this approach can provide useful information for evaluating possible opportunities for new research and development activities in a specific field of study. Although the individual methods have been applied before in the context of foresight studies, the main contribution of this study lies in their combined integration in one framework.

With the application of text mining along with the development of a technology hierarchy and cluster analysis, relevant keywords and their interactions in scientific literature are tracked during the development trajectory of the research field. By combining this (quantitative) automated analysis (i.e. presence or absence of co-occurrence and correlation) with (qualitative) expert input, knowledge gaps in this field are discovered in an efficient way, and potential areas for future research and development are specified (Section 4.3). The element that links these

three steps together is the research methodology of the scientific publications. First, it allows to gain insight into the influence of methodological trends on the characteristic development and scientific attention over a technology's life-cycle. Moreover, it efficiently points the attention of analysts towards areas of research that have not previously been explored.

5.2 Reflection and future perspectives

The application potential of this framework was demonstrated for the case of wind catchers; a sustainable natural ventilation system for buildings. However, the approach was designed in a generic way that also enables its use for science foresight activities in many other fields of science and engineering. Due to the size of the database (92 papers) it was possible to manually crosscheck some of the results.

Given the importance of the impact of textual data on the accuracy of text mining, a sensitivity analysis was carried out for three cases, where (i) title, (ii) title, abstract and keywords, and (iii) full-text of the papers were considered. The results show that text mining using full-texts identifies the research methodology of the papers with an accuracy of about 93%, while that is about 68% for the combination of title, abstract and keywords, and about 13% for just the title. It is worth noting that the majority of recently published text mining studies (Table 1) are not based on the full-text of publications. There are other considerations, such as data storage, computational requirements, and the lack of easy access to full-texts that may favour the use of only abstracts, titles and keywords. Nevertheless, even though descriptive terms (keywords) are more sparsely spread in full articles compared to the compressed format of abstracts (Shah, Perez-Iratxeta, Bork, & Andrade, 2003), it seems worth considering in cases when high accuracy is desired.

Although the relatively small database of the case study was very useful for the development and testing of the science foresight framework, it can also be argued that its scope is somewhat small for highlighting the true practical value of algorithm-driven science foresight approaches. It is expected that the benefits of science foresight become more pronounced for domains with a larger volume of publications. In such applications, the network of connections between sub-domains is more complex, and their interactions are more difficult to oversee for human beings. Automated clustering analysis would then be more likely to expose previously unexplored relationships.

Results in the present study rely to a large extent on the tight coupling between automated analysis and expert input. The quality of the results will thus be influenced by the knowledge of domain experts. For example, selecting the relevant keywords and developing the dictionary or defining the optimal number of clusters in k-means algorithm depends strongly on experts' input. In addition, after the clusters are finalized, an analysis is needed to be done on each cluster by experts to determine their characteristics. To make the methodology more robust, it would be interesting to investigate ways of making it less dependent on the subjective bias that is introduced under the influence of human decision-makers. For example in the SOM method, the optimal number of clusters can be determined by heuristic models. Besides, in this method, the characteristics and features of each cluster could be determined automatically.

Acknowledgements

Hamid Montazeri is currently a postdoctoral fellow of the Research Foundation – Flanders (FWO) and is grateful for its financial support (project FWO 12M5316N).

References

Abbott, P. A., Foster, J., Marin, H. d. F., & Dykes, P. C. (2014). Complexity and the science of implementation in health IT—Knowledge gaps and future visions. *International Journal of Medical Informatics, 83*(7), e12-e22. doi: http://dx.doi.org/10.1016/j.ijmedinf.2013.10.009

Andrade, M. A., & Bork, P. (2000). Automated extraction of information in molecular biology. *FEBS letters, 476*(1), 12-17.

Arroyabe, M. F., Arranz, N., & de Arroyabe, J. C. F. (2015). R&D partnerships: An exploratory approach to the role of structural variables in joint project performance. *Technological Forecasting and Social Change, 90*, 623-634.

Bahadori, M., Mazidi, M., & Dehghani, A. (2008). Experimental investigation of new designs of wind towers. *Renewable Energy, 33*(10), 2273-2281.

Bahadori, M. N. (1985). An improved design of wind towers for natural ventilation and passive cooling. *Solar Energy, 35*(2), 119-129.

Bahadori, M. N. (1994). Viability of wind towers in achieving summer comfort in the hot arid regions of the middle east. *Renewable Energy, 5*(5–8), 879-892. doi: http://dx.doi.org/10.1016/0960-1481(94)90108-2

Baloglu, S., & Assante, L. M. (1999). A content analysis of subject areas and research methods used in five hospitality management journals. *Journal of Hospitality & Tourism Research, 23*(1), 53-70.

Bansal, N., Mathur, R., & Bhandari, M. (1994). A study of solar chimney assisted wind tower system for natural ventilation in buildings. *Building and environment, 29*(4), 495-500.

Bengisu, M., & Nekhili, R. (2006). Forecasting emerging technologies with the aid of science and technology databases. *Technological Forecasting and Social Change, 73*(7), 835-844. doi: http://dx.doi.org/10.1016/j.techfore.2005.09.001

Berloznik, R., & Van Langenhove, L. (1998). Integration of technology assessment in R&D management practices. *Technological Forecasting and Social Change, 58*(1), 23-33.

Bezdek, J. C., & Pal, N. R. (1998). Some new indexes of cluster validity. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 28*(3), 301-315. doi: 10.1109/3477.678624

Blocken, B. (2014). 50 years of Computational Wind Engineering: Past, present and future. *Journal of Wind Engineering and Industrial Aerodynamics, 129*, 69-102.

Calautit, J. K., Hughes, B. R., & Ghani, S. A. (2013). A numerical investigation into the feasibility of integrating green building technologies into row houses in the Middle East. *Architectural Science Review, 56*(4), 279-296.

Calautit, J. K., O'Connor, D., & Hughes, B. R. (2014). Determining the optimum spacing and arrangement for commercial wind towers for ventilation performance. *Building and environment, 82*, 274-287.

Campani, M., & Vaglio, R. (2014). A simple interpretation of the growth of scientific/technological research impact leading to hype-type evolution curves. *arXiv preprint arXiv:1410.8685*.

Chemchem, A., & Drias, H. (2015). From data mining to knowledge mining: Application to intelligent agents. *Expert Systems with Applications, 42*(3), 1436-1445.

Choi, S., Park, H., Kang, D., Lee, J. Y., & Kim, K. (2012). An SAO-based text mining approach to building a technology tree for technology planning. *Expert Systems with Applications, 39*(13), 11443-11455. doi: http://dx.doi.org/10.1016/j.eswa.2012.04.014

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., & Carrillo, P. M. (2009). The needs and benefits of Text Mining applications on Post-Project Reviews. *Computers in Industry, 60*(9), 728-740. doi: http://dx.doi.org/10.1016/j.compind.2009.05.006

Clarke, J., & Hensen, J. (2015). Integrated building performance simulation: Progress, prospects and requirements. *Building and environment, 91*, 294-306.

Coates, J. F. (1985). Foresight in federal government policy making. *Futures Research Quarterly, 1*(2), 29-53.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology, 63*(8), 1609–1630. doi: 10.1002/asi.22688

Coccia, M. (2009). What is the optimal rate of R&D investment to maximize productivity growth? *Technological Forecasting and Social Change, 76*(3), 433-446. doi: http://dx.doi.org/10.1016/j.techfore.2008.02.008

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change, 73*(8), 981-1012.

de Miranda Santo, M., Coelho, G. M., dos Santos, D. M., & Fellows Filho, L. (2006). Text mining as a valuable tool in foresight exercises: A study on nanotechnology. *Technological Forecasting and Social Change, 73*(8), 1013-1027.

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications, 34*(3), 1707-1720. doi: http://dx.doi.org/10.1016/j.eswa.2007.01.035

Dubarić, E., Giannoccaro, D., Bengtsson, R., & Ackermann, T. (2011). Patent data as indicators of wind power technology development. *World Patent Information, 33*(2), 144-149.
Ernst, H. (1997). The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. *Small Business Economics, 9*(4), 361-381.
Ghadiri, M. H., Lukman, N., Ibrahim, N., & Mohamed, M. F. (2013). Computational Analysis of Wind-Driven Natural Ventilation in a Two Sided Rectangular Wind Catcher. *International Journal of Ventilation, 12*(1), 51-61.
Ghazinoory, S., Ameri, F., & Farnoodi, S. (2013). An application of the text mining approach to select technology centers of excellence. *Technological Forecasting and Social Change, 80*(5), 918-931. doi: http://dx.doi.org/10.1016/j.techfore.2012.09.001
Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management, 41*(6), 1548-1572. doi: http://dx.doi.org/10.1016/j.ipm.2005.03.021
Greenacre, M., & Hastie, T. (2010). Dynamic visualization of statistical learning in the context of high-dimensional textual data. *Web Semantics: Science, Services and Agents on the World Wide Web, 8*(2–3), 163-168. doi: http://dx.doi.org/10.1016/j.websem.2010.03.007
Hassan, A. M., & Lee, H. (2014). A theoretical approach to the design of sustainable dwellings in hot dry zones: A Toshka case study. *Tunnelling and Underground Space Technology, 40*, 251-262.
Hsu, F.-C., Trappey, A. J., Trappey, C. V., & Hou, J.-L. (2006). Technology and knowledge document cluster analysis for enterprise R&D strategic planning. *International Journal of Technology Management, 36*(4), 336-353.
Huang, C., Notten, A., & Rasters, N. (2011). Nanoscience and technology publications and patents: a review of social science studies and search strategies. *The Journal of Technology Transfer, 36*(2), 145-172.
Hughes, B. R., Calautit, J. K., & Ghani, S. A. (2012). The development of commercial wind towers for natural ventilation: A review. *Applied Energy, 92*, 606-627.
IEA. (2013). Transition to Sustainable Buildings - Strategies and Opportunities to 2050. *International Energy Agency*.
Iniyan, S., & Sumathy, K. (2003). The application of a Delphi technique in the linear programming optimization of future renewable energy options for India. *Biomass and Bioenergy, 24*(1), 39-50.
Jun, S., Park, S.-S., & Jang, D.-S. (2014). Document clustering method using dimension reduction and support vector clustering to overcome sparseness. *Expert Systems with Applications, 41*(7), 3204-3212. doi: http://dx.doi.org/10.1016/j.eswa.2013.11.018
Kajikawa, Y., Yoshikawa, J., Takeda, Y., & Matsushima, K. (2008). Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. *Technological Forecasting and Social Change, 75*(6), 771-782. doi: http://dx.doi.org/10.1016/j.techfore.2007.05.005
Kalantar, V. (2009). Numerical simulation of cooling performance of wind tower (Baud-Geer) in hot and arid region. *Renewable Energy, 34*(1), 246-254. doi: http://dx.doi.org/10.1016/j.renene.2008.03.007
Karakatsanis, C., Bahadori, M. N., & Vickery, B. (1986). Evaluation of pressure coefficients and estimation of air flow rates in buildings employing wind towers. *Solar Energy, 37*(5), 363-374.
Khan, N., Su, Y., & Riffat, S. B. (2008). A review on wind driven ventilation techniques. *Energy and buildings, 40*(8), 1586-1604.
Kidwell, D. K. (2013). Principal investigators as knowledge brokers: A multiple case study of the creative actions of PIs in entrepreneurial science. *Technological Forecasting and Social Change, 80*(2), 212-220.
Kim, C., Choe, S., Choi, C., & Park, Y. (2008). A systematic approach to new mobile service creation. *Expert Systems with Applications, 35*(3), 762-771. doi: http://dx.doi.org/10.1016/j.eswa.2007.07.044
Kolokotsa, D., Rovas, D., Kosmatopoulos, E., & Kalaitzakis, K. (2011). A roadmap towards intelligent net zero- and positive-energy buildings. *Solar Energy, 85*(12), 3067-3084.
Kostoff, R. N. (2008). Literature-Related Discovery (LRD): Introduction and background. *Technological Forecasting and Social Change, 75*(2), 165-185. doi: http://dx.doi.org/10.1016/j.techfore.2007.11.004
Kostoff, R. N. (2011). Literature-related discovery: Potential treatments and preventatives for SARS. *Technological Forecasting and Social Change, 78*(7), 1164-1173. doi: http://dx.doi.org/10.1016/j.techfore.2011.03.022
Kostoff, R. N., Johnson, D., Bowles, C. A., Bhattacharya, S., Icenhour, A. S., Nikodym, K., . . . Dodbele, S. (2007). Assessment of India's research literature. *Technological Forecasting and Social Change, 74*(9), 1574-1608. doi: http://dx.doi.org/10.1016/j.techfore.2007.02.009
Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *Engineering Management, IEEE Transactions on, 48*(2), 132-143. doi: 10.1109/17.922473
Kundu, A., Jain, V., Kumar, S., & Chandra, C. (2015). A journey from normative to behavioral operations in supply chain management: A review using Latent Semantic Analysis. *Expert Systems with Applications, 42*(2), 796-809. doi: http://dx.doi.org/10.1016/j.eswa.2014.08.035

Larsen, P. O., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*(3), 575-603.
Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation, 29*(6–7), 481-497. doi: http://dx.doi.org/10.1016/j.technovation.2008.10.006
Leydesdorff, L., Cozzens, S., & Van den Besselaar, P. (1994). Tracking areas of strategic importance using scientometric journal mappings. *Research Policy, 23*(2), 217-229. doi: http://dx.doi.org/10.1016/0048-7333(94)90054-X
Li, L., & Mak, C. (2007). The assessment of the performance of a windcatcher system using computational fluid dynamics. *Building and environment, 42*(3), 1135-1141.
Liew, W. T., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry, 65*(3), 393-400. doi: http://dx.doi.org/10.1016/j.compind.2014.01.004
Loonen, R., Singaravel, S., Trčka, M., Cóstola, D., & Hensen, J. (2014). Simulation-based support for product development of innovative building envelope components. *Automation in Construction, 45*, 86-95.
Martin, B. R. (1995). Foresight in science and technology. *Technology analysis & strategic management, 7*(2), 139-168.
Martin, B. R. (2010). The origins of the concept of 'foresight' in science and technology: An insider's perspective. *Technological Forecasting and Social Change, 77*(9), 1438-1447. doi: http://dx.doi.org/10.1016/j.techfore.2010.06.009
Meyer, P. S., Yung, J. W., & Ausubel, J. H. (1999). A primer on logistic growth and substitution: the mathematics of the Loglet Lab software. *Technological Forecasting and Social Change, 61*(3), 247-271.
Montazeri, H. (2011). Experimental and numerical study on natural ventilation performance of various multi-opening wind catchers. *Building and Environment, 46*(2), 370-378.
Montazeri, H., & Azizian, R. (2008). Experimental study on natural ventilation performance of one-sided wind catcher. *Building and Environment, 43*(12), 2193-2202.
Montazeri, H., & Azizian, R. (2009). Experimental study on natural ventilation performance of a two-sided wind catcher. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 223*(4), 387-400.
Montazeri, H., & Blocken, B. (2013). CFD simulation of wind-induced pressure coefficients on buildings with and without balconies: validation and sensitivity analysis. *Building and Environment, 60*, 137-149.
Montazeri, H., Blocken, B., Derome, D., Carmeliet, J., & Hensen, J. (2015). CFD analysis of forced convective heat transfer coefficients at windward building facades: influence of building geometry. *Journal of Wind Engineering and Industrial Aerodynamics, 146*, 102-116.
Montazeri, H., Blocken, B., & Hensen, J. (2015). Evaporative cooling by water spray systems: CFD simulation, experimental validation and sensitivity analysis. *Building and environment, 83*, 129-141.
Montazeri, H., Blocken, B., & Hensen, J. L. (2015). CFD analysis of the impact of physical parameters on evaporative cooling by a mist spray system. *Applied Thermal Engineering, 75*, 608-622.
Montazeri, H., Montazeri, F., Azizian, R., & Mostafavi, S. (2010). Two-sided wind catcher performance evaluation using experimental, numerical and analytical modeling. *Renewable Energy, 35*(7), 1424-1435.
Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology, 54*(13), 1237-1249.
Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications, 42*(3), 1314-1324. doi: http://dx.doi.org/10.1016/j.eswa.2014.09.024
Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*(16), 7653-7670.
No, H. J., An, Y., & Park, Y. (2014). A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining. *Technological Forecasting and Social Change*(0). doi: http://dx.doi.org/10.1016/j.techfore.2014.04.007
Nouanégué, H., Alandji, L., & Bilgen, E. (2008). Numerical study of solar-wind tower systems for ventilation of dwellings. *Renewable Energy, 33*(3), 434-443.
Ogawa, T., & Kajikawa, Y. (2015). Assessing the industrial opportunity of academic research with patent relatedness: A case study on polymer electrolyte fuel cells. *Technological Forecasting and Social Change, 90, Part B*, 469-475. doi: http://dx.doi.org/10.1016/j.techfore.2014.04.002
Pearlmutter, D., Erell, E., Etzion, Y., Meir, I., & Di, H. (1996). Refining the use of evaporation in an experimental down-draft cool tower. *Energy and buildings, 23*(3), 191-197.

Pereira, J. C., & Escuder, M. M. L. (1999). The scenario of Brazilian health sciences in the period of 1981 to 1995. *Scientometrics*, *45*(1), 95-105.
Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719-745.
Robinson, D., Ruivenkamp, M., & Rip, A. (2007). Tracking the evolution of new and emerging S&T via statement-linkages: Vision assessment in molecular machines. *Scientometrics*, *70*(3), 831-858.
Saadatian, O., Haw, L. C., Sopian, K., & Sulaiman, M. (2012). Review of windcatcher technologies. *Renewable and Sustainable Energy Reviews*, *16*(3), 1477-1495.
Saritas, O., & Burmaoglu, S. (2015). The evolution of the use of Foresight methods: a scientometric analysis of global FTA research output. *Scientometrics*, *105*(1), 497-508.
Shah, P. K., Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2003). Information extraction from full text scientific articles: Where are the keywords? *BMC bioinformatics*, *4*(1), 20.
Singh, N., Hu, C., & Roehl, W. S. (2007). Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development. *International Journal of Hospitality Management*, *26*(1), 131-147.
Soutullo, S., Sanchez, M., Olmedo, R., & Heras, M. (2011). Theoretical model to estimate the thermal performance of an evaporative wind tower placed in an open space. *Renewable Energy*, *36*(11), 3023-3030.
Soutullo, S., Sanjuan, C., & Heras, M. R. (2012). Energy performance evaluation of an evaporative wind tower. *Solar Energy*, *86*(5), 1396-1410. doi: http://dx.doi.org/10.1016/j.solener.2012.02.001
Su, H.-N., & Lee, P.-C. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, *85*(1), 65-79.
Su, Y., Riffat, S. B., Lin, Y.-L., & Khan, N. (2008). Experimental and CFD study of ventilation flow rate of a Monodraught™ windcatcher. *Energy and buildings*, *40*(6), 1110-1116.
Sunikka, A., & Bragge, J. (2012). Applying text-mining to personalization and customization research literature – Who, what and where? *Expert Systems with Applications*, *39*(11), 10049-10058. doi: http://dx.doi.org/10.1016/j.eswa.2012.02.042
Thorleuchter, D., & Van den Poel, D. (2013). Web mining based extraction of problem solution ideas. *Expert Systems with Applications*, *40*(10), 3961-3969. doi: http://dx.doi.org/10.1016/j.eswa.2013.01.013
Trappey, C. V., Trappey, A. J., & Wu, C.-Y. (2010). Clustering patents using non-exhaustive overlaps. *Journal of Systems Science and Systems Engineering*, *19*(2), 162-181.
Trappey, C. V., Wu, H.-Y., Taghaboni-Dutta, F., & Trappey, A. J. C. (2011). Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics*, *25*(1), 53-64. doi: http://dx.doi.org/10.1016/j.aei.2010.05.007
Trappey., C. V., Wu, H.-Y., Taghaboni-Dutta, F., & Trappey, A. J. C. (2011). Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics*, *25*(1), 53-64. doi: http://dx.doi.org/10.1016/j.aei.2010.05.007
Wang, M.-Y., Fang, S.-C., & Chang, Y.-H. (2015). Exploring technological opportunities by mining the gaps between science and technology: Microalgal biofuels. *Technological Forecasting and Social Change*, *92*(0), 182-195. doi: http://dx.doi.org/10.1016/j.techfore.2014.07.008
Weiss, S. M., Indurkhya, N., & Zhang, T. (2010). *Fundamentals of predictive text mining*: Springer Science & Business Media.
Wu, W., Tang, X.-P., Yang, C., Liu, H.-B., & Guo, N.-J. (2013). Investigation of ecological factors controlling quality of flue-cured tobacco (Nicotiana tabacum L.) using classification methods. *Ecological Informatics*, *16*, 53-61.
Yoon, B., Park, I., & Coh, B.-y. (2014). Exploring technological opportunities by linking technology and products: Application of morphology analysis and text mining. *Technological Forecasting and Social Change*, *86*(0), 287-303. doi: http://dx.doi.org/10.1016/j.techfore.2013.10.013
Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change*, *72*(2), 145-160. doi: http://dx.doi.org/10.1016/j.techfore.2004.08.011