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

M. Rezaeian a,⁎, H. Montazeri b,c, R.C.G.M. Loonen b

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

Abstract

Science foresight comprises a range of methods to analyze 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 analyzing 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.

1. Introduction

Effective management and planning of research and development activities require strategic allocation of available resources (Arroyabe et al., 2015; Berloznik and 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 et al., 2008; Kostoff, 2008; Kostoff and Schaller, 2001; Leydesdorff et al., 1994; Ogawa and 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; Yoon and Park, 2005). 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 et al., 2003; Porter and Rafols, 2009). In addition, the information that is documented in patents, reports and research papers continues to grow in size at an exponential rate (Kajikawa et al., 2008; Kostoff and Schaller, 2001; Larsen and Von Ins, 2010; Bengisu and Nekhili, 2006). 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 (Martin, 1995; Martin, 2010). It has successfully been implemented in different fields, such as economy (Nassirtoussi et al., 2014), environmental science (Dubarri et al., 2011; Inyian and Sumathy, 2003), foresight (Saritas and

http://dx.doi.org/10.1016/j.techfore.2017.02.027
0040-1625/© 2017 Elsevier Inc. All rights reserved.

Please cite this article as: Rezaeian, M., et al., Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.027
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 maturity stage is entered. The saturation stage indicates limited growth with only few additional patent applications (Trappey et al., 2011a).

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 (Trappey et al., 2011a; Campani and Vaglio, 2014; Trappey et al., 2010).

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 et al., 1999).

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

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

where \(K\) is the asymptotic limit that the growth curve approaches and shows the saturation level of the growth, \(\Delta t\) 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).

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, \(\Delta t\), \(t_m\) and \(K\) can be determined. Further information is presented by Meyer 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):

\[
N(t) = N_1(t) + N_2(t)
\]

2.2. 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 et al., 2009; Delen and Crossland, 2008; Ghazinoory et al., 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 et al., 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.

2.2.1. 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 (Dubarí et al., 2011; de Miranda Santo et al., 2006; Trappey et al., 2011a; Daim et al., 2006; Trappey et al., 2011b; Andrade and Bork, 2005). 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.

The sensitivity analysis is carried out in three steps: First, a dictionary of 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).

\[ T_{ij} = \sum_{m=1}^{T(KP_j)} KPF_{jm} \]  

where \( T_{ij} \) is the frequency of research methodology \( i \) in paper \( j \), \( T(KP_j) \) is the number of methodology keywords (related to research methodology \( j \)) in paper \( j \), and \( KPF_{jm} \) is the frequency of keyword \( m \) (related to research methodology \( j \)) 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.

2.2.2. 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.

2.2.2.1. 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 conducted, the tokens are filtered by the experts to determine which tokens are relevant and meaningful keywords. These keywords are used to make a dictionary.

| Year | Ref. | Field of study | Database | Text mining technique |
|------|------|---------------|----------|-----------------------|
| 2005 | (Glänzel, Jansens, & De Moor, 2005) | Scientometrics | Paper | Bibliometric analysis, categorization, clustering |
| 2005 | (Yoon & Park, 2005) | TFT-LCD | Patent | Factor analysis |
| 2006 | (de Miranda Santo, Coelho, dos Santos, & Fellows Filho, 2006) | Nanotechnology | Paper | Bibliometric analysis, text analysis, visualization |
| 2006 | (Hsu, Trappey, Trappey, & Hou, 2006) | Hand tool industry | Patent | Information extraction, clustering |
| 2007 | (Singh, Hu, & Roehl, 2007) | Human resource management | Paper | Clustering, bibliometric analysis |
| 2007 | (Kostoff et al., 2007) | Technical articles | Paper | Clustering, clustering, bibilometric analysis |
| 2008 | (Kim, Choi, 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 |
| 2009 | (Delen & Crossland, 2008) | Management information system | Paper | Clustering |
| 2009 | (Choudhary, Ouikpe, Harding, & Carrillo, 2009) | Construction | Post project reports | Information retrieval and extraction, text analysis, categorization, summarization, visualization |
| 2010 | (Su & Lee, 2010) | Technology foresight | Paper | Visualization |
| 2010 | (Greenacre & Hastie, 2010) | Journal vaccine | Paper | Clustering, visualization |
| 2011 | (Kostoff, 2011) | Severe acute respiratory syndrome | Paper | Clustering, auto-correlation mapping, factor analysis |
| 2011 | (Trappey, Wu, Taghaboni-Dutta, & Trappey, 2011a) | 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 | (Sumilka & 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, & Jiang, 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 |
2.2.2. Expert’s judgment

2.2.2.2.1. 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.

2.2.2.2.2. 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 and 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.

2.2.2.3. 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, \( F_{ij} \) is the frequency of keyword \( i \), in paper \( j \).

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 Trappey et al. (2011b), 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 \( f_{ik} \) 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 \( d_f \) 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 (Trappey et al., 2010).

\[
NTF_{ik} = \frac{f_{ik}}{d_f} \times \frac{1}{WN_k} \times \log_2 \left( \frac{n}{d_f} \right)
\]  

2.2.2.4. Cluster analysis

2.2.2.4.1. 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 (Dubarič et al., 2011; de Miranda Santo et al., 2006; Trappey et al., 2011a; Daim et al., 2006; Andrade and Bork, 2000). 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 Table 2.
in Section 4.3. 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 Rand is the correlation between keyword i, and the group of keywords related to methodology j.

2.2.2.4.2. 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 (de Miranda Santo et al., 2006; Chemchem and Drias, 2015). 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 and 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 (Kolokotsa et al., 2011; IEA, 2013). Research and development of innovative building systems is expected to play an important role in facilitating this transition towards sustainable building design (Loonen et al., 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 et al., 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 (Saadatian et al., 2012; Montazeri and Azizian, 2009; Montazeri, 2011; Montazeri and Azizian, 2008; Montazeri et al., 2010; Bahadori et al., 2008; Soutullo et al., 2011; Hassan and Lee, 2014; Su et al., 2008; Calautit et al., 2013; Calautit et al., 2014). In addition to academic contexts, industry has also paid attention to the development of these systems. These technology transfer activities have led to the introduction of multiple new commercial wind catcher systems that meet the requirements of modern-day building design (Hughes et al., 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. Database

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 by 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; consisting 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 101 is equal to Δt, and the time at 106 is the point of inflection (tmm). 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 5% 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.

4.2. Text mining

4.2.1. Impact of textual data on text mining

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

| Methodology 1 | Methodology 2 | … | Methodology N |
|---------------|---------------|---|---------------|
| KP 1          | R1,1          | … | R1,N          |
| KP 2          | R2,1          | … | R2,N          |
| KP 3          | R3,1          | … | R3,N          |
| …             | …             | … | …             |
| KP N          | Rm,1          | … | Rm,N          |

Please cite this article as: Rezaeian, M., et al., Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.027
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

Please cite this article as: Rezaeian, M., et al., Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.027

Fig. 5. Wind catcher-publication hierarchy.
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 (Bézdek and Pal, 1998; Wu et al., 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.

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. 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 2.2.2.2.2. 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 multi-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 to 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.

#### 4.2.3.1. 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 only performed in a limited set of points in space (Montazeri and 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. Montazeri et al., 2010; Ghadiri et al., 2013). 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.

### Table 5

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

#### Table 6

| Cluster | Methodology | Representative phrases |
|---------|-------------|------------------------|
| 1       | WT          | Cylindrical, Three-sided, Six-sided, Twelve-sided |
| 2       | FS          | Plastic, Residential, Day, Thermal comfort, Night, Mild, Evaporative cooling, Water spray, Air quality, Shadow, Warm, Glass, Lighting |
| 3       | CFD         | Apartment, Rectangular, Louvre, Turbulence intensity, Buoyancy, Natural convection |
| 4       | BPS, AFN & AN | Humid, Conduction, Cold |
| 5       | WT + CFD    | Cubic, Damper, Metal, Industrial, Four-sided, Wind speed, Wind velocity, Velocity, Incompressible |
| 6       | WT + CFD + (BPS, AFN & AN) | Rural, Wood, Two-sided, One-sided, Solar panel, PV panel |
| 7       | WT + CFD + (BPS, AFN & AN) | Air flow, Pressure, Cp, 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 |

---

Please cite this article as: Rezaeian, M., et al., Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.027
4.2.3.2. 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 and 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 cooling” 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 (Montazeri et al., 2015a; Montazeri et al., 2015b).

4.2.3.3. 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 “louvres” (Cluster 3) or “damper” (Cluster 5) is widely investigated using CFD.

4.2.3.4. 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 and 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 et al., 2015c). 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 favor 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 et al., 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). Andrade, M.A., Bork, P., 2000. Automated extraction of information in molecular biology. FEBS Lett. 476, 12–17.

Aroyave, M.E., Anzana, N., de Aroyave, J.C.F., 2015. R&D partnerships: an exploratory approach to the role of structural variables in joint project performance. Technol. Forecast. Soc. Chang. 90, 623–634.

Bahadori, M.N., 1985. An improved design of wind towers for natural ventilation and passive cooling. Sol. Energy 35, 119–129.

Bahadori, M.N., 1994. Viability of wind towers in achieving summer comfort in the hot arid regions of the Middle East. Renew. Energy 5, 879–892.

Bahadori, M., Mazidi, M., Dehghani, A., 2008. Experimental investigation of new designs of wind towers. Renew. Energy 33, 2273–2281.

Baloglu, S., Assante, L.M., 1999. A content analysis of subject areas and research methods used in five hospitality management journals. J. Hosp. Tour. Res. 23, 53–70.

Bansal, N., Mathur, R., Bhardari, M., 1994. A study of solar chimney assisted wind tower system for natural ventilation in buildings. Build. Environ. 29, 495–500.

Bengisu, M., Nekhli, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technol. Forecast. Soc. Chang. 73, 835–844.

Berloznik, R., Van Langenhove, L., 1998. Integration of technology assessment in R&D management practices. Technol. Forecast. Soc. Chang. 58, 23–33.

Bezek, J.C., Pal, N.R., 1998. Some new indexes of cluster validity. IEEE Trans. Syst. Man Cybern. B Cybern. 28, 301–315.

Blocken, B., 2014. 50 years of computational wind engineering: past, present and future. J. Wind Eng. Ind. Aerodyn. 129, 89–102.

Calautit, J.K., Hughes, B.R., Ghanii, S.A., 2013. A numerical investigation into the feasibility of integrating green building technologies into row houses in the Middle East. Archit. Sci. Rev. 56, 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. Build. Environ. 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.

Chechem, A., Drias, H., 2015. From data mining to mining knowledge: application to intelligent agents. Expert Syst. Appl. 42, 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 Syst. Appl. 39, 11443–11455.

Choudhary, A.K., Ohukum, P.I., Harding, J.A., Carrillo, P.M., 2009. The needs and benefits of text mining applications on post-project reviews. Comput. Ind. 60, 728–740.

Clarke, J., Jensen, J., 2015. Integrated building performance simulation: progress, prospect and requirements. Build. Environ. 91, 294–306.

Coates, J.F., 1985. Foresight in federal government policy making. Futur. Res. Q. 1, 29–53.

Collazo-Díez, F., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F., 2012. SciMAT: a new science mapping analysis software tool. American Society for Information Science and Technology—J. Am. Soc. Inf. Sci. Technol. 63, 1609–1630.

Coccia, M., 2009. What is the optimal rate of R&D investment to maximize productivity growth? Technol. Forecast. Soc. Chang. 76, 433–446.

Daim, T.U., Rueda, G., Martin, H., Cerdasi, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. 73, 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. Technol. Forecast. Soc. Chang. 73, 1013–1027.

Delen, D., Crossland, M.D., 2008. Seeding the survey and analysis of research literature with text mining. Expert Syst. Appl. 34, 1707–1720.

Dubart, E., Guinasso, D., Bengtson, R., Ackermann, T., 2011. Patent data as indicators of wind power technology development. World Patent Inf. 33, 144–149.

Ernst, H., 1997. The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. Small Bus. Econ. 9, 361–381.

Ghadiri, M.H., Lukman, N., Ibrahim, N., Mohamed, M.F., 2013. Computational analysis of wind-driven natural ventilation in a sided rectangular wind catcher. Int. J. Vent. 12, 51–61.

Ghazinoori, S., Ameri, F., Farnoodi, S., 2013. An application of the text mining approach to see technology centers of excellence. Technol. Forecast. Soc. Chang. 80, 918–931.

Glenisson, P., Glänzel, W., Janssens, F., De Moor, B., 2005. Combining full text and bibliometric information in mapping scientific disciplines. Int. Process. Manag. 41, 153–157.

Greenacre, M., Hastie, T., 2010. Dynamic visualization of statistical learning in the context of high-dimensional textual data. Web Semant. Sci. Serv. Agents World Wide Web 8, 163–168.

Haralick, R.M., Lee, H., 2014. A theoretical approach to the design of sustainable dwellings in hot dry zones: a Toska case study. Tunn. Undergr. Space Technol. 40, 251–262.

Hsu, F.-C., Treppey, A.J., Treppey, C.V., Hou, J.-L., 2006. Technology and knowledge document cluster analysis for enterprise R&D strategic planning. Int. J. Technol. Manag. 36, 336–353.

Huang, C., Notten, A., Rasters, N., 2011. Nanoscience and technology publications and patents: a review of social science studies and search strategies. J. Technol. Forecast. Manag. 36, 145–172.

Hughes, B.R., Calautit, J.K., Ghanii, S.A., 2012. The development of commercial wind towers for natural ventilation: a review. Appl. Energy 92, 606–627.

IEA, 2013. Transition to Sustainable Buildings - Strategies and Opportunities to 2050. International Energy Agency.

Inyany, S., Sumathy, K., 2003. The application of a Delphi technique in the linear programming optimization of future renewable energy options for India. Biomass Bioenergy 24, 39–50.

Jun, S., Park, S.-J., Jiang, D.-S., 2014. Document clustering method using dimension reduction and support vector clustering to overcome sparseness. Expert Syst. Appl. 41, 3204–3212.

Please cite this article as: Rezaeian, M., et al., Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.027
Morillo, F., Bordons, M., Gómez, I., 2003. Interdisciplinarity in science: a tentative typology.

Montazeri, H., Montazeri, F., Azizian, R., Mostafavi, S., 2010. Two-sided wind catcher performance: a systematic review. Energ. Buildings 40, 1586–1604.

Kidwell, D.K., 2013. Principal investigators as knowledge brokers: a multiple case study of the Canadian Science Policies of 4Ns in entrepreneurial science. Technol. Forecast. Soc. Chang. 78, 80–212.

Kim, C., Choe, S., Choi, C., Park, Y., 2008. A systematic approach to new mobile service creation. Expert Syst. Appl. 35, 762–771.

Kolokota, D., Rovas, D., Kosmatopoulos, E., Kalaitzakis, K., 2011. A roadmap towards intelligent net-zero and positive-energy buildings. Sol. Energy 85, 3067–3084.

Kostoff, R.N., 2008. Literature-related discovery (LRD): introduction and background. Technol. Forecast. Soc. Chang. 75, 165–185.

Kostoff, R.N., 2011. Experimental and numerical study on natural ventilation performance of one-sided wind catcher. Build. Environ. 43, 2193–2199.

Kundu, A., Jain, V., Kumar, S., Chandra, C., 2015. A journey from normative to behavioral operationalizing in supply chain management: a review using latent semantic analysis. Expert Syst. Appl. 42, 796–809.

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, 575–603.

Lee, S., Yoon, B., Park, Y., 2005. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29, 481–497.

Leydesdorff, L., Cozzens, S., Van den Besselaar, P., 1994. Tracking areas of strategic importance using scientometric journal mappings. Res. Policy 23, 217–229.

Li, L., Mak, C., 2007. The assessment of the performance of a windcatcher system using computational fluid dynamics. Build. Environ. 42, 1135–1141.

Liew, W.T., Adhiyia, A., Sinivasan, R., 2014. Sustainability trends in the process industries: a text mining-based analysis. Comput. Ind. 65, 393–400.

Loonen, R., Singaravelu, S., Trèka, M., Côstola, D., Hensen, J., 2014. Simulation-based support for product development of innovative building envelope components. Autom. Constr. 45, 86–95.

Martin, B.R., 1995. Foresight in science and technology. Tech. Anal. Strat. Manag. 7, 139–168.

Martin, B.R., 2010. The origins of the concept of ‘foresight’ in science and technology: an insider’s perspective. Technol. Forecast. Soc. Chang. 77, 1438–1447.

Meyer, P.S., Jung, J.W., Ausubel, J.H., 1999. A primer on logistic growth and substitution: the mathematics of the Loglet Lab software. Technol. Forecast. Soc. Chang. 61, 247–271.

Montazeri, H., 2011. Experimental and numerical study on natural ventilation performance of various multi-opening wind catchers. Build. Environ. 46, 370–378.

Montazeri, H., Azizian, R., 2009. Experimental study on natural ventilation performance of one-sided wind catcher. Build. Environ. 43, 2193–2199.

Montazeri, H., Azizian, R., 2008. Experimental study on natural ventilation performance of a two-sided wind catcher. Proc. Inst. Mech. Eng. Part A: J. Power Energy 223, 387–400.

Montazeri, H., Blocken, B., 2013. CFD simulation of wind-induced pressure coefficients on buildings with and without balconies: validation and sensitivity analysis. Build. Environ. 60, 147–150.

Montazeri, H., Montazeri, F., Azizian, R., Mostafavi, S., 2010. Two-sided wind catcher performance evaluation using experimental, numerical and analytical modeling. Renew. Energy 35, 1424–1435.

Montazeri, H., Blocken, B., Hensen, J.L., 2015a. CFD analysis of the impact of physical parameters on evaporative cooling by a mist spray system. Appl. Therm. Eng. 80, 608–622.

Montazeri, H., Blocken, B., Hensen, J.L., 2015b. Evaporative cooling by water spray systems: CFD simulation, experimental validation and sensitivity analysis. Build. Environ. 83, 129–141.

Montazeri, H., Blocken, B., Derome, D., Carmeliet, J., Hensen, J., 2015c. CFD analysis of forced convective heat transfer coefficients at windward building facades: influence of building geometry. J. Wind Eng. Ind. Aerodyn. 146, 102–116.

Morillo, F., Bordons, M., Gómez, I., 2003. Interdisciplinarity in science: a tentative typology of disciplines and research areas. American Society for Information Science and Technology. –> AM. Soc. Inf. Sci. Technol. 54, 1237–1249.

Moritz, 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 Syst. Appl. 42, 1314–1324.

Nassirzadeh, A.K., Aghabozorgi, S., Wah, T.Y., Ngo, D.C.L., 2014. Text mining for marketing applications. Expert Syst. Appl. 41, 7633–7670.

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.