Technological Forecasting & Social Change 117 (2017) 25–37

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

TeknoRoadmap, an approach for depicting emerging technologies

Iñaki Bildosola [a,*], Rosa María Río-Bélver [b], Gaizka Garechana [a], Ernesto Cilleruelo [a]

[a] Technology Foresight Management (TFM) Group, Industrial Organization and Management Engineering Department, University of the Basque Country (UPV/EHU), C/Alameda de Urquijo, s/n, 48013 Bilbao, Spain
[b] Technology Foresight Management (TFM) Group, Industrial Organization and Management Engineering Department, University of the Basque Country (UPV/EHU), C/Nieves Cano, 12, Vitoria, Spain

ARTICLE INFO

Article history:
Received 10 October 2016
Accepted 17 January 2017
Available online 8 February 2017

Keywords:
Technology forecasting
Bibliometrics
Technology roadmapping
Data mining
Web content mining
Cloud computing

ABSTRACT

One of the biggest challenges for current enterprises is the adoption of emerging technologies as soon as these provide competitive improvements. In this sense, several types of technology forecasting and surveillance activities are present in their daily activity. From the academic point of view, technology forecasting activities involve the combination of methods of a diverse nature, with which the technology is depicted and its potential future paths are discussed. Within this conceptual framework, the present work aims at describing a novel approach, known as TeknoRoadmap, which combines bibliometrics and technology forecasting methods to depict emerging technologies. Thus, this contribution aims to widen the scope compared to those provided by previous works within the field, and to that end, the depiction of emerging technologies is provided based on two main elements, namely: the profile of the research activity; and a complete technology roadmap. The approach combines consolidated methods such as text mining and roadmapping, and novel ones such as web content mining, with special attention given to forecasting activities. The work provides a detailed description of the steps on which the approach is structured, as well as the results of one specific application to a cutting edge emerging technology: cloud computing.

1. Introduction

The contribution of the present work falls within the framework of two different research fields: *technology forecasting* and *bibliometrics*. Nowadays, the importance of technology forecasting activities is given by the fact that they are present in many forms of the current society. For example, big companies require it in several crucial aspects, such as prioritizing research and development paths, planning new products developments and taking strategic decisions such as technology license agreements (Firat et al., 2008). In the case of small and medium enterprises (SMEs), these kinds of initiatives are less common and usually need someone to help them initiate the change and adoption of new technologies (Major and Cordey-Hayes, 2000). Together with this, governments use technology forecasting for making choices between competing alternatives in science and technology, and for linking science and technology more closely to the nation's economic and social needs (Martin and Johnston, 1999). Forecasting initiatives have relied heavily and are still mainly based on qualitative methods. These kinds of methods present several shortcomings and a strong tendency towards the use of quantitative methods is establishing itself. In fact, many authors suggest the use of multiple methods to compensate in any one approach (Salerno et al., 2008). In regards to quantitative methods included in forecasting exercises, researchers have long realized that technology monitoring and forecasting aims can be served by bibliometric analyses. Thus, bibliometrics provide a powerful source of information on emerging technologies and their potential (Porter and Detampel, 1995).

Based on the above, the present work aims combining both fields based on a novel approach, known as *TeknoRoadmap* (*TKRM*), with which a complete depiction of emerging technologies can be obtained. This approach contains methods of a quantitative and qualitative nature. Quantitative methods are represented by *bibliometrics*, which more than a method, can be regarded as a research field; *data mining*, in terms of *text mining* and *web content mining*; as well as *trend analysis*. Semi-quantitative methods are represented by *technology roadmapping* (*TRM*). Finally, qualitative methods are integrated by means of *expert assessment*, which introduces a counterpoint to the quantitative methods. The approach is based on eight steps which are divided in two phases and the specific objectives can be summarized as follows:

- Collect and organize scientific and non-scientific information related to an emerging technology.

* Corresponding author.
E-mail address: inaki.bildosola@ehu.es (I. Bildosola).

http://dx.doi.org/10.1016/j.techfore.2017.01.015
0040-1625/

- Describe the research activity profile of the technology. Specifically, the *literature profile and research community profile*; and *the state of the research and its evolution*.
- Depict the prior evolution of the technology and forecast potential paths of the short and medium-term future, identifying the necessary elements for this purpose: sub-technologies, technology applications, and links among both.
- Integrate all the information of the evolution of technology in a TRM.

The embryonic idea of the present approach was presented in Bildosola et al. (2015), where the combination and naming of the steps on which the approach is structured was motivated. In addition to this, further description of the first part of the approach (first three steps), and partial results from their application were provided in Bildosola et al. (2017). In that work, the selected technology to be depicted was cloud computing (CC), a cutting edge emerging technology with huge impact in the current enterprise. Following the logical path, the present work relies on the mentioned works and aims to delve deeper into the second part of the approach, detailing the steps that constitute it and the results of their application to CC technology, as well as presenting the complete approach as a whole. The structure of this paper includes the following parts: in "Research objective" a brief summary of the specific goals of the approach are provided. The "Background" section analyses the research fields which constitute the framework for the approach. The "Research approach: TeknoRoadmap (TKRM)" section provides a comprehensive explanation of the approach, with special attention to its second part. In this sense, each step is described thoroughly in terms of input information, output outcome and methods and tools used. In "Approach result: depicting cloud computing", the depiction of the CC technology is provided. Finally, the "Conclusions and future work" section is dedicated to the interpretation of the obtained results, as well as to describing the limitations and future lines of study that may result from this work.

2. Background

The main goal of the TKRM approach is to depict emerging technologies, with particular attention to its forecasting. Martino (1993) defined technology forecasting as a prediction of the future characteristics of useful machines, procedures, or techniques. From its origins, there have appeared many overlapping ways of forecasting technology developments and their subsequent impact, including technology intelligence, forecasting, roadmapping, assessment, and foresight (Firat et al., 2008). The existing diversity of methods belonging to forecasting activities has led to attempts to summarize and structure the field. In this sense, Popper (2008), created a Foresight Diamond which identifies 33 methods and distributes them within a diamond, based on seven attributes. It is analyzed from one side depending on whether the methods are based on evidence, expertise, interaction or creativity, and from the other side analyzing the nature of each methodology, their proximity given to be qualitative, semi-quantitative or quantitative. This schematization is a clear indicative of the diversity of methods involved in this field.

Focusing on research fields and methods included in the TKRM, bibliometrics is used for the structuring of information, which is the base for a consistent forecasting exercise. Bibliometric analysis is based on three basic principles (Kongthon, 2004): measurement of the activity by counting publications; measurement of the impact by the analysis of the citations; and measurement of the citations by the analysis of the co-citations and the use of keyword article by article. Currently, specific tools that are used in this field encompass authors, affiliations, conceptual maps, clusters and factor analysis, citation and co-citation analysis (Daim et al., 2006). For instance, bibliometrics can help researchers in mapping and profiling their entire research domain (Börner et al., 2003). Analyzing the connection between bibliometrics and technology forecasting, it should be highlighted that both fields have experienced an important expansion within the last years and this has resulted in the proliferation of numerous methods within them. In addition, it can be said that both fields have suffered an interlinking process, especially in terms of technology-foresight-approaches using bibliometrics-methods. As an indicator of the connection between the two fields the following graph (Fig. 1) shows the number of works, contained in SCOPUS and WoS databases, which contain at least the word *bibliometrics* or *scientometrics*, and the word *technology foresight* or *technology forecasting* within their title, abstract or keywords. Even though the search is not consistent enough to draw solid conclusions, the observed evolution allows the increase in the combination of both fields to be described.

With regard to the methods included in the TKRM, text mining is used to complete the bibliometrics analysis. Text mining can be defined as the process of obtaining information from text-based data. Moreover, whereas bibliometrics focuses on measurement of scientific activity to find patterns and trends, text mining goes beyond processing the content of publications (Kostoff and Geisler, 1999). The importance of text mining and its applicability is supported by the appearance and growth of dedicated software options (Mikut and Reischl, 2011). Text mining tools can be used to examine research trends and patterns in the fields of technology management, using software developed specifically for these types of knowledge mining applications (Porter et al., 2003). In this context, numerous papers can be found where text mining and bibliometrics analysis methods construct the basis for some kind of forecasting exercise (see for example: Daim et al., 2006; Bengisu and Nekhili, 2006; Kajikawa et al., 2008).

TKRM makes also use of a novel method within the data mining field: web content mining. Gök et al. (2015) presented a work which examined the practicalities and effectiveness of web content mining as a research method for innovation studies. The authors concluded that website data offers additional insights when compared with other traditional research methods. However, it is hard to find actual web content mining applications for technology forecasting, with few examples such as that presented by Thorleuchter and Van den Poel (2013). In any case, it is worth noting that this method has demonstrated its potential to provide useful forecasting information for topics such as the price of mobile phones selling online (Zhu et al., 2011). Therefore, it can be said that technology forecasting and web content mining promises to be a profitable combination, however, it is only starting to be approached.

The final outcome of the TKRM approach is a complete TRM, which is used to integrate all the information in a single visualization element. Generally speaking, TRMs have a very varied use in both scientific and professional fields. Bray and Garcia (1998) underscore the major uses and benefits derived from technology roadmapping, and within the framework of technology forecasting they highlight that roadmapping provides a mechanism to help experts forecast S&T developments within targeted areas. Regarding the use of TRMs and its combination with other methods, the usefulness of bibliometrics to support technology roadmapping was already stated by Kostoff et al. (2004), and this connection has since grown. Thus, several works which combine bibliometrics and TRMs have been carried out (see for example: Zhang et al., 2013; Yoon and Phaal, 2013; Huang et al., 2014), although not all of them perform technology forecasting.

All the results obtained by applying the approach are finally evaluated by experts. Expertise has been widely used when it comes to forecast the evolution of technologies. This is justified by several reasons: some of the key behavioral elements are included; the variety of inputs and thereby the quality of results will increase; it will lead to broader support for the results and it may contribute to the democratic character of the process (Porter et al., 2004). In fact, many of the works cited above make use of it somehow (Daim et al., 2006; Bengisu and Nekhili, 2006; Thorleuchter and Van den Poel, 2013; Zhang et al., 2013; Huang et al., 2014).

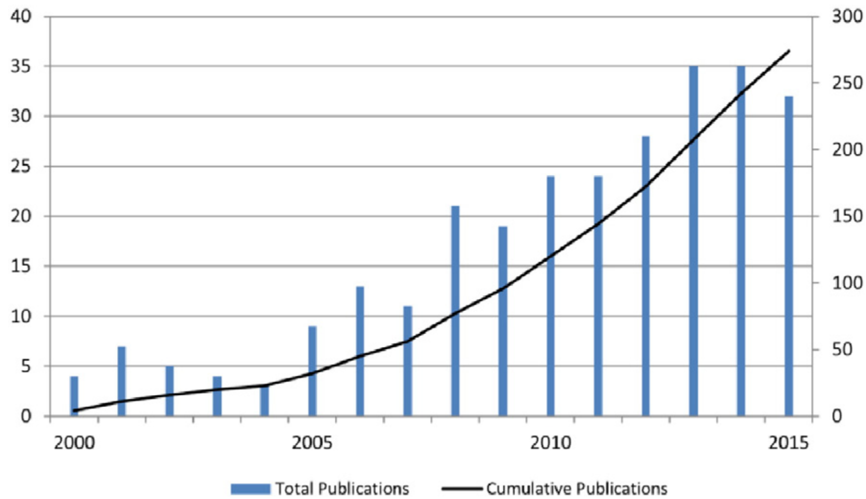

Fig 1. Number of papers which combine bibliometrics and technology forecasting research field methods for the period 2000–2015.

3. Research approach: TeknoRoadmap (TKRM)

The approach is based on eight specific steps which are designed to generate two main elements: the profile of the technology, which more accurately should be named as research activity profile, and the TRM. The first three steps are focused on obtaining the former element, i.e. answering questions such as who, where, how or what is being or has been researched and developed related to the technology. The last five steps are focused on obtaining the TRM, which is the most important outcome of the approach. Each step has a specific input and output, thus a flow of information is generated on which the approach is structured. Fig. 2 shows the whole process, identifying the specific methods and/or techniques used in each step, as well as the connection existing among the steps and the information flow through the approach. The reason for combining diverse methods is based on experts' opinions, who agree that different methods should be used when it comes to forecasting technologies. The primary reason for this is to attempt to offset the weaknesses of one method with the strengths of another (Martin and Daim, 2012). Moreover, by doing this, the forecaster avoids problems of trying to select the single best method for a particular application (Martino, 1993).

It can be observed in the figure how the first three steps are applied one after another until the profile is generated. On the other hand, those steps focused on obtaining the TRM can be applied in parallel, as they make use of the information already generated, with the exception of step six which needs to be applied after the fifth step. Finally, the final step is applied once all other steps have been completed, since it focuses on a full assessment of the work. The specific tasks to be done in each

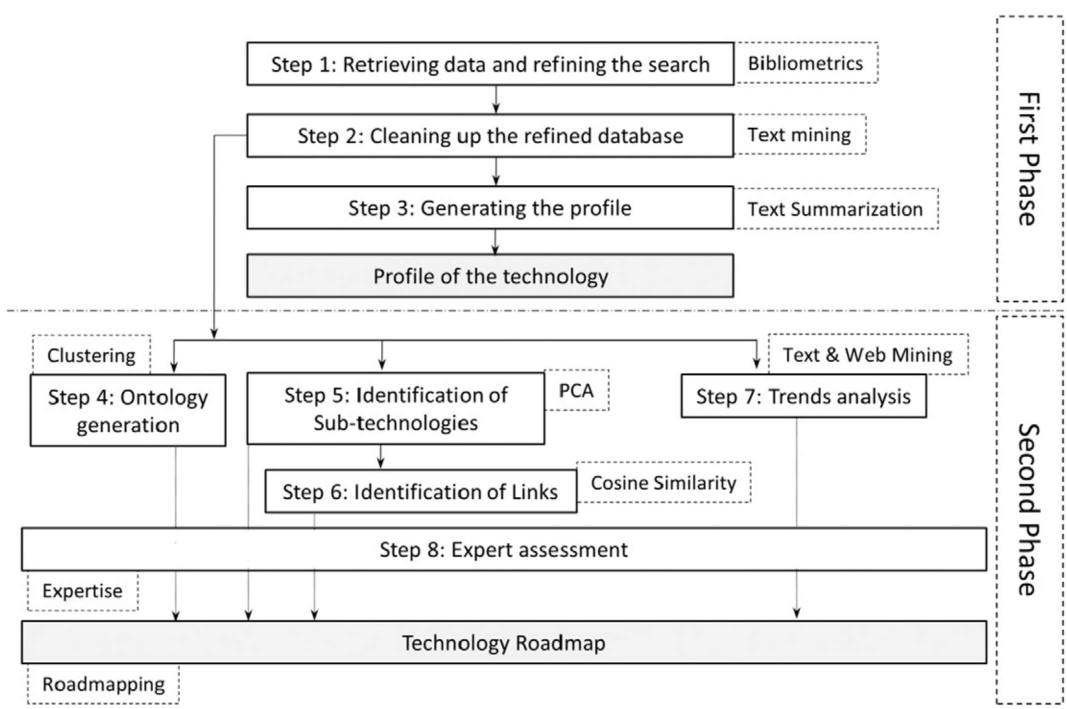

Fig 2. Approach's framework, step by step breakdown of the process and included methods.

step and the information flow are described in the following sections. It should be noted that for ease of understanding, continuous references are made to the application for the approach to CC technology; however, the goal of the approach is the ability to be applied to any kind of emerging technology.

3.1. Profile generation

The first three steps are already described in detail in Bildosola et al. (2017), therefore this work only mentions the idea behind them. All the tasks required in this phase have been carried out with the *VantagePoint* text mining tool. The first two steps have been done twice in order to generate two databases. The first of these contains scientific publications directly related with the technology basic research, whereas the second contains documents which provide specific applications of the technology. These databases will be exploited in the second phase of the approach. The main tasks required in each step can be summarized as follows:

Step 1: *retrieving the data and refining the search*. The first task is to generate a specific database consisting of scientific publications directly related to the technology being depicted. For the case of CC, the scientific databases from which the specific database was generated were Web of Science (WoS) and Scopus. This approach is inspired by the work of Porter et al. (2008), which describes how to build a set of queries which define the boundaries of the science of nanotechnology. In this case, the refinement was performed through Boolean conditions implemented by the aforementioned databases' own tools. The time range was 2008–2016.

Step 2: *Cleaning up the refined database*. The second task involves the use of a text mining tool and some of its basic functionalities. Scientific publications obtained from the previous step have to be integrated in a single database and *fuzzy matching* has to be applied to specific fields within the documents: authors, affiliations and author's keywords. This process takes all the variations which express exactly the same concept (plural, acronyms, equivalent expressions, etc.) as a single word.

Step 3: *Generating the profile*. The profile is divided in two parts: *literature profile and research community profile*; and the *state of the research and its evolution*. The first of these is centered on describing the research activity in terms of academic, private or government publication distribution, journal or conference publications, as well as top countries, authors or institutions, among others. The types of elements generated are top-ten lists, co-occurrence maps and auto-correlation maps. On the other hand, the analysis of the second part of the profile goes through the identification of the most frequent keywords, as well as the most increased ones. As regards to keywords, it should be noted that the analysis was performed with author's keyword field, discarding options such as automatically generated keywords. This field contains the keywords that, in the opinion of the author, best characterize the research performed, creating a powerful tool with which to characterize the research (Garechana et al., 2015).

3.2. TRM generation

The second phase of the approach is centered on the depiction of the technology by means of completing a TRM. A three-layer roadmap has been selected for the present approach, as the interest lies in discovering the disruption of *technologies* in their embryonic state (first layer), their evolution toward real *applications* (second layer), and the solutions that are reaching the *market* (third layer). With regard to the time-axis, the interest lies in the past (placed close to the present since the goals are emerging technologies), present, short and medium-term future phases. The steps included in this phase generate the elements needed to obtain the roadmap, namely:

- *Principal layers*: technology layer, application layer and market layer.
- Vertical breakdown of the principal layers: *main fields*.
- Past, present and future *sub-technologies* belonging to the main fields.
- Specific *terms* within the sub-technologies.
- *Links* among sub-technologies belonging to each of the principal layers.

3.2.1. Step 4: ontology generation

The inputs of this step are both databases arising from the second step. The task to be done is to generate the hierarchical structure of the technology and its applications, separately. This information makes it possible to identify the *main fields* belonging to each layer, which shall be no more than the elements for the first level of the ontology (see Fig. 4 of next chapter). This step is conducted using four specific tasks which are applied one after the other. In addition to this, the set of tasks is repeated iteratively and each of the iterations generates a deeper subdivision of the ontology, creating more sub-clusters and more levels of the hierarchical structure, until the ontology is generated.

(1) *Co-occurrence matrix generation*. This is generated based on the conjoint appearance of the authors' keywords throughout the documents. Keyword reduction is performed by taking only the most frequent keywords - those with more than three appearances - into account, in order to remove as much noise as possible from the data. In addition to this, terms such as “cloud”, “CC” or “computing” have been excluded. This kind of reduction (infrequent and keywords with too general a meaning) is a common practice in coword analysis (Bhattacharya and Basu, 1998), as cluster analysis is quite sensitive to outliers (Garechana et al., 2015). All the work was carried out with *VantagePoint*.

(2) *Distance matrix generation*. The clusters are obtained from similarity or distance matrices, which describe how similar (or different) the keywords are in terms of their co-occurrences within the documents. The construction of this kind of matrix is based on similarity measures, which need to go through a normalization process. In this sense, Eck and Waltman (2009) discuss the properties of existing measures, concluding that probability-based measures do have the desirable properties. Taking this into account, the selected measure was the indirect Salton's cosine, as defined in Van Eck and Waltman (2008): $Cos_{i,j} = \frac{\sum_{q=1}^{N} C_{iq}\ C_{jq}}{\sqrt{\sum_{q=1}^{N} C_{iq}^2\ \sum_{i=1}^{N} C_{jq}^2}}$, where C_{iq} and C_{jq} represent the co-occurrence values adopted by i and ?? terms (keywords) with the rest of the terms. In the case of this work, all the process required the exportation of the co-occurrence matrices to *R* software, where the selected measure calculation was programmed and the distance matrix was obtained.

(3) *Clustering*. The distance matrix makes it possible to identify similar keywords in the above exposed terms, thus it can be used as an input for a clustering algorithm. *R* software offers various algorithms to perform this clustering process. So, in the case of the present approach, the Agnes package (Kaufman and Rousseeuw, 2009) with *Ward* clustering method was selected, which has been used in a wide range of work related to term grouping (Janssens et al., 2006; Leydesdorff, 1987).

(4) *Naming*. Finally, the terms contained within the clusters have to be analyzed in order to name them by means of the most representative term. This is a critical task as these denominations are what appear in the ontology itself. Moreover, as previously mentioned, the naming of the clusters obtained in the first iteration (Fig. 4), generates the vertical structure of the roadmap: the *main fields*

(Figs. 6, 7, 8 and 9). The decisions taken in the present task represent the first item to be assessed by experts in step eight.

3.2.2. Step 5: identification of sub-technologies

The identification of the sub-technologies differs for each of the layers of the roadmap. For the case of the first two, the inputs are both databases arising from the second step. The first task to be done is the identification of different sub-technologies or research fields which have had a prominent presence in the research activity throughout the years studied. In this case, the generation of the groupings differs from the previous step in several points, as discussed below.

Whereas the previous clustering process was performed to all the keywords contained in the dataset (with the mentioned restrictions), for the present task the dataset has to be divided into years and the identification of the sub-technologies done year by year. The generation of the sub-technologies for each year is performed with Principal Components Analysis (PCA), which determines how frequently terms occur together in a dataset. Consequently, these sub-technologies consist in grouping terms, i.e. authors' keywords. Obtaining and representing these factors was carried out in *VantagePoint*, which makes use of the Multidimensional Scaling (MDS) technique in order to place the factors on a map and represent their relationships (Hout et al., 2013).

Once the sub-technologies are identified, the second task is to define their location within the TRM. As the x axis represents the time dimension, the horizontal placement is straightforward: each sub-technology has to be located on the corresponding year. However, vertical placement requires the analysis of the characteristics of the terms included in each sub-technology, in order to place them on the corresponding main field of each layer. Finally, when one or more sub-technologies represent the same concept, these are merged in a unique sub-technology. As a result of that, the sub-technologies represented on the roadmap have different durations, which is a logical representation of reality (Figs. 6, 7, 8 and 9). The decisions taken in the present step suppose the second item to be assessed by experts in the step eight. Regarding the market layer, all the information within it comes from the *web content mining* task, which is described in detail in step 7.

3.2.3. Step 6: identification of links

The inputs of this step are the sub-technologies of the previous step. The task to be done is to identify existing links among the sub-technologies belonging to the technology layer with those belonging to the application layer. This makes it possible to identify which basic research activities (sub-technologies of the technology layer) have given rise to important applications (sub-technologies of the application layer). Moreover, those sub-technologies which do not present links with applications should be considered either as research fields without the capacity to evolve into practical applications, or research fields with potential applications yet to be *discovered*. The links are represented by colored lines in the TRM.

The task is initiated by generating the keyword-vector for each sub-technology. Thus, the similarity of two sub-technologies corresponds to the correlation between these vectors. This is quantified as the cosine of the angle between vectors, that is, the so-called *cosine similarity*. This similarity measure is a common procedure in numerous information retrieval applications (Huang, 2008). It is calculated as follows: $\textit{cosine similarity}\ (\overrightarrow{t_a}, \overrightarrow{t_b}) = \frac{\overrightarrow{t_a}.\overrightarrow{t_b}}{|\overrightarrow{t_a}|\, x\, |\overrightarrow{t_b}|}$, where $\overrightarrow{t_a}$ and $\overrightarrow{t_b}$ are vectors of the m dimension belonging to pairs of sub-technologies. The result is bounded to the interval [0,1], where 0 indicates no relation and 1 total similarity. It should be noted that few references have been encountered related to this task, thus as a first approach a threshold of 0.4 for cosine similarity has been set in order to accept the existence of a strong link. The script used for this task was generated in Python.

3.2.4. Step 7: trends analysis

The inputs of this step are from one side, the list of the most-increased-keywords generated in the third step, and from the other side, the list of terms obtained from the application of the *web content mining* method. It should be pointed out that even though the information retrieved from web pages can be located in all the layers of the TRM, this method provides clear advantages identifying technology applications and market solutions.

The task to be done is to identify potential future paths of the technology for the short and medium term future, i.e. the technology forecasting. In this case, instead of obtaining terms grouped in a cluster, as occurring in the fourth step, or in a factor, as occurring in the fifth, these are picked from simple lists and have to be located coherently along the roadmap. Once again, this kind of decision shall be assessed by experts. Trend identification is separated into two phases:

- *Short term future*. This time range is formed by n and $(n+1)$ years, where n represents the year following the year in which the research is being conducted. In order to discover possible short-term future paths, those topics which have raised more interest among researchers during recent years should be identified. To do this, the most-increased-keywords lists, which were generated in the third step (profile generation) have to be studied. These lists contain those keywords which, in addition to being among the most frequent, have also suffered sharp increments in such frequencies compared with the previous year. In this sense, it can be said that the most increased keywords allow identifying potential paths of the technology. More specifically, the list of the $(n-2)$ year is used to retrieve the terms to be placed in the n year of the roadmap, whereas the list of the $(n-1)$ year is used to retrieve the terms to be placed in the $(n+1)$ year. In the case of the present approach the number of keywords retrieved for each year was 20.
- *Medium term future*. This time range is formed by $(n+1)$, $(n+2)$ and $(n+3)$ years. So year $(n+1)$ is considered as the nexus of both time ranges and is completed with information stemming from both phases. In this phase the foresight has a greater degree of uncertainty and is based on a *web content mining* process. The source of information is constituted by web pages of two kinds of entities with know-how in the studied technology: market research companies and technology providers. In this regard, what is being obtained is first-hand information, in terms of current existing sub-technologies and especially applications that are being marketed. This task presents several sub-tasks:
  - Several web pages should be selected and these pages have to be representative of market research companies and provider companies; this selection has to be assessed by experts. In the case of this paper, three market research companies: Forrester, International Data Corporation (IDC), and Gartner; in addition to three Cloud providers: Amazon Web Services (AWC), LunaCloud and CloudSigma were selected.
  - Web content mining initial task has to be done by a *web crawler*, with which manageable text files that contain existing information within the web pages can be generated. This task requires the management and treatment of different URLs for each website, which in turn requires careful analysis of the individual pages themselves.
  - The content of these text files should be analyzed using either the web crawler, assuming it possesses the required functions, or exporting the text files to a text mining tool.
  - *Natural Language Processing* (NLP) technique has to be applied to the content of the text files in order to identify nouns representing concepts. Along with this, lists containing terms with high frequency of appearance need to be generated. These terms represent those topics that are raising interest among practitioners, as well as researchers not uniquely involved in scientific publications.
  - The identified high frequency terms have to be placed on the roadmap. More specifically, it has to be decided whether a term

goes into an already existing sub-technology or if it is creating, along with other terms, a new sub-technology. In the case of the present approach, 10 terms were retrieved from each web page.

The first task belonging to the web content mining method is carried out using the *web crawler* included in the *IBM Watson* (Ferrucci et al., 2010) platform, specifically in the *Content Analytics* product (Zhu et al., 2014), which also provides a complete analytics package. However, this was only used for crawling HTML content and generating the TXT files; once the text files were obtained these were exported to VantagePoint software in order to analyze them via NLP and generate the above mentioned lists.

3.2.5. Step 8: expert assessment

The inputs of this step are all those decisions taken by the researcher which require a certain degree of expertise in the technology being depicted. Although it has been mentioning what are the points to be discussed with the experts, these can be summarized as follows:

- Both ontologies generated in the fourth step. Specifically the first level of the ontologies, the denomination of all the clusters and the complete structure.
- The TRM. Specifically the vertical structure, the naming of the sub-technologies, the coherency of the term collections within the sub-technologies and the logic of the identified links.
- Selected web pages for web content mining. Their representativeness and technological know-how.

In order to apply expert assessment for the selected methodology interviews with three experts were arranged. The profile of the experts was low-level hierarchy but highly specialized in the field. The order of the day was previously proposed and the meetings were limited to 60 min, being mainly focused on an explanation of the work and details of items to be assessed. Afterwards, a period of ten days was given to receive a complete assessment and all of the comments were cross-checked and included in the outcomes. This method does not have a quantitative approach for measuring the impact of the assessments, but it can be said that the final outcomes possess a combined nature, quantitative and qualitative.

4. Approach results: depicting cloud computing

4.1. Introduction to cloud computing

One of the most used definitions for CC is from the NIST (Mell and Grance, 2010) which states: "CC is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction". The most important features can be summarized as follows: on-demand self-service; broad network access; resource pooling; rapid elasticity; measured service. In addition to these five essential characteristics, the cloud community has extensively used the following three service models to categorize the cloud services (Miller and Veiga, 2009): *infrastructure as a Service* (*IaaS*): a flexible infrastructure of distributed data center services connected via Internet style networks; *platform as a Service* (*PaaS*): middleware providing application services and/or runtime environment for cloud applications; *software as a Service* (*SaaS*): top-layer applications delivered on demand.

Cloud technology is having a huge impact on current enterprises. In this sense, studies such as that conducted by the Cisco Systems (2016), revealed some impressive figures concerning Cloud-related data traffic in the near future. In addition to this, CC has the potential to play a major role in addressing inefficiencies and make a fundamental contribution to the growth and competitiveness of organizations mainly for SMEs (Sahandi et al., 2013). However, Cloud industry to an extent has struggled to grow among SMEs due to the reluctance and concerns expressed by them (Khan and Al-Yasiri, 2016). Thus, any attempt to increase the awareness about this technology will result in benefit for many companies facilitating its implementation. The present work supports this view, generating a helpful profile and a TRM for CC technology.

4.2. First phase: cloud computing technology profile

As previously mentioned, partial results of the profile of CC technology were provided in Bildosola et al. (2017). However, in the case of that work the focus was centered on the description of the *literature profile and research community profile*. Thus, what it was provided was the identification of the top journals and conferences of the field; top countries, organizations and authors; as well as author-keyword maps. In addition to this, some insights about what has been and will be researched was described, related to the *state of the research and its evolution*. For the purpose of delving deeper into this concept, additional descriptive elements are provided here.

When it comes to describing the evolution of research activity, it is worth describing the evolution of the number of publications related to the field, as well as the number of institutions involved in it. In this sense, Fig. 3 shows this information, where the publications are disaggregated in journals and conferences, as well as entities in government, academic and private companies. The graphs provide meaningful information about the evolution of the research activity of the field. As soon as the first publications related to CC (2008) appeared, the field entered a phase of sharp increment in terms of number of publications and number of entities researching it (2719% increase for publications and

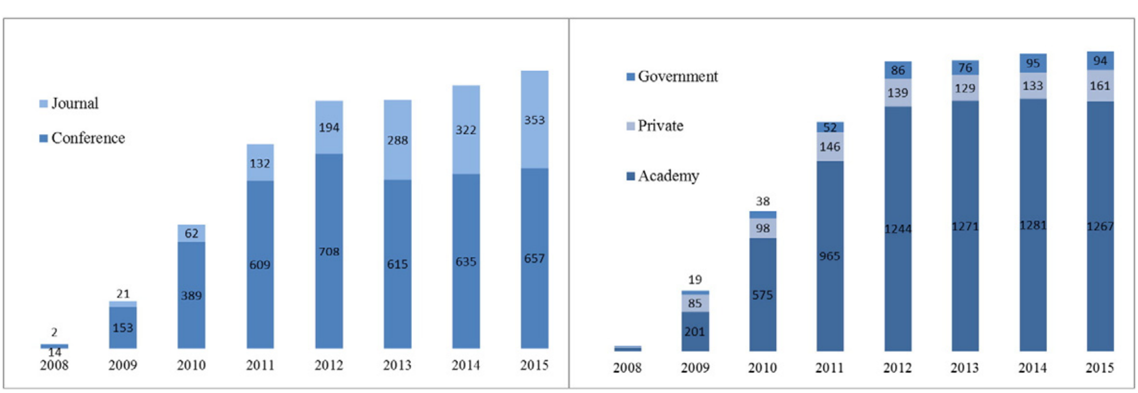

Fig. 3. Evolution of research, number of publications and institutions involved year by year.

Table 1
Most frequent keywords 2008–2015 (the overall frequency in brackets).

2008	2009	2010	2011
Distributed computing (5.6)	Grid computing (2.8)	CC security (2.4)	CC security (2.9)
Virtual computing (5.6)	Service oriented architecture (2.8)	Virtualization 2.4	Virtualization (1.9)
CC security (5.6)	Virtualization (2.6)	Privacy (1.9)	Software as a service (1.5)
	Software as a service (2.3)	Distributed computing (1.8)	Infrastructure as a service (1)
	Distributed computing (2.3)	Software as a service (1.5)	Service level agreement (1)
	CC security (1.8)	Service oriented architecture (1.3)	Privacy (1)
	Data center (1.5)	Infrastructure as a service (1.2)	Distributed computing (0.9)
	Web services management (1.5)	Platform as a service (1)	Platform as a service (0.9)
	Privacy (1.3)	Data center (1)	Mobile CC (0.8)
	Utility computing (1.3)	Grid computing (1)	Access control (0.7)
2012	2013	2014	2015
CC security (2.6)	CC security (8.8)	Mobile CC (2.3)	Big Data (0.9)
Virtualization (1.7)	Virtual machine (4.9)	CC security (1.2)	Mobile CC (0.7)
Virtual machine (1.3)	Mobile CC (4.7)	Virtualization (1)	Digital storage (0.7)
Privacy (1.3)	Privacy (4.5)	Load balancing (1)	Quality of service (0.5)
Service level agreement (1.2)	Distributed computing (4.3)	Quality of service (0.8)	Virtual machine (0.5)
Infrastructure as a service (1.1)	Virtualization (4.2)	Resource allocating (0.8)	Web services (0.5)
Software as a service (1)	Quality of service (3.3)	Access control (0.7)	Energy efficient (0.5)
Energy saving (0.9)	Energy saving (3.3)	Software as a service (0.7)	Task scheduling 0.5
Distributed computing (0.9)	Software as a service (3.2)	Task scheduling (0.7)	Cryptography (0.4)
Quality of service (0.9)	Infrastructure as a service (2.9)	Fine grained access control (0.6)	Customer relationship manager (0.4)

2991% for entities in the period 2008–2010). Over the next two years the increase continued with less intensity (100% increase for publications and 106% for entities in the period 2010–2012). However, after four years of this behavior a plateau was reached in 2012, entering in a phase of maintained activity (12% increase for publications and 13.6% for entities in the period 2012–2015). Even though these graphs are not enough to perform an analysis related to Technology Life Cycle indicators in the sense of Watts and Porter (1997), it may help to understand its position in the technology life-cycle, indicating whether a technology is in an embryonic, growing or declining phase (Yoon and Phaal, 2013). For the case of CC in 2015 the declining phase had not yet been reached. In addition to this, more insights can be obtained from these graphs, i.e. the shift in the nature of the publications from conference contributions to journal contributions. When it comes to emerging technologies, initial contributions usually come from conferences, where more embryonic works are presented. However, once the research on the technology matures, more contributions come from journals, where published research tends to have a longer work-period lag. This behavior is easy to see in Fig. 3, where data show 12.1% journal contributions in 2009 compared to 34.9% in 2015.

The characterization of the technological profile is completed by top-keywords lists. Thus, the list of the most frequent keywords (Table 1) allows those hot topics which have led the field year by year to be identified. Thus, according to the results, *security* topic and technology provision models (*Software as a Service, Platform as a Service* and *Infrastructure as a Service*) have appeared in the top positions since the first years. It is clear that security is one of the most prolific research activities, as it is a major concern for companies when it comes to adopting CC (Takabi et al., 2010). Moreover, although initial works included just the term *security* within the keywords, most recent publications have opted to use more specific terms such as *fine grained access control*. This evolution is applicable to the whole set of keywords, in the sense that more generic terms topped the lists initially (*distributed computing*, *grid computing data center*, *privacy*, etc.), which were clearly referring to foundational research activities, whereas more specific fields have appeared little by little (*energy saving*, *mobile cloud computing*, *customer relationship management* or *load balancing*). This term-shifting phenomenon describes the thematic specialization of research, another significant indicator of technology research maturity.

Continuing with the analysis of keywords, it is also worthwhile identifying the most-increased-keywords year by year. This information can be seen in Table 2, where keywords are sorted by frequency increment compared with the previous year. In order to avoid including residual keywords with sharp increments, only those keywords with more than 4 appearances were considered. The fact that a certain connection between the tables exists is worth noticing. Some of the most-increased-keywords have been able to catch the researchers' attention and have appeared among the most frequent ones. This is the case for *mobile CC*, a new paradigm in CC which provides cloud resources for mobile devices. It appeared among the most increased in 2011 and

Table 2
Most-increased-keywords for the period 2010–2015 (the increase percentage with respect to the previous year in brackets).

2010	2011	2012
Service level agreement (700)	e-Governance (900)	Reliability (1200)
Risk management (700)	Eucalyptus (900)	Adoption of CC (600)
Trust computing (700)	Mobile CC (850)	Game theory (600)
Infrastructure as a service (650)	Secure CC (700)	Digital forensic analysis (500)
Network management (600)	Dependability analysis (700)	Middleware (500)
Hadoop (600)	Green cloud (500)	Replication (500)
Private CC (600)	Accounting (500)	Sensor network (400)
Geographic information system (500)	Data placement (500)	Bayesian (400)
Load balance (500)	OpenNebula (500)	Noise generation (400)
Ubiquitous computing (500)	Real-time CC (500)	Fault-tolerant (300)
2013	2014	2015
Attributed based access control (500)	Analytic hierarchy process (800)	Information management (1400)
Log analysis (500)	Performance analysis (800)	Cost analysis (1200)
Outsourced data (500)	Computation offloading (700)	Online social networks (1100)
TCCP (500)	Elastic cloud (600)	Complex networks (1100)
Customer relationship management (400)	Mobile device (600)	Mobile telecommunications (900)
Fine-grained access control (400)	Energy aware (600)	Java programming language (900)
Remote access (400)	Heuristics (600)	Cloud Application (850)
Hypervisor (400)	Homomorphic encryption (600)	Monitoring (800)
Stochastic Petri net (400)	Identity management system (600)	Artificial intelligence (800)
Data analysis (300)	Knowledge management systems (600)	Authentication (800)

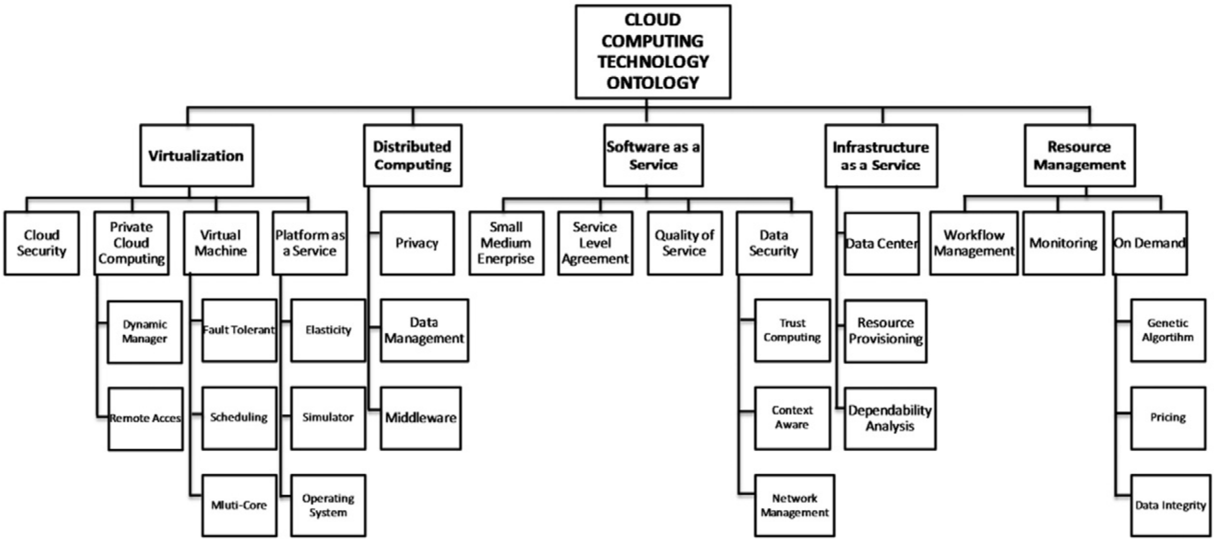

Fig 4. Cloud computing technology ontology.

among the most frequent in 2013 and 2014. Together with this, terms related to access management appeared among the most increased in 2013, with examples such as *Attributed based access control* and *fine-grained access control*, and this had a direct consequence as *fine-grained access control* and *access control* appeared among the most frequent in 2014. In the same line, *energy aware* appeared within the most increased in 2014, and *energy efficient* within the most frequent in 2015. Of course not all the most-increased-keywords manage to appear in the most frequent top ten after a few years, however, this kind of information is valuable when it comes to foresighting potential future paths of a technology, and this is the case for the present approach. Thus, the *trends identification* step makes use of the terms included in Table 2 and combines them with other forecasting information sources.

4.3. Second phase: technology roadmap

4.3.1. Ontology

The vertical structure of the TRM (see Figs. 6, 7, 8 and 9) is directly obtained from the ontologies' first line. Ontology generation has been explained in the description for step 4 and it is derived from an iterative clustering process. The whole process is based on the analysis of keywords contained in the publications and therefore the obtained ontology does not purport to be the best possible hierarchical descriptive element, but rather one that is strictly derived from the aforementioned analysis and which possess a valid descriptive capacity. Fig. 4 shows the generated ontology, which specifically results from applying the above process to the publications related to basic research (first database from step 2). Considering the first line of the ontology, it can be observed that it is made up of two of the delivery models (*software as a service* and *infrastructure as a service*), basic technology research concepts (*virtualization* and *distributed computing*), and one of the main challenges of CC technology, *resource management*, especially in terms of energy saving (Beloglazov et al., 2012). This is the result of the clustering analysis and naming process, and the opinions of the consulted experts. It provides a significant picture of the technology and highlights the structure of research related to CC.

A second ontology was generated (Fig. 5). In this case applying the process described in step 4 to the database containing scientific

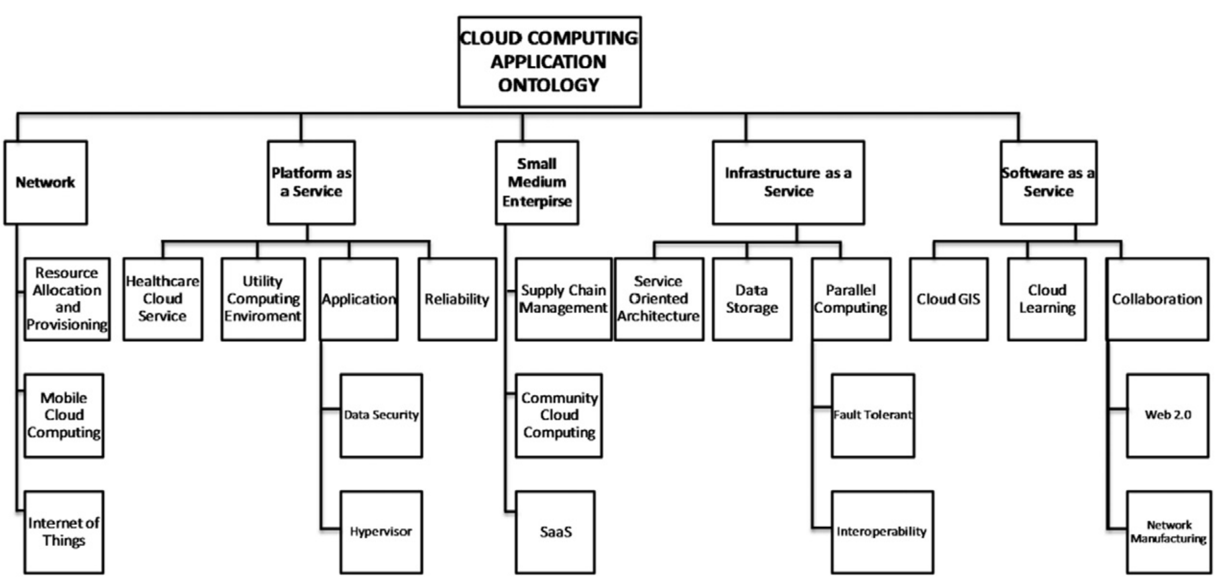

Fig 5. Cloud computing application ontology.

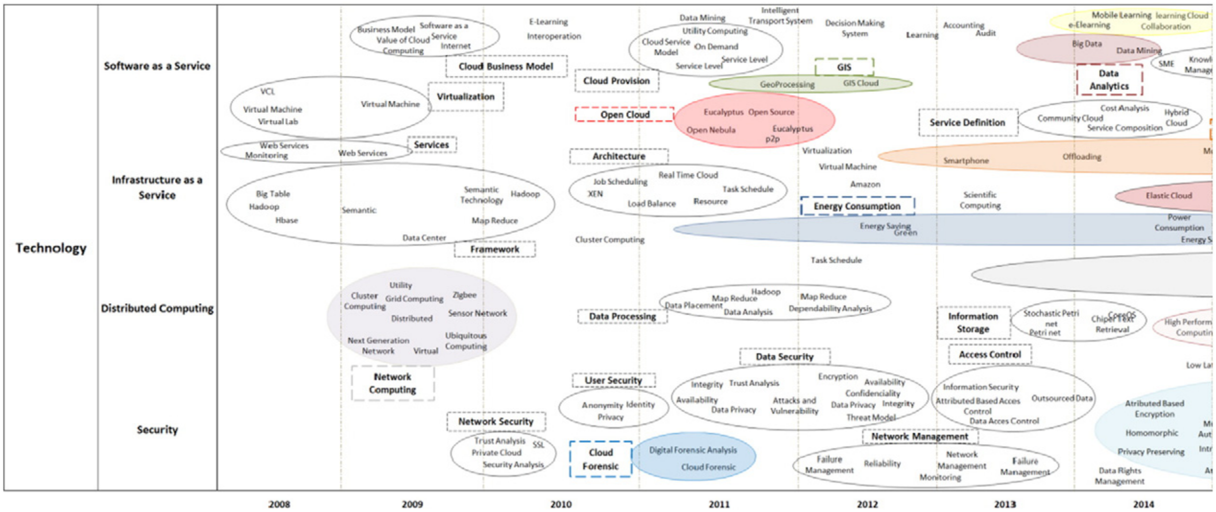

Fig 6. Zoom of the TRM which represents the period 2008–2014 for the technology layer.

documents related with applications of the technology. The first line is structured with three delivery models (*platform as a Service*, *infrastructure as a Service* and *software as a Service*); a concept that describes the nature of Cloud technology (*network*); and the type of company that can get more out of the technology (*Small and Medium Enterprise*). In this case the experts reported that the represented structure provides an extensive image of technology applications. Thus, both ontologies help to identify the main sub-technologies within CC technology and generate a complete scheme which allows the hierarchical relationships among them to be identified. This information will be useful when generating the roadmap, as it directly defines the vertical structure of the TRM and helps name the sub-technologies that appear throughout it.

4.3.2. Technology roadmap

Based on the description of the approach, the time range of the roadmap is 2008–2020 for the case of CC. In order to simplify the analysis, the TRM has been divided in two similar periods: 2008–2014 and 2015–2020; and each layer is described separately: technology layer, application layer and market layer. All the elements located within the TRM stem directly from step 5 (identification of sub-technologies), step 6 (identification of links) and step 7 (trend analysis). These all elements have been reviewed with experts. Fig. 6 shows a zoom of the TRM which pertains to the initial period (2008–2014) and to the *technology* field. The first analysis should include observation of the sub-technologies (identified as ellipses) and the terms contained within them. Even if this initial period presents low density in terms of sub-technologies, several foundational research activities start to appear, such as *virtualization* or *network computing*. These initial research fields gradually give way to more specific topics such as *open cloud*, *data processing* or *data security*. A bigger picture of the roadmap (Fig. A1) shows how density increases as it moves forward in time.

In the case of the application layer (Fig. 7), the behavior is pretty similar but the specialization of topics is harder to identify. This is because in this case all the ellipses are representing applications of the technology. However, it is worth noticing how more generic applications such as *high performance computing* or enterprise related concepts such as *provider selection* have evolved to some very specific activities represented by *image processing* or *transport management*.

Comparing both layers a marked difference in the size of the sub-technologies can be observed. Those belonging to the technology layer present, on average, a longer life with respect to the application's sub-technologies. This matter is mainly explained by the fact that basic

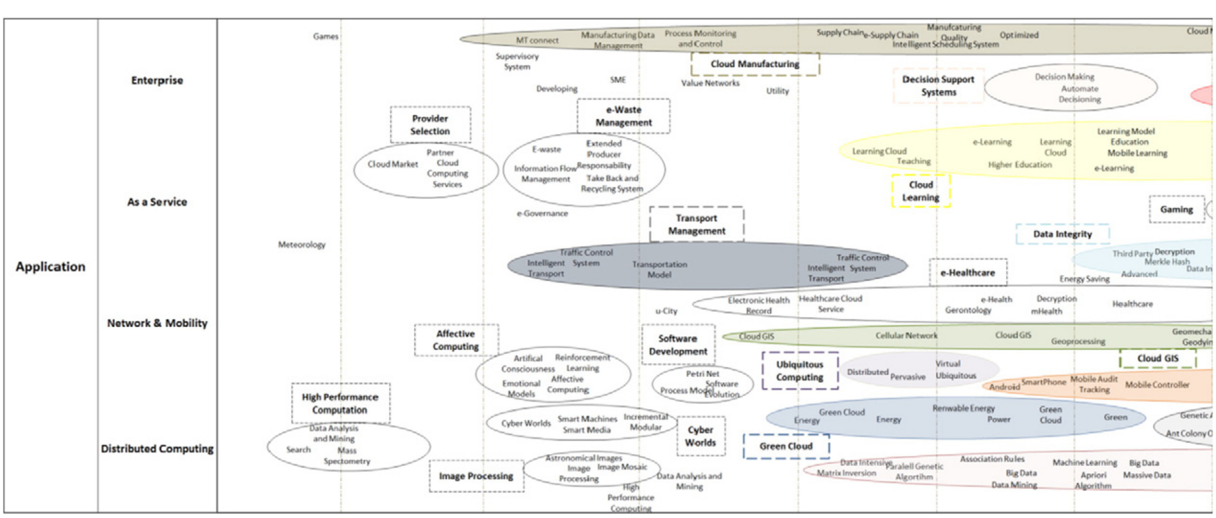

Fig 7. Zoom of the TRM which represents the period 2008–2014 for the application layer.

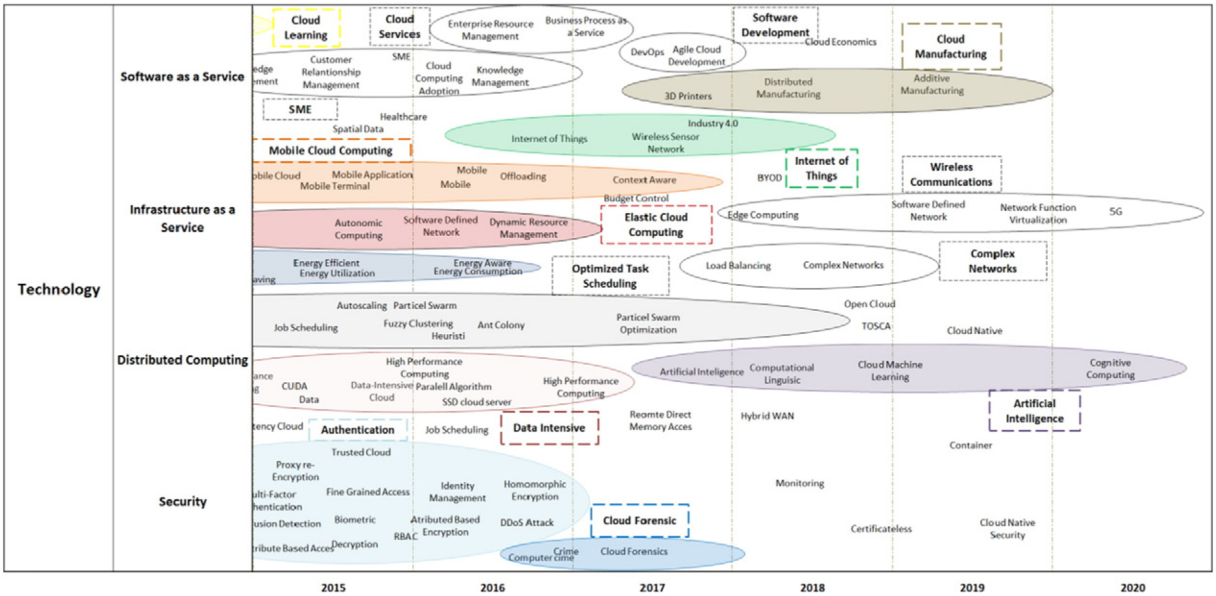

Fig 8. Zoom of the TRM which represents the period 2015–2020 for the technology layer.

research activities require more prolonged work, whereas researches focused on applications usually make use of the previous work and targets more specific fields, resulting in more diverse and shorter works.

Entering in the second period (2015–2020) for the technology layer (Fig. 8), it can be seen how the density has increased sharply and consequently numerous sub-technologies reflecting an equal number of research activities can be found. So even if, as previously discussed, the research activity reached its peak in 2012 in terms of publications and entities related to it, it was not until approximately 2015 when it reached its maximum in terms of diversity (number of different live sub-technologies). Going into detail, this period includes both the present (2016–2017) and the future (2018–2020), and in that sense two kinds of sub-technologies can be identified: those that simply maintain their activity and thus expand their presence throughout these periods, as is the case with *optimized task scheduling* or *internet of things*; and those which represent new activities, such as *complex networks* or *wireless communications*. In addition to this, single-term sub-technologies such as *container* and *cloud native security* should be included within this last group. These single-term examples represent topics that could be of future importance but which are not diverse enough to generate a multi-term sub-technology on their own.

When it comes to the application layer (Fig. 9), similar conclusions are derived from the analysis. Once again the density is higher in this second phase, even though the increase is less accentuated as it comes from a higher density compared with the technology layer. Some of the sub-technologies already appeared earlier than 2015 and are maintaining the activity for the 2017–2018 period. However, there are certain others which are expected to reduce in activity and therefore fade off the TRM, as is the case with *e-Healthcare*, a huge CC application which lasted for six years. When it comes to identifying emerging

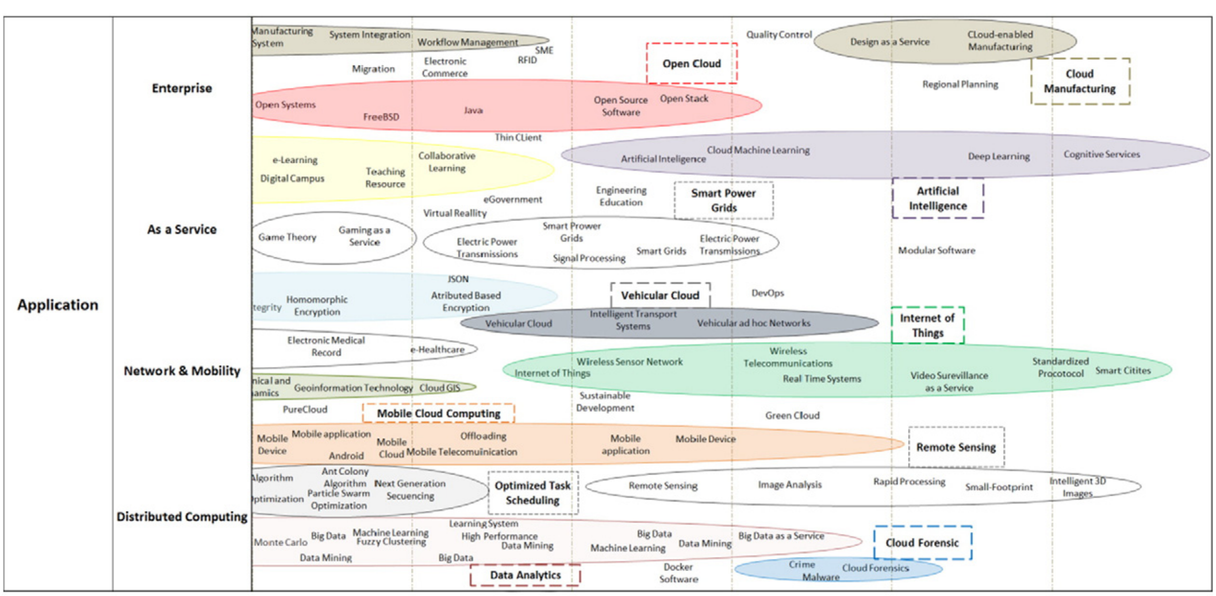

Fig 9. Zoom of the TRM which represents the period 2015–2020 for the application layer.

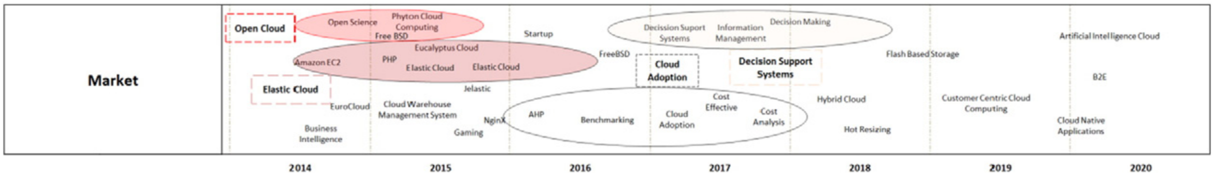

Fig 10. Zoom of the TRM which represents the period 2014–2020 for the market layer.

sub-technologies, clear examples are *artificial intelligence*, which is supposed to bring the technology to a new era with the proliferation of cognitive services; and *internet of things*, based on the interconnection capacity of diverse type of devices, heavily reliant on CC.

The last layer (Fig. 10) contains the most important sub-technologies which have reached the market. This information stems directly from the web content mining task, and it is of great interest for those decision makers who need to know what the technology will bring in the short or medium term. The capacity to search and retrieve the content of CC related web pages enriches the roadmap significantly, since it allows generating the market layer and completing the rest of the layers, with completely current information. Analyzing this information, it is worth remarking the significant expansion of benchmarking solutions for CC providers and services, which will ease the process of migration or adoption, provided that these solutions are known. In respect of specific applications, the cloud *business to employee* (*B2E*) will change remarkably the classical employee portal, as it will encompass everything that businesses do to attract and retain well-qualified staff. This is a change in which all the companies must be involved unavoidably. Together with this, this layer provides more examples about emerging solutions, which must be analyzed by those who are interested on taking advantage of the opportunities of CC technology, such as *hot resizing* solutions and *customer centric cloud computing*, among others.

Finally, it is worth studying the links that have been included along the TRM. These links are identified with different colors and relate sub-technologies located within different layers. There are clear cases such as *mobile cloud computing*, named identically in technology and application layers and colored in orange. However, this is not the case for all of them, as can be seen by linked sub-technologies identified as *energy consumption* (technology) and *green cloud* (application), colored in blue. Several more links were identified and consequently included on the TRM. The identification of these links enhances the provided information, generating connections among the layers and integrating all the evolution of the technology. All the mentioned elements can be observed together in the Appendix A where the complete TRM is provided (Fig. A1). In this way the evolution of the density of sub-technologies and the specialization of the topics are visible. Moreover, the coloration resulting from the identification of links provides clearer identification of existing synergies between the layers.

5. Conclusions and future work

The work has presented an approach (TKRM) in which bibliometrics and technology forecasting methods have been merged coherently. Moreover, the approach has been applied to a cutting edge emerging technology, as is CC. Thus, CC research activity has been depicted by means of bibliometrics and in addition to this, structuring of information contained in publications has been obtained, providing profitable access to a solid database. From there, several methods have taken advantage of previous work and have generated a detailed TRM in addition to other meaningful elements such as double ontology. The results of the work confirm how the combination of selected methods enhances the understanding of the analyzed technology, providing a complete depiction of it, with special emphasis on its evolution forecasting. The inclusion of emerging methods, such as web content mining, and its combination with consolidated methods, has made it possible to identify sub-technologies that will maintain their activity in the short-term future, others which will not, and those which will come into being within that period.

When it comes to analyzing the usefulness of the obtained depiction of the technology as such, there are important conclusions to be drawn. Firstly, the profile of the scientific publications has enabled to detect the research maturity level. Secondly, the ontologies obtained via clustering process have provided a complete picture of the structure of the technology, in terms of existing sub-technologies and the relationships among them. Finally, the roadmap goes one step further and introduces the time dimension to the depiction, divides the technology into three main fields and combines two kinds of information sources: scientific publications and web content. Going into detail, the obtained picture allows identifying the foundational sub-technologies, which are placed throughout the initial years. Examples of this type are *virtualization* and *network computing* for the case of CC. In the following years it can be seen how these initial sub-technologies have changed toward other more specific ones, as the technology was reaching a higher maturity level. Following the timeline, the roadmap provides information about current sub-technologies which are leading the field of research and those solutions available on the market, in addition to the identification of emerging sub-technologies which shall be the spearhead of the technology in the years to come. Examples of this type are *artificial intelligence* and *remote sensing* in the application layer and *B2E* in the market layer for CC technology.

Further conclusions can be drawn from the roadmap analysis. However, it should be noted that the aim of the approach when it is applied to an emerging technology, is to obtain a complete depiction of it, which shall act as a starting point for interested researchers, practitioners or decision makers, who may take advantage of the information provided for the purpose of improving their understanding of the technology and its evolution, with the objective of applying it accordingly to their goals. For the case of CC, the full TRM is included in the Appendix A (Fig. A1) in order to provide a complete picture for such purposes.

Finally, it should be underlined that the approach presents some limitations that shall be tackled in the near future. The first of these is the management of link identification. In this particular case a threshold was established in order to determine whether two sub-technologies are strongly connected or not, thus applying a binary perspective (whether to draw the link or not). However, deeper analysis can be carried out and broader background literature needs to be studied within this field. Even though almost no references have been encountered, the possibility to introduce some kind of granularity among the links is a clear option. By way of example, the identified links could be classified as weak, medium and strong connections, based on different thresholds, or even a more detailed classification could be used. Another aspect to be revised and improved is the web content mining process. The *IBM Watson* platform has been merely used as a web crawler, and has not taken advantage of its analytical potential to apply it to the retrieved web content. Thus, future applications of the approach should deepen into the functionalities offered by the platform and make the most of it. Taking all this into account, a new application of the approach to a different emerging technology is considered as the best option in order to include these, and probably further improvements made to it.

Appendix A

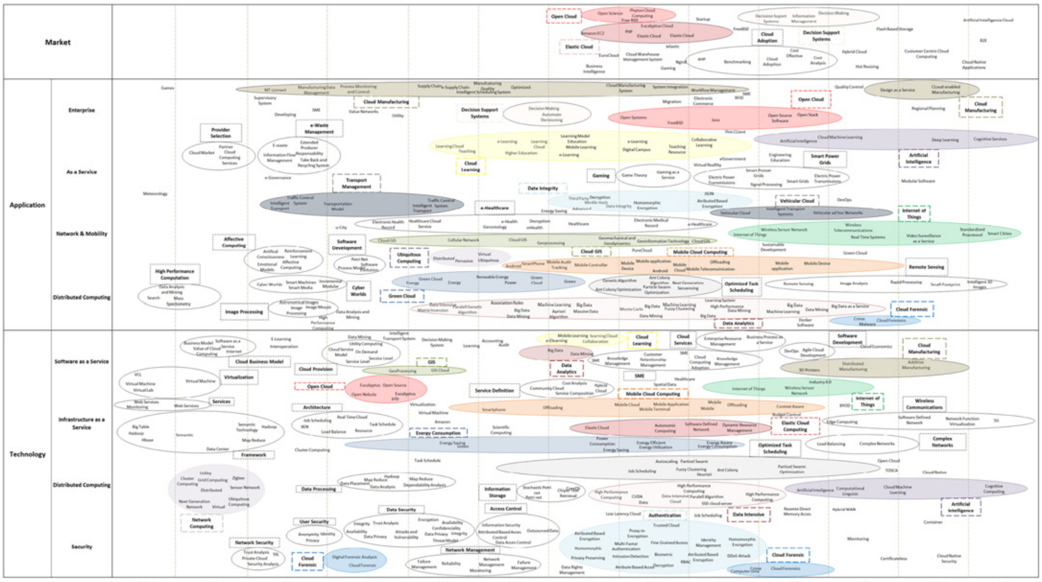

Fig. A1. Technology roadmap for cloud computing technology, 2008–2020 period.

References

Beloglazov, A., Abawajy, J., Buyya, R., 2012. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Futur. Gener. Comput. Syst. 28 (5), 755–768.

Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technol. Forecast. Soc. Chang. 73 (7), 835–844.

Bhattacharya, S., Basu, P., 1998. Mapping a research area at the micro level using co-word analysis. Scientometrics 43 (3), 359–372.

Bildosola, I., Rio-Bélver, R., Cilleruelo, E., Gavilanes, J., 2017. Forecasting cloud computing: producing a technological profile. Engineering Systems and Networks. Springer International Publishing, pp. 23–31.

Bildosola, I., Rio-Bélver, R., Cilleruelo, E., 2015. Forecasting the big services era: novel approach combining statistical methods, expertise and technology roadmapping. Enhancing Synergies in a Collaborative Environment. Springer International Publishing, pp. 371–379.

Börner, K., Chen, C., Boyack, K.W., 2003. Visualizing knowledge domains. Ann. Rev. Inform. Sci. Technol. 37 (1), 179–255.

Bray, O.H., Garcia, M.L., 1998. Technology Roadmapping: Approaches. Tools & Lessons Learned, Sandia National Laboratories, Albuquerque, NM.

Cisco Systems, 2016. White paper: Cisco Global Cloud Index: forecast and methodology, 2014–2019.(Available):. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html (accessed on 30 September 2016).

Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. 73 (8), 981–1012.

Eck, N.J.V., Waltman, L., 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. J. Am. Soc. Inf. Sci. Technol. 60 (8), 1635–1651.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A.A., ... Schlaefer, N., 2010. Building Watson: an overview of the DeepQA project. AI Mag. 31 (3), 59–79.

Firat, A.K., Woon, W.L., Madnick, S., 2008. Technological forecasting–a review. Composite Information Systems Laboratory (CISL). Massachusetts Institute of Technology.

Garechana, G., Rio Belver, R., Cilleruelo, E., Larruscain Sarasola, J., 2015. Clusterization and mapping of waste recycling science. Evolution of research from 2002 to 2012. J. Assoc. Inform. Sci. Technol. 66 (7), 1431–1446.

Gök, A., Waterworth, A., Shapira, P., 2015. Use of web mining in studying innovation. Scientometrics 102 (1), 653–671.

Hout, M.C., Papesh, M.H., Goldinger, S.D., 2013. Multidimensional scaling. Wiley Interdiscip. Rev. Cogn. Sci. 4 (1), 93–103.

Huang, A., 2008. Similarity measures for text document clustering. Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand, pp. 49–56 (April).

Huang, L., Zhang, Y., Guo, Y., Zhu, D., Porter, A.L., 2014. Four dimensional science and technology planning: a new approach based on bibliometrics and technology roadmapping. Technol. Forecast. Soc. Chang. 81, 39–48.

Janssens, F., Leta, J., Glänzel, W., De Moor, B., 2006. Towards mapping library and information science. Inf. Process. Manag. 42 (6), 1614–1642.

Kajikawa, Y., Yoshikawa, J., Takeda, Y., Matsushima, K., 2008. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy. Technol. Forecast. Soc. Chang. 75 (6), 771–782.

Kaufman, L., Rousseeuw, P.J., 2009. Finding Groups in Data: An Introduction to Cluster Analysis vol. 344. John Wiley & Sons.

Khan, N., Al-Yasiri, A., 2016. Framework for cloud computing adoption: a road map for Smes to cloud migration. arXiv Preprint arXiv:1601.01608.

Kongthon, A., 2004. A Text Mining Framework for Discovering Technological Intelligence to Support Science and Technology Management. (Doctoral dissertation). Georgia Institute of Technology.

Kostoff, R.N., Geisler, E., 1999. Strategic management and implementation of textual data mining in government organizations. Tech. Anal. Strat. Manag. 11 (4), 493–525.

Kostoff, R.N., Boylan, R., Simons, G.R., 2004. Disruptive technology roadmaps. Technol. Forecast. Soc. Chang. 71 (1), 141–159.

Leydesdorff, L., 1987. Various methods for the mapping of science. Scientometrics 11 (5-6), 295–324.

Major, E.J., Cordey-Hayes, M., 2000. Engaging the business support network to give SMEs the benefit of foresight. Technovation 20 (11), 589–602.

Martin, B.R., Johnston, R., 1999. Technology forecasting for wiring up the national innovation system: experiences in Britain, Australia, and New Zealand. Technol. Forecast. Soc. Chang. 60 (1), 37–54.

Martin, H., Daim, T.U., 2012. Technology roadmap development process (TRDP) for the service sector: a conceptual framework. Technol. Soc. 34 (1), 94–105.

Martino, J.P., 1993. Technological Forecasting for Decision Making. McGraw-Hill, Inc.

Mell, P., Grance, T., 2010. The NIST definition of cloud computing. Commun. ACM 53 (6), 50.

Mikut, R., Reischl, M., 2011. Data mining tools. Wiley Interdiscip. Rev. Data Min. Knowl. Disc. 1 (5), 431–443.

Miller, H.G., Veiga, J., 2009. Cloud computing: will commodity services benefit users long term? IT Prof. Mag. 11 (6), 57.

Popper, R., 2008. 3. Foresight methodology. The Handbook of Technology Forecasting: Concepts and Practice, p. 44.

Porter, A.L., Detampel, M.J., 1995. Technology opportunities analysis. Technol. Forecast. Soc. Chang. 49 (3), 237–255.

Porter, A.L., Ashton, W.B., Clar, G., Coates, J.F., Cuhls, K., Cunningham, S.W., et al., 2004. Technology futures analysis: toward integration of the field & new methods. Technol. Forecast. Soc. Chang. 71 (3), 287–303.

Porter, A.L., Watts, R.J., Anderson, T.R., 2003. Mining PICMET: 1997–2003 papers help you track management of technology developments. Proceedings, Portland International Conference on Management of Engineering and Technology (PICMET).

Porter, A.L., Youtie, J., Shapira, P., Schoeneck, D.J., 2008. Refining search terms for nanotechnology. J. Nanopart. Res. 10 (5), 715–728.

Sahandi, R., Alkhalil, A., Opara-Martins, J., 2013. Cloud computing from SMEs perspective: a survey based investigation. J. Inf. Technol. Manag. 24 (1), 1–12.

Salerno, M., Landoni, P., Verganti, R., 2008. Funded research projects as a tool for the analysis of emerging fields: the case of nanotechnology. SPRU 40th Anniversary Conference–The Future of Science, Technology and Innovation Policy vol. 6.

Takabi, H., Joshi, J.B., Ahn, G.J., 2010. Securecloud: towards a comprehensive security framework for cloud computing environments. Computer Software and Applications Conference Workshops (COMPSACW), 2010 IEEE 34th Annual. IEEE, pp. 393–398 (July).

Thorleuchter, D., Van den Poel, D., 2013. Weak signal identification with semantic web mining. Expert Syst. Appl. 40 (12), 4978–4985.

Van Eck, N.J., Waltman, L., 2008. Appropriate similarity measures for author co-citation analysis. J. Am. Soc. Inf. Sci. Technol. 59 (10), 1653–1661.
Watts, R.J., Porter, A.L., 1997. Innovation forecasting. Technol. Forecast. Soc. Chang. 56 (1), 25–47.
Yoon, B., Phaal, R., 2013. Structuring technological information for technology roadmapping: data mining approach. Tech. Anal. Strat. Manag. 25 (9), 1119–1137.
Zhang, Y., Guo, Y., Wang, X., Zhu, D., Porter, A.L., 2013. A hybrid visualisation model for technology roadmapping: bibliometrics, qualitative methodology and empirical study. Tech. Anal. Strat. Manag. 25 (6), 707–724.
Zhu, Q., Cao, S., Ding, J., Han, Z., 2011. Research on the price forecast without complete data based on web mining. Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2011 Tenth International Symposium on. IEEE, pp. 120–123.
Zhu, W.D.J., Foyle, B., Gagné, D., Gupta, V., Magdalen, J., Mundi, A.S., ... Triska, M., 2014. IBM Watson Content Analytics: Discovering Actionable Insight from Your Content. IBM Redbooks.

Iñaki Bildosola Telecommunications Engineer and Master in “Project Management” (Higher Technical School of Engineering of Bilbao, University of the Basque Country (UPV/EHU)), Master in “Economics: Economic Analysis Tools” (Faculty of Economics and Business Studies at the University of the Basque Country (UPV/EHU)). PhD from the University of the Basque Country (Industrial Organization and Management Engineering). His current research focuses on Technological Forecasting and Information Structuring related to emerging technologies. He applies different methods ranging from Bibliometrics and Scientometrics, to Data Mining in terms of Text Mining and Web Mining. He is also involved in the generation of Technology Roadmaps.

Rosa María Río-Bélver has a PhD in Industrial Engineering, specializing in Organization. MSc in Environmental Sciences and Technology (UPV/EHU) and Msc in Innovation Management (AENOR). She is a University professor at the University of the Basque Country UPV/EHU; teaching subjects such as Operations Research, Industrial and Technological Politics and Competitiveness-Innovation. Her main research interests are focused on innovation management, competitive intelligence, technology maps, emerging technologies, tech mining, knowledge engineering, and foresight. She is the author of numerous papers and is involved in national and European projects.

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

Ernesto Cilleruelo has a PhD in Industrial Engineering, Msc in Innovation Management (AENOR) and he is a University Professor of the knowledge area of management. He has developed his academic and research career at the Faculty of Engineering at Bilbao, University of the Basque Country UPV/EHU. His research interest is innovation in management, a research line that has been working with for twenty years.