UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial,
Aeroespacial i Audiovisual de Terrassa

Titulació:

Master's Degree in Technology and Engineering Management

Alumne:
Lorenzo Capelli

Enunciat TFM:
Project for building an information system for generating systematic reviews of literature

Director del TFM:
Vicenç Fernandez Alarcon

Convocatòria de lliurament del TFM:

15/01/2020

REPORT

I declare that,

the work in this Master Thesis is completely my own work,

no part of this Master Thesis is taken from other people's work without giving them credit,

all references have been clearly cited,

I understand that an infringement of this declaration leaves me subject to the foreseen disciplinary actions by *The Universitat Politècnica de Catalunya - BarcelonaTECH*.

Lorenzo Capelli 15/01/2020

Title of the Thesis: Project for building an information system for generating systematic reviews of literature

TABLE OF CONTENTS

1. ACKNOWLEDGEMENTS

Foremost, I would like to express my deepest gratitude to my supervisor Prof. Vicenç Fernández of the "Universitat Politècnica de Catalunya" for the continuous support during the development of the project. He encouraged and steered me in the right direction every time the road got tough.

Ringrazio mia mamma e mio papà per tutti gli sforzi che hanno compiuto per permettermi di svolgere un master all'estero. Grazie per aver sempre supportato le mie scelte e per avermi accompagnato in tutto il cammino universitario. Grazie per le attenzioni e i consigli, grazie per aver permesso di realizzarmi e per essere diventati il mio modello da seguire. Ringrazio il mio piccolo ma grande fratello, per il supporto, i consigli, i regali e la presenza.

I would like to thank my old friends Andrea, Denis, Enrico, Lorenzo, Luigi, Martin, Mattia, Nicolas, Silvio, Stefano and Sushanth. Thanks for the support, the presence, the discussions and advices. You are my second family. Thanks to Anna, Elisa, Erika, Francesca and Sara.

I would like to thank Jacopo, Lorenzo, Marika, Roberto and Simone who left Italy and spent the first year of the master with me. Thanks for having shared indelible moments and amazing experiences. I owe a special thanks to my friend and roommate Roberto, who is still putting up with me. My sincere thanks to Alessandro, Carlo, Chiara, Giorgio, Laura, Sara, Valentina and Valerio from the University of Bergamo with whom I shared the majority of my university career.

Last but not least, I would like to thank all the people I met during this experience in Barcelona. From the other Italians, Emanuele and Riccardo, to the German guys, Steffen and Christoph. Thanks to Ana Clara, Andrea Liviere, Chiara and Marta. Thanks also to all the others who I do not mention directly. You have all enriched my life.

2. INTRODUCTION

The "Talent Analyzer" application is a tool that allows a user to perform bibliometric and content analysis on a dataset of academic papers, focusing on talents. The application does not require any specific knowledge of programming. The user has just to know some basic concepts about the two types of analysis that are carried on by the platform. The application is programmed in Shiny and it is mainly divided into four parts: the first part of data entry, a second part of data management, a third part of analysis and a final part of reporting. For the first and second parts, tables are used to store and edit data. For the third part, graphs and tables are used as outputs of the analysis. The last part generates a structured report in HTML or WORD that the user can export. The information system is linked with a MySQL database, where data are stored and extracted. The information system realized has been tested with a sample of 23 papers which is also available to the user when he installs it. The 23 papers used are mentioned in the reference chapter (subchapter: "*ACADEMIC PAPERS USED FOR THE SAMPLE*").

The main goal of this project is to realize a friendly-user and working information system which gathers information from a database and which generates and reports systematic reviews of literature in the field of talent management. The platform has to guarantee the possibility of being used by everyone, even those who do know R and R programming: the goal is to provide to that kind of people the opportunity to realize graphs and tables without the requirement of knowing the code of the realization of them. The platform has to work on different operative systems and it does not have to require difficult operations and a huge amount of time for the installation: the goal is to give to the user the minimum number of files and instructions to install the application. Another important goal that the informatic system has to fulfil is to provide to the user the possibility to download all the outputs of the analysis done, either singularly either placed in a structured report, created with a markdown script.

The report is structured as follows: after this introductory chapter, the figures index is provided. After that, a literature review about the academic databases, the bibliometric analysis and the systematic literature review for the content analysis is done. Then, a chapter describing the IT architecture of the informatic system is included. It describes in detail the software used for building the application and the reasons why they have been

chosen for this project. After the description of the IT architecture, the report deals with the set-up of the project. In particular, in this chapter are described the three files which are necessary for the installation of the information system, focusing also on the explanation of the structure of the database and the R packages required for the correct functioning of the platform. The study keeps on with the detailed description of the application and its structure, enriched with figures which provide a good example for the reader. Finally, some personal conclusions and a list of possible improvements are provided. The report ends with the chapters of the references and the Annex.

In this project there is not any profit, therefore an economic analysis has not been performed. The only cost of the project has been the development, as reported in the Budget.

3. FIGURES INDEX

4. LITERATURE REVIEW

The literature review of this report deals with three different aspects:

1. The three main academic databases
2. Bibliometric analysis
3. Systematic Literature Review

Each one of them is well analysed in the next chapters.

4.1 ACADEMIC DATABASES

There are three main academic databases, which are Scopus, Web of Science and Google Scholar. Scopus is provided by Elsevier Science and it has been launched in 2004. It covers more than 36,000 titles from more than 11,000 publishers. There are three types of sources: journals, trade journals and book series. The covered disciplines are life sciences, social sciences, physical sciences and health sciences. All the journals included in Scopus are yearly reviewed to ensure high-quality standards. For each title, Scopus provides for types of quality measures: H-index, CiteScore, SCImago Journal Rank and Source Normalized Impact per Paper. None of them have been used in the application.

Web of Science was produced by the Institute for Scientific Information (ISI) but now is managed by Clarivate Analytics. The discipline coverage includes sciences, social sciences arts and humanities, while the temporal coverage arrives up to 1900. Since 2008, it also hosts some regional databases, such as the Chinese Science Citation Database. Among the indexed literature, there are scholarly books, peer reviews journals, editorials and abstracts. The Web of Science database consists of six citation databases, which are the Science Citation Index Expanded, the Social Sciences Citation Index, the Arts & Humanities Citation Index, the Emerging Sources Citation Index, the Book Citation Index and the Conference Proceedings Citation Index.

Google Scholar is a free-access bibliographic database, launched in 2004. It contains academic journals, books, conference papers, theses and dissertations, patents and other scholarly literature. It contains more than 389 million documents. It has been strongly

criticized because it does not review journals and it includes predatory journals. Like the previous two databases, it provides citation indexing.

There are some differences between the three databases. The most important one regards the different criteria for the inclusion of documents: Scopus and Web of Science use a set of source selection criteria that are applied by expert editors. Google Scholar instead, includes all the documents that can find on the academic web. Each approach has different advantages and disadvantages: relying on selection criteria allows a more curated collection of documents. On the other hand, the selection criteria are sensitive to subjectivity and biases. It is observed that in Scopus and Web of Science, the Social sciences and Humanities documents are not well covered. The same happens for literature that is not written in the English language and documents which do not belong to journal articles. Including any academic document can lead to technical errors, like duplicate entries, incorrect information and wrong selection of documents.

For providing an example of the different coverage of the three databases, two papers have been analysed: Martín-Martín, A., Orduna-Malea, E., Thelwall, M., &, López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Infometrics, 12*, 1160-1177 and Martín-Martín, A., Orduna-Malea, E., & López-Cózar, E. D. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics, 116*, 2175-2188. The authors use a sample of more than 2,500 very highly cited documents across 252 subject categories from Google Scholar and released in 2017. The aim is to check if the documents are covered also in the other databases. The results show that a considerable amount of Social Sciences and Humanities are not covered by the other two databases. Then, the complete list of citations has been added to the analysis to see the percentage of coverage of them in the databases. Results show that Google Scholar was able to find 94% of citations to Social Sciences articles, while Web of Science and Scopus respectively 35% and 43%. Then, the authors analyse this difference and they discover that almost 60% of the citations found only by Google Scholar come from non-journal sources. Notwithstanding, the percentage of documents not covered by Scopus and Web of Science remains high. Therefore, a trade-off seems to be present: order and reviewed documents of Scopus and

Web of science or large coverage and consideration of not well-known sources of Google Scholar.

4.2 BIBLIOMETRIC ANALYSIS

The evaluation of the scientific production of researchers has become an important question either for National Governments either for individual institutions in the last years. National Governments are mainly interested in the assessment of performances to find out the most efficient and effective method for the allocation of public funding to the individual institutions. Individual institutions, instead, are more interested in evaluating their performances. Indeed, they can understand which are those fields where they excel and consolidate them, and they can identify which are those fields where they suffer and try to catch talents to boost them up.

The literature presents two techniques to perform the research evaluation: peer review and bibliometric methodology. The peer-review method is based on the presence of a panel of experts which provides a qualitative judgment according to a quantitative rating. They consider the main criteria, which are rationality, reliability, efficiency, effectiveness and impartiality. Depending on the type of research institutions, a maximum number of products realized within a certain period is chosen by the institution and submitted to the commission for the evaluation. This process should guarantee that only the best products are chosen to be evaluated. The result of the assessment process is a score for each product which allows the comparison among different fields and different institutions. The main advantages of the peer-review process are the following:

1. Low-quality products are weeded out directly by institutions.
2. Feedbacks received by experts help institutions in understanding how they can improve their production.

The main disadvantages of the peer-review process are:

1. The evaluation process requires a considerable amount of time and cost.
2. The objectiveness of measurement is not guaranteed: experts may produce unfair reviews.

The second approach used to measure the research activity is the bibliometric methodology. It is also called "*Scientometrics*" and it is defined as the application of quantitative analysis and statistics to publications. It is used by different institutions for different purposes. One reason is for the evaluation and strategic planning of the research institution. Other reasons are for finding new staff and improving funding decisions, for publicizing universities and capture students and for helping libraries to identify which are the most important journals to subscribe to.

The way of performing a bibliometric analysis is leveraging on some specific indicators which focus on the published scientific literature such as articles or conference proceedings and their accompanying citation counts. It does not exist an indicator which is suitable and perfect for all the previously described purposes of bibliometric analysis. Therefore, the starting point of the bibliometric analysis is to identify the research question. This is an extra effort but it is really necessary to optimize resources and reduce the possibility of mistakes and misunderstandings. For instance, for the development of the "*Talent Analyzer*" application, the interest of a bibliometric analysis is on the evaluation of the productivity of authors and the comparison of the importance of journals where they publish. To do this, two bibliometric indicators will be used: the number of papers published and the Impact Factor (IF). In addition to these indicators, information about the number of authors per year and the number of journals per year will be provided by the application. If the purpose of the application had been the analysis of the relative impact, the bibliometric indicators used would have been the percent cited/uncited papers and the crown indicator (c-index). If the purpose had been the analysis of the collaboration, metrics that include rates of co-authorship for pairs of authors would have been exploited.

The *Total Number of Publications* is one of the simplest bibliometric indicators which counts the number of articles that have been published by an author in a journal. This indicator can be computed at different levels, considering, for instance, the total number of publications of a certain author, in a certain field, for all the author of a certain university or in a certain country. It is possible to obtain this indicator directly from the database where papers are selected. The great shortcoming of this measure is that it is difficult to be interpreted: a higher number of publications does not always mean better. The indicator gives information about the productivity of the subject analysed but it does

not indicate the quality of the publication. A possible way to integrate the quality dimension into the *Total Number of Publications* measure is to consider only those articles which are published in the most important journals. However, in this case, the limitation regards to determine through which criteria a journal is objectively better than another.

To try to solve this problem, the *Impact Factor* measure has been introduced. It is computed as the average number of citations that articles published by a journal in the previous two years have received in the current year. The *Impact Factor* gives a functional approximation of the mean citation rate per citable item. The interpretation is the following: If a journal has an impact factor equal to 1.50 it means that, on average, the articles which have been published two or one years ago, have been cited 50% more than the average. The value of this index is available in the Web of Science Database. It allows the comparison among journals and the creation of a hierarchy within subject areas. It is important to bear in mind that a publication in a journal that has a high impact factor does not mean a high citation count. One of the main limitations of it is that it is easy to be manipulated with self-citation. To provide a more general overview, a modification of the *Impact Factor* has been introduced. It is called *5-Year Impact Factor* and it considers the articles published six or less years ago.

All the bibliometric indicators have been characterized by strong improvements and variation in recent years and this confirms the fact that bibliometric analysis has become the most widespread methodology for the evaluation of academic production. The types of indicators used in a bibliometric analysis are strongly influenced by the field considered and the research goals. They go from the simplest techniques, like the number of citation received by a paper, to the most recent and complicate techniques which try to include uncertainty or some other external factors like the size of the institution considered. Moreover, some indicators have been proposed directly to solve the shortcoming of the previous. Therefore, it is impossible to find the advantages and disadvantages that are valid for all the indicators but they have to be related to the indicator considered. On the other hand, it is possible to identify which are the advantages and disadvantages of bibliometric analysis. The advantages are:

1. Bibliometric analysis is done with quantitative indicators and therefore is objective. The results can be reproduced by everyone if the same method is used.
2. It requires a low amount of time and a low amount of cost. This is valid either for the computation either for the reproduction.
3. It can be applied at different levels: individual level (for instance, a single author), institutional level (for instance, a single university), national or international level.
4. It can be combined with peer-review techniques to provide a more complete analysis.

The disadvantages of performing a bibliometric analysis are:

1. The indicators used for bibliometric analysis can be affected and authors may cheat. An example is the phenomenon of self-citations.
2. Indicators need to be interpreted. For instance, a higher number of citations does not mean that the paper is good. It may be cited several times as a bad example.
3. They are not independent of the subject area: there are areas of study where is more used to publish and others where is not so common.

4.3 SYSTEMATIC LITERATURE REVIEW

A systematic review is a type of literature review that exploits systematic methods for the collection of secondary data and research studies and for synthetizing findings qualitatively or quantitatively. They are transparent and rigorous methods for generating robust and empirically derived answers to a focused research question. The systematic review was applied the first time in the field of medical sciences in the 1970s to study the effectiveness of health-care interventions. Nowadays, it is used in different fields, from social sciences to business and economics researches. When the research question is formulated, related studies are identified for providing a complete summary of the current evidence. When a systematic review uses statistical techniques to join the results of suitable studies, it is called a meta-analysis.

A systematic review has to achieve the following outcomes:

1. Find robust conclusions.

2. Criticize and synthesize one or more literature. Recognize relations and contradictions and analysing the reasons for these.
3. Develop a new theory or improve an existing one.
4. Give implications for practice and policy. Outline roads for future researches.

There are different reasons why a systematic literature review should be performed when is possible:

1. It is more comprehensible and allows a higher quality output compared to other types of literature review. For these two reasons, it has a higher probability to be published and cited.
2. Its high quality and high transparency ensure a relatively safe bet with journal peer reviewers.
3. It is based on five steps which make the process of the review tangible, less stressful and more traceable.
4. It fosters scientific rigor and it is reproducible.

A systematic review may be quantitative or qualitative. Whether the first or the second is the most appropriate, will depend on the existing literature, the question and the theoretical or empirical problems. If the aim is to put together studies that have tested the same hypothesis, a quantitative systematic review is used and is called meta-analysis. When the information is not quantitative, a qualitative systematic review is performed. There are two types of qualitative systematic review: a narrative review, which is used when there are quantitative studies that have used different methodologies and when the aim is to link together studies on different topics to find a new theory. The second type of quality systematic review is called qualitative meta-synthesis and it is used when the aim is to summarize qualitative studies for identifying key themes or theories which interest the review.

The process for realizing a systematic review can be split into five steps:

1. SCOPING: the starting point of a systematic literature review is the identification of the scope. It is translated into the definition of one or more research questions which clarifies what the reviewer wants to know and about which topics, who is the audience of the review and which are the possible findings that can help for

answering the question. Then, it is important to define the width of the review, considering the following trade-off: the smaller is the review, the simpler it is but the narrower the conclusions are. The width depends on several factors like cost, time, resources, the possibility of collaboration and so on. Another important step that belongs to the definition of the scope regards the check on the uniqueness of the review. If a review already exists, it may require an update. This step is linked with the last part of scoping which regards the familiarization with the literature. It is done with the aim of understanding if the research question may bring a contribution to the scientific knowledge.

2. PLANNING: here the systematic review is carefully planned. The starting point of the planning process is the division of the research question into search terms, which are analysed to find all the potential relevant and related works. When the user searches a term, he has also to consider different terminology of it: it is very useful to search synonymous, singular or plural terms, or change the spelling of the search term. This is done because relevant and related works may exist but they may be labelled with a similar name and not with the search term identified. The next planning step regards the definition of the borders of the review which is done with the identification of inclusion and exclusion criteria, according to the research question formulated. The reason whether an element is included or excluded has always to be present in the review. These criteria should be defined before the literature is reviewed and they should be revisited after the literature examination. If a borderline case is found, the inclusion or exclusion strongly depends on the type of literature review: if a meta-analysis is taking place, the case could be included and tested. All the decisions taken in this phase, as well as their results and consequences, should be stored in a record-keeping system. This is helpful either for the reviewer either for the reader. The reviewer can exploit the already consolidated guidelines which stress how to report a systematic review. An example is the PRISMA checklist.
3. IDENTIFICATION: here the methodical literature search is performed. The starting point is to find all the works which are related to the research question, looking in more than one database, according to the topic area. It is suggested that this step is conducted by two reviewers who finally have to agree to the works

included. A process to manage the case of disagreements has to be previously defined in the previous phase. When the search is concluded, the results have to be analysed, to check if the inclusion or exclusion boundaries have been respected and if the search terms were defined properly. Then, the reviewer has to be sure that not only all the related published works, but also the non-published works have been included or excluded. In fact, the reviewer has to consider the phenomenon of "*Publication bias*" that is defined as the tendency of publishing only studies which have statistically significant results. However, it is really difficult to find related non-published works. One way could be to contact researchers who have a related publication and ask if they have other related works not published.

4. SCREENING: in this phase, the results found are screened for deciding if they can be included or not. The results should be given to a citation manager software to delete possible duplicates. Then, the next suggested step is to read the abstract of the paper to understand if they still fit the inclusion criteria or not. If the abstract respects the inclusion criteria, the next step for the reviewer is to obtain the full text. The reviewer has always to keep track of the rejected articles.
5. ELIGIBILITY: it is the last step and it regards the reading of the full-text version of the article for deciding whether to include it or not. When an article has been definitively included, the analysis of contents begins. The aim is to extract the relevant information according to the topic and the research question. All the information extracted should be tabulated to remember them. It is important to remember that the inclusion and exclusion criteria detect the relevant or irrelevant works and do not provide any indication about the quality of them. This has to be assessed by the reviewer, using some specific available tools. This is not a straightforward process and there are several articles that discuss the study quality tools and their application.

In these steps, some elements which will be used in the "*Talent Analyzer*" application have been introduced. For instance, it is clearly stated that the systematic literature review is the tool used if a content analysis has to be performed. The topic which interests the application, as the name suggests, is "*Talent Management*". Another example of a clearly

recognizable concept which is present also in the application is the analysis of the presence of the "*Research Question*" and the "*Research Aim*".

The reasons why a systematic literature review should be performed have already been discussed previously. This chapter ends with some weaknesses of the systematic literature review:

1. There is the possibility that it requires an update, even in the same year of the publication.
2. It is difficult to find inherent nonpublished articles.
3. It is difficult to skim the low-quality articles

5. IT ARCHITECTURE

The "Talent Analyzer" platform is built using three open-source software: R, MySQL and Shiny. All of them are well described in the next subchapters.

5.1 R

R is a language and an environment for statistical computing and graphics, developed by Robert Gentleman and Ross Ihaka and which belongs to the GNU project. It is free software, available for different operative systems, such as Unix, Linux, macOS and Microsoft Windows. R is considered an implementation of S, a statistical programming language developed in 1976 by John Chambers, Rick Becker and Allan Wilks in the Bell Laboratories. The first R version was released on the 22nd of November 1999 but the first version considered stable by the authors for the production use was released on the 29th of February 2000.

R is endowed with two features: statistical and programming features. The statistical and graphical techniques involve, for instance, linear and nonlinear modelling, statistical tests, time-series analysis, clustering and classification. Users can write C, C++ and Fortran code to compute intensive tasks and they can modify R objects writing in C, C++, Java or Python. R gives also the possibility to produce high-quality graphs which may include mathematical symbols. Graphs produced can be static or interactive.
R is also an interpreted language, meaning that instructions are executed directly, without previously compiling a program (this case is defined compiled language). It allows matrix arithmetic and it accepts vectors, matrices, arrays, lists and data frames. It provides the possibility to write and manipulate functions which are characterized by a "lazy" behaviour because the function arguments are evaluated only when the function is used and not when it is called.

R can be easily expanded with packages that can be created by users and developed in R or also in Java, C or C++. The aim is to increase R capabilities, allowing, for instance, specialized statistical techniques or reporting tools (like Rmarkdown). With the installation of R, a small core set of packages is provided. However, there are more than

15,000 packages which are mainly available at the Comprehensive R Archive Network (CRAN) or at GitHub. Finally, R has its LaTeX-like documentation format, used to produce comprehensive online or offline documentation.

The main advantages and the main reasons for the choice of using R for the development of the platform are:

1. It is an open source software. This means that R can be downloaded and changed by everyone and this is valid also for the platform. The unique obligations imposed by the GNU General Public License is that if there is a change in the R source code, it must be available for anybody. Another direct benefit is that, since any user can modify and improve the code, the R software is becoming more and more stable and reliable.
2. It is available for Windows, Unix systems and Mac. For the platform, this is really important because it means that anyone can use it, independently from the operative system.
3. It has a well-developed and active community, spread all over the world. This means that many people can help and be helped by other people for solving their problems. The community usually meets online, through websites like Stack Overflow. However, there is also a growing number of R events, such as the official annual gathering of R users, called "userR!", which has been recently held in Toulouse (France) and the next year will be in St. Louis (Missouri). Finally, there is also a journal, called "The R Journal" which contains articles about the use and development of R. For the development of the platform the presence of advanced R users in Stack Overflow has been fundamental because it allowed to solve doubts with posts previously made by other users and to obtain easily the necessary information.
4. It connects with other languages, databases and applications. This is really important for the development of the platform because it stores data in a MySQL database. The connection is done with the R Package "*RMySQL*".
5. It is endowed with extensive libraries for data wrangling and computation of statistics. The first one is time-consuming and very important because the cleaning of data affects all the results of the analysis. Therefore, it has to be performed carefully and with the right tools. R provides different packages to do it, among

which there is "*dplyr*". The statistical analysis is continuously made easier by the development of new packages and improvement of the existent ones, operated by the well-consolidated R community.

6. It provides extensive libraries for data visualization, with high-quality graphs that allow gaining insights about the data and hidden patterns. R has packages either for the realization of a simple and immediately comprehensible plot, either for the realization of 3D and complex graphs. One of the most famous packages is "*ggplot2*" which is becoming the standard for realizing graphs. It has been really useful during the development of the platform thanks to its immediacy even for the realization of complex graphs.

The version of R software used for the development of the application is 3.6.1. The user must have a version equal or more recent for the correct functioning of the application.

For the development of the platform, RStudio has been used. It is an integrated development environment (IDE) for R, which has been developed by the enterprise "RStudio, Inc.", which has no formal connection with the founder of R. It is available in two versions: RStudio Desktop, which is a desktop application, and RStudio Server, which exploits a web browser and a remote Linux server. It is mostly developed in Java and the first release was the 28th of February 2011. It is made by a console, a syntax-highlighting editor, which enables the immediate code running, tools for plotting, history, debugging and workspace management. The main advantages of using RStudio are:

1. It makes easier the creation of scripts. It allows seeing at the same time the script, the execution of it and the results. Moreover, while the user writes the command, potential syntax options are displayed.
2. It makes easier the interaction with objects present in the environment, highlighting the type of object and some related information.
3. It makes easier the set-up of the working directory and the interaction with documents present in the computer, by dedicating a specific window for this.

The version of RStudio used for the development of the application is 1.2.5001. The user must have the same or a more recent version for the correct functioning of the application.

5.2 MySQL

MySQL is an open-source relational database management system, which is defined as a software system that allows a user to define, create keep and control the access to a database. It has been developed by the Swedish company "MySQL AB" and released for the first time on the 23rd of May 1995. It is available for many system platforms, among which there are Microsoft Windows, macOS and Linux. It is used for many database-driven web applications, such as WordPress, and many popular websites, such as Facebook, Twitter and YouTube. MySQL is based on a client-server architecture, which is a relationship where one party (the client) asks for a service or a resource from another party (the server). The centre of MySQL is the server, called MySQL server, which manages the database commands. The server is available also as a library which can be linked with other external applications. The commands are sent from MySQL client. Usually, MySQL is set up on a computer that sends it to different locations, to which users can access from different interfaces. These interfaces have the goal to send to the server the commands and display the results. Data can be stored in different storage engines, such as InnoDB or CSV, and the user can access through standard SQL commands. MySQL is developed in C and C++ and it supports different formats of data, like "text", "date", "float" or "integer". The main MySQL client program, which is also the one used for the development of the platform, is MySQL Workbench. It is the official integrated environment for MySQL and it allows the user to design the database structure and manage databases. There is a subscription-based version of MySQL dedicated to enterprises and called MySQL Enterprise Server, which is enriched with a series of extensions.

The main advantages and reasons for using MySQL Server and MySQL Workbench for the development of the platform are:

1. It is the most secure database management system, also used by famous web applications, such as Facebook and YouTube. It uses an access privilege and encrypted password system which permits host-based verification.
2. It guarantees high flexibility and high performance. Independently from the quantity of data stored, the footprints left by the system is trifling. Therefore, it guarantees high performance either if it is used by an eCommerce website which

gets millions of queries every day, either if it is a high-speed transactional processing system.

3. It is simple to be installed and used and it assures an uptime equal to 24 hours per 7 days. The simplicity is guaranteed by the possibility of install third-party tools and the use of the standard SQL language.
4. It is an open-source system with a large community. As for R, members of the community can help each other to solve their problems. Moreover, the open-source characteristic assures fast upgrades and easy maintenance.
5. It allows immediate access to organized data. Data are stored into tables which are characterized by different types of relationship. If the user has permission, he can access the data immediately with simple SQL commands. Tables can be easily modified, updated, deleted, visualized and managed from the Object Browser.

The main reason for the use of a database and a database management system for the development of the platform is the possibility of storing a huge amount of data precisely and have quick access to them. The user of the platform does not have to import or export excel files. As will be described later, the data added by the user are directly stored in the database and any modification performed by the user through the platform directly affects the data stored in the database.

The version of MySQL Workbench used for the development of the application is 8.0.13. The user must have the same or a more recent version for the correct functioning of the application.

5.3 SHINY

Shiny is an R package developed to build interactive web applications easily and directly from R. The application can be run locally in the computer with RStudio or create as a standalone. This second option can be obtained in two ways: deploying the application to a Shiny Server or use a hosting service. The Shiny package contains built-in examples which show how the application works. A Shiny application is made by two components:

1. The User Interface (UI): it is an object responsible for the appearance and the layout of the application. In the UI are placed two elements: the reactive outputs,

defined as elements (which can be tables or graphs) which respond to the user actions, and which are triggered by the control widgets. The second elements are the control widgets, defined as web elements which send messages to the application and with whom the user can interact.

2. The Server: it is a function which contains the instructions that the computer needs for building the application and for reacting to the changes that user do through the UI.

One of the most famous functions used for building the UI is "*fluidPage*". It enables to generate a display that automatically adapts to the user's browser window. Two other famous functions are "*titlePanel*" and "*sidebarLayout*". The first one allows giving a title to the page while the second one is used to allocate space for control widgets and reactive outputs. This last function has two arguments: "*sidebarPanel*", where the control widgets are placed, and "*mainPanel*", where the reactive outputs are placed. An example is provided by the figure below:

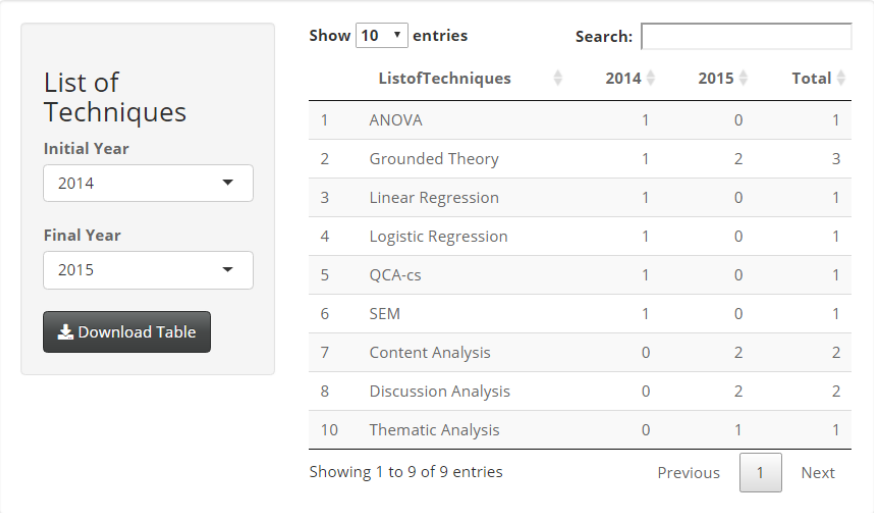


Fig.1: *Example of "sidebarPanel" (left) and "mainPanel" (right). The first one is placed to the left and it contains three control widgets, two for choosing the year of the analysis and one for downloading the table. The second one is placed to the right and contains one reactive output that, in this case, is a datatable.*

The commands which are responsible for creating the reactive objects and for updating them according to the choices of the user are placed in the server function. The element created in the server function and the widget placed in the UI object are linked by a character string which is the name of the reactive element. In the figure below there is the

example of the code placed in the server function needed to compute the table above and the code placed in the UI object needed to display the table for the user:

```
output$tablot = DT::renderDataTable(datatable({
  listoftechniques<-listoftechniquesinput()
  listoftechniques<-listoftechniques [!duplicated(listoftechniques[c(1,2,3)]),]
  listoftechniques<-separate_rows(listoftechniques, ListofTechniques, sep = "; ", convert = TRUE)


  listoftechniques <- listoftechniques[,-c(1)]
  rownames(listoftechniques) <- seq(length=nrow(listoftechniques))
  listoftechniques <- group_by(listoftechniques, Year, ListofTechniques )
  listoftechniques <- as.data.frame(tally(listoftechniques))
  colnames(listoftechniques)[3] <- "Frequency"


  listoftechniques1<-reshape(listoftechniques,timevar="Year",idvar="ListofTechniques",direction="wide")
  names(listoftechniques1) <- substring(names(listoftechniques1), 11)
  index<-listoftechniques1[1]
  listoftechniques1[1]<-NULL
  listoftechniques1[is.na(listoftechniques1)]<-0
  listoftechniques1$Total <- rowSums(listoftechniques1)
  listoftechniques1<-cbind(index, listoftechniques1)
  colnames(listoftechniques1)[1] <- "ListofTechniques"
  listoftechniques1}, rownames = TRUE))
```

Fig.2: *R code placed in the Server function needed to build the table displayed in Fig.1*

```
mainPanel(
  dataTableOutput("tablot"))
```

Fig.3: *R code placed in the UI object needed to display the table for the user.*

From **Fig.3** is possible to notice that the name in the "*dataTableOutput*" function (which is a reactive output) is equal to the character string used in the server in **Fig.2** for building the reactive element. In this way, the link among server and user interface is established.

All the application has been built following the logic described by the previous example. The main advantages of using R Shiny package for building the application are:

1. It is used to create an interactive application that allows people to explore a dataset and present to the end-users results of an analysis with graphs and tables, permitting the interaction with them. The application developed with Shiny is developed directly for the web, removing the old necessity of relying on a web developer to convert the work from R into GUI experiences.
2. It guarantees a beautiful user interface that adapts to user's changes.
3. The application can be hosted online and it is shared easily using platforms like "*shinyapps.io*".

4. It allows the creation of attractive interactive visualizations with no need to use HTML or JavaScript.
5. It gives the possibility to the user to produce common reports.
6. The developer can benefit from the previously described advantages of R and RStudio.

6. SET UP OF THE PROJECT

To run the application, three files are necessary:

1. A MySQL script needed to create the database, create the tables for storing data, create the user and grant permissions and finally populate the tables with the first sample of data used to provide an example of how the application works.
2. An R script needed to run the application. It contains the UI object, the server function and the list of the packages required for the right execution of the application.
3. An R markdown document needed to allow the creation of the report from the application.

In this chapter are presented all the steps which are necessary to set up the project. In particular, a lot of attention is put on the MySQL script because it contains the instructions for the creation of the database and the tables which are the first steps to prepare the application. Then, the attention is shifted to the R script and, in particular, to the R packages without which the application cannot work. Finally, the structure of the R markdown document is briefly presented.

6.1 CREATION OF THE DATABASE AND TABLES

The databased is composed of 9 tables:

1. **tbl_project**: this table provides the list of the projects that are ongoing by the user of the platform. It is made by two columns:
    1) *id_keyword*: it contains the ID of the project and it is the primary key of the table. A primary key is defined as that column that uniquely identifies the records in the specific table. The type of data stored in this column is an integer. A maximum space of 20 characters is reserved.
    2) *tx_keyword*: it contains the name of the project. The type of data stored in this column is a string of characters. A maximum space of 100 characters is allocated.

2. **tbl_journal**: this table contains the list of the academic journals which are analysed by the user. The table is composed on three columns:
    1) *id_journal*: it is the ID of the journal and the primary key of the table. The data stored is an integer and a space of 20 characters is allocated.
    2) *tx_journalname:* it contains the name of the journal. The data stored is a character string, with a maximum space of 250 characters.
    3) *tx_ISSN*: it contains the ISSN code of the journal. It is an eight-digit number used to uniquely identify a serial publication. The data is a character and the space allocated is 20.

3. **tbl_keyword**: this table contains the lists of keywords. It is composed of two columns:

    1) *id_keyword*: it is the ID of the keyword list and the primary key of the table. The data type is an integer and the space reserved for the data is equal to 20 characters.
    2) *tx_keyword*: it contains the list of the keywords. The data type is character and the maximum number of characters admitted is 1,000.

4. **tbl_affiliation**: this table contains the list of the affiliations of the authors of the papers. It is composed on three columns:

    1) *id_affiliation*: it is the ID of the affiliation and the primary key of the table. The data type is an integer and the space reserved for the data is equal to 20 characters.
    2) *tx_country:* it contains the name of the country where the affiliation is. The data type is a character string and the maximum number of characters admitted is 100.
    3) *tx_affiliation:* it contains the name of the affiliation. The data type is a character string and the maximum possible dimension of it is 1,000 characters.

5. **tbl_paper**: this table contains the list of the papers analysed. It is composed on the following columns:

1) *id_paper*: it is the ID of the paper and the primary key of the table. The data type is an integer and the space reserved for the data is equal to 20 characters.
2) *tx_title:* it is the column that contains the title of the paper. The data type is a character string with a maximum dimension of 1,000 characters.
3) *id_ext_journal:* this column contains the ID of the journal where the paper is published. It is an integer with a maximum space of 20 characters.
4) *ext_journal:* this column contains the name of the journal where the paper is published. It is a character with a maximum dimension of 250.
5) *tx_volume:* it is a character that contains the volume of the journal where the paper is published. Maximum space of 45.
6) *tx_pages:* it is a character that contains the pages of the journal where the paper is published. Maximum space of 45 characters.
7) *int_year:* it is an integer that is used to store the year of the journal where the paper is published. The maximum length of the integer is 11 characters.
8) *fl_if:* this column is used to store the impact factor of the journal. This is an annual journal value which measures the number of citations received in a specific year by a journal divided by the total number of possible citable items in the same journal during the period analysed. The type of the variable is float.
9) *tx_abstract:* in this column is placed the abstract of the paper. The type is a character with a maximum value of 10,000.
10) *int_nauthors:* this column stores the number of authors of the analysed paper. It is an integer with a maximum length of characters of 11.
11) *bin_nauthinter:* this is a binary variable. 1 means that the paper is written by authors of different nationalities, otherwise the value is 0.
12) *id_ext_project:* it is an integer variable that stores the ID of the project at which the paper belongs. The maximum length is 20.
13) *ext_project:* this last variable contains the name of the project at which the paper belongs. It is a character with a maximum dimension of 100 characters.

6. **tbl_author**: this table provides the list of all the authors that have written at least one paper that is present in the application. It is made by five columns:
    1) *id_author*: it contains the ID of the author and it is the primary key of the table. The type of data stored in this column is an integer. A space of maximum of 20 characters is reserved.
    2) *tx_surname:* it contains the surname (or surnames) of the authors. The data type is a character with a maximum dimension of 100.
    3) *tx_forename:* it contains the name (or names) of the authors. The data type is character, maximum dimension of 100.
    4) *bin_academic:* it is a binary variable: 1 means that the author is academic, 0 otherwise.
    5) *tx_education:* it is a character variable that has been introduced if the user wants to write some comments about the education of the author. The maximum length available is 100 characters.

7. **tbl_paper_author**: this table provides details about the authors and the papers published. It connects the **tbl_paper, tbl_affiliation** and the **tbl_author** previously described. It is made by the following columns:
    1) *id_paper_author*: it contains the ID of the observation, which represents a specific author that has published a specific paper. The type of data stored in this column is an integer. A space of a maximum of 20 characters is reserved.
    2) *id_ext_paper*: it contains the ID of the paper that has been published by the author. The data type is an integer with a maximum length of 20 characters.
    3) *ext_paper:* it contains the title of the paper that has been published by the author. The type of the column is a character, with a maximum dimension of 1,000.
    4) *id_ext_author:* it contains the ID of the author that has published the specific paper. The type of column is an integer, with a maximum dimension of 20 characters.

5) *ext_author*: it stores the name and the surname of the author which has written the specific paper. The type of the variable is a character, with a maximum dimension of 200.
6) *id_ext_affiliation1*: it contains the ID of the first affiliation at which the author belongs. It is an integer of a maximum length of 20 characters.
7) *ext_affiliation1:* it contains the name of the first affiliation at which the author belongs. It is a character of a maximum length of 1,000.
8) *id_ext_affiliation2*: it contains the ID of the possible second affiliation at which the author belongs. It is an integer of maximum length of 20 characters.
9) *ext_affiliation2:* it contains the name of the possible second affiliation at which the author belongs. It is a character of maximum length of 1,000.
10) *int_position:* this variable is used to record the position of the author among all those who have published the same paper. If, for instance, the name of the specific author is the first which appears in the paper, the value of the variable is equal to 1. The type of variable is an integer with a maximum dimension of 20 characters.
11) *bin_institution:* this is a binary variable: 1 means that the author belongs to an institution, 0 otherwise.

8. **tbl_paper_keyword**: this table provides details about the papers and the relative keyword. It connects the **tbl_paper** and the **tbl_keyword** previously described. It is made by the following columns:
    1) *id_paper_keyword*: it contains the ID of the observation, which represents a specific paper with the corresponding keyword. The type of data stored in this column is an integer. A space of a maximum of 20 characters is reserved.
    2) *id_ext_paper*: it contains the ID of the paper. The data type is an integer with a maximum length of 20 characters.
    3) *ext_paper:* it contains the title of the paper. The type of the column is a character, with a maximum dimension of 1,000.

4) *tx_keyword:* it contains the text of the keyword for the corresponding paper. The type of the variable is character, with a maximum dimension of 1,000 characters.

9. **tbl_content**: this table provides details about the content of the paper. It is made by the following columns:
    1) *id_content*: it contains the ID of the content that will be analysed. One and only one content is associated with one paper. The type of data stored in this column is an integer. A space of maximum 20 characters is reserved.
    2) *id_ext_paper*: it contains the ID of the paper. The data type is an integer with a maximum length of 20 characters.
    3) *ext_paper:* it contains the title of the paper. The type of the column is a character, with a maximum dimension of 1,000.
    4) *tx_study_type:* it contains information about the type of study of the specific paper. Three options are available for the user to be chosen: "qualitative", "quantitative" or "mixed" study. The type of variable is a character with a maximum length of 45.
    5) *tx_methodology:* this variable gives information about the methodology used for the study of the specific paper. The methodology can be "inductive" or "deductive". The type of variable is a character with a maximum length of 45.
    6) *tx_method_list:* this variable contains the list of methods used during the study of the specific paper. The character " ; " is used to separate the methods in the list. The type of variable is a character with a maximum length of 100.
    7) *tx_technique_list:* this variable contains the list of techniques used during the study of the specific paper. The character " ; " is used to separate the techniques in the list. The type of variable is a character with a maximum length of 100.
    8) *tx_level_list:* this variable contains the list of levels of analysis of the specific paper. The character " ; " is used to separate the levels in the list. The type of variable is a character with a maximum length of 100.

9) *tx_level_notes:* this variable contains the comments done about the level of the analysis of the specific paper. The type of variable is a character with a maximum length of 500.
10) *tx_data_source_list:* this variable contains the list of data sources used during the study of the specific paper. The character " ; " is used to separate the sources on the list. The type of variable is a character with a maximum length of 100.
11) *int_size_source_list:* this variable contains the list of the size of the samples of the specific paper. The character " ; " is used to separate the different samples in the list. The type of variable is a character with a maximum length of 100.
12) *txt_focus_group:* in this column is stored the Focus group of the content analysis of the specific paper. The variable is a character, with a maximum dimension of 100.
13) *tx_country_type:* this variable defines the number of different countries present in the study of the specific paper. The values this variable can assume are single or multiple. The type of the variable is a character with a maximum number of characters equal to 45.
14) *tx_country_list:* this variable represents the list of the different countries present in the study of the specific paper. The character " ; " is used to separate the countries in the list. The type of variable is a character with a maximum length of 100.
15) *tx_continent_type:* this variable defines the number of different continents present in the study of the specific paper. The values this variable can assume are single or multiple. The type of the variable is a character with a maximum number of characters equal to 45.
16) *tx_continent_list:* this variable represents the list of the different continents present in the study of the specific paper. The character " ; " is used to separate the continents in the list. The type of variable is a character with a maximum length of 100.
17) *tx_sector_type:* this variable defines the type of sector present in the study of the specific paper. The values this variable can assume are "single" or

"multiple". The type of the variable is a character with a maximum number of characters equal to 45.

18) *tx_sector_list:* this variable represents the list of the different sectors present in the study of the specific paper. The character " ; " is used to separate the sectors in the list. The type of variable is a character with a maximum length of 100.
19) *tx_company_size_type:* this variable is used to classify the type of companies by size of the specific paper. The values this variable can assume are "single" or "multiple". The type of the variable is a character with a maximum number of characters equal to 45.
20) *tx_company_size_list:* this variable represents the list of the different sizes of the companies of the specific paper. The character ; is used to separate the sizes in the list. The type of variable is a character with a maximum length of 100.
21) *tx_company_owner_type:* this variable is used to classify the type of companies by the owner of the specific paper. The values this variable can assume are "single" or "multiple". The type of the variable is a character with a maximum number of characters equal to 45.
22) *tx_company_owner_list:* this variable represents the list of the different owners in the company of the specific paper. The character " ; " is used to separate the type of owners on the list. The type of variable is a character with a maximum length of 100.
23) *tx_company_scope_type:* this variable is used to classify the type of companies by scope for the specific paper. The values this variable can assume are "single" or "multiple". The type of the variable is a character with a maximum number of characters equal to 45.
24) *tx_company_scope_list:* this variable represents the list of the different scopes of the company of the specific paper. The character " ; " is used to separate the type of scopes in the list. The type of variable is a character with a maximum length of 100.
25) *tx_company_nonprofit:* this column is used to address the following question: is the company a non-profit company? Two answers are

available: “yes” or “no”. The type of the variable is a character with a maximum number of characters equal to 45.

26) *memo_extra_notes:* this variable stores the extra notes that the user can add on the sample of the specific paper. The type of the variable is a character with a maximum length equal to 250.
27) *tx_rq_exist:* this column is used to address the following question: does the research question exist? Two answers are available: “yes” or “no”. The type of the variable is a character with a maximum number of characters equal to 45.
28) *memo_rq_description:* this variable stores the comments about the research question that the user can add for the specific paper. The type of the variable is a character with a maximum length equal to 250.
29) *tx_aim_exist:* this column is used to address the following question: does the research aim exist? Two answers are available: “yes” or “no”. The type of the variable is a character with a maximum number of characters equal to 45.
30) *tx_aim_description:* this variable is used for the description of the research aim of the specific paper. The type of the variable is a character with a maximum length of 100.
31) *memo_aim_description:* this variable stores the comments about the research aim of the specific paper that the user can add. The type of the variable is a character with a maximum length equal to 250.
32) *tx_focus_list:* this variable represents the list of different focus of the specific paper. The character “ ; ” is used to separate the values in the list. The type of variable is a character with a maximum length of 100.
33) *tx_focus_notes*: this variable stores the description of the focus of the specific paper. The type of the variable is a character with a maximum dimension of 100.
34) *tx_theory_exist:* this column is used to address the following question: does the academic theory exist? Two answers are available: “yes” or “no”. The type of the variable is a character with a maximum number of characters equal to 45.

35) *tx_theory_list:* this variable represents the list of different academic theories of the specific paper. The character " ; " is used to separate the values in the list. The type of variable is a character with a maximum length of 100.
36) *tx_talent_exist:* this column is used to address the following question: has the talent been defined? Three answers are available: "yes-direct", "yes-indirect" or "no". The type of the variable is a character with a maximum dimension of 45.
37) *memo_talent_definition:* this variable stores the comments about the definition of the talent that the user can add for the specific paper. The type of the variable is a character with a maximum length equal to 250.
38) *tx_talent_source:* in this column the source of the definition of the talent of the specific paper is stored. The type of the variable is a character with a maximum length equal to 100.
39) *tx_tm_exist:* this column is used to address the following question: has talent management been defined? Three answers are available: "yes-direct", "yes-indirect" or "no". The type of the variable is a character with a maximum dimension of 45.
40) *memo_tm_definition:* this variable stores the comments about the definition of the talent management that the user can add for the specific paper. The type of the variable is a character with a maximum length equal to 250.
41) *tx_tm_source:* in this column the source of the definition of the talent management of the specific paper is stored. The type of the variable is a character with a maximum length equal to 100.
42) *tx_tmp_exist:* this column is used to address the following question: has the talent management practice been defined? Two answers are available: "yes" or "no". The type of the variable is a character with a maximum dimension of 45.
43) *tx_tmp_list:* this variable represents the list of the different talent management practices. The character " ; " is used to separate the values in the list. The type of variable is a character with a maximum length of 100.

44) *memo_tmp_notes:* this variable stores the comments about the talent management practices that the user can add. The type of the variable is a character with a maximum length equal to 250.
45) *memo_final_notes:* this variable stores the final comments on the specific paper that the user can add. The type of the variable is a character with a maximum length equal to 250.

All the variables of all the tables can assume a null value with the exception of all the primary keys which cannot be null by definition.

An example of the commands used in the MySQL script necessary to create the database and to create the previously described tables is shown in the figure below:

```
DROP SCHEMA IF EXISTS dbsystem;
CREATE SCHEMA IF NOT EXISTS  `dbsystem` ;
use dbsystem;

CREATE TABLE IF NOT EXISTS `dbsystem`.`tbl_project` (
    `id_project` BIGINT NOT NULL,
    `tx_nameproject` VARCHAR(100) NULL,
    PRIMARY KEY AUTO_INCREMENT (`id_project`)
);
```

Fig.4: *MySQL code for creating the database and the tbl_project.*

The "*Drop*" function, which is possible to see in **Fig.4**, is used to delete any previous redundant schema if there is one. The "*Create*" function is used to create the database and call it "dbsystem". Then, the "*Use*" command is used to set the database that has just been created as the one to work with. Finally, the "*Create*" function is used again to create the table with the attributes previously described. The primary key is set with the option of autoincrement to avoid that the user, when a new observation is recorded, has to specify also the ID.

Then, an example of the code used for the creation of the user and the filling on the tables is provided in the figure below:

```
CREATE USER IF NOT EXISTS 'userplat'@'localhost' IDENTIFIED BY 'userplatpassword';
GRANT ALL PRIVILEGES ON dbsystem.* TO 'userplat'@'localhost';
FLUSH PRIVILEGES;
ALTER USER 'userplat'@'localhost' IDENTIFIED WITH mysql_native_password BY 'userplatpassword';

INSERT INTO tbl_project
VALUES ('1','Overview of Academic Talent');
```

Fig.5: *MySQL code used for the creation of the user and the filling of the tbl_project.*

The "*Create*" function, which is possible to see in **Fig.5**, is used to create the new user, which is identified by a password. Then, the "*Grant*" function is used to assign to the new user created all the privileges on the database. Among them, there is the privilege to use the functions "*Update*", "*Insert*" and "*Delete*". Then, the function "*flush privileges*" is required by some old version of MySQL to reload the grant tables when a function like "*Update*", "*Insert*" or "*Delete*" is used. This command has been included for precaution. Then, "*alter user*" is necessary to specify the authentication plugin. An authentication plugin is a programmable piece of code used to demonstrate that a MySQL account is owned by whoever it claims. Finally, the filling of the table is done with the function "*insert into*", where is necessary to specify the table, and "*values*", where is required to write the values to be added in the columns of the table.

6.2 R PACKAGES

In this chapter are enlisted and explained the R packages necessary for the correct execution of the application. An R package is a collection of functions, data and compiled code that are stored in the R environment. The packages used are:

1. **shiny:** it is the package that allows the creation and execution of the web application with R. It includes the prebuilt widget used for the interaction of the user with the platform.
2. **DT:** this package is used to render the data objects introduced in R as HTML tables.
3. **DTedit:** this package allows the creation of editable data tables for a shiny application. On the package are included three call-back functions that provide a straight mechanism for the user to add, edit or delete a row in the data table.

Another advantage of this package is that it allows direct and easy interaction with a permanent data storage scheme. The package has fewer options for modifying the aspect of the data table compared to those provided by the package **DT**. The download from CRAN is not available, because the package is still under development. To install this package, the user has to use the **devtools** package and run the following code:

devtools::install_github('jbryer/DTedit')

All the tables of the application have been rendered with **DTedit** package, apart from those who belong to the page "*List of Papers*" because, as it will be explained in the chapter *"7.2 Managing"*, those tables have to interact among them. They are built with the package **DT**.

4. **RMySQL**: this package is necessary to create the legacy DBI interface to MySQL. It is indispensable for the application because it allows sending SQL queries from R to MySQL database if a connection among them is established. Therefore, for instance, any time a user inserts a new row of data, the R code sends the SQL query to the database, which is updated.
5. **dbConnect**: this package is responsible to set the connection between R and MySQL. Without this connection, the two software cannot communicate.
6. **data.table**: this package is required by the application for managing and modifying data in data tables, in a fast and efficient way. It is really important to do not confuse this package with the package **DT**. This package is used to provide all the functions required to modify the data stored in the table, while **DT** (or **DTedit**) is used only to display the table as an HTML element.
7. **dplyr**: this package is necessary for data manipulation. It contains essential functions that allow to filter, group and reorder data.
8. **rowr**: also this package is used for data manipulation. In particular, it provides functions that work on all R objects as if they were arranged in rows.
9. **plyr**: it provides a set of tools for splitting a big data structure into small pieces, apply a function to each piece and then combine everything back together. It is still connected to data manipulation.

10. **tm**: this package is essential for the manipulation of data that will be used for the creation of the word cloud for the bibliometric analysis. It provides all the tools for text mining.
11. **wordcloud**: this package contains the function which is used to create a pretty word cloud.
12. **ggplot2**: this package provides all the tools for the realization of all the graphs present either in the bibliometric analysis either in the content analysis, with the unique exception of the word cloud which required a specific function. This package has been preferred to the classic and pre-installed package for graphs because it gives more tools for the management of axis, labels, legends and it also allows to create more beautiful and complex graphs, like the world map for the country analysis present in the content analysis.
13. **rmarkdown**: this package is necessary for the creation of the report from a markdown file, enabling to choose the format among HTLM or WORD.
14. **knitr**: this package provides some tools for generating dynamic reports. In particular, it has resulted fundamental for the realization of the report because it has permitted the realization of tables both for HTML and WORD format, with the function "*kable*".
15. **tidyr**: this package is required for data tidying, which is the process to structure data in a way that makes easier the analysis and use of them. In particular, this package provides functions that have been used to reshape the data taken from MySQL tables and realize the majority of the tables of the content analysis.
16. **sp**: this package (and the following one) is used only to create the world map for the country study in the content analysis. This package, in particular, contains classes and methods for spatial data. It has been useful for retrieving coordinates.
17. **maps**: this package is used for displaying the world map of the country study in the content analysis, realized through the **sp** package.

All these packaged, apart from **DTedit**, are installed from CRAN, with the following standard command. Among the quotation marks, the name of the package is placed:

install.packages("shiny")

Each package has to be loaded for providing the required functions. This is done with the following standard command. Among the quotation marks, is placed the name of the package:

library("shiny")

6.3 R MARKDOWN DOCUMENT

The R Markdown document is named "*report.Rmd*". This file must not be renamed and must stay in the same folder with the R script. It contains all the chunks that are executed when the user creates the report (see chapter "*7.4 Reporting*") of the analysis done. It realizes a WORD or HTML document (according to the format chosen by the user) that can be stored anywhere and named as the user wants. The structure of this document is straightforward: initially, there is one part where are printed the details that the user specifies, which are the title of the report, the author, the date, some comments and the period of the analysis. Then, there is a number of chunks equal to the number of tables and graphs that are available for the bibliometric and content analysis. Each chuck, if selected by the user, contains the code which simply prints the graph or table that has been realized in the previous analysis. The chuck does not compute another time the graph or the table, but it takes those realized in the previous sections of the application. This will be better explained in the chapter "*7.4 Reporting*".

7. DESIGN OF THE APPLICATION

This chapter contains a description of the application with all its features. The application has seven pages, which are shown in the figure below:

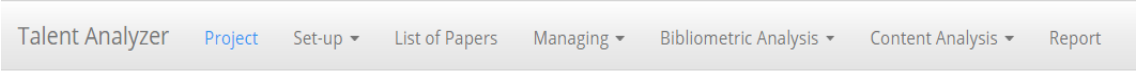


Fig.6: *The seven pages of the "Talent Analyser" application.*

The seven pages can be grouped as follow:

1. *Set-up of the project*: it contains the first and second pages, which are used to create the project (if it does not exist) and add all the information regarding journals, authors and affiliations.
2. *Managing*: it contains the third and fourth pages. The first one is a summary page of the papers and the related information. The second one is composed of a group of subpages that give to the user the possibility to manage the information of the papers and the related authors, keywords and contents.
3. *Analysis*: it contains the fifth and sixth pages. The first one groups all the subpages responsible for the bibliometric analysis whilst the second one groups all the subpages responsible for the content analysis.
4. *Reporting*: it regards the last page, which is used to give to the user the possibility to download a report of the analysis just performed.

In the next subchapters, all the pages will be described one by one.

7.1 SET-UP OF THE PROJECT

This group contains the page "*Project*" and the page "*Set-up*".

7.1.1 PROJECT

As far as the page "*Project*" is concerned, it is the direct representation of the data stored in the MySQL table "*tbl_project*". The column "*ID*" shows the id associated with the

project. The column "*Name*" shows the name given to the project. The figure below shows how the table looks and the different buttons that the user can use.

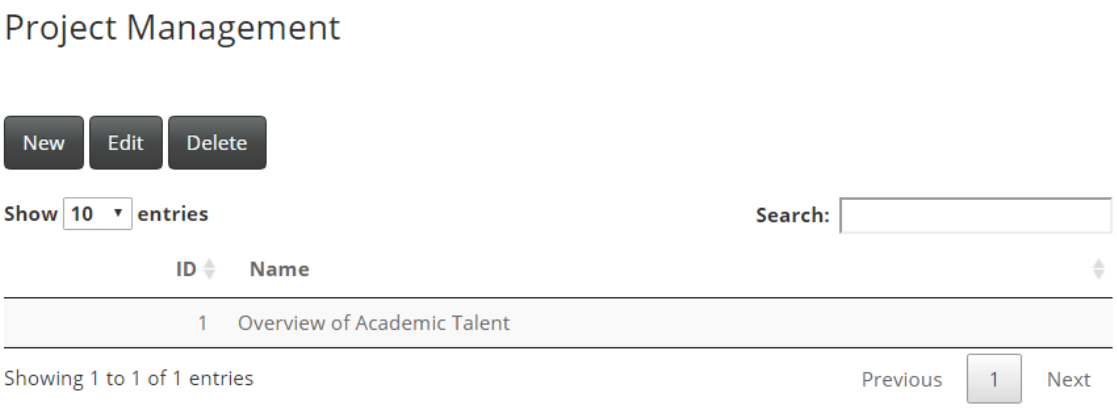


***Fig.7**: Table present on the page "Project".*

As it is possible to see from **Fig.7**, the first thing which appears is the title of the page, which is "*Project Management*". Then, there are three buttons that give to the user the possibility to perform three different actions. In particular, they can add a new observation, edit an already existing observation or delete one. When the "*New*" button is clicked, the window, displayed in the figure below, appears:

***Fig.8**: Window which appears when the "New" button is clicked.*

As it is possible to notice from **Fig.8**, when the "*New*" button is clicked, a window that contains a space for writing the name of a new project appears. When the user has written the name of the new project he wants to add, he has two possibilities: add the project,

clicking the button "*Save*", or go out from the window without any change, clicking the button "*Cancel*". In the first case, the application automatically computes the ID of the new project as the maximum numeric value of the ID of all the projects that are in the "*tbl_project*" plus 1. Doing so, the user does not have to specify any ID and the risk that he adds an already existing ID (which would generate a warning from the application) is completely avoided.

The "*New*" button works as soon as the user click it. For the "*Edit*" and "*Delete*" instead, the user has first to select the row he wants to modify or edit and then click on the action button. When a row is selected, it is highlighted in light blue, as it is shown in the figure below:

***Fig.9**: Appearance of a selected row.*

When a user has selected the row, he can select the action button. If the "*Edit*" button is selected, the window which appears is the following:

***Fig.10**: Window displayed when the "Edit" button is clicked.*

The user can cancel the old name of the project and replace it with a new one. If he wants to save the change he clicks "*Save*". Then, the new name will appear on the table. If he wants to go out from the window without any change, he has to click "*Cancel*".

If, after the row selection, the "*Delete*" button is clicked, the following window appears:

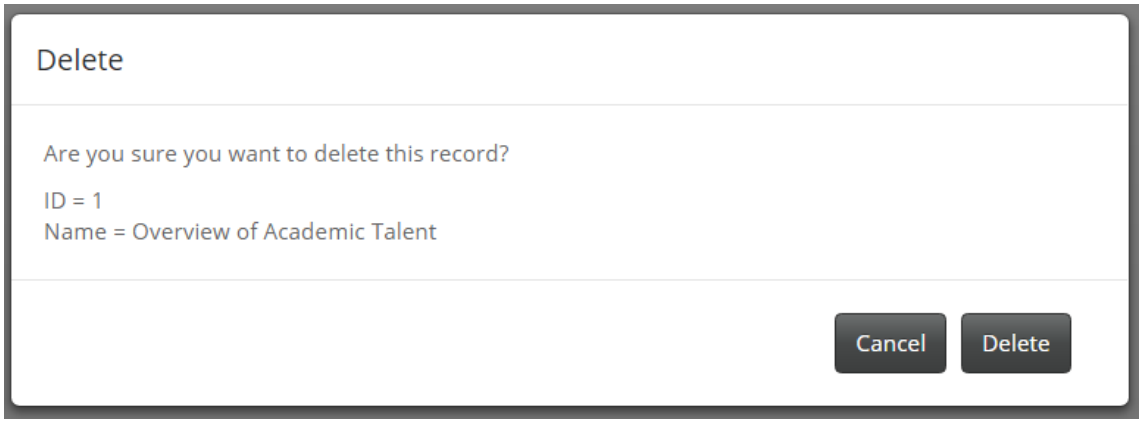


***Fig.11**: Window which is opened when the "Delete" button is clicked.*

In the window, a message asking the user if he is sure he wants to delete the selected row appears. If he clicks "*Delete*", the observation is deleted and no more available in the table. If the user clicks "*Cancel*", the action is cancelled.

All these actions directly affect the data stored in "*tbl_project*" thanks to the connection established among R and MySQL.

Then, the user can change the number of entries displayed in the table on one page, by modifying the number in the small window placed to the left-top of the table. By default, 10 entries are shown in the table per page. The user can filter the value stored in the table by writing a portion of text or a number in the space reserved in the top-right of the table, highlighted by the word "*search*". If the table has more than 10 observations, it is expanded in more pages. The user can navigate from one page to another using the buttons "*Previous*" and "*Next*" placed to the bottom-right of the table. Clicking on the small arrows which are present next to the name of each column, the user can sort by ascending or descending order the observations. These features are common to all the tables that are in the application.

7.1.2 SET-UP

The second page which belongs to the "*Set-up of the project*" is "*Set-up*". It groups three different pages, which are "*Affiliation*", "*Author*" and "*Journal*". They are explained one by one in this chapter.

AFFILIATION

The page "*Affiliation*" displays the MySQL table "*tbl_affiliation*". The figure below shows how the page looks.

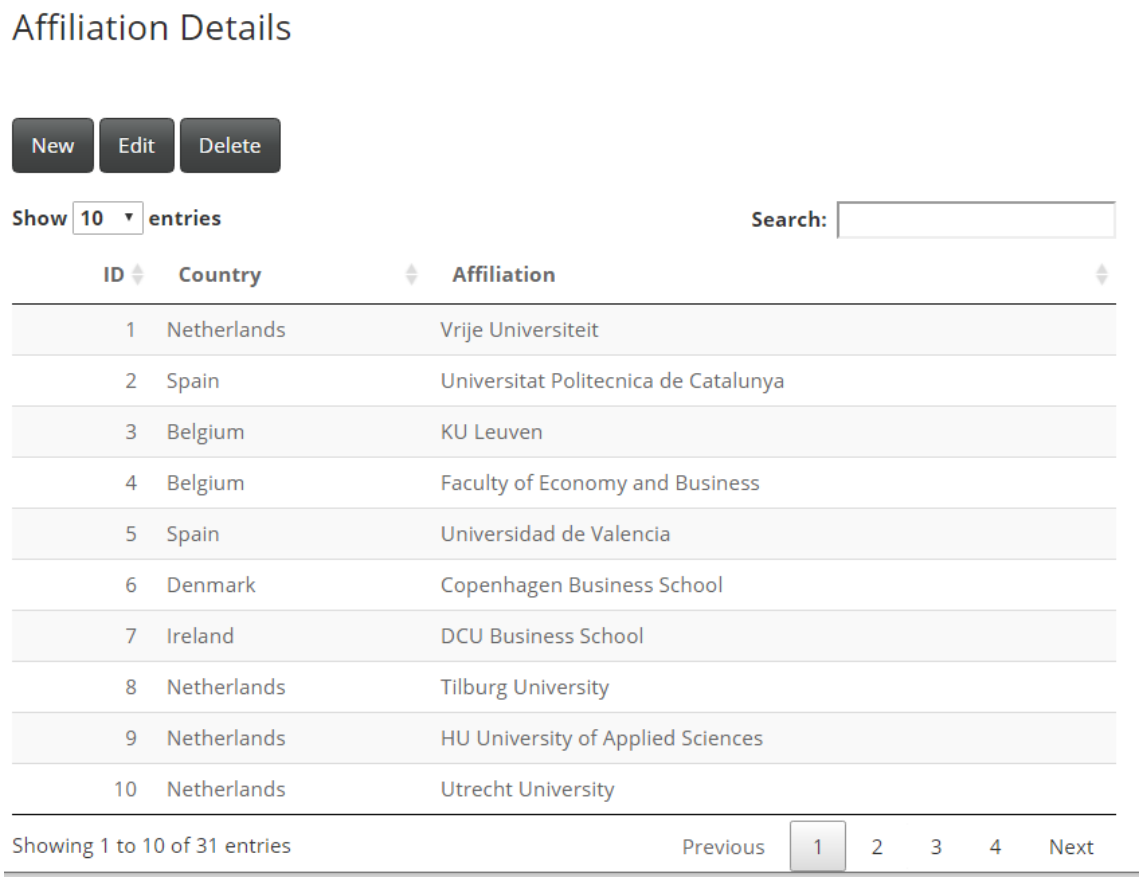


***Fig.12**: "Affiliation" page.*

The title of the page is "*Affiliation Details*". The columns shown are the variables already described in the chapter "*6.1 Creation of the database and tables*" for the table "*tbl_affiliation*". The "*New*", "*Edit*" and "*Delete*" buttons work as previously described for the "*Project*" page. When the user adds a new observation, he has to fill all the spaces that the window provides. In the case the user does not fill all, the application adds the

observation with the values specified and inserts "NA" in the columns not filled by the user. The user can select, edit the row, remove "NA" and add the values he needs. The same mechanism of the automatic computation of the ID is still valid, and it is applied to all the tables of the platform when a new observation is added. The user can never edit the column of the ID. The unique different thing compared to the table of the previous page is that now, below the table and between the option "*Previous*" and "*Next*", 4 numbers are displayed. These are created automatically by the application because there are 31 entries and, by default, a maximum of 10 entries are shown per page. Those numbers allow the user to jump from one page to another of the table. If the user changes the number of entries shown per page and set, for instance, 50, the numbers disappear because all the entries (31) are shown in one page, as the table in the page "*Project*".

AUTHOR

The second page of "*Set-up*" is "*Author*", which is the representation of the MySQL table "*tbl_author*". The figure below shows how the page appears:

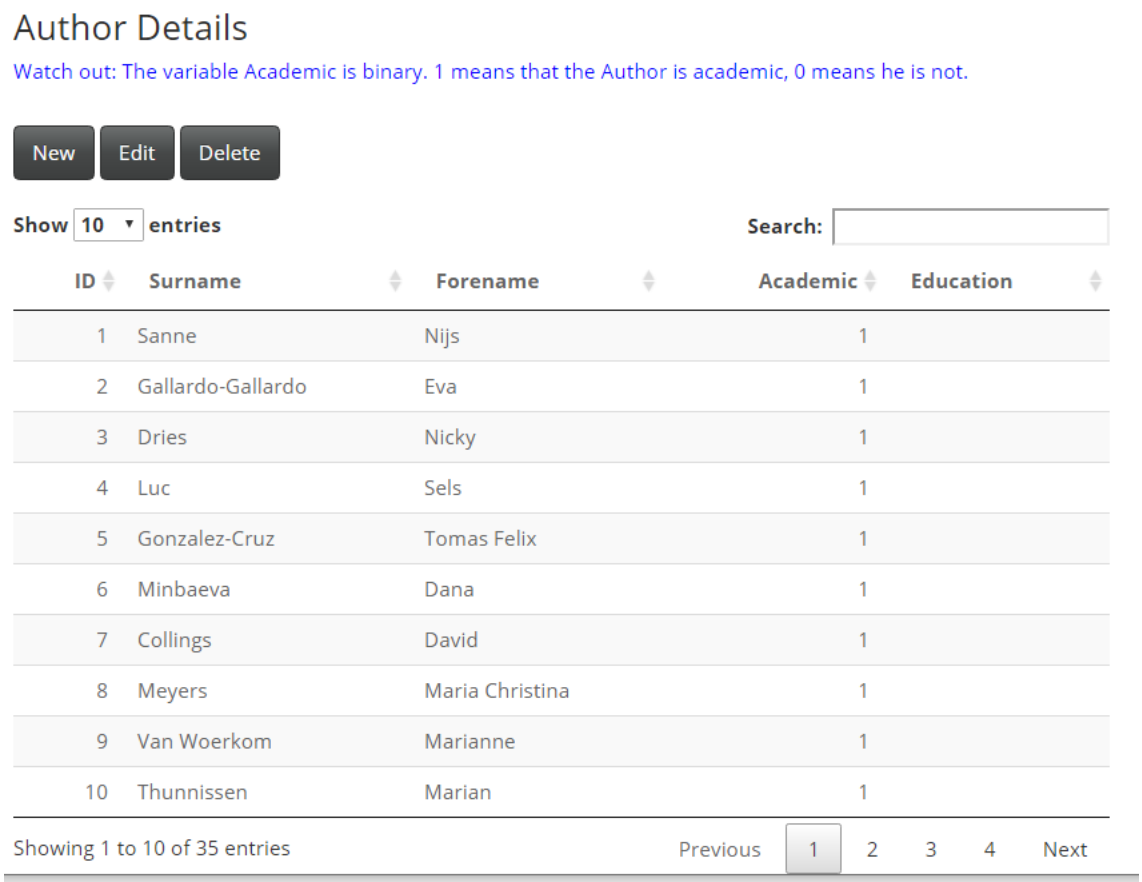


***Fig.13**: Appearance of the page "Author".*

The first difference which is possible to notice is that there is a warning below the title of the page "*Author Details*". It reminds the user that the variable "*Academic*" is binary and it explains what means when the value is 1 and when the value is 0. This warning is present in any table which contains a binary variable. It has been written with a different colour so it can catch directly the attention of the user. The second peculiarity of the table on this page is that the variable "*Education*" contains no text. This variable contains an additional description of the author that the user can add. If nothing is written, the application substitutes "NA" and it appears as shown in **Fig.13**. The user can modify and add the text at each row whenever he wants.

JOURNAL

The last page of the group "*Set-up*" is "*Journal*", which reflects the MySQL table "*tbl_journal*". The page is displayed by the figure below:

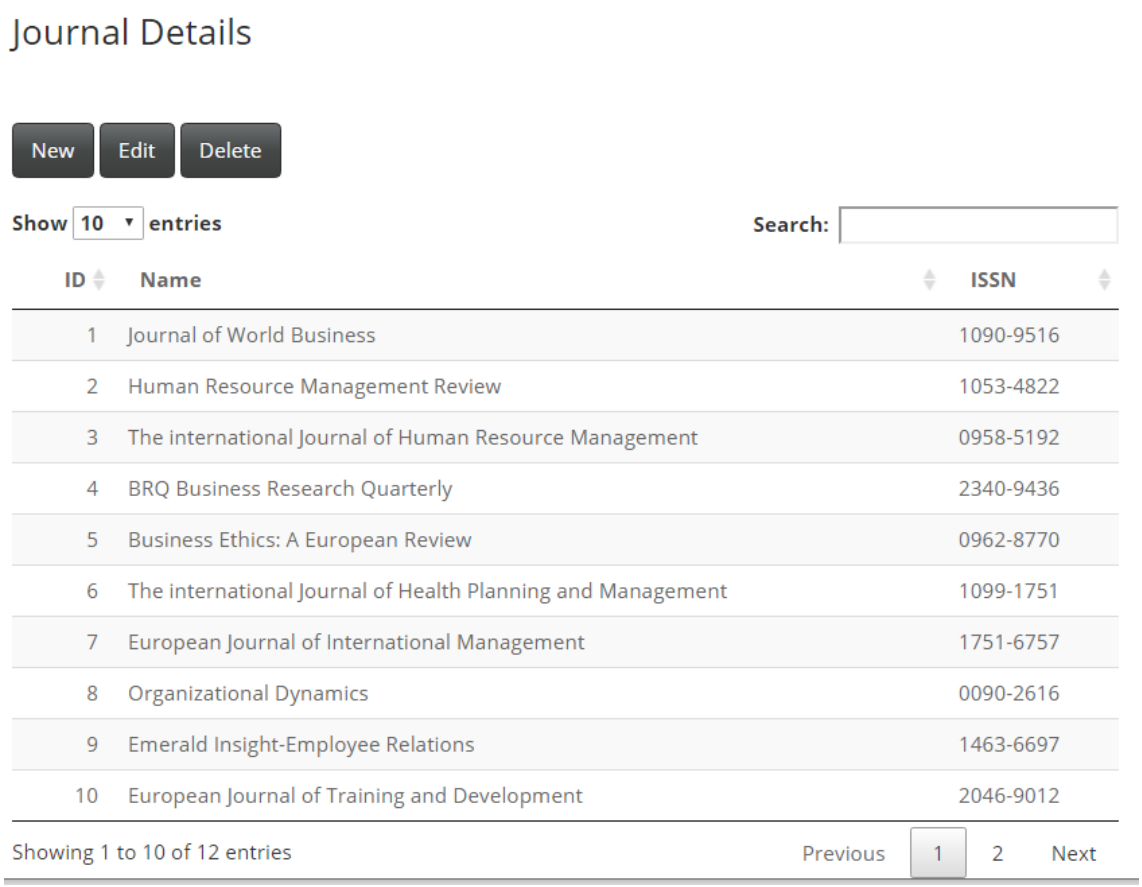


***Fig.14**: Page "Journal".*

All the features of this page have been already explained in the previous ones. There are the classic three action buttons ("*New*", "*Edit*" and "*Delete*"), the possibility to filter ("*Search*"), the possibility to order observations (the double arrows next to the name of each column) and the possibility to show more entries in the same page.

7.2 MANAGING

This group contains the page "*List of Papers*" and the group of pages "*Managing*".

7.2.1 LIST OF PAPERS

As far as the page "*List of Papers*" is concerned, it contains three different tables which are respectively the representation of the MySQL tables "*tbl_paper*", "*tbl_paper_author*" and "*tbl_paper_keyword*". The overall page looks as displayed in the figure below:

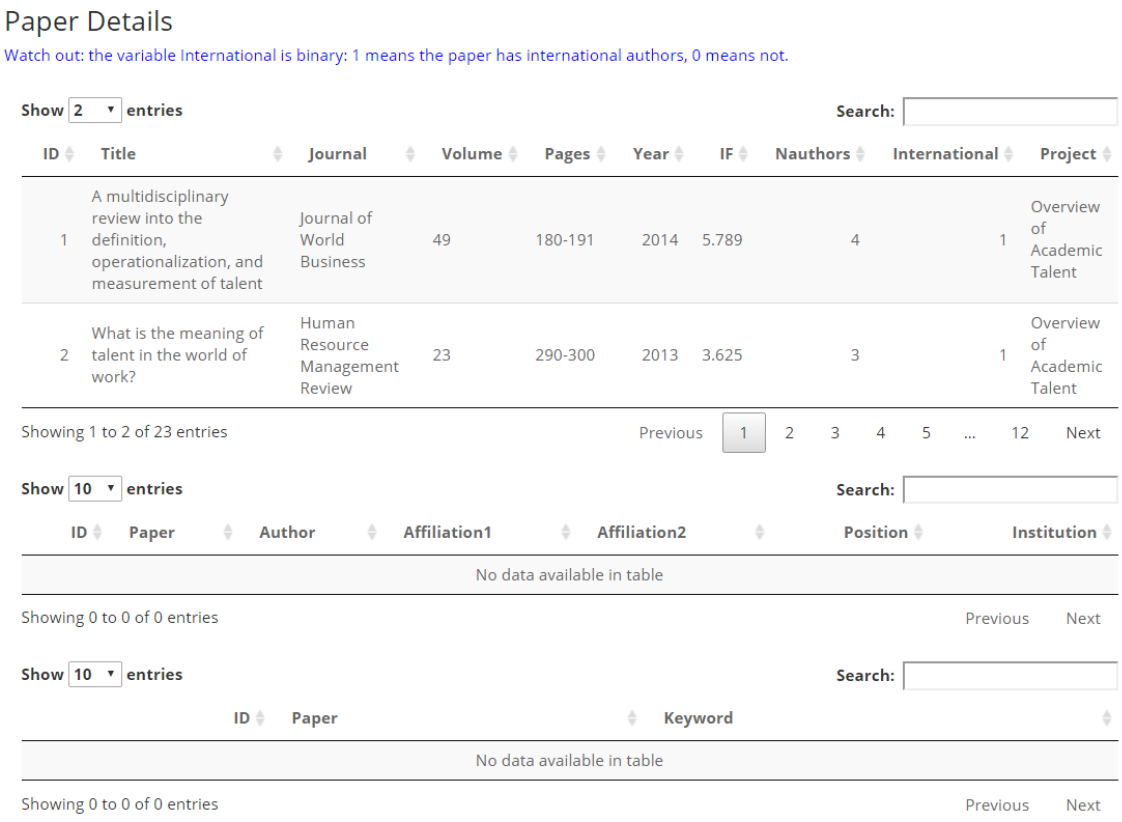


***Fig.15**: The three tables displayed on the page "List of Papers". For simplicity and for a better fitting of the figure to the report, only the first two records (papers) are displayed in the table in the figure. By default, the application always shows 10 entries for each page of the table.*

The first thing which is possible to notice is that there are not the three classic buttons for adding, editing and deleting an observation. If the user wants to perform one of these three actions on the paper, he has to go on the page "*Managing Papers*", which is part of the group "*Managing*" and which will be explained later. The "*List of Papers*" page has been developed as a summary page that contains all the information regarding papers, related keywords and authors that are stored in the database. Then, also in this case, is printed a warning which informs the user about the binary nature of the variable "*International*", explaining the meaning of 0 and 1.

It is possible to notice that among the three tables, two do not show any data initially. This is because both the second and the third tables react to user actions. In particular, when a user selects one paper (one row of the first table), in the second and third table are respectively displayed the information of the authors and the keywords related to the paper selected. An example is provided in the figure below:

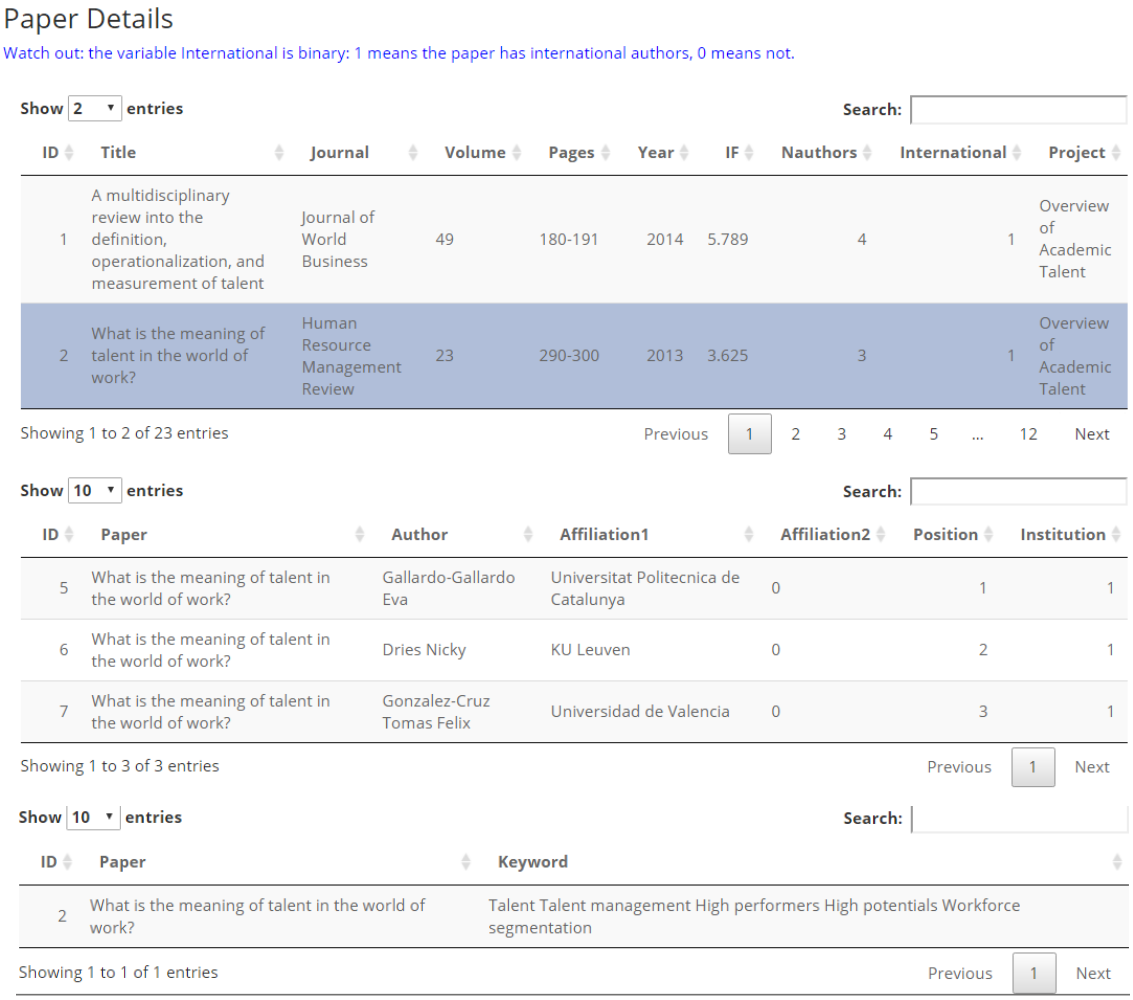


***Fig.16**: Reaction of the three tables when the user selects one paper.*

As it is possible to see from **Fig.16**, when the user selects a row, the second and third tables are filled with the information of the paper. In particular, the second table represents the MySQL table "*tbl_paper_author*" and displays the information of all the authors that have written the selected paper. There are two columns for the affiliation because one author may have two affiliations. If the author has only one affiliation, the second column of affiliation (*Affiliation2*) is filled with a 0 (this is exactly what is represented in the example provided in **Fig.16**). The third table corresponds to the MySQL table "*tbl_paper_keyword*" and displays the information regarding the keywords of the paper selected. The user can select more than one paper and the information of all the papers selected is shown in the other tables. This feature gives to the user the possibility to compare in glance keywords and authors of different papers. For these three

tables, not all the columns of the respective MySQL tables have been shown. In particular, all the columns regarding the IDs (apart from the ID which works as primary key) are not shown because it is believed that it is not an important information for the user of the platform. Moreover, for the first table, the column which stores the abstract of the paper is hidden, because it would occupy too much space. It is shown on the page "*Managing Paper*", where the user can also modify it.

7.2.2 MANAGING

The group of pages "*Managing*" contains the page "*Managing Papers*", "*Managing Authors*", "*Managing Keywords*" and "*Managing Contents*". The overall purpose of these pages is to provide to the user the possibility to modify respectively the information regarding papers, authors, keywords and contents.

MANAGING PAPERS

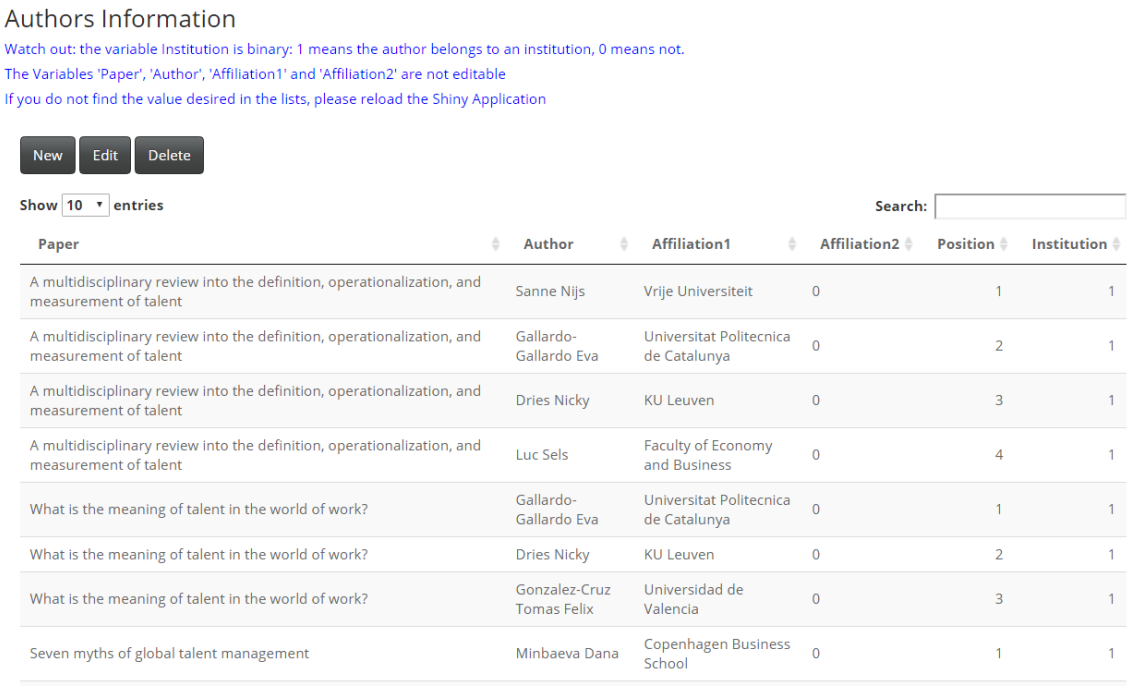
The page "*Managing Papers*" is really similar to the page "*List of Papers*". The main differences are that, in "*Managing Papers*" only one table is shown (which is the first table displayed in "*List of Papers*") and that there is the possibility to add, edit or delete one observation. The aim of this page is to give the user the possibility to add, edit or delete records that regard papers. An overview of the page is provided by the figure below:

Fig.17: *Appearance of the page "Managing Papers". For simplicity and for a better fitting of the image to the report, the table is captured with only one observation.*

As it is possible to notice from **Fig.17**, initially there are four warnings that help the user to correctly utilize the application. Then, there are the three classic buttons "*New*", "*Edit*" and "*Delete*", which work as already described in the previous tables. Compared to the table displayed in "*List of Papers*", the table represented in **Fig.17** contains the same information with the unique addition of the column "*Abstract*", which shows the abstract of the paper. One feature of the application which is present in this page is that, when the user deletes one paper, after reloading the application, the connected observations of the same paper in the other pages of the group "*Managing*" ("*Managing Authors*", "*Managing Keywords*", "*Managing Contents*") are deleted as well. At the same time, when a new paper is added, the application automatically creates one observation with the title of the paper in the tables of the page "*Managing Keywords*" and "*Managing Contents*". Therefore, when the user goes in these pages to add the keywords and the contents of the new paper, he does not have to create a new paper and choose it from the list of papers available, but he only has to edit the already created observation, adding the desired information. In order to see the new observation in the other table, the user must reload the Shiny Application with the appropriate button placed on the top of the application. Both these two features are time-saving for the user.

MANAGING AUTHORS

The second page of this group is "*Managing Authors*". On this page, the user can modify all the information related to the authors of the papers. Any change made on the author's information will be reflected in the second table of the page "*List of Papers*" The figure below shows the page:


Authors Information

Watch out: the variable Institution is binary: 1 means the author belongs to an institution, 0 means not.
The Variables 'Paper', 'Author', 'Affiliation1' and 'Affiliation2' are not editable
If you do not find the value desired in the lists, please reload the Shiny Application

New Edit Delete

Show 10 entries Search:

Paper	Author	Affiliation1	Affiliation2	Position	Institution
A multidisciplinary review into the definition, operationalization, and measurement of talent	Sanne Nijs	Vrije Universiteit	0	1	1
A multidisciplinary review into the definition, operationalization, and measurement of talent	Gallardo-Gallardo Eva	Universitat Politecnica de Catalunya	0	2	1
A multidisciplinary review into the definition, operationalization, and measurement of talent	Dries Nicky	KU Leuven	0	3	1
A multidisciplinary review into the definition, operationalization, and measurement of talent	Luc Sels	Faculty of Economy and Business	0	4	1
What is the meaning of talent in the world of work?	Gallardo-Gallardo Eva	Universitat Politecnica de Catalunya	0	1	1
What is the meaning of talent in the world of work?	Dries Nicky	KU Leuven	0	2	1
What is the meaning of talent in the world of work?	Gonzalez-Cruz Tomas Felix	Universidad de Valencia	0	3	1
Seven myths of global talent management	Minbaeva Dana	Copenhagen Business School	0	1	1

***Fig.18**: Appearance of the page "Managing Authors". For a better fitting and comprehension of the figure, not the entire page is reported in the figure.*

The table of **Fig.18** is the representation of the MySQL table "*tbl_paper_author*". Not all the columns of those tables are represented. Indeed, the columns containing the ID of the observation, the ID of the paper, the ID of the author and the ID of the two affiliations are not displayed, because they are not useful for the user. It is possible to notice that also this table has three buttons for the classic three operations that the user can do. Moreover, a warning clarifying to the user how the application works and the meaning of a binary variable is printed.

MANAGING KEYWORDS

The third page of the group "*Managing*" is "*Managing Keywords*". The table placed on this page reflects the MySQL table "*tbl_paper_keyword*". In this page, the user can add, edit or delete the keywords of a paper. Any operation in the table of this page will be reflected in the third table of the page "*List of Papers*". An overview of the page is provided by the figure below:

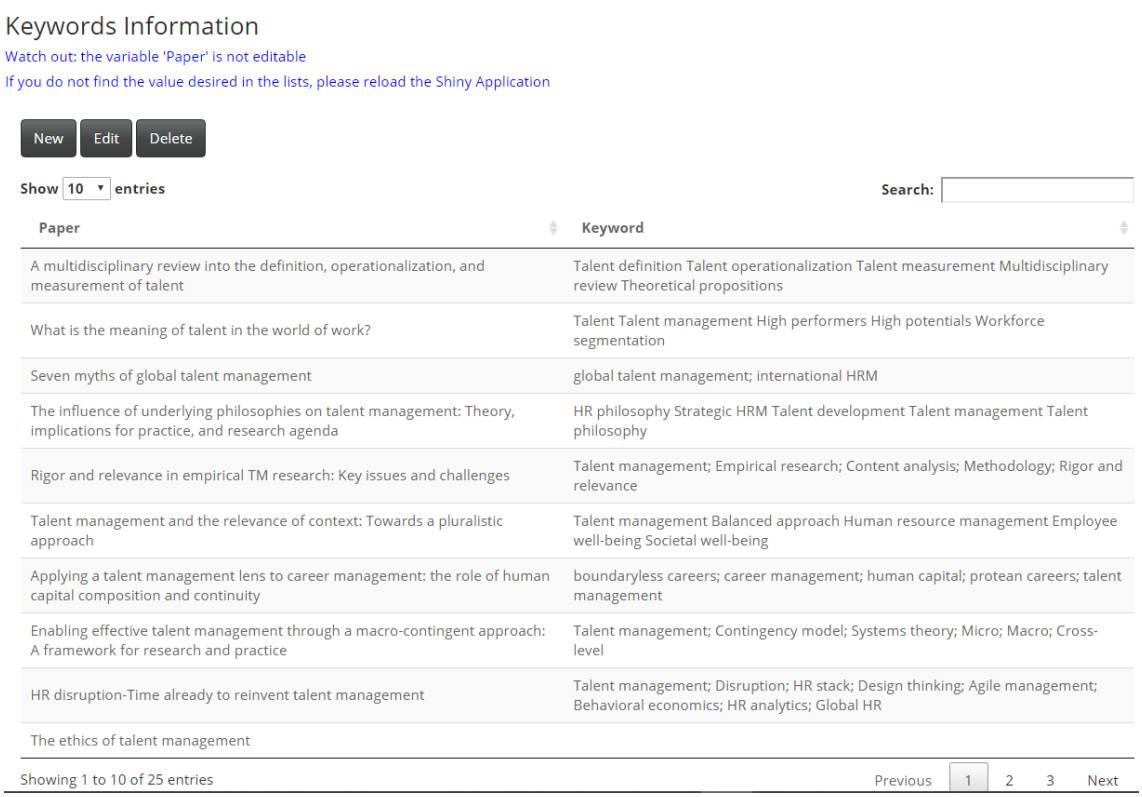
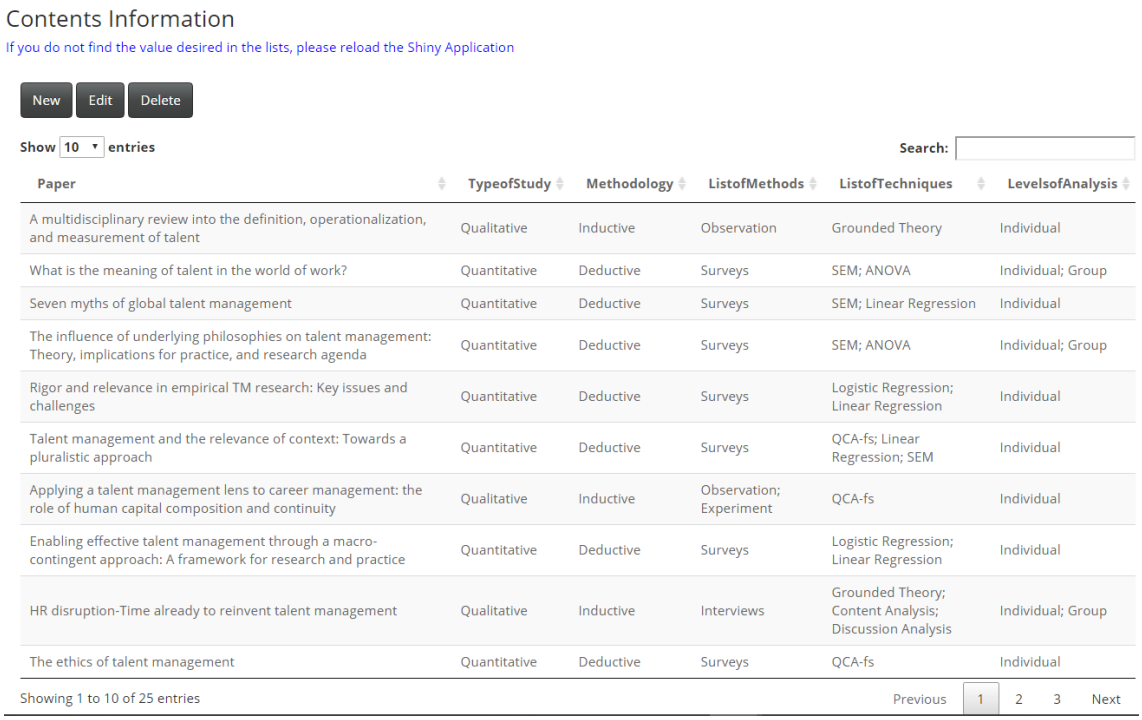


***Fig.19**: Appearance of the page "Managing Keywords".*

As it is shown in **Fig.19**, the structure is the same as all the other pages of this group previously described. Also in this case, the information about the IDs are hidden. The user can write the keywords separating the characters with " ; " or with space. In terms of the analysis that will be performed in the next group of pages with the word cloud, this has no difference. Then, if a paper has no keyword, the user can leave the observation of the paper and avoid filling the column "*Keyword*". Also this has no impact on the word cloud analysis.

MANAGING CONTENTS

The last page of this group is "*Managing Contents*". It contains the table which corresponds to the MySQL table "*tbl_content*". The structure of this page is the same as the previous ones of the group "*Managing*" and it is displayed in the figure below:



Contents Information

If you do not find the value desired in the lists, please reload the Shiny Application

New Edit Delete

Show 10 entries Search:

Paper	TypeofStudy	Methodology	ListofMethods	ListofTechniques	LevelsofAnalysis
A multidisciplinary review into the definition, operationalization, and measurement of talent	Qualitative	Inductive	Observation	Grounded Theory	Individual
What is the meaning of talent in the world of work?	Quantitative	Deductive	Surveys	SEM; ANOVA	Individual; Group
Seven myths of global talent management	Quantitative	Deductive	Surveys	SEM; Linear Regression	Individual
The influence of underlying philosophies on talent management: Theory, implications for practice, and research agenda	Quantitative	Deductive	Surveys	SEM; ANOVA	Individual; Group
Rigor and relevance in empirical TM research: Key issues and challenges	Quantitative	Deductive	Surveys	Logistic Regression; Linear Regression	Individual
Talent management and the relevance of context: Towards a pluralistic approach	Quantitative	Deductive	Surveys	QCA-fs; Linear Regression; SEM	Individual
Applying a talent management lens to career management: the role of human capital composition and continuity	Qualitative	Inductive	Observation; Experiment	QCA-fs	Individual
Enabling effective talent management through a macro-contingent approach: A framework for research and practice	Quantitative	Deductive	Surveys	Logistic Regression; Linear Regression	Individual
HR disruption-Time already to reinvent talent management	Qualitative	Inductive	Interviews	Grounded Theory; Content Analysis; Discussion Analysis	Individual; Group
The ethics of talent management	Quantitative	Deductive	Surveys	QCA-fs	Individual

Showing 1 to 10 of 25 entries Previous 1 2 3 Next

***Fig.20**: Structure of the page "Managing Contents".*

Also this page is characterized by the same structure of all the previous ones. There are only two things which are important to highlight: first at all, "*tbl_content*" is made of 45 columns. Obviously, it is impossible to represent all of them in the table. Therefore, the variables shown are the first six (excluding, as in the previous cases, the IDs). However, when the user adds or edits an observation, he has the possibility to fill all the variables, still excluding the IDs, which are computed automatically (as in the tables of the other pages). The second important observation is about the way of adding the data for all those variables which admit more than one character (they are highlighted by the wording "*list*"). As it is shown in **Fig.20**, the characters must be separated by a " ; " followed by a space. If the user separates the observation only with " ; " forgetting the space, this will affect the content analysis because the separator set in the R code for the analysis is a semicolon followed by a space.

Before moving to the next chapter, a final consideration has to be done. The application is developed with the idea that the user adds one overall observation at a time. Therefore, the user has to follow this path: first, check if the project exists. If not, he can create a new one. Then, he checks if the affiliation, journal and authors are already in the database.

Otherwise, he can create new records. Then, he moves to "*Managing Papers*" and he can add information about the paper. After this, he moves to the "*Managing Keywords*" and "*Managing Contents*" pages to fill the already created row with the values desired. If the user follows this path, there will not be any problem with the application.

7.3 ANALYSIS

Two different types of analysis are performed by the application: bibliometric analysis and content analysis. They are explained in the next subchapters.

7.3.1 BIBLIOMETRIC ANALYSIS

The bibliometric analysis group contains five pages, the first one realizes a word cloud while all the remaining four provide a scatterplot. Each page is well analysed in this chapter.

WORD CLOUD

A word cloud is a visual representation of keywords or tags used in a website or in a generic text file. The single words are represented in different font sizes: the bigger is the word represented the higher is its importance/frequency. The word cloud is a very powerful tool that gives to the user the possibility to perceive immediately the most important terms. The application realized a word cloud for the keywords of all the papers that are published during a certain period, selected by the user. The figure below shows how the page appears to the user:

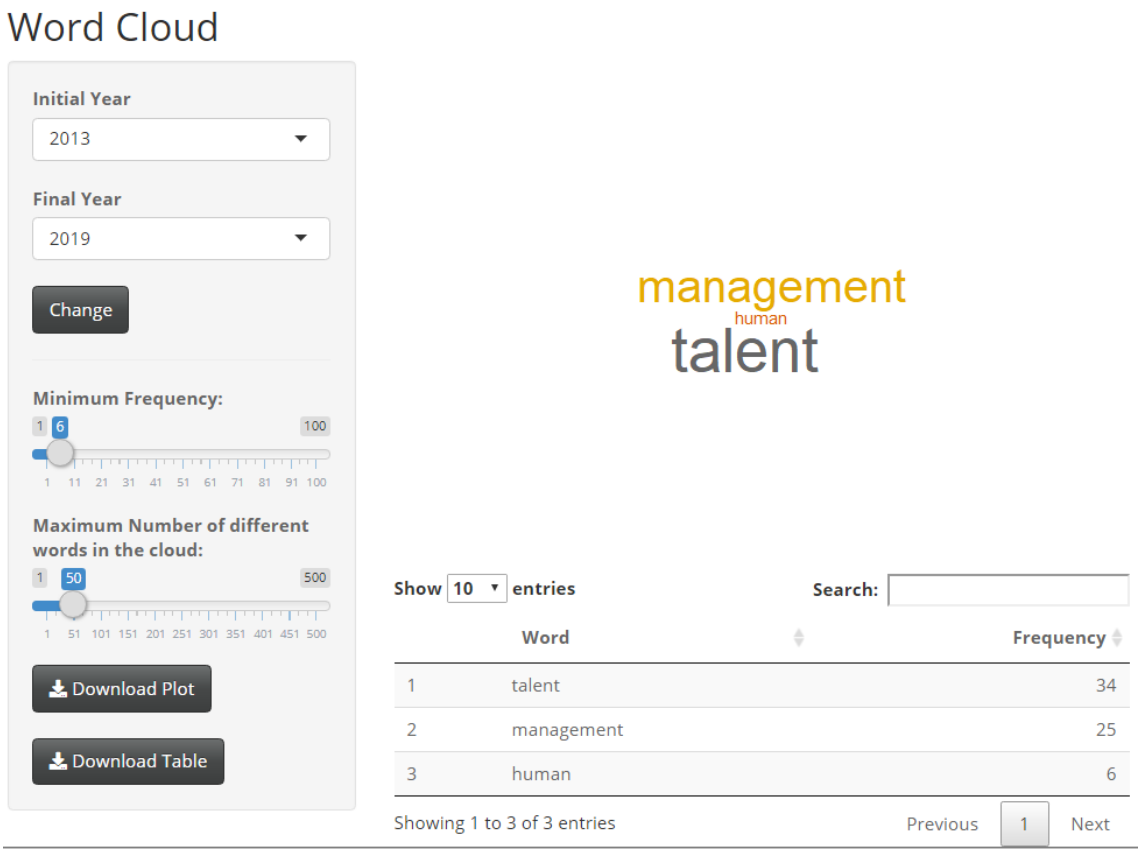


***Fig.21**: "Wordcloud" page.*

The goal of the analysis is to provide the user an immediate perception of the most frequent words in the papers published during a certain period. As it is possible to see from **Fig.21**, at the left-hand side of the page the user finds the "*sidebarPanel*", which contains all the parameters that the user can set. In particular, first, the user sets the initial and the final year of the period he wants to analyse. Then, he can select the minimum frequency that a word must have to appear in the word cloud. Next, the user can choose the maximum number of different words that will be part of the word cloud. When the user clicks the action button "*Change*" the graph is plotted. Below the graph, a table that shows the words that are part of the word cloud (with their frequencies) is provided. The table has the same style and the same features of all the other tables already described for the previous groups of pages, apart from the three action buttons "*New*", "*Edit*" and "*Delete*". All the tables that are part of both the analysis share the same style and features. The user can change the minimum frequency by moving the value of the first slider. He can change the maximum number of words plotted in the cloud by moving the value of

the second slider. If the user selects a minimum frequency higher than the values computed, a warning message appears above the slider of the minimum frequency, stating: "*Please select a minimum frequency lower than the maximum value in the table*". The word cloud and the table immediately change when the user moves one value of the two sliders. If the user wants to change the period of analysis, after selecting the new years he has to click on the button "*Change*" to update the graph and the table. Finally, the user has the possibility to download the word cloud realized, by clicking on the action button "*Download Plot*" and to download the table containing the words and frequencies, in a CSV format, by clicking on the action button "*Download Table*". From the example in **Fig.21**, it is possible to see that the plot and the table are part of the "*mainPanel*" and placed to the right-hand side. Since the word "*talent*" is the biggest, it has the highest importance (highest frequency). This can also be read from the table below the word cloud.

The structure of all the different tools for both the analysis is always the same. The user will find on the left-hand side of the page all the instruments for setting the parameters of the analysis and for downloading the outputs. On the right-hand side of the page, he will find the outputs of the analysis (graphs and tables). The constant structure of the tools for the analysis allows the user an easier and quicker understanding of how the application works.

SCATTER PLOT PAPER

A scatter plot is defined as a graph where the values of typically two variables are plotted along two axes, using Cartesian coordinates. The values are displayed with points. A regression line or a simple line which links all the points can be added. The scatter plot is used to highlight the possible trend and relationship among data. In all the scatter plots realized by the platform, the variable on the x-axis is the period of analysis. The value of the y-axis depends on the variable analysed. All the scatter plots provided by the application have the same style: blue points connected by a thinner blue line.

The goal of the analysis through the scatter plot of papers is to understand how the number of papers published changes over time. The figure below shows the aspect of the page:

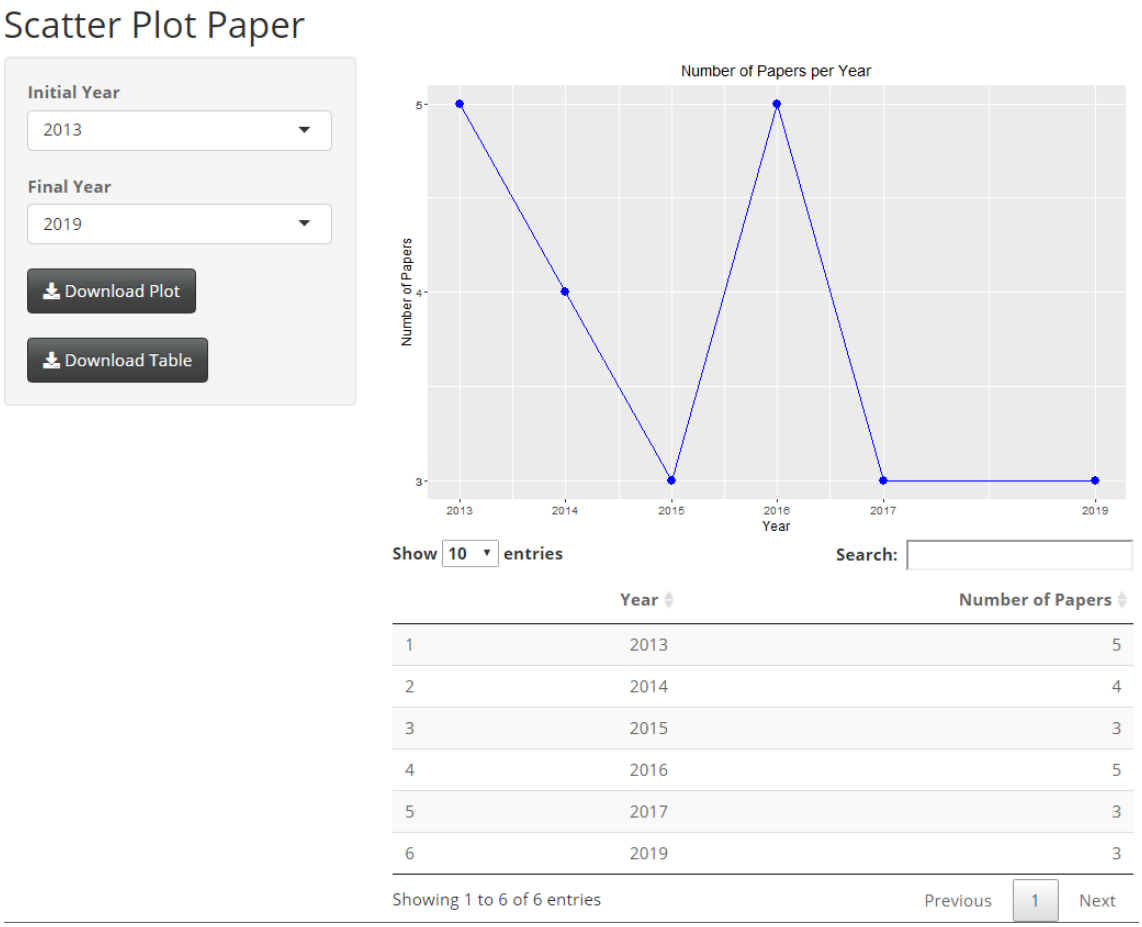


***Fig.22**: Appearance of the page "Scatter Plot Paper".*

As it is possible to notice from **Fig.22**, the page has the same layout of the "*Word Cloud*" previously described: outputs of the analysis on the right-hand side, parameters that have to be set and download buttons on the left-hand side. When the user selects the period analysed, using the "*Initial Year*" and "*Final Year*" select boxes, two outputs are created: the first one is the scatter plot where, as the title suggests, is represented the trend of the number of papers published during the period analysed. The x-axis shows the years of the analysis while the y-axis shows the number of papers published each year. The second output is the table placed below the scatter plot, which shows the same information about the graph: for each year is reported the number of different papers published. One row corresponds to one year. When the user changes the period of analysis, both the plot and the table are immediately updated. Finally, there are two buttons which provide to the user the possibility to download the scatter plot ("*Download Plot*") and to download the table in CSV format ("*Download Table*").

SCATTER PLOT AUTHOR

On this page is analysed the number of different authors which publish a paper during a period specified by the author. From now on, for the remaining three pages of the group "*Bibliometric Analysis*", the figure which shows the structure of the page is included only in the Annex (**Fig.39**) because it is completely the same as the page "*Scatter Plot Paper*".

On the left-hand side, the user selects the initial and the final year of the analysis. On the right-hand side, the classic two outputs are displayed. First, the scatter plot, with the title "*Number of Different Authors per Year*" is presented. On the x-axis of the graph, there are the years of the time period, while on the y-axis is displayed the number of different authors per year. The second output is the table placed below the graph, which shows different information: it gives to the user an indication about the number of different papers published by each author which belongs to the analysis done. By default, authors in the table are sorted by alphabetic order. Finally, the user has the possibility to download both the scatter plot and the table with the two buttons placed on the left-hand side of the page.

SCATTER PLOT JOURNAL

The goal of this analysis is to understand how the overall number of different journals where authors publish change over time. On the left-hand side, the user sets the beginning and the end of the analysis. On the right-hand side of the page, a scatter plot and a table are automatically printed. The scatter plot shows the trend of the number of different journals. On the x-axis, there are the years of the period of reference, while on the y-axis are represented the number of journals. The table below the graph shows the number of different papers that are published in each journal which falls into the period of analysis. The table is sorted by alphabetical order of journals. The user has the possibility to download the plot and the table, using, respectively, the "*Download Plot*" and "*Download Table*" buttons placed on the left-hand side of the page. The figure which provides an example of this page is put in the Annex (**Fig.40**).

SCATTER PLOT IMPACT FACTOR

The goal of this analysis is to provide the user insight into the change of the trend of the average impact factor during a specific period. On the left-hand side, the user selects the

period he desires, with the classic select boxes. On the right-hand side, the scatter plot and the table are automatically printed. The x-axis of the scatter plot contains the years of the analysis. The y-axis shows the average of the impact factor of all the journals in a specific year, rounded at 2 significant digits. The table below the graph presents all the journals, which fall in the period analysed, and their impact factor. It is possible to notice that one journal may appear more than one time. This is correct, because one journal may have improved or worsened its impact factor during the period analysed, since the impact factor is a yearly measure. The observations in the table are sorted in descending order by the value of the Impact Factor (IF) of the journal. The user has, as always, the possibility to download both the plot and the table using the buttons placed on the left-hand side of the page. A figure showing an example of this page is present in the Annex (**Fig.41**).

7.3.2 CONTENT ANALYSIS

The analysis of the content of papers is divided into five pages: "*Methodology*", "*Data I*", "*Data II*", "*Research Definition*" and "*Talent*". Each page contains some graphs and tables which are all built from the variables of the MySQL table "*tbl_content*". Each page is well analysed in this chapter. The structure of the style of the graphs and the layout of the page is the same in all the content analysis and consistent with the "*Bibliometric Analysis*".

METHODOLOGY

This page focuses on the analysis of the methodologies used. It contains five different analysis:

Type of study

The goal of this analysis is to represent the different types of study of the papers published in a certain period, exploiting a bar plot. A bar plot is defined as a graph which represents categorical data with rectangular bars of different heights, depending on the value that they represent. An example of this analysis is provided by the image below:

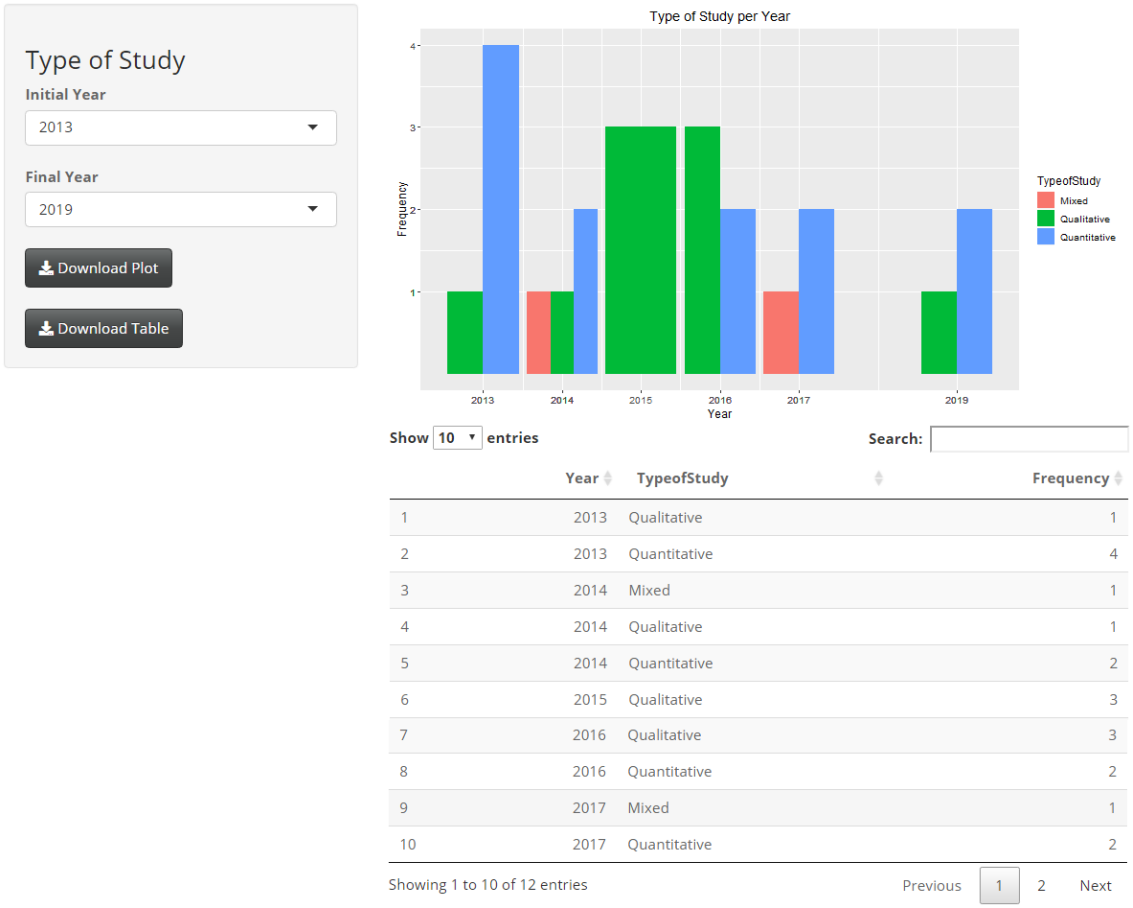


***Fig.23**: "Type of Study" analysis.*

As it is possible to notice from **Fig.23**, on the left-hand side the user has the classic panel where he can set the years, for determining the period of the analysis, and the two buttons, for downloading the desired output. On the right-hand side, the bar plot is placed. On the top, there is the title of the graph ("*Type of study per year*"). On the x-axis, there are the years which are part of the analysis, while on the y-axis is represented the number of times (frequency) a specific study is done. Then, on the right-hand side of the graph, a legend explaining the meaning of each bar is placed. An example of the interpretation of the graph is the following: considering, for instance, the year 2014, one paper has a type of study "*Mixed*", one paper has a type of study "*Qualitative*", while the remaining two papers published in 2014 have a type of study of "*Quantitative*". Below the graph, a table that summarizes the frequency of each type of study per year is placed. By default, it is sorted in ascending order by year. As in all the previous analysis, the first thing the user has to do is to define the time period with the two select boxes. Then, the graph and the

tables are automatically printed. As soon as the user changes one value of the years, the two outputs immediately react and update themselves.

Methodology

The goal of this analysis to show how many times a type of methodology is used per year, through a barplot. The methodology can be of two types: "*Deductive*" or "*Inductive*". The style of the analysis and the type of outputs are the same as the previous analysis. Therefore, only the description of the analysis is present here, while a figure showing an example of the analysis is placed in the Annex (**Fig.42**). The user has, on the left-hand side, the boxes for selecting the period and the buttons to download the outputs. On the right-hand side, the graph and the table are placed. The title of the graph is "*Methodology per year*". The x-axis and the y-axis are the same as the previous analysis (respectively "*Year*" and "*Frequency*"). A legend which helps the user to distinguish the two type of methodology is placed on the right-hand side of the graph. Below the graph, the classic table which summarizes the methodology per year, with its frequency, is provided. It is sorted by ascending order of years.

List of methods

The goal of this analysis is to provide the user insight into the most used methods of content analysis during a specific period. In this case, the output is only one table which is built from the column "*ListofMethods*". It contains the different methods applied to one paper, separated by a semicolon and a space. An example of this analysis is provided by the figure below:

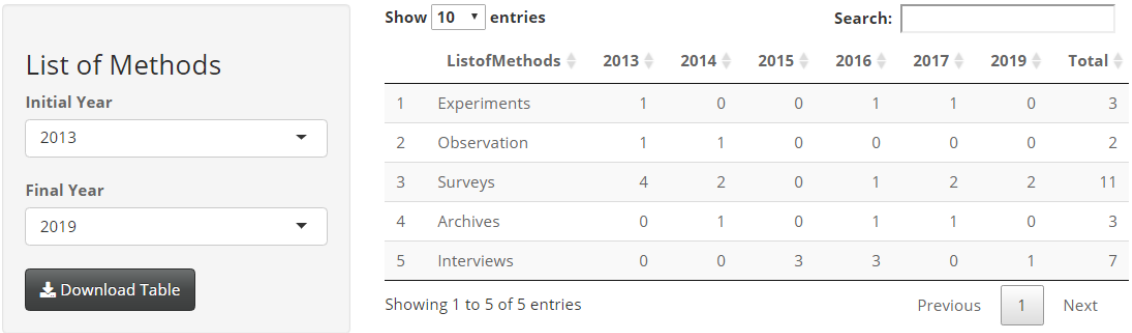


***Fig.24**: Example of the "List of Methods" analysis.*

As it is possible to notice from **Fig.24**, on the left-hand side the classic panel is placed. On the right-hand side, the unique output, the table, is printed. The table has, as the first column, the name of the methods used by the papers which fall into the period analysed. The other columns are the years of the analysis. The last column, "*Total*", gives to the user an indication about how many times the specific method has been used over the period considered. The rows represent the methods considered in the analysis. Only one download button is placed because only one output is realized by this analysis.

List of techniques

The goal of this analysis is to provide the user insight into the most used techniques for the analysis of the content of one paper, during a specific period of time. The style of this analysis is identical to the previous "*List of Methods*". First, the user selects the time period of the analysis. Then, a table is printed. Rows represent the type of techniques used. The first column enlists the name of the techniques, while the remaining columns are the years of the analysis, with the unique exception of the last column, which is used to compute the total number of times a technique has been used in a certain period. An example of this analysis is provided in the Annex (**Fig.43**).

Levels of Analysis

The goal of this analysis is to give to the user an indication about the most used levels of analysis of papers, during a specific period of time. The style of this analysis is identical to the previous "*List of Methods*" and "*List of Techniques*". First, the user selects the time window of the analysis. Then, a table is printed. Rows represent the levels of analysis used. The first column enlists the name of them, while the remaining columns represent the years of the analysis, with the unique exception of the last column, which is used to compute the total number of times a certain level of analysis appears in a certain period. An example of this analysis is provided in the Annex (**Fig.44**).

DATA I

This page provides four types of analysis which are now explained in detail.

Source of Data

The goal of this analysis is to provide the user insight into the most common source of data of one paper, during a specific period of time. The style of this analysis is the classic panel for inputs and a unique table as output. First, the user selects the time period of the analysis. Then, the table is printed. Rows represent the data source used for the papers analysed. The first column shows the name of the data source, while the remaining columns are the years of the analysis, with the unique exception of the last column, which is used to calculate the total number of times a data source appears in a certain period. Since this analysis is a standard one, an example of it is provided in the Annex (**Fig.45**).

Size of the Sample

The goal of this analysis is to provide the user a boxplot of the distribution of the different sizes of samples during the period analysed. The boxplot is built using five summary numbers:

1. The minimum value of data, excluding outliers (in this case, an observation is an outlier if it is lower than the first quartile minus 1.5 times the difference between the third quartile and the first quartile).
2. The first quartile, defined as that number which represents 25% of the ordered data.
3. The median, which is the middle value of the ordered dataset.
4. The third quartile, defined as that number which represents 75% of the ordered data.
5. The maximum value of data, excluding outliers (in this case, an observation is an outlier if it is higher than the third quartile plus 1.5 times the difference between the third and the first quartile).

An example of this type of analysis is provided by the figure below:

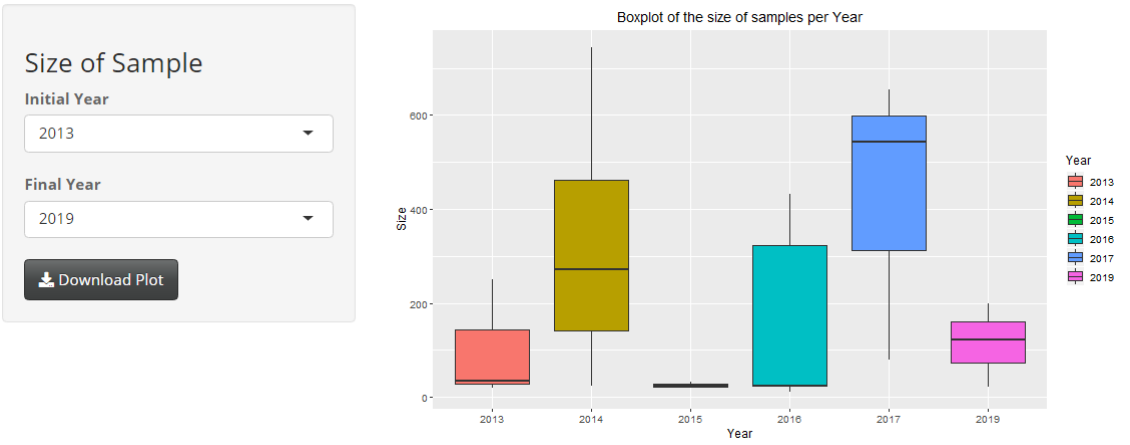


Fig.25: Representation of the boxplot analysis for the size of the sample.

As it is possible to notice from the **Fig.25**, the plot contains one boxplot for each year analysed, placed on the x-axis. Each boxplot has a different colour, in order to distinguish one year from another. The meaning of each colour is communicated through the legend placed on the right-hand side of the plot. When the user changes the years of the analysis, the boxplots are immediately updated. The numbers represent the size of the sample, as it is stated on the y-axis. The length of the lower tail is equal to the value computed for understating if there are outliers or not (the first quartile minus 1.5 times the difference of the third and first quartile). The same for the length of the higher tail (computed as the third quartile plus 1.5 times the difference between the third quartile and the first quartile). The lower line of the box is the first quartile, the thicker line in the middle of the box is the median and the higher line of the box is the third quartile. As always, the user has the possibility to download the plot with the appropriate button.

Country

The goal of this analysis is to provide the user insight into the country of the talent. In order to do this, three outputs are used: a classic bar plot, a classic table and one map which highlights with different colours, according to the frequency, the countries of the talent. An example is provided by the figure below:

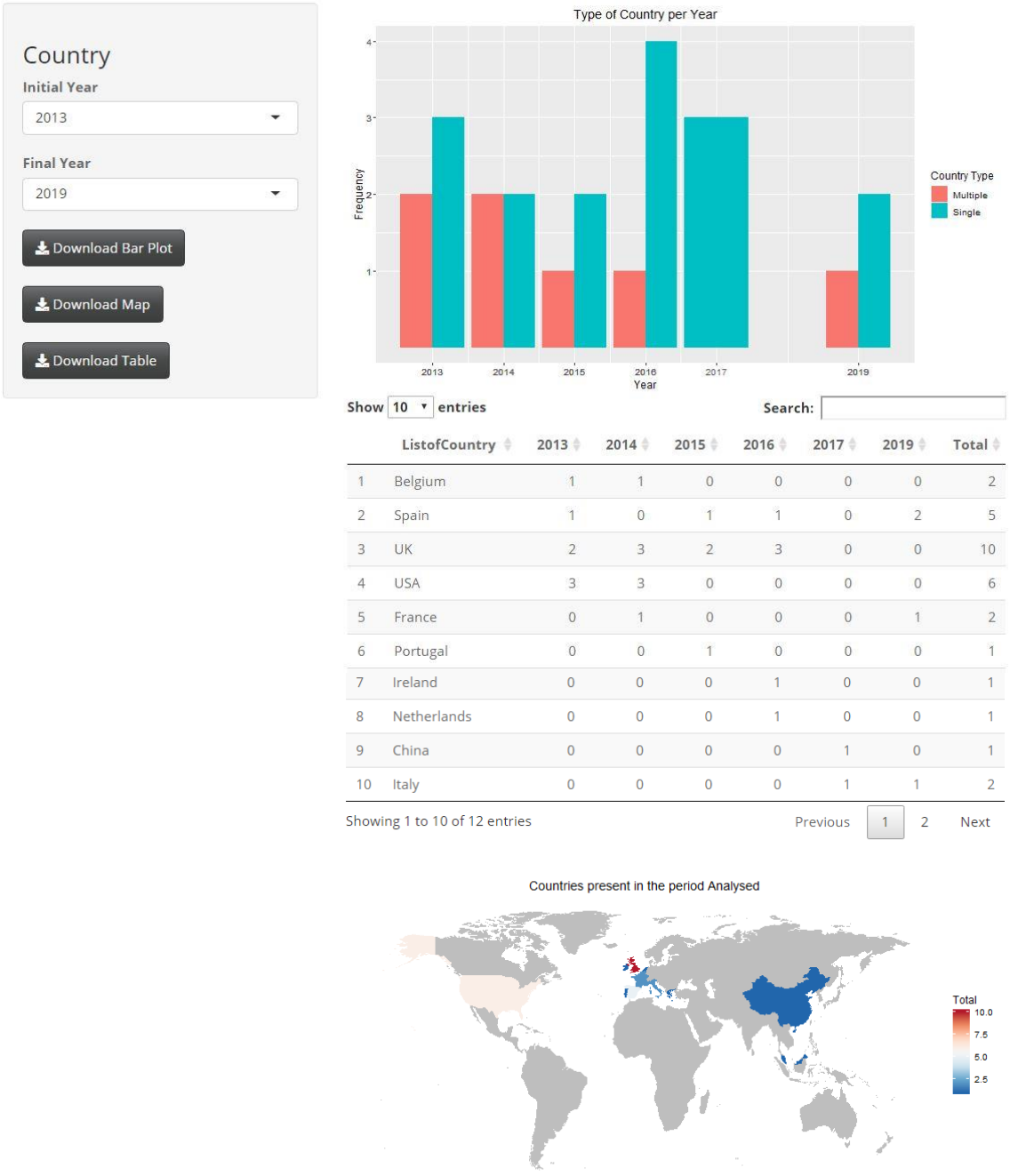


***Fig.26**: Example of the analysis of "Country".*

Fig.26 shows the three outputs realized: the bar plot, the table and the map. The starting point is always the same: on the left-hand side, the user selects the period of the analysis. The bar plot is used to give to the user an indication about the type of the country of the talent: it can be "*Single*" or "*Multiple*". It represents the frequency (y-axis) of each type,

for all the years of the analysis (x-axis). Then, below the graph is printed a table which gives to the user an insight into the name of the countries and how many times they appear per year and in total, during the period considered. Finally, a world map is printed. The countries which are present in the analysis are coloured by a gradient that allows distinguishing which country appears the most and which country appears the least. For example, from **Fig.26**, it is easy to understand that the UK is the most present country in the analysis, while China is one of the least. In this case, the user finds three download buttons, one for each output of the analysis.

Continent

The goal of this analysis is practically the same as the goal of the "*Country*" but referred to continents. An example is provided by the figure below:

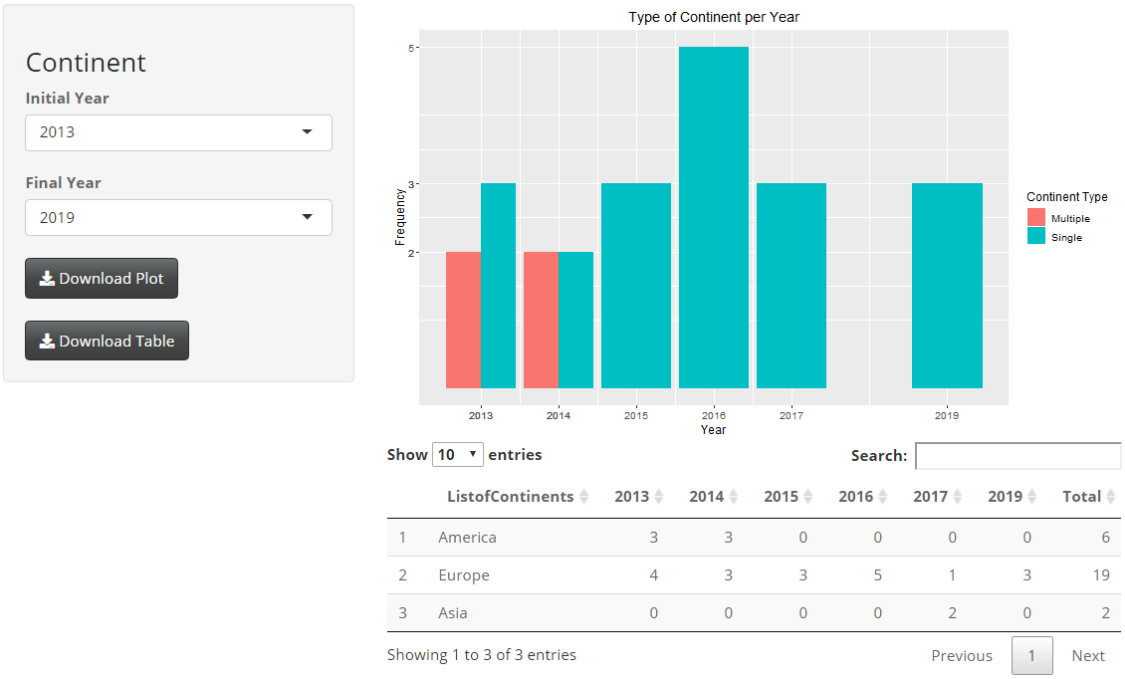


***Fig.27**: Example of the "Continent" analysis.*

The bar plot provides an indication of the type of continent: it can be "*Multiple*" or "*Single*". On the x-axis, there are the years of the analysis, while on the y-axis the frequency. Below the plot, there is the table which shows the frequency of appearance of each continent per year and in total. The starting point is always the set-up of the period by the user with the classic select boxes. The two download buttons are present.

DATA II

This is the third out of five pages of the content analysis. It provides five studies that are now well explained:

Sector of the Company

The goal of this analysis is to provide to the user the information about the type of the sector of the company and the list of the sectors, with an indication about the frequency of appearance in a certain period of time. The figure below shows an example:

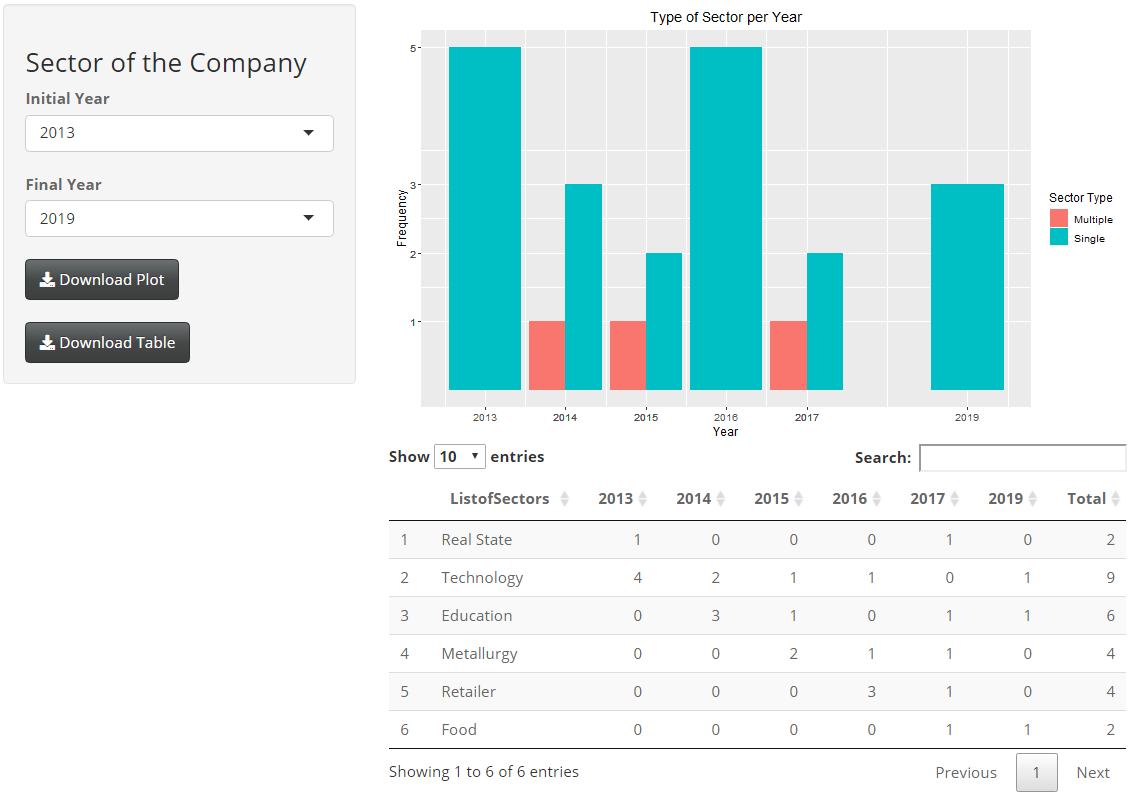


Fig.28: Example of the "Sector" analysis.

As it is possible to notice from **Fig.28**, the two outputs of the analysis are the classic ones: a bar plot for the type of the sector, with years on x-axis and frequency on y-axis, and a table with the list of the name of the sectors, which include also a "*Total*" column. The user can select the period of the analysis and download the two outputs with the appropriate buttons.

Size of the Company

The goal of this analysis is to show to the user the different types of the size of the company (if it is "*Single*" or "*Multiple*") and to understand which are the most common sizes in the period analysed. The first goal is reached with the classic bar plot, while the second goal is reached with the classic table. They are the same type of those previously described for the analysis of the "*Sector of the Company*". An example of it is placed in the Annex (**Fig.46**).

Type of the owner of the Company

The goal of this analysis is to show to the user the different types of the owner of the company (if it is "*Single*" or "*Multiple*") and to understand which are the most common types in the period analysed. The first goal is reached with the classic bar plot, while the second goal is reached with the classic table. They are the same type of those previously described for the analysis of the "*Sector of the Company*". An example of it is placed in the Annex (**Fig.47**).

Type of the Scope of the Company

The goal of this analysis is to show to the user the different types of the scope of the company (if it is "*Single*" or "*Multiple*") and to understand which are the most common types in the period analysed. The first goal is reached with the classic bar plot, while the second goal is reached with the classic table. They are the same type of those previously described for the analysis of the "*Sector of the Company*". An example of it is placed in the Annex (**Fig.48**).

Non-Profit Company

The goal of this analysis is to provide an answer to the following question: *"Is a company no profit ?"*. In order to reach this goal, a bar plot is used, as it is shown in the example below:

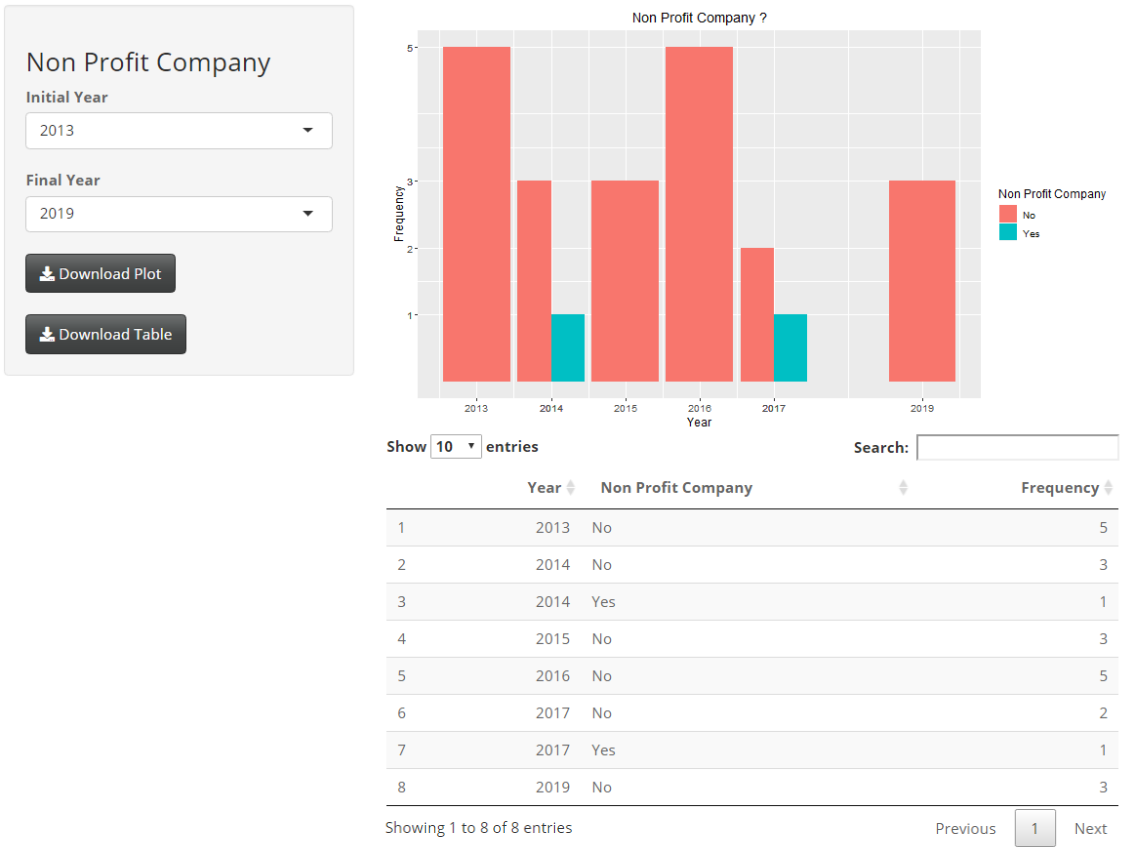


Fig.29: Example of the analysis of the question: "Is the company non-profit?".

As it is possible to notice from the **Fig.29**, an answer to the question can be easily found looking at the bar plot: the pink bars represent the number of profit companies, while the light-blue bars represent the non-profit companies. As in all bar plots, on the x-axis, there are the years while on the y-axis there are the frequencies. Below the bar plot, a table showing the frequency is printed.

RESEARCH DEFINITION

This is the second to last page of the content analysis. It allows the user to perform four analysis:

Explicit Research Question

The first analysis is about answering the following question: "*Is there an explicit research question?*". The possible answers are: "*Yes*" or "*No*". An example of this analysis is provided by the figure below:

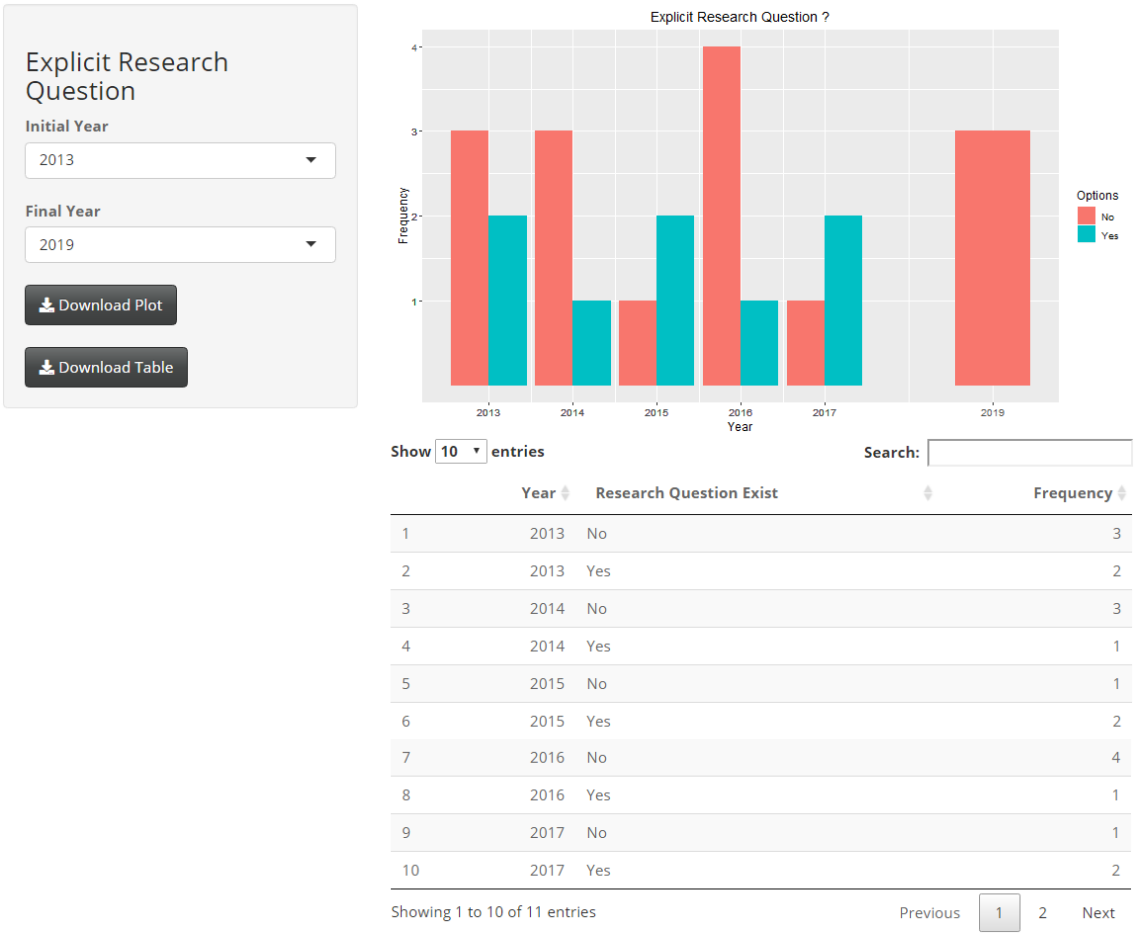


***Fig.30**: Example of the analysis for answering to the question: "Is the research question explicit?".*

As it is possible to see from the **Fig.30**, for the user is enough a look at the bar plot to find the answer to the question and how many times a certain answer appears per year. Moreover, below the graph, there is the classic table which provides the same information but in a different way. The rows of the table are sorted by ascending order of the year. The user has to select the years of the analysis and can download the plot and the table in CSV format (as in all the analysis).

Research Aim

The goal of this analysis is to answer the following question: "*Is there a research aim?*". Two answers are available: "*Yes*" or "*No*". The tools used for answering the question are the same as those previously described for the analysis of the "*Explicit Research Question*": a bar plot and a table. Therefore, an example is provided only in the Annex (**Fig.49**).

Focus of the Research

The aim of this analysis is to provide the user indication of which are the most frequent focus of the research for a certain period. In order to do this, a table is used. It is represented in the figure below:

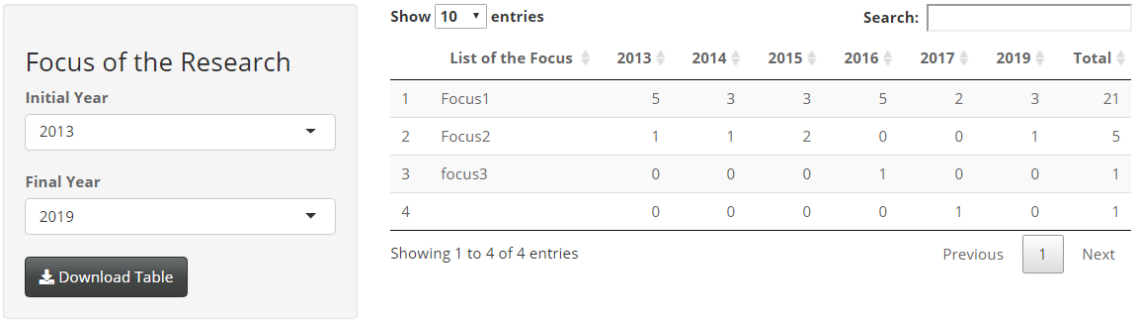


***Fig.31**: Example of the analysis of the focus of the research.*

The first peculiarity which is possible to notice from **Fig.31** is that one row has a "NA" as the value of the "*List of the Focus*" column. This does not mean that there is a mistake. It means that one observation of the content table placed in the page "*Content Management*" of the group "*Management*" has no value for the column "*List of the Focus*". Therefore, the user should find that observation and add the value at that column, if he wants to remove the space. Then, the structure of the table and the main panel placed on the left-hand side are the typical ones.

Academic Theories

The first goal of this analysis is to answer the following question: "*Does the academic theory exist?*". Two answers are available "*Yes*" or "*No*". The second goal of the analysis is to provide to the user the list of the academic theories which are in the analysis, with the indication of their frequencies per year and in total. An example is provided by the figure below:

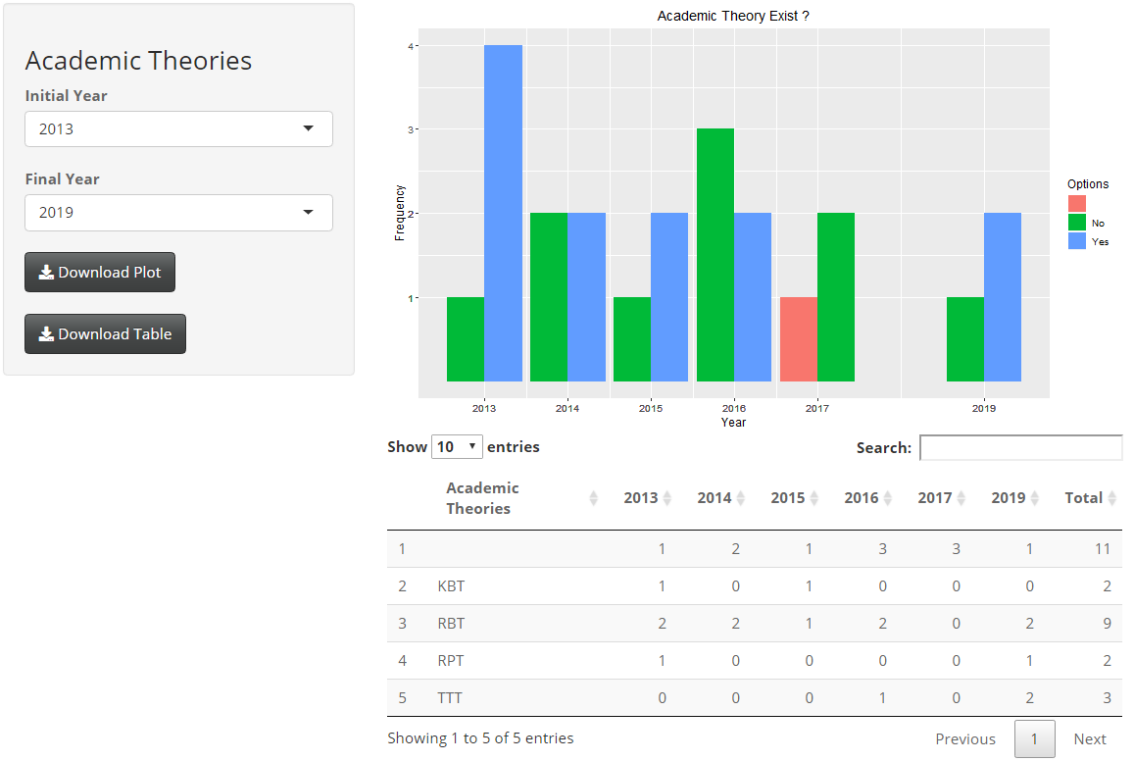


Fig.32: Example of the study done on academic theory.

As it is possible to notice from **Fig.32**, the structure is the classic one: the user selects the years and then the bar plot is printed for answering to the first question, while the table is shown for addressing the second goal. It is possible to notice from the table that there are 11 observations where it seems that there is no academic theory. However, 10 out of 11 are because of that reason, while 1 out of 11 is because there is no answer to the question. It is possible to notice this by looking at the bar plot in the year 2011: it is supposed to have only two bars (one for "*Yes*" and one for "*No*"). However, a third, pink and not specified bar is represented. Also in this case the user should go back to the table of contents in "*Managing Content*" in the group "*Managing*" for fixing the observation.

TALENT

This is the last page which belongs to the "*Content Analysis*". It provides six analysis on the aspects which strictly regard the talent and its management:

Definition of Talent

The first analysis is done to answer the question: "*Does the talent exist?*". Three answers are possible: "*Yes-direct*", "*Yes-indirect*" and "*No*". The tool used to answer to this question is the classic bar plot, as it is shown in the example below:

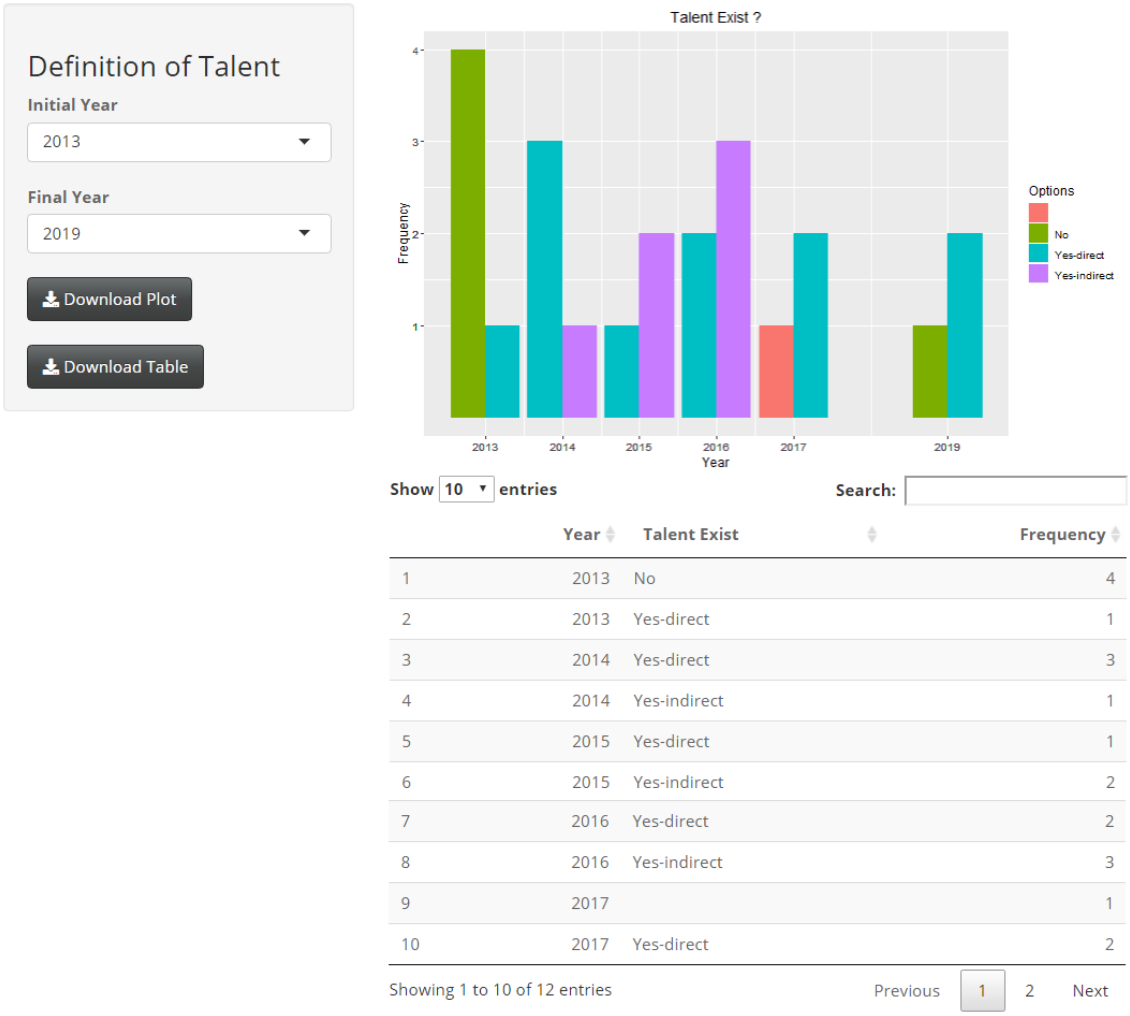


***Fig.33**: Example showing the tool used to the question: "Does the talent exist?".*

As it is possible to see from the **Fig.33**, there is still an observation which has not been defined well. The user can find an answer to the question by looking at the bar plot or looking at the table.

Source of Definition

The aim of this analysis is to provide to the user an indication about the source of the definition and how many times a specific source appears in the time window selected by

the user. In order to address this aim, a classic table is provided, as it is shown in the example below:

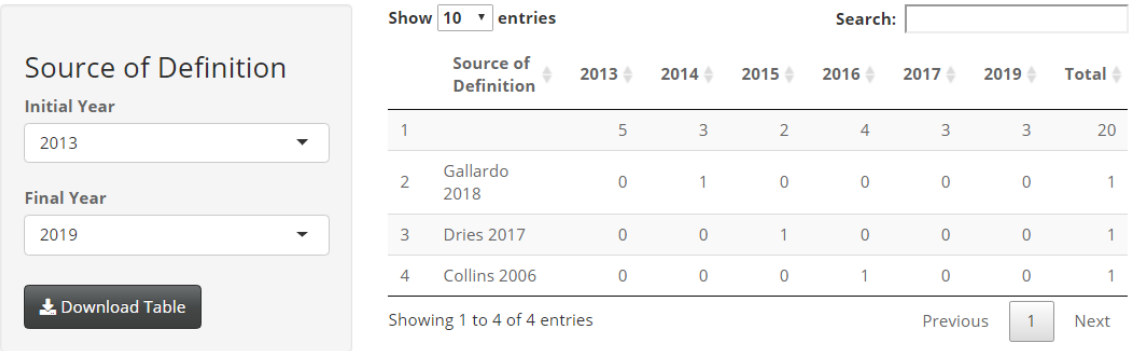


***Fig.34**: Example of the analysis on the source of the definition of talent.*

After the selection of the period by the user, done with the two select boxes on the left-hand side, the table of all the sources of the definition is printed. It provides an indication of the source of definition used each year and during the period considered. The user can download the CSV table with the appropriate download button.

Definition of Talent Management

The aim of this analysis is to answer the following question: "*Does the talent management exist?"*. Three answers are possible: "*No*", "*Yes-direct*" and "*Yes-indirect*". The tool used and the appearance of it is completely equal to the one used for the "*Definition of Talent*". An example is placed into the Annex (**Fig.50**).

Source of Definition of Talent Management

The aim of this analysis is to provide to the user the list of all the sources used for the definition of talent management which fall into a certain period of time. The tool used and its appearance is the same as the one used for the "*Source of Definition*" described previously. Therefore, the example is placed only in the Annex (**Fig.51**).

Discussion of Practices

The aim of this analysis is to answer the following question: "*Does the talent management practice exist?"*. Two answers are possible: "*No*", and "*Yes*". The tool used and the

appearance of it is completely equal to the one used for the "*Definition of Talent*". An example is placed only in the Annex (**Fig.52**).

Source of Definition of Talent Management

The aim of this analysis is to provide to the user the list of all the practices used for talent management, which fall into a certain period of time. The tool used and its appearance is the same as the one used for the "*Source of Definition*" described previously. Therefore, the example is placed only in the Annex (**Fig.53**).

7.4 REPORTING

This is the last page of the application. It enables the user to download a report of the analysis done through the section "*Analysis*". The page is divided into four blocks:

SET-UP INFORMATION

The first part of the "*Report*" page gives to the user the possibility to specify some basic information that will be part of the document that will be created. The figure below shows the possible fields the user can fill:

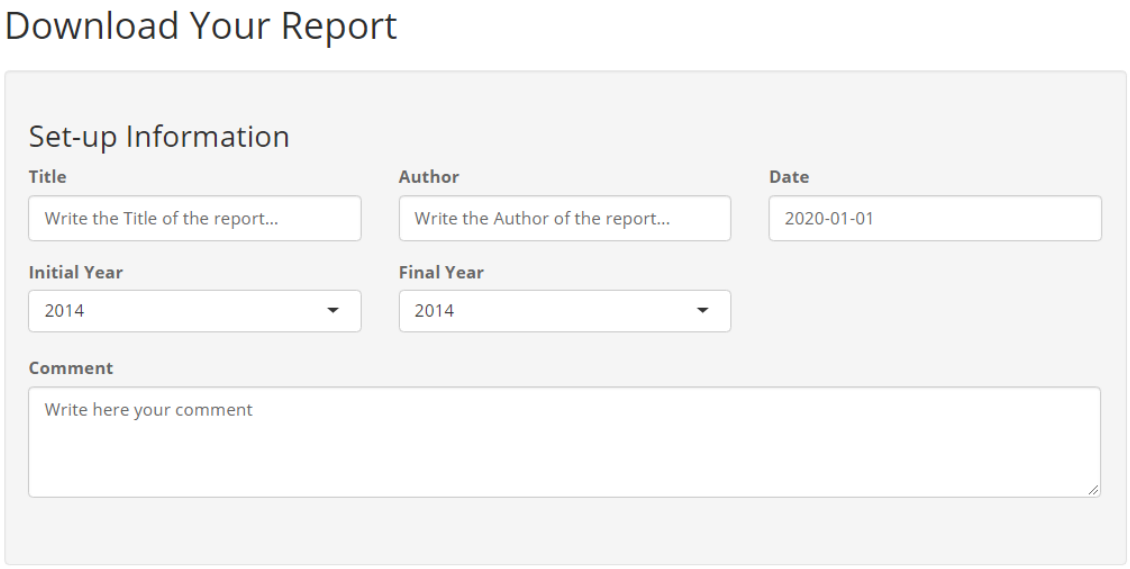


***Fig.35**: Example of the set-up information the user can add to the report.*

As it is possible to see from **Fig.35**, the first thing the user can add is the title of the report. Then, he can specify the author, the date of realization of the analysis, the time period

that the analysis considers and finally he can add a long comment about the analysis performed. A very important thing is that, when the user selects the time period of the analysis, it must be consistent with the period selected for the realization of graphs in the "*Bibliometric Analysis*" and "*Content Analysis*". If the two periods are different, the graph or table included considers the period as the one used for the creation of the output on the appropriate page. As it will be explained better in the next two subchapters, when a user selects a graph (or table) to be included in the report, the application goes to the page where the graph (or table) has been previously realized and it prints it in the report. The page "*Report*" does not compute any graph, but it simply recalls, if selected, those already done in the previous analysis.

BIBLIOMETRIC ANALYSIS

This is the second part of the "*Report*" page. It consists of a panel of checkboxes that give to the user the possibility to tick the graphs or tables he wants to include in the report. The figure below provides an example of this part:

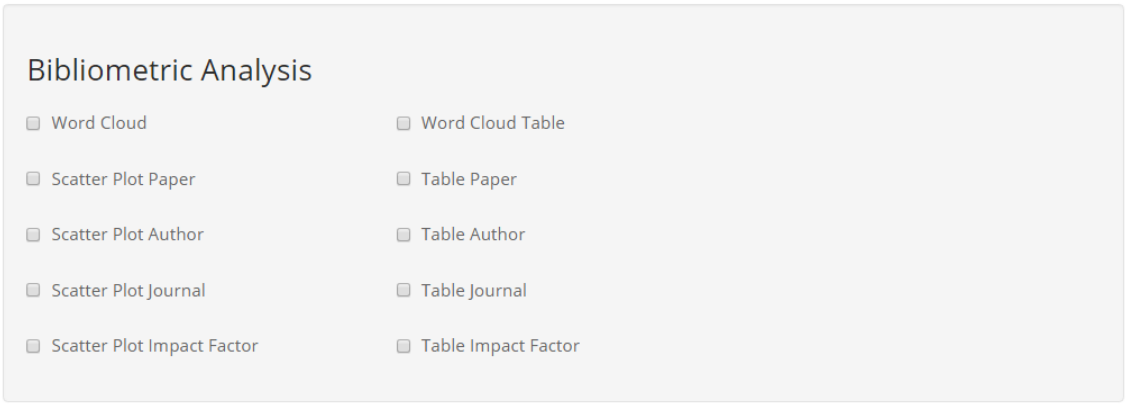


***Fig.36**: Example of the second part of the "Report" page.*

On the left-hand side of the panel shown by **Fig.36**, there are all the checkboxes that regard the graphs realized in the bibliometric analysis. On the right-hand side of the panel, there are the checkboxes of all the tables realized in the bibliometric analysis. The user can select the checkbox of the graphs or tables he wants to include in the report. When a graph or table is ticked, the application checks if it has been realized in the page "*Bibliometric Analysis*". If the application finds the graph or the table, it is included in the report. Otherwise, nothing is printed. Therefore, the user must be aware of two things:

1. In order to be included in the report, a graph or a table must be previously computed in the page "*Bibliometric Analysis*".
2. The table or graph included is the one created in the page "*Bibliometric Analysis*". Therefore, the parameters of the graph (or table), such as the period of analysis, the frequency of the word cloud of the maximum number of words are those specified when the graph (or table) has been realized.

CONTENT ANALYSIS

This is the third part of the "*Report*" page. It consists of a panel of checkboxes that give the user the possibility to tick the graphs or tables he wants to include in the report from the "*Content Analysis*" page. The figure below provides an example of this part:

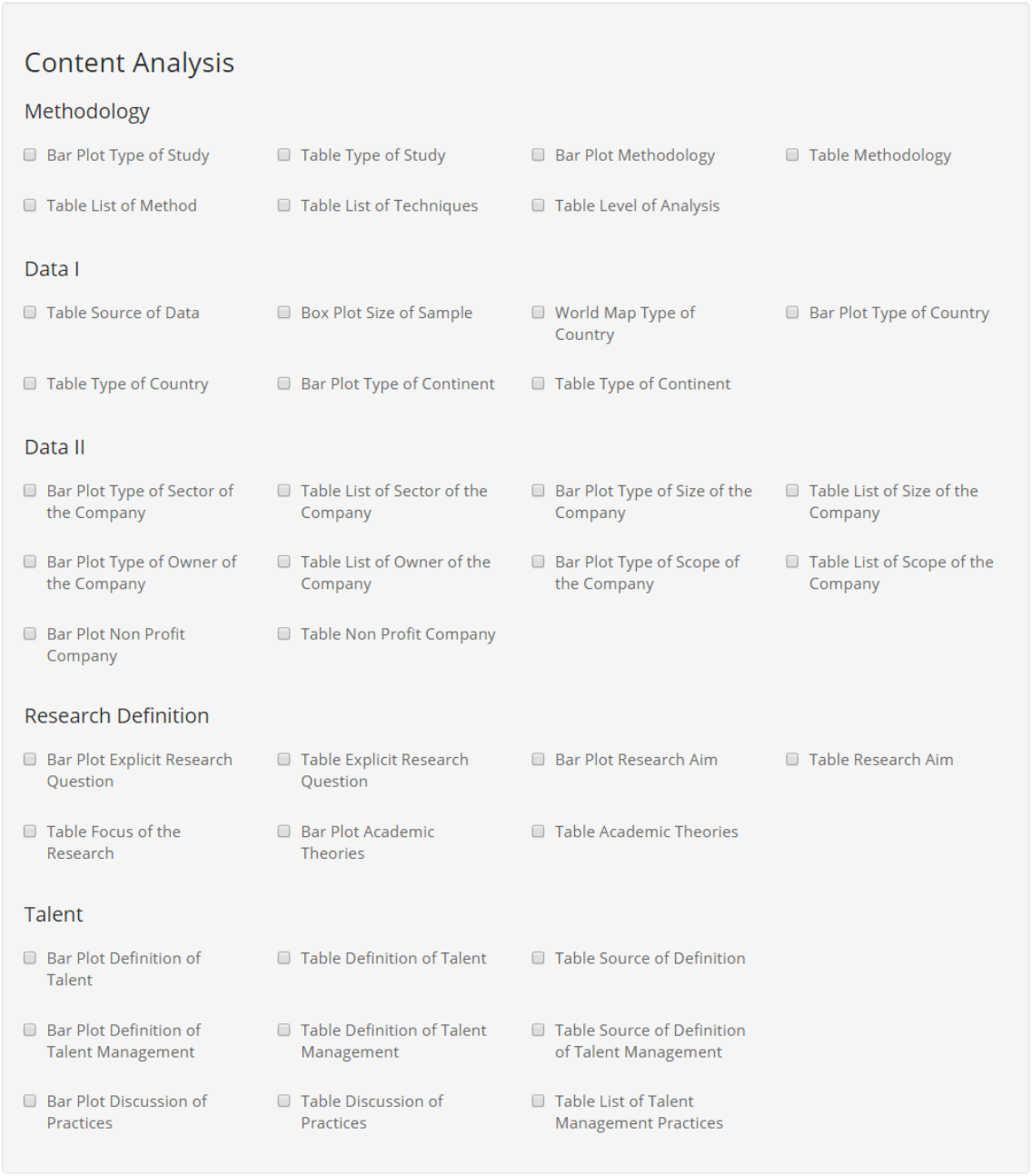


***Fig.37**: "Content Analysis" panel of the page "Report".*

The panel of **Fig.37** is divided into five parts, one for each page of the group "*Content Analysis*". There are all the graphs and tables that the user can realize in the "*Content Analysis*". The functioning is the same described for the previous panel: if a checkbox is ticked, the corresponding output is included in the report, if it has been previously realized. The two considerations done for the "*Bibliometric Analysis*" are still valid here:

1. In order to be included in the report, a graph or a table must be previously computed in the page "*Content Analysis*".
2. The table or graph included is the one created in the page "*Content Analysis*". Therefore, the parameters of the graph (or table), such as the period of analysis, are those specified when the graph (or table) has been realized.

FORMAT

The last part of the "*Report*" page is the panel shown in the figure below, which gives to the user the possibility to choose the format of the report:

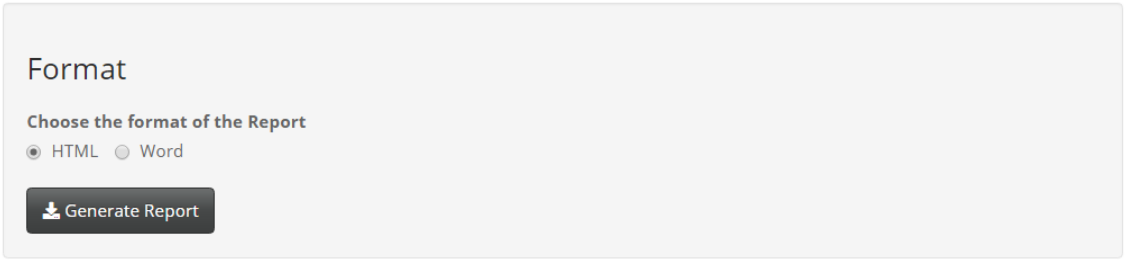


***Fig.38**: Last part of the page "Report".*

As it is possible to see from the **Fig.38**, the user can choose between two formats: "*HTML*" or "*Word*". When the user has chosen the format, he can download the report clicking on the download button "*Generate Report*". The generation of the report could require a certain amount of time, which depends on the number of graphs and tables selected and on the computational power of the laptop.

8. CONCLUSION

The "Talent Analyzer" application provides the possibility to perform two types of analysis also for users who do not have specific knowledge of R and programming. Personally speaking, the development of the platform helped me to improve the knowledge of R software that I have acquired during the master's degree at the "Università degli Studi di Bergamo" and, above all, during the master's degree at the "Universitat Politècnica de Catalunya". I had the possibility to dig into the field of bibliometric analysis which I did not face during my university route. Moreover, I explored more in detail the concept of "Data Analysis" which I consider really interesting, also for future employment. When I was choosing the master's thesis, I was looking for a project which gave me the possibility to develop something concrete, and which was not only a study. I can say that I reached my target with the development of this platform. Apart from the technical skills regarding R programming and MySQL, during the development of the platform I really improved also my soft skills, above all the problem solving: I faced several problems that I solved in different ways, which goes from the dialogue with my supervisor with the goal to find another path, to the deep researches in the R and MySQL community websites. I did not limit myself to execute the project of the platform that my supervisor gave it to me, but I always tried to find some improvements, proposing alternatives and additional features. I am really satisfied with myself about this project which I consider the perfect crowning to my university career.

For concluding this work, I would like to suggest some possible improvements for the application, which can be grouped into two categories: those regarding the programming side and those regarding the use of the application.

Starting from the first category, the first improvement which I suggest is to streamline the code which realizes the application, introducing the so-called "*Modules*". If someone wants to see and analyse the code, he will find around 5,000 lines of code. Therefore, the interpretation process may result difficult. A solution is to use "*Modules*" that are defined as small pieces of the application which are called by the main R script when it is necessary. In this way, the user can analyse better each function and see how the different modules interact among each others.

Then, in order to improve the speed of the application, would be interesting to understand which parts of the Shiny App are the slowest, in order to set a priority list. To do this the packages "*profvis*" provides a function with the same name which generates a report made of two graphs: one which shows how many time and memory each line of the code requires to execute and a second one which is a visual representation of the results of the profiler.

Considering now the application, one limitation is that the columns which take values from other tables of other pages are not editable, as specified in the blue warnings printed above some tables. For instance, if the user adds a paper, makes a mistake in choosing the journal and he recognizes it after he saved, he has to delete the record and create it again. This represents a waste of time for the user which can be avoided by allowing the editing of those types of columns.

Another possible small improvement regards the binary variables. It is common knowledge that a binary variable can assume only two values, 1 and 0. However, there could be the possibility that not all the users are aware of it. Therefore, would be better to display the meaning of 1 and 0 in the table or, at least, limit the range of numbers that can be added in the binary fields. In fact, the binary columns admit integer numbers and a user could insert a value like "2" or "10" and no mistake or warning would appear.

When, for instance, a user adds a new journal, he moves to the page "*Managing Paper*" for adding a new paper and from the drop-down menu of the column "*Journal*" he searches the new observation previously added, he will not find it. As it is specified by the warnings, when an observation is added and it regards also other tables in other pages, in order to see it, the user has to reload the Shiny Application. When the Shiny Application is reloaded, it brings the user to the first page ("*Project*"). A first improvement would be to leave the user on the page where he is when he reloads the application. A second improvement would regard the elimination of the necessity of reloading the Application.

For the correct functioning of the application, is necessary that the user follows a specific path: first, he checks if there is the project in the table. If not, he creates a new one. Then he checks if the affiliation, author and journal are already in the application. If not, he creates what he needs. Then he can add first the paper, then the keywords and finally the

content. It is recommended that one overall record is complete before a new one starts. This means that the user cannot add first all the journals, then all the authors and then all the papers. This could bring problems to the application. This path may represent a limitation for some users.

One possible improvement regards the realization of the word cloud in the "*Bibliometric Analysis*". If the minimum frequency selected through the slider is higher than the maximum frequency displayed in the table, the word cloud is created by adding all the terms. Therefore, it would be better that a warning message is printed, rather than a wrong word cloud. This is a problem related to the package that creates the word cloud.

All the select boxes which allow the user to select the period of the analysis in the page "*Bibliometric Analysis*" and "*Content Analysis*" are created in the same following way: the column "*Year*" of the table of papers in "*Managing Papers*" is taken and it is used as the list of values of the select boxes. However, the main problem is that the list of values follows the order of the papers added to the table. Therefore, the years are not ordered. I did not find a way of ordering them and this is a point that could be improved.

The last improvement which I suggest regards an expansion of the power of analysis of the application. For instance, the bibliometric analysis could include also a study about the number of papers published by the author. Moreover, if some other information regarding the papers, such as the number of citations or the H-Index, are collected, the bibliometric analysis could provide even more outputs to the user.

9. REFERENCES

ACADEMIC PAPERS USED FOR LITERATURE REVIEW

[1] Merigó, J. M., & Yang, J. B. (2017). A bibliometric analysis of operations research and management science. *Omega*, *73*, 37-48.

[2] Martín-Martín, A., Orduna-Malea, E., Thelwall, M., &, López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Infometrics, 12*, 1160-1177.

[3] Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics, 106*, 787-804.

[4] Martín-Martín, A., Orduna-Malea, E., & López-Cózar, E. D. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics, 116*, 2175-2188.

[5] Siddaway, A. P., Wood, A. M., & Hedges, L.V. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology, 70*, 747-770.

ACADEMIC PAPERS USED FOR THE SAMPLE:

[6] Nijs, S., Gallardo-Gallardo, E., Dries, N., & Sels L. (2014). A multidisciplinary review into the definition, operationalization, and measurement of talent. *Journal of World Business, 49*, 180-191.

[7] Gallardo-Gallardo, E., Dries, N., & González-Cruz, T. F. (2013). What is the meaning of 'talent' in the world of work?. *Human Resource Management Review, 23*, 290-300.

[8] Minbaeva, D., & Collings, G. D. (2013). The International Journal of Human Resource Management, *The International Journal of Human Resource Management, 24, 9*, 1762-1776.

[9] Meyers, M. C., & van Woerkom, M. (2014). The influence of underlying philosophies on talent management: Theory, implications for practice, and research agenda. *Journal of World Business, 49*, 192-203.

[10] Thunnissen, M., & Gallardo-Gallardo, E. (2019). Rigor and relevance in empirical TM research: Key issues and challenges. *BRQ Business Research Quarterly*, *22*, 171-180.

[11] Thunnissen, M., Boselie, P., & Fruytier, B. (2013). Talent management and the relevance of context: Towards a pluralistic approach. *Human Resource Management Review*, *23*, 326-336.

[12] De Vos, A., & Dries, N. (2013). The International Journal of Human Resource Management. *The International Journal of Human Resource Management, 24, 9,* 1816-1831.

[13] King, K. A., & Vaiman, V. (2019). Enabling effective talent management through a macro-contingent approach: A framework for research and practice, *BRQ Business Research Quarterly*, *22*, 194-206.

[14] Claus, L. (2019). HR disruption - Time already to reinvent talent management. *BRQ Business Research Quarterly*, *22*, 207-215.

[15] Swailes, S. (2013). The ethics of talent management. *Business Ethics: A European Review*, *22,1*, 32-46.

[16] Malik, A. R., & Singh, P. (2014) 'High potential' programs: Let's hear it for 'B' players. *Human Resource Management Review, 24*, 330-346.

[17] Powell, M. (2014). The snakes and ladders of National Health Service management in England. *The International Journal of Health Planning and Management, 29*, 260-279.

[18] Gallardo-Gallardo, E., Nijs, S., Dries, N., & Gallo, P. (2015). Towards an understanding of talent management as a phenomenon-driven field using bibliometric and content analysis. *Human Resource Management Review, 25*, 264-279.

[19] Morley, J. M., Scullion, H., Collings, D. G., & Schuler, R. S. (2015). Talent management: a capital question. *European Journal International Management, 9,1,* 1-8.

[20] Schuler, R. S. (2015). The 5-C framework for managing talent. *Organizational Dynamics, 44,* 47-56.

[21] Gallardo-Gallardo, E., & Thunnissen, M. (2016). Standing on the shoulders of giants? A critical review of empirical talent management research. *Employee Relations, 38(1),* 31-56.

[22] Swailes, S., & Blackburn, M. (2016). Employee reactions to talent pool membership. *Employee Relations, 38(1),* 112-128.

[23] Tansley, C., Hafermalz, E., & Dery, K. (2016). Talent development gamification in talent selection assessment centres. *European Journal of Training and Development, 40, 7,* 490-512.

[24] Tatoglu, E., Glaister, A. J., & Demirbag, M. (2016). Talent management motives and practices in an emerging market: A comparison between MNEs and local firms. *Journal of World Business, 51,* 278-293.

[25] Wiblen, S. (2016). Framing the usefulness of eHRM in talent management: A case study of talent identification in a professional services firm. *Canadian Journal of Administrative Sciences, 33,* 95-107.

[26] Gallardo-Gallardo, E., Moliner, A. L., & Gallo, P. (2017). Mapping collaboration networks in talent management research. *Journal of Organizational Effectiveness, People and Performance, 4,4,* 332-358.

[27] Khoreva, V., Vaiman, V., & Van Zalk, M. (2017). Talent management practice effectiveness: investigating employee perspective. *Employee Relations, 39(1),* 100-111.

[28] King, A. K. (2017). Creating an organisational context supportive of sustainable talent development through implementation of a strong talent system. *Journal of Organizational Effectiveness People and Performance, 4,4,* 298-314.

WEBSITES

[29] Learn Shiny. (n.d). Retrieved from https://shiny.rstudio.com/tutorial/

[30] Articles. (n.d). Retrieved from https://shiny.rstudio.com/articles/

[31] Jbryer/dtedit. (n.d). Retrieved from https://github.com/jbryer/DTedit

[32] DT: An R interface to the DataTables library. (n.d). Retrieved from https://rstudio.github.io/DT/

[33] Attali, D. (2017). Persistent data storage in Shiny apps. Retrieved from https://shiny.rstudio.com/articles/persistent-data-storage.html#mysql

[34] Blackard, J. (2018, March 03). Introduction to MySQL with R. Retrieved from https://programminghistorian.org/en/lessons/getting-started-with-mysql-using-r#create-an-r-script-that-connects-to-the-database

[35] Attali, D. (2016, August 29). Shiny tips & tricks for improving your apps and solving common problems. Retrieved from https://deanattali.com/blog/advanced-shiny-tips/

[36] Karimeddini, F. (n.d.). Word Cloud. Retrieved from https://shiny.rstudio.com/gallery/word-cloud.html

[37] Chang, W. (n.d). Generating downloadable reports. Retrieved from https://shiny.rstudio.com/articles/generating-reports.html

[38] Chang, W. (n.d). Generating reports. Retrieved from https://shiny.rstudio.com/gallery/generating-reports.html

[39] RStudio, Inc. (n.d). Downloading knitr Reports. Retrieved from https://shiny.rstudio.com/gallery/download-knitr-reports.html

[40] R (programming Language). (n.d.). Retrieved from https://en.wikipedia.org/wiki/R_(programming_language)

[41] What is R? (n.d.). Retrieved from https://www.r-project.org/about.html

[42] de Vries, A., & Meys, J. (n.d.). The Benefits of Using R. Retrieved from https://www.dummies.com/programming/r/the-benefits-of-using-r/

[43] 6 Reasons: Why Choose R Programming for Data Science Projects?. (2017, September 18). Retrieved from https://www.newgenapps.com/blog/6-reasons-why-choose-r-programming-for-data-science-projects

[44] MySQL. (n.d.). Retrieved from https://en.wikipedia.org/wiki/MySQL

[45] Rouse, M. (n.d.). MySQL. Retrieved from https://searchoracle.techtarget.com/definition/MySQL

[46] Top Reasons to Use MySQL. (n.d.). Retrieved from http://www.databasequest.com/index.php/product-service/mysql-dbquest

[47] Branson, T. (2016, November 16). 8 Major Advantages of Using MySQL. Retrieved from https://www.datamation.com/storage/8-major-advantages-of-using-mysql.html

[48] Mack, J. (2014, May 28). Five Advantages & Disadvantages Of MySQL. Retrieved from https://www.datarealm.com/blog/five-advantages-disadvantages-of-mysql/

[49] RStudio. (n.d). Retrieved from https://rstudio.com/products/rstudio/

[50] RStudio. (n.d.). Retrieved from https://en.wikipedia.org/wiki/RStudio

[51] Love, K. (n.d.). The Advantages of RStudio. Retrieved from https://www.theanalysisfactor.com/the-advantages-of-rstudio/

[52] 9 reasons to use RStudio. (2012, October 16). Retrieved from https://benjaminlmoore.wordpress.com/2012/10/16/9-reasons-to-use-rstudio/

[53] The basic parts of a Shiny app. (n.d). Retrieved from https://shiny.rstudio.com/articles/basics.html

[54] Przytula, P. (2019, September 10). Why You Should Use R Shiny for Enterprise Application Development. Retrieved from https://appsilon.com/why-you-should-use-r-shiny-for-enterprise-application-development/

[55] Pylvainen, I. (n.d). Why would I use Shiny instead of Tableau, Spotfire, Qlikview or similar BI tools ?. Retrieved from https://support.rstudio.com/hc/en-us/articles/218294727-Why-would-I-use-Shiny-instead-of-Tableau-Spotfire-Qlikview-or-similar-BI-tools-

[56] Gott, A. (2018, June 8). Why Bother with Shiny?. Retrieved from https://www.r-bloggers.com/why-bother-with-shiny/

[57] Scatter plot. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Scatter_plot

[58] Bar chart. (n.d). Retrieved from https://en.wikipedia.org/wiki/Bar_chart

[59] Box plot. (n.d). Retrieved from https://en.wikipedia.org/wiki/Box_plot

[60] Tag cloud. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Tag_cloud

[61] Liberto, D. (n.d.). Quartile. Retrieved from https://www.investopedia.com/terms/q/quartile.asp

[62] Optimizing your Application. (n.d.). Retrieved from https://bookdown.org/hadrien/how_to_build_a_shiny_app_from_scratch/optimizing-your-application.html

[63] Scopus. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Scopus

[64] Web of Science (n.d.). Retrieved from https://en.wikipedia.org/wiki/Web_of_Science

[65] Google Scholar. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Google_Scholar

[66] Using Bibliometrics: a guide to evaluating research performance with citation data. (n.d.). Retrieved from http://ips.clarivate.com/m/pdfs/325133_thomson.pdf

[68] Pros and cons of bibliometrics. (n.d.). Retrieved from http://www.open.ac.uk/library-research-support/bibliometrics/pros-and-cons-bibliometrics

[69] Systematic review. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Systematic_review

[70] The Role of Systematic Literature Reviews. (n.d.). Retrieved from http://rtcil.org/role-systematic-literature-reviews

[71] Piper, J. R. (n.d.). How to write a systematic literature review: a guide for medical students. Retrieved from http://sites.cardiff.ac.uk/curesmed/files/2014/10/NSAMR-Systematic-Review.pdf

R PACKAGES

[72] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

[73] Winston Chang [aut, cre], Joe Cheng [aut], JJ Allaire [aut], Yihui Xie [aut], Jonathan McPherson [aut]. shiny: Web Application Framework for R, version 1.4.0. https://cran.r-project.org/web/packages/shiny/index.html

[74] Yihui Xie [aut, cre], Joe Cheng [aut], Xianying Tan [aut], JJ Allaire [ctb], Maximilian Girlich [ctb], Greg Freedman Ellis [ctb], Johannes Rauh [ctb]. DT: A Wrapper of the JavaScript Library 'DataTables', version 0.11. https://cran.r-project.org/web/packages/DT/index.html

[75] Jeroen Ooms [aut, cre], David James [aut], Saikat DebRoy [aut], Hadley Wickham [aut], Jeffrey Horner [aut]. RMySQL: Database Interface and 'MySQL' Driver for R, version 0.10.18. https://cran.r-project.org/web/packages/RMySQL/index.html

[76] Dason Kurkiewicz, Heike Hofmann, Ulrike Genschel. dbConnect: Provides a graphical user interface to connect with databases that use MySQL, version 1.0. https://cran.r-project.org/web/packages/dbConnect/index.html

[77] Matt Dowle [aut, cre], Arun Srinivasan [aut]. data.table: Extension of 'data.frame', version 1.12.8. https://cran.r-project.org/web/packages/data.table/index.html

[78] Hadley Wickham [aut, cre], Romain François [aut], Lionel Henry [aut], Kirill Müller [aut]. dplyr: A Grammar of Data Manipulation, version 0.8.3. https://cran.r-project.org/web/packages/dplyr/index.html

[79] Craig Varrichio. rowr: Row-Based Functions for R Objects, version 1.1.3. https://cran.r-project.org/web/packages/rowr/index.html

[80] Hadley Wickham [aut, cre]. plyr: Tools for Splitting, Applying and Combining Data, version 1.8.5. https://cran.r-project.org/web/packages/plyr/index.html

[81] Ingo Feinerer [aut, cre], Kurt Hornik [aut]. tm: Text Mining Package, version 0.7-7. https://cran.r-project.org/web/packages/tm/index.html

[82] Ian Fellows. wordcloud: Word Clouds, version 2.6. https://cran.r-project.org/web/packages/wordcloud/index.html

[83] Hadley Wickham [aut, cre], Winston Chang [aut], Lionel Henry [aut], Thomas Lin Pedersen [aut], Kohske Takahashi [aut], Claus Wilke [aut], Kara Woo [aut], Hiroaki Yutani [aut]. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, version 3.2.1. https://cran.r-project.org/web/packages/ggplot2/index.html

[84] JJ Allaire [aut], Yihui Xie [aut, cre], Jonathan McPherson [aut], Javier Luraschi [aut], Kevin Ushey [aut], Aron Atkins [aut], Hadley Wickham [aut], Joe Cheng [aut], Winston Chang [aut], Richard Iannone [aut]. rmarkdown: Dynamic Documents for R, version 2.0. https://cran.r-project.org/web/packages/rmarkdown/index.html

[85] Yihui Xie ORCID iD [aut, cre]. knitr: A General-Purpose Package for Dynamic Report Generation in R, version 1.26. https://cran.r-project.org/web/packages/knitr/index.html

[86] Hadley Wickham [aut, cre], Lionel Henry [aut]. tidyr: Tidy Messy Data, version 10.0. https://cran.r-project.org/web/packages/tidyr/index.html

[87] Edzer Pebesma [aut, cre], Roger Bivand [aut], Barry Rowlingson [ctb], Virgilio Gomez-Rubio [ctb], Robert Hijmans [ctb], Michael Sumner [ctb], Don MacQueen [ctb], Jim Lemon [ctb], Josh O'Brien [ctb], Joseph O'Rourke [ctb]. sp: Classes and Methods for Spatial Data, version 1.3-2. https://cran.r-project.org/web/packages/sp/index.html

[88] Richard A. Becker and Allan R. Wilks. maps: Draw Geographical Maps, version 3.3.0. https://cran.r-project.org/web/packages/maps/index.html

10. ANNEX

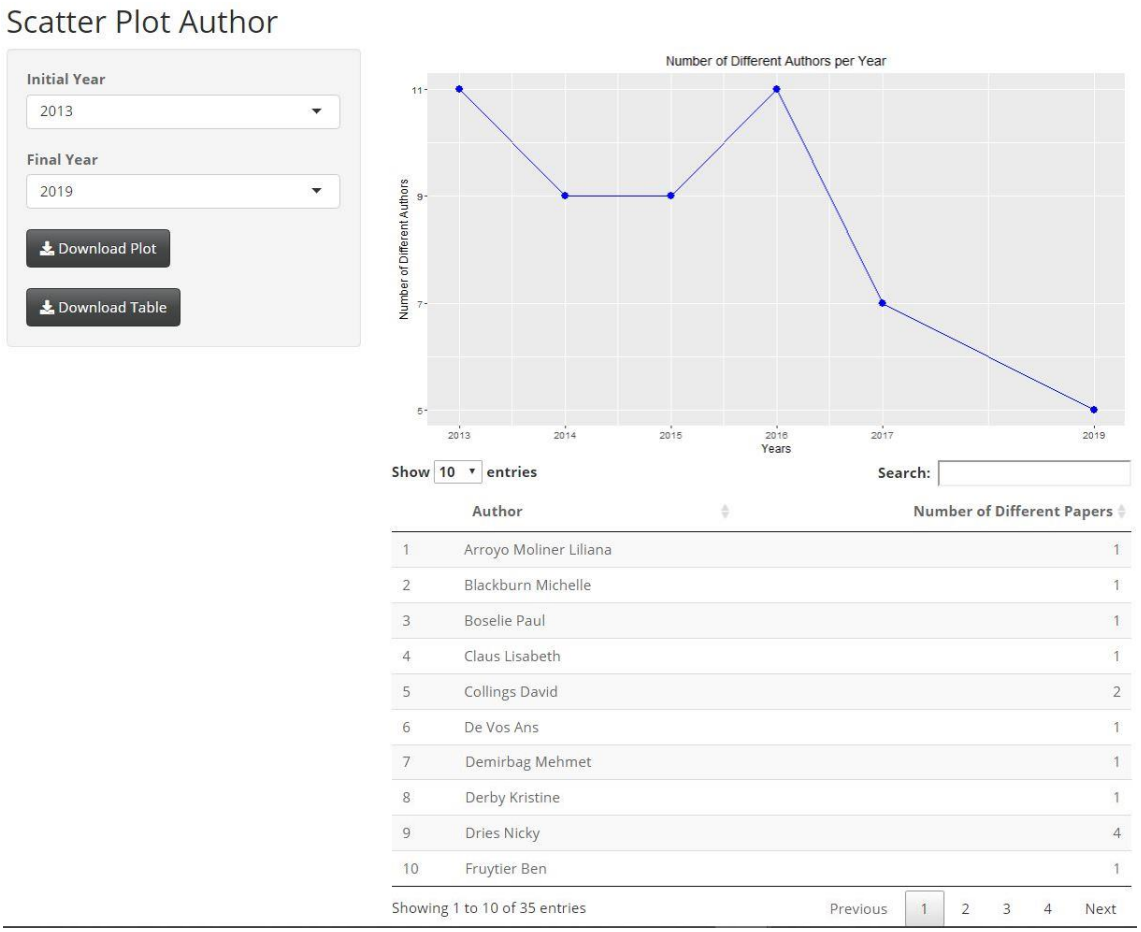


***Fig.39**: Example of the outputs created during the analysis of the authors.*

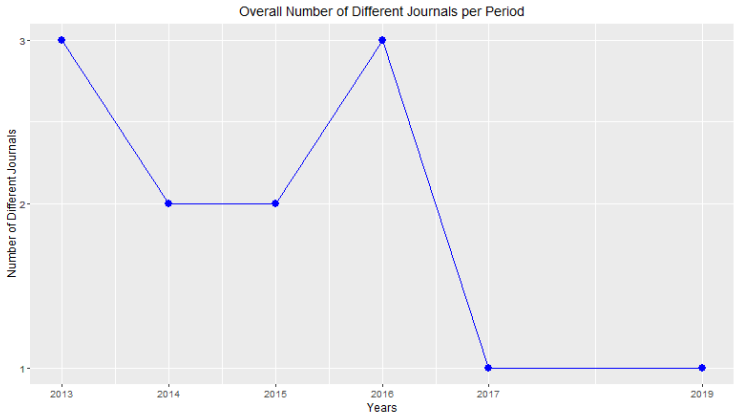
Scatter Plot Journal

Initial Year
2013

Final Year
2019

Download Plot

Download Table



Show 10 entries Search:

	Journal	Number of Different Papers
1	BRQ Business Research Quarterly	3
2	Business Ethics: A European Review	1
3	Canadian Journal of Administrative Sciences	1
4	Emerald Insight-Employee Relations	3
5	European Journal of International Management	1
6	European Journal of Training and Development	1
7	Human Resource Management Review	4
8	Journal of Organizational Effectiveness: People and Performance	2
9	Journal of World Business	3
10	Organizational Dynamics	1

Showing 1 to 10 of 12 entries Previous 1 2 Next

***Fig.40**: Example of the outputs realized with the analysis on the journals.*

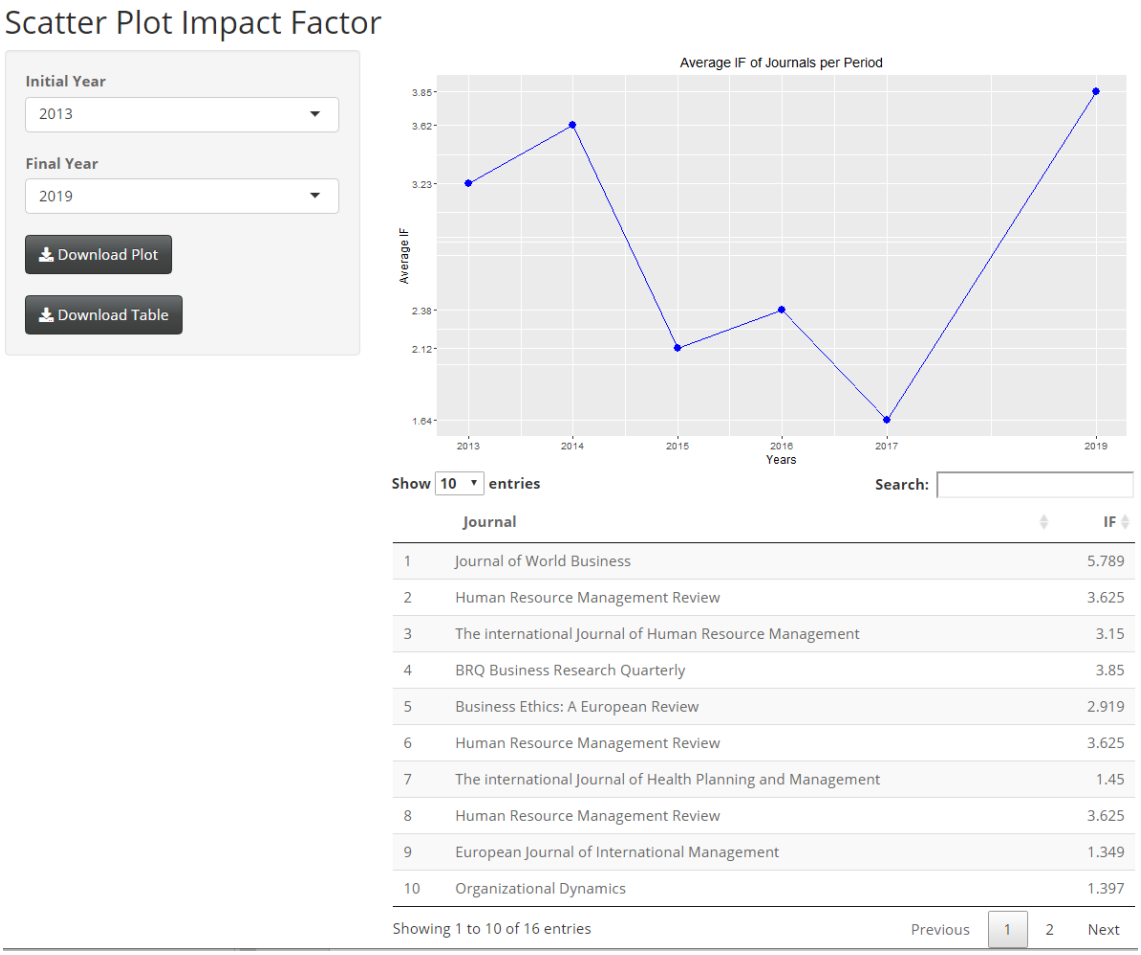


***Fig.41**: Example of the outputs that can be realized for the analysis of the impact factor.*

Methodology

Initial Year

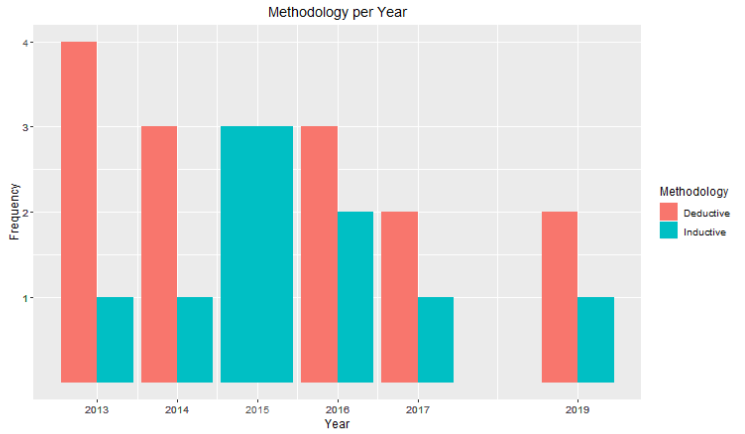
2013

Final Year

2019

Download Plot

Download Table



Show 10 entries Search:

	Year	Methodology	Frequency
1	2013	Deductive	4
2	2013	Inductive	1
3	2014	Deductive	3
4	2014	Inductive	1
5	2015	Inductive	3
6	2016	Deductive	3
7	2016	Inductive	2
8	2017	Deductive	2
9	2017	Inductive	1
10	2019	Deductive	2

Showing 1 to 10 of 11 entries Previous 1 2 Next

***Fig.42**: Example of the analysis done on the type of "Methodology".*

List of Techniques

Initial Year

2013

Final Year

2019

Download Table

Show 10 entries Search:

	ListofTechniques	2013	2014	2015	2016	2017	2019	Total
1	ANOVA	1	1	0	0	0	0	2
2	Linear Regression	2	1	0	2	1	2	8
3	QCA-fs	3	0	0	0	0	0	3
4	SEM	3	1	0	0	3	0	7
5	Grounded Theory	0	1	2	2	0	1	6
6	Logistic Regression	0	1	0	2	1	2	6
7	QCA-cs	0	1	0	0	0	0	1
8	Content Analysis	0	0	2	2	0	1	5
9	Discussion Analysis	0	0	2	3	0	1	6
10	Thematic Analysis	0	0	1	2	0	0	3

Showing 1 to 10 of 10 entries Previous 1 Next

***Fig.43**: Example of the analysis of the possible techniques.*

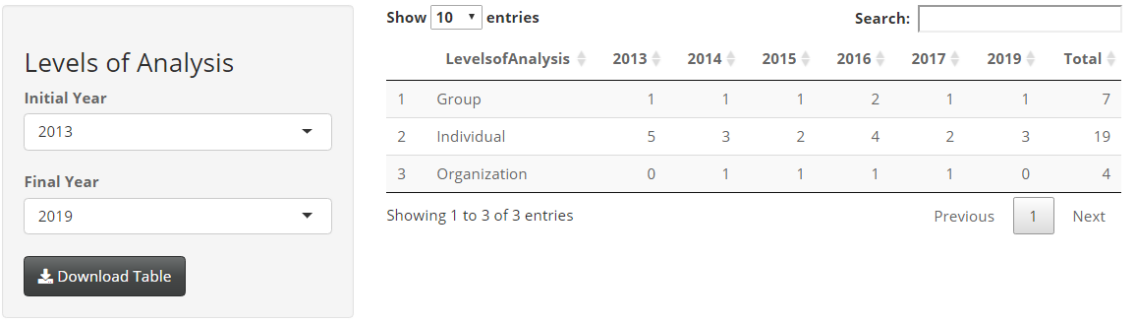


Fig.44: Example of the study done on the possible levels of the analysis.

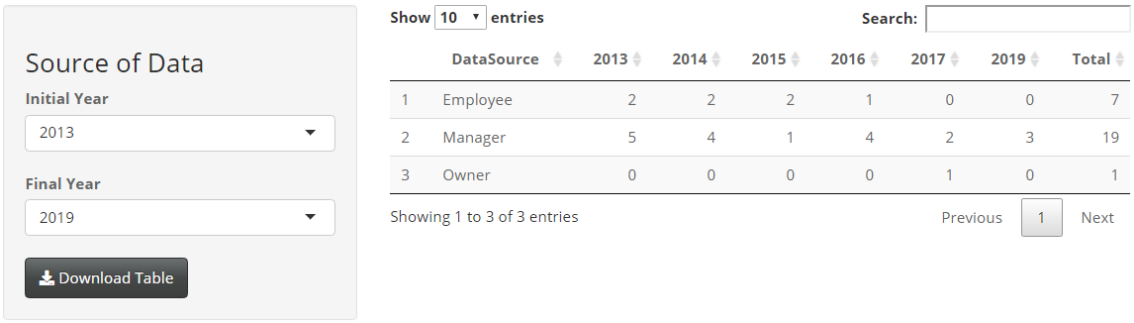


Fig.45: Example of the study on the possible source of data.

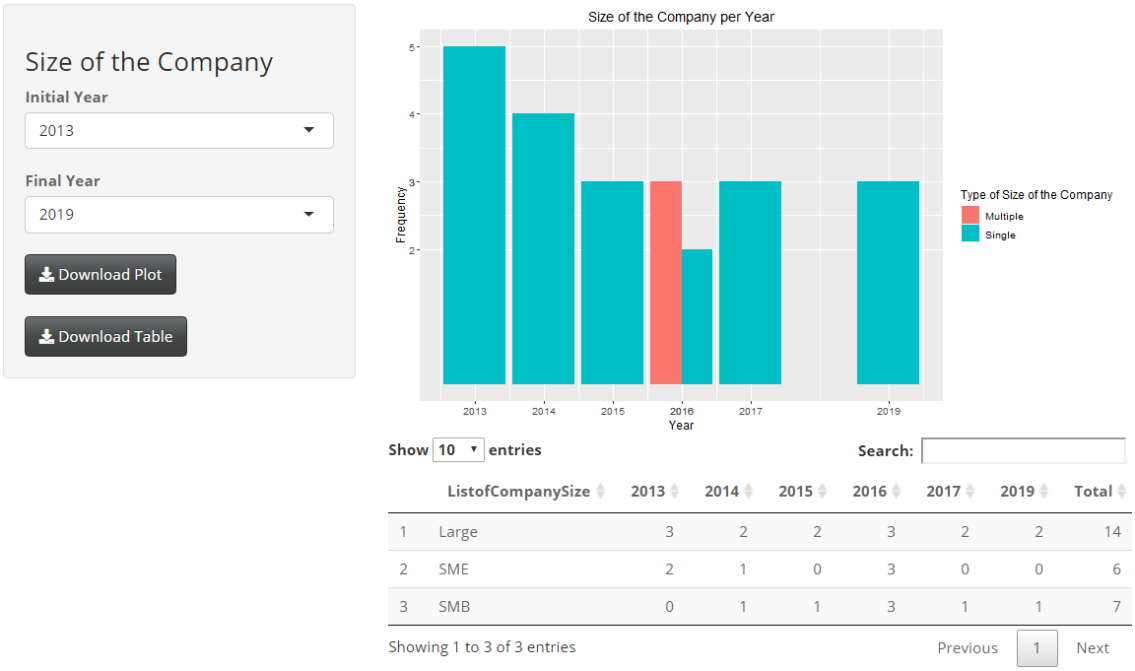


Fig.46: Example of the analysis done on the type of size of the company.

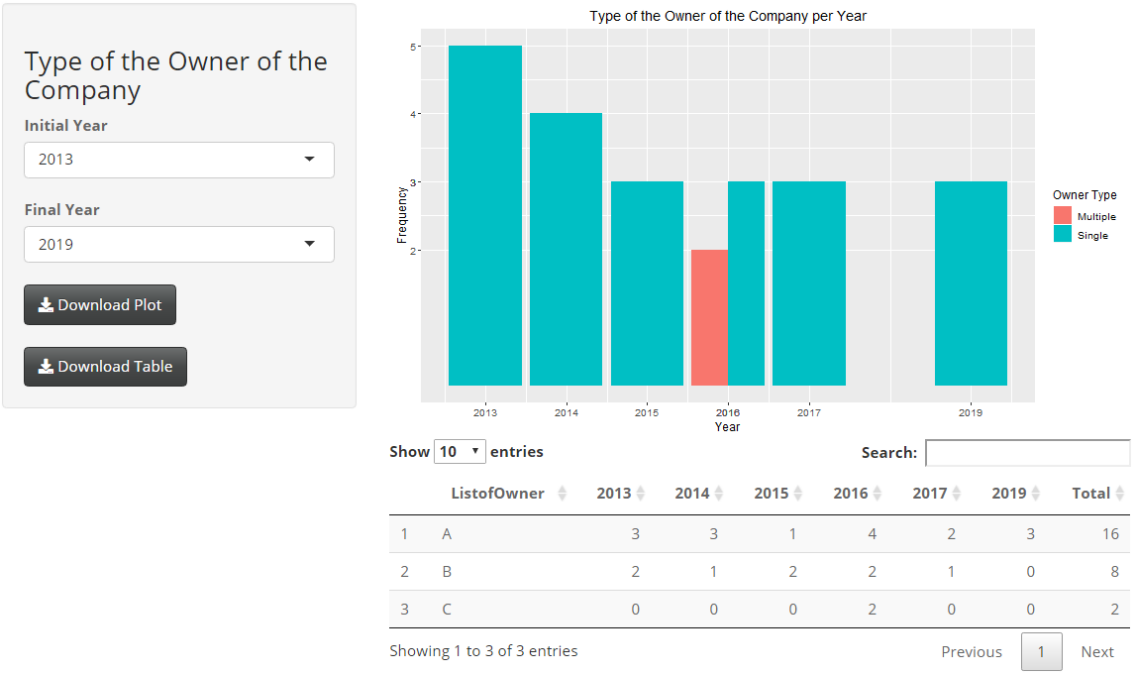


Fig.47: Example on the analysis done on the type of owner.

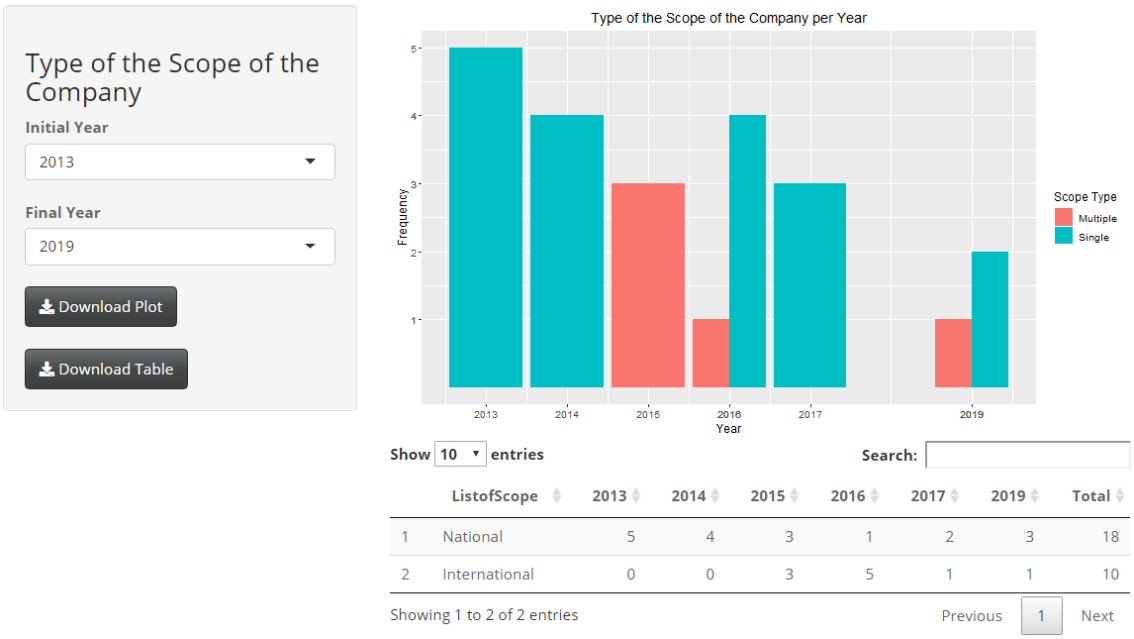


Fig.48: Example of the analysis done on the type of scope of the company.

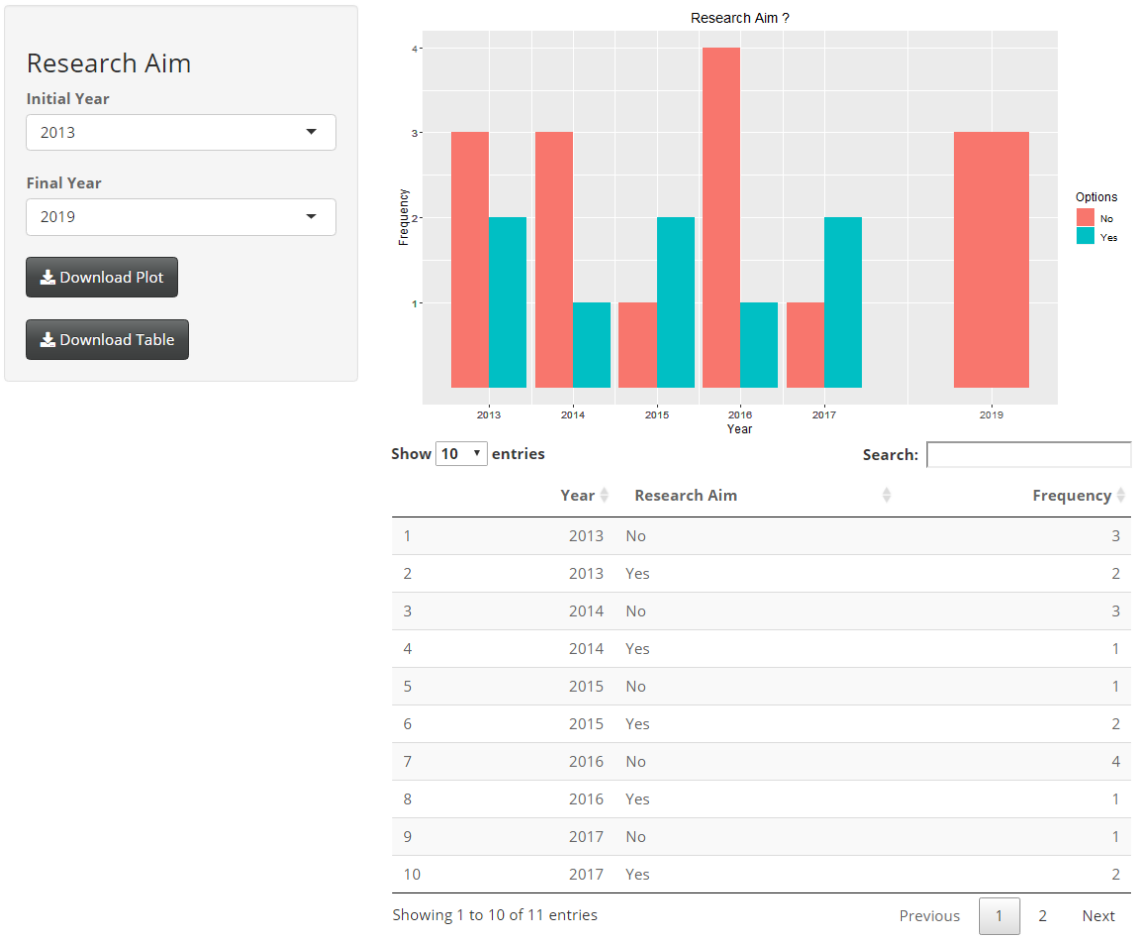


***Fig.49**: Example of the analysis done on the existence of the research aim.*

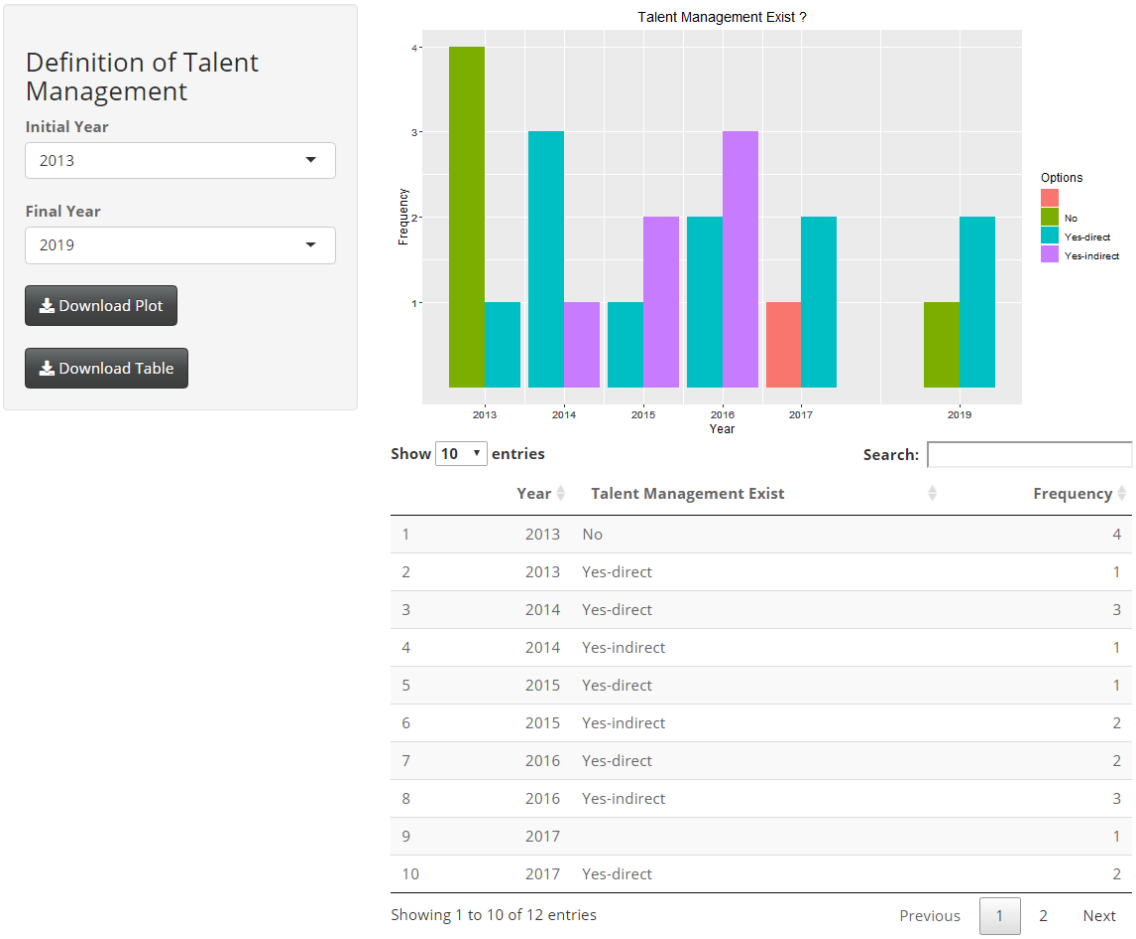


***Fig.50**: Example on the analysis done on the existence of the talent management.*

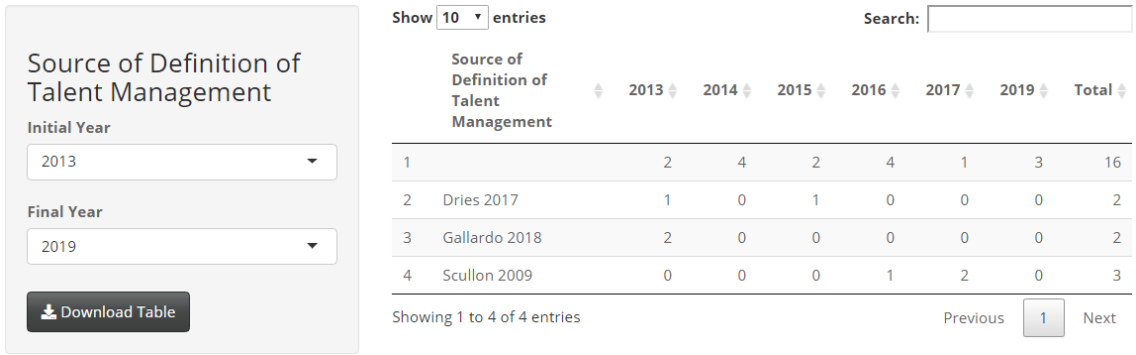
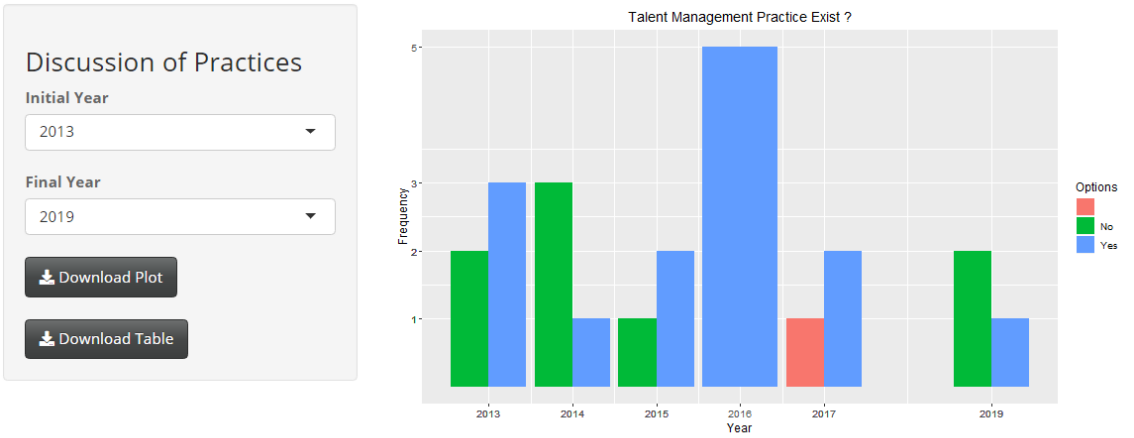


***Fig.51**: Example of the analysis on the source of the definition of the talent management.*

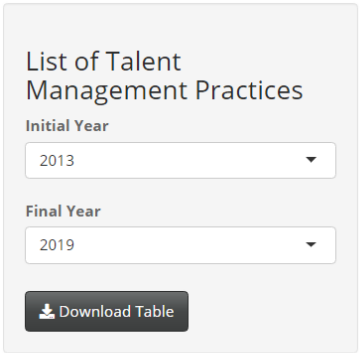


Show 10 entries Search:

	Year	Talent Management Practice Exist	Frequency
1	2013	No	2
2	2013	Yes	3
3	2014	No	3
4	2014	Yes	1
5	2015	No	1
6	2015	Yes	2
7	2016	Yes	5
8	2017		1
9	2017	Yes	2
10	2019	No	2

Showing 1 to 10 of 11 entries Previous 1 2 Next

***Fig.52**: Example of the answers to the question: "Does the talent management practice exist?".*



Show 10 entries Search:

	List of Talent Management Practices	2013	2014	2015	2016	2017	2019	Total
1		2	3	1	0	1	2	9
2	Identification	3	1	0	0	0	1	5
3	Selection	1	1	0	0	0	1	3
4	Performance	0	0	1	4	0	0	5
5	Retain	0	0	1	4	0	0	5
6	Training	0	0	1	0	0	0	1
7	Promotion	0	0	0	1	2	0	3

Showing 1 to 7 of 7 entries Previous 1 Next

***Fig.53**: Example of the analysis done on the types of talent management practices.*