Journal of Innovation Management
JIM 8, 2 (2020) 1-10
HANDLE: https://hdl.handle.net/10216/128300
DOI:https://doi.org/10.24840/2183-0606_008.002_0001

Phrasing the giant: on the importance of rigour in literature search process

João José Pinto Ferreira
jjpf@fe.up.pt | INESC Technology and Science, Faculty of Engineering, University of Porto, Portugal
Anne-Laure Mention
anne-laure.mention@rmit.edu.au | RMIT University, Australia
Marko Torkkeli
marko.torkkeli@lut.fi | Lappeenranta University of Technology, Finland

> *Literature is the noblest of all the arts. Music dies on the air, or at best exists only in memory; oratory ceases with the effort; the painter's colors fade and the canvas rots; the marble is dragged from its pedestal and is broken into fragments.*
>
> Elbert Hubbard

At a very early age, we start to develop a sense of playfulness. We touch things, we build things, we break them apart. Soon after we begin to utter words. We babble, we squeal, we try to imitate. Music begins to inform our bodily movements. What develops last and continues to develop throughout our waking lives is connections of words. The essential and characteristic features of words used to describe things within and around us are the hardest to grapple with. The same word can be expressed in different ways and could mean different things in different contexts. Literature, being the written expression of words in its various forms, has progressively shaped our world view.

Liberal news outlets around the world have been stressing recurrently that words matter, as the imagination of some politicians' is set loose and boundaries to what one may say seem not to exist. However, despite this current societal struggle to adhere to facts, namely amid the current pandemic, science has remained irreducible in its systematic approach supported by the scientific method where facts and doubt do co-exist as a process towards the discovery and construction of new knowledge.

As the years go by, time flies, and suddenly, as a grownup doing research, one needs to select keywords in order to find the right information to extend our world view. We have all been there:

ISSN 2183-0606
http://www.open-jim.org

sitting in front of our digital devices and rapidly searching for information using keywords, say "COVID-19. First, you keep it simple with just one keyword, but soon realise that the results return some relevant articles. So, you extend the search terms, say "COVID-19 Europe", but still the results are wide and varied. So, you limit the search further "COVID-19 Europe Statistics". Now you have some relevant information. You randomly pick few articles (perhaps those with catchy titles), consult them, and be happy that in some arbitrary way you added value to your knowledge store. This is fine for everyday searches but is not what is preferred or generally accepted when it comes to literature search composing one's research paper or thesis. Those papers should indeed be the right ones, as the objective is to write a paper or thesis. Get the wrong papers, or have essential papers missing and trouble is just around the corner! So, we need the right keywords, "As any good library or information worker knows the accurate and consistent application of keywords can serve to enhance the content representation and retrieval of literature." (Grant, 2010, p.173). Reviewing the literature represents an "essential first step and foundation when undertaking a research project" (Baker, 2000, p. 219). It is well established that literature search seeks to reveal relevant information on a topic and make a contribution towards scientific rigour (Baker, 2000; Cooper, 1998; Garfield, 1977). Rigour is achieved when the search process effectively avoids the investigation and already well researched topic and allows for composition of extant knowledge base.

Still looking at words, scientific literature can be analysed using several techniques. These may help us "understand global research trends or see links and patterns amongst scientific documents" (Isenberg et. al., 2016), and examples of these techniques are co-citation analysis, co-word analysis, co-author analysis, word frequency analysis. The new digital sources have enabled researchers to count words based on proximity of their appearance in a text (e.g. Nicholson, 2012; Guldi, 2012), visualise the results using Ngrams or word clouds (e.g. Holmes, 2016) and even depict the strength of disciplinary networks or the extend of a topic (e.g. Randhawa, Wilden & Hohberger, 2016). However, will a novice researcher have to cope with these techniques to start his/her research? Surely not! So, how should we do it?

Setting the Scene

The following paragraphs outline an iterative method that may be used by both the novice and the seasoned researcher in the process of finding the keywords that best fit the search for the information they want.

For the sake of this illustration, let us start by defining a scenario:

- We do not know exactly which keywords to use.
- It is our first time looking up information in this knowledge domain, and, as a result, we just have a feeling about the broad keywords.
- We want to be able to explore; we want to try alternative paths and do it efficiently.
- We would like to have an easy way to cope with the overwhelming amount of information available.

For the sake of this scenario, let us pretend we want to look into the literature in the area of business, entrepreneurship, and innovation.

ISSN 2183-0606
http://www.open-jim.org

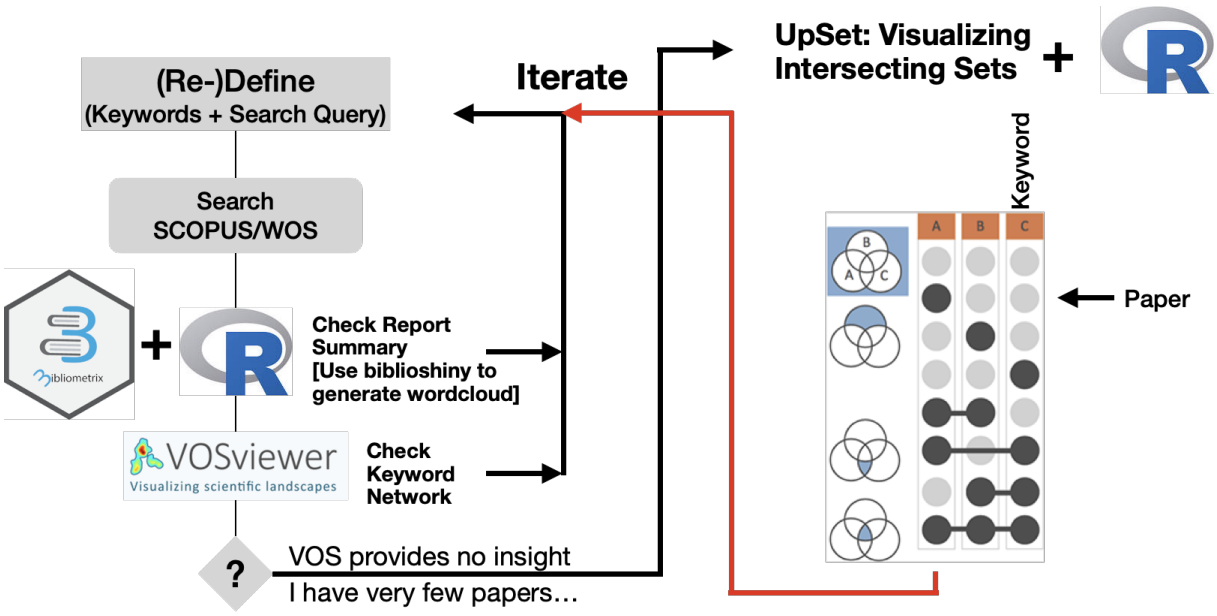


Fig. 1. The keyword exploration process (uses logos for the different tools and the picture extracted from the UPSet website that was adapted to explain the intersecting sets of keyword / paper)

The Approach

When we are looking for literature in a database such as SCOPUS or WoS, we define the keywords, we define the search query which is a specific combination of keywords to be used in the search and, as results, the databases will provide us with a list of hundreds or even thousands of records. So, what do you do? Typically, one would output the data to a Microsoft Excel worksheet and have a look at the records, one at a time. This is very slow and time consuming and the question one may ask is, are there tools out there that could help us with this job?

A natural way of thinking about this issue is, it would be great if we could visualize it! Can we do it? If so, how?

The first thought was to look for tools used to support research in the area of bibliometrics. This search revealed two most interesting and powerful tools: Bibliometrix (Aria & Cuccurullo, 2017) and VOSviewer - Visualization of Similarity (Van Eck & Waltman 2007). Another tool was added latter, the so-called "UpSet: Visualizing Intersecting Sets" (Lex et.al. 2014).

Fig. 1 outlines the iterative process to be detailed in the next section. The objective is to start with a search query, visualize the outputs of that query and decide if it represents what we are looking for or, if not, pick up new keywords that may be worth exploring. The process starts all over again, until we reach a point where we feel that we got what we want.

The tools used are the following:

- R is a free software environment for statistical computing and graphics. R is a very powerful in handling huge amounts of data.

ISSN 2183-0606
http://www.open-jim.org

- Bibliometrix is a wonderful tool for handling and processing massive amounts of data. Bibliometrix is a R Package. This means the user may build on other R functions, Packages and scripting possibilities to enhance functionalities and automate frequent tasks. It's fast and easy to export whatever we like to VOSviewer.
- VOSviewer is a wonderful tool for visualization. VOS imports bibligraphic data (in this process, we use data that we from R), and enables different types of analysis, involving for example, keywords (e.g.: co-occurrence) and references (e.g.: co-citation/bibliographic coupling). Graphics are easy to generate, and navigate.
- UpSet: Visualizing Intersecting Sets is an R Package. This means that we may visualize data available in R, we just have to transform this data into the correct format for generating the graphics. This tool plots a graphic built from a sparse matrix where in each line has a "1" if the keyword (column) occurs for that paper corresponding to that line.

The full process, Step-by-Step

This sequence of steps builds on the assumption that Bibliometrix and UpSet are installed in the R platform. Also install VOSviewer in the computer.

Step 1. Load the libraries

library(bibliometrix)

library(UpSetR)

Step 2. Do the search in SCOPUS or WOS. In this example we did the search in SCOPUS using the following query:

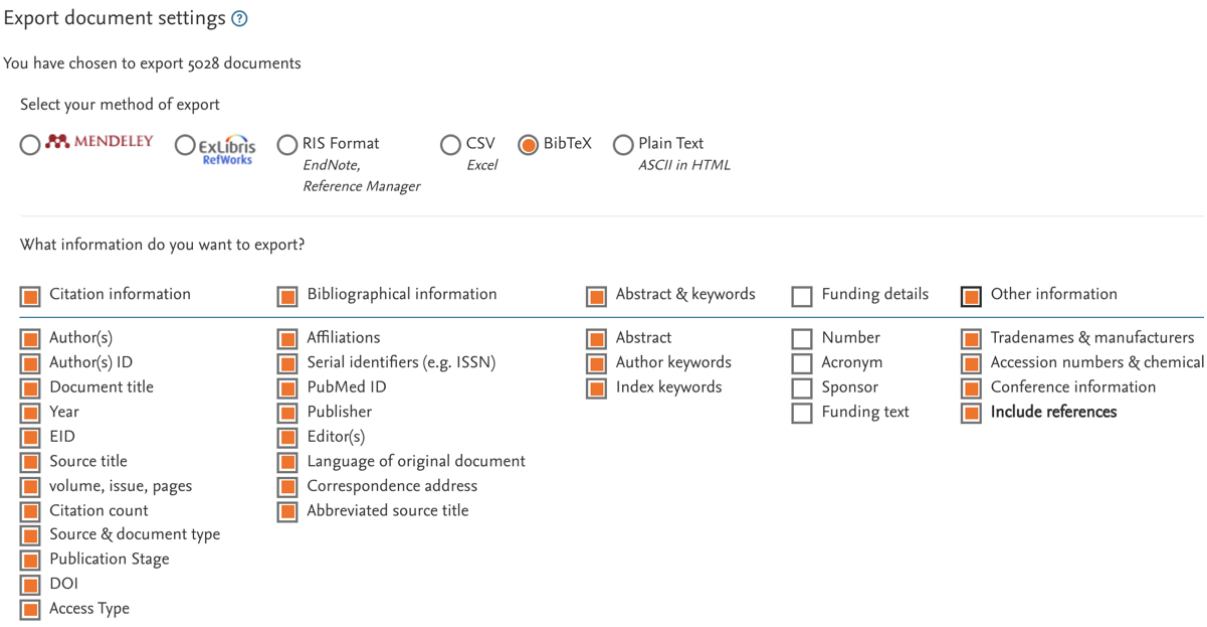


Fig. 2. SCOPUS Export method.

KEY (business AND entrepreneur* AND innovation)

The results were then exported in the BibTeX format as illustrated in Fig. 2. The file was saved as “scopus.bib”.

Step 3. Import all records to R using the Bibliometrix functions and convert the imported data structure to a dataframe. The result is saved in the vaiable M_SCOPUS0.

D_scopus0 <- readFiles("scopus.bib")

M_SCOPUS0 <- convert2df(D_scopus0, dbsource="scopus",format="bibtex")

Remark: if we import records from both SCOPUS and VOS, Bibliometrix provides the function mergeDbSources that merges the dataframes from the two sources by removing duplicates.

Step 4. Have a first glimpse into the contents by using the function biblioAnalysis. The actual results may then be checked using the summary function. One may also plot to picture the numeric results.

results <- biblioAnalysis(M_SCOPUS0)

summary(results, k=10, pause=F, width=130)

plot(x=results, k=10, pause=F)

The summary provides interesting information, namely the number of papers by “Document type” (Article, Book Chapter, ...), the “Annual Scientific Production”, the “Most Productive Authors”, the “Top manuscripts per citations”, the “Most Relevant Sources”, and the “Most Relevant Keywords”.

Step 5. We would now like to visualize what we have in the database. Our proposal is to use VOSviewer. The easy way to do it is to just export the dataframe to a Comma Separated Values (CSV) file as a text document using the function write.csv:

write.csv(M_SCOPUS0,"for_VOS.txt", na="")

Note: na="" replaces the not available (NA) contents to null char.

Step 6. In the application VOSviewer, the file "for_VOS.txt" should be opened as a bibliographic bibliographic data. Then we may choose to generate a map of co-occorrence of “All keywords”,”Author keywords” or “KeyWords Plus”. For the sake of this example, we will focus on “Author keywords”.

Fig. 3 illustrates the resulting keyword co-occurrence map. In the process, the user has to select the threshold for the minimum number of occurrences of a keyword.

This map gives an interesting perspective of the knowledge stores in the database. It also shows that authors’ used in some cases the keyword “business model” and in other cases “business models”.

Step 7. Let us now suppose we want to dig deeper and explore the word business model. We could repeat the whole process with the new search query which returned 30 documents:

KEY (business and entrepreneur* and innovation AND "business model")

ISSN 2183-0606
http://www.open-jim.org

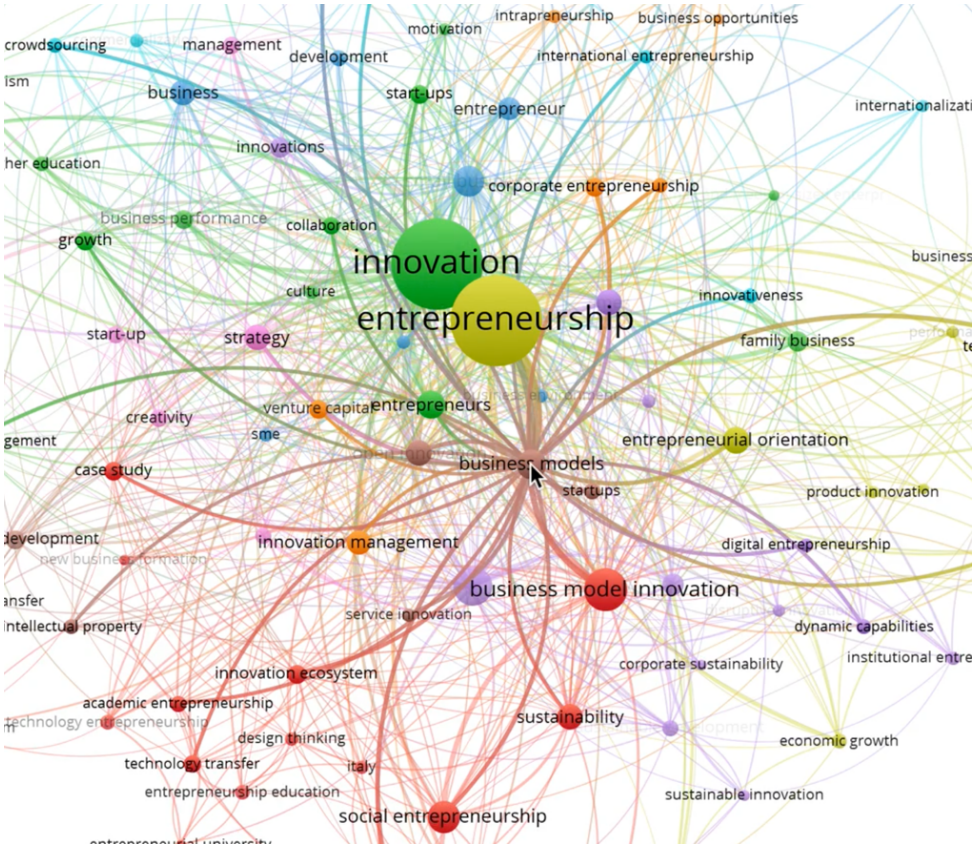


Fig. 3. Keyword co-occurrence map. The selection of one node highlights the node connections for easier visualization.

An alternative would have been to use an R command to create a new dataframe from the original M_SCOPUS0 by just selecting the records that matched the keyword "business model". This means that, with some practice, one is able to load a big databased into R and manipulate and generate new ones by selecting records matching a specific condition. This accelerates the process and saves us regular visits to SCOPUS and WOS.

Fig. 5 illustrates to new keyword map and it seems clear not useful information may be extracted regarding the contents of the selected papers. So, we go to step 8.

Step 8. This step is about getting further detail from the papers we have in the database. To this end we will use the R package UpSet: Visualizing Intersecting Sets and also a bibliometrics function cocMatrix that generates the Co-occurrence matrix. The command below generates the Co-occurrence matrix for the Field "DE" that corresponds to the Author Keywords. The second command converts the matrix to a dataframe and the third, removes the line labels (paper reference).

co_de <- cocMatrix(M_SCOPUS0, Field = "DE", type = "matrix", sep = ";", binary = TRUE)

co_de.df <- as.data.frame(co_de)

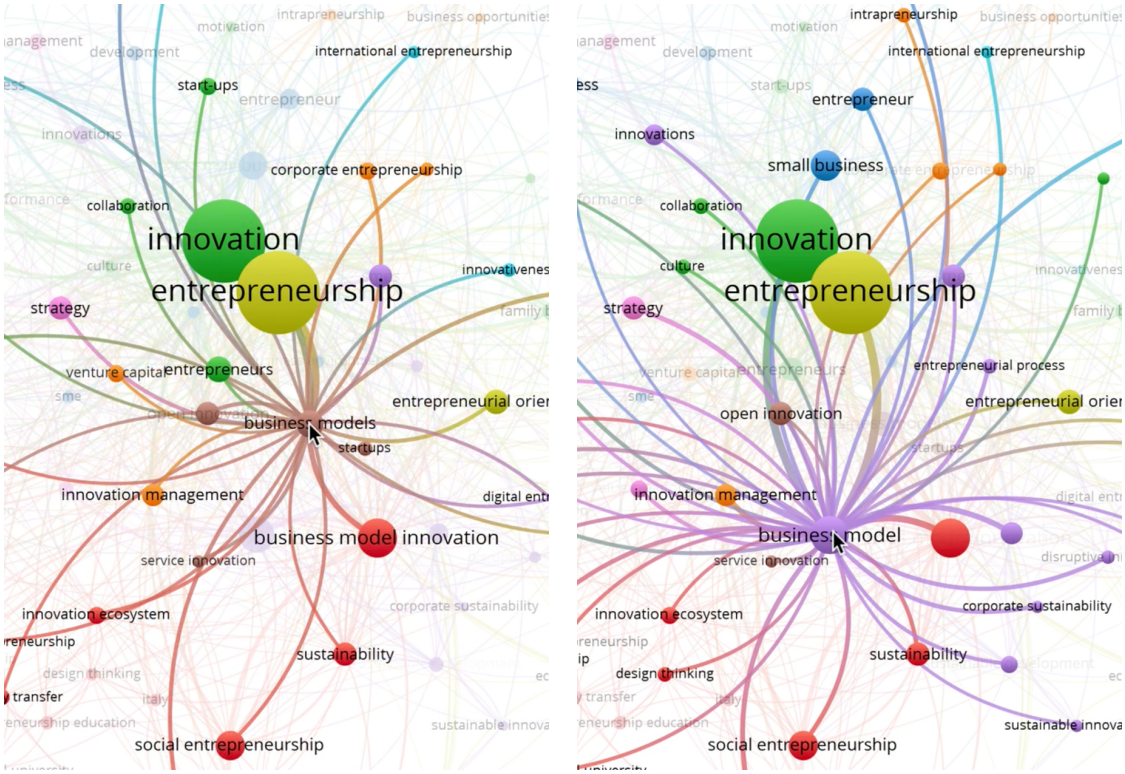


Fig. 4. Showing the keywords "business model" and "business models"

row.names(co_de.df) <- NULL

All is now set to run the UpSet command and generate the visualization of the intersecting sets:

upset(co_de.df, nsets = 40, nintersects = 60, mb.ratio = c(0.3, 0.7), order.by = c("degree","freq"), decreasing = c(TRUE,FALSE), show.numbers = TRUE)

This visualization provides an interesting insight into what one has in the selected papers. Remember that we are exploring what we have, and we may realize that the keyword "social franchise" seems to be interesting and we see that appears together with the other keywords highlighted in the red circles in one paper. In R, using the RStudio (ref) interface it is quite easy to define a filter with the keyword "social franchise". In less than a minute have all the information about the paper we need, and we can retrieve it from google scholar (provided you are within a VPN that grants the needed permission).

Conclusion

The process herein described, may be used to find the adequate keywords to start any research. This a possible approach that may be used to handle the overwhelming amount of information one gets whenever looking for any bibliographic material. The process allows exploration of the

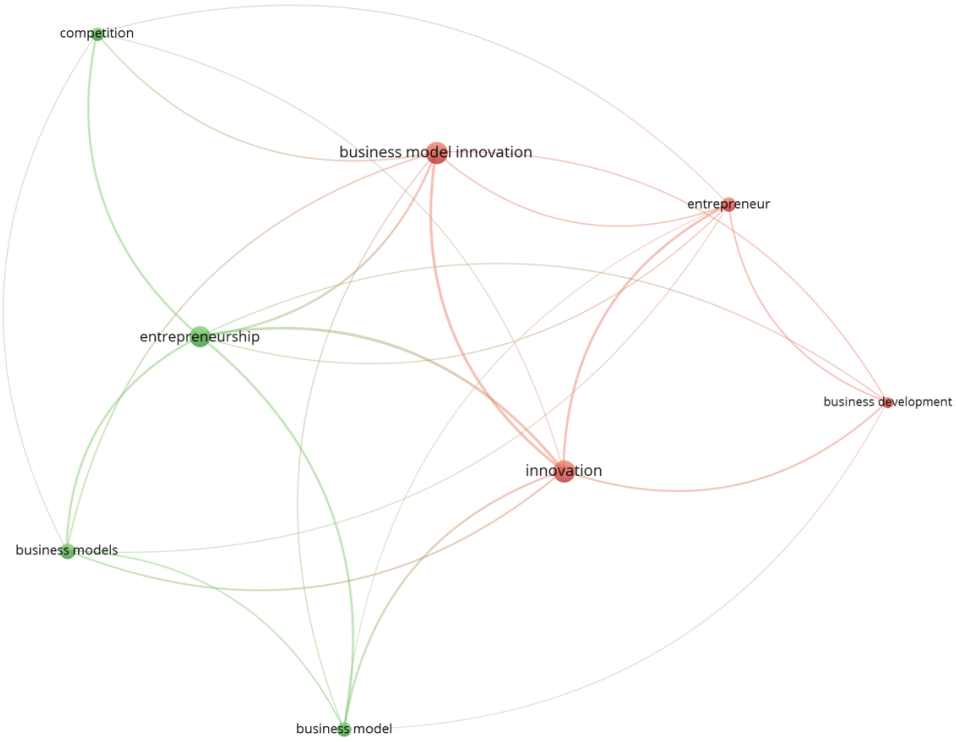


Fig. 5. The new map of keywords

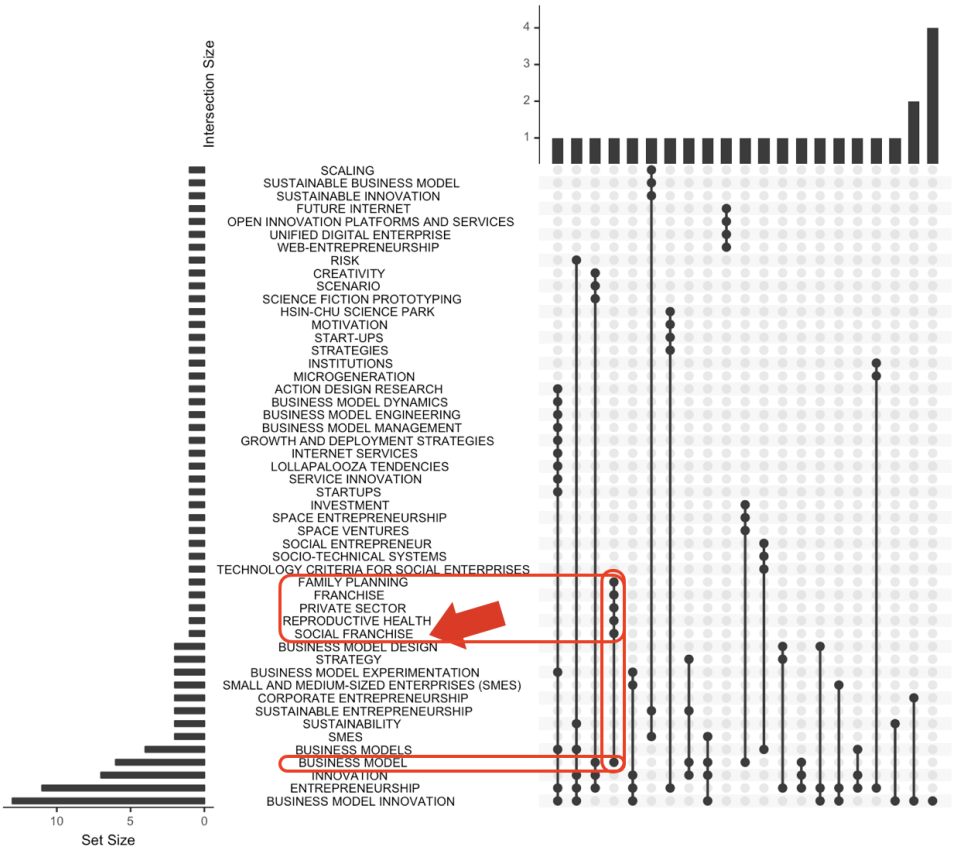
keyword space and use successive zoom's into areas that may seem interesting, exploring them fast, and getting back up if what we find does not seem promising.

The above tools, both VOSviewer and Bibliometrix provide other great functionalities such as co-citation and bibliographic coupling analysis (Boyack & Klavans, 2010). For a novice researcher, as soon as the right keywords are found, the co-citation analysis will provide them with the most co-cited papers by the documents stored in the database.

Remark: This co-citation analysis may not be fully correct, in the sense that, in order to do it properly, one would need to make sure that cited papers referring to the same document have exactly the same text. The issue is that, different papers may refer to the same cited paper using a slightly different text, and the system will look at them and consider that these papers are not the same.

It seems to be quite useful to have a systematic approach that enables the exploration of the papers extracted from a repository, even before reading any of them. This exploration should unfold, keeping in mind the actual final goal. The researcher will want to know which way to go! Is research he/she is thinking about needed? Should one read the articles in full? Considerations of relevance and significance of the phenomenon should always remain central to the process. These will eventually guide the decision on whether to proceed with the research.

ISSN 2183-0606
http://www.open-jim.org



HEALTH POLICY AND PLANNING; **17**(2): 121–130 © Oxford University Press 2002

Review article

Franchising of health services in low-income countries

DOMINIC MONTAGU
DrPH Candidate, University of California at Berkeley, USA

Grouping existing providers under a franchised brand, supported by training, advertising and supplies, is a potentially important way of improving access to and assuring quality of some types of clinical medical services. While franchising has great potential to increase service delivery points and method acceptability, a number of challenges are inherent to the delivery model: controlling the quality of services provided by independent practitioners is difficult, positioning branded services to compete on either price or quality requires trade-offs between social goals and provider satisfaction, and understanding the motivations of clients may lead to organizational choices which do not maximize quality or minimize costs. This paper describes the structure and operation of existing franchises and presents a model of social franchise activities that will afford a context for analyzing choices in the design and implementation of health-related social franchises in developing countries.

Key words: franchise, social franchise, family planning, business model, private sector, reproductive health

Fig. 6. Visualizing intersecting sets (sets of keywords appearing together in papers) (above); The selected papers and the keywords found in the intersecting sets (below)

Wishing you a great experience in exploring the literature in the search for the best words,

Innovatively yours,

The Editors.

References

Aria, M. & Cuccurullo, C. 2017, 'Bibliometrix: An R-tool for comprehensive science mapping analysis', *Journal of Informetrics*, 11(4), pp 959-975, Elsevier, DOI: 10.1016/j.joi.2017.08.007

Baker, M.J., 2000, 'Writing a Literature Review', *Marketing Review*, 1 (2), 219-247.

Boyack, K. W., & Klavans, R., 2010, 'Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?', *Journal of the American Society for information Science and Technology*, 61(12), 2389-2404.

Cooper, H.M., 1988, 'Organizing knowledge syntheses: A taxonomy of literature reviews', Knowledge in Society, 1, 104-126.

Garfield, E., 1977, 'Proposal for a new profession: scientific reviewer', Essays of an Information Scientist, 3, 84-87.

Grant, M. J. 2010, 'Key words and their role in information retrieval', Health Information & Libraries Journal, 27(3), 173-175.

Guldi, J., 2012, 'The history of walking and the digital turn:

Stride and lounge in London', The Journal of Modern History, 84, 116–44.

Holmes, D., 2016, 'A new chapter in innovation', Nature, 533(7602), S54-S55.

Isenberg, P., Isenberg, T., Sedlmair, M., Chen, J., & Möller, T. 2016, 'Visualization as seen through its research paper keywords', IEEE Transactions on Visualization and Computer Graphics, 23(1), 771-780.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. 2014, 'UpSet: visualization of intersecting sets', IEEE transactions on visualization and computer graphics, 20(12), 1983-1992.

Nicholson, B. 2012, 'Counting culture; or, how to read Victorian newspapers from a distance', Journal of Victorian Culture 17, 238–46.

Randhawa, K., Wilden, R. and Hohberger, J., 2016, 'A bibliometric review of open innovation: Setting a research agenda', Journal of Product Innovation Management, 33(6), 750-772.

Van Eck, N.J., & Waltman, L. 2007, 'VOS: a new method for visualizing similarities between objects'. In H.-J. Lenz, & R. Decker (Eds.), Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society (pp. 299-306). Springer.