How undertake a literature review through bibliometrics. An example with review about “user innovation”

Blanca de-Miguel-Molina, María de-Miguel-Molina and José Albors

Abstract

In this paper we explain how to carry out a literature review using Bibliometrics. We use as example the literature review about user innovation, and we explain every step followed, from the query in Web of Science database, until the visualisation of networks, and the use of Social Network Analysis to determine citations and keywords more important in the literature of this field. We include information about the different software and tools used in the paper for cleaning data, visualise them and analyse. We explain the importance in the queries and how to improve them to better incorporate all the literature in a field.

Keywords: user, innovation, Bibliometrics.

Introduction

Bibliometrics is a method broadly used to draw the big picture (Porter et al. 2002) in a literature review. It start with the definition of questions to be answered, which include some question like Who, What, Where, When, With Whom (Börner and Polley, 2014). Who refers to authors, What to keywords, Where to countries or other geographical locations, When to years or periods defined, and With Whom to the cooperation in research, which can indicate authors or affiliations.

Social Network Analysis (SNA) is a methodology very used in Bibliometrics to determine the importance of keywords, authors and citations in networks formed by these types of nodes. Among the measures form SNA utilised for determining the most important nodes are eigenvector centrality and betweenness. Eigenvector centrality denotes important nodes connected with other nodes that are also important (Newman 2010), while betweenness indicates that if one node were not there, literature on that path would not have been developed.

Concerning the field we have selected to applied the literature review, Von Hippel et al. (1999) pointed out that many comercial products were initially thought or prototyped by users rather than manufacturers, and that such products tended to be developed by “lead users”, which defined as “companies, organisations, or individuals that are well ahead of marked trends and have needs that go far beyond those of the average user”. From this perspective, examples about involvement of users on innovation has been explained for cases like medicine (von Hippel et al. 1999) or computer chip industry (Thomke and von Hippel 2002).

Literature review

The first step in the literature review consist on making some queries in databases of scientific literature. This first step needs to be repeated some times until we obtain the query which would include all the papers (or patents) in a field. We will show how to improve queries. In our example, we have repeated the query one time, basing the second query in a figure like Figure 2, in which we can find all the keywords that refer to concepts in the field we review.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0)
To analyse literature related to users in innovation, we made the following first queries in the Web of Science database on April 2015:

a) “user led innovation” OR
b) “end user innovation” OR
c) “user driven innovation”

These queries were undertaken in topic, that is, title, abstract and keywords. Results were 64, which included articles and proceedings. In this step, because we usually work with much more data, we select only papers.

Once we have the results in Web of Science, we download them in text file to be imported in other software. For cleaning the data we use the software for tech-mining VantagePoint (the Search Technology, Inc. USA). It is a commercial software which allows us to work with thousand of data. There exist free software, but we use them for visualisation and some Social Network Analysis, and some results can be obtained from the Web of Science query directly. However, our experience is that cleaning the data with VantagePoint is more secure.

In order to detect if the queries we made were correct, we cleaned the keywords and we elaborated Figure 2, which represents the keywords which have been used by authors, when they appear first, and which are more important. We can observe in the figure the keywords that authors have used to express users participation on innovation. We compare the keywords with our first queries, in which we took into account the terms “user led innovation”, “end user innovation” and “user driven innovation”. However, in Figure 2 we observe other terms like “lay users”, “user involvement”, “lead-user innovation”, “co-creation”, “co-design”, “co-development”, “co-innovation”, “collective creativity”, “inclusive design”, and “living lab”. Therefore, the next step would be to repeat the queries including more terms.

Figure 2. First queries. Evolution in the keywords used by authors in relation to participation of customers on innovation. Cleaning and visualization with software VantagePoint
In the second queries we undertook, we tried to include more terms, but we wanted they referred to innovation in all the cases. The queries were:

a) “user innovat*” OR  
b) “customer innovat*” OR  
c) “consumer innovat*” OR  
d) “lay user innovat*” OR  
e) “user-led innovat*” OR  
f) “lead-user innovate*”

In this case, in which we have made some changes in the queries, results obtained are 371 articles. We clean keywords and we obtain Figure 3, which gives us more information about the terms that literature has used to refer to users innovation. With the two queries we would elaborate the final query, which would be:

a) “end user innovation” OR  
b) “user driven innovat*” OR  
c) “user innovat*” OR  
d) “customer innovat*” OR  
e) “consumer innovat*” OR  
f) “lay user innovat*” OR  
g) “user led innovat*” OR  
h) “lead user innovat*” OR  
i) “user entrepreneurship” OR  
j) “community based innovation” OR  
k) “collaborative innovation” OR  
l) “living lab” OR  
m) “user co-creation” OR  
n) “user co-design” OR  
o) “user co-innovation” OR
How undertake a literature review through bibliometrics. An example with review about “user innovation”

Once we obtain the results on the third query, we will follow the next steps:

1. Elaborating a figure which presents the evolution in the number of works published. This step is useful for dividing in periods the results and elaborate all the analysis based in different periods.
2. We obtain the countries involved in the development of the field.
3. We obtain the keywords used by authors in their works.
4. We analyse the keywords with Social Networks Analysis, which gives us how important is every keyword in relation to others. This analysis can be made through the use of software like UCINET6 (Borgatti et al. 2002) or Gephi (Bastian et al. 2009). We usually use UCINET6 to transform matrixes in excel format to networks in net format.
5. We identify the literature framework, that is, the references more cited by authors in the field.
6. We use SNA to identify the central references.
7. We analyse the institutions involved in the increase of papers in the field.

Conclusions

In this paper we have carried out a literature review through the use of Bibliometric methodology to explain the different steps authors follow in undertaking reviews with this methodology. To explain it, we have used an example with a query based on users and innovation.

Explanation have started with the queries in the Web of Science database, to continue to cleaning of data, visualisation and analysis.
Variables included in the example were countries, keywords, citations and author’s affiliations. All data were cleaned through the software VantagePoint (The Search Technology, Inc.), while visualisations by years were made through this software and VOSviewer was used for networks.

Analysis of data, especially analysis of networks was made using Social Networks Analysis (SNA), through the software UCINET6 and Gephi. This analysis allowed us to indicate the keywords and references more important.

To obtain a precise literature review is necessary to repeat some of the steps explained in this paper. In our opinion, the best way consist on taking the picture of keywords, observe which keywords were forgotten in the queries, and repeat the query adding these keywords.

References


