HAL open science

Assessment of the applied orientation of a researcher's production: An informetric approach based on a content analysis

Ivana Roche, Nathalie Vedovotto, Claire François, Dominique Besagni, Marianne Hörlesberger, Dirk Holste, Edgar L. Schiebel, Pascal Cuxac

To cite this version:

Ivana Roche, Nathalie Vedovotto, Claire François, Dominique Besagni, Marianne Hörlesberger, et al.. Assessment of the applied orientation of a researcher's production: An informetric approach based on a content analysis. Collnet Meeting 2012, Oct 2014, Séoul, South Korea. hal-00956450

HAL Id: hal-00956450

https://hal.science/hal-00956450v1

Submitted on 17 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Assessment of the applied orientation of a researcher's production: An informetric approach based on a content analysis[1]

Ivana Roche(*), Nathalie Vedovotto(*), Claire François(*), Dominique Besagni(*), Marianne Hörlesberger(**), Dirk Holste(**), Edgar Schiebel(**) , Pascal Cuxac(*)

*ivana.roche@inist.fr; nathalie.vedovotto@inist.fr; claire.francois@inist.fr; dominique.besagni@inist.fr; pascal.cuxac@inist.fr
INIST-CNRS, 2 allée du Parc de Brabois, 54519 Vandoeuvre-les-Nancy Cedex, France

**marianne.hoerlesberger@ait.ac.at; dirk.holste@ait.ac.at; edgar.schiebel@ait.ac.at
AIT, Austrian Institute of Technology GmbH, Donau-City-Strasse 1, 1220 Vienna, Austria

Extended abstract

One way of making the distinction between pure and applied research was introduced by Donald Stokes (Stokes, 1997) who defined a two dimensions chart, "the Pasteur's Quadrant". It is a label given to a class of scientific research developments that both seek fundamental understanding of scientific problems, and, at the same time, seek to be eventually beneficial to society. The works of Louis Pasteur, a French chemist and physicist, pioneer of the microbiology, are thought to exemplify this type of method, which bridges the gap between "basic" and "applied" research. The Pasteur's Quadrant characterizes three distinct classes of research:

- pure basic research, illustrated by the work of Niels Bohr, early 20th century atomic Danish physicist;
- pure applied research, exemplified by the work of Thomas Edison, North-American inventor and businessman;
- use-inspired basic research, described as "Pasteur's Quadrant".

A classic solution to determine the applied orientation of the works of a researcher is to search for the possible patents in the development of which he(she) took part (e.g. Glänzel, 2003; Moed, 2004; Glänzel, 2011). Usually, the submission of a patent is the outcome of an initiative of technology transfer and can be indeed considered as the practical fulfilment of research works presenting determinedly applied characteristics. Another possible approach consists on directly examining the researcher's works published in the S&T (scientific and technological) literature and characterizing their content in applied or fundamental.

In the context of a previous European project (Holste, 2012), we modelled this two criteria by means of indicators calculated as:

- the enumeration of the patents in which the researcher has contributed. This indicator gets integer values included in the interval [0, ∞ [. Unfortunately, the number of patents is often very low, which involves a lack of accuracy of the related indicator.

[1] The authors wish to acknowledge the support of the IDEAS specific programme of the EU's 7th Framework Programme for Research and Technological Development in which context this work is partially accomplishing (DBF project - reference n° 240765 - within the Coordination and Support Actions).

- the ratio of the researcher publications appearing in journals which content is categorized as applied. This indicator gets real values between 0 and 1.

For both indicators we guessed that the higher the value, the more asserted is the applicability of the researcher works.

Even if such an approach could sound pragmatic, it conceals a few weaknesses in particular because of the journals categorization step. Indeed, by using a binary categorization "applied or fundamental", it seems *a priori* easy to automatically transpose the journal's category to all the articles that are published in it. A first difficulty appears when we have to determine the criteria defining this binary categorization of the S&T journals. If works have been done supplying hierarchical tree classifications, more or less detailed, of the scientific domains, the problem persists when we wish to determine which ones can be determinedly considered applied or fundamental.

In addition, the category of a journal can be not unique but variable, according to the scientific domain of each researcher that publishes his(her) work in it. Let us consider, for instance, the Biology domain as *a priori* fundamental. All the journals classed in this domain get then the category "fundamental", as well as all the articles published in them. Yet this does not always correspond to reality. Indeed, if this remains true for the biologists' publications in these journals, that of an IT specialist who would bring a development software to Biology should receive the "applied" category.

And so in this work, instead of examining directly the researcher's production, we propose to assess its applied orientation by analyzing the S&T literature citing his(her) publications. The citation expresses its exploitation, in different ways and at different degrees of importance, and we consider that this corpus of bibliographic records referring to the researcher's production represents a real and pragmatic information source about the utilization of his(her) former works by the scientific community in new developments. The content analysis approach applied to this corpus gives us the means to appreciate the degree of application of the researcher's scientific production. This way, we can detect potentially applicable works whose results could be integrated in more applied issues.

So, in a first step, we extract from a bibliographic database the list of the researcher's publications in S&T literature. This list enables us to determine the set of publications citing at least one of the extracted publications. This corpus can be considered as an image of the scientific landscape of citing papers that are based on the past work of this researcher.

Then a data mining step, based on NLP (natural language processing) techniques, is used to obtain an assisted indexing of the records by combining them with keywords. A clustering is then applied to this enriched corpus. The software tool uses a non-supervised and non-hierarchical clustering algorithm, the axial K-means, inspired by Kohonen's self-organizing maps formalism. This method considers the keywords as indicators of the content of bibliographic records, which in their turn are considered as indicators of the research themes. This first step is followed by a principal component analysis leading to a 2D-mapping of the clusters. Thematic networks emerge from the relations between clusters and, according to a geographical metaphor, build a map of the corpus research landscape.

At this stage, a scientific expert performs an analysis of the clusters in terms of their content and of their position and relations in the map. For this analysis, he(she) must adopt a particular point of view, in order to evaluate whether the content of a corpus is essentially applied or essentially fundamental: he(she) looks therefore at the content of each cluster by considering the bibliographic records title and keywords, in order to evaluate how much the content of the clusters can be considered as applied. We call this the level of "applicateness" of the cluster.

Finally, we weight the "applicateness" value of each cluster with the help of a parameter coming from the clustering results, in order to take into account the global context of the corpus research landscape represented by the map of thematic clusters and, in particular, by analyzing the links between them. The combination of the weighted "applicateness" of each cluster gives us the level of the applied orientation of the researcher's production.

In this paper, we will firstly detail the approach used to evaluate the applied orientation of a researcher's production. Secondly, we will describe the data processing, and finally we will, on one hand, present and discuss the obtained results for a few case studies and, on the other hand, look towards the future by proposing possible improvements of our methodology.

References

Glänzel W., Meyer M. (2003). Patents cited in the scientific literature: An exploratory study of "reverse" citation relations. Scientometrics, 58, pp. 415-428.

Glänzel W., Zhou P. (2011). Publication activity, citation impact and bi-directional links between publications and patents in biotechnology. Scientometrics, 86, pp. 505-525.

Holste D., Roche I., Hörlesberger M., Besagni D., Scherngell T., François C., Cuxac P., Schiebel E. (2012) submitted to Scientometrics.

Moed H.F., Glänzel W., Schmoch U. (2004). Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems. Kluwer Academic Publishers.

Stokes D. (1997). Pasteur's Quadrant - Basic Science and Technological Innovation, Brookings Institution Press.