



OWC 2020 Paper Submission - Science Forum

Topic 3 - Transition towards organic and sustainable food systems

OWC2020-SCI-1208

ORGANIC FOOD AND FARMING SCALING: A SEARCH STRATEGY TO IDENTIFY RELEVANT LITERATURE

Guillaume Ollivier* ¹, Servane Penvern¹

¹Ecodéveloppement, INRAE, Avignon, France

Preferred Presentation Method: Oral or poster presentation

Full Paper Publication: Yes

Abstract: While Organic food and farming (OFF) sector grows, this development is gaining interest in the scientific arena. As part of its new program on organic scaling (METABIO), INRAE wishes to monitor and analyze the evolution of the scientific literature on the conditions, determinants and consequences of this transition. But, building a search query, *ie* a combination of keywords to get relevant literature from a bibliographical database, is yet challenging on such emergent and fuzzy topic.

We present here the search strategy we developed based on both expert knowledge and an iterative and inductive process of lexical query improvement.

We show that our sophistication approach allows to capture literature with a good quality. Moreover, we detail the constraints weighting on such a strategy and the advantages of the iterative approach in the acquisition of knowledge on organic scaling.

Introduction: OFF is developing worldwide [1]. This movement out of the niche touches all the dimensions of the agri-food system at all scales and raises original scientific questions on agroecological transitions. Based on the assumption of a major development of organic farming and consumption in France, INRAE has launched a new scientific program called METABIO to study the consequences and the ways to support such a generalization. Indeed, in the last decades the OFF development has already been the subject of many academic publications allowing to identify and understand the various determinants and bottlenecks of organic scaling. This knowledge is useful for the OFF sector to design and evaluate different development scenarios. To contribute to the many tasks of the METABIO program (e.g. production of states-of-the-art and meta-analysis, research agenda, project consortia, information of researchers...), it is crucial to build a bibliographical query allowing to monitor the scientific literature on OFF scaling. But, this task is often underrated [2]. Since we postulate that a simple query does not produce satisfactory results, our purpose is to build a query that captures the many aspects of OFF scaling, going from the conditions, pathways and consequences from resources management to food consumption. The task is thus complex, particularly in the case of domains rapidly evolving, without stabilized lexicon or with polysemic phrasing [2].

Starting with a query we already built on OFF research [3], our challenge was thus to focus on papers related to scaling aspects. This requires to identify and assess the relevance of terms to be implemented in a search query. This contribution aims to describe the process we developed, to show its interest and discuss the issues of building such a query.

Material and methods: We used Web of Science (WoS), a major academic bibliographical database that enables a multiplicity of further analysis (e.g. citation analysis).

Many methods are suggested in the literature on search strategy namely on emergent, complex and interdisciplinary research domains [2]: citation-based searches, citation expansion from a core lexical query, Natural Language Processing... But these approaches are complicated to implement. We rather used a more parsimonious search strategy, mostly based on human control, rather than a statistically driven approach.

Figure 1: The iterative and inductive process of query building

The figure 1 describes the iterative and inductive process we developed. Based on articles we knew to be relevant on the topic, and those emerging during the process, we iteratively built a set of terms and combination of terms. Each candidate term, or combination, was assessed as regard to the *Precision* of the intermediary resulting papers, i.e. the ratio of relevant papers on a sample of 15 to 30 papers randomly selected among all the papers found. If too much noise is generated (*Precision* <85%), we used some tactics to be more domain-specific and narrow too generic or polysemic terms. The first one consisted in identifying specific phrases that can be added directly in the query (e.g. "change of practice" rather than the very polysemic word "change"). Another tactic consisted in using the boolean (AND, OR) or proximity operators (NEAR/x). For instance, the phrases "growth of the organic market" or "market growth" can be captured by the syntax (growth NEAR/2 market).

After many iterations and terms added to the query, we performed a global evaluation on the total corpus generated to measure Precision and Sensitivity (i.e. the ratio of relevant papers found among all the relevant ones we knew). As the total set of expected documents is unknown by definition, it should be noted that this knowledge is inevitably partial and based on relevant articles found during our iterations. Search strategy is still a trade-off between these two indicators: focusing on Precision allows for the exclusion of irrelevant papers but with the risk of omitting relevant ones and inversely, focusing on Sensitivity allows to capture most of the relevant papers but with the risk of also capturing irrelevant ones. After many iterations, we reached a "relevance plateau", ie the cost of improvement goes higher and higher for too little improvements.

Results and discussion: We know that our query still captures some irrelevant papers while it omits relevant ones. But at this stage, the estimated Precision is 89% and Sensitivity 93% (fig. 1), which is a good compromise. Our method of sophistication resulted in a search query composed of two crossing blocks of terms. This sophistication greatly increases the volume of literature recovered. The [OFF] block, composed of 550 terms, returned 20,990 papers (oct. 2019), almost 2.5 more than a simple query such as ["organic farm*" OR "organic agriculture"] (n=8265) used in Sigmeier et al. [4]. Likewise, the [SCALING] block, adding 443 terms, returned 4001 papers while a simple query with the term "scaling" and its common derivation like "up-scaling" or "out-scaling" only returns 72 papers.

Our iterative method also allows a progressive delineation of the perimeter of the search topic. While we initially focused on scenarios of OFF massification, it seemed necessary to us to broaden the scope - from the description of past trends to prospective views on the conditions and determinants of this change, whether individual or collective.

The iterative process, thanks to our expertise and examination of intermediary results generated at each iteration, also allows the decomposition of OFF scaling into sub-topics (see examples in Tab. 1).

Table 1: Examples of some sub-topics composing the [SCALING] block.

Some of these sub-topics are simple to translate into a lexical query. For instance, the Conventionalization or the “feed the world” debates, use very specific and quite standardized lexicon. Sometimes, relevant papers are not eliciting scaling into terms. For instance, the paper of Buck et al. [5] never mentions “conventionalisation” although being the seminal paper of this sub-topic and thus is a good example of relevant but not found papers. Other sub-topic queries are more complex to build, because of their use of generic or polysemic terms. For instance, the [scaling] sub-topic mention terms like “growth”, “development” or “change” also used in the context of biological phenomena. The sophistication of the search query, ie adding and excluding search terms, needs thus a strong attention to the meanings of terms, to the way authors articulate them into language and to the explored literature.

Conclusion: The systematic search of relevant literature on OFF scaling is challenging due to the plurality of related sub-topics and the polysemy and non-specificity of the terms currently used. A large acceptance of the topic is proposed to capture a maximum of relevant papers. We developed an iterative process to refine the perimeter of the search topic. Without being totally free of noises and gaps, our expert-based approach and the immersion in the literature allowed us to better characterize and acquire a good knowledge of the literature on OFF scaling. The process is continuous and iterations are still performed namely with emerging papers to gain in precision and sensitivity. As perspectives, the corpus can now be better defined and be used to feed the INRAE program and reflections of sector stakeholders. We also suggest the research community to use standardized lexicon to affiliate their work to the “organic food and farming scaling” topic. Our perspective is to test some ex-post corpus cleaning methods before further bibliometric analysis (e.g. dynamics of publications, main authors and institution, thematic and/or co-citation analysis) allowing to better understand the various incentives and barriers of an OFF scaling.

References:

1. Willer, H.; Lernoud, J. (2019) *The World of Organic Agriculture. Statistics and Emerging Trends 2019*. FiBL and IFOAM Organics International.
2. Huang, Y.; Schuehle, J.; Porter, A.; Youtie, J. (2015) A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for ‘Big Data’. *Scientometrics*, 1-18.
3. Ollivier, G.; Bellon, S.; Penvern, S. (2011) *Thematic and citation structure dynamics of Organic Food & Farming research*, 3rd ISOFAR Scientific Conference, Neuhoff, D., Ed. Gyeonggi Paldang.
4. Siegmeier, T.; Blumenstein, B.; Mühlrath, D.; Möller, D. (2014) *Structure and development of scientific journal publications on organic agriculture: A scientometric review*, 4th ISOFAR Scientific Conference.
5. Buck, D.; Getz, C.; Guthman, J. (1997) From farm to table: The organic vegetable commodity chain of northern California. *Sociologia Ruralis*, 37, 3-20.

Image:

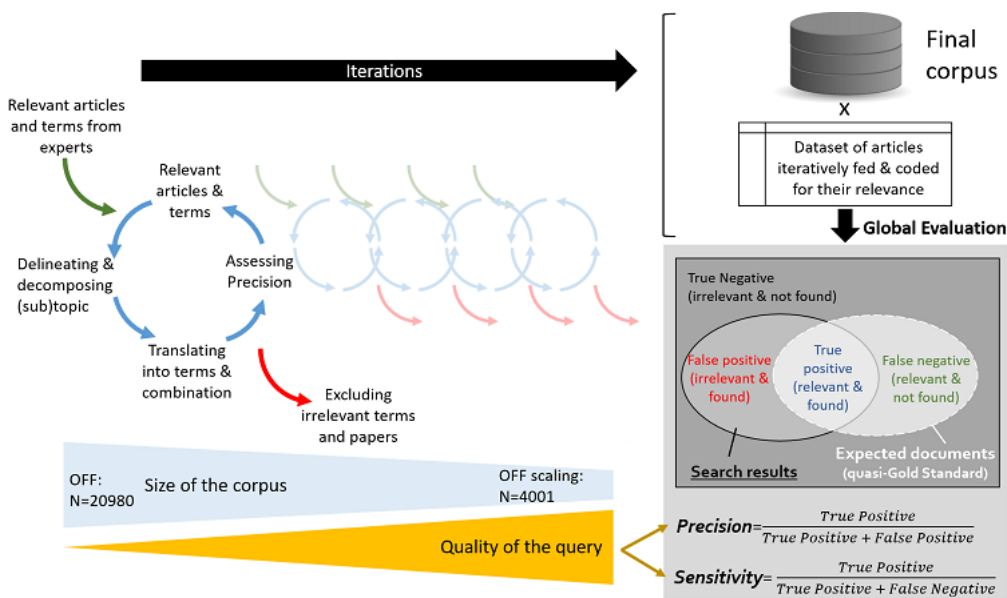


Image 2:

Sub-topic	Query extract	# papers	Precision
[scaling]: Generic terms related to change applied to organic qualified item, or to economic or behavioral processes	TS={(((upscal* OR "up scal*" OR [...] OR massification OR [...]) NEAR/2 (organic* OR demand OR consumption OR offer OR market* OR [...] determinant* OR barrier* OR lever* [...])) NOT ("vegetative growth" OR "plant growth" [...]))	1948	90%
[Transition/Conversion]: terms related to transition, adoption or conversion crossed with terms related to processes, mechanisms and determinants...	TS={ ((diffusion) NEAR/2 (innovati* OR technolog* OR econom* [...])) OR ("transition" [...] OR conversion* [...] OR adoption* [...]) AND (determinant* OR reasons OR "key factor*" OR willingness OR attitud* OR motivation* [...] OR barrier* [...] OR "government support*" [...]))	1962	90%
[Conventionalisation]: Hypothesis on the pattern of evolution of OFF developed after Buck et al., 1997 [5]	TS=(Conventionalization OR [...] "industrialization of organic*" [...])	111	100%
["Feed the world" debate]	TS=("food for the planet" OR "Feed* the world*" OR "feed a world population" [...] OR "global conversion" OR [...] OR "Feed the Region*" OR [...] "regional self-reliance" [...])	85	86%
[Prospectives]: considering horizon scanning of OFF development according to different methods	TS=((scenari* OR projection* [...]) NEAR/4 (development OR change OR growth [...]) OR "Delphi analysis" OR "foreseeable future*" OR "future* of organic*" [...])	260	95%
[Large conversion scenario]	TS=(("50% organic*" OR "100% organic*" OR "completely converted") NEAR/5 (world* OR region* OR nation* OR countr* [...]) OR "widespread conversion" OR "widespread adoption" OR "100% conversion" [...])	74	89%
Final corpus		4001 (19% of the OFF corpus)	89.2%

Disclosure of Interest: None Declared

Keywords: bibliometrics, information retrieval, organic scaling, query building, research planning