



## 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<sup>\* 1</sup>, Servane Penvern<sup>1</sup>

<sup>1</sup>Ecodéveloppement, INRA, Avignon, France

**Preferred Presentation Method:** Oral or poster presentation

**Full Paper Publication:** Yes

**Abstract:** As part of the construction of its new (meta)program on organic scaling, INRA wishes to be able to monitor and analyze the evolution of production and research communities of this topic. But, the capture of the relevant literature via a bibliographic query is challenging.

We thus propose a search strategy capturing the relevant literature on organic scaling. This method is based on both expert knowledge and an iterative and inductive process of lexical query improvement.

We show that our approach leads to a good quality corpus (Precision=89% and Sensitivity=90%). Moreover, we detail the constraints weighting on such a strategy and the advantages of the iterative approach in the acquisition of knowledge of sub-topics, and on the lexicon used by researchers to discuss organic scaling.

**Introduction:** Organic food and farming (OFF) is developing worldwide [1]. Based on the assumption that organic farming will be the majority in France in 2030, INRA has launched a new scientific (meta)program called METABIO to study the consequences and the ways to support such a generalization. To clarify the state of the art on this topic, we worked on a bibliographic query that should also allow us to identify, analyze and monitor the evolution of research communities, the theoretical frameworks used, and the scientific issues to target and design new research projects.

Since a simple query does not produce satisfactory results, our purpose is to build a query that captures the many aspects of OFF scaling. The construction of queries is a crucial, but often underrated, task and particularly challenging in the case of domains with rapid evolution, without stabilized lexicon or polysemic phrasing [2]. This contribution aims to describe the process and issues of construction of such a query.

**Material and methods:** We used the database Web of Science (WoS), a major academic reference database, to enable a multiplicity of further analysis.

To capture the literature on the topic of OFF scaling, we enhanced a query we already built to analyze the OFF research domain [3]. Our main contribution concerns the query dedicated to scaling aspects, i.e. the conditions, pathways and consequences of a development of OFF at all levels, from resource management to food consumption, and from individuals to collective entities.

Many methods are suggested in the literature on search strategy namely on emergent, complex and interdisciplinary research domain [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.

The figure 1 describes our iterative and inductive process of query building. Using already known relevant articles, we iteratively built a set of terms or their combination using Boolean operators. For each candidate terms or combination, we assessed the results to determine if it provides relevant information or noise. To do this iterative evaluation, we sampled 15 to 30 papers returned by the query to measure Precision (fig 1). If too much noise is generated (Precision <85%), we used some tactics to be more domain-specific. The first one consisted in identifying specific phrases containing these polysemic terms that can be added directly in the query (eg. "change of practice"). Another tactic consisted in narrowing generic terms using the Boolean or proximity operator (NEAR/x) which allows to deal with distance between targeted

terms and to capture more papers. For instance, the phrases “growth of the organic market” or “market growth” can be summarized in the syntax (growth NEAR/2 market).

Figure 1 : The iterative and inductive process of query building

Once we achieved a reasonable Precision rate (~85%), we performed a global evaluation.

It was performed thanks to the dataset of papers we iteratively sampled via our iterative evaluation plus the relevant papers identified by experts. This dataset, containing the WoS identifiers of sampled papers as well as their hand-coded relevance (T/F), is crossed with the final corpus to detect found and not found papers which enables the measure of Precision and Sensitivity. As the set of expected documents is unknowable by definition, it should be noted that the number of false negatives, and thus Sensitivity, is only approximated by the iterative sampling.

Indeed, search strategy is a trade-off between these two complementary indicators: focusing on Precision allows to capture fewer irrelevant papers with the risk of omitting relevant ones outside the corpus and inversely, focusing on Sensitivity allows to capture most of the relevant papers but with the risk of also capturing irrelevant ones.

**Results:** After many iterations, we reached a “relevance plateau”, ie the cost of improvement goes higher and higher for too poor enhancements. We then decided to assess globally the query (tab 1). We know that not 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 quite good compromise.

The process allows the progressive delineation of the perimeter of the search topic. Thus, we finally decided to consider the topic in a rather large sense, focusing on the lexical field of change - from description of past trends to prospective views both at individual and also collective scales - associated with the conditions and determinants of change (barriers or drivers). These aspects may be central in the paper but sometimes peripheral referring to organic scaling only in the context or in the discussion parts of the abstract (about 13% in our sample). Considering the functionalities of WoS search engine, it is however not possible to discriminate central from peripheral mentions.

The iterative process, thanks to our expertise and analysis of the intermediary corpora generated at each iteration, also allows the topical decomposition in sub-topics related to organic scaling. We resulted in two main blocks of terms: [OFF] AND [SCALING]. [3] The OFF block is a sophisticated query, based upon 550 terms, which returned 20990 papers (18 oct. 2019), almost 2.5 more than a simple query such as [TS=("organic farm\*" OR "organic agriculture")] (n=8265) used in Sigmeier et al. [4]. This advocates for a query sophistication. Indeed, the [SCALING] block also needs sophistication, since the term “scaling” (and its derivation like “up-scaling” or “out-scaling”) only returns 72 papers. This block is the combination (with the OR Boolean) of terms (n=443) related to different sub-topics (see table 1).

Table 1: Examples of some sub-topics composing the scaling block

The optimization of the search query, ie adding and excluding search terms, needs a particular attention to the meanings of term and to the way authors articulate them into language. Some sub-topics are quite 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 false negative paper. Some other sub-topics are related to more generic or polysemic terms. For instance, the terms “growth”, “development” or “change”, important to deal with scaling up, are also used in the context of biological phenomena, hence the sophistication of the query.

**Discussion:** 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 of a maximum of relevant papers, even though it captures some irrelevant ones. We developed an iterative process to refine the perimeter of the search topic. Our expert-based approach and the immersion in the literature also allowed us to better characterize and acquire a good knowledge of the literature on organic scaling.

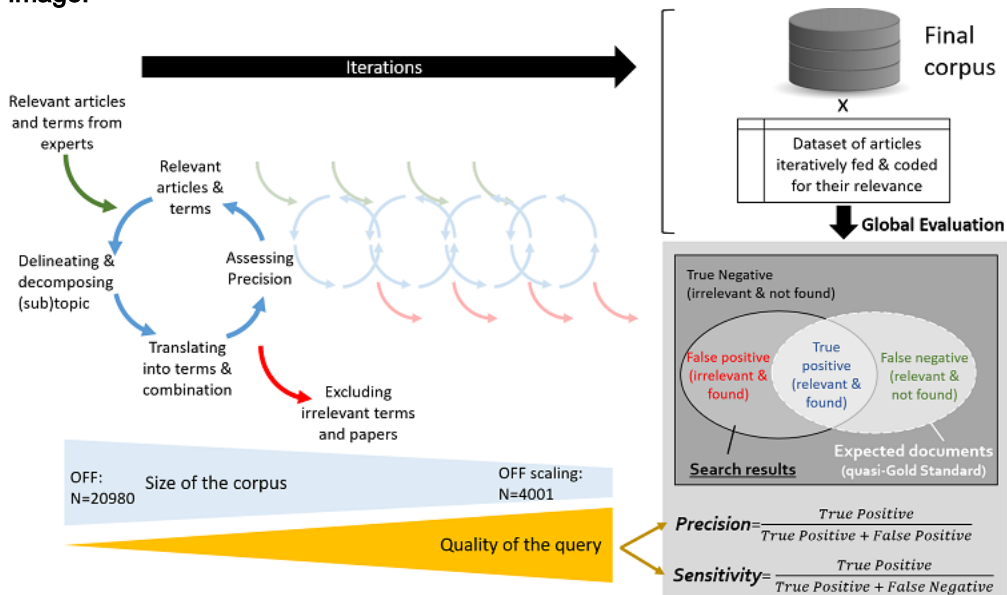
As perspectives, the corpus can now be better defined and be used to feed the INRA program. We also suggest the research community to better elicit and specify a more standardized lexicon describing their work (eg. “organic food and farming scaling”). Finally, we intend to test some ex-post corpus cleaning methods (lexical supervised learning filtering, bibliographical coupling...) for further bibliometric analysis.

**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