

7-1-2012

Identifying emerging research topics in Wind Energy research using author given keywords

Gali Halevi
Elsevier

Henk F. Moed Dr

Follow this and additional works at: <https://www.researchtrends.com/researchtrends>

Recommended Citation

Halevi, Gali and Moed, Henk F. Dr (2012) "Identifying emerging research topics in Wind Energy research using author given keywords," *Research Trends*: Vol. 1 : Iss. 29 , Article 5.

Available at: <https://www.researchtrends.com/researchtrends/vol1/iss29/5>

This Article is brought to you for free and open access by Research Trends. It has been accepted for inclusion in Research Trends by an authorized editor of Research Trends. For more information, please contact r.herbert@elsevier.com.

Section 4: Behind the data

Identifying emerging research topics in Wind Energy research using author given keywords

Gali Halevi MLS, PhD and Dr Henk Moed

The value of well constructed thesauri as means for effective searching and structuring of information is something a seasoned searcher is very familiar with. Thesauri are useful for numerous information management objectives such as grouping, defining and linking terms, and identifying synonyms and near-synonyms as well as broader and narrower terms. Searches based on thesauri terms are considered better in terms of both recall and precision (1,2,3).

Yet the construction of a comprehensive thesaurus is a laborious task which often requires the intervention of an indexer who is expert in the subject. Terms incorporated in a thesaurus are selected carefully and examined for their capability to describe content accurately while keeping the integrity of the thesaurus as a whole. Terms incorporated in a thesaurus are referred to as controlled vocabulary or terms. Uncontrolled vocabulary on the other hand, consists of freely assigned keywords which the authors use to describe their work. These terms can usually be found as a part of an abstract, and appear in most databases as "author keywords" or "uncontrolled vocabularies". In today's fast moving world of science where new discoveries and technologies develop rapidly, the pace by which thesauri capture new areas of research may be questioned, and so the value of now using author keywords in retrieving new, domain-specific research should be examined.

This study sought to examine the manners by which thesauri keywords and author keywords manage to capture new and emerging research in the field of "Wind Energy". The research questions were as follows:

1. Do author keywords include new terms that are not to be found in a thesaurus function?
2. Can new areas of research be identified through author keywords?
3. Is there a time lapse between the appearance of a keyword assigned by an author and its appearance in a thesaurus?

Methodology

In order to answer these questions we analyzed controlled and uncontrolled terms of 4000 articles grouped under the main heading "Wind Power" in Compendex captured between the years 2005–2012. Compendex is a comprehensive bibliographic database of scientific and technical engineering research available, covering all engineering disciplines. It includes millions of bibliographic citations and abstracts from thousands of engineering journals and conference proceedings. When combined with the Engineering Index Backfile (1884-1969), Compendex covers well over 120 years of core engineering literature.

In each Compendex record a list of controlled and uncontrolled terms are listed and can be searched on. Over 17,000 terms were extracted from the Compendex records and sorted by frequency. Two separate files were created; one depicting all the controlled terms and the second depicting the author given keywords (i.e. uncontrolled terms). For each term a count of the number of times they appear in each year from 2005-2012 and the total number of articles in which each term appears was recorded. In addition, a simple trend analysis compared the number of the times each term appears on average in papers published during the years 2009–2012 with the same measure calculated for 2005–2008. This trend analysis allowed for a view of terms that increase in usage in the past 3 years, compared to the overall time period.

To answer the research questions, the following steps were taken:

- All author keywords that appear 100 times or more were collected.
- The author keywords were searched in the Compendex Thesaurus: if an author keyword appeared, the year in which it was introduced was recorded.
- The author keyword was then searched for in Compendex across all years and the year in which it first appeared was recorded.
- The author keywords that appeared more than 100 times were grouped into themes. In addition these author keywords were searched for in Compendex in order to identify their corresponding articles and the topics they cover.

Findings

Table 1 shows the most recurring uncontrolled terms. The terms were categorized in 4 groups as follows:

Topic Group	Environment	Mechanics	Integration	Computerization
Uncontrolled Terms	Renewable energies Renewable energy source Wind energy Wind speed Wind Resources	Doubly-fed induction generator Offshore wind farms Permanent magnet Synchronous generator Wind farm(s) Wind turbine generators Wind generators Wind generation Wind energy conversion system	Control strategies Power grids Power output Grid-connected	Simulation result

Table 1 – Most recurring uncontrolled terms in the retrieved articles. Source: [Engineering Village](#)

Looking at the corresponding literature within Compendex, there were three main topics that emerged from the author key words which indicate specialized areas of research within the overall ‘wind power’ main heading. These terms did not appear in the Compendex thesaurus.

Wind Farms: This term first appeared as an uncontrolled term (i.e. Author keywords) in 1985 in an article by NASA researchers (4). The term refers to large areas of land on which wind turbines are grouped. Some examples of such wind farms are The Alta Wind Energy Center (AWEC) which is located in the Tehachapi-Mojave Wind Resource Area in the USA and the Dabancheng Wind Farm in China. This research includes a wide variety of topics ranging from agriculture, turbines mechanics, and effects on the atmosphere and power grid integrations. The term has shown substantial growth in use as an author keyword between 2006 and 2012 with peak of 757 articles in 2011 (see Figure 1).

In the thesaurus, however, this term is included under “Farm buildings” which also contains livestock buildings and other structures that are to be found in farms.

Offshore wind farms: This term first appeared as an uncontrolled terms in 1993(5) and refers to the construction of wind farms in deep waters. Some examples of such wind farms include [Lillgrund Wind Farm](#) in Sweden and [Walney](#) in the UK. In the thesaurus articles with this keyword are assigned to the term “Ocean structures”. This of course includes other structures such as ocean drilling, gas pipelines and oil wells. The use of this term has been steadily growing (see Figure 2) with substantial increase between 2008 and 2011.

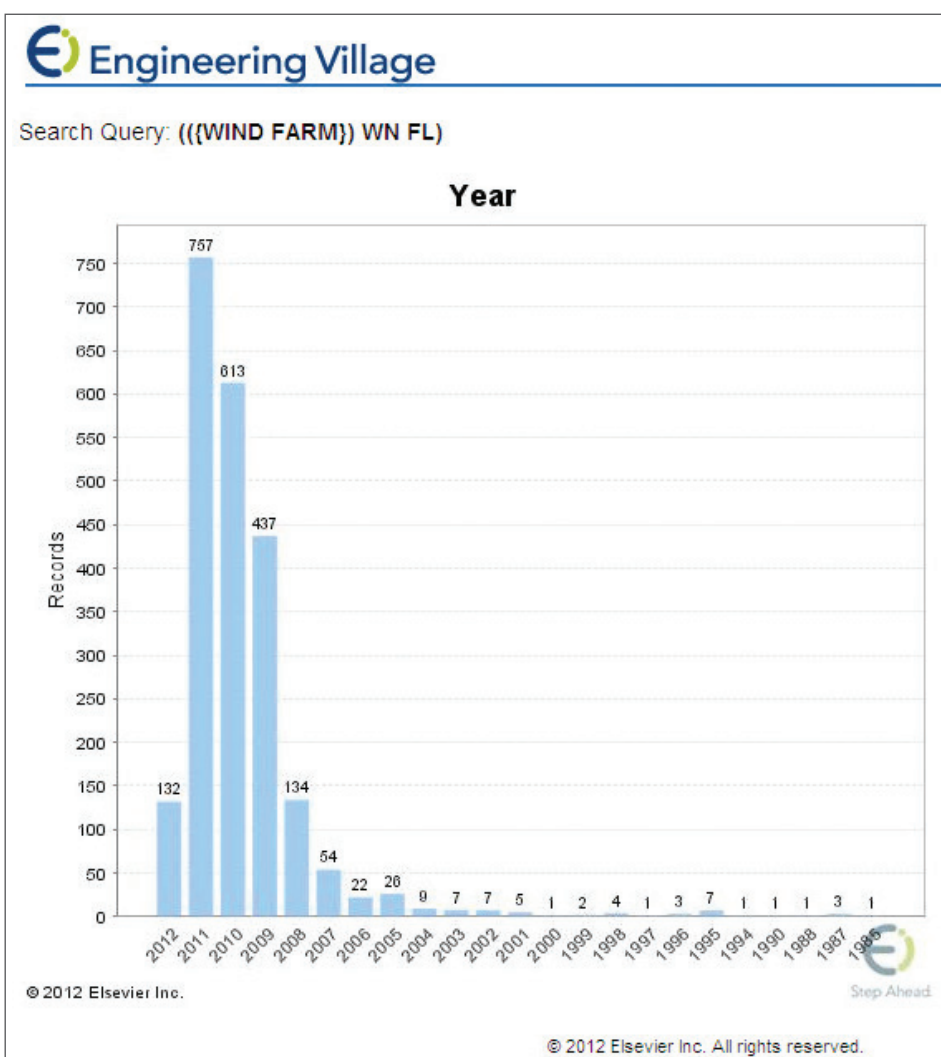


Figure 1 – Use of keyword Wind Farm by authors. Source: [Engineering Village](#)

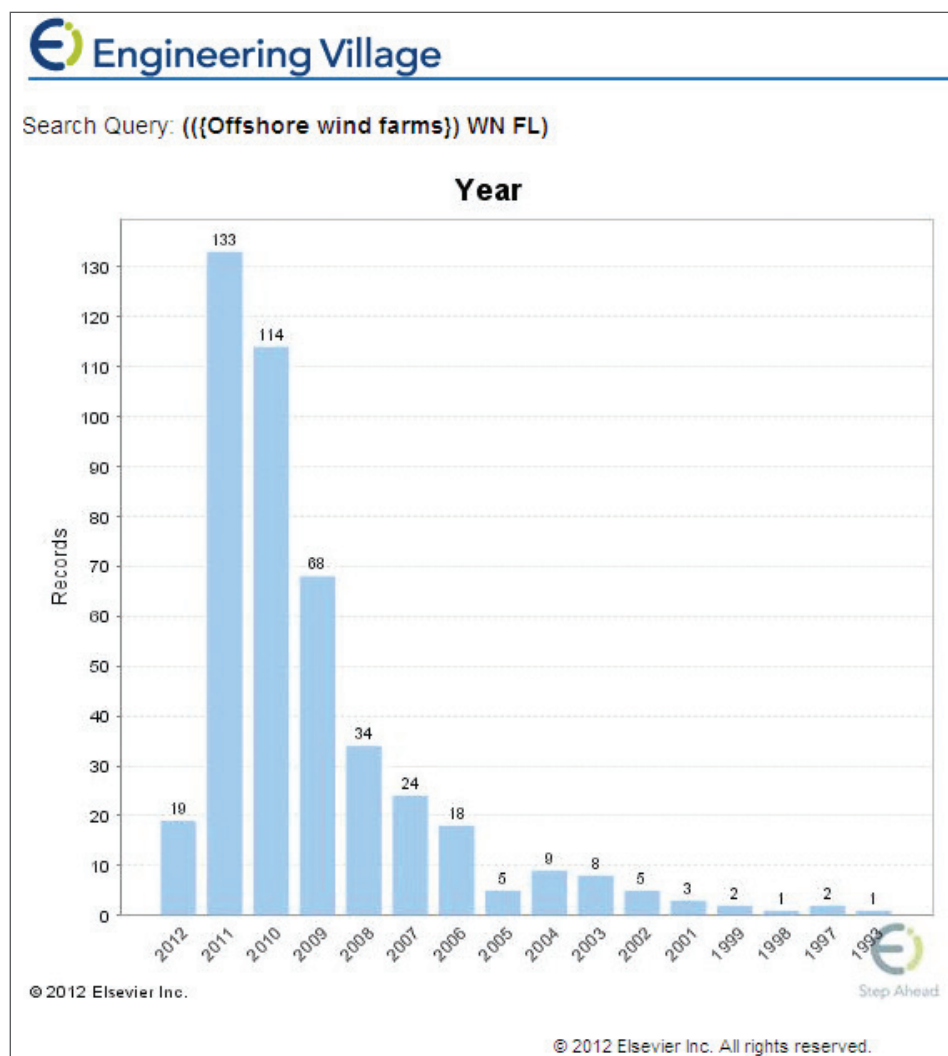


Figure 2 – Use of keyword Offshore Wind Farms by authors. Source: Engineering Village

Most surprisingly, however, is the fact that the term Wind energy itself doesn't appear in the thesaurus at all. The topic as a whole appears under "Wind Power" which also applies to damages caused by wind, wind turbulences, wind speed and so forth. The term has been used by authors since 1976 and first appeared in an article by the Department of the Environment, Building Research Establishment of UK Government (6) and has seen constant growth between 2006 and 2012 (see Figure 3).

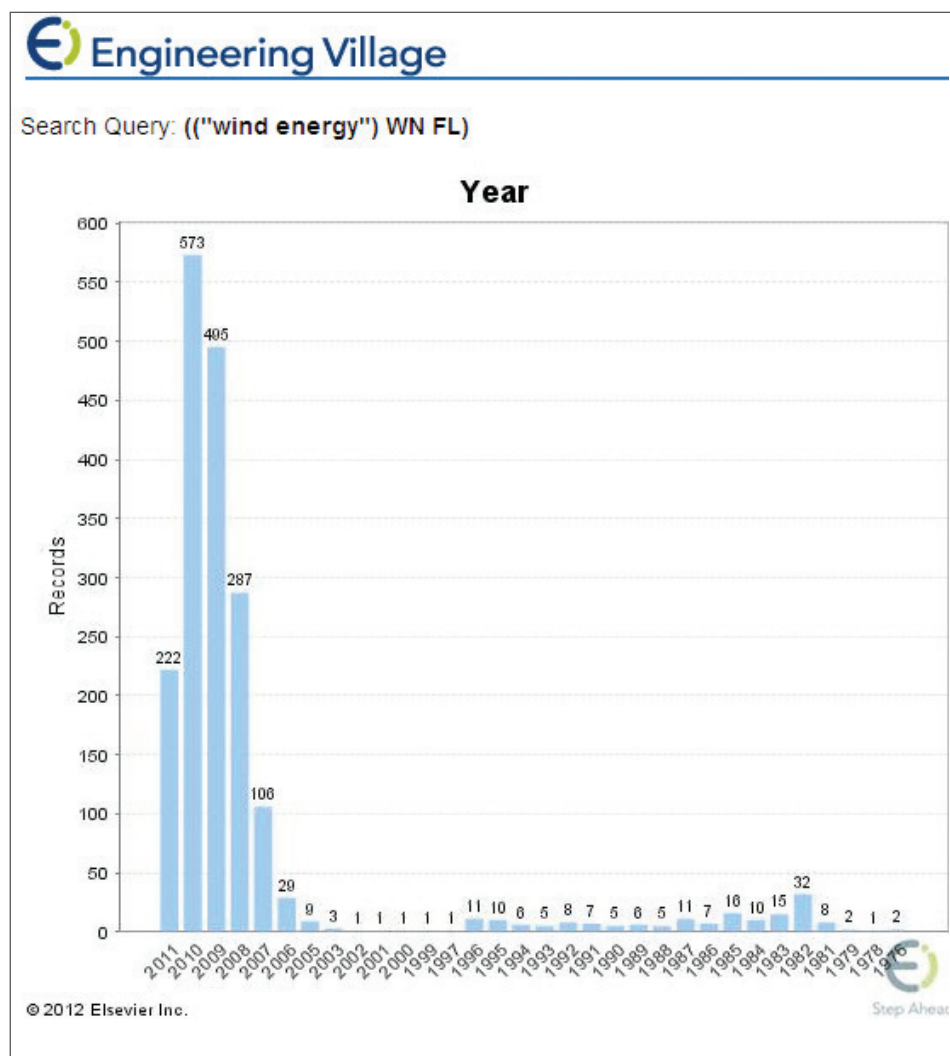
Other emerging topics include: wind energy integration into power grids, effects of wind farms on the atmosphere, wind farms and turbines computer simulations and control software. In addition, comparing the uncontrolled and controlled terms that appeared most commonly there are apparent differences in foci as they emerge from the vocabulary. While the uncontrolled vocabulary highlights Wind speed and farms, the controlled vocabulary features Wind power, Electric utilities, and Turbomachine blades. This could be due to the fact that the Compendex thesaurus is engineering focused, thus giving the mechanics of wind power conversion prominent descriptors. In this case, the author given keywords are valuable and they provide a supplementary view on these topics by depicting the environmental aspects of these research articles. Table 2 illustrates the different foci of the keywords.

Discussion

Wind energy is by no means a new area of exploration, yet in the past 4 to 5 years this area has seen a considerable growth in research output especially in wind turbines technology and wind harvesting. Although the data sample analyzed is small and covers one subject field only, our findings illustrate that author keywords may indeed include new terms that are not to be found in a thesaurus function. The use of thesauri terms is usually recommended as a part of precision strategy in searching. Yet, in our case controlled terms have a more general scope. Table 3 below summarizes some of our major conclusions as they pertain to the properties of using author-given keywords and controlled terms in the search process. Our findings show that the use of author given keywords as a search strategy will be beneficial when one searches for more specific technologies and applications or new research areas within the overall topic (see Table 3).

Uncontrolled terms	Controlled terms
Wind speed (43 papers, 10%)	Wind power (172 papers, 41%)
Wind farm (37, 9%)	Wind turbines (171, 41%)
Wind farms (22, 5%)	Computer simulation (74, 18%)
Wind turbine blades (17, 4%)	Mathematical models (73, 18%)
Fatigue loads (12, 3%)	Aerodynamics (72, 17%)
Wind energy (12, 3%)	Electric utilities (63, 15%)
Wind turbine wakes (12, 3%)	Turbomachine blades (58, 14%)
Control strategies (11, 3%)	Wind effects (49, 12%)
Offshore wind farms (11, 3%)	Rotors (48, 12%)
Power systems (11, 3%)	Wakes (45, 11%)

Table 2 – Most common controlled and uncontrolled terms on search. Source: Engineering Village



Our analysis showed, for example, that strongly emerging areas identified in our sample are wind farms and offshore wind farms. These terms, although appearing in the author given keywords for over 20 years have not entered the Compendex thesaurus. This could be due to the fact that the Compendex database is engineering-focused and built to serve engineers therefore grouping these articles under terms that are mechanical in nature. However, this might hinder a broader understanding of the topics in context.

In this case using the thesaurus as basis for searching Wind Energy articles would create broader results sets. Depending on what the purpose of the search is, this could be viewed as a positive or negative outcome. Our analysis shows that the two types of terms have different properties and serve different purposes in the search process. In the analysis of emerging topics author-given keywords are useful tools, as they enable one to specify a topic in a way that seems difficult to carry out when one uses only terms from a controlled thesaurus.

Figure 3 – Use of keyword Wind Energy by authors. Source: Engineering Village

	Controlled Terms	Uncontrolled Terms	Notes
Recall	✓		Using controlled terms retrieves a larger number of articles since they are lumped under broader descriptors
Precision		✓	Uncontrolled terms are very specific and enable retrieval of detailed topics
Discoverability	✓	✓	Uncontrolled terms enable the discovery of the new topics and can serve as indicators of the latest discoveries made in this field. Controlled terms enable the clustering of such topics thus enabling connections between larger numbers of articles and topics
Serendipity	✓		Controlled terms are broader thus retrieving a larger amount of article and enabling serendipity through browsing
State of the Art		✓	Uncontrolled terms depict the latest descriptors of methods, applications and processes in a certain topic

References:

- Sihvonen, A., Vakkari, P. (2004) "Subject knowledge, thesaurus-assisted query expansion and search success", Proceedings of RIAO2004 Conference, pp. 393-404.
- Sihvonen, A., & Vakkari, P. (2004) "Subject knowledge improves interactive query expansion assisted by a thesaurus", Journal of Documentation, 60(6), 673-690.
- Shiri, A.A., Revie, C., Chowdhury, G. (2002) "Thesaurus-enhanced search interfaces", Journal of Information Science, Volume 28, Issue 2, 2002, Pages 111-122.
- Neustadter, H. E., & Spera, D. A. (1985) "Method for Evaluating Wind Turbine Wake Effects on Wind Farm Performance", Journal of Solar Energy Engineering, Transactions of the ASME, 107(3), 240-243.
- Olsen, F., & Dyre, K. (1993) "Vindeby off-shore wind farm - construction and operation", Wind Engineering, 17(3), 120-128.
- Rayment, R. (1976) "Wind Energy in the UK", Building Services Engineer, (44), 63-69.

Table 3 – Evaluation of the impact of controlled and uncontrolled terms on search. <https://www.researchtrends.com/researchtrends/vol1/iss29/5>