

Agrovoc Descriptors: organic agriculture, information retrieval, information science, information storage, data collection, data processing, documentation, terminology

Agris Category Codes: P01, C30, A01

COBISS Code 1.01

Scientific and technical information on organic farming: assessment of selected bibliographic indicators in database CAB Abstracts

Tomaž BARTOL¹, Špela DRNOVŠEK², Majda ČERNIČ-ISTENIČ³

ABSTRACT

We scientometrically examined selected publishing patterns in organic farming as indexed by database CAB Abstracts (CABA) in the period 1973-2003. Some characteristics of CABA were also established. 4,170 records of all types (2,740 journal articles, 1,165 conference papers etc.) were harvested with the employment of descriptors derived from the CAB Thesaurus. The journal articles were further analyzed for yearly growth, languages, authorship, and geographic location. Growth has accelerated during the last ten years. English prevails among the 31 retrieved languages, with some 50 % citations, followed by German, Italian, French, and Japanese. A few highly productive authors (among the 4,304 different authors), with some publishing as many as 18 articles, accounted for the core of documents, whereas as many as 3,438 authors contributed only one article in the entire period. Germany was most frequently indexed as a geographic-location-descriptor, followed by Italy, U.K., Denmark, and Switzerland. This descriptor field, however, is not employed consistently so the geographic data must be interpreted with caution. This analysis is based mostly on descriptor Organic farming so some other terms could possibly also be employed to retrieve more documents related to this field of research.

Keywords: organic farming, bibliographic databases, bibliographic data, information retrieval, information scatter, data collection, data processing, indexing, documentation, information science, bibliometrics, scientometrics, terminology

ZNANSTVENE IN STROKOVNE INFORMACIJE O EKOLOŠKEM KMETIJSTVU: OCENA IZBRANIH BIBLIOGRAFSKIH INDIKATORJEV V ZBIRKI CAB ABSTRACTS

IZVLEČEK

Scientometrično smo ovrednotili značilnosti objav za ekološko kmetijstvo po podatkih iz zbirke CAB Abstracts (CABA) v obdobju 1973-2003. Pri tem smo ocenili smo tudi nekatere značilnosti same zbirke. Zbrali smo 4.170 zapisov vseh tipov (2.740 člankov iz revij, 1.165 prispevkov s posvetovanj ipd.) s pomočjo deskriptorjev iz tezavra CAB. Članke smo nadalje

¹ Ass. Prof., Ph.D., Univ. of Ljubljana, Biotechnical Fac., Agronomy Dept., Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia

² Ass. Prof., Ph.D., *ibid.*

³ B. Sc., *ibid.*

This article is part of Bachelor of Science thesis of Špela Drnovšek. Supervisor: Tomaž Bartol, Ph.D., Co-supervisor: Majda Černič-Istenič, Ph.D.

analizirali glede na letno rast, jezik, avtorstvo in geografsko lokacijo. Rast se pospešuje zadnjih deset let. Med 31 različnimi jeziki prevladuje angleščina (50 % dokumentov); sledijo nemščina, italijanščina, francoščina in japonščina. Nekaj posameznih visoko produktivnih avtorjev (med 4.304 različnimi avtorji), od katerih je en napisal kar 18 člankov, je prispevalo pomemben delež vseh dokumentov, kar 3.438 pa je bilo takih, ki so prispevali le po en članek. Najpogosteje rabljeni deskriptor pri geografski lokaciji je bila Nemčija, sledila je Italija, Združeno kraljestvo, Danska in Švica, vendar pa se to deskriptorsko polje v zbirko ne vnaša konsistentno, zato je treba geografske podatke interpretirati s pazljivostjo. Analiza temelji pretežno na deskriptorju Organic farming, zato je potrebno upoštevati, da je v zbirki verjetno najti še precej drugih dokumentov, ki bi vsebinsko tudi ustrezali temu pojmu, a z njim niso bili indeksirani.

Ključne besede: ekološko kmetijstvo, bibliografske podatkovne zbirke, bibliografski podatki, iskanje informacij, razkropljenost informacij, zbiranje podatkov, obdelava podatkov, indeksiranje, znanstvena informatika, bibliometrija, scientometrija, terminologija

1 INTRODUCTION

Last decade has seen a significant growth of activities in the field of organic farming. This has been linked to an increased awareness regarding both environmental protection and food safety. These activities have reflected in agricultural practice and scientific exploration. Such exploration usually also results in some form of published research which gets, in turn, indexed by major scientific databases.

In our study we wish to scientometrically examine some basic publishing patterns in the field of organic farming with the assistance of database CAB Abstracts (CABA) as the principal database for agriculture and related sciences. We intend to observe growth of organic-farming-related records through the entire period of the existence of this database. We wish to identify some principal bibliographic features, such as the input by authors and by languages. We wish to identify the most prolific authors in this field. We also wish to assess some geographic patterns, such as participation of European countries.

By observing the above data we also wish to look into some characteristics of the database CABA and determine how these characteristics can impact retrieval of records. Namely, each bibliographic database may possess particular features which may affect the retrieval results so these characteristics need to be identified in order to prevent erroneous search queries, and also to put some limitations on retrieval expectations.

Scientometric literature regularly addresses also agricultural topics. Such research frequently focuses on a specific journal as a scientometric object. It can also focus on a specific subject within the broader area of agriculture. Alfaraz and Calvino (2004) investigated scientific production in the field of food science and technology in Spain and Latin America. Authorship of doctoral theses in agriculture and veterinary sciences in India was studied by Krishna and Kumar (2004). Authorship in Egyptian agricultural journals was examined by Farahat (2002). Plant and crop science in general was investigated with regard to documents published in Slovenia (Bartol, 2002). Apple scab literature was bibliometrically evaluated by Stopar (2002). Growth of literature in the field of ethnobotany, and some authorship patterns, were studied by

Dhiman and Sinha (2001). Scrub literature has been bibliometrically investigated with regard to worldwide publications, and with emphasis on growth of documents, document types and authors (Botello *et al.*, 1998). Different aspects of literature on garlic were addressed and focus was placed on South America. Garlic was investigated as an alternative crop for export (Paredes and Portela, 1997). International literature on fruit and vegetable economics was investigated by Codron and al. (1995). Bibliometric methods were used by Olsen (1994) to identify monographs and core journals in seven sub-disciplines of agriculture (agricultural economics and rural sociology, agricultural engineering, animal science and diseases, soil science, crop science, food science and human nutrition, and forestry and agroforestry).

Database CAB Abstracts has also been used as a source of scientometric material in some of the above studies. This database was employed in investigating global trends in fruit cultivation research (Stopar *et al.*, 2004). Primary agriculture journals, as indexed by CABA and some other databases, were identified by Kawasaki (2004). Seven databases, including CABA, were employed by Fernandes and Antunes (2002) in order to identify records on medicinal plants. CABA was used to identify trends in research on mites as biological control agent (Gerson, 2001), and to identify records related to medicinal plant *Salvia* (Bartol and Baričević, 2000).

2 MATERIALS AND METHODS

We selected database CAB Abstracts as the leading database in the field of agriculture and related sciences. The database is produced by CAB International and presently contains over 4,5 million records from 1973 onwards with current yearly addition of some 220,000 records. Presently it indexes 6,000 journals and 3,500 documents of other types such as books, book contributions, conferences proceedings etc. In the year 2003 more than 150,000 new journal articles were added. Documents are contributed by 140 countries and are written in 50 original languages. Common bibliographic data such as titles, subject headings and abstracts are always displayed in English.

The subject of organic agriculture comes about also under other terms or phrases so we first examined the principles of subject indexing and terminology in this database. A thesaurus (CAB Thesaurus) is used as the source of controlled Heading Words (descriptors).

CAB Thesaurus employs the following *Heading Words* or *subject fields*:

SH:	Subject Headings
GL:	Geographic Location
ID:	Identifiers
OD:	Organism Descriptors
BT:	Broad Terms

General concepts or practices such as agriculture, farming, husbandry etc. are arranged in the field SH (*Subject Headings*). There exists an elementary subject tree with references to different narrower or related concepts within the broader category of farming. The thesaurus directs to the term of ORGANIC FARMING which is then used as a single descriptor for this topic, and which stands also for non-descriptor terms of *eco-agriculture*, *organic culture*, and *organic gardening*. By employing descriptor *organic farming* we retrieved and consequently extracted 4,146 records for the entire period of 1973- 2003. First records indexed with these terms occurred only in 1976.

We also identified another term: ALTERNATIVE FARMING, which is, according to the CAB thesaurus, not related to the concept of ORGANIC FARMING but rather to the concept of SUSTAINABILITY. This concept, however, represents many other topics and can't be used in *sensu stricto* to represent the concepts related to organic agriculture. As the occurrences of records for *alternative farming* were rather few we nevertheless browsed also through these records and manually extracted a few relevant records which could also have been indexed with the term *organic farming*. We also extracted a few relevant records based on descriptor ORGANIC CULTURE which were also associated with organic farming but were not indexed as such. We thus obtained additional 24 records. Our final collection consequently amounted to 4,170 records.

We then created a separate experimental database where we downloaded the 4,170 records. We merged all the above *Heading Words* fields into a single operational field to facilitate summarization of the terms. This database was then used for further analyses.

We sorted the records according to the *Document Type* as defined by the database CABA. Yearly growth and other data can best and most consistently be assessed with *Journal Articles*. We thus created a sub-database to conduct further analyses for *Journal Article* document type. We monitored the following details:

yearly growth of records, language of document, (co)authorship, geographic location

Language field was uniformly standardized so the analyses could be conducted with simple counting and sorting of languages. Author field was more complex and sometimes contained data for ten authors or more in a single record so these records had to be accordingly divided to enable further analysis. Also, this field was not fully standardized. There occurred also some spelling errors so some authors could come about under several different variants, such as *Ren HuiFeng = Ren, H. / Bidappa, C.C.= Biddappa, C.C. etc.* As we wished to evaluate the authorship for each particular author we had to identify the different variants and merge them as single authorship when applicable.

The geographic location is represented by separate field so in theory it could be identified easily. By way of browsing we carried out a preliminary analysis to determine some possible patterns. We found out, however, that this field lacked in consistency what will be explained in next chapter.

3 RESULTS AND DISCUSSION

3.1 Document types

Descriptor-based retrieval in database CABA generated 4,170 records related to the concept of *organic farming* during the period of 1976-2003. Journal articles account for 66 % of all records (2,740), followed by conference papers. Other document types are less frequent. Some records may be defined as more than one type, e.g. a *conference paper* that may have been issued also as a *journal article*. The total number of types in Table 1 therefore exceeds the total of 4,170 documents

Table 1. Document Type of records retrieved in CABA during 1976-2003 with regard to descriptor Organic farming.

Document type	Records	%	Document type	Records	%
Journal article	2,740	65,7	Annual report	39	0,9
Conference paper	1,165	27,9	Journal issue	24	0,6
Book chapter	317	7,6	Thesis	21	0,5
Miscellaneous	274	6,6	Standard	13	0,3
Book	126	3,0	Patent	2	0,1
Conference proceedings	97	2,3	Abstract only	1	0,1
Bulletin	94	2,3	Correspondence	1	0,1

Journals usually apply a more consistent publishing policy so journal articles generally offer better possibility as to the assessment of bibliographic patterns. Our further analysis therefore considered only the 2,740 retrieved journal articles.

3.2 Yearly growth of journal articles

Table 2 and Figure 1 present growth of documents indexed with the term organic farming. The numbers in the first decade are rather low. First two records come about only in 1976. There is a rather high growth in a more recent period, followed by a slight decline in the last two years under observation. This decline can possibly be explained by some delay of input in the database, and not by decline of publications what can, however, only be more thoroughly examined in the near future.

Table 2. Yearly occurrences of records (journal articles) indexed with descriptor Organic farming in CABA during 1976-2003.

Year	Rec.	Year	Rec.	Year	Rec.	Year	Rec.
1976	2	1983	11	1990	56	1997	234
1977	2	1984	14	1991	57	1998	170
1978	8	1985	8	1992	64	1999	276
1979	16	1986	16	1993	64	2000	298
1980	15	1987	19	1994	76	2001	364
1981	11	1988	33	1995	123	2002	321
1982	1	1989	46	1996	149	2003	286

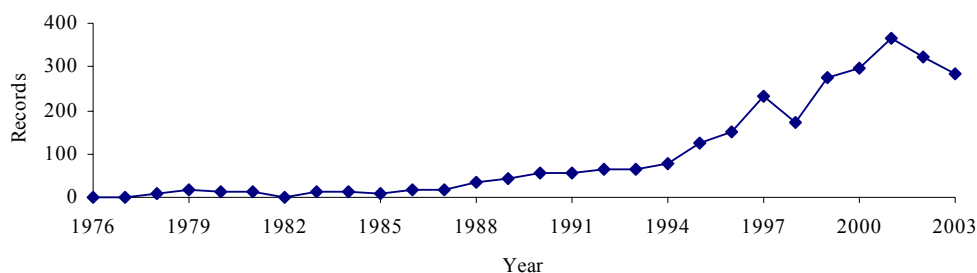


Figure 1. Yearly occurrences of records (journal articles) indexed with descriptor Organic farming in CABA during 1976-2003.

3.3 Language of journal articles

In further exploration language of journal articles was identified. Roughly 50 % all articles (1,366) were published in English, followed by German, and Italian. Other languages account for fewer than 200 occurrences. Altogether there were 31 languages. It has to be stressed that these data reflect publication by languages, and not publication by countries. Many journals in non-English speaking countries are namely published in English. Data in Table 3 present languages that were used in more than ten documents. Other languages include Turkish (9), Slovak (8), Estonian (7), Lithuanian (7), Romanian (5), Bulgarian (3), Croatian (2), Persian (2), Ukrainian (2), Finnish (1), Korean (1), Latvian (1), Macedonian (1), Serbian (1).

Table 3. Language distribution in records indexed with descriptor Organic farming in CABA during 1976-2003.

Language	Records	Language	Records	Language	Records
English	1,366	Portuguese	59	Russian	33
German	465	Danish	57	Swedish	25
Italian	240	Chinese	46	Czech	17
French	122	Dutch	42	Hungarian	16
Japanese	64	Polish	38	Slovenian	10
Spanish	61	Norwegian	35	<i>Other</i>	53

As was the case with the previous analysis of document types some documents were assigned more than one language, usually English and some other language, so the total number of language occurrences slightly exceeds the total of 2,740 articles.

3.4 Productivity of authors

The productivity of authors was assessed on the base of a preceding detailed analysis of author-field with the technique of separation of multi-author fields into single author data. We examined authorship of all 2,740 articles. The articles were published by 4,304 authors, i.e. 4,291 different *individual authors* and 13 *corporate authors* (i.e. institutions where a particular author was not cited). In 78 articles the author data were not available. The reason could not be ascertained. 3,438 authors (80 %) contributed an only article in the entire period, whereas 14 highly productive authors (co)published more than 10 articles each, with the leading author contributing as many as 18 articles (Table 4, Figure 2). Figure 2, which presents the data for 4,304 authors, exhibits a well defined and typical bibliometric curve, what is in accordance with Lotka's bibliography (1926). A few distinct authors were thus highly productive, whereas a majority of authors (3,438 or 539 respectively) published only one or two documents on the subject of organic farming.

Table 4. Number of articles per individual authors in CABA with regard to records (journal articles) related to descriptor Organic farming in 1976-2003.

No. of Articles	18	17	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
No. of Authors	1	1	1	1	1	1	2	3	4	7	16	26	45	57	161	539	3,438

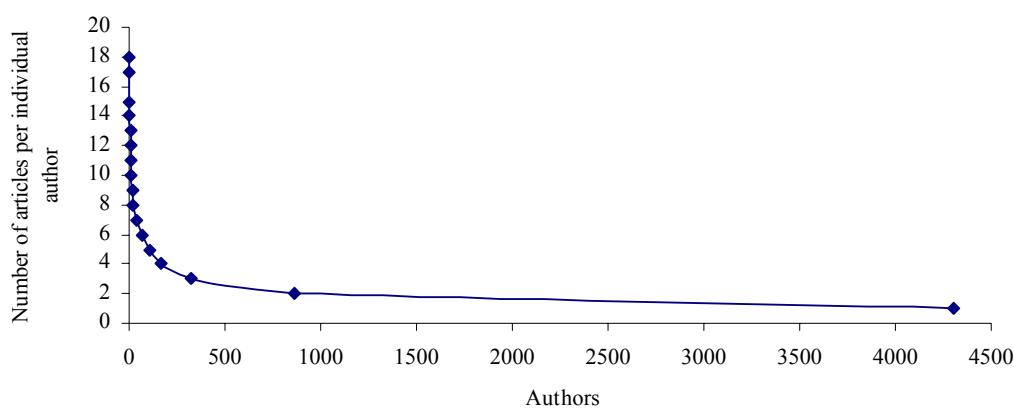


Figure 2. Number of articles per individual authors in CABA with regard to records (journal articles) related to descriptor Organic farming in 1976-2003.

The above numbers address the authorship regardless of the author's share in a document. 2,662 articles (among the 2,740) were authored. The authored articles contained data for 5,963 different author names. There were in fact 4,304 different authors (4,291 different individual authors and 13 corporate authors). The authors-per-article patterns again exhibit a very specific distribution which is related to the above distribution of articles-per-author. We can observe that 1.094 (40%) of the total of 2,740 articles were written by one author, 673 articles (25%) were written by two, 454 by three etc. A record holder is an article with 20 authors, followed by two articles with 15 authors etc. (Table 5, Figure 3).

Table 5. Number of (co)authors per article in CABA with regard to records (journal articles) related to descriptor Organic farming in 1976-2003.

Authors/Article	20	15	12	11	10	9	8	7	6	5	4	3	2	1	0
No of Articles	1	2	1	2	6	3	8	18	52	96	252	454	673	1,094	78

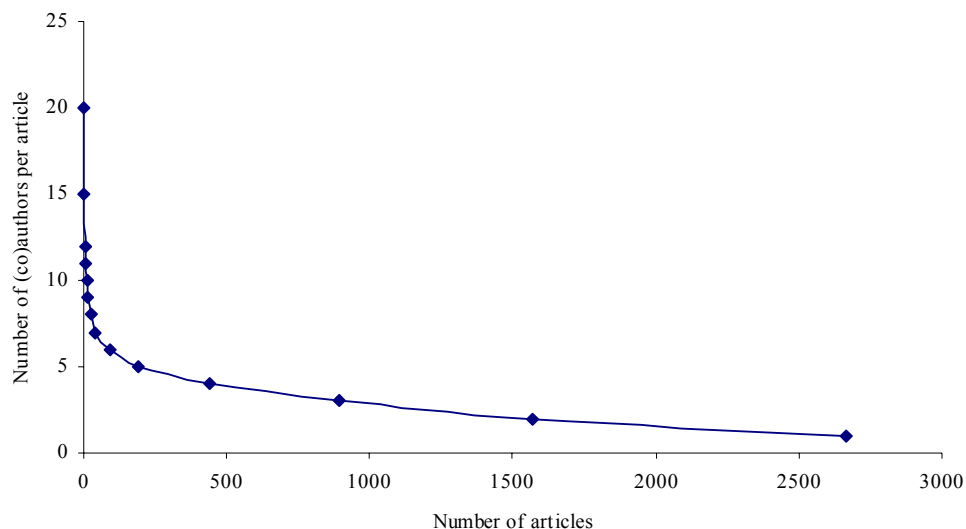


Figure 3. Number of (co)authors per article in CABA with regard to records (journal articles) related to descriptor Organic farming in 1976-2003.

3.5 Geographic location

We used the existing descriptor field *Geographic location* to identify location of research. We placed emphasis on European data. We checked if it is possible to group geographic location also according to the principles other than a country or a nation-state. In CAB Thesaurus there exist very specific descriptors for different levels of geographic location which follow some form of geographic hierarchy.

For example, selected European countries can all be supplied with the following *Broader Terms*: *Developed Countries*, *European Union Countries*, *OECD Countries*. Some countries may acquire additional *Broader Terms*: *Mediterranean Region*, *Southern Europe*, *Scandinavia*, *Western Europe*, *Central Europe* etc. However, Sweden can only be *Scandinavia* but not *Western Europe*. Italy can be *Southern Europe*, France can be *Western Europe* etc. There exist also *Geographic location* descriptors for provinces or other units, such as Tuscany, Lombardy, Bavaria, Isle of Man, Northern Ireland etc.

When it comes to practical record retrieval, however, the above descriptors turn out not to be assigned consistently. On many instances a particular record contains no geographic location at all. Also, a record may include a descriptor representing only a province. On some other instances descriptors will represent both the province and the country. On many instances only some broader territorial units are specified. Based on these limitations we decided to unify all the data according to a country (nation-state) if possible.

Table 6 presents only those occurrences where a country could be identified as relevant to a particular investigation. There were several other broader geographic units where the country was not referred to or was not relevant, such as Mediterranean region, Scandinavia, Central Europe etc. Countries with fewer than ten occurrences

were the following: Estonia (9), Bulgaria (6), Romania (5), Croatia (4), Albania (3), Ukraine (3), Yugoslavia (2), Portugal (2), Belarus (1), Iceland (1), Latvia (1), Macedonia (1).

Table 6. Records (journal articles) index with descriptor Geographic location in CABA in relation to descriptor Organic farming in 1976-2003.

Country	Records	Country	Records	Country	Records
Germany	288	Sweden	70	Greece	19
Italy	214	Norway	59	Turkey	18
U.K.	182	Poland	49	Belgium	17
Denmark	138	Spain	34	Slovakia	15
Switzerland	114	Czech R.	28	Lithuania	13
Austria	90	Russia	26	Slovenia	11
Netherlands	79	Hungary	23	Ireland	10
France	70	Finland	22		

These figures thus represent only such data as derived from the existing fields in the particular record and can consequently serve only as a very rough approximation of activities related to a particular country. With regard to a country or place of research it has to be underscored that there exist no clear principles as to when a country ought to be assigned as a descriptor term. Relevance of country in a particular research is namely difficult to specify. The activity of organic farming can nevertheless be quite specifically related to specific conditions in a particular country. The existing indexing patterns in the database, however, indicate that the field *Geographic location* can not be used as a very reliable tool to identify the place and regional conditions of research.

As a comparison we present the factual data with regard to the area under organic cultivation in ten leading countries of Europe (Figure 4).

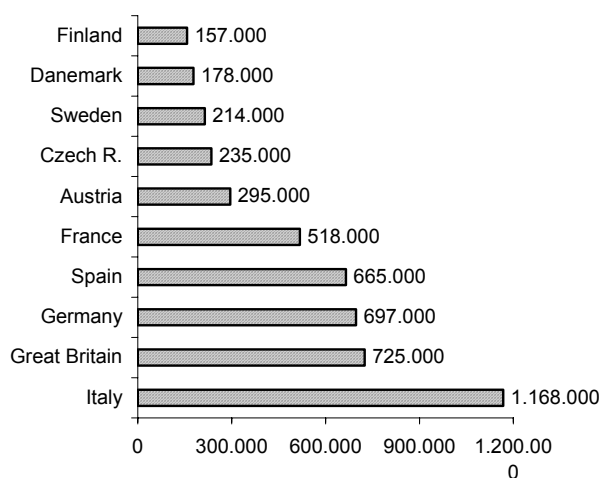


Figure 4. Area under organic cultivation (in ha) in five leading European countries (Willer and Yussefi, 2004).

4 CONCLUSIONS

We examined the database CABA as the principal global database for the broader area of agriculture with regard to the publishing patterns in the field of organic farming. It is possible to observe a steady growth of documents throughout the period of the last ten years. The assessment of this growth is based on the presence of descriptor *Organic farming* as an indexing element of the database. It can, on the other hand, also suggest that the indexing with this term has been more thorough on account of an increased awareness of these activities.

Journal articles account for the majority of documents. We employed only journal articles for further more detailed analyses. Language distribution proves the primacy of English. This dominance, however, is not as pronounced as in some other more theoretical subjects therefore probably indicating the significance of practical research and activities more closely related to a particular geographic area.

Author participation indicates a very typical bibliographic curve where a small group of highly active and productive authors publishes a very significant share of documents, and where, on the other hand, hundreds of different authors participate only one document in the entire period under observation. It is therefore possible to effectively identify the most central authors in the field of organic farming.

Geographic patterns are more difficult to ascertain. In CAB Thesaurus there exist not only very specific, but also very general geographic descriptors. These descriptors are, however, less consistently assigned. Geographic location can thus not serve as a very reliable feature in relation to a particular document.

As a conclusion we need to point out that our analysis tackled the employment of descriptor *Organic farming*. This descriptor is in this database designated to denote the particular subject matter of organic farming and related activities. However, as elucidated in the preceding chapters there exists in the CABA yet another descriptor *Alternative farming*, which is not, according to the CAB Abstracts, related to *organic farming* but rather to the topic of *sustainability*. This later term may also cover several environment-oriented and possibly organic aspects of farming. These have, however, not been indexed with the term *organic farming*, and could thus not have been a subject of our investigation.

These facts reveal some of the problems that can occur during information retrieval and need to be taken into consideration. The retrieved documents usually indicate only the presence of a particular term in a document and not necessarily all the documents of possible interest.

5 REFERENCES

- Acea Nebril, B.; Figueira Moure, A.; Gómez Freijoso, C. 2000. authorship, interval between acceptance and publication, and bibliographic references. *Revista Espanola de Enfermedades Digestivas*, 92, 2: 573-585.

- Alfaraz, P. H.; Calviño, A. M. 2004. Bibliometric study on food science and technology: Scientific production in Iberian-American countries (1991-2000). *Scientometrics*, 61, 1: 89-102.
- Bartol, T. 2002. Plant and crop science, production and protection: bibliometric evaluation of bibliographic data in documents published in the period 1994-2000. *Zb. Bioteh. fak. Univ. Ljublj., Kmet.*, 79, 1: 5-18.
- Bartol, T.; Baričević, D. 2000. In: Sage. Kintzios, S.E. (ed.). Amsterdam, Harwood Academic: 269-288
- Botello, A.; Cabezas, J.; Pulgarin, A.; Escudero, J.C. 1998. A bibliometric analysis of worldwide publications on scrub. *Rangelands*, 20, 2: 24-27.
- Codron, J.M.; Bedu, O.; Cibanel, C. 1995. A bibliometric analysis of international literature on fruit and vegetable economics from 1975 to 1989. *Acta-Horticulturae*, 340: 35-42.
- Dhiman, A. K.; Sinha, S C. 2001. Impact of research collaboration on growth of literature in ethnobotany: a bibliometric study. *Journal of Information Management*, 38, 1: 53-62.
- Farahat, H. 2002. Authorship patterns in agricultural sciences in Egypt. *Scientometrics*, 55, 2: 157-170.
- Fernandes, L. R.; Antunes, A. S. 2002. Strategic information on medicinal plants obtained from online databases.. *Agro Food Industry Hi-Tech.*, 13, 6: 13-18.
- Gerson, U. 2001. Trends in research on acarine biocontrol agents.. *Acarology: Proceedings of the 10th International Congress*. CSIRO Publishing, Collingwood, Australia: 473-475.
- Kawasaki, Jodee L 2004. Agriculture journal literature indexed in life sciences databases. *Issues in Science and Technology Librarianship*, 39 (No page numbers).
- Krishna, K M; Kumar, S. 2004. Authorship trends in agriculture research: a bibliometric analysis. *Journal of Information Management*, 41, 2: 229-234.
- Lotka, A.J. 1926. The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences*: 317-323.
- Olsen, W. C. 1994. The current core literature of the agricultural sciences.. *Journal of Agricultural & Food Information*, 2, 3: 99-105.
- Paredes, M.I.; Portela, J.A. 1997. Evolution of the garlic crop in the southern cone of South America from a bibliometric focus. *Avances en Horticultura*, 2, 1: 36-46.
- Stopar, K. 2002. Literatura s področja jablanovega škrlupa: bibliometrična raziskava = Apple scab literature: a bibliometric study.. *Zb. Bioteh. fak. Univ. Ljublj., Kmet.*, 79, 1: 115-125.
- Stopar, M.; Lokar, V.; Mikulic, M. 2004. Global trends in fruit cultivation research, *SAD*, 15, 4: 7-9.
- Willer H., Yussefi M. 2004. In: IFOAM publication. 170 p. (<http://www.soel.de>; Jul. 2004).