

# Microlevel Infometric Investigation in Phytochemistry

L. Pohl,<sup>†,‡</sup> L. Quoniam,<sup>†</sup> S. Quazzotti,<sup>§,⊥</sup> and H. Dou<sup>\*,†</sup>

CRRM,<sup>||</sup> Centre Saint-Jérôme, 13397 Marseille Cedex 13, France, and CRP Henri Tudor, 1359 Luxembourg-Kirchberg, Luxembourg

Received June 30, 1996<sup>⊗</sup>

In this paper we present a computer-based methodology in answer to a specific question in the field of phytochemistry. This study, prompted by an academic question, was undertaken on a voluntary basis but may be used as an aid in a business decision. The study is divided into three parts: first describing the scientific background motivating this analysis, second, the method for collecting the data, and third, the statistical study of the data and analysis of the results. Throughout this paper it is shown how a practical question can be resolved by the use of information technology. The study was carried out by automating a series of data processes, from data collection and analysis to final presentation of the results in a user-friendly manner. Manual checking was completed after each step. The aim of the exercise was to identify chemists who are specialists in natural compounds for high-quality perfumes. We have divided the question into three subjects which characterise the abilities needed to comprehensively cover the field of natural high-quality perfumes. We based our analysis on the supposition that to be considered a “specialist” is related to the publication activity of scientific articles in the three different areas of competence. The main difficulty of this study was to find authors who wrote articles on the three subjects, but not necessarily in the same article. The originality of the approach consists in data collection from several databases without downloading the bibliographic references.

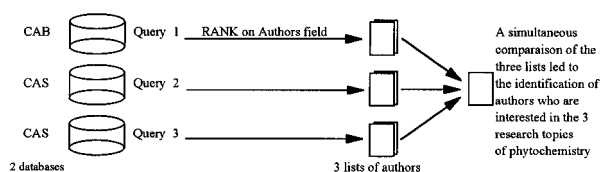
## INTRODUCTION

One of the constraints in competitive intelligence consists of combining rapidity and reliability in order to solve a specific information problem or, in other words, how to obtain more information from databases than with a classical search using Boolean operators AND, OR, or NOT.<sup>1</sup> In this example, we utilize two different databases, in which the authors' names (object of the search) may be written in different ways (with initial or full name) (see Figure 1).

## OBJECT

In this study, we were asked to identify key authors in the area of the production of natural compounds for high-quality perfumes in the fundamental field of analytic phytochemistry. Identifying international personalities in a specific field for science policy purposes has already been done.<sup>2,3</sup>

Questions from decision makers are rarely expressed as a query for databases, and the first step consists of transforming the decision maker's question into a database query. In this case, to be a “specialist” in the production of high-quality natural perfumes, a phytochemist should be competent in all of the steps in the process for obtaining natural compounds of high quality in good yields. This way of describing the object of our investigation is so general that we decided, in agreement with phytochemists, to split it into



**Figure 1.** Global scheme of the information system developed for this action.

three separate subjects: the production of plants with optimized characteristics; the best methods of extracting essential oils and balms from these plants; the analysis of the natural products in order to first evaluate the result of the two previous steps and then to improve them.

Thus, in this investigation, a specialist should first be concerned with the production of plants for essential oils and balms (e.g. the selection of the plants in order to obtain the best yields from the natural compounds). The selection of a plant depends, among other things, on its particular variety and on the date of its harvesting. This is the reason why we qualify the first subject: “plant production, harvesting, and varietal selection”.

The second step is described as extraction and purification of the natural compounds of interest from the plants. Most of the time a chemist will use a distillation or a solvent extraction or a concrete trapping with grease or heavy oils. This second step is identified as “extraction, purification, and distillation” in the field of phytochemistry.

The last step concerns the control of product quality, generally achieved by classical methods such as chromatography and nuclear magnetic resonance (NMR). The description of this step is “analysis, chromatography, and NMR”.

When the quality of the extracted natural products is known, it is possible to change one or another of the plant's

<sup>†</sup> CRRM.

<sup>‡</sup> Ph.D. between CRRM (Marseille) and CRP Henri Tudor (Luxembourg) with a grant from the Education Ministry of the Grand Duchy of Luxembourg. E-mail address: laurence.pohl@crpht.lu.

<sup>§</sup> CRP Henri Tudor.

<sup>⊥</sup> E-mail address: serge.quazzotti@crpht.lu.

<sup>||</sup> E-mail address: crrm@crmm.univ-mrs.fr.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, August 1, 1997.

characteristics or its preparation in order to improve the quality. It is necessary to check and verify each of these steps to control the complete production.

We shall also consider the fact that infometric<sup>4</sup> macroanalyses have already been done for each of these subsubjects: in analytical phytochemistry (and extraction of natural compounds) and in agricultural sciences for food and cosmetics.<sup>5,6</sup> However, to our knowledge, these fields have not been compared to verify a microlevel analysis. This is what we present here.

After the identification of the specialists as described above, the second part of the investigation gives a bibliographic description of each author, so that one can rapidly identify and find him/her in databases. To improve both on the speed of analysis and quality of presentation of the results, this step has also been done by computer.

## RESEARCH OF BASIC INFORMATION

Answering such a question for an information science laboratory involves the use of external databases. It is clearly impossible to have internal databases on every industrial subject. The same methodology could be used in another specific field. The first step of this type of study is to determine which commercial databases contain enough information and to elaborate the data collection strategy. This should be done with the help of a specialist in the field of study. Step two involves using a maximum of on-line indicators<sup>7</sup> (such as the Dialog RANK<sup>8</sup> (Dialog Information Services, Inc.; Palo, Alto, CA) statistical command) in order to minimize the number of downloaded references. In this way, the methodology leads to a considerable reduction in both time spent and cost.

**Data Collection Strategy.** We now show the method we will use to access public information. Other forms of information retrieval have already been developed elsewhere.<sup>9</sup> Here we use different parameters to evaluate our results.

In this study, the three areas for investigation deal with natural fragrances and perfumes. A specific vocabulary, defined to give some bibliographic homogeneity, was applied for the purpose of this work. The three queries executed all include the element

[(essential(w)oil??) or resinoid?? or fragrance??]

Then, each area is further defined with the use of keywords commonly assigned to scientific publications dealing with phytochemistry. The result is presented in Table 1.

**Selection of the Different Databases.** (a) **Query on Dialog Server using the "Dialindex" Database.** As we have already stressed, our investigation is separated into three domains. Hence, the resulting challenge resides in collecting bibliographic references that are as homogeneous as possible. This is required in order to be able to effect the subsequent data treatment. It was evident that our three domains were too disparate to be covered by a single database. In order to select the right databases, the Dialog server's database—DIALINDEX—was used. This option indicates how many references about each domain are in each Dialog database. It is then possible to determine the most relevant databases for a specific subject. The requests applied are those described above.

The result of this investigation is shown in Table 2.

**Table 1.** Requests Used for Collecting Information on the American Server Dialog

subject	request
agricultural production of the plants and varietal selection	[( <b>ESSENTIAL(W)OIL??</b> ) or <b>RESINOIDS</b> or <b>FRAGRANCE??</b> ] and [( <i>PLANT(W)PRODUCTION??</i> ) or ( <i>VARIETAL(W)SELECTION??</i> ) or <i>HARVESTING??</i> ]
essential oils and balms production	[( <b>ESSENTIAL(W)OIL??</b> ) or <b>RESINOIDS</b> or <b>FRAGRANCE??</b> ] and [ <i>EXTRACTION??</i> or <i>PURIFICATION??</i> or <i>DISTILLATION??</i> ]
products analysis	[( <b>ESSENTIAL(W)OIL??</b> ) or <b>RESINOIDS</b> or <b>FRAGRANCE??</b> ] and [ <i>ANALYSIS</i> or <i>CHROMATOGRAPHY</i> or <i>NMR</i> ]

(b) **Clues for Choosing the Databases.** The choice of the databases is based on four criteria: numbers of answers, scope of the databases, "supplementary indexing" and time period. From these criteria, two databases were selected, and these are shown in Table 3.

(a) First, the databases which contain the largest set of answers are selected. That is why the "Dialindex" described above is used. Nevertheless, if only this criterion is considered, three different databases should be kept and, thus, three types of data from a bibliographical point of view. This single criterion is not enough. We are looking for sets of references in order to use them in automated treatments, so data as homogeneous as possible is required. That is why databases which cover the same kinds of documents (i.e. only bibliographic references or only patents) must be kept.

(b) We wish to answer a question which belongs to a scientific field, with more or less technical considerations. Our investigation therefore has to rely on *general* scientific literature. In this way with a database such as the "US Pat. Fulltext" (US Patent Fulltext: Knight Ridder Information, Inc., Dialog customer services, Mountain View, CA) only treating patents is not suitable for the domain extraction, distillation, and purification. The kind of database we need to consider should contain bibliographic references of scientific articles, conference papers, and some patents.

(c) Our investigation aims at identifying the authors of scientific literature who fit into the three criteria cited above. Therefore, the analysis is run on the "Authors" bibliographical field of the references. We have stated that we want to use the on-line command "RANK". This is only possible if the DIALOG server has added a "supplementary indexing" to the database on the Authors field. This supplementary indexing has been added to some databases, and others not. If it has been, it is applied to different fields. We then need to check if the Authors field received it. That is why we prefer choosing the "CAB Abstract" (CAB International; Wallingford, Oxon, United Kingdom) database over "Agricola" (Agricola: Information Systems Division, NAL-USDA, Beltsville, MD), in which there is no supplementary index on the Authors field.

(d) The last criterion we shall consider for selecting the necessary databases is the time period covered by the database. We wish to consider similar periods of time, which run up to the present.

## AUTOMATED TREATMENT OF DATA

In this step of our study, we shall try to identify the authors common to the three preceding subjects.

**Table 2.** Result of the Query in Dialindex, To Identify the Databases from Which We Should Extract the Bibliographic Set

Ref.	Items	File
((ESSENTIAL(W)OIL??) OR /RESINOID?? OR FRAGRANCE??) AND ((VARIETAL(W)SELECTION??) OR HARVESTING?? OR (PLANT(W)PRODUCTION??))		
N1	303	10: AGRICOLA_1979-1995/Feb
N2	301	50: CAB ABSTRACTS_1972-1995/Feb
N3	221	203: AGRIS International_1974-1995/Jan
N4	76	648: Trade & Industry ASAP (TM)_1983-1995/Mar W1
N5	59	653: US Pat.Fulltext_1980-1989
((ESSENTIAL(W)OIL??) OR (RESINOID??) OR (FRAGRANCE??) AND (EXTRACTION OR DISTILLATION OR PURIFICATION)		
N1	1534	653: US Pat.Fulltext_1980-1989
N2	919	654: US Pat.Full_1990-1995/Mar 14
N3	872	399: CA SEARCH(R)_1967-1995/UD=12212
N4	855	652: US Patents Fulltext_1971-1979
N5	762	51: Food Sci&Tech.Abs_1969-1995/Mar
((ESSENTIAL(W)OIL??) OR (RESINOID??) OR (FRAGRANCE??) AND (ANALYSIS OR CHROMATOGRAPH OR NMR)		
N1	3281	399: CA SEARCH(R)_1967-1995/UD=12212
N2	2154	545: Investext(R)_1982-1995/Mar 21
N3	1683	144: Pascal_1973-1994/Aug
N4	1676	5: BIOSIS PREVIEW(R)_1969-1995/Mar W3
N5	1610	653: US Pat.Fulltext_1980-1989

**Table 3.** Final Selection of the Databases for Data Collection

subject	chosen base	no. of answers
agricultural production and varietal selection	CAB ABSTRACTS_1972-1995/Feb	340
essential oils and balms production	CA SEARCH(R)_1967-1995	938
products analysis	CA SEARCH(R)_1967-1995	3427

**Table 4.** Number of Downloaded Items (Names of Authors) per Subject

subject	no. of refs	no. of downloaded items
agricultural production and varietal selection	340	646
essential oils and balms production	938	2014
products analysis	3427	6506

**Extraction of the Three Lists of Authors, As Defined by the Relevant Subjects.** During the search in the databases, we utilize the on-line statistical command RANK allowed by the Dialog server on the Authors field for the CAS (Chemical Abstract Service (CAS), Columbus, OH) and CAB Abstract databases. After extracting the lists of the authors from the sets of references without downloading them, three lists of authors working in the following fields are obtained: the production and selection of plants; the extraction of essential oils and balms; the analysis of these oils. The names of the different authors which appeared in these lists are numerous (Table 4).

**Homogenization of the Data.** Our aim is to identify authors common to the three lists. Here the names present in the three lists are so numerous (Table 4) that we require specialized software to carry this. Moreover the names present in these lists are not homogeneous on two levels, at least from a bibliographic point of view.

(i) First, the names are classified in a decreasing frequency order. So in each list the same names are not in the same order (Table 5a). Also, they are not always common to the three lists. It is clear that you cannot manually compare the three lists of 6506, 2014, and 646 different names.

(ii) By the same token, if ever the name of an author is written, once with the complete first given name and another time with only the initial of the first given name, our

automated treatment will mistakenly consider the name(s) of this same person as two different items (Table 5b).

This is the reason why the first step of the analysis gives a homogeneous construction of the terms to the three lists: the name followed with the initial of the first given name, as shown in the Table 5c. This step is performed with a commercial software package for reformatting data (Infotrans from I+K).

It is important to note here a possible inconvenience of this reformatting step that could lead to a risk of increasing synonymy.

If this treatment of the data was not carried out, then the software would not recognize the same author's name written in different ways. Many correspondences could otherwise be missed, resulting in a situation which could not be remedied later.

Authors like "WATANABE, YOHYA" and "WATANABE, YOSHIHIRO" would both be represented by the form "WATANABE, Y." This means that from an infometric point of view one name could represent two different people. If these two authors appear in two different lists, too much data will be generated. A method for checking and validating the results will be discussed later.

Then comes the removal from this lists of elements which appear several times. This step was performed with the aid of MS World macrocommands. This resulted, after reformatting the names, in a slightly decreased number of items in the lists, shown in Table 6.

**First Step: Identification of Interesting Authors. (a) Automated Treatment.** We now discuss our method of analyzing the data. It is important here to keep in mind that the primary aim of this study is to identify the chemists who are experts in each of the processes of production and analysis of essential oils and balms. "Authors" who are qualified as "interesting" are those who have published articles on these three subjects:

\* the agricultural production of the plants and their varietal selection

AND \* the extraction of oils and balms from these plants

Table 5

(a) Samples of the Information Downloaded from the Dialog Server								
agricultural production and varietal selection			oils and balms production			products analysis		
no.	ranked	term	no.	ranked	term	no.	ranked	term
1	12	PALEVITCH, D.	1	15	SHLYAPNIKOV, V. A.	1	26	SUZUKI, SUKEJI
2	10	PUTIEVSKY, E.	2	12	GEORGIEV, E.	2	22	MOSANDL, ARMIN
3	10	SIMON, J. E.	3	8	SHLYAPNIKOVA, A. P.	3	21	KIJIMA, KEIJI
4	8	MATHE, A.	4	7	BALINOVA-TSVETKOVA, A.	4	21	SHCHEDRINA, M. M.
5	7	GORINI, F.	5	7	PERINEAU, FRANCIS	5	20	GAGLIARDI, L.
6	5	BORDOLOI, D. N.	6	6	DELMAS, MICHEL	6	20	MASLOWSKA, JOANNA
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
646	1	ZWANEPOL, S.	2014	1	ZVYAGINTSEV, A. YA.	6506	1	ZWETKOV, R.

(b) Raw Information Downloaded from the Dialog Server								
no.	ranked	term						
1	26	SUZUKI, SUKEJI			one complete given name			
2	22	MOSANDL, ARMIN C.			one complete given name followed by the first letter of the second name			
⋮	⋮	⋮						
4	21	SHCHEDRINA, M. M.			first letter of two given names			
5	20	GAGLIARDI, L.			first letter of one given name			

(c) Example of the Reformed Upper List <sup>a</sup>								
	26	SUZUKI, S.		21				SHCHEDRINA, M.
	22	MOSANDL, A.		20				GAGLIARDI, L.
	21	KIJIMA, K.		⋮				⋮

<sup>a</sup> The transformation consists of keeping the first letter of the first name.

Table 6. Comparison of the Numbers of Downloaded Names and of Reformatted Forms of These Names for Each Subject

subject	no. of downloaded names	no. of reformatted items
agricultural production and varietal selection	646	637
essential oils and balms production	2014	1880
products analysis	6506	5665

AND \* the analysis of the oils and balms

The use of Boolean operators such as AND or OR evidently cannot be applied here to answer this question. There are two main difficulties in using these operators. In the first instance, the AND boolean operator will just give the rare papers dealing with the three subjects. Specialists who wrote several papers (one in each subject) could not be identify with this method. In the second instance, the OR boolean operator will give all specialists in each of the three subjects. It may be possible to achieve part of the comparison by using another method. First, a list of authors is made on-line on a first subject, and then this list is transferred to another database to build queries on a second and finally on a third subject. This is clearly time-consuming and may be beyond the memory capacity of the host. Moreover, as the authors' names are sometimes written differently from one database to another, many correspondences may be lost.

A simpler method was therefore chosen. We compared the three lists, taking out the names of the authors who worked in the three domains. This action was performed by software developed in our laboratory.<sup>10</sup>

An example of the results is presented in Table 7.

These kinds of bibliometric indicators, resulting from a comparison of lists, have already been performed with keyword lists.<sup>11</sup>

Only 30 authors are represented in all three lists. With the three sets of names containing 5665, 1880, and 637

Table 7. Example of the Resulting File after an Automated Treatment<sup>a</sup>

name	presence in list 1	presence in list 2	presence in list 3
AGARWAL, S	1	1	1
ATAL, C	1	1	1
BALINOVA-TSVETKOVA, A	1	1	1
BHATTACHARYYA, S	1	1	1
BICCHI, C	1	1	1
CHANG, L	1	1	1
COLLIN, G	1	1	1
COTRONEO, A	1	1	1
DUGO, G	1	1	1
FLEISHER, A	1	1	1
HANOVER, J	1	1	1
HETHELYI, I	1	1	1
⋮	⋮	⋮	⋮
ADZET, T	1	0	1
⋮	⋮	⋮	⋮
BHATTACHARYA, A	1	1	0
⋮	⋮	⋮	⋮
AIT IGRI, M	0	1	1
⋮	⋮	⋮	⋮
ABBOTT, P	1	0	0
⋮	⋮	⋮	⋮
total no. of names/list	637	1880	5665
total no. of names in all lists: 7425			

<sup>a</sup> In the last three columns, you can observe the presence or the absence of an author's name form in one of the three lists, specified by 0 or 1.

different forms, respectively, it is clear that comparison cannot be performed manually in a reasonable time.

**(b) Validation of the Results.** As indicated above, the problem of synonymy has to be considered. It has been assumed that statistical treatments involve errors or that statistical representations give a view of reality though still containing errors. In this investigation, we need to validate our results because of the nature of the data. How can one be sure if a form (like WATANABE, Y.) represents only one person in our set of bibliographic references or several?

**AUTHOR: BALINOVA-TSVETKOVA A****Affiliation Laboratory:**

Institut za Zashchita na Rastenyata, Kostinbrod, Bulgaria.

**Worked in collaboration with the following laboratories:**

Aromat. Med. Cult.;Inst. Oil-Bear. Roses;Kazanlik;Bulg.  
 Central Experimental Station of Selection and Cultivar Maintenance of Diagn.;Plovdiv;Bulg.  
 Department of Toxicology, Plant Protection Institute;2230;Kostinbrod;  
 Higher Inst. Food Ind.;4000;Plovdiv;Bulg.  
 Inst. Oil-Bear. Rose, Ethereal Oil Plants Med. Plants;Kazanluk;Bulg.  
 Inst. Plant Prot.;Sofia;Bulg.  
 Institut po Maslodainata Roza, Kazanluk, Bulgaria.  
 Institute of Plant Protection;2230; Kostinbrod, Bulgaria.  
 Opitna Stantsiya po Yagodoplodni Kulturi, Kostinbrod, Bulgaria.  
 Tsentralna Opitna Stantsiya po Sortopodd"rzhane i Diagnostika,  
 Vissh Inst. Khranit. Vkusova Prom.;Plovdiv;Bulg.

**Publication Years:**

1972 to 1992

**Used Keywords: (freq > 2)**

analysis	metolachlor
application	Metribuzin
benomyl	plant residues
chemicals	residues
composition	simazine
crops	small fruits
damage	soil
dill	strawberries
essential oils	Tomatoes
fennel	usage
herbicides	vegetables
Lenacil	weed control

**CAS sections:**

CA011XXX_Plant Biochemistry	CA205001_Agrochemical Bioregulators
CA062002_Essential Oils and Cosmetics	CA219XXX_Fertilizers, Soils, and ...

**Papers and congresses:**

Cpt. Rendus de l'Acad. Bulg. des Sci.	Khimiya i Industriya Pesticides Abstracts
Dokl. Bolg. Akad. Nauk Bulg.	Khimiya v Sel'skom Khozyaistve
Dokl. S-kh. Akad., Sofia	Ovoshcharstvo
Flavour Fragrance J.	Pochvoznanie, Agrokimiya i Rastitelna Zashchita
Fresenius. J. Anal. Chem	Rasteniev'dni Nauki
Gradinarska i Lozarska Nauka	Rastenievud. Nauki
Int. Congr. Essent. Oils, (Pap.), 6th & 7th	Riv. Ital. EPPoS
J. Chromatogr.	Weed Res

**Figure 2.** Example of an author bibliographic check list.

To answer this question, we first came back to the lists in their downloaded forms ranked in alphabetical order. Sometimes the first names were fully written, so forms which corresponded to several authors could be removed. From the 30 names issued from the comparison of the lists, 24 were retained. An expert in phytochemistry was then able to identify certain authors recognized as specialists in this field. The validation of the remaining unknown names needed the creation of an internal database containing more information about each person. In this step, the references related to the selected 24 authors were downloaded and crossed with the three subjects. Instead of downloading at the beginning about 4705 (=3427 + 938 + 340) references, less than 200 references were downloaded and paid for (this represents about 25 times less than the original set).

With this internal database, different indicators can be used to find out if one interesting name represents one or several authors. First, one consults the complete bibliographic data to see if the initial of the first name corresponds to the same

first name. If it is the case, then some other indicators may help (e.g. the dates of publication: scientific articles are written by an individual over a periods of, at a maximum, 35 years).

Another indicator is the nationality of an author: here we can consider that an author mainly works in the same country. If ever he changes country, we can consider that in one period he worked in one country and in another period, in another country, but not in both at the same time. On this point, one should mention that during a period of teaching practice or a sabbatical, the affiliation laboratory may change. Also only the first laboratory will appear in the bibliographic data even though several are mentioned in the article. Here, personal intuition and judgement may be necessary.

Two people with exactly the same last name and first name work in the same laboratory, it is clear this will not be revealed through informatic data analysis.

We now focus on the form "Dugo, G". From a study of the lists this name obviously identifies two authors: "Dugo, Giovanni" and "Dugo Giacomo". They work in the same laboratory, with the same co-workers on the same subjects. The difference appears in the period on which they work: Dugo, Giovanni worked from 1967 until 1994 and Dugo, Giacomo published articles on a period of 4 years only. An interpretation of this could be that Giacomo worked as a student in the same laboratory as Giovanni. Even if there were obviously two experts under the form "Dugo, G.", we decided to keep this form as identifying "Dugo, Giovanni". Indeed we can say that he worked on the three subjects of interest. Otherwise, publication activity of Dugo, Giacomo is not as relevant compared to that of Dugo, Giovanni.

This kind of validation of results must be performed manually with the help of an expert in the field. Obviously, it is easier to do this for a short list. If the common part to the three themes seems to short, it is always possible to work with two field lists. Notice that these lists include a total of **7425** different authors (see Table 7).

**Second Step: Authors Bibliometric Check List.** As mentioned above, we analyzed 4705 references at low cost, of which less than 200 references downloaded were really relevant to the investigation. Thus, we can run another analysis on our internal database references. Indeed, always in order to identify the specialists of "natural fragrances", we need additional information which is accessible in the complete references.

Now we can describe an author from a bibliographic point of view: in which laboratory he mainly works; who are his co-workers; in which period he works; which keywords he uses; in which section of the CAS database his articles appear; in which journals and congresses he usually publishes his works.

All of this information is automatically extracted from the references downloaded from the Dialog server, with the DATAVIEW software.<sup>12</sup> Then, we automatically present this new elaborated information for each author on a user-friendly check list (see Figure 2), by using Winword "macrocommands". This is a very good example of elaborated information obtained from publicly available information.

## CONCLUSION

We have presented a methodology and the results of a specific infometric survey applied to scientific research in phytochemistry. This investigation was performed with chemists in order to identify and describe authors who were specialists in the production of high-quality natural perfumes. There were two major interests from an infometric point of view in this investigation. The first was the original way of splitting the investigation into three subjects in order to compare them. The second one was the totally automated chain of processing, from the collection of publicly available information to the presentation of the reduced information of higher value.

One aim of this study was to answer a specific question at low cost price. We demonstrated how we reduce the overall cost of the investigation by using on-line tools of a commercial server. This allowed us to target less than 200 relevant references downloaded selectively from a set of 4705 references.

This investigation was performed at a microlevel but involved a network of experts. Indeed, we needed experts

of scientific information management to collect the data, experts of the automated information treatments, and also specialists of the studied fundamental field in phytochemistry.

Only 30 authors who combine the three criteria of selection were identified. This is due to the fact that few people write articles on the three subjects. The specialist who asked for this study may consider our investigation too specific. However, the same methodology can be applied to only two criteria—chosen by a phytochemist—and the number of answers (names of authors) will increase. On the other hand, our results can be regarded as a starting point for more advanced investigations. A second step could consist of contacting the identified authors, as they will probably know other scientists in the same field.

From the methodological point of view, we have demonstrated that our starting supposition was a valid one. There are indeed authors who publish in all three areas, even if the results are surely incomplete. We used statistical tools, and we are not expected to provide exact representations of reality.

Moreover, the methodology based on only three areas of expertise can be increased to  $n$  parameters of selection in another investigation or field of application. It should be noted however that increasing the set of conditions will probably decrease the number of selected items.

This methodology can be applied for different purposes, e.g. for identifying interesting journals covering several subjects or for identifying laboratories with particular expertise.

## ACKNOWLEDGMENT

We would like to thank Dialog and Jonathan S. Keogh for the gracious help they provided during this work.

## REFERENCES AND NOTES

- (1) Dou, H.; Hassanaly, P.; Quoniam, L.; La Tela, A. Strategic patent clusters obtained from non boolean logic. *World Pat. Inf.* **1990**, *12* (4), 221–225.
- (2) Bordons, M.; Fernandez, M. T.; Gomez, I. Scientific cooperation intra-muros, extra-muros and extra-frontiers. Which is the goodness of each? *Fourth International Conferences on Science and Technology Indicators*, Anvers, Belgium; Oct 5–7, 1995; DSWO-Press: Leiden, 1996; pp 71–74.
- (3) Hinze, S.; Grupp, H. Application of bibliometric methods for science and technology policy purposes: Bibliometric mapping of microsystems engineering and international actors in the field. *Fifth international conference of the international society for scientometrics and informetrics*, Chicago, IL, June 7–10, 1995; Learned Information, Inc.: Medford, NJ, 1995; pp 225–234.
- (4) Egghe, L. Bridging the gap: Conceptual discussions on informetrics. *Scientometrics* **1994**, *30* (1), 35.
- (5) Munoz de la Pena, A.; Pulgarin, A. Research in analytical chemistry in Spain today, through an evaluation of the scientific literature. *Anal. Proc.* **1992**, *29* (12), 517–520.
- (6) King, J. A review of bibliometric and other science indicators and their role in research evaluation. *J. Inf. Sci.* **1987**, *13* (5), 261–276.
- (7) Christensen, F. H.; Ingwersen, P. Fundamental methodological issues of data set creation online for the analyses of research publications. *Fifth international conference of the international society for scientometrics and informetrics*, Chicago, IL June 7–10, 1995; Learned Information Inc.: Medford, NJ, pp 103–112.
- (8) *Dialog Pocket Guide*; Dialog: Palo Alto, CA, 1994; p 37.
- (9) Salton, G.; McGill, M. J. *Introduction to Modern Information Retrieval*; McGraw Hill Company: New York, NY, 1988.

- (10) Dou, H.; Hassanaly, P.; Quoniam, L.; La Tela, A. Post Processing of Online Searches. *Cah. Document. Belg.* **1990**, 3, 51–70.
- (11) Sawson, D. R.; Smalheiser, N. R. Computer-assisted knowledge discovery in bibliographic databases. *Fifth biennial conference of the international society for scientometrics and informetrics*, Chicago, IL, June 7–10, 1995, Learned Information Inc.: Medford, NJ.
- (12) Rostaing, H.; Nivol, W.; Quoniam, L.; La Tela, A. The DATAVIEW software and its use as tool for concurrency evaluation. *4th days on elaborated information, congrès of Ile Rousse, French infometric journal.* **1993**, 12, 360–387.

CI9601067