Domain Analysis for the Construction of a Conceptual Structure: A Case Study

BSc. Ricardo Arencibia-Jorge[1]
BSc. Rosa Lidia Vega-Almeida[2]
MSc. Yohannis Martí-Lahera[3]

[1] Network of Scientometric Studies for Higher Education. Department of Scientific Information. National Center for Scientific Research. Havana City, Cuba.
ricardo.arencibia@cnic.edu.cu
[2] Health Analysis and Trends Unit. Ministry of Public Health. Havana City, Cuba.
vega.rosa@infomed.sld.cu
[3] Faculty of Communication. University of Havana. Havana City, Cuba.
yohannis.marti@fcom.edu.cu

Abstract

This work describes the construction of a conceptual structure, as a result of the domain analysis of the current research on Benign Prostatic Hyperplasia (BPH), through different techniques based on the citations and semantic analysis. A sample which covers all published articles about the disease in mainstream journals was chosen. A total of 1,968 articles from 2000 through 2004 were retrieved from Science Citation Index. Author Co-citation Analysis (ACA) techniques and semantic analysis were used and the results were represented by different informatic programs. A macrostructure of the domain of the current research on BPH was identified through bibliometric techniques. The text mining techniques allowed validation of the identified macrostructure and the obtainment of the most frequent words in the text. The semantic analysis of the most cited reviews on BPH during the studied period allowed the definition of the categories to be used in the structure. Finally, a conceptual structure to be used as controlled (structured) language in the information retrieval, inside of a specialized information system on the approach of the disease is shown.

Keywords: *domain analysis, knowledge organization, conceptual structure, bibliometrics, citation analysis, semantic analysis, text mining, Prostatic Benign Hyperplasia.*

Introduction

The idea that the knowledge organization is based on the active role of the community of users who are involved in cognitive strategies for a good information retrieval in concrete social environments, leads to the fact that both informational processes should circumscribe themselves to contexts or specific domains. In this sense, it is a necessary consideration that the conceptual world of the individual, the acquisition of concepts and the *linguistic work* should be explained by the common social division of labour, as expressed by Hjørland and Albretchen (1995).

That is why the homogeneous groups are especially determinants in their own domains, which are conceived like social-cultural spaces where integrity, singularity and unit are observed, and where users ascribe and become owners of *socialized* meaning.
Therefore, the need to build these micro-spaces of interaction through communication defined as *collision of semantic structures* (Nazaretian, 1976), which provokes the fusion of cognitive horizons when it comprises the informational object as well as the users community should be admitted.

Then, to understand how knowledge and its meaning are built implies focus not only on the cognitive dimension but also on the social dimension. Congruously, the replacement of *knowledge* by *demanded knowledge*, and the understanding of the representation – organization – process in the diverse fields of knowledge, especially in connection with the behaviour of members of the productive community, the specialized language of the area, and the information users (Hjørland, 2000; Broughton Hansson, Hjørland & Lopez-Huerta., 2005) is important to recognize.

However, to incorporate the user's perspectives and their realities, implies new challenges for the information management, and the design of information retrieval systems. This suggests, as strategy for the document analysis, the combination of bibliometric techniques and speech analysis, with the incorporation of semiotic approaches and *critical-hermeneutical* approaches among others, for the study of a specific thematic domain (Glenisson, Glänzel, Janssens & Moor, 2005; Schneider & Borlun, 2004; Broughton et. al., 2005). The main objective of these strategies is the representation of knowledge without isolating it from the social context (Lafuente, 2001).

The bibliometric techniques, applied in different researches for the creation of knowledge organization systems, can be used in combination with other analysis techniques, with the aim to achieve a pertinent conceptual structure characterized by its flexibility, plasticity and efficacy.

Different authors have used bibliometric methods for construction of knowledge organization schemes (Rees-Potter, 1989; Pao & Worthen, 1989; Pao, 1993; Schneider, 2004; Glenisson et. al., 2005; Broughton et. al., 2005). The common argument of their work is that terminology used in the citation context of citing papers reflects concepts of a given specialty area (Schneider & Borlun, 2004). Usually, the employed methods aim to cluster topically related documents through co-ocurrence analysis of references and citations; which have the potential to cluster documents of the same topic that differ in choice of terminology. Some critics, however, do not feel the ACA is a useful tool for mapping *subject matter specialties* of scientific research; King (1987) stressed that co-citation analysis implies the lost of relevant papers, inclusion of non relevant papers, overrepresentation of theoretical papers, lack of emergent topics in co-citation maps, and subjectivity inherent in the setting of thresholds levels, as these thresholds levels strongly affect size and content of clusters. That is why, different authors like Braam, Moed & Van Raan (1991), and Small (1997) suggest the use of co-word in addition to co-citation.

Co-words analysis was developed by Callon, Law, Rip, Latour , Courtial and Whittaker in the later 1980s. Since its introduction, co-words analysis has been applied to detect topics in a given research area, the relationship between these topics, the extent to which they are central to the whole area, and the degree to which they are internally structured (Schneider, 2004). Co-words analysis allows the obtainment of a proximity map which shows the connection between ideas and concepts hidden behind the central themes in a research area (Callon, Law & Rip, 1986). It is possible because co-words analysis makes use of words frequency pairs in the whole set of papers. Therefore, this kind of analysis is a complement to co-citation analysis in the construction of a knowledge organization scheme.

Nevertheless, the use of co-words analysis has also been criticized. Leydersdorff (1997) argues that words and co-words cannot map the development of science, because words change positions not only in terms of the dimensional scheme of *theoretical*, *methods* and *observation results*, but also change in meaning from one text to another. In opposition to Leydersdorff's opinion, Courtail (1998) argues that what is important for co-words analysis is not the exact meaning or definition of the words, but the fact that these words are linked to others. Courtail's comment is based on the definition of the semantic analysis as a phase of natural language processing, following parsing, which involves extraction of context-independent aspects of a sentence's meaning, including the semantic roles of entities mentioned in the sentence, and quantification information, such as cardinality, iteration, and dependency (*Science and Technology Dictionary*, 2003).

In order to recognize the meaning of the words in a set of analyzed documents, it is important to judge some aspect of the words, which are only implicit in context (full text). The examination of the utilization context of the candidate words to be included into knowledge organization schemes, not only allows the observation of the specificity and uniformity of the meaning of words assigned by authors, but also the construction of representative noun phrases from the domain. When there is a huge amount of documents for full text analysis purposes, it is suggested that only the most cited should be selected. This consideration is based on the idea that highly cited documents symbolize concepts to those who cited them (Small, 1978).

All of the methods presented above could be used in the generation of conceptual structures, a special kind of knowledge organization scheme used in online information retrieval systems.

The identification of a cognitive structure related to the analysis of a scientific discipline or determined field of knowledge, in this case the research on Benign Prostatic Hyperplasia (BPH), to determine the lines of work developed in the world for their study, as well as the basic conceptual structures, should take into account not only the conceptual dynamic related to the study of the disease, but also the different approaches related to the methodology for the diagnosis, follow-up and treatment.

This paper attempts to identify a conceptual structure related to the current research on BPH, based on the domain analysis theoretical perspective and through the fusion of bibliometric and semantic analysis techniques. The main objective is its possible use as a controlled (structured) language for the information retrieval within a specialized information system in dealing with the disease for researchers from the National Scientific Research Centre (CNIC).

Method

For the construction of the conceptual structure, the domain of the most current research on BPH was examined. As an analysis sample, the articles published in mainstream journals were chosen, and techniques to identify citation and semantic patterns were combined.

As an information source, the *Science Citation Index* database was used, through the *Web of Science* on-line service offered by the Institute for Scientific Information (ISI), placed in Philadelphia, in United States. *Science Citation Index* comprises of over 5 000 titles from serial publications. The choice was mainly based on the need to analyze citation patterns, which is impossible to carry out in *Medline*, the main database from biomedical domain. However, the core journals which publish articles on BPH are comprised in both databases.

The sample articles were obtained in May 2004, through a search strategy which consisted of the identification of the words *Benign*, *Prostatic* and *Hyperplasia* in every available field, and which allowed the retrieval of 1 968 articles corresponding to the period between January 1st, 2000 and May 15th, 2004.

The number of articles corresponding to 2004, according to the date when the search was made, only comprises 4 months of the year, although this factor did not affect the aim of the current study.

A bibliographic database using the *EndNote* program was created. The *EndNote* program, developed by ISI, allowed the importation of retrieved registers for normalization (error control) of the fields observed in the study; and the analysis of articles comprised in the different clusters, obtained by means of bibliometric techniques used.

Once the fields were normalized, we began to create the lists of years and journals which were most productive, as well as the authors and articles which were most cited by the scientific production 2000-2004, and the articles published in that period with a greater amount of citations. The lists were exported to text files which were processed with the Microsoft *Excel* program in order to generate the corresponding tables and figures.

The identification of the research fronts about BPH was carried out through the Authors Co-citation Analysis (ACA), following the traditional methodology proposed by White and McCain (1998). Those authors with 100 or more citations were analyzed using the *Bibexcel* program (developed by Olle Persson at Umea University, Sweden) for the obtainment of the co-occurrence matrix from the authors in the field *Cited references*. Besides, the programs *STATISTICA 5* and Microsoft *Excel* were used for graphical representation of the matrix, obtaining visualizations by means of multidimensional scaling techniques (MDS) and cluster analysis.

The semantic analysis of the articles was developed by means of the *RefViz*^{TM} program, also developed by ISI, which allowed the obtainment of a proximity map of the articles according to the frequency of words appearing in the text, as well as the creation of the lists of primary descriptors identified in the most remarkable topic groups created by the program. The construction of the noun phrases for the structure, is based on the determination of syntactic, semantic and pragmatic relations from the most frequent terms obtained with the program, through the analysis of the full text from the five most cited reviews about BPH during the studied period (Barry & Roehrborn, 2001; McConnell, Bruskewitz & Walsh, 1998; Gormley, Stiner & Bruskewitz., 1992; Chokkalingam et. al., 2003; Guess, 2001).\

The proposed methodology has the following steps:

I. *Definition of the macrostructure:*

1. Selection of the most cited authors in the scientific production corresponding to the evaluated period.
2. Co-citation analysis of the most cited authors (ACA). Obtaining the hierarchic dendrogram that expresses the proximity of the authors.
3. Analysis of the obtained author's clusters. Determination of the main research fronts that identify each cluster. Definition of the macrostructure.

II. *Validation of the macrostructure and determination of the thematic area's weight (triangulation method)*

4. Selection of the most cited articles by the rest of the articles during the studied period as well as the most cited articles produced in the period. Identification of the shift trends and established pattern in the domain.
5. Text mining in order to identify thematic clusters within the scientific production. Obtaining lists of the most frequent terms in each cluster.
6. Determination of the main topics in the most productive clusters. Checking the correspondence between main topics obtained by text mining and those obtained through ACA. Validation of the macrostructure (triangulation method).

III. *Construction of the conceptual structure*

7. Full-text analysis of the most cited reviews during the period. Examination of the context where the most frequent terms from the most productive clusters were employed.
8. Identification of the syntactic, semantic and pragmatic relations between the most frequent terms. Generation of noun phrases and its placement in the macrostructure.

Results and Discussion

The search strategy retrieved 1,968 articles related to BPH, in which it is obtainable the predominance of original articles (Table 1a), and the traditional hegemony of the English language (Table 1b).

a) Tipology	Articles	%	b) Languages	Articles	%
Original articles	1,604	81.50	English	1913	97.21
Reviews	181	9.20	French	29	1.47
Congress Abstracts	80	4.06	German	19	0.97
Editorials	72	3.66	Spanish	3	0.15
Letters to the editor	24	1.22	Russian	1	0.05
Corrections	7	0.36	Polish	1	0.05
			Portuguese	1	0.05
			Chinese	1	0.05

Table 1. a) Tipology of articles on BPH in the Web of Science 2000-2004.
b) Languages identified.

The 1,968 articles were published in 429 serial publications from biomedical topics, observing that the 12 most productive journals, mainly specialized in Urology (Table 2), constitute the 50.2 % from all the articles.

Journals	Country	Articles	%
Journal of Urology	USA	240	12.20
Urology	USA	186	9.45
European Urology	SUIZ	138	7.01
BJU International	UK	137	6.96
Prostate	USA	84	4.27
Journal of Endourology	USA	45	2.29
Urologia Internationalis	SUIZ	33	1.68
International Journal of Urology	JAP	27	1.37
Cancer Research	USA	26	1.32
Prostate Cancer and Prostatic Diseases	UK	25	1.27
Progres En Urologie	FRA	24	1.22
Scandinavian Journal of Urology and Nephrology	NOR	23	1.17

Total of Journals: 429

Total of Articles: 1,968

2.9 % of the journals (12) publish the 50.2 % of the articles.

Table 2. The most productive journals in the scientific production on BPH during the period 2000-2004.

Authors Co-citation Analysis (ACA)

According to the total amount of citations received by their articles, the authors represented in Table 3 constitute those who most influenced the research related to BPH during the five years comprised in the study.

Author	Articles 2000-04	Cited articles	Articles where he/she is cited	Citations	Self citations	Total of citations
Lepor H	12	20	327	620	30	590
Barry MJ	11	16	314	485	28	457
Roehrborn CG	52	23	245	478	106	372
McConnell JD	16	15	269	349	15	334
Abrams P	5	28	215	335	2	333
Chapple CR	11	17	175	265	5	260
Djavan B	19	10	179	282	48	234
Catalona WJ	0	15	110	219	0	219
Oesterling JE	1	10	163	205	1	204
Kirby R	15	20	157	186	9	177
Jacobsen SJ	13	12	113	191	33	158
McNeal JE	7	20	106	169	12	157
Berry SJ	0	4	146	150	0	150
Boyle P	23	14	141	173	23	150
DelaRosette JJMCH	43	29	127	193	43	150
Kaplan SA	23	14	118	148	7	141
Lukacs B	6	10	88	157	16	141
Garraway WM	0	4	126	140	0	140
Mebust WK	0	9	126	135	0	135
Schulman CC	11	15	103	126	7	119
Narayan P	7	13	104	124	8	116
Gormley GJ	2	8	105	119	4	115
Madersbacher S	22	9	97	146	34	112
Caine M	0	10	83	110	0	110
Girman CJ	12	9	88	123	13	110
Debruyne FMJ	14	9	99	111	3	108
Michel MC	12	11	66	146	39	107
Partin AW	11	15	78	106	1	105
Buzelin JM	1	5	68	105	1	104
Stamey TA	7	18	75	117	17	100

Table 3. The most cited authors by the scientific production on BPH during the period 2000-2004.

The majority of the most cited authors also produced a high number of articles during the period studied. This suggests that the main contributions for the study of the disease are very recent. The research around BPH, therefore, has a full up-to-date nature, which is based on the high incidence and progression of this disease at world level (Barry & Roehrborn, 2001).

The co-citation analysis among the 30 authors who received 100 or more citations during the analyzed period, allowed clear identification of two research fronts in which the scientific production is focused (Fig. 1). A graph was obtained, using the multidimensional scaling technique, where two perfectly defined clusters can be observed.

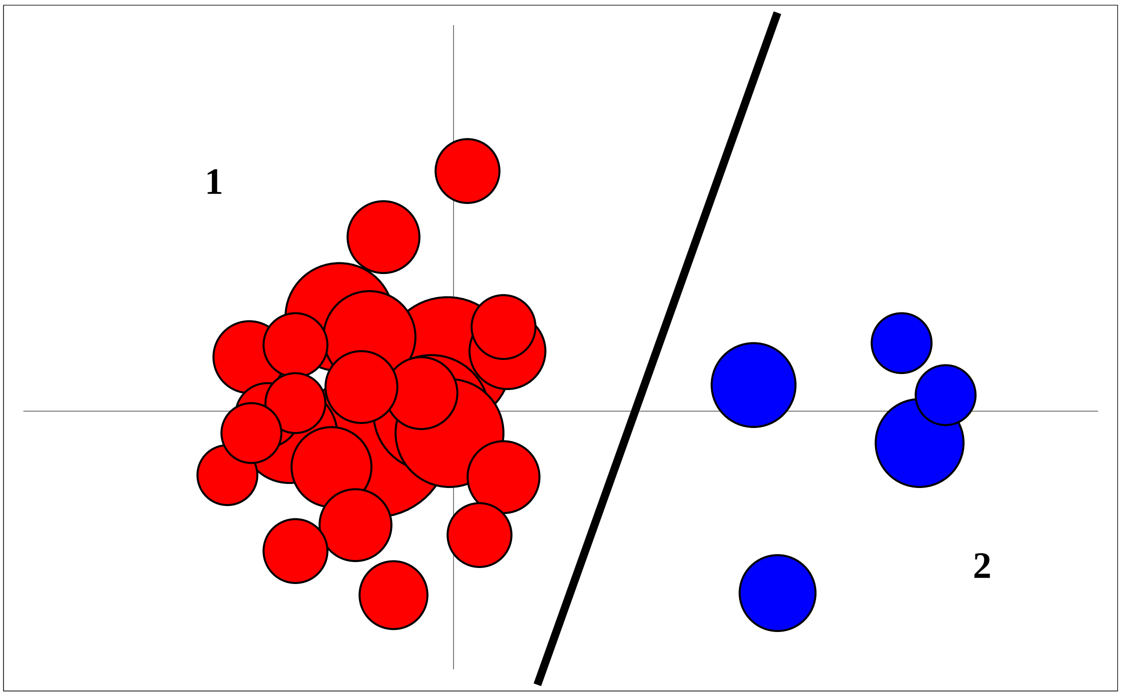

Figure 1. Obtained clusters from the most cited authors using multidimensional scaling techniques (STATISTICA 5, Microsoft EXCEL).

The majority of the most cited authors are grouped in sector 1. They are referenced in articles dedicated to the specific study of the disease, its ethiology and diagnosis, as well as the different therapeutical approaches for their treatment. In sector 2, the most co-cited authors in articles which comprise the relation between BPH and Prostate Cancer (PC) are gathered; identifying risk factors and developing genetic expression profiles and molecular markers for the early diagnosis of malign manifestations of the disease.

A clearer idea about the structure of the research on HPB stems from the cluster analysis formed by the 30 most cited authors, according to the hierarchical dendrogram presented in Figure 2.

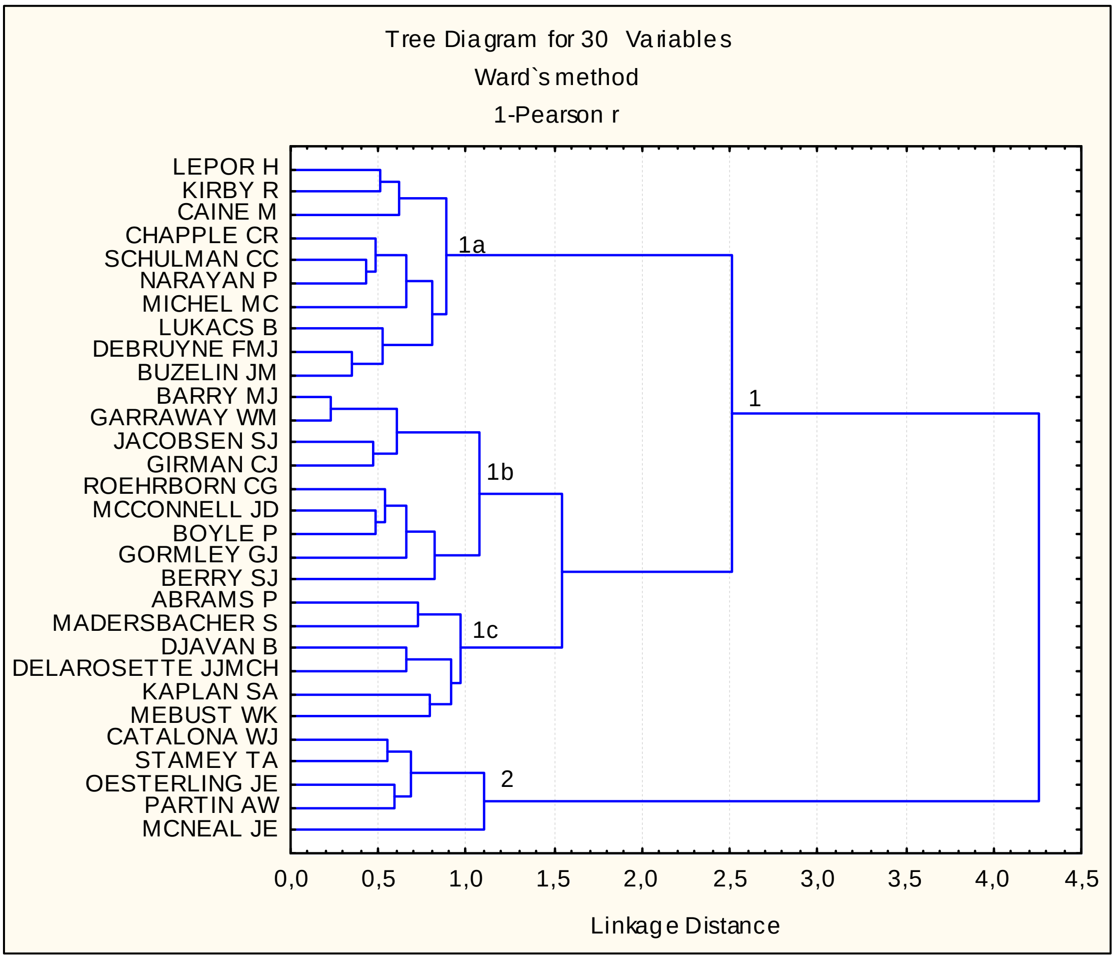

Figure 2. Hierarchical dendrogram of the most cited authors using techniques of cluster analysis (STATISTICA 5).

The most represented research front is subdivided into three well-defined sectors. Sector 1a is going to gather nine co-cited authors in articles which try to comprise different pharmacological therapies for fighting against BPH. In these articles, the study of drugs which are 5-α reductase inhibitors, such as Finasteride, and α-blockers such as Doxazosin, Terazosin, Tamsuzolin, Alfuzosin and Prazosin, and in a smaller number, drugs derived from natural products such as Saw Palmetto extract, β Sitosterol plant extracts, and rye grass pollen extracts. Among the most cited authors by these articles are doctors H. Lepor, from the Medical Center of the New York University in the United States, C.R. Chapple, from the Real Hospital of Hallamshire in England, and the also British R. Kirby, from the St. George Hospital, at London University.

Sector 1b concentrates research on diagnosis, ethiology, prevalence, and different epidemiological approaches on the disease, having as fundamental references other 9 authors, among them, Doctor M.J. Barry, from the General Hospital of Massachusetts in the United States is outstanding, along with other North Americans C.G. Roehrborn and J.D. McConnell, both from the University of Texas.

Finally, sector 1c gathers the researches on surgical therapeutical methods, where Transurethral Resection Prostatectomy (TURP), Transurethral Microwave Thermotherapy (TUMT), Transurethral Needle Ablation (TUNA), and Transurethral Incision of the Prostate (TUIP) are outstanding, among other non pharmacological methods that are more or less invasives. This sector is going to be represented by 6 authors headed by doctors P. Abrams, from the Urology Institute of Bristol, England, and B. Djavan, from Viena University in Austria.

The five remaining authors, headed by doctors J.W. Catalona and J.E. Oesterling, from St. Johns Hospital in Detroit, United States, are going to concentrate on the second research front, previously mentioned, devoted to the PC study and its link with BPH, emphasizing on the risk factors and on the use of genetic expression profiles and molecular markers for its diagnosis (Fig. 3).

1. **Benign Prostatic Hiperplasia**

   1a) Pharmacological Therapy

   - **Finasteride, Doxazosin, Terazosin, Tamsuzolin, Alfuzosin, Prazosin, etc.**
     Lepor, Kirby, Caine, Chapple, Schulman, Narayan, Michel, Lukacs, Debruyne, Buzelin.

   1b) Diagnostic and Epidemiology

   - **Etiology and Diagnostic, Primary Care, Prevalence, Risk Factors, Progression, Incidence, Prevention, Quality of Life, Costs, etc.**

   Barry, Garraway, Jacobsen, Girman, Roehrborn, McConnell, Boyle, Gormley, Berry.

   1c) Non Pharmacological Therapy

   - **Transurethral Resection of the Prostate, Transurethral Microwave Thermotherapy, Transurethral Incision of the Prostate, Transurethral Needle Ablation, Laser Prostatectomy, Electronic vaporization, Transrectal Ultrasound, etc.**
     Abrams, Madersbacher, Djavan, DelaRosette, Kaplan, Mebust.

2. **Prostate Cancer and Benign Prostatic Hyperplasia**

   - **Risk Factors, Diagnosis, Antigens, Genetic Expresión Profiles, Molecular Markers.**
     Catalona, Stamey, Oesterling, Partin, McNeal.

Figure 3. Macrostructure obtained through Author Co-citation Analysis (ACA).

Analysis of most cited articles

The articles on BPH with more visibility, and with a higher impact on the scientific community, according to the total amount of citations received, were divided into 2 groups. On one side the most cited articles during the period studied were included, this is, those articles that in a greater number of times, appeared as references in the articles produced between 2000 and 2004, standing as forced reference articles for the study of the disease (Table 4).

No.	Most cited articles	Times cited	%*
1	Barry, M.J.; Fowler, F.J.; O´Leary, M.P.; *et al.* The American Urological Association symptom index for benign prostatic hyperplasia. J Urology 1992;148:1549-1557.	143	0.22
2	McConnell, J.D.; Bruskewitz, R. and Walsh, P. The effect of finasteride on the risk of acute urinary retention and the need for surgical treatment among men with benign prostatic hyperplasia. *New Engl J Med* 1998;338:557-563.	119	0.18
3	Berry, S.J.; Coffey, D.S.; Walsh, P.C.; Ewing, L.L. The development of human benign prostatic hyperplasia with age. J Urology 1984;132:474-479.	114	0.18
4	Lepor, H. The efficacy of terazosin, finasteride, or both in benign prostatic hyperplasia. *N Engl J Med* 1996;335:533-539.	103	0.16
5	Mebust, W.K.; Holtgrewe, H.L.; Cockett, A.T.K., Peters, P.C.; and writing committee. TURO: immediate and post operative complications. A cooperative study of 13 participating institutions evaluating 3885 patients. J Urology 1989;141:243-247.	93	0.14
6	Garraway, W.M.; Collins, G.W.; Lee, R.J. High prevalence of benign prostatic hypertrophy in the community. Lancet 1991;338:469-471.	79	0.12
7	Gormley, G.J.; Stoner, E. and Bruskewitz, R.C. The effects of finasteride in men with benign prostatic hyperplasia. *N Engl J Med* 1992;327:1185-1191.	77	0.12
8	Chute C.G.; Panser, L.A. ; Girman, C.J. ; *et al.* The prevalence of prostatism: a population-based survey of urinary symptoms. *J Urol* 1993;150:85-89.	60	0.09
9	Lepor, H. Phase III multicenter placebo-controlled study of tamsulosin in benign prostatic hyperplasia. Tamsulosin Investigator Group. *Urology* 1998;51:892–900.	58	0.09
10	Djavan, B. and Marberger, M. A. Meta-analysis on the efficacy and tolerability of alpha1-adrenoceptor antagonists in patients with lower urinary tract symptoms suggestive of benign prostatic obstruction. *Eur Urol* 1999;36:1-13.	57	0.09
11	Jacobsen S.J.; Jacobson, D.J. and Girman, C.J. Natural history of prostatism: risk factors for acute urinary retention. *J Urology* 1997; 158:481-487.	57	0.09
12	Boyle, P.; Gould, A.L. and Roehrborn, C.G. Prostate volume predicts outcome of treatment of benign prostatic hyperplasia with finasteride: meta-analysis of randomised clinical trials. *Urology* 1996;48:398-405.	56	0.09
13	Debruyne, F.M.J.; Jardin, A. and Colloi, D. Sustained-release alfuzosin, finasteride and the combination of both in the treatment of benign prostatic hyperplasia. *Eur Urol* 1998; 34:169-175.	53	0.08
14	Stenman, U.H., J. Leinonen, H. Alfthan, S. Rannikko, K. Tuhkanen and O. Alfthan. A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. Cancer Res 1991;51:222.	52	0.08

* Percentage from the total of citations.

Total of articles: 1,968. Total of citations: 64,892. Average of references by article: 32.97

Table 4. Most cited articles about BPH during the period 2000-2004.

On the other hand, the articles published in the studied period with a greater amount of citations were included, which are outstanding because of the new scientific contributions and updates about the study of the disease and the research on BPH in general (Table 5).

No.	Most cited articles	Times cited	%*
1	Luo, J., Duggan, D. J., Chen, Y. D., Sauvageot, J., Ewing, C. M., Bittner, M. L., Trent, J. M., Isaacs, W. B. Human prostate cancer and benign prostatic hyperplasia: Molecular dissection by gene expression profiling. *Cancer Research* 2001;61(12): 4683-88. USA	105	1.25
2	Adam, B. L., Qu, Y. S., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z. D., Wright, G. L. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Research* 2002;62(13):3609-14. USA	78	0.93
3	Lau, K. M., LaSpina, M., Long, J., Ho, S. M. Expression of estrogen receptor (ER)-alpha and ER-beta in normal and malignant prostatic epithelial cells: Regulation by methylation and involvement in growth regulation. *Cancer Research* 2000; 60(12):3175-82. USA	76	0.91
4	Yoshimura, R., Sano, H., Masuda, C., Kawamura, M., Tsubouchi, Y., Chargui, J., Yoshimura, N., Hla, T., Wada, S. Expression of cyclooxygenase-2 in prostate carcinoma. *Cancer* 2000;89(3):589-96. JAP-USA	66	0.78
5	Kenny, A. M., Prestwood, K. M., Gruman, C. A., Marcello, K. M., Raisz, L. G. Effects of transdermal testosterone on bone and muscle in older men with low bioavailable testosterone levels. *Journals of Gerontology Series A-Biological Sciences and Medical Sciences* 2001;56(5):M266-72. USA	65	0.77
6	Irvine, R. A., Ma, H., Yu, M. C., Ross, R. K., Stallcup, M. R., Coetzee, G. A. Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Human Molecular Genetics* 2000;9(2):267-74. USA	62	0.74
7	Guimaraes, S., Moura, D. Vascular adrenoceptors: An update. *Pharmacological Reviews* 2001;53(2):319-356. Review. POR	61	0.73
8	Parks, L. G., Ostby, J. S., Lambright, C. R., Abbott, B. D., Klinefelter, G. R., Barlow, N. J., Gray, L. E. The plasticizer diethylhexyl phthalate induces malformations by decreasing fetal testosterone synthesis during sexual differentiation in the male rat. *Toxicological Sciences* 2000;58(2):339-49. USA	54	0.64
9	Matsumoto, A. M. Andropause: Clinical implications of the decline in serum testosterone levels with aging in men. *Journals of Gerontology Series A-Biological Sciences and Medical Sciences* 2002;57(2):M76-99. Review. USA	50	0.59

* Percentage of the citations received by the articles published during the period 2000-2004.

Total of articles: 1,968	Total of citations received: 83,97
Total of cited articles: 1,216	Average of citations by article: 4,27
Percentage of cited articles: 61.79 %	Average of citations by cited article: 6,90

Table 5. Most cited articles about BPH published during the period 2000-2004

The citations received from these articles, not only arise from those which form the sample studied, but from all the articles taken by the *Science Citation Index*.

The analysis from Table 4 allows us to reach the conclusion that the most cited articles by the scientific production evaluated have as their main topic the study of tolerability and efficacy of drugs for the treatment of BPH. The decade of the 1990s is a period in which the majority of the most cited articles deals with the use of alternative therapies for prostatectomy, specifically the administration of pharmacological agents among which Finasteride is outstanding, constituted the most active research line (Hamilton & Sharp, 2004; Clifford & Farmer, 2000). This fact was clearly reflected in the citation patterns. This result allows the obtainment of a more objective dimension from sector 1a obtained in ACA, granting a certain protagonism (Figures 2 and 3).

On the other hand, Table 5 analysis shows two emergent tendencies on the research about BPH. Taking into account that BPH symptomatology as well as PC, which is locally advanced, are similar (Roehrborn et. al., 2001), an important number of the researchers at present is devoted to study the relationship between both sufferings, and particularly the development of diagnosis methods which allow the early detection of both. Although it is true that the non-existence of an association between BPH and cancer has been proved (Simpson, 1997), many times both diseases have existed together. That is why without the diagnosis of BPH revealing the existence of a cancer the result is the evolution of the cancer continuous and the patient's condition worsing against the doctor's will (Stoevelaar & McDonnell, 2001). The research to solve this problem which is very common in medical practice is manifested in part of sector 1b, and mainly in sector 2 from Figures 2 and 3. The second tendency identified is directed to clearing up the biological mechanisms which influence the appearing of the disease, and is closely related to the development of new methods of diagnosis and therapeutical ones.

Semantic analysis

The semantic analysis of the 1968 articles comprised in the study allowed the complement of the results obtained through the analysis of the citation patterns. By means of software *RefViz*^{TM}, a figure which represents the scientific production on BPH was obtained in the evaluated period (Figure 4).

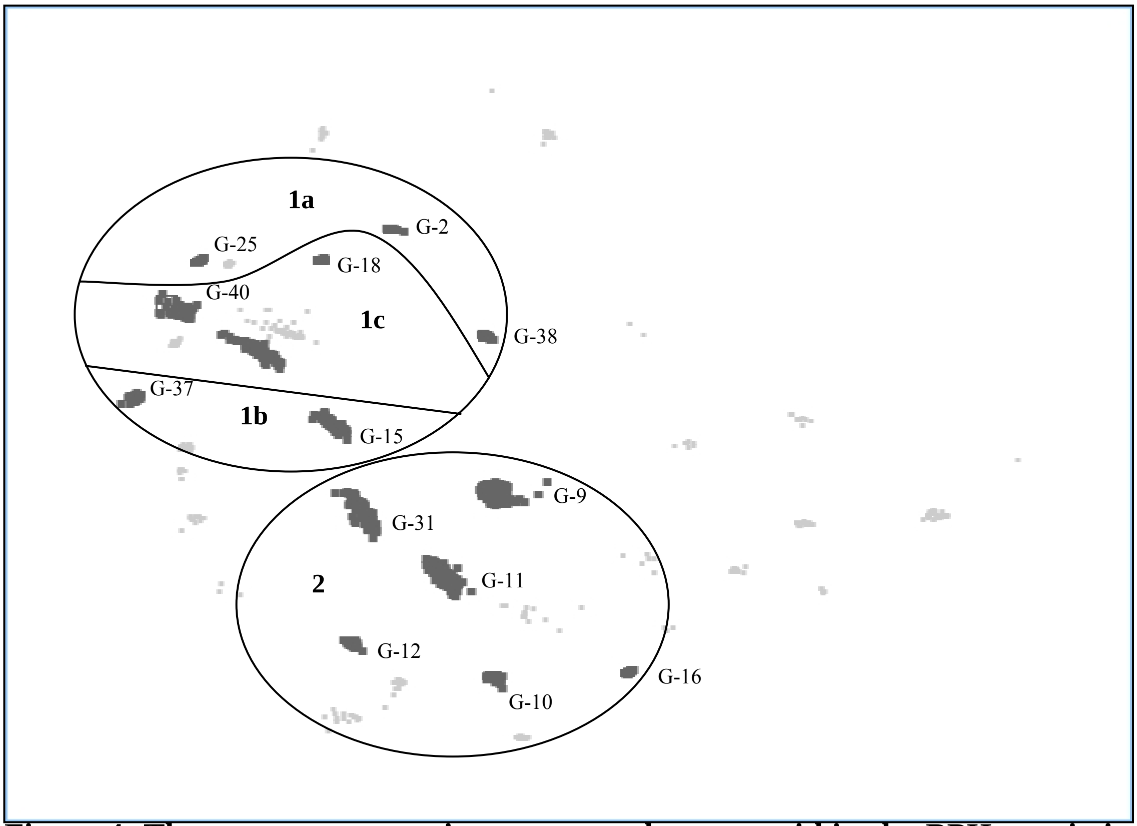

Figure 4. The most representative groups and sectors within the BPH proximity map (Text Minning, RefViz^{TM}).

The figure in itself constitutes a galaxy or proximity map formed by a total of 40 groups of articles, grouped according to the coincidence of words in the text. Each document is a point within the galaxy. The egg form of the galaxy implies the multiple varieties of topics from the scientific production on the disease. The groups with a darker grey constitute the topic groups with a greater amount of articles. The groups which are more productive were analyzed according to the primary descriptors with a higher frequency in the articles, which can be observed in Table 6.

Group 38	f	Group 9	f	Group 16	f	Group 11	f	Group 18	f	Group 4	f	Group 25	f
Tissue	48	tissue	70	cell	121	cancer	113	symptom	46	symptom	79	symptom	97
Symptom	32	cell	53	expression	80	risk	46	volume	34	score	58	score	59
Cancer	29	cancer	45	tissue	78	gene	42	score	26	volume	57	luts	37
Volume	28	expression	39	epithelial	73	tumor	37	risk	22	bladder	49	baseline	36
Growth	24	growth	39	growth	66	cell	35	cancer	21	obstruction	42	bladder	32
testosterone	24	receptor	38	normal	65	serum	34	bladder	19	urine	31	volume	31
cell	21	activity	31	cancer	60	ratio	31	prostatectomy	19	post-operative	27	placebo	28
finasteride	21	epithelial	31	stromal	55	tissue	29	finasteride	17	laser	24	void	26
androgen	20	normal	31	protein	52	psa	27	serum	17	baseline	23	obstruction	24
bladder	20	androgen	29	receptor	39	normal	23	turp	15	ipss	22	tamsulosin	23
concentration	17	protein	27	carcinoma	37	polymorphism	23	void	15	outlet	21	urine	21
normal	17	rat	27	culture	36	androgen	22	baseline	14	prostatectomy	21	alfuzosin	18
serum	17	testosterone	27	activity	32	growth	20	ratio	14	void	21	dose	18
muscle	16	antagonist	26	stain	32	grade	19	chronic	12	urodynamic	20	antagonist	17
ratio	16	inhibitor	24	epithelium	31	genotype	18	sexual	12	turp	19	ipss	17
receptor	16	carcinoma	21	mrna	31	carcinoma	17	tissue	12	ablation	17	pressure	17
antagonist	15	stromal	21	tumor	29	receptor	17	alpha-blocker	11	risk	17	risk	17
inhibitor	15	alpha	19	androgen	25	allele	16	dysfunction	11	microwave	16	sexual	15
plasma	15	apoptosis	18	line	25	expression	16	psa	11	thermotherapy	16	outlet	14
score	15	epithelium	18	alpha	24	biopsy	14	tumor	11	ultrasound	16	alpha-blocker	13

Group 10	f	Group 15	f	Group 31	f	Group 12	f	Group 40	f	Group 37	f	Group 2	f
cancer	92	symptom	67	cancer	62	cancer	64	symptom	58	symptom	50	symptom	38
expression	76	risk	30	psa	54	psa	63	score	54	score	35	placebo	29
cell	72	score	30	serum	49	serum	46	volume	33	luts	21	antagonist	22
tissue	66	luts	28	biopsy	23	free	44	thermotherapy	27	bladder	20	finasteride	21
tumor	57	dysfunction	20	concentration	22	ratio	31	microwave	24	void	17	dose	20
gene	55	sexual	17	volume	20	biopsy	26	urine	24	volume	17	tamsulosin	18
protein	44	prevalence	16	ratio	19	concentration	17	bladder	21	ipss	15	score	17
normal	42	volume	15	free	18	tumor	17	ipss	20	obstruction	14	pressure	16
line	34	bladder	14	tumor	17	zone	15	qol	19	incontinence	12	doxazosin	15
carcinoma	32	erectile	14	normal	16	normal	13	baseline	18	item	10	inhibitor	15
mrna	29	activity	13	prostatectomy	16	pca	13	post-operative	18	urine	10	risk	15
growth	27	elder	13	tissue	14	roc	13	obstruction	17	dysfunction	9	baseline	14
gleason	22	finasteride	13	cell	12	mug	12	tumt	17	prevalence	9	volume	12
grade	22	inhibitor	13	ultrasound	12	carcinoma	10	urodynamic	17	alpha	8	alfuzosin	11
epithelial	20	cancer	12	carcinoma	11	risk	10	prostatectomy	15	baseline	8	alpha	10
microarray	17	serum	12	risk	11	tissue	10	turp	15	chronic	8	alpha(1)-adrenoceptor	10
prostatectomy	16	antagonist	11	pca	10	tpsa	10	laser	11	outlet	8	formulation	9
receptor	15	obstruction	11	symptom	10	fpsa	9	luts	11	pain	8	terazosin	9
androgen	14	tamsulosin	11	zone	10	cutoff	8	void	11	sexual	8	sexual	8
rna	14	testosterone	11	grade	9	density	8	ipss	10	storage	8	5alpha-reductase	7

Table 6. Primary descriptors most frequently used in the RefViztm groups with more than 50 articles (f = Frequency of terms appearing).

From the analysis of the co-occurrence of the descriptors represented in Table 6, and the revision of articles contained in each group, the main research lines which control scientific production could be pointed out in the graphic, which coincide with those obtained through ACA. For this reason they were identified with the same nomenclature.

The scientific production of the 14 most prolific groups gathers a total of 1332 articles, which constitute 67.7 % of all. From the 1,332 articles, 622 (46.7 %) are grouped in the sector devoted to the study of PC and its relationship with BPH (groups G-9, G-10, G-11, G-12, G-16 and G-31). This figure offers a new vision of sector 2, represented in Figures 1, 2 and 3, according to the data presented in Table 6, making it the emergent research profile in the study of BPH.

The conceptual structure resulting from the conjunction of techniques employed is shown in Figure 5, which reflects the current research lines about the disease.

Figure 5. BPH Conceptual structure obtained by fusion of bibliometric techniques and semantic analysis

- Prostatic Disease
  - Prostatitis
  - Benign Prostatic Hyperplasia (BPH)
    - Epidemiology
      - Etiology
      - Progression
      - Risk factors
        - Biochemical factors
          - Dihydrotestosterone levels
          - Other biochemical factors
        - Racial factors
        - Social factors
        - Genetic factors
        - Other factors
      - Prevalence
        - Age-specific prevalence
      - Incidence
      - Prevention
      - Quality of life
      - Cost
    - Diagnosis
      - Primary care
        - Mandatory investigation
          - Full medical history
          - Urinary symptom review
          - Symptom score
            - American Urological Association Symptom Index

-
  -
    -
      -
        - International Prostate Symptom Score (IPSS)
        - Boyarsky Index
        - Madsen-Iversen Index
  - Basis
    - Benign Prostatic Enlargement (BPE)
    - Urinary flow
    - Lower Urinary Tract Symptoms (LUTS)
      - Voiding symptoms *or* Obstructives symptoms
        - Hesitancy in the irritation of micturitia
        - Weak force of stream
        - Stopping and re-starting of the stream
        - Interruption of the stream
        - Terminal dribbling
        - Post-micturitian dribbling
      - Filling symptoms *or* Irritative symptoms
        - Nocturia
          - Frequency
          - Urgency
          - Dysuria
          - Sensation of incomplete voiding
        - Urge incontinence
          - Pain
          - Hematuria
          - Acute retention
        - Other symptoms
  - Tests
    - Prostatic Specific Antigen (PSA)
    - Uroflowmetry
    - Residual urine
    - Pressure flow studies
    - Serum creatinine
    - Urinary analysis
    - Digital Rectal Examination (DRE)
- Therapy
  - Medical treatment
    - Pharmacological therapy
      - α-Blockers
        - Terazosin
          - Pharmacology*
          - Benefits*
          - Harms*
          - Prescription*
          - Dossage*
          - Adverse effects*
          - Clinical trials*
        - Tamsulosin
        - Alfuzosin
        - Doxazosin
        - Prazosin
        - Indoramin
      - 5-α reductase inhibitors
        - Finasteride
    - Alternative therapy
      - Phytotherapy
        - Saw Palmetto plant extracts
          - Preparation**
          - Benefits**
          - Harms**
          - Adverse effects**
          - Clinical trials**

- β Sitosterol plant extracts
- Rye grass pollen extracts
- Non medical treatment
  - Surgical therapy
    - Transurethral Resection of the Prostate (TURP)
      - Indications***
        - Acute urinary retention
        - Chronic urinary retention
        - Elevated creatinine/urea
        - Suspected malignancy
        - Hematuria
        - Bladder stones
      - Complications***
        - Bladder Outflow Obstruction (BOO)
        - Recurrent urinary tract infection
        - Pyelonephritis
        - Dilatation
        - Hydronephrosis
        - Acute urinary retention post TURP
        - High residual urine volume
        - Renal azotaemia
        - Bladder stones post TURP
        - Persistent incontinence
        - High symptom score
        - Bleeding requiring intervention
        - Erectyle dysfunction
        - Retrograde ejaculation
        - Incontinence
      - Morbidity***
        - Bleeding
        - Failure to void
        - Urinary Tract infection (URI)
        - TURP syndrome
        - Myocardial arrhythmia
        - Indwelling catheters
      - Mortality***
      - Trials***
  - Alternative surgical therapy
    - Transurethral Microwave Thermotherapy (TUMT)
    - Transurethral Needle Ablation (TUNA)
    - Transurethral Incision of the Prostate (TUIP)
    - Laser therapy
    - Laser ablation
    - Electronic vaporization
- Cancer of the Prostate (PC)
  - *Related to BPH*
  - Epidemiology
    - Etiology
    - Risk factors
    - Prevalence
    - Incidence
    - Prevention
    - Quality of life
  - Diagnosis
    - Prostate Specific Antigen (PSA)
      - Detection
      - PSA density (PSAD)
      - PSA transition zone density (PSATZ)
    - Prostate-specific Membrane Antigen (PSM)
    - Gene expression profiles

Gene amplification
Protein expression
Protein profiling technologies
Osteoprotegerin protein expression
Tumor-asociated protein
Molecular markers
Biological markers

** Categories standard to pharmacological therapies*
*** Categories standard to alternative therapies*
**** Categories standard to surgical therapies*

Ongoing work and development based on BPH conceptual structure

Conceptual structures have multiples uses. The most well-known is knowledge organization use oriented to information retrieval; however, there are others like knowledge domain mapping. In science and technology the information access and retrieval are as important as the opportunity to know the research topics evolution and its trends.

Recognizing the previous idea, it becomes a requirement the implementation of information services oriented to show the domain dynamics. In fact, this work is the starting point to implement a new online information service on BPH. This service will be addressed to researchers from the Cuban National Scientific Research Center (CNIC).

The main information service characteristics demanded the use of a conceptual structure to overcome the limitations of thesaurus and traditional classification schemes. This is possible because the conceptual structures are more flexible and allows the systematic update through the continual introduction of new terms. These new terms reflect the emergent research topics in the scientific literature. In addition, as part of the information service implementation, the design of graphic interfaces for the visualization of different bibliometric indicators were considered. These indicators will be calculated upon the records included in the CNIC's BPH Repository.

Other advantages derived from BPH conceptual structure are the opportunity to develop an ontology and the opportunity to nurture the collaborative use of knowledge between Cuban researchers and international scientific community specialized in this disease.

Final considerations

The analysis of the community which produces knowledge in concrete domains from the use of the bibliometric method, complemented with the semantic analysis, allows the construction of representative conceptual structures from the domain.

The current work presented a methodology which is a hybrid of terminological extraction which intends to fuse linguistic and non-linguistic methods, using bibliometrics as an instrument for the identification of a macrostructure or global thematic pattern from the analyzed domain, in this case, the most current research on BPH, and the semantic analysis for the selection of the categories in any depth level, representative from the domain. For this reason we can asseverate that the obtained conceptual structure reflects the implicit cognitive structure of a concrete social group (scientific community on BPH) on the selected scientific production.

These kinds of methodologies, which combine bibliometric techniques and text mining, are alternatives which can contribute to the strengthening of qualitative dimension in the processes of documentary analysis, applicable either to the evaluation systems from research in a specific scientific domain, or to the processes of construction for systems of knowledge representation and organization.

However, although it is important the employed method, the essential aim of the cognitive structures design for information systems, and the main criterion to validate its qualitative dimension, is the system efficiency for the information retrieval. The construction of conceptual structures, such as the one obtained from this study, will require the always necessary validation by an expert committee with the aim of a later implementation as controlled language (structured) in any kind of online information service developed for the use of specialists and researchers devoted to the deep and systematic study of the disease.

Acknowledgements

To Professor Jesús A. Núñez Romay, from the National Center for Scientific Research, Cuba; Dr. Víctor Herrero Solana, from the University of Granada, Spain; and Danielle Charles, for their help in the revision and correction of this paper.

References

Barry, M.J., & Roehrborn, C.G. (2001). Extracts from *Clinical Evidence*: benign prostatic hyperplasia. *British Medical Journal,* 323, 1042-6.

Braam, R.R., Moed, H. & Van Raan, A.F.J. (1991). Mapping of science by combined co-citation and word analysis. II. Dynamical aspects. *Journal of the American Society for Information Science*, *42*(4), 252-66.

Broughton, V., Hansson, J., Hjørland, B., & López-Huerta, M.J. (2005) Knowledge organization. In: Leif Kajberg & Leif Lørring (Eds.). *European Curriculum Reflections on Library and Information Science Education.* Copenhagen: Royal School of Library and Information Science, 133-148.

Callon, M., Law, J. & Rip, A. (1986). *Mapping the Dynamics of Science and Technonolgy: Sociology of Science in the real World*. London: The Macmillan Press Ltd.

Chokkalingam, A.P., Nyren, O., Johasson, J.E., et al. (2003). Prostate carcinoma risk subsequent to diagnosis of benign prostatic hyperplasia: a population-based cohort study in Sweden. *Cancer, 98*, 1727-34.

Clifford, G.M., & Farmer, R.D.T. (2000). Medical therapy for benign prostatic hyperplasia: a review of the literature. *European Urology, 38,* 2-19.

Courtail, J. P. (1998). Comments on Leydesdorff's article. *Journal of the American Society for Information Science*, *49*(1), 98.

Glenisson, P., Glänzel, W., Janssens, F.A.L., & Moor, B de. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management*, *41*(6): 1548-1572.

Gormley, G.J., Stoner, E., & Bruskewitz, R.C. (1992). The effects of finasteride in men with benign prostatic hyperplasia. *New England Journal of Medicine, 327,* 1185-91.

Guess, H.A. (2001). Benign prostatic hyperplasia and prostate cancer. *Epidemiology Review, 23,* 152-8.

Hamilton, W., & Sharp, D. (2004). Symptomatic diagnosis of prostate cancer in primary care: a structured review. *British Journal of General Practice*, *54*, 617-21.

Hjørland B. (2000). Documents, Memory institutions and Information Science. *Journal of Documentation, 56*(1), 27-41.

Hjørland, B., & Albretchen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society of Information Sciences, 46*(6), 446-60.

King, J. (1987). A review of bibliometric and other science indicators and their role in research e valuation. *Journal of Information Science*, *13*(5), 261-76.

Lafuente López, R. (2001). Sobre el análisis y representación de documentos. *Investigación Bibliotecológica, 15*(30), 163-13.

Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, *48*(5), 418-27.

McConnell, J.D., Bruskewitz, R., & Walsh P. (1998). The effect of finasteride on the risk of acute urinary retention and the need for surgical treatment among men with benign prostatic hyperplasia. *New England Journal of Medicine, 338,* 557-63.

Nazaretian, A.P. (1976). Dos formas de enfocar la construcción de la Teoría General de la Comunicación. *Actualidades de la Información Científico-Técnica, 61*(11), 1-21.

Pao, M.L. (1993). Term and citation retrieval: a field study. *Information Processing & Management, 29,* 95-112.

Pao, M.L., & Worthen, D.B. (1989). Retrieval effectiveness by semantic and pragmatic relevance. *Journal of the American Society of Information Sciences, 40,* 226-35.

Rees-Potter, L.K. (1989). Dynamic thesaural systems: a bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management, 25,* 677-91.

Roehrborn, C.G., Bartsch, G., Kirby, R., *et al*. (2001). Guidelines for the diagnosis and treatment of benign prostatic hyperplasia: a comparative, international overview. *Urology, 58,* 642-50.

Science and Technology Dictionary. (2003). New York: McGrawHill Professional.

Schneider, J.W. (2004). *Verification of Bibliometric Methods' Applicability for Thesaurus Construction.* PhD dissertation. Aalborg: Royal School of Library and Information Science. Available: http://biblis.db.dk/uhtbin/hyperion.exe/db.jessch04

Schneider, J.W., & Borlun, P. (2004). Introduction to bibliometrics for construction and maintenance of thesauri: methodological considerations. Verification of bibliometric methods´ applicability for thesaurus construction. *Journal of Documentation, 60*(5), 524-49.

Simpson, R.J. (1997). Benign prostatic hyperplasia. *British Journal of General Practice, 47,* 235-40.

Small, H. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*, 327-40.

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics, 38*, 275–293.

Stoevelaar, H.J., & McDonnell, J. (2001). Changing therapeutic regimens in benign prostatic hyperplasia: clinical and economical consideration. *Pharmacoeconomics, 19,* 131-53.

White, H.D., & McCain, K.W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society of Information Sciences, 49,* 327-55.