Evaluation of controlled vocabularies by interindexers consistency

Concha Soler-Monreal
Universitat de València, m.concepcion.soler@uv.es

Isidoro Gil-Leiva
Universidad de Murcia, isgil@um.es

Abstract:

Introduction. Several controlled vocabularies are used for indexing three journal articles to check if with a list of descriptors are achieved better or equals of consistency rates that with a standard thesaurus and augmented thesaurus.
Method. A set of terminology of Library and Information Science was used to build a list of descriptors with equivalence relations (USE and UF), a standard thesaurus and a augmented thesaurus (all the descriptors have scope notes). Subsequently, three articles were indexed by selected indexers who had varying degrees of experience – on the one hand Library and Information Science students and on the other, professionals from various documentation centres. Hooper's measure to find the consistency between pairs of novice indexers and experts has been applied.
Analysis. Data were tabulated and analysed systematically according pairs of novice indexers and experts has been applied.
Results. The tool with the best results is the list of descriptors (39.5% consistency), followed by the augmented thesaurus (29.8%) and, with an almost identical value, the standard thesaurus (27.5%).
Conclusion. It is concluded that the list of descriptors in both groups returns better indexing consistency but we need more research.

Keywords: Controlled vocabularies evaluation; thesaurus evaluation; list of descriptors evaluation; indexing; indexing consistency; interindexer consistency

1. INTRODUCTION

Vocabulary control has been revealed as an essential procedure in the organization and retrieval of information. The most significant contributions in this field of work are many and varied although the main ones taken here are those from Gil Leiva, 2008 (p. 118-154). The first contribution was the work done by Charles Ammi Cutter in his famous *Rules for a printed dictionary catalogue* published in 1876. It is here that the first rules appear that have full effect today, such as the principle of economy, the definition and use of the term *headings* for both matter and for place and form; resending for synonyms and antonyms; the problem of homonymy; the structure of the subject headings (simple and complex); word inversion; syntax (See, See also, etc.) and punctuation marks (commas, brackets, etc.).

The second contribution was the building of lists of subject headings. Shortly after the contributions of Cutter, the *American Library Association* (ALA) published in 1895 the *List of Subject Headings for Use in Dictionary Catalogs* as an indexing tool for small and medium sized libraries with non specialized stocks. The first *Subject Headings*

Used in the Dictionary Catalogues of the Library of Congress appeared in 1909 and took as its main references the contributions mentioned above. Although it came into being for internal use by the Cataloguers in the Library of Congress, it would soon become a reference tool used in indexing in large public and academic libraries and it was translated or totally or partially adapted to other countries and languages, for example, Brazil (1948), Canada (1967), Greece (1978), South Africa (1992) or Egypt (1995) among others.

The third contribution comes from Calvin Mooers who at the beginning of the 1950s introduced the word *descriptor* to communicate ideas, so distancing himself from particular terminological uses employed in documents and thus specifying the subject of the information in an information retrieval context. A follow on of this was the construction of the first lists of descriptors and the first thesauri, like the Dupont Thesaurus (Engineering Information Centre Du Dupont, 1959), the Thesaurus of Astia Descriptors (United States Departament of Defense, 1960), or the Chemical Engineering Thesaurus (American Institute of Chemical Engineers, 1961), among others.

The fourth contribution is the provision of national and international norms. Work in this sphere got underway early in France since in 1957 the AFNOR Z 44-070 *Catalogue alphabétique de matières* was presented, which was devoted to establishing and providing rules for the choice and presentation of subject headings. The first norms for thesauri were the French AFNOR Z 47-100-1973 (Norme experimental. Regles d'établissement des thèsaurus monolingues), the ISO 2788-1974 (Documentation. Guides for the establishment and development of monolingual thesauri) and the ANSI Z39.19-1974 (American National Standard guidelines for thesaurus structure, construction and use). Since then, other countries and the ISO itself have been working on and extending the norms until the unification of the ISO 2788-1986 and ISO 5964-1985 in the new ISO/DIS 25964-1:2010 (Information and documentation -- Thesauri and interoperability with other vocabularies (Part 1: Thesauri for information retrieval; Part 2: Interoperability with other vocabularies).

The evaluation of controlled vocabularies is an issue of concern for professionals and researchers in the area. The evaluation can be performed with the aim of the analysis being the controlled vocabularies themselves so as to study their structure, the thematic fields or facets, scope notes, semantic relations, degree of specificity, etc., (intrinsic evaluation) or by studying its impact on the information systems which use them both in indexing and retrieval (extrinsic evaluation).

The first evalaution of import was carried out by Cyril W. Cleverdon in the Cranfield Projects (Proposals for an investigation into the efficiency of various retrieval systems, 1956; ASLIB Cranfield Research Project: report on the first stage of an investigation into the comparative efficiency of indexing systems, 1960, etc.). Cleverdon compared the efficiency of the Universal Decimal Classification, an alphabetical index of subjects, a faceted classification scheme and the indexing through uniterms of eighteen thousand documents analysed by three indexers. There have been many and varied subsequent studies to evaluate controlled vocabularies, both subject headings and thesauri. We have for example the works by Henzler (1978), Fidel (1991 and 1992), Betts and Marrable (1991), Ribeiro (1996), Gil Urdiciaín (1998) and Gross and Taylor (2005), who studied

the advantages and drawbacks of indexing and retrieving documents in natural language and in controlled language.

Another way of evaluating controlled vocabularies, mainly thesauri, is to compare them with each other. Kishida, et al. (1988) compared the MeSH, the ERIC thesaurus, the INSPEC and the Root thesaurus, among others, taking as their reference the construction principles, their structure and the information they contributed. In contrast, Weinberg and Cunningham (1985) studied the semantic proximity between the MeHS and the Medline, while Pozhariskii (1982) proposed quantifying the capacity or semantic strength of a thesaurus in terms of flexibility, economy and universality. Elsewhere, Larsen (1988) analyses the capacities for use of a thesaurus for indexing a certain collection of documents. Soler Monreal (2009) evaluates three controlled vocabularies (a list of descriptors, a standard thesaurus and a augmented thesaurus in which all the descriptors have scope notes) in order to find out if with a list of descriptors are obtained consistency scores higher than a standard thesaurus and augmented thesaurus.

Indexing consistency can be studied as a reference to a single indexer or to several. When a professional indexes the same document at different moments in time we speak of intraconsistency or intraindexer consistency. And when several people indexing a document to compare the results or the result of indexing a document by two indexers is compared we speak of interconsistency or interindexer consistency.

Since the 1960s, numerous and diverse investigations are carried out on indexing consistency. The main conclusion which can be drawn from the tests is that inconsistency is an inherent feature of indexing more than a sporadic anomaly. Although the tests carried out are very diverse in their methodology we can speak of that achieved indexing consistency ranges from 10% to 60% approximately. The vast majority of the tests carried out from 1960 until the present time cannot be homogenized by the used methodological diversity. We only point out here some of the variables that hinder their homogenization and only a sample of the test carried out:

> Measures: To find the consistency scores between groups of indexers Slamecka & Jacoby (1965) and Iivonen (1990) proposed their respective measures. In contrast, for pairs of indexers others measures were outlined as in Hooper (1965), Lancaster (1968), Rolling (1981) or Saarti (2002), although the most commonly used in most tests is that of Hooper, c / (a + b - c), where c are the common terms between the two indexers, a are the terms proposed by the indexer a and b are the terms proposed by the indexer two.

> Novice versus expert indexers: In some occasions was employed expert indexers as in Lancaster (1968), Leonard (1975); other times with novice indexers Hudon (1998a and 1998b), Gil Leiva (2002); or with experts and novices (Bertrand and Cellier , 1995; Saarti, 2002; Soler Monreal, 2009).

> Number of indexers: Lancaster (1968) worked with 3 indexers; Bertrand and Cellier (1995) 25 indexers; Hudon (1998a and 1998b) 25 indexers; Gil Leiva (2002) 27 indexers; Saarti (2002) 30 indexers; and Soler Monreal (2009) 63 indexers.

Material used: Sometimes has worked with journal articles (Lancaster, 1968; Leonard 1977; Funk and Reid, 1983; Middleton, 1984; Siever and Andrews, 1991; Iivonen and Kivimäki, 1998; Leininger, 2000; Gil Leiva, 2002), sometimes with books (Tonta, 1991; Bertrand and Cellier, 1995; Gil Leiva, 2001; Saarti, 2002, Neshat and Horri, 2006; Gil Leiva, Polsinelli and Spotti, 2008; Chen, 2008), at other times with visual material (Markey, 1984 and Gil Leiva, 2002). And finally, have also been used summaries of journal articles (Hudon, 1998; Soler Monreal, 2009).

Number of documents indexed: Lancaster (1968) uses 16 articles; Tarr and Borko (1974) 15 items; Leonard (1975) 100 articles; Funk and Reid (1983) 760 articles; Markey (1984) 100 documents; Iivonen (1990) 10 documents; Siever and Andrews (1991) 71 articles; Iivonen and Kivimäki (1998) 49 documents; Leninger (2000) 60 documents; Tonta (1991) 82 books; Bertrand and Cellier (1995) 8 books; Gil Leiva (2001) 11 books; Saarti (2002) 5 books; Gil Leiva, Polsinelli and Spotti (2008) 10 books; Chen (2008) 3307 monographs; Hudon (1998a and 1998b) 12 abstracts; Monreal Soler (2009) 3 abstracts.

Hawthorne effect: Individuals when they know that they are being studied behave differently that when they are not observed. In some studies indexers knew their product would be evaluated (Lancaster, 1968, Leonard, 1975, Bertrand and Cellier 1995, Hudon (1998a and 1998b); Gil Leiva, 2002; Saarti, 2002; Soler Monreal, 2009); and in other studies as the result of indexing the same documents is compared but in different information systems could not be this possibility as, for example, two bibliographic databases: Middleton (1984) compare the indexing of ERIS / APAIS and AEI / APAIS; Iivonen and Kivimäki (1998) databases KINF and LISA; or compare the indexing library catalogs as Tonta (1991) the Library of Congress and the British Library; Gil LeIva (2001) in 31 catalogs of public libraries; Neshat and Horri (2006) National Library of Iran and 12 academic libraries; Gil Leiva, Polsinelli and Spotti (2008) 30 university library catalogs; Chen (2008), the National Library of China and the China Academy Library & Information System. Finally, duplicate records in information systems (Funk and Reid (1983) use 760 articles indexed in Medline twice; Siever and Andrews (1991) work with 71 duplicates of the database Information Science Abstracts (ISA); Leininger (2000) compares 60 duplicates of the PsycINFO database.

Concepts versus terms: In most of the studies mentioned above are compared with each indexing terms or descriptors derived from a controlled vocabulary, while sometimes to find consistency comparisons were made with the concepts taken directly from the documents such as Iivonen (1990) or Gil Leiva (2002) that makes both the concepts used as descriptors from Eurovoc Thesaurus.

2. MATERIALS AND METHODS

For this study we built three controlled vocabularies on Information Science : a list of descriptors with control for synonymy, a standard thesaurus and a thesaurus in which all the descriptors have scope notes (augmented thesaurus).

At the time of initiating this research did not exist in Spanish a thesaurus published on this subject, hence began to refine a list of descriptors consisting of 2756 terms which were in use in the design and maintenance of an automatic indexing system (Gil Leiva, 1997 and 2008). Finally, the list was a total of 2455 terms, of which 1436 are descriptors and 1019 non descriptors. From this list of descriptors was constructed standard thesaurus. This thesaurus has an alphabetic display, another hierarchical one and other types KWOC permuted index. In Annex A shows the first terms of the three tools built.

The thesauri were built with thesaurus management software "MultiTes" and following the spanish norm UNE 50-106-90 (equivalent to ISO 2788-1986).

Centralized acquisition
TC: J02
UP: Centralized Purchases
TG1: Acquisition of documents
 TG2: Development of collections
  TG3: Documental process

Topographic catalogues
TC: F03
TG1: Catalogues (Information sources)
 TG2: Secondary sources
  TG3: Information sources

Figure 1: Descriptors of the standard thesaurus.

Finally, specialized dictionaries are used to add the scope notes to all the descriptors to built augmented thesaurus.

Centralized acquisition
TC: J02
NA: Purchase of documental stocks by an institution which also distributes them to other centres so as to economize on resources.
UP: Centralized purchases
TG1: Acquisition of documents
 TG2: Development of collections
  TG3: Documental process

Topographic catalogues
TC: F03
NA: Catalogues in which the bases follow the order of the place occupied by the documents in the collection or on the shelves, coinciding with the order of the topographic library number.
TG1: Catalogues (Information sources)
 TG2: Secondary sources
  TG3: Information sources

Figure 2: Descriptors from augmented thesaurus.

After building the three controlled vocabularies, an intrinsic (qualitative and quantitative) evaluation is carried out to check that they comply with the recommendations for the compilation of thesauri. The compilation was carried out following the parameters proposed by Lancaster (2002), Gil Urdiciaín (2004) and Gil

Leiva (2008). It was confirmed that the thesauri meet the traditional requisites for compilation of thesauri.

Later, we decided that the material indexed were three abstracts of journal articles since it is concise, well structured and understandable information sources (Annex B). We then worked on the selection of the indexers who were going to use the three indexing languages to indexing three abstracts of Information Science articles. Finally, we decided that the indexers should have different levels of experience.

Group 1: Second year Information Science students.
Group 2: Fourth year Information Science students.
Group 3: Fifth year Information Science students.
Group 4: Experienced professionals in document indexing.

The three groups of students already had some theoretical and practical knowledge of indexing and use of controlled vocabularies. Each group comprised 18 people and was divided into three subgroups of six indexers for each of the three tools. The exception was Group 4, which was made up of 9 professionals for whom indexing is a habitual task. The professionals work in documentation centres in Public Administration (3), communication (3) and technological institutes (3). These were also subdivided into nuclei of three indexers per tool. None of the indexers were familiar with the indexing languages constructed for the tests, although both the novice and the expert indexers had used indexing languages from other spheres. Finally, it should be mentioned that it was difficult to find more professionals who were available to participate in these types of tests.

The results of the indexing of the three abstracts were compared pair wise, so novice indexers were compared fifteen times for each of the three articles and for each of the three tools being compared – giving a total of 137 comparisons. As regards the expert indexers, three comparisons were obtained for each for each of the three articles and the three tools under comparison – giving a total of 27 comparisons.

We used a relaxed, and not an exact, system of coincidence to calculate consistency between indexers, as was done in Gil Leiva (2001) and Gil Leiva, Polsinelli Rubi and Spotti Lópes Fujita (2008). A coincidence of 1 (100%), 0.5 (50%) or 0 (0%) was considered. For example, if one indexer consigns *librarians* and another *reference librarians*, a consistency of 0.5 is recorded. As a general norm, it was considered that a score of 0.5 should be awarded to those non coincident terms that were, however, specific of another one, while 1 was given to very similar concepts.

Table 1: Table of relaxed equivalences between descriptors

Indexer 1	Indexer 2	Agreement
Biomedical journals	Scientific journals	0,5
Librarianship techniques	Librarianship	1
Databases	Bibliographical databases	0,5
Librarians	Librarians of reference	0,5
Scientific journals	Scientific publications	1

Since their beginnings, tests on indexing consistency have used various formulas, among which the most important are those used by Hooper (1965) and Rolling (1981). Gil Leiva (1997 and 2001), Gil Leiva, Polsinelli Rubi and Spotti Lópes Fujita (2008)

and Soler Monreal (2009) we have used extensively Hooper's measure of indexing consistency adapted as follows:

$$C_i = \frac{T_{co}}{(A+B)-T_{co}}$$

where C_i is the consistency between two indexings, T_{co} is the number of terms in common between the two indexings, A is the number of terms used by Indexer A, B is the number of terms used by Indexer B, and T_{co} is the number of terms they use in common.

Example:

Indexer 1	Indexer 2	Agreement
1. Librarianship	1. Librarianship	= 1
2. Cite frequency	2. Medical documentation	
3. Biomedical journals	3. Impact factor	
	4. Scientific journals	= 0,5

$$C_i = \frac{1,5}{(3+4)-1,5} = 0,2727 \text{ x } 100 = 27,27 \text{ \% consistency}$$

3. RESULTS AND DISCUSSION

The results from the comparisons carried out to ascertain the consistency for each of the indexing languages constructed are organized by groups for the sake of presentation and analysis as can be seen in the Annex C.

The results of the tests with novice indexers are summarized in the means table below:

Table 2: Mean of the results of the novice indexers with the three tools as %

	List of descriptors	Augmented thesaurus	Standard thesaurus
Second year students	29,5	25,9	29,6
Fourth year students	39,8	34,1	23,7
Fifth year students	33,3	35,8	26,3
Mean	**34,2%**	**31,9%**	**26,5%**

However, data for consistency among indexers experts are:

Table 3: Mean of the results for the expert indexers with the three tools

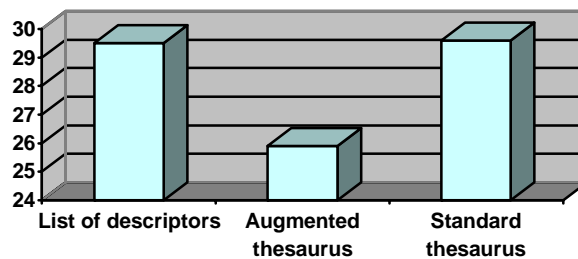|	List of descriptors	Augmented thesaurus	Standard thesaurus
Expert indexers	55,7 %	23,7 %	31,3 %

We have also obtained the mean for all the consistency obtained, both for expert and novice indexers, as can be seen in Table 4.

Table 4: Mean of the results for all the indexers with the three tools as %

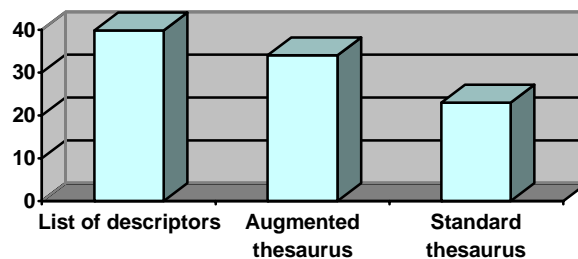|	List of descriptors	Augmented thesaurus	Standard thesaurus
Second year students	29,5	25,9	29,6
Fourth year students	39,8	34,1	23,7
Fifth year students	33,3	35,8	26,3
Expert indexers	55,7	23,7	31,3
Mean	**39,5%**	**29,8%**	**27,7%**

For group 1 (Second year students) the standard thesaurus and the list of descriptors return the same levels (29.6% and 29.5%), followed by the augmented thesaurus (25,9%).

Figure 3 GRUP 1

In group 2 (fourth year students) of indexers the list of descriptors returned the best consistency results, 39.8%, versus 34.1% for the augmented thesaurus and 23,7% for the standard thesaurus, as is reflected in the figure below:
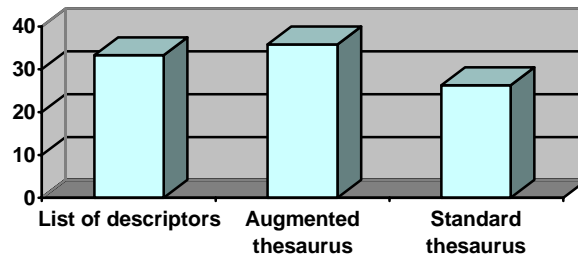
Figure 4 GRUP 2

For group 3 (fifth year students) the augmented thesaurus returns the best results (35.8%), followed closely by the list of descriptors (33.3%), as can be observed in Figure 5.
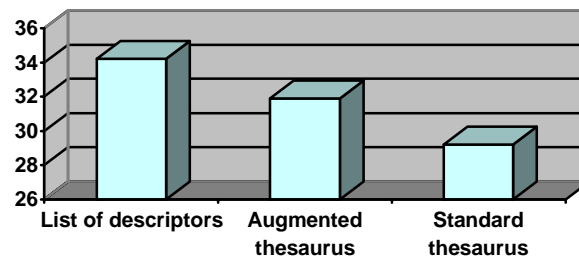
Figure 5 GRUP 3

From the results it can be stated that the list of descriptors provides the highest indexing consistency among all novice indexers, with 34.2% coincidence, versus 31.9% for the augmented thesaurus and 29. 2 % for the standard thesaurus, as Figure 6 shows.

Figure 6 All novice indexers

Expert indexers (Group 4) also obtain their maximum consistency index with the list of descriptors, 55.7%. In second place are the results obtained with the standard thesaurus, 31.3%. The lowest consistency 23,7% were returned with the augmented thesaurus. This may be due to a lack of previous knowledge of this tool or to the fact that the scope notes annotate the meaning of the terms, leading the indexer to choose certain descriptors on the basis of the definition given and not according to previously conceived ideas.

Figure 7 Expert indexers
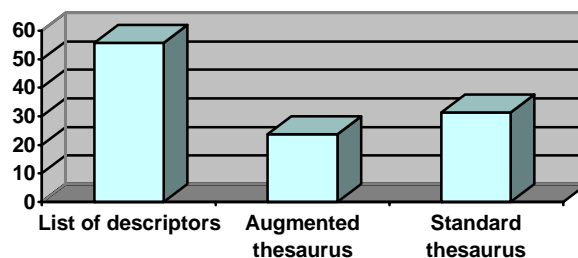
We have also found the mean for all the consistency obtained for both expert and novice indexers. It is clearly observed that the tool with the best results is the list of descriptors (39.5% consistency), followed by the augmented thesaurus (29.8%) and, with an similar value, the standard thesaurus (27.7%), as can be seen in Figure 8.

Figure 8 All indexers

With the exception of two cases, the highest consistency occur for abstract 2.

The results obtained in this study fall within the margins of consistency obtained in other previous studies ranging from 10% to 60%. We also said in the introduction to this article that we can compare our results with data is achieved by Michèle Hudon (1998a and 1998b). Hudon also used three versions of a thesaurus (a) thesaurus standard, b) standard thesaurus with definitions for all their descriptors and c) stripped thesaurus whit definitions, but hierarchical and associative relationships between terms) to see if the definitions including all descriptors of a thesaurus can raise levels of consistency among novice indexers.

Hudon's results show that a) for the selection of all descriptors (main and minors) indexers who worked with the augmented thesaurus didn't obtain better consistency that those that worked with the standard thesaurus; b) indexers that used the stripped thesaurus were same or more consistent to each other that those that used the standard thesaurus. In contrast, for the selection of main descriptor indexers who worked with the augmented thesaurus consistency scores were better than the standard thesaurus in 7 of the 12 documents, and that the stripped thesaurus indexers were more consistent than the standard thesaurus indexers in 8 of the 12 documents. Hudon concludes that the availability of definitions in a thesaurus does not increase the consistency in indexing of novice indexers in the selection of main and minor descriptors, and the availability of definitions in a thesaurus may lead to novice indexers achieve acceptable levels of consistency in the selection of the main descriptor when using a stripped thesaurus.

In our research has become clear that the list of descriptors has achieved the highest levels of consistency in indexing for novice and expert indexers; though in the results obtained for Hudon with stripped thesaurus with respect to others there are some indications to suggest that the lists of descriptors can achieve similar results to standard thesauri and augmented thesauri.

Anyway, to corroborate these results it would be necessary to carry out more research and thus mark out an important limitation of this research. In the two comparative studies we don't know how much percent of inconsistency is due to the complexity and subjectivity of indexing (reading and analysis of the document and selection of the appropriate keywords) as to the later selection of the descriptors of the controlled vocabulary (conversion of the selected keywords to descriptors of the controlled vocabulary); or how much percent of inconsistency is due to indexing languages used.

4. CONCLUSIONS

The inconsistency is an inherent feature of the indexing as we have seen from the data obtained in researches conducted since the 1960s to the present day. Precisely because the disparity of variables used in the investigations perhaps may be appropriate to carry out a systematic review and or meta-analysis of the relevant literature on indexing consistency to provide more light on this issue. Similarly, as already been suggested, more research is needed on the properties of the lists of descriptors compared to standard thesauri or augmented thesauri with application notes or notes of definition, because in an information system small is always easier to build a list of descriptors that a thesaurus. To further examine the properties of the lists of descriptors compared to thesauri could continue working on this line but including the technique of verbal protocols or 'thinking aloud' during the process of indexing the documents with thesauri. This technique will allow to gather valuable information on the use that indexers do of the associative and hierarchical relations.

5. REFERENCES

BERTRAND, A. y CELLIER, J.M. Psychological approach to indexing: effects of the operartor's expertise upon indexing behaviour. *Journal of Information Science*, 1995, vol. 21, nº 6, p. 459-472.

BETTS, R. y MARRABLE, D. Free text vs. controlled vocabulary – retrieval precision and recall over large databases. *Online Information 91:Proceedings of the Fifteenth International Online Information Meeting, London, 10-12 December 1991*, p. 153-165.

CLEVERDON, C.W. ASLIB Cranfield Research Project: report on the first stage o fan investigation into the comparative efficiency of indexing systems, sept. 1960. Texto disponible en http://hdl.handle.net/1826/1122 [Consultado: 02/12/2009].

CLEVERDON, C.W. Proposals for an investigation into the efficiency of various retrieval systems, 1956. Disponible en: http://hdl.handle.net/1826/1367 [Consultado: 02/12/2009].

CHEN, X. Indexing consistency between online catalogues. Berlín: Universidad Humboldt, 2008, PhD.

FIDEL, R. Searchers' selection of search keys: 2. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science*, 1991, vol. 42, nº 7, p. 501-514.

FIDEL, R. Who needs controlled vocabulary? *Special Libraries*, 1992, vol. 83, nº 1, p. 1-9.

FUNK, M.E., REID. C.A. Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association*, 1983, vol. 71, p. 176-183.

GIL LEIVA, I. La automatización de la indización. Propuesta teórico-metodologícia: Aplicación al área de Biblioteconomía y Documentación. Murcia: Universidad de Murcia, 1997 (PhD Thesis).

GIL LEIVA, I. Consistencia en la asignación de materias en las Bibliotecas Públicas del Estado. *Boletín de la Asociación Andaluza de Bibliotecarios*, 2001, nº 63, p. 69-86.

GIL LEIVA, I. Consistencia en la indización de documentos entre indizadores noveles. *Anales de Documentación*, 2002, vol. 5, p. 99-111. Retrieved 15 juin, 2010 from http://webs.um.es/isgil/

GIL LEIVA, I. *Manual de indización: teoría y práctica.* Gijón: Trea, 2008.

GIL LEIVA, I, POLSINELLI RUBI, M. y SPOTTI LOPES FUJITA, M.. Consistência na indexçao em bibliotecas universitárias brasilerias. *TransInformação*, 2008, vol. 20, nº 3, p. 233-253.

GIL URDICIAÍN, B. Evaluación del rendimiento de los tesauros españoles en sistemas de recuperación de información. *Revista Española de Documentación Científica*, 1998, vol. 21, nº 3, p. 286-302.

GIL URDICIAÍN, B. *Manual de lenguajes documentales*. Gijón: Trea, 2004.

GROSS, T. y TAYLOR, A.G. What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries*, 2005, vol. 66, nº 3, p. 212-230.

HOOPER. R.S. Indexer consistency test: origin, mesaruments, results and utilization. Bethesda, MD: IBM Corp., 1965.

HENZLER, R.G. Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. *International Classification*, 1978, vol. 5, nº 1, p.21-26.

HUDON, M. An assessment of the usefulness of standardized definitions in a thesaurus through interindexer terminological consistency measurements. University of Toronto, 1998 (PhD Thesis).

HUDON, M. The usefulness of standardized definitions in thesauri: An assessment through interindexer consistency measurements. In TOMS, E.G., CAMPBELL, D.G. y DUNN, J. (ed.). 26th Annual Conference of the Canadian-Association-for-Information-Science. 1998, p. 221-230. Retrieved 15 juin, 2010 from http://www.cais-acsi.ca/proceedings/1998/Hudon_1998.pdf.

IIVONEN, M. The impact of the indexing environment on interindexer consistency. In FUGMANN, R. (ed.). Tools for Knowledge organization and the human interface, 1990, vol. 1, p. 259-266.

IIVONEN, M. y KIVIMAKI, K. Common entibies and missing properties: similarities and differences in the indexing of concepts. *Knowledge Organization*, 1998, vol. 25, nº 3, p. 90-102.

KISHIDA, K. et al. A comparative evaluation of thesauri: performance of 'presentation of a body of knowledge' [in Japanese]. *Joho no kagaku to gijutsu*, 1988, vol. 38, nº 10, p. 565-572.

LANCASTER, F.W. *Evaluation of the MEDLARS demand search service.* Washington, D.C. National Library of Medicine, 1968.

LANCASTER, F.W. *El control del vocabulario en la recuperación de la información.* Valencia: Universidad de Valencia, 2002.

LARSEN, M.S. Indexering med eudised thesaurusen (Indexing with the EUDISED thesaurus). *Biblioteksarbedje*, 1988, vol. 9, nº 23/24, p. 95-114.

LEININGER, K. Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science,* 2000, vol. 32, nº 1, p. 4-8.

LEONARD, LE. Inter-indexer consistency and retrieval effectiveness: measurement of relations. Champaign: University of Illinois, 1975 (PhD Thesis).

LEONARD, L.E. Inter-indexer consistency studes, 1954-1975: A review of the literature and summary of study results. University of Illinois, 1977.

MARKEY, K. Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research,* 1984, vol. 6, nº. 2, p. 155-177.

MIDDLETON, M.R. A comparison of indexing consistency and coverage in the AEI, ERIC, and APAIS databases. *Behavioral & Social Sciences Librarian*, 1984, vol. 3, nº 4, p. 33-43.

NESHAT, N. y HORRI, A. A study of subject indexing consistency between the National Library of Iran and Humanities Libraries in the Area of Iranian Studies. *Cataloging & Classification Quarterly*, 2006, vol. 43, nº 1, p. 67-76.

POZHARISKII, I.F. Opredelenie semanticheskoi sily informatsionnopoiskogo tezaurusa (Determining the semantic strength of an information-retrieval thesaurus). *Nauchno- Tekhnicheskaya Informatsiya*, Series, 1982, vol. 2, nº 6, p. 21-25.

RIBEIRO, F. Subject indexing and authority control in archives: the need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists*, 1996, vol. 17, nº 1, p. 27-54.

ROLLING, L. Indexing consistency, quality and efficiency. *Information Processing and Management,* 1981, vol. 17, nº 2, p. 69-76.

SAARTI, J. Consistency of subject indexing of novels by public library professionals and patrons. *Journal of Documentation,* 2002, vol. 58, nº 1, p. 49-65.

SIEVERT, M.E. y ANDREWS, M.J. Indexing consistency in Information Science Abstracts. *Journal of the American Society for Information Science*, 1991, vol. 42, nº1, p. 1-6.

SLAMECKA, V., and JACOBY, J.J. Consistency of human indexing. In "*The Coming Age of Information Technology*". Bethesda, MD: Documentation Inc., 1965, p. 32-56.

SOLER MONREAL, M.C. Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores. Universidad Politécnica de Valencia, 2009 (PhD Thesis).

TARR, D. & BORKO, H. Factors influencing inter- indexer consistency. IN Ingår i Zunde, Pranas, (eds.). In *Proceedings of the 37th annual meeting of the American Society for Information Science,* 13-17 October, Atlanta, Georgia. Washington, D.C.: Asis, 1974, p. 50-55.

TONTA, Y. A Study of indexing consistency between Library of Congress and British Library Catalogers. *Library Resources and Technical Services*, 1991, vol. 35, nº. 2, p. 177-185.

WEINBERG, B.H. y CUNNINGHAM, J.A. The relationship between term specificity in MeSh and online postings in MEDLINE. *Bulletin of the Medical Library Association*, 1985, vol. 73, nº 4, p. 365-372.

Annex A: First terms of the three tools built

Listo f descriptors	Thesaurus	Augmented thesaurus
3W USE: **World Wide Web**	3W USE: **World Wide Web**	3W USE: **World Wide Web**
AACR USE: **Reglas de catalogación**	AACR USE: **Reglas de catalogación**	AACR USE: **Reglas de catalogación**
Abstracts USE: **Resúmenes**	Abstracts USE: **Resúmenes**	Abstracts USE: **Resúmenes**
Accesibilidad USE: **Acceso a la información**	Accesibilidad USE: **Acceso a la información**	Accesibilidad USE: **Acceso a la información**
Accesibilidad de la información USE: **Acceso a la información**	Accesibilidad de la información USE: **Acceso a la información**	Accesibilidad de la información USE: **Acceso a la información**
Accesibilidad universal a la información USE: **Disponibilidad Universal de Publicaciones**	Accesibilidad universal a la información USE: **Disponibilidad Universal de Publicaciones**	Accesibilidad universal a la información USE: **Disponibilidad Universal de Publicaciones**
Acceso a bases de datos	**Acceso a bases de datos** SC: 4000 BT1: Acceso a la información BT2: Derecho a la información BT3: Derecho BT4: Ciencias y técnicas auxiliares	**Acceso a bases de datos** SC: 4000 SN: Obtención de un dato de una base o banco de datos. BT1: Acceso a la información BT2: Derecho a la información BT3: Derecho BT4: Ciencias y técnicas auxiliares
Acceso a la documentación USE: **Acceso al documento (Archivos)**	Acceso a la documentación USE: **Acceso al documento (Archivos)**	Acceso a la documentación USE: **Acceso al documento (Archivos)**
Acceso a la información UP: Accesibilidad Accesibilidad de la información … … …	**Acceso a la información** SC: 4000 UP: Accesibilidad UP: Accesibilidad de la información BT1: Derecho a la información BT2: Derecho BT3: Ciencias y técnicas auxiliares NT1: Acceso a bases de datos NT1: Acceso a los materiales NT1: Acceso remoto RT: Acceso al documento (Archivos) RT: Acceso al documento (Bibliotecas) RT: Derecho de la información RT: Difusión de la información RT: Fuentes de información … … …	**Acceso a la información** SC: 4000 SN: Facilidad para acceder y utilizar un servicio o instalación. UP: Accesibilidad UP: Accesibilidad de la información BT1: Derecho a la información BT2: Derecho BT3: Ciencias y técnicas auxiliares NT1: Acceso a bases de datos NT1: Acceso a los materiales NT1: Acceso remoto RT: Acceso al documento (Archivos) RT: Acceso al documento (Bibliotecas) RT: Derecho de la información RT: Difusión de la información RT: Fuentes de información …

Annex B: Abtracts

Abstract 1

ARAUJO RUÍZ, J.A., ARENCIBIA JORGE, R. y GUTIÉRREZ CALZADO, C. Ensayos clínicos cubanos publicados en revistas de impacto internacional: estudio bibliométrico del período 1991-2001. *Revista Española de Documentación Científica*, 2002, vol. 25, nº 3, p. 254-266.

Con el objetivo de evaluar el alcance de los estudios de investigación clínica generados por las instituciones científicas cubanas, se realizó una búsqueda retrospectiva de los ensayos clínicos publicados en revistas indizadas por las bases de datos MEDLINE y Science Citation Index, y se recuperaron 172 referencias de trabajos publicados con la anuencia de centros de investigación del país. Se identificaron un total de 653 autores de origen cubano y 175 extranjeros. El promedio de autores por artículo fue de 7,16, y los colectivos de autores más comunes estuvieron integrados por más de seis especialistas. 82 ensayos clínicos fueron producto de la colaboración entre varias instituciones, donde participaron 83 centros de investigación, 47 de ellos cubanos y 36 extranjeros. 96 publicaciones periódicas de 17 países se encargaron de publicar los 172 ensayos clínicos, y los artículos publicados en lengua inglesa constituyeron el 74,4 % del total. 63 productos, técnicas o procedimientos terapéuticos fueron ensayados en los distintos tipos de pacientes, con el objetivo de tratar 41 padecimientos. Los adultos humanos, con relativo equilibrio entre hombres y mujeres, fueron los sujetos que con mayor frecuencia se estudiaron. El análisis bibliométrico permitió confirmar los avances de Cuba en cuanto a la realización de ensayos clínicos para legitimar los productos generados por su industria médico-farmacéutica, así como definir los centros que marchan a la vanguardia en ese sentido.

The aim of this work is to assess the scope of the clinical research performed by Cuban scientific institutions. A retrospective search about clinical trials published by journals indexed in *MEDLINE* and *Science Citation Index* was carried out, and 172 references to works published with the participation of Cuban research centers were retrieved. A group of 653 Cuban and 175 foreign authors were identified. The average of authors by article was 7,16, and the most common author groups were made up of more than six specialists. A total of 82 clinical trials were the result of collaborations between scientific institutions; 83 research centers took part in the trials, 36 of them from others countries. The reports about the 172 clinical trials were published in 96 journals from 17 countries, and the 74,4 % of the articles were written in English. Sixty-three therapeutic products, technics and procedures were tested in different types of patients, and 41 disorders were treated. Human adults, with a relative balance between women and men, were the subjects most frequently studied. The bibliometric study made possible to confirm the cuban advances as regards to the clinical trials execution for the authentication of products reached by the medical-pharmaceutical industry, as well as to define the research centers in the vanguard regarding this subject.

Abstract 2

CARO CASTRO, C., CEDEIRA SERANTES, L. y TRAVIESO RODRÍGUEZ, C. La investigación sobre recuperación de información desde la perspectiva centrada en el usuario: métodos y variables. *Revista Española de Documentación Científica*, 2003, vol. 26, nº 1, p. 40-50.

La incorporación del usuario a los estudios sobre recuperación de información ha supuesto el desarrollo de una línea de investigación que centra su atención en cómo se formalizan las necesidades de información y se plantean las búsquedas, cómo interactúa el usuario con el sistema, cómo se valoran los resultados obtenidos y qué características individuales influyen en el proceso. En este estudio exploratorio se analizan 25 trabajos de investigación original que se enmarcan en la perspectiva centrada en el usuario. Se ha establecido una clasificación para las variables cuyas categorías principales son: características de los usuarios, características del sistema, entorno y proceso de búsqueda, y resultados. La coincidencia en la utilización de las técnicas de recogida de datos, métodos de análisis y variables ha servido para comprobar la similitud de las investigaciones y para obtener una representación gráfica de las diferentes orientaciones que se pueden observar entre ellas.

User has been included in the research in information retrieval, and this factor has implied a new research approach. This perspective is focused on new issues related to searching process, such as formulation of queries, interaction between user and system, evaluation of the results obtained and influence of some personal characteristics. This exploratory paper examines 25 original research works of the user-centered perspective. A classification of the variables according to the following categories has been established: user characteristics, searching environment and process, and results. The coincidence in data collecting techniques, analysis methods and variables has served for checking the similarity between the research works analysed. Finally, a graphical representation of the different trends observed in these works is presented.

Abstract 3

ALCAIN, Mª D., et a. Evaluación de las bases de datos ISOC a través de un estudio de usuarios. Homenaje a José María Sánchez Nistal. *Revista Española de Documentación Científica*, 2001, vol. 24, nº 3, p. 275-288.

El objetivo del presente estudio es realizar un acercamiento al estado de la cuestión de los estudios de usuarios en la gestión y evaluación de calidad de bases de datos y aplicar los modelos existentes a un caso real, la base de datos ISOC (CSIC). Para ello se diseñan dos cuestionarios, uno para usuarios finales y otro para bibliotecarios referencistas. Los resultados muestran diferencias entre ambos colectivos tanto en el uso, como en motivos de consulta, objetivos y nivel de satisfacción. También se observan objetivos y grado de satisfacción diferentes entre los dos grupos de usuarios más destacados, el mayoritario de 'investigadores y docentes universitarios' y el menos numeroso de 'estudiantes'. Los datos obtenidos permiten establecer una radiografía sobre el uso de la base ISOC y obtener algunos indicadores de valoración de la misma. Se concluye que ISOC es muy utilizada y bien valorada en términos generales y se

considera importante seguir realizando este tipo de estudios para poder hacer un seguimiento de las demandas de los usuarios.

The objective of this work is to approach the state of the art of the user's studies about Quality Management and Evaluation of data bases, and to apply the existing models to a real case: the ISOC data base. To this end, two questionnaires have been designed: one addressed to end users and the other to reference librarians. The results show the differences between the two groups in the use, reasons for consultation, objectives and satisfaction. A difference in objectives and level of satisfaction has also been found between the two main users: researchers and professors on the one hand, and students, on the other. The results allow to establish a map of the use of ISOC data base and to obtain some value indicators. It is concluded that ISOC is very used and generally well valued. This kind of studies have proved to be necessary in order to follow up users' requirements.

Annex C: Results

GROUP 1: Second year students

Controlled vocabulary used: **List of descriptors**

Table 1: Consistency for group 1 with the list of descriptors as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	0	42	0
Indexer 1 versus Indexer 3	50	28	18
Indexer 1 versus Indexer 4	20	42	08
Indexer 1 versus Indexer 5	50	30	09
Indexer 1 versus Indexer 6	20	14	09
Indexer 2 versus Indexer 3	0	75	33
Indexer 2 versus Indexer 4	0	100	33
Indexer 2 versus Indexer 5	0	46	40
Indexer 2 versus Indexer 6	0	20	27
Indexer 3 versus Indexer 4	50	75	66
Indexer 3 versus Indexer 5	50	42	28
Indexer 3 versus Indexer 6	20	25	20
Indexer 4 versus Indexer 5	50	46	28
Indexer 4 versus Indexer 6	20	20	20
Indexer 5 versus Indexer 6	20	12	23
Mean	**23,3%**	**41,1%**	**24,1%**

Overall mean: 29,5%

Controlled vocabulary used: **Augmented thesaurus**

Table 2: Consistency for group 1 with the augmented thesaurus as %

	Abstract	**Abstract**	**Abstract 3**
Indexer 1 versus Indexer 2	60	50	30
Indexer 1 versus Indexer 3	20	33	40
Indexer 1 versus Indexer 4	28	06	0
Indexer 1 versus Indexer 5	16	75	50
Indexer 1 versus Indexer 6	28	0	50
Indexer 2 versus Indexer 3	20	12	16
Indexer 2 versus Indexer 4	16	28	07
Indexer 2 versus Indexer 5	25	33	18
Indexer 2 versus Indexer 6	27	25	18
Indexer 3 versus Indexer 4	14	23	0
Indexer 3 versus Indexer 5	20	40	16
Indexer 3 versus Indexer 6	23	28	75
Indexer 4 versus Indexer 5	20	16	0
Indexer 4 versus Indexer 6	33	28	16
Indexer 5 versus Indexer 6	20	14	50
Mean	**24,6%**	**27,4%**	**25,7%**

Overall mean: 25,9%

Controlled vocabulary used: **Thesaurus**

Table 3: Consistency for group 1 with the thesaurus as %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	14	40	22
Indexer 1 versus Indexer 3	20	20	28
Indexer 1 versus Indexer 4	16	50	50
Indexer 1 versus Indexer 5	14	66	20
Indexer 1 versus Indexer 6	20	40	14
Indexer 2 versus Indexer 3	14	16	27
Indexer 2 versus Indexer 4	12	40	37
Indexer 2 versus Indexer 5	25	20	37
Indexer 2 versus Indexer 6	33	33	44
Indexer 3 versus Indexer 4	16	20	12
Indexer 3 versus Indexer 5	33	25	12
Indexer 3 versus Indexer 6	20	16	22
Indexer 4 versus Indexer 5	12	66	50
Indexer 4 versus Indexer 6	16	40	33
Indexer 5 versus Indexer 6	60	50	60
Mean	**21,6%**	**36,1%**	**31,2%**

Overall mean: 29,6%

GROUP 2: Fourth year students

Controlled vocabulary used: **List of descriptors**

Table 1: Consistency for group 2 with the list of descriptors as %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	42	50	25
Indexer 1 versus Indexer 3	23	80	22
Indexer 1 versus Indexer 4	33	66	33
Indexer 1 versus Indexer 5	20	60	25
Indexer 1 versus Indexer 6	28	57	18
Indexer 2 versus Indexer 3	50	66	28
Indexer 2 versus Indexer 4	25	57	40
Indexer 2 versus Indexer 5	33	80	33
Indexer 2 versus Indexer 6	40	50	27
Indexer 3 versus Indexer 4	50	57	36
Indexer 3 versus Indexer 5	20	80	28
Indexer 3 versus Indexer 6	38	50	20
Indexer 4 versus Indexer 5	25	60	27
Indexer 4 versus Indexer 6	36	42	48
Indexer 5 versus Indexer 6	16	37	12
Mean	**32%**	**59,4%**	**28,1%**

Overall mean: 39,8%

Controlled vocabulary used: **Augmented thesaurus**

Table 2: Consistency for group 2 with the augmented thesaurus as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	66	25	18
Indexer 1 versus Indexer 3	40	14	25
Indexer 1 versus Indexer 4	11	60	11
Indexer 1 versus Indexer 5	23	100	50
Indexer 1 versus Indexer 6	50	75	25
Indexer 2 versus Indexer 3	33	14	30
Indexer 2 versus Indexer 4	10	26	30
Indexer 2 versus Indexer 5	20	33	18
Indexer 2 versus Indexer 6	40	29	41
Indexer 3 versus Indexer 4	30	42	27
Indexer 3 versus Indexer 5	29	50	25
Indexer 3 versus Indexer 6	28	50	40
Indexer 4 versus Indexer 5	16	60	11
Indexer 4 versus Indexer 6	09	50	40
Indexer 5 versus Indexer 6	17	75	25
Mean	**28,1%**	**46,8%**	**27,6%**

Overall mean: 34,1%

Controlled vocabulary used: **Standard thesaurus**

Table 3: Consistency for group 2 with the thesaurus as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	8	22	60
Indexer 1 versus Indexer 3	16	10	18
Indexer 1 versus Indexer 4	36	27	14
Indexer 1 versus Indexer 5	30	20	16
Indexer 1 versus Indexer 6	41	33	21
Indexer 2 versus Indexer 3	16	20	14
Indexer 2 versus Indexer 4	33	12	20
Indexer 2 versus Indexer 5	20	16	25
Indexer 2 versus Indexer 6	25	16	33
Indexer 3 versus Indexer 4	0	12	14
Indexer 3 versus Indexer 5	16	0	66
Indexer 3 versus Indexer 6	10	0	17
Indexer 4 versus Indexer 5	60	25	25
Indexer 4 versus Indexer 6	20	66	33
Indexer 5 versus Indexer 6	25	33	23
Mean	**23,7%**	**20,8%**	**26,6%**

Overall mean: 23,7%

GROUP 3: Fifth year students

Controlled vocabulary used: **List of descriptors**

Table 1: Consistency for group 3 with the list of descriptors as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	50	100	50
Indexer 1 versus Indexer 3	14	57	20
Indexer 1 versus Indexer 4	20	16	28
Indexer 1 versus Indexer 5	16	50	66
Indexer 1 versus Indexer 6	33	80	57
Indexer 2 versus Indexer 3	11	37	18
Indexer 2 versus Indexer 4	14	16	42
Indexer 2 versus Indexer 5	12	50	37
Indexer 2 versus Indexer 6	23	80	29
Indexer 3 versus Indexer 4	42	11	8
Indexer 3 versus Indexer 5	22	50	16
Indexer 3 versus Indexer 6	25	33	15
Indexer 4 versus Indexer 5	12	50	22
Indexer 4 versus Indexer 6	33	14	26
Indexer 5 versus Indexer 6	12	42	44
Mean	**22,6%**	**45,7%**	**31,8%**

Overall mean: 33,3%

Controlled vocabulary used: **Augmented thesaurus**

Table 2: Consistency for group 3 with the augmented thesaurus as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	33	75	30
Indexer 1 versus Indexer 3	42	44	14
Indexer 1 versus Indexer 4	66	50	35
Indexer 1 versus Indexer 5	12	50	38
Indexer 1 versus Indexer 6	12	37	15
Indexer 2 versus Indexer 3	42	44	0
Indexer 2 versus Indexer 4	25	50	50
Indexer 2 versus Indexer 5	12	50	57
Indexer 2 versus Indexer 6	12	22	33
Indexer 3 versus Indexer 4	60	57	10
Indexer 3 versus Indexer 5	16	83	11
Indexer 3 versus Indexer 6	16	42	0
Indexer 4 versus Indexer 5	16	66	44
Indexer 4 versus Indexer 6	16	28	25
Indexer 5 versus Indexer 6	71	50	28
Mean	**31,7%**	**49,8%**	**26%**

Overall mean: 35,8%

Controlled vocabulary used: **Standard thesaurus**

Table 3. Consistency for group 3 with the thesaurus as %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	60	22	18
Indexer 1 versus Indexer 3	33	10	16
Indexer 1 versus Indexer 4	14	27	36
Indexer 1 versus Indexer 5	16	20	30
Indexer 1 versus Indexer 6	16	33	41
Indexer 2 versus Indexer 3	20	20	16
Indexer 2 versus Indexer 4	50	28	45
Indexer 2 versus Indexer 5	25	16	33
Indexer 2 versus Indexer 6	25	16	25
Indexer 3 versus Indexer 4	20	0	12
Indexer 3 versus Indexer 5	66	0	16
Indexer 3 versus Indexer 6	25	0	10
Indexer 4 versus Indexer 5	25	25	60
Indexer 4 versus Indexer 6	25	66	33
Indexer 5 versus Indexer 6	33	33	25
Mean	**30,2%**	**21%**	**27,7%**

Overall mean: 26,3%

GROUP 4: Expert indexers

Controlled vocabulary used: **List of descriptors**

Table 1: Consistency for group 4 with the list of descriptors as %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	42	75	27
Indexer 1 versus Indexer 3	42	60	25
Indexer 2 versus Indexer 3	71	75	87
Mean	**51%**	**70%**	**46,3%**

Overall mean: 55,7%

Controlled vocabulary used: **Augmented thesaurus**

Table 2: Consistency for group 4 with the augmented thesaurus as %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	27	28	33
Indexer 1 versus Indexer 3	0	20	42
Indexer 2 versus Indexer 3	14	28	22
Mean	**13,6%**	**25,3%**	**32,3%**

Overall mean: 23,7%

Controlled vocabulary used: **Standard thesaurus**

Table 3: Consistency for group 4 with the standard thesaurus as %

	Abstract 1	**Abstract 2**	**Abstract 3**
Indexer 1 versus Indexer 2	28	33	23
Indexer 1 versus Indexer 3	28	60	20
Indexer 2 versus Indexer 3	20	60	11
Mean	**25%**	**51%**	**18%**

Overall mean: 31,3%