Evaluation of controlled vocabularies by inter-indexer consistency

Concha Soler Monreal
Universidad de Valencia, Facultad de Medicina,Avenida Blasco Ibañez, 13, 46010 Valencia, Spain
Isidoro Gil-Leiva
Universidad de Murcia, Facultad de Comunicación y Documentación, Campus de Espinardo s/n, 30100 Murcia, Spain

Introduction. *Several controlled vocabularies are used for indexing three journal articles to check if better or equal consistency rates are achieved with a list of descriptors than with a standard thesaurus and augmented thesaurus.*
Method. *A terminology set for library and information Science was used to build a list of descriptors with equivalence relations (USE and UF), a standard thesaurus and an augmented thesaurus (all the descriptors have scope notes). Subsequently, three articles were indexed by selected indexers with 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 to pairs of novice indexers and experts.*
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 more research is required.*

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: 118-154. The first contribution was the work done by Charles Ammi Cutter in his famous *Rules for a printed dictionary catalog* published in 1876. It is here that the first rules appear that are in effect today, such as the principle of economy, the definition and use of the term headings for both matter and for place and form; re-sending 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 Catalogs 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 for 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 either 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 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 to 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 Department 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 the impact on the information systems which use them both in indexing and retrieval (extrinsic evaluation).

The first evaluation of import was carried out by Cleverdon in the Cranfield Projects (1956; 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 (Medical Subject Headings), 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 MeSH and Medline, while Pozhariskii (1982) proposed quantifying the capacity or semantic strength of a thesaurus in terms of flexibility, economy and universality. Elsewhere, Larsen (1988) analysed the capacities for use of a thesaurus for indexing a certain collection of documents. Soler Monreal (2009) evaluated three controlled vocabularies (a list of descriptors, a standard thesaurus and an augmented thesaurus in which all the descriptors have scope notes) in order to find out if consistency scores higher than a standard thesaurus and augmented thesaurus are obtained with a list of descriptors.

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 intra-consistency or intra-indexer 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 inter-consistency or inter-indexer consistency.

Since the 1960s, numerous and diverse investigations have been carried out on indexing consistency. The main conclusion which can be drawn from the tests is that inconsistency is an inherent feature of indexing, rather than a sporadic anomaly. Although the tests carried out are very diverse in their methodology, we can say that achieved indexing consistency ranges from approximately 10% to 60%. The vast majority of the tests carried out from 1960 until the present time cannot be homogenized because of the methodological diversity used. We only point out here some of the variables that hinder their homogenization and only a sample of the tests 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, other 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 is the common terms between the two indexers, a is the terms proposed by indexer one, and b is the terms proposed by indexer two.
- Novice versus expert indexers: On some occasions expert indexers were employed as in Lancaster (1968), Leonard (1975); on others it was 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 three indexers; Bertrand and Cellier (1995) twenty-five indexers; Hudon (1998a and 1998b) twenty-five indexers; Gil-Leiva (2002) twenty-seven indexers; Saarti (2002) thirty indexers; and Soler Monreal (2009) sixty-three indexers.
- Material used: Sometimes work has been with journal articles (Lancaster 1968; Leonard 1977; Funk and Reid 1983; Middleton 1984; Sievert 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 *et al.* 2008; Chen 2008) and at other times with visual material (Markey 1984 and Gil-Leiva 2002); summaries of journal articles have also been used (Hudon 1998; Soler Monreal 2009); social tagging on delicious.com Kipp 2009) and social tagging on CiteULike.org (Wolfram *et al.* 2009).
- Number of documents indexed: Lancaster (1968) used sixteen articles; Tarr and Borko (1974) fifteen items;

Leonard ([1975](#)) 100 articles; Funk and Reid ([1983](#)) 760 articles; Markey ([1984](#)) 100 documents; Iivonen ([1990](#)) ten documents; Siever and Andrews ([1991](#)) seventy-one articles; Iivonen and Kivimäki ([1998](#)) forty-nine documents; Leninger ([2000](#)) sixty documents; Tonta ([1991](#)) eighty-two books; Bertrand and Cellier ([1995](#)) eight books; Gil-Leiva ([2001](#)) eleven books; Saarti ([2002](#)) five books; Gil-Leiva *et al.* ([2008](#)) ten books; Chen ([2008](#)) 3,307 monographs; Hudon ([1998a](#) and [1998b](#)) twelve abstracts; Monreal Soler ([2009](#)) three abstracts.

- Hawthorne effect: Individuals, when they know that they are being studied behave differently from when they do not know they are being 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](#)) compares the indexing of ERIS/APAIS and AEI/APAIS; Iivonen and Kivimäki ([1998](#)) databases KINF and LISA; or compare the indexing library catalogues like Tonta ([1991](#)) who compared the Library of Congress and the British Library; Gil-Leiva ([2001](#)) in thirty-one catalogues of public libraries; Neshat and Horri ([2006](#)) National Library of Iran and twelve academic libraries; Gil-Leiva *et al.* ([2008](#)) thirty university library catalogues; 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](#)) worked with seventy-one duplicates of the database Information Science Abstracts; Leininger ([2000](#)) compared sixty duplicates of the PsycINFO database.

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

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 there did not exist in Spanish a thesaurus published on this subject. Hence, we began to refine a list of descriptors consisting of 2,756 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 2,455 terms, of which 1,436 are descriptors and 1,019 non-descriptors. A standard thesaurus was constructed from this list. This thesaurus has an alphabetic display, another hierarchical one and other types KWOC permuted index. [Appendix A](#) shows the first terms of the three tools built.

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

Centralized acquisition	Topographic catalogues
TC: J02 UP: Centralized Purchases TG1: Acquisition of documents TG2: Development of collections TG3: Documental process	TC: F03 TG1: Catalogues (information sources) TG2: Secondary sources TG3: Information sources

Table 1: Descriptors of the standard thesaurus

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

Centralized acquisition	Topographic catalogues
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	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

Table 2: Descriptors from augmented thesaurus

After building the three controlled vocabularies, an intrinsic (qualitative and quantitative) evaluation was 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 to be indexed was to be three abstracts of journal articles since these are concise, well structured and understandable information sources (Appendix B). We then worked on the selection of the indexers who were going to use the three indexing languages to index 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 eighteen 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 nine 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 fields. 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 twenty-seven comparisons.

We used a relaxed, and non exact, system of coincidence to calculate consistency between indexers, as was done in Gil-Leiva (2001) and Gil- Leiva *et al.* (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.

Indexer 1	Indexer 2	Agreement
Biomedical journals	Scientific journals	0.5
Librarians' techniques	Librarianship	1
Databases	Bibliographical databases	0.5
Librarians	Librarians of reference	0.5
Scientific journals	Scientific publications	1

Table 3: Table of relaxed equivalences between descriptors

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 *et al.* (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 2. Cite frequency 3. Biomedical journals	1. Librarianship 2. Medical documentation 3. Impact factor 4. Scientific journals	= 1 =0.5

Table 4: Equivalences between descriptors

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 Appendix C.

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

	List of descriptors	Augmented thesaurus	Standard thesaurus
Second year students	29.5	25.9	29.6
Fourth year students	39.5	34.1	23.7
Fifth year students	33.3	35.8	26.3
Means	**34.2 %**	**31.9 %**	**26.5 %**

Table 5: Means of the results of the novice indexers
with the three tools as %

However, data for consistency among indexers experts are:

	List of descriptors	Augmented thesaurus	Standard thesaurus
Expert indexers	55.7 %	23.7 %	31.3 %

Table 6: Means of the results for the expert indexers
with the three tools as %

We have also obtained the mean for all the consistency obtained, for both expert and novice indexers, as can be seen in Table 7.

	List of descriptors	Augmented thesaurus	Standard thesaurus
Second year students	29.5	25.9	29.6
Fourth year students	39.5	34.1	23.7
Fiftht year students	33.3	35.8	26.3
Expert indexers	55.7	23.7	31.3
Means	**39.5%**	**29.8%**	**27.7%**

Table 7: Means of the results for all indexers with the
three tools as %

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%).
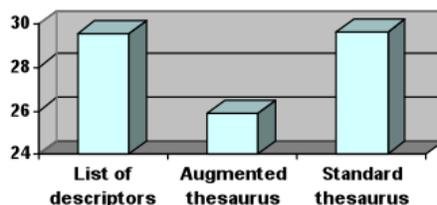

Figure 1: Group 1 - Second year students

In group 2 (Fourth year students) 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 shown in Figure 2 below:
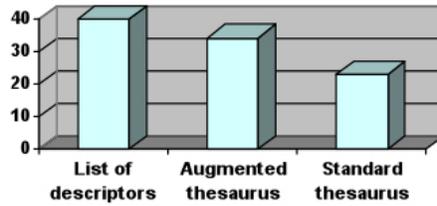
Figure 2: Group 2 - Fourth year students

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 shown in Figure 3.
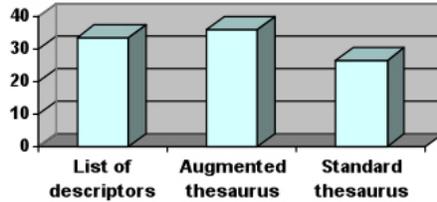

Figure 3: Group 3 - Fifth year students

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 4 shows.
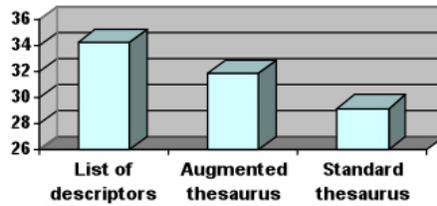

Figure 4: All novice indexers

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%, was 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.
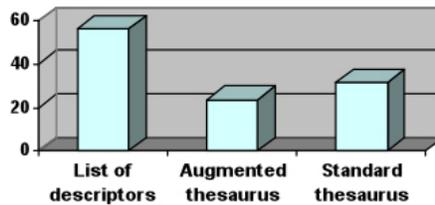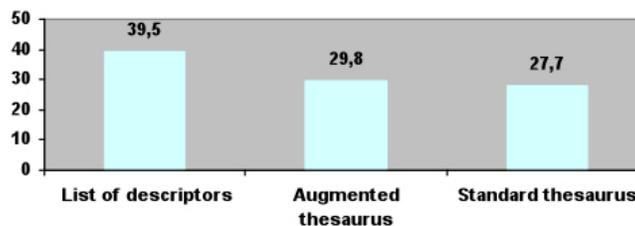

Figure 5: Expert indexers

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 a similar value, the standard thesaurus (27.7%), as can be seen in Figure 6.

Figure 6: All indexers

With the exception of two cases, the highest consistency occurs for abstract Number Two.

The results obtained in this study fall within the margins of consistency obtained in other previous studies, ranging from approximately 10% to 60% (Lancaster 1968; Leonard 1975; Funk and Reid 1983; Markey 1984; Middleton 1984; Tonta 1991; Sievert and Andrews 1991; Iivonen and Kivimäki 1998; Leininger 2000; Gil-Leiva 2001 and 2002; Saarti 2002; Neshat and Horri 2006; Gil-Leiva *et al.* 2008 and Kipp 2009). We also said in the introduction to this article that we can compare our results with the data achieved by Hudon (1998a and 1998b). Hudon also used three versions of a thesaurus - a) standard thesaurus, b) standard thesaurus with definitions for all their descriptors and c) stripped thesaurus with 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 minor), indexers who worked with the augmented thesaurus did not obtain better consistency than those who worked with the standard thesaurus; b) indexers that used the stripped thesaurus were equally or more consistent with each other than those who 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 seven of the twelve documents, and the stripped thesaurus indexers were more consistent than the standard thesaurus indexers in eight of the twelve 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 achieving acceptable levels of consistency in the selection of the main descriptor when using a stripped thesaurus.

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

In any case, further research is required to corroborate these results. One important limitation of this research that should be pointed out is that in the two comparative studies we do not know what percentage of inconsistency is due to the complexity and subjectivity of indexing (reading and analysis of the document and selection of the appropriate keywords) or 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 inconsistency is due to the indexing languages used.

Conclusions

Inconsistency is an inherent feature of indexing, as we have seen from the data obtained in research conducted since the 1960s to the present day. Precisely because of the disparity of variables used in the research, it 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 has 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 a small information system is always easier to build a list of descriptors than a thesaurus. Further examination of the properties of the lists of descriptors compared to thesauri could be a line of study to follow but with the inclusion of techniques using verbal protocols or 'thinking aloud' during the process of indexing the documents with thesauri. This technique will allow valuable information to be gathered on the use that indexers make of associative and hierarchical relations.

Acknowledgements

About the authors

Concha Soler Monreal gained her PhD in 2009. She works as an information manager in a public television company and has written several research papers. She has been part-time professor in the University of Valencia. She can be contacted at: solermonreal@telefonica.net\

Isidoro Gil-Leiva gained his PhD in 1997 in Philosophy and Arts and is Professor of Information and Library Science at the University of Murcia. He has written several academic handbooks and research papers and has participated in projects in the field of library and information science. Currently is the editor in chief of the journal *Anales de Documentación*. He can be contacted at: isgil@um.es and http://webs.um.es/isgil/

- Bertrand,A. & Cellier, J.M. (1995). Psychological approach to indexing: effects of the operartor's expertise upon indexing behaviour. *Journal of Information Science*, **21**(6), 459-472

- Betts, R. & Marrable, D. (1991). Free text vs. controlled vocabulary—retrieval precision and recall over large databases. In David I. Raitt, (Ed.). *Online Information 91: Proceedings of the Fifteenth International Online Information Meeting, London, 10-12 December 1991*, (pp. 153-165). Oxford: Learned Information.

- Cleverdon, C.W. (1960). ASLIB Cranfield research project: report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield, UK: College of Aeronautics. Retrieved 25 November, 2011 from https://dspace.lib.cranfield.ac.uk/bitstream/1826/1122/1/1960b%20ASLIB%20.pdf (Archived by WebCite® at http://www.webcitation.org/63T0U0JTg)

- Cleverdon, C.W. (1956). Proposals for an investigation into the efficiency of various retrieval systems. Cranfield, UK: College of Aeronautics. Retrieved 25 November, 2011 from https://dspace.lib.cranfield.ac.uk/bitstream/1826/1367/1/1956.pdf (Archived by WebCite® at http://www.webcitation.org/63T0hOUZs)

- Chen, X. (2008). *Indexing consistency between online catalogues*. Unpublished doctoral dissertation, Humboldt University, Berlin, Germany. Retrieved 25 November, 2011 from http://edoc.hu-berlin.de/dissertationen/chen-xu-2008-05-14/PDF/chen.pdf (Archived by WebCite®at http://www.webcitation.org/63T12WuHC)

- Fidel, R. (1991). Searchers' selection of search keys: 2. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science*, **42**(7), 501-514

- Fidel, R. (1992). Who needs controlled vocabulary? *Special Libraries* **83**(1), 1-9

- Funk, M.E. & Reid. C.A. (1983). Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association*, **71**(2), 176-183

- Gil-Leiva, I. (1997). *La automatización de la indización. Propuesta teórico-metodológíca: aplicación al área de biblioteconomía y documentación.* [The automation of indexing. Theoretico-methodological proposal: applied to the area of librarianship and documentation.] Unpublished doctoral dissertation, Universidad de Murcia, Murcia, Spain.

- Gil-Leiva, I. (2001). Consistencia en la asignación de materias en las Bibliotecas Públicas del Estado. [Consistency in the assignment of materials in state public libraries.] *Boletín de la Asociación Andaluza de Bibliotecarios*, 63, 69-86. Retrieved 15 June, 2010 from http://webs.um.es/isgil/Consistencia%20indizacion%20bibliotecas%20libraries%20indexing%20consistency%20Gil%20Leiva.pdf (Archived by WebCite® at http://www.webcitation.org/63T1gx53n)

- Gil-Leiva, I.(2002). Consistencia en la indización de documentos entre indizadores noveles. [Consistency in the indexing of documents between novice indexers] *Anales de Documentación***5**, 99-111. Retrieved 25 November, 2011 from http://revistas.um.es/analesdoc/article/download/2211/2201 (Archived by WebCite® at http://www.webcitation.org/63T1ziJQr)

- Gil-Leiva, I. (2008). *Manual de indización: teoría y práctica.* [Manual of indexing: theory and practice.] Gijón, Spain: Trea

- Gil-Leiva, I., Polsinelli Rubi, M. & Spotti Lopes Fujita, M. (2008). Consistência na indexaçao em bibliotecas universitárias brasilerias. [Consistency of indexing in Brazilian University libraries.] *TransInformação* **20** (3),233-253

- Gil Urdiciaín, B. (1998). Evaluación del rendimiento de los tesauros españoles en sistemas de recuperación de información. [Performance evaluation of Spanish thesauri in information retrieval systems.] *Revista Española de Documentación Científica*, **21**(3), 286-302

- Gil Urdiciaín, B. (2004). *Manual de lenguajes documentales*. [Manual of documentary languages.] Gijón, Spain: Trea

- Gross, T.& Taylor, A.G. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. College & Research Libraries **66**(3), 212-230

- Henzler, R.G. (1978). Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. *International Classification*, **5**(1), 21-26

- Hooper. R.S. (1965). Indexer consistency test: origin, measurements, results and utilization. Bethesda, MD: IBM Corp

- Hudon, M. (1998a). *An assessment of the usefulness of standardized definitions in a thesaurus through inter-indexer terminological consistency measurements.* Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

- Hudon, M. (1998b). The usefulness of standardized definitions in thesauri: An assessment through interindexer consistency measurements. In Toms, E.G., Campbell, D.G. & Dunn, J. (ed.). *Proceedings 26th Annual Conference of the Canadian Association for Information Science*, (pp. 221-230). Toronto, Canada: Canadian Association for Information Science. Retrieved 25 November, 2011 from http://www.cais-acsi.ca/proceedings/1998/Hudon_1998.pdf (Archived by WebCite® at http://www.webcitation.org/63T2hHQHD)

- Iivonen, M. (1990). The impact of the indexing environment on interindexer consistency. In Robert Fugmann, (Ed.). *Tools for knowledge organization and the human interface. Proceedings : 1st International ISKO-Conference, Darmstadt, 14-17 August, 1990*, Volume 1, (pp. 259-266). Würzburg, Germany: Egon-Verlag.

- Iivonen, M. & Kivimaki, K. (1998). Common entibies and missing propoerties: similarities and differences in the indexing of concepts. *Knowledge Organization*, **25**(3), 90-102

- Kishida, K., Mushakoji, S., Matsuyama, N., Higuchi, K., Inagaki, K. & Ueda, S. (1988). A comparative evaluation of thesauri: performance of 'presentation of a body of knowledge' [in Japanese]. *Joho no kagaku to gijutsu*, **38**(10), 565-572

- Kipp, M.E.I. (2009). *Exploring measures of inter-tagger consistency*. Poster presented at the SIG-CR Workshop, Annual Meeting of the American Society for Information Science and Technology, Vancouver, British Columbia, Canada. Retrieved 26 November, 2011 from http://eprints.rclis.org/bitstream/10760/13794/1/sigcrposter2009.pdf (Archived by WebCite® at http://www.webcitation.org/63TvjAw1G)

- Lancaster, F.W. (1968). *Evaluation of the MEDLARS demand search service.* Washington, D.C. National Library of Medicine

- Lancaster, F.W. (2002). *El control del vocabulario en la recuperación de la información.* [Vocabulary control in information retrieval.] Valencia, Spain: Universidad de Valencia
- Larsen, M.S. (1988). Indexering med eudised thesaurusen. [Indexing with the EUDISED thesaurus]. *Biblioteksarbedje*, **9**(23/24), 95-114
- Leininger, K. (2000). Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science,* **32**(1), 4-8
- Leonard, LE. (1975). *Inter-indexer consistency and retrieval effectiveness: measurement of relations.* Unpublished doctoral dissertation, University of Illinois, Champaign, Illinois, USA.
- Leonard, L.E. (1977). *Inter-indexer consistency studes, 1954-1975: a review of theliterature and summary of study results.* Champaign, IL; Graduate School of Library and Information Science. University of Illinois. (Occasional papers no. 131). Retrieved 26 November, 2011 from https://www.ideals.illinois.edu/bitstream/handle/2142/3885/gslisoccasionalpv00000i00131.pdf?sequence=1 (Archived by WebCite® at http://www.webcitation.org/63TwQO5q6)
- Markey, K. (1984). Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research,* **6**(2), 155-177
- Middleton, M.R. (1984). A comparison of indexing consistency and coverage in the AEI, ERIC, and APAIS databases. *Behavioral & Social Sciences Librarian*, **3**(4), 33-43
- Neshat, N. & Horri, A. (2006). A study of subject indexing consistency between the National Library of Iran and humanities libraries in the area of Iranian studies. *Cataloging & Classification Quarterly*, **43**(1), 67-76
- Pozhariskii, I.F. (1982). Opredelenie semanticheskoi sily informatsionnopoiskogo tezaurusa. [Determining the semantic strength of an information-retrieval thesaurus]. *Nauchno-Tekhnicheskaya Informatsiya*, **2**(6), 21-25
- Ribeiro, F. (1996). 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,* **17**(1), 27-54
- Rolling, L. (1981). Indexing consistency, quality and efficiency. *Information Processing and Management,* **17**(2), 69-76
- Saarti, J. (2002). Consistency of subject indexing of novels by public library professionals and patrons. *Journal of Documentation,* **58**(1), 49-65
- Sievert, M.E. & Andrews, M.J. (1991). Indexing consistency in Information Science Abstracts. *Journal of the American Society for Information Science*, **42**(1), 1-6
- Slamecka, V. & Jacoby, J.J. (1965). Consistency of human indexing. In *The coming age of information technology.* (pp. 32-56). Bethesda, MD: Documentation Inc.
- Soler Monreal, M.C. (2009). *Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores.* [Evaluation of controlled vocabulary indexing of documents using indices of consistency between indexers.] Unpublished doctoral dissertation, Universidad Politécnica de Valencia, Valencia, Spain.
- Tarr, D. & Borko, H. (1974). Factors influencing inter-indexer consistency. In *Proceedings of the 37th annual meeting of the American Society for Information Science, 13-17 October, Atlanta, Georgia.* (pp. 50-55). Washington, DC: ASIS.
- Tonta, Y. (1991). Study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources and Technical Services*, **35**(2), 177-185
- Weinberg, B.H. & Cunningham, J.A. (1985). The relationship between term specificity in MeSh and online postings in MEDLINE. *Bulletin of the Medical Library Association*, **73**(4), 365-372
- Wolfram, D., Olson, H.A. & Bloom, R. (2009). Measuring consistency for multiple taggers using vector space modeling. *Journal of the American Society for Information Science and Technology*, **60**(10), 1995-2003

Find other papers on this subject

Check for citations, using Google Scholar

Bookmark This Page

Appendices

Appendix A: First terms of the three tools built

List of 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
UP: 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 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

Appendix B: Abstracts

Abstract 1

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, techniques 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 it 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.>

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.

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 us to establish a map of the use of ISOC data base and to obtain some value indicators. It is concluded that ISOC is widely used and generally well valued. These kinds of studies have proved to be necessary in order to follow up users' requirements.

Appendice C:

Group 1: Second year students

Controlled vocabulary used: **List of descriptors**. Consistency in %

	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	0.8
Indexer 1 versus Indexer 5	50	30	0.9
Indexer 1 versus Indexer 6	20	14	0.9
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**. Consistency in %

	Abstract 1	Abstract 2	Abstract 3
Indexer 1 versus Indexer 2	60	50	30
Indexer 1 versus Indexer 3	20	33	40
Indexer 1 versus Indexer 4	28	0.6	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	0.7
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: **Standard thesaurus**. Consistency in %

	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

	Abstract 1	Abstract 2	Abstract 3
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**. Consistency in %

	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**. Consistency in %

	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	11	60	11

4			
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	0.9	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**. Consistency in %

	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**. Consistency in %

	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**. Consistency in %

	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

	Abstract 1	Abstract 2	Abstract 3
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**. Consistency in %

	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**. Consistency in %

	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**. Consistency in %

	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**. Consistency in %

	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 %**		