Consistency between indexers in the LILAC database (Latin American and Caribbean Health Science Literature)

Luis Miguel Moreno Fernández
Department of Information and Documentation, University of Murcia, Spain.
Mónica Izquierdo Alonso
Department of Phylology, Communication and Documentation, University of Alcalá, Spain.
Antonio Maurandi López
Statistical Support Office, University of Murcia, Spain.
Javier Vallés Valenzuela
Neurobiology Institute, UNAM, Mexico.

Abstract

Introduction. *Consistency in indexing of Literatura Latinoamericana y del Caribe en Ciencias de la Salud is analysed and some features are compared with those of MEDLINE, ISA and PsycInfo.*
Method. *The study was carried out with 194 duplicate entries chosen from 8,547 entries, which is analogous to the process followed in MEDLINE, ISA and PsycInfo. Analysis. Consistency was calculated using the Hooper and Rolling formulas.*
Analysis. *Consistency was calculated using the Hooper and Rolling formulas. The Mann-Whitney U Test and the Student T were used to study the relationships between different variables.*
Results. *Consistency increased on separating out the quailifiers for secondary terms. The quailifiers are used more consistently than secondary terms. Secondary terms are used with very low frequency in the description of documents. In general, the more terms used to describe the documents, the lower the level of consistency.*
Conclusions. *Indexing consistency in LILAC is shown to be substantially less than that offered by MEDLINE, ISA and PsycINFO..*

Introduction

Consistency or coherence among indexers, i.e. the degree of agreement or concordance between two or more indexers when choosing the terms which represent the informative content of the documents is, perhaps, one of the most controversial of the elements concurring in indexing (along with correction, exhaustiveness, specificity). This is due to the existence of a number of discrepancies amongst researchers in the subject (Lancaster 1991, 1998, 2003). The main argument hinges on the thesis that consistency may be an indicator of quality in indexing (Cooper 1969, Zunde-Dexter 1969, Rolling 1981, Fugmann 1985, Lancaster 1991, Soergel 1994, White-Griffith 1987, Braam&Bruil 1992). The majority accept this link between coherence and quality of indexing and the link between the latter and Information Retrieval Systems efficiency (Moreno Fernández 2003). These authors are not a stand-alone group, but rather ones who have produced works of varying reach while, in general, accepting the thesis of the positive correlation between coherence and search outcome.

Consistency in indexing can be studied from different angles. Studies can be divided into two large categories, which simplifies the issue:

1. Those studies that deal with consistency using the terms chosen by the indexers to represent the documents; e.g. by comparing the indexing terms used to describe the same entries in different databases (Tonta 1991, Lind Blackwell 1994) and contrasting the indexing terms with respect to a *model* or previously established *gold standard* (Uren 2000). Another recent piece of research (Soler and Gil-Leiva 2011) looks at the relation between the type of indexing language (list of quailifiers, augmented thesaurus and the standard thesaurus) and consistency.

2. Those studies that focus on the choice of terms and concepts people use when retrieving information in different environments (Iivonen 1995). This perspective has been broadened in

an attempt to build a model that considers the choice of search terms used as the navigation of different discourses ([Iivonen & Sonnenwald 1998](#)).

The first group comprises researchers who choose to use existing duplicate entries in the same database ([Funk and Reid 1983](#) for MEDLINE, [Sievert amd Andrews 1991](#) for Information Science, [Leininger 2000](#) for PsycINFO). We consider this approach to be very suitable for analysing a hypothetical quality of indexing, since we presuppose that within a same context and with the same tools, the indexers would coincide in the characteristic description of the document. This procedure has the added advantage that it reveals the indexing of documents in a real environment within an information system, i.e. it goes beyond the constraints imposed by an experiment or test.

Given the complex nature of indexing consistency, we do not limit ourselves to merely calculating its levels. In order to gain further knowledge of its fundamentals, we study the relationship between consistency and the number of terms used in the description of the document. We also investigate some of the factors which might modify or in someway affect it, e.g. the language the indexed document is written in, the subject matter, whether or not it included an abstract, and the type of document analysed. Funk and Reid ([1983](#)) considered the depth (exhaustiveness) of indexing, the priority or preference assigned to indexed journals, the language in which the document was written, its length and the thematic areas of high and low consistency. Lancaster ([1991](#)) included even more factors affecting consitency. We have analysed the effect of consistency of those appearing in the entries.

This entanglement of interdependencies enables us to identify the causes of a higher or lower coefficient for consistency which, in short, influence the quality of the indexing and the efficacy in retrieving information through subjects or topics. Since some of the methodology used here is analogous to that in studies of MEDLINE, ISA and PsycInfo, we are in a position to compare figures and percentages and to draw conclusions, something which, unfortunately, is not always viable in studies on consistency. From data of international character, LILAC (Literatura Latinoamericana y del Caribe en Ciencias de la Salud), it is possible to use duplicate entries in our study, thanks to the way in which the documents making up this database are catalogued and indexed. There are three more reasons in support of the work performed: (a) there are no studies to date on consistency in this important database, for which the demand is constantly increasing; (b) given LILACs' growing importance and use, it is necessary to explain why the consistency rate of the indexing is lower than that of other databases, and c) this will identify which factor or factors are having a negative influence and enable inconsistencies to be solved and so ensure greater accuracy in the retrieval of information.

LILAC is a cooperative database on Health Sciences in which Latin-American and Caribbean documentation centres collaborate. It gathers literature published since 1982 in Latin-American countries by Latin-American writers in the field of health sciences which is not included in other international databases. Twenty-seven Latin-American and Caribbean countries participate in the project, which connects almost 600 libraries. The database describes and indexes books, theses, communications and talks at congresses and conferences, scientific and technical reports, government publications and articles from journals, from some 860 publications in this field, including e-journals ([Jiménez Miranda 1998](#)). BIREME-LILACS is a collective regional effort. BIREME is a centre of the Pan American Health Organization (PAHO) which, in turn, is an office of the World Health Organization (WHO), and it serves the American continent. The name BIREME was changed to Latin American and Caribbean Centre on Health Sciences Information, which better reflects its aims and functions. However, the old acronym is still widely used. BIREME trains staff in different countries in the use of work tools like the DeSC thesaurus. The national coordinating centers (NCC) use cooperating centers to provide LILAC database with their own information resources. Cooperative Centers selects the literature of their respectives countries. In order to coordinate the selection procedure they uses the Guide to select documents of the LILAC database (Guía para seleccionar documentos). They aren't publishers, but documentation centers. The components of the system also participate in policies and technical management through national advisory committees and NCC representatives CCN in the technical group of the Latin American network, which guides BIREME as to the modifications to be introduced in the methodology and, in particular, in the ongoing updating of vocabulary (DeCS) and its adaptation to the semantic characteristics of each country and to specific subjects..." ([Armenteros Vera 2002](#)).

Methodology

Duplicate entries should not exist in databases, but it is a fact that they do and although it is not the aim of this study to analyse the reason for this phenomenon, it seems clear that the cause stems from the cooperative nature of the LILACS project, in which different entities possess and describe the same

document, which may even be analysed by different people in the same centre ([BIREME 2006](#)). In order to locate duplicate entries in LILAC for various subjects, we started the document search using the toponym *Mexico*, which figures as a descriptor in a wealth of documents. We used Mexico instead of any other word to select information sources recovered in the field of health sciences. We thought that the number of documents obtained allows at least to establish a tendency in the consistency between indexers in LILAC. The search was made under terms, since if we had carried it out under country or year of publication, the number of documents returned would have exceeded 27,000. We obtained 8,547 entries, of which 194 were duplicated. Subsequently, we sifted through the documents carefully, ordering them alphabetically, and found the 97 pairs

Other authors have acted similarly. Funk and Reid ([1983](#)) detected 760 articles indexed twice in Medline; Sievert and Andrews ([1991](#)) worked with 496 entries of the database Information Science Abstracts and, more recently, Leininger ([2000](#)) in PsycINFO located 60 pairs of entries, i.e, 120 documents in total. The percentage of duplicate entries with respect to the total obtained is 2.32. These 194 entries do not, strictly speaking, constitute a sample of the whole database because they have not been taken as a sample, rather we use all the ones we have found which are repeated (the whole population available, as sociologists would say). The data managed are sufficient for significant results to be obtained, or at least a trend, as happens in the studies cited above.

We calculate the percentage of duplicate entries using the formula used by Sievert and Andrews ([1991](#)) and which can be expressed as follows: ***Percentage of duplicate entries = 100 T / (N-T)***

The coherence or consistency of the indexing is a measure of concordance. Several formulas have been proposed to measure coherence. Perhaps the most common is the now classic one proposed by Hooper, expressed simply as the ratio AB/(A+B), although some authors have subsequently opted to express it mathematically as ***C = 100 N / (A + B ? N)***

In the Hooper formula it is implicit that choosing the term twice is a necessary condition for coherence to be attained. To compensate for what he considers low weighting, Rolling ([1981](#)) included a modification in the Hooper formula: C = 2c / A + B, where 2c represents the number of terms in which there is agreement, multiplied by two. A + B represents the total number of terms assigned by both indexers.

Table 1 presents the overall data for consistency, obtained by applying the formulas of Hooper and Rolling. Therefore, do not include data on indexing terms considered as principal or as secondary within the field of quailifiers, as are, respectively, the quailifiers themselves and the above-mentioned qualifiers. Qualifers are subordinate (are secondary terms) to the descriptor. Nor are descriptor and qualifier considered a single term; in principle, we consider them independent. Hence, both types of term are treated separately in each pair of entries. The reason for this lies in the fact that the second word ('qualifier') does not form part of the descriptor, unless the indexer chooses to link them with a forward slash (/), once it has been chosen from a restricted set of possible qualifiers which can be used. If we took descriptor and qualifier as a single expression, the level of consistency would be reduced to a greater extent.

Tables 2, 3, 4 and 5, in contrast, present the levels of consistency of the principal and secondary terms. Within the second block -Tables 2,3,4 and 5- we further distinguish between qualifiers and limits, which do not appear in table 1 because these form a different field from that made up by the quailifiers. This enables us to compare the consistency of the overall indexing with the consistency obtained from detailing the same in principal terms, or quailifiers, and secondary terms, in which qualifiers and limits are included, respectively. This is what Funk and Reid ([1983](#)) did in Medline, as did Sievert and Andrews ([1991](#)) in Information Science Abstracts, when they differentiated between Main Headings, Subheadings, and so on. Next, we investigate the nature of the consistency by analysing the variables which influence it. Many factors can condition the degree of concordance between indexers. Some of these were detailed by Lancaster ([1991](#), [1998](#), [2003](#)): the number of terms assigned to a document by the indexers; the use of controlled or natural language (*free text indexing*); the size and specificity of the documental language used; the characteristics of the matter analysed and its terminology; the factors which can affect the indexer; the tools the indexer has available and the length of the document analysed. Funk & Reid ([1983](#)) analyze some of these in MEDLINE.

We try here to show how other new factors affect consistency and, moreover, to consider them as interrelated. Thus, the determinants analysed whose correlation with consistency we wish to establish are not all of those referred to by Lancaster, but those found in the entries retrieved and which had not been studied before, although some do coincide, as is the case of the subject or topic described, or the number of terms used. In any case, it would not make sense to include other aspects mentioned in

previous studies, such as the size of the language of the document, its degree of specificity, control, etc., because in LILACs the same indexing tool is always used. The variables we have considered as possibly having an effect on consistency are exhaustiveness of the indexing (absolute total of quailifiers and absolute total of secondary terms), the language of the document, the type of publication (article or monograph or report), author of the indexing (entity responsible), the subject or topic described, and whether the document includes an abstract or not. We calculate the Pearson between-variables linear correlation coefficient, the Mann-Whitney U test, the student T test and the Spearman Rho non parametric test to study the relations between the different variables of the study, depending on the different conditions of normality and homoscedasticity presented by the samples (applying the Kolmogorov–Smirnov test to check for normality and the Levene test for homogeneity of variances). In one case it was possible to apply a one-way ANOVA.

The quantitative variables estimated to observe the hypothetical relationship between exhaustiveness of indexing and consistency of indexing (tables 7, 8, 9, 10, 11, 12 and 13) are: Total number of quailifiers; Hooper consistency quailifiers; Rolling consistency quailifiers; Total of secondary terms. The category variables we analyse (tables 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29) for relations or to see if they affect consistency are: Language; Type of public ation; Entity (entity responsible for the indexing); Subject (classified according to the micro-disciplines which make up the thesaurus, such as Anatomy, Organisms, Illnesses, etc.); and Abstract (whether the document indexed includes an abstract or not). These five category variables were selected as factors after analyzing their possible influence on the marginal distributions of the continuous variables. The Kruskal-Wallis test has been applied to category variable and type of publication in order to verify hypotheses on differences of variables with groups.

Results and Discussion

In general, indexing consistency in LILAC is low compared to MEDLINE, ISA and PsycINFO. Taking principal and secondary terms together, the mean for the consistency in ISA is 48.12%, using the Hooper formula; and it is similar, although slightly higher, for PsycINFO at 50.40% (we do not have the mean for MEDLINE). In PsychINFO, the Rolling coefficient pushes the figure up to 60.83%. Consistency levels in LILAC are considerably reduced, and they would be even lower had we not included the Mexico toponym, which is included in almost all the documents, in the count. In Table 1 we observe that when the total number of terms used in indexing each document is considered (i.e. main terms and secondary ones), the mean consistency is no higher than 31.57% (Hooper), and reaches a maximum of 43.20% (Rolling). Consistency is higher in this case because we multiply the number of terms common to both indexers by 2, in order to give them more specific weight than the discordants. However, it is the median that gives the most precise information of how reduced the consistency is: 23% for Hooper and 38% for Rolling. The typical coherency deviation is very similar for both types of calculation: 25.23% for the first and 25.90% for the second. The most repeated coherency percentage, the mode, is 50% (Hooper) and 67% (Rolling).

	Total number of terms	Hooper consistency in total number of terms	Rolling consistency in total number of terms
No. valid	97	97	97
No. missing	0	0	0
Mean	12.28	0.3157	0.4320
Median	10.00	0.2300	0.3800
Mode	9	0.50	0.67
Std. Desviation	6.875	0.25234	0.25908

Table 1: Indexing consistency: total number of terms (main terms or quailifiers and secondary ones)

The total number of terms across the 194 entries is 1.191. The mean number of terms per document is 12.28 (median = 10); typical deviation stands at 6.87 and the most frequent number of terms assigned to a document is 9 quailifiers. Each document therefore has a high number of terms assigned. This means a high level of exhaustiveness and that the informational content of the documents is widely described. Tables 2 and 3 give more details of the characteristics of the indexing in LILAC. They highlight the contrasts between the nature of the indexing terms and the use made of these to describe documents. The main terms or quailifiers are distinguished from the secondary ones, which are the qualifiers and the limits.

	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
No. valid	97	97	97
No. missing	0	0	0
Mean	9.55	0.3664	0.4856
Median	9.00	0.3300	0.5000
Mode	9	0.50	0.67
Std. Desviation	3.614	0.26879	0.26610

Table 2: Indexing consitency referred only to quailifiers, secondary terms (qualifiers or limits) not included

The mean coherence of the quailifiers now improves slightly and it is greater than the total number of terms, since the median increases to 33% (Hooper) and 50% (Rolling). Although the typical deviation reveals more favourable consistency percentages (26.87% and 26.61% respectively), we can see that this moves further away from that of MEDLINE (61.10%), ISA (52.25%) and PsycINFO (43.24% with the Hooper procedure and 56.09% with Rolling). The total number of qualifiers assigned to the documents is 926, and the mean of expressions per document does not reach 9.55. The mode is slightly lower (9 terms) and the standard deviation is 3.61, lower than when main and secondary terms are not differentiated. In any case, the number of qualifiers used in the indexing is much higher than that of secondary terms (i.e., qualifiers and limits. BIREME 2005b). The secondary terms (qualifiers and limits) make up 22.25% of the total number of terms, i.e., 265 of the 1.191. The mean number of terms per document is 2.73 (median = 1.00). Typical deviation is 4.64 and the number of secondary terms most frequently used to describe a document is 0. Very few secondary terms are therefore used in document indexing. However, of the 265 secondary terms found, 157 (59.24%) are qualifiers and the remaining 108 (40.75%) are limits (see Tables 4 and 5). There is a certain equilibrium between the two classes, although qualifiers are more frequent. The mean number of qualifiers per document is 1.62, while that of limits stands at 1.11. It should be remembered that the mean number of qualifiers or main terms per document is 9.55.

In Tables 3, 4 and 5 we observe that the degree of consistency is much higher than that of the total number of terms and that of the main terms or quailifiers. The consistency median is 45% (Hooper) and 62% (Rolling); the typical deviation of these variables (Table 3) is in both cases 0.48, which is very low. The consistency of the secondary terms is somewhat lower than that of MEDLINE (54.90%) but almost the same as that of ISA (45.54%).

	Total number of secondary terms	Hooper consistency in secondary terms	Rolling consistency in secondary terms
No. valid	97	97	97
No. missing	0	0	0
Mean	2.73	0.4962	0.5076
Median	1.00	0.4500	0.6200
Mode	0	1.00	1.00

Table 3: Indexing consitency referred only to secondary terms (includes only qualifiers and limits)

	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
No. valid	97	97	97
No. missing	0	0	0
Mean	1.62	0.5833	0.59485
Median	0.00	1.0000	1.0000
Mode	0	1.00	1.000
Std. Desviation	3.101	0.47397	0.477106

Table 4: Indexing consistency referred only to qualifiers

	Total number of limits	Hooper consistency in limits	Rolling consistency in limits
No. valid	97	97	97
No. missing	0	0	0
Mean	1.11	0.7633	0.7672
Median	0.00	1.0000	1.0000
Mode	0	1.00	1.000
Std. Desviation	2.618	0.41956	0.41919

Table 5: Indexing consistency referred only to limits

From the data in Tables 4 and 5 it could be wrongly inferred that the resulting degree of consistency is much greater in the secondary terms than in the main ones or the quailifiers. This is because the calculations given for the secondary terms include those documents for which the indexers are either in agreement or coincide in not assigning indexing terms. In other words, we have chosen to consider as total coincidence the circumstance in which two or more indexers opt not to use any secondary term to describe the document. Perhaps this is what occurs in MEDLINE or ISA, although there is no indication of this. What these figures really reveal is that the number of secondary terms used in indexing is very low. This is shown by biasing the data we have just presented (see Table 6). That is, once the entries for which there are no secondary terms have been removed, and if we do not include the tacit agreement not to describe these documents in those terms, the results change significantly. There is a notable decrease in the consistency of the indexing with both Hooper and Rolling (9.5% and 11.5% respectively (median = 0)). The standard deviation for the total number of secondary terms is high at 5.3. When we break down the secondary terms into qualifiers and limits, we have that the mean number of qualifiers used in documents is 3.49, and the consistency continues to be very low: 11% (Hooper) and 13% (Rolling).

Limits are used somewhat more (4.32 per document), but the consistency is even lower: 8% (Hooper) and 9% (Rolling). The standard deviation is relatively high and very similar in the total number of qualifiers (3.7) and limits (3.6). However, the standard deviation is higher in the consistency which affects the qualifiers (Hooper (0.25) and Rolling (0.29)) than that affecting the limits -0.22 for Hooper and 0.26 for Hooper (0.25) and Rolling (0.29).

Factors	No. valid	No. missing	Mean	Median	Mode	Mimimun	Maximun	Sum
Total number of secondary terms	54	43	4.91	3.00	1	1	29	265
Hooper consistency in secondary terms	54	43	0.0950	0.0000	0.00	0.00	1.00	5.13
Rolling consistency in secondary terms	54	43	0.1156	0.0000	0.00	0.00	1.00	6.24
Total number of qualifiers	45	52	3.49	2.00	1	1	18	157
Hooper consistency in qualifiers	45	52	0.1111	0.0000	0.00	0.00	1.00	5.00
Rolling consistency in qualifiers	45	52	0.13444	0.00000	0.00	0.000	1.000	6.050
Total number of limits	25	72	4.32	3.00	1	1	11	108
Hooper consistency in limits	25	72	0.0816	0.0000	0.00	0.00	0.80	2.04
Rolling consistency in limits	25	72	0.0968	0.0000	0.00	0.00	0.89	2.42

Table 6: Indexing consistency referred to secondary terms with biased data

This allows the following hypothesis to be inferred: 'When it is decided to index with secondary terms, the consistency is notably reduced, but is greater than that of the quailifiers if we consider the circumstances in which the indexers are in agreement in not assigning secondary terms. When secondary terms are used or when indexers are in agreement in not assigning secondary terms, the consistency of indexing is notably reduced, but is still greater than when quailifiers are used". We will now look at the possible relation between the exhaustiveness of the indexing (number of indexing terms) and the degree of consistency.

Number of terms (exhaustiveness) and consistency

The Pearson correlation coefficient, which is denoted by r and the significance, commonly called p-value, of table 7 shows a significant negative relation between the *Total number of terms* and the *Hooper consistency quailifiers* , r = -0.272, p(bilateral) <0.05; and between the *Total number of terms*, and the *Consistency Rolling quailifiers* , r = -0.280, p(bilateral) <0.05.

	Pearson correlation coefficient	Total number of terms	Hooper consistency in total number of terms	Rolling consistency in total number of terms
Total number of terms (n=97)	*r*	1	-0.272 (**)	-0.280 (**)
	Sig. (2-tailed)		0.007	0.005
Hooper consistency in total number of terms	*r*	-0.272 (**)	1	0.980
	Sig. (2-tailed)	0.007		0.0000
Rolling consistency in total number of terms	*r*	-0.280 (**)	0.980 (**)	1
	Sig. (2-tailed)	0.005	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 7: Pearson correlation coefficient (*r*) between total terms and Hooper consistency and Rolling consistency

	Pearson correlation coefficient	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
Total number of quailifiers (n=97)	*r*	1	-0.391 (**)	-0.405 (**)
	Sig. (2-tailed)		0.000	0.005
Hooper consistency in quailifiers	*r*	-0.391 (**)	1	0.978
	Sig. (2-tailed)	0.000		0.0000
Rolling consistency in quailifiers	*r*	-0.405 (**)	0.978 (**)	1
	Sig. (2-tailed)	0.000	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 8: Pearson correlation coefficient *r* between number of quailifiers and Hooper and Rolling consistency in quailifiers

Tables 7 and 8 show that there is a weak negative correlation between the number of terms (Total number of terms) and the consistency coefficients given by the Hooper formula and the Rolling formula. If we break down the data, we observe, Table 7, that the negative correlation (Pearson Coefficient) between the total number of terms and the values for Hooper and Rolling are, respectively, -0.272 and -0.280. In Table 7 the estimated negative correlation (Pearson Coefficient) for the Hooper value is -0.391, and -0.405 for the Rolling values; that is, the negative correlation is stronger or greater. In fact, both Tables serve to corroborate the truth of the generally accepted hypothesis which states that the higher the number of terms chosen to describe a document, the lower its consistency. In both these tables, the confidence level is higher than 99%, which indicates a significant correlation of 0.01. There is, moreover, a positive correlation between the Hooper consistency variables and those of Rolling.

	Pearson correlation coefficient	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
Total number of secondary terms	*r*	1	-0.417 (**)	-0.372(**)
	Sig. (2-tailed)		0.000	0.000
Hooper consistency in secondary terms	*r*	-0.417 (**)	1	0.997
	Sig. (2-tailed)	0.000		0.000
Rolling consistency in secondary terms	*r*	-0.372(**)	0.997 (**)	1
	Sig. (2-tailed)	0.000	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 9: Pearson correlation coefficient *r* between total number of secondary terms and Hooper and Rolling consistency in secondary terms

	Pearson correlation coefficient	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
Total number of qualifiers (n = 97)	*r*	1	-0.417 (**)	-0.421 (**)
	Sig. (2-tailed)		0.000	0.000
Hooper consistency in qualifiers	*r*	-0.464 (**)	1	0.997 (**)
	Sig. (2-tailed)	0.000		0.000
Rolling consistency in qualifiers	*r*	-0.421 (**)	0.997 (**)	1
	Sig. (2-tailed)	0.000	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 10: Pearson correlation coefficient *r* between total number of secondary terms and Hooper and Rolling consistency in secondary terms

	Pearson correlation coefficient	Total number of quailifiers	Hooper consistency in quailifiers	Rolling consistency in quailifiers
Total number of limits (n=97)	*r*	1	-0.622 (**)	-0.596 (**)
	Sig. (2-tailed)		0.000	0.000
Hooper consistency in limits	*r*	-0.622 (**)	1	0.999 (**)
	Sig. (2-tailed)	0.000		0.000
Rolling consistency in limits	*r*	-0.596 (**)	0.999 (**)	1
	Sig. (2-tailed)	0.000	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 11: Pearson correlation coefficient (*r*) between limits and Hooper and Rolling consistency limits

As regards the secondary terms when considered together (qualifiers and limits), they all show a higher negative correlation than that of the descriptors, with a confidence level of more than 99% (Table 9). Individually, Tables 10 and 11 show that the Pearson Coefficient between the qualifiers and the Hooper consistency values is -0.464 and -0.421 for Rolling. For the limits, the Pearson Coefficient is -0.622 for Hooper and -0.596 for Rolling. This means that in the case of the qualifiers a mean negative correlation is not reached. In the case of the limits, however, we do have a mean negative correlation. The important finding here is that the correlations between the number of terms selected for indexing documents and the indexing consistency show that the use of many indexing terms causes the level of

consistency to fall. Yet consistency is greater in the descriptors than in the secondary terms, which supports the general hypothesis, which states that indexers show a greater degree of disagreement, i.e. less consistency when representing the secondary aspects of the documents than the main ones. This is because consistency of indexing is lower when secondary terms are used.

However, if we look at the *biased* data on correlations between the use of secondary terms and the degree of consistency (table 12), i.e. again removing those entries in which indexers do not assign any indexing term, the correlation between secondary terms and consistency is now highly positive and significant for Hooper ($r = 0.306$) and Rolling ($r = 0.987$)

	Pearson correlation coefficient	Total number of secondary terms (n=54)	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Total number of secondary terms	*r*	1	0.237	0.306 (**)
	Sig. (2-tailed)		0.085	0.025
Hooper consistency in secondary terms	*r*	0.237	1	0.987
	Sig. (2-tailed)	0.085		0.000
Rolling consistency in secondary terms	*r*	0.306 (*)	0.987 (**)	1
	Sig. (2-tailed)	0.025	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

Table 12: Pearson correlation coefficient (*r*) with biased data between the secondary terms and consistency between Hooper and Rollin

The Spearman Rho coefficient (***P***) has also been calculated for these data, and the correlation between the Hooper consistency and the total number of terms is significant (Rho = 0.355*).

	Spearman's Rho	Total number of secondary terms	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Total number of secondary terms (n=54)	**P**	1	0.355 (**)	0.355 (**)
	Sig. (2-tailed)		0.009	0.009
Hooper consistency in secondary terms	**P**	0.355 (**)	1.000	1.000 (**)
	Sig. (2-tailed)	0.009		0.000
Rolling consistency in secondary terms	**P**	0.355 (**)	1.000 (**)	1.000
	Sig. (2-tailed)	0.009		

** Correlation is significant at the 0.01 level (2-tailed).

Table 13: Non parametric correlation coefficient using Spearman's Rho (P) with biased data between the secondary terms and consistency between Hooper and Rolling

This allows a complementary hypothesis to be added to the one above: in the few cases that secondary terms are used, there is a certain degree of moderate or low correlation with consistency, which means that when qualifiers and limits are used, this is not done inconsistently. This may be because the qualifiers and limits are related only to specific descriptors in the wider context of the language of the document, and are either used with these or are not used. They therefore make up a small subset of terms whose use leaves little scope for subjectivity on the part of the indexer, other than that of deciding whether or not to use them to describe documents. Below, we present the results for the remaining variables studied in terms of interrelations and we offer their median as a measure of central trend, since this is better than the mean in non parametric tests.

The Language variable

If we look at the row Exact Sig (1-tailed) we observe that there is a difference in means between the two groups of languages for the variables Hooper consistency and Rolling consistency for the total number of terms, and also for the descriptors. For the remaining variables, these cannot be said to be different.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Total number of terms	Total number of descriptors	Total number of secondary terms
Mann-Whitney U	321.000	323.500	341.500	341.500	540.500	557.500	536.000
Wilcoxon W	3807.000	3809.500	3827.500	3827.500	645.500	662.500	641.000
Z	-2.672	-2.646	-2.465	-2.465	-0.417	-0.243	-0.485
Asymp. sig. (2-tailed)	0.008	0.008	0.014	0.014	0.676	0.808	0.628
Exact Sig. (2-tailed)	**0.007**	**0.007**	**0.013**	**0.013**	**0.682**	**0.812**	**0.634**
Exact sig. (1-tailed)	0.003	0.004	0.006	0.006	0.341	0.406	0.320
Point probability	0.000	0.000	0.000	0.000	0.002	0.002	0.002

Table 14: Test statistics (grouping variable: Spanish and other languages)

Measure	Other languages	Spanish	Total
Hooper consistency in total number of terms	0.5000	0.20000	0.2300
Rolling consistency in total number of terms	0.6700	0.3300	0.3800
Hooper consistency in descriptors	0.5850	0.2900	0.3300
Rolling consistency in descriptors	0.7350	0.4400	0.5000
Hooper consistency in secondary terms	1.0000	0.2500	0.4500
Rolling consistency in secondary terms	1.0000	0.4000	0.6200
Total number of terms	10.50	10.00	10.00
Total number of descriptors	8.50	9.00	9.00
Total number of secondary terms	0.50	1.00	1.00

Table 15: Central trend measures: medians (Grouping variable: language)

The mean is lower for the Spanish in all the variables. The variable Hooper consistency in total number of terms differs in mean for Other languages, Median = 0.5 than for Spanish, Median = 0.2, U= 321, significant, p= 0.003<0.05. The variable Rolling consistency in total number of terms differs in mean to a greater degree for Other languages, Median = 0.6 than for Spanish, Median = 0.3, U= 323, significant, p= 0.004<0.05. The variable Hooper consistency in descriptors differs in mean for *Other languages*, Median = 0.5 than for Spanish, Median = 0.2, U= 341, significant, p= 0.006<0.05. The variable Rolling consistency in descriptors differs in mean for *Other languages*, Median = 0.7 than for Spanish, Median = 0.4, U= 341, significant, p= 0.006<0.05. This enables us to put forward a hypothesis stating that documents indexed in Spanish show a lesser degree of consistency than those described in other languages (Portuguese and English).

The *entity responsible* variable

If we look at the row Exact Sig (1-tailed) we observe that there is a difference in means between the two groups of those responsible (Same country, Different countries) for the variables Hooper consistency and Rolling consistency (for the total number of terms, as well as for the descriptors. For the remaining variables, these cannot be said to be different.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Total number of terms	Total number of descriptors	Total number of secondary terms
Mann-Whitney U	396.000	394.000	559.000	559.900	818.500	1027.000	668.000
Wilcoxon W	2476.000	2639.000	2639.000	2639.000	1379.500	3107.000	1229.000
Z	-5.031	-3.795	-3.795	-3.795	-1.815	-0.223	-3.102
Asymp. sig. (2-tailed)	0.000	0.0000	0.000	0.000	0.070	0.824	0.002
Exact sig. (2-tailed)	**0.000**	**0.000**	**0.000**	**0.000**	**0.070**	**0.826**	**0.002**
Exact sig. (1-tailed)	0.000	0.000	0.000	0.000	0.035	0.413	0.001
Point probability	0.000	0.000	0.000	0.000	0.000	0.001	0.000

Table 16: Test statistics

Measure	Entity responsible		
	Same country	Different countries	Total
Hooper consistency in total number of terms	0.5000	0.1700	0.2300
Rolling consistency in total number of terms	0.6700	0.2900	0.3800
Hooper consistency in descriptors	0.5000	0.2500	0.3300
Rolling consistency in descriptors	0.6700	0.4000	0.5000
Hooper consistency in secondary terms	1.0000	0.0000	0.4500
Rolling consistency in secondary terms	1.0000	0.0000	0.6200
Total number of terms	9.00	11.00	10.00
Total number of descriptors	9.00	9.00	9.00
Total number of secondary terms	0.00	1.50	1.00

Table 17: Central trend measures: medians

The mean is lower for Different countries in all the variables. The variable *Hooper consisteny total number of terms* differs in mean for *Same country*, Median= 0.5 than for *Different countries*, Median= 0.1, U= 396, significant, p= 0.0<0.05. The variable *Rolling consistency total number of terms* differ in mean to a greater degree for *Same country*, Median= 0.6 than for *Different countries*, Median= 0.2, U= 394, significant, p= 0.0<0.05. The variable *Hooper consistency descriptors* differs in mean for *Same country*, Median.= 0.5 than for *Different countries*, Med.= 0.2, U= 559, significant, p= 0.0<0.05. The variable *Rolling consistency descriptors* differs in mean for *Same country*, Median= 0.6 than for *Different countries*, Median= 0.4, U= 559, significant, p= 0.0<0.05 This allows the hypothesis to be put forward which states that consistency of indexing increases if the documents are indexed in the same country, while it decreases when the document is analysed in different countries. If we link this variable to language, one interpretation could be that indexers of different languages who describe the document are less consistent, even though they use the same documental language, in this case trilingual.

The *subject* variable

If we look at the row Exact Sig (1-tailed) we observe that there is only a difference in means between the two groups (Other subjects and Public health) for the Hooper consistency and the Rolling consistency variables for the total number of terms. For the remaining variables, we cannot state that these are different.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Total number of terms	Total number of descriptors	Total number of secondary terms
Mann-Whitney U	946.000	948.500	945.500	946.000	759.000	971.000	700.000
Wilcoxon W	3091.000	3093.500	3090.500	3091.000	2904.000	3116.000	2845.000
Z	-0.722	-0.703	-0.727	-0.723	-2.164	-0.534	-2.739
Asymp. sig. (2-tailed)	0.470	0.482	0.467	0.470	0.030	0.593	0.006
Exact Sig. (2-tailed)	0.473	0.485	0.470	0.473	0.030	0.596	0.006
Exact sig. (1-tailed)	0.237	0.243	0.235	0.236	0.015	0.298	0.003
Point probability	0.001	0.001	0.001	0.001	0.000	0.001	0.000

Table 18: Test statistics (grouping variable: subject)

Measure	Subject		
	Other subjects	Public health	Total
Hooper consistency in total number of terms	0.2400	0.2100	0.2300
Rolling consistency in total number of terms	0.3900	0.3500	0.3800
Hooper consistency in descriptors	0.3300	0.3300	0.3300
Rolling consistency in descriptors	0.5000	0.5000	0.5000
Hooper consistency in secondary terms	0.1300	1.0000	0.4500
Rolling consistency in secondary terms	0.2200	1.0000	0.6200
Total number of terms	12.00	10.00	10.00
Total number of descriptors	9.00	9.00	9.00
Total number of secondary terms	2.00	1.00	1.00

Table 19: Central trend measures: medians (Grouping variable: subject)

The variable *Hooper consistency total number of terms* differs in mean for *Other subjects*, Median= 12 than for *Public health*, Median= 10, U= 759, significant, p= 0.01<0.05. The variable *Rolling consistency total number of secondary terms* differs in mean for *Other subjects*, Median= 2.0 than for *Public health*, Median= 1.0, U= 700, significant, p= 0.003<0.05. This allows the hypothesis to be put forward which states that the subject of the documents does not condition or influence indexing consistency.

The *abstract* variable

If we look at the row Exact sig. (1-tailed) we observe that there is only a difference in means between the two groups (Do have an abstract and Do *not* have an abstract) for the Hooper consistency and Rolling consistency variables for the total number of terms, as well as for the total number of descriptors. The remaining variables cannot be said to be different.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Total number of terms	Total number of descriptors	Total number of secondary terms
Mann-Whitney U	463.000	462.000	366.000	366.000	776.500	627.500	672.500
Wilcoxon W	3389.000	3398.000	3292.000	3292.000	1007.500	858.500	3598.500
Z	2.938	-2.943	-3.794	-3.794	-0.189	-1.507	-1.154
Asymp. sig. (2-tailed)	0.003	0.003	0.000	0.470	0.850	0.132	0.248
Exact sig. (2-tailed)	0.003	0.003	0.000	0.473	0.853	0.133	0.251
Exact sig. (1-tailed)	0.001	0.001	0.000	0.236	0.427	0.067	0.126
Point probability	0.000	0.000	0.000	0.001	0.002	0.001	0.001

Table 20: Test statistics (grouping variable: abstract)

The variable *Hooper consistency total number of terms* differs in mean for *documents that Do have an abstract*, Median.= 0.3 than for *documents that Do Not have an abstract*, Median= 0.1, U= 463, significant, p= 0.001<0.05. The variable *Rolling consistency total number of terms* differs in mean for *documents that Do have an abstract*, Median= 0.5 than for *documents that Do Not have an abstract*, Median= 0.3, U= 462, significant, p= 0.001<0.05.

Measure	Abstract		
	Yes	No	Total
Hooper consistency in total number of terms	0.3300	0.1800	0.2300
Rolling consistency in total number of terms	0.3300	0.3050	0.3800
Hooper consistency in descriptors	0.5000	0.2600	0.3300
Rolling consistency in descriptors	0.5000	0.4100	0.5000
Hooper consistency in secondary terms	0.6700	0.5100	0.4500
Rolling consistency in secondary terms	0.1300	0.6750	0.6200
Total number of terms	0.2200	10.00	10.00
Total number of descriptors	9.00	9.00	9.00
Total number of secondary terms	2.00	1.00	1.00

Table 21: Central trend measures: medians

The variable *Hooper consistency descriptors* differs in mean for *documents that Do have an abstract*, Median= 0.5 than for *documents that Do Not have an abstract*, Median. = 0.2, U= 366, significant, p=0.00<0.05. The variable *Rolling consistency descriptors* differs in mean for *documents that Do have an abstract*, Median= 0.6 than for *documents that Do Not have an abstract*, Median= 0.4, U= 366, significant, p=0.00<0.05. This means we can put forward the hypothesis that those documents which include an abstract are indexed more consistently than those that do not.

The *type of publication* variable

We apply a Kruskasl-Wallis test to check the hypothesis on differences in variables with two groups. In this case, we have: Group 1= Proceedings or Anthology, Group 2 = Journal article, Group 3 = Report or Statistics Annual or Monograph, Group 4 = Anonymous work.

Measure	Proceedings or anthology	Journal Article	Report or Statistic Annual or Monograph	Anonymous work	Total number of secondary terms
Hooper consistency in total number of terms	0.6000	0.3200	0.1800	0.1500	0,2300
Rolling consistency in total number of terms	0.7500	0.4900	0.3100	0.2700	0.3800
Hooper consistency in descriptors	0.6000	0.4250	0.3300	0.2000	0.3300
Rolling consistency in descriptors	0.7500	0.6000	0.5000	0.3300	0.5000
Hooper consistency in secondary terms	1.0000	0.0000	1.0000	0.0000	0.4500
Rolling consistency in secondary terms	1.0000	0.0000	1.0000	0.0000	0.6200
Total number of terms	9.00	13.00	10.00	11.00	10.00
Total number of descriptors	9.00	9.00	8.00	9.00	9.00
Total number of secondary terms	0.00	2.50	1.00	1.00	1.00

Table 22: Central trend measures in the different types of publications

Measure	Chi-squared	df	Asymp. sig.	Exact sig.	Point probability
Hooper consistency in total number of terms	20.181	3	0.000		
Rolling consistency in total number of terms	20.372	3	0.000		
Hooper consistency in descriptors	17.206	3	0.001		
Rolling consistency in descriptors	17.206	3	0.001		
Hooper consistency in secondary terms	12.627	3	0.006	0.004	0.000
Rolling consistency in secondary terms	12.627	3	0.006	0.004	0.000
Total number of terms	2.667	3	0.446		
Total number of descriptors	3.789	3	0.285		
Total number of secondary terms	12.647	3	0.285		

(a) Kruskal Wallis Test (b) Grouping Variable: Type of publication (c) Some or all exact significances cannot be computed because there is insufficient memory.

Table 23: Kruskal-Wallis test: differences between the levels of the type of publication

According to the Kruskasl-Wallis test, there are significant differences between some levels of the variable for the Hooper consistency and the Rolling consistency variables of the total number of terms, the descriptors and the secondary terms, i.e. for the first six variables. In order to ascertain between which types of publication there are differences, we carried out *post hoc* tests.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	45.500	45.500	61.500	61.500	34.000	34.000
Wilcoxon W	181.500	181.500	197.500	197.500	170.000	170.000
Z	-2.577	-2.577	-1.889	-1.889	-3.335	-3.335
Asymp. sig. (2-tailed)	0.010	0.010	0.059	0.059	0.001	0.001
Exact sig. [2*(1-tailed Sig.)]	0.009(a)	0.009(a)	0.062(a)	0.062(a)	0.001(a)	0.001(a)
Exact Sig. (2-tailed)	0.009	0.009	0.060	0.060	0.001	0.001
Exact sig. (1-tailed)	0.004	0.004	0.030	0.030	0.000	0.000
Point probability	0.000	0.000	0.002	0.002	0.000	0.000

(a) Not corrected for ties. (b) Grouping Variable: Type of publication

Table 24: Test statistics for Group 1= Proceedings or anthologies and Group 2 = Journal article

To avoid an increase of type 1 error, we apply the Bonferroni correction (multiple comparisons) in the post-hoc tests. Thus, our level of significance has to be 0.0083, and not the usual 0.005. The results of the Mann-Whitney U test, when significance is corrected using the Bonferroni multiple contrasts fit method , allow significant differences to be appreciated between Proceedings or Anthology (Group 1) and Journal article (Group 2). In Tables 22 and 24 it is observed that the variable Hooper consistency in total number of terms differs in mean for Proceedings or Anthology, Median = 1.0 than for Journal article, Median = 0. U= 34 significant, p<0.0083. It is also observed that the variable Rolling consistency in secondary terms differs in mean for Proceedings or Anthology, Median = 1.0 than for Journal article, Median = 0. U= 34, significant, p<0.0083.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	110.500	110.500	133.500	133.500	185.500	185.500
Wilcoxon W	114.500	1145.500	1168.500	1168.500	1168.500	1168.500
Z	-3.399	-3.399	-2.972	-2.972	-2.302	-2.302
Asymp. sig. (2-tailed)	0.001	0.001	0.003	0.003	0.021	0.021
Exact sig. (2-tailed)	0.000	0.000	0.002	0.002	0.000	0.000
Exact sig. (1-tailed)	0.000	0.000	0.001	0.001	0.009	0.009
Point probability	0.000	0.000	0.000	0.000	0.002	0.002

Table 25: Test statistics for Group 1= Proceedings or anthologies and Group 3 = Report or Statistics annual or Monograph (Grouping variable: Type of publication)

The results of the Mann-Whitney U test, when significance is corrected using the Bonferroni multiple contrasts fit method, shows significant differences *Proceedings or Anthology* (Group 1) and *Report or Statistics Annual or Monograph* (Group 3). In Tables 22 and 25 it is observed that: the variable *Hooper consistency total number of terms* differs in mean more for *Proceedings or Anthology*, Median.= 0.6 than for *Report or Statistics Annual or Monograph*, Median= 0.1, U= 110 significant, p=0.0<0.0083

The variable *Rolling consistency total number of terms* differs in mean for *Proceedings or Anthology*, Median= 0.7 more than for *Report or Statistics Annual or Monograph*, Median= 0.3, U= 110 significant, p= 0.00<0.0083. The variable *Hooper consistency descriptors* differs in mean for *Proceedings or Anthology*, Median.= 0.6 more than for *Report or Statistics Annual or Monograph*, Median= 0.3, U= 133 significant, p= 0.001<0.0083. The variable *Rolling consistency descriptors* differs in mean for *Proceedings or Anthology*, Median.= 0.7 more than for *Report or Statistics Annual or Monograph*, Median= 0.5, U= 133 significant, p= 0.001<0.0083

The variable *Rolling consistency total number of terms* differs in mean for *Proceedings or Anthology*, Median= 0.7 more than for *Report or Statistics Annual or Monograph*, Median= 0.3, U= 110 significant, p= 0.00<0.0083. The variable *Hooper consistency descriptors* differs in mean for *Proceedings or Anthology*, Median.= 0.6 more than for *Report or Statistics Annual or Monograph*, Median= 0.3, U= 133 significant, p= 0.001<0.0083. The variable *Rolling consistency descriptors* differs in mean for *Proceedings or Anthology*, Median.= 0.7 more than for *Report or Statistics Annual or Monograph*, Median= 0.5, U= 133 significant, p= 0.001<0.0083

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	35.000	34.500	49.500	49.500	68.000	68.000
Wilcoxon W	311.000	310.500	325.500	325.500	344.000	344.000
Z	-3.779	-3.795	-3.306	-3.306	-3.307	-3.307
Asymp. sig. (2-tailed)	0.001	0.000	0.001	0.001	0.002	0.002
Exact Sig. [2*(1-tailed Sig.)]	0.000(a)	0.000(a)	0.001 (a)	0.001 (a)	0.006 (a)	0.006 (a)
Exact sig. (2-tailed)	0.000	0.000	0.001	0.001	0.002	0.002
Exact sig. (1-tailed)	0.000	0.000	0.000	0.000	0.001	0.001
Point probability	0.000	0.000	0.000	0.000	0.000	0.000

(a) Not corrected for ties.

Table 26: Test statistics for Group 1= Proceedings or anthologies and Group 4 = Anonymous work

The results of the Mann-Whitney U test, when significance is corrected using the Bonferroni multiple contrasts fit method, allows significant differences to be appreciated between *Proceedings or Anthology* (Group 1) and *Anonymous work* (Group 4). In Tables 22 and 26 it is observed that: the variable *Hooper consistency total number of terms* differs in mean for *Proceedings or Anthology*, Median= 0.6 more than for *Anonymous work*, Median= 0.1, U= 35 significant, p= 0.0<0.0083. The variable *Rolling consistency total number of terms* differs in mean for *Proceedings or Anthology*, Median= 0.7 more than for *Anonymous work*, Median= 0.2, U= 34 significant, p= 0.00<0.0083. The variable *Hooper consistency descriptors* differs in mean for *Proceedings or Anthology*, Median= 0.6 more than for *Anonymous work*, Median= 0.2, U= 49 significant, p= 0.00<0.0083. The variable *Rolling consistency descriptors* differs in mean for *Proceedings or Anthology*, Median= 0.7 more than for *Anonymous work*, Median= 0.3, U= 49 significant, p= 0.00<0.0083. The variable *Hooper consistency secondary terms* differs in mean for *Proceedings or Anthology*, Median= 1.0 more than for *Anonymous work*, Median= 0.0, U= 68 significant, p= 0.001<0.0083. The variable *Rolling consistency secondary terms* differs in mean for *Proceedings or Anthology*, Median= 1.0 more than for *Anonymous work*, Median= 0.0, U= 68 significant, p= 0.001<0.0083.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	266.500	264.000	263.500	263.500	273.500	273.500
Wilcoxon W	1301.500	1299.000	1298.500	1298.500	409.500	409.500
Z	-1.535	-1.576	-1.588	-1.588	-1.571	-1.571
Asymp. sig. (2-tailed)	0.125	0.115	0.112	0.112	0.116	0.116
Exact Sig. (2-tailed)	0.127	0.117	0.114	0.114	0.119	0.119
Exact Sig. (1-tailed)	0.063	0.058	0.057	0.057	0.065	0.065
Point probability	0.001	0.001	0.001	0.001	0.000	0.000

Table 27: Test statistics for Group 2= Journal article and Group 3= Report or Statistics Annual or Monograph

The results of the Mann-Whitney U test, when significance is corrected using the Bonferroni multiple contrasts fit method, indicate that there is no significant difference between Journal article (Group 2) and Report or Statistics Annual or Monograph (Group 3).

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	93.000	92.000	88.000	88.000	179.500	179.500
Wilcoxon W	369.000	368.000	364.000	364.000	315.500	315.500
Z	-2.603	-2.631	-2.573	-2.573	-0.146	-0.146
Asymp. sig. (2-tailed)	0.009	0.009	0.006	0.006	0.884	0.884
Exact sig. [2*(1-tailed sig.)]	0.009 (a)	0.008 (a)	s0.005 (a)	0.005 (a)	0.899 (a)	0.899 (a)
Exact sig. (2-tailed)	0.008	0.008	0.005	0.005	0.880	0.880
Exact sig. (1-tailed)	0.004	0.004	0.003	0.003	0.448	0.448
Point probability	0.000	0.000	0.000	0.000	0.007	0.007

(a) Not corrected for ties.

Table 28: Test statistics for Group 2= Journal articles group and Group 4 = Anonymous work

The results of the Mann-Whitney U test when significance is corrected using the Bonferroni multiple contrasts fit method, allows significant differences to be appreciated between *Journal article* (Group 2) and *Anonymous work* (Group 4). In Tables 22 and 28 it is observed that: the variable *Hooper consistency total number of terms* differs in mean for *Journal article*, Median= 0.3 more than for *Anonymous work*, Median= 0.1, U= 93 significant, p= 0.004<0.0083. The variable *Rolling consistency*

total number of terms differs in mean for *Journal article*, Median= 0.4 more than for *Anonymous work*, Median= 0.2, U= 92 significant, p= 0.004<0.0083. The variable *Hooper consistency descriptors* differs in mean for *Journal article*, Median= 0.4 more than for *Anonymous work*, Median= 0.2, U= 88 significant, p= 0.003<0.0083. The variable *Rolling consistency descriptors* differ in mean for *Journal article*, Median= 0.6 more than for *Anonymous work*, Median= 0.3, U= 88 significant, p= 0.003<0.0083.

Measure	Hooper consistency in total number of terms	Rolling consistency in total number of terms	Hooper consistency in descriptors	Rolling consistency in descriptors	Hooper consistency in secondary terms	Rolling consistency in secondary terms
Mann-Whitney U	392.000	391.500	379.500	379.500	428.000	428.000
Wilcoxon W	668.000	667.500	655.500	655.500	704.000	704.000
Z	-1.629	-1.635	-1.792	-1.792	-1.313	-1.313
Asymp. sig. (2-tailed)	0.103	0.102	0.073	0.073	0.189	0.189
Exact Sig. (2-tailed)	0.104	0.103	0.073	0.073	0.183	0.183
Exact Sig. (1-tailed)	0.052	0.052	0.037	0.037	0.090	0.090
Point probability	0.001	0.001	0.001	0.001	0.014	0.014

Table 29: Test statistics for Group 3= Reports or Statistics annual or Monograph and Group 4 = Anonymous work

The results of the Mann-WhitneyU test when significance is corrected using the Bonferroni multiple contrasts fit method show no significant difference between *Report or Statistics Annual or Monograph* (Group 3) and *Anonymous Work* (Group 4).

Post hoc tests

The Bonferroni multiple comparisons show a significant effect for the type of publication for the *Hooper consistency descriptors* variable between *Proceedingss or Anthologies* and *Report or Statistics annual or Monograph* and between *Proceedings or Anthologies* and *Anonymous works*. We can, therefore, put forward the following hypothesis: the Hooper consistency descriptors are greater for Minute and Anthologies than for journal articles, Reports or Statistics annual or Monographs or Anonymous work.

Conclusions

Indexing consistency in LILAC is shown to be substantially less than that offered by MEDLINE, ISA and PsycINFO for both principal and secondary terms. Descriptor consistency in LILACs is 33% (Hooper) and 50% (Rolling). However, secondary terms (qualifiers and limits), which are used very little in indexing (Median = 1), show a higher degree of consistency: 45% (Hooper) and 62% (Rolling), which is very similar to that of MEDLINE and ISA. This is because the indexers coincide in not assigning secondary terms to documents. By biasing the data and excluding those entries that lack secondary terms, the consistency is significantly reduced: 9.5 (Hooper) and 11.5 (Rolling). In short, consistency is higher in the descriptors than in the secondary terms, which are used very little. When they are used, though, their consistency is acceptable, probably because they are taken from a very limited context, and qualifiers and limits are applied only to very specific descriptors, which leaves little margin for subjective use. The nature of indexing means that it is not possible to distinguish between indexing term, concept and aspect, as was done by Iivonen (1990), although here again the consistency among indexers from different centers was significantly lower.

As for the relationship between the number of indexing terms used (exhaustiveness), in general terms the higher the number of words used to describe a document, the lower the rate of consistency. Since consistency is higher for descriptors than for secondary terms (when weighting the data, as mentioned), this implies that there is less agreement in describing secondary aspects of documents than primary ones.

In the case of those category variables that are susceptible to mediatizing the the degree of consistency, that the following points are noted: a) documents written in Spanish get a lower consistency rate than those in English or Portuguese; while, for MEDLINE, language has no significant differences; b) documents indexed in the same country show higher levels of consistency than those indexed in different countries; c) the subject matter of the documents does not affect indexing consistency; d) documents including an abstract return higher levels of consistency; e) documents

such as Proceedings and anthologies present more consistent indexing than reports, statistics annuals, monographs or anonymous works.

Why do documents indexed in Spanish show a lesser degree of consistency than those described in other languages (Portuguese and English)? Because it has been proved that consistency of indexing increases if the documents are indexed in the same country, while it decreases when the document is analysed in different countries. The reason for this lies in the fact that LILACS is the product of cooperative efforts. The participating entities (libraries) have not always had professionals with knowledge of the descriptors to be applied. Not all the staff from the various countries consistently share and use the semantic matrices of the terms included in the DeCS vocabulary. Common training of staff in knowledge and management of DeCS is fundamental for there to be a common description of documents and to improve information retrieval.

There is also a need for an organisation that would review the consistency of the descriptors assigned by the staff of the countries making up BIREME. The work done in the different countries needs to be supervised but the resources to contract qualified staff for such tasks are generally not available. Ideally, projects of this type would be run from a single organisation devoted to assigning and revising the descriptors, as happens in other activities. This organisation would have to improve the uniformity of the terms of the DeCS Thesaurus so that the language would respond better to the international standards of the main medical information databases.

DeCS management needs to be updated and carried out with wider perspectives so as to enhance its interoperability, with the emphasis on transforming it into a controlled vocabulary for representation of content objects in knowledge organization systems. DeCS Display, construction, testing, maintenance, and management should be in accordance with the new Guidelines for the Construction, Format and Management of Controlled Vocabulaires (ANSI/NISO Z39.19:2005; BS 8723-3: 2007; BS 8723-4: 2007; ISO 25964-1: 2011 and ISO 25964-2: 2011).

The indexing system is more consistent in MEDLINE, ISA and PsycINFO. The consistency coefficients are higher and present, within logical differences, homogeneity among themselves. Indeed, our study highlights that the single underlying macrotendency in the indexing process common to these databases lies in the fact that the indexers have a higher level of agreement when designating the words to describe the principal aspects of documents than the secondary ones.

About the authors

Luis Miguel Moreno Fernández is Lecturer in Indexing Languages in the Department of Information and Documentation of the School of Communication and Information Studies in the University of Murcia. He is the author of some of the main monographies written in Spaish into the field of Subject Cataloguing. He can be reached at morfedez@um.es

Mónica Izquierdo Alonso is Lecturer in Information Systems Planning and Management at the Department of Philology, Communication and Information of the Information Studies School in the University of Alcalá de Henares. Her current research is centered in analysing the scientific production. She can be reached at monica.izquierdo@uah.es

Antonio Maurandi López works in the Statistical Office of the Research Support Service in the University of Murcia. He is Associate Teacher in the Education School. He is author of several papers and a ebook related with the statistical foundation in research activities. He can be reached in amaurandi.um.es

Javier Vallés Valenzuela is the Director of the Library Campus of the Academic and Cultural Centre of Jariquilla, sited in Querétaro (Mexico). This institution is affiliated to the National and Autonomus University of Mexico (UNAM), the main academic institution of Latin America. He can be reached at biblio@teljuriquilla.unam.mx

References

American National Standards Institute. (2005). *ANSI/NISO Z39.19. Guidelines for the construction, format and management of monolingual controlled vocabularies.* Bethesda, MD: NISO Press.

Armenteros, V.I. (2002). Procedimientos de trabajo para LILACS. *ACIMED*, **4**. Retrieved 17 April, 2013 from http://bvs.sld.cu/revistas/aci/vol10_4_02/aci050402.htm (Archived by WebCite® at http://www.webcitation.org/6Lidf9eRk)

BIREME. (2005). *Manual de indización de documentos para la base de datos LILACS (2nd. ed.).* Sao Paulo, Brazil: BIREME/OPS/OMS.

BIREME. (2006). *Guía de selección de documentos para la base de datos LILACs.* Sao Paulo, Brazil: BIREME/OPS/OMS.

Braam, R.R. & Bruil, J. (1992). Quality of indexing information: authors' views on indexing of their articles in Chemical Abstracts online CA-file . *Journal of Information Science*, **18**(5), 399-408.

British Standards Institution. (2007). *BS 8723-3. Structured vocabularies for information retrieval: guide: vocabularies other than thesauri*. London: British Standards Institutions.

British Standards Institution. (2007). *BS 8723-4. Structured vocabularies for information retrieval: guide: interoperability between vocabularies*. London: British Standards Institution.

Cooper, W.S. (1969). Is interindexer consistency a hobgoblin? *American Documentation*, **20**(3), 268-278.

Fugmann, R. (1985). The five-axiom theory of indexing and information supply. *Journal of the American Society for Information Science*, **36** (2), 116-129.

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

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

Iivonen, M. (1990). Interindexer consistency and the indexing environment. *International Forum on Information and Documentation*, **15**(2), 16-21.

Iivonen, M. (1995). Consistency in the selection of search concepts and search terms. *Information Processing & Management*, **31**(2), 173-190.

Iivonen, M. & Sonnenwald, D. H. (1998). From translation to navigation of different discourses: a model of search term selection during the pre-online stage of the search process. *Journal of the American Society for Information Science*, **49**(4), 312-326.

International Standards Organization. (2011). *ISO 25964-1. Thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval.* Geneva, Switzerland: International Standards Organization.

International Standards Organization. (2012). *ISO 25964-1. Thesauri and nteroperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneva, Switzerland: International Standards Organization.

Jiménez Miranda, J. (1998). Acceso a MEDLINE y LILACS mediante el MeSH y el DeCS. *ACIMED*, **6**(3), 153-162.

Lancaster, F.W. (1991). *Indexing and abstracting in theory and practice*. London: The Library Association.

Lancaster, F.W. (1998). *Indexing and abstracting in theory and practice.* (2nd ed.). London: The Library Association.

Lancaster, F.W. (2003). *Indexing and abstracting in theory and practice.* (3nd ed.) Champaign, Ill.: University of Illinois, Graduate School of Library and Information Science.

Leininger, K. (2000). Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science*, **32**(1), 4-8.

Lind Blackwell, M. (1994). *Three library and information science databases revisited: currency, coverage and overlap, interindexing consistency*. (Unpublished doctoral dissertation, Kent State University, Ohio, USA).

Moreno Fernández, L.M. (2003). La consistencia de la indización: II. Estado de la cuestión y tendencias de la investigación. *AIBDA: Asociación Interamericana de Bibliotecarios, Documentalistas y Especialistas en Información Agrícola*, **24**(1-2), 31-66.

Rolling, L. (1981). Indexing consistency, quality and efficiency. *Information Processing and Management*, **17**(1), 69-76.

Sievert, M.C. & Andrews, M.J. (1991). Indexing consistency in Information Science Abstracts. *Journal of the American Society for Information Science*, **42**(1), 1-6.

Soergel, D. (1994). Indexing and retrieval performance: the logical evidence. *Journal of the American Society for Information Science*, **45**(8), 589-599.

Soler Monreal, C. & Gil-Leiva, I. (2011). Evaluation of controlled vocabularies by inter-indexer consistency. *Information Research* **16**(4), paper 502. Retrieved 11 December, 2013 from http://

informationr.net/ir/16-4/paper502.html (Archived by WebCite® at http://www.webcitation.org/6LnOZJ6D1)

Tonta, Y. (1991). A study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources and Technical Services*, **35**(2), 177-185.

Uren, V. (2000). An evaluation of text categorisation errors. In *Proceedings of the one-day Workshop on Evaluation of Information Management Systems, 15 September 2000*, (pp. 79-87). London: Queen Mary and Westfield College.

White. H.D. & Griffith, B.C. (1987). Quality of indexing in online databases. *Information Processing and Management*, **23**(3), 211-224.

Zunde, P. & Dexter, M.E. (1969). Indexing consistency and quality. *American Documentation*, **20**(4), 259-267.

How to cite this paper