Indexing Consistency and its Implications for Information Architecture: A Pilot Study

Hope A. Olson & Dietmar Wolfram
School of Information Studies
University of Wisconsin-Milwaukee
P.O. Box 413
Milwaukee, WI 53201
Phone: 414 229-4707
Fax: 414 229-6699
E-mail {holson, dwolfram}@uwm.edu

Abstract

Consistency in the assignment of indexing terms has been studied on the small scale for many years. As opportunities increase for large numbers of people to contribute to indexing of public documents on the World Wide Web, consistency on the large scale becomes problematic. This pilot study examines inter-indexer consistency on a larger scale than other studies. Consistency in the assignment of indexing terms to a document follows an inverse shape, although it is not strictly power law-based, as has been observed for many social phenomena. The authors discuss the challenges and potential benefits of having this diversity of indexing access points to documents and their implications for effective information architecture.

Introduction

Consistency in indexing has long been considered essential for effective retrieval. In the process of indexing, indexers choose what topics to represent and what to call those topics. The goal is to select and name topics consistently so that all of the material about any given topic will be found together. Decades of research on consistency between indexers and by the same indexer at different times has documented medium to high levels of inconsistency. People simply do not choose the same concepts or the same words for those concepts. This body of research has identified various factors that contribute to consistency, but it has not addressed the underlying nature of consistency. Can consistency be achieved or is inconsistency inevitable? Does it follow predictable patterns? Consistency is also taken to be essential to good information architecture. When developing access points to content, reliance on commonly used terminology is desired for effective access.

One thing we know about consistency is that indexers usually agree on a core of topics, but vary in representing peripheral ones. This suggests that the choices might adhere to a power law that can predict the distribution of topics. If the distribution of topics indexed is predictable it might be possible to minimize the variation or to develop interfaces for

searching databases that would take that distribution into account. It could also be that the variation is a positive factor. This study explores these possibilities.

Previous Research

Interindexer consistency studies illustrate consistent results in their consistent findings of inconsistency. The 1960s was a sort of golden age of consistency studies culminating in Zunde and Dexter's analysis (1969) of earlier data (Schultz, Schultz, & Orr, 1965) pointing toward "power laws" and fuzzy sets as explanatory devices and Cooper's (1969) questioning of the importance of consistency. Later studies were variations on these themes, branching out to other contexts. Markey's (1984) meta-analysis of 25 studies, largely from the 1960s and 1970s, noted not only the ubiquity of inconsistency from levels of 82% consistency down to 4%, but also looked at factors such as exhaustivity (i.e., indexing depth within a document) and vocabulary size that influence those levels. Markey's own study (1984) explored consistency in indexing of images. Other studies also examined different contexts. Chan (1989) tested consistency in a library catalog context, refining definitions of a match and partial match to suit a precoordinate vocabulary, with similar results. Reich and Biever (1991), Sievert and Andrews (1991), Giral and Taylor (1993) and Leininger (2000) looked at indexing in agriculture, information science, and psychology respectively. These four studies, as well as Chan's (1989) and Tonta's (1991) each grew from the existence of duplicate records either within or between databases, so they examined terms assigned to two each of multiple titles. More recently, Saarti (2001) tested consistency in indexing and searching for fiction using a controlled vocabulary. Numerous other perspectives also have been taken. For example, David et al (1995) explored the dynamics of inconsistency as a cognitive issue. Further examples generally follow the same patterns of investigation, either multiple records from a serendipitous duplication or experiments involving multiple indexers or searchers and few records. Looking in a different direction, Zunde and Dexter (1969, 266) identify the implication of their data that a power law is operating in group indexing, suggesting that further informetric exploration could be fruitful. Maron's (1977) concept of retrieval aboutness or *R-about*, suggests that the ideal index terms are those that a given population who would find a document useful would use to search for that document. Maron was calling for a consensus that is obviously not found in previous interindexer consistency studies.

The research described in this article uses informetric methods to study inter-indexer consistency. Informetrics is the quantitative study of recorded discourse. It examines regularities in the way information is produced and used. Research in informetrics has explored many information-related social phenomena, including patterns in authorship within disciplines, the scatter of literature on a given topic over different publication sources, word usage in text corpora, the growth of information sources over time, and citation patterns among authors and sources. Informetric research has also examined regularities in document indexing, both in terms of exhaustivity and specificity (or how frequently and how many different terms are used across documents within a document set). Observed regularities for most of these phenomena follow patterns where many sources, whether authors, words, or documents, contribute few items or occurrences.

When grouped based on frequencies, many of these phenomena have been observed to follow so-called "power laws" based on the formulaic representation of the relationship. A popular example of this is Zipf's Law (Zipf, 1949), first used to describe word frequency occurrence, which takes the form:

$$y = a / r^b$$

where r is the rank of a given word based on its frequency of occurrence within a text, y is the word's frequency, and a and b are parameters whose values are determined by the characteristics of the observed distribution. Variations of this "law" exist for other social phenomena and have been named after their discoverers, including Lotka's Law (Lotka, 1926) for author productivity and Pareto's Law (Pareto, 1897), for the distribution of wealth within a population. More recently, it has been reported that power laws apply to Web-based characteristics as well (Huberman, 2001; Nielsen, 1997).

Method

Previous studies that have explored interindexer consistency generally have data from only a handful of indexers or a mix of indexers or librarians and searchers (e.g. Schultz, Schultz, and Orr 1965 and Saarti 2001). That is where the current study differs. It gathers data, even in this pilot study, from a substantial group of people all playing the role of indexer. The data were derived from responses by MLIS students in an indexing exercise. Data were collected from two sections (one online, one onsite) of a course on Information Organization. Students were asked to provide five terms that describe what an assigned conference paper on the topic of indexing is about. Students were asked to submit their terms without identifying themselves. Aggregate data were stored and tabulated in a Microsoft Access database based on the frequency of occurrence of terms assigned. Terms were not from a controlled vocabulary and minimal regularization (e.g. collapsing singular and plural occurrences) was employed. The frequency of co-occurrence of terms by a given indexer was also determined to see if there was much agreement on the pairing of terms used among indexers. Exploratory analysis was also used to determine if there were any relationships in the way terms were used.

Findings

Thirty-three usable responses were collected from the two class sections. Only five responses were submitted by students in the online section of the class, undoubtedly due to the voluntary nature of the exercise for which there was no "in-class" time set aside due to the asynchronous nature of the instruction, unlike the onsite students who had class-time to complete the exercise.

Initial findings indicate that over a reasonably large number of people, consistency in identifying key concepts is small. In fact, the frequency distribution of common topics follows a strong inverse pattern, with few terms being agreed upon by many participants, and many uniquely identified terms. The relationship, although not classically Zipfian (i.e., following a classic power law) based on visual inspection of a logarithmic

transformation of the data set, demonstrates the wide degree of scatter in terminology used by participants in the study. This is demonstrated in Figure 1.

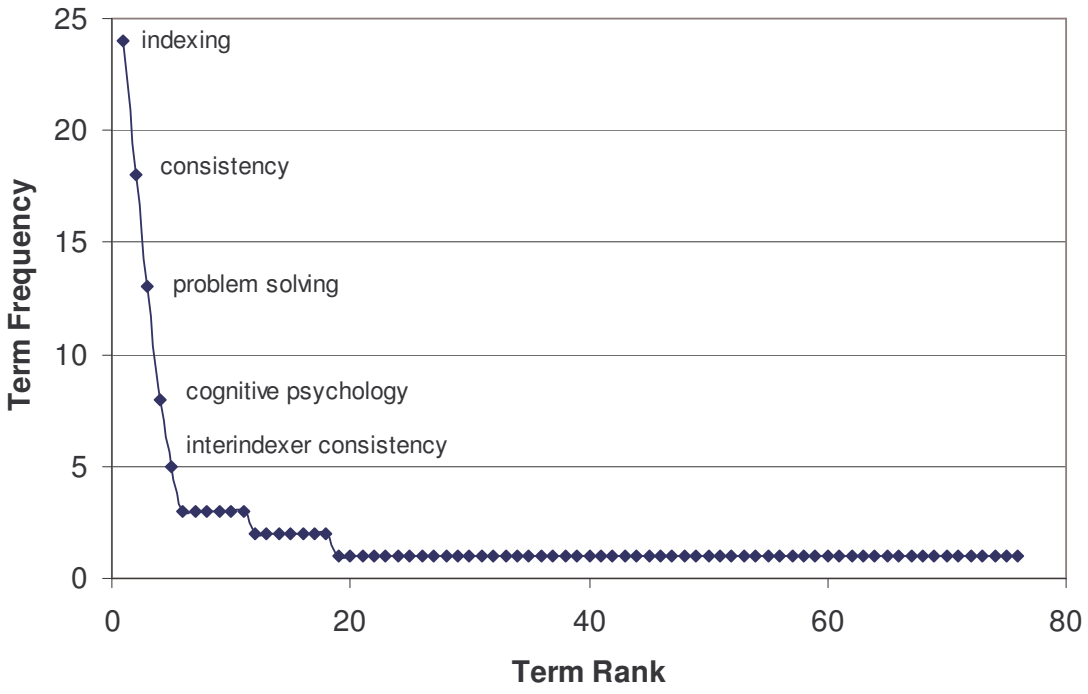

Figure 1. Rank-Frequency Plot of the Frequency of Occurrence of Different Indexing Terms

The distribution of term co-occurrences also follows a strong inverse shape, demonstrating that there is little agreement on the use (Figure 2). In fact, 89% of co-occurring term pairs (214 of 241) occur only one time.

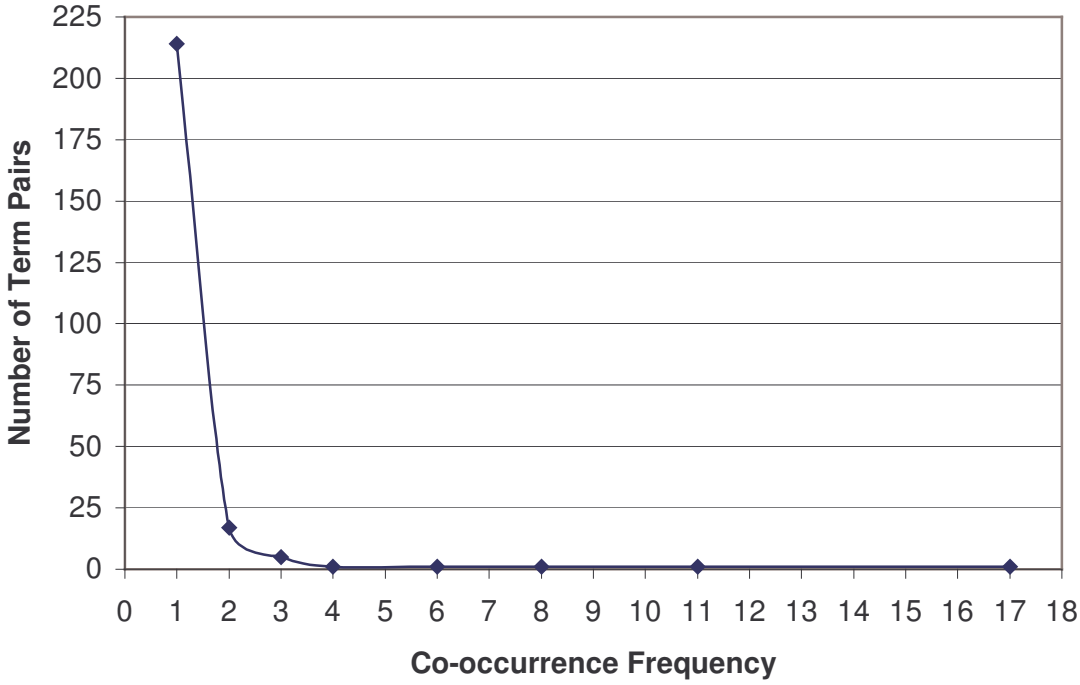

Figure 2. Size-Frequency Plot of the Frequency of Co-occurrence of Different Pairs of Terms

To determine if there were any relationships in term assignment and their frequency of use, pair-wise correlations between the frequencies of co-occurrence of the terms occurring three or more times were calculated using Spearman's rank correlation coefficient. The skewness of the data distribution made the use of a standard Pearson's *r* correlation less reliable. The resulting correlation values are summarized in Table 1. Only nine of the term pairs (shaded in Table 1) involving six of the twelve most frequently used terms (indexing, study, cognitive psychology, problem solving, consistency, descriptors) resulted in a significant positive correlation, indicating a relationship in their use by the participants. The investigators also initially attempted to use more exploratory data analysis techniques to determine if there were any hidden relationships among the indexing terms. The relatively small size of the data set for this pilot study and poor outcomes in the assessment using both multi-dimensional scaling and cluster analysis made the use of these techniques inappropriate.

Table 1.
Correlation Matrix of Term Co-Occurrences for Terms Occurring >= 3 Times

		indexing	study	cognitive psychology	consistency	indexing behavior	interindexer consistency	problem solving	exhaustivity	cataloging	evaluation	descriptors
indexing	ρ	1.00										
	Sig.	.										
study	ρ	.527	1.00									
	Sig.	.096	.									
cognitive psychology	ρ	.842	.608	1.00								
	Sig.	.001	.047	.								
consistency	ρ	.883	.560	.764	1.00							
	Sig.	.000	.073	.006	.							
indexing behavior	ρ	.168	.522	.192	.122	1.00						
	Sig.	.621	.099	.572	.720	.						
interindexer consistency	ρ	-.142	.023	-.184	-.455	.409	1.00					
	Sig.	.677	.946	.588	.160	.212	.					
problem solving	ρ	.738	.219	.684	.666	.082	-.091	1.00				
	Sig.	.009	.517	.020	.025	.812	.790	.				
exhaustivity	ρ	.414	-.041	.231	.159	.193	.495	.543	1.00			
	Sig.	.205	.904	.494	.640	.570	.122	.084	.			
cataloging	ρ	.497	.118	.320	.535	.106	-.173	.686	.292	1.00		
	Sig.	.120	.730	.338	.090	.756	.610	.020	.383	.		
evaluation	ρ	.179	.304	.161	.419	.006	-.081	.100	.098	.230	1.00	
	Sig.	.599	.364	.637	.199	.985	.814	.771	.774	.495	.	
descriptors	ρ	.746	.271	.384	.778	.165	-.099	.573	.339	.450	.366	1.00
	Sig.	.008	.420	.244	.005	.628	.772	.065	.308	.165	.268	.

Discussion & Conclusions

The findings of this preliminary study demonstrate the variability inherent in the indexing process. The size of the data set and resulting frequency distribution of term occurrences do not allow the investigators to conclude that the observed behavior of large-scale indexer consistency follows a power law. Although this has been shown to apply in other social environments, particularly for relatively small and well-defined areas, recent larger scale studies of different Internet content and usage phenomena indicate that observed behavior may not be best modeled by a strictly Zipfian distribution (Ajiferuke & Wolfram, 2004; Ajiferuke, Wolfram, & Xie, 2004). It follows from the scatter of the assignment of index terms that searchers who try to identify query terms for searching for information will also rely on different terminology to describe a similar concept of interest. It is, therefore, desirable to have many access points to information sources to

accommodate the diversity of terms likely to be used by searchers. Iivonen suggests that "variety in the selection of search terms can also be seen as an opportunity" using many terms to increase the likelihood of matching searchers' terms (1995, 188). Iivonen goes on to suggest, as others have done, use of a search thesaurus. However, a range of access points can also serve as entry points to a conventional controlled vocabulary and navigation through its syndetic structure to core terms under which most material on a topic will be indexed. In this way, indexers applying a controlled vocabulary do the intellectual labor of connecting terms, thereby saving searchers' time.

Recognition of the differences in indexing assignment has implications for the development of public information systems. Multiple access points that accommodate the different ways that users interpret content are needed so that users may be guided to relevant content despite using different terminology. By studying patterns in user behavior regarding the assignment of topicality of system content, systems developers may incorporate an understanding of these patterns into information system design.

Professional indexing of Web-based content to accommodate broad search needs is, undoubtedly, expensive and time consuming. This is not to suggest that professional indexers are no longer needed. The structure and quality control provided by information professionals ensure a level of subject access, particularly to formally published content that should not be undervalued. These initial findings merit further investigation along several lines. By comparing the most popular indexing terms assigned by a large general audience with those of an experienced indexer, the reliability of large scale indexing by consumers of information instead of relying on costly and time-intensive professional indexing can be assessed. In the case of the study used in the current investigators' indexing exercise, the three major descriptors assigned to the study by ERIC match the first, third, and fourth top terms generated by the student indexers: "indexing," "problem solving," and "cognitive psychology." The ERIC record also includes a major identifier (a term not in the ERIC thesaurus), "interindexer consistency," that matches the fifth term and includes the second term, "consistency." Of the four minor descriptors and one minor identifier only one was used by a student indexer and it occurred only once. This comparison of ERIC to the data gathered from students again follows the general pattern of a power law with a focused core and diffuse periphery. Future research might ask indexers to indicate major and minor terms to add more understanding of the balance between terms.

With the growing popularity of "folksonomies" and the demonstrated reliability of communally developed information resources like Wikipedia (Giles, 2005), at least for some disciplines, communally developed indexes based on group consensus may have merit. Communal subject indexing could be applied to popular content available on the Web to supplement more standard keyword searching environments currently in use.

As a pilot study with limited data, the findings of this study cannot be generalized, but suggest that further investigation is warranted. Participants were not expert in indexing, so findings may vary with more experienced indexers. Other factors will influence outcomes. The shape of the resulting term and co-occurrence distributions will be

influenced by the level of exhaustivity permitted. Up to five terms were allowed for the current study. A higher level of exhaustivity would allow opportunities for additional terms to be used. The next phase of this research will involve a larger number of users to permit a more comprehensive investigation of selection and co-occurrence patterns using additional exploratory and analytical methods.

References:

Ajiferuke, I., & Wolfram, D. (2004). Informetric modeling of Internet search and browsing characteristics. *Canadian Journal of Information and Library Science*, 28(1), 1-16.

Ajiferuke, I., Wolfram, D., & Xie, H. (2004). Modelling Website Visitation and Resource Usage Characteristics by IP Address Data. In *Proceedings of the 32nd Annual Conference of the Canadian Association for Information Science.* Available from: http://www.cais-acsi.ca/proceedings/2004/ajiferuke_2004.pdf

Chan, L.M. (1989). Inter-indexer consistency in subject cataloging. *Information Technology and Libraries, 8*(4), 349-358.

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

David, C., et al. (1995). Indexing as problem solving: A cognitive approach to consistency. *Proceedings of the ASIS Annual Meeting,* 32, 49-55.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438, 900-901.

Giral, A., & A.G. Taylor. (1993). Indexing overlap and consistency between the Avery Index to Architectural Periodicals and the Architectural Periodicals Index. *Library Resources & Technical Services, 37*,(1), 19-44.

Huberman, B. A. (2001). The laws of the Web: Patterns in the ecology of information. Cambridge, MA: MIT Press.

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

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

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science, 16*(12), 317-323.

Markey, K. (1984). Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library & Information Science Research, 6*, 155-177.

Maron, M.E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science, 28*(1), 38-43

Nielsen, J. (1997). *Zipf Curves and Website popularity*. Available: http://www.useit.com/alertbox/zipf.html [2006, January 10th].

Pareto, V. (1897). *Cours d'economie politique* (Vol. 2). Lausanne: Université de Lausanne.

Reich, P, & E.J. Biever. (1991), Indexing consistency: The input/output function of thesauri. *College & Research Libraries 52*, 336-342.

Saarti, J. (2001). Consistency of subject indexing of novels by public library professionals and patrons. *Journal of Documentation, 58*(1), 49-65.

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

Schultz, C.K., W.L. Schultz, & R.H. Orr. (1965). Comparative indexing: Terms supplied by biomedical authors and by document titles. *American Documentation, 16*(4), 299-312.

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

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.

Zunde, P, & M.E. Dexter. (1969). Indexing consistency and quality. *American Documentation, 20*(3), 259-267.