

Martin Frické — SIRLS, The University of Arizona. Email: mfricke@u.arizona.edu

Logic and the Organization of Information: An Introduction

Abstract

The paper considers how logic might be used in the organization of information, in particular with: indexing, concepts, synonyms and homographs, directed acyclic graphs of topics, faceting, and information navigation.

Introduction

This paper is a paean to the use of logic in the organization of information (see also (Shera 1965) p.105 and (Rogers 1960a, b) for other enthusiasts).

Indexing

Indexing is the establishment of an ordered association list of pairs of keys and values (Frické 2012). The values are Information Objects (IOs), or some means of identifying them, and the keys are typically string tokens (i.e. natural language words or phrases). In so-called ‘derived indexing’ the keys identify string tokens which are instances of patterns to be found in the source IO texts; in contrast, in ‘assigned indexing’ the keys identify something else, in the view of this paper they identify uses of concepts.

There is a tradition that is suspicious of the surface form of source texts (Gardin 1973; Frické 2012; Fugmann 1982, 1980, 1993; Stock 2010). The reason is that it is possible to say the same thing in different ways, and, for the most part, indexing is interested in that ‘same thing’ and not in the particular words used on particular occasions to say it. A possible move here is to invoke the familiar Triangle of Meaning (Frické 2012; Stock 2010; Hjørland 2007)) and then to suggest that the index keys (on one vertex of the Triangle) are just names of concepts (on another vertex of the Triangle). This accommodates synonyms, homographs, and it allows for the indexing a keys which does not themselves appear in the source texts (see also (Weinberg 1996)).

Concepts and Logic

Concepts can be understood as abstract objects (Frické 2012) (see also (Hjørland 2009; Szostak 2011)). And, in turn, some symbolic logic can be used to identify those concepts (Frické 2012) (see also (Gnoli 2006; Stock 2010)).

There is the abstraction or comprehension or intensional abstraction notation.

$\{x: \Phi(x)\}$

which appears in naïve set theory as ‘set builder notation’. In an intensional abstraction $\{x: \Phi(x)\}$, the $\Phi(x)$ itself denotes an ‘open sentence’ (a formula of the Predicate Calculus, usually with a free occurrence of x .) So the three formulas

$\{x: \text{Iron}(x)\}$

$\{y: \text{Aluminum}(y)\}$

$\{z: \text{Ingot}(z) \& \text{Aluminum}(z)\}$

are examples of abstractions which we understand as straightforward notation for concepts (see also (Bealer 1982, 1998)).

How does this formalization of concepts help? First it reveals many of the *a priori* interrelations between concepts. In an ordinary back of the book index, for example, there usually would be hierarchical entries such as

```
Ingots
    Aluminum
    Iron
```

But straight derivation, or computer theorem proving, would reveal how the concepts

```
{x:Ingot(x)&Aluminum(x)}
{x:Ingot(x)&Iron(x)}
{x:Ingot(x)}
```

are related one to another. (This is a matter of the deductive relations between the scopes of the abstractions.)

Using abstractions for concepts also helps with synonyms and homographs.

Synonyms and Homographs

Synonyms and homographs—and generalized synonymy and generalized homography—are problems for librarianship and information retrieval for they reduce precision and recall (Harter and Hert 1997; Salton 1992).

A solution is to do all the important operations using concepts. In turn, working with concepts can be done using symbolic logic. Symbolized formulas are the scaffolding or reserve currency or universal translation language lying in the background (see also (Gardin 1973; Fugmann 1982, 1980, 1993; Stock 2010)).

Synonyms would be collected into synsets (Fellbaum 1998). But the label or identifier for each synset is not a headword or preferred term, rather it is a concept which in turn is a logical formula. Managing preferred terms, and ‘lead-in’ terms, which would occur in controlled vocabularies, would be done via the concepts. Homographs would be similar except that the relationship between word and concept is one-to-many. This is pretty well the approach of the Simple Knowledge Organization System (SKOS) (Isaac and Summers 2009; Miles et al. 2005; W3C 2010) apart from SKOS does not give the concepts any internal structure. Generalized synonyms and generalized homonyms invite the use of considerable internal structure in logic, and they would also likely need humans for correct identification and rendering in logic (Frické 2012).

The eventual uses of underlying concepts to produce string representations of themselves, for the index display keys, is both flexible and powerful. It is relatively easy to generate from a logical formula a representation in any string syntax and vocabulary that is considered desirable (i.e. for different audiences, children and adults, and different natural languages).

Directed Acyclic Graphs (DAGs) of Concepts or Topics

Information retrieval has a special interest in subjects or topics (Cutter 1876, 1904). The view here is that topics are just concepts (indeed, typically the ones used as index keys). And the all important relationship between topics is whether one topic is broader or narrower than another. In fact, this relationship may be the most important in information retrieval. For example, that is the basic axiom of the Classification Research Group’s 1955 Paper (Classification Research Group 1955).

A topic can obviously have more than one narrower topic child, which suggests that a topic graph might be tree-like or hierarchy-like. And a topic does not want to be broader or narrower than itself. But what about a topic's broader parent topic (or topics)? Many topics should have more than one parent topic. For example, the topic 'women poets' should have the one broader parent topic 'poets', generalizing on the women, and another broader parent topic 'women authors', generalizing on the poets. If topics can have multiple parents, the topic or concept graph is a Directed Acyclic Graph (not a tree or a hierarchy) (cf. (Kwasnik 1999)).

Logic and inference can provide some of the graph structure, the *a priori* part. But many of the links between topics do not come from logic alone. Many come from science or the empirical structure of our world. That 'Europe' is a more general topic than 'France' is not a matter of pure logic. It is related to how our world is. A different example of a mechanism to link topics comes from education or learning. Within any discipline, or field of learning, say Physics or Carpentry, there are ideas of how the component ideas, or topics, should fit together to provide a reasonable learning experience to a community of students. Hence there will be DAGs of topics related to learning. Partitive links, or hierarchies, may also be used to construct DAGs; for example, for some Users and some purposes, we may wish to link the concept carburetor to the concept engine.

Indexing, Search, and Faceting

Search is a counterpart to indexing. There are many kinds of searches and much of the routine kinds can be done as database retrievals or string-in-string pattern matches on text or metadata. But the most important and challenging kind of search involves subjects or topics i.e. concepts and logic.

Many, or even most, of the concepts for indexing are going to be compounds. So, from a logical point of view, they will be synthesized from components, as and when needed. Synthesis goes along with postcoordination. This suggests that many of the searches for multiterm strings will be postcoordinate searches. To date, attempts with semantic postcoordination have been restricted to Boolean constructions and they have characteristic shortcomings. But First Order Logic goes beyond Boolean constructions and it opens new opportunities.

Additionally, often the components of a synthesized compound are, or can be, categorized or faceted. For example, the topic '18th Century France' is composed of a time period and a place. One component is of the category period and the other of the category place. There is a focus from a period facet, and a focus from a place facet. There are kinds of concepts (Austin 1984; Foskett 1977; Lambe 2007; Morville and Rosenfeld 2006; Willetts 1975; Vickery 1960, 1966; Cheti and Paradisi 2008; Slavic 2008). Hence there is the very important faceted classification (Broughton 2006; Classification Research Group 1955; Perreault 1969; Beghtol 2008) and its counterpart faceted search (Broughton 2004, 2006; Buchanan 1979; Ranganathan 1959, 1967; Wilson 2006; Gnoli 2008; Vickery 1960, 2008, 1966; La Barre 2006, 2010; Foskett 1996; Foskett 2003; Gardin 1965).

Search using logic can exploit these features: the facetting and the synthesizing.

Information Navigation

Many of the retrieval patterns involve following trails of ‘bibliographical relationships’ (Tillett 1987, 1991b, a, 2001). One bibliographical relationship is that different IOs address related topics. This amounts to traveling a DAG of topics.

Many of these bibliographical relationships, in so far as they can be computed and assembled into paths, involved logic either as database retrievals or topic DAG manipulations.

Elaine Svenonius tells us of the need for what she calls ‘subject languages’ for collocation, of IOs on the same topics, and for navigation, around the bibliographical universe (Svenonius 2000; Wilson 2001). The suggestion of the present paper that symbolic logic is valuable for collocation, and DAGs of concepts provide the navigation.

Conclusion

Knowledge organization might well benefit from the greater use of symbolic logic.

References

- Austin, D. 1984. *PRECIS: a manual of concept analysis and subject indexing*. 2 ed. London: The British Library.
- Bealer, G. 1982. *Quality and Concept*. Oxford: Oxford University Press.
- Bealer, G. 1998. Intensional entities. In: Craig, E. (ed.) *Routledge Encyclopedia of Philosophy*. Vol. 4. London: Routledge, pp.803-807.
- Beghtol, C. 2008. From the Universe of Knowledge to the Universe of Concepts: The Structural Revolution in Classification for Information Retrieval. *Axiomathes* 18 (2), pp.131-144.
- Broughton, V. 2004. *Essential Classification*. New York: Neal-Schuman.
- Broughton, V. 2006. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives* 58 (1/2), pp.49-72.
- Buchanan, B. 1979. *Theory of Library Classification*. London: Clive Bingley.
- Cheti, A. and Paradisi, F. 2008. Facet analysis in the development of a general controlled vocabulary. *Axiomathes* 18 (2), pp.223-241.
- Classification Research Group 1955. The need for a faceted classification as the basis of all methods for information retrieval. *Library Association Record* 57 (7), pp.262-268.
- Cutter, C. A. 1876. *Rules for a Printed Dictionary Catalogue*. Washington: Government Printing Office.
- Cutter, C. A. 1904. *Rules for a Dictionary Catalog*. 4 ed. Washington, DC.: U.S. Government Printing Office.
- Fellbaum, C. (ed.) 1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Foskett, A. C. 1977. *Subject approach to information*. 3 ed. London: Clive Bingley.
- Foskett, A. C. 1996. *Subject approach to information*. 5 ed. London: Facet Publishing.
- Foskett, D. J. 2003. Facet analysis. In: Drake, M. A. (ed.) *Encyclopedia of Library and Information Science*. 2 ed. New York: Marcel Dekker, pp.1063-1067.
- Frické, M. 2012. *Logic and the Organization of Knowledge*. New York: Springer.
- Fugmann, R. 1980. On the Practice of Indexing and its Theoretical Foundations. *International Classification* 13 (7), p.13.
- Fugmann, R. 1982. The Complementarity of Natural and Indexing Languages. *International Classification* 9 (3), pp.140-144.
- Fugmann, R. 1993. *Subject analysis and indexing. Theoretical foundation and practical advice*. Frankfurt/Main: Indeks Verlag.

- Gardin, J.-C. 1965. Free classifications and faceted classifications: their exploitation with computers. In: Atherton, P. (ed.) *Classification research: proc. int. conf. Elsinore 1964, Munksgaard, Copenhagen*, 1965. pp.161-176.
- Gardin, J.-C. 1973. DOCUMENT ANALYSIS AND LINGUISTIC THEORY. *Journal of Documentation* 29 (2), pp.137-168.
- Gnoli, C. 2006. The meaning of facets in nondisciplinary classifications. In: *9th ISKO conf. Vienna*, pp.11-18.
- Gnoli, C. 2008. Facets: A Fruitful Notion in Many Domains. *Axiomathes* 18 (2), pp.127–130.
- Harter, S. P. and Hert, C. A. 1997. Evaluation of information retrieval systems: approaches, issues and methods. In: Williams, M. E. (ed.) *Annual Review of Information Science and Technology*. Vol. 32 Medford, NJ: Information Today, Inc., pp. 3–94.
- Hjørland, B. 2007. Semantics and knowledge organization *Annual Review of Information Science and Technology*. Vol. 41 Medford, NJ: Information Today, Inc., pp.367-405.
- Hjørland, B. 2009. Concept theory. *Journal of the American Society for Information Science and Technology* 60 (8), pp.1519–1536.
- Isaac, A. and Summers, E. (eds.) 2009. *SKOS Simple Knowledge Organization System Primer*.
- Kwasnik, B. H. 1999. The Role of Classification in Knowledge Representation and Discovery. *Library Trends* 48 (1), pp.22-47.
- La Barre, K. 2006. *The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design*. PhD dissertation. Indiana University.
- La Barre, K. 2010. Facet Analysis. In: Cronin, B. (ed.) *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, Inc, pp.243-286.
- Lambe, P. 2007. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Oxford, England: Chandos Publishing.
- Miles, A. et al. 2005. SKOS Core: Simple Knowledge Organisation for the Web. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Madrid, Spain, 12-15 September 2005. p. 5-13. Retrieved April 15, 2006, from: <http://www.slais.ubc.ca/PEOPLE/faculty/tennis-p/dcpapers/paper01.pdf>.
- Morville, P. and Rosenfeld, L. 2006. *Information Architecture for the World Wide Web*. Sebastopol, CA: O'Reilly.
- Perreault, J. M. 1969. *Towards a theory for UDC*. London: Archon Books & Clive Bingley.
- Ranganathan, S. R. 1959. *Elements of Library Classification*. 2 ed. London: Association of Assistant Librarians.
- Ranganathan, S. R. 1967. *Prolegomena to Library Classification*. Available at: <http://dlist.sir.arizona.edu> (Accessed: October 10 2010).
- Rogers, F. B. 1960a. *Medical Subject Headings. Preface and Introduction*. Washington D.C.: U.S. Department of Health, Education, and Welfare, pp.i-xix.
- Rogers, F. B. 1960b. Review of Taube, Mortimer. Studies in coordinate indexing. *Bull. Med. Libr. Assoc.* 42 (July 1954), pp.380-384.
- Salton, G. 1992. The state of retrieval system evaluation. *Information Processing and Management* 28 (4), pp.441–449.
- Shera, J. H. 1965. Classification: Current functions and applications to the subject analysis of library materials. *Libraries and the organization of knowledge*. Hamden, CT: Archon, pp.97-111.
- Slavic, A. 2008. Faceted Classification: Management and Use. *Axiomathes* 18 (2), pp.257-271.
- Stock, W. G. 2010. Concepts and semantic relations in information science. *Journal of the American Society for Information Science and Technology* 61 (10), pp.1951–1969.
- Svenonius, E. 2000. *The Intellectual Foundation of Information Organization*. Cambridge, MA: M.I.T. Press.
- Szostak, R. 2011. Complex concepts into basic concepts. *Journal of the American Society for Information Science and Technology* 62 (11), pp.2247-2265.

- Tillett, B. B. 1987. *Bibliographic relationships: Toward a conceptual structure of bibliographic information used in cataloging*. Ph.D. dissertation. University of California, Los Angeles.
- Tillett, B. B. 1991a. A Summary of the Treatment of Bibliographic Relationships in Cataloging Rules. *Library Resources and Technical Services* 35 (4), pp.393-405.
- Tillett, B. B. 1991b. A Taxonomy of Bibliographic Relationships. *Library Resources & Technical Services* 35 (2), pp.150-158.
- Tillett, B. B. 2001. Bibliographic Relationships. In: Bean, C. A. and Green, R. (eds.) *Relationships in the organization of knowledge*. Dordrecht: Kluwer Academic, pp.19-35.
- Vickery, B. C. 1960. *Faceted classification: a guide to construction and use of special schemes*. London: Aslib.
- Vickery, B. C. 1966. Faceted classification schemes. In: Artandi, S. (ed.) *Rutgers Series on Systems for the Intellectual Organization of Information*. Vol. 5. New Brunswick, NJ: Graduate School of Library Science at Rutgers University.
- Vickery, B. C. 2008. Faceted Classification for the Web. *Axiomathes* 18 (2), pp.145-160.
- W3C 2010. *Introduction to SKOS*. Available at: <http://www.w3.org/2004/02/skos/intro> (Accessed: October 10 2010).
- Weinberg, B. H. 1996. Complexity In Indexing Systems - Abandonment And Failure: Implications For Organizing The Internet. Available at: <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html> (Accessed: October 12 2010).
- Willets, M. 1975. An investigation of the nature of the relation between terms in thesauri. *Journal of Documentation* 31 (3), pp.158-184.
- Wilson, P. 2001. The Intellectual Foundations of Information Organization. *College & Research Libraries* 62 (2), pp.203-204.
- Wilson, T. 2006. The strict faceted classification model. Available at: <http://facetmap.com/pub/> (Accessed: October 10 2010).