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 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: Ψ(x)\}
which appears in naïve set theory as ‘set builder notation’. In an intensional abstraction \{x: Ψ(x)\}, the Ψ(x) itself denotes an ‘open sentence’ (a formula of the Predicate Calculus, usually with a free occurrence of x.) So the three formulas
\{x:Iron(x)\}
\{y:Aluminum(y)\}
\{z:Ingot(z)&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 straightforward derivation, or computer theorem proving, would reveal how the concepts

\{x: \text{Ingots}(x) \land \text{Aluminum}(x)\}
\{x: \text{Ingots}(x) \land \text{Iron}(x)\}
\{x: \text{Ingots}(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 faceting 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


