150

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

Relationships in Knowledge Organization

Rebecca Green

OCLC Online Library Computer Center, Inc., Dewey Editorial Office,

Decimal Classification Division, Library of Congress,

101 Independence Ave. S.E., Washington, DC 20540-4330,

<greenre@oclc.org>

Rebecca Green is an assistant editor of the *Dewey Decimal Classification*. She came to this position in 2007 from the College of Information Studies, University of Maryland, where she had been a member of the faculty since 1989. Her interests include knowledge organization, database design, classification theory, knowledge representation, and linguistic applications in information retrieval. Dr. Green has degrees in computer science, library and information studies, linguistics, and music. She serves on the editorial board of *Knowledge Organization*.

ABSTRACT: Relationships that interconnect entity classes of import to knowledge organization (knowledge, documents, concepts, beings, information needs, language) include both non-subject bibliographic relationships (document-to-document relationships, responsibility relationships) and conceptual content relationships (subject relationships, relevance relationships). While the MARC format allows the recording of most bibliographic relationships, many of them are not expressed systematically. Conceptual content relationships include, in turn, interconcept and intraconcept relationships. The expression of interconcept relationships is covered by standard thesaural relationships, which typically do not distinguish fully between the underlying lexical relationship types. The full expression of complex intraconcept relationships includes indication of the basic nature of the relationship (including a set of semantic roles), the set of entities that participate in the relationship, and a mapping between participants and semantic roles. Knowledge organization schemes seldom express these relationships fully.

1. Introduction

Relationships are at the very heart of knowledge organization. Whether seeking a specific document or seeking information, users often address search systems on the basis of relationships. For example, a document may be wanted because of the subject matter it addresses, because of the person or persons who authored it, or because of its relationship (e.g., commentary, sequel) to some other document; information is sought because of its relationship to an information need. From the knowledge organization professional's perspective, many of the same types of relationships undergird organization of materials and information. This is so because of the need to organize materials and information in the same ways in which they need to be retrieved. But it is also so because without relationships there can be no organization; every organizing principle is constituted of

one or more relationships. This central role of relationships within knowledge organization is explored in Bean and Green (2001); Green et al. (2002) address relationships within the larger arena of knowledge representation and reasoning.

After discussing the general nature of relationships, the paper examines the entity classes of import to knowledge organization. Based on these entity classes, the paper then investigates, first at a high level and then at more detailed levels, the types of relationships that interconnect these entity classes in the knowledge organization context. The paper then turns its attention to the expression of these relationships in knowledge organization schemes.

2. The nature of relationships

In the broadest sense, a relationship is an association between two or more entities or between two or mo-

re classes of entities. To specify a relationship, we may first designate all the parties bound by the relationship (hereafter referred to as the participants in the relationship) and then specify the nature of any relationship that binds them together. Alternatively, starting from the other end, we may identify a relationship generically as a set of semantic roles; then for a specific relationship instantiation, we identify both the participants of the relationship and a mapping between the participants and the roles involved in the relationship.

Both starting points have merit. Many papers on semantic relations start, naturally enough, with the relationships themselves. If relationships are being discussed without regard to a specific context, that is, indeed, all one can do. Semantic relation types are thought to be universal (Murphy 2003). Because they operate on a conceptual level, they are not language-specific; further, core relationship types—such as generalization/specialization, part/whole, type/token, cause/effect—are relevant across domains. Exploring relationships by first looking at relationship types allows one to take in the greatest extent of the landscape with the least effort.

But if one is interested in relationships in the context of a specific domain, starting with the classes of entities that are important in that domain and examining their interrelationships is the only way to discover the relationships unique to the domain. Relationships involving entity classes unique to a domain and relationship types unique to the domain (the two often coincide) give a domain its character. We can't know a domain well without exploring its relationships in this way.

In exploring relationships in knowledge organization, the undertaking of both approaches is imperative. On the one hand, knowledge organization tools are used to organize the materials and information of all domains. General knowledge organization tools will rely heavily on the core relationship types relevant across domains. Starting from relationship types makes sense for such tools. On the other hand, we are also interested in knowledge organization tools for specific domains. In addition to using core relationship types, these tools should also take into account relationship types that are wholly unique or relatively unique to the domain. Further, we are specifically interested in knowledge organization as a context. To develop an understanding of relationships in knowledge organization, we must identify relationships that are unique to knowledge organization. Starting from participant types makes sense in this context.

3. Entity classes and relationships of import to knowledge organization

Thellefsen (2004, 68) sets out the ("obvious") purpose of knowledge organization as "organizing, systematizing and structuring knowledge or information in order to facilitate information retrieval." In his introductory discussion, he goes on to refer to the "retrieval of purposefully relevant documents," to "concepts [that] are complex constructs intelligibly shaped by human need to understand and create meaning of reality," and to the complementarity of concepts and their lexicalization as terms, which are central to knowledge organization. Although not writing with the intent of listing entity classes of import to knowledge organization, Thellefsen has come close to producing just such a list: knowledge/information, documents, concepts, humans, [information] needs, reality, terms. With a little massaging, the list can be transformed into a slightly more useful one:

– knowledge (also encompassing both information and reality [in that we'll assume that all knowledge reflects some reality]; both public, as conveyed by documents, and private, as held in human minds),
– documents (encompassing text, video, audio, multimedia, etc.; both intellectual units, e.g., texts, and physical units, e.g., books, serials, Web pages),
– concepts (including subjects; as with language below, encompassing both simple concepts and concepts of arbitrary complexity),
– beings (generalizing from humans and also including organizations),
– information needs (also encompassing the need for knowledge), and
– language (as with concepts above, encompassing words/terms both as single units and as utterances).

How are these entity classes interrelated? The following list is merely illustrative of the extensive set of possible relationships:

– Document <is related to (e.g., is part of, describes, cites, links to)> Document
– Document <contains> Language
– Language <expresses> Concept
– Document <is about> Concept
– Being <has> Information need
– Document <conveys> Knowledge
– Knowledge <addresses> Information need
– Knowledge <extends> Knowledge

152

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

Overall, among such relationships in the organization of knowledge, we may distinguish between document-to-document relationships and conceptual content relationships. The latter set of relationships includes both subject relationships and relevance relationships.

3.1 Entity classes and relationships in FRBR

IFLA's *Functional Requirements for Bibliographic Records* (1998; material drawn primarily from section 3) has identified three groups of entity types involved in basic bibliographic relationships, all of which correspond to and specialize one of the entity classes posited above: Group 1, Work, Expression, Manifestation, Item maps to Document; Group 2, Person, Corporate Body maps to Being; and Group 3, Concept, Object, Event, Place maps to Concept. In addition to integral units at the heart of Group 1, *FRBR* also recognizes the membership of both components (e.g., a chapter of a book, a division of a company) and aggregates (e.g., a web site as a collection of web pages, a CD as a collection of individual recordings) in Group 1.

The *FRBR* model identifies three high-level relationships that interconnect its entity classes.

– First, a set of relationships chain together the members of Group 1:
  - Work <is realized through> Expression. Example: Shakespeare's *Hamlet* has been translated into dozens of languages, has been made into at least half a dozen films, has been performed thousands upon thousands of times. Each interpretation (e.g. translation, performance) represents a different expression of the one work.
  - Expression <is embodied in> Manifestation. Example: The English text of *Hamlet* has been physically produced by numerous publishers. Each publication set represents a different manifestation of the single expression. (Note: Multiple expressions may be embodied in a manifestation, for example, an anthology. Thus, the cardinality of the Expression <is embodied in> Manifestation relationship is many-to-many.)
  - Manifestation <is exemplified by> Item. Example: A specific publication of *Hamlet* is usually mass produced; each copy of the manifestation is a separate item.

– Second, a set of "responsibility" relationships relate members of Group 1 and Group 2:
  - Work <is created by> Person / Organization. Example: Shakespeare authored *Hamlet*.
  - Expression <is realized by> Person / Organization. Example: Kenneth Branagh directed and played the title role in a 1996 film version of *Hamlet*.
  - Manifestation <is produced by> Person / Organization. Columbia TriStar Home Video distributed the Kenneth Branagh film *Hamlet*.
  - Item <is owned by> Person / Organization. A public library owns a specific copy of the Columbia TriStar Home Video distribution of Kenneth Branagh's *Hamlet*.

– Third, a set of "subject" relationships binds members of Group 1 to other members of Group 1 or to members of Group 2 or Group 3:
  - Work <has as subject> Work / Expression / Manifestation / Item. Example: A journal article is about Shakespeare's *Hamlet* / a German translation of *Hamlet* / the Riverside edition of *Hamlet* / the manuscript copy of Hamlet held by the late Mary Crapo Hyde, Viscountess Eccles
  - Work <has as subject> Person / Corporate body. Example: A film is about William Shakespeare / the Lord Chamberlain's Men.
  - Work <has as subject> Concept / Object / Event / Place. Example: A book is about revenge in Shakespeare's plays / props used in a performance of *Hamlet* / the 400th anniversary of the founding of the Globe Theatre / the Globe Theatre.

3.2 Document-to-Document Relationships

We turn now to document-to-document relationships that extend beyond *FRBR*'s relationship chain that links work to expression to manifestation to item. (*FRBR* includes these additional relationships, but does not further organize them.) In her analysis of these relationships, Tillett (2001) groups document-to-document relationships into primary relationships (these comprise the relationship chain found in FRBR), content relationships, whole-part and part-to-part relationships, and shared characteristic relationships. Included among her content relationships are equivalence relationships (e.g., Manifestation <is a reprint of> Manifestation, Item <is photocopy of> Item), derivative relationships (e.g., Work <is a translation of> Work, Work <parodies> Work), and

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

153

descriptive relationships (e.g., Work <reviews> Work, Work <provides commentary on> Work). Part-whole relationships that involve aggregates and components may occur at any of the four levels, Work / expression / manifestation / item <has part> Work / expression / manifestation / item. Examples include a single volume within a multivolume monograph, the soundtrack for a film, an article within a journal issue. Sequential (e.g., [Prequel]Work <has successor> [Sequel]Work) and accompanying relationships (e.g., Work <accompanies / complements> Work, as with a teacher's guide or workbook accompanying a text book) round out the whole-part and part-to-part relationships. Lastly, shared characteristic relationships hold between works that have any useful attribute in common, for example, language, date of publication, format.

Related to, but not encompassed by, the whole-part relationships included by *FRBR* and Tillett are (internal) text structure relationships. Many texts, such as this one, are explicitly divided into sections intended to form a logical overall text structure. Such textual units form coherent semantic units that interrelate to form a whole text and express a complete exposition of a subject in much the same way that words, properly chosen and ordered, interrelate to form a sentence and express a complete thought. Components of the overall structure yield iteratively to decompositional processes to form smaller and smaller text units. Text relationships operate on all levels of the breakdown, from macrostructures to microstructures. A vast text structure literature exists, among which Halliday and Hasan (1976) and Crombie (1985; see especially chap. 2, "Semantic relations between propositions: An outline") may be singularly mentioned for their coverage of the relationship types that undergird text structure. Such relationships make a text work, by rendering it cohesive and coherent. While some portion of text structure operates on the surface of the text (for example, sections, paragraphs, sentences), the relationships that establish cohesion and coherence are strictly semantic. As knowledge organization strives to live up to its name—organizing knowledge and information and not just containers of knowledge and information—it will need to incorporate sensitivity to text structure, without which full comprehension of a text's meaning is not possible.

Another document-to-document relationship not included in either *FRBR*'s or Tillett's treatment of bibliographic relationships is the citation relationship. Although there are many reasons why a work cites another work, of particular interest to knowledge organization is the possibility—some would say probability—that a citation reflects a subject relationship between citing and cited work. If so, we could use citation relationships either to substitute for or to supplement more conventional means (e.g., subject indexing, keyword searches) in subject searching. Several studies have investigated the correlation between citation and subject similarity. For example, Ali (1993) investigated the overlap between the words in the titles of citing and cited works, while Harter, Nisonger, and Weng (1993) examined semantic relationships between citing and cited documents by looking at the overlap between the subject descriptors assigned to them. Trivison (1987) found that documents bound by a citing relationship had a significantly higher document similarity, as measured by term co-occurrence within the titles and abstracts of the documents, than documents without any citation relationship. Unfortunately, all three studies limited their sense of semantic relationship to the co-occurrence of specific words or phrases, thus ignoring the possible occurrence of such semantic relationships as synonymy and hyponymy. In reality we know very little about the range of semantic relationships between citing and cited documents.

3.3 Conceptual content relationships

As previously seen, some document-to-document relationships are also content relationships. For example, text structure relationships—whether signaled by section, paragraph, and sentence boundaries or not—are content relationships. Likewise citation relationships are based on the conceptual content of both citing and cited documents.

But our prototypical concern with conceptual content lies in the dual challenge of assigning appropriate conceptual concept tags (e.g., a descriptor from a thesaurus, a class notation from a classification scheme) to documents and of identifying appropriate tags (e.g., words, descriptors, class notations) to search under. In such cases we are most often concerned with the semantic relationships between the tags assigned to documents and the tags used for searching. We will refer to these relationships as subject relationships. We are also concerned with conceptual content relationships between documents, without regard to the assignment of subject descriptors. We will refer to these relationships as relevance relationships.

3.3.1 Subject relationships

The single most important variable in information retrieval today is subject. Unfortunately, searching by subject is inherently difficult. For one thing, documents are almost never about a single subject only. For another thing, it is not uncommon for a user statement of need to fail to specify exactly the subject or subjects of materials that can actually help satisfy his or her need. For yet another, there is a seemingly infinite set of subjects. What keeps this situation from being utterly hopeless is the perception that all of these many subjects are related to each other and that this web of subject relationships, if understood, can help users navigate between their needs and the resources that can help meet their needs.

Typically we express concepts through the use of words. Such lexicosemantic relationships as hyponymy, meronymy, synonymy, antonymy, and series relationships (minimally, generalized to hierarchical and equivalence relationships) should thus constitute the backbone of our web of subject relationships. But beyond this well-recognized and closed class of (paradigmatic) semantic relationships is a much larger and open class of (syntagmatic) relationships, including, for example, material/product, process/product, instrument/process, and attribute/measuring tool. This class includes many of the relationships referred to in the LIS world as associative relationships; being open, the membership of the class can never be fully enumerated. These statements may be seen to imply that paradigmatic and syntagmatic relationships together exhaust the inventory of relationships and that relationships are either paradigmatic or syntagmatic. But some of the most important associative relationships, such as the dependence relationship of integrative level theory (Gnoli et al. 2007), possess some characteristics of paradigmatic relationships (the relationships are *a priori* in the world), but lack others (the words that express concepts linked by the dependence relationship often fail the substitutability condition associated with paradigmatic relationships).

A promising use of subject relationships, especially associative relationships, arises in literature-based knowledge discovery. Swanson (1986, 1990, 1993) and Davies (1989) lay out the basic premise that hitherto undiscovered knowledge may be gleaned from bringing together literatures whose subject matter (including, for example, assertions, arguments, evidence) is related in certain ways, although they are not bibliographically related through citation. Because "combinations of potentially related segments

of literature can grow at a rate far higher than the capacity of the [scholarly] community to identify and assimilate such relatedness," "the fragmentation of knowledge inevitably will spawn the most important information problems of the future, problems that also are opportunities to create new knowledge by discovering new relationships" (Swanson 1993, 606, 619). Or, as Davies (1989, 275) puts it, "every time a new item of knowledge is created there will be a vast number of potential relationships with existing items and those that prove valid … will constitute more knowledge awaiting discovery." Davies points out that Swanson's work has almost exclusively focused on knowledge discovery based on inferences from transitive (causal) relationships of the form "*A* causes *B*" and "*B* causes *C*," while other relationship types (for example, finding apparent conflict between theories and reported data) would apply to other categories of literature relatedness, potentially leading to new knowledge. Swanson (1991, 282-283) specifically lays out the logical structure of relatedness leading to one of his literature-based knowledge discoveries.

3.3.2 Relevance relationships

The relationality of relevance has long been recognized. For example, Saracevic characterizes "relevance as a measure of effectiveness between a source and a destination in a communication process. A measure is a relation. Relevance is also a relation" (1976, 91; emphasis in original omitted). In similar fashion, Lancaster and Warner (1993, 47) characterize relevance and utility in terms of relationships between a document and a user, a request, and/or a need. More specifically we may say that relevance relationships refer to the relationships between a user and his or her need, on the one hand, and those sources relevant to the need (that is, those sources of potential usefulness in the resolution of the need), on the other.

There is general consensus that the single most important aspect of the user's need and of the document within the relevance relationship is topicality, although this is by no means a sufficient condition to generate relevance. But topicality relationships do not begin to exhaust the relationality involved in relevance. A host of other criteria also contribute to (the perception of) relevance. Studies by, for example, Halpern and Nilan (1988), Nilan, Peek, and Snyder (1988), Schamber (1991), and Barry (1994) have isolated numerous factors beyond topicality that affect users' judgments of document relevance. Barry, for

instance, identified twenty-three extra-topical criteria, which she grouped into seven broad classes of criteria, namely, those pertaining to the information content of documents (e.g., scope, validity, clarity, recency), the user's previous experience and background (e.g., ability to understand, content novelty), the user's beliefs and preferences (e.g., subjective validity, affectiveness), other information and sources within the information environment (e.g., consensus, external verification), the sources of document (e.g., source quality, source reputation), the document as a physical entity (e.g., obtainability, cost), and the user's situation (e.g., time constraints, relationship with author). Each of these factors may be involved in unique types of associations relating relevant material to the user need. However, the relational aspects of these associations have not been addressed in depth.

While it has generally been agreed that a topical relationship exists between a need and the material that can (help) resolve it, discussions of topicality have often wrongly made the simplifying assumption that relatedness means sameness (Green 1995). Harter (1992, 602-603) points out the oddity of equating topical relevance with simply being on the same topic. After noting that the equation of relevance with "on the [same] topic" diverges from the everyday meaning of *relevance*, he gives two quite apt examples that hammer home the point: "That there was a drought in South Dakota in 1985 was relevant to my vacation plans there that year. Developments in computer technology are relevant to the future careers of students enrolled in schools of library and information science." Clearly the topics of the two parts of each of those examples are conceptually related, but are not the same. Indeed, the relationships involved in topical relevance may be quite complex and are likely to range across the full array of semantic relationship types (Green and Bean 1995).

4. The expression of relationships in knowledge organization tools

So far we have briefly reviewed the array of relationships of import to knowledge organization. We turn now to consider the expression of such relationships in knowledge organization tools. In doing so we will first examine the expression of relationships in bibliographic catalogs, and then we will consider the expression of conceptual relationships in other contexts.

4.1 The expression of relationships in bibliographic catalogs

Present-day bibliographic catalogs already take explicit account of some responsibility and subject relationships, but this is limited by not heeding all the distinctions that *FRBR* makes with regard to work, expression, manifestation, and item. For example, on the responsibility side, creators of works and producers of manifestations are usually recorded in specific fields of the MARC bibliographic format (Library of Congress 2000) and of specific metadata elements of the Dublin Core (2001). Persons or corporate bodies responsible for realizing an expression may or may not be recorded; however, if they are recorded, it is unlikely that they can be easily identified and retrieved automatically. Subject relationships are recorded in specific fields of the MARC bibliographic format and of the Dublin Core. MARC provides for distinguishing among person, corporate body, meeting, uniform title (which would generally correspond to a work or maybe an expression), time, space, and topic/concept; these distinctions are compatible with *FRBR* only in part. At the same time, the Dublin Core provides only for distinguishing spatial or temporal characteristics of the content from other subject relationships.

As intimated above, present-day catalogs generally do not make the same distinctions that *FRBR* does between works, expressions, manifestations, and items. At the same time, the wisdom of recognizing these distinctions and using them to collocate related bibliographic records is acknowledged by such projects as the *FRBR* Display Tool, developed within the Network Development and MARC Standards Office at the Library of Congress (2004), and a group of projects undertaken in the Office of Research at OCLC, of which FictionFinder is an exemplary representative (Vizine-Goetz 2007). *FRBR* and its companion, *Functional Requirements for Authority Data* (FRAD) (IFLA 2007), provide the conceptual models that underlie AACR2's successor, Resource Description and Access (RDA), which is due to be released in 2009 (Oliver 2007).

4.2 Expression of conceptual content relationships

Two kinds of relationships involving conceptual content exist: On the one hand, we may have relationships between two concepts; these are interconcept, paradigmatic relationships. On the other hand, we may have relationships between the components of a

156

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

single complex concept; these are intraconcept, syntagmatic relationships.

4.2.1 Expression of interconcept relationships

The expression of interconcept relationships is readily exemplified in the relational structure of a thesaurus. Typically, a thesaurus expresses hierarchical, equivalence, and associative relationships. Hierarchical relationships—or in thesaurus parlance, broader term/narrower term relationships—typically include generalization/specialization (hyponymy), type/token (instantiation) and whole-part (meronymy) relationships. Some thesauri distinguish among these subtypes of hierarchical relationships using BTG/NTG (generic), BTP/NTP (partitive), and BTI/NTI (instance) notation, while others lump them together under the all-inclusive BT/NT notation.

With equivalence relationships, synonymy (e.g., dog, canine), quasi-synonymy (e.g., lexical relationships, paradigmatic relationships), and occasionally antonymy (e.g., good, evil) are expressed by choosing one of the set of terms as an authorized descriptor and using it in lieu of all others. The authorized descriptor is used for (UF) the unauthorized terms, while entry at the unauthorized terms should direct the user to use (USE) the authorized term. The set of unauthorized terms with USE instructions constitute the lead-in vocabulary of the index language.

So-called associative relationships may be expressed as related terms (RT) of each other. Associative relationships come in a variety of flavors (e.g., product-material, process-instrument, creator-creation). Some of the relationships that underpin associative interconcept relationships also operate as intraconcept relationships.

4.2.2 Expression of intraconcept relationships

As noted previously, important aspects of specifying relationships are the designation of the participants that are bound by the relationship, some means of setting forth the semantics of the relationship (for example, by identifying a set of semantic roles), and some means of integrating the two (for example, through a mapping between the participants and the roles). We will begin our exploration of the expression of conceptual content relationships by considering how the semantics of the relationship might be expressed. As we will see, this can be done with varying degrees of explicitness.

4.2.2.1 Expression of intraconcept relationships by implication

One approach forgoes any attempt to identify the nature or semantics of the relationship, but counts on the enumeration of participants or participant types to imply an underlying relationship. The assumption is that only one sensible relationship binding the participants together exists, or if more than one sensible relationship exists, one relationship type is more likely to occur than others and this greater likelihood is accessible to the human or machine making use of the relationship.

The keyword access typical of modern search engines is of this nature; indeed much full text searching is of this nature. Search failures in this scenario can be ascribed to many underlying reasons (for example, the inability to disambiguate which word senses are intended, ignorance of search conventions); one reason that is seldom fully appreciated is the failure of the presumption that only one relationship exists between or among (the concepts behind) the search terms. This approach generally expresses only the participants of the relationship; since the nature of the relationship is only implied, no mapping between participants and roles can be expressed.

This approach is also evident, for example, in the typical use of standard subdivision—08 in the *Dewey Decimal Classification* (*DDC*) (Chan and Mitchell 2003, 91), which is used to show treatment of a topic with respect to groups of people. In most situations the role played by the group of people with respect to the topic is only implied.

The approach is also evident in most Library of Congress subject headings that use subdivisions. For example, in the subject heading Detergents—Biodegradation, we are left to infer that the topic is the biodegradation of detergents and not the biodegradation of something else as facilitated by detergents.

4.2.2.2 Expression of intraconcept relationships by convention

Closely related to this approach is a second one in which a relationship type between entity classes holds by convention. Beghtol (2001, 101-102) notes, for example that in Ranganathan's PMEST formula, the S(pace) and T(ime) facets are understood to situate the topic of the writing as taking place in the specified Space and Time and not to situate the actual writing of the document, for example, in that

Space and Time. This approach is also evident in some natural language use, as when the relationships intended by a noun-noun phrase (e.g., dog food, book cover) are built into the phrase's meaning: *Dog food* means food for dogs, not food made of dogs; *book cover* means the cover of a book, not the use of a book for cover. Again, since this approach generally expresses only the participants of the relationship and the nature of the relationship is only implied, no mapping between participants and roles can be expressed.

4.2.2.3 Expression of intraconcept relationships by naming

A third approach simply names the relationship type that holds—which has been the practice of this paper, e.g., Document < is part of > Document. (The actual set of roles that comprise the relationship are not enumerated, and so, once again, a mapping between participants and roles is not made explicit.) This approach makes the assumption that the user is familiar with the relationship type ("< is part of >") through his or her own personal experience and can access an understanding of the semantics of the experience on the basis of a natural language label. Natural language usage that makes conceptual relationships explicit (for example, through the use of prepositional phrases) exemplifies this approach, as can be seen in the Library of Congress subject heading, Midwives—Supervision of. It is clear in this subject heading that midwives are being supervised and not doing the supervision.

We noted previously that the use of standard subdivision —08 in the *DDC* typically exemplifies approach 1. However, there are times when the classification scheme defines the role of the groups of people with respect to the topic. For example, the schedules explicitly indicate that the standard subdivision —08 within 253 Pastoral office and work (Pastoral theology) is to be used for pastoral care performed *by* specific groups of people and not for pastoral care *of* specific groups of people.

4.2.2.4 Formal expression of intraconcept relationships

The most explicit approach uses a formal language to set forth the semantics of the relationship type. Such an approach will commonly be found in knowledge representation systems that incorporate relationship types. Ultimately the symbols of that formal language must be transformed into understanding on the basis of personal experience and probably using the medium of natural language. The third and fourth approaches are thus closer than they may appear on the surface. However, this fourth approach is the only approach that, with its formal expression of relationship semantics, sets out the roles that comprise the relationship and that can express a mapping between participants and roles.

We should note that the identification of semantic roles does not, by itself, ensure the full expression of an intraconcept relationship. Let us consider, for example, the topic, the sale of ethanol to the United States by Brazil, and assume the use of the following relatively standard set of thematic roles: agent, counteragent, theme, result, instrument, source, goal, experiencer, beneficiary, possessive, time, locative (Fillmore 1968). The intraconcept relationship could be (partially) expressed as follows: Brazil (source), United States (goal), ethanol (theme). The first issue is the absence of some mechanism to link together exactly these three topic-(role) pairs and none others. Even after such a mechanism is provided, we have yet to specify the basic nature of the relationship, that of a sale, and, not for instance, that of a gift. The same issues arise in the (limited) use of role indicators made by *MeSH* (*Medical Subject Headings*) (e.g., Acupuncture /Adv[erse] eff[ects]). Faceted thesauri and classification schemes typically provide the necessary mechanism to link together all interrelated concepts in a single unit, but often fail to specify the nature of the relationship except at a general overall subject level.

Semantic frames provide a possible solution to the intraconcept relationship expression conundrum. A frame is a representational structure that includes a label (which names the general nature of the relationship) and a set of frame elements or slots (which identify the semantic roles in the relationship). The slots are filled by descriptors (thus providing a mapping between participants and semantic roles). A frame representation of the ethanol sale might be as follows:

COMMERCIAL EVENT:
    buyer   [United States]
    seller   [Brazil]
    goods   [ethanol]
    money   [ ]

158

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

5. Conclusion

Relationships are at the heart of knowledge organization: We attempt to locate information that relates to a user's need; we attempt to locate documents containing words that might be used in conveying the sought-for information; and/or we retrieve documents indexed by a relevant subject descriptor or classed in a relevant subject category. We use relationships to navigate among subjects. We help to locate documents that together say more than the sum of their parts.

Despite the centrality of relationships, their expression in knowledge organization schemes seldom rises to full and systematic expression. To the disappointment of the researcher, studies often show that the use of deeper semantics, as would be provided by the full and systematic expression of relationships, does not in fact improve retrieval effectiveness. And yet, in our everyday lives the full expression of relationships is often crucial to effective comprehension, reasoning, and communication. We are left with the impression that we are "missing something." Either we are not implementing relationships properly, or we don't understand the fundamental properties of the context in which we are working—or perhaps both!

References

Ali, S. Nazim. 1993. Subject relationship between articles determined by co-occurrence of keywords in citing and cited titles. *Journal of information science* 19: 225-32.

Barry, Carol L. 1994. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science* 45: 149-59.

Bean, Carol A. 1996. Analysis of non-hierarchical associative relationships among Medical Subject Headings (MeSH): Anatomical and related terminology. In Green, Rebecca, ed., *Knowledge organization and change: Proceedings of the Fourth international ISKO Conference 15-18 July 1996 Washington, DC.* Würzburg: Ergon, 80-86.

Bean, Carol A. and Green, Rebecca. 2001. *Relationships in the organization of knowledge*. Dordrecht: Kluwer Academic Publishers.

Beghtol, Clare. 2001. Relationships in classificatory structure and meaning. In Bean, Carol A. and Green, Rebecca, eds., *Relationships in the organization of knowledge.* Dordrecht: Kluwer Academic, 99-113.

Chan, Lois Mai and Joan S. Mitchell. 2003. *Dewey Decimal Classification: principles and application*. Dublin, Oh.: OCLC.

Crombie, Winifred. 1985. *Process and relation in discourse and language learning.* London: Oxford University Press.

Davies, Roy. 1989. The creation of new knowledge by information retrieval and classification. *Journal of documentation* 45: 273-301.

Dublin Core Metadata Element Set: An American National Standard. 2001. National Information Standards Organization (NISO). Available at http://www.techstreet.com/cgi-bin/pdf/free/335284/Z39.85-2001.pdf.

Fillmore, Charles J. 1968. The case for case. In Bach, Emmon W. and Harms, Robert T., eds., *Universals in linguistic theory*. New York: Holt, Rinehart, and Winston, 1-88.

Gnoli, Claudio, Bosch, Mela, and Mazzocchi, Fulvio. 2007. A new relation for multidisciplinary knowledge organization systems: dependence. In Rodríguez Bravo, Blanca and Alvite Díez, M.a Luisa, eds., *La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico = Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge: actas del VIII Congreso ISKO-España: León, 18, 19 y 20 de Abril de 2007*. León: Universidad de León Secretariado de Publicaciones, 399-409.

Green, Rebecca. 1995. Topical relevance relationships: why topic matching fails. *Journal of the American Society for Information Science* 46: 646-53.

Green, Rebecca and Bean, Carol A. 1995. Topical relevance relationships: an exploratory study and preliminary typology. *Journal of the American Society for Information Science* 46: 654-62.

Green, Rebecca, Bean, Carol A., and Myaeng, Sung Hyon. 2002. *The semantics of relationships: an interdisciplinary perspective*. Dordrecht: Kluwer Academic Publishers.

Halliday, M. A. K. and Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Halpern, David and Nilan, Michael. 1988. A step toward shifting the research emphasis in information science from the system to the user: an empirical investigation of source-evaluation behavior information seeking and use. In Borgman, Christine L. and Pai, Edward Y. H., eds., *ASIS '88 proceedings of the 51st ASIS Annual Meeting, Atlanta, Ga., October 23-27, 1988.* Medford, N.J.: Published for the American Society for Information Science by Learned Information, Inc., 169-76.

Knowl. Org. 35(2008)No.2/No.3
R. Green. Relationships in Knowledge Organization

159

Harter, Stephen. P. 1992. Psychological relevance and information science. *Journal of the American Society for Information Science* 43: 602-15.

Harter, Stephen. P., Nisonger, Thomas E,. and Weng, Aiwei. 1993. Semantic relationships between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science* 44: 543-52.

IFLA Study Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2007. *Functional requirements authority data: A conceptual model.* Draft 2007-04-01. Available at http://www.ifla.org/VII/d4/FRANAR-ConceptualModel-2ndReview.pdf.

IFLA Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional requirements for bibliographic records: final report.* München: K.G. Saur. Available at http://www.ifla.org/VII/s13/frbr/frbr.pdf.

Lancaster, F. Wilfrid and Warner, Amy J. 1993. *Information retrieval today.* Arlington, Va.: Information Resources Press.

Library of Congress. 2000. *MARC 21 concise format for bibliographic data.* [Washington, D.C.]: Library of Congress, Network Development and MARC Standards Office. Available at http://www.loc.gov/marc/bibliographic/.

Library of Congress. Network Development and MARC Standards Office. 2004.

FRBR display tool. Version 2. Available at http://www.loc.gov/marc/marc-functional-analysis/tool.html.

Murphy, M. Lynne. 2003. *Semantic relations and the lexicon antonymy, synonymy, and other paradigms.* Cambridge, UK: Cambridge University Press.

Nilan, Michael S., Peek, Robin. P., and Snyder, Herbert W. 1988. A methodology for tapping user evaluation behaviors: an exploration of users' strategy, source and information evaluating. In Borgman, Christine L. and Pai, Edward Y. H., eds., *ASIS '88 proceedings of the 51st ASIS Annual Meeting, Atlanta, Ga., October 23-27, 1988.* Medford, N.J.: Published for the American Society for Information Science by Learned Information, Inc., 152-59.

Oliver, Chris. 2007. Changing to RDA. *Feliciter* 53:250-253. Available at http://www.collectionscanada.gc.ca/jsc/docs/felicitervol53no7p250-253.pdf.

Saracevic, Tefko. 1976. Relevance: a review of the literature and a framework for thinking on the notion in information science. *Advances in librarianship* 6: 79-138.

Schamber, Linda. 1991. *Users' criteria for evaluation in multimedia information seeking and use situations.* Ph.D. dissertation, Syracuse University.

Swanson, Don R. 1986. Undiscovered public knowledge. *Library quarterly* 56: 103-18.

Swanson, Don R. 1990. The absence of co-citation as a clue to undiscovered causal connections. In Borgman, C. L.. ed., *Scholarly communication and bibliometrics.* Newbury Park, Calif.: Sage, 129-137.

Swanson, Don R. 1991. Complementary structures in disjoint scientific literatures. In *SIGIR '91,* proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development. New York: ACM, 280-9.

Swanson, Don R. 1993. Intervening in the life cycles of scientific knowledge. *Library trends* 41, no.4: 606-31.

Thellefsen, Martin. 2004. Concepts and terminology reflected from a LIS perspective. How do we reflect meanings of concepts? In *Knowledge and change*: *Proceedings of the 12th Nordic Conference on Information and Documentation, Aalborg, Denmark, September 1-3, 2004*, 68-75. Available at http://www2.db.dk/NIOD/thellefsen.pdf.

Tillett, Barbara B. 2001. Bibliographic relationships. In Bean, Carol A. and Green, Rebecca, eds., *Relationships in the organization of knowledge.* Dordrecht: Kluwer Academic, 19-35.

Trivison, Donna. 1987. Term co-occurrence in cited/citing journal articles as a measure of document similarity. *Information processing & management* 23:183-94.

Vizine-Goetz, Diane. 2007. FictionFinder: imagining what a "next generation" catalog might look like." Presentation given at the OCLC Members Council Research and New Technologies Interest Group meeting, 6 February 2007, Québec, QC (Canada) and Dublin, Ohio (USA). Available at http://www.oclc.org/research/memberscouncil/2007-02/vizine-goetz.pdf [The FictionFinder prototype itself is available at: http://fictionfinder.oclc.org/].