Furner, J. (2002). Bibliographic Relationships, Citation Relationships, Relevance Relationships, and Bibliographic Classification: An Integrative View. 13th ASIS SIG/CR Classification Research Workshop, 29-37. doi:10.7152/acro.v13i1.13833

Bibliographic Relationships, Citation Relationships, Relevance Relationships, and Bibliographic Classification: An Integrative View

Jonathan Furner
Department of Information Studies
University of California, Los Angeles
jfurner@ucla.edu

1. INTRODUCTION

In the interests both of theoretical consistency and of practical potential, it is suggested that citation relationships and relevance relationships of certain kinds --- ordinarily considered separately from (topic-oriented) bibliographic classification in general, and from bibliographic relationships in particular --- would more profitably be treated by the classification-research community as (in Tillett's terminology) shared-characteristic relationships. On this basis, designers of library catalogs should work to offer users the opportunity to exploit, in an integrated manner, all sources of evidence of document relatedness.

2. BIBLIOGRAPHIC RELATIONSHIPS

In the library cataloging and classification research communities, the term "bibliographic relationship" (BR) is commonly used to refer to a relationship between two or more bibliographic entities or their representations (Green, 2001), such entities in turn being identifiable variously as works, expressions, manifestations, or items, following terminology formalized by IFLA's Study Group on the Functional Requirements for Bibliographic Records (IFLA, 1988). Informally, the term "document" is often used imprecisely to refer to individual bibliographic entities of different kinds.

In Tillett's classic formulation (1987; 1991; 2001), several types of BR are identified. These include the following: shared-characteristic (S-C) relationships, which link multiple entities observed to share a common author, title, subject, or other property; derivative relationships, each of which links a source entity to an entity observed to be a modification of that source; and descriptive relationships, each of which links a source to an entity that describes or evaluates that source.

3. TOPICAL, CITATION, AND RELEVANCE RELATIONSHIPS

As a practical tool for modeling the structure of library catalogs --- both actual and proposed --- Tillett's taxonomy of BRs, and subsequent versions and revisions of it, have proved very useful (Leazer, 1993; Carlyle & Fusco, 2002). But the scope of these schemes has largely remained deliberately narrow, in two respects.

In the first place, BR taxonomies focus on those relationships that may be identified through analysis of the kind that is undertaken by descriptive catalogers. Such relationships are commonly considered to be objective and permanent, and their identification is a matter on

ISSN: 2324-9773

which empirical evidence (of certain, typically uncontested, kinds) may convincingly be brought to bear. The S-C relationship holding between two works by the same author, and the descriptive relationship holding between a work and a review of that work, are two examples.

There is, however, another class of BRs that, in contrast, has received rather less emphasis in BR taxonomies. Subject analysis, of the kind that is normally undertaken in the course of bibliographic classification, is the source of expressions of relationships among bibliographic entities that are subjective, temporary, and the product of personal preference. The S-C relationship holding between two works judged to be about the same topic is an obvious example.

Moreover, there are other productive sources of expressions of such relationships --- sources other than indexers, subject catalogers, and classifiers, that is. The authors of works are also prolific identifiers of relationships between the bibliographic entities that they personally create and those produced by other authors. These relationships are commonly expressed in the form either of citations (in scholarly works) or of hyperlinks (in web pages) (Borgman & Furner, 2002).

In the second place, BR taxonomies focus on those relationships that hold between bibliographic entities, strictly-defined as documents or their representations. People, organizations, topics, places, etc., are treated by the schemes only as attributes of entities, and not as entities that themselves may be considered at the same level as, and directly relative to, documents. Thus, relationships of the kind that may be identified as holding between a reader or information-seeker and a document --- or between a particular expression of an information need and a document --- are not usually incorporated in BR taxonomies. Relationships of this latter type have instead been studied separately as "relevance relationships" (Bean & Green, 2002), to contrast with BRs and citation relationships.

It may be argued, however, that to continue this separate treatment of bibliographic, relevance, and citation relationships is to hamper the very improvement of catalog design that was intended by the original taxonomists of BRs. It is instead suggested (a) that certain kinds of topical indexing, relevance, and citation relationship are more appropriately considered as sub-types of S-C BRs; and (b) that, indeed, indexing, relevance, and citation relationships may be treated in general as types of BRs. To develop this argument, it is necessary to clarify the conceptual connection between BRs and bibliographic classification.

4. BIBLIOGRAPHIC CLASSIFICATION

Bibliographic classification is typically defined quite simply, as the grouping together of bibliographic entities that are alike or similar in some respect. It is undertaken through analysis of the properties of those entities: entities that share the same properties, that are similar in respect of those properties, are identified as members of the same class.

It will also be clear that, in assigning a class label to a given bibliographic entity, the classifier is recording their identification of a relationship between that entity and all others that have previously been, or will subsequently be, similarly classified. In this sense, the act of classification may be interpreted as the identification and/or expression of the relationships that

ISSN: 2324-9773

hold between entities, such that related entities may be collectively identified as the members of a given class. In the prevailing Aristotelian perspective (pace "prototype" theory), the members of any single class may be said to be linked by a particular "shared-characteristic" relationship, in virtue of their sharing a common property. If we accept this view, that the act of identifying relationships between bibliographic entities is a core component of the process of bibliographic classification, we will surely conclude that the study of BRs is inseparable from the study of bibliographic classification.

The aim of bibliographic classification, the fundamental reason or motivation for embarking on the task, is to allow information-seekers to access and retrieve entities at the class level. Imposition of a classified structure on a set of entities allows information-seekers to browse among classes, and to base their decisions to access individual entities on their perceptions of the degree to which such browsing activity is successful.

Given this statement of the aim of bibliographic classification --- a statement that focuses on the instrumental quality of classification as a tool for improving the performance of retrieval systems --- it may further be concluded that the topic or subject of a bibliographic entity is not necessarily the only property on whose basis that entity may be classified.

The products of descriptive cataloging --- assertions of more-or-less objective properties of a bibliographic entity and of the process by which it was created, such as the name of its creator, the date and place of its creation, and so on --- may additionally be considered as class labels that may themselves be structured with a view to supporting information-seekers' navigation of the entity set. One of the primary contentions of those who promote the exploitation of bibliographic relationships in library catalogs is that descriptive cataloging data be used in this way.

There are properties of yet other kinds, however, on which the classification of bibliographic entities may be based, although consideration of these tends to be pushed to one side in discussions of the role that may be played by BRs in support of information retrieval.

5. INDEXERS, CITERS, AND RELEVANCE JUDGES

One way to highlight the distinctive nature of these other properties is to focus on the identity of the classifying agent. In the cases of statements of the kind found in conventional library catalogs --- statements of the topic of a document, and of its physical and historical characteristics --- the classifying agent may normally be identified as a cataloger or indexer, responsible for the description of multiple documents of varying origin.

The citations and hyperlinks produced by the authors or creators of documents may also be viewed as properties --- i.e., properties both of citing (source) and cited (target) documents --- that may be analyzed in order to classify such documents. Just as indexers assign documents to terms or headings that represent the topical classes to which those documents are determined by the indexers to belong, authors may be seen to assign cited documents to citing documents that represent the classes (or "citation contexts") to which the cited documents are determined by the authors to belong. In the former case, indexed documents are represented by subject headings

ISSN: 2324-9773

(and vice versa); in the latter, cited documents are represented by citing documents (and vice versa).

In a third case, the relevance judgments produced by the readers of documents may also be interpreted as properties on which classification may be based. We may consider that documents can be assigned to readers on the basis of the degree of approval for those documents expressed (implicitly or explicitly) by those readers, and that readers' information needs (expressed or unexpressed) may thus be taken to represent the classes to which the documents are determined, through examination of the history of judgments of approval, to belong.

6. CO-CITATION, CO-RELEVANCE, AND CO-INDEXING

Another way to focus on some interesting analogies between the processes of indexing, relevance judgment, and citation, is to develop a more formal model of the elements involved.

Given (i) two bibliographic entities, a and b, related in the sense that they are both perceived to be members of the same class x, and (ii) the class x, of which the entities are perceived to be members, there are at least two ways in which the inter-entity relationship may be modeled. The choice of method will likely depend on the nature of the entities in question, and the purpose of the modeler.

On the one hand, we may decide to emphasize the single action of a judge j, who, at time t, expresses the opinion that entity a is related to entity b, with respect to class x. The membership of class x enjoyed by each entity is inferred from the judge's statement of the direct relationship between the two entities. For example: we may wish to describe the decision of citer j that cited document a is related to citing document b, with respect to context x; or the decision of information-seeker j that document a is related to query b, with respect to information need x; or the decision of indexer j that document a is related to subject heading b, with respect to topic x (see Fig. 1). We may refer to relationships of the first kind as citation relationships; of the second kind as relevance relationships; and of the third kind as indexing relationships.

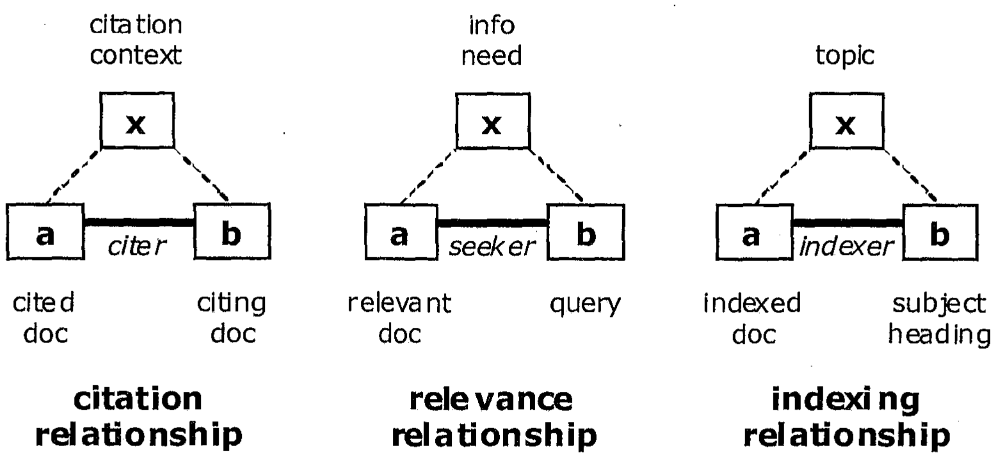
Any given relationship of one of these kinds may be considered as an event and defined by a quintuple, specifying the two entities a and b, the class x, the judge j, and time t. In this model, it should be noted that any specification of a relationship that is missing a single one of these five values is incomplete. This formalizes the notion, for example, of topicality or aboutness as an event rather than as an inherent property of a particular document.

It will also be noted that the definition of bibliographic entity used here is somewhat broader than that typically used in discussions of BRs. Here, subject headings and queries, as well as documents, are treated as bibliographic entities; in this model, indeed, any statement of content, any representation or container of a message, would be treated thus.

ISSN: 2324-9773

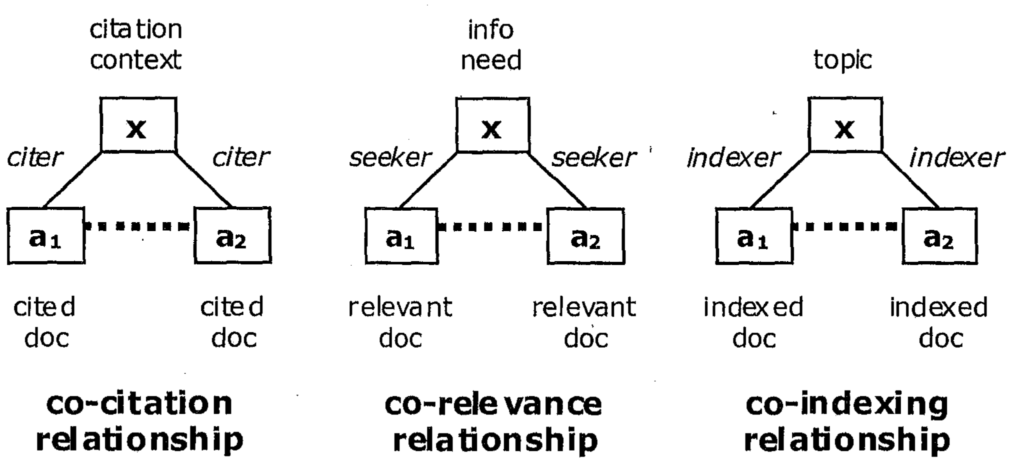
Furner, J. (2002). Bibliographic Relationships, Citation Relationships, Relevance Relationships, and Bibliographic Classification: An Integrative View. 13th ASIS SIG/CR Classification Research Workshop, 29-37. doi:10.7152/acro.v13i1.13833

Figure 1. Citation, relevance, and indexing relationships.



On the other hand, we may decide to emphasize the two actions, carried out on separate occasions either by a single judge or by two different judges, in which two entities a_1 and a_2 are each reported to be members of class x. In this case, the relationship between the two entities is inferred from the judges' statements of class membership. For example: we may wish to describe the decision of citer(s) j that documents a_1 and a_2 are both worthy of citation in context x; or the decision of information-seeker(s) j that documents a_1 and a_2 are both relevant to information need x; or the decision of indexer(s) j that documents a_1 and a_2 are both about topic x (see Fig. 2). We may refer to the inferred relationships of the first kind as co-citation relationships; of the second kind as co-relevance relationships; and of the third kind as co-indexing relationships.

Figure 2. Co-citation, co-relevance, and co-indexing relationships.



What is the characteristic that is "shared" by documents a_1 and a_2 in each of these cases? We might argue that, strictly speaking, the shared characteristic is not a property of the documents themselves, but of events in which they have both been involved. Imagine we have two events

ISSN: 2324-9773

defined respectively by $\langle \underline{a}_1,\underline{b}_1,\underline{x}_1,\underline{j}_1,\underline{t}_1 \rangle$ and $\langle \underline{a}_2,\underline{b}_2,\underline{x}_2,\underline{j}_2,\underline{t}_2 \rangle$. If we can determine that $\underline{x}_1 = \underline{x}_2$, then we can conclude that $\underline{a}_1$ and $\underline{a}_2$ are related on that basis. But $\underline{x}_1$ and $\underline{x}_2$ are characteristics of events not of documents: they owe their existence to the actions of particular people at particular times. Two co-cited documents are deemed to be citation-worthy in the same context (e.g., in the same citing document); two co-relevant documents are judged to be relevant to the same information need; two co-indexed documents are determined to be about the same topic.

On the basis of the equivalencies depicted here, we suggest (a) that relationships of each of these last three kinds --- co-citation, co-relevance, and co-indexing relationships --- may be treated, in the terminology of BR taxonomies, as S-C relationships. Moreover, (b) since citation relationships, relevance relationships, and indexing relationships are themselves identifiable as relationships between bibliographic entities (defined broadly), it might seem sensible to treat these, too, as BRs (albeit not as instances of the S-C sub-type). At the very least, we might argue for explicit recognition both of the basis on which the three kinds of relationship may be said to be equivalent, and of the common means by which we may derive comparable S-C relationships from each of them.

7. INTEGRATING SOURCES OF EVIDENCE OF DOCUMENT RELATEDNESS: A RESEARCH AGENDA

So far, we have argued that co-indexing, co-relevance, and co-citation relationships may be treated as S-C relationships. But why should they be? One way of justifying this normative claim would be to continue to appeal to rationalist principles, and to argue for the logic, consistency, completeness (and so on) of broadening the scope of S-C relationships in the suggested manner. A different approach would be to take an instrumentalist perspective, and to point to empirical evidence demonstrating the improved performance of retrieval systems that are designed to treat S-C relationships more broadly. To obtain such evidence, we might be moved to conduct an experiment, testing the hypothesis that retrieval systems which offer the information-seeker the opportunity to exploit knowledge of S-C relationships of multiple kinds carry out their functions more successfully than those which do not. Such an experiment would be situated in the tradition of IR evaluation that began with the precursors to the Cranfield tests of the 1960s. Tests of this kind have often been criticized for the low degree to which they are perceived to successfully simulate the interaction, between seeker and system, that is a defining characteristic of searches for information; recent experimental designs, however, have taken more appropriate account of seekers' decision-making during search sessions.

On these lines, we may suggest an agenda for further research as follows:

1. Build a retrieval system that treats co-indexing, co-relevance, and co-citation relationships as S-C relationships. What would such a system look like? It would be one that offers functionality of the following kinds (inter alia):

(a) creation, storage, and maintenance of a collection of documents;

(b) elicitation of query documents, through the provision of an interface supporting the information-seeker's activity of query-formulation;

ISSN: 2324-9773

Furner, J. (2002). Bibliographic Relationships, Citation Relationships, Relevance Relationships, and Bibliographic Classification: An Integrative View. 13th ASIS SIG/CR Classification Research Workshop, 29-37. doi:10.7152/acro.v13i1.13833

(c) identification of the manually-generated expressions of indexing, citation, and relevance relationships perceived to exist among documents, classes, and judges; this would involve:

- identification of metadata fields containing classification codes, subject headings, index terms, descriptors, etc.;

- automatic identification of the occurrence, within stored documents, of citations and/or sources of hypertext links, and the automatic linking of citing to cited documents; and

- automatic identification of the decisions --- explicit and implicit --- made by relevance judges;

(d) creation and storage of vectors representing, for each document, the indexing, citation, and relevance relationships identified in (c);

(e) analysis of the vectors created in (d), in order to identify the occurrence and strength of co-indexing, co-citation, and co-relevance relationships;

(f) ranking of the documents in the collection on the basis of the degree to which each is co-indexed, co-cited, co-relevant, etc., with each query document;

(g) presentation to the information-seeker of multiple rankings of documents, obtained by the various methods in response to each query; and

(h) recording of the judgments (provided by the information-seeker or by subject experts) of the relevance of documents in presented rankings.

2. Evaluate the performance of such a system. This might be done in roughly the following manner:

(a) setting the system up in direct comparison with a number of other systems that do not exploit S-C relationships of multiple kinds;

(b) observing, and recording representations of, the use that is made of all systems;

(c) analyzing these data, to produce values for metrics (e.g., recall and precision) that are understood to indicate system performance; and

(d) comparing these values.

Specifically, the judgments obtained from the information-seeker in 1.(h) may be analyzed in order to determine the extent to which each automatic ranking method successfully predicts the information-seeker's personal preference ranking. Thus, we may determine not only (i) whether the provision of multiple ranking methods, based on the analysis of multiple kinds of S-C

ISSN: 2324-9773

relationships, improves on the retrieval performance of a system that exploits only co-indexing relationships, but also (ii) the extent to which any individual ranking method outperforms another. If the performance of particular ranking methods is observed to correlate with aspects of the retrieval situation such as the purpose for which information is being sought, the topic of the information need, the style of the searcher --- or with any other characteristic of the information-seeker or information need --- then it may turn out to be possible to suggest how the selection of optimum ranking method may be made automatically by the system. An alternative, however, would be to recognize the importance of allowing the human user to maintain control over this selection act, and simply to provide the searcher with the opportunity to exploit as many different sources of evidence of relevance as possible.

The activity of identifying documents' class memberships --- by groups of agents in various roles, e.g., citers, relevance judges, and indexers --- may be viewed both as the source of evidence of the degree to which documents are related to each other, and as the source of evidence of the degree to which documents are relevant to individual information-seekers. The computational techniques, reliant on graph theory and matrix algebra, that are implemented in retrieval systems so that such evidence may be exploited by the system user are well-understood, and can be modeled in such a way that the correspondences between classification by context (citing), classification by approval history (relevance rating), and classification by content (indexing), are highlighted (Furner, 2002). Yet conventional library catalogs are designed to exploit one in particular of these sources of evidence --- that supplied by professional catalogers. Once the correspondence in function of relationships of co-citation, co-relevance, and co-indexing (and others of similar derivation) is recognized, the prospect arises of hybrid systems that provide information-seekers with the opportunity not only to exploit multiple sources of evidence of document relevance in an integrated environment, but also to maintain control over the ways and combinations in which such sources are exploited in any given context. It may at least be hypothesized that all S-C relationships are worthy of exploitation in support of effective information retrieval in the library catalog, just as they are recognized to be so in recommender systems such as Amazon.com and e-print retrieval systems such as NEC's CiteSeer.

REFERENCES

Bean, C. A., & Green, R. (2001). Relevance relationships. In C. A. Bean, & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 115-132). Dordrecht, The Netherlands: Kluwer.

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual review of information science and technology: Vol. 36* (pp. 3-72). Medford, NJ: Information Today.

Carlyle, A., & Fusco, L. M. (2002). Equivalence in Tillett's bibliographic relationships taxonomy: A revision. In M. J. López-Huertas (Ed.), *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference* (Granada, Spain, July 10-13, 2002; pp. 258-263). Würzburg, Germany: Ergon.

Furner, J. (2002). A unifying model of document relatedness for hybrid search engines. In M. J. López-Huertas (Ed.), *Challenges in knowledge representation and organization for the*

ISSN: 2324-9773

21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference (Granada, Spain, July 10-13, 2002; pp. 245-250). Würzburg, Germany: Ergon.

Green, R. (2001). Relationships in the organization of knowledge: An overview. In C. A. Bean, & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 3-18). Dordrecht, The Netherlands: Kluwer.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1988). *Functional requirements for bibliographic records: Final report.* München, Germany: K. G. Saur.

Leazer, G. H. (1993). *A conceptual plan for the description and control of bibliographic works.* Unpublished D.L.S. dissertation, Columbia University, New York, NY.

Tillett, B. B. (1987). *Bibliographic relationships: Toward a conceptual structure of bibliographic information used in cataloging.* Unpublished Ph.D. dissertation, University of California, Los Angeles, CA.

Tillett, B. B. (1991). A taxonomy of bibliographic relationships. *Library Resources & Technical Services, 35*: 150-158.

Tillett, B. B. (2001). Bibliographic relationships. In C. A. Bean, & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 19-35). Dordrecht, The Netherlands: Kluwer.

ISSN: 2324-9773