Image descriptions and their relational expressions: a review of the literature and the issues

Allen C. Benson
U.S. Naval War College, Newport, Rhode Island USA

Received 16 July 2013
Revised 6 January 2014
Accepted 7 January 2014

Abstract

Purpose – The purpose of this paper is to survey the treatment of relationships, relationship expressions and the ways in which they manifest themselves in image descriptions.

Design/methodology/approach – The term "relationship" is construed in the broadest possible way to include spatial relationships ("to the right of"), temporal ("in 1936," "at noon"), meronymic ("part of"), and attributive ("has color," "has dimension"). The intentions of these vaguely delimited categories with image information, image creation, and description in libraries and archives is complex and in need of explanation.

Findings – The review brings into question many generally held beliefs about the relationship problem such as the belief that the semantics of relationships are somehow embedded in the relationship term itself and that image search and retrieval solutions can be found through refinement of word-matching systems.

Originality/value – This review has no hope of systematically examining all evidence in all disciplines pertaining to this topic. It instead focusses on a general description of a theoretical treatment in Library and Information Science.

Keywords Library systems, Archives, Indexing, Knowledge organizations

Paper type Literature review

Introduction

This review attempts to lay down the interesting intellectual activity evident in the literature that concerns relationships and their role in organizing information, especially visual information expressed in photograph descriptions. It negotiates through both the generalities and specifics as they pertain to how people describe and represent relationships in photograph descriptions within the context of library information systems. It considers, in Mitchell's (1994) terms, the ways in which image content in photographs may be turned into language.

There is a significant literature devoted to naming and categorizing image attributes and visual primitives, but few empirical studies try to account for the relationships humans express when describing visual information, so the potential for Library and Information Science (LIS) contributions to this debate could be substantial. This fundamental gap in the literature can be attributed in part to the influence of well-established bibliographic standards and the traditional approach to information seeking problems as indexing and subject analysis problems. Michael Krause illustrates this point when he describes how indexers locate the meaning of a photograph within a catalog by breaking apart its meaning and assigning the photograph one word or very few words as authorized headings called entry points (Krause, 1988).

Characterizing the various semantic functions performed by relationships depicted in pictorial content and expressed in text is a subject of much debate. There are a number of relationships that we intuitively express when we describe photographs. A description that states, "digital countdown clock partially visible in upper left corner

Journal of Documentation
Vol. 71 No. 1, 2015
pp. 143-164

0022-0418
DOI 10.1108/JD-07-2013-0093

of photo," refers to a spatial relationship between the depiction of a clock and its location on the two-dimensional surface of a print. Some of the things we say concerning photographs express extrinsic relationships. For instance, "located in the National Archives & Records Administration, Still Picture Branch, Panama Canal Collection Series" denotes the photograph's relationship to a collection and to a geographic location in Washington, DC. Rather than express what we know about photographs we can describe what their images denote, for example, "star in window denoting the home of serviceman." In addition to descriptions of what relationships in visual information refer to and denote we classify relationships by reference to their semantic properties. The predicate during in "Outside the Farm Security Administration grant office during the pea harvest," is a type of temporal interval and the predicate taken on in "taken on May 21, 1936" describes a temporal moment (Allen, 1984).

The purpose of this paper, therefore, is to give an overview of the relationship problem as it has been examined in the LIS literature. The focus of the review is on semantic relationships that we express in our descriptions of photographs applied to visual information and linguistic representations. The following two modes of description are proposed as a framework for guiding and limiting this discussion of the literature: machine-readable description and phenomenal description.

This paper proceeds as follows: image content and relationships in the context of machine-readable descriptions is defined and examined in three studies. The review then turns to phenomenal description focussing on nine research papers concerned with image searcher behavior. The review closes with a discussion of the findings and conclusions.

Machine-readable description

Structuring visual information descriptions in machine-readable records is motivated by a desire to make photographs and photographic image data accessible and retrievable in online library systems. The phrase "machine-readable" refers to bibliographic and other information that is structured, represented, and communicated in MARC format. MARC, an acronym for MAchine-Readable Cataloging, defines a data format used for building the bibliographic records found in most library catalogs. It is a specification that enables computers to interpret, use, and exchange the bibliographic information found in cataloging records. While this mode of description may be described as readable by machines and humans, predicates and word senses represented in bibliographic database records are only meaningful to humans.

Machine-readable descriptions typically involve formalized standards that fall into one of three categories: data structure, data content, or data values (Walch and Matters, 1994). Well-known examples include the Library of Congress Subject Headings, controlled vocabularies such as the Library of Congress Thesaurus for Graphic Materials and metadata standards such as the Dublin Core Metadata Initiative. Librarians adopt traditional subject cataloging, descriptive cataloging, and finding aids practices when describing image content in this manner.

The main focus of research in this category of description has been on bibliographic entities, but there are some studies that engage with relationships in general and with image content specifically. The following sections examine Green's entity relationship model published in 2001, which touches on the role of relationships in data value standards; Enser's generic-specific continuum published in 2008; and finally, Jaimes and Chang's (1999), developers of a conceptual subject indexing model.

Green's entity relationship model

Green (2001) provides a practical place to begin a discussion of the literature as it pertains to relationships in LIS and their role in defining data value standards. In LIS there are a limited number of relationships that are named, defined, categorized, and given some kind of organizational structure and symbolic representation. To begin understanding what these are, Green proposes a specification model – a means of explaining relationships and their semantics – by designating three conditions: what entities are bound by the relationship, what an entity's role is within the relationship, and a determination of whether the relationship is abstract or concrete. Green's specification encompasses subject relationships, properties of relationships and entities and instances of relationships.

Green argues that the specification model demonstrates that knowledge of the entities being linked together in a relationship provides clues as to the nature of the semantic relationship and thus whether it is abstract or concrete. To illustrate this point she offers this example:

(1) Person < born in > Place.

(2) Michelangelo < born in > Caprese.

(3) Raphael < born in > Urbino.

Green classifies assertion 1 above as an abstract < born in > relationship because it joins together universals – the set of all persons with the set of all places as opposed to named individuals. She classifies assertions 2 and 3 as concrete < born in > relationships because they form associations between proper names, or what she calls "specific entities."

Green further refines semantic relationships by drawing a correspondence between parts of speech and class distinction, extending it to relationships constituting closed and open classes. A closed class can be enumerated whereas an open class cannot. Prepositions and conjunctions, for example, are classified as closed classes and adjectives and nouns are classified as open classes. It is not likely that someone will invent a new preposition (closed class), but it is not unusual to see new nouns and adjectives entering our language (open classes).

The character of closed class relationships helps define the nature of spatial relationships describing visual information, which are closely related to prepositions. This has important implications for the relationship research and it provides for an interesting space to investigate knowing that in terms of linguistic expressions, at least, the set of possibilities is most likely finite.

Returning to Green's earlier distinctions between abstract and concrete relationships, applied to the bibliographic universe, bibliographic entity classes operate with abstract relationships, but catalog records operate with concrete relationships. Catalogs form a complex domain that not only considers indexing relationships, but also relationships that link two or more bibliographic entities together. For example, Tillett (1991) classic shared-characteristic relationship that links together entities that share a common author, title, subject, or other attribute. To extend the categories of concrete and abstract relationships to machine-readable records ultimately returns to the problem of knowledge representation and the kind of knowledge, in this instance, a < born in > relationship holds.

Turning to the field of knowledge representation for an answer, Sowa (2000) notes that abstract forms may be embodied in physical entities and that the same abstract form may be embodied in many different physical objects, especially in library information systems. Take for example, Dorothea Lange's photograph "Migrant Mother." We use the

predicate "v-photograph," which stands for the verb form of the word "photograph," to relate Lange to the photograph Migrant Mother, as in "Lange photographed *Migrant Mother*."

Using Green's formulation, an abstract relationship is expressed in proposition 1 below and a concrete relationship is expressed in 2.

(1) Person < v-photograph > Person.

(2) Dorothea Lange < v-photograph > "Migrant Mother".

Applying Sowa's analysis, the name "Migrant Mother" could refer to an abstract form – the work conceived by Dorothea Lange – or it could refer to its embodiment in subsequent photographic prints, or to a pattern of bits encoded for computer processing. When the name is coded in MARC and the photograph is scanned for processing in the library's catalog, multiple instances of "Migrant Mother" – both physical and abstract – occur.

Sowa's view of repeatable abstract forms differs from the librarian's view of the world modeled in the Functional Requirements for Bibliographic Records (FRBR) (Plassard, 1998). FRBR, which is grounded in Lubetzky's (1953, 1960, 1969) work, who built upon and expanded the ideas of Panizzi (1841) and Cutter (1904), defines work as, "a distinct intellectual or artistic creation" (Plassard, 1998, p. 12). It is an abstract entity that is multiply instantiable as an expression, which is another type of abstract entity defined as, "the intellectual or artistic realization of a work," (Plassard, 1998, p. 12).

Sowa argues that in computational environments confusion results when clear distinctions are not drawn between these many entities – multiple abstract forms characterizing physical entities, each form with a different name and names that take different forms. He does not draw a line of demarcation between the abstracta work and expression. Instead, the thought is that the same abstract form may exist in multiple physical embodiments. This author will not attempt here to clarify the matter further, but merely stress that it is no simple matter to specify the precise nature and semantics of relationships between entities or enumerate the types of entities that participate in or imply certain kinds of relationships.

Enser's generic-specific continuum

Green's proposal that relationships be classified as either abstract or concrete follows earlier generic-specific models central to Shatford's (1986) subject oriented organization scheme for pictures, Armitage and Enser's (1997) mode/facet matrix and Jaimes and Chang's (2000) Pyramid. Enser (2008) proposes moving away from these dichotomous models instead proposing a generic-specific continuum, adopting Shatford Layne's (1994) composite model of "subject" made up of four attributes: object, spatial, temporal, and activity/event.

Enser's central argument is that a continuum model accommodates a hierarchy of gradations, but he fails to specify what these gradations might be. His model adopts Shatford Layne's (1994) specification of "subject" as a composite of object, spatial, temporal,# and activity/event facets. Modeling the object attribute of subject, for example, Enser makes clear a four-category model consisting of: Generic Object Instance (e.g. "tree"), Generic Object Class Hierarchy (e.g. "maple tree"), Specific Named Object Class (e.g. "Japanese maple"), and Specific Named Object Instance (e.g. the Japanese maple growing at 611 SW Kingston Avenue, Portland, Oregon).

The four labeled-categories raise the questions, is Enser's model ontologically neutral and did he intend to accommodate all possible relationship types expressed in image descriptions? If not, what are the limitations of this model? For example, he informally defines what he means by "generic location" and "specific location." Locative expressions in the continuum, however, are limited to geographic locations and do not include other spatial expressions such as paths. For instance, "bear running around a tree." In addition, he is missing a distinction between events in the four dimensions of space and time and descriptions of events represented in images. This brings into light the challenges of evolving from natural language descriptions read by humans to precise, disambiguated word senses and predicates read by machines. Enser's continuum also illustrates the indistinction between signifiers (words) and the objects they stand in for (signifieds).

Jaimes and Chang's ten-level indexing pyramid

Turning now to Jaimes and Chang's (2000) conceptual subject indexing model, what is most noteworthy about their research is the lenses through which they choose to view meaning in image content to assist in indexing visual and non-visual information. In examining what to index and how to carry it out, Jaimes and Chang propose looking for general concepts and visual concepts and then ask what meanings emerges along these dimensions. The model they arrive at is a pyramidal structure built up from ten levels. The first four levels in the pyramid are syntactic and refer to how the content is arranged and the next six levels are semantic attributes concerned with the meaning of objects and how they are organized. In photographic images, for example, black and white photographs could be a type, which is a Level 1 concept. Global distribution, Level 2, refers to attributes that are global to the image, for instance grayscale histogram. Level 3, local structures, is concerned with local elements like lines, shadow and texture and Level 4, global composition, refers to how these elements are arranged. Applying semantic categories, elements can be described as generic (cypress tree), specific (Árbol del Tule, the Montezuma Cypress in Santa María del Tule) and abstract (stout). These same semantic levels can be applied to scenes.

Arrows along the top of the pyramid indicate more knowledge is required as you move down through the levels. For computer scientists relying on automated image retrieval, this means it is easier for a machine to recognize a generic tree the individual tree *Árbol del Tule* and it would require a human to describe and interpret entities at the very lowest level (the most detailed semantic level) of meaning.

In addition to the ten-level indexing pyramid, the authors also consider syntactic and semantic relationship types. Syntactic relationships can be generic or specific and hold among image elements at any level in the pyramid. They classify spatial, temporal, and visual relationships as syntactic types of relationships. Semantic relationships can be generic, specific, or abstract and occur only at the six semantic levels in the pyramid. Lexical and predicative relationships are examples of semantic relationships.

In a demonstration of spatial relationships, the authors present a photograph showing Z. Jiang standing near B. Yeltsin. The predicate $<$standing near$>$ is explained as an instance of orientation relationship or possibly a subtype of that class. During the course of their discussion, a simple taxonomy of relationships and instances of relationships emerge. An orientation relationship is a type of spatial relationship and a spatial relationship is a type of syntactic relationship. A class labeled visual relationship may be too general and ambiguous to count as a relationship type

and deserves clarification. Examples include the predicates < darker than > and < bluer than > .

In 2001, Jörgensen, Jaimes, Benitez and Chang test the Pyramid's conceptual structure and usefulness as a tool for capturing attribute classification. Their research questions focus on how well the Pyramid assists student indexers classify terms describing image attributes. In Appendix A of the study, "relationships" is listed as an attribute and two terms "brothers" and "romance" are given as examples of relationships. Other than this hint at connections or associations that may exist between members of the class "persons," there is no other reference to relationships.

This study and other work in this area remains focussed on indexing image attributes (Fidel *et al.*, 1994; Shatford Layne, 1994), the range of attributes used for describing images (Jörgensen, 1998), and analysis of image search logs measuring user behavior based on number of sessions, searches per session, terms per query, and so on (Jörgensen and Jörgensen, 2005). Recognition and analysis of relationships among objects and their properties, for the most part, remains an elusive subject.

Phenomenal descriptions

Phenomenal descriptions evolve during research studies examining image user behavior (Jörgensen, 2003) and during analysis of concepts and relationships included in controlled vocabularies (Bean and Molholt, 1996) or data sets extracted from online image databases (Tribble, 2010). This is a less direct link to machine-readable processes. Researchers develop categories of concepts and relationship types and apply various systems of representing these entities within the context of their research papers, which may or may not coincide with existing bibliographic machine-readable standards. Phenomenal descriptions are often times more formal than natural language, but not always. More important, this section explores research on the nature of description within the framework of image searchers' questions. Writers include Enser (1993) on unique and non-unique image properties, Enser and McGregor's (1993) adaptation of the Gibbs-Smith and Keystone classification schemata, Enser *et al.* (2005) image taxonomy, Keister (1994) on the language of queries, Armitage and Enser (1997) on the classification of queries, Collins' (1998) pre-iconographical description, Jörgensen (2003) on user behavior, and Greisdorf and O'Connor (2002) on viewer's percepts.

Enser's unique/non-unique dichotomy

Enser (1993) examines the nature of user demand for visual information by closely looking at the form and content of user requests received by the Hulton Deutsch Collection Limited. Enser's research is important to this review because it recognizes that designing effective retrieval systems rely in part on understanding the nature of image searchers' queries and it examines the semantic content of a set of image requests that contain complex, natural language statements expressing image attributes and relationships.

The Hulton Collection contains over ten million images ranging in format from photographic prints and negatives to cartoons, maps, and engravings. The majority of queries comprising the data set for this study are telephone queries where a Hulton picture researcher mediates and elicits a subject statement and records the client's requests on an "Internal Enquiry Form." It is not clear what part of the requests reflect the client's own words and what part, if any, is interjected or interpreted by the picture researcher. Nonetheless, turning to written queries as data sources for research into image descriptions is significant.

As part of this study, Enser tries to assess the extent to which the Collection's Gibbs-Smith and Keystone classification schemata can represent the level of detail found in the uniquely defined subject requests. The Gibbs-Smith schema is introduced in the next section Enser & McGregor's Hulton Collections analysis.

Enser brings together a test collection of 1,000 client requests that are selected and stratified according to six user types including "other." A total of 2,722 individual image requests were extracted from the 1,000 request forms. Enser analyzes the test collection by classifying image content according to two of Erwin Panofsky's levels of understanding and interpretation: pre-iconographical description, which Panofsky describes as the first level of interpretation that a viewer apprehends "by identifying pure *forms*, that is: certain configurations of line and color, or certain peculiarly shaped lumps of bronze or stone, as representations of natural *objects* such as human beings, animals, plants, houses, tools and so forth" (Panofsky, 1939, p. 5). Enser equates this with a level of meaning found in generic subject queries, and iconographical description, which Panofsky describes as a level of interpretation gained through knowledge of literary sources and practical experience. For example, ascribing politeness or a polite greeting to the lifting of a hat (Panofsky, 1939, p. 4). Enser has difficulty applying this dichotomous classification system, however, acknowledging that while the initial request might be stated as a generic topic – pre-iconic subject – the picture that is finally selected and retrieved for the client has iconic properties associated with it. He illustrates this problem with a generic request for a pre-iconographic image of "the first microscope." Enser explains that in order to find this image, the picture researcher performs interpretation, factoring in other unique properties not explicitly stated by the client. The retrieved image, he argues, is essentially an image that is iconic in nature and concludes that Panofsky's pre-iconographical and iconographical classifications used for fine art do not work effectively in the general commercial environment of the Hulton collections.

Enser sets out devising a new image property called uniqueness, which he defines as "a request for the visual representation of an entity, the desired, particular occurrence of which can be differentiated from every other occurrence of the same entity type" (Enser, 1993, p. 29). He settles on four categories of image requests: non-unique, non-unique refined, unique, and unique refined, and adds sub-categories of time, location, action, event, and technical specification. For example, the request "shell shock" would be classified as a unique subject and "shell shock after First World War" as a unique subject with the refiners of time period added. The query "5-6 year old boy trampolining, in mid-air, in silhouette" is considered by Enser to be a non-unique subject, a boy, refined by age, event, and technical specification (in silhouette).

Enser discovers that refiners play an important role in characterizing visual information. In 34 percent of the unique and non-unique queries, the target entity is expressed in the context of a given time period or era. This has implications for accommodating the capture and representation of temporal relationships in future studies. Enser also discovers that 69 percent of all requests sought unique entities associated with people, objects, locations, or events, especially in the case of requests coming from newspaper and magazine publishers. An example of a request in this category is "crying, distress, must be over 16, good focus on individual." Given the extent to which refiners are added to unique and non-unique subjects, Enser concludes that clients rely heavily on the picture researchers' roles as intermediary and that the Gibbs-Smith scheme "can function only as a blunt pointer to regions of the Hulton collections where pertinent material might be co-located" (Enser, 1993, p. 35).

Enser and Mcgregor's Hulton collections analysis

In developing subject classification for use in cataloging the pictures in the Hulton image collections, Enser and McGregor (1993) introduce the Gibbs-Smith and Keystone classification schemata. This formal research report is a more detailed analysis of the research project introduced in the preceding section and it is reviewed here to see what, if any, relationships are revealed in Hulton's image classification systems. The report includes valuable illustrations showing samples of prints from the collections, request forms, a cross reference sheet, over four hundred image searcher's queries, and a "pictorial request" where the researcher sketches the image he is seeking.

The scope of the Hulton picture collections span the picture contents of the Hulton Group's periodicals, the best known being Picture Post. The predicted client base and the potential range of picture topics are unlimited: "every 'picturable' subject and activity on earth […] throughout history to the present day" (Gibbs-Smith, 1950). The original intentions were that the classification scheme would work in partnership with a well-informed picture researcher knowledgeable of the collections and that successful retrieval would engage both directed searching and browsing.

Enser and McGregor describe the Gibbs-Smith scheme as hierarchically structured dividing knowledge into four broad categories. They present it formally as a set of terms representing classes and subclasses. The diagrammatic view presented in Figure 1 is a literal representation of the authors' narrative explanation, illustrated here as a concept map with nodes and unlabeled links arranged hierarchically in a top-down fashion. Viewing the schema in this manner makes clear some of the strengths and weaknesses of the model.

Four of the more interesting observations are presented here:

(1) The relationships among entities are not made explicit and do not appear to reflect a hierarchical framework as suggested by the authors. This less formal structure invites a broad interpretation of headings and may explain in part how the staff succeeded in manually cataloging some ten million images – the estimated size of the collection in 1993.

(2) Every top-level node could represent two or more concepts and relationships. The resulting ambiguity may make processing and subject analysis less cumbersome and time consuming.

(3) A second-level node or subclass appears to represent attributes or roles associated with members of their superclasses. This reflects a general viewpoint expressed throughout the literature that at this level of modeling—natural language narrative – it is not necessary to be explicit or precise in distinguishing between properties and relationships.

(4) A distinction is not drawn between a person's name (a string) and the person (a human being).

The model supports three or more levels of specificity. As the level of specificity goes up, more attributes are associated with the broader class. For example, the third level of specificity for the class labeled "Portraits" includes Surname, Forename, Date of Birth, Date of Death and Occupation/Title. Other evidence of semantic relationships in this system are found in the cross-referencing. Enser and McGregor describe three relationships that guide users to other relevant parts of the collection: here, away, and

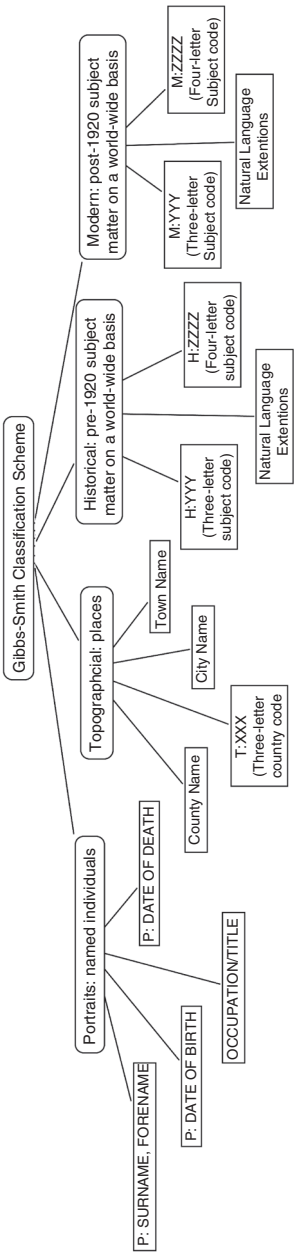

Figure 1.
Diagrammatic view of the Gibbs-Smith classification scheme

see also, which are roughly equivalent to the traditional thesauri relationships broad, narrow, and related terms.

Enser, Sandom, and Lewis's image taxonomy

Enser *et al.* (2005) are interested in the semantic gap and image retrieval in real-world applications. Their attention focusses on surveying still image types, image users, and image metadata with hopes of providing insight into the scope and significance of the semantic gap. In traditional content-based image retrieval (CBIR), the semantic gap refers to regions of information that lie between the low-level, automatically extracted image primitives such as shape, color, and texture, and the higher-level processes that humans apply during searching and describing tasks such as interpretation or that require recognizing or labeling abstract entities such as point of view and mood.

The authors survey image retrieval literature to develop a simple taxonomy of image types illustrated in Figure 2.

Enser, Sandom, and Lewis ground their arguments in the belief that, "the retrieval utility of visual images is generally realized in terms of their inferred semantic content" (Enser *et al.*, 2005, p. 177). They claim inferential reasoning arises from semiotic distinctions drawn between denotation and connotation of image content. The CBIR community has attached the label "semantic image retrieval" to the formulation and resolution of information needs that engage this process. The authors conclude, from research conducted by Enser (1995), Armitage and Enser (1997), Ornager (1995), and Enser *et al.* (2005), that, "*identification* is dependent upon prior existence – and knowledge by the user – of a defining linguistic label" (p. 180). The authors add that certain image components may be recognizable by shape alone – for example, an image of a refrigerator – while certain other attributes may rely on textual annotations–such as limiting search results to refrigerators manufactured in the 1950s.

A brief look at Enser (2008) shows him extending this model three years later, describing a visual information domain – still and moving images – as a dichotomous community with one group working in image retrieval and another engaged in curatorial image management. The former group is made up of researchers and the later practitioners. Enser is concerned that there is limited communication between these two groups and that image retrieval researchers know very little about image searchers' needs or the logistics of managing picture collections. This results in procedures and practices that are technologically feasible, but that may serve little useful purpose in practice.

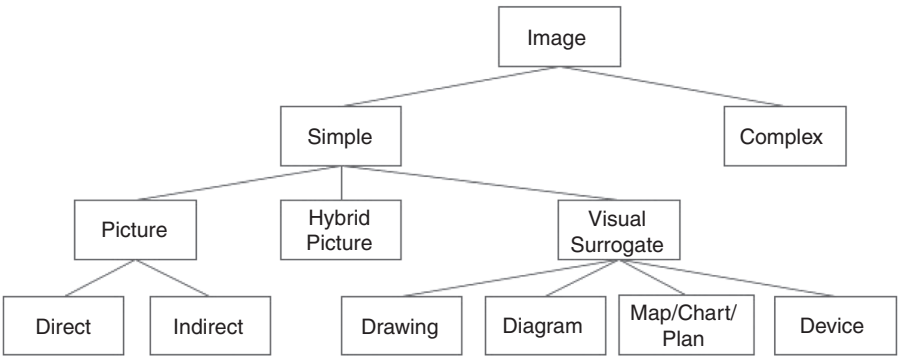

Figure 2.
Enser *et al.*'s (2005) taxonomy of still images

Keister on the language of queries

The question of how textual description brings meaning to images and the recognition that meaning is embedded in image searchers' queries was raised by Keister (1994) in her retrospective analysis and re-analysis of users' queries at the Prints and Photographs Collection of the National Library of Medicine (NLM). This data set consists of the reference query log for the year 1984 and 291 additional queries from 1991.

Keister draws a distinction between image searchers from the health profession, museum, and art community noting, for instance, that picture professionals think visually and use a query language full of jargon. She infers a relationship between types of image searchers and the language of the query, but stops short of identifying and discussing what those relationships might signify.

Keister considers that image searchers express more than just image attributes when submitting a request. For example, one query requests an action shot of George Papanicolau. The phrase "action shot of" is a *functional expression t*hat denotes a single entity when applied to the name "George Papanicolau." That is, it designates some unique individual photograph. Keister also held that searchers sometimes constructed images with words and called these image construct queries. For example, the query "poor people, especially children, may be on a city street, lame or crippled, with canes" constructs a corresponding image in the mind of the reader. She reports that one-third to one-half of all image requests in her library end up being this type of image construct query.

Sometimes the words in an image's catalog record do not match the description provided by a searcher. Keister describes queries requesting a well-known image by Benjamin Rush titled "Tranquilizing Chair." Researchers generally ask for, "the man sitting in the chair with a box on his head." Her basic contention is that searchers cannot find pictures using their descriptions if the words they use are not accessible in the catalog record. It seems fair to say that for Keister, the meaning an image invokes can be found not only in a library's catalog, but in the searcher's image construct and that this can influence how we build catalog records. As a solution she proposes cataloging images at the item level and including a surrogate image in the catalog record itself.

Bean on the relationships among anatomic entities

While at the Department of Medical Informatics, Columbia University, Bean and Molholt (1996) tried to determine the character, structure, and organization of implicit associative relationships relevant to clinical anatomy. She adopts analytic tools for determining the nature of terminological relationships between concepts. Bean, along with others (Bean and Molholt, 1996; Green, 1996) proposed applying terminological relationships to information retrieval and thesauri construction, specifying the properties of the link in order to better extend non-hierarchical relationships.

Bean examines a total of 256 term pairs that she organizes into five broad categories: procedures, (other) anatomic entities, functions, disruptions to functions, and chemical agents. Her investigation shows that the most common category is neoclassical compounds joined with a combining form that designates a therapeutic, preventative, or diagnostic procedure. For example, Bronchi:bronchoscopy. In Bean's terminology, this example illustrates a semantic category of procedures that relate terms naming a particular procedure (bronchoscopy) performed on a focal anatomic entity (bronchi). Her goal is to identify the associative relationship joining these two terms. In some

cases the related term (second term) is not known. Bean predicts that this happens when the procedure is specific only to a particular anatomic entity, for instance, gonads: castration. In other words, castration is a procedure specific to the focal term in this term pair. Herskovits (1986) maintains that general and specific world knowledge about a particular domain or context – in this instance, knowledge of medical subject headings and procedures – enables making pragmatic inferences, adding facts to a description that go beyond the original utterance. Bean considers relationships such as removing and placing as low-level reality. That is, they are instances of applying general principles to specific instantiations. Her high-level model is concerned with relationships that can be generalized across any domain.

Bean concludes there is a need for a set of distinctive relationships that reflect precise actions of procedures on anatomical entities and intends to eventually pursue using a simple pattern-making program to classify relationships based on string matching features of terms.

Armitage and Enser on the categorization of queries

Armitage and Enser (1997) provide an analysis of a cross section of query types collected from seven libraries whose archival holdings in still and moving images cover a wide variety of topics. Their goals are to advance the understanding of visual information needs and to inform the interface designs developed for accessing image collections. This research is an extension to earlier research by Enser and McGregor (1993) and Enser (1993) – studies concerned with analysis of information needs in the visual domain.

Armitage and Enser choose a sample of 1,749 images in consultation with library staff and examined their descriptions in the context of image queries. The data set was organized broadly according to query type and image content. Questions were grouped into mediated, recorded questions, or unmediated questions. The former were questions jotted down by reference librarians during reference transactions. The later, extracted from image request files in public libraries, were queries expressed by image searchers in written communications. Image queries were also categorized according to: image content, identification/attribution/provenance checking, accessibility to work and ownership/viewing availability, and miscellaneous. Queries in category 1 were further subdivided by requests for named artists, known items, unique subjects, and non-unique subjects. The miscellaneous category presents interesting possibilities for semantic analysis of the outliers that they consider to be, among other things, "unusable queries" and "requests for administrative procedures."

Armitage and Enser build a model for analyzing levels of meaning in image content based on Panofsky's modes of image analysis (Panofsky, 1939) and facet analysis introduced by Markey (1983) and refined by Shatford (1986). The focus here is on the utility of their mode/facet matrix (Table I) and the extent to which it meets their promise of representing and characterizing both queries and image content. It functions by generating 12 categories of subject content by combining one of Shatford's four facets (who, what, where, when) with one of Panofsky's modes (iconography, pre-iconography, iconology).

A theme throughout this review is the problem of interpreting concepts consistently and determining in what subject category to place borderline and ambiguous concepts, compound terms, and entire sentences. This problem surfaces for Armitage and Enser as they struggle to pigeonhole words and phrases into rigid categories.

The issue of finding the right category in which to place a thing aligns closely with what Dorr refers to as the special composition question: "Under what circumstances do

	Iconography (specifics)	Pre-iconography (generics)	Iconology (abstracts)
Who?	Individually named person, group, thing (SI)	Kind of person or thing (Gl)	Mythical or fictitious being (Al)
What?	Individually named event, action (S2)	Kind of event, action, condition (G2)	Emotion or abstraction (AZ)
Where?	Individually named geographical location (S3)	Kind of place: geographical, architectural (G3)	Place symbolized (A3)
When?	Linear time: date or period (S4)	Cyclical time: season, time of day (G4)	Emotion, abstraction symbolized by time (A4)

Table I. Armitage and Enser's matrix for coding categories of subjects in image content

several things compose something?" (Dorr, 2005, p. 234). The "several things" in this instance could be the multiple facets present in a query and the entity that is composed of these parts could be the query's subject category. For example, Armitage and Enser claim a hierarchical relationship between unique and non-unique subjects in the visual domain. As they state it, "an entity can always be interpreted into an hierarchy of related super-concepts and sub-concepts" (p. 290). Their concern is at what level in the hierarchy does an entity cease being unique and instead become non-unique. This is a common problem in systems of classification – a problem explored earlier in Enser's (2008) continuum of general to specific. In Dorr's terms this question could be framed as, "Under what circumstances is there a subject category having each of several facets as component parts, every facet of which is related to one of them?."

The special composition question can be applied to a specific example where Armitage and Enser try to code the query, "cheetahs running on a greyhound course in Haringey in 1932." They conclude based on their coding matrix that the subject is S2. That is, a specific individually named event. They answer the compositional question in part by saying an individually named entity does not necessarily have to consist of a pronoun. It can be a generic reference. Moreover, they claim that the essence of the query can be conveyed by a single facet or notion of an event without making the remaining concepts and relationships in the query explicit in the coding.

A useful example of what happens to relationships when applying the mode/facet matrix is illustrated in the single facet versus multi-facet image description. Consider the three queries in Table II.

Query 1 is a single facet generic query; query 2 is a multi-facet query consisting of a unique, named geographic location plus a non-unique subject; query 3 is a multi-faceted query consisting of a unique, named geographic location plus a specific date. The question they raise is how to classify 2 when its component parts are both a

Query	Code (see Table II)
1. Carnivals	G2
2. Rio carnivals	S3+G2
3. Rio Carnival, 1986	S2+S4

Table II. Three queries applying combinations of Armitage and Enser's mode/facet codes

specific geographic location and non-unique event. Their solution is to introduce the concept of refinement, that some subjects can be non-unique, but refined with a modifier or "refiner." Thus, "carnival" is non-unique (G2) refined by location "Rio," which is unique (S3).

If the concept "carnival" is expressed as having a date and location function Armitage and Enser's approach could begin serving as a template for explaining any number of events expressed in image descriptions comprised of an event name, event location and date. In its current form, however, query categorization does not have a mechanism for representing the relationships between entities that make up events. In this case geographic location and points in time.

Collins' pre-iconographical description

Collins (1998) provides another example of applying Panofsky's pre-iconographic description of images to indexing visual materials. Like others before her (Markey, 1988; Shatford, 1986), Panofsky's analytic framework serves as her basis for subject analysis during the description process. She continues the practice of treating relational statements as primitives that need no explanation.

In this research, Collins focusses attention on patrons' visual information needs in historical photographic collections. She investigates whether there is a need to inventory and index more of the elementary, factual aspects of images and whether some objects are more important to name than others. Another overarching question Collins raises is whether detailed item level descriptions of photographs are better than collection level descriptions.

Before proceeding further, if we are to continue moving toward effective representation of photograph descriptions in computer-based information systems, two critical points deserve close attention: there is an important difference not made explicit here between analyzing the image content of photographs and analyzing photograph descriptions. The former is concerned with picture elements and entails interpretation on the part of the researcher. The later focusses on linguistic expressions humans use to represent image content. The "subject" is usually treated as a primitive in image indexing, but intuition alone suggests that the index term or "subject" of a photograph can be complex and not easily reduced to controlled vocabulary terms. Even the simplest term engages some level of pragmatic inference. Humans rely on general and specific world knowledge and knowledge about communication to infer more facts than what is apparent in the utterance of a single index term.

Collins asserts catalogs usually limit image descriptions to secondary subject matter corresponding to Panofsky's iconographic level of analysis and argues for more primary subject matter, or pre-iconographical description. To support this argument Collins offers us a scenario where an image requires a higher, more interpretive level of description. Consider the caption, "Maj. Jesse Marcel holding debris from Roswell crash." This iconographic level of description, she claims, prevents both specialists and non-specialists from searching at the primary pre-iconographic level on generic phrases such as "people holding debris" and "men in uniform." This example – people holding debris – reinforces Collins' claim that primary subjects describe objects and events. A forceful objection to all approaches that invite us to think of phrases like "this person holding debris" as being indexable, however, is that this ignores the relational statement. There are two entities being joined together in this phrase by the relationship "holding," which is an instance of a particular kind of activity.

Jörgensen on user behavior

By far Corinne Jörgensen is the most influential scholar providing us with insight into how indexing image attributes influences access to images. Jörgensen (1995) establishes classes of image attributes based on user behavior while carrying out three categories of tasks: describing, sorting, and concept searching. Analysis of the results reveals 12 distinct categories of image attributes with "objects" being the most prevalent category of image description in all three tasks.

In Jörgensen (2003) she reexamines the data produced by participants in her earlier study to consider the nature and structure of the relationships among individual image attributes and classes of attributes. She adopts Graesser and Goodman's (1985) view that in text analysis both implicit and explicit knowledge structures contribute to meaning, which leads her to identify two relationship types in her analysis: hierarchical and something she calls figure/action structure. She defines the later as "a main figure or object and an action performed by or upon the main figure or object"(Jörgensen, 1995, p. 173). Jörgensen (1995) also describes images as having focal areas and names the primary focal area the "figure" and the area that remains in the background as the "ground." She brings these two entities into a relationship with one another, but does not state explicitly what the nature of the relationship is or how it is expressed in participants' descriptions.

Her analysis of hierarchical relationships measures what percentage of terms are basic level, subordinate, and superordinate. As an example, if "gun" represents a basic level term, then "rifle" represents a subordinate term and "weapons" a superordinate term. She determines, for example, that in the descriptive viewing task 76.6 percent of the terms are basic level, 11.8 percent are superordinate, and 11.6 percent are subordinate.

In the conceptual search task, participants are shown the same six images used during the other tasks, but this time are asked to describe a search statement that might successfully retrieve each image. The analysis reveals that participants describe images in a narrative style using a wide variety of relationships. No comprehensive attempt is made to predict what all the relationships are beyond object-figure/action relationships found in 64 percent of the descriptions. These are characterized by an object or figure term followed by a verb, in turn followed by one or more nouns or prepositional phrases.

Greisdorf and O'Connor viewer's percepts

The last review in the phenomenal category of descriptions examines Greisdorf and O'Connor (2002) who consider what image viewers think about when they evaluate images. They base their study on three assumptions. First, that prior research indicates there are seven descriptive categories of image attributes that describe viewers' percepts, including color, shape, texture, object, location, action, and/or affect. Second, they support a hierarchy of perception model for image search, retrieval, and evaluation based on Panofsky's (1939) three levels of meaning in Renaissance art. Greisdorf and O'Connor describe these three levels as: primitive features (e.g. color, shape, and texture), objects (e.g. person/thing, place/location, and action), and of particular interest to the researchers inductive interpretation. That is, things not pictured in the content of the image. Finally, they perceive image retrieval as a process of reconceptualization where concepts are initially embodied in the context of the query and then emerge during the context of viewing.

Nineteen participants were given 27 pre-selected query terms and ten grayscale images drawn from the NOAA government database. Participants were also given the opportunity to develop their own descriptors. The results of the study suggest that affective/emotion-based query terms are an important descriptive category in image retrieval. Greisdorf and O'Connor appear to be claiming in this regard that these aspects of image need to be indexed along with the objective descriptions. An objection to studies that invite participants to use pre-selected terms is that the chosen words and images can bias the selection process. For instance, a participant describes an image of a large body of water surrounded by a forest using the pre-selected descriptors green, boat, and fishing even though the image is grayscale, there are no boats and no one can be seen fishing.

In contrast to textual indexing, they suggest classifying and categorizing images according to metaphoric, analogical, metonymic, or synecdochic relationships. They propose that these aspects of an image should be conveyed to image searchers who are uncertain about the actual topic they seek.

In a cursory discussion attempting to delineate these different figures of speech and relate them to images, the authors explain that metaphoric relationships are used when images or picture elements within an image stand for something else and analogy conveys something about the story an image tells. There are countless examples of photographs picturing businessmen wearing suits and carrying an attaché while climbing a ladder – a picture that could be interpreted as a metaphor for climbing the ladder of success. The authors admit that this meaning could be represented by existing retrieval systems through captions attributed to photographs by the artist, but claim that the artist's labeling may not match the searcher's needs.

They define metonymy as "a figure of speech that consists of using the name of one object or concept for that of another to which it is related" (pp. 20-21), such as "lend me your ear" to mean "pay attention." A more useful metonymic relationship in the domain of photographs might be a picture of Washington, DC, described as a metonym for the US Government. Subtypes of metonymic relationships could include cause and effect, container and contents, possessor and the thing possessed, and an occupation and its sign. In the example given earlier, Greisdorf and O'Connor identified the terms boat and fishing as metonymic characteristics of the image picturing a large body of water.

Closely related to metonymy are synecdochic relationships, which involve using a part to represent a whole as in "red coat" standing in for "soldier in the British Army." The authors also consider the participants in their study assigning color attributes to black and white images as evidence of synecdoche.

Discussion

The diversity of arguments and positions on this topic of relationships should be noted. We may begin by revisiting Green's assertion that the semantics of relationships are determined by the nature of the entities they link together. Doing so sheds light on one central aspect of the relationship problem in LIS. It is often proposed that the entities that exist in the world and the relationships that hold between them need only labeling and categorization. No further explanation is necessary. Such an account neglects the role that humans play in reasoning over relationships and how systems might supply a means of supporting reasoning. It is also objected that while understanding the semantics of relationships is considered a key factor in information systems, many writers fail to distinguish representing relationships and their properties vs representing keywords and subject terms. For example, Greisdorf and O'Connor

argue that by describing image content within a framework of "figures of speech" traditional indexing systems can be extended by generating an array of associations and meanings between entities not visible in the image. The researchers' goal, in effect, is to represent in image descriptions what image searchers "see." They claim that by applying these relationship types to search and indexing systems searchers can achieve more meaningful and relevant search results.

Green acknowledges our source of confusion in this area admitting there is a disconnection between humans' intuitive understandings of relationships and the problems of representation and processing relationships in computational environments. It will be through reasoning over relationship types, green contends, that the discovery process can enable searchers to find information that would otherwise go unnoticed. Bean's conclusions also point beyond relationship recognition arguing the importance of determining how relationships work. For example, if a relationship's properties are transitive, a system can potentially gain storage capacity and retrieval efficiency by applying subsumption and transitive inheritance principles.

In 1993, Ensor questions how researchers' queries describe images and claimed this should determine indexing and subject classification in information systems. While existing technologies lack the expressive power to represent the rich semantic relationships Enser hints at making explicit, especially in the semantic field of spatial expressions, his continuum model is an extension of Shatford Layne's finer-grained approach to subject indexing. Future research could devote considerable attention to spatial relationships because, as Gruber's (1965) thematic relationships theory makes evident, their semantics hold the key to understanding a wide variety of other semantic fields.

Instead, Enser and McGregor (1993), who make evident a rich variety of relationships present in the Gibbs-Smith classification scheme and cataloging practices of the Holton collections, their work defends only the three transitional cross-reference relationships found in thesauri: broader, narrower, and related terms.

Keister, like Enser, provides an empirically based account of image searchers' reliance on semantic relationships when constructing queries. He defends a model that employs a combination of controlled vocabulary keyword searching on the first pass and then visual browsing of surrogate images on the retrieved record set to see which images convey "most effectively the desired message." Recognizing the problem of conveying image content in words, Keister introduces the notion of an "image construct query," which views queries as attempts to build images with words.

Jörgensen (1995) analyzed the types of words and phrases used by image viewers to describe images, drawing upon research in cognitive science and theories of human image understanding. She concluded that the choice of image attributes is determined in some part by whether participants are assigned describing, sorting, or searching tasks. This suggests that an information retrieval system interface would have to provide some mechanism to capitalize on this facet of use. Her analysis provides 47 attribute categories organized into twelve higher-level categories. Although she recognizes prepositions indicate relationships, the spatial and temporal meanings of English prepositions, and other ordered internal relationships in descriptions, are not considered during data analysis. This is problematic for many reasons, not the least of which is a lack of research examining the range of relationship types associated with image descriptions that might prove useful in image retrieval systems.

Finally, Armitage and Enser (1997) provide an empirical study that gathers together a semantically rich data set expressing both image attributes and relationships among

attributes. They introduce the mode/facet matrix as a function that generates codes for explaining categories of image content, claiming that there are only 12 possible codes that are available for classifying all image content. While relationship types are not explicitly examined, they exist implicitly in the matrix and they include a generous sample of mediated and unmediated queries from all seven libraries in the appendix of their paper.

Conclusions and directions for future research

It remains that little is known about the nature and scope of relationships expressed in the contexts of describing, searching, and retrieving photographs or about the intellectual problems posed by these activities. It seems likely that a complete understanding of relationships represented in image descriptions and pictorial content must include the activities of catalogers, image searchers, and curators. While the problem of representation in computational systems has not been dealt with systematically in these studies, the reviewed work implies that the semantic relationships we express in our descriptions of photographs applied to visual information and linguistic representations places hard requirements on our information systems in terms of searching and retrieving relevant images in large collections of images.

Situating relationship research in the social milieu of archives and other picture collections should not prevent engagement with other disciplines, including linguistics, philosophy, and artificial intelligence that are concerned with describing concepts or entities in the world. Like machine-readable descriptions, conceptual descriptions are made up of data structures. They differ significantly, however, in the fact that the data structures represent something. They are symbolic representations – propositions that hold knowledge about things in the world. They can be reasoned over and used for drawing inferences. These finer distinctions amplify the importance of bringing into the discussion the viewpoints of disciplines that consider how relationships are represented in machine environments.

There is no denying that when we adopt a more analytic approach and attempt describing, in any detail, semantic relationships expressed in photograph descriptions we quickly find ourselves explaining linguistic phenomenon. Recent trends in LIS show movement in this direction, formalizing relationships as objects of study and examining their symbolic representation and inference in knowledge organization (Benson, 2011a, b; Green, 2008); classifying web terms into thesauri relationships (Milonas, 2012); surveying attempts to classify relationships (Szostak, 2012); examining how relationships are represented in FRBR (Arsenault and Noruzi, 2012; Picco and Repiso, 2012); how FRBR can be used to define and clarify relationships in knowledge organization systems (Žumer *et al.*, 2012); and mapping FRBR's bibliographic relationships to Tillett's taxonomy of bibliographic relationships (Noruzi, 2012).

Important questions to ask are to what extent do relationships occur in the discourse surrounding picture archives and what types of relationships emerge when searchers, curators, and catalogers describe image content. One line of research might reanalyze query logs as factual assertions to determine to what extent relationship expressions in queries convey visual information through linguistic means. For example, where a traditional approach might identify "economic development" and "World War II" as subject terms or keywords in a query, analysis of the larger semantic structure, "[I want a photograph showing] economic development after World War II," would examine the meaning of the terms and identify the relationship "after" as a type of temporal relationship. This information should inform the future design of search and retrieval

systems, especially knowledge-base systems describing actions, events, time and time durations represented in images.

Research that investigates what image attributes are needed for affective image indexing (Jörgensen, 1995) suggests that adopting an expressive representation system that combines linguistic and semantic knowledge may be useful, but research needs to be done to fine-tune what knowledge-matching rules would be most effective and what data structures would best support reuse, simple inference and search. One approach would be frame based. Minsky (1975) described the frame as a "data-structure representing a stereotyped situation" (p. 212). The ideas is that by adopting frames for describing entities and situations represented in scenes, a single concept representing something in particular brings background knowledge to the surface. In the photograph description "Tent with Inuit woman standing in doorway, in Arctic regions," application of an Inuit knowledge or geospatial semantic frame to this subject permits reasoning over a larger set of semantically related entities linked to the frame. It remains to be explored what default values should be encoded into information systems to fill in or anticipate information not present in a photograph or its description.

Finally, a significant amount of time and labor has already gone into building image indexing systems used in large-scale enterprises. Two of the better-known controlled vocabularies are the Library of Congress Thesaurus of Graphic Materials and the Art and Architecture Thesaurus. Among the problems of representing these systems in structures that represent knowledge are limitations on the number and types of relationships allowed in controlled vocabularies and problems minimizing ambiguity. These limitations may be eliminated by the use of ontologies in knowledge-base structures that remove the ambiguity found in natural language and where there are no limits on the number of relationships that can be defined.

Prior research in ontological engineering and capture (Gruber, 1995; Perakath *et al.*, 1994; Grüninger and Fox, 1995; Uschold and King, 1995; Uschold and Grüninger, 1996; Fernández *et al.*, 1997) could serve as a foundation for developing formal specifications of the concepts expressed in these controlled vocabularies. While much theory exists on relationships, little empirical work has been undertaken and therefore there are large gaps in the body of knowledge in this area. Additional, in-depth studies are needed for the relationship types that may emerge during ontology capture. These should be grounded within the domain of LIS if practical solutions are to be found for improving description, organization, and retrieval in library information systems using ontology-based representations.

References

Allen, J.F. (1984), "Towards a general theory of action and time", *Artificial Intelligence*, Vol. 23 No. 2, pp. 123-154.

Armitage, L.H. and Enser, G.B. (1997), "Analysis of user need in image archives", *Journal of Information Science*, Vol. 23 No. 4, pp. 287-299.

Arsenault, C. and Noruzi, A. (2012), "Analysis of work-to-work bibliographic relationships through FrBR: a Canadian perspective", *Cataloging & Classification Quarterly*, Vol. 50 Nos 5-7, pp. 641-652.

Bean, C.A. and Molholt, P. (1996), "A relational model of clinical anatomy: symbolic knowledge base design", *AMIA Annual Fall Symposium, October, Washington, DC.*

Benson, A.C. (2011a), "Capturing relationships expressed in image descriptions: a preliminary investigation", *BiD: Textos Universitaris De Biblioteconomia I Documentació*, No. 27.

Benson, A.C. (2011b), "Ontophoto and the role of ontology in organizing knowledge", *Knowledge Organization*, Vol. 38 No. 2, pp. 79-95.

Collins, K. (1998), "Providing subject access to images: a study of user queries", *The American Archivist*, Vol. 61 No. 1, pp. 36-55.

Cutter, C.A. (1904), *Rules for a Dictionary Catalog*, 4th ed., Government Printing Office, Washington, DC.

Dorr, C. (2005), "What we disagree about when we disagree about ontology", in Kalderon, M. (Ed.), *Fictionalism in Metaphysics*, Oxford University Press, Oxford, pp. 234-286.

Enser, P.G.B. (1993), "Query analysis in a visual information retrieval context", *Journal of Document and Text Management*, Vol. 1 No. 1, pp. 25-52.

Enser, P.G.B. (1995), "Progress in documentation: pictorial information retrieval", *Journal of Documentation*, Vol. 51 No. 2, pp. 126-170.

Enser, P.G.B. (2008), "Visual image retrieval", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Information Today, Medford, NY, pp. 3-42.

Enser, P.G.B. and McGregor, C.G. (1993), "Analysis of visual information retrieval queries", British Library R&D Report 6104, The British Library Board, London.

Enser, P.G.B., Sandom, C.J. and Lewis, P.H. (2005), "Surveying the reality of semantic image retrieval", in Bres, S. and Laurini, R. (Eds), *Visual Information and Information Systems: 8th International Conference, Revised Selected Papers (Lecture Notes in Computer Science, 3736)*, Springer-Verlag, Berlin, pp. 177-188.

Fernández, M., Gómez-Pérez, A. and Juristo, N. (1997), "Methontology: from ontological art towards ontological engineering", *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering, Stanford University*, pp. 33-40.

Fidel, R., Hahn, T.B., Rasmussen, E.M. and Smith, P.J. (Eds) (1994), *Challenges in Indexing Electronic Text and Images*, Learned Information, Medford, NJ.

Gibbs-Smith, C.H. (1950), "The hulton picture post library", *Journal of Documentation*, Vol. 6 No. 1, pp. 12-24.

Graesser, A.C. and Goodman, S.M. (1985), "Implicit knowledge, question answering, and the representation of expository text", in Britton, B.K. and Black, J.B. (Eds), *Understanding Expository Text: A Theoretical and Practical Handbook for Analyzing Explanatory Text*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 109-171.

Green, R. (1996), "Development of a relational thesaurus", *Advances in Knowledge Organization*, Vol. 5, pp. 72-79.

Green, R. (2001), "Relationships in the organization of knowledge: an overview", in Bean, C.A. and Green, R. (Eds), *Relationships in the Organization of Knowledge*, Kluwer Academic Publishers, Boston, MA, pp. 3-18.

Green, R. (2008), "Relationships in knowledge organization", *Knowledge Organization*, Vol. 35 Nos 2-3, pp. 150-159.

Greisdorf, H. and O'Connor, B. (2002), "Modeling what users see when they look at images: a cognitive viewpoint", *Journal of Documentation*, Vol. 58 No. 1, pp. 6-29.

Gruber, J.S. (1965), "Studies in lexical relations", doctoral dissertation, MIT, Cambridge, MA.

Gruber, T.R. (1995), "Toward principles for the design of ontologies used for knowledge sharing", *International Journal of Human-Computer Studies*, Vol. 43 No. 5, pp. 907-928.

Grüninger, M. and Fox, M.S. (1995), "Methodology for the design and evaluation of ontologies", in IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal.

Herskovits, A. (1986), *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*, Cambridge University Press, New York, NY.

Jaimes, A. and Chang, S.-F. (1999), "Conceptual framework for indexing visual information at multiple levels", *Proc. IS&T/SPIE Internet Imaging, San Jose, CA, Vol. 3964*, pp. 2-15.

Jörgensen, C. (1995), "Image attributes: an investigation", doctoral dissertation, Syracuse University, Syracuse, NY.

Jörgensen, C. (1998), "Attributes of images in describing tasks", *Information Processing & Management*, Vol. 34 Nos 2/3, pp. 161-174.

Jörgensen, C. (2003), *Image Retrieval: Theory and Research*, Scarecrow Press, Lanham, MD.

Jörgensen, C. and Jörgensen, P. (2005), "Image querying by image professionals", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 12, pp. 1346-1359.

Keister, L.H. (1994), "User types and queries: impact on image access systems", in Fidel, R. *et al.* (Eds), *Challenges in Indexing Electronic Text and Images*, Learned Information, Medford, NJ, pp. 7-22.

Krause, M. (1988), "Intellectual problems of indexing picture collections", *Audiovisual Librarian*, Vol. 14 No. 2, pp. 73-81.

Lubetzky, S. (1953), *Cataloging Rules and Principles*, Processing Dept., Library of Congress, Washington, DC.

Lubetzky, S. (1960), *Code of Cataloging Rules: Author and Title Entry*, American Library Association, Chicago, IL.

Lubetzky, S. (1969), *Principles of Cataloging*, Institute of Library Research, University of California, Los Angeles, CA.

Markey, K. (1983), "Computer-assisted construction of a thematic catalog of primary and secondary subject matter", *Visual Resources*, Vol. 3 No. 1, pp. 16-49.

Markey, K. (1988), "Access to iconographical research collections", *Library Trends*, Vol. 37 No. 2, pp. 154-174.

Milonas, E. (2012), "Classifying web term relationships: an examination of the search result pages of two major search engines", *Proceedings of the Twelfth International ISKO Conference, Mysore, August 6-9*, pp. 201-205.

Minsky, M. (1975), "A framework for representing knowledge", in Winston, P.H. (Ed.), *The Psychology of Computer Vision*, McGraw Hill, New York, NY, pp. 211-277.

Mitchell, W.J.T. (1994), *Picture Theory*, The University of Chicago Press, Chicago, IL.

Noruzi, A. (2012), "FRBR and tillett's taxonomy of bibliographic relationships", *Knowledge Organization*, Vol. 39 No. 6, pp. 409-416.

Ornager, S. (1995), "The newspaper image database: empirical supported analysis of users' topology and word association clusters", *Proceedings of the 19th Annual Conference on Research and Development in Information Retrieval, SIGIR 95'*, pp. 190-197.

Panizzi, A. (1841), "Rules for the Compilation of the Catalogue", *British Museum*, Vol. 1, The Catalogue of Printed Books in the British Museum, London, pp. [v]-ix.

Panofsky, E. (1939), *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*, Oxford University Press, New York, NY.

Perakath C.B. Menzel, C.P., Mayer, R.J., Fillion, F., Futrell, M.T., deWitte, P.S. and Lingineni, M. (1994), *IDEF5 Method Report*, Knowledge Based Systems Inc., Texas.

Picco, P. and Repiso, V.O. (2012), "The contribution of FRBR to the identification of bibliographic relationships: the new RDA-based ways of representing relationships in catalogs", *Cataloging & Classification Quarterly*, Vol. 50, pp. 622-640.

Plassard, M.-F. (1998), "Functional requirements for bibliographic records: final report", *IFLA Study Group on the Functional Requirements for Bibliographic Records*, KG Saur Verlag GmbH & Co. KG, München.

Shatford Layne, S. (1994), "Some issues in the indexing of images", *Journal of the American Society for Information Science*, Vol. 45 No. 8, pp. 583-586.

Shatford, S. (1986), "Analyzing the subject of a picture: a theoretical approach", *Cataloging & Classification Quarterly*, Vol. 6 No. 3, pp. 39-62.

Sowa, J.F. (2000), *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole, Pacific Grove, CA.

Szostak, R. (2012), "Toward a classification of relationships", *Knowledge Organization*, Vol. 39 No. 2, pp. 83-94.

Tillett, B.B. (1991), "A taxonomy of bibliographic relationships", *Library Resources & Technical Services*, Vol. 35 No. 2, pp. 150-158.

Tribble, A. (2010), "Textual inference for retrieving labeled object descriptions", doctoral dissertation, Carnegie Mellon University.

Uschold, M. and Grüninger, M. (1996), "Ontologies: principles, methods, and applications, *Knowledge Engineering Review*, Vol. 11 No. 2, pp. 93-155.

Uschold, M. and King, M. (1995), "Towards a methodology for building ontologies", Artificial Intelligence Applications Institute, The University of Edinburgh, Edinburgh, p. 15.

Walch, V.I. and Matters, M.E. (1994), *Standards for Archival Description: A Handbook*, The Society of American Archivists, Chicago, IL, available at: www.archivists.org/catalog/stds99/index.html

Žumer, M., Zeng, M.L. and Mitchell, J.S. (2012), "FRBRizing KOS relationships: applying the FRBR model to versions of the DDC", *Proceedings of the Twelfth International ISKO Conference, Mysore, August 6-12*, pp. 190-194.

Further reading

Eakins, J.P. (1996), "Automatic image content retrieval – are we getting anywhere?", *Proceedings of Third International Conference on Electronic Library and Visual Information Research (ELVIRA3), De Montfort University, Milton Keynes*, pp. 123-135.

Eakins, J.P. (1998), *Techniques for Image Retrieval. Library & Information Briefings, 85*, Library Information Technology Centre, South Bank University, London.

Eakins, J.P. and Graham, M.E. (1999), "Content-based image retrieval:a report to the JISC Technology Applications Programme", Institute for Image Data Research, University of Northumbria at Newcastle, Newcastle.

Mazzocchi, F., Tiberi, M., De Santis, B. and Plini, P. (2007), "Relational semantics in thesauri: an overview and some remarks at theoretical and practical levels", *Knowledge Organization*, Vol. 34 No. 4, pp. 196-213.

Svenonius, E. (2000), *The Intellectual Foundation of Information Organization*, MIT Press, Cambridge, MA.

Corresponding author

Dr Allen C. Benson can be contacted at: acbenson473@gmail.com