

LBSC 670 Information Structure Glossary and Concept Maps

Compiled by Christine Conway
May 2009

anaphoric reference. Use of a pronoun or general noun to refer to an object, action or thought identified earlier in a text.

AND, logical. A Boolean search operator. When used to connect two aspects of a query formulation in a search retrieves a set of documents in which both aspects must be present; an intersection of sets.

Boolean operations. Use of the logical operators AND, OR, and NOT to limit or broaden search terms in a query formulation. Allows searcher to broaden or narrow search results.

authority list. Rules that specify what entities are allowed in a conceptual data schema and the identifiers used for those entities.

bibliographic database. A database that focuses on documents as main entity type.

catalog. A database containing descriptive data about documents, objects, or data sets with one main entity type in which the main entities are members of a defined collection.

category, basic level. Most useful level of categorizing an object for classification. Includes most shared concepts about the object and an optimum functionality for understanding the object. The most inclusive level of classification.

checklist technique of indexing. Process used by an indexer to check each possible relevant descriptor in an index language in turn to decide whether the descriptor is relevant to an entity.

checklist descriptors. A subset of descriptors from an index language that is particularly useful to an organization with an ISAR. Used as a first step in checklist technique.

class. A grouping or defining of entities by shared essential attributes.

classification. A logically coherent framework for the arrangement of knowledge.

classificatory structure. In terminological control, the set of all concepts as represented by the preferred terms, together with the relationships among these concepts. The main relationships are Broader Term (BT), Narrower Term (NT), and Related Term (RT).

coherence. In text structure, the semantic sense and logic of a text. Does the text make sense?

cohesion. In text structure, the elements of a text are properly linked grammatically through properly structures sentences and inter-sentence relationships.

concept. An idea of something--include individual, class and category concepts, abstract ideas, mass and count, and false stereotype ideas.

concept combination. A combination of subject descriptors used to search classifications and classification hierarchies to retrieve information relevant a query.

concept map. A knowledge representation structure organized in the fashion of a semantic network and consisting of concept nodes linked by relationships among them.

conceptual data schema. Defines the types of data included in an information system; defines the intellectual and logical content of a database; defines the rules for entering and extracting information.

compactness. Characteristic of and evaluation criterion for data schemas. Is information storage space maximized in relationship to intellectual content of a schema?

copy, individual. A bibliographic unit—the physical copy of an edition.

data elements. The elemental components of a record.

data structure. Arrangement of database files and indexes to provide access to data within the entity store of an ISAR system.

database. A system of one or more relations or files. May refer to Information System as a whole or to the entity store of an ISAR system.

descriptor. Any authorized entity identifier that can be used to retrieve other entities; for concepts one uses subject descriptors.

descriptor-find index. An index of subject descriptors and their synonyms and related terms. Used to find authorized descriptors in a thesaurus or classification scheme. Leads from elemental concepts to compound concepts.

disambiguation. The process of determining the particular meaning of a homonym within a text.

discrimination. A measurement the success of an ISAR system in rejecting irrelevant documents for a search. A measurement of how many irrelevant documents are rejected out of all irrelevant documents available.

distributed relatives. Precombined descriptors that share an elemental concept but are scattered in a classification or thesaurus arrangement.

document. Any presentation (written or spoken text, images, music, etc.) of any information in any media (print, screen, radio, etc.).

document template. A frame with a slot for each part of a document. Slots may have procedures attached to obtain information from a database.

domain. For each entity type, what entities are covered in a conceptual data schema.

edition. A bibliographic unit—arguably, the method of rendering a work by means of graphical image or sound, taken in the abstract. More frequently defined as a set of printings of a document from the same type image.

efficiency. Characteristic of and evaluation criterion for data schemas. How much effort is needed to collect, enter, integrate, communicate, and display data in a schema?

elemental descriptor. An elemental concept used as a descriptor.

entity. (1) A word or term (noun) representing the basic concept of a thing or idea. (2) A work or physical object contained in a collection.

entity identifiers. A descriptor designating a type of entity. For instance, a person identifier such as a social security number or a document identifier by acquisition number.

entity-oriented indexing. Indexing that focuses on the entity itself and its faithful description, often with the requirements of shelf arrangement in mind. Concepts useful to users for search retrieval, but not suggested from the entity itself are often not included in indexing.

entity-relationship modeling.

entry. In bibliographic control, an element, such as author name, title, or subject descriptor under which a document may be found in a catalog or index.

essential attributes. Attributes used in the definition of a class. All entities in the class must have these attributes to be members of the class.

exhaustivity. Characteristic of and evaluation criterion for data schemas. How well does the schema cover all the entities and relationships necessary to retrieval and display of expected search requests?

equivalency. In terminological control, having terms that designate closely related or widely overlapping concepts. Terms also called quasi-synonyms.

expert system. An information system that incorporates in its retrieval process procedures that simulate expert reasoning.

extended frame. A frame with slots for more than one relationship.

facet. Aspects or viewpoints from which entities such as documents may be analyzed

fallout. A measurement of how many irrelevant documents are retrieved out of all possible irrelevant documents available for a search.

file. In data structures, a set of records or actual entities. May consist of a pairs relation or many data elements arranged in groupings on several levels.

fixed-field record. Essentially a character string with sections in the string that represent data elements.

flexibility. Characteristic of and evaluation criterion for data schemas. Can the schema adapt to variability between entities?

frame. A data structure that uses hierarchical inheritance for data input and storage. A template provides slots for attributes of frame concepts.

free-text searching. Searching by examining all words in a document to match search terms supplied by a user. Keyword searching.

global performance. Performance of an ISAR system over all search requests. Use as basis of ISAR design.

granularity. A measure of the levels and size of components that make up a system.

hierarchical inheritance. In a classification scheme, where a subclass inherits the attributes of a broader class or class higher in the structure.

homonym. A single word or term that has more than one meaning.

hospitality. Characteristic of and evaluation criterion for data schemas. Can entity and relationship types be added as new needs become known? Can additional levels of detail to the schema?

HTML. A markup language that uses tags in a document direct the display of the document, specify links to other documents to include in the display, or to provide links to other documents.

hypertext. Web-based or computer-based text that uses text blocks as links to other web-based text or documents and organized so that related terms of information are connected.

hyponymy. (1) In lexical relationships, a word or term, having the semantic relationship of hierarchy, being subordinate to a term classified higher. (2) In narrow definition, two sets of words or phrases that have the same spelling or pronunciation. (3) Same word has two different meanings.

identifiers. Entity values represented by symbols, such as

inclusive searching. Searching all entities that are indexed by a descriptor itself or any of its narrower descriptors.

index. A file that facilitates access to records or entities in another file or main file.

index language. Subject descriptors and lead-in vocabulary used in a thesaurus of an information retrieval system to allow searching.

information retrieval. Matching of items or information (i.e. documents, images, objects) in an information system with a user's needs. That is, the act of finding and making available to users a subset of information relevant to solving a user's problems.

information storage and retrieval system (ISAR or IR). A system within a larger Information System. For each input query formulation of user needs the system matches entities or information contained in a collection.

information system. A structure used for the storage, organization, and retrieval of information. Includes such components as type of information included in a collection of information, how the information is structured, who the users of the system are, the technology that provides access to the information, and a system to match information entities with user needs.

importance threshold. How important is it to find this document under this descriptor? A high importance threshold means having few descriptors that are each highly relevant to a search. A low threshold means having more descriptors that are marginally relevant to a search.

input record. A form with blanks to fill for putting information in an ISAR system.

ISAR or IR. *See* information storage and retrieval system.

knowledge representation. An expression of knowledge or information through a system of symbols or signs. For example, words, icons, numbering formats. The “heart” of an Information System.

language processing. Parsing sentences or other text for terms and concepts in order to store and extract information by computer.

lead-in term. A term in a thesaurus that is not a subject descriptor and not a selected preferred term, but leads a user looking for a concept to the chosen subject descriptor.

lead-in vocabulary. The lead-in terms for an index language, including non-selected preferred terms and non-preferred terms.

local performance. Performance of an ISAR system for the one request at hand.

markup. The insertion of codes or tags into a document’s text to specify instructions on how the text should display or how the text should be structured. Used in formatting of text for computer use and display.

mental model. A form of mental visualization of the “state of affairs” of a situation. As a person’s mind receives and interprets new information from the environment or from the imagination, a mental image may be actively reshaped, resulting in an updated image. This process is also called cognition.

metadata. Data used to describe other data. Structured information on the attributes of items like documents—MARC records, for example.

metadata schema. A system for recording metadata for cataloging information or structuring descriptive records.

minimal frame. A frame where each slot represents one point of a relationship and all slots are needed to make the frame complete.

monohierarchy. A tree-shaped hierarchy in which each concept may have only one broader concept.

natural language processing.

NOT, logical. A Boolean search operator. Used to exclude an aspect of a query formulation in a search.

noun phrase. A phrase that functions as the subject of a verb or object of a verb or preposition, formed by a noun and all its modifiers and determiners.

OR, logical. A Boolean search operator. When used to connect two aspects of a query formulation retrieves those documents in which either aspect is contained; a set union.

output record. A display or display format of information retrieved from an ISAR system; should be easily understood by user.

parse. Map a sentence to its syntactic structure and logical form.

pertinent. Document described as pertinent if it is topically relevant to a search and appropriate for the particular user searching.

polyhierarchy. A hierarchy structure in which each concept may have more than one broader concept.

polysemy. The same word or phrase has multiple, related meanings all of which can be traced to a common core of meaning.

postcombination system. An ISAR system that uses elemental descriptors which are then combined in query formulations after indexing is completed.

precision. A measurement of how many relevant documents to a search are retrieved out of all documents retrieved.

precombination system. An ISAR system that uses precombined descriptors so that concepts are combined prior to indexing.

precombined descriptor. A compound concept used as a descriptor.

preferred term. A term chosen from synonyms or quasi-synonyms that is used for searching and that is the focal point where all information about a concept is collected.

primary concept. An idea or concept derived from sensory and motor experiences of the outside world.

proximity operator. Search syntax designating that a word or search term must be within a certain distance of another word or search term within a text.

quasi-synonyms. Terms that have closely related or widely overlapping concepts and are not quite true synonyms. Also called equivalent terms.

query. The statement of a need; the topic of a request containing information informally described in the query statement. It may be one or a combination of several conceptual components such as descriptors and often used with a syntax such as Boolean operations.

query language. The rules for extracting information from the conceptual data schema.

query statement. The statement of a need expressed in natural language terms.

ranking. A measurement of expected relevance of retrieved documents.

recall. A measurement of the success of an ISAR system in retrieving relevant documents or entities for a search. A measurement of how many relevant documents are retrieved out of all relevant documents available.

record. A simple frame with a list of data elements pertaining to a particular focal entity.

record format. Determines how different pieces of information are arranged in a record; specifies which data elements are to be included in a record.

relation. A set of ordered pairs of entities within a data structure.

request-oriented indexing. Indexing based on analyzing the anticipated queries of a set of users of an information system.

relevance. How well retrieved documents meet the information needs of a user.

reliability. Characteristic of and evaluation criterion for data schemas. How free from human or machine error are the processes in a data schema?

request-oriented indexing. Using anticipated queries and the user's point of view as a frame of reference for analyzing entities for indexing.

retrieval. Returning from a collection of entities a subset that serves a user's purpose. Can mean finding documents that meet the requirements of a Boolean search. More broadly, an information system should use all available evidence to predict the degree of relevance of an entity to a particular user in a particular situation. Evidence might include title words, abstracts, syntactic relationships, etc.

schema. A conceptual structure that arranges parts in some order, showing interrelationships.

secondary concept. Concepts extracted from other concepts.

semantic analysis. In linguistic analysis, disambiguating synonyms and homonyms to extract elementary concepts.

semantic factoring. The process of analyzing a compound concept into its elementary components.

SGML. A markup meta-language developed for the publishing industry that helped in creation of HTML and to generate XML.

specificity. Characteristic of and evaluation criterion for data schemas. Are entities and relationships defined at the appropriate level of detail for retrieval and display of useful information?

spreading activation. As a topic in an information network is activated, topics directly linked are also activated and so on as activation flows across relationships in a network...

statement. An association between two or more entities, in which the entities are in a given relationship.

synonym. A word or phrase with a meaning that is the same as, or very similar to, another word or term. One concept may be represented by different words or terms.

syntactic analysis. Analysis of text for parts of speech and to resolve anaphoric references.

template. A preset format for a document or markup language, essentially a frame with slots for entity values. Used so that a document format or other entity does not need to be recreated each time it is used.

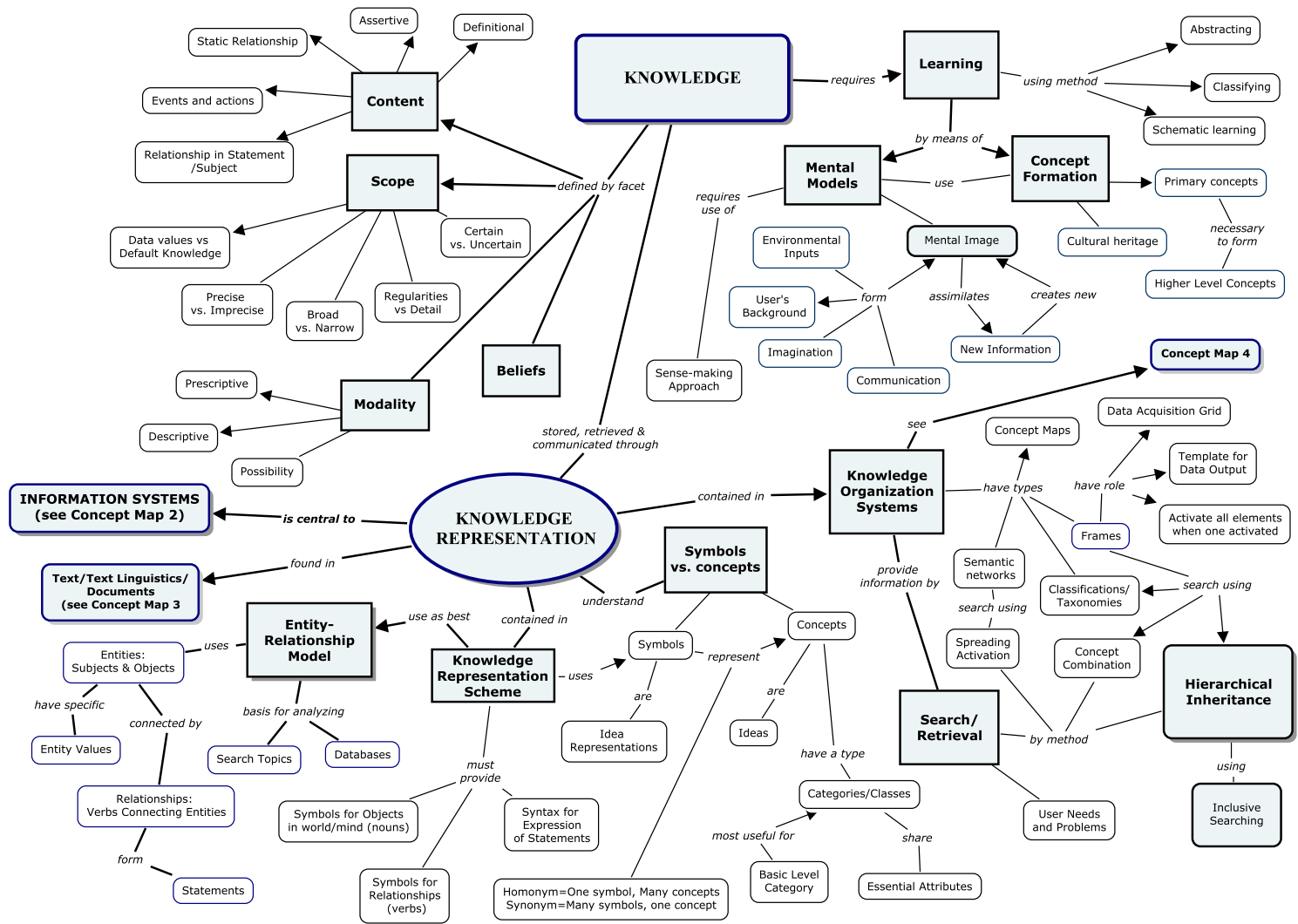
terminological control. Using controlled vocabularies with predefined, authorized terms to organize knowledge for retrieval in indexes, subject headings, classifications and thesauri.

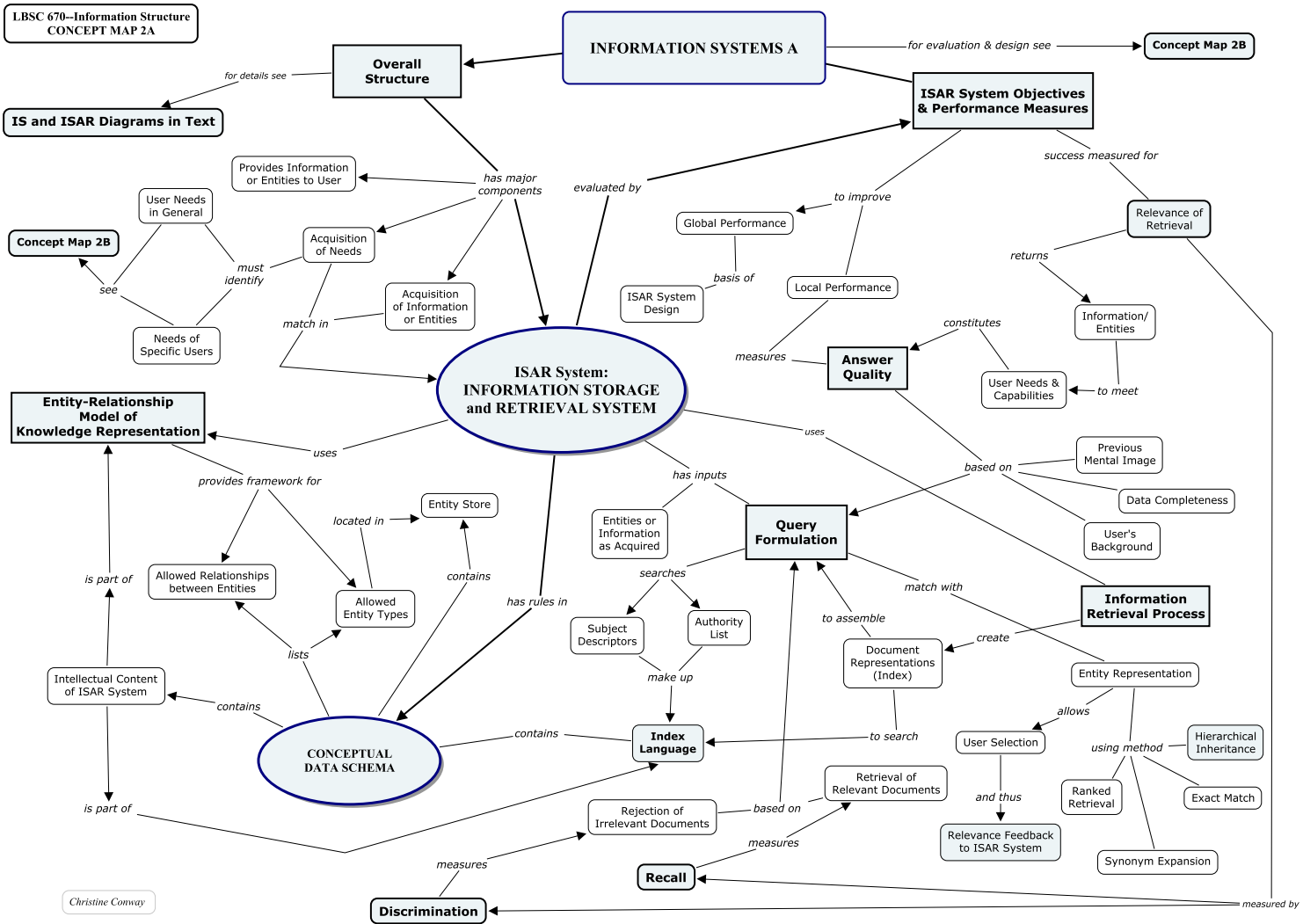
thesaurus. An aid for searching and for controlled vocabulary for indexing that includes subject descriptors selected from preferred terms. Also includes remaining preferred terms and non-preferred terms as lead-in vocabulary. A thesaurus groups terms that are synonymous and disambiguates homonyms.

weight. The relative importance granted to something.

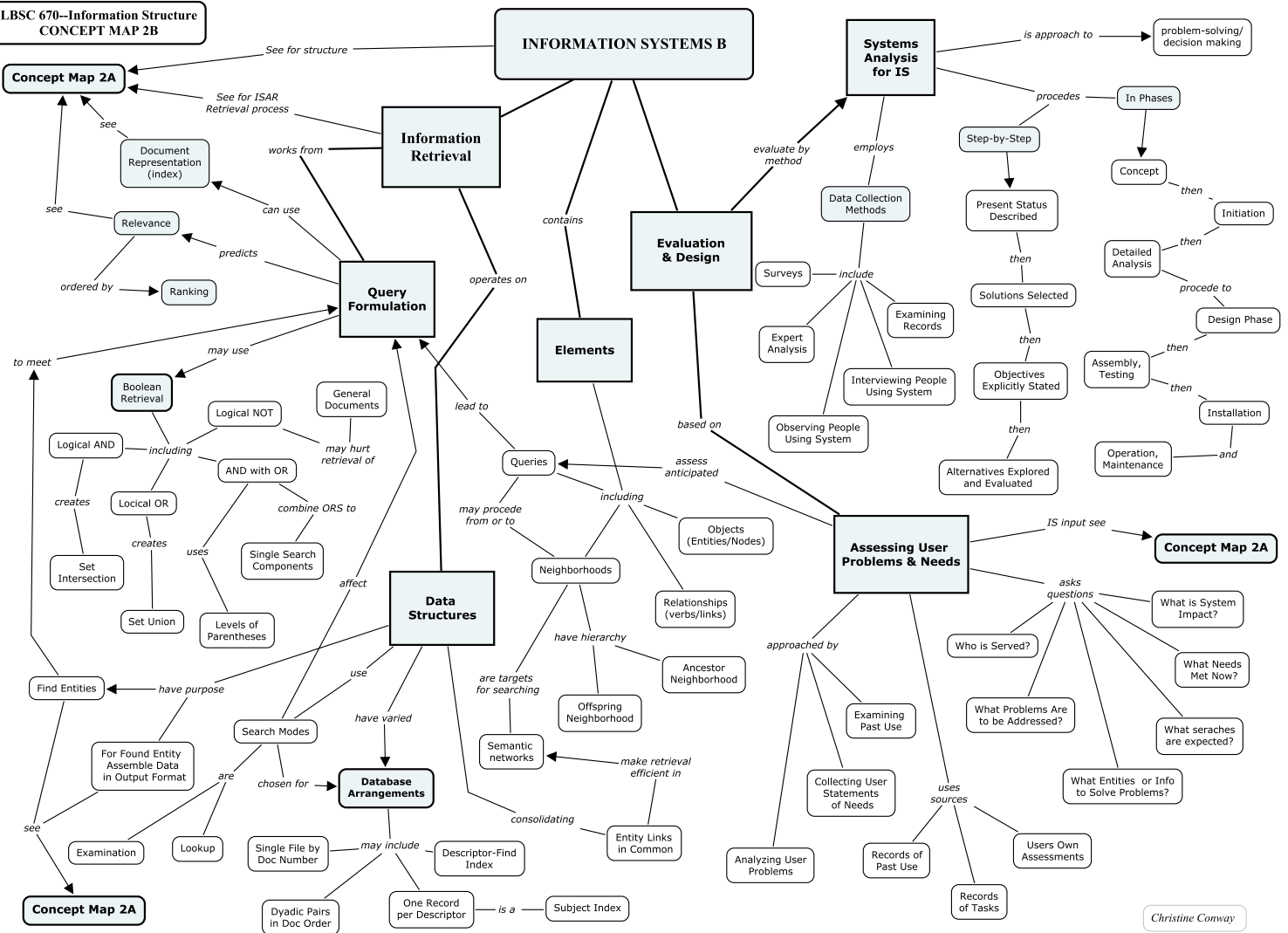
work. A bibliographic unit—arguably, an artistic or intellectual conceptual entity comprised of the essence of the creative process of a text, image, music composition, etc. Has been more frequently defined as a set of related texts with a common origin and context.

XML. A markup meta-language used to define one's own tags to create one's own markup language.





**LBSC 670--Information Structure
CONCEPT MAP 2B**



Christine Conway

