

Glossary

abbreviation

A shortened form of a name or term (e.g., *Mr.* for *Mister*). *See also* **acronym** and **initialism**.

access point

An entry point to a systematic arrangement of information, specifically an indexed field or heading in a work record, a vocabulary record, or another content object that is formatted and indexed in order to provide access to the information in the record.

acronym

An abbreviation or word formed from the initial letters of a compound term or phrase (e.g., MoMA, for Museum of Modern Art). *See also* **abbreviation** and **initialism**.

ad hoc query

Also called a **direct query**. A query or report that is constructed when required and that directly accesses data files and fields that are selected only when the query is created. It differs from a **predefined report** or querying a database through a **user interface**.

administrative data

In the context of cataloging art, information having to do with the administrative history and care of the work and the history of the catalog record (e.g., insurance value, conservation history, and revision history of the catalog record). *See also* **descriptive data**.

administrative entity

In the context of a geographic vocabulary, a political or other administrative body defined by administrative boundaries and conditions, including inhabited places,

empires, nations, states, districts, and townships. *See also* **physical feature**.

algorithm

In the context of this book, an algorithm is a procedure, a formula, or the rules in a computer program or set of programs, often expressed in algebraic notation, that follow a logical, unambiguous step-by-step process to retrieve a set of results, solve a problem, make a decision, manipulate or alter data, or achieve some other result or state. Although a computer program may be considered one large algorithm, in common usage in computer science, the term typically refers to a small procedure applied recurrently. *See also* **computer program**.

alphanumeric classification scheme

A set of controlled codes (letters or numbers or both) that represent concepts or headings and generally have an implied taxonomy that can be surmised from the codes (e.g., the Dewey Decimal Classification system number 735.942). *See also* **chain indexing**.

alternate descriptor (AD)

A variant form of a descriptor available for use; usually a singular form or a different part of speech than the descriptor (e.g., *lithograph* is an alternate descriptor for the plural descriptor *lithographs*). In thesauri, the relationship indicator for this type of term is AD.

ancestor

In a hierarchy, any record that is a broader context for the record at hand, including parents, grandparents, and all other broader contexts at higher levels; any node in the succession of parent nodes on a path all the way up to the root. *See also* **descendant**.

antonym

A term that is the opposite in meaning of another term (e.g., *roughness* is an antonym for *smoothness*).

application

Also called an **application program**. A software program designed to accomplish a task for an end user (e.g., word processing or project management), as distinguished from the **operating system** program that runs the computer itself.

application programming interface (API)

In the context of this book, an online system, source code, and interface that a data provider (e.g., a vocabulary provider or library) employs to allow users to have access to the data. It may be language dependent (designed for a specific programming language) or language independent (works with multiple programming languages).

architect

A person or firm involved in the design or creation of structures or parts of structures that are the result of conscious construction, are of practical use, are relatively stable and permanent, and are of a size and scale appropriate for—but not limited to—habitable buildings.

architectural work

See **built work**.

architecture

Refers to the built environment that is typically classified as fine art, meaning it is generally considered to have aesthetic value, was designed by an architect, and was constructed with skilled labor. See also **built work**.

archival group

See **group**.

art

In the context of this book, refers to the **visual arts** such as painting, sculpture, drawing, printmaking, photography, ceramics, textiles, and decorative arts of the type and caliber generally collected

by museums. Performance art is also included, but the performing arts are not. Note that these are works of visual art of the *type* collected by art museums. The objects themselves may actually be held by an ethnographic, anthropological, or other museum, or owned by a private collector.

artist

Any person or group of people involved in the design or production of visual arts that are of the type collected by art museums.

ascending order

In the context of a string of hierarchical parents, refers to the display of parents from narrowest to broadest (e.g., *Columbus (Bartholomew county, Indiana, United States)*). See also **descending order**.

ASCII

Acronym for the American Standard Code for Information Interchange, a 7-bit character code defining 128 characters used for information interchange, data processing, and communications systems.

associative relationship

In a thesaurus, the relationship between concepts that are closely related conceptually, but the relationship is not hierarchical because it is not whole/part or genus/species. The relationship indicator for this relationship is RT (for **related term**). See also **equivalence relationship** and **hierarchical relationship**.

asymmetric relationship

In the context of a thesaurus, refers to a reciprocal relationship that is different in one direction than it is in the reverse direction—for example, BT/NT (for **broader term/narrower term**). See also **symmetric relationship**.

authoritative source

A published source that is based on reliable documentary evidence that is accepted as true by most experts and used as a standard source in a given discipline.

authority file

Also called simply an **authority**. A file, typically electronic, that serves as a source

of standardized forms of names, terms, titles, etc. Authority files should include references or links from variant forms to preferred forms. The main purpose of an authority is to enforce usage, often requiring users to use only the preferred term for a given concept. Any type of vocabulary can be used as an authority. *See also* **controlled vocabulary** and **local authority**.

authority heading

A preferred, authorized heading used in a vocabulary, particularly in a bibliographic authority file that typically includes a string of names or terms, with additional information as necessary to allow disambiguation between identical headings (e.g., *United States—History—Civil War, 1861–1865—Battlefields* and *United States—History—Civil War, 1861–1865—Campaigns*). The types of authority headings used by the Library of Congress are the following: subject, name, title, name/title, and keyword authority headings. *See also* **heading**.

authorization

In the context of vocabularies, the process by which the creators of a vocabulary or an oversight group regulate the selection of terms and establishment of relationships in a controlled vocabulary. *See also* **warrant**.

automatic indexing

In the context of online retrieval, indexing by the analysis of text or other content using computer algorithms. The focus is on automatic methods used behind the scenes with little or no input from individual searchers, with the exception of relevance feedback. The results tend to be broad and imprecise, as contrasted to human **indexing**. *See also* **co-occurrence mapping**.

autoposting

See **up-posting**.

batch load

In the context of populating or contributing to vocabulary systems or other databases, refers to moving or manipulating a group of records as a single unit for the purpose

of data processing, typically accomplished by the computer without user interaction, as contrasted to entering records manually, one at a time. *See also* **load** and **processing**.

batch processing

See **processing**.

best match

Also called a **weighted term ranking**. Refers to a variety of electronic term-matching and ranking methods that attempt to predict the potential relevance of query results by assigning relevance scores and ranking based on comparing search terms to the indexing terms of the target database. *See also* **exact match**.

blind reference

In the context of a vocabulary that is being used for indexing or retrieval on a defined data set, refers to a term in the vocabulary that is not linked to any content in the data set. End users should typically not receive blind references in a retrieval situation because they result in a failed search; however, these terms should be retained in structured vocabularies that are used for indexing because they may be needed in the future or in another context.

Boolean operators

Logical operators used as modifiers to refine the relationship between terms in a search. The four most commonly used Boolean operators are AND, OR, NOT, and ADJ (adjacent). They may be used with parentheses and other punctuation to form logical groupings of criteria in queries (e.g., *Castillo OR Rancho AND Diego*).

bound term

A compound term representing a single concept, characterized by the fact that the words almost always occur together and the meaning is lost or altered if the term is split into its component words. *See also* **compound term** and **lexical unit**.

brand name

A trade or proprietary name for a thing or process (e.g., *Super Glue*).

broadcast searching

See **federated searching**.

broaden results

To adjust criteria in a search in order to retrieve a larger number of results, typically because the searcher did not find what he or she wanted in an initial narrower search. See also **narrow results**.

broader term (BT)

Also called a **broader context**. A vocabulary record to which another record or multiple records are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT. Variations on the notation include BTG (broader term generic), BTP (broader term partitive), BTI (broader term instance), BT1 (broader term level 1), BT2 (broader term level 2), etc.

browsing

The process whereby a user of a system or Web site visually scans and maneuvers through navigation lists, results lists, hierarchical displays, or other content in order to make a selection, as contrasted to the user entering a search term in a search box. See also **searching**.

built work

An instance of architecture, which includes structures or parts of structures that are the result of conscious construction, are of practical use, are relatively stable and permanent, and are of a size and scale appropriate for—but not limited to—habitable buildings. Built works in the context of art information are manifestations of the built environment typically classified as fine art, meaning it is generally considered to have aesthetic value, was designed by an architect (whether or not his or her name is known), and constructed with skilled labor. See also **architecture** and **movable work**.

candidate term

Also known as a **provisional term**. A term under consideration for admission into a controlled vocabulary because of its potential usefulness. See also **contribution**.

cataloger

In the context of this book, the person who records information in records for works. See also **end user** and **indexer**.

cataloging

In the context of this book, the process of describing and indexing a work or image, particularly in a collections management system or other automated system. Cataloging involves the use of prescribed fields of information and rules (e.g., the rules described in *CCO* and *CDWA*).

cataloging rules

See **editorial rules**.

cataloging tool

A system that focuses on content description and labeling output (e.g., wall labels or slide labels), often part of a more complex **collection management system**.

chain indexing

Also called **chain procedure**. A technique for indexing that uses a numeric or **alpha-numeric classification scheme**—for example, the Dewey Decimal Classification system—where the entries have meaning beyond simple numeric sequencing (e.g., in Dewey number 735.942, 735 means *sculpture after the year 1400 CE*, 9 means *geographic area*, 4 means *Europe*, and 2 means *England*).

child

See **narrower term**.

classification

In the context of this book, the process of arranging works or other content objects systematically in groups or categories of shared similarity according to established criteria and using terms to identify the classes.

classification notation

In a vocabulary, a numeric, alphabetic, or alphanumeric code in a system of codes used to classify or categorize entries; may be used in a hierarchical arrangement to impose a display or sorting order on the lines or levels in the hierarchy (e.g., V, V.PC, V.PE). See also **notation**.

classified display

See **hierarchical display**.

clustering

In the context of automated data, usually refers to the process of grouping or classifying items or data through automatic or algorithmic means rather than incorporating human judgment.

code

See **computer code**.

collection

In the context of cataloging art, refers to multiple works that are physically or conceptually arranged together, including the entire set of objects curated by a given museum or other repository.

collection management system (CMS)

A type of database system that allows an institution to control various aspects of its collections, including description (artist, title, measurements, media, style, subject, etc.) as well as administrative information regarding acquisitions, loans, and conservation information.

complex term

A single phrase denoting more than two distinct concepts, which could be broken out and used independently, as defined by the Library of Congress. See also **bound term**, **compound term**, and **heading**.

component

In the context of cataloging art and architecture, a part of a larger item. A component differs from an item in that an item can stand alone as an independent work, but a component typically cannot or does not stand alone (e.g., a panel of a polyptych or a façade of a basilica). See also **group** and **item**.

compound term

A term consisting of two or more words. In the context of this book, mention of compound terms generally refers to bound terms, which are compound terms that represent a single concept (e.g., *flying*

butresses). See also **bound term**, **complex term**, and **lexical unit**.

computer code

Also called **code**. The machine-readable form, arrangement of data, and instructions of a computer program that are created when a computer program, which was written by a human programmer, is converted into binary code that can be read by a computer.

computer program

Also called a **program**. A specific set of instructions for ordered operations that result in the completion of a task by the computer; a computer program consists of **computer code**. While the program is technically a type of data, computer programs are generally considered as separate from the data to which the programs refer (e.g., data would be the terms, scope notes, etc., in a vocabulary record). A program is *interactive* if it acts when prompted by an action or information supplied by a user, or *batch* if it automatically runs at a certain time or under certain conditions and then stops after the task is completed. A program is written in a **programming language**. See also **processing**.

computer system

See **system**.

concept

In the context of the *AAT* and other thesauri comprising generic terms, the subject of the vocabulary record (i.e., the concept to which the terms refer), including abstract concepts; physical attributes such as shape, pattern, and color; style or period; activities; terms for performers of activities; materials; objects; and visual and verbal communication forms. See also **discrete concept**.

concept record

See **record**.

conceptual data model

An abstract model or representation of data for a particular domain, business enterprise, field of study, etc., independent of

any specific software or information system; usually expressed in terms of entities and relationships. *See also* **logical data model**.

content object

In the context of a database, any entity that contains data. A content object can itself be made up of content objects. For example, a journal is a content object made up of individual journal articles, which are themselves content objects. *See also* **information object**.

contribution

In the context of controlled vocabularies, a term or record that is submitted for admission into a thesaurus or other vocabulary by an agency or individual outside the group responsible for maintaining the vocabulary; contributions are typically made by users of the vocabulary. *See also* **candidate term**.

controlled field

In the context of this book, a field in a record that is not free text, meaning it is specially formatted and often linked to controlled vocabularies (authorities) or controlled lists to allow for successful retrieval. *See also* **free-text field**.

controlled format

Rules applied to the field regarding the types of values that may be included (e.g., a controlled measurement's value field would allow only numbers). Fields may have controlled format in addition to being linked to controlled vocabulary, or the controlled format may exist in the absence of any finite controlled list of valid values.

controlled list

A simple list of terms used to control terminology. In a well-constructed controlled list, terms should be unique, members of the same class, not overlapping in meaning, equal in granularity/specificity, and arranged alphabetically or in another logical order. A type of **controlled vocabulary**.

controlled vocabulary

An organized arrangement of words and phrases used to index content and/or

to retrieve content through browsing or searching. A controlled vocabulary typically includes preferred and variant terms and has a limited scope or describes a specific domain.

co-occurrence mapping

Also called **co-occurrence clustering**. An automated method of compiling groups of terms that tend to occur together in certain contexts and are therefore presumed to be related in some way; the resulting groups of terms are considered to be loosely related and may be used to automatically broaden a user's search or to suggest alternative search terms to users in order to improve search results. *See also* **automatic indexing**.

core fields

Also called **core elements**. In the context of this book, the set of fields representing the fundamental or most important information required for a minimal record, whether the record is a work record or a vocabulary record. *See also* **required fields**.

corporate body

In the context of vocabularies discussed in this book, an organized, identifiable group of individuals working together in a particular place and within a defined period of time, whether or not they are legally incorporated (e.g., architectural firms, artist studios, and art repositories).

criteria

In the context of this book, a specific set of limiting conditions used to create a query or select a subset of entries (e.g., a *WHERE statement* in SQL). *See also* **variable**.

cross-database searching

See **federated searching**.

cross-reference links

See **syndetic structure**.

crosswalk

A chart or table (visual or virtual) that represents the semantic or technical mapping of fields or data elements in one database, metadata framework, standard, or schema to fields or data elements that have a

similar function or meaning in one or more other databases, frameworks, standards, or schemas (e.g., the *artist* element in one standard may map to the *creator* element in another). *See also* **mapping**.

cultural heritage

The total corpus of activities and the artifacts of activities that provide a record of the life of a culture. *See also* **material culture**.

cultural works

In the context of this book, art and architectural works and other artifacts of cultural significance, including both physical objects and performance art. In related disciplines, the scope could be broader, also including the performing arts.

data

In common usage in computer science, this term is used as a singular noun to refer to information that exists in a form that may be used by a computer, excluding the program code. In other uses, *datum* is the singular and *data* is the plural, referring to facts or numbers in a general sense.

database

A structured set of data held in computer storage, especially one that incorporates software to make it accessible in a variety of ways. A database is used to store, query, and retrieve information. It typically comprises a logical collection of inter-related information that is managed as a unit, stored in machine-readable form, and organized and structured as records that are presented in a standardized format in order to allow rapid search and retrieval by a computer. *See also* **system**.

database field

Also called a **data field**. A placeholder for a set of one or more adjacent characters comprising a unit of information in a database, forming one of the searchable items in that database. It is a portion of a structured record, especially a machine-readable record, containing a particular category of information (e.g., *term* and *scope note*

would be fields included in a vocabulary record). *See also* **field**.

database index

Also called a **data index**. A particular type of data structure that improves the speed of operations in a table by allowing the quick location of particular records based on key column values. Indexes are essential for good database performance. The concept is distinguished from **indexing** (human indexing) and **automatic indexing**.

database normalization

See **normalization**.

database record

See **record**.

data content

The organization and formatting of the words or terms that form **data values**.

data elements

The specific categories or types of information that are collected and aggregated in a database.

data preprocessing

See **preprocessing**.

data processing

See **processing**.

data structure

A given organization of data, particularly the data elements, the logical relationships between data elements, and the storage allocations for the data.

data table

Sets of data that are organized in a grid or matrix comprising rows and columns.

data values

In the context of this book, the terms, words, or numbers used to populate fields in a work or vocabulary record. *See also* **data content**.

decoordination

In the context of a thesaurus, the splitting of a **compound term** into its component words to stand as individual terms. This would typically happen if a compound term

had been added to the thesaurus but was later determined not to be a **bound term**.

deep Web

See **hidden Web**.

derivation

Also called **modeling**. In the context of this book, the process of building a new vocabulary based on an existing vocabulary. In this approach, an appropriate controlled vocabulary is selected as a model for developing controlled terminology for local use, so that the local terms will be interoperable with the larger original vocabulary. See also **local authority** and **microcontrolled vocabulary**.

descendant

Also often spelled **descendent** in the disciplines of computer science and thesaurus construction. In a hierarchy, any record that is a narrower context for the record at hand, including children, grandchildren, and all other narrower contexts at all lower levels; any node in the succession of parent nodes on a path all the way down to the tips (leaves) of the hierarchies. See also **ancestor**.

descending order

In the context of a string of hierarchical parents, the display of parents from broadest to narrowest (e.g., *Columbus (United States, Indiana, Bartholomew county)*). See also **ascending order**.

descriptive data

In the context of cataloging art, data intended to describe and identify a work, as contrasted to information necessary for administrative, technical, or accounting purposes. See also **administrative data**.

descriptor (D)

In a thesaurus, the term recommended to represent the concept in displays and indexing. Also called the **main term**, **postable term**, or **preferred term** in a monolingual thesaurus. A multilingual thesaurus may have multiple descriptors (one in each language represented) but may possibly have only one preferred term for

use as a default in displays. In thesauri, the relationship indicator for this type of term is D.

diacritics

Also called **diacritical marks**. Signs or accent marks found over, under, or through alphabetic letters in many languages (e.g., the umlaut in German, *München*), used to indicate emphasis or pronunciation, often to distinguish different sounds or values of the same letter or character without the diacritical mark.

digital asset management system (DAMS)

A type of system for organizing digital media assets, such as digital images or video clips, for storage and retrieval. Digital asset management systems sometimes incorporate a descriptive data cataloging component, but they tend to focus on managing workflow for creating digital assets and for managing asset rights, requests, and permissions.

direct mapping

In the context of interoperability of vocabularies, refers to the matching of terms one-to-one in two controlled vocabularies. While the vocabularies need not be the same size or cover exactly the same content, where overlap exists, there should be the same meaning and level of specificity between the two terms in each controlled vocabulary. See also **switching**.

direct query

See **ad hoc query**.

disambiguation

In the context of creating and displaying a vocabulary, the use of qualifiers, headings, or other methods to clarify and remove ambiguity between homographs (e.g., *Smith, John (English printmaker, 1654–1742)* and *Smith, John (English architect, 1781–1852)*). See also **word sense disambiguation**.

discrete concept

In the context of a generic concept vocabulary, a discrete thing or idea as opposed to

a subject heading, which often concatenates multiple terms or concepts together in a string. *See also* **concept**.

displayed index

An index that is visible and available to end users for browsing. *See also* **nondisplayed index**.

display field

In the context of this book, a field intended for viewing by the end user, typically showing data in natural language that is easily read and understood and that can convey nuance and ambiguity. Display information may, in some cases, be concatenated from controlled fields; in other cases, this information is best recorded in free-text fields. *See also* **indexing**.

document

In the context of search and retrieval, the combination of a defined, primarily self-contained, machine-readable text or other information and the format in which it is housed.

dominant language

In the context of multilingual vocabularies, the more prominent or original language to which terms in other languages are mapped and in which other fields in the record (e.g., scope notes or date notes) are written. In a purely multilingual vocabulary, no language is dominant, but in a rich and complex vocabulary (e.g., the *AAT*), a dominant language may be required for practical purposes.

download

See **load**.

editorial rules

In the context of this book, written rules and guidelines for creators or editors of vocabulary records that dictate how to populate fields and choose or interpret data. They should include which fields are required, how to choose appropriate values for various fields (e.g., how to choose a preferred term), how to choose hierarchical positions, the format and syntax for each field, authorized sources, etc. Analogous

rules for catalogers of works are called **cataloging rules**.

end user

In the context of this book, usually the searcher, client, or patron who retrieves, views, and uses the data in a vocabulary or work record, as distinguished from the editors or catalogers. In the context of systems design, the term refers to any client for whom a database system is designed and used; from that perspective, it could include the editors or catalogers for whom an editorial or cataloging system has been designed.

end-user thesaurus

A thesaurus designed for direct access by searchers rather than for use by indexers. Instead of controlling the terminology, the purpose of an end-user thesaurus is to help searchers find useful terminology for improving, narrowing, and broadening their queries. *See also* **indexer thesaurus**.

entity

In the context of computer science, a self-contained piece of data that can be referenced as a unit. In a more general sense, the term is used in this book to refer to a distinct person, place, or concept in a vocabulary.

entity-relationship model

A type of conceptual data model that represents structured data in terms of entities and relationships. An **entity-relationship diagram** can be used to visually represent information objects and their relationships. Because the constructs used in the entity-relationship model can easily be transformed into relational tables, this type of model is often used in database design.

entry array

A type of display, often used for headings, in which any two or more entries that have the same broader heading (e.g., *Religious art—Ancient Egyptian, Religious art—Christian, Religious art—Hindu*, etc.) are grouped together vertically under the broader heading. While this is not a true hierarchical display, it may resemble a hierarchical display through use of indentation.

equivalence relationship

In a thesaurus, the relationship between synonymous terms or names for the same concept, typically distinguishing **preferred terms** (descriptors) and **nonpreferred terms** (variants or UFs). *See also* **associative relationship** and **hierarchical relationship**.

equivalent term

A term that is considered an equivalent in search-and-retrieval, including not only true synonyms but possibly also near-synonyms and any other terms that are considered closely enough related to be useful in broadening a query; to narrow a query, exact equivalents could be used instead.

exact equivalence

The relationship between synonyms in one language and terms in different languages that have the same usage and meaning. *See also* **inexact equivalence** and **nonequivalence**.

exact match

Electronic term-matching that produces a result that precisely matches the user's query term and does not implement automatic Boolean operators, truncation, proximity ranges, or stemming. In a strictly applied exact match, **normalization** is not used, so that differences in punctuation, spacing, and diacritics are maintained in the match. *See also* **best match**.

exhaustivity

In the context of cataloging and indexing, the degree of depth and breadth that the cataloger uses in assigning indexing terms or writing a description. Measures of greater exhaustivity include the use of a greater number of optional fields and the assignment of a greater number of indexing terms for each field. *See also* **specificity**.

expansion

See **query expansion**.

explode a hierarchy

To retrieve and display all the descendants of any given node, typically in a graphic display.

extension vocabulary

A thesaurus that is created with the intention of, or is later adapted for, linking to another vocabulary that is larger, broader, or more generic; the extension vocabulary is typically linked through node linking, rather than being integrated at many points in the original vocabulary. *See also* **micro-controlled vocabulary**, **node linking**, and **satellite vocabulary**.

external node

See **leaf node**.

facet

Also called a **faceted display**. A fundamental, homogeneous, and mutually exclusive category of information in a thesaurus (e.g., the *AAT* has seven facets: Associated Concepts, Physical Attributes, Styles and Periods, Agents, Materials, Activities, and Objects).

facet indicator

A **node label** that designates a **facet**.

false hit

Also called a **false drop**. In search and retrieval, an entry in a list of results that does not comply with the user's intended results.

federated searching

Also called **broadcast searching**, **cross-database searching**, **metasearching**, and **parallel searching**. Performing queries simultaneously across resources that are in different domains and created by different communities. Federated searching may involve searching across multiple databases, different platforms, and varying protocols, thus requiring the application of **interoperability** between resources and vocabularies.

field

In the context of this book, an area (often mapping to a metadata element in a metadata element set) in the user interface of a system where a discrete unit of information is displayed or the cataloger can enter information. Note: In this context, *field* is not necessarily equivalent to a **database field**.

filing rules

A set of guidelines that determine how letters, numbers, spaces, and special characters should be processed when assembling an alphabetical or other listing. *See also* **sorting**.

first name

Also called a **given name**. In Western tradition, the name of a person that identifies that individual, typically unique in the immediate family and used with a last name (e.g., *Richard* in *Richard Meier*). *See also* **last name** and **middle name**.

flat-file database

A database with a data model designed around a single table, often a single file containing many records that all have exactly the same fields. It is a simpler model than the more highly structured relational and object-oriented models.

flat format

In the context of a thesaurus, an alphabetical display in which only one level of broader contexts and one level of narrower contexts are displayed for each focus record. *See also* **generic structure**.

focus

Also known as a **head noun** for terms and a **trunk name** for proper names.

In the context of a compound term, the noun component that identifies the class of concepts to which the term as a whole refers (e.g., *buttresses* in the term *flying buttresses*). In the context of a modified name such as a place name, the part of the name that is not a modifier (e.g., *Etna* in *Mount Etna*). *See also* **modifier**.

folksonomy

A neologism referring to an assemblage of concepts, which are represented by terms and names (called *tags*) that are compiled through **social tagging**, generally on the Web. A folksonomy differs from a **taxonomy** in that it is not structured hierarchically, and the authors of the folksonomy are typically the casual users of the content rather than professional indexers

following standard protocols and using standardized controlled vocabularies.

format

Used in two senses in this book. In the context of cataloging art, the configuration of a work—including technical formats—or the conventional designation for the dimensions or proportion of a work (e.g., *cabinet photograph* or *IMAX*). In the context of computer science, the physical layout of a data storage device or the logical structure or composition of a file.

format control

See **controlled format**.

free-text field

A field that may contain data entered without any vocabulary control or system-defined structure. It may be used to express ambiguity, uncertainty, and nuance in a note. *See also* **controlled field** and **text**.

generic concept

In the context of this book, a concept in a vocabulary that is described by terms other than proper nouns or names (e.g., the type of artwork, such as *amphora*, or a material, such as *terracotta*). Generic concepts do not include proper names of persons, organizations, geographic places, named subjects, or named events.

generic posting

In controlled vocabularies, the use of narrower terms as *used for* terms for a descriptor that is really a broader term in the same vocabulary record. A generic posting is typically used as a time-saving strategy rather than making separate records for all the terms and linking them hierarchically. *See also* **up-posting**.

generic structure

A display format for a thesaurus in which all hierarchical levels are displayed by using indentation, codes, or punctuation marks. *See also* **flat format**.

genus/species relationship

Also called a **generic relationship**. A hierarchical relationship in which all

children must be a *kind of, type of, or manifestation of* the parent. The genus/species relationship is the most common hierarchical relationship in thesauri and taxonomies, because it is applicable to a wide range of topics. *See also* **instance relationship** and **whole/part relationship**.

given name

See **first name**.

gloss

See **qualifier**.

grandparent

In a thesaurus, the level immediately above the parent of the focus record (e.g., in the following example, *Indiana* is the grandparent of *Columbus: Columbus, Bartholomew county, Indiana, United States*).

granularity

See **specificity**.

group

Also called an **archival group** or **record group**. In the context of cataloging works, refers to an aggregate of items that share a common provenance. *See also* **component** and **item**.

group-level cataloging

Describing and assigning indexing terms for a group of works as a whole, typically focusing on the most important or most frequently occurring characteristics in the items of the group. *See also* **item-level cataloging**.

guide term

A **node label** that is not a **facet**, but is created as a hierarchical level to provide order and structure to thesauri by grouping narrower terms according to a given logic. Guide terms are not used for indexing and are often enclosed in angled brackets or otherwise distinguished from other terms in displays (e.g., *<photographs by form>*).

hardware

The physical components of a computer system, including those that are mechanical, electronic, magnetic, and electrical

such as disks, disk drives, chips, electronic circuitry, keyboards, monitors, modems, and printers. *See also* **software**.

harmonization

In the context of vocabularies and standards, the process of preventing, minimizing, or eliminating technical and content differences and contradictions between standards or vocabularies that have the same or similar scope or that must work interchangeably or in concert.

heading

Also called a **label**. A string of words comprising a term combined with other information that serves to modify, disambiguate, amplify, or create a context for the main term in displays. Examples include the listing of qualifiers and/or broader contexts for terms (e.g., *rhyta (<vessels for serving and consuming food>, containers)*), place types and administrative broader contexts for place names (e.g., *Dayr al-Bahri (deserted settlement) (Qinā governorate, Egypt)*), or biographical information for people's names (e.g., *Francesco Aliunno (Italian calligrapher, active 15th century)*). *See also* **authority heading**, **name authority**, and **subject heading list**.

head noun

See **focus**.

hidden Web

Also called the **deep Web** or **invisible Web**. The sum of the Web pages that are not accessible to Web crawlers or robots, usually because they are either dynamically generated by a user querying a database or are password protected or subscription based.

hierarchical display

Also called a **classified display** or **systematic display**. In a thesaurus, a graphic arrangement of terms showing broader/narrower relationships through the use of indentation, codes, or another method.

hierarchical relationship

The broader and narrower (parent/child) relationship between two entities

in a thesaurus, namely whole/part (e.g., *Montréal* is part of *Québec*), genus/species (e.g., *bronze* is a type of *metal*), or instance relationships (e.g., *Montréal* is an instance of a *city*). It is the basic structure that creates a **hierarchy**.

hierarchy

An organization of records related by levels of superordination and subordination. Each record in the hierarchy, except the root, is a narrower context of the record above it. *See also monohierarchy, polyhierarchy, and subfacet.*

historical term

Also called a historical name. In the context of the vocabularies discussed in this book, a term or name that was used to refer to a person, place, subject, or concept in the past, but in current usage has been replaced with a different term or name (e.g., historical names for *St. Petersburg, Russia*, are *Leningrad* and *Petrograd*).

hits

See results list.

homograph

A term that is spelled the same as another term, but the meanings of the terms are different (e.g., *drums* can have at least three meanings: components of columns, membranophones, or walls that support a dome). Homographs exist whether or not the terms are pronounced alike. Terms are generally considered homographs despite differences in capitalization, punctuation, or diacritics. *See also qualifier.*

homophone

A term that is pronounced like another term but spelled differently (e.g., *bows* and *boughs*). Homophones are not typically labeled in traditional controlled vocabularies.

human indexing

See indexing.

hyperlink

Also called a hypertext link. In the context of online information, an embedded link that connects different parts of an online document or data set to other parts

of the document or to other documents. It is usually indicated by color or other emphasis applied to a word, phrase, icon, or symbol.

hypertext database

A dataset that resides as a collection of online documents with links joining various parts to each other, with access provided via an interactive browser.

Hypertext Markup Language (HTML)

A markup language used to create the layout and presentation of documents for World Wide Web applications.

image

In the context of cataloging art, a visual representation of a *work*, usually existing in a photomechanical, photographic, or digital format. In a typical visual resources collection, an *image* is a slide, photograph, or digital file.

indentation

Also called indention. In the context of printing or other displays of typed words or texts, refers to the white or blank space of a fixed width on a row along the right or left margin of a display, as commonly used to indicate the first line in a new paragraph of text. Graduated indentation is used to indicate relationships between parents and their descendants in hierarchical displays of thesauri.

indexer

A person who assigns indexing terms for a work or image, typically the same person as the cataloger. *See also cataloger.*

indexer thesaurus

A thesaurus designed to control terminology and guide indexers in the choice of terms. *See also end-user thesaurus.*

indexing

Also called human indexing and manual indexing. In the context of this book, the process of evaluating information and designating indexing terms by using controlled vocabulary that aids in finding and accessing the cultural work record. Refers to indexing done by human

labor, not to the automatic parsing of data into a database index (**automatic indexing**), which is used by a system to speed up search and retrieval.

inexact equivalence

The relationship between synonyms in one language or terms in different languages that have similar or overlapping meaning and usage but are not true synonyms (e.g., *floating* and *flying*). *See also* **exact equivalence**, **nonequivalence**, and **partial equivalence**.

information object

A digital unit or group of units, regardless of type or format, that a computer can address or manipulate as a single discrete object. *See also* **content object**.

information processing

See **processing**.

information retrieval database

Also called an **IR database**. Any database designed primarily for discovering and retrieving information. The systems that work with IR databases provide the following: a search interface to permit users to compose queries, methods for searching through the target data, viewable or behind-the-scenes indexes, and results displays.

initialism

A set of initials that stand for the full form of a name (e.g., *MFA*, for *Museum of Fine Arts*). *See also* **abbreviation** and **acronym**.

instance relationship

A hierarchical relationship in which all children must be an example of a broader context, most commonly seen in vocabularies where proper names are organized by general categories of things or events (e.g., if the proper names of mountains and rivers are organized under the general categories *mountains* and *rivers*). *See also* **genus/species relationship** and **whole/part relationship**.

interactive processing

See **processing**.

internal node

See **nonleaf node**.

interoperability

In the context of controlled vocabularies, the ability of two or more vocabularies and their systems or components of their systems to map to each other's data, with the goal of exchanging information or enhancing discovery.

inverse document frequency (IDF)

An automatic ranking method often used in a formula with **term frequency** in information retrieval and text mining to estimate how important a term is to a set of data and how useful it will be in retrieval.

inverted form

Also called an **inverted index**. In the context of a controlled vocabulary, the indexing form of a multiple-word name or term, where the last name or trunk portion of the term is listed first, followed by a comma and the descriptive word (e.g., *Wren, Christopher*, or *buttresses, flying*). *See also* **natural order form** and **permuted index**.

invisible Web

See **hidden Web**.

ISO (International Organization for Standardization)

A worldwide voluntary, nontreaty network of national standards institutes of approximately 160 countries. The standards bodies work in partnership with international organizations, governments, industry, business, and consumer representatives to reach consensus, set standards, and promote their use with the goal of facilitating trade and meeting the broader needs of society.

item

In the context of cataloging art, an individual object or work. *See also* **component** and **group**.

item-level cataloging

Describing and assigning indexing terms for individual items in a collection of works. *See also* **group-level cataloging**.

jargon

A characteristic terminology of a particular group or discipline that is typically not understood by a more general audience.

keyword

In the context of vocabularies, a **verbal unit** or word of a term that may be used in a search expression (e.g., for the place name *Sena Julia*, *Sena* is one keyword and *Julia* is another). In the broader context of online retrieval, any significant word or phrase in the title, subject headings, or text associated with an information object.

Keyword in Context (KWIC)

A type of automatic indexing in which each word in a text, title, subject heading, string of words, or term becomes an entry word in the index, with the exception of words in **stop lists**. Variations on KWICs are KWOCs (Keyword Out of Context) and KWACs (Keyword Alongside Context).

keyword index

An index based on individual words (keywords) found in a vocabulary term, text, or other content object.

label

See **heading**.

language model

A type of automatic indexing based on term weighting and relevance prediction that attempts to predict probable query search terms based on term frequencies within documents and the inverse document frequency of terms across the target data. It is similar to the **probabilistic model**.

last name

Also called a **surname**. In Western tradition, the family name used with a first name to identify a person (e.g., *Meier* in *Richard Meier*). See also **first name** and **middle name**.

latent semantic indexing (LSI)

A form of automatic indexing based on the co-occurrence clustering of terms in combination with content that is associated with these clusters; it attempts to

partially address the problem of the variety of terms that can be used to express similar concepts.

Latin 1

A character set (consisting of 191 characters) that is part of a series of ASCII-based character encodings defined in *ISO/IEC 8859-1:1998: 8-Bit Single-Byte Coded Graphic Character Sets—Part 1*.

latinization

See **romanization**.

lead-in term

See **used for term**.

leaf linking

See **node linking**.

leaf node

Also called an **external node**. In a thesaurus, a node that has no children, as with the ends or tips of hierarchical trees.

lexeme

A fundamental unit of the words of a language, around which may be clustered a set of words that are different forms of the same word (e.g., *paint* is the lexeme for *paints*, *painted*).

lexical unit

Also called a **lexical item**. One or more words that refer to a single concept (e.g., *flying buttresses* or *bills of sale*). See also **bound term** and **compound term**.

lexical variant

A term that is a different word form for another term, caused by spelling differences, grammatical variation, or abbreviations (e.g., *watercolor* and *water-colour*). Lexical variants are considered as and grouped with synonyms in a vocabulary record, but they technically differ from synonyms in that synonyms are different terms for the same concept. See also **synonym**.

link

In the context of this book, any relationship between two vocabulary records, two works, a work and image, or a work or image and an authority. Compare to **hyperlink**.

literary warrant

Justification for the inclusion of a term in a vocabulary based on published evidence that is sufficient to prove that the form, spelling, usage, and meaning of the term are widely agreed upon in authoritative sources. *See also* **organizational warrant**, **source**, and **user warrant**.

load

The process of moving or transferring files or software from one disk, computer, or server to another disk, computer, or server. To **upload** means to transfer from a local computer to a remote computer; to **download** means to transfer from a remote computer to a local one.

loan word

In the context of a given language, a word that is taken directly from another language (e.g., *sotto in su*, an Italian phrase used in English to mean painted in correct perspective as if viewed from below).

local authority

An authority developed for local use. Although often compiled from one or more standard authoritative published vocabularies, a local authority enforces preferences and usage pertinent for the local setting. *See also* **authority file** and **derivation**.

locator

In a bibliographic index, the part of an index entry that indicates the location of the book, page, or other resource. In an online index, it may be a hyperlink to the source.

logical data model

A data model that includes all entities and the relationships among them based on the structures identified in a conceptual data model, and that specifies all attributes for each entity. The data is described in as much detail as possible, without regard to how it will be implemented in a specific database. *See also* **conceptual data model**.

logical record

See **record**.

main term

See **descriptor**.

manual indexing

See **indexing**.

mapping

A set of correspondences between terms, fields, or element names used for translating data from one standard or vocabulary into another, or as a means of combining terms or data for search and retrieval. *See also* **crosswalk**.

markup language

A formal way of annotating a document or collection of digital data using embedded encoding tags to indicate the structure of the document or data file and the contents of its data elements. This markup also provides a computer with information about how to process and display marked-up documents. HTML, XML, and SGML are examples of standardized markup languages.

material culture

A term referring to art together with the broad realm of physical objects and edifices produced by a culture. *See also* **cultural heritage**.

metadata

A structured set of descriptive elements used to describe a definable entity. This data may include one or more pieces of information, which can exist as separate physical forms. In the context of art information, metadata includes data associated with information about the creation, physical characteristics, history, location, administration, or preservation of the work.

Metaphone

A phonetic algorithm for matching terms and names by sound, as pronounced in English, by translating words into a standard code or representation. It was developed by Lawrence Phillips to address the perceived deficiencies in the Soundex algorithm. Metaphone and its later improvements are available as built-in operators in a number of systems. *See also* **Soundex**.

metasearching

See **federated searching**.

microcontrolled vocabulary

Also called a **microthesaurus**. A controlled vocabulary that is limited in the range of topics covered but fits within the domain of a larger, broader, or more generic controlled vocabulary. It typically contains highly specialized terms that are not necessarily in the broader controlled vocabulary but that map to the hierarchical structure of the broader controlled vocabulary. See also **derivation**, **extension vocabulary**, and **satellite vocabulary**.

middle name

In Western tradition, any name for a person placed before the last name (surname) but after the first name (e.g., *Alan* in *Richard Alan Meier*). See also **first name** and **last name**.

minimal description

In the context of cataloging art, a record containing the minimum amount of information in the minimum number of fields or metadata elements.

modeling

See **derivation**.

modifier

In a compound term or name, the adjectival component that modifies the noun (e.g., *flying* in *flying buttresses*; *Mount* in *Mount Etna*). See also **focus**.

monohierarchy

A hierarchy in which each child has only one immediate parent. Distinguished from a **polyhierarchy**.

monolingual

Expressed in a single language, as distinguished from **multilingual**. In a monolingual thesaurus, the terms and names are expressed in only one language.

movable work

In the context of cataloging art, any tangible object capable of being moved or conveyed from one place to another, as opposed to

real estate or other buildings. Distinguished from **built work**.

multilingual

Expressed in more than one language, as distinguished from **monolingual**. In a *multilingual thesaurus*, terms and other information may be expressed in more than one language.

name authority

An authority containing proper names, most often personal names. See also **subject heading list**.

narrower term (NT)

Also called **narrower context** or **child**. A record to which another record or multiple records are superordinate in a hierarchy (e.g., *Brewster chair* is a narrower term to *armchair*). In thesauri, the relationship indicator for this type of term is NT. Variations on the notation include NTG (narrower term generic), NTP (narrower term partitive), NTI (narrower term instance), NT1 (narrower term level 1), NT2 (narrower term level 2), etc.

narrow results

To adjust criteria in a search in order to retrieve a smaller number of more precise results that better match the intention of the searcher. See also **broaden results**.

natural language

Spoken or written texts, as distinguished from fielded data and **controlled vocabulary**.

natural order form

In the context of a controlled vocabulary, the form of a multiple-word name or term, where the name or term appears in the form that would be used in speech or a written text (e.g., *Christopher Wren* or *flying buttresses*), rather than inverted (as may be appropriate for an index). See also **inverted form**.

navigation

In the context of search and retrieval, the facility that allows users to move through a controlled vocabulary or other content

object by using preestablished links or relationships.

near synonymy

*Also called **quasi-synonymy**.* The characteristic of a term with meaning that is regarded as different from another term, but both the terms are treated as equivalents for the purposes of broadening retrieval. *See also **synonym** and **true synonymy**.*

neologism

A term that has been newly invented, or an existing term to which a new meaning is applied, often arising in the professional literature of a discipline.

nickname

A familiar, affectionate, derogatory, or humorous name that is used to refer to a person, place, or corporate body as a replacement for, or in addition to, the real or official name (e.g., *Masaccio*, meaning "big Tom," is a nickname for the painter *Tommaso Guidi*). (In the case of *Masaccio*, in the *ULAN* it is the preferred name based on literary warrant.) *See also **pseudonym**.*

NISO (National Information Standards Organization)

A nonprofit association that is accredited by the American National Standards Institute (ANSI) and identifies, develops, maintains, and publishes technical standards to manage information.

node

In the context of a thesaurus, any point or record in the hierarchy that is a location at which a branch or individual record (leaf) is attached; thus, the basic conceptual unit used to build hierarchies.

node label

A word or phrase inserted into a hierarchy to indicate the logical classification of the terms beneath it. *See also **facet indicator** and **guide term**.*

node linking

*Also called **leaf linking**.* In the context of combining multiple vocabularies, a method that uses various nodes in the hierarchical

structure of a source controlled vocabulary to link to more detailed controlled vocabularies that are applicable to a single node of the parent hierarchy. The vocabulary linked to a broader vocabulary in this way is often called an **extension vocabulary**.

nondisplayed index

A machine-readable index that is not displayed for browsing or other direct access of end users, but is used behind the scenes to improve accuracy or speed in search and retrieval. Such indexes may be created beforehand or on the fly at the time of the query. *See also **displayed index**.*

nonequivalence

In mapping one vocabulary to another, the situation where there is no exact match, no term in the second language has partial or inexact equivalence, and there is no combination of descriptors in the second language that would approximate a match. *See also **exact equivalence** and **inexact equivalence**.*

nonleaf node

*Also called an **internal node**.* In a hierarchy, a node that links to one or more narrower contexts. *See also **leaf node**.*

nonpreferred parent

In a polyhierarchical thesaurus, any parent that is not flagged as preferred for use as a default in displays. *See also **preferred parent**.*

nonpreferred term

*Also called a **nonpreferred name**.* Any term in a vocabulary record that is not the **preferred term**, which is the term flagged as preferred for use as default in displays.

normalization

In the context of vocabulary retrieval, normalizing terms through a process of converting a term to its simplest form by removing case sensitivity, spaces, punctuation, and diacritics. It differs from **database normalization**, which is the process of reducing a complex data structure into its simplest structure, a technique used to eliminate data redundancy by

converting Unicode text into a standardized form, among other things.

notation

For a thesaurus, the alphabetic code used to express term types (D, AD, UF), associative relationship (RT), hierarchical relationships (BT, NT, BTG, NTG, BTP, NTP, BTI, NTI, BT1, BT2, NT1, NT2), and scope notes (SN), among others. *See also* **classification notation**.

object

See **work**.

object-oriented database

A data model where the universe is divided into a framework of classes, with each class containing instances or members (called "objects"). Classes can contain subclasses, members of which inherit the properties of the parent or "superclass." Rules and algorithms for processing the data are integrated with the data.

online catalog

In the context of art information, a type of system used by end users to search for and view data and images.

ontology

A formal, machine-readable specification of a conceptual model, in which concepts, properties, relationships, functions, constraints, and axioms are all explicitly defined. While an ontology is not technically a controlled vocabulary, it uses one or more controlled vocabularies for a defined domain and expresses the vocabulary in a representative language that has a grammar for using vocabulary terms in an automated way to express something meaningful.

operating system

Also called an **operating system**

program. A software program that runs a computer, as distinguished from an **application** program, which is designed to accomplish a task for an end user (e.g., word processing).

operational specificity

Also called **postings specificity**. An automated method that attempts to predict the

specificity of terms in a domain based on the number of postings or links to that term in a content object (e.g., a term that is linked to very few content objects is predicted to be highly specific).

organizational warrant

Justification for the inclusion of a term in a vocabulary based on the specialized requirements or jargon of the group or organization that is creating or sponsoring the vocabulary. *See also* **literary warrant** and **user warrant**.

orphan term

In a thesaurus, a record that has no associative or hierarchical relationship to any other term in the thesaurus.

orthography

Correct or proper spelling and form of a word or words, including capitalization, diacritics, and punctuation, based on standard usage or convention.

paradigmatic relationship

Also called a **semantic relationship**.

A relationship between terms or concepts that is permanent and based on a known definition.

parallel searching

See **federated searching**.

parent

See **broader term (BT)**.

parenthetical qualifier

A **qualifier** placed in parentheses for display.

parent string

The display of hierarchical parents in a horizontal string, as distinguished from vertical indented displays or displays using **notation**.

parsing

In processing data, a process where data is broken or filtered into smaller, more distinct units.

partial equivalence

The relationship between terms in two vocabularies where one term has a broader

scope but is partially synonymous with the other term. *See also* **exact equivalence** and **inexact equivalence**.

partitive relationship

See **whole/part relationship**.

patronymic

Also called a **patronym**. A word or words used with a given name to identify a person; common in early Western personal names when last names were uncommon (e.g., *Bartolo di Fredi* means "Bartolo, son of Fredi"); may also refer to a surname derived from a paternal ancestor (e.g., *Robinson* means "son of Robin").

permuted index

A type of index where individual words of a term are rotated to bring each word of the term into alphabetical order in the term list. *See also* **inverted form**.

phonetic matching

A process by which terms are matched to other terms that are presumed to sound like the original term, in an attempt to compensate for users' misspellings or general variation in spelling of names or terms (e.g., *Meier* and *Meyer* are pronounced alike).

Phonetic algorithms—such as Soundex, Metaphone, and others—are used for indexing words by their pronunciation.

physical feature

In the context of geographic information, a characteristic of the earth's surface that has been shaped by natural forces, including continents, mountains, forests, rivers, and oceans. *See also* **administrative entity**.

pick list

A user interface feature that allows the user to select from a preset list of terms and is typically used to control vocabulary for indexing or to provide options in a query. A pick list is generally populated with a controlled list.

polyhierarchy

A thesaurus in which any record may be linked to multiple parent records. *See also* **hierarchy**.

polyseme

A word or lexical unit (e.g., a compound term) with multiple meanings; known as a **homograph** in written language and a **homophone** in spoken language.

postable term

See **descriptor**.

postcoordination

The process of combining two or more terms at the time of retrieval rather than at the indexing stage; usually uses the Boolean operators AND, OR, or NOT (*Baroque* AND *cathedral*) in formulating a query. *See also* **precoordination**.

posting

In the context of indexing, any instance of a given indexing term having been assigned to records, documents, or other content objects. Formulas used for predicting the usefulness of terms or methods of retrieval may count the number of postings relative to the target content objects or use the numbers of postings in other statistics.

postings specificity

See **operational specificity**.

precision

A measure of a search system's effectiveness in terms of retrieving only relevant results; expressed as the ratio of relevant records or documents retrieved from a database to the total number retrieved in response to the query. A high-precision search means that most of the results retrieved will be relevant; however, a high-precision search will not necessarily retrieve all relevant results. Recall and precision are inverse ratios (when one goes up, the other goes down). *See also* **recall**.

precoordination

The formulation of a compound term or multiword heading at the time of indexing, rather than at the time of retrieval. An example of a precoordinated term is *Baroque cathedrals*; an example of a precoordinated heading is *United States—History—Civil War, 1861–1865*. *See also* **postcoordination**.

predefined report

A report for which the query and the output have been written and made available for repeated use by users; users may be allowed to enter variables that are plugged into the report. *See also* **ad hoc query**.

preferred flag

A designation indicating that a term or other data instance is preferred over others of the same type in a record. In addition to a preferred term for the record overall, there may be a **preferred indexing name flag** for the inverted order version of the term, a **preferred display name** for the natural order form of the name, a **preferred role** or **preferred place type** flagged among a list of roles or place types, and so on.

preferred parent

In a polyhierarchical thesaurus, the broader context that is chosen as conceptually preferred; or, to serve as the default in hierarchical displays. *See also* **nonpreferred parent**.

preferred term

Also called a **preferred name**. The term designated among all synonyms or lexical variants for a concept to be used as the default term to represent the concept in displays and other situations. In a monolingual thesaurus, the preferred term is also the only descriptor in the record. In a multilingual thesaurus, there may be a descriptor for every language, but there is often only one preferred term for the record as a whole. *See also* **descriptor**.

preprocessing

Also called **data preprocessing**. Preliminary processing or transformation of data in order to facilitate further processing, parsing, etc.

probabilistic model

An automatic relevance and weighting method in which terms in a text or other content object are modeled as random variables so that term frequency and distribution are used to predict the probability of relevance. *See also* **language model**.

procedure

Also called a **subprogram** or **subroutine**. A relatively independent portion of computer code within a larger computer program that performs a specific task in a series of steps.

processing

Also called **data processing** or **information processing**. The manipulation or transformation of data through a series of operations. In **batch processing**, the operations are grouped together in batches and performed automatically; in **interactive processing**, the operations are prompted by input from a human programmer or user. *See also* **computer program**.

program

See **computer program**.

programming language

A formal language defined by syntactic and semantic rules and used to write instructions that can be translated into machine language and then executed by a computer (e.g., SQL, C++, C#, Java, Perl).

provisional term

See **candidate term**.

pseudonym

A false or fictitious name, especially one assumed by an artist, author, or other person to maintain anonymity or to designate an identity for a particular activity, among other reasons (e.g., *Le Corbusier* is a pseudonym assumed by the architect *Charles Édouard Jeanneret*). *See also* **nickname**.

punctuation

In the context of vocabulary terms, the marks from standard written communication used to clarify, organize, or indicate how a word or words should be read (e.g., hyphen, comma, period, quotation marks, parentheses).

qualifier

A word or phrase used to distinguish a term in a vocabulary from otherwise identical terms that have different meanings. A

qualifier is separated from the term, usually by parentheses. It is also called a **gloss**; although, strictly speaking, a qualifier should be used only with homographs, and a gloss has a more general meaning in the field of linguistics. *See also* **homograph**.

quasi-synonymy

See **near synonymy**.

query

Also called a **search**. In the context of retrieval, a command to look in a database and find records or other information that meet a specified set of criteria (e.g., *select subject_id from term where normalized_term like 'A%' and historic_flag = 'H'*);. The most precise queries are those that return the fewest false hits.

query expansion (QE)

Reformulating a query in order to return a broader or more comprehensive set of results (e.g., adding synonyms to the user's search term).

recall

A measure of a search system's effectiveness in terms of retrieving all results that are possibly relevant, expressed as the ratio of the number of relevant records or documents retrieved over all the relevant records or documents. A high recall search retrieves a comprehensive set of relevant results; however, it also increases the likelihood that marginally relevant content objects will also be retrieved. Recall and precision are inverse ratios. *See also* **precision**.

reciprocity

In reference to vocabulary records, the characteristic of a two-way relationship in which both entities have mutual dependence, action, or influence on each other. Semantic relationships in controlled vocabularies must be reciprocal, meaning each relationship from one record to another must also be represented by a reciprocal relationship in the other direction. Reciprocal relationships may be **symmetric** (e.g. RT/RT) or **asymmetric** (e.g. BT/NT).

record

Also called a **logical record**. In the context of this book, a conceptual arrangement of fields referring to a vocabulary concept or a work. This is different from a **database record**, which is one row in a database table or another set of related, contiguous data. *See also* **concept record**.

record group

See **group**.

related term (RT)

A concept that is associatively (not hierarchically) linked to another concept in a thesaurus. In thesauri, the relationship indicator for this type of term is RT. *See also* **associative relationship**.

relational table database

Also called a **relational database**. A database in which data is organized into columns and rows according to specific defined relationships (e.g., in a vocabulary database, a table of terms may be linked to a table for languages).

relationship

In the context of this book, a link between two types of data, records, files, or any two entities of the same or different types in a system or network. *See also* **link**.

relationship indicator

A word, code, or other device used in thesauri to identify the semantic relationship between terms (e.g., UF), other fields (e.g., SN), or records (e.g., BT).

relevance

The extent to which information retrieved in a search is judged by the user to meet the criteria of the query.

relevance ranking

Ranking and sorting of query results, typically estimated by an algorithm that calculates the number and weight of occurrences of the search term in the targeted data.

report

An organized set of data presented in a format suitable for viewing or printing,

typically produced by a preestablished query that may or may not have variables that are manipulated by the user.

repository

In the context of art and related disciplines, refers to an institution, agency, or individual that has physical or administrative responsibility for an art object, work of architecture, or cultural object.

required fields

Fields or data elements that are required to meet a standard or the requirements of a system's operations. *See also* **core fields**.

reserved characters

Letters, numbers, or symbols that have special uses or meanings in a programming or querying language.

results list

The records or other data retrieved in response to a query and presented online or in a system in an organized display.

retrieval

In the context of this book, the activity of using a search or other method to find records or other data in a database. *See also* **query**.

romanization

Also called **latinization**. The conversion of a character or word expressed in a non-Roman alphabet or writing system (e.g., Cyrillic or Korean) into the Roman alphabet by means of transcription, transliteration, or a combination of the two methods.

root

Also called **root node** or **top term**. The highest level of the hierarchy, from which all branches descend.

rotated listing

See **permuted index**.

satellite vocabulary

A thesaurus that is created with the intention of, or is later adapted for, linking to another vocabulary that is larger, broader, or more generic; it may be integrated at many points in the original vocabulary. *See*

also **extension vocabulary**, **microcontrolled vocabulary**, and **node linking**.

schema

Also called a **scheme**. In the context of this book, the organization, structure, and rules for a set of data (e.g., the set of tables, views, indexes, and descriptions for columns in a database, or the organization and description of an XML document).

scope note (SN)

A note explaining the coverage, specialized usage, and meaning of terms. In thesauri, the relationship indicator for this note is SN.

search

See **query**.

searching

Operations or algorithms intended to determine if one or more data items meet defined criteria or possess a specified property.

see also reference

A type of cross-reference, usually in a printed index, directing the reader to a related term or entry. A *see also* reference differs from a *see* reference in that the *see also* reference is not made between synonyms, but between terms or headings that are more peripherally related.

see reference

A type of cross-reference, usually in a printed index, directing the reader from a nonpreferred term or subject heading to the preferred term or subject heading for the same concept. The term or subject heading at the *see* reference is a synonym for the preferred term or heading.

semantic linking

A method of linking terms in a vocabulary or larger database according to the meaning of the terms and relationships between terms.

semantic relationship

See **paradigmatic relationship**.

SGML (Standard Generalized Markup Language)

International Standards Organization standard *ISO/IEC 8879:1986*; a markup

language first used by the publishing industry, for defining, specifying, and creating digital documents that can be delivered, displayed, linked, and manipulated in a system-independent manner. XML and HTML are derived from SGML.

sibling

A concept that shares the same immediate broader context (one level higher) as other concepts. Siblings are subordinate to the same broader concept and are at the same hierarchical level.

single-to-multiple term equivalence

In the context of mapping terms from different vocabularies to each other, the situation that occurs when a term in one vocabulary has no direct match in the second vocabulary, but instead must be mapped to a combination of terms.

social tagging

The decentralized practice and method by which individuals and groups create, manage, and share tags (terms, names, etc.) to annotate and categorize digital resources in an online “social” environment. *See also* **folksonomy**.

software

The components of a computer system that are not physical, including programs, procedures, algorithms, and documentation pertaining to the operation of a system and the performance of specific tasks, such as word processing, Web browsers, photo editing, and art cataloging or vocabulary editing. *See also* **hardware**.

sorting

In the context of this book, the automated process of organizing a results list, data elements in a record, or other data in a particular sequence based on established criteria or attributes of the data—for example, alphabetically, by parent string, or by an associated date. There may be primary sort criteria and secondary sort criteria (e.g., an algorithm can be formulated to first sort place names in a results list alphabetically, and then—for homo-

graphs in the list—to sort by the parent string). *See also* **filing rules**.

Soundex

A phonetic algorithm for matching terms and names by sound, as pronounced in English, by translating words into a standard code or representation. It was developed by Robert Russell and Margaret Odell and patented in 1918 and 1922. The National Archives and Records Administration (NARA) maintains the current rule set for the official implementation of Soundex used by the U.S. Government. *See also* **Metaphone**.

source

In the context of building vocabularies, a citable reference to a term in the literature that helps establish its form, spelling, usage, and meaning. *See also* **literary warrant**.

source authority

In the context of this book, a bibliographic authority file used to control the citations providing warrant for terms in a vocabulary or information in a work record.

source language

In the context of translating or mapping one vocabulary to a vocabulary in another language, the language of the original vocabulary. *See also* **target language**.

specialized vocabulary

See **microcontrolled vocabulary**.

specifications

In the context of designing an information system, the formal, detailed description of user and technical requirements, including specific descriptions of procedures, functions, screens, reports, materials, other features, and hardware. *See also* **user requirements**.

specificity

In the context of indexing, the degree of precision or **granularity** used in assigning terms. Measures of greater specificity include the use of the narrowest applicable indexing term rather than a broader, more generic term. *See also* **exhaustivity**.

SQL (Structured Query Language)

A standard command language used with relational databases to perform queries and other tasks.

standard

A vocabulary, set of rules, code of practice, or description of characteristics and parameters that is documented, established by experts, or approved by an authoritative body and widely recognized or employed as an authoritative exemplar of correctness or best practice; used within a discipline or domain in order to promote interoperability and efficiency.

statistical specificity

See **operational specificity**.

stemming

In the context of mapping terms for search and retrieval, the alteration of a term by automatically truncating or removing common suffixes, word endings, or prefixes in order to find a match, usually applied to sets of related words that are derived from a common root and appear in a variety of grammatical forms (e.g., *paint*, *painting*, *painted*).

stop list

In the context of search and retrieval, words in a vocabulary or target data that are ignored in searching or matching because they occur too frequently or are otherwise of little value in retrieval for a given domain. Common stop lists for a text contain articles, conjunctions, and prepositions, although these words are typically not included in a stop list for a vocabulary.

string syntax

Also called **string indexing**. The creation of headings by computer algorithm, characterized by headings that are more consistent than the typically idiosyncratic headings created by hand (e.g., the automatic concatenation of a parent string in a heading for a geographic place, such as *San Gimignano (Siena province, Tuscany, Italy)*).

structure

See **data structure**.

subfacet

A major conceptual division of a thesaurus that is located near the top of the tree but under a facet. Also called a **hierarchy** in the *AAT*, although *hierarchy* has a more general meaning as well.

subject

In the context of this book, the focus concept of a vocabulary record (e.g., the subject of a ULAN record is a person). Also used to refer to the subject matter (often iconographical content) of what is depicted in or by a work of art or the content of a text.

subject heading list

An alphabetical list of words or phrases used to indicate the content of a text or other thing; characterized by **precoordination** of terminology, meaning that several unique concepts are combined in a string (e.g., *Archaeology and art—China—History—20th century*). A type of controlled vocabulary. See also **authority heading** and **heading**.

subject indexing

A term typically used in the context of bibliographic cataloging but also applicable to cataloging art; refers to the application of indexing terms to the content of the document, as contrasted to a description of its physical characteristics.

subprogram

See **procedure**.

subroutine

See **procedure**.

surface Web

See **visible Web**.

surname

See **last name**.

switching

In the context of mapping one vocabulary to another, refers to the use of a third vocabulary (a **switching vocabulary**) that itself can link to terms in each of the two original controlled vocabularies; useful when the original two vocabularies do not map well

directly to each other. *See also* **direct mapping**.

symmetric relationship

In the context of a thesaurus, a reciprocal relationship that is the same in both directions (e.g., RT/RT). *See also* **asymmetric relationship** and **reciprocity**.

syndetic structure

Also called **cross-reference links**. In the context of a vocabulary, refers to the linking of equivalent, broader, narrower, and other related terms so that they can be used as cross-references to each other and to related headings for the purpose of access.

synonym

A term having a different form but exactly or very nearly the same meaning as another term. *See also* **near synonymy** and **true synonymy**. Compare **lexical variant**.

synonym ring list

A type of controlled vocabulary containing terms that are considered equivalent for the purposes of retrieval but do not necessarily have true synonymy.

synonymy

A type of semantic relation in which two words or terms have the same or very similar meaning. *See also* **near synonymy** and **true synonymy**.

syntax

In the context of this book, the structure of elements in a compound term or name (e.g., last name first, comma, first name, middle initial) or heading; also used to refer to the structure of elements in a search query (e.g., rules for the placement of the Boolean operators OR, AND, or NOT between terms); and analogous to the linguistic structure of elements in a sentence.

synthesis note

A brief preliminary finding, example, or recommendation. This expression was used in the original print publication of the *AAT* to refer to bottom-of-page notes throughout each subfacet (or hierarchy) that suggested ways in which descriptors from that subfacet could be combined in post-

coordination with other descriptors (these recommendations are now found in the *AAT Editorial Manual*).

system

Also called a **computer system**. A number of interrelated hardware and software components that work together to store and convert data into information by using electronic processing. In the context of this book, a system for building and maintaining vocabularies, cataloging art, or performing search and retrieval. *See also* **database**.

systematic display

See **hierarchical display**.

table

See **data table**.

target language

In the context of translating or mapping one vocabulary to a vocabulary in another language, the language into which the original vocabulary is being translated. *See also* **source language**.

taxonomy

A classification organized into a hierarchical structure and applicable to a defined domain. Often used to refer to the classification of living organisms according to physical characteristics, but the term and principles can be applied to classification in any discipline. Unlike thesauri, taxonomies do not typically include synonyms and associative relationships. *See also* **folksonomy**.

term

A word or group of words representing a single concept; a vocabulary record comprises terms and other information, including relationships, scope notes, sources, etc. Additionally, in the jargon of thesaurus construction, the word *term* is often used as shorthand to refer to the concept that is represented by that *term* (e.g., BT and NT actually refer to the relationships between concepts). The distinction between a *term* in the strict sense and *term* meaning a *record* must often be inferred from the context of the discussion.

term frequency (TF)

An automatic ranking method often used in a formula with **inverse document frequency** in information retrieval and text mining to measure how important a term is to a set of data and how useful it will be in retrieval.

term record

In the jargon of thesaurus construction, the collection of information associated with a descriptor, including the history of the term, its relationships to other terms and records, etc. In this book, it is referred to as a **record** (or a **concept record**) in order to distinguish it from the information that is actually associated only with the term table in a relational database model (e.g., language of the term, contributor of the term).

text

In the context of this book, data that is not vocabulary controlled and generally unstructured beyond the common structure of standard language expressions of characters, words, sentences, or paragraphs. *See also* **free-text field**.

thesaurus

A controlled vocabulary arranged in a specific order and characterized by three relationships: equivalence, hierarchical, and associative. Thesauri may be monolingual or multilingual. Their purposes are to promote consistency in the indexing of content and to facilitate searching and browsing.

top term (TT)

See **root**. In thesauri, the relationship indicator for this type of term is TT.

transcription

In the context of cataloging art, the process of recording a term or text word-for-word and letter-for-letter, including accurately copying capitalization, punctuation, spacing, line breaks, illegible passages, and all other possible aspects of the original (e.g., to accurately express the nuances of an artist's signature or an ancient architectural inscription). Transcriptions in this context are typically semidiplomatic or

seminormalized transcriptions, meaning both substantive and accidental features of the original are retained, but abbreviations are spelled out using brackets or other punctuation to distinguish the original from the editorial content.

translation

The process of changing a term or text from one language into another by interpreting the meaning of the original (source) term and expressing it as an equivalent in the second (target) term (e.g., *copper mines* in English is translated as *mines de cuivre* in French).

transliteration

The process of rendering the letters or characters of one alphabet or writing system into the corresponding letters or characters of another alphabet or writing system, generally based on phonetic equivalencies. While a common noun will often be translated, a proper name in a non-Roman alphabet is more often transliterated. There are often multiple standards for transliterating from one writing system to another, thus producing multiple variant names.

tree structure

A controlled vocabulary display format in which the complete hierarchy of records is shown or accessible by clicking. The tree structure may be constructed by assigning a tree number or line number to each record, or by another method. *See also* **hierarchical display**.

true synonymy

The characteristic of terms or names that have meanings that are identical or as nearly identical as is possible with language. The purpose of enforcing true synonymy in a vocabulary is to increase precision in indexing and retrieval. *See also* **near synonymy and synonym**.

truncation

In searching and matching, the action of cutting off characters in a search term in order to find all terms with a certain common string of characters; typically involves the user employing a wildcard

symbol to search for a string of characters no matter what other characters follow (or sometimes, precede) that string (e.g., searching for *arch** will retrieve *arch*, *arches*, *architrave*, *architecture*, *architectural history*, etc.).

trunk name

See **focus**.

typography

The font style and size, and arrangement, appearance, and layout of words and texts on a page; in the context of this book, one of the critical elements in designing an end-user display of vocabulary records.

Unicode

A 16-bit character encoding scheme and standard for representing letters, characters, and diacritical marks in most of the world's modern scripts.

unique identifier

A number or other string that is associated with a record or piece of data, exists only once in a database, and is used to uniquely identify and disambiguate that record or piece of data from all others in the database.

upload

See **load**.

up-posting

Also known as **autoposting**. The automatic generation of search terms or indexing terms by adding broader terms to the specific term requested by a searcher or used by the indexer. *See also* **generic posting**.

used for term

Also called a **UF**. In thesaurus jargon, a term that is not a descriptor and not an alternate descriptor. If the thesaurus is being used as an authority, a *used for* term is not authorized for indexing. *Used for* terms typically comprise spelling or grammatical variants of the descriptor or have true synonymy with the descriptor.

user

See **end user**.

user interface (UI)

The portion of the design and functionality of a cataloging, editorial, search and retrieval, or other system or Web site with which end users interact, including the arrangement of displays, menus, clickable text or images, pagination, etc. A user interface that is easy for users to utilize is called **user friendly**.

user requirements

In system design, the initial formal explanation of functionalities, displays, and reports expressed from the point of view of the users' needs and expectations. *See also* **specifications**.

user warrant

Justification for a term in a controlled vocabulary based on the frequency of user queries that employ the term. User warrant may be used for terms intended for retrieval but is typically not sufficient warrant for posting a term in a thesaurus used for indexing. *See also* **literary warrant** and **organizational warrant**.

variable

In a query, criteria or factors that may be changed to produce different results (e.g., as may be expressed in a *where* clause, as the relationship type code in this query: *select distinct subjecta_id from associative_rels where rel_type_code = '2110'*); *See also* **criteria**.

variant term

In a vocabulary, a term that is not the preferred term but refers to the same concept, including *used for* terms and alternate descriptors.

vector-space model

A method of automatic weighting in retrieval where an algebraic model is used for term frequency and distribution, creating representative vectors in multiple dimensional space; when compared to the vectors of an incoming query, the relevance of results may be predicted.

verbal units (VU)

In linguistics and computer science, the phonemic, morphemic, or grammatical

clauses or units of language or texts, corresponding in part to syllables, letters, or words.

visible Web

The subset of the World Wide Web that is visible to Web browsers and can be indexed by search engines' Web crawlers or robots, in contrast to pages that are impenetrable by search engines or to data that is generated dynamically.

visual arts

See art.

vocabulary

See controlled vocabulary.

vocabulary control

The process of enforcing the use of certain terminology with the goal of providing consistency and improving retrieval.

warrant

In the context of vocabularies, sources that provide justification for the spelling and usage of a term to refer to a particular usage for a concept, including warrant of publications, common usage by experts of a discipline, or other sources.

Web browser

A software application that enables users to view and interact with information and media files on the Web (e.g., Internet Explorer, Mozilla Firefox, and Safari).

Web site

A collection of related electronic pages (**Web pages**), generally formatted in HTML and found at a single address where the server computer is identified by a given host name.

weighted term ranking

See best match.

whole/part relationship

*Also called a **partitive relationship**.*

A hierarchical relationship between a larger entity and a part or component. In the context of cataloging art, it typically refers to a relationship between two work records or two records in a thesaurus (e.g., *Florence* is part of *Tuscany*). *See also **genus/species relationship** and **instance relationship**.*

wildcard

*Also called a **wildcard character** or **wildcard symbol**.* In searching, a character or symbol, such as an asterisk or percent sign, that is used to represent any other character or characters in a Boolean query or other string (e.g., the asterisk in *Buonar**).

word sense disambiguation (WSD)

In automatic search and retrieval, the problem of determining in which sense a homograph is intended in a given data set or text. *See also **disambiguation**.*

work

In the context of this book, a creative product, including architecture; artworks such as paintings, drawings, graphic arts, sculpture, decorative arts, and photographs that are considered to be art; and other cultural artifacts. A work may be a single item or may be made up of many physical parts.

XML (Extensible Markup Language)

A simple, flexible markup language derived from SGML. Originally designed for large-scale electronic publishing, but now playing an increasingly important role in the publication and exchange of a wide variety of data on the Web.