International Standard

ISO/IEC 5394

Information technology — Criteria for concept systems

Technologies de l'information — Critères pour les systèmes de concept

First edition
2024-01

Reference number
ISO/IEC 5394:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

A concept system is a "set of concepts structured in one or more related domains according to the concept relations among its concepts" according to ISO 1087:2019. A concept is defined as a unit of knowledge created by a unique combination of characteristics which are the abstraction of a property of an object or of a set of objects. Concept systems are used to support semantic interoperability and integration in domains, information classification and organization, indexing, retrieval, etc.

Concept systems are used in the description of semantics of data in the ISO/IEC 11179 series of standards. ISO/IEC 11179-1 introduces data element concept, object class, property, conceptual domain and value meaning as concepts. They can be organized through the use of relations among them into concept systems. A classification scheme is represented as a concept system in ISO/IEC TR 11179-2.

There are various types of concept systems, ranging from the simplest concept systems with simple relations among concepts to ontologies with rich formal semantics.

The construction of most controlled vocabularies is based on concept systems. On the basis of the relevant concept system, they add the relationships among terms and establish the relationships among terms and concepts on the basis of semantic characteristics. Concepts are mainly represented by terms. Therefore, the discussion of the types of concept systems is inseparable from the discussion of vocabularies.

Ontologies comprise an important kind of concept system. The goals of ontologies are to capture the knowledge of one or several subject fields and to provide a common understanding. Also, ontologies serve to determine the common terms in the subject field, and to provide a clear understanding of the relations among the relevant concepts based on various levels of formal patterns.

The development of artificial intelligence technology and ontology technology has expanded both the content and the application of the scope of concept systems. The issues regarding the structure, classification, description and application of concept systems are becoming more and more important.

Information technology — Criteria for concept systems

1 Scope

This document provides the criteria for effective concept systems. It provides the requirements for components, formation, representations, structural levels and management of concept systems. Concept systems are used in the description of semantics of data in ISO/IEC 11179 standards.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 704, *Terminology work — Principles and methods*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
object
anything perceivable or conceivable

Note 1 to entry: Objects can be material (e.g. ‘engine’, ‘sheet of paper’, ‘diamond’), immaterial (e.g. ‘conversion ratio’, ‘project plan’) or imagined (e.g. ‘unicorn’, ‘scientific hypothesis’).

[SOURCE: ISO 1087:2019, 3.1.1]

3.2
property
<terminology>feature of an *object* (3.1)

EXAMPLE 1 ‘Being made of wood’ as a property of a given ‘table’.

EXAMPLE 2 ‘Belonging to person A’ as a property of a given ‘pet’.

EXAMPLE 3 ‘Having been formulated by Einstein’ as a property of a given ‘$E=mc^2$’.

EXAMPLE 4 ‘Being compassionate’ as a property of a given ‘person’.

EXAMPLE 5 ‘Having a given cable’ as a property of a given ‘computer mouse’.

Note 1 to entry: One or more objects can have the same property.

[SOURCE: ISO 1087:2019, 3.1.3, modified — added the domain <terminology>.]

3.3
characteristic
abstraction of a *property* (3.2)

EXAMPLE 'Having a cable for connecting with a computer' as a characteristic of the concept 'cord mouse'.

Note 1 to entry: Characteristics are used for describing concepts.

[SOURCE: ISO 1087:2019, 3.2.1]

3.4
concept
unit of knowledge created by a unique combination of *characteristics* (3.3)

Note 1 to entry: Concepts are not necessarily bound to particular natural languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

Note 2 to entry: This is the concept 'concept' as used and designated by the term "concept" in terminology work. It is a very different concept from that designated by other domains such as industrial automation or marketing.

[SOURCE: ISO 1087:2019, 3.2.7]

3.5
domain
subject field
field of special knowledge

Note 1 to entry: The borderlines and the granularity of a domain are determined from a purpose-related point of view. If a domain is subdivided, the result is again a domain.

[SOURCE: ISO 1087:2019, 3.1.4]

3.6
designation
designator
representation of a *concept* (3.4) by a sign which denotes it in a *domain* (3.22) or subject

Note 1 to entry: A designation can be linguistic or non-linguistic. It can consist of various types of characters, but also punctuation marks such as hyphens and parentheses, governed by domain-, subject-, or language-specific conventions.

Note 2 to entry: A designation can be a term including appellations, a proper name, or a symbol.

[SOURCE: ISO 1087:2019, 3.4.1]

3.7
synonymy
relation between *designations* (3.6) in a given natural language representing the same *concept* (3.4)

EXAMPLE Synonymy exists between "deuterium" and "heavy hydrogen", between "United Nations" and "UN".

Note 1 to entry: Designations in the relation of synonymy are called "synonyms".

[SOURCE: ISO 1087:2019, 3.4.23]

3.8
definition
representation of a *concept* (3.4) by an expression that describes it and differentiates it from related concepts

[SOURCE: ISO 1087:2019, 3.3.1]

3.9
term
designation (3.6) that represents a general *concept* (3.4) by linguistic means

EXAMPLE "laser printer", "planet", "pacemaker", "chemical compound", "3/4 time", "Influenza A virus", "oil painting".

Note 1 to entry: Terms may be partly or wholly verbal.

[SOURCE: ISO 1087:2019, 3.4.2]

3.10
vocabulary
terminological dictionary that contains *designations* (3.6) and *definitions* (3.8) from one or more *domains* (3.22) or subjects

Note 1 to entry: The vocabulary may be monolingual, bilingual or multilingual.

[SOURCE: ISO 1087:2019, 3.7.5]

3.11
controlled vocabulary
CV
vocabulary (3.10) for which the entries, i.e. *definition* (3.8) /*term* (3.9) pairs, are controlled by a Source Authority based on a rulebase and process for addition/deletion of entries

Note 1 to entry: In a controlled vocabulary, there is a one-to-one relationship of definition and term.

EXAMPLE The contents of "Clause 3 Definitions" in ISO/IEC standards are examples of controlled vocabularies with the entities being identified and referenced through their ID code, i.e., via their clause numbers.

Note 2 to entry: In a multilingual controlled vocabulary, the definition/term pairs in the languages used are deemed to be equivalent, i.e., with respect to their semantics.

Note 3 to entry: The rule base governing a controlled vocabulary may include a predefined concept system.

Note 4 to entry: Source Authority is defined in ISO/IEC 15944-2:2015, 3.109. Its definition is "Person recognized by other Persons as the authoritative source for a set of constraints"

[SOURCE: ISO/IEC 15944-5:2008, 3.34, modified — added Note 4 to entry]

3.12
semantic spectrum
range of increasingly precise *definitions* (3.8)

Note 1 to entry: Generally, the semantic spectrum includes glossaries, classification schemes, taxonomies, terminologies, subject heading schemes, thesauri, ontologies, etc. according to their semantic precision.

3.13
concept relation
relation between *concepts* (3.4)

[SOURCE: ISO/IEC 11179-3:2023, 3.2.8]

3.14
hierarchical relation
hierarchical concept relation
generic relation or partitive relation

[SOURCE: ISO 1087:2019, 3.2.12]

3.15
associative relation
associative concept relation
pragmatic relation
non-hierarchical *concept relation* (3.13)

EXAMPLE An associative relation exists between the concepts 'education' and 'teaching'.

[SOURCE: ISO 1087:2019, 3.2.23]

3.16
equivalence relationship
relationship between two *designations* (3.6) that both represent the same or similar *concept* (3.4)

3.17
concept system
system of concepts
set of *concepts* (3.4) structured in one or more related *domains* (3.22) according to the *concept relations* (3.13) among its concepts

[SOURCE: ISO 1087:2019, 3.2.28]

3.18
glossary
terminological dictionary that contains *designations* (3.6) from one or more *domains* (3.22) or subjects together with equivalents in one or more natural languages

Note 1 to entry: In English common language usage, glossary can refer to a monolingual list of designations and definitions in a domain or subject.

[SOURCE: ISO 1087:2019, 3.7.6]

3.19
classification scheme
descriptive information for an arrangement or division of *objects* (3.1) into groups based on criteria such as *characteristics* (3.3), which the objects have in common

EXAMPLE Origin, composition, structure, application, function, etc.

[SOURCE: ISO/IEC 11179-3:2023, 3.2.5]

3.20
taxonomy
type of hierarchy which deals with generalization/specialization relationships

[SOURCE: ISO/IEC 11179-32:2023, 3.20, modified — Note 1 to entry deleted]

3.21
terminology
set of *designations* (3.6) and *concepts* (3.4) belonging to one *domain* (3.22) or subject

[SOURCE: ISO 1087:2019, 3.1.11]

3.22
subject heading scheme
subject heading language
subject heading list
SHL
structured *vocabulary* (3.10) comprising *terms* (3.4) available for subject indexing, plus rules for combining them into pre-coordinated strings of terms where necessary

[SOURCE: ISO 25964-2:2013, 3.77]

3.23
thesaurus
controlled vocabulary (3.11) and structured *vocabulary* (3.10) in which *concepts* (3.4) are represented by *terms* (3.9), organized so that relationships between concepts are made explicit, and preferred *terms* are accompanied by lead-in entries for synonyms or quasi-synonyms

[SOURCE: ISO 5127:2017, 3.8.3.01, modified — Notes to entry deleted.]

3.24
ontology
collection of *terms* (3.9), relational expressions and associated natural-language *definitions* (3.8) together with one or more formal theories designed to capture the intended interpretations of these definitions

[SOURCE: ISO/IEC 21838-1:2021, 3.14, modified — Note 1 to entry deleted]

4 Abbreviations

4.1 Abbreviations relating to terminology

BT Broader Term

NT Narrower Term

RT Related Term

UF Used For

4.2 Abbreviations relating to computing

CSV Comma Separated Values

JSON Java Script Object Notation

OWL Web Ontology Language

RDF Resource Description Framework

UML Unified Modeling Language

XML eXtensible Markup Language

5 Requirements for a concept system

5.1 Overview

A concept system is defined as a set of concepts structured in one or more related domains according to the concept relations among its concepts according to ISO 1087. Concept systems can be classified as a semantic spectrum by their increasingly precise definitions. Concept systems can be developed using different approaches.

The development of information technology and artificial intelligence technology aims to shift as much complex intelligent work as possible to computer systems. As a consequence, the core issue of artificial intelligence systems involves the creation of complete knowledge systems coupled with the implementation of knowledge-based reasoning skills. Using concept systems for abstracting and organizing knowledge has many benefits. A concept system serves to model concepts and relations among them based on specialized knowledge of a subject field, helps to form new concepts and terms, forms the basis for a uniform and standardized terminology, facilitates the implementation of automated reasoning based on certain rules and facilitate the indexing, retrieval, information organization and navigation of data resources.

5.2 Components of a concept system

5.2.1 General

This subclause provides a framework for the use of general concept systems in the information technology field. This framework describes the basic components and relations making up a general concept system, so as to assist the construction of different types of concept systems. The UML[8][9] diagram shown in Figure 1 shows the framework of a general concept system.

A concept system usually consists of (1) a list of concepts, (2) characteristics of concepts, (3) relations among concepts, and possibly (4) axioms.

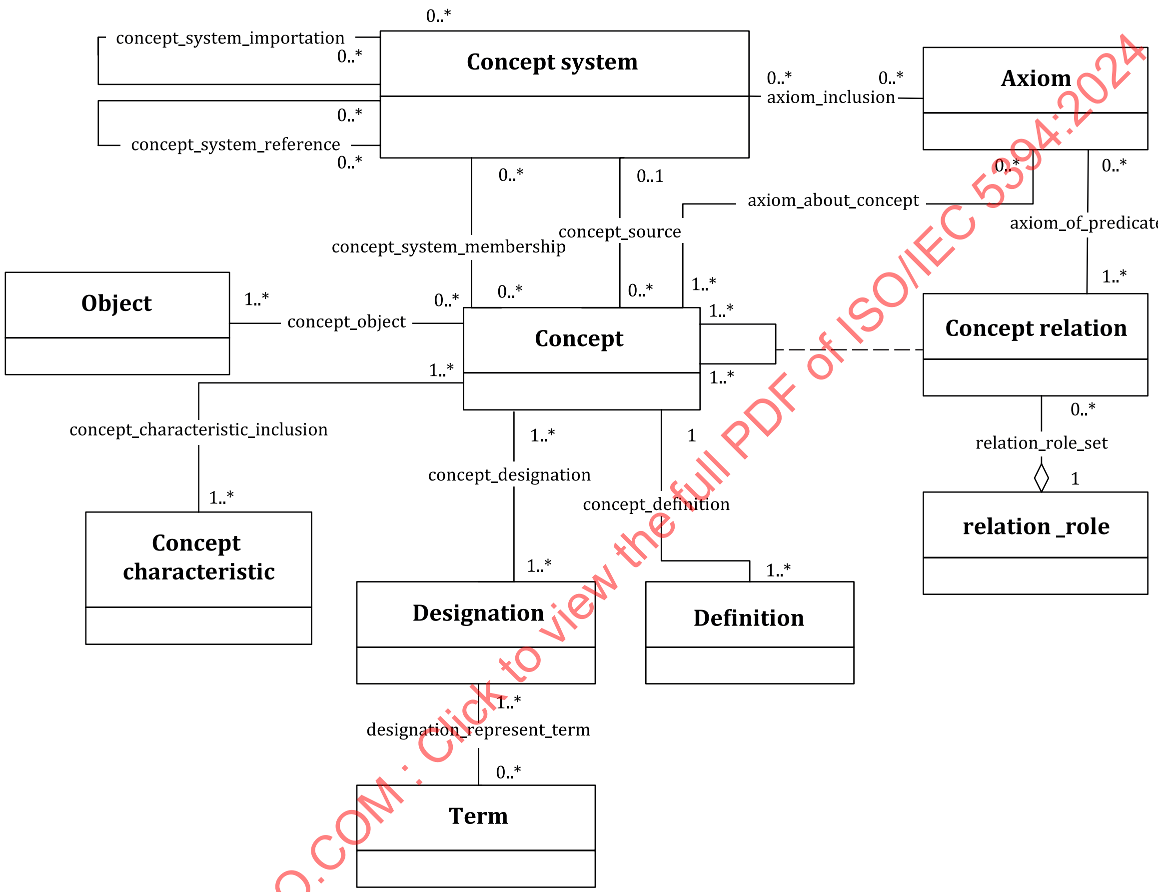


Figure 1 — Concept system framework

5.2.2 Concepts

Concepts are the core elements of any concept system. A concept is defined as a unit of knowledge created by a unique combination of characteristics which are each the abstraction of a property of an object or of a set of objects according to ISO 1087. An object is defined as anything perceived or conceived. Through observation and a process of abstraction, objects are categorized into classes, which correspond to units of knowledge called concepts.

The essential properties of the object are abstracted to form characteristics associated with the mental representation of the concept. Together, the set formed by these characteristics comprises the intension. The set of objects conceptualized as a concept is known as the extension of the concept.

Concepts are represented by different kinds of designations, such as terms, abbreviations, appellations, or even symbols and formulas. In addition, some semantic elements, such as definitions, are used to help to remove ambiguity, distinguish the concept from others, or to indicate the scope of the concept in concept systems.

5.2.3 Concept relations

Different concepts can have various concept relations according to the differences of their generalized objects, scope and degree. According to their intension and extension, concepts can not only be divided into different categories, but also can be assigned to different levels and series, which together constitute the concept system. ISO 704 and many other works on concepts specify a number of different types of relations among concepts. There are mainly two types of concept relations used to model a concept system: (1) hierarchical relation is the relation between two concepts which can be either a generic relation or a partitive relation, and (2) associative relation is a relation between two concepts derived from underlying relations between objects.

In a generic relation, the intension of one of the concepts includes that of the other concept and at least one additional delimiting characteristic. For example, 'vehicle' and 'car'.

In a partitive relation, one of the concepts constitutes the whole and the other concept a part of that whole. For example, 'hand' and 'finger', where a finger is a part of hand.

An associative relation illustrates concepts that are interrelated for sequence, space, causality and other factors. For example, 'production' and 'consumption' are in sequential relation. For example, 'action' and 'reaction' are in causal relation. The collection concept relation is also a kind of associative relation in this document. The collection concept relation that specifies a concept is an element of a collection of concepts. This relation does not specify the nature of the concept collection nor information about relations among the concepts that are elements in the concept collection. This relation can be used in glossaries, dictionaries, etc.

The relations among concepts constitutes the structure of a concept system. Some concept systems are mainly structured by hierarchical relations. Other concept systems can have different kinds of relations, incorporating generic, partitive, and associative relations within the same overall structural network.

5.2.4 Terms

When naming concepts, it is necessary to form or use appropriate, unambiguous, concise, linguistically accurate terms (designations) for each concept. The method of naming concepts shall be in accordance with ISO 704. When a concept is named, it is necessary to establish the relationship between the concept and the resulting term.

5.2.5 Definitions

Definitions are used to describe concepts. The method of writing definitions shall conform to ISO 704. A definition does not describe all the characteristics of the concept, but rather should capture the most essential characteristics required to differentiate the concept from other concepts.

5.3 Principles for forming concept systems

5.3.1 General

The following principles should be followed in the formation of a concept system:

— human and computer readability;

— comprehensiveness;

— clarity;

— extensibility;

— stability;

— retrievability;

— standardization.

5.3.2 Human and computer readability

A concept system should not only satisfy human understanding, but also be easy to read by computers. A concept system should be intuitive for humans to read and understand and available in at least one computer representation that can be interpreted by computers such as JSON, CSV, XML or UML[8][9].

5.3.3 Comprehensiveness

There should be no repetition or obvious omissions of concepts in the subject field that can be relevant to the targeted audience. A concept system should be complete including all or nearly all relevant concepts needed.

5.3.4 Clarity

A concept system should be understandable to users. The concept relations and criteria of subdivision used in creating the concept system should be clearly described for users. If the system is too complex it will lose its explanatory power. Complexity can be avoided by reducing the number of concepts and concept relations.

5.3.5 Extensibility

A concept system should be easy to expand and modify (see ISO 704), such that new concepts can be introduced to an appropriate position in the concept system. It should also be easy to establish new concept relations with existing concepts. There should be an organization or individual responsible for extensions of the concept system if any omissions are found or new requirements are identified.

5.3.6 Stability

A concept system should be stable enough that it will not be easily negated when new theories emerge. It should be possible to always maintain the relative stability and dynamic balance of the whole system.

5.3.7 Retrievability

The formation and organization of a concept system should be convenient for humans and information systems to retrieve the concepts they need, such as by using a search engine or information retrieval system.

5.3.8 Standardization

Use standards for principles and methods for terminology development including methods for concept designations and definitions according to ISO 704. Related standards should be checked to ensure that existing terms and definitions are used wherever possible.

5.4 Representations

A concept system can be expressed in various styles, such as: text in natural languages, tables, tree graphs, atom graphs, concept diagrams, UML[8][9], RDF, Common Logic[15] or OWL 2[12][13], etc.

Concept systems differ with respect to the expressive power of the language used in the specification. Some conceptualizations require a highly expressive language to define the concepts, whereas others can be specified using less expressive language depending on the target audience and its needs. An informal concept system can be expressed only as a list of terms and definitions.

Logical languages have both a formal syntax and a model-theoretic semantics. Examples include RDF, OWL[12][13] and Common Logic[15]. Semi-formal languages, such as XML and EXPRESS, have a formal syntax but sometimes lack model-theoretic semantics. There are also many kinds of concept systems whose terms and definitions are specified only in natural language. Where a concept system exists in graph-theoretical form, the concepts occupy the vertices of the graph. The edges of the graph then represent concept relations between the concepts. A node in a graph can also be called a concept system node.

5.5 Level of structure

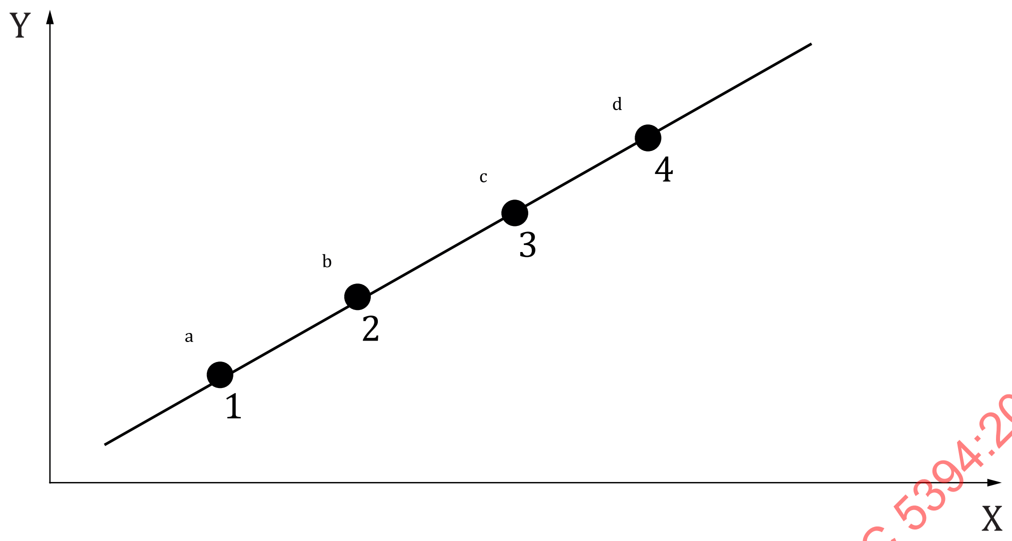
The level of structure associated with a given concept system corresponds with the strength and richness of the concept relations among concepts. Concept systems can be classified by the complexity of their structure, scaled starting with: weakly structured concept systems, hierarchical concept systems, hybrid concept systems and reason-based concept systems with machine-readable semantics, as shown in Figure 2.

— Weakly structured concept system: a concept system in which concepts are collected as simple sets of concepts or the concept relation is not obvious, and the terms assigned to concepts are mostly displayed in a simple way or alphabetical order.

— Hierarchical concept system: a concept system in which all the concepts relate to each other hierarchical, using generic relation, partitive relation and so on.

— Hybrid concept system: a concept system constructed using a combination of concept relations, including hierarchical relations and associative relations.

— Reason-based concept system: a concept system in which all the concepts have precise mathematical descriptions, the concepts and their concept relations are clearly and precisely defined. The concept system can be used for reasoning, especially for automatic machine reasoning, usually dependent on some sort of semantic representation (Common Logic[15], OWL[12][13], RDF, UML according to ISO 24156-1[14], etc.).



Key

X richness of concept relation types
Y semantic explicitness
1 weakly structured concept system
2 hierarchical concept system
3 hybrid concept system
4 reason-based concept system
[a] Example: Glossary ,Classification scheme.
[b] Example: Taxonomy.
[c] Example: Terminology, Subject heading scheme, Thesauri.
[d] Example: Ontology.

Figure 2 — Level of structure of concept systems

5.6 Relations between concept systems and vocabularies

Concept systems shall be represented by some kind of vocabulary.

Concept systems and vocabularies are interrelated and distinguished from each other. Concept systems are the basis for constructing vocabularies. Many vocabularies are based on concept systems. However, there are some differences between concept systems and vocabularies as follows.

— Components. The main component of a concept system is the concept, while the main component of vocabularies is the designation including terms, appellations, a proper name, or a symbol. A concept is the basic unit of knowledge, the summary of human cognition at a certain stage, and the crystallization of human cognition results. Most concepts are represented by terms. Language is the material shell of thought, while a term is the external language form of a concept. Terms also have phonetic form and grammatical function, but concepts do not.

— Language neutrality. Terms and concepts are represented in a specific natural language, but a concept system can be independent of a specific natural language.

— Types of relations. In addition to the concept relations, vocabularies always include the relationships among terms, such as synonymy, antonymy and equivalence, and the relationships between concepts and terms, such as mononymy, monosemy, homonymy and polysemy.

5.7 Concept system management

5.7.1 Developing concept systems

In the process of concept system construction, there are several issues to be solved.

a) Analyse requirements. Is there a requirements document? Are the requirements clear? Can requirements be broken down? What are the sources of the requirements? Do two or more requirements conflict with each other? Is this the latest version of the requirements? What are the old versions of requirements? Why change requirements? Does the change affect the consistency and integrity of the requirements document, etc.?

b) Define the domain and scope of the concept system. What is the domain covered by the concept system? What is the purpose of the concept system? What kind of information does the concept system provide? Who uses and maintains the concept system, etc.?

c) Consider reusing an existing concept system. Refine and expand existing resources for specific areas or tasks. If a concept system applies to multiple applications, it can be expedient to reuse the concept system.

d) Enumerate important concepts and terms in the concept system. Analyse the characteristics of these concepts, determine which properties comprise essential characteristics, and then build definitions of concepts. Select the preliminary concepts to be treated by taking into account the subject field, the user group and their needs.

e) Define concepts and relations among concepts. Determine the transitivity of the relations. Avoid circular and redundant statements or logic. Analyse coordinate concepts for the concept under consideration. Determine multiple inheritance relations if they exist.

f) Define the constraints for the properties. Specify the value type, and value domain of any properties.

5.7.2 Management of concept systems

Issues can to be considered in managing concept systems:

— concept system governance;

— policy governing interaction with users;

— policy for making and approving changes to the concept system;

— concept system versioning policy;

— policy for creation and maintenance of identifiers.

5.7.3 Integration of concept systems

Issues can to be considered in integrating concept systems:

— concept system mapping;

— concept system merging;

— concept system alignment.

Annex A
(informative)

Examples of representing various forms of concept systems

A.1 Glossaries

A.1.1 General

A glossary is a terminological dictionary which contains a list of designations and their definitions from a subject field, and can contain equivalents in one or more languages. The traditional aim of glossaries is to improve the clarity of communication in spoken or written discourse. Traditionally, a glossary appears at the end of a book and includes terms used in that book that are either newly introduced, uncommon, or specialized. In the 1990s, it became increasingly common for enterprises to build glossaries for applications across their entire range of business activities in order to improve the consistency and quality of information sharing and publishing. Glossary work focuses on creating a complete list of the terminology of domain-specific terms and acronyms.

A.1.2 Semantic components and relations

The main semantic components of a glossary are designations and definitions.

In a glossary, the “concept” is represented by “designations”. When cases of synonymy arise, one of them becomes established as a “preferred term”, while acceptable synonyms are simply “admitted terms” and unacceptable ones become “deprecated terms”.

Some glossaries also associate a definition with each concept. The definition helps to clarify the semantic boundaries of a concept in order to distinguish this concept from others within the domain.

Glossaries are usually sorted alphabetically for retrieval purposes. Concepts in the glossary usually do not need to be organized using complex concept relations.

A.1.3 Sample glossary

The Glossary of Unicode Terms[23] is an example of a glossary, which is updated periodically to stay synchronized with changes to various standards maintained by the Unicode Consortium. Translations of Unicode[25] and ISO/IEC 10646[3] terminology are also available.

A.2 Classification schemes

A.2.1 General

A classification scheme defines descriptive information for the arrangement or division of objects (ideas) into groups based on criteria such as the properties associated with the objects (ideas).

The traditional purposes of classification schemes are to organize information resources, whether in hard copy or electronic, and to allow retrieval by browsing the shelves or by browsing a classified catalogue. In the metadata field, classification schemes can provide a sound conceptual basis for the development of metadata having enhanced semantic purity and design integrity. A classification scheme enables users to find a single object from among a large collection of objects in a metadata registry, and facilitates the discovery, administration, and analysis of a collection of objects.

Classification schemes have many sub-types. Some types include classification schemes used for records management, taxonomies and subject classification schemes, etc.

For more on methodologies for developing classification schemes, see ISO 22274[13] and ISO 25964-2[17].

A.2.2 Semantic components and relations

The basic approach of classification is to organize concepts into classes. The main semantic components of a classification scheme are notations and captions. A notation is a set of symbols representing a concept or a class in classification schemes. A notation can also take the form of a class code. The caption is the text label of the class represented by a notation in a classification scheme.

In subject classification schemes, concepts are not usually arranged in hierarchical concept relations. For example, “chemistry” and “chemical elements” are not represented in any hierarchical concept relations. However, in a library it is convenient to be able to find books about chemical elements in the chemistry section. Therefore, the class “chemical elements” can be the sub-class of “chemistry” in subject classification schemes.

A.2.3 Sample classification scheme

The Library of Congress Classification (LCC)[24] is an example of a classification scheme. LCC has been widely used by many libraries. LCC is a classification scheme which was originally designed for organizing books based on their content. LCC uses Latin letters and Arabic numerals to number the categories. A major class uses a single letter, such as B for Philosophy/Psychology/Religion, and P for Languages and literature. In general, two letters are used for subordinate categories, such as PG for Soviet literature and PQ for Published literature. A real number from 1 to 9999 is added after the letter. For instance, the sub-class for Spermatophyta under Botany is QK474.8-495.

A.3 Taxonomies

A.3.1 General

The term “taxonomy” is used widely. Taxonomies range from the very simple to the very complex. In the simplest taxonomies the categories are not necessarily divided into subcategories, while in the complex ones multiple hierarchical levels can be found. However, in this clause, taxonomy is a collection of controlled vocabulary terms organized into a hierarchical structure which deals with hierarchical relations such as genus-species, part-whole, etc. It has some of the same features described in classification schemes, or in thesauri.

The typical taxonomy is used for classifying or categorizing, organizing, browsing, navigating, searching and/or filtering any type of content in networked environments. A common use case is to support navigation, especially by hierarchical organization and browsing through a broad set of electronic resources, e.g. websites, intranets, portals, wikis. To complement the navigational features with a search capability, taxonomies can include synonyms operating behind the scenes as entry terms, and “See also” references between related categories in the hierarchy.

Some taxonomies allow poly-hierarchies, which means that a term can have multiple parents. This means that if a term appears in multiple places in a taxonomy, then it still remains the same term and applies to the same concept. Specifically, if a term has children in one place in a taxonomy, then it has the same children in every other place where it appears.

For more on methodologies for developing taxonomies, see ISO 25964-2[17].

A.3.2 Semantic components and relations

The basic units of a taxonomy are usually known as “categories” or “taxa”. A category can encompass either a single concept, e.g. “Human rights” or a combination of concepts, e.g. “Government, citizens and rights”. A combination like this is often created as a presentational device to group several more specific categories.

Each category is given a category label, which in some taxonomies is unique but in others is not disambiguated. Category labels often have synonyms. These terms can include near-synonyms, abbreviations, acronyms and lexical variants.

The hierarchical relation is the most important relation in taxonomies. However, some taxonomies incorporate associative relations between related categories by using “See also” references.

A definition or scope note or both can be attached to a category where necessary to clarify its scope.

Optionally, rules for automatic categorization, automatic querying and personalization can be attached to the categories.

A.3.3 Sample taxonomies

The National Center for Biotechnology Information (NCBI) Taxonomy[25] is an example of a taxonomy, which includes organism names and classifications for every sequence in the nucleotide and protein sequence databases of the International Nucleotide Sequence Database Collaboration (INSDC). The NCBI provides a framework for clustering elements within other domains of NCBI web pages, for internal linking between domains of the Entrez Molecular Sequence Database System and for linking out to taxon-specific external resources on the web and relevant publications. Each entry in the NCBI Taxonomy database is a “taxon”, also referred to as a “node” in the database, and can be associated with secondary names of several types – synonyms, scientific name, common names and so on. The “root node” (taxid1) is at the top of the hierarchy. The path from the root node to any other particular taxon in the database is called its “lineage”; the collection of all of the nodes beneath any particular taxon is called its “subtree”. Fielded Boolean queries are used in the NCBI Taxonomy to answer many questions. For example, how many amphibian taxa have appeared in NCBI Taxonomy for the first time this year?

A.4 Terminology

A.4.1 General

A terminology is a set of designations and concepts belonging to one domain or subject, each designation representing a concept by means of a sign, symbol, term or appellation. Terminological data can be presented in various formats, for example as term banks, termbases, glossaries (see A.1) or other resources. An important aim of a terminology is to support consistency in the use of terms. The range of approaches commonly covered by terminologies comprises a continuum of resources reflecting increasing degrees of control and prescriptive activity, from uncontrolled vocabulary to standardized terminologies created as consensus-based resources for vocabulary used, for instance, in documentation of various kinds.

Terminologies can be useful in the construction or maintenance of a thesaurus, as a source of concepts and/or of terms. They can also help with writing scope notes, guiding the choice of preferred terms, and supplying term definitions if needed.

An uncontrolled vocabulary comes from natural language, which is not standardized and effectively controlled by human beings, and its quantity is not strictly controlled. Therefore, an uncontrolled vocabulary is not a type of concept system. However, an uncontrolled vocabulary can support the expansion and maintenance of a concept system.

In terminology, a designation is the representation of a concept by a sign which denotes it. There are several types of designations such as abbreviations, terms, and appellations. An abbreviation is a kind of designation formed by omitting words or letters from a longer form and designating the same concept. A term is a verbal designation of a general concept in a specific subject field. An appellation is a verbal designation of an individual concept.

There are three main types of designation relations between designations:

— Synonymy - relation between or among terms in a given language representing the same concept

— Antonymy - relation between two terms in a given language representing opposite concepts

— Equivalence - relation between designations in different languages representing the same or similar concept. Equivalencies can be exact or inexact, total or partial, single to multiple, one to part of one, etc

There are four main types of relations between designations and concepts:

— Mononymy - relation between designations and concepts in a given language in which one concept has only one designation

— Monosemy - relation between designations and concepts in a given language in which one designation only relates to one concept

— Homonymy - relation between designations and concepts in a given language in which one designation represents two or more unrelated concepts

— Polysemy - relation between designations and concepts in a given language in which one designation represents two or more concepts sharing certain characteristics

Definitions (concept definitions) are used to represent the concept by a descriptive statement or a formal expression which serves to differentiate it from related concepts.

A.4.2 Semantic components and relations

The main semantic components of a terminology are designations, definitions and relations.

In a terminology, each "concept" is represented by one or more "designations". When cases of synonyms arise, one of them becomes established as a "preferred term", while acceptable synonyms are simply "admitted terms" and unacceptable ones become "deprecated terms".

In a terminology, the definition serves to clarify the semantic boundaries of a concept in order to distinguish this concept from others within the domain.

The three main types of relations found in terminologies are:

a) Hierarchical relation between two concepts;

b) Associative relation between two concepts;

c) Synonym relations between terms in the same language.

A.4.3 Sample terminology

The World Health Organization Lexicon of alcohol and drug terms[26] is an example of a terminology which aims to provide a set of definitions of terms concerning alcohol, tobacco and other drugs. It is useful to clinicians, administrators, researchers, and other interested parties in this field. Main diagnostic categories in the field are defined, as are key concepts in scientific and popular use. Social as well as health aspects of drug use and problems related to use are covered. The lexicon does not provide comprehensive coverage of every term; areas that have been excluded are, for example, production and marketing terms and slang terms.

A.5 Subject heading schemes

A.5.1 General

A subject heading scheme is a type of controlled vocabulary. It represents concepts in the form of terms or phrases, and provides syntactic rules for combining terms into pre-coordinated strings that represent more complex concepts and topics.

A primary function of subject heading schemes is to bring together related topics so that collections of information resources can be organized systematically on the basis of their subject content, and to facilitate browsing and navigation around a subject domain.

Subject heading schemes were created at the end of the 19th century as a tool used to systematize subject access to information resources, when cataloguers began to create lists of subject terms to ensure indexing consistency within their own institution's catalogue. Since then, various lists of subject headings have been established and used, such as Library of Congress Subject Headings (LCSH)[27], Medical Subject Headings (MeSH)[33], etc.

For more on methodologies for developing subject heading schemes, see ISO 25964-2[17].

A.5.2 Semantic components and relations

The main components of a subject heading scheme are headings, subdivisions and relations between them. The scheme can also include or be accompanied by rules for when and how to combine these components.

Headings can consist of a term representing a single concept (e.g. “Arithmetic”) or a combination of distinct concepts (e.g. “Mines and mineral resources”). The form and meaning of each heading are controlled, albeit less strictly than is the case in a thesaurus. Subject headings can be usable for indexing (accepted headings) or not usable for indexing (rejected headings). Homographs are disambiguated by a qualifier.

A subject heading can be simple, or it can be compound, if one or more subdivisions are added to the initial heading to form a pre-coordinated string. The function of subdivisions is to represent the perspective, point of view, form, etc. A double dash generally connects the subdivision to its initial heading.

The three main types of relations found in most subject heading schemes are:

a) Hierarchical relation between two concepts [indicated by level of indentation, by typographic devices, or by an explicit tag such as BT (broader term) and NT (narrower term)];

b) Associative relation between two concepts [indicated by cross-references such as See also or by the tag RT (related term)]; and

c) Equivalence relationship as it applied in both monolingual and multilingual situations [indicated by cross-references such as See, from, or by the tags USE and UF (used for)].

A.5.3 Sample subject heading schemes

The Library of Congress Subject Headings (LCSH)[27] is an example of a subject heading scheme, which remains the most widely used subject heading scheme in the world. It has been translated into many languages, and is used around the world by libraries large and small. LCSH has been actively maintained since 1898 to catalog materials held at the Library of Congress. Proposals for additions and changes are reviewed regularly at staff meetings in the Policy and Standards Division (PSD) and an approved list is published.

In LCSH, headings are listed in boldface type, e.g., Nuclear physics, Home-based businesses**.** A heading may be followed by the legend (May Subd Geog), which shows that the heading may be subdivided by places according to the rules in the Manual, and by class numbers (e.g. [QC770-QC798]). Scope notes giving guidance on the meaning or application of a heading may follow the heading Home-based businesses in separate paragraphs, followed by subdivisions of the subject headings (e.g. Law and legislation etc.). References associated with the headings are then listed in groups marked with explicit tags such as “UF”, “BT”, “NT” etc.

A.6 Thesauri

A.6.1 General

A thesaurus is a collection of controlled vocabulary terms organized into a hierarchical structure and preferred terms are accompanied by lead-in entries for synonyms.

The traditional aim of a thesaurus is to guide the indexer and the searcher to choose the same term for the same concept. In order to achieve this, a thesaurus should first list all the concepts that might be useful for retrieval purposes in a given domain. The concepts are represented by terms, and for each concept, one of the possible representations is selected as the preferred term. Secondly, the thesaurus should establish relationships between terms in a structured display.

For more on methodologies for developing thesauri, see ISO 25964-1[16].

A.6.2 Semantic components and relations

The main components of a thesaurus are terms, definitions and relations between them.

The terms selected to represent concepts can be single-word terms or multi-word terms. All thesaurus terms are expressed as unambiguously as possible. Sometimes more than one term in the same language is available to represent the same concept. When multiple terms are included corresponding to the same concept, one of these is designated as the preferred term and the others as non-preferred terms.

When homographs are needed as thesaurus terms, the qualifier is used to indicate the context or subject area to which the concept belongs. It forms part of the term and does not serve as a scope note.

When the meaning of the preferred term in ordinary discourse can be interpreted too broadly or too narrowly, a scope note is used to clarify the boundaries of the concept. Unlike the qualifiers, a scope note is not regarded as forming part of the term to which it is attached.

A full definition is not usually required to clarify the way in which a preferred term is used. The source of each definition is recorded alongside the definition itself.

The purpose of establishing and displaying relations is likewise to guide users (or intelligent agents acting on their behalf) to choose the most appropriate term(s) for expressing a given concept.

The three main types of relations found in most thesauri are:

a) equivalence relationship as applied in both monolingual and multilingual contexts;

b) hierarchical relation between superordinate concept and subordinate concepts; the broader term represents a superordinate concept, and the narrower term represents a subordinate concept;

c) associative relation.

A.6.3 Sample thesaurus

The UNESCO Thesaurus[28] is an example of a thesaurus, which is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. It is continuously enriched and updated. The thesaurus is available in English, French, Russian and Spanish. Concepts are grouped into 7 broad subject areas.

A.7 Ontologies

A.7.1 General

In computer science, more specifically in the field of knowledge engineering and artificial intelligence, an ontology is defined as collection of terms, relational expressions and associated natural-language definitions together with one or more formal theories designed to capture the intended interpretations of these definitions[20][23]. An ontology has a precise mathematical description, in which the precision of concepts and their relationships is clearly defined. A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints.

The main fundamental purposes of an ontology are:

— inferring class membership for individuals;

— inferring relationships between classes and properties; and

— checking the consistency of a knowledge base.

Unlike other concept systems described in this document ontologies are not designed for information retrieval by index terms or class notation, but for making assertions about individuals, e.g. about real persons or abstract things such as a process.

Ontologies can be distinguished by the degree of specificity, the domain covered or the application purpose. Thus, top-level ontologies distinguish the most general, domain-independent categories of existence.

Unlike thesauri, ontologies necessarily distinguish between classes and individuals, in order to enable reasoning and inferencing. The concepts documented in a thesaurus and the classes categorized in an ontology represent meaning in two fundamentally different ways. Thesauri express the meaning of a concept through terms, supported by adjuncts such as a hierarchy, associated concepts, qualifiers, scope notes, a precise definition or some combination of these all directed mainly to human users. Ontologies, in contrast, convey the meaning of classes through machine-readable membership conditions.

For more on methodologies for developing ontologies, see ISO 25964-2[17].

A.7.2 Semantic components and relations

The main components to consider are classes, properties, axioms, individuals (instances) and assertions which adopt Web Ontology Language (OWL[12][13]) terminology.

An ontology class is a construct with a set of property constraints that establish the criteria for membership.

Axioms are statements specifying the basis of the classes, properties and other entities in an ontology. An axiom can be as simple as a statement that a specific class or property exists. Other examples of axiomatic statements describe the properties of a class (e.g. the class “Planet” is required to have an attribute “hasMass”). Class hierarchies in ontologies are created through sub-class axioms between classes. In the sub-class relation, all axiomatic statements of a parent class also apply to its child classes.

Individuals are the objects of discourse in a particular domain and are sometimes referred to as instances. They are the fundamental things that an ontology makes statements about. Examples of individuals are a specific person, abstract things, a process or event, etc.

Assertions, a particular group of axioms, are statements about individuals in a domain. The “class assertion”, in particular, states that an individual is a member of a class. Other assertions can state that two individuals are the same or that they have certain property constraints. Assertions and individuals are generally brought in as part of the ontology application rather than including them in the ontology itself.

Classes and properties, as well as individuals, all have identifiers. Although natural language labels are not strictly necessary in ontologies, they are often provided to enhance readability for humans.

A.7.3 Sample ontologies

The Cell Ontology (CL)[29] is an example of an ontology, which is designed as a structured controlled vocabulary for cell types. This ontology was constructed for use by the model organism and other bioinformatics databases, where there is a need for a controlled vocabulary of cell types. This ontology is not organism-specific: it covers cell types from prokaryotes to mammals. However, it excludes plant cell types, which are covered by the Plant Ontology (PO)[36].

A.8 Comparison between various concept systems

We analyse the function, semantic components, relation types and examples of each of them as shown in Table A.1.