

Available Article

Author's final: This draft is prior to submission for publication, and the subsequent edits in the published version. If quoting or citing, please refer to the proper citation of the published version below to check accuracy and pagination.

Cite as: Bergman, M. K. Glossary. in *A Knowledge Representation Practionary: Guidelines Based on Charles Sanders Peirce* (ed. Bergman, M. K.) 435-449 (Springer International Publishing, 2018). doi:10.1007/978-3-319-98092-8

Official site: <https://link.springer.com/book/10.1007/978-3-319-98092-8>

Full-text: <http://www.mkbergman.com/publications/akrp/glossary.pdf>

GLOSSARY

Throughout this book many terms have been presented, often defined in context. This section attempts to gather all of these definitions. Sources include my online glossary¹ on my working blog, but many sources have also contributed.^{2 3 4 5} Note that *italicized* entries refer to other terms defined in the glossary. Also, see the glossary at the end of Brent's biography.⁶

A

ABox

An ABox (for assertions, the basis for A in ABox) is an "assertion component"; that is, a fact associated with a terminological vocabulary within a knowledge base. ABox are TBox-compliant statements about instances belonging to the concept of an ontology. Instances and instance records reside within the Abox

Abductive reasoning

Abduction (or abductive reasoning) is a mode of symbolic inference that involves the screening and selection from a domain D of the possible explanation paths to an outcome O possibly involving any element E of D, with the selection of candidate paths for inductive testing based on plausibility, economy, and potential impact; abduction does not produce probable results, only winnowed candidates.

Access control

Access control is the protection of resources against unauthorized access; a process by which use of resources is regulated according to a security policy and is permitted by only authorized system entities according to that policy; see further RFC 2828

Actions

Actions are reactions to perceptions or stimuli, are energetic, or are thought, as understood to be broadly construed; actions reside in Secondness

Activities

Activities are sustained actions over durations of time; activities may be organized into natural classes

Accuracy

A statistical measure of how well a binary classification test correctly identifies or excludes a condition. It is calculated as the sum of true positives and true negatives divided by the total population

Adaptive ontology

An adaptive ontology is a conventional knowledge representational ontology that has added to it a number of specific best practices, including modeling the ABox and TBox constructs separately; information that relates specific types to different and appropriate display templates or visualization components; use of preferred labels for user interfaces, as well as alternative labels and hidden labels; defined concepts; and a design that adheres to the open world assumption

Administrative ontology

Administrative ontologies govern internal application use and user interface interactions

Annotation

Annotations are indexes and the metadata of the KB; these cannot be inferred over and do not participate in reasoning or coherency testing. But, they can be searched, and language features can be processed in other ways. Annotations may be grouped for convenience, but may not be typed

API

An application programming interface (API) defines how communication may take place between applications. Implementing APIs that are independent of a particular operating environment (as are the W3C DOM Level 2 specifications) may reduce implementation costs for multi-platform user agents and promote the development of multi-platform assistive technologies. Implementing conventional APIs for a particular operating environment may reduce implementation costs for assistive technology developers who wish to interoperate with more than one piece of software running on that operating environment

A KNOWLEDGE REPRESENTATION PRACTICIONARY

Architectonic

Architectonic is a governing philosophy expressed iconically as a design, schema or structure

Architecture

Architecture, as limited to the use herein, is the structure of a knowledge base (or bases) written in a knowledge representation formalism, and embedded in a general knowledge management platform. The architecture combines knowledge artifacts with software program(s) or computing system(s), the relationships among them, and the conditions on their use

Artificial intelligence

AI is the use of computers to do or assist complex human tasks or reasoning. There are many, broad sub-fields from pattern recognition to robotics and complex planning and optimizations.

Aspects

Aspects are aggregations of an entity type that are grouped according to features or views different from the type itself. As examples, the type of 'music composer' may have an aspect of being from the 19th century, or 'authors' may have the aspects of being Russian or writing novellas. The organization of aspects closely parallels that for SuperTypes

Assertion

Assertion is a statement, wherein a fact or logical expression with consequences is made

Attributes

Attributes are the intensional characteristics of an object, event, entity, type (when viewed as an instance), or concept. The relationship is between the individual particular and its attributes and characteristics, in the form of A:A. Attributes may be intrinsic characteristics or essences of single particulars, adjunctual or accidental happenings to the particular, or contextual in time or space or situations. Collectively known as depth, comprehension, significance, meaning or connotation

Attribute type

Attribute types are an aggregation (or class) of multiple attributes that have similar characteristics amongst themselves (for example, colors or ranks or metrics). As with other types, attribute types do not have attributes

Axiom

An axiom is a premise or starting point of reasoning. In an ontology, each statement (assertion) is an axiom

B

Base concept

Base concepts are all of the *classes* in the overall KBpedia, comprised of the KBpedia Knowledge Ontology and all of its official typologies

Belief

Belief is a state of evidence sufficient to enable action

Binding

Binding is the creation of a simple reference to something that is larger and more complicated and used frequently. The simple reference can be used instead of having to repeat the larger thing

Blank node

Also called a bnode, a blank node in *RDF* is a *resource* for which a *URI* or literal is not given. A blank node indicates the existence of a thing, implied by the structure of the *knowledge graph*, but which was never explicitly identified by giving it a *URI*. Blank nodes have no meaning outside of their current graph and therefore cannot be mapped to other resources or graphs

C

Cardinality

Cardinality is where the number of members in a class or type is set or limited, such as `hasBiologicalParent` being set to two

Class

Class is a collection of one or more *instances* or *classes* that share the same potential *attributes* or *relations*; *concepts* and *entity types* are both classes

Closed world assumption

CWA is the premise that what is not currently known to be true is false. CWA is the most common logic applied to relational database systems, useful for transaction-type systems. In knowledge management, CWA is used in at least two situations: 1) when the *knowledge base* is known to be complete, and 2) when the knowledge base is known to be incomplete, but a 'best' definite answer must be derived from incomplete information. See contrast to the *open world assumption*

GLOSSARY

Coherence

Coherence is the state of being coherent for logic systems, where the knowledge base (domain) is consistent and has a high degree of conjunction for non-deductive assertions

Collection

See *class*

Complete

Complete is an evaluative criterion for logic systems where all *statements* which are true in all models are provable

Concept

See *class*

Cyc

Cyc is a common-sense *knowledge base* that has been under development for over 20 years backed by 1000 person-years of effort. The smaller OpenCyc version is available in OWL as open source; a ResearchCyc version of the entire system is available to researchers. The Cyc platform contains its own logic language, CycL, and has many built-in functions in areas such as *natural language processing*, search, inferencing and the like. *UMBEL* is based on a subset of Cyc

Cycle

A cycle is where, in a graph, a path from a given *node* may reach itself (such as $A \rightarrow B \rightarrow C \rightarrow A$). In a subsumption hierarchy, this is an error

D

Data integration

Data integration is the bringing together of data from heterogeneous and often physically distributed data sources into a single, coherent view

Dataset

Dataset is a combination of one or more *records*, transmitted as a single unit (though it may be split into parts due to size)

Datatypes

Datatypes are pre-defined ways that attribute values may be expressed, including various literals and strings (by language), URIs, Booleans, numbers, date-times, etc. See XSD (XML Schema Definition) for more information

DBpedia

A project that extracts structured content from *Wikipedia*, and then makes that data available as *linked data*. There are millions of entities characterized by DBpedia in this way. As such, DBpedia is one of the largest – and most central – hubs for *linked data* on the Web

Deductive reasoning

Deductive reasoning extends from premises known to be true and clear to infer new facts

Description logics

Description logics and their semantics traditionally split *concepts* and their *relationships* from the different treatment of *instances* and their *attributes* and roles expressed as *fact assertions*. The concept split is known as the *TBox* and represents the schema or *taxonomy* of the *domain* at hand. The TBox is the structural and *intensional* component of conceptual relationships. The second split of instances is known as the *ABox* and describes the attributes of instances (and individuals), the roles between instances, and other assertions about instances regarding their class membership with the TBox concepts

Disjoint

Disjoint is where membership in one *class* specifically excludes membership in another; this is a useful *property* in that it allows large, well-designed *knowledge bases* to be 'sliced-and-diced' for more effective processing or analysis

Distant supervision

A method to use *knowledge bases* to label *entities* automatically in text through *machine learning*, which is then used to extract features and train a machine learning classifier. The knowledge bases provide coherent positive training examples and avoid the high cost and effort of manual labeling

Documents

Documents (or articles or *records*) may be in the form of articles or data records. Whatever the form, extractions are needed to convert source information into the *triples* useful to the *knowledge base*

Domain

Domain is the bounded scope of real-world considerations that may contribute to the *knowledge representation* or information queries at hand. Scoping the domain is one of the first activities undertaken in a new KR project

A KNOWLEDGE REPRESENTATION PRACTITIONARY

Domain (property)

Domain, as applied to a property, is a statement that declares the classes or types from which the subject of the assertion must be drawn

Domain ontology

Domain (or content) *ontologies* embody more of the traditional ontology functions such as information interoperability, inferring, reasoning and conceptual and knowledge capture of the applicable domain

E

Edges

Edges, in a graphical representation of a *knowledge graph*, are the connections or the *relations* between *subjects* and *objects*; edges are the linking *properties* in a “triple” *statement*

Entailment

Entailment is a consequence arising from a statement deemed to be true based on some underlying logic. The logical consequence is said to be *necessary* and *formal*; necessary, because of the rules of the logic (the conclusion is the consequent of the premises); and formal because the logical form of the statements and arguments hold true without regard to the specific *instance* or content

Entity

Entities are the basic, real things in our domain of interest; they are nameable things or ideas that have an identity, are defined in some manner, can be referenced, and may be related to *types*; entities are most often the bulk of an overall *knowledge base*. An entity is an *individual object* or *instance* of a *class*, a Secondness; when affixed with a proper name or label it is also known as a *named entity* (thus, named entities are a subset of all entities). Entities are described and characterized by *attributes*. Entities are connected or related to one another through *external relations* and are referred to, signified or indexed by *representations*. An entity may be independent or separate or can be part of something else. Entities cannot be *topics* or *types*

Entity recognition

The use of *natural language processing* to identify specific *entities* in text. Often used in conjunction with named entities, where it is abbreviated NER

Entity type

Entity types are the aggregations or collections or *classes* of similar entities, which also share some essence; entity types have the *attributes* (but not the same values) of instances of the type

Essence

The *attribute* or set of attributes that make an entity what it fundamentally is; it is a unique or distinguishing attribute that helps define a *type*

Event

Events are nameable sequences of time, are described in some manner, can be referenced, and may be related to other time sequences or types. Events are like entities, except they have a discrete time beginning and end. Events are a Secondness and may be typed

Extensional

The extension of a *class*, concept, idea, or sign consists of the things to which it applies, in contrast with its *intension*. For example, the extension of the word “dog” is the set of all (past, present and future) dogs in the world. The extension is most akin to the *attributes* or characteristics of the *instances* in a set defining its class membership

External linkages

External linkages (or *mappings*) are any of the relational properties may be used to map external datasets and schema to KBpedia. In its base form, which can be expanded, KBpedia has mappings to more than 20 external sources

External relations

External relations are assertions (relationships) between an *object*, *event*, *entity*, *type*, or *concept* and another particular or general. An external relationship has the form of *A:B*. External relations may be simple ones of a direct relationship between two different instances; may be copulative by combining objects or asserting membership, quantity, action or circumstance; or mediative to provide meaning, context, relevance, generalizations, or other explanations of the subject with respect to the external world. External relations are extensional

F

Fact

A basic *statement* or *assertion* within a *knowledge graph* or *knowledge base*

GLOSSARY

Fallibility

Fallibility is the doctrine that truth is a limit function, unknowable in the absolute, and provides the logical basis for questioning premises

False negative

An error where a test result indicates that a condition failed, while it actually was successful. That is, the test result indicates a negative when the correct result should have been positive. Also known as a false negative error or *Type II error* in statistics. It is abbreviated FN

False positive

An error where a test result indicates that a condition was met or achieved, while it actually should have failed. That is, the test result indicates a positive when the correct result should have been negative. Also known as a false positive error or *Type I error* in statistics. It is abbreviated FP

Feature

A feature is a measurable property of the system being analyzed, equivalent to what is known as an explanatory variable in statistics

Feature engineering

Feature engineering is a process of creating, generating and selecting the *features* to be used in *machine learning*, based on an understanding of the underlying data and choosing features based on their likely impact on learning results and effectiveness

Firstness

Firstness is possibility, the essences of what may be, the unexpected chance occurrence

Folksonomy

A folksonomy is a user-generated set of open-ended labels called *tags* organized in some manner and used to categorize and retrieve Web content such as Web pages, photographs, and Web links

Function

Function is any algebraic or logical expression allowable by the semantics and primitives used in the KR language where an input is related to an output

G

Generals

Generals are the mediating, continuous, vague and indeterminate aggregations of instances into concepts, *classes*, *types*, collections or sets. Generals are in Thirdness. Generals may often be considered real, and their understanding and identification may be shared through *representations*

GeoNames

GeoNames integrates geographical data such as names of places in various languages, elevation, population, and others from various sources

Gold standard

A gold standard is a reference, benchmark test set where the results are already scored and known, with a minimum (if not zero) amount of *false positives* or *false negatives*; good gold standards also include *true negative* results

I

Identifier

Identifier is a reference pointer, a sign pointing to an *object*, but not the *object* itself. Identifiers should not be confused with the naming or defining label for the *object*; in practice, it is often a unique string assigned to the object. In RDF and KBpedia this identifier is a *URI*

Individual

Individual in RDF and OWL (indeed, commonly in description logics) is synonymous with an *instance* or *entity*; we try not to use this term because of general terminological confusion; see *instance*

Inductive reasoning

A method of reasoning where lines of possible evidence are weighed to determine probable outcomes

Inference

Inference is the act or process of deriving logical conclusions from premises known or assumed to be true. The logic within and between *statements* in an *ontology* is the basis for inferring new conclusions from it, using software applications known as inference engines or *reasoners*

A KNOWLEDGE REPRESENTATION PRACTICIONARY

Instance

Instances are individual *entities* or *events*, the ground level components of a *knowledge base*. Instances may include concrete objects such as people, animals, tables, automobiles, molecules, and planets, as well as abstract instances such as numbers and words; instances are in Secondness. An instance is also known as an *individual*, with *member* and *entity* also used somewhat interchangeably

Instance record

An *instance* with one or more *attributes* also provided

Intensional

The intension of a class is what is intended as a definition of what characteristics or properties its members should have. Intension is most akin to the *attributes* or characteristics of the *instances* in a set defining its class membership. It is therefore like the key-attribute pair aspects of an instance (or *ABox*) in an *ontology*

Inverse

Inverse is when a *property*, say, `hasParent`, can be defined as the inverse property of `hasChild`

K

Key-value pair

Also known as a *name-value pair* or *attribute-value pair*, a key-value pair is a fundamental, open-ended data representation. All or part of the data model may be expressed as a collection of tuples `<attribute name, value>` where each element is a key-value pair. The key is the defined *attribute*, and the value may be a reference to another object or a literal string or value. In *RDF triple* terms, the subject is implied in a key-value pair by nature of the *instance record* at hand

Kind

Used synonymously herein with *class*

Knowledge base

A knowledge base (abbreviated KB or kb) is a special kind of database for knowledge management. As used in KBpedia, a KB includes *instances* and *classes* related to each other via triple statements

Knowledge-based artificial intelligence

Knowledge-based artificial intelligence, or KBAI, is the use of large statistical or knowledge bases to inform feature selection for machine-based learning algorithms used in AI

Knowledge graph

See *ontology*

Knowledge management

Knowledge management, or KM, is the practice of creating, sharing, finding, annotating, connecting, and extending information and knowledge for a given *domain*

Knowledge representation

A field of *artificial intelligence* dedicated to representing information about the world in a form that a computer system can utilize to solve complex tasks

Knowledge supervision

A method of *machine learning* to use *knowledge bases* in a purposeful way to create features, and negative and positive *training sets* in order to train the classifiers or extractors. *Distant supervision* also uses knowledge bases, but not in such a purposeful, directed manner across multiple machine learning problems

L

Leaf nodes

Leaf nodes are terminal nodes in a tree structure, often representing instances (but not always so)

Linkage

A specification that relates an *object* or *attribute* name to its full *URI* (as required in the *RDF* language)

Linked data

Linked data is a set of best practices for publishing and deploying *instance* and *class* data using the *RDF* data model, and uses *uniform resource identifiers* (URIs) to name the data objects. The approach exposes the data for access via the HTTP protocol while emphasizing data interconnections, interrelationships, and context useful to both humans and machine agents

GLOSSARY

Lists

Lists are unordered *members* or *instances*, with or without gaps or duplicates, useful for bulk assignment purposes. Lists generally occur through a direct relation assignment (e.g., `rdf:Bag`)

M

Machine learning

The construction of algorithms that can learn from and make predictions on data by building a model from example inputs. A wide variety of techniques and algorithms ranging from *supervised* to *unsupervised* may be employed

Mapping

A considered correlation of *objects* in two different sources to one another, with the relation between the objects defined via a specific *property*. Linkage is a subset of possible mappings

Member

Used synonymously herein with *instance*

Metadata

Metadata are *annotations* and provide information about one or more aspects of the content at hand such as means of creation, purpose, when created or modified, author or provenance, where located, topic or subject matter, standards used, or other descriptive characteristics. In contrast to an *attribute*, which is an individual characteristic intrinsic to an *instance*, metadata is a description about that data

Metamodeling

Metamodeling is the analysis, construction, and development of the frames, rules, constraints, models, and theories applicable and useful for modeling a predefined class of problems

Microdata

Microdata is a proposed specification used to nest semantics within existing content on web pages. Microdata is an attempt to provide a simpler way of annotating HTML elements with machine-readable tags than the similar approaches of using *RDFa* or *microformats*

Microformats

A microformat (sometimes abbreviated `µF` or `uF`) is a piece of mark up that allows expression of semantics in an HTML (or XHTML) web page. Programs can extract meaning from a web page that is marked up with one or more microformats

N

Natural language processing

NLP is the process of a computer extracting meaningful information from natural language input and/or producing natural language output. NLP is one method for assigning structured data characterizations to text content. NLP applications include automatic summarization, coreference resolution, machine translation, named entity recognition, question answering, relationship extraction, topic segmentation and recognition, word segmentation, and word sense disambiguation

Named entity

See *entity*

Named entity recognition

See *entity recognition*; also called NER

Negation

Negation is a unary operation that produces a value of *true* when its operand is false and a value of *false* when its operand is true

Nodes

Nodes, in a graphical representation of a *knowledge graph*, are the *subjects* and *objects* in a “triple” *statement*; they are connected to one another via *relations* (or *edges* in a graphical representation)

O

OBIE

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. Ontology-based information extraction (OBIE) is the use of an ontology to inform a “tagger” or information extraction program when doing natural language processing. Input ontologies thus become the basis for generating metadata tags when tagging text or documents

A KNOWLEDGE REPRESENTATION PRACTIONARY

Object

An object is an *entity, event, class, concept* or *property* that can be referred to via an *identifier* of some sort; in KBpedia, every object has a URI *identifier*

Ontology

An ontology is a data model that represents a set of *concepts* or *instances* within a *domain* and the *relationships* between those concepts. Loosely defined, ontologies on the Web can have a broad range of formalism, or expressiveness or *reasoning* power

Ontology-driven application

Ontology-driven applications (or ODapps) are modular, generic software applications designed to operate in accordance with the specifications contained in one or more *ontologies*. The relationships and structure of the information driving these applications are based on the standard functions and roles of ontologies (namely as *domain ontologies*), as supplemented by UI and instruction sets and validations and rules

Open Semantic Framework

The open semantic framework, or OSF, is a combination of a layered architecture and an open-source, modular software stack. The stack combines many leading third-party software packages with open source *semantic technology* developments from Structured Dynamics

Open world assumption

OWA is a formal logic assumption that the truth-value of a *statement* is independent of whether or not it is known by any single observer or agent to be true. OWA limits the kinds of *inference* and deductions to those that are known to be true. OWA is useful when we represent knowledge within a system as we discover it, and where we cannot guarantee that we have discovered or will discover complete information, typical of knowledge. See contrast to the *closed world assumption*

OWL

The Web Ontology Language (OWL) is designed for defining and instantiating formal Web *ontologies*. An OWL ontology may include descriptions of *classes*, along with their related *properties* and *instances*. There are also a variety of OWL dialects

P

Particulars

Particulars are all *entities* and *events*; they are in Secondness

Pragmatic maxim

Pragmatic maxim is the understanding of a topic or object by an apprehension of all of the practical consequences potentially arising from it

Pragmatism

Pragmatism, what Peirce came to term pragmatism because of what he felt was a misappropriation of his idea, is the consideration and weighing of available alternatives or explanations in order to pick the most likely ones with a return

Precision

The fraction of retrieved documents that are relevant to the query. It is measured as *true positives* divided by all measured positives (true and false). High precision indicates a high percentage of true positives in relation to all positive results

Predicate

See *Property*

Preferred label

Preferred Label (or *prefLabels* or *title*) is the readable string (name) for each *object* in KBpedia. The labels are provided as a readable convenience; the actual definition of the object comes from the totality of its description, *prefLabel*, *altLabels*, and connections (placement) within the *knowledge graph*. Labels of all kinds are *representations* and reside in Thirdness

Precision

Precision, or its verbs *prescind* or *prescinded from*, is the process of comparing two items and seeing if either may exist independent of the other. If so, we say the independent one is *prescinded from* the dependent one; it is one way to determine a subsumption relationship.

Property

Property is an official term in RDF and OWL that includes what we term *attributes*, *external relations*, and *representations*; we try to use the term sparingly, generally when only referencing those items in relation to RDF or OWL

Punning

In computer science, punning refers to a programming technique that subverts or circumvents the type system of a programming language, by allowing a value of a certain type to be manipulated as a value of a different type. When used for *ontologies*,

GLOSSARY

it means to treat a thing as both a *class* and an *instance*, with the use depending on context

Q

Query

Query is a request for information from an agent using a suitable knowledge representation

R

Range (property)

Range (property) declares the *classes* or data types from which the *object* data or *types* of an *assertion* must be drawn

RDF

Resource Description Framework (RDF) is a data model with a syntax that allows *statements* about *resources* in the form of *subject-predicate-object* expressions, called *triples* in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object

RDFa

RDFa uses attributes from meta and link elements and generalizes them so that they are usable on all elements allowing annotation markup with semantics. RDFa 1.1 is a W3C Recommendation that removes prior dependence on the XML namespace and expands HTML5 and SVG support, among other changes

RDF Schema

RDFS or RDF Schema is an extensible *knowledge representation* language, providing basic elements for the description of *ontologies*, otherwise called RDF vocabularies, intended to structure RDF resources

Reasoner

A semantic reasoner, reasoning engine, rules engine, or simply a reasoner, is a piece of software able to infer logical consequences from a set of asserted facts or *axioms*. The notion of a semantic reasoner generalizes that of an inference engine, by providing a richer set of mechanisms

Reasoning

Reasoning is one of many logical tests using *inference* rules as commonly specified by means of an ontology language, and often a description language. Many reasoners use first-order predicate logic to perform reasoning; inference commonly proceeds by forward chaining or backward chaining

Recall

The fraction of the documents that are relevant to the query that is successfully retrieved. It is measured as *true positives* divided by all potential positives that could be returned from the corpus. High recall indicates a high yield in obtaining relevant results

Record

As used herein, a shorthand reference to an *instance record*

Reference concept

Reference concepts (or *RefConcepts* or *RCs*), the *base concepts* in KBpedia, are any of the noun *objects* within KBpedia and abbreviated as RC. An RC may be either an *entity*, *entity type*, *attribute*, *attribute type*, *relation*, *relation type*, *topic* or abstract *concept*. RCs are a distinct subset of the more broadly understood ‘concept’ such as used in the SKOS RDFS controlled vocabulary 7 or formal concept analysis or the very general or abstract concepts common to some ontologies. The KBpedia *knowledge graph* is a coherently organized structure of the nearly 60,000 RCs in KBpedia. All RCs are OWL *classes*

Referent

The *object* referred to by an *identifier*

Reflexivity

Reflexivity is when every element of X is related to itself, every *class* is its own subclass, such as every person is a person

Reinforcement learning

Reinforcement learning (RL) is a method of *machine learning* wherein actions are evaluated, most often as a Markov decision process, in accordance with stated performance objectives via a reward function to help converge to those desired goals.

Relation

Relations are the way we describe connections between two or more things; *attributes*, *external relations*, and *representations* are all *instances* of the relations *class*

Relation type

An aggregation (or *class*) of multiple *relations* that have similar characteristics amongst themselves. As with other types,

A KNOWLEDGE REPRESENTATION PRACTICIONARY

shared characteristics are subsumed over some *essence(s)* that give the *type* its unique character

Representations

Representations are signs, and the means by which we point to, draw or direct attention to, or designate, denote or describe a particular *object*, *entity*, *event*, *type* or *general*. A representational relationship has the form of *re:A*. Representations can be designative of the subject, that is, be icons or symbols (including labels, definitions, and descriptions); indexes that more-or-less help situate or provide a traceable reference to the subject; or associations, resemblances and likelihoods in relation to the subject, more often of an indeterminate character

Root

Root is the name for the top-level node in a *taxonomy*, *knowledge graph*, *ontology* or *typology*

RSS

RSS (an acronym for Really Simple Syndication) is a family of web feed formats used to publish frequently updated digital content, such as blogs, news feeds or podcasts

S

Satisfies

Satisfies means that all statements have an interpretation that can be shown to be true

schema.org

Schema.org is an initiative launched by major search engines to create and support a common set of schema for structured data markup on web pages. schema.org provided a starter set of schema and extension mechanisms for adding to them. schema.org supports markup in microdata, microformat and RDFa formats

Semantic enterprise

An organization that uses *semantic technologies* and the languages and standards of the *semantic Web*, including *RDF*, *RDFS*, *OWL*, *SPARQL*, and others to integrate existing information assets, using the best practices of *linked data* and the *open world assumption*, and targeting knowledge management applications

Semantic technology

Semantic technologies combine software and semantic specifications to encode meanings separate from data and content files and separate from application code. This approach enables machines as well as people to understand, share and reason with data and specifications separately. Semantic technologies provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes

Semantic Web

The Semantic Web is a collaborative movement led by the World Wide Web Consortium (W3C) that promotes common formats for data on the World Wide Web. By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web of unstructured documents into a "web of data." It builds on the W3C's Resource Description Framework (RDF)

Semset

Semsets (or *synsets* or *alternative labels* or *altLabels*) are collections of alternate labels and terms to describe a *concept* or *entity* or *event*. These semset alternatives include true synonyms, but may also be more expansive and include abbreviations, acronyms, aliases, argot, buzzwords, cognomens, derogatives, diminutives, epithets, hypocorisms, idioms, jargon, lingo, metonyms, misspellings, nicknames, non-standard terms (see Twitter), pejoratives, pen names, pseudonyms, redirects, slang, sobriquets or stage names; in short, any term or phrase that can be a reference to a given *instance* or *class*

Sequences

Sequences are ordered *members* or *instances*, with or without gaps or duplicates, useful for bulk assignment purposes. Sequences generally occur through a direct relation assignment (e.g., *rdf:Seq*)

SKOS

SKOS or Simple Knowledge Organisation System is a family of formal languages designed for representation of thesauri, classification schemes, *taxonomies*, subject-heading systems, or any other type of structured controlled vocabulary; it is built upon *RDF* and *RDFS*

Sound

An evaluative criteria in logic systems where all provable statements are true in all models

SPARQL

SPARQL (pronounced "sparkle") is an *RDF* query language; its name is a recursive acronym that stands for SPARQL Protocol and RDF Query Language

GLOSSARY

Statement

A statement is the standard and most basic expression in *RDF* and *OWL*. A statement is comprised of a “triple” (*subject - property - object/value*)

Subject

A subject is either a *concept*, *entity*, *event* or *property* (also collectively known as a “resource” in *RDF*), and is the item currently under consideration or focus; it is equivalent to the linguistic *subject*

Subject extraction

Subject extraction is an automatic process for retrieving and selecting subject names from existing *knowledge bases* or data sets. Extraction methods involve parsing and tokenization, and then generally the application of one or more information extraction techniques or algorithms

SuperType

SuperTypes (also Super Types) are a collection of (mostly) similar *reference concepts*. Most of the SuperType classes have been designed to be (mostly) disjoint from the other SuperType classes. SuperTypes are synonymous with the *typologies* used in *KBpedia*. SuperTypes (typologies) provide a higher-level of clustering and organization of *base concepts* for use in user interfaces and for reasoning purposes

Supervised learning

A *machine learning* task of inferring a function from labeled training data, which optimally consists of positive and negative *training sets*. The supervised learning algorithm analyzes the training data and produces an inferred function to determine the class labels for unseen instances correctly

Symmetric

Symmetric is when A relates to B exactly if it relates B with A

Synechism

Synechism is a philosophical doctrine that space, time, and law are continuous and form an essential Thirdness of reality in contrast to existing things and possibilities

T

Tag

A tag is a keyword or term associated with or assigned to a piece of information (*e.g.*, a picture, article, or video clip), thus describing the item and enabling keyword-based classification of information. Tags are usually chosen informally by either the creator or consumer of the item

TBox

A TBox (for terminological knowledge, the basis for T in TBox) is a “terminological component”; that is, a conceptualization associated with a set of *facts*. TBox *statements* describe a conceptualization, a set of *concepts* and *properties* for these concepts. The TBox is sufficient to describe an *ontology*. Best practice often suggests keeping a split between instance records, the *Abox*, and the TBox schema

Taxonomy

In the context of knowledge systems, taxonomy is the hierarchical classification of *entities* of interest of an enterprise, organization or administration, used to classify documents, digital assets and other information. Taxonomies can cover virtually any type of physical or conceptual entities (products, processes, knowledge fields, human groups, etc.) at any level of granularity

Topic

The topic (or theme) is the part of the proposition that is being talked about (predicated). In *topic maps*, the topic may represent any *concept*, from people, countries, and organizations to software modules, individual files, and events. Topics are in Thirdness

Topic Map

Topic maps are an ISO standard for the representation and interchange of knowledge. A topic map represents information using *topics*, associations (similar to a predicate relationship), and occurrences (which represent relationships between topics and information resources relevant to them), quite similar in concept to the *RDF triple*

Training set

A set of data used to discover potentially predictive relationships. In *supervised learning*, a positive training set provides data that meet the training objectives; a negative training set fails to meet the objectives

Transitivity

Transitivity is when item A is related to item B, and item B is related to item C, then A is also related to A; this is the critical *property* for establishing inheritance chains

A KNOWLEDGE REPRESENTATION PRACTIONARY

Triple

A basic *statement* in the *RDF* language, which is comprised of a *subject – property – object* construct, with the subject and property (and object optionally) referenced by *URIs*

True negative

A correct result, but one which fails (is negative) to meet the test objective. It is abbreviated TN

True positive

A correct result, and one which succeeds (is positive) to meet the test objective. It is abbreviated TP

Tychism

A philosophical doctrine that absolute chance is real and operative in the world; it is the source of irregularity and variety and the underlying force of evolution

Type

Types are the hierarchical classification of natural kinds of *instances* as determined by shared *attributes* (though not necessarily the same values for those attributes) and some common *essence*, which is the defining determinant of the type. All types may have hierarchy. Types are in Thirdness

Typology

Typologies are a natural organization of natural *classes* or *types*, with the most general types at the top of the hierarchy, the more specific at the bottom. All types contained in a typology are children (sub-classes) of the *root* type that is the basis for the character of the typology. Typologies provide a modular basis for expanding or collapsing the coverage of similar *instances* within a *knowledge base*. Typologies are central architectural components of KBpedia

U

UMBEL

UMBEL, short for Upper Mapping and Binding Exchange Layer, is an *upper ontology* of about 35,000 reference concepts, designed to provide common mapping points for relating different ontologies or schema to one another, and a vocabulary for aiding that ontology mapping, including expressions of likelihood relationships distinct from exact identity or equivalence. This vocabulary is also designed for interoperable *domain ontologies*

Unsupervised learning

A form of *machine learning*, this approach attempts to find meaningful, hidden patterns in unlabeled data

Upper ontology

An upper ontology (also known as a top-level ontology or foundation ontology) is an *ontology* that describes very general concepts that are the same across all knowledge domains. An important function of an upper ontology is to support very broad semantic interoperability between a large number of ontologies that are accessible ranking “under” this upper ontology

URI

A uniform resource identifier is a Web-accessible address (string) for a specific piece of data; it is a more generalized form of a URL, which points to a page or resource location

V

Value

Value is a string, literal or data value that pairs a numerical quantity, or quality or utility of a *subject* in relation to the meaning of its associated *attribute*, separate from the subject (but in association with it), these are known as key-value pairs; a value has no meaning or context absent its paired *attribute*

Vocabulary

A vocabulary, in the sense of knowledge systems or *ontologies*, is a *controlled vocabulary*. Vocabularies provide a way to organize knowledge for subsequent retrieval. They are used in formal declarations, subject indexing schemes, subject headings, thesauri, *taxonomies* and other forms of knowledge organization systems

W

Web-oriented architecture

Web-oriented architecture, WOA, is a subset of the *service-oriented architectural* (SOA) style, wherein discrete functions are packaged into modular and shareable elements (‘services’) that are made available in a distributed and loosely coupled manner. WOA uses the representational state transfer (REST) style, geared to the HTTP model

Wikidata

This is a crowdsourced, open *knowledge base* of (currently) about 18 million structured *entity records*. Each record consists of at-

GLOSSARY

tributes and values with robust cross-links to multiple languages. Wikidata is a key entities source

Wikipedia

Wikipedia is a crowdsourced, free-access and free-content *knowledge base* of human knowledge. It has nearly 5 million articles in its English version. Across all Wikipedias, there are nearly 35 million articles in 288 different language versions

WordNet

WordNet is a lexical database for English that groups words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. WordNet provides a combination of dictionary and thesaurus to support text analysis and artificial intelligence applications

Y

YAGO

“Yet another great ontology” is a WordNet structure placed on top of Wikipedia

Section Notes

1. Bergman, M. K., “Acronyms & Glossary,” *AI3::Adaptive Information* Available: <http://www.mkbergman.com/acronyms-glossary/>.
2. “KBpedia - KKO Structural Components,” *Cognonto* Available: <http://kbpedia.com/docs/kko-structural-components/>.
3. Bergman, M. K., “Glossary of Semantic Technology Terms,” *AI3::Adaptive Information* Available: <http://www.mkbergman.com/1017/glossary-of-semantic-technology-terms/>.
4. Bergman, M. K., “A Data Model of Web Data Models: Part I,” *AI3::Adaptive Information*, Oct. 2007.
5. Bergman, M. K., “Conceptual and Practical Distinctions in the Attributes Ontology,” *AI3::Adaptive Information*, Mar. 2015.
6. Brent, J., *Charles Sanders Peirce: A Life*, Indiana University Press, Bloomington, 1998.
7. *SKOS Simple Knowledge Organization System Reference: W3C Recommendation*, World Wide Web Consortium, 2009.