The LexiCon: structuring semantics

Author:
Kees Woestenenk

Institution:
STABU foundation

E-mail:
kwoestenenk@stabu.nl

Abstract:
The ISO/PAS 12006-3:2000 "Framework for object oriented information exchange" provides a meta-model for structuring data relevant to the construction industry. The LexiCon, developed by STABU, is an implementation of that model. The LexiCon consists of a set of concepts, grouped into categories according to 12006-3. Each concept should be defined by a set of other concepts and is placed in a specialization hierarchy with other concepts of the same category. Ultimately, a set of basic concepts exists, from which all the other concepts are being derived, but which can only be described by a verbal definition.

The concepts in the LexiCon represent the language of the construction industry. The terms of this language are extracted from numerous documents on several levels of formality, such as standards, building codes, specification systems, product information systems, technical dictionaries and classification systems. From these terms found in external documents concepts are then defined in the LexiCon with formal definitions. Terms from external documents can be in any human language, thus the LexiCon will be multi-lingual.

Putting the concepts in the LexiCon in a specialization hierarchy has been a process of trial and error, so far. In principal, the number of possible hierarchies is endless, leading to also endless discussions. We are now trying to use a more systematic approach, based on the idea that, when a concept is defined by a set of other concepts, there will be a subset of these defining concepts that will dedicate where the described concept will fit into a hierarchy. A consistent hierarchy will be obtained when the subsets of defining concepts are subsets of defining concepts on a lower level of that hierarchy. It will then be possible to reorder the hierarchy into another hierarchy, by using different sets and subsets of defining concepts.

Keywords:
Concepts, Semantic model, Object library

Context for data

The LexiCon is a structure for the storage of data in such a way that – within a certain context – the meaning of these data is assured. The LexiCon therefore can be regarded as a semantic system, defining the context for data explicitly and defining contexts within broader contexts.

Example: The fire resistance of a construction is 60 minutes. The statement '60 minutes fire resistance' can only be interpreted correctly within the context defining the meaning of 'fire resistance', how to measure '60 minutes' and what kind of constructions are dealt with.

The context of the LexiCon as a whole is limited to the built environment, from the point of view of the construction sector that erects this built environment, maintains and ultimately demolishes it, produces and supplies materials and products, as well as from the point of view of the users of this built environment. In the LexiCon, everything part of the built environment, including materials and products, is grouped under the heading Subject, everything dealing with the production process, the maintenance, demolishment and use, is grouped under the heading Activity. The LexiCon only contains generalizations of Subjects and Activities, therefore, Subjects and Activities should be interpreted as Subject types and Activity types.

Example: A Lighting fixture is a Subject. A Lighting fixture installed in a certain space is part of the built environment. A Lighting fixture is a product too, however, produced and marketed by a manufacturer of

Lighting fixtures. Installing and Connecting the Lighting fixture are Activities, changing the Lamp of a Lighting fixture is an Activity as well. The Lamp is a Subject in itself. Lighting fixture and Lamp are indicators referring to a type, not to a single instance. Likewise, Installing and Connecting refer to types of Activities.

Example: Habitation is a User activity, facilitated by the Subject Dwelling as part of the built environment.

Description and definition

Lexical description

The LexiCon is meant for the lexical representation of data. Hence a lexical description of the structure, possibly illustrated with images. Lexical descriptions are language dependent. The LexiCon is basically multi-lingual, guaranteeing a consistent meaning of what is described over the languages.

Example: 'Vloer' is the Dutch term for a Subject indicating a part of the built environment on which you can walk. The English term for this Subject is Floor. However, in the English language, Floor might also mean Story. The LexiCon must resolve this ambiguity, making it explicit which interpretation to use when this term pops up.

In the LexiCon, terms used as indicators are grouped under the heading Name. A name may consist of several words. Names in natural languages, such as Dutch and English, often can have multiple interpretations (homonyms). On the other hand, natural languages have synonyms as well. In the Lexicon homonyms and synonyms are allowed, as long as the Lexicon resolves the ambiguity, by making it explicit which meaning is meant within a given context.

Example: The name Door may indicate a Door leaf as well as a Door set, or might even apply to the abstract concept of an Opening in a Wall that can be closed. The LexiCon allows the use of this name for all of these meanings, as long as it is made clear that this name points to different concepts, of which one should be selected. If these different concepts have different names, such as Door leaf, Door set and Door opening, then the name Door may be used as a synonym.

In the LexiCon, abbreviations of names and lexical symbols that are used as indicators are grouped under the heading Short name. Short names are also language dependent, synonyms and homonyms are allowed.

Example: In the English language the short name for the name metre is m.

Concepts in the LexiCon can be further explained with one or more descriptions. A description tries to clarify what is really meant with the concept. Descriptions are of course language dependent. Descriptions should not be contradictory to each other, but may differ in the level of preciseness and completeness. Descriptions are similar to definitions, but do not need to be as strict as a formal definition.

Example: A Dutch description of Door: "Een te openen en te sluiten toegangsconstructie". A similar English description: "Construction for closing an opening, intended primarily for access".

The LexiCon allows for images to be used as illustrations. There is no restriction on the number of Images, and the same Image might be used with different concepts, when applicable.

Formal description of Subjects and Activities

Besides descriptive identifiers by means of names, short names, descriptions and images for clarification, Subjects and Activities are formally defined in the LexiCon by associated properties and – as far as applicable – by typical composition.

A Property is a quality or quantity that can be associated to a Subject or Activity as a predicate. A Property forms the direct context for data. The actual data content is named a Value in the LexiCon. A Value is related to a Measure that may have an associated Unit. A Value can be numeric, or belong to a list (enumeration), or it can be a description. A Value might be single, or it might consist of a set of values. Providing Values is often not needed in the LexiCon, it suffices to only provide a Property without a Value. This way a Property is generalized into a Property type, which can act as a context for Values to be provided on instantiation. When a Value has to be selected from a list of Values then the LexiCon will contain the whole list of possible Values, however. Associating generalized Property types to Subjects and

Activities, make these Subjects and Activities Subject types and Activity types.

Example: Length is a quantity that has Linear dimension measure as a Measure. A possible Unit for this Measure is metre, which requires a numeric Value. As long as the Value is not provided Length will be a Property type, as soon as a Value is provided this type will be instantiated. Hence a length of 2.15 m is an instance of Property type Length. Alternative to metre the Unit inch might be used, which is another Linear measure. Values using metres or inches as a Unit are transferable to each other.

Example: Fire resistance is a quality that uses a Fire resistance measure as its Measure. The Dutch standard NEN 6069 describes the method of determining the fire resistance related to separation function, using minute as the Unit. The Building Code requires certain Values for Fire resistance, dependent on the kind of construction, referring to the Dutch standard NEN 6068, which in turn refers to NEN 6069. Other countries use other methods to determine the Fire resistance, or provide Fire resistance classes that can be referred to. With qualities it is not always certain that Values depending on determining methods are transferable to each other.

Example: The Property Moving characteristic contains a list of the possible ways a Door or Window can move, such as Side hung, open in, left closing, Side hung, open in, right closing, Side hung, open out, left closing, Horizontal sliding, Tilting sliding, etc. Each of these list items represents a possible Value, of which one and only one has to be chosen.

The LexiCon allows multiple associations of a single Property type to a Subject or Activity. In this case the Name of the Property is extended with a distinctive description, which is called an Extension in the LexiCon.

Example: A Door set has three times the Property Breadth, namely the Coordinating width, the Door set width and the Joint clearance.

The typical Composition describes the parts of which a Subject or an Activity is assembled. Every part of a Subject is a Subject in its own right, likewise, every part of an activity is an Activity again. The list of parts describing a composition does not need to be exhaustive, nor is it needed that every described part is always present. Composition could be used, however, to distinct one Subject or Activity from another.

Example: A Fluorescent lighting fixture is typically composed of a Casing, a Lamp fitting, one or more Tube lights, a Choke, a Reflector and a Light screen. An Incandescent lighting fixture typically consists of a casing, a Lamp fitting, one or more Light bulbs, a Reflector and a Light screen. Both are Lighting fixtures that can be distinguished by their composition. Although the Reflector and the Light screen are typical components, it is not necessary that these are always present. In addition, both Lighting fixtures might have extra components, such as an Emergency lighting fixture or a Protection basket.

The Lexicon makes a distinction between Subjects and Activities of which a typical Composition cannot be given on the one hand, and on the other Subjects and Activities for which this would be possible. The first case is called a Functional concept, the second one a Solution concept. A Functional concept is defined by the Function(s) that have to be fulfilled, or, in other words which Role(s) has (have) to be performed and which performance criteria have to be met. A Functional concept is characterized by its extrinsic Properties (directed outwards). This excludes the description of the (inner) structure. A Solution concept on the other hand is a break down of the Functional concept, which includes the description of the structure. A Solution concept performs like the Functional concept, but uses for this the set of Components, acting together to provide the functionality.

Example: A Lighting fixture is a Functional concept. A Fluorescent lighting fixture and an Incandescent lighting fixture are both Solution concepts, each having their own technical way to act as a Lighting fixture as a Functional concept.

In the Lexicon components that are mentioned as part of a Composition (hence describing a Solution concept), are themselves Functional concepts, each replaceable by one of the possible Solution concepts on further break down.

Example: An Electrical lighting installation is a Solution concept, composed of parts such as Lighting fixtures, Electrical power supply en Switchgear. Of these, the Electrical lighting fixture is a Functional concept that can be substituted by one of the possible Solution concepts, as described above.

Specialization hierarchy

As mentioned earlier the descriptions of the Built environment and related Processes are grouped under the headings Subject and Activity. Subject and Activity are very abstract concepts. They form the top of a hierarchical structure – and are therefore the ultimate summaries of all the nodes within that structure. The more downwards, the more concrete this structure becomes. This is called a Specialization hierarchy: nodes on the lower levels are specialized versions of the nodes on the higher levels. The hierarchy evolves from abstract (global, summarized) to concrete (specific, detailed).

Example: Built environment is an abstract concept characterized more or less solely by the notion that is conceived by construction activities, contrarily to Natural environment. Environment is even more abstract, combining Built and Natural environment in a single concept. Space is another abstract concept, which exists as the counterpart of Matter. One could say that Environment consists of Space and Matter, Environment therefore is a Composition of Space and Matter. All concepts derived from Environment are then Compositions of Space and Matter. A Cave is a specialization of a Natural environment, and a Tunnel a specialization of Built environment. Both Cave and Tunnel are examples of Spaces inside Matter.

Example: Construction activity and User activity are both abstract concepts, but still specializations of the even more abstract concept Activity. Brick laying, Assembling, Demolishing and Painting are specializations of Construction activities; Living, Sleeping and Working are specializations of User activities. Construction activities always deliver a Result: the Built environment (in the case of Demolishing the Result is negative). User activities require certain environmental conditions.

The Specialization hierarchy shows on which level of abstraction an item is positioned. The higher level then provides a more general concept, whereas the lower levels lists the more specialized concepts.

Once a Property is associated with an item in the Specialization hierarchy, all the items on the lower levels of that hierarchy will have that Property as well. This is called Inheritance of properties.

In the LexiCon Properties, Measures and Units are also structured in a Specialization hierarchy.

Collections

Previously, two kinds of Collections have already been discussed: Composition and Specialization. A Composition is a Collection of Subjects or Activities, which together form a new Subject or Activity with its own characteristics. Specialization is a Collection of similar concepts (hyponyms), which are summarized in a single – more general – concept. Apart from these two types of Collections the LexiCon provides the possibility of Collections that are neither Compositions nor Specializations. These are Subject collections, Activity collections and Property collections. The concepts within these Collections may also belong to other Collections.

Example: Ironmongery is a Collection of Subjects, which has no type relationship, but is grouped together because they belong to a specific suppliers assortment.

Example: Maintenance activities is a group of non-related activities, grouped by their association with the life phase of Subjects that have to be maintained.

Example: Fire safety can be a Collection of Properties, grouped by their involvement in a certain function.

Collections are grouped in a Specialization hierarchy in the LexiCon.

References

External documents may contain data relevant to the contents of the LexiCon. The LexiCon provides links to these external documents through References. There are two main groups of Reference documents: documents containing descriptions or definitions of terms, and documents that group terms based on characteristics.

Specialization hierarchy for Subjects

There are many possibilities setting up a specialization hierarchy for Subjects. Arguments can be provided in favour of each of these possibilities, and a discussion about the preferred hierarchy could be endless.

Furthermore, the hierarchy is only important inside the LexiCon, because each Subject, on any level will always present the full set of associations to the outside world, implying that associations defined on the higher level are included.

Example: Subject (the highest level) has an association with the Property collection Lifecycle related properties (consisting of Properties like Endurance and Operational durability). This implies that all types derived from Subjects will be associated with this Property collection.

Rules for setting up the hierarchy are formulated as guidelines in the following.

Guideline: A Subject on a lower level of the hierarchy is more specialized than Subjects on the higher levels.

Example: A Wastewater pump is a specialized Pump, hence the Wastewater pump will be on the lower (more specialized) level of the hierarchy. This also follows from the name: a Wastewater pump is a Pump, not every Pump however is a Wastewater pump.

Example: Subject is on the highest level of the Specialization hierarchy of Subjects en thus has to be regarded as the most generic of all Subjects. All types derived from Subject can be indicated with the term Subject, hence both Pump and Wastewater pump can be indicated with the Name Subject.

Guideline: A Subject described with regard to its behaviour or functionality (indicated as Functional concept) is on a higher level in the hierarchy than a similar Subject described from its internal structure as an assembly (indicated as Solution concept). Therefore, when setting up the hierarchy, Functional concepts have preference above Solution concepts, and Solution concepts are interpreted as specializations of their functional counterparts.

Example: A Pump (or Pumping system) is a specialization of a Water transport system. A Water transport system is characterized by its Water transporting function. A Pump realizes this function by creating a difference of pressure in a Piping system and is therefore a Solution concept for the Water transport system, the Pump fulfils the Water transport function. An alternative for a Pump is a Natural drop Piping system that uses gravity to transport the fluid.

Example: A Plunger pump is a Solution concept for a Pump (system), using a Piston to create the difference of pressure. A Piston is a typical component of a Plunger pump, which name indicates how it is constructed. An alternative for a Plunger pump would be a Centrifugal pump that uses an Impeller.

Guideline: Subjects derived from a single Subject are differentiated by a single characteristic.

Example: A Plunger pump and a Centrifugal pump are differentiated by their method of creating a pressure difference in a Piping system. Based on this characteristic they can act as Solution concepts for a Pump.

Example: A Wastewater pump, a Clean water pump and a Faeces pump are distinguished by the type of fluid that is transported, and therefore may be regarded as distinct specializations for a Pump.

Example: A Wastewater pump and a Centrifugal pump cannot be distinct alternatives derived from Pump, because their difference is based on different characteristics. A Centrifugal pump could be a Solution concept for a Wastewater pump, and a Wastewater pump might be of the type Centrifugal pump.

Guideline: When choosing a characteristic for specialization preference is given to a functional characteristic.

Example: A Water transport system could be specialized based on the type of fluid (functional characteristic) or on the transport method (solution characteristic). Based on the Guideline and the earlier examples, Water transport system will first be specialized in Wastewater transport system, Clean water transport system and Faeces transport system. Each of these systems then could be further specialized based on the transport method (solution): Pumping system and Natural drop system.

Guideline: A specialization level will be added only when the specialized Subjects have an extra characteristic.

Example: The specialization of Water transport system into Waste-water transport system, Clean water transport system and Faeces transport system use a single Property: Type of water. If the choice of water type only influences the size of the free passage opening there is still no extra characteristic involved,

because Free passage opening will be an associated Property in all cases. One could argue that the relevant difference between these systems only lies in the quantity and the size of the particles in the fluid, which can be described with the Properties Particle percentage and Particle size. These Properties are only relevant for contaminated water, which allows the specialization to be limited to Wastewater transport system. Clean water transport system would then only be a specialized type if it would be needed against Wastewater transport system to balance the hierarchy.

The specialization hierarchy ultimately resulting from the above examples will look like the following:

- Water transport system
    - Clean water transport system
        - Natural drop clean water transport system
        - Clean water pumping system
            - Plunger pump
            - Centrifugal pump
    - Waste-water transport system
        - Natural drop waste-water transport system
        - Waste-water pumping system
            - Waste-water plunger pump
            - Waste-water centrifugal pump

Following is an example starting at the top of the Subject hierarchy.

- Subject
    - Space
    - Matter
        - Shapeless matter
            - Gas
            - Liquid
            - Solid
        - Shaped matter
            - Shaped matter without cohesion
            - Environment (Components: Space + Matter)
                - Natural environment
                - Built environment

This hierarchy follows from the example under Specialization hierarchy in the first chapter. Stated there was that the LexiCon aims at Subjects from the Built environment. Built environment can be generalized into Environment, and the counterpart of Built environment is Natural environment. Environment can be described as a Collection of Space and Matter, so Space and Matter both have to be Subjects, otherwise they cannot be part of a Collection.

Space and Matter cannot be generalized without loss of characteristics, and therefore follow immediately the level of Subject. Matter and Space are each other's counterparts and therefore can form a specialization level. Because of their complementary nature, other derived Subjects may not be added to this level. Space exists within Matter, and Matter occupies Space. People are Matter, deploying their activities in well-equipped Spaces.

Matter can be specialized into Shaped matter and Unshaped matter. Unshaped matter is natural matter without any cohesion, which might fill a Space. Shaped matter has undergone some labour by natural or human activities.

Environment is Cohesive shaped matter with Space in between. Natural environment and Built environment are distinguished by their way of creation. Natural environment is created without human interference, and Built environment is created through human activity, aimed at shaping Matter.

Re-arranging the hierarchy

As we have seen in the above discussion, Subjects are known by their names, but are defined by their set of characteristics. For Activities a similar reasoning would apply. We also saw that for each level of specialization a single characteristic will be used to distinguish between sibling types. Revisiting the Water transport system, the characteristics used to build the hierarchy were:

- Function (water transport)
- Type of fluid (clean water, contaminated water)
- Transport method (forced, natural drop)
- (For the Pump) Pumping method (using a Piston or an Impeller).

The Pump inherits all of the characteristics of the higher level, the order of which should not be important. Hence, if we would swap the Type of fluid with the Transport method in the hierarchy, this would not make any difference for the Pump, but the hierarchy would look as follows:

- Water transport system
    - Natural drop water transport system
        - Natural drop clean water transport system
        - Natural drop waste-water transport system
    - Pumping system
        - Clean water pumping system
            - Clean water plunger pump
            - Clean water centrifugal pump
        - Waste-water pumping system
            - Waste-water plunger pump
            - Waste-water centrifugal pump

By selecting another set of characteristics we could easily re-arrange the hierarchy. We could even use characteristics not used so far.

Conclusion

Building a specialization hierarchy is not easy, for each new level a decision has to be made for the characteristic to be used as a discriminator. As long as each level of the hierarchy uses only one discriminator, the hierarchy will be valid, and can easily re-arranged into another hierarchy, without any consequences for the nodes in that hierarchy. The ultimate choice of the hierarchy will then be based on usefulness of a certain super type from a particular point of view.