
Development of an Ontology for the Document Management Systems for Construction

Alba Fuertes^{a,1}, Núria Forcada^a, Miquel Casals^a, Marta Gangolells^a and Xavier Roca^a

^aConstruction Engineering Department. Technical University of Catalonia, Spain.

Abstract. This paper describes the development of an ontology for the AEC/FM projects' documentation management that allows the classification of the documents along the lifecycle of AEC/FM projects. This ontology is aimed at reducing the interoperability and information exchange problems, inherent nowadays in AEC/FM projects, establishing a hierarchical structure of the different areas that conform the lifecycle of AEC/FM projects and an interrelationship system between them. Therefore, all the documentation created along a project could be classified in the different areas of the project lifecycle and related to them by this hierarchical structure. Moreover, metadata like identifier, creation date,... have been incorporated to documents in order to be completed and modified by the author to facilitate users' understanding. Therefore, this ontology is the first step to improve the Document Management Systems in AEC/FM projects and their interoperability limitations.

Keywords. Ontology, document management, interoperability, construction.

1 Introduction

The architecture, engineering, construction and facility management industry (AEC/FM) is fragmented due to the many stakeholders and phases involved in a construction project as well as to the complexity and diversity of their projects. This fact has led to a huge amount of organizational information formalized in unstructured documents.

Electronic Document Management Systems (EDMS) is an Information and Communication Technology (ICT) application that has started to be used in the construction industry as a tool to reduce some of the problems generated by fragmentation creating an environment that allows the centred stored of the documentation on a server. However, EDMS have also some limitations, most of them related to the interoperability, the ability for information to flow from one computer application to the next throughout the life cycle of a project, that

¹ Construction Engineering Department. Technical University of Catalonia, Spain. C/Colom,11, Building TR-5. 08222 Terrassa (Barcelona, Spain); Tel:+34 93 7398947; Fax: +34 7398101; Email: alba.fuertes@upc.edu; nuria.forcada@upc.edu

becomes difficult to achieve because of its dependence on the development and use of common information structures throughout all the different EDMS users. In order to solve this problem, different standards, aimed at establishing internationally recognised information classification structures, projects based on them and ontologies, another kind of system of representation of the concepts that a domain contains and the relations that exist between them, are being developed. As example ISO 12006 [3] and Industry Foundation Classes (IFC) [4], standards that establish a structure for the classification of objects in AEC/FM sector, Lexicon [5], a project based on ISO 12006 that provides general information such as building regulations, product information, cost data and quality assessments in a common and standardized language, and eCognos [2], a project that establishes and deploys a domain ontology for knowledge management in the AEC/FM sector, between others. From the study of these different initiatives it could be observed that most of them are object-oriented and define an information classification structure in different fields of the domain of AEC/FM projects, however none of them are oriented to the document management.

2 Research objective, scope and methodology

The objective of this research is to develop an Ontology for the Document Management Systems used in AEC/FM with the aim of reducing the interoperability and information exchange problems from the establishment of a hierarchical structure of the different areas that conform the lifecycle of AEC/FM projects and an interrelationship system between them. Thus, it is claimed:

- To make a literature review of existing EDMS, the interoperability limitation, possible solutions, and existing Information Classification Systems in Construction sector.
- To make a literature review of ontologies' definitions, characteristics, applications, methodology, editors and languages.
- To develop the Ontology for the DMS for Construction, including the definition of a Concept Model of the Documentation Flow from the literature review of existing theories about projects lifecycle.

3 Background

Classification systems that attempt to organize the knowledge base of national construction industries have a long history. The Swedish SfB system has been under development since 1945 and although it has long been superseded in Sweden itself it remains the basis for many existing knowledge classification systems such as CI/SfB [6], which is widely used in the UK.

The growing experience with classification systems and the development of ICTs has led to the development of the ISO 12006 series [3], which is aimed at establishing internationally recognized classification principles. Projects such as Uniclass [1], Lexicon [5], etc are examples of adaptations of ISO 12006.

Currently, the technical committee ISO/TC59/SC13 is working on the development of guidelines connected to IFD (International Framework for Dictionaries)-object libraries based on ISO 12006-3 with the aim to improve the quality of object libraries and to link and integrate them with other libraries.

Other research projects developed basic conceptual structures in the building construction domain, for example, the e-Construct taxonomy and e-Cognos ontology process-centered system for knowledge management in construction [2].

4 Development of the Ontology for DMS for Construction

4.1 Classification of information

Along this paper an ontology is understood as a system of representation of a domain or area of knowledge, where a method is applied in order to obtain a formal representation of the concepts that it contains and the relations that exist between them.

The development of the ontology implies the previous definition of the domain, the important terms that it may contain and their classification system. Consequently, the Concept Model of concepts that may be included in the ontology and their classification system is defined.

The concepts that the ontology may contain are determined by its domain (*All the documents generated along a construction project*) and its future usage (*as an ICS in a DMS for Construction*). Therefore, concepts such as documents and construction project lifecycle must be analyzed.

Documentation Flows created along a construction project depend on the actors involved in it as well as the roles they play in the organization chart defined by the kind of contractual arrangement. Consequently, actors and contractual arrangement are considered basic specifications for the selection of the documents to study.

In reference to the type of contractual arrangement, the Traditional procurement arrangement is the one selected bearing in mind that is the most used for ordinary projects of moderate size and complexity, in which the owner wishes to retain the maximum amount of control for the design-construction process. On the other hand, three different actors, pretending to include all the possible actors involved in a construction project, are defined: Client, Designer and Contractor.

Up to this point, all the documents generated in a construction project can be identified. In reference to the ontology proposed, the documents included in it are the ones commonly created along Spanish construction projects, as well as the terminology used.

All this documentation identified is created along the lifecycle of a project as a result of an activity. Therefore, from the literature review the most useful and understandable way to classify documents is to locate them along the construction project lifecycle and as a consequence of an activity. Different theories and models have been studied and the construction project lifecycle has been divided in Phases, the period in the duration of a construction project identified by the overall

character of the processes which occur within it, and Stages, defined as sub-processes of the project Phases. See Figure 1.

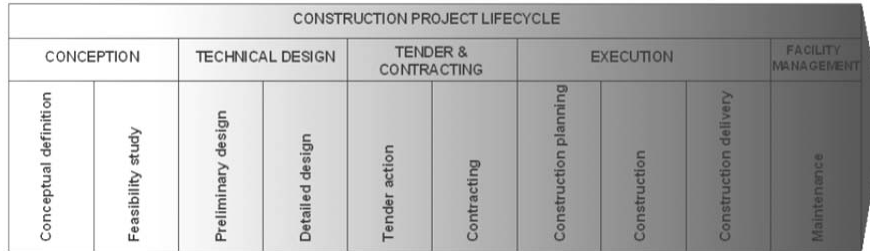


Figure 1. Phases and Stages of a Construction Project lifecycle

On the other hand, documents are a result of different activities and subactivities occurred along the project. By this way, Activities, defined as a working area of the project, and Subactivities, understood as the type of information of special importance in a project, are included in the Concept Model. Therefore, the Activities defined are: Advance, Changes, Contractings, Costs, Environment, Project, Quality and Safety & Health. And the Subactivities identified are: Communication, Documentation, Control and Planning.

As a result, with the Stages and Phases, that complete the lifecycle of the project, the documents listed and the Activities and Subactivities identified, all the relationships between them can be defined. These relationships will classify the documents along the lifecycle of the project bearing also in mind the activity and subactivity from where they come from. Therefore, the concepts and the relationships that constitute the Concept Model of the Documentation flow are already identified.

These concepts are classified in two kinds of metadata related to each document. Content-related metadata will be understood as the metadata that relates the documents with the phases, stages, activities and subactivities in order to situate them along the project lifecycle. This metadata is inherent in the document. On the other hand Content-properties metadata is related to what the document contains or is about, thus providing to users and applications useful hints to help document search and retrieval and to improve the reuse of documented information. This metadata is not compulsory and depends on the author needs. As example the name of the creator or the receiver, the type of format, the creation date, the version,... See Figure 2.

From the definition of this Concept Model of the Documentation flow, documents are classified as the result of the intersection of an Activity and a Subactivity that take place along one or more Stages that are part of a Phase. By this way documents are located along the lifecycle of the construction project basing on a three dimensional model. See Figure 3.

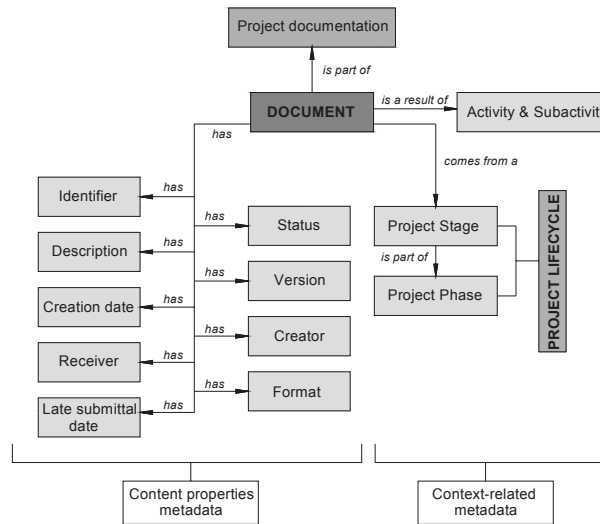


Figure 2. Concept Model of Documentation flow. Document metadata

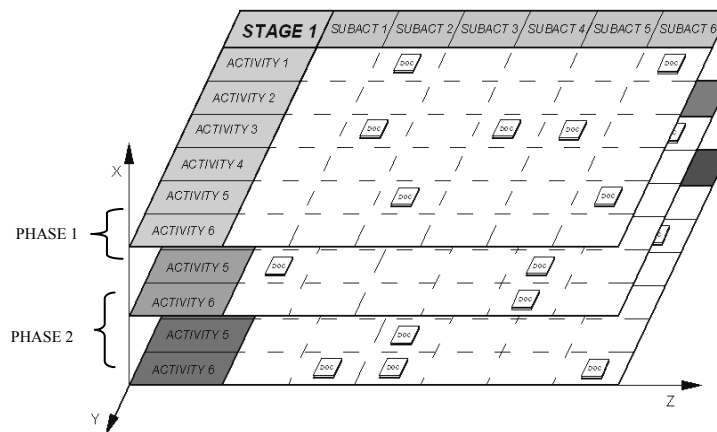


Figure 3. Representation of documents along the Construction Project lifecycle

4.2 Implementation

The implementation of the ontology has been carried out in Protégé editor, for its opened code, free access and simplicity, working with OWL DL language.

Reusing ontologies is an important point to bear in mind before the development of a new ontology. Taking into account that ontologies related to DMS haven't been found reusing has not been possible, however terminology to describe the phases, stages, authors, activities and subactivities in AEC/FM has been extracted from existing standards.

From the concepts identified in the Concept Model, classes, subclasses and properties are defined. Classes describe concepts in the domain and subclasses *kinds of* the already defined classes. In this proposed ontology the classes defined are shown in Figure 4. By the same way are represented the subclasses of each class.

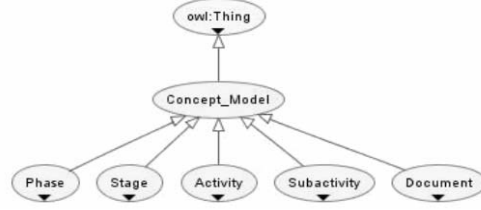


Figure 4. Classes of the Ontology for DMS for Construction exposed in Protégé

To classify the documents in the hierarchical structure created properties of classes and subclasses are defined. Three kinds of Protégé properties have been used:

- Object-properties to relate the class Document and their subclasses to the classes Stage, Activity and Subactivity. *is a result of* property relates each document with the intersection of an Activity, a Subactivity and a Stage, *is part of* relates each Stage with its particular Phase, and finally, *is composed by* relates each Phase with all its Stages;
- Datatype-properties to add the content-properties metadata already defined such as the creator, the receiver, the format,... to each document;
- Annotation-properties that, in this case, have been used to provide multi-lingual names for ontology elements (Spanish, Catalan and English) and some comments to make easier their understanding.

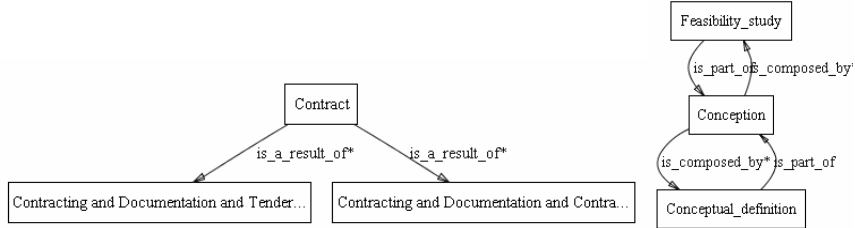


Figure 5. Examples of Object-properties

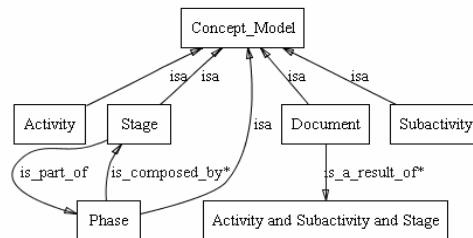


Figure 6. Concept Model and Object-properties

Documents are related to the Stages, Activities and Subactivities from which they come from by logic expressions such as *and* and *or*, that express intersection and union, respectively. Therefore, *and* expression is used to state that a document is a result of the intersection between an Activity, a Subactivity and a Stage (the Phase is already related with the Stage), and *or* expression is used to state that this document can be located in different locations (intersections) along the project lifecycle. See Figure 7.

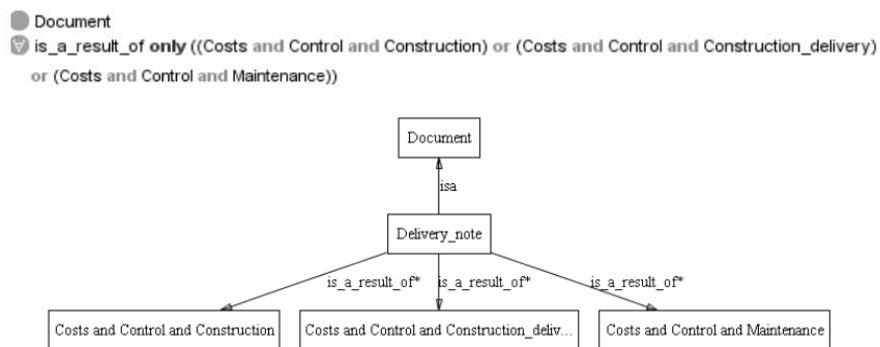


Figure 7. *and* and *or* expressions

Up to this point, the hierarchical structure that allows the documentation classification of a general construction project is defined. The following and the last step depends on the particular project and users that work with the DMS where the ontology is applied. This step consists on the definition of *individuals*, that are considered as particular classes. These individuals are created by the users, who add the information to the already defined properties of each document to facilitate its research and understanding. An example is shown in the Figure 8.

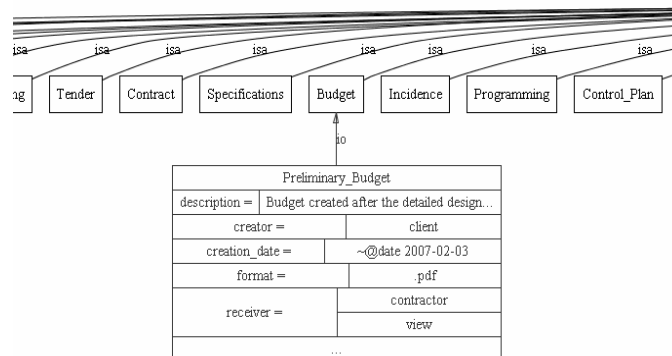


Figure 8. Example of an individual of the class *Budget*

5 Conclusions and future work

By the development of the Ontology applicable to a DMS for Construction all the documentation generated during a construction project is classified along its lifecycle as a result of the intersection between an Activity, Subactivity and Stage, improving, by this way, the existing indexing, search, retrieval and reuse of documents' limitations. This ontology provides a context-related metadata, that locate the documents along the project lifecycle, and a content-properties metadata, that add information to the document to facilitate its research and understanding.

The future work of this research would be the application of the developed Ontology as the internal information classification system of a tool such as BSC, a web page,... allowing users' interaction and the exchange of information by the web. Moreover, the creation of *Actors*, *Type of contractual arrangement* and *Related documents* classes would improve the classification of the documents and facilitate users' research. Finally, a more ambitious step would be to consider documents as a group of information and not as an entity.

6 References

- [1] Construction Industry Project Information Committee (CIPIC). Available online at: <<http://www.productioninformation.org/>> Accessed on Jan. 25th 2007.
- [2] Ei-Diraby TE, Lima C and Feis B. Domain taxonomy for construction concepts: Toward a formal ontology for construction knowledge. *Journal of Computing in Civil Engineering*, Vol. 19, Issue 4, 2005, pp. 394-406.
- [3] International Organization for Standardization ISO 12006-22. International Organization for Standardisation. Building construction - Organization of information about construction works. Part 2: Framework for classification of information, 2001.
- [4] Liebich T and Wix J Eds. IFC Technical Guide. International Alliance for Interoperability, October 27, 2000.
- [5] Woestenek K. From lexicon to XTD. In Z. Turk & R. Scherer (eds), *European Conference on Product and Process Modelling in the Building and Related Industries; Proc. eWork and eBusiness in Architecture, Engineering and Construction*, Portoroz, Slovenia, 9-11 September, 2002. The Netherlands: Balkema
- [6] Wright T. *Classifying Building Information: A Historical Perspective*. Technical report, Construction Information Systems Australia Pty Ltd., 1998.