Retrieval of Project Knowledge from Heterogeneous AEC Documents

Sabine Reul, Raimar J. Scherer [1]

Abstract

A knowledge system for the retrieval of project knowledge from civil-engineering documents will be described from different views. The views demonstrate the research process, the knowledge acquisition process and the recovery structure of the knowledge system. The described knowledge system should provide the possibility to reuse construction knowledge locked in documents. The documents will be classified using text analysing methods combined with clustering techniques. A lexicon which includes important parts of knowledge like construction key words, synonyms etc. for the specific domain of the building process will be combined with a document class structure and project class structure. During the process of general text analyses the lexicon will be completed with the frequency of words structured in different clusters of the building process. The knowledge system built up, will be a long-term storage of project experiences for the building industry.

Introduction

Although recorded in civil-engineering documents (such as construction diaries, order forms, delivery notes, discussion memos, etc.), most of the knowledge experienced from a project turns useless without the interpretation by those people who were directly involved in the project, because they are the only ones who remember the context and inter-connections between the single elements of documents. Indeed, the single civil-engineering documents are fragmentary. What is the situation now: the building industry will lose important information like experiences with the end of each building process, because nobody has the time to search in old unstructured civil-engineering documents.

The knowledge equation usually defined as $K=(P+I)^S$ (K = Knowledge, P = People, + = Technology, I = Information, S = Sharing) takes effect in civil

[1] Dresden University of Technology, Department of Civil Engineering, Institute of Applied Informatics in Civil Engineering, Germany, 01062 Dresden

engineering in a very strong way. Information in civil-engineering documents are fragmentary because context information and interconnections between the documents are not given explicitly. The objective is to extract as much context knowledge as possible from these fragmented documents. If a way could be found to retrieve the context knowledge from all the single stored documents, the entire knowledge and experience or at least a great amount of it would be available to everyone and could be used for planning and carrying out future projects. Hence, the digital storage of the civil-engineering documents together with the possibility to retrieve the context knowledge will be the long-term storage of project experience.

Architecture of the retrieval system

The architecture comprises 3 levels. First the document domain, second the design product model domain and third the human expert domain. An overview of the architecture of the knowledge system is given in fig. 1.

On level 1, a preliminary object-oriented knowledge network will be extracted from the documents by applying general text analysis methods combined with clustering techniques. The documents are parsed for civil-engineering key words. As a result different relationships between the key words determined by the investigated project are obtained.

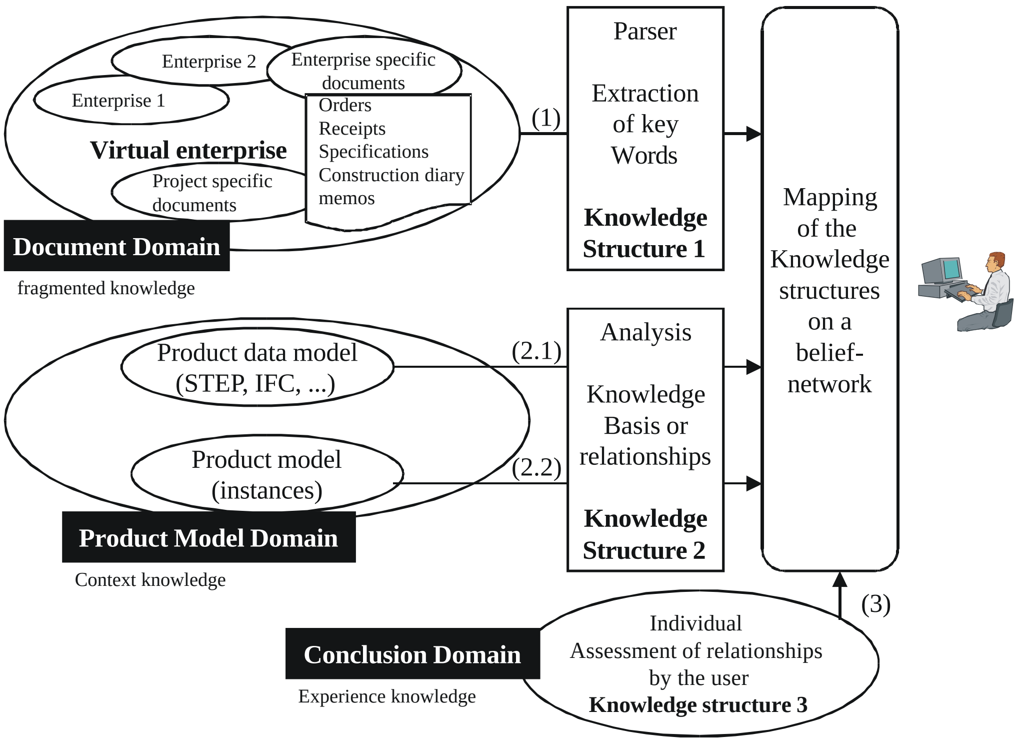

Figure 1: The three-level architecture of the retrieval system

The necessary lexicons are determined by the specific domain of AEC (Architecture, Engineering, Construction). Text analysis methods, e.g. used by the Alta Vista search engine and the Online Analytical Processing Methods, and which are for instance implemented in ConText (© Oracle Corporation) are used for a initial estimate of the inherent information structure in the document base related to a specific request.

On level 2, the Product Model Domain, knowledge contained in the design product data model and the instantiated product model will be used in order to improve the still vague results obtained by these generally linguistically based methods. Context knowledge is extracted from the knowledge structure of the design product model and design product data model. The context knowledge from documents can be verified and improved with the context knowledge of the design product model. From this, the context knowledge is to be mapped onto the initial document knowledge network in order to refine it and enhance its reliability. This kind of knowledge can be deduced and transformed in a formalized way by applying interoperability methods, e.g. mapping and matching methods [Katranuschkov].

Nevertheless some uncertainty and fuzzyness will remain. This should be attached on level 3 by modelling and limiting this uncertainty explicitly by fuzzy network methods. This allows first that uncertainties are explicitly propagated during reasoning and second that the user gets aware of these uncertainties. A neuronal network will be created with normalized data from documents and product models. Fuzzy rules will be extracted from this neuronal network. This fuzzy rule network can be used for finding answers in a specific situation. The requesting construction engineer can modify and store relationships between the fuzzy rules according to this expert knowledge from past experiences into the knowledge system.

The final result of a query can be a list of documents representing this requested project knowledge, ranked by their reliability or an answer generated by the fuzzy rules. The relationship between the documents can be visualized by knowledge maps, taking the reliability values as zoom factors, for instance.

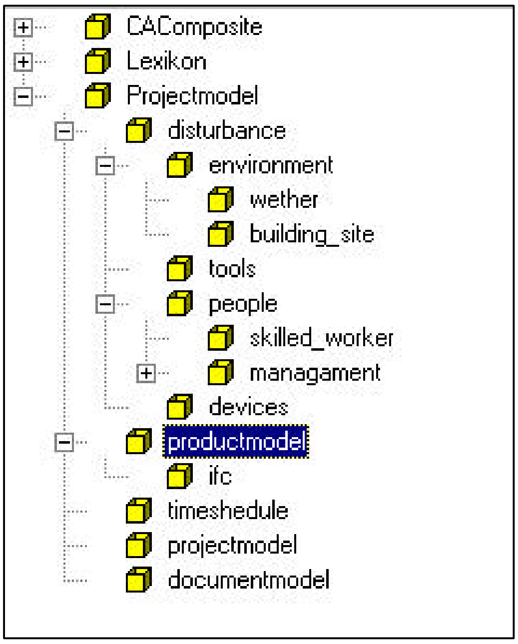

Figure 2: Part of project model class structure

Document Domain

A lexicon, which includes important parts of knowledge like construction key words, synonyms and homonyms for the specific domain of AEC will be determined by a document class structure, product model structure and project class structure.

During the process of general

text analyses the lexicon will be improved as a side effect with the occurence frequency of words structured in a hierarchy determined by the building process. The knowledge system will be completed with a product model structure including product model and product data model, a project model class structure (as shown in fig. 2) and document class structure. All parts of the knowledge system will be influenced by the other parts and have a reverse connection to the general text analyses with the lexicon.

For instance, impacts on the building process can be analysed too and described with their influences on costs, quality of the building and time schedule. By the way, the relationship between the key words "cost", "quality" and "time schedule" will be described in the knowledge system.

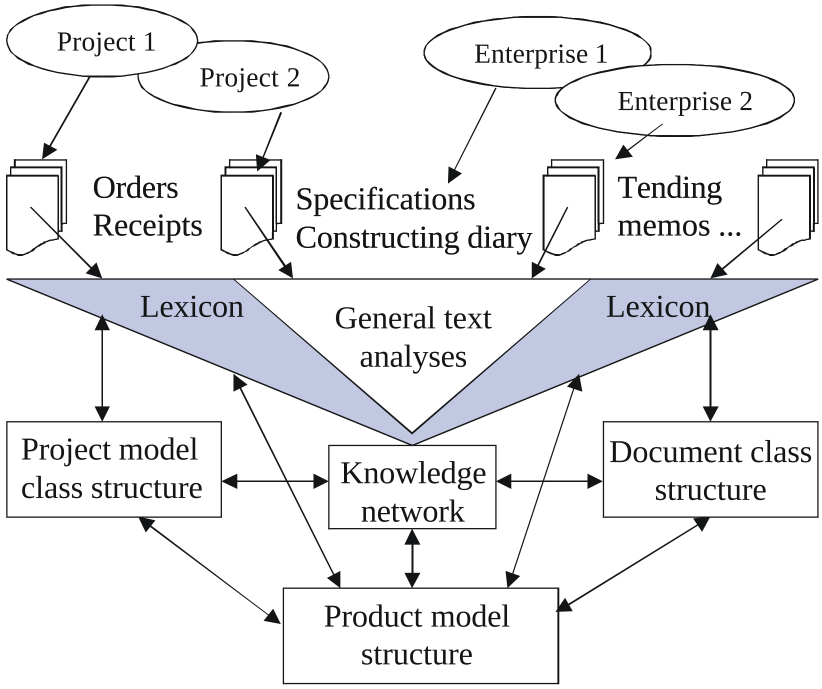

Figure 3: Document domain: Extraction of context knowledge

Knowledge from different projects and enterprises can be compared to enhance the reliability of relationships between the civil-engineering key words, as shown in fig. 3. Each knowledge acquisition process influences the knowledge structure 1, shown in fig. 1.

During the process of general text analyses, the concepts pruning and fuzzy conclusion combined with monotonic conclusion will be investigated. The non monotonic conclusion is not necessary, because the knowledge processing take place after the end of the building process. In the following up step we will check the heuristic search to increase the performance of conclusion process. Heuristics will be developed with investigation of different specific cases and forms and orders.

Different inference strategies are possible: either forward chaining or backward chaining with either weight search or depth search. Using backward chaining, a hypothesis will be located with available facts in the knowledge base. Forward chaining uses available facts to find out a hypothesis in the knowledge base. Weight search investigates all nodes of one layer before starting in the next layer of knowledge structure. The depth search starts with one branch to the last layer to locate a hypothesis in the knowledge base. The combination of this inference strategies will be investigated to find out the best fitting methods for knowledge processing.

Product Model Domain

Different product models and data product models are used in civil engineering, for example domain models of architecture and structural framework planning. For knowledge extraction these models have to be first combined or equalized. For this different methods are known:

- Build harmonious model parts by intentional integration or model definition (STEP [ISO], IFC [IAI]).
- Integration by definition comparable with multi data base systems [Motro].
- Integration by communication, for instance with object oriented model management OOMM [Rüppel].

These kinds of solutions demand the application of new models, definitions and restrictions are therefore slowly accepted by civil engineers.

In our case we have to combine one dynamic knowledge structure with product data model, described in EXPRESS-C. By that way, equalizing or harmonisation is not possible. Therefore we will apply mapping and matching methods [Katranuschkov] to merge the dynamic knowledge structure with the product data model.

With this methods we get all necessary knowledge included in different models and structures of the construction process into one model and we can describe semantic, structural and functional differences or inconsistence of the knowledge just as the relationships between key words.

Conclusion Domain

Using the result of mapping and matching we can extract the knowledge of the investigated building process and create new fuzzy rules in the knowledge network. This knowledge can be used for the verification of knowledge extracted from the documents. This gives us the possibility to reduce or limit the uncertainty and fuzzyness step by step.

The conclusion domain, shown in fig. 1, includes the fuzzy system and an interaction tool. The interaction tool gives the possibility to change the conclusions of the fuzzy rules by civil engineers. At first, based on the rules and the conclusion the fuzzy system generates a draft determined by the extracted knowledge in an automatic process. The result of the conclusion describes the relationship between the construction key words and will be visualized by a knowledge map with level lines. For each construction key word, which has to be investigated, a linguistic variable will be developed. A linguistic variable describes human experiences with mathematic rules in a good way. As an illustration three linguistic variables (see fig. 4) and a short cut-out (shown in Table 1) will be discussed.

Premise 1		**Premise 2**		**Conclusion**
if (project flow) = start	and	if (activity on time) = late	then	cost development = high
if (project flow) = late	and	if (activity on time) = late	then	cost development = small
if (project flow) = late	and	if (activity on time) = medium	then	cost development = very small
if (project flow) = start	and	if (activity on time) = very late	then	cost development = very high
if (project flow) = early	and	if (activity on time) = very late	then	cost development = high

Table 1: Short cut-out of fuzzy rules in the knowledge network

This example shows, that late activity in an early phase of project has more influence on the cost development than late activity in a late phase of the project. The fuzzy rules will be written in C++ . Different rules describing the building process can be created with this knowledge system. With this method key words can be used for the communication between the civil engineer and the knowledge system. This will improve the acceptance of our knowledge system in the civil engineering world.

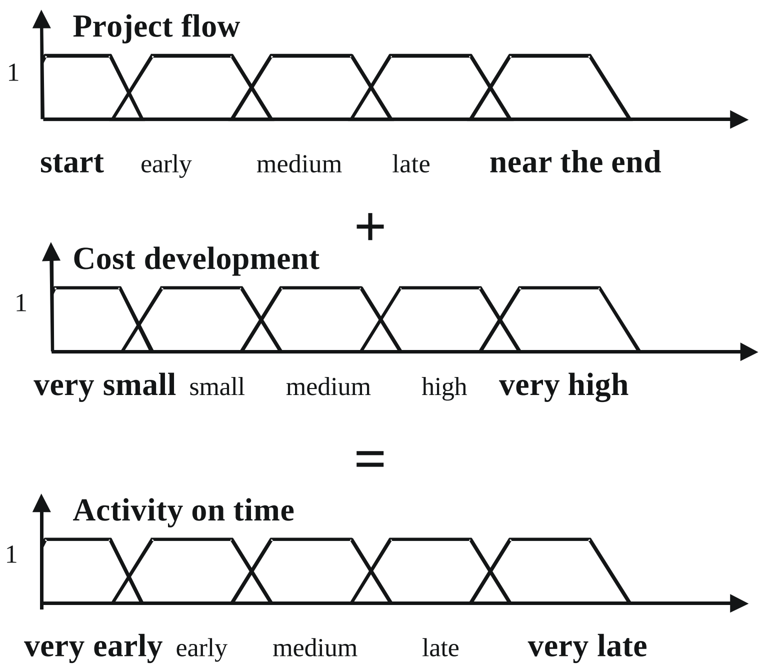

Figure 4: Fuzzy sets for linguistic variables

Clustering

The development of the knowledge system follows the architecture of the retrieval system, shown in figure 1. The part of the knowledge system will be structured in clusters, see in fig. 5. The cluster structure follows the idea of recovery systems. Clusters will be developed for a better overview and structuring of the knowledge system. This cluster network is a new view of our knowledge system. Each cluster is determined by a different structure and content. They will influence each other via the cluster network and will be additionally influenced by time schedule and enterprises.

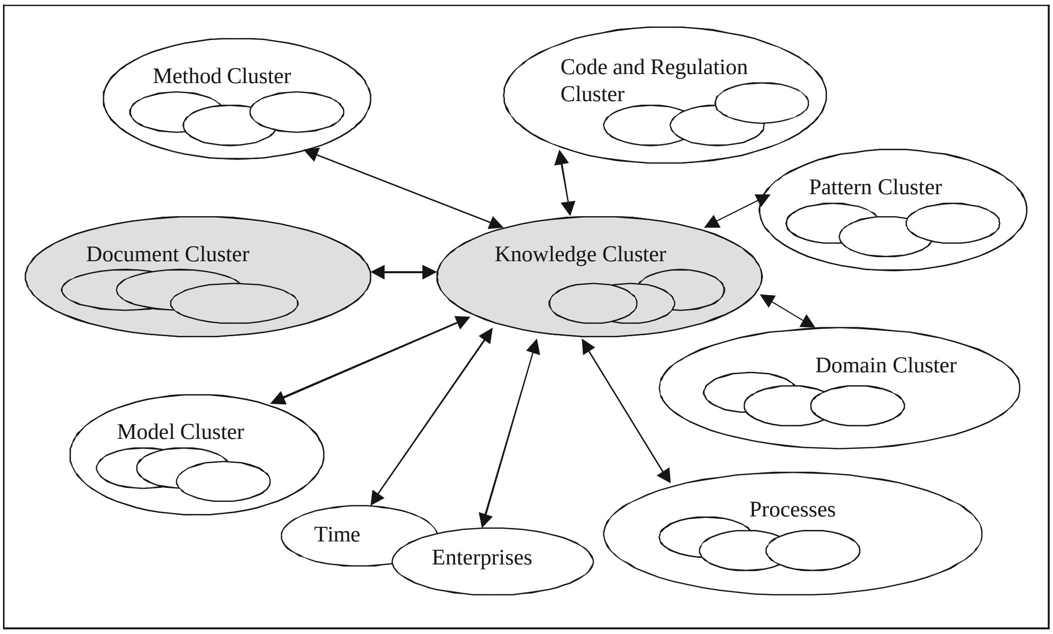

Figure 5: Cluster network

Clusters are different in content and structure and concentrate on a specific element of the knowledge system such as:

- The **method cluster** includes all methods of the knowledge system as
    - Knowledge extraction methods using documents
    - Text analyses methods
    - Artificial intelligence methods as fuzzy logic, neuronal networks, linguistic
    - Mapping and matching methods
    - Mapping operations etc.

  It is not hierarchically structured, comparable with the document class structure.
- The **document cluster** includes the document class structure.
- The **pattern cluster** gives an overview of pattern and pattern structure.
- The **knowledge cluster** includes the knowledge network, shown in fig. 2. This will be determined with an object oriented data base and artificial knowledge presentations.
- The **code and regulation cluster** shows the relationships between construction codes and regulations such as the DIN, HOAI, EU-Codes.

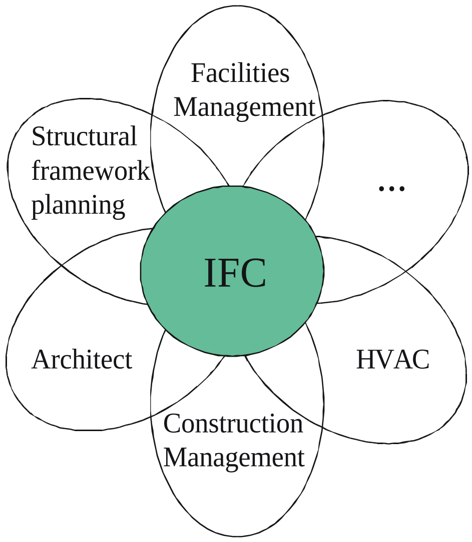

Figure 6: Model cluster

- The **domain cluster** describes the relationships between the different construction domains such as architecture domain, construction management domain, facility management domain and HVAC domain.
- The **model cluster** includes models from architecture, structural framework planning, construction management etc. and describes the influences between the different models and which parts are shared by other models (see fig. 6).
- **Time schedule and enterprises** can grow to a cluster during our future research process.

Conclusions

Different concepts of knowledge acquisition are investigated. We have excluded data mining methods, neuronal networks and deductive learning because of the specific situation in civil engineering and the availability of documents from selected projects. Object oriented project model class structure and the lexicon are already developed as prototypes by the investigation of knowledge acquisition and processing.

References:

Motro A. (1987): *Superviews: Virtual Integration of Multiple Databases*, IEEE Transactions on Software Engineering 13(7), S. 785-798.

IAI (1999): *An Introduction to the International Alliance for Interoperability and the Industry Foundation Classes*, IFC Release 2.0, IAI Publ., Oakton, VA., 21 S.

Rüppel U. (1994): *Object orientierted Management of Product models of structural framework planning.* dissertation, report 1/94, Inst. of numeric methods and computer science in civil engineering, Darmstadt University of Technology.

ISO 10303 -1, -11, -21, -22, -41 bis 49 (1994-96): *Product Data Representation and Exchange - Parts 1, 11, 21, 22, 41-45*. ISO TC184/SC4, Geneva.

Katranuschkov P., Scherer R.J. (1996): *Schema Mapping and Object Matching: A STEP-Based Approach to Engineering Data Management in Open Integration Environments*, in Turk Ž. (ed.) “Construction on the Information Highway” Proc. CIB-W78 Workshop, Bled 10-12 June, CIB Proceedings Publ. 198, Univ. of Slovenia, Ljubljana, pp 335-346.