A Model of Dispersed Historic Architectural Knowledge Base
Exploring the potential of a data warehouse structure

Michal Kabata [1],Krzysztof Koszewski [2]
[1]Polish Japanese Institute of Information Technology,
[2]Faculty of Architecture, Warsaw University of Technology
http://www.pjwstk.edu.pl   http://www.arch.pw.edu.pl

Abstract. *This paper is based on the experience with creation of a small knowledge base about the between-war architecture of one of the Warsaw districts.*
Design, as a process of creation, combines processing of procedural and declarative knowledge. There is a vast amount of declarative knowledge of different kinds to be collected even before the design process starts. Advances in ICT (Information and Communications Technologies), particularly in such field as databases, data warehousing and knowledge engineering, make it a lot easier to design complex systems, which will allow to combine procedural and declarative knowledge. We used historic-architectural knowledge as an example of the second kind mentioned here. The sources of this kind of information are dispersed, the data is gathered in various formats, using different standards and for various purposes. Past experiences with creating detailed architectural heritage inventories in Poland led us to a conclusion that such a subject specific knowledge base may be a part of larger hierarchical structure, which still needs to be built. These are the reasons for adopting a data-warehouse-like structure, which responds with it's tools to such needs.
The assumptions for such system are presented and the context-based structure is discussed.
During our work we also came for some more general conclusions. These concern a need for disseminating an OpenSource Society ideas through all the keepers of information related to architectural heritage.

Keywords: *information and knowledge management, database systems, architectural heritage, data warehouse technology.*

Introduction

Data warehousing

The eighties' paradigm of architectural design process was defined as knowledge based activity focused on formalization of such knowledge. Soon it became obvious that there was a big "reality gap" between the potential of knowledge representation for architectural systems and its practical use (Björk 2001). Having all the work that has been done in mind and observing instant development of knowledge management systems, the question we want to pose is: is it possible to achieve a new quality of information-intense tasks at early design stage. Most of earlier systems usually dealt with procedural knowledge, as designing is a process of creation. However, there is a vast amount of declarative knowledge of different kinds to be collected even before the process starts. Advances in ICT (Information and Communications Technologies), particularly in such field as databases, data warehousing and knowledge engineering, make it a lot easier to design more complex systems, which will allow to combine procedural and declarative knowledge. Collecting information and enriching it with knowledge supplied by experts is a task that can be discussed in context of knowledge engineering. Storage of such large amount of diverse data pleads for distributed and flexible database solution, based on data warehousing technology. And last but not least, searching and filtering such database requires both data warehousing and data mining techniques.

Rediscovery

Sometimes a solution to a simple and isolated problem leads to the development of more general procedures, knowledge and ideas. The same scheme had place in our work. The main task of the project (an addendum to the Ph.D. dissertation of one of the authors) was to design a model of subject-specific knowledge base about architecture of Saska Kepa, a single family housing district of the between-war Warsaw. Some of the most significant examples of modern movement in polish architecture can be found there. The idea of a knowledge base for this district is justified by the need for documenting it's unique architectural heritage. The large part of the original building documentation and iconography is still in private hands and vanishes almost on daily basis. The rest of related documents was lost during the world war II or dispersed through various institutions: archives, museum, libraries. Every attempt to synthesize any aspect of the knowledge concerning the architecture of Saska Kepa forces to do the whole research almost from the beginning. The need for the mentioned knowledge base (KB) has lead us to a general solution and even to several other ideas, mostly rediscovered, but seemingly very underestimated or perhaps unwanted and avoided altogether by some.

It also led us to some thoughts about the model of knowledge-base which can support the design process at its early stages. The base we had in mind should contain architectural-historic knowledge, so the first general question has become:

Why and how do we use architectural-historic knowledge?

In general, this kind of knowledge is useful in three fields of activity associated with architecture: design, science and education (including dissemination of this knowledge in the society). Having in mind the most interesting area: design, we can observe a change of attitude throughout the history of architecture. Here is its brief outline. One of the approaches, with its roots back to Vitruvio (or rather the Renaissance rediscovery of his work), leads to treating architecture or architectural design more like knowledge or even science than art. The base of this science was architectural-historic knowledge, used more or less directly in the design process in the form of various catalogs and templates. During the XIX century more analytical approach was introduced, like Viollet-le-Duc or

Choisy works, but it was still based on the interpretation of past decades of architecture. This attitude dramatically changed at the beginning of the XX century with the dawn of modernism, which denied all the relations to the past. But this approach radically evolved. The importance of architectural tradition shifted towards contextual concepts, and in this way influenced the design process and its results in a more indirect way. These contexts play such an important role that they can be used as a basis for creating rules which are a vital part of a knowledge base structure.

A hundred years after the first signs of its beginnings, we may say that modernism has become a part of history of architecture. It is a decision of contemporary designers if the architectural-historic knowledge, including modern movement, is a foundation for defining design context or it acts more like a source of forms and paradigms, playing a role which it denied history to play. Such kind of consideration is even more appropriate, when we speak about specific knowledge-base for the architecture of Saska Kepa, Warsaw enclave of modern architecture.

A biased (step one) approach

Every part of surrounding reality may be considered as a separate system. The condition is that it has to be delimited and there must be some kind of relationship between parts of it. We can describe this system in many various ways, changing description detail level or selecting the data we want to use. We can even change system boundaries. But all these changes, reflecting the way the system is described, depend on the purpose of description (Wrona et al. 1996 p. 146). So, by defining the objectives of the description, we define the system and the models we apply to the derived data.

The system - as a part of reality - is a source of information that can be used in various ways. While speaking about the early stages of design we must admit that in every design case the system itself is different. That is simply because the design task varies from case to case. On the basic level of generic information this seems to be consistent with a paradigm of situated computing (Gero 2003), even though the tool does not necessarily encode the information about itself. This could be exemplified by a large database of building materials, which may be used in different cases resulting in different solutions.

In process of retrieving information we intuitively divide the whole area of interest into subsystems (applying certain rules). This way, we select the information and then produce knowledge we can use. And although we sometimes rely on certain tools to do it, the process itself does not really depend on the tools used. Belief, that the “search space” for the design solution is unlimited thanks to the use of computational techniques leads us to yet another of the “reality gaps” (Björk 2001).

Such kind of approach to information in architectural-historic domain could end with conclusion, that it is necessary to store and to render accessible all the information we have and all that is still to be gathered. Is it really necessary and possible at all?

The Polish experience

The history of recording Polish monuments was strongly influenced by the history of Polish state. Such kind of activity was very problematic until 1918, when Poland had restored its independence. After this date a strong effort was made to accomplish detailed monument inventories. Unfortunately, before World War II only 5 out of 279 districts had inventories released in print. Considering the great amount of work and investment needed, in 1954 a decision was made to abandon detailed inventories and to discontinue the series of handbook catalogs. The idea of thematic inventories was brought up. These inventories, limited in their scope, could fill the gap between raw catalogs and the missing detailed listings and descriptions. This concept was only partially implemented.

When we started to work on the idea of knowledge base model for architectural heritage, the example we have chosen was such a thematic fragment, limited both spatially and temporally. And the reason was the same as for the described 1954 decision: a larger base was not possible for us to handle considering the size of the project it was a part of. It would be very difficult to create and experiment with it, even if we had only its model in mind. Further consequences are that for the base describing special fragment of reality - like the mentioned Warsaw district - we do not really need all kinds of descriptions, vocabulary and rules. As far as the basic interoperability of records is provided, we can use reduced schemes and thesauri.

So the answer to the question: do we really need all of the information and complete structure for broad-style description is: no. But the information gathered and transformed into knowledge is still useful not only in the area of interest. The warehouse-like structure assures interoperability and there is an instant possibility of adopting new kinds of sources due to the open framework of this model. This leads to the hierarchical structure of knowledge based systems describing a chosen section of reality, usually the more complicated the less specialized. And this in turn may lead to a system, which is capable of delivering all of the knowledge needed during the design process (Fig. 1).

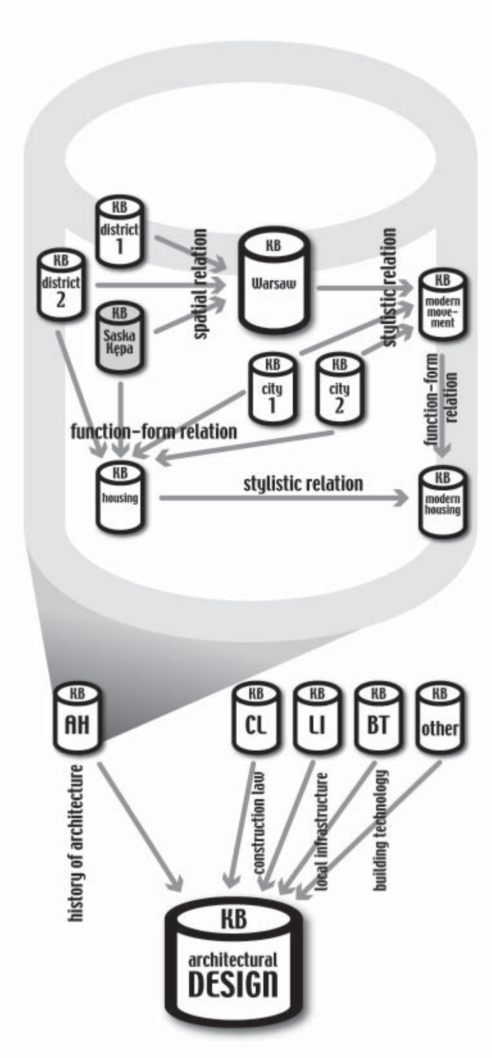

Figure 1. An example of a hierarchical structure of possible subject-specific knowledge bases

Knowledge base

Design

Any creative process, as design, uses both kinds of knowledge - procedural and declarative. Demand for each differs in various processes, both in a single process and also in its phases. The demand for purely declarative knowledge is far bigger at the early stages of the design process then in its final phases.

Knowledge base may supply both kinds of knowledge, but in a classical understanding KB is usually restricted to declarative knowledge. On the other hand, knowledge processing inside KB is in fact just another form of procedural knowledge. In our solution we have decided to gather only declarative knowledge since we want to enrich design by supplying facts and not procedures at this stage,

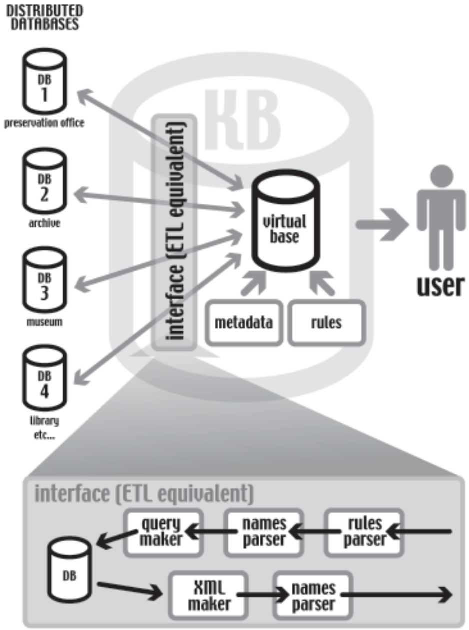

Figure 2. Diagram showing the knowledge base structure.

especially concerning the historical aspects of architecture. Although there is no limitation from KB regarding its contents, one or more of KB sources may be a database of procedural knowledge, such as guidelines or standards of procedures, without any redesign of the base.

The assumptions

The proposed solution should:

1. provide knowledge about historic buildings showing them in many contexts

2. provide the most complete knowledge based on all available information sources

3. be utilized by a wide range of users, providing them with the knowledge they need

Therefore, the system should:

1. contain expert's knowledge expressed in the form of rules derived from contexts applied to gathered information

2. rely on distributed, heterogeneous (in semantic and syntactic meaning) sources of information

3. be able to profile the knowledge considering three fields of activity in the domain of architecture: design, learning and research.

Basing on those assumptions we have chosen to use data warehouse as a foundation to our solution. First of all, we are not trying to build a single database, but rather a intelligent interface to many distributed databases. Second, ETL (extraction-transform-load) tools have almost one-to-one translation on remote database interface from our base point of view.

Characteristics: federation, contexts, solutions

The designed knowledge base is based on data warehouse with federation structure, in which the data itself is distributed along many remote and local databases (further called sources). In a warehouse solution, central database consists only of metadata, our solution enriches this database with rules, which turn pure information into knowledge, and allow for better searching (Fig. 2).

For convenience we have used XML for storage of metadata and rules as well as for communication with other databases. This presents a two-fold convenience. Firstly, the ease of programming, operation on XML files and other implementation related activities, and secondly, its ease of cooperation with sources. Most database systems are capable of returning results as XML, as web services extensively use XML (so we can easily talk to Web Services and not only to databases). XML is increasingly becoming accepted and even required as a standard for information exchange, therefore from our point of view designing of a system not only using but based on XML is very convenient.

We have used UNIX-style approach to the design of interfaces to sources. The main knowledge base needs only to be aware of existence of the source, while the rest of the processing is done by filters. A filter is a very small and simple program, designed for one purpose - to change the input in some way, and to produce an output.

The path from the knowledge base to the source consists of several such filters (in ETL this is an equivalent of extraction with elements of transformation):

1. Rules parser - changes rules names to query form (i.e. “around (built,1920,5)” to
“built >= 1915 and built <= 1925”)

2. Names parser - changes the names of database fields from generic ones used in the knowledge base to names adequate for a given source

3. Query maker - changes a generic query to a form understood by the source (MySQL, MSSQL, Oracle or anything other then SQL)

The source answer must be processed in a similar manner (in ETL this is both transformation and load combined):

1. XML maker - if the results are not in XML, they must be changed, and if they are already XML they might need some simplifying

2. Names parser - this is exactly the same filter as mentioned before

Such a design, using filters, allows for faster changes, easier adoption of new sources and easier maintenance. It also removes almost any requirements regarding sources. It would, of course, be best to use a source via XML over TCP interface, but we can, with almost the same ease, use an image of that source stored in Berkeley DB formatted file on a local drive, just by changing one small filter.

Rules

Information processing is a key difference between database and the knowledge base. Metadata, gathering rules, processing rules, output rules, context rules, search rules and so on... these are some of the possibilities to change pure information into useful knowledge.

We based our solution on several of them:

metadata - used to describe sources, relations between fields in different sources

context rules - used to enrich searching, use of fuzzy or rough set based rules allow searching of similar objects.

The contexts of historic building resulting in sets of rules are:

1. spatial
2. temporal
3. formal/functional
4. stylistic
5. defined by person/corporate body
6. historic
7. bibliographic
8. iconographic

Fuzzy-like rules allow non crisp searching over dates, and other numerical fields. example: “around(1000,5)” makes tolerance of 5% around value 1000, resulting in SQL “x >= 950 and x <= 1050”

Rough set based rules give even more possibilities; we can build almost any similarity relation, basing on any field (numerical and symbolic)

example: “young_Corbu()” may result in query selecting objects that contain any of authors names (signatures), tolerance of years when he was young, and location or style - those values may be entered by hand or may be taken from the objects existing in a database. Second solution may result in very interesting relations between objects, and allows to define such things as style basing on objects, not on expert knowledge and assumptions.

Rules define contexts of objects in knowledge base, and contexts define relations between these objects. As such, they are one of the core components of the architectural design.

User profiles (bias step two)

Despite our open approach to information, we see a need to restrict access to some information for general security reasons. There is no need to allow everyone to gain access to detailed building plans, as they may be used in wrong ways. Also, for users convenience and clarity, the amount of displayed (but not the total available information) should be limited. For example, a person interested in purely historical knowledge may find all the technical information about buildings or formats and resolutions of pictures quite confusing.

On the other hand, such profiles might provide useful differentiation of users regarding rules - an engineer and a designer may have a bit different understanding of some relations - and thanks to the profile-based rules we are able to provide each of them with the information they need and in the context suitable for them.

Of course management of such profiles, especially restricting access to information is quite a difficult task and it is beyond the scope of our project at the current stage.

OpenSource society

Such approach to design and information gathering needs very specific approach to information itself.

In an information-based society, every bit of information can exist in two forms: open and closed (restricted). Restricted information represents value, usually quite easily materialized, but in scientific understanding those two approaches to information have quite different outcomes. While open systems lead to improvement and survival of that information, restricted information may degrade over time, and eventually be lost. Restricted valuable information looses much of its value, and what is even worse, it limits the ability of others to improve upon it, and come up with something new. So as a result of “greed” one may stop progress. Sharing information one has been searching for many years might sound foolish to some, but in the end, isn’t it just the very root of all science? The choice is simple, either to keep all information to yourself or to share it with the world, possibly gaining respect of others, and really contributing to scientific community, or even society as a whole.

The concern is not only about science. All the data kept in all kinds of public archives, all cultural property gathered in the museums, vast amount of information held in libraries - these can be useful in all the fields of activity. Also the activity in the field of architecture.

OpenSource society and old hackers (best programmers, not to be mistaken with today understanding of that word), have proven that sharing of knowledge (not only pure information) and giving feedback to other people about their inventions and expertise may only improve society and lead to a continuous progress.

It’s a noble idea that can be made real, without any cost other that a small change of attitude to ourselves and our possessions. Inappropriate understanding of the concept of intellectual rights and property, promoted mostly by agents, may lead to impaired dissemination of knowledge.

Conclusions

1. The architectural-historic information plays an important role in various fields of activity associated with architecture. Design, research and educational tasks require instant access to the sources of such knowledge. Recent advances in knowledge engineering may result in new approaches to gathering, managing and especially retrieving such information.

1. The bias-based approach to the concept of architectural-historic knowledge base, resulting in thematic repository may be considered as one of the effective ways of building such a structure. Data warehousing technology offers tools, which can be very useful in implementation of this idea. The other facet of the idea is knowledge profiling by modifying the scope of information and the rules applied according to the users’ needs.

2. Due to the broad scope of the sources possibly useful for the architectural-historic knowledge base there is a need for using distributed sources of information. Digital records, very vulnerable to loss, must be maintained by the holders of con-

tent, usually institutions responsible for preserving the heritage. As a part of their mission the sources should be accessible for such kind of retrieval. Again, data warehousing with the set of ETL tools seems to be an appropriate solution to be applied here.

3. Federational model of the knowledge base results in independent sources and hierarchical structure. Such an idea relies on possibility of using any base as a source for another, which can build a framework for repository of knowledge used in the design process.

4. The idea of knowledge base using distributed sources relies on common access to the information related to the architectural heritage. This requires constant discussion about the necessity and possibility of making such information available. Subjects such as intellectual law and property must be based upon a well designed compromise derived from the needs of creators, holders and users of the considered materials.

References

Björk, B., 2001, Open Source, Open Science, OpenCourseWare, in: Architectural Information Management, 19th eCAADe Conference Proceedings Helsinki 2001, pp. 13-17.

Blaise, J. and Dudek, I., 1999, SOL: Spatial and Historical Web-Based Interface for On Line Architectural Documentation of Krakow's Rynek Glowny, in: Architectural Computing from Turing to 2000, eCAADe Conference Proceedings, Liverpool, pp. 700-707.

Gero, John S., 1986, An Overview of Knowledge Engineering and its Relevance to CAAD, in: CAAD Futures Conference Proceedings, Delft 1985, pp. 107-119.

Gero, John S., 2003, Siutated Computing. A New Paradigm for Design Computing, in: Proceedings of the 8th International Conference on Computer Aided Architectural Design Research in Asia Bangkok 2003, pp. 579-590.

Jarke. M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P., 2003, Fundamentals of data Warehouses, Berlin – Heidelberg.

Klos, J., Miller, D., Wrona, S. K., 1996, Rola informacji w projektowaniu architektonicznym, Warszawa.

Martens, B., 2004, CAAD-Repositories for Scientific Information Exchange – An Overview on SciX-Pilots Related to CAAD, in: Architecture in the Network Society, 22nd eCAADe Conference Proceedings, Copenhagen, pp. 262-268.

Maver, T. and Petric, J., 1999, Virtual Heritage: Is There a Future for the Past?, in: III Congreso Iberoamericano de Grafico Digital, SIGRADI Conference Proceedings, pp. 482-487