ar:searchbox

Knowledge management for architecture students

Christoph Langenhan[1], Arne Seifert[2], Astrid Teichert[3], Frank Petzold[4]
[1,4]Technische Universität München (TUM), Germany, [2,3]TUM University Library, Germany
[1,4]http:/ai.ar.tum.de,[2,3]http://www.ub.tum.de
[1]langenhan@tum.de, [2]teichert@ub.tum.de, [3]seifert@ub.tum.de, [4]petzold@tum.de

Abstract. *As media-orientation and access to media becomes increasingly widespread in society, so too is the availability of architectural designs on the internet. In most cases these are published in the form of raster images of plans, elevations and perspective drawings together with written descriptions on architecture databases and platforms such as archINFORM or nextroom, as well as on the homepages of the respective architecture offices. Knowledge is generally regarded as useful information. However, the literature does not elaborate clear differentiations between what is knowledge, and what is information and data. In our view it is the preparation of information in data structures that makes it useful as knowledge. Knowledge management systems are therefore intelligent information systems in which knowledge is presented and made useful through representation and modelling methods (Abeckerand Decker, 1999).*
Keywords. *Knowledge management; ontology; information retrieval.*

INTRODUCTION

During the design process architects and students often study the plans of buildings that have already been built or designed. Such information is available conventionally as a form of collective memory in architectural monographs and journals as well on the internet. For the purposes of targeted research, however, the accessibility of these sources is hampered by an inconsistent use of terminology and a lack of structured, non-subjective metadata.

In the design of buildings, a variety of tools and strategies are employed which can depend on the designer as well as the task at hand. Gänshirt (2007) notes that rigidly prescribed terminology is of little help to designers, whereas criteria, examples and the results of prior design work provide the designer with different potential courses of action. The study of buildings in a similar context or that are based on a similar initial premise is seen as way of approaching a design problem and developing a possible course of action. Ritteland Webber (1973) differentiates between "tame" or well-defined problems, such as those that scientists and engineers solve, and so-called "wicked" problems, to which he counts design and planning problems.

The ar:searchbox project aims to link and network information as a source of reference for the design of buildings in the early design stages and to elaborate fundamental principles for the use of metadata and related research strategies. Figure 1 illustrates the basic information needed in the early design stages.

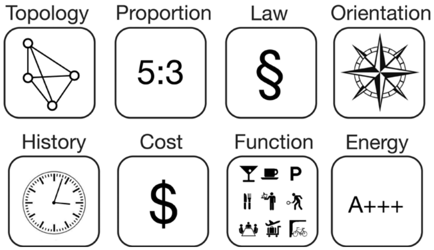


Figure 1
Information needed in early design stages.

As a web-based system, it provides a platform especially tailored to the needs of architecture studentswho are able to enter content themselves. By storing building projects in a central database, the project aggregates previously disparate information and thereby overcomes barriers between media. Each reference project refers back to the original information sources so that students can consult the original sources if need be. An initial intellectual consideration of the building project results in a detailed description, including its special characteristics and precise categorisation. This way of capturing the data helps the student understand the project at several levels: on the one hand at a verbal descriptive level, and on the other in terms of its structural classification within a predefined categorisation system. At the start of the project a data structure and metadata scheme was designed based on data collected from model case study projects. This structure is flexible and can be adapted or extended at any time to reflect changing requirements. The selection of projects and manual entry (Figure 2) by the students serves to restrict the selection to relevant high-quality building projects and offers a form of quality control for the descriptive information.

The continuing expansion of the catalogue gives architecture students quick access to a pool of reference design projects (Figure 2). This new knowledge-management system offers a series of different research tools and assists the designer in the creative design process. For example, searching building projects within the existing categories can give rise to valuable ideas or stimuli for own design projects through a process of free association. The targeted metadata also provide a better understanding of the digital search mechanisms and offer ideas on how these could be applied for use with other research instruments.

RELATED WORK

In the early stages of the architectural design process, architects are only rarely able to specify the required information. Case-based reasoning (CBR) is an area of Artificial Intelligence (AI) and describes a knowledge management process based on conclusion by analogy. It attempts to assess similarities according to the basic premise that similar problems have similar solutions. In CBR a case consists of a problem and solution description. By entering a new problem description to obtain similar solutions the CBR system first searches for an old problem description. Figure 3 illustrates the basic concept of CBR, where similar problems have similar solutions. Aamodt and Plaza (1994) described this adaptation of the thinking process inside the CBR cycle with the verbs retrieve, reuse, revise and retain.

Since the middle of the 1990s the approach of applying CBR to design and architectural tasks has been known as Case-Based Design (CBD). The casebase contains information on buildings that have already been built or designed, enabling the computer to adapt solutions accordingly, on its own or with help from the architects. Table 1 provides a brief overview of some CBD systems based on two studies published by Heylighen et al. (2001)and by Richter et al. (2007) regarding the proposed approach.

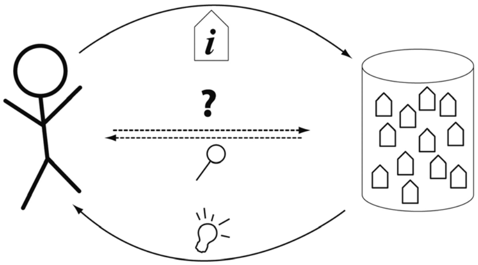
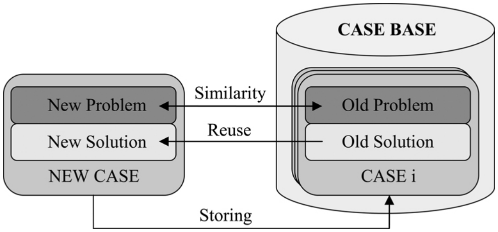


Figure 2
Information cycle of ar:searchbox.

Figure 3
CBR basic concept.



Six of the CBD prototypes CADRE (Hua et al., 1996), FABEL (Schaafand Voss, 1995), IDIOM (Smith et al., 1995), SEED (Flemming et al., 1994), SL_CB (Lee, 2002) and TRACE (Mubarak, 2004) aim to partially or completely automate the generation of building layouts by applying the retrieved solution. Two of these prototypes CADRE and IDIOM leave the selection of the reference project to the user. The remaining four FABLE, SEED, SL_CB and TRACE apply the solution to the given architectural problem automatically and generate building layouts independently with very little user input. The more state-of-the-art approaches aim to support users during the design process, like Archie-II (Kolodner, 1993), PRECEDENTS (Oxman and Oxman, 1993), CaseBook(Inanc, 2000), MONEO (Taha, 2006), CBA (Lin and Chiu, 2003) and DYNAMO (Heylighen, 2000).

The study by Richter et al. in 2007 identifies an acquisition bottleneck in putting complete case descriptions (problem and solution) into the case-base. We assume this is due to a lack of adequate input strategies, indexing methods and knowledge management procedures. For example at the Bauhaus-Universität Weimar, a housing database was developed that catalogued building projects according to so-called innovative criteria. The criteria used for the database are, however, not generally applicable and do not adequately and unequivocally describe spatial qualities.

The “Probado” research project, a joint project by the German National Library of Science and Technology at the University of Hanover, the Graz University of Technology, Austria, and the University of Bonn, examines methods of processing non-textual documents in libraries. The principle focus lies on

Table 1
Overview CBD systems.

CBD application and supported feature	Data Storage			Input System			Output System						
	Floor plans + text	Abstraction	Topology	Graphic	Verbal	Adaptation	Reference projects	Applying solutions	Graphical Information	Learning	Sub-problems	Semantic net	Analogy
Archie-II	X	X			X		X		X		X	X	
CADRE	X	X	X	X	X	X		X	X	X	X		
FABEL	X	X	X	X	X	X	X	X	X		X		X
IDIOM		X	X	X	X	X		X	X				
PRECEDENTS	X	X			X		X		X		X	X	
SEED		X			X	X	X	X	X		X		
SL_CB	X	X			X	X	X	X	X				
TRACE		X	X	X	X		X	X	X				
CaseBook	X		X	X	X		X						
MONEO	X	X			X		X		X				X
CBA	X	X			X		X					X	
DYNAMO	X	X		X	X		X		X	X		X	

content acquisition, description, search, presentation and storage (Berndt et al., 2010). In the context of digital architectural models, the aim is to achieve a semi-automatic indexing method and an intuitive visual search mechanism. The index creation is based on three-dimensional geometric data. In addition to search queries using examples (3D models) and text input, it is possible to visually search 3D models using the "Princeton Shape Benchmark" (Shilane et al., 2004). By using a "Room Connectivity Graph" (Berndt et al., 2010; Wessel et al. 2008) spatial configurations (topologies) can be drawn in sketch form to search for 3D models with a similar configuration. "Subgraph Matching" is used to determine model similarity.

MEDIATUM | AR:SEARCHBOX

The TUMünchen (TUM) uses an open-source software package called "mediaTUM" as a central media server for the TUM. The system is developed, operated and maintained by the University's Library. In addition to providing general functionality as a repository for the individual university departments and facilities, mediaTUM can manage specific collections and projects and present these via a public interface.

The ar:searchbox project is a joint cooperation between the Chair for Architectural Informatics and the TUM University Library. A special extension of the repository functionality offers students a custom interface through which they can present project data as well as establish relationships between architectural projects. The aim was to provide students with a platform for storing, managing and researching architectural projects.

To this end a special metadata scheme was developed that is specific enough to describe the building projects in detail but also flexible enough to cover a broad spectrum of projects. The metadata also includes attributes that allow different sorting criteria and categorisation. This makes it possible to search the data via more fine-grain criteria. It is also possible to search within hierarchical structures. Using a classification system, projects are stored in a flexible ordering system that allows one to browse the data set according to a particular topic. mediaTUM (Leiss et al., 2010) offers a variety of views, including tools for zooming into details of high resolution plans and images.

In addition to the metadata scheme, the objects that contain the architectural data are also adapted to the needs of the students. A project folder serves as a container for core information about a project and can contain an unlimited number of different plans, drawings and verbal descriptions. Each of these individual pieces of detailed data can be attributed with its own metadata so that it can be retrieved via the search function. Different views are generated for each object, for example thumbnail images, so that one has a quick overview of the information contained in a project folder (Figure 4).

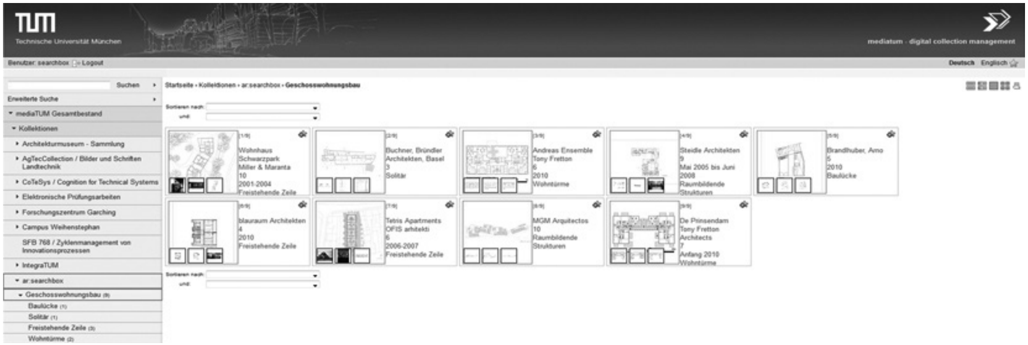


Figure 4
The user interface of mediaTUM/ar:searchbox.

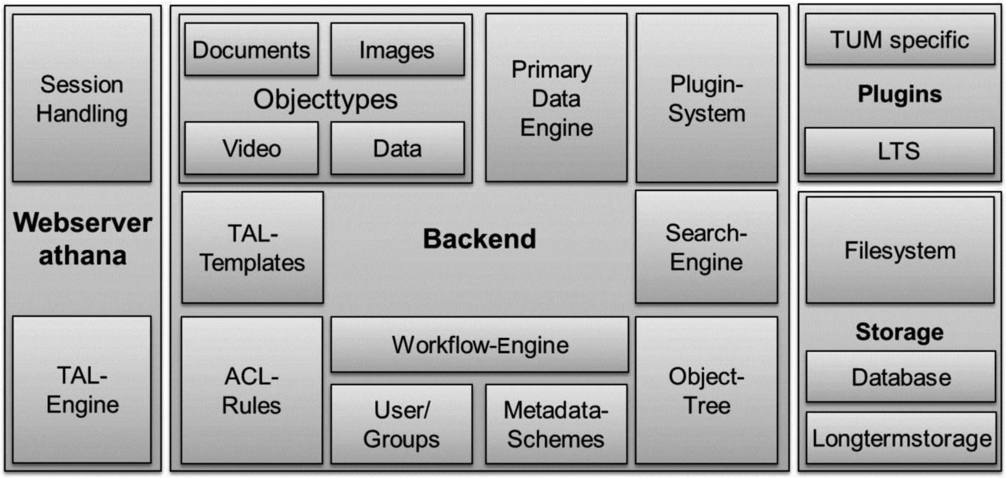


Figure 5
Technical structure - The core components of media TUM.

Through the decentralised administration concept of the media server, content can be entered in parallel from multiple locations and is immediately available for others to reference via ar:searchbox. A sophisticated user access control mechanism ensures that editing and viewing rights are available only to defined sets of users. The media server is accessed entirely through a browser and requires no further software.

TECHNICAL DETAILS AND INTERFACES

MediaTUM is an open source software under GNU General Public License. It is implemented in python and provides all management features via a web interface. The open software architecture with plugin concept enables an easy connection of different program extensions. MediaTUM [Figure 5] consists of four basic components webserver, backend, plugins,storage and archiving (Seifert, 2010).

With "athana" an own webserver component is included, that is responsible for sessionhandling and for generating HTML-output via a TAL-engine. The backend is the core component of mediaTUM. It includes tools for the administration of the different object types such as images,documents and videos. Furthermore it contains methods for building up the data structure and for the configuration of the stored data. A workflow engine, a search engine and components for digital rights management and user administration are also elements of the back end. Extensions and specific applications can be integrated via the plugin system. mediaTUM uses tools with different interfaces for the long term storage of data. The basic prerequisite for the connection of different systems is the implementation of an open API. For the long term preservation of documents and data mediaTUM uses the interface to the Leibniz Rechenzentrum. This local computer centre and provider of scientific data network offers the Tivoli Storage Manager of IBM as technical infrastructure for data preservation.

There are two different ways of storing objects in mediaTUM. Metadata are stored within a relational database (normally mySQL or SQLite). The digital objects are deposited into the file system on a configured position. The media server offers different interfaces for interoperation with external systems. MediaTUM facilitates for example the smooth exchange of data via OAI Protocol for Metadata Harvesting (OAI-PMH), web services or Z39.50.

DISCUSSION

During the process of designing, planning, building and the management of projects, a large amount of data is created. The information is stored in various forms, sources, platforms and different formats such as text documents (e-mails, technical reports, contracts, etc.), 3D models, BIMs, CAD drawings, diagrams, schemes, pictures and so on. Classifications are made in terms of types, morphology, similarity or patterns but the quantitative and qualitative comparison of functional as well as structural features is as yet not possible. The configuration of space and the relations between physical structures are hard to represent using keywords, in fact transforming these structural configurations into verbally expressed typologies tends to result in unclear and often imprecise descriptions of architecture.

A universal description and query language is indispensable for storing descriptive metadata independent of file type and source as well as structural, graphical or textual information. First of all, a user interaction should support the graphical sketch-based workflow of architects combined with textual, schematic and tabular input strategies. Secondly, an appropriate indexing strategy is needed in contrast to the overall data storage method used. Sketches are widely used in engineering and architectural fields as they are a familiar, efficient and natural way of expressing certain kind of ideas.

Paths through space, adjacency or the orientation of rooms are topological aspects that can be stored for further processing as an information subset in a so-called semantic fingerprint we proposed (Langenhan and Petzold, 2010). A semantic fingerprint stores various aspects of a building (e.g. topology, taxonomy, energy, geometry) based on ontologies so that this can be retrieved using semantic research strategies.

CONCLUSION

The paper presents a technological and organisational approach for the development of metadata schemes and the establishment of a knowledge management system based on reference projects that serve as a source of knowledge and inspiration for approaches to architectural design problems.

Using the metadata schemes developed for the ar:searchbox project and the defined criteria for recording a quality-oriented data set, a search agent software tool could be developed that searches the internet for verbal descriptions of building projects and uses the metadata to automatically create an index of reference projects. Using ontological text analysis, potential sites of useful information can be identified and with the help of corresponding semantic information, the found raster graphics could, for example, be identified as floor plans. Based on a "prior selection" of relevant raster graphics, an image analysis method using artificial intelligence methods could be trained to efficiently analyse images (Ahmed et al., 2011) and extract verbal, topological as well as geometric metadata from them for saving in the semantic fingerprint. The ar:searchbox project prepares the technological, methodological and organisational basis for such a system.

REFERENCES

Aamodt, A and Plaza, E. 1994, 'Case-Based Reasoning - Foundational Issues, Methodological Variations, and System Approaches', *AI Communications,* Vol. 7, Nr. 1, S. 39–59.

Abecker, A and Decker, S. 1999, 'Organizational Memory - Knowledge Acquisition, Integration and Retrieval Issues', *Knowledge-based Systems: Survey and Future Directions, Proceeding of the 5th German Conf. on Knowledge-based Systems*.

Ahmed, S, Liwicki, M, Weber, M and Dengel, A 2011, 'Improved Automatic Analysis of Architectural Floor Plans', *International Conference on Document Analysis and Recognition (ICDAR-2011)*.

Berndt, R, Blümel, I and Wessel, R 2010, 'PROBADO3D – Towards an automatic multimedia indexing workflow for architectural 3D models', *International Conference on Electronic Publishing,* 14.

Flemming, U, Coyne, R and Snyder, J 1994, 'Case-Based Design in the SEED System', *Automation in Construction,* Volume 3, S. 123–133.

Gänshirt, C 2007, 'Werkzeuge für Ideen - Einführung ins ar-

chitektonische Entwerfen' - Einführung ins architektonische Entwerfen. Basel: Birkhäuser.
Heylighen, A and Neuckermans, H 2000, 'DYNAMO - A Dynamic Architectural Memory On-line', *Educational Technology & Society,* 3.
Heylighen, A and Neuckermans, H 2001, 'A case base of Case-Based Design tools for architecture', *Computer-Aided Design,* 33.
Hua, K, Faltings, B and Schmith, I 1996, 'CADRE - case-based geometric design', *Artificial Intelligence in Engineering,* 10.
Inanc, BS 2000, 'CaseBook - An information retrieval system for housing floor plans', *CAADRIA*.
Kolodner, JL 1993, '*Case-based reasoning*', San Mateo, CA: Morgan Kaufmann Publishers.
Langenhan, C and Petzold, F 2010, 'The Fingerprint of Architecture - Sketch-Based Design Methods for Researching Building Layouts Through the Semantic Fingerprinting of Floor Plans', In: *International electronic scientific-educational journal: Architecture and Modern Information Technologies,*4 (13).
Lee, JH 2002, '*Integrating Housing Design and Case-Based Reasoning*', Dissertation. Carnegie Mellon University, Pittsburgh.
Leiss, J, Pretz, E and Seifert, A 2010, 'mediaTUM: Der zentrale Medienserver der Technischen Universität München', In: *Informationsmanagement in Hochschulen*, S. 365–377.
Lin, C.J and Chiu, ML 2003, 'Smart Semantic Query of Design Information in a Case Library', *CAAD Futures*.
Mubarak, K 2004, '*Case Based Reasoning for Design Composition in Architecture*'. Dissertation, Carnegie Mellon University, Pittsburgh.
Oxman, R and Oxman, R 1993, 'PRECEDENTS - Memory Structure in Design Case Libraries', *CAAD Futures*.
Richter, K, Heylighen, A and Donath, D 2007, 'Looking Back to the Future - An Updated Case Base of Case-Based Design Tools for Architecture', *eCAADe,* 25.
Rittel, HWJ and Webber, MM 1973, 'Dilemmas in a general theory of planning', In: *Policy Sciences*,1973 (4), S. 155–169.
Schaaf, JW and Voss, A 1995, 'Retrieval of Similar Layouts in FABEL using AspecT', *CAAD Futures*.