THE CATEGORISATION OF DATA FROM SMART BUILDINGS FOR A POSITIVE SUSTAINABLE APPLICATION

Information retrieval in smart environments - or "how to make a Google™ in the House"

SCHOCH, ODILO
Schoch Architectural Services Ltd., 8004 Zurich, Switzerland,
www.schoch-architecture.com
info@schoch-architecture.com

Abstract. This paper describes an architecture oriented approach of a categorization of information in so called computer integrated buildings (CIB). It follows an approach which seeks for the suitability of information categorization in order to achieve an energetically sustainable operation of CIBs. The CIB and its environment may exist as physical objects or as a simulation. The underlying question of this paper is: how can data of high temporal and spatial density be made accessible within a CIB with the goal of quickly delivering the most suitable answers for a specifical application. The subtitle „Google™ in the House" was chosen due to the achievable relation of a model with simple user interfaces, extendable pre-valuable data basis and high quality answers is being developed. The model's foundation is the potential of any built context that it is very easy to position data in physical space by adding its position by Cartesian coordinates. This architectural way of pre-sorting data is defined as logical arranging. Further on, known mechanisms of information retrieving and mapping are added to the model. A preliminarily summary concludes that energetically sustainable operations are backed through means of computational categorization models.

1. Thematical focus and delimitation

From an architectural point of view a building can be regarded as a container of numerous hybrid topics such as functions, activities, contents and objects. Examples are e.g. protection from climatical extremes, warranty of material safety, exercising activities like sleeping, working, eating, managing, loving, talking, etc. Walls, windows, sun orientation, geoposition, etc. are suitable when used with their attributes respectivly physics and content are used in order to describe the built context.

The work is not yet finished, which is the reason why fundamental questions on the paper can not be answered. One of those fundamental questions is: Is the emphasis focused on prognosis, assessment, categorization, classification, combination, arranging or on the quality of search results.

The definition of the term “information” as “data in context” provokes the obvious connection between digital data and architectural context in so called computer integrated buildings. At the same time the contents of the definition for the “semantic web” appears to be helpful on the search of information: “Knowledge management is concerned with acquiring, maintaining, and accessing the knowledge of an organization.” (Berners-Lee, 2003).

In a document concerning the way Google™ search engines work you can find the statement ‘design to act "like a human being would"’ (googlerank.com : November 2006). This statement is a basis of the model.

2. Aim

The goal of the theoretical work is to develop a general model for digital information management within the context of CIB. The prime application of such information categorization is an energetically sustainable operation of CIBs (Schoch, 2006a). This perimeter is applied in order to prevent from being lost in the broadness and complexity of information in physical reality.

Two approaches justify the focus on energetically sustainable building usage beyond the social necessity of taking care of resources and a general aim of optimization. 1) Only since recent years technology does allow intensive spatial acquisition of data. It allows new applications for building usage. 2) The processing of these large amounts of heterogenic data is predestined for the resource optimized usage of buildings. In this context resources are people, energy flows, spaces, infrastructures, etc.

3. Method

Aspects of built space, of well known internet search engines and of principles of information categorization are being examined by their qualification for information retrieval in CIB. Tools of knowledge management are being integrated into the scheme as they are auxiliaries of systematic solution finding. A model for information retrieval will be developed that enables the accessibility of hybrid information. The subject’s validity and target value will be proof-checked using two realized computer integrated buildings that were set up by the author.

4. Analyis

4.1 BUILT SPACE

With its ability of geometrical positioning and physical values, built space offers high and precise density of hybrid information. Hybrid information arises in particular when including changes of time, users and positions. The quality of content of such information is very high as a result of its diversity (Wang, 1997). Since the advent of digitally networked sensors, data about various issues can be recorded and/or processed instantly or with very slight

time lag. In addition, built space can often be categorized into typologies. Achten (1997) states that “The building type is a knowledge structure that is recognised as an important element in the architectural design process.”

Those links between building and knowledge management is taken as the foundation of the pure model developed. Although focussing on generic graphical representation, Achten continues to describe the potentials of an informed building model: “Therefore it is necessary to identify the elements of the graphic representation that play a role in decision making.” (Achten, 1997). He uses an approached based on ‘graphic units’. In data models such as semantic stacks, this modular thinking is resumed.

4.2 SEARCH ENGINES

The information offered by the vastness of the internet represents an enormous amount of knowledge which – for humans - seems to be only a mouse click away (Gates, 1990). Search engines offer retrieval functions that apparently or truly deliver the desired information as target link.

Fundamental schemes have already been formulated in patents since the 1980´s such as the statement that an „Interactive data retrieval apparatus“ can exist (Beaven, 1981). Another work stats that error tolerance and thematical proximity are possible as the system must „understand near matches and misspellings“ (Beaven, 1981). Weissmann et. al. stated in their patent in 1999 “A semantic space is created by a lexicon of concepts and relations between concepts.” As well as “Search results on the input query are presented where the target data elements that are closest in meaning, based on their determined semantic distance, are ranked higher.” Those issues are topics of information retrieval within CIBs.

In order to improve information retrieval advanced technologies such as digital learning systems (see Wical, 1996), and sorting algorithms that permanently merge the relevant data for a dedicated request. Systems seem to be generated by the support of application that follow schemes comparable with the patents mentioned above. The system seems to gain quick access to categorized information packages in advance (see Lindstrom et. al., 1983).

Interestingly enough contemporary works on „real space management“ associate the functionality of self-generated systems with the tasks of data management. In his article on parking space retrieval, Andres Sevtsuk mentions the agent based information retrieval (Sevtsuk, 2006). In most cases his conclusions allow faster search results and better usage rates as long as a system is utilized that is based on social behaviour patterns.

4.3 CONTEXTUAL INDEXING, ASSESSMENT AND THE SEMANTIC WEB

Analysing the descriptive assignment of information in digitally networked physical buildings concerning the logic and/or the context is, beside others, a discussion of informational content. Compare the computer integrated building to a permanently active human servant. A search request to the building as something like “are my users feeling climatically well?” The servant knows that the definition of “climatic well being of people” shows hybrid parameters.

Papamichael (2003) describes in the chapter ‘What Computers Cannot Do’ that the computationally impossible task “is related to the evaluation of qualitative performance parameters, which cannot be delegated and is certainly not computable.” This results in the difficulty in avoiding algorithmic situation where qualitative performance parameters might occur.

The assessment of single information bits, information groups or results is central for determining the importance respectively non-importance of the theme considered. The assumption is quite obvious that a principle importance of one single information bit may be identified by the amount of services that use those single informations, similar to the PageRank™-system of Google™ (Brin et. al., 1998). Immediate and known critique is the issue of quality rather than quantity of search results in the Internet. In the case of a CIB particularly anomalies are of interest and importance as they might cause serious threat, damage or costs. An artificial importance needs to be created if necessary in order to stay in the ranking system.

The resource description framework (RDF) of the semantic web (http://www.w3.org/2001/sw/ : December 2006) is the technological foundation of the semantic web. According to Berners-Lee „the Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation “(Berners-Lee et. al., 2001). This basis does allow the fixture of hybrid information in a digitally linked building. Information can be grouped by logical themes in combination with an universal description of dependencies within single topics (see Achten, 1997). This leads towards high quality as well as fast and consistent information retrieval.

4.4 3D/4D GIS – GEOGRAPHIC INFORMATION SYSTEMS POTENTIALS

In relation to GIS handling, Agnolotti states: “Although our life takes place in three dimensions, commercial Geographic Information Systems (GIS) were mostly limited to handling data of two dimensions” (http://www.geoit.ethz.ch/research/mdgis_en.html, December 2006). “GIS, by means of its capability of handling topology, provides each building with the necessary information about the built and natural environment surrounding it and the consequent influence on its thermal performances.” (http://www.geoit.ethz.ch/research/simgis_en.htm, December 2006) The general logics of GIS are expected to work in small scale indoor applications as well, which leads to an approach based on 3D/4D GIS. Further on Agnolotti’s states that “The concept for a coupling of GIS and simulation programs bears good potential for optimizing both energy efficiency and user comfort in buildings.” (Agnolotti et. al., 2005).

4.5 MAPPING

If several instances with the hierarchical and branched structure of semantic stack (http://www.w3.org/2003/Talks/05-gartner-tbl/slide29-0.html, November 2006) are layered with overlapping attributes, a topology develops that can be compared with Rhizom (Deleuze G. et al., 1980). Unifying element in CIB is the ‘location’ (see figure 1). The “knowledge map” is introduced. This kind of cartography enables search engines for

building relevant data to find important information in a target oriented way with fixed schemes. Caldwell defines three classes of knowledge map: “The ‘procedural knowledge map’, the ‘conceptual knowledge map’ and the ‘competency knowledge map’ (Caldwell, http://www.destinationkm.com/articles/default.asp?ArticleID=1041 :2005). In this manner it seems possible to develop schemes that enable the rapid and context specific information retrieval within a CIB.

Cartography in the classical sense describes a pictorial interpretation and abstraction of the reality for a specific aim. Interestingly the rules of abstraction are designed by human specialists. This process of design is being applied semi-automatically at present in order to achieve high quality and verity. For application within CIB, both rules of abstraction and processes should be worked out by humans in order to achieve high quality.

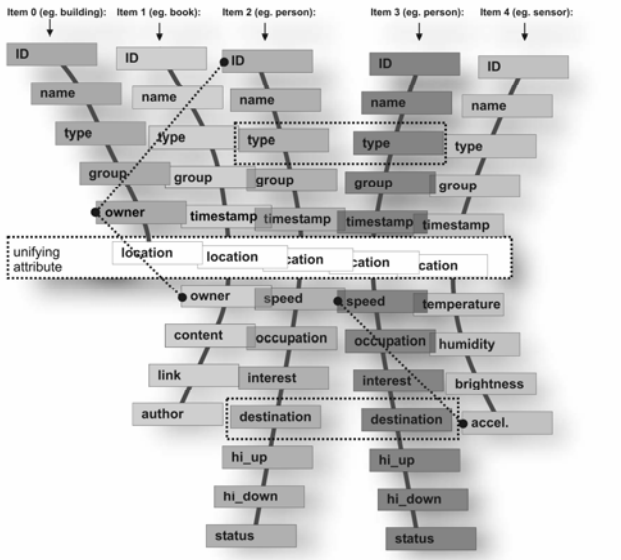

Figure 1: Map of Hybrid descriptors of objects in CIB and their unifying attributes like ‘location’, ‘type’, etc.

5. Model

Chapter 4 deals with research of processes, systems, technologies and requirements. It points out that there are obviously a number of solutions available, which should allow high quality information retrieval within CIB.

The aspects are summarized:

1. Due to their diversity the content of hybrid information is of high quality.
2. Agent based information retrieval is promising.
3. Eventually an artificial importance needs to be created in order to stay in the system of ranking.

4. The semantic web allows high quality, thus rapid and consistent information finding.
5. No complex communication at building automation level is needed.
6. The impact of time, culture, objects, individuals, etc. must be integrated.
7. Rules of abstraction designed by humans create quality.
8. Processes of mapping lead by humans create accuracy.

These aspects formulate the basic structure of the model for information categorization in CIB´s. The linking element is the four-dimensionality of a built environment, which is applied to the semantic data space. This application of four-dimensionality seems possible with most data. An automatic pre-sorting emerges within the classical architectural context that allows the placement of a measured value in the local context – and at the same time pre-filtering the information flood. As preliminarily results the amount of data is reduced and exceptions are identified.

A process chart in the style of a Google™-search request will probably look like this: several linked instances deal with tasks of acquisition, subscription and linkage. "Google™bot, a web crawler that finds and fetches web pages. An indexer that sorts every word on every page and stores the resulting index of words in a huge database. The categorization of spacial position and theme. The processing of a semantic map with the help of semantic descriptions. The query processor, which compares your search query to the index and recommends the documents that it considers most relevant." (http://www.Googleguide.com/Google_works.html, November 2006).

6. Proof-Checks

For this paper the question of a systematical approach for assessing building relevant data and information may be answered as follows: The aim of information categorization is increased user comfort and an economical building operation. It is assumed that sustainable building operation is of economical benefit, be it short-, medium- or long-term (Schoch, 2006a). An account of success concerning comfort is difficult to achieve. Therefore the aspects of economic building operation are discussed. The basis of the assessment is cost amortisation, cost-benefit and broadness of service (social sustainability). The verification took place on selected installations and themes and practical with a test installation for fine grained data acquisition and modification of the room temperature in built environment. Two examples are described in detail:

1. A conceptional proof check (simulation): The installation of a digitally networked building at ETH Zurich proved: optimized usage of space resources and reduced thermal energy flow can be achieved by hybrid information retrieval. The real buildings offers lecture halls, displays, students, staff and databases with information on activities and enrolments. By linking room specific information (room status like 'room at present/soon occupied', room functions like 'which use does this room offer?', room

capacity like ‘space for how many persons?’) and event specific information (time of event, number of participants, necessary functions) short term room assignment in digital building operation can be optimized. This assignment can be performed up to a few minutes before the event is due to begin. Simulations have shown that 10% of current spaces are needless, when hybrid information management is applied. Precisely one out of 10 lecture halls is redundant and therefore was never needed to be built. In addition, the room’s air conditioning can be pre-adjusted according to the expected use. The aim of this climatic room optimization is to prevent energy intensive peaks and to create a situation of improved comfort. By categorizing and logically networking information an automatic control is made possible that is more efficient than controlled by humans. The semantic description of information delivers clear areas of combination for a correct process of analysis.

2. A real world installation with on site data handling. With the help of 35 wireless sensors a fine-grained temperature range in the entrance lobby of a university building at ETH Zurich was monitored over the period of one week (Schoch, 2006b). Sensor ID, time of measuring, temperature and brightness were recorded. In combination with the buildings event calendar the possible movement and distribution of people was estimated. The resulting scheme permits the successful estimation of local changes of room climate in the near future by simulations. The differentiation takes place by overlapping information at the relevant positions. The semantic description quickly leads to logical overlapping of points or areas. Points are to be understood in means of space and topic. Additional influences like thermal behaviour can be integrated into the scheme and serve as one further optimization. This practical test shows both optimization of resources and significant enhancement of comfort by knowledge mapping. Nevertheless, a permanent implementation of the system would acquire larger tolerance concerning reaction time and definition of comfort.

7. Outlook

Future steps will include the setup of a software-based framework within computer integrated spaces. The framework will integrate modular elements that allow the modification of single issues of information categorisation. In addition, a benchmarking system is to be applied in order to give qualified feedback on the performance of the system. The performance is evaluated primarily within the system and with real world data.

Acknowledgements

For the infrastructure and real world installations I would like to thank Professor Dr. Ludger Hovestadt, head of Chair of CAAD at ETH Zurich, Professor Gerd Folkers, Collegium Helveticum of ETH Zurich and Zurich University and the Villa Garbald Foundation.

Remark

At the point of time of writing this paper, the author is in no cooperation with Google™ Inc. or its subsidiaries apart from being a user of their services. The subtitle of the work is based on the background that the term “to google” is used generally as synonym of research with the help of the medium internet, for searching digitalized data provided by Google™ Inc. All Trademarks are respected by the author.

References

Achten, Henri H.: 1997, “Generic representations: an approach for modelling procedural and declarative knowledge of building types in architectural design”; PhD thesis, http://alexandria.tue.nl/extra2/9703788.pdf, p. 208.

PhDAgnolotti, V., Giger, Ch.: 2005, SIMGIS: GIS for Thermal Simulations in teh Built Environment, *in: Proceedings of 25th ESRI International User Conference*, San Diego, CA, 26-29 July 2005

Beaven, P.: 1981/1984, “Interactive data retrieval apparatus”, International Business Machines Corp., *US Pat. No. 4433392*

Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, *in Scientific American*, May 17, 2001

Berners-Lee, T.: 2003, Foreword, *in* Berners-Lee T. (ed.), Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., *The Semantic Web*, John Wiley & Sons, Chichester, 2003, p.1

Brin, S., Page, L.: 1998, http://infolab.stanford.edu/~backrub/google.html : December 2006).

Deleuze, G., Guattari, F.: 1977, Preface *in* Deleuze, G., Guattari, F., Massumi, B., *A Thousand Plateaus*, University of Minnesota Press, 1987

Gates, B.: 1990, http://www.microsoft.com/billgates/speeches/IndustryandTech/ iayf2005.aspx : July 2006

Lindstrom, E., Vitter, J., 1983, “External sorting using key value distribution and range formation”, International Business Machines Corporation, *US patent 4575798*

Papamichael, K.: 2003, The Role of Computers in the Building Life Cycle - What Computers Can and Cannot Do, *in: Proceedings of the 8th International Conference on Computer Aided Architectural Design Research in Asia*, Bangkok, p.913.

Schoch, O.: 2006a, Applying Digital Smart Technologies for a Sustainable Architecture, *in CAADRIA 2006, Proceedings of the 11th International Conference on Computer Aided Architectural Design Research in Asia*, Kumamoto (Japan), p.261-268.

Schoch, O.: 2006b, Energy Management in Buildings with Sensor Networks, David Garcés, Albert Krohn, Odilo Schoch, *in:* Kay Römer, Holger Karl, Friedemann Mattern (Eds.), *Wireless Sensor Network, Technical Report TR 507*, ftp://ftp.inf.ethz.ch/pub/ publications/ tech-reports/5xx/507.pdf

Sevtsuk, A.: 2006, Real-Time Space Management, *in* Pae, K., Ahas, R., Mark U. (eds.), *Joint space : open source on mobile positioning and urban studies*, Positium OÜ 2006, Tallin, p. 31ff.

Wang, K., Franklin, M.: 1997, Highly Accurate Data Value Prediction using Hybrid Predictors, *in Proceedings of the 30th annual ACM/IEEE international symposium on Microarchitecture*, Research Triangle Park , p. 281

Weissmann et al.: 1999/2002 “Meaning-based information organization and retrieval”, Applied Semantics, Inc., *US Patent 6453315*

Wical, K., Conklin, J.: 1996, “Learning system for classification of terminology”, Oracle Corporation, *US patent 6101515*