ON-TO-CAAD

Speculations towards the Knowledge contained in the Corpus of CAAD

ANAND BHATT, BOB MARTENS
CEPT University, Ahmedabad, India
Vienna University of Technology, Vienna, Austria
anand.bhatt@ab-a.net, b.martens@tuwien.ac.at

Abstract. The work presented here discusses an ongoing research in constructing a knowledge representation system founded on CAAD document repositories and investigating the network of concepts constructed by CAAD research. The approach does not rely on conceptual models or theories about CAAD research, rather a formal semantics amenable to decidable reasoning on the corpus of research in CAAD. A web-based demonstration application is illustrated in the paper in which we discuss two propositions about CAAD research discovered in the process of our investigations.

1. Introduction and background

The discipline of Computer-Aided Architectural Design (CAAD), which was once considered a specialist discipline within the domain of architectural education has been steadily gaining ground in the context of the wider discipline of Architectural Education and Research. CAAD is regarded as a discipline capable of transforming the futures of architectural design as it provides architects with the foundations of knowledge on which they can conceptualize possibilities, invent new techniques to deal with familiar problems, and discover innovative ways of practising architecture.

CAAD research has long been pre-occupied with the representation of knowledge in architectural thinking, especially through the design process, as seen in numerous models, knowledge formalisms and theories proposed in the past where the tangible objects in architecture have often been explored to a degree of satisfaction. However, it is the *axiological systems in architecture* which have evolved through architectural practice that remain a largely underdeveloped field of research. The process of knowledge acquisition in architecture is grounded in the dyad master-apprentice, which is characteristic of a practice in the humanities, where procedural knowledge and tacit (or implicit) knowing have traditionally occupied a central position.

We observe increasing investments in computing by mainstream architectural education and research institutions, as indicated by the rapidly increasing number of degree courses, research centers, international peer-reviewed conferences, and an international journal. CAAD educators, who were closer to the 'mother' discipline of computing rather than architecture *per se* have moved on to higher positions in academia affirming the potential of CAAD research. As a result, the native concerns of CAAD have been projected into the mainstream. This, however, demands the *formalized representations of boundaries of CAAD as a research discipline*, which were commonly understood by CAAD researchers. At first glance the Universe of Discourse (UoD) constructed by CAAD *per se* is not evident in the context of architectural education and research mainstream.

There is a perceived need in the community of CAAD researchers for a Knowledge Representation (KR) system, for reasons such as: (i) Computing applications are deployed in practically all aspects of architectural education and research with the topics of CAAD being scattered throughout the mainstream and infused with procedural concerns (ii) The research is conducted by individual scientists in relatively small research groups in scattered geo-spatial locations, necessitating environments for storage and retrieval of knowledge (iii) Experience of the formative years of the research is not recorded, and there are already observable gaps in institutional memories or current practice. (Maver 1995; Kvan 2004) (iv) CAAD educators lack standard textbooks and often need to consult the *Corpus of CAAD research* stored in multiple repositories.

KR would assist CAAD researchers in identifying pertinent areas; enable institutions in discovering newer areas of CAAD activity or reinterpret 'traditional' areas; leverage research funding by helping in analysis of research outcomes. KR can assist other educators in estimating how CAAD articulates their field of specialization. A KR system in CAAD could provide a template for KR in related fields in a process leading to upper ontologies.

2. Outline of the Project

In applying off-the-shelf machine learning and data mining techniques to the 3,000 records available in CumInCAD in the late 1990s, Turk et al (2001) found that "*the machine-made clusters* [of research documents] *are much different to the clusters created by the humans. It seems hard for a machine without the background deep knowledge to cluster a topic on its own and, in this way autonomously define what the topics of CAAD are*." Further, "*This* *confirms our belief that the way we understand the topics of CAAD, and to which this or that paper belongs, is subjective and based on one's current interests and perspectives. What defines a scientific community is, that its members to a large extent share a similar deep understanding of the topic*."

Following Turk's observations, we propose to construct a KR system with components characteristic of a KR formalism. These are: (i) a *Declarative semantics* which ought to be defined regardless of the procedural programmes that operate the knowledgebase (in our case, the corpus of CAAD research expressed by CAAD document repositories), (ii) an *"intelligent" retrieval mechanism* capable of deducing implicit knowledge contained in the corpus and (iii) a *structured representation of knowledge* (Baader 1997). It is also felt that (iv) a project making claims about the identity of a research discipline must evolve in a consultative manner. Given these characteristics, our strategy towards obtaining this formalism comprises four aspects:

- Evaluation: The *first aspect* evaluates the Corpus of CAAD and ascertains the needs of CAAD research and education community, with a view to discovering the knowledge-structures including the existent micromodels, microtheories and propositions about CAAD and design knowledge constructed in the last two decades of research, communication and publication.
- Acquisition and modelling: The *second aspect* of the project establishes Knowledge Acquisition techniques and knowledge modelling techniques by which the corpus can be processed.
- Representation and retrieval: The *third aspect* consists of sets of procedures for extracting knowledge from the Corpus of CAAD Research documentation and capturing organizational memory through expert participation. Means of retrieval of knowledge by the community of researchers, including latent semantic searches are also planned.
- Refinement: The *fourth aspect* presents the outcome of the research to experts and various user groups for evaluation, enrichment and testing. A Simple Knowledge Organisation System (SKOS) encapsulating the concepts entailed in CAAD research will be refined by experts at this stage.

Our initial tentative incursions about the first three aspects are discussed below. Section 3 discusses our approach towards knowledge acquisition and modelling, and Section 4 discusses our strategies in representing and excavating the Corpus of CAAD Research. The fourth aspect is not discussed at this early stage.

3. Towards the Knowledgebase: CumInCAD and CAADRIA 2005 databases in OWL-DL

An Open Access database related to CAAD research output is available through the Cumulative Index of CAAD (CumInCAD - cumincad.scix.net) since 1998. A user can browse through almost 5,000 full texts, which are linked to 7,500 citation index records. CumInCAD is constructed with SciX Open Publishing Services (SOPS) which is an Open Archives Initiative (OAI)-compliant digital library system (Clifton 2003). Some simple Knowledge Management (KM) capabilities can be built on top of the SOPS architecture, allowing for summarization and similarity-based knowledge extraction.

SOPS is built around a dozen concepts which are connected in different ways following the SciX objectives, demonstrating a business model for scientific journal and conference proceedings publication process.

A central concept in the SciX model is an article (see Figure 1). The model is compatible with Dublin Core Metadata (DCMI). Various user actions are associated with the article, both as part of the publication/review workflow in ascertaining the quality level of the article and in a community of practice situation where articles may be annotated and discussed.

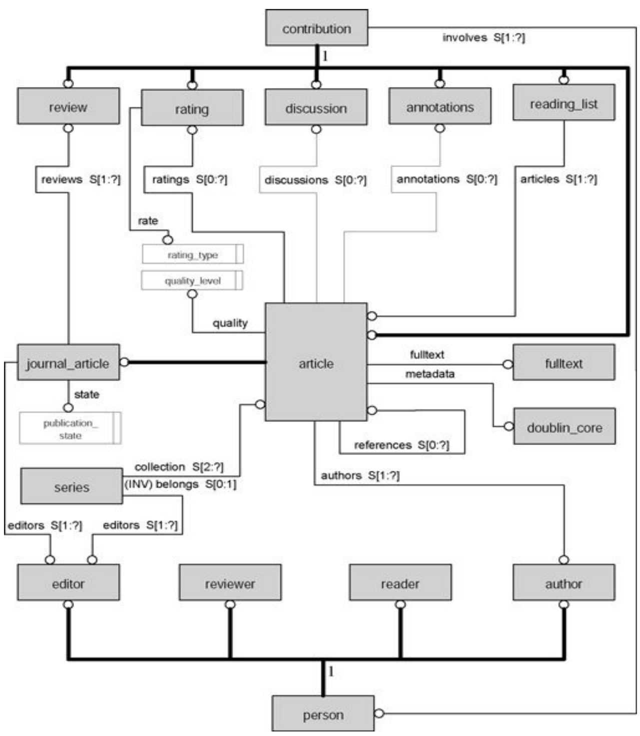

Figure 1. Abstract diagram of concepts supported by the SciX architecture (Clifton, 2003).

3.1 NORMALIZATION OF THE CUMINCAD DATABASE

The first aim in constructing the proposed knowledgebase is knowledge normalization, i.e. converting the data stored in the CumInCAD database into represented knowledge about CAAD research. Given the scope of the project at this stage, the focus is placed only on the terminological knowledge contained in the database.

Following Baader (1997), we first normalized the available datasets in such a way that not only were the connections of concepts constituting the database made explicit, meaning(s) of the connections were represented in an explicit way. This type of connection can be represented by Horn clauses:

directly-connected $(x,y) \rightarrow$ connected (x,y),
directly-connected (x,z) ^ connected (z,y) → connected (x,y).

In addition to the KR modeling requirements a pragmatic constraint in the knowledge normalization process is that the knowledge retrieval process must rely on the syntactic form of entities stored in CumInCAD (as represented in Figure 1), without drastically changing the behaviour of SOPS. Given the OWL-based Web Ontology Language for services (OWL-S) there are many ways of bridging the database and the knowledgebase, using web services technologies and semantic web querying languages such as SPARQL. Programming a loosely coupled architecture between databases and documents stored online, and a knowledgebase is rather straightforward.

SOPS also stores Dublin Core metadata about each record in the database; CumInCAD data structures comply with Dublin Core standard where appropriate. It is also possible to represent the Dublin Core vocabulary as a micromodel of the knowledge contained in CumInCAD. Figure 2 expresses the meaning of relationships between the various database entities in CumInCAD as a DCMI micromodel.

Bridging through the DCMI micromodel is significant in another direction, as SOPS - following the SciX project goals - contains a series of selections concerning the basic logic of storing, classifying, connecting and retrieving scientific research literature. CumInCAD makes further selections within the SciX framework, and as such, provides the user with a series of domain-specific assertions about the taxonomy of CAAD research documentation. For example, CumInCAD only stores information about education level of the user which can be expressed in DCMI, so that the relationship of the actors in CAAD research is expressed only in terms of academic qualification.

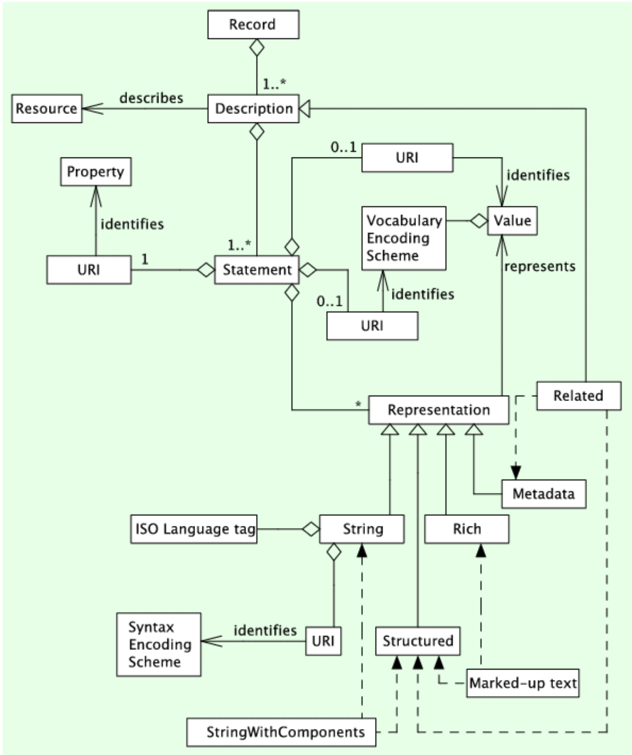

Figure 2. Dublin Core standard abstract model.

3.2 CONVERTING CUMINCAD INTO A KNOWLEDGE REPRESENTATION

Web Ontology Language, OWL-DL is chosen as a markup language so as to express CumInCAD in a description logic (DL) which is amenable to decidable reasoning (Buchheit et al 1993). OWL-DL supports those users who want maximum expressiveness in an ontology while retaining computational completeness, i.e., where all conclusions are guaranteed to be computable; and decidability, i.e., all computations will finish in finite time. OWL, which is a language meant for open-world computing, is chosen as the project is conducted in a consultative/collaborative framework.

The conversion was performed in several stages. In the first stage we recast the CumInCAD database structure in Resource Description Framework (RDF) while making explicit the structure of CumInCAD database and its representation in DCMI shown in Figure 3.

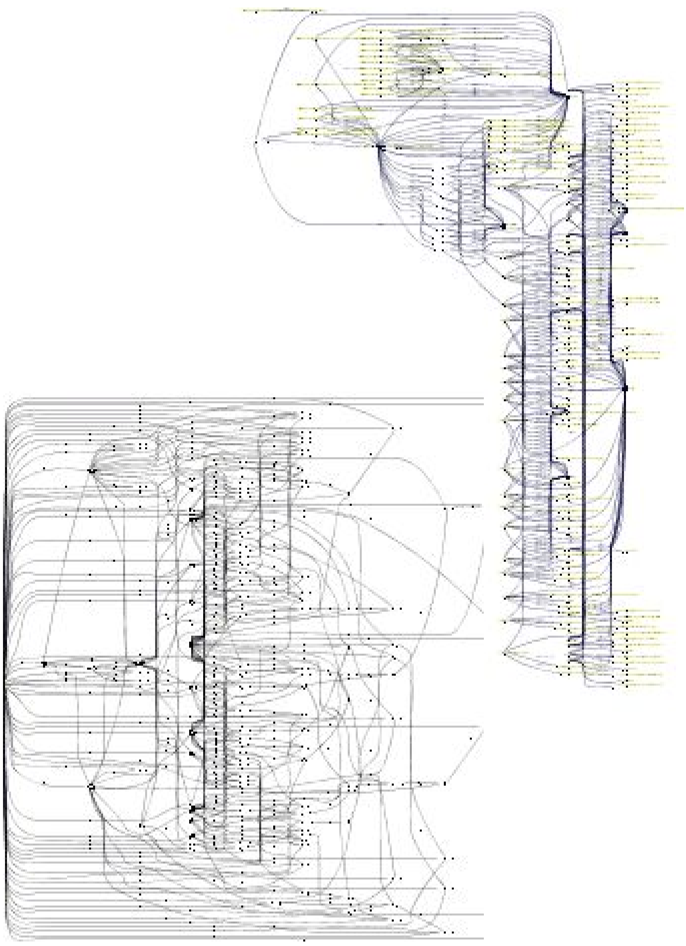

Figure 3. Two views of CumInCAD database structure in DCMI with CAADRIA 2005 data

This model is based on the equivalence relations amongst the database structure and the structure of its descriptors. Dublin Core is represented in OWL-DL as annotation properties, which are disjointed from ontology properties for semantic reasons. While revealing the meanings of the connectedness asserted in CumInCAD, the annotation properties serve at once as the referent for constructing frames necessitated by the Knowledge Representation System, and as a disjoint by which it becomes possible to normalize assertions by other CAAD repositories and micromodels in the knowledgebase while maintaining CumInCAD assertions.

In the second stage, the RDF data was ported into an OWL-DL ontology. Relevant ontology properties were constructed and related to the annotation properties provided by CumInCAD. The disjoint was exploited further to compensate for the uneven quality of metadata supplied by authors and the specific assertions made in CumInCAD database structures.

3.3. COMPENSATING FOR CUMINCAD ASSERTIONS

Assertions contained in CumInCAD are specific to SciX research objectives, and there exist other database structures (Lin 2004), such as the CAADRIA paper review system (caadria-paper.org). The OWL-DL is augmented with CAADRIA 2005 data by following a procedure similar to the CumInCAD conversion. This augmentation had several advantages. Information about rejected papers is stored in the CAADRIA databases showing the propensities of the review process.

Variations in quality of the user-supplied metadata were observed within CumInCAD as metadata is sourced from multiple origins, and papers are tagged from differing standpoints over a large period of time. Metadata tagging habits typical of scientific literature authors were accounted for. A simple SKOS thesaurus was constructed by analyzing the sentence and phrase structures from the annotation properties pertaining to the DCMI. The terms were extracted using quantitative/statistical methods and related using generic similarity ranking software. Thus, the SKOS does not pertain to a UoD unique to CAAD at this stage, and needs further refinement.

This representation is written in RDF equivalent Notation-3 and can be accessed at Architexturez Subject Gateway (architexturez.in), where RDF/Schema inferencing with reasoning in OWL-DLP and Querying in standard query languages is possible.

4. Exploring the Knowledgebase

Architecture is not an exact science, and it is significant to explore the disjoints created by micromodel(s) and theories in CAAD. The objects of knowledge often need to be studied from the standpoint of a certain school of thought, opinion or bias. Micromodels concerning architectural data gain special significance as they affect a number of tacit and inter-subjective understandings characteristic of a condition where axiological systems are dominant. These go on to create the classification-and-inferencing procedures in a knowledge-programming environment where microtheories in CAAD represent local evolutions in space and time, and distinctions across the system of ideas in CAAD thinking. These are significant, given the molecular 'scattering' of CAAD research locations. Our research intends to evolve rather more organic representations of knowledge in CAAD as they provide an alternative to a humongous KR.

What seems to be the structure of CAAD as a discipline? Our initial merging of two sets of assertions, while populated with a fraction of data, already shows that topics and generic agents researching CAAD are quite evenly distributed but in a wide and shallow pattern. Some concepts have a greater density around them, as seen in Figure 4. A horizontal bar diagram suggests a structure where CAAD spans the gamut of architecture research, connecting architecture activity at multitudinal levels whereby topics in conventional architecture are related in new ways by CAAD research.

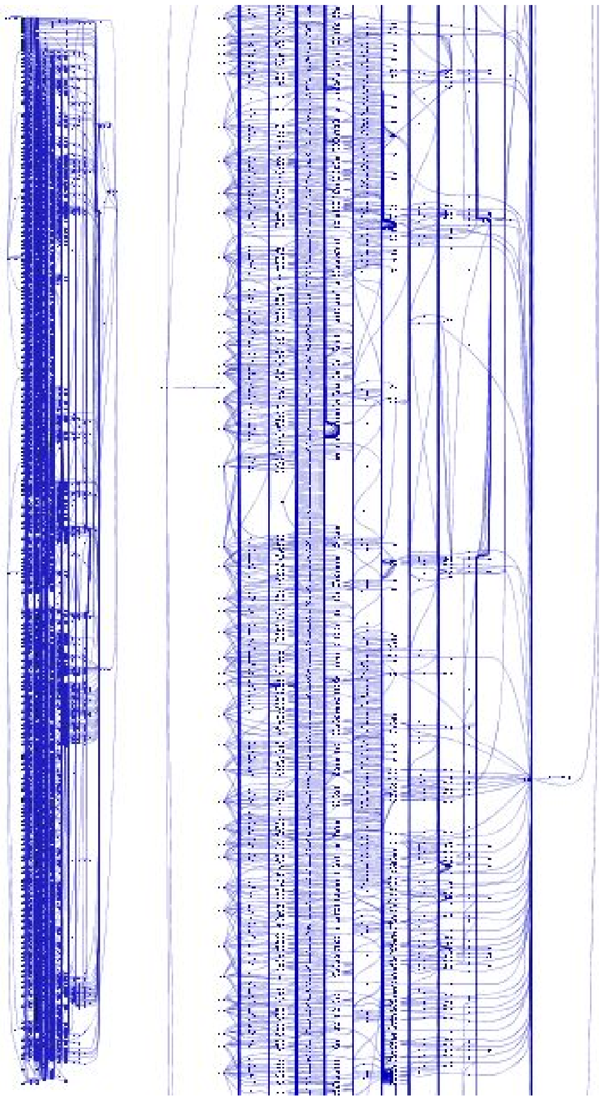

Figure 4. Two views of the augmented Database, with a bird's-eye view on left

Further research is warranted in the boundary condition of CAAD as an academic discipline, to check if it is structurally coupled with conventional architectural practices and in exploring the Universe of Discourse constituted by both. There are indications that some generic agents in CAAD research extend intuitions, insights or micro theories which are significant enough to affect the taxonomy. Several of our queries show that some established agents are classed not only in the context of other generic agents such as researchers and institutions but also *as concepts*. That is, certain concepts are further identified "in the sense of this agent". This is also expected in a field which is being invented, so particular agents are associated with particular concepts in a specific sense.

The question, "What are the main interpretations or self-understandings of CAAD as a discipline occurring in the corpus of CAAD research?" becomes interesting in this regard. Apart from evident pedagogical value, such a question leads to a characterization of the UoD in CAAD, which in itself is a subject of further research.

We have extended several research claims throughout this paper which need to be evaluated by the community. Our preferred focus is on claims concerning the combination of multiple structures of assertions on problems concerning the Universe of Discourse characteristic of CAAD research. Evaluation at this stage will assist us in definition of the boundaries of CAAD research while keeping the descriptors of the internal structures of CAAD research fluid.

Our approach leaves questions concerning the positive, "what is", identity of CAAD open to interpretation. We believe this is a necessary condition for any open, collaborative enterprise such as ours, and it is possible to obtain this condition by developing the approaches outlined in this paper.

The project at this stage limits itself to the extraction, formalization and management of terminological knowledge contained within CAAD Research document repositories, and it is programmed amenable to decidable reasoning. This will change as partial, subjective and time-dependent knowledge is built into the formalization.

References

Baader, F.: 1997, Logic Based Knowledge Representation, *Artificial Intelligence Today, Recent Trends and Developments,* Springer Verlag, pp. 13-41.

Buchheit, M., Donini, F. and Schaerf, A.: 1993, *Decidable Reasoning in Terminological Knowledge Representation systems*, Research Report RR-93-10, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany.

Clifton, B., Cooper, G.: 2003, *D9a: Architecture Synthesis*, available at http://www.scix.net.

Lin C.-J. and Chiu, M.-L.: 2004, Design Knowledge Discovery in Cases - The Machine View vs. the Human View, *CAADRIA 2004*, Seoul, pp. 265-274.

Maver, T. W.: 1995, CAAD's Seven Deadly Sins, *Sixth International Conference on Computer-Aided Architectural Design Futures*, Singapore, pp. 21-22.

Kvan, T.: 2004, The Dual Heritage of CAAD Research, *International Journal of Architectural Computing* **2**(1), pp. 11-17.

Turk, Z., Cerovsek, T. and Martens, B.: 2001, The Topics of CAAD. A Machine's Perspective, *Proceedings of the Ninth International Conference on Computer Aided Architectural Design Futures*, Eindhoven, pp. 547-560.