Studying Corpus Changes in CumInCAD

Anand Bhatt [1], AVV Kishore [2]
ABA-NET/Archtiexturez Imprints, Kiami
[1] http://www.ab-a.net, [2] http://www.kiami.com
[1] anand.bhatt@ab-a.net, [2] kishore@kiami.com

Abstract. *We discuss several experiments conducted with the Corpus of CAAD research, where we focus on the change in the CAAD Ontology. These experiments are representative of the investigations which could be conducted through an ontology driven, online application designed to allow the research community investigate the nature, structure and evolution of our discipline.*
Keywords: *Knowledge Modelling, Modelling View, Semantics and Change, CAAD Ontology, Representation.*

Outline

This paper extends our previous experiments in structuring the corpus of CAAD, where we have added a set of formal semantics to the corpus by learning explicit semantics from the corpus stored in CumInCAD [cumincad.scix.net]. These experiments have generated domain ontologies from textual resources by applying natural language processing and machine learning techniques and extended the modeling view [Clancey, 1993] of the corpus to articulate questions raised in previous research, namely, (i) as to how the CAAD research community's self-understanding of their discipline has transformed over time; and (ii) concerning the tacit, implicit or 'imprecise' nature of Architectural knowledge.

We use strategies in ontology learning and data-driven change detection to re-address the questions encountered in our previous research. In particular, we focus on two types of change: structure-change in the CAAD ontology created previously, and data-driven or 'bottom-up' inflexions inducing changes in the structure of the ontology over time. Data-driven changes help us capture the various states the corpus of CAAD research has undergone through its his lifecycle, and provide measures of changes that have occurred. A third type of change, usage (or user)-driven change, allowing for the influence of human agents, and how they create affects capable of influencing the structures of knowledge in CAAD is not studied at the moment.

Precedence

As the corpus has grown considerably in size and complexity, the need for computational support in representing and analyzing the corpus is observed. We have already seen several experiments where CAAD researchers have analyzed the significance of concepts and emergent structures reflexive of the community's understandings of the field.

Turk (2001) attempted to organize the CumInCAD corpus using statistical methods; Chiu (2002) used ontology-assisted data mining techniques to cluster documents; Cerovsek (2004), while focusing on the networked scholarly discourse aspects in CAAD, used citation indexing and other techniques to discover the life-cycle of concepts; Bhatt and

Martens (2005) modeled the CumInCAD data structures in description logics using extent model-theoretic semantics available for Dublin Core specification and Web Ontology Language (OWL). However, all of these experiments remain somewhat limited as they rely heavily on asserted metadata about the documents contained in the library, and they do not rely on standards for composing interoperable text and multimodal analytics.

Methodology

Our research questions require experiments with scholarly sense-making of CAAD as a field of study. Following Boland (1995), this type of experimentation necessarily requires "tools to track ideas and results in a field, and to express, analyze and contest their significance". Firstly, the process of structuring the corpus as it is processed by applications in a semi-automated manner at higher levels of detail. The corpus was analyzed to assign application-specific semantics to the existing structured representations. This required an explicit re-factoring of ontologies to express them as core ontologies, or structures, at a formal level.

Secondly, out of methodological aspects of change discovery at a semantic level, aiming at generating implicit requirements by inducing ontology changes from existing data. This required a complete re-engineering of the interface between at CumInCAD and the knowledgebase. The new interface represents the acquired ontological structures at a meta-level, in the form of modeling primitives, before storing them in the formal description logics which are submitted to specific applications for analysis.

Structuring the Corpus

Following Gruber (2008) have previously defined ontologies as "explicit specifications of conceptualization processes", and worked on an ontological level as described by Guarino, (1995); where "a KR formalism is constrained in such a way that its intended models are made explicit", as the goal is, "to restrict the number of possible interpretations, characterizing the meaning of the basic ontological categories used to describe the domain." These definitions in the past required us to undergo processes entailing high levels of generalization, and required us to constrain ourselves to application-specific semantics, as demanded by the research questions at hand.

However this is an ontology learning process, focusing on change-capture and measurement, and it is inherent in such processes that the acquired ontologies represent 'imprecise', uncertain and possibly contradicting knowledge. Identification, and storing of such knowledge is vital to our understanding of CAAD as a discipline, and for this reason, an acquisition pipeline is created incorporating modelling primitives. Following Stumme (2001), this is built around a core definition of 'ontology as a structure':

$$C: = (C, < C, R, < R, \sigma);$$

Where, C: set of concept identifiers, R: set of relation identifiers, <C partial order on C (concept hierarchy), <R: partial order on R (relation hierarchy), Signature: σ: R $\rightarrow$ C +, Mathematical definition of extension of concepts [c] and relations [r], L-Axiom System: α (disjoint (c, c')) = $\forall e \in [c] \rightarrow e' \notin [c']$.

Different extensions are then introduced over this definition, to progressively expand the core definition to logical languages and lexicons. The introduction of this layer allows us to adapt earlier, application-specific ontologies such as the CAAD ontology version 2, stored in a description-logics conformant OWL-DL; and the CAAD lexicon stored in a SKOS thesaurus [Figure-1] to the newer tasks at hand. These tasks comprise of discovering changes at a semantic level in the CAAD Corpus, for specific types of conditions, of storing discovered changes in ontologies, and reconciling the discovered knowledge with knowledge already stored in the system per se.

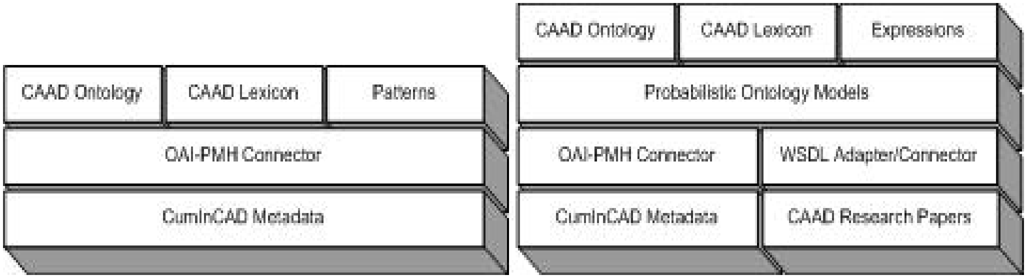


Figure 1
CAAD Ontology, version 2.0 and 3.0

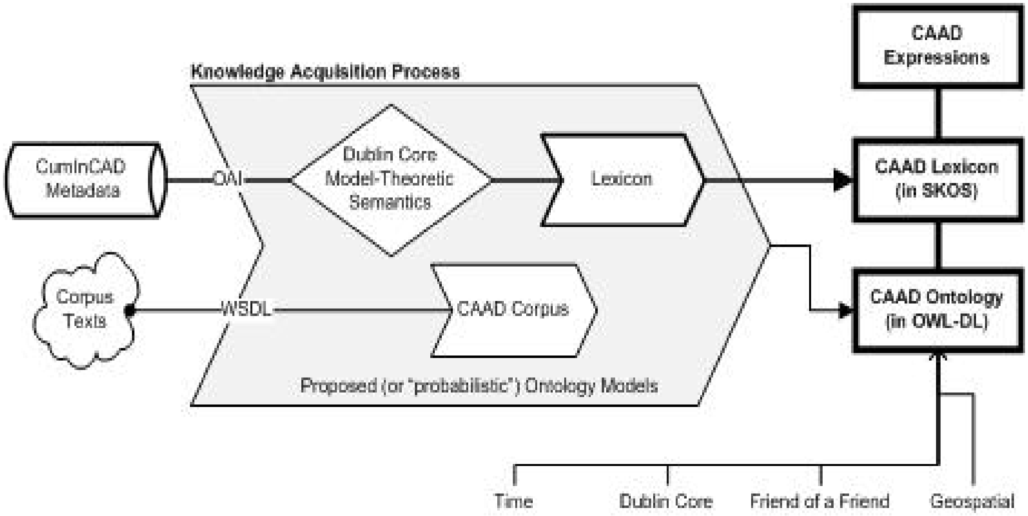


Figure 2
The Knowledge Acquisition Pipeline

Change Discovery

Change Discovery is not change detection or change capture. Change detection can be accomplished by asynchronously assimilating knowledge from multimodal sources; change capture will measure changes against a pre-existing set of rules or axioms. The both approaches do not suffice for our tasks, as we are focused on knowledge as it is construed by a certain process, or a certain aspect of CAAD activity. The formal representation of knowledge must arise at the very moment we encounter new sets of information. Figure-2 shows the new artifacts added to facilitate the experiments.

Acquisition and Storage:

The knowledge acquisition pipeline is constructed around a change discovery mechanism (Text2Onto, ontoware.org/projects/text2onto), which provides a Modelling Primitives Library (MPL) allowing for the definition of modelling primitives in a declarative fashion. Collections of these modelling primitives, stored in the knowledge-acquisition pipeline in our system result in Probabilistic Ontology Models (POM). The POM are not probabilistic in a mathematical sense, but rather, collections of relations typical of an ontology which are instantiated in the pipeline, and are all primitives with an assigned value indicating the confidence/score for a given primitive.

Controller	Concept	Instance Of	Domain Of	Range Of
Concept Extraction	0.0			Process
Pattern Extraction	0.001639344262295082			
Controller, Conference				
Concept, eCAADe 01	0.0	2.1094821221390148E-4	Pedagogy	Process
Concept, eCAADe 07	1.0		Architectural Pedagogy	

Table 1
Concept | Pedagogy | in four different POM.

Text2Onto provides several sets of algorithms which can be combined to produce different types of POM. These algorithms are initialized by a controller whose purpose is to execute the algorithms in a specific order [Cimiano and Volker, 2005]. The concrete knowledge stored in the acquisition pipeline depends on the way these algorithms are invoked in the pipeline.

Table-1 shows the concept | Pedagogy | in several data acquisition experiments, with different algorithms and different datasets. Here a concept extraction algorithm, stores the text references (Pedagogy is a self-referential concept for eCAADe 2001 corpus); whereas a pattern extractor stores a score based on occurrences of the pattern. The differing ('situated') interpretations are stored as separate models in a RDF database, where they are invoked by dispatchers or a special type of reasoners that do not perform any reasoning by themselves, but route goals to other reasoners.

Structure-Driven Changes:

Ontologies are structure storing information, and structure changes normally occur when (i) the modeled domain has changed, (ii) when the perspective from which the model is viewed or (iii) when we discover flaws in the way the domain is modeled.

Figure-3 shows the 30 dominant concepts identified as central concepts in CAAD [Turk (2001)] for two conferences, eCAADe 2001 and 2007. These concepts form the core of the CAAD Lexicon ontology, which is expressed in an SKOS Core vocabulary. In an application-specific system, the terms will always appear as an invariable network, with the asserted relations and rules always controlling the reasoning process. Here, as we acquire several POM for each dataset, and dispatch to reasoners, the CAAD Lexicon undergoes a structural transformation for each

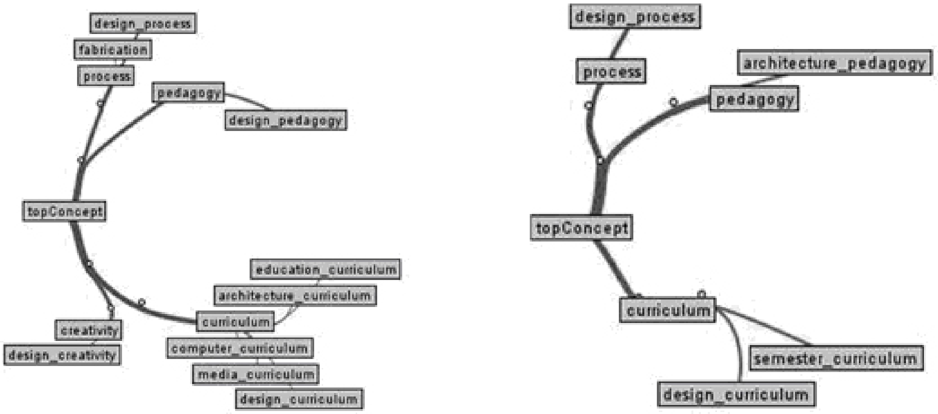
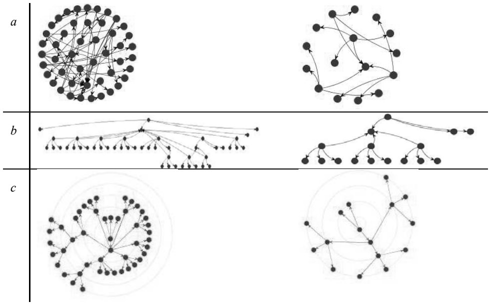


Figure 3
CAAD Lexicon, changes in structure between eCAADe 2001 and 2007

Figure 4
Minimum spanning trees showing concept | Virtual Reality | in eCAADe 2001 and 2007



given dataset. Especially the arbitrary L-Axiom systems, and mathematical extensions of concepts will change in a dramatic manner by relying on scores illustrated in Table-1. For example, word-concept | Pedagogy | has undergone several changes, (i) whereas in 2001, it was a self-reflexive concept, and domain of Design Pedagogy in 2007 it is the domain of Architectural Pedagogy, (ii) whereas in 2001, it was in the range of Process, in 2007, it is subsumed by a concept inheritance, and acquires the properties associated with the concept | Architectural Pedagogy |, or (iii) specialization of the concept may have to do with CAAD researchers revising, re-modelling or reviewing the concept.

Data-Driven Changes:
Data-driven change discovery provides methods for automated or semi-automatic adoption of ontology according to modifications being applied to the underlying data set, and transform the OWL models so as to map the events in a networked, scholarly, discourse contained in the corpus. Data-driven changes are measured across boundaries, and these boundaries can be synchronous (across specific conditions in knowledge, for example, or representing certain processes in knowledge creation) or diachronous. MPL are generated for any given state under observation, with typical modelling primitives such as concepts, concept inheritance or instantiation, property-relations, domain and range restrictions, metrological relations and equivalence. Specific instantiations of the Modelling Primitives Library are sufficient in themselves for certain types of operations.

Figure-4 shows minimum spanning trees starting with concept | Virtual Reality | in eCAADe 2001 and 2007 conferences. A directed subgraph of all concepts related to virtual reality is extracted using the POM generated at runtime (level a). The subgraph is then expanded to the required levels, and induced into an undirected graph (level b). This graph is processed to yield the minimum spanning tree (level c). The tree can then be used to determine various kinds of cluster structures.

Observations

The experiments are grounded in previous research, on CAAD, as a branch of architectural studies per se, and conservation architecture, a largely axiological pursuit where modeling architectural knowledge presents us with very special problems in both cases. Unlike problem encountered normative knowledge acquisition processes, which may deal with structures of common knowledge or well-formed terminological knowledgebases, the structures within Architectural Knowledge systems remain rather fluid. Architectural knowledge is constructed in a series of engagements between the deterministic order of things, and the pure experience of order, it is situated in a series of what Michel Foucault terms as the middle practices. They remain in a tacit, tactical way: left alone, would any machine, from a machine's perspective, reconcile our discipline, | Computer-Aided Architectural Design | with the art of making buildings?

Previous researches have noted the limitations of standard methods in clustering the topics of CAAD (Turk, 2001 and Cerovesk 2004). Our context dependent subgraphs illustrating various states indicate this may well have to do with the way topics are framed in the corpus of CAAD.

Future Work

An online infrastructure for making many more types of experiments is deployed at www.architexturez.in. CAAD researchers can investigate the corpus of CAAD as-a-whole. The experiments already set-up are not limited to concepts (or topics) of CAAD, they also expand the research already demonstrated by Cerovsek [2004] and Chiu [2002], insomuch as they allow investigation of semantic networks, and the social-semantic aspects of our discipline. Apart from time-dependent (or diachronic) views, it is possible to construct structure-dependent and synchronic views of the discipline.

In the next stages, we aim to conduct more types of experiments, at a larger scale, and set-up new experiments, as required by the CAAD research community. The overall objective of this, future research, is in providing tools to aid in comprehension CAAD research as a large-span, asynchronous discourse. Where we also intend to evaluate the research claims put forth by CAAD research, and generate very high-level views of the discipline analyzed as a unified social-semantic graph allowing for studies of affect, or usage-driven changes in the structure of CAAD research.

References

Boland, R. J. J. and Tenkasi, R. V.: 1995, Perspective making and perspective taking in communities of knowing, Organization Science, 6(4), pp. 350-372.

Bhatt, A.: 2007, Modeling Conventional Architectural Processes: Maintaining knowledge created by institutions situated in a historic context in CAADRIA 2008, Proceedings of the 13th International Conference on Computer Aided Architectural Design Research in Asia, Forthcoming.

Bhatt, A. and Martens, B.: 2006, ON-TO-CAAD: Investigating the Knowledge Contained within the Corpus of CAAD Research, in CAADRIA 2006, Proceedings of the 11th International Conference on Computer Aided Architectural Design Research in Asia, Kumamoto (Japan), March 30th - April 2nd 2006, pp. 413-424.

Cerovsek, T. and Martens, B.: 2004, On the Extended Use of Citations in CAAD, in J. P. Van Leeuwen and H.J.P. Timmermans (eds.), Developments in Design & Decision Support Systems in Architecture and Urban Planning, Eindhoven University of Technology, Eindhoven, pp. 1-17.

Chiu, M. L., Lin, C. J., Jeng, T. S. and Lee, C. H.: 2002, Re-Searching The Research Problems with CAAD: Datamining in i-CAADRIA, in CAADRIA 2002, Proceedings of the 7th International Conference on Computer Aided Architectural Design Research in Asia, Cyberjaya (Malaysia), 18–20, pp. 31-38.

Cimiano, P. and Voelker, J.: 2005, Text2onto - a framework for ontology learning and data-driven change discovery, Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB'2005), june 2005.

Clancey, W. J.: 1993, The Knowledge Level Reinterpreted: Modelling Socio-Technical Systems. International Journal of Intelligent Systems, 8, pp. 33-49.

Gruber, T. R.: 2008, Ontologies, in Ling Liu and M. Tamer Özsu (eds.), The Encyclopedia of Database Systems, Springer-Verlag.

Guarino, N. and Poli, R. (eds.): 1995, Formal Ontology in Conceptual Analysis and Knowledge Representation. Special issue of the International Journal of Human and Computer Studies, 43(5-6), Academic Press.

Stumme, G.: 2001, Using ontologies and formal concept analysis for organizing business knowledge, in Proc. Referenzmodellierung.

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.