

Seamless Structured Knowledge Acquisition

Päivikki Parpola

Helsinki University of Technology, Finland

INTRODUCTION

Some parts of this text, namely “Co-operative Building, Adaptation, and Evolution of Abstract Models of a KB” and most subsections in “Performing Reasoning in SOOKAT According to a KB”, have appeared in an article (DOI:10.1007/s10115-004-0181-6) published in the ‘Knowledge And Information Systems’ journal (Parpola, 2004).

A knowledge base (KB) contains data and instructions for using it (e.g., as a rule base). A KB containing knowledge possessed by experts can be used in an expert system. It can solve problems requiring expert knowledge, explain its decisions and deal with uncertainty. An expert system can be used as a basis for a larger system, called a knowledge-based system (KBS).

Knowledge acquisition (KA) that is the development and maintenance of KBs, (e.g. an expert system), can be divided into several phases, performed sequentially and iteratively. Some phases may be performed in parallel with other phases. The most commonly recognised phases are requirements definition, analysis, design, and implementation.

Disintegration, or the gap between phases of development, especially between abstract and executable descriptions, was recognised during the early stages of KA (Marcus, 1988a; Motta, Rajan and Eisenstadt, 1988). It complicates the development of KBs and hinders traceability between parts of abstract and executable descriptions. •

BACKGROUND

Seamless Structured Knowledge Acquisition (SeSKA) (Parpola, 1998; Parpola, 1999a; Parpola, 1999b; Parpola, 2000) is a methodology for as well the development and maintenance of KBs as performing reasoning in them. It is designed to enhance integration of the KA process.

During KB construction, a series of models, including the (combined) dependency graph, the domain

model, the inference model, together with analysis, design and implementation descriptions, is created and possibly modified. The structure of the knowledge base is based on the logical structure of the domain which has been noticed to be more stable than the component structure (Jacobson et al., 1992).

Related work concerning the use of metaobjects and metalevels in KA includes the following:

- Protégé-2000 (Fridman Noy, Ferguson and Musen, 2000) uses a metaobject protocol (Steele, 1990; Kiczales, des Riviers and Bobrow, 1991) to describe a model, for example, the CommonKADS model of expertise (Schreiber, Crubézy and Musen, 2000). This allows applications to be presented as instantiations of the model.
- OIL (Ontology Inference Language) (Fensel, van Harmelen, Decker, Erdmann and Klein, 2000) is a proposal, based on OKBC (Open Knowledge Base Connectivity), XOL (Ontology Exchange Language), and RDF (Resource Description Framework), for a joint standard for specifying and exchanging ontologies over the Internet. Modelling ontologies in OIL distinguishes three separate layers, the object level the first metalevel, and the second metalevel. The structure consists of several components. Rule bases, classes and slots, and types, as well as slot constraints and inheritance, are used. OIL is a frame-based system, using, for example, rule bases.

MAIN FOCUS OF THE CHAPTER

Models used in SeSKA

A domain model (DM) contains a somewhat stable componential structure of a domain. Knowledge is described through a network of relations between domain or abstract concepts with attributes. These attributes in the DM are selected according to what is needed in the dependency graph (DG).

Initial dependency graphs (DG) are acquired from different sources. DGs present inferential dependencies between attributes of DM concepts. Descriptions can be attached to dependencies. The actual DG is a combination of initial DGs. A DG contains dynamic knowledge described through a network of concept attributes and dependencies.

An inference structure (IS) presents the structure of possible inference sequences performed. The IS is shared among three sets of descriptions. Collections of analysis, design and implementation descriptions are attached to inferences in the IS. The result is called the inference model (IM).

These models can be described in terms of ontologies and natural language analysis (Parpola, 2000). The DM and the DG can be produced using several different KA techniques (Parpola, 1999b). Heterogeneous vocabulary can also be harmonised. The domain model and inference model can be instantiated to form the value model and execution model. This enables performing inferences.

Managing Change through Seamless Transformations

When constructing a KB with SeSKA, integration of a structured set of models can be produced through seamless transformations that is predefined ways of getting from objects in one model to objects in another model (Jacobson et al., 1992; Parpola, 1998). The KB structure is also maintained using the constructed shared skeleton: The inference structure (IS) describes the structure of possible inference sequences through a network of roles and inference steps. The former refer to concept attributes, and the latter have attached analysis, design, and implementation descriptions:

- The major logical components of abstract descriptions,
- Their formal descriptions, and
- Executable rules or functions, respectively.

The collections of different descriptions of all inference steps, in combination with the inference structure, form the analysis, design, and implementation models. The possibility of performing inferences, described in the models, requires instantiation of domain and inference models.

The idea of being able to describe a KB via models is proposed in the SeSKA methodology and implemented in the SOOKAT tool, described later.

The Knowledge Base Construction Process

Initial Formation of the Models Describing the Domain

The initial domain model and dependency graphs are formed on the basis of default value suggestions for, and dependency suggestions between, concept attributes, acquired from several knowledge sources that may give differing values.

Combining Dependencies and Attribute Values

Complementary dependency graphs can be processed using joining and simplification rules (Parpola, 1998). These rules allow different fragments of knowledge to be brought together, even before building a KB, and make it possible to show how they might be combined. Combination rules may accelerate the construction of a KB. To cope with contradictory or multiple attribute values, SeSKA defines combination heuristics (Parpola, 1999b).

Network of Roles and Inferences

A role in the IS is formed of concept attributes that a certain attribute depends on. Inferences between roles and descriptions associated with the inferences can be created on the basis of dependencies between attributes.

Several different dependency graphs can produce the same analysis model. One way to form such a dependency graph is to take the roles connected by an inference and set all the concepts referenced by the conclusion role to depend on all concepts referenced by the premise role. The analysis description of the inference can be attached to all dependencies.

Analysis descriptions are formalised to implementation descriptions, possibly via semi-formal design descriptions. The process is iterative and modular.

Managing Change through Seamless Transformations

Often, a need for change is acknowledged during the development or maintenance of a KB through imple-

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/seamless-structured-knowledge-acquisition/11050

Related Content

Multi-Group Data Classification via MILP

Fadime Üney Yüksektepe (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1365-1371).

www.irma-international.org/chapter/multi-group-data-classification-via/10999

Participatory Literacy and Taking Informed Action in the Social Studies

Casey Holmes and Meghan McGlinn Manfra (2020). *Participatory Literacy Practices for P-12 Classrooms in the Digital Age* (pp. 40-56).

www.irma-international.org/chapter/participatory-literacy-and-taking-informed-action-in-the-social-studies/237412

Data Cube Compression Techniques: A Theoretical Review

Alfredo Cuzzocrea (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 367-373).

www.irma-international.org/chapter/data-cube-compression-techniques/10846

Data Mining in Genome Wide Association Studies

Tom Burr (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 465-471).

www.irma-international.org/chapter/data-mining-genome-wide-association/10861

Spatio-Temporal Data Mining for Air Pollution Problems

Seoung Bum Kim, Chivalai Temiyasathit, Sun-Kyoung Park and Victoria C.P. Chen (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1815-1822).

www.irma-international.org/chapter/spatio-temporal-data-mining-air/11065