

Re-usable Knowledge: Development of an Object Oriented Industrial KBS and a Collaborative Domain Ontology

Paul Crowther¹, Gerd Berner² and Raymond Williams²

¹School of Computing and Management Sciences, Sheffield Hallam University, Howard St, Sheffield S1 1WB, UK.

e-mail: P.Crowther@shu.ac.uk

²School of Computing, University of Tasmania, PO Box 1214 Launceston, Tasmania 7250, Australia.

e-mail: R.Williams@utas.edu.au

Abstract. Since the early eighties, there has been a gradual shift in the focus of development of knowledge based systems away from the rapid prototyping techniques that had previously prevailed, toward more structured methodologies, including model based reasoning and modeling of knowledge domains. The default standard for the development of these systems has become the CommonKADS methodology. This paper will assess the feasibility of applying the CommonKADS methodology to the knowledge base of an existing legacy system, the subsequent re-useability of knowledge and domain schema that results from the process and the development of collaborative domain ontologies. Currently the CommonKADS methodology embraces reuse by providing standard inferencing primitives and a set of generic task models. The resulting model set is assessed to determine its suitability to form the basis of an ontology for electroplating in the manufacturing industry.

1. Introduction

This study investigates the benefits and feasibility of using the CommonKADS methodology for reengineering an existing knowledge base which was developed using rapid prototyping without a formal methodology for structuring it. Knowledge bases of this nature are described by Wielinga and Schreiber [23] as often being idiosyncratic and having little or no generality. The new model formed the basis of an ontology.

Having developed a CommonKADS model of the system, a further set of activities were undertaken to determine how much of the knowledge could be reused in other parts of the system being used as the basis of the study. The methods used for finding a candidate process and the amount of reuse are presented.

The problem domain for this study was the Overlay Plating Section (figure 1) of ACL Bearing Company, located at Rocherlea in Northern Tasmania. ACL Bearings is a highly specialised manufacturer which recognises they are susceptible to loss of organisational expert knowledge due to domain experts leaving. The overlay plating section consists of a series of processes designed to apply a layer of alloy to engine bearings. The original system concentrated on one of the plating tanks, the P78 flash tank.

The P78 bath is unique in that its primary function is cosmetic, while also offering an additional degree of corrosion protection. Despite the fact that shell bearings are manufactured to be bolted inside an engine and never looked at again until they are due for replacement, the markets serviced by ACL Bearings are very sensitive to products that do not have a high quality of finish.

The P78 solution contains lead and tin, which are plated using the same process as for nickel plating, P1 flash and P77 plating (figure 2).

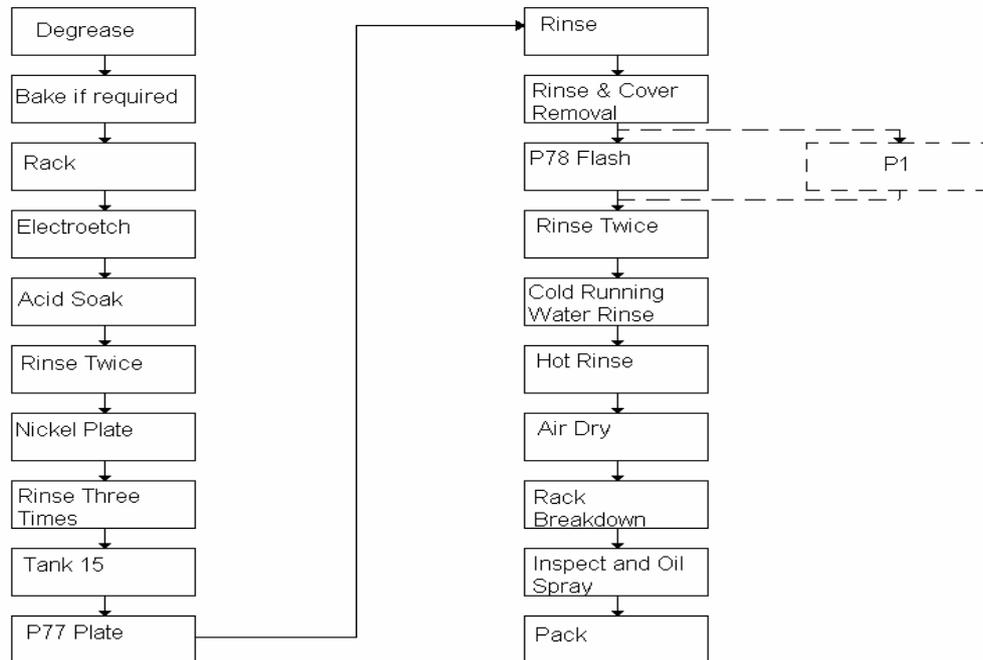


Figure 1. The general arrangement of the ACL Bearing Overlay Plating Section. ACLE Mark 1 was based on the P78 Flash tank process

The P1 flash was the process used as the basis of the knowledge reuse exercise. Unlike the P78 flash, the P1 flash does not form part of the existing knowledge base. P1 is an alternative flash plate that is used approximately every three months, for specific orders that require a level of finish that is above and beyond what is normally required. When this occurs, the solution is pumped from the P78 tank into a holding tank and the P1 solution is introduced. The plating line returns to normal operation, the main difference being that the P1 bath has no dissolved lead. The resulting flash plate has a higher percentage of tin than when P78 is used and has a shinier appearance.

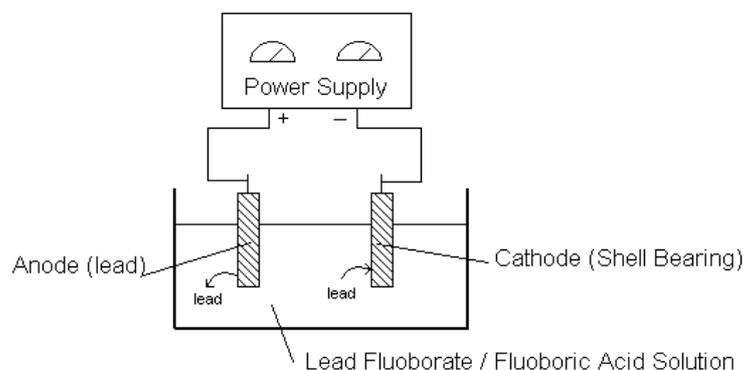


Figure 2. Diagram of a Plating Bath (for example P77, P78 and P1)

A first generation knowledge system, referred to here as ACLE Mark 1, was designed and constructed using rapid prototyping, and was successfully completed to specification but without any facility or structure being implemented for reuse of the domain knowledge. The system incorporated a set of production rules that chain together to form a series of categorised binary

18/08/2006

trees. At the root of each tree is the general category of problem which is selected by the user of the system upon beginning a consultation.

ACLE Mark 1 operationalises the somewhat outdated approach, typical of rule based systems, whereby the original notion of task comprised both the task specification and the knowledge base. Effectively, the problem solving method is fixed and forms an integral part of the implementation [15]. ACLE Mark 1 could therefore be described as having no function-data decoupling [18]. The system also fails to meet the guidelines set out by Moller [14] with regard to declarative, modular knowledge bases and easy explanations of the knowledge stored within the system. It is in fact doubtful as to whether the original ACLE Mark 1 system can be classified as an expert system, when considering a common definition of expert systems [5].

2. Theoretical Background

2.1 Introduction

The core challenges to the knowledge system community according to Doyle and Dean [3] include the development of broad and deep knowledge of large domains. This will be achieved through the continuation of current goals that seek to discover expressive and efficient forms and methods for representing information about all aspects of the world. These methods will be used to compile explicit, formal catalogues of knowledge that represent the efforts of domain experts from a variety of sources. These catalogues or libraries are known as ontologies.

The modern approach to analysis for the construction of a knowledge system is in stark contrast to traditional metaphors like 'Mining nuggets of information from the expert's head'. Schreiber et al. [19] declare that a KBS is not a container filled with knowledge extracted from an experts head, but an operational model that exhibits some desired behaviour which can be observed in terms of real-world phenomena.

2.2 CommonKADS

KADS was initiated as a "Structured methodology for the development of knowledge based systems"[19]. The limitations of production rules, combined with their inherent non-reusability contributed significantly to the impetus to develop methodologies like KADS. The two central principles that underlie the KADS approach are the introduction of multiple models as a means of coping with the complexity of the knowledge engineering process, and the use of knowledge-level descriptions as an intermediate model between expertise data and system design.

Motta [15] coins the term "knowledge modeling revolution", which refers to the paradigm switch from symbol level (rule based) approaches to knowledge level task centred analysis. This heralded the necessary decoupling of the task specification and the problem solving method.

There has also been the development of methodologies for the re-use of components of knowledge models. These allow efficient development of subsequent systems using reusable libraries of domain ontologies. For this to occur there needs to be standardisation of defining complex domains from the full spectrum of human knowledge [3].

2.3 Ontologies

Ontology has been defined as "...a view of what can possibly exist, concerned with possibility, necessity and contingency, but on an abstract level, rather than dealing with rules and

constraints.” [20]. This definition has been extended for artificial intelligence, such that an artificial intelligence ontology is a theory of what entities can exist in the mind of a knowledgeable agent [23]. An agent may be a human operator, or a software system.

A knowledge base can be viewed as a model of some part of the world, that allows for reasoning to take place in that world model, given some inference mechanism. The model is described in a particular language and has a vocabulary and a syntax. An ontology defines the constraints of possible objects expressed in the model in addition to the constraints imposed by the syntax [23].

In order to make statements and ask queries about a subject domain, Farquhar et al [4] explain that an ontology must use a conceptualisation of that domain which names and describes the entities that may exist in that domain and the relationships among them. It therefore provides a vocabulary for representing and communicating knowledge about the domain.

Using ontologies in the development of a knowledge based system allows for a disciplined design to be carried out and facilitates sharing and reuse [6] [21]. In the construction of an ontological model, consideration should be given to adherence to the 5 design criteria, as described by Gruber [10]. The criteria include “clarity, coherence, extendibility, minimal encoding bias and minimal ontological commitment”.

Knowledge sharing and reuse will require a common framework to support interoperability of independently created ontologies [6]. Two possible frameworks that to a certain degree are interchangeable are the CommonKADS library [22] and the Ontolingua Server.

According to Farquhar et al [4], reusable ontologies are becoming increasingly important for tasks such as information integration, knowledge level interoperation, and knowledge base development. The Ontolingua Server, hosted by Stanford University has been built to support the process of achieving consensus on common shared ontologies by geographically distributed groups. These tools make use of the world wide web to enable wide access and provide users with the ability to publish, browse, create, and edit ontologies stored on an ontology server. One additional feature of this service is the ability to create an ontology from a library of components.

3. Methodology

3.1 Prototyping

The ACLE mark 1 system was developed using a rapid prototyping methodology. It displays similar shortcomings to that of XCON [1], built as one of the original proof of concept configuration expert systems in 1979 by the Digital Corporation. Like XCON the ACLE mark 1 knowledge system suffered from indiscriminate representation of different types of knowledge using the same declarative formalism, that is, production rules [5]. Specifically, the implicit coding of a sequence of rules to provide a structure to the diagnostic process was identified as a problem by Fensel [5] and Schreiber et al [19]. Utilising this type of approach to knowledge system development impairs the acquisition and refinement of knowledge throughout the system lifecycle and hinders or prevents reuse of components of the system [19].

The attributes used in the ACLE mark 1 system shows an inconsistent level of specificity in their definition. This is indicative of an undisciplined knowledge acquisition process that did not form part of a structured methodology. Another reason for these inconsistencies is that much of the knowledge acquisition was carried out from diagnostic manuals of the system without interaction with the domain experts. Schmalhofer [17] suggests that knowledge acquisition primarily from

18/08/2006

texts is a most imprecise exercise. Texts tend to contain strategies rather than algorithms and should only be used to produce weak production rule systems.

The ACLE mark 1 knowledge base does not incorporate any intermediate reasoning steps (state dependency rules in CommonKADS terminology), a critical mistake, that hampers maintainability and reuse capability. This can have serious implications for the life cycle of the system. A system that is difficult to update and modify as the domain model changes, will quickly become redundant and will cease to be used.

Fensel [5] has identified some of the disadvantages and pitfalls of the rapid prototyping approach which include:

- The development team perform many tasks simultaneously including designing, analysing, coding and evaluation.
- Generally previous prototypes form the basis for the successive prototype and maintaining documentation over these incarnations is difficult. The result can be a completed system where the documentation must be generated from the system.
- When trying to complete a 'rapid' prototype, it is difficult to maintain a clear division between task and domain knowledge. There is a significant risk of these being intermingled, as has happened with ACLE mark 1.
- When using rapid prototyping, it is increasingly difficult to prevent implementational bias being introduced into the system.
- Additional bias may be introduced in the case of rapid prototyping, when issues of efficient programming collide with issues of efficient knowledge representation.

Rapid prototyping may have its place in system development, but it would appear from the evidence that it is not the development methodology to be used for modern knowledge systems. A structured methodology seeks to avoid these problems by presenting the programmer with a full set of implementation independent models as a starting point. This prevents the problems identified by Moller [14] of a restricted domain structure or a fixed problem solving method.

One of the most critical decisions to be made in commissioning a new knowledge system is associating the knowledge with the right level of expertise. The domain experts at ACL Bearings have a vast knowledge of the overlay plating section, but only a fraction of this knowledge is represented in the ACLE mark 1 knowledge base. It would seem that the level of system expertise could have been a good deal higher.

3.2 Knowledge Base Development and Reuse Strategy

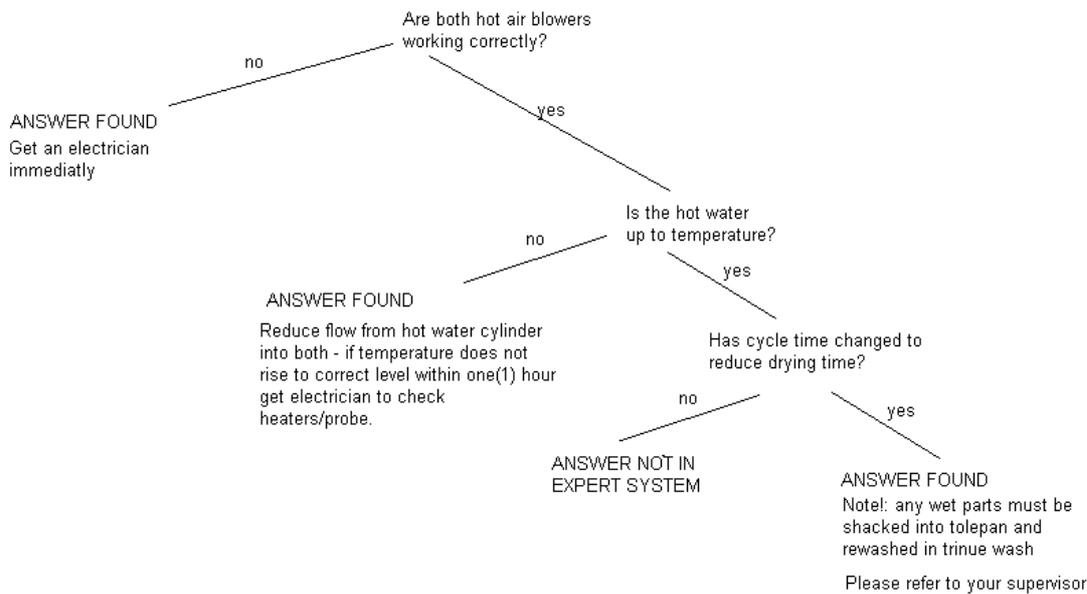
The idea of knowledge reuse is central to the principles of MIKE (Model based and Incremental Knowledge Engineering), a paradigm that recognises the cyclic nature of the knowledge system lifecycle and encompasses the CommonKADS methodology [16]. This is in keeping with Fensel's [5] description of knowledge acquisition as an iterative process that is infinite and approximate. One of the principle goals of the process of the CommonKADS modeling is to provide a catalogue of artefacts that can be reused within the current application, or in others.

Rather than build a new system completely from scratch, knowledge from ACLE mark 1 was reused in consultation with the domain experts (to avoid Schmalhofers [17] objections) and put into object oriented format implemented in COOL (CLIPS Object Oriented Language)[8]. Once this had been completed those components which could be reused were identified. This strategy had three stages.

- The knowledge was converted to a structure that is conducive to allowing comparison of various objects and concepts that exist in the domain. This structure was supplied by the CommonKADS methodology [18]
- A clustering exercise [2] was carried out on the completed CommonKADS models, to allow for the choice of reuse candidates of domain schema and knowledge base. The data for this was obtained during interviews with the domain expert. This was computer facilitated to enable the knowledge engineer and the domain expert to collaboratively assign values of closeness between every pair of objects that exist in the system. This approach eliminated the need to present the domain expert with a broad sheaf of paper, requiring multiple numerical entries [9].
- Reuse candidates were chosen from the existing knowledge base and schema. These were presented to the domain expert in the form of a reuse questionnaire, for assessment of their applicability to the P1 Flash Plate section (which was not considered in the ACLE mark 1 system).

3.3 Construction of CommonKADS Models

The first steps were to construct the organisational, agent and task models for the P78 schema. This modelled the system requirements. The details of these models are beyond the scope of this paper however it must be noted that it is essential they be before the knowledge model could be developed.



Visual Defect - Wet Parts

Figure 3 A decision tree from ACLE Mark 1

By carrying out an examination of the categories of knowledge and the decision trees (figure 3) held in the ACLE Mark 1 system, it was possible to identify the individual physical objects (figure 4). These objects were conceptualised in the CommonKADS model set [18]. Once the complete set of objects in the domain were identified, they were defined to the required level of abstraction by the addition of schema. Schema definition for physical objects in the domain was carried out by examining each internal node in the tree structure of the ACLE Mark 1 system.

```
concept P78_Flash_plate_section;
```

```

sub-type-of: plating_line;
properties:
  lead_count_greater_or_equal_to_20ppm: {yes, no} ;
  lead_count_still_greater_or_equal_to_20ppm: {yes, no} ;
  copper_level_of_P78_greater_than_20ppm: {yes, no} ;
  P78_bath_clean_and_free_of_contamination: {yes, no} ;
  P78_ammeter_setting_high: {yes, no} ;
  lead_and_tin_level_in_P78_bath_within_specification: {yes,no};
  robber_bars_straight_and_positioned: {yes, no} ;
  P78_bath_chemistry_within_specifications: {yes, no} ;
  bath_and_flight_contacts_in_P78_bath_clean: {yes, no} ;
  dummys_clean_and_rust_free: {yes, no} ;
  cover_removed_from_rack_prior_to_entry_to_P78: {yes, no} ;
  covers_left_on_the_racks_while_in_P78_bath: {yes, no} ;
  tin_and_resourcinol_levels_in_spec: {yes, no} ;
  P78_bath_low_in_gelatine: {yes, no} ;
  particle_size_of_gelatine_too_large: {yes, no} ;
  any_plating_baths_been_agitated: {yes, no} ;
  resourcinol_levels_in_spec: {yes, no} ;
  fluoboric_acid_level_within_spec: {yes, no} ;
  Pfaudler_evaporator_operating_correctly: {yes, no} ;
  bath_cloudy: {yes, no} ;
  P78_bath_within_specification: {yes, no} ;
  extra_or_increased_gelatine_added_to_plating_baths: {yes,no} ;
  current_settings_correct: {yes, no} ;
end concept P78_Flash_plate_section;

```

Figure 4: CommonKADS description of the P78 Flash plate section

```

concept air_dry_section;
  sub-type-of: plating_line;
  properties:
    both_hot_air_blowers_working_correctly: {yes, no} ;
end concept air_dry_section;

```

Figure 5: CommonKADS description of the first node of the decision tree.

The knowledge base was constructed by instantiating each leaf node into a diagnosis (problem-defect-object (figure 5)) and, from each leaf node, backtracking to the root. Each internal node that was encountered on the path to the root node was included in the set of antecedents that apply to that particular diagnosis node. The diagnosis stores the question text, the required value and the domain object to which the question refers. In addition to this, each classification of problem or defect was included as a concept [18].

black-areas-on-p1-defect	current_diagnosis	Candidate number 6
p1_Flash_plate_section	P1_bath_clean_and_free_of_contamination	yes
	note: black areas on P1 could be described as blemish	no
		Organic breakdown products do accumulate in the bath must be removed otherwise black streaks on the OD will occur. Treat p1 filter with activated carbon, followed by an addition of 0.5g/l of gelatine
Enter a rating of applicability of this diagnosis to the P1 bath		3/5
What additional factors could be added to make the diagnosis more applicable to the P1 bath	check filtration and RTL addition and run Hull cell in diagnoses	
what modifications are required to the diagnosis to reflect its use in P1	There is no gelatine in P1: The diagnosis should read: Check filtration and RTL addition, run hull cell test	
What existing factors in the diagnosis do not apply to the P1 bath and should be removed to make the diagnosis more applicable to the P1 bath		none

Figure 7: Completed reuse questionnaire showing a partial reuse match

The domain expert was encouraged to consider the reuse candidate diagnosis with regard to potential reuse in the P1 Flash Plate Section, The domain expert was then requested to give a rating of applicability of this diagnosis to the P1 bath (a rating between 1 and 5 where 5 is a perfect match with no changes necessary). Reuse candidates were classified as, full matches, partial matches adapted, partial matches scrapped, outright misses. Hutchinson and Hindley [11] specifically encourage the approach of adapting partial matches for the purpose of reuse of domain knowledge [16].

4. Results and Discussion

4.1 ACLE Mark 2 and CommonKADS

The decision to build a system that embodied the remodeled system and knowledge base, including the reused knowledge as applies to the P1 flash plate bath is supported by Fensel [5], in that it allows evaluation of the resulting knowledge. The ACLE Mark 2 system is effectively a fully instantiated model of a knowledge domain, with sufficient information to perform a reasoning process, with a problem solving strategy that has been decomposed to an atomic level [23]. The ACLE Mark 2 system is also consistent with the 5 design criteria for knowledge models, described by Gruber [10]: clarity, coherence, extendibility, minimal encoding bias and minimal ontological commitment.

The structure that has been used in the design and construction of the ACLE Mark 2 system is in fact a "reusable abstract domain-independent problem solving strategy" [13], which is the essence of the CommonKADS methodology.

4.2 Reuse of Knowledge

The results showed that 56% of the reuse candidate diagnoses from the P78 process were directly applicable to P1 process without modification. A further 25% of the reuse candidates could also be applied to P1 with a minor modification such as dropping or adding an antecedent, or

modifying the wording of a diagnosis slightly. The final 19% of reuse candidates were totally non-applicable to the P1 plating tank.

Of the total of 32 schema attributes that are used to describe the P1 flash plate section, 56% were derived from the reuse experiment, and could be applied directly without modification. There was 1 schema attribute that required some modification, but was still able to be included in the P1 flash plate schema. In addition to this, there were an additional 3 schema attributes derived from the reuse questionnaires, in the form of a required additional antecedent. This is an extremely significant ratio of reuse, with a total of 68% of the total domain schema for the P1 flash plate section being supplied by the reuse exercise.

4.3 A Formal Ontology for Electroplating

Some shortcomings of the ACLE Mark 2 knowledge base would currently prevent it from being used as the basis for a standardised domain ontology, but with the investment of some time by a knowledge engineer and the domain expert, the potential is certainly there. The principle deficiency that should be addressed prior to any ontology being released is the inconsistent 'grain size' of attributes. Some have been decomposed sufficiently, and others have not [18]. There are also the confidentiality issues to be addressed, with respect to the release of information that is proprietary to ACL Bearings.

5. Conclusion

This research has demonstrated that a legacy system, developed using rapid prototyping and built from a foundation of no formal knowledge modeling can still form the basis for a reuse strategy centred on the CommonKADS methodology. This being so, it follows that legacy systems developed using more rigorous decomposition and modeling techniques, but still employing rapid prototyping, would be more amenable to this technique. The capacity for reuse of knowledge that is provided by CommonKADS, such that new objects that are added to an existing knowledge domain might be modeled and added to the knowledge system with a minimum of effort and expense, has been demonstrated.

The results of applying these techniques to the ACLE Mark 1 system proved very successful. There was also a large number of candidates that were reused with minor changes. This is a very significant result when considering that the documented knowledge base that existed before the reuse exercise represents only 25% of the total knowledge base after the reuse exercise.

These results should be tempered with the knowledge that the P78 and P1 baths are mutually exclusive in the system but occupy the same space and perform very similar functions. This should not discourage the practitioner of knowledge reuse, as many domains in all aspects of human endeavour, particularly manufacturing, embrace the principle of redundancy, whereby a knowledge domain may contain numerous objects of similar characteristics [12] [7].

References

- [1] Barker, V. & O'Connor, D. (1989). Expert Systems For Configuration at Digital: XCON and Beyond, *Communications of the ACM*, 32(3), 298-317.
- [2] Borgatti, S. P., 1994, *How to Explain Hierarchical Clustering*. <http://analytictech.com/networks/hiclus.html>.
- [3] Doyle, J. and Dean, T., 1997, Strategic Directions in Artificial Intelligence. *AI Magazine*, V18(1), pp. 87-101.

- [4] Farquhar, A., Fikes, R. and Rice, J., 1996, *The Ontolingua Server: a Tool for Collaborative Ontology Construction*. Stanford Knowledge Systems Laboratory. <http://www-ksl-svc.stanford.edu:5915/doc/project-papers.html>
- [5] Fensel, D., 1995, *The Knowledge Acquisition and Representation Language, KARL*, London: Kluwer Academic Publishers.
- [6] Fridman Noy, N. and Hafner, C. D., 1997, The State of the Art in Ontology Design: A Survey and Comparative Review. *AI Magazine* V18(3), pp53 - 74.
- [7] Gallagher, S., 1999, *Scalability Through Redundancy*, Information Week, CMP Publications inc.
- [8] Giarratano, J. and Riley, G. 1998b. *CLIPS Reference Manual Volume II: Advanced Programming Guide*. PWS Publishing Company, Boston.
- [9] Gottlieb, M. 1986. *Interview*, Longman Publishers, New York.
- [10] Gruber, T.R., 1993, *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. Stanford Knowledge Systems Laboratory. Available: <http://www-ksl-svc.stanford.edu:5915/doc/project-papers.html>
- [11] Hutchinson, J. W. and Hindley, P. G., 1988, A Preliminary Study of Large Scale Software Reuse. *Software Engineering Journal*, V3(5), pp. 208 - 212.
- [12] Lawley, M., Reveliotis, S., Ferreira, P., 1997, *Correctness and Scalability*, Institute of Industrial Engineers.
- [13] Menzies, T., 1995, *Limits to Knowledge Level-B Modeling (and KADS)*. <http://www.cse.unsw.edu.au/~timm/pub/docs/papersonly.html>
- [14] Moller, J., 1995, *Operationalisation of KADS models by using Conceptual Graph Modules*. [Online] [2000, June 27]<http://nats-www.informatik.uni-hamburg.de/~jum/papers/95/CG-KADS.html>
- [15] Motta, E., 1997, Trends in Knowledge Modeling: Report on the 7th KEML Workshop. *The Knowledge Engineering Review*, V12(2), pp. 209-21
- [16] Pressman, R. S., 1997, *Software Engineering: A practitioner's approach*, The McGraw-Hill Companies, USA.
- [17] Schmalhofer, F. (1998). *Constructive Knowledge Acquisition: a Computational Model and Experimental Evaluation*. London: Lawrence Erlbaum Associates Publishers.
- [18] Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W. and Wielinga, B., 2000, *Knowledge Engineering and Management, The CommonKADS Methodology*, Cambridge: MIT Press.
- [19] Schreiber, G., Wielinga, B. and Breuker, J., 1993, *KADS: A Principled Approach to Knowledge-Based System Development*, London: Academic Press.
- [20] Thro, E., 1991, *The Artificial Intelligence Dictionary*, Microtrend Books, California, USA.
- [21] Van der Vet, P. E. and Mars, N. J. I., 1998, Bottom-Up Construction of Ontologies. *IEEE Transactions on Knowledge and Data Engineering*, V10(4), pp. 513 - 526.
- [22] Valente, A, Breuker, J and Van De Velde, W., 1998, The CommonKADS Library in Perspective. *International Journal of Human-Computer Studies*, V49, pp. 391 – 416.
- [23] Wielinga, B. J. and Schreiber, A. 1993. *Reusable and Sharable Knowledge Bases: A European Perspective*. Proceedings of the International Conference on Building and Sharing of Very Large-Scaled Knowledge Bases '93.