A Knowledge Acquisition Systematic within the Domain Analysis Context

Mônica Zopelari Roseti, Cláudia Maria Lima Werner

COPPE/UFRJ – Computer Science Department
Federal University of Rio de Janeiro
Caixa Postal 68511 - CEP 21945-970
Rio de Janeiro - RJ - Brazil
E-mail: { zopelari,werner }@cos.ufrj.br

Abstract. This work presents a systematic approach to knowledge acquisition within the domain analysis context, which grants a well defined process for eliciting and representing domain information that is sensitive to the kind of knowledge to be explored. Within the domain analysis context, knowledge acquisition encompasses aspects that are similar to the ones found in Requirements Engineering, although it is necessary to identify, elicit, and register requirements that belong to a family of systems instead of a unique system. Besides domain requirements, every kind of information that is related to the given domain must be identified, elicited and registered, supporting its comprehension and future generation of applications.

Keywords: Knowledge acquisition, Domain Analysis, Software Reuse, Knowledge acquisition systematic.

1. Introduction

Knowledge acquisition[1] is a very important activity within the Domain Analysis (DA) context, since it is necessary to organize the available knowledge about an application domain in a coherent manner. A DA process must help to "identify and organize the knowledge about a certain group of problems to support their description and solution. This process comprises learning, with an underlying concern about capturing, gathering, organizing, and modeling the knowledge within a certain domain." (Arango, 1991)
In the DA context, knowledge acquisition encompasses aspects that are similar to the ones found in Requirements Engineering, although it is necessary to identify, elicit, and register requirements that belong to a family of systems instead of a unique system. Besides domain requirements, every kind of information that is related to the given domain must be identified, elicited and registered, supporting its comprehension and future generation of applications.
Since DA deals with knowledge identification, elicitation, and representation of software products, techniques that systematize these tasks are necessary. One can assume that this is a well resolved situation, since there are several DA approaches found in the technical literature (Prieto, 1987,1993,1994) (Cohen, 1991) (Neighbors, 1992) (Leite, 1994) (Gomaa, 1994,1995) (Arango *et al*, 1994) (Simos, 1994) (Stars, 1996) (Jacobson *et al*, 1997) (Griss *et al*, 1998). However, when considering the issue of knowledge acquisition, these approaches do recognize the importance of this activity, but they do not provide a systematic process for performing it.
This work presents a systematic approach to knowledge acquisition within the domain analysis context, which grants a well defined process for eliciting and representing domain information that is sensitive to the kind of knowledge to be explored. It is assumed that we are engaged in exploring, as much as we can, the existing knowledge about a certain domain. Therefore, it is necessary to deal with different kinds of knowledge (e.g., procedural, logical, etc.), hence requiring different techniques and abstractions to elicit and represent them.

[1] The knowledge acquisition process encompasses not only the identification and gathering of information but also its organization and storage, always considering the continuous and evolutionary nature of this process (Scott *et al*, 1991).

This paper is organized in five sections: section 2 displays an evaluation of the knowledge acquisition issue within current domain analysis approaches; section 3 describes the proposed systematic; section 4 presents a case study that uses the proposed systematic; section 5 encompasses the conclusion.

2. Knowledge Acquisition within current DA approaches

The analysis of current DA approaches was performed considering a simplified perspective of this activity, as shown in Figure 1. First, the domain knowledge sources are identified. These are selected in terms of relevance to the domain, explored and registered in a suitable and coherent representation.
Under this perspective, two issues are important in order to evaluate the DA approaches: knowledge elicitation and its representation. The analyzed approaches were: Facets and Sandwich Method (Prieto, 1987, 1994); Feature Oriented Domain Analysis (Cohen, 1991, 1994), DRACO (Neighbors, 1992) (Leite, 1994); Evolution Domain Life Cycle (Gomaa, 1994, 1995) Arango's DA Generic Process (Arango *et al*, 1994), ODM (Simos, 1994) (Stars, 1996), and RSEB (Jacobson *et al*, 1997) (Griss *et al*, 1998).

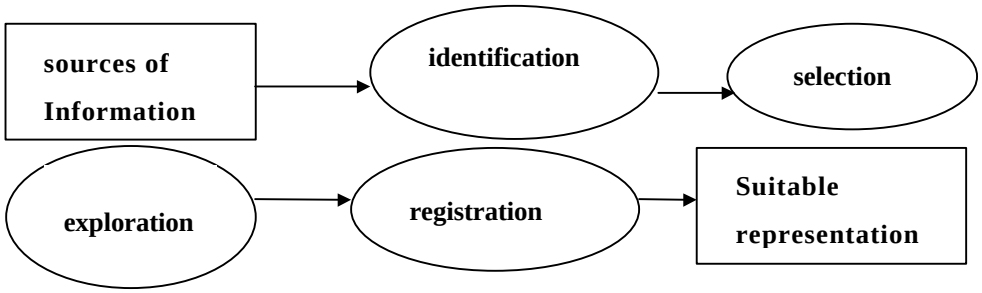

Fig. 1. Simplified Perspective of Knowledge Acquisition within DA

We were able to observe that most approaches include an information source identification activity, which can be either specialists or existing systems, and some also indicate technical references about the subject. However, the necessary procedures to elicit the information are not precisely described, nor how to conduct them in an adequate way.
The individual roles are generally well defined, but a script for playing these roles is not provided, and on that account there is no description of a technique that aids people to organize their thought at a certain point while performing their activities.
The specialist is the one who is expected to provide the directives, issue tips on information sources, and explain his knowledge about the domain. Only in Arango's generic process one can find suggestions for interaction techniques with the specialist, such as: interviews, questionnaires, and meetings. Nonetheless, there is not a precise description available about these techniques, and no indication whatsoever about the appropriate time to use them.
Therefore, we can state that the knowledge elicitation issue has had its importance recognized by all approaches. However, a systematic and disciplined way to conduct it within the DA context has not yet been defined. On the other hand, there are several techniques developed for Knowledge Based Systems (KBSs), or even for conventional system analysis, that can and should be used within the DA context. However, DA activity has unique features regarding the project as a whole (individuals involved, context, aims, and products). Therefore, we adopt the hypothesis that specific techniques to develop this activity are necessary, even if they represent just a revision of existing ones.
Regarding the representation issue, the fact that modeling must formalize knowledge, domain concept meanings, modularization of decisions, and domain semantics is shared by all approaches. The *domain model* is the major result of all DA approaches. It should be able to capture the critical knowledge that remains within the analyst's mind, to provide a deep understanding among producers and re-users, to aid its evaluation, and its trustworthy evolution.
To fulfill these requirements the model must be represented in a coherent way. Several representation means are presented: facets, glossaries, domain languages, E-R models, functions, objects, features diagrams, hierarchy models, OO diagrams, and state transition diagrams, among others.
Many approaches emphasize only one kind of representation (i.e., facets), which most of the time is not enough to completely capture the knowledge that is acquired within a DA process. Others support the

construction of a domain model with several kinds of specific representations (e.g., Feature Oriented Domain Analysis) directing the DA process to the construction of these representations.
As mentioned before, we assume that it is important to be able to deal with several kinds of knowledge, but we do not necessarily know it before we have a concrete DA project. Therefore, the representation needs to be determined according to the kind of knowledge that we are going to deal with in a specific DA project. There is no DA approach to our knowledge that considers this issue.

3. A Systematic approach to Domain Knowledge Acquisition

Our proposal consists of a DA knowledge acquisition systematic that provides well-defined phases, activities and corresponding techniques, which are selected according to certain directives. This systematization is a combination of Arango's proposed DA generic process (Arango *et al.*, 1994) with the philosophy established by SCOTT *et al* (1991) for knowledge acquisition within KBSs.
From Arango's generic process (Arango *et al.*, 1994), we kept the phases and some of its activities. From Scott's *et al* (1991) knowledge acquisition process, we applied the philosophy of constantly revising and refining the acquired knowledge throughout the process.
According to Scott et al (1991), one must identify an initial phase in which efforts must be well directed, so that it is possible to understand the domain, sparing as much as possible the specialist's total time. This initial phase corresponds to the Data Gathering phase from Arango's generic process (Arango *et al*, 1994), involving the following activities: Available Literature Revision, Existing Systems Evaluation and Specialist Knowledge Elicitation. We included a Captured Knowledge Revision activity at the end of this phase to preserve Scott et al's philosophy about constantly revising and refining the acquired knowledge.
Scott *et al* (1991) consider that a detailed phase must follow, when a deep investigation is performed. This corresponds to Arango's Data Analysis and Components Classification phases (Arango *et al*, 1994). Consequently, we grouped these two phases into one, keeping only the activities listed in Table 1. We included a Captured Knowledge Validation activity in the beginning of this phase, and a Model Revision activity at the end, in order to keep Scott et al's revision philosophy.

1- Domain Characterization Phase- Preliminary Phase
1.1- Domain selection 1.2- Domain Description 1.3- Relevant Data Identification - structured interviews 1.4- Data Inventory Creation - structured interviews 1.5- Process Planning
2- Data Gathering Phase – Adapted according to Scott et al's philosophy
2.1-Available Literature Revision- LEL technique 2.2- Existing Systems Evaluation - Protocol Analysis 2.3- Specialist Knowledge Elicitation - Brainstorming 2.4- Captured Knowledge Revision - Self-explanation
3- **Data Analysis and Components Classification Phase- Adapted according to Scott et al's philosophy**
3.1- Captured Knowledge Validation - structured interview/Delphi technique 3.2- Initial Model Development – Scenarios 3.3 - Similarities, Variations and Combination Analysis - Scenarios 3.4- Model Revision - Self-explanation
4- Model Evaluation and Availability - Adapted according to Scott et al's philosophy
4.1- Model Validation with the Specialist - structured Interview/Delphi technique 4.2-Model Evaluation and Availability - structured interview/Delphi technique

Table 1. Knowledge acquisition systematization within the DA context

At the Model Evaluation and Availability phase, we included two activities: Model Validation with the Specialist, and Model Evaluation and Availability.

For each activity there is an associated knowledge acquisition technique, as shown in the Table 1. Also, each acquisition technique is associated to a certain source of information with its corresponding representation (Table 2).

While specifying and refining the systematic presented in Table 1, we conducted three case studies. These helped us to experience in practice the need for a knowledge acquisition systematic within the DA context. We were able to experience the existing difficulties while performing domain knowledge acquisition without such systematic and had to propose adequate techniques for managing them. The results allowed us to evaluate the benefits generated by the use of such systematic. Despite the fact that these case studies did not have the scientific rigors of formal experiments (Kitchenham *et al*, 1995), they provided us with enough information on how helpful it was to use a disciplined process.

The case studies were carried out within research projects established among the Software Reuse Group from the Computer Science Department at COPPE/UFRJ and three entities: Macromolecules Institute at UFRJ, exploring the polymeric chemistry domain; a medium size *software house*, exploring the storage control domain; and a municipal legislative office, exploring the legislative process domain.

During these case studies, we also verified that it was necessary to adapt the systematic to specific DA projects. Therefore, some directives to identify the different contexts were developed, based on the works of Cima *et al* (1995) and Maiden *et al* (1996). These directives allow the adaptation of the systematic regarding the phases, activities and techniques that should be used in a certain project. The domain analyst identifies the context of his specific project by analyzing each of the directives provided, establishing which phases and activities, as well as the corresponding techniques, from the knowledge acquisition process described in Table 1, will his knowledge acquisition process be composed of.

Technique	***Sources***	***Representation***
Protocol Analysis	Existing systems	Facets
Scenarios	Existing systems Refinement	Itemized Cards
LEL[2]	Written references Domain Vocabulary	Itemized Text
Brainstorming	Specialists Users	Numbered Text
Structured interviews	Specialists Users	Structured Text
Delphi technique	Specialists Users	Previous Formatted Report
Self-explanation	Domain analyst Inferences and deductions	Previous representations

Table 2. Associations among techniques, sources and representation

The directives cover issues, such as:

1. ***Adopted Domain Analysis Point of View:*** problem analysis or solution analysis.
2. ***Domain Analysis Objectives:*** acquisition objectives must comply with domain analysis objectives (e.g., construction of reusable generic architectures, organization and establishment of domain specific function libraries, planning of domain application development, or the organization of a certain domain knowledge so it can be further researched).
3. ***Kind of Explored Knowledge:*** the kind of knowledge that is to be captured during each phase of the acquisition process must be determined beforehand (e.g., procedural, logical, etc).
4. ***Execution Conditions:*** some factors, which together form a context to perform knowledge acquisition and set restrictions to the free choice of techniques, must be determined. These factors include: estimated duration time, available individuals, information sources (e.g., specialists, users, written references, and existing systems), hardware, documentation resources, and spare parts (e.g., video, tapes).
5. ***Inter-dependency between techniques:*** the sequence of techniques must be taken into account, considering their results as input for other techniques.

[2] LEL is a technique that aims to elicit the domain application vocabulary at first, because of the belief that the formalization must be a gradual process to create a formal model which can reflect more accurately the reality (Leite, 1993), (Leite, 1995).

A tool that allows the inclusion of other systematic besides the one described in this paper, by adding new processes, techniques and directives in its database, was developed (Roseti *et al*., 1998). This tool presents to the domain analyst the knowledge acquisition processes available in its database. Based on this list, he can select the one that he wants to adapt for his specific DA project. A questionnaire is presented to him. Each question relates to one of the mentioned directives. By answering this questionnaire the tool is able to customize the selected process to his context.

4. The Case Study of the Legislative Process Domain

In this section, we briefly describe one of the case studies previously mentioned, which was conducted in a municipal legislative office, exploring the legislative process domain. This use case was the last and most conclusive one, allowing us to verify the usefulness of the proposed knowledge acquisition systematic within the DA context.

The Software Reuse Group from the Computer Science Department at COPPE/UFRJ was interested in evaluating the proposed systematic, and the municipal legislative office was interested in verifying the possibility of developing projects cooperatively with other legislative offices, by providing a domain model which could be shared among them.

First of all, we identified the objective of the domain analysis, the available sources of information and the kind of knowledge to be explored. The objective of the domain analysis was to organize the available knowledge about the legislative process, in order to provide a model to be shared by the different legislative offices interested in cooperative projects. The kind of knowledge to be explored was mostly procedural. The sources of information were technical manuals about legislative process, installed systems, user manuals of those systems, domain specialists, and system developers. This information (directives) is used to adapt the systematic regarding the phases, activities and techniques that should be used in a certain project. For instance, if there is no written material available there is no need for the Available Literature Revision activity of the Data Gathering Phase.

All phases and corresponding activities were monitored. Tables 3, 4, 5, and 6 present the results and time evaluation of the domain characterization phase, the data gathering phase, the data analysis and components classification phase, and the model evaluation and availability phase, respectively.

Domain Characterization phase
1-Domain selection
The legislative domain was chosen because of the easy access of COPPE researches to the municipal legislative office, which gave us the chance to identify systems with high potential for reuse.
2-Domain description
To speed up the work we executed this activity in the same interview for the next activity.
3-Relevant Data Identification
Technique: Interview with the specialist. *Results and time evaluation:* we spent three hours in an interview where we defined the global objectives and the domain scope. We also identified the sources of information: technical manuals of the legislative process, installed systems, users manuals, domain specialists, and system developers. A context diagram was also built.
4-Data Inventory Creation
Results and time evaluation: a textual catalog containing a description of each source of information its location, media, and kind was created.
5-Process Planning
Results and time evaluation: The planning was done according to the proposed knowledge acquisition systematic for DA context and it proved to be very useful.

Table 3. Monitoring of the Domain Characterization Phase

Data Gathering Phase
① Available literature revision *Technique:* LEL *Results and time evaluation:* this activity took two days, applying the LEL technique to the system manuals and the technical manuals of the legislative to obtain the domain vocabulary.
②Existing system evaluation *Technique:* Protocol analysis *Results and time evaluation:* The systems were verified in two days, resulting in a functional faceted scheme. By using the vocabulary domain, it was easy to understand domain particularities found.
③ Specialist knowledge elicitation *Technique:* Brainstorming *Results and time evaluation:* During three hours, specialists and developers talked about the legislative process, which resulted in a list of common terms often used in this domain, represented in the same way as proposed by the LEL technique. So we were able to compare the terms found in the literature with those used regularly by people, minimizing some possible vocabulary misunderstandings.
④Captured knowledge revision *Technique:* Self-explanation *Results and time evaluation:* The domain analyst revised the domain vocabulary and the functional faceted scheme built, to verify his own understanding and learning of the domain. The doubts and inferences were noted down to be presented to the specialist.

Table 4. Monitoring of the Data Gathering Phase

Data Analysis And Components Classification Phase
① Captured knowledge validation *Technique:* Structured Interview *Results and time evaluation:* An interview of two hours took place with the specialists and developers, to verify and correct the domain vocabulary and the functional faceted scheme. Also the doubts and inferences of the previous activity were presented.
② Initial model development *Technique:* Scenarios *Results and time evaluation:* From the functional faceted scheme, it was possible to create behavioral scenarios of the domain systems, and from these scenarios it was possible to develop an O.O. model. This activity was performed during two days.
③Similarities, Variations and Combination Analysis *Technique:* Structured interview based on scenarios *Results and time evaluation:* Based on the scenarios built, the similarities and variations were identified in order to produce a feature diagram as proposed by the FODACom approach (Griss *et al.*, 1998).
④Model revision *Technique:* Self-explanation *Results and time evaluation:* The domain analyst revised the O.O. model to consolidate and test his learning and understanding. Again, the doubts and inferences about the model were noted down.

Table 5. Monitoring of the Data Analysis and Component Classification Phase

Model Evaluation and Availability Phase
① Model Validation with the Specialist *Technique:* Structured interview *Results and time evaluation:* The O.O. model together with the doubts and inferences noted down on the previous activities were presented to the specialist to the final validation and acceptance of the final version of the model. This activity took three hours.
② Model Evaluation and Availability *Technique:* Structured interview *Results and time evaluation:* The domain model was composed by the context diagram, the domain vocabulary, the functional faceted scheme, the O.O. model and the feature diagram.

Table 6. Model Evaluation and Availability Phase

Figures 2, 3, 4, and 5 present the functional faceted scheme, a scenario, the O.O. diagram and the feature diagram, respectively.

Agent	***Operation***	***Elements***
Author	Present	Autograph
Committee	Register	Dispatch
Official Press	Follow up	Day order
Directory	Discuss	Counsel
Mayer	Develop	Proposal
Parlamentarian	Sent	Original Composition
President	Publish	Intermediate Composition
Directory Secretary	Promulgate	Final Composition
Account Court	Sanction	Vote
	Forbid	Forbidding
	Vote	Legal procedure

Figure 2. Functional faceted scheme

Name:	**Proposal Presentation**
Type:	**Compulsory - variable**
Actors:	**Author, Directory secretary, <variant2>**
Variant1:	**Project, Amendment, Sub-amendment, Substitute**
Variant2:	**President, Directory**
Use case:	
When author is Account Court or Mayer The <variant1 > is direct to <variant2> When author is a Parlamentarian, Directory or a Committee The author presents the <variant1 > to the Directory Secretary The Directory Secretary protocols the <variant1> The <variant1> is sent to <variant2> The <variant2> defines which Committees will develop a counsel about <variant1> The <variant2> develops the dispatch The <variant1> is numbered by the Directory Secretary The <variant1> is published by the Official Press The <variant1> is sent to the Committees to develop the counsel	

Figure 3. Proposal Presentation Scenario

5. Conclusion

This paper presented a knowledge acquisition systematic to be used within a DA process. It provides explicit acquisition activities and a set of knowledge elicitation techniques, each with a corresponding representation scheme. The main premise is that it is necessary to explore all kinds of knowledge during a DA process, and there should be specific elicitation techniques and representation for each kind.

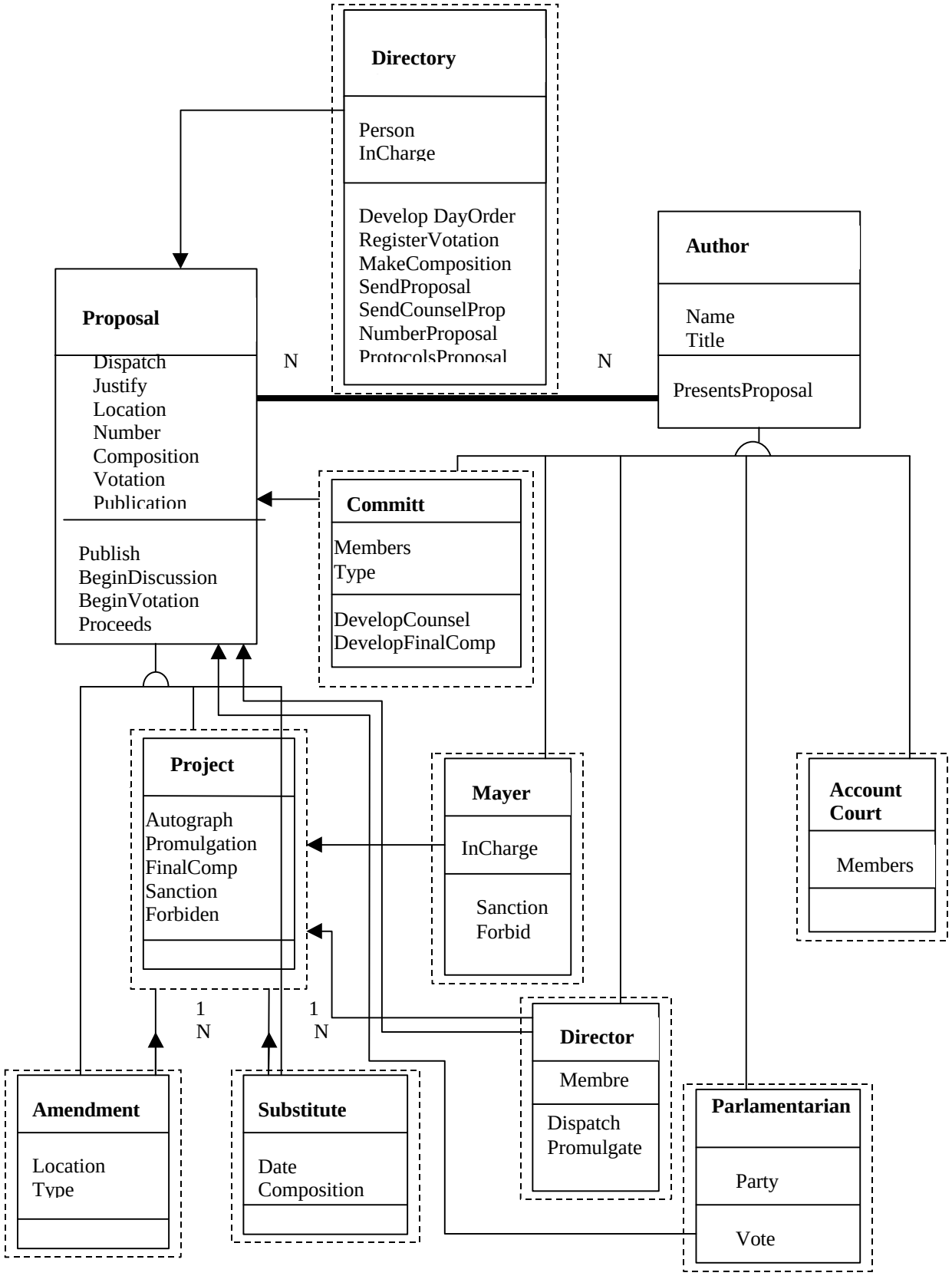

Figure 4. O. O. Diagram

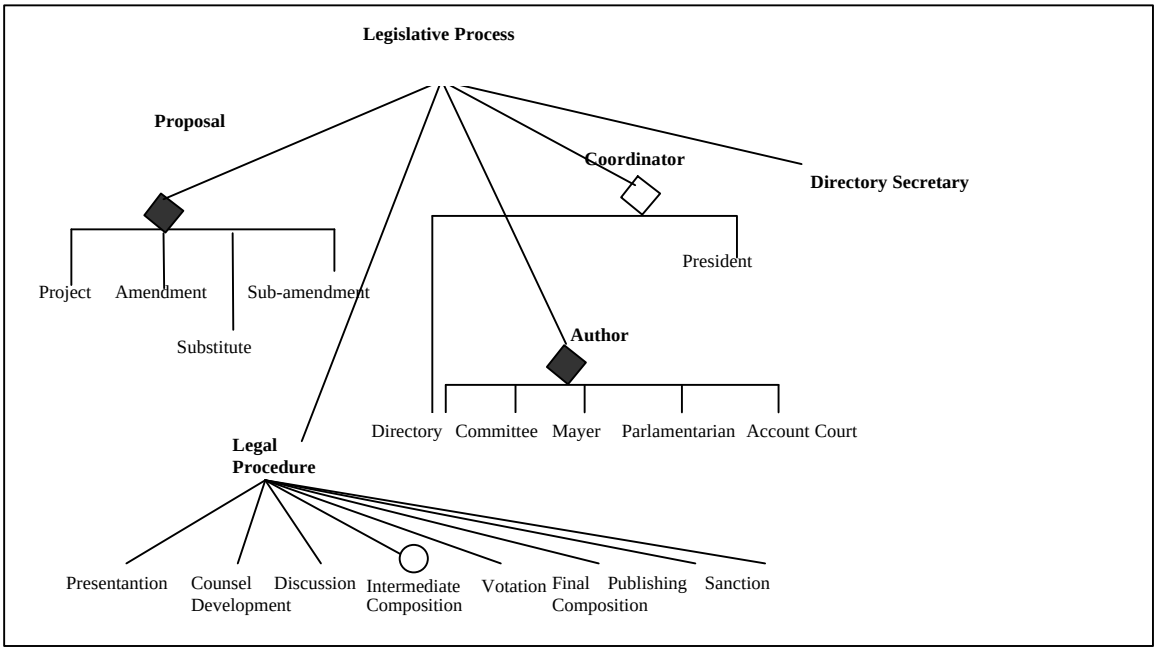

Figure 5. Feature Diagram

Considering the results of the performed case studies, our conclusions were:

- There is no doubt that systematization is necessary for knowledge acquisition within the DA context.
- We really face different kinds of knowledge, during a DA process.
- Consequently, there is the need to deal with different sources of information: written references, and human resources playing different roles (i.e., application developers, domain specialists, and users).
- For each kind of knowledge and information source, there should be a technique to better explore it.
- Once presenting well defined, orderly and inter-related phases, the directives for the selection of adequate techniques, according to the kind and nature of the information source, the systematic served as an important support to domain analysts who were able to avoid situations, such as:
  - Difficulty in organizing the thought process during DA knowledge acquisition.
  - Lack of direction while performing the activity.
  - Tendency to initial inertia while facing an unknown and/or complex domain.
  - Not being fluent in the domain language.
  - Difficulty to extract information from identified sources and to evaluate them in an adequate manner.
  - Difficulty to sort out relevant information from the domain.
  - Lack of knowledge about the technique that is adequate at a given time.
  - Lack of an organized sequence of steps to be performed.
  - Lack of specialist's availability.

The proposed systematic fulfills the knowledge acquisition planning activity, that was the first difficult situation faced during the DA case studies, so all the efforts were directed to solve it. However, we consider that it is important to conduct further studies to solve the difficulties that are related to knowledge acquisition per say, and its follow-up.

Also, the implemented tool should be extended to support the execution of acquisition activities, providing adequate guidance for each eliciting technique and electronic storage means for its corresponding representation. This would yield an increased precision of the final results.

Considering the consistency among representations and the validation of models in general, which are currently dependent of a specialist's evaluation, one should consider the identification of set of criteria and/or metrics to support this activity in a more concrete basis. This is a topic for future research.

References

Arango, G., Prieto, R., 1991, "Domain Analysis Concepts and Research Directions", in: *Domain Analysis And Software System Modeling*, 1st ed., California, IEEE Computer Society Press Tutorial, pp. 09-25.

Arango, G., Basili, V.R., Borstler,J., *et al.*, 1994, *Software Reusability*, 1st. ed, Great Britain, Ellis Horword.

Cima,A. M., Werner, C., 1995, *The Selection Process of a Domain Analysis Method*, Technical Report, Computer Science Department, COPPE/UFRJ, Rio de Janeiro, Brazil (in Portuguese).

Cohen,S., 1991, "Process and Products for Software Reuse and Domain Analysis", in: *Proceedings of the Fourth Annual Workshop on Software Reuse*, Syracuse, New York.

Cohen,S., 1994, "Feature-Oriented Domain Analysis: Domain Modeling", in: *Tutorial Notes of the Third International Conference on Software Reuse*, Rio de Janeiro, Brazil, Nov.

Gomaa, H., 1994, "A Prototype Domain Modeling Environment for Reusable Software Architectures", in: *Proceedings of the Third International Conference on Software Reuse*, pp. 74-83, Rio de Janeiro, Brazil, Nov.

Gomaa, H., 1995,"Domain Modeling Methods and Environments", in: *Proceedings of the Symposium on Software Reusability*, pp. 256-258, Seattle, Washington, Apr.

Griss,M., Favaro,J., Alessandro, M., 1998, "Integrating Feature Modeling with the RSEB", in: *Proceedings of the Fifth International Conference on Software Reuse*, Victoria, Canada, Jun.

Kitchenham, B., Pickard, L., Pfleeger, S.L., 1995, "Case Studies for Method and Tool Evaluation", *IEEE Software*, Jul, pp. 52-62.

Jacobson, I., Griss, M., Jonsson, P., 1997, *Software Reuse – Architecture, Process and Organization for Business Success*, 1st ed., ACM Press, New York.

Leite, J.C.S.P., 1993, *Eliciting Requirements Using a Natural Language Based Approach: The Case of the Meeting Scheduler Problem*, Texts on Computer Science, Computer Science Department, PUC-Rio, Rio de Janeiro, Brazil.

Leite, J.C.S.P., 1994, "Draco-Puc: A Technology Assembly for Domain Oriented Software Development" , in: *Proceedings of the Third International Conference on Software Reuse*, pp. 94-100, Rio de Janeiro, Brazil.

Leite, J.C.S.P., 1995, "Recovering Business Rules from Structured Analysis Specifications", in: *Proceedings of the Second Working Conference on Reverse Engineering*, pp. 13-21, Ontario, Canada, Jul.

Maiden, N.A.M., Rugg, G., 1996, "ACRE: Selecting Methods for Requirements Acquisition", *Software Engineering Journal*, May, pp. 183-192.

Neighbors, J.M., 1992, "The Commercial Application of Domain Analysis", in: *Proceedings of the Fifth Annual Workshop on Software Reuse*, Palo Alto, California.

Prieto, R., 1987, "Domain Analysis for Reusability", in: *Proceedings of the Eleventh Annual International Computer Software and Applications Conference*, pp. 23-29, Tokyo, Japan, Oct.

Prieto, R., 1993, "Software Reusability, Classification and Domain Analysis", in: *Mini Tutorial of VII SBES*, Rio de Janeiro, Brazil, Oct.

Prieto, R., 1994, "Software Reuse: From Concepts to Implementation", in: *Tutorial Notes of the Third International Conference on Software Reuse*, Rio de Janeiro, Brazil, Nov.

Roseti, M.Z., Murta, L.G.P., Werner, C.M.L., 1998, "A Tool for Configuring a Knowledge Acquisition Process in the Domain Analysis context", XII Brazilian Symposium on Software Engineering, 1a. Brazilian Exhibition of Academic and Commercial Software, Maringa, Brazil, Oct. 1998 (in Portuguese).

Scott, A.C., Clayton, J. E., Gibson, E., 1991, *A Practical Guide to Knowledge Acquisition*, 1st. ed., Addison-Wesley.

Simos, M.A., 1994, "An Introduction to Organization Domain Modeling", Tutorial Notes in: *Proceedings of the Third International Conference on Software Reuse*, Rio de Janeiro, Brazil, Nov.

STARS, 1996, *Organization Domain Modeling Guidebook*, Technical Report, USA, Jun.