

ISSN 0103-9741

Monografias em Ciência da Computação nº 13/93

Eliciting Requirements Using a Natural Language Based Approach: The Case of the Meeting Scheduler Problem

Julio Cesar S. P. Leite

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22453-900
RIO DE JANEIRO - BRASIL

PUC RIO - DEPARTAMENTO DE INFORMÁTICA

ISSN 0103-9741

Monografias em Ciência da Computação, $\,\mathrm{N}^{2}\,$ 13/93

Editor: Carlos J. P. Lucena

June, 1993

Eliciting Requirements Using a Natural Language Based Approach: The Case of the Meeting Scheduler Problem*

Julio Cesar S. P. Leite

* This work has been sponsored by the Secretaria de Ciência e Tecnologia da Presidência da República Federativa do Brasil.

In charge of publications:

Rosane Teles Lins Castilho
Assessoria de Biblioteca, Documentação e Informação
PUC Rio — Departamento de Informática
Rua Marquês de São Vicente, 225 — Gávea
22453-900 — Rio de Janeiro, RJ

Brasil

Tel. +55-21-529 9386

Telex +55-21-31048

Fax +55-21-511 5645

E-mail: rosane@inf.puc-rio.br

techrep@inf.puc-rio.br (for publications only)

Eliciting Requirements Using a Natural Language Based Approach: The Case of the Meeting Scheduler Problem

Julio Cesar Sampaio do Prado Leite*

Departamento de Informática

Pontifícia Universidade Católica do Rio de Janeiro

R. Marquês de S. Vicente 225 Rio de Janeiro 22453-900

e-mail: julio@inf.puc-rio.br

Brasil

March 1993

Abstract

Requirements elicitation is a process in which the software engineer works toward understanding and discovering what is expected from a planned software system. Although it is common belief that it is a hard task, several researchers are attacking the problem using different frameworks. One of these frameworks is based on the use of natural language. One strong reason for a natural language based framework is to avoid the introduction of artificial languages at the beginning of the software development process. We believe that, although using this framework, it is possible to achieve more precision. Our group has been investigating one approach, called the Language Extended Lexicon (LEL), that fits in such framework. We have used the LEL mainly for eliciting application vocabulary directly from actors in an application universe of discourse. In this study we report in a variation of our approach. We show how the LEL can help the elicitation of requirements when one is departing from a written document. The approach is supported by a hypertext system specially designed for the LEL.

1 Introduction

The overload of the term requirements is a reality. Different proposals and different views have been exposed regarding requirements. Our research has been paying special attention to the task of elicitation. We have proposed a viewpoint based validation scheme [10] to help the understanding of requirements¹ and have been studying the elicitation of application vocabulary [12]. The natural language framework [17], [13], [16] for dealing with specifications is observed as a long-standing minority [4]. Frameworks dealing with artificial languages or

^{*}This work is supported by CNPq

¹Reubenstein properly classified our proposal as a debriefing strategy [15].

representations have been prevailing [10] in the literature. It is important to point out that a natural language approach does not necessarily mean natural language understanding, and this is our case. Our usage of natural language in the elicitation of application vocabulary aims to rapidly reconnoiter requirements [3].

Critiques of natural language descriptions are frequent in the literature. One of these critiques is Meyer's [14] paper where he lists the seven sins of natural language specifications. Although no one would argue against the benefits of formalization, one can achieve more precision without relying on artificial languages. Formalization should be a gradual process to deal with the wide gap between reality and a formal model. We have showed that it is possible to impose more rigor on natural language descriptions without giving up on its presentation benefits [11].

Our group has been working on an application vocabulary acquisition strategy [5] based on what is called, the Language Extended Lexicon (LEL). The Language Extended Lexicon is a representation of the symbols in a language. We have used the LEL mainly for eliciting application vocabulary directly from actors in an application universe of discourse. In this study we report in a variation of our approach. We show how the LEL can help the elicitation of requirements when one is departing from a written document. We used as our case study the *The Meeting Scheduler Problem: Preliminary Definition* [6].

We organized our report in four more sections. One details the ideas behind vocabulary acquisition and the LEL. Section 3 describes the process used in applying LEL to a written document. Section 4 shows how the LEL raised issues related to the elicitation of the problem. We conclude pointing out aspects of integration of this work with other work in the area, highlighting the great difficulty in establishing the limits of a software system.

2 Language Extended Lexicon

Recently, Jackson and Zave [9] pointed out the importance of paying attention to the reality that is being described. They also observed the pitfalls of specifications that do not separate between a description of a property of a software system from the property of its environment. Their article refines the idea presented in JSD [8], where the important aspect is to model the reality, and **not** the functions of a demanded software system. In particular, Jackson and Zave propose a description structure, called identification, that aims to state explicitly what is being described. In building an identification Jackson and Zave observe:

... This place a salutary obligation on the specifier to choose to identify the most directly recognizable phenomena as the basis for the specification. Recognition of these phenomena then allows the user of the specification to orient it correctly with respect to the real domain, . . .

It is our point of view that what Jackson and Zave call recognizable phenomena is, very much similar to Carnap's observation language [1]. That is, the way we can ground further formalization of anything is when we anchor our formalization into something recognizable or observable in a given Universe of Discourse².

The main objective of vocabulary acquisition is helping the identification of recognizable phenomena. In our approach we work with the hypothesis of a closer relation between culture

²We believe that Universe of Discourse is a better name than Domain or Environment and it is an established name in the conceptual modeling of information systems literature

and language [1], [2]. As such, a given Universe of Discourse is bounded to have a proper language. Consequently if we identify the vocabulary of this language, we are identifying the recognizable phenomena that occurs in this Universe of Discourse.

Our approach to vocabulary acquisition draws its founding concepts from areas of knowledge unrelated to computer science, philosophy [1] and Semiotics [2]. Carnap's idea of the existence of an observation language and Eco's work on theory of codes are the basis for our strategy on vocabulary acquisition. The motto behind the strategy is:

understand the language of the problem, without worrying about understanding the problem.

As such, vocabulary acquisition focus on the language and not on the problem.

Central to our strategy is the task of lexicon construction, which requires both product level knowledge and process level knowledge. In terms of product, we centered our attention on what is called the Language Extended Lexicon (LEL). The Language Extended Lexicon is a representation of the symbols in the language. At the process level we have grouped a series of heuristics for eliciting, from the Universe of Discourse, the symbols and their meaning. We also proposed a validation strategy [12]. The LEL is a set described by the following postulates:

- (1) $LEL_{UofD} = \{ Entry \}^+$
- (2) Entry = Title { Notion }+ { Behavioral Response }+
- (3) Title = Sign | Title

More than one Sign in Title shows synonym.

- (4) Notion = $\{ \text{ Sign}, \neg \text{Sign} \}^+$
- (5) Behavioral Response = { Sign, ¬Sign }+
- (6) Sign $\cap \neg$ Sign $= \emptyset$

Sign is a word of the application language.

(7) $\neg \text{Sign} \in \mathcal{NL}$

 \mathcal{NL} is a small subset of a natural language dictionary. This subset is the "COBUILD Wordlist", words that are used ten times or more in explanations in the Collins COBUILD Dictionary.

Below we show an example of a lexicon entry.

meeting date

Notion:

- 1) The calendar date and the time period where participants will be together. Behavioral Response:
- 1) The calendar date of a meeting date belongs to the date range.
- 2) A meeting date is an element of the preference set of all participants.

This set oriented description is based on a very simple system of codes, proposed by Eco [2], with three different entities: signs³, notions and behavioral responses. Notion is the denotation, that is, the intended meaning. Behavioral response is the connotation, that is, an additional meaning for a word (a sign). In describing notions and behavioral response for a sign we established anenforcing rule that tries to maximize the use of signs in the meaning of other signs.

This recursive flavor of the representation scheme naturally drives the whole set to be as complete as possible, thus improving the chances of discovering missing signs. We call this the *principle of circularity*, since we maximize the use of signs of the LEL when describing a notion or a behavioral response.

The idea of a LEL being self-contained is also enforced by another rule which demands that external vocabulary be minimized and reduced to the smallest set as possible. We call this the *principle of minimal vocabulary*, and use as external vocabulary the words in the COBUILD Wordlist, a larger but well defined set.

Although we have pointed out that sign identification could be performed by reading or scanning a document, most of our work have been using unstructured interview and/or observation for finding signs. Sign meaning, that is assigning notions and behavioral responses, has been performed by means of structured interviews with actors in the Universe of Discourse.

In the case study reported here, a single text is the only information source for sign identification and for sign meaning. Next we will describe the main heuristics used when a written text is the only source available.

3 LEL Construction from Written Documents

The situation described here is one in which a textual description is the only source of information. The major guideline for this process is that the actor performing this LEL construction should restrain the use of previous knowledge (be impartial). The software engineer performing this task should assume the text as the Universe of Discourse. It is obvious that the description will not answer all the questions the lexicon builder may have, but it will help in the construction of an agenda for further investigation. We quickly detail the strategies for sign identification and sign meaning.

3.1 Sign Identification

Sign identification can be performed either manually or with the help of a software. An example of help is the program phrase finder [13], another one is the use of the regular

³Here we should not make a confusion between the usual denotation of the word sign – notice publicly displayed for advertising purposes or for giving directions – and the denotation – a symbol or a thing – used in this paper.

expression searching facility provided by Reads [18]. More intelligent schemas like lexical affinity [13] or contextual dependencies [7] can improve the automation of sign identification.

Although automation will help in the process, we strongly believe that the quality of the task is still strong dependent on whom does the job. Performing sign identification manually is a hard task, it requires much attention and time. Sign identification, by means of text reading, has three basic steps:

- read paragraph by paragraph,
- within each paragraph; read sentence by sentence,
- choose the words or phrases by paragraphs.

It is important to stress that we are using a local policy. We choose signs as they appear and not after reading the whole text.

3.2 Sign Meaning

Since our source of information is a closed and static system, we must gather all the sign meaning from within the text. The lexicon builder has to be careful and not mix notion and behavioral response. He/she should maximize the use of identified signs in the descriptions, and use the minimal vocabulary. Most important, the lexicon builder has always to make sure that new signs are consistent with previous entries. To ease the burden of constructing a lexicon we have designed and implemented HyperLex [12], a hypertext system that implements the LEL structure. To attach meaning to each elicited sign we are using the following heuristics [11].

- 1. A sign may have more than one notion and zero or more behavioral responses.
- 2. Each notion behavioral response has to be described as a short and direct simple sentence.
- 3. Notions or behavioral response for a given sign, may represent different viewpoints or may be complementary.
- 4. The short and direct simple sentences for notion or behavioral response description must obey the principle of circularity and the principle of minimal vocabulary.
- 5. For a sign that is a subject in a sentence, its notions have to make clear who the subject is. The behavioral responses should register the actions it executes.
- 6. For signs with a verb role, the notions should say who executes the action, when it happens, and the procedures involved in the action. For the behavioral response, they should identify: the constraints on the happening of an action, what the actions triggered in the environment are, and what the phrases (situations) caused by the actions are.
- 7. If a sign is the object in a sentence, the notion should define the object and identify other objects with whom it has a relationship. For behavioral responses, the actions that may be applied to this object should be described.
- 8. For the signs that express a state, the notions have to makeclear what it means, and the actions which triggered that sign. For behavioral responses, the description should identify other states and actions that can happen from this specific situation.

9. Signs that share the same meaning, that is that are viewed assynonyms in the UofD, are just an entry in the LEL. These signs are listed with the "/" symbol as a separation mark.

3.3 Informal Verification

Since we are assuming a closed world, where the only elements are the lexicon builder and the text, it is not possible to perform validation⁴. Since the text is not a formal entity it is also not possible to perform verification. Although these postulates are correct, the lexicon builder may found out, when attaching meaning to signs, missing signs, and as such needs to add new signs to the list of signs. We are calling this phenomena informal verification.

4 LEL and the Meeting Scheduler Problem

Our basic hypothesis is that LEL helps the elicitation of requirements. This is true even when using a single source of information, and as such limited and not much rich, but good for a first or initial debriefing. In this Section, we report on the observations made in using the LEL for understanding the Meeting Scheduler problem. We will give a taxonomy of problems found during the case study.

Using the strategy above, sign identification, we had identified sixty four (64) signs. They are: meeting scheduler, meeting date, meeting location, meeting request, participants, organizing meeting, meeting initiator, personal agenda, set of dates participants can not attend, exclusion set, set of dates participants would prefer, preference set, calendar date, time period, data range, provide equipment requirement, state preferences about meeting location, important participants, proposed meeting date, date conflict, weak conflict, extend the date range, remove dates from exclusion set, withdraw from the meeting, add new dates to preference set, meeting room, preferred location, new round of negotiation, representative, replan a meeting, propose a meeting date, propose a meeting location, modify preferred location, more important meeting, resolution policies, communicate requests, get replies, reacting promptly, make participants aware, authorized user, physical constraints, elapsed time, submission of a meeting, request, submission of date, location, communicate a meeting, date, location, actual date of, meeting, non-experts, professional meeting, private meeting, evolving data, status, priority, explicit priority, explicit dependencies, delegation, partial attendance, date format, address format, interface language, course schedules.

Using this list, we start to describe notions and behavioral responses for each sign. Of course in this process we had to alter the list due to missing signs or to exclusion of signs (informal verification). Due to the characteristics of this report, we will not show the entire LEL. We will show some entries, and explain how it helps the understanding of a given text, as well as sets an agenda for further investigation, or decision.

Meeting Scheduler was the first sign we described. Detailing its meaning, we found out that the sign strong conflict was not in our original list. We also found out, later, when describing organizing meetings, that it did not fit as a sign, since its meaning was too broad. As such, we later modify the Notion of meeting scheduler and marked only the sign meeting, since organizing belongs to the minimal vocabulary⁵.

⁴Validation is an empirical process where a social context accepts a given product as it satisfies its demands.
⁵A minimal vocabulary may be any set of well defined words, so if there is a set shorter than the COBUILD list and that satisfies the problem, great!

meeting scheduler:

Notion:

A system for organizing meetings.

Behavioral Response:

- 1) Produces a meeting date and a meeting locations for a meeting request.
- 2) Produces a strong conflict or a weak conflict for a meeting request.
- 3) Produces a new round of negotiation if no meeting location is available.

When describing **meeting location**, we had our first encounter with a kind of problem that is widespread over the text, vocabulary. There are cases in which, given the text and our *impartial* observation, it is reasonable to list two or more signs as synonym. On the other hand there are cases in which further investigation is necessary.

meeting location / meeting room / location

Notion:

- 1) An address for a meeting.
- 2) A meeting location should belong to the set of preferred location of important participants

Behavioral Response:

- 1) A meeting room should meet the requirements given by provide equipment requirement.
- 2) An important participant may modify preferred location.
- 3) A meeting room must be available for a selected meeting date of a meeting.

As an example of LEL construction, we will try to provide the rationale used for describing meeting location.

- 1. We scan the text looking for the appearance of the sign under scrutiny.
- 2. The sign is first referred as location on the first paragraph of the text. It does not provide much information at this point.
- 3. We used common sense⁶ to define the notion, that is, we used theminimal vocabulary sign address.
- 4. At paragraph four there is a reference to meeting location linking it to the signs: important participants, provide equipment requirement, and preferred location. As such we described a notion (2) showing the relation between location, participants and preferred location. A behavioral response (1) is described since the object location can be affected by the action provide equipment requirement.
- 5. At paragraph six (6) we found a constrain linking meeting, meeting date and meeting location. Although this relationship could be stated as notion, given our heuristics (Section 3), it is really a strong constraint. Because of that we choose to list it as behavioral response⁷.
- 6. At paragraph eight (8) we see that, important participants can modify preferred location, and list it as a behavioral response (2).

⁶At this point, we have a good example of inference. Although the text does not provide more information, the actor may use its intelligence to infer certain information. This attitude is not contrary to the general rule of being *impartial*, since this is a direct inference

⁷We should observe that we are dealing with flexible heuristics, there is no sense in being too hard on heuristics since the spirit of LEL is helping the requirements engineer without cutting his/her freedom

Overall, we had found four classes of problems in constructing a LEL for the Meeting Scheduler Problem that would require further investigation. These problems cannot be reasonable resolved without using other sources of information: vocabulary, lack of information, inconsistency, and confusion between the environment and the system. We will detail some problems found in each of these classes.

4.1 Vocabulary

Let's examine, for instance, the sign meeting request. One of the first listed signs was not described until we found out that it could be the same as: request a meeting, submission of a meeting and plan a meeting⁸.

Another example is the case of authorized user. When trying to describe it we found out that the text has referred to the signs, not listed firsthand, user and client. There is also a sign non-expert that is linked to these signs. Are those all the same? It is not possible to tell. It is also impossible to tell the relationship of these signs with participant, important participant and meeting initiator.

4.2 Lack of Information

When describing date range we asked ourselves if as meeting date should it also include the time period? We decided that it does not have a time period, but this information could be explicitly stated in the text.

date range

Notions:

1) A calendar date interval in which a meeting should occur.

Behavioral Response

- 1) The meeting initiator sets a date range for a meeting.
- 2) The meeting initiator extends the date range if a date conflict occurs.

The text mention resolution policies but nowhere we found another reference to this sign, and as such could not make an entry in the LEL for this sign. Another confusing sign is actual date of the meeting. What is exactly new round of negotiation? What is an active participant? It is also impossible to tell the relationship of this sign with participant, important participant and meeting initiator. Although it seems clear that the meeting scheduler is just for meetings within a calendar date, nowhere it is said that meetings could not last for more than one day.

4.3 Inconsistencies

In general the text is very much free of inconsistencies, much more than it could be expected of a free text, if you regard some of them as a vocabulary problem. One inconsistency we found wasthe case of the requirement regarding future use of status and priorities, eventhough the text makes a distinction between **ordinary participants** and **important participants**.

The set of participants is not well defined, it is not clear that all non ordinary participants are important participants.

⁸Although plan a meeting was not listed as a sign at first, when describing replan a meeting we noted that plan a meeting was not referred in the text, since using common sense it is obvious that if you replan then you must had planned before.

4.4 Confusion Between the Environment and the System

In several places, we had problems in understanding what is the environment in which such a meeting scheduler should work. Let's, for instance, look at the sign personal agenda.

personal agenda

Notion:

- 1) A set of calendar date and time period.
- 2) Each participant has a personal agenda.

Behavioral Response:

- 1) A personal agenda gives a exclusion set for a meeting.
- 2) A personal agenda gives a preference set for a meeting.

Using this description we are opening the possibility for a highly automated system, that is all the information about dates and times are already available for the meeting scheduler. On the other hand, the text mentions that the meeting initiator asks all potential meeting attendees for the information, but it does not mention any procedure or policy for doing this.

The text does not inform how people are chosen for meetings (invitation?). However the text assumes that participants are well integrated, "manage all interactions among participants required during the organization of a meeting" (paragraph 8.7). Another example is seen on the sign below.

provide equipment requirements

Notion:

- 1) Performed by active participants.
- 2) Active participants give the meeting initiator the equipment requirements to be available on the meeting location.

Behavioral Response:

- 1) Meeting initiator asks active participants to provide equipment requirements.
- 2) A new round of negotiation may be required if the meeting location does not have the equipment requirements.

Although the sign description is informative, several questions related to the relation with the outside world are pending. For instance: how does the participant give the information? do they send the information to the meeting initiator and then this person enters it in the system? Is the negotiation done on line?

Overall the description does not deal with the distinction between the environment and the system. In the only two sentence that deal with interface, they deal from the perspective of the user, and not from the perspective of the environment. In one it says that the system should be usable by non-experts, and in the other it talks about "interface language".

5 Conclusion

Our basic hypothesis is that LEL helps the elicitation of requirements. This is true even when one is just using a single source of information, and as such, limited information, but good for a first or initial debriefing. We have shown, by using the Meeting Scheduler description, that our hypothesis holds. The use of LEL for the task of an initial debriefing helps the understanding of a given description, and, as a side effect, organizes the requirements as a hypertext document.

Comparing this experience with previous case studies [Leite 92], we observe that here there was a lack of consistency in the vocabulary used. Eliciting vocabulary from the actors in the Universe of Discourse we noted that:

The actors in a Universe of Discourse naturally follow the principle of circularity as they try to explain the meaning of a sign.

This observation is valid from different Universe of Discourse, from a stock trader firm to a doctors' office. This is true because, usually, in such social settings there is an established language, and as such a vocabulary. A reason for the non existence of a well defined vocabulary may be that the description of the Meeting Scheduler was made by outsiders, that is people whose day to day business is not the organization of meetings. Another possible reason may be the fact that in writing usually the vocabulary is richer than in speaking. We believe that this case study has shown, again, the utility of a natural language based elicitation method for rapidly reconnoitering requirements.

Our work with LEL has already devised a partial automated scheme that derives a conceptual model from a given LEL [12], such that from a rapid reconnoiter to a detailed conceptual model we can count on automated support. We are studying the application of viewpoint resolution [10] as a validation strategy for the validation of a LEL.

We would like to conclude saying that we very much agree with Jackson and Zave [9], when they say that

We consider that a careful and explicit description of the environment is an essential first stage in a serious development.

In our study of the Meeting Scheduler, we had this problem and classified it as a major problem. We believe that there is a need for investigating methods and techniques for helping the explicit description of the environment.

References

- [1] R. Carnap, "The Methodological Character of Theoretical Concepts", Feigl, H. Scriven, M. Minnesota Studies in the Philosophy of Science. Minneapolis, 1956.
- [2] U. Eco, A Theory of Semiotics, University Press, Bloomington, 1979.
- [3] M. S. Feather, Requirements Reconnoitering at the Juncture of Domain and Instance, In Proceedings of the IEEE Internatinal Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [4] S. Fickas and A. Finkelstein, Requirements Engineering 93, In Proceedings of the IEEE International Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [5] A.P.M. Franco, Métodos e Representações de Suporte a Aquisição de Linguagens da Aplicação, Tese de Mestrado, Departamento de Informática, PUC-RIO, 1992.
- [6] A. Lamsweerde, R. Darimont, Ph. Massonet, The Meeting Scheduler Problem: Preliminary Definition, *Université Catholique de Louvain, Belgium*, 1993.

- [7] Y.Ishiara, H.Seki, T. Kasami, A Translation Method from Natural Language Specification into Formal Specifications Using Contextual Dependencies, In Proceedings of the IEEE Internatinal Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [8] M.A. Jackson, Systems Development, Prentice-Hall, 1983.
- [9] M.A. Jackson and P. Zave, Domain Descriptions, In Proceedings of the IEEE International Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [10] J.C.S.P. Leite and P.A. Freeman, Requirements Validation Through Viewpoint Resolution, In IEEE Trans. on Software Engineering, Vol. 17, N. 12, Dec. 1991.
- [11] J.C.S.P. Leite, Enhancing the Semantics of Requirements Statements, *Proceeding of the XII International Conference of the SCCC*, Santiago, Chile, 1992.
- [12] J.C.S.P. Leite and A.P.M. Franco, A Strategy for Conceptual Model Acquisition, In Proceedings of the IEEE Internatinal Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [13] Marrek and D. Berry, The Use of Lexical Affinities in Requirements Capture, 5th IWSSD, Pittsburgh, 1989.
- [14] B. Meyer, On Formalism in Specifications, In IEEE Software, Jan. 1985.
- [15] H.B. Reubenstein and R.C. Waters, The Requirements Apprentice: Automated Assistance for Requirements Acquisition. *In IEEE Trans. on Software Engineering*, Vol. 17, N. 3, Mar. 1991.
- [16] K. Ryan, The Role of Natural Language in Requirements Engineering, In Proceedings of the IEEE International Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.
- [17] M. Saeki, H. Horai and H. Enomoto, Software Development Process from Natural Language Specification, In 11th International Conference on Software Engineering, IEEE Computer Society, Pittsburgh, 1989.
- [18] T.J. Smith, READS: A Requirements Engineering Tool, In Proceedings of the IEEE Internatinal Symposium on Requirements Engineering, RE-93, IEEE Computer Society, San Diego, Cal. Jan. 1983.