CREWS Report Series 99 - 19

EVALUATING THE CREWS-*L'ECRITOIRE* REQUIREMENTS ELICITATION PROCESS

Mustapha Tawbi, Fernando Velez, Carine Souveyet,
Camille Ben Achour

CRI, Université de Paris1- Sorbonne, 90, rue de Tolbiac, 75013 Paris

Submitted to ICRE2000 Fourth IEEE International Conference on Requirements Engineering, Schaumburg, Illinois, USA, June 2000

Evaluating the CREWS-*L'Ecritoire* Requirements Elicitation Process[1]

Mustapha Tawbi, Fernando Velez, Camille Ben Achour, C. Souveyet
Centre de Recherche en Informatique
90, rue de Tolbiac
75013, Paris - France

{ tawbi, velez, camille, souveyet}@univ-paris1.fr

Abstract *: the CREWS 21.903 long term research ESPRIT project has proposed several strategies to support requirements elicitation through textual scenarios analysis. In the CREWS-L'Ecritoire approach, guidelines are proposed to systematise these strategies. However these guidelines have never been evaluated so far. To evaluate the effectiveness of the CREWS-L'Ecritoire requirements elicitation guidelines, an empirical study was undertaken : several subjects were asked to apply the CREWS requirements elicitation strategies with and without guidance. The results of this experiment indicate that : (i.) subjects apply the CREWS-L'Ecritoire guiding rules with different rates of efficiency, (ii.) in average, all the guiding rules improve the subjects' ability to elicit correct requirements, and (iii.) each of the guiding rule has a different rate of efficiency. The paper presents the protocol used to conduct the experiment, and details the evaluation of these results for each of the three CREWS scenario based requirements elicitation strategies.*

1 Introduction

Several proposals *coupling goals and scenarios* have been recently made to overcome some of the deficiencies and limitations of goal-based and scenario-based requirements elicitation approaches used in isolation [Framework][CREWSRep97-10][Holbrook 90] [Potts 94] [Cockburn 95] [Anton 96] [Leite 97] [Haumer 98] [Van Lamsweerede 98] [Rolland 98]. Within this family of approaches, the CREWS-*L'Ecritoire* approach [Rolland 98] proposes to exploit a bi-directional coupling between goals and scenarios to support requirement elicitation. On the one hand, when a goal is discovered, the approach proposes to author a scenario to illustrate it ; this applies the coupling in the forward direction. On the other hand, the approach proposes to analyse every scenario to yield new goals ; this applies the coupling in the backward direction. Starting from a high level problem statement, the CREWS-*L'Ecritoire* approach guides the top-down discovery of a complete hierarchy of goals illustrated by scenarios. To systematise the guidance of this process a set of guidelines has been proposed. The guidelines consist first in *automated rules* to guide goal discovery, and second in *writing guidelines* and *linguistic analysis and verification rules* to guide scenario authoring.

So far, the CREWS-*L'Ecritoire* approach was described into details [Rolland98] [Rolland99] [BenAchour98] [TheseCBA99], its process modelled [SiSaid98][Benjamen99], implemented in a software prototype [Tawbi99], and the scenario authoring guidelines were evaluated [Ben Achour 99]. In addition, the CREWS-*L'Ecritoire* goal discovery rules were already partly evaluated. First, two workshops were conducted with 52 requirements engineers and software engineers from French companies who were asked to evaluate the usefulness of the

[1] The work presented in this article is partly funded by the European Community within the framework of the ESPRIT LTR project CREWS

different features of the CREWS-L'Ecritoire approach. A very positive evaluation (reported in [TSE99]) have emerged from the workshops. Second, the CREWS-*L'Ecritoire* approach has been applied to study the introduction of new businesses in an electricity supply company within the framework of the European market deregulation [Nurcan 97][Rolland 98]. Third, several one-day tutorials were undertaken with graduate students of the University of Paris 1 – Sorbonne. During these tutorials the subjects were asked to use the *Ecritoire* tool environment implementing the CREWS-*L'Ecritoire* strategies. Longitudinal observation and discussion with the subjects showed at first some difficulties to use the guiding rules, but an increasing efficiency all along the use of the tool. Globally these experiments have shown positive results concerning the effectiveness of the approach, and have demonstrated the scalability of the overall approach to real size problems. Nevertheless, there is still little evidence concerning the effectiveness of each individual CREWS-*L'Ecritoire* guiding rule. This paper reports the results of an empirical study undertaken with 41 subjects to evaluate the effectiveness of the goal discovery support within CREWS-*L'Ecritoire*.

The outline of the paper is the following : the next section overviews the CREWS-*L'Ecritoire* requirements elicitation approach, section 3 defines the working hypotheses taken on the CREWS-*L'Ecritoire* features. Beyond a simple description of the experiment protocol, section 4 sets up the rationale of the protocol design, presents the experienced subjects, and defines the marking scheme of the subjects' responses. The results of the evaluation are reported in section 5. Possible impacts of these results on the design of the CREWS-*L'Ecritoire* approach and software prototype are foreseen in the concluding section.

2 Guiding requirements elicitation in CREWS-*L'Ecritoire*

This section first presents some of the key concepts and terminology of the CREWS-*L'Ecritoire* approach and then provides an overview of its process. The purpose here is to give an overview of the elements which were evaluated during the experiment. An in-depth description of the process can be found in [TSE98].

2.1 Concepts and terminology

A *Requirement Chunk* (RC) is a pair <G, Sc> where G is a goal and Sc is a scenario. Since a goal is intentional and a scenario is operational in nature, a requirement chunk is a possible way of achieving the goal.

A *goal* is defined as *"something that some stakeholder hopes to achieve in the future"* [ICSP98]. In the CREWS-*L'Ecritoire* approach, goals are expressed as clauses composed of a main verb and several parameters, where each parameter plays a different role with respect to the verb. For example the goal :

Provide $_{\text{verb}}$ (efficiently) $_{\text{quality}}$ (electricity) $_{\text{object}}$ (from PPC producer) $_{\text{source}}$ (to our non eligible customers) $_{\text{beneficiary}}$ (using PPC network) $_{\text{means}}$ (in a normal way) $_{\text{manner}}$

Is composed of a verb and of six parameters, each playing a different role in the goal. CREWS-*L'Ecritoire* proposed a list of 12 roles, which description, definition and use example can be found in [Prat97].

(Cooperative Requirements Engineering With Scenarios) 21.903.

A *scenario* is *"a possible behaviour limited to a set of purposeful interactions taking place among several agents"* [ICSP98]. Every scenario is thus composed of one or more *actions*, an *action* being an interaction *from* one agent *to* another. In addition, the combination of actions in a scenario describes a unique path. In CREWS-*L'Ecritoire* scenarios are formally defined according to a conceptual an a linguistic view [TheseCBA][TSE98]. One important aspect of the scenario definition is that every scenario describes a single path within the set of behaviours the agents it involves can have. In the CREWS-*L'Ecritoire* approach, goals can be stated in a way precise enough to identify individually each scenario it is attached to.

Requirement chunks classification and *levels of abstraction* : CREWS-*L'Ecritoire* introduces three levels of abstraction called the *contextual*, the *functional*, and the *physical* level. The contextual level identifies the services that a system should provide to an organisation and their rationale. The functional level focuses on the interactions between the system and its user to achieve the needed services. Finally, the physical level deals with the actual performance of the interactions. Each level corresponds to a type of RC. As a result, the collection of requirements can be organised in a three level abstraction hierarchy.

Relationships between requirement chunks : to organise scenarios, three types of relationships among Requirement Chunks (RC) namely, the composition, the alternative, and the refinement relationship. The first two of these are extensions of conventional AND/OR relationships between goals leading to a horizontal AND/OR structure between RCs. AND relationships among RCs link together those chunks that require each other to define a completely functioning system. RCs related through OR relationships represent alternative ways of fulfilling the same goal. The refinement relationship relates requirement chunks at different levels of abstraction; thus, it establishes a vertical link between requirement chunks.

2.2 The requirements elicitation process

The CREWS-*L'Ecritoire* process aims at discovering/eliciting requirements by exploiting a bi-directional coupling between goals and scenarios. This coupling allows movement from goals to scenarios and vice-versa. When a goal is discovered, a scenario is authored to illustrate it. In this sense, the goal-scenario coupling is exploited in the forward direction from goals to scenarios. Once a scenario has been authored, it is analysed to yield goals. This leads to goal discovery by moving along the goal-scenario relationship in the reverse direction. The exact sequence of steps of the CREWS-*L'Ecritoire* requirements elicitation process is :

1. Initial Goal Identification
repeat 2. Goal Analysis
3. Scenario Authoring
4. Goal Discovery
until all goals have been elicited

It can be seen that goal elicitation and scenario authoring are complementary steps and goals/requirements are incrementally discovered by repeating the goal-analysis, scenario-authoring, goal-discovery cycle. As mentioned in the introduction, the objective of the CREWS-*L'Ecritoire* approach is to guide this process. For this purpose, each of the three steps of the cycle is supported by systematic mechanisms. These are detailed in the next subsection.

2.3 Detailed presentation of the CREWS-L'Ecritoire goal discovery strategies

The CREWS-*L'Ecritoire* guidance mechanism for goal analysis is based on a linguistic analysis of goal statements. It helps in reformulating a narrative goal statement to conform to a *goal template*. The goal template defines a goal structure according to the formalism introduced in section 2.1. The mechanism for scenario authoring combines style/content guidelines and linguistic devices. The former advise authors on how to write scenarios whereas the latter provides semi-automatic help to check, correct, conceptualise, and complete a scenario [EuroJap98]. Finally, for goal elicitation, CREWS-*L'Ecritoire* proposes rules based on a systematic analysis of the authored scenarios. Three different goal discovery strategies are proposed, namely the *refinement* strategy, the *composition* strategy, and the *alternative* strategy. The first of these helps discovering goals at a lower level of abstraction than a given one ; the second discovers goals ANDed to the original one ; the last discovers goals ORed to the original goal.

Each of the three next subsections presents two rules respectively proposed in the CREWS-*L'Ecritoire* approach to support each of the three goal discovery strategies. The presentation of the guiding rules is in two parts : first, the guiding rule is named and decomposed into steps, then, the steps are commented. Further comments and examples illustrating the application of the guiding rules can be found in [TSE98].

2.3.1 CREWS-L'Ecritoire rules supporting the refinement strategy

- **Refinement guiding rule (R1)**

The principle of the refinement guiding rule R1 is to use every atomic action (or interaction) in a given scenario Sc illustrating a given goal G, and to consider it as a goal at a lower level of abstraction. The rule is composed of four steps :

> Step1 : Associate a goal Gi with every atomic action Ai in Sc. Gi refines G
>
> Step2 : Complement Gi by the manner 'in a normal way'
>
> Step3 : Evaluate the resulting menu of goals Gi and select the goals of interest
>
> Step4 : Requirement chunks corresponding to these selected goals are ANDed one another, and refine G

The refinement mechanism underlying the rule treats every interaction between two agents in the scenario Sc as a goal for the next lower level of abstraction (step 1). This requirs a rephrasing which is achieved at step 2. The goals of interest which can be selected from the resulting menu of goals (step 3) are associated to requirement chunks and integrated into the requirement chunk hierarchy at hand according to step 4.

- **Refinement guiding rule (R2)**

The refinement guiding rule R2 aims at refining a given requirement chunk <G, Sc> by suggesting new actions in Sc that could be looked upon as goals refining G.

> Step1 : Type Sc actions according to the classification (information provision/request, service provision/request, condition evaluation action/constrained flow of actions),
>
> Step2 : Construct action pairs for each of the three types,
>
> Step3 : Detect missing actions and update Sc accordingly,
>
> Step4 : Consider Refined goals to G for every added action,
>
> Step5 : Evaluate and name the selected goals.

The rule R2 uses three classes of dependent action pairs : (service request, service provision), (information request, information provision) and (condition evaluation action, constrained flow of actions). The rule proposes to use these classes to type the actions of the analysed scenario. There exists a dependency among the actions of each of the three proposed pairs : any service request implies at least one service provision, an information request implies at least one information provision and a constrained flow of actions implies an action which evaluates the condition. Using these dependencies, pairs of actions are constructed (step 2), and every incomplete pair suggests a missing action (step 3) which can be inserted in the scenario. In step 4, it is additionally recommended to consider new actions as goals refining the one at hand; the resulting menu of goals has to be evaluated and the selected goals named correctly (step 5).

2.3.2 CREWS-L'Ecritoire rules supporting the composition strategy

- **Composition guiding rule (C1)**

The composition guiding rule C1 helps discovering goals that complement an initial goal G, based on the inclusion of final and initial states of the associated scenario Sc.

> Step1 : Check inclusion of Sc initial states in Sc final states
>
> Step2 : For every initial state *Is* that is not included in the final states
> point out the final state *Fs* hindering the reaching of the initial state
>
> Step3 : Identify a recovery scenario having *Fs* as part of its initial states and *Is* as part of its final states. Name, and select the associated recovery goal

According to the inclusion property, the initial states of a scenario must be included in its final states for ensuring a self-contained functioning, i.e every scenario execution should leave the concerned agents in a state which permits the repeated execution of the same scenario. The rule suggests to check if the inclusion property holds (step 1). If this is not the case then, this means that an unexpected state has been reached (step 2). Once the rule has detected the needed recovery scenarios (step 3), it proposes to qualify them by goals to be ANDed to the initial goal G.

- **Composition guiding rule (C2)**

The guiding rule C2 uses a classification of the interactions described in a given scenario Sc to support the discovery of goals ANDed to (i.e. which are complementary to) the associated goal G.

> Step1 : Identify the interaction objects in Sc that correspond to resources
>
> Step2 : Construct interaction pairs (*Consume*, *Produce*) for each identified resource
>
> Step3 : Suggest a new goal ANDed to G for every incomplete pair (i.e. in which either the Consume interaction or the Produce interaction is missing)
>
> Step4 : Name and select the relevant goals

The composition guiding rules C2 proposes first to look for interactions which objects are physical resources (step 1). Applying for each resource the producing/consuming principle, the rule suggests to construct pairs of interactions in which the resource is once consumed, once produced (or conversely produced then consumed) by the engineered system (step 2). Every incomplete pair originates a new goal (step 3) that can be accepted or not as part of the collection of goals and integrated accordingly into the hierarchy (step 4).

2.3.3 CREWS-L'Ecritoire rules supporting the alternative strategy

- **Alternative guiding rule (A1)**

The guiding rule A1 uses the goal G of an initial requirement chunk <G, Sc>, to discover goals ORed to it.

> Step1 : Identify the structure of the goal G according to the goal template
>
> Step2 : Provide possible alternatives for each parameter of the goal G
>
> Step3 : Compute all possible combinations of parameters
>
> Step4 : Evaluate, rephrase and select the goals of interest

This discovery rule exploits the goal structure defined by the goal template mentioned in section 2.1. First, the parameters of the goal have to be identified. For example, the goal *'Provide cash to our bank customers from ATM'* has the following structure :

> *'Provide (cash)*$_{Result}$ *(to our bank customers)*$_{Destination}$ *(from account)*$_{Source}$ *(with a card based ATM)*$_{Means}$*'*

This leads in the above example, to introduce the source (*from account*)$_{Source}$ and to specialise the means (*with a card based ATM*)$_{Means}$. Then, alternative values must be proposed for each parameter (step 2). All possible combinations (step 3) of the alternative values of each parameter correspond to potential goals; a cartesian product is thus computed. The resulting goals have to be evaluated and those of interest selected (step 4) The resulting collection of goals correspond to design alternatives of G; all of them are thus ORed.

- **Alternative guiding rule (A2)**

The alternative rule A2 exploits the flow structure of a given scenario Sc. As mentioned in the scenario definition (section 2.1), every scenario describes a unique path within the set of possible behaviours. Given a requirement chunk <G, Sc>, the rule A2 helps identifying alternative ways of achieving the goal G.

Step1 : Scan Sc to construct a graph identifying the path of actions it describes

Step2 : Complete the graph using information from scenario descriptions associated to goals having the same parameters as G except the manner

Step3 : Compute all possible missing paths

Step4 : Select the cases of interest and associate each of them with a specific manner to achieve G

The goals discovered using the guiding rule A2 are similar to the initial goal G : their verb, and most of their parameters are the same. However, every goal discovered by the guiding rule A2 identifies an alternative manner of achieving G. Therefore, to each of them is associated a different manner. For example, *'Withdraw cash from ATM in a normal way'*, *'Withdraw cash from ATM in a normal way with code error correction'* are goals that rule A2 aims at discovering. The graph constructed by the rule represents all possible behaviours already identified in the scenario of the initial requirement chunk (step 1) and in the scenarios associated to all the alternative requirement chunks (step 2). Every path is characterised in the graph by *zero* to *n* nested flow conditions. The graph is considered incomplete when there exists a flow condition without a path tackling its negation. The rule suggests to compute all the combinations of negated conditions, and to investigate them as possible missing paths (step 3). Each suggested missing path can be characterised by a manner to deal with the initial goal G; the corresponding goals can thus be stated and selected if they are found relevant (step 4).

The next section presents the hypotheses underlying the CREWS-*L'Ecritoire* goal discovery rules, and goal analysis support. These hypotheses are then taken into account in the design of the experiment presented in section 4.

3 Experimental hypotheses

Each of the hypotheses underlying the theoretical research undertaken in CREWS-*L'Ecritoire* relates to one of the features presented in the previous section. The aim of the experiment is to evaluate these hypotheses to check whether CREWS-*L'Ecritoire* is effective or not.

H1 : the goal template helps formulating *more precise* goals

H2 : the guiding rule A1, at the contextual level, helps *better envisioning* the future system

H3 : the alternative strategy helps finding *more variations* than an ad hoc process

H4 : the alternative strategy helps better finding *variations which are at the same level of abstraction*

H5 : the alternative strategy helps *better separating concerns* of alternative behaviour descriptions

H6 : the composition strategy helps finding *more system functions* than an ad hoc process

H7 : the composition strategy helps better finding *system functions* which are *at the same level of abstraction*

H8 : the composition strategy helps better *separating system functions*

H9 : the three predefined levels of abstraction help better *preserving the consistency of action descriptions* within a scenario

The following section describes the protocol of the experiment, and details the marking scheme applied to evaluate empirically these hypotheses.

4 Experiment

41 software engineers were requested each to elicit requirements for the design of a Digital Video Disk distribution Machine (DVDM in short). Expected requirements included delivering a customer's club card, checking the validity of the card, filling-in the machine with DVDs, reporting transactions to the central information system, printing a receipt and controlling the stock replenishment level. This problem domain was chosen for its novelty, thus maximising the potential envisionment of how the system could behave.

4.1 Subjects of the experiment

The 41 subjects (26 Male and 15 Female) were full-time or part time post graduate students in Information System engineering at the University Paris 1 - Sorbonne, France. In both cases, the students had professional experience in Information Systems. Their knowledge extended from object-oriented methods such as UML [Fowler 1997], OBJECTORY [Jacobson 1995], OMT [Rumbaugh et al. 1991], O* [Brunet 1993], or Remora [Rolland 1988]. The subjects, aged between 22 and 46 with an average of 25, volunteered their service and received no financial reward. None of the subjects had prior knowledge of the CREWS-*L'Ecritoire* approach. However, all had received a half day seminar on use cases.

4.2 Experiment protocol

All the subjects were requested to follow a requirements elicitation process comprising five *steps*. Each step was dealing with a specific issue, namely :

- **S1** : formulating goals
- **S2** : envisioning design options that are alternative to a given one
- **S3** : eliciting variations of use of a given system functionality
- **S4** : eliciting functions that are complementary to a given one
- **S5** : organising and tracking scenarios through different levels of abstraction

At each step, the subjects were requested to perform one or several tasks. All subjects went twice through each task of the five steps ; and were requested to answer the same question twice. In the first round, the subjects were not guided. In the second round, they were provided with the appropriate CREWS-*L'Ecritoire* guiding rules. Therefore they dealt with the same issue twice. This protocol was chosen (rather than dividing the subjects into two groups with and without the CREWS-*L'Ecritoire* guiding rules) : (a) the small number of subjects did not allow to divide them into two subgroups of significant size, (b) this protocol allows to evaluate a larger sample of responses, and (c) the CREWS-*L'Ecritoire* guiding rules help performing tasks which can also be performed manually in the working context; our goal was to evaluate the *improvements* due to the CREWS-*L'Ecritoire* guiding rules rather than compare the CREWS-*L'Ecritoire* approach to a non guided requirements elicitation way of working. It must in particular be noticed that all the results reported in the following emphasise the improvements brought by CREWS-*L'Ecritoire*, as they would be obtained in a practical requirements engineering context.

The experiment was scheduled as follows : prior to undertaking the task, experimental instructions (see [webAppendix]) were read by the experimenter ; for example, the subjects were requested to avoid backtracking from the second round to the first one. At the beginning of each step, subjects received a document (see [webAppendix]). Then they were : (i.) requested to achieve the first round of tasks in a given slot of time, (ii.) requested to read the CREWS-*L'Ecritoire* guiding rules proposed to systematise the work, and (iii.) asked to apply them for the same duration as (i.).

Hypotheses H1 to H9 were evaluated using the subjects responses obtained at the end of the different tasks performed during the steps S1 to S5. Table 1 presents the scheduling of tasks performed by the subjects during the five steps. Notice that some tasks (e.g. T3 and T5) were used to evaluate several hypotheses at the same time.

Table 1 : scheduling of the tasks performed by the subjects during the experiment

Step	Task	Hypothesis
S1 – formulating goals	T1 - check goal correctness and rephrase when necessary	H1
S2 – envisioning design options that are alternative to a given one	T2 - identify design alternatives from a given goal	H2
S3 - eliciting variations of a given system functionality	T3 - identify the possible variations of a given scenario T3 - identify the possible variations of a given scenario T4 - check the absence of several paths in a scenario	H3 H4 H5
S4 - eliciting functions that are complementary to a given one	T5 - identify complementary system functions from a goal and a scenario T5 - identify complementary system functions from a goal and a scenario T6 - check that a scenario describes the use of a single system function	H6 H7 H8
S5 – organising and tracking scenarios through different levels of abstraction	T7 - check the homogeneity of the level of abstraction in a scenario	H9

The slots of time allocated to perform each round of each task in each step were determined during a pilot study. The whole duration of the experiment was of 2 hours ½. This excludes the distribution and collection of the work documents. Each of the tasks T1 to T7 is described in turn in the next subsection.

4.3 Tasks performed by the subjects during the experiment

- **Task T1**

During this task, a set of 9 requirements was proposed to subjects. These requirements were at different levels of abstraction. However, some of these requirements were formulated according to the goal structure in Crews-*L'Ecritoire* and some were not. Subjects were asked (a) to rephrase requirements as goals, and then to select from this set, the goals that the system should satisfy, and (b) to eliminate incorrect requirements.

For example, requirement like *'DVD location'* was considered as incorrectly formulated. Subjects were expected to rephrase it correctly for example as *' Loan DVDs from a DVDM with a special card'*. Requirements like *' to be near to the Video club'* were considered as non relevant because they express non functional requirements and mismatch the goal template expected in Task T1.

As for all activities, subjects had to realise the task T1 twice. The first round they had to use their own knowledge and goals. The second round, they were requested to use the CREWS-L'Ecritoire goals templates and example of goal statement. At both rounds, the answers had to be motivated.

- **Task T2**

T2 aimed to evaluate if using the rule A1 helps maximising the potential envisionment of a system. During this Task subjects were provided with a single system goal:

> *' Provide location services to users from an interactive DVDM with a card'.*

Then, they were asked to identify several design alternatives (i.e. system goals expressing other ways to build the future system). For example the goal *'Provide location and sale services to users from DVDM with a credit card'* was considered as a correct design alternative of the initial goal. At the second round, subjects were provided with rule A1 to realise the task

- **Task T3:**

The aim of task T3 was to evaluate both hypotheses H3 and H4. During this Task the following scenario was provided to subjects :

Initial State :

> The client has a card
> The DVDM is in service

The client inserts a card into the *DVDM*, the *DVDM* verifies the card validity. If the card is valid, the *DVDM* verifies the client validity. If the client is valid, the *DVDM* displays a message asking the client to enter the title of the desired movie. The client communicates the movie title to the *DVDM* and the *DVDM* verifies if there is an available DVD. If one DVD is available, the *DVDM* records the DVD location, the *DVDM* ejects the card and the DVD to the client. The *DVDM* prints a receipt to the client. If the DVD store is not empty, the *DVDM* displays the message 'In service' to the next client.

Final State :

> The client has a card
> The client has a DVD
> The *DVDM* is in service

Subjects were asked to propose variants of this scenario. At the second round, subjects were provided with the rule A2. According to the rule A2 four alternatives can be considered. These alternatives correspond to the cases where :

> *'the card is not valid'*
>
> *'the card is valid and the client is not'*

'the card and the client are valid but the requested DVD is not available'
'the card and the client are valid and the requested DVD is available but the DVD store is empty'

- **Task T4**

During this task subjects were provided with a scenario describing the use of the *DVDM* system to loan a DVD. The scenario was, on purpose, incorrectly written; indeed, two cases were considered at the same time. Subjects were excepted to identify two paths and thus, to decompose the material in two scenarios, each describing one of the two cases. Between the two rounds of T4, the CREWS-*L'Ecritoire* scenario definition was provided to subjects.

- **Task T5**

The task T5 was designed to evaluate both of hypotheses H6 and H7. During this Task, subjects were provided with the same scenario provided in Task T3 , then they were asked to find functions complementing to the one described in the scenario at hand. At round1, subjects had no guidelines to perform the Task; at round 2, they were provided with the CREWS-*L'Ecritoire* rules C1 and C2. According to rules 'C1' and 'C2' two system functions could be discovered :

' Fill in the DVDM with DVDs'
'Fill in the DVDM with receipt paper'

- **Task T6**

During task T6, subjects were provided with a scenario illustrating the use of three different system functions. Subjects were asked to evaluate the scenario validity and to justify their answers. Between the two rounds of the task, the CREWS-*L'Ecritoire* scenario definition was provided to subjects.

- **Task T7**

Task T7 aimed to evaluate if the three predefined CREWS-*L'Ecritoire* levels of abstraction helps in organising scenarios. Subjects were asked to classify the actions of a scenario they were given according to the level of concern. A justification of the answer was requested from the subjects.

Between the two rounds a definition of the three CREWS-*L'Ecritoire* levels of abstraction was provided and at round 2, subjects were asked to classify each action into one of the three predefined levels.

For the sake of place, the whole set of documents distributed during the experiment are not provided here, but they can be found in the electronic annex [webAppendix]. The following subsection describes the marking scheme adopted to evaluate whether the subjects' results were confirming the hypotheses H1 to H9 or not.

4.4 Marking scheme

To evaluate the aforementioned set of hypotheses, each subject received different scores. The following describes how the different scores were calculated for each of the evaluated hypotheses.

Hypothesis H1 : Two scores were affected to each subject to evaluate the aid provided by the goal template during goal formulation : (i.) a *goal template applicability* score, and (ii.) a *goal formulation preciseness improvement* score.

To measure the *goal template applicability score*, it was necessary to count for each subject : (a) how many incorrect goals provided in the task T1 the subject was able to identify with the goal template but not without, and (b) how many incorrect goals the subject was able to rephrase more precisely with the goal template than without.

The precision of goal formulation was measured by the number of correct parameters provided by the subject in the goal statement. For example, starting from the goal :

'The client renews his subscription'

given to the subjects in the task T1, the rephrased goal :

'Renew clients subscriptions'

was considered as more correctly stated than the goal :

'Clients could renew their subscriptions'

Indeed, the former contains more relevant parameters than the latter. In addition, the goal template was considered as properly applied for each goal that was rephrased in the second round (with the goal template) but not in the first round (without the goal template).

To measure the *goal formulation preciseness improvement score*, each subject's response was checked for : (a) the difference between the number of goals precisely rephrased with the goal template, and the number of goals precisely rephrased without the goal template, and (b) the difference between the number of goals incorrectly rephrased even with the goal template, and the number of goals incorrectly rephrased even without the goal template.

Hypothesis H2 : the *quality of the envisionment of the future system* by the subjects was evaluated by counting the number of correct design options proposed by the subjects as alternatives of a contextual goal given to them in the task T2. The subjects were requested to state these design options as goals. The subject were supposed to be aware of the way to formulate goals correctly since they had already been provided with a goal formulation template in the task T1 (step 1). Therefore, the correctness of goals was evaluated : (a) with respect to their form (i.e. by checking that they respected the goal template), and (b) with respect to their content (i.e. by checking that they were really suggesting design options of the provided goal). The more correct goals were proposed by the subjects, the highest was scored the quality of the future system envisionment.

Hypotheses H3 and H4 : the variations of the scenario given to subjects in task T3 were evaluated with respect to their correctness. A subject's variation was considered as correct if : (a) it identified alternative behaviours of the system and its users. These behaviours were excepted to lead to the achievement of the goal proposed in task T3 ; (b) it was correctly stated, i.e. it identified the cause of the alternative behaviour and not its effect. For example, the subject's variation *the card is not valid'*

was considered as a possible cause of alternative behaviour, and thus counted as a correct variation, whereas the variation :

'*the card is not ejected'*

was considered as a consequence, and thus counted as an incorrect variation. Indeed, the card may not be ejected for several reasons ; each of these correspond to different requirements that have to be identified independently ;

it was at the same level of abstraction as the scenario proposed to subjects in task T3. For example a system dysfunction corresponds to a possible cause of alternative behaviour at the system internal level. It was thus, considered as an incorrect variation of the system interaction level scenario provided to the subjects in task T3.

Hypothesis H5 : the ability of subjects to separate concerns of alternative behaviour descriptions was evaluated by checking whether subjects were able to identify the two paths described in the scenario they were given in task T4.

Hypothesis H6, H7 : the system functions identified by the subjects as complementary functions to the one of given in task T5 were evaluated with respect to their correctness. A subject's function was considered as correct if : (a) it did not correspond to a variation of the scenario provided to the subjects in task T5 , and (b) it was at the same level of abstraction as the scenario proposed in T5. As the initial scenario was at the system interaction level, the subjects' functions were expected to be at the system interaction level too. They had thus, to correspond to services provided by the future system to its users. In consequence, the function :

'*Verify the card validity'*

was counted as incorrect.

Hypothesis H8 : the ability of subjects to separate concerns of complementary system functions was evaluated by checking whether subjects were able to identify the three system functions described in the scenario they were given in task T6.

Hypothesis H9 : two scores were affected to each subject to evaluate the aid provided by the predefined levels of abstraction during the action description consistency checking : (a) a score of *applicability of the predefined levels of abstraction*, and (b) a score of *performance in the identification of inconsistent action descriptions*.

The applicability score was measured by the number of subjects able to identify actions belonging to different levels of abstractions. The performance score was measured by the number of actions given in task T7 that the subjects were able to classify in the correct level of abstraction.

5 Evaluation results

The evaluation of hypotheses H1 to H9 was driven by a quantitative measurement of the difference between the subjects' results at tasks T1 to T7. The following presents the main findings for each of the aforementioned hypothesis.

5.1 Detailed evaluation of hypothesis H1

The subjects' results for task T1 validate significantly hypothesis H1 in two respects : first the goal template does help differentiating goals from other kinds of requirements and second, it helps reformulating goals correctly. These two aspects are dealt with in turn in the following.

During task T1, the subjects were first asked to identify correct goals from a list of given statements. At the end of the first round (i.e. without guidance), 65% of subjects were able to identify at least one correct goal. In round 2 (i.e. using the goal template) this proportion is increased to 90%. In average, the goal template improved the subjects' ability to identify correct goals by 25%.

Moreover, the subjects' results show that the goal template improve the ability to correctly reformulate goals that were incorrectly stated. Indeed, 82% of the subjects have correctly reformulated at least one more goal in round 2 than in round 1 of task T1. In average, their ability to reformulate correctly goals is increased by 46%.

Figure 1 compares the proportion of subjects having correctly reformulated from 0 to more than 6 goals during the two rounds of task T1. The figure shows that owing to the goal template, 44% of subjects were able to correctly reformulate more goals than the best subject could do without guidance.

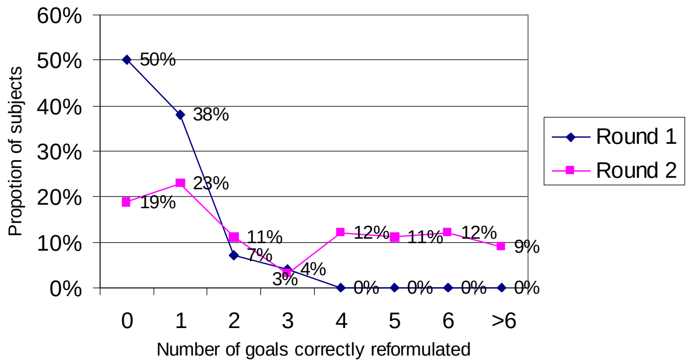

Figure 1 : subjects' performance in goal reformulation

5.2 Detailed evaluation of hypothesis H2

During task T2, the subjects were asked to find design alternatives with and without the guiding rule A1. Figure 2 compares the number of correct design alternatives found by subjects between round 1 (i.e. without the guiding rule A1) and round 2 (i.e. using the guiding rule A1) of task T2. The figure shows that whereas the subjects were only able to find a little number of design alternatives on their own, the guiding rule helped them finding high numbers of correct design alternatives.

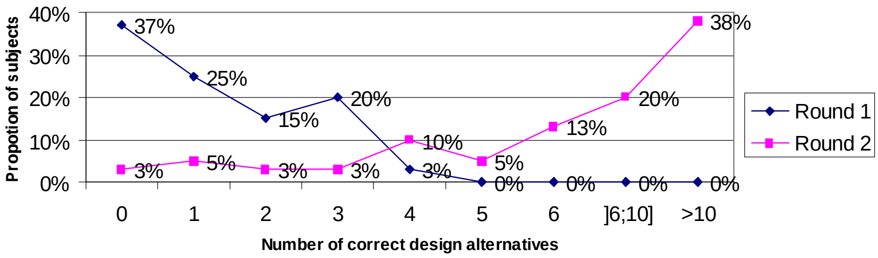

Figure 2 : subjects' performance in finding design alternatives

We observed that 76% of the design alternatives found at round 1 were relevant. This average increases up to 96% at round 2. However, in absolute values the number of correct design alternatives discovered by the subjects was multiplied by 8 owing to the guiding rule A1. Moreover, 95% of subjects did actually improve their performance. Therefore, the guiding rule A1 helped most of the subjects to identify significantly more design alternatives which were in proportion more correct. Hypothesis H2 is thus, very strongly validated.

5.3 Detailed evaluation of hypotheses H3 and H4

The results collected from the first round of task T3 show that the subjects were already able to identify a large number of scenario alternatives without guidance. However, these alternatives were in large proportion (42%) incorrect. In the second round of the task A2 (i.e. when applying the guiding rule A2), most of the incorrect scenario alternatives were either corrected or removed. Our initial observation is thus that (i) hypothesis H3 ("the guiding rule 'T2' improves the number of alternatives") is not strongly validated as such, but (ii) the guiding rule A2 helps improving the quality of discovered alternatives. It seems thus, that the guiding rule A2 improves the proportion of correct alternatives.

Figure 3 compares the distribution of subjects who have discovered given proportions of correct/incorrect alternatives during the two rounds of task T3.

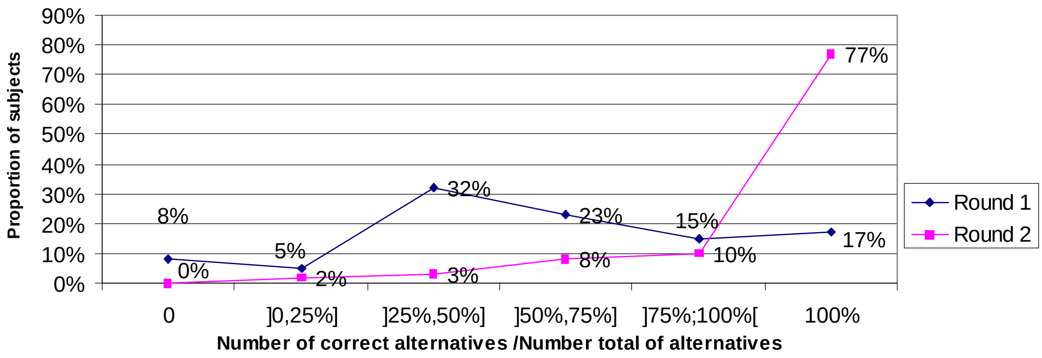

Figure 3: subjects capacity to identify correct variant between round 1 and 2

The figure shows that, using the guiding rule A2, the subjects were in average able to increase the proportion of correct alternatives by 35 %. Indeed the proportion of correct alternatives was of 58% at the first round where and of 93% at the second round. Let's notice that most of the 60% of incorrect alternatives identified at round 1 were due to an invalid level of abstraction. At round 2, the proportion of errors due to incorrect levels of abstraction is decreased to 7%. Owing to rule A2, 82% of subjects were able to eliminate all the variants initially identified at the wrong level of abstraction.

5.4 Detailed evaluation of hypothesis H5

The subjects' results at the first round of task T4 show that a small proportion (11%) of subjects were able to identify without guideline alternative behavioural descriptions from a given scenario.

Figure 4 compares the proportion of subjects having correctly separated different paths from the given scenario during the two rounds of task T4. The figure shows that 39% of subjects were able to separate concerns correctly using the guiding rule A2, whereas they were not without guidance. Hypothesis H5 is thus validated.

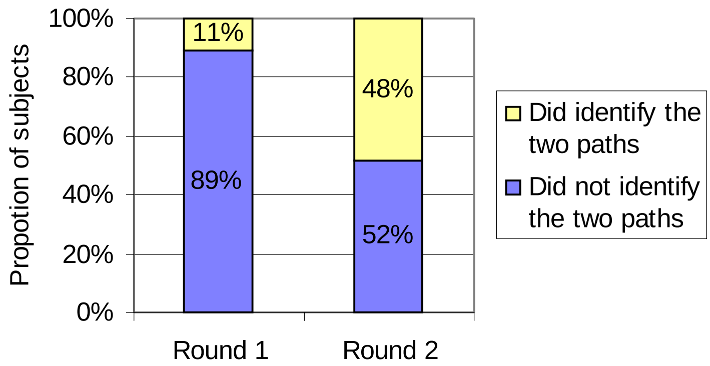

Figure 4: proportions of subjects able to identify multiple paths from a given scenario

5.5 Detailed evaluation of hypotheses H6 and H7

From the subjects' results at task T5, we initially observed that the guiding rules C1 and C2 which support the composition strategy help : (i) finding more system functions, and (ii) finding system functions which are more homogenous with respect to the level of abstraction.

Most often, the functions discovered without guidance were incorrect. Using the guiding rules C1 and C2 more than 80% of subjects have corrected the collection of system functions they proposed. Actually, 96% of subjects could provide at least one correct system function using the guiding rules. Without guidance, 92% of subjects did not provide any correct system function. Incorrect system functions were due to an inadequate level of abstraction, and to confusions between complementary and alternative system functions.

Using the guiding rules C1 and C2, the subjects have identified more correct complementary system functions than without guidance. Hypothesis H6 is thus confirmed. The main error that the guiding rule C1 and C2 helped avoiding is the confusion between alternative and complementary system functions. Indeed, at round 1, 35% of subjects did the confusion, at least once. At round2, 92 % of subjects did never make this error.

Figure 5 compares the distribution of subjects having identified predetermined numbers of system functions stated at an incorrect level of abstraction during the two rounds of task T5. The figure emphasises that at round 1 (without guidance), most of the subjects have stated complementary system functions at different levels of abstraction. On the contrary, using the guiding rules C1 and C2, only a small proportion of subjects have made this error. Hypothesis H7 is thus validated too.

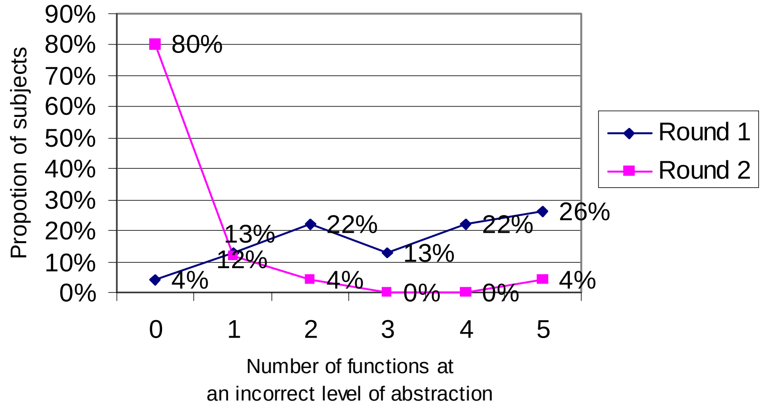

Figure 5 : distribution of subjects having identified given proportions of incorrect system functions

5.6 Detailed evaluation of hypothesis H8

Let's recall that to evaluate hypothesis H8, the subjects were requested during task T6 to identify several system functions hidden within a single scenario. Three system functions were expected.

A relatively small proportion of subjects (15%) was able to identify the expected system functions without guidance; we observed a significant improvement with the guiding rule R1. Indeed, as shows Figure 6, 39% of subjects who did not distinguish the 3 expected system functions without guidance were able to identify all of them using the guiding rule R1.

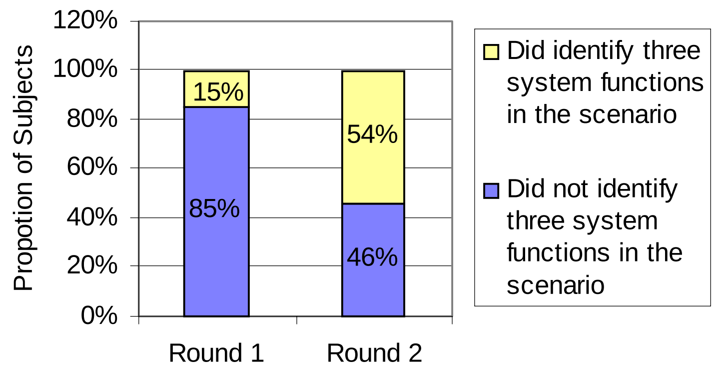

Figure 6 : subjects ability to identify multiple system functions in scenarios

5.7 Detailed evaluation of hypothesis H9

During task T7, the subjects were asked to check inconsistencies in the level of abstraction of the actions described in a given scenario. At the first round, 80% of subjects were not able to detect that the proposed scenario was containing actions at different levels of abstractions. On the contrary, using the three predefined levels of abstraction, all of them were able to tell that the scenario was inconsistent. Hypothesis H9 is thus validated. However, we observed different results for each of the three predefined levels :

At the behavioural level : within the set of actions classified as belonging to the behavioural level 47% were correctly classified; the other way round, 72% of the subjects have proposed an incorrect set of actions at the behavioural level.

At the functional level : 84% of the classified of actions at the functional level were correct; additionally 56% of subjects had proposed a correct set of functional level actions.

At the physical level : the set of actions classified at the physical level was correct at 96%. Moreover, 88% of subjects did propose an entirely correct set of physical actions.

Hypothesis H9 is thus validated, but with different results for each of the three predefined levels of abstraction.

6 Conclusions

Globally, we obtained some very positive results in this experiment. For example, hypotheses like H1 or H2 are strongly validated. In so far as H2 is concerned, the goal analysis strategy for guiding the discovery of design alternatives seems to be highly effective; a better vision of the future system is thus given.

However, not all the hypotheses could be so positively validated. For example the rules provided by CREWS-*L'Ecritoire* to guide the alternative and composition goal discovery strategies were expected to help identifying more goals (hypotheses H3 and H6). If the rules help indeed identifying more goals, the difference is not significant. Indeed, the subjects were already able to identify many goals without guidance (good user performance without guidelines). However, the proportion of correct goals discovered without guidance was also very low. Most of the subjects made a confusion between alternative and complementary goals, and had difficulty to identify the goals at the expected level of abstraction. On the contrary, the CREWS-L'Ecritoire guidelines helped the subjects

These results seem to tell that some of the CREWS-L'*Ecritoire* guidelines have more a corrective power than a generative power, and thus, should be used in verification/correction strategies rather than in pro-active elicitation strategies. However, let's recall that in the CREWS-L'*Ecritoire* tool environment the guidelines are partly automated whereas during this experiment they were applied manually. In addition to the experiment reported here, several tutorial with use of the CREWS-*L'Ecritoire* tool have been undertaken. Discussion with the subjects after this complementary experiment, together with our observations during tool use by the subjects, seem to tell that the automated guiding rules help users correcting their result while they are performing goal discovery tasks. Therefore, when automated the CREWS-*L'Ecritoire* guiding rule seem to have the desired corrective effect reported here (the requirements discovered are more correct), as well as the pro-active generative effect which is as well desired during requirements elicitation (more requirements are discovered).

Acknowledgement : we thank warmly the 41 students of the DESS at IAE who participated to this study.

References

[Anton 96] A. I. Anton, *Goal based requirements analysis*. Proceedings of the 2nd International Conference on Requirements Engineering ICRE'96, pp. 136-144, 1996.

[Ben Achour 98] C. Ben Achour, *Guiding Scenario Authoring*, Proceedings of the Eight European - Japanese Conference on Information Modelling and Knowledge Bases. May 25-29, Vamala, Finland.

[Ben Achour 99a] C. Ben Achour, *Extraction des Besoins par Analyse de Scénarios Textuels*, Phd thesis, Université Paris 6, France, 1999.

[Ben Achour 99b] C. Ben Achour, N.A.M. Maiden. *Empirical Study of Use Case Authoring : Experimental Material*. Internal Report, Centre de Recherche en Informatique, Université de Paris 1 - Sorbonne, http://www.univ-paris1.fr/CRINFO/users/benachour/ ESEM/index.html, 1999.

[Benjamen 99] *Towards a dynamic construction of process model*, Proceedings of MFPE'99 : 2nd International Workshop on the Many Facets of Process Engineering, Gammarth, Tunisia, 12-14, May 1999.

[Brunet 1993] J. Brunet, *Analyse Conceptuelle orientée-objet*, PhD Thesis, University of Paris 6, 1993.

[Cockburn 95] A. Cockburn, *Structuring Use Cases with Goals*. Technical report. Human and Technology, 7691 Dell Rd, Salt Lake City, UT 84121, HaT.TR.95.1, http://members.aol.com/acockburn/ papers/usecases.htm, 1995.

[CREWS 97] CREWS Team. *Scenario Use in European Software Organizations -Results from Site Visits and Questionnaires*.

[Haumer 98] Peter Haumer, Klaus Pohl, Klaus Weidenhaupt, *Requirements Elicitation and Validation with Real World Scenes*, to appear in IEEE Transactions on Software Engineering, Vol. 24, No. 12, Special Issue on Scenario Management, December 1998.

[Holbrook 90] C. H. Holbrook III, *A Scenario-Based Methodoly for Conducting Requirement Elicitation*, ACM SIGSOFT Software Engineering Notes, Vol. 15 N° 1, pp. 95-104, 1990.

[Jacobson 1995] I. Jacobson, *The use case construct in object-oriented software Engineering.* In ''Scenario-based design: envisioning work and technology in system development'', John M. Carroll (ed.), John Wiley and Sons, 309-336, 1995.

[Leite 97] J.C.S. do Prado Leite, G. Rossi, F. Balaguer, A. Maiorana, G. Kaplan, G. Hadad and A. Oliveros, *Enhancing a requirements baseline with scenarios*. In Third IEEE International Symposium On Requirements Engineering RE'97, Antapolis, Maryland, IEEE Computer Society Press, pp. 44-53, 1997.

[Nurcan 97] Selmin Nurcan, Georges Grosz, Carine Souveyet, *Describing Business Processes with a Guided Use Case Approach*, Proceedings of the 10th International Conference CAiSE'98, B. Lecture Notes in Computer Science 1413, Pernici, C. Thanos (Eds), Springer. Pisa, Italy, June 1998

[Plihon 98] V. Plihon, J. Ralyté, A. Benjamen, N.A.M. Maiden, A. Sutcliffe, E. Dubois, P. Heymans. *A reuse-oriented approach for the construction of scenario based methods*. Proceedings of the International Software Process Association's 5th International Conference on Software Process (ICSP'98), Chicago, Illinois, USA, 14-17 June 1998.

[Potts 94] C. Potts, K. Takahashi and A.I. Anton, *Inquiry-based Requirements Analysis*. in IEEE Software 11(2) (1994) 21-32.

[Prat 97] N. Prat, *Goal formalisation and classification for requirements engineering*. Proceedings of the Third International Workshop on Requirements Engineering: Foundations of Software Quality REFSQ'97, Barcelona, pp. 145-156, June 1997.

[Rolland 88] C. Rolland, O. Foucaut, G. Benci, Conception de Systèmes d'Information, La methode REMORA, Editions EYROLLES, 1998.

[Rolland 98] C. Rolland, C. Souveyet, C. Ben Achour. *Guiding Goal Modelling using Scenarios,* IEEE Transactions on Software Engineering, Special Issue on Scenario Management, Vol. 24, No. 12, 1055- 1071, Dec. 1998.

[Rolland 99] C. Rolland, A. Benjamen , *A Multi-Model View of Process Models*, Requirement Engineering Journal, 1999.

[Rumbaugh 91] J. Rumbaugh, G. Booch, *'Unified Method, Notation Summary'* Version 0.8, Rational Software Corporation, 1996.

[SiSaid 98] S. Si Saïd, C. Rolland. *Formalising Guidance for the CREWS Goal-Scenario Approach to Requirements Engineering*. Proceedings of the Eight European - Japanese Conference on Information Modelling and Knowledge Bases. May 25-29, Vamala, Finland, 1998.

[Tawbi 99] M. Tawbi, C. Souveyet, *Guiding Requirement Engineering with a Process Map*, Proceedings of MFPE'99 : 2nd International Workshop on the Many Facets of Process Engineering, Gammarth, Tunisia, 12-14, May 1999.

[Van Lamsweerde 98] Van Lamsweerde A., Willemet L., *'Inferring Declarative Requirements Specification from Operational Scenarios'*, IEEE Transactions on Software Engineering, Special Issue on Scenario Management, Vol. 24, No. 12, 1089- 1114, Dec. 1998.

[WebAppendix] http://www.univ-paris1.fr/CRINFO