A Goal-Oriented Strategy for Supporting Commercial Off-The-Shelf Components Selection

Claudia Ayala, Xavier Franch

Universitat Politècnica de Catalunya
UPC-Campus Nord (Omega), 08034 Barcelona, Spain
{cayala,franch}@lsi.upc.edu

Abstract. The use of Commercial Off-The-Shelf (COTS) components is becoming a strategic need because they offer the possibility to build systems at reduced costs and within shorter development time. Having efficient and reliable COTS components selection methods is a key issue not only for exploiting the potential benefits of this technology, but also for facing the problems and risks involved. Searching COTS components requires to overcome several obstacles: the growing size and evolvability of the COTS marketplace, the dependencies from the components to be selected with others, and the type of descriptions currently available for those components. In this report, we present a goal-oriented strategy for an effective localization, analysis and structuring of COTS components information. Our proposal is the GBTCM+ method, which provides methodological support to the construction of taxonomies. We present the seven activities that conform this method, which are illustrated with the case of real-time synchronous communication tools.

1. Introduction

Nowadays, the construction of systems based on pre-packaged solutions, usually known as Off-The-Shelf (OTS) components, is becoming an economic and strategic need in a wide variety of different application areas. OTS-based systems, either developed from the scratch or by migrating from existing ones, are built by assembling and integrating OTS components that may be of different nature. The potential benefits of OTS technologies are mainly the reduced cost and shorter development time of OTS-based systems, while maintaining their quality [1]. Nevertheless, many challenges, ranging from technical to legal issues, must be faced for adapting the traditional software engineering activities with the aim of exploiting these benefits.

One of the most critical activities in OTS-based systems development is the selection of the components that must be integrated therein. Selection is basically composed of two main activities, namely search of candidates and their evaluation with respect to system requirements. Most of the different existing methods for COTS selection [2], [3] focus on evaluation instead of search. This lack of specific proposals is a serious drawback that impacts in selection reliability: no matter how good is the evaluation process, selection may be wrong if the candidates chosen to be evaluated are not the right ones.

Searching candidate OTS components is not an easy task, especially in the case of Commercial-Off-The-Shelf (COTS) components, i.e. components that are acquired for a fee. On the one hand, COTS components are a class of reusable components, and it is well-known that one of the essential problems in reusing software components is locating and retrieving them from a large collection [4]. On the other hand, COTS search must cope with some challenging characteristics:

1. *Growing size of the COTS marketplace*: New and improved products and technologies are continuously offered. Basically this means that existing market segments offer more and more products, and also that new segment markets are continuously emerging. Mobile technologies are a good example of both situations.

2. *Rapid changes in the COTS marketplace*: New versions of existing products are released every few months. And market segments frontiers move slightly over the years, making products to offer services that initially were seen as belonging to different segments. For instance, current mail server systems usually provide instant messaging facilities, even video-conferencing services.

3. *Dependencies among COTS components*: COTS components are not designed to work isolated, but in collaboration with others. Therefore many dependencies among them exist, either for enabling, enhancing or complementing their functionality [5]. For instance, document management systems need document imaging tools for scanning and storing paper documents.

4. *Type of descriptions available for COTS components*: COTS components suppliers do not provide the kind of structured information that would allow performing automated or at least assisted search. Moreover, it is not realistic to think that the situation will change in the future. This is especially true for coarse-grained COTS components such as ERP, CRM or CMS systems. The situation is aggravated by the fact that supplier information of course tends to highlight strengths and hide weaknesses of the licensed components.

Consequently, when carrying out a particular searching process, some practical questions may arise: Which are the market segments of interest for this particular context? Which are the relationships among the identified market segments and which are their implied needs? How can structured and trustable information be obtained for the COTS components available in the marketplace?

In this report, we claim that an effective COTS search strategy shall rely on a thorough description of the COTS marketplace whose nature adapts to the above mentioned characteristics (diversity, size, evolvability, interoperability, lack of structure and subjectivity) and therefore provides real answers to the questions above. As a result, we present a method aimed at building a reuse infrastructure that may be used in COTS search processes by arranging marketplace segments as a taxonomy. The nodes of this taxonomy are characterized by means of goals and their relationships declared as dependencies. The method includes a domain analysis phase which faces the prob-

lem of unstructured and not validated information coming from lots of information sources.

The rest of the report is organized as follows. In section 2 we present our research method and previous work, and related work in section 3. The core of the proposal, the GBTCM+ method, is presented from sections 4 to 11, illustrated with a case study on the category of real-time synchronous communication tools. Finally, in section 12 we give the conclusions and some future work.

2. Research Method and Previous Work

Our proposal relies on several industrial experiences which have been undertaken under action-research premises [6], as well as literature survey and grounded theory [7]. Furthermore, we have formulated in early work some preliminary proposals. This section provides details about both points.

The first industrial experience taken was in the context of an academic record management information system development which was planned to include some strategic business functionalities. We undertook a thorough analysis of the domain and experimented the problems mentioned in the introduction. As a result, we presented a first paper [8] proposing the use of taxonomies to structure the COTS business application marketplace. After this, we had other collaborations in the field of requirements management tools, telephony systems and others. We complemented these real cases with some academic ones. Therefore, we incorporated the notion of goal to formalize the meaning of the nodes in the taxonomy making it domain-independent. There exists some evidence that goals are quite stable with respect to changes [9]. In addition, goal refinement provides a natural mechanism for structuring and exploring many alternatives [10]. Finally, we presented a goal-oriented method called GBTCM (*Goal-Based Taxonomy Construction Method)* which added the process dimension to our previous work [11]. It was inspired on GBRAM (Goal-Based Requirements Analysis Method) [10], a widespread method in the requirements engineering discipline.

Although GBTCM was an improvement of our previous work, we have recently encountered some method design flaws, some due to the use of GBRAM in a different context, others due to our method as such. The flaws are:

- GBRAM is a requirements acquisition method; therefore the sources of information are mainly human beings, which is not the case in the COTS context.
- Furthermore, GBRAM lacks of proper mechanisms to deal with the huge amount of unstructured information of the COTS marketplace.
- GBTCM does not give the required importance to the analysis of the domain, which is more difficult than in a non-COTS context because expertise is needed not only on the domain itself but also on how this domain is represented in the marketplace.

- GBRAM is a one-shot method, with no orientation to knowledge reuse.
- GBTCM focuses on the market segments but did not consider the COTS components themselves.
- GBTCM definition was not oriented to having tool-support.

The method presented here, GBTCM+, aims at overcoming these flaws.

3. Related Work

Due to the highly applicable nature of the subject of our research, we find related work not only concerning scientific proposals but also in the way that the COTS marketplace is really organized nowadays. Profit and non-profit organizations define categories of services, products, and knowledge, usually structured in a hierarchical form. This type of organizations can be classified as follows:
- IT consultant companies such as Gartner [12] or Forrester [13] use these categories to structure their reports and services on IT technology.
- Commercial web-based companies such as ComponentSource [14] and Genium [15] group the products commercially available for facilitating the web browsing.
- Professional societies such as INCOSE [16] use hierarchies to organize systems engineering knowledge (often not related specifically to COTS issues).
- Portals with different registration procedures offer white reports, user's opinions [17], [18] or technical products from research projects [19].

In the academic world, organizations, teams and individuals have presented their own proposals that range from specific of one domain [20] to a wide range [4] [21] or even a field [22], being the extreme case proposals such as SWEBOK that acts as a body of knowledge of a particular discipline [23]. However it is well-know that the effort devoted to these activities is more valuable if the attributes can be reused; in this sense a wide range of works about COTS characterization exists [24-27].

However, such proposals do not provide proper mechanisms for facing the characteristics of the COTS marketplace mentioned in the introduction. In [28] a survey of different approaches for classifying COTS is given and also emphasizes some of the problems. Furthermore, sometimes, the meaning of a particular domain is not clear without further examining the items, especially if the domain is absolutely unknown to the user. Consequently the understanding, use, evolution, extension, and customization of the categorization proposal may be difficult.

We have experienced in details these drawbacks in the case of Gartner and INCOSE, whose classifications were used as starting points in [8] and [11], respectively. To sum up, we consider that there is a gap between the mentioned proposals and their applicability in the COTS search context that can be considered an important open issue [29].

4. An Overview of the GBTCM+ Method

GBTCM+ has been structured into seven activities:

1. Exploration of information sources.
2. COTS marketplace domain analysis.
3. Identification, refinement, and statement of goals.
4. Determination of dependencies.
5. Goal taxonomy structuring.
6. Taxonomy validation.
7. Knowledge base management.

Although presented as sequential for clarity, the activities may in fact be intertwined and iterated as required to obtain the target infrastructure incrementally. Furthermore, GBTCM+ does not depend on the extent and characteristics of the taxonomy built (e.g., a small part of the COTS marketplace such as photo processing software, or a huge portion like business applications).

The ultimate goal of GBTCM+ is to populate a knowledge base with data according to the UML [30] conceptual model firstly sketched in Fig. 1 and refined in the subsequent sections (since they are not essential to the method, OCL constrains are not included for the sake of brevity). At the heart of this model lies the taxonomy composed of two types of nodes, market segments and categories, which are characterized by their goals. Market segments are the leaves of the taxonomy, whilst categories serve to group related market segments and/or subcategories (e.g., the category of communication infrastructure systems or financial packages).
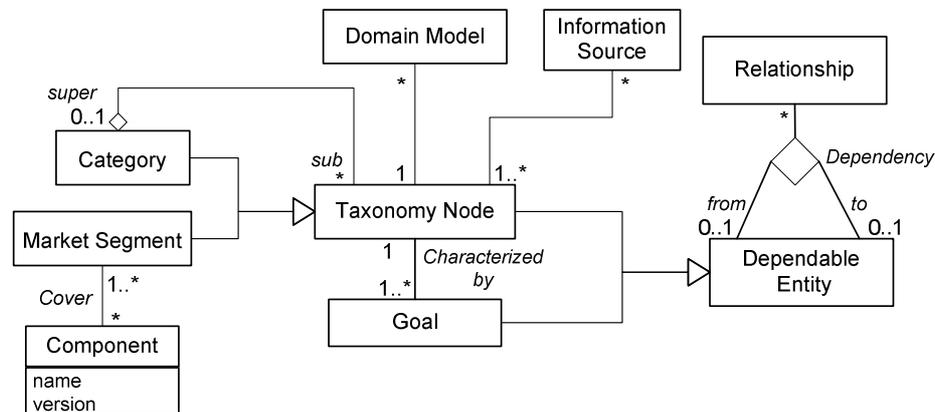

Fig. 1. Conceptual model for goal-oriented COTS taxonomies: overview.

From a semantic point of view, market segments stand for the basic types of COTS components available in the marketplace (e.g., the domain of anti-virus tools or spreadsheet applications), i.e. atomic entities covering a significant group of function-

ality such as their decomposition would yield to too fine-grained domains. As a consequence, COTS components are associated with market segments and not with categories (although an indirect relationship exists, because market segments belong to categories). Components may cover more than one market segment. For simplification purposes, we are not distinguishing at the moment versions of components; two different versions are treated as two different products.

Dependencies among nodes provide a comprehensive view of the marketplace. In the case of dependencies among market segments, they stand for interoperability needs (e.g. mail server systems depend on anti-virus tools to support integrity). Concerning categories, more abstract relationships are modeled. In addition to taxonomy nodes, dependencies may involve goals, when the relationship can be established more accurately. The *Dependable Entity* superclass allows modeling this situation comfortably. Note that dependencies are represented by a ternary association, because they involve two elements (depender and dependee) and the relationship itself.

Finally, nodes have auxiliary artifacts bound, which are built during the domain analysis activity. Their construction is a result of the analysis of some information sources which are gathered, analyzed, and prioritized according to several characteristics. In following sections, we provide details of the GBTCM+ activities. To illustrate our approach, we use the Real-Time Synchronous Communication (RTSC) category. It means the various tools and technologies used to enable communication and collaboration among people in a "same time-different place" mode.

5. Exploration of Information Sources

This activity must be able to locate as much relevant information as possible, dealing with the diversity of its type, supporting media, cost, etc. We distinguish three related subactivities:

- *Gathering of sources*. Identification of the potential information sources for the domain of interest using information acquisition techniques (e.g., literature review, web screening, etc.). We have identified the following types of sources: existing hierarchies, taxonomies and ontologies; standards in the domain; vendor information; independent reports (scientific, divulgation and technical); oral information; test of tools and systems; experiences on the field; others.

- *Analysis of sources*. Some techniques, are applied to determine the relevant criteria to be used to rank the identified sources: reliability of the information; availability of the source; acquisition cost; timeliness; scope covered; and time needed to process the enclosed information. These criteria move along three dimensions: information source type, organization or people that created the information, and particular item of information.

- *Prioritisation of sources*. The analysed sources are ranked according to several characteristics of the taxonomy construction project, mainly: expected frequency of taxonomy use in future selection processes; resources allocated to the project, especially deadline, money and person/months; current and future knowledge of the domain and technical skills of the conformed team; and expected criticality of the domain (and therefore required accuracy and completeness of the solution).

Table 1. Information sources for the RTCS case.

Information Source			Information Type	Language	Examples	Utility
Existing Taxonomies and Ontologies			Classifications; Categories; Glossaries	Natural Language (NL); Tree-like diagrams	Gartner, IDC, eCOTS, ComponentSource	They help not only for understanding domains and refining goals, but also for getting insights for organising goals.
Related Standards			Descriptions; Glossaries	NL	IETF-SIP ITU H.323 ISO 9126-1	They are considered the most confident of the sources, so the high-level goals are based on them.
Vendors Information			Brochures; Evaluation forms; Benchmarks	NL; Values for attributes	Microsoft	They are helpful to know functionalities, trends and interactions among components
Domain Descriptions	Scientific	Academic Events, Jounals Textbooks	Precise and rigorous descriptions	NL; Models; Formulas; Schemas	ICCBSS, ICSE, TSE	The information enclosed in this kind of sources, generally helps to understand domains and refining goals into sub goals.
	Divulgation	Magazines, Forums and Web-sites	Descriptions and tips for the general public	NL; Schemas; Tables	PCWorld, IEEE Software, COCOTS web-site	
	Technical	White Papers, Surveys and Comparatives	Papers, Comparative tables	NL, Tables; Figures	Gartner, INCOSE, eCOTS	
Oral Information	Interviews		Knowledge; Tips; Practical Info.	NL	ICCBSS panels, SEI courses, Business luncheons	
	Talks, seminars and courses					
Test of Tools and Systems			Test results; User's manuals	Visual data; NL	ICQ, MSN Messenger, CommuniGate	
Experiences on the field			Knowledge; Technical reports	Knowledge; NL	Past projects made	
Others			Any	Any		

At the end of this phase we have a knowledge acquisition program which will allow extracting knowledge from the domain by reconciling the characteristics of the available sources with those of the taxonomy construction process. Table 1 is an excerpt of the information sources considered for the RTSC case and shows details of their utility and the kind of information therein. Fig. 2 shows an excerpt of some mechanisms and artifacts we used for analyzing sources. Fig. 3 extends the departing conceptual model presented in Fig. 1. The different types of information sources are defined as specializations. Also, the author of the information is modeled explicitly. We show with a particular kind of characteristic, namely reliability of the source, the way in which we assign values in the above mentioned three different dimensions. The rest of attributes are not shown for simplicity of the drawing.

Id	Name	Type	Author	Cost	...
1	Session Initiation Protocol	Standard	Engineering Task Force	Free	
2	H.323	Standard	International Telecommunication Union	±80€	
3	IMTC	Independent Report	International Teleconferencing Consortium	Free	
4	RTC-Gartner	Hierarchy	Gartner	Free	
:					

Example of Questionnaire

Determining Author reliability

Is it a reputable author or organization?
☐ Excellent ☐ Good
☐ Satisfactory ☐ Weak
Did you see this source listed in other sources?
☐ Yes ☐ No
...

Fig. 2. Examples of the artifacts used for the information sources analysis.

Information Source

Id
name
reliability

Type — * ... 1

Information Source Type

reliability
name:TypeOfSource

Author

Id
name
reliability

* ... 1..*

<<enumeration>>
TypeOfSource

Hierarchy
Standard
VendorInfo
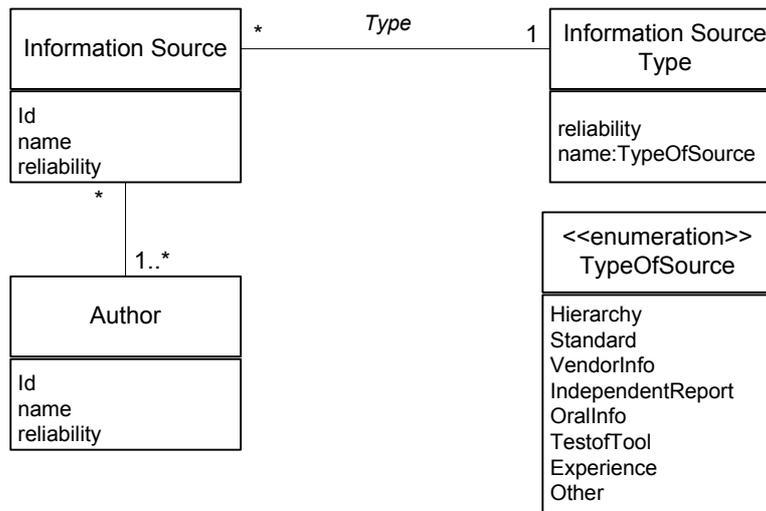IndependentReport
OralInfo
TestofTool
Experience
Other

Fig. 3. Conceptual model for goal-oriented COTS taxonomies: information sources

6. Domain Analysis

Domain analysis has been identified as a major factor in the success of software reusability [31]. Its goal is to acquire and consolidate information about a given domain so that reusable software infrastructure can be designed reliably. More precisely, it helps to identify the basic elements of the domain, organize an understanding of relation-

ships among these elements, and represent this understanding in a useful way. Domain analysis is especially crucial in our approach because of two main reasons:

- Using domain analysis principles we avoid syntactic and semantic discrepancies common in the COTS marketplace. For instance: one of the most endangering points in COTS domains is the lack of standard terminology, the same concepts are named different by different vendors or even worse, the same name may denote different concepts in different COTS components.

- The core elements of a domain and the relationships among them usually remain more stable, while the technologies and implementation environments are in continuous evolution. Hence, domain analysis models remain valid for long periods.

Several proposals of domain analysis available in the literature may differ in the type of artefacts proposed to record the knowledge. In this work, we propose the following four artefacts:

- *Use Case Specification*. A UML use case diagram [30], arranged in packages if necessary, to provide an overall view of the services that the COTS components in the market segment or category offer. Individual specifications of use cases are recommended to be very abridged, for different reasons (evolvability of marketplace, avoid committing to behaviour of particular COTS components, etc.).

- *Class Diagram*. To keep track of the fundamental concepts in the domain, their attributes, associations and taxonomic relationships. Also in UML [30].

- *Quality Model*. A hierarchical representation of the quality factors applicable to the domain, such as those referring to efficiency and integrity, together with their metrics. For standardization issues, we propose the use of the ISO/IEC 9126 quality standard [32].

- *Glossary of terms*. It includes at least the names of elements in the class diagram and the quality model. The glossary must not include overloaded terms, although many definitions may exist for a single term (which should be semantically equivalent). We propose to use the Language Extended Lexicon (LEL) [33] for capturing the meaning and fundamental relationships of the particular symbols (words or phrases) of the domain.

Fig. 4 shows the UML representation of these auxiliary models. The classes *Use Case Specification* and *Class Diagram* are linked to the UML metamodel and therefore not refined in the drawing. The quality model has been defined complied with ISO/IEC 9126. The glossary includes terms together with an association to identify synonymous that has demonstrated to be useful. For traceability purposes, an association among glossary terms (and due to the integrity constraints mentioned above, class diagram and quality model elements) and information sources has been introduced.

It is important to remark that the models present some relationships when considering the nodes in the taxonomy, enforced by integrity constraints. In particular:

- The class diagram of a category is defined as the composition of the class diagrams of its heirs.

- The quality model of a category is inherited by all of its heirs (this means that common quality factors and metrics may be defined in the upper parts of the taxonomy facilitating thus their reuse).

- The glossary of terms for a category is the union of the glossaries of its heirs.

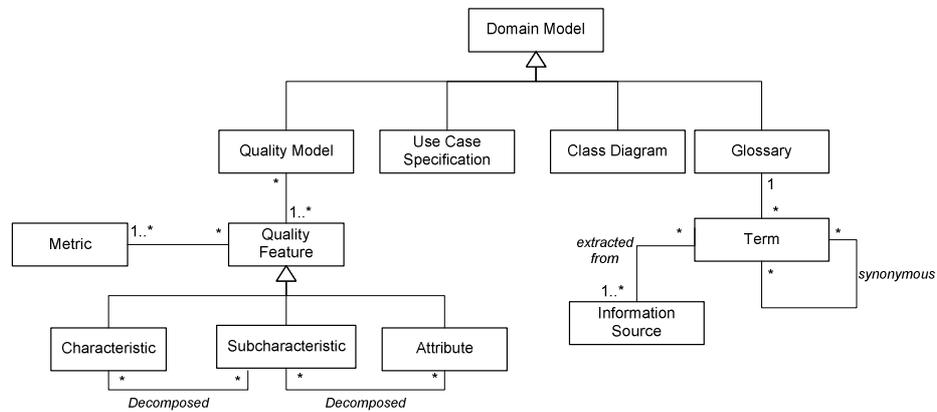- Use case specifications apply only to market segments.

Fig. 4. Conceptual model for goal-oriented COTS taxonomies: domain analysis

Contradictions when composing or joining models may arise and of course they should be detected and reconciled. In Fig. 5 we show excerpts of the four types of artifacts for the RTSC case.
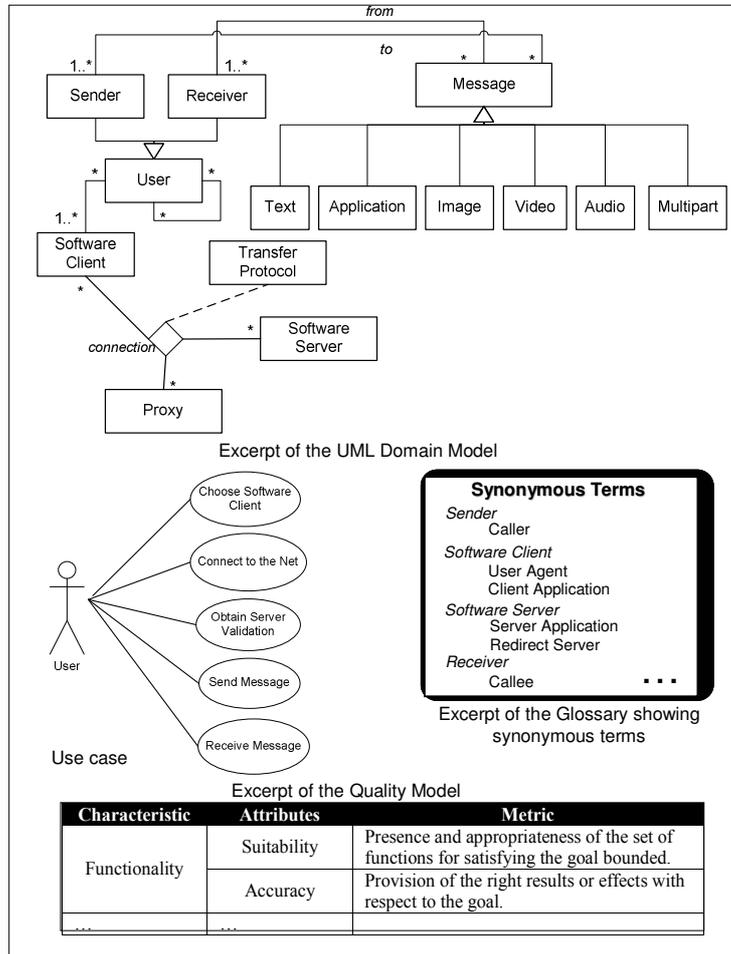
from

to

1..* Sender | Receiver 1..* | * * Message

User * | *

1..*

Software Client

Transfer Protocol

*

connection

Software Server *

Proxy *

Text | Application | Image | Video | Audio | Multipart

Excerpt of the UML Domain Model

Choose Software Client

Connect to the Net

Obtain Server Validation

Send Message

Receive Message

User

Use case

Synonymous Terms

Sender
    Caller
Software Client
    User Agent
    Client Application
Software Server
    Server Application
    Redirect Server
Receiver
    Callee   . . .

Excerpt of the Glossary showing synonymous terms

Excerpt of the Quality Model

Characteristic	Attributes	Metric
Functionality	Suitability	Presence and appropriateness of the set of functions for satisfying the goal bounded.
	Accuracy	Provision of the right results or effects with respect to the goal.
…	…	

Fig. 5. Excerpt of models built for the RTCS case.

7. Identification, Refinement and Statement of Goals

A goal is an objective that should be achieved that may be formulated at different levels of abstraction, ranging from high-level strategic concerns to low-level technical concerns [9]. Goal formulation refers to intended properties to be ensured. The activities performed in this stage are iterative and they have the next objectives:

- *Identification* aims at extracting goals from available information sources applying different goal-acquisition mechanisms and techniques [34] as heuristics, the Inquiry Cycle (IC) approach [35] (consisting of a series of questions and answers designed to pinpoint where and when the information needs arise) and scenarios for

idenfity new goals. Table 2 is an example of the use of scenarios for obtaining goals in the RTSC case.

- *Refinement* entails the process of refinement of the goal set considering possible obstacles and constructing scenarios to uncover hidden goals as well as mechanisms to discover synonymous or duplicated goals. Table 3 shows an excerpt of techniques applied to uncover goals.

- *Statement* consists on expressing the goals in a systematic way. We propose the use a pre/post style for specifying these goals, i.e. stating which conditions are met when others hold, as showed in Table 4.

Table 2. An scenario excerpt of the RTSC case study

Action Initiator	Goal	Consumed Resources	Produced Resources	Action Addressed
Human User (Sender)	Message Sent	Message	Message, Receiver address	Requesting to Software Client
Software Client	Sent Request to the Server	Message, Receiver address	Sender address	Requesting to Software Server
Software Server	Messages Routed	Message, Sender and Receiver address	Routed Receiver address	Sending to Software Client (Receiver)
Software Client	Message Delivered	Message, Sender address	Message	Deliver to a Human User (Receiver)
Human User (Receiver)	Message Received	Message, Sender address	Message	Answering

Table 3. An excerpt of techniques applied to discover goals

Goals	Goal Obstacles	Scenarios
G1: RTSC Established	1.- No RTSC Established 2.-There is no infrastructure available 3.-Users are no connected at the same time	1.-Establishing RTSC
...		

Table 4. An example of goal statement.

Goal:	Multiuser Textual Communication Established
Type	Achievement
Description	Provide RTSC in a Text Multi-user Environment
Agent	Software Client
Stakeholder(s)	Software Client, Software Server, Sender, Receiver
Precondition(s)	1) Users Communicated in Real Time; 2)Session Established; 3) Number of users >=2
Postcondition(s)	Multiuser Textual Communication Established
Subgoal(s)	1) Software Client Provided; 2) Software Server Provided

8. Establishment of Dependencies

The need for identifying dependencies among COTS market segments has already been mentioned. More precisely, we have identified that a COTS component may need another for:

- Enabling its functionality. For instance, in order to follow document life-cycles, document management tools need workflow technology to define them.

- Complementing its functionality with an additional feature, not originally intended to be part of its suitability. For instance, a web page edition tool can complement a web browser to facilitate the edition and modification of web pages.

- Enhancing its quality attributes. For instance, resource utilization can be improved significantly using compression tools.

Relationships are identified analyzing the goal information obtained in previous activity as well as domain models and information of sources that yields to this kind of knowledge (e.g. test of tools, experience reports, etc.). The way in that we perform this analysis leads us to identify relationships gradually. These relationships are declared as dependencies using goal-oriented models, specifically *i** models [36]. Using this notation as proposed in [5], we represent market segments and categories as *i** actors, and establish dependencies that may be of four different types: goal dependencies, when an actor depend on another to attain a goal; task dependencies, when an actor requires another to perform an activity in a given way; resource dependencies, when an actor depend on another for the availability of some data; and soft goal dependency, when an actor depends on another to achieve a certain level of quality of service. Fig. 6 shows some dependencies for the RTSC case.
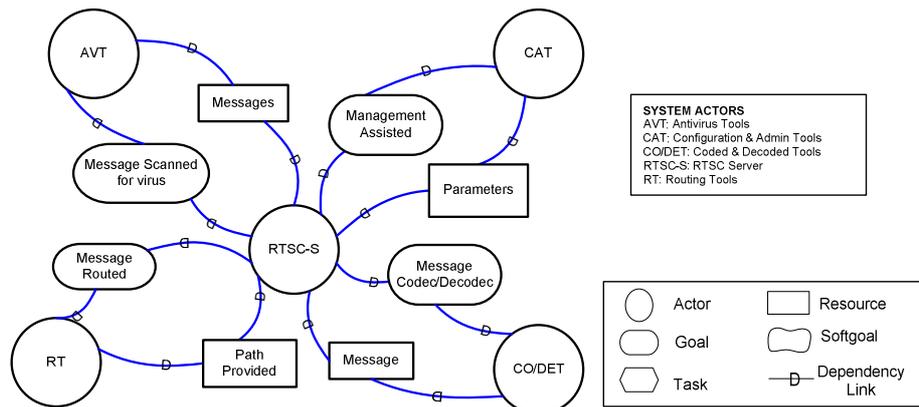

Fig. 6. *i** SD model representing the main high level dependencies identified in the RTSC case.

9. Goal Taxonomy Structuring

Taxonomic classification in the form of decision tree is the intellectual tool that helps us to organize goals in order to establish a structure and the locate/retrieve mechanisms. Our taxonomies are goal-driven, which means we provide semantics to the nodes expressing goals, giving a rationale for the decisions taken. The organization of goals comes from the analysis of pre and postcondicions stated for each goal.

Goals are operationalized in terms of variables which, in the case of categories, represent classifiers (e.g., number of users of the system, data processing profile, …). These classifiers may take values (e.g., for data processing profile, values are Acquisition, Storage, Preparation, Analysis), and for each possible value, a subcategory or market segment applies. Thus, *Goals* are defined over a set $X =\{x_k\}_n$ of independent variables that characterize the taxonomy. *Goal satisfaction* is defined by means of assignment to the variables, therefore for each assignment $ass = (x_1 \leftarrow v_1, \ldots, x_n \leftarrow v_n)$, the expression $sat_{ass}(G)$ yields true if the goal G evaluates to true for this assignment, otherwise false.

Fig. 7 shows the extension of the conceptual model from Fig. 1 with variables and assignments for goals. An association for relating categories with their heirs as a result of goal satisfaction is included.
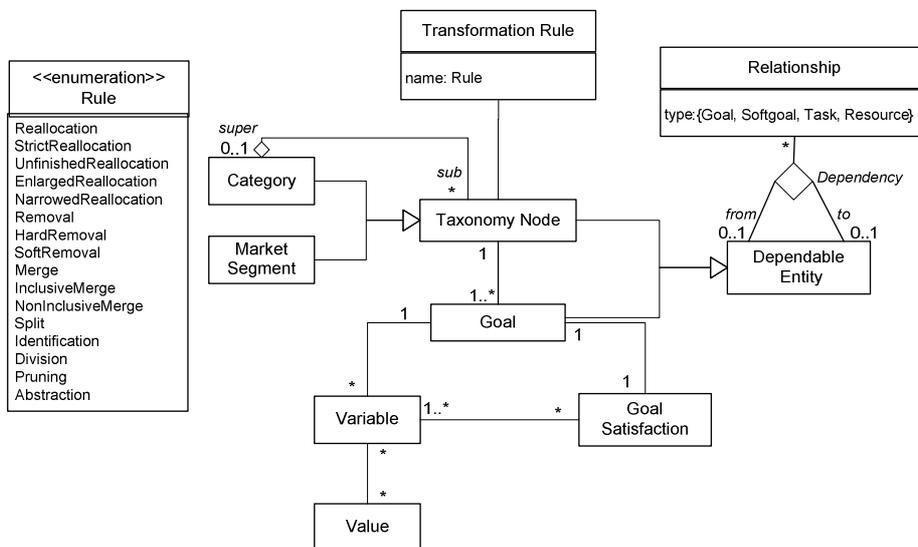

Fig. 7. Conceptual model for goal-oriented COTS taxonomies: goals and dependencies

Table 5 shows an excerpt of the departing goal hierarchy for the RTSC case as well as its variables assignment, considering that all the assignments are inherited downwards the hierarchy.

Table 5. Excerpt of the departing goal-oriented taxonomy for RTSC case

Goal /SubGoal	Variable	Satisfaction Values
Users Communicated in Real Time	TypeOfConnection	TypeOfConnection←RealTime
Intra-organizational Communication Established	Infrastructure	TypeOfConnection←RealTime Infrastructure ←Intranet
Global Communication Established	Infrastructure	TypeOfConnection←RealTime Infrastructure←Internet/WAN
…		

10. Taxonomy Validation

In order to be useful for driving COTS search processes, we require three conditions to the taxonomy: to be consistent, to be complete and to be not ambiguous. Also, we aim at leveraging its nodes to get similar levels of abstraction in the nodes of the same level. We have defined the process of taxonomy validation as the repeated application of some stated transformation rules (defined in terms of the goals pre and postconditions) over the nodes to manipulate the hierarchy until reaching a stop condition. These transformations rules shall satisfy a precondition to be applied until completeness and correctness conditions with respect to the involved goals is assured, in such a way that a goal-oriented taxonomy is said to be correct and complete if it satisfies these invariant conditions. Specifically, this process has 4 steps each of them is aimed to ensure each condition:

- *Step 1* ensures the hierarchy of nodes is well-formed, which means that satisfaction of the goal of a node implies satisfaction of its parent goal.

- *Step 2* that the variable assignation provides a unique way for classifying COTS components, which means that there is no variable assignment which makes two siblings satisfy their goals simultaneously.

- *Step3* that any COTS related with the domain can always be classified using the taxonomy, which means that the taxonomy covers all the possible assignment of variables.

- *Step 4* was added for applying transformation rules in order to tailor the taxonomy to the particular (and subjective) taste of the designer with respect to the level of detail and organizational concerns.

Table 6. Partial view of the RTSC Taxonomy.

Categories				Domains
Level 1	Level 2	Level 3	Level 4	
a. Intranet Communication	d. Multi-user Communication	…	…	Multi-User Shared Applications
		j. Chat	o. Chat Client	Multi-user We-Based Chat Client
				Multi-user No Web-Based Chat Client Application
				Multi-user Chat Server Application
		k. Video&Audio		Multi-user Video Application
				Multi-user Audio Application
	e. One-to-One Communication	…	…	One-to-One Shared Application
		i. Chat	p. Chat Client	One-to-One We-Based Chat Client
				One-to-One No Web-Based Chat Client Application
				One-to-One Chat Server Application
		m.Video&Audio		One-to-One Video Application
				One-to-One Audio Application
b. Internet Communication	f. Multi-user Communication	…	…	…
	g. One-to-One Communication	…	…	…
c. WAN Communication	h. Multi-user Communication	…	…	…
	i. One-to-One Communication	…	…	…

Table 7. An excerpt of the variables and satisfaction values browsing the taxonomy

Level	Category	Classifier Value	Variable/Satisfaction Values
1	Root	Infrastructure	Infrastructure ←Intranet Infrastructure ←Internet Infrastructure ←WAN
2	a	Number of Users	NumberUser←Multi-user NumberUser←One-to-One
3	d	Type Of Data	TypeOfData←Application TypeOfData←Textual TypeOfData←Audio&Video
4	j	Architecture Element	ArchitectureElement←Client ArchitectureElement←Server
5	o	Technology	Technology←WebBased Technology←NoWebBased
…			

This process and applicable transformation rules are detailed in [37]. Through this process we manipulated the nodes in a formal way to obtain the resulting taxonomy. For instance, in Table 5 we can see that the goal *Users Communicated in Real Time* was implying as subgoals 2 ways in which we can establish a communication, however in the resulting taxonomy showed in Table 6, 3 nodes are stated because the goal *Global Communication Established* was mixing 2 different concepts and functionalities that seems to be fashionable requirements demanded in the marketplace: *Internet Communication Established* and *WAN Communication Established*; thus we applied a rule for showing explicitly this value (as showed in table 7) preserving correctness and completeness properties. As a result of the process we have a high quality taxon-

omy in which the rationale for the classification is very clear and correctness and completeness are ensured by construction.

11. Knowledge Base Management

Many studies refer that it is necessary to build a body of knowledge towards a knowledge-based framework for COTS components identification [29], [1]. GBTCM+ as defined in this report provides an efficient mechanism to maintain a repository of the obtained knowledge due to the UML class diagram that defines the form that this repository exhibits. This knowledge base is the infrastructure support not only for an easy evolution and maintaining of taxonomies, but also for their suitability to specific organizational concerns.

In this report we have presented a method, GBTCM+, for facing COTS components search that is based on the notion of goal for building abstract, well-founded and stable taxonomies, which may evolve as the marketplace does. GBTCM+ is defined in a rigorous way, with a conceptual model that introduces all the concepts needed. GBTCM+ has been presented as a way to overcome the characteristics of the COTS marketplace mentioned in section 1:

- *Growing size of the COTS marketplace*. Proliferation of information is taken into account by prioritizing information sources in the bases of given criteria (time, money, reliability, …). Appearance of a new market segment is easier to handle than in other approaches, since it requires to locate its place in the taxonomy using the defined classifiers, and once there even some useful artifacts are inherited (e.g., quality models and glossaries).

- *Rapid changes in the COTS marketplace*. We use a goal-oriented approach, in the belief that goals are stable concepts [9]. Also, the fact that taxonomy nodes do not stand for types of COTS components available but for related groups of functionalities, makes the taxonomy more robust with respect to the segment barriers movement effect mentioned in the introduction.

- *Dependencies among COTS components*. We represent explicitly these dependencies with a model built with *i**, a widespread and accepted notation in some other disciplines (e.g., requirements engineering, agent-oriented development).

- *Type of descriptions available for COTS components*. We have identified two activities for collecting information sources and carrying out domain analysis to cope with the diversity, lack of structure and lack of reliability of information about COTS components. Also, our resulting taxonomy provides an external view that is: well-founded (with a clear rationale of the proposed structure), validated (sound, complete, pair-wise disjoint and balanced) and ready to browse (using the defined classifiers).

It is worth to think about applicability of the method. Basically GBTCM+ requires the following characteristics to be applicable:

- The taxonomy addresses a category of market segments that is of general interest. This means that a great deal of organizations need to select COTS components from these segments. Some examples are: communication infrastructure (including the RTSC case used in this report), ERP systems, security-related systems, etc. In these contexts, the number of selection processes that take place will be high and then reusability of the models likely to occur.

- The addressed market segments offer COTS components of coarse-grained granularity. This makes domain understanding more difficult, time-consuming and cumbersome and therefore domain analysis and taxonomy construction are helpful. Market segments such as CRM and ECM systems are typical examples, whilst time or currency converters are not. In these cases, having knowledge available and classifiers to know when a market segment is of interest is a great help. This last point is especially appealing in those selection contexts in which the organization that is interested in the selection does not have clear requirements about the kind of system needed.

- The COTS components search activity is monitored by an organization that accumulates experience from past selection processes. This organization will find valuable to have means to transfer knowledge from one experience to another and to assist their clients in the maintenance of their COTS-based software systems.

As a result, diverse actors may benefit from our approach:

- IT consultant companies offering assessment for business automation may structure their services better.
- Commercial web-based companies or portals may structure their offering in well-founded categories with a clear rationale behind.
- Medium- and large-size companies with their own IT department may be more confident on their own selection processes.
- Software engineers which usually carry out COTS components selection may structure better their knowledge and may aim at a better return of investment.

At the time being, we have experimented our GBTCM+ method in the following fields: Real-Time Synchronous Communication Systems, Message-based Communication Systems, some sub-categories of Enterprise Applications (with emphasis with those related to Content Management) and Requirements Engineering Tools. The results are promising from the academic point of view, but we have not had the chance yet to make a proper validation involving an industrial partner, by means of some action-research collaboration as we have done in the past. Industrial validation is our main aim for future work. We also are going to tackle immediately development of tool support starting from the UML conceptual model presented in this report.

References

1. Li, J., Conradi, R., Slyngstad O.P., Bunse, C., Khan, U., Torchiano, M., Morisio, M. "Validation of New Thesis on Off-The-Shelf Component-Based Development". *11th IEEEInternational Software Metrics Symposium, 2005*.

2. Ruhe, G. "Intelligent Support for Selection of COTS Products". In *Proceedings Web Databases and Web Services 2002*. LNCS 2593, pp. 34-45, 2003.

3. Morisio, M., Seaman, C., Basili, V., Parra, A., Kraft, S., Condon, S. "COTS-based software development: Processes and open issues" *Journal of Systems and Software* 61(3): 189-199 (2002)

4. Prieto-Díaz, R., Freeman, P. "Classifying Software for Reusability" IEEE Software. January 1987 pp. 6-16.

5. Franch, X., Maiden, N. "Modelling Component Dependencies to Inform their Selection". Proceedings 2nd International Conference on COTS-Based Software Systems. 2003.

6. Baskerville R., Wood-Harper A.T. "Diversity in Information Systems Action Research Methods. European Journal on Information Systems. Vol. 7 No.2, June 1998.

7. Martin, P., Turner, B. "Grounded Theory and Organizational Research". Journal of Applied Behavioral Science.1986; 22: 141-157.

8. Carvallo, J.P., Franch, X., Quer, C., Torchiano, M. "Characterization of a Taxonomy for Business Applications and the Relationships among Them". Proceedings of 3rd International Conference on COTS-Based Software Systems, LNCS 2959, 2004.

9. van Lamsweerde, A. "Goal-Oriented Requirements Engineering: A Guided Tour" Proceedings 5th IEEE International Symposium on Requirements Engineering, August 2001.

10. Antón, A. I.. "Goal Identification and Refinement in the Specification of Software-Based Information Systems". Ph.D. thesis, Georgia Institute of Technology, June 1997.

11. Ayala, C.P, Botella, P., Franch, X., "On Goal-Oriented COTS Taxonomies Construction" Proceedings 4th International Conference on COTS-Based Software Systems, LNCS 3412, February 2005.

12. Gartner Inc. www.gartner.com.

13. Forrester Research Inc. www.forrester.com

14. ComponentSource componentsource.com

15. Genium Software Development http://www.genium.dk/index. xml

16. INCOSE www.incose.org.

17. IT products guide. http://productguide.itmanagersjournal.com

18. eCOTS www.ecots.org

19. CBSE www.cbsenet.org/pls/CBSEnet/ecolnet.home

20. Arranga, E. "Cobol Tools: Overview and Taxonomy". IEEE Software, 17(2): 59-61, 2000.

21. Llorens, J., Astudillo, H. "Automatic Generation of Hierarchical Taxonomies from Free Text Using Linguistic Algorithms". Proceedings of Advances in Object-Oriented Information Systems: OOIS 2002 Workshops, 2002.

22. Glass, R.L., Vessey, I. "Contemporary Application-Domain Taxonomies" IEEE Software July 1995.

23. SWEBOK www.swebok.org.

24. Ochs, M.A., Pfahl, D., Chrobok-Diening, G., Nothhelfer-Kolb, B. "A Method for Efficient Measurement-based COTS Assessment and Selection- Method Description and Evaluation Results" Proceedings IEEE 7th International Software Metrics Symposium, 2001.

25. Carney D., Long F. "What Do You Mean by COTS? Finally a Useful Answer" IEEE Software, 17 (2), March/April 2000

26. Bianchi, A; Caivano, D; Conradi, R; Jaccheri, L; Torchiano, M; Vissagio, G. "COTS Products Characterization: Proposal and Empirical Assessment". Proceedings ESERNET 2001-2003. LNCS 2765, 2003.

27. Erofeev, S., DiGiacomo, P. "Usage of Dynamic Decision Models as an Agile Approach to COTS Taxonomies Construction". Proceedings International Conference on COTS-Based Software Systems 2006.

28. Cechich, A., Réquilé-Romanczuk, A., Aguirre, X., Luzuriaga, J. "Trends on COTS Component Identification and Retrieval" Proceedings International Conference on COTS-Based Software Systems 2006.

29. Réquilé-Romanczuk, A; Cechich, A; Dourgnon-Hanoune, A; Mielnik, J-C. "Towards a Knowledge-Based Framework for COTS Component Identification". *Proceedings ICSE-MPEC 05, St Louis, Missouri, USA. ACM 2005*.

30. UML Specifications http://www.uml.org/

31. Prieto-Díaz, R., Arango, G. "Domain Analysis and Software Systems Modelling". IEEE Computer Society Press, 1991. p. 300.

32. ISO/IEC International Standard 9126-1 "Software Engineering-Product Quality-Part 1: Quality Model" 2001.

33. Leite, J.C.S.P. "Application Languages: A Product of Requirements Analysis" Informatics Department PUC-/RJ (1989).

34. Regev, G. "Where do Goals Come from: the Underlying Principles of Goal-Oriented Requirements Engineering". 13th IEEE Requirements Engineering Conference 2005.

35. Potts, C., Takanashi, K., Antón, A. "Inquiry-Based Requirements Analysis", IEEE Software, 11 (2), March 1994.

36. Yu, E. "Modelling Strategic Relationships for Process Reengineering" PhD Thesis, University of Toronto, 1995.

37. Ayala, C., Franch, X., "Transforming Software Package Classification Hierarchies into Goal-Based Taxonomies" 16th International Conference on Database and Expert Systems Applications. August 2005.