Construction of a Taxonomy for Requirements Engineering Commercial-Off-The-Shelf Components

Claudia Ayala, Pere Botella and Xavier Franch
Universitat Politécnica de Catalunya, UPC-Campus Nord (Omega)
08034 Barcelona, Catalunya, Spain.
{cayala,botella,franch}@lsi.upc.edu
http://www.lsi.upc.edu/~gessi

ABSTRACT

This article presents a procedure for constructing a taxonomy of COTS products in the field of Requirements Engineering (RE). The taxonomy and the obtained information reach transcendental benefits to the selection of systems and tools that aid to RE-related actors to simplify and facilitate their work. This taxonomy is performed by means of a goal-oriented methodology inspired in GBRAM (Goal-Based Requirements Analysis Method), called GBTCM (Goal-Based Taxonomy Construction Method), that provides a guide to analyze sources of information and modeling requirements and domains, as well as gathering and organizing the knowledge in any segment of the COTS market. GBTCM claims to promote the use of standards and the reuse of requirements in order to support different processes of selection and integration of components.

Keywords: COTS components, Goal-Oriented, Taxonomy, GBTCM method, Knowledge reuse, RE

1. INTRODUCTION

The use of Commercial Off-The-Shelf (COTS) components (hereafter, COTS) as parts of larger systems has grown steadily [1, 2]. The process of developing systems from COTS is an economic and strategic need in a wide variety of different application areas. As a result, a huge amount of COTS have become accessible in the market. This gives raise to a new problem: how to organize the knowledge about these COTS in such a way that searching the market becomes a feasible task.

In [3] we proposed to use *taxonomies* as a way to organize the COTS market (see fig. 1). The general idea was to construct a decision tree, the leaves of the tree representing *COTS domains*; a COTS domain encloses a significant group of functionality (e.g., the domain of anti-virus tools or mail servers systems). Domains are grouped into *categories* (e.g., the category of communication infrastructure systems or financial packages), which may be grouped at their turn. We proposed the use of *characterization attributes* [4] to discriminate among different categories or domains. We bind questions and answers to these attributes as a way for browsing the taxonomy. Dependencies among domains that belong to the taxonomy are included in the hierarchy itself (e.g., mail server systems depend on anti-virus tools to support integrity). As an additional point, we also bind *quality models* to nodes in the taxonomy, each describing the quality factors that are of interest for the particular category or domain; quality models are inherited downwards the taxonomy.

Although the main ideas of our proposal were satisfactory enough, it turned out that the way to identify the discriminating characterization attributes (which capture the relevant information for discriminating categories and domains) was not properly defined. Therefore, we carried out a research work to discuss the applicability of goal-based approaches for generalizing, formalizing, improving and clarifying the process of identifying and evaluating characterization attributes in a formal way [5] instead of only common sense [3]. Thus, we specifically evaluated the GBRAM (Goal-Based Requirements Analysis Method) proposed by Annie I. Anton [6] in the field of software requirements.

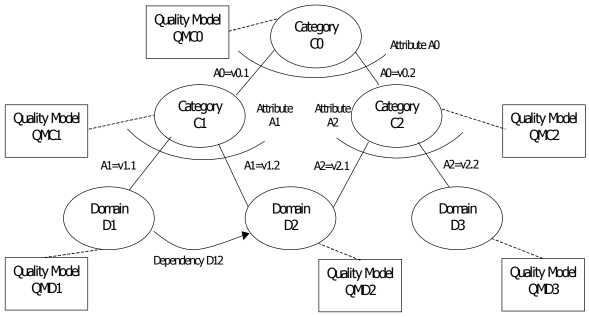


Fig 1. The fundamental elements of a taxonomy

In [7] we depart from the general idea of the GBRAM activities and adapt it to the COTS context, obtaining the GBTCM method (*Goal-Based Taxonomy Construction Method)* that provides a guide to analyze sources of information and modeling requirements and domains, as well as gathering and organizing the knowledge in any segment of the COTS market. In addition GBTCM contribute and enhance the requirements and knowledge reuse in different processes of selection and integration of components.

We use the notion of *goal* as introduced in the context of requirements engineering [8, 9]. Goals are the rationale to identify characterization attributes and therefore COTS categories and domains. In general, goals are very stable with respect to changes, and goal refinement provides a natural mechanism for structuring and exploring many alternatives in the COTS market.

The purpose of this article is to describe the application of GBTCM to a particular segment of the COTS market, that is, the systems and tools for supporting the various activities embraced by the RE phase, obtaining as a result a general taxonomy and the knowledge organization in that area.

2. GBTCM: CONTRUCTION OF COTS TAXONOMIES

Fig. 2 shows the activities (ovals) and artifacts (inclined rectangles) involved in GBTCM. The two high level phases are: Goal Analysis (concerns the exploration of available information sources for goal identification followed by the organization and classification of goals) and Goal Refinement (concerns the evolution of goals from the moment they are first identified to the moment that they are translated into requirements).

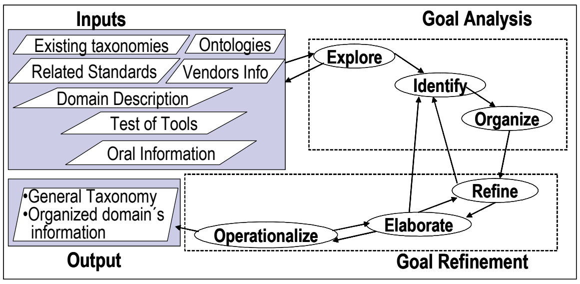


Fig. 2 General Activities of GBTCM

The general activities are: *Explore* (entails the analysis of available information), *Identify* (aims at extracting goals applying techniques as heuristics), *Organize* (involves the classification and organization of goals according to goal dependency relationships), *Refine* (entails the actual pruning of the goal set), *Elaborate* (refers to the process of more refinement, analyzing the goal set by considering possible obstacles and constructing scenarios to uncover hidden goals and requirements), and *Operationalize* (refers to translating goals into requirements for the final taxonomy). In table 1 we can realize that the output of each activity is the input of the next.

Activity	Outputs
Explore	Information sources qualified; Some goals
Identify	Set of goals; Stakeholders and agents; Auxiliary models and artifacts
Organize	Matching of goals from different information sources Dependency relationships among goals Goal hierarchy
Refine	Refined goal set
Elaborate	Scenarios Constraints
Operationalize	Hierarchical structure of Goals Associated information and models and artifacts Characterization attributes for constructing the taxonomy of the domain

Table 1. Main activities and their outputs

Certainly, the information sources (inputs) are the basis for obtaining goals. The final result of GBTCM is not only a taxonomy of COTS in a specific area, but also a set of information and knowledge (repository) that contains: all the activities performed in that area (expressed as goals) and the dependency relationships among COTS domains, in order to assess the impact of changes among domains.

Basically, GBTCM guides to gather, manage, and generalize information related with a domain (e.g. existent taxonomies, standards in the field, and vendors). It also may include artifacts and models (e.g. UML class diagrams, *i** models) that permit to ensure consistency and evolution of the repository of knowledge. This repository can be used during different selection process. It is also the source for constructing a general taxonomy and both (the repository and the general taxonomy) could be the basis for organizations to build up the most suitable taxonomy to their processes.

We refer to [7] for more details in GBTCM.

3. A COTS TAXONOMY FOR THE REQUIREMENTS ENGINEERING FIELD

It is widely acknowledged within the software industry that software projects are critically vulnerable when the activities of software requirements are performed poorly; also reports exist supporting this hypothesis [10]. Therefore, to improve the efficiency of the activities performed in the area, COTS technology aid RE-related actors to simplify and facilitate their work.

Our purpose in this section is to apply GBTCM to the field of requirements engineering in order to propose a general COTS components taxonomy for that field.

This taxonomy helps Software Engineers (which usually carry out COTS components selection and integration) to structure and reuse better their knowledge for their repeated use during different selection process; on the other hand, we argue that the taxonomy and the information and knowledge obtained (repository) could be the basis for organizations to build up the most suitable taxonomy of requirements engineering COTS components to their processes.

Information Sources

As it can be expected for a topic such this, lots of information sources exist and many of them were gathered [11-22]. Table 2 summarizes the most important ones we considered in this case study and shows details in a general manner of the kind of information therein.

We can note that most information is textual, available in printed form or the web, issued by different organizations or people. Sources such as domain experts and tools demos still remain, but they play a secondary role.

It is important to remark that we pretend to represent information that does not depend on a concrete project or software system, but also to create a general repository with as much information as possible related with the RE area, that permits its later applicability and suitability in different organizations.

Analyzing Infomation Sources and Identifying Goals and Objectives

The use of one or another information source is determined by several qualities, among which we mention: reliability of the information, availability of the source, acquisition cost, timeliness, scope covered and time needed to process the enclosed information. These qualities depend on three factors: information source type, organization or people that creates the information, and particular item of information.

It is considered as a good practice to fundament goal analysis on the most solid and confident of the sources for extracting the main high level goals in order to assure the consistency of the set of goals, and then extracting sub-goals from the remaining sources.

Due to the standard nature of SWEBOK in the field, we started with this source for obtaining the high-level goals that guide the whole process (even considering that SWEBOK is not tool-oriented, on the contrary of other sources). For example, consider the following description in natural language from SWEBOK: “The next topics breakdowns for RE discipline are generally accepted in that they cover areas typically in texts and standards: activities such as Requirements Engineering Process, Requirements Elicitation, and Requirements Analysis, along with Requirements engineering-specific descriptions. Hence, we identify Requirements Validation and Requirements Management as separate topics”. By examining the statement and asking “what goal(s) does this statement/fragment exemplify?” some goals become evident from the description. We present some of these goals in the first column of table 3. These goals are going to be decomposed in sub-goals by means of the analysis of other information sources and then, applying refinement techniques. We will further describe the refinement process in the next subsections.

Type of source	Source organization	Information enclosed	Comments
Existing taxonomies	INCOSE	Classification of Software Engineering tools	This section is available free and widely accepted
Related standards	SWEBOK	Main RE areas stakeholder types	Available free, widely accepted
	IEEE std 830-1998	Software activities related with RE	Subscription/payment needed
	IEEE/EIA 12207.1-1997		
	ISO/IEC 12207		
Vendors information	IBM-Rational	Capabilities of products and trends	Exhaustive description of products
	ComponentSource	Capabilities of products and trends focused in platforms	Available free, widely accepted
Tools	RequisitePro	Capabilities of a real RMT	Included in the IBM-Rational Suite
	IRqA		Tool used often in our projects
	EasyWinWin	Capability of a research tool for requirements negotiation	Some tutorials attended and contacts with authors
Academic sites	eCOTS	Trends	Available free, widely accepted
Scientific items	RE-related conferences	Timely state of the art	Subscription/payment needed
	RE&SE textbooks	Areas of RE	
Magazines	Requirements Engineering	Trends and timely state of the art	
WebSites	Volere	RE resources	Available Free
Technical	INCOSE	Trends and concepts in RE	Subscription/payment needed
	Gartner		
Own experiences	Academic records management	Use of RE-oriented tools in a real project	CMM-2 compliant requirements management

Table 2. Main sources of information used in the RE taxonomy

Goals	Agents	Stakeholders
G1:Process of Software Requirements Defined	(RE)	RE, PM,QAM
G2:Requirements Elicitation Performed	RE	RE, Stakeholders
G3:Requirements Analysis Performed	RE	RE, Stakeholders
G4:Requirements Specification Done	RE	RE, users/customer, QAM
G5:Requirements Validation Performed	RE	RE, users/customer, Tester
G6:Requirements Management Done	RE	RE, SCM
G6.1:Change Management in Requirements Controlled	RE	RE
G6.2:Requirements Attributes Defined	RE	RE, SCM
G6.3:Requirements Tracing Controlled	RE	RE, SCM

Table 3. Some goals obtained from SWEBOK

Identifying Stakeholders and Agents

At this stage, we aim at determining who are the stakeholders involved in the achievement of goals. Once the goals and stakeholders are specified, the goals must be assigned to their responsible agent(s).

A stakeholder is any representative affected by the achievement or prevention of a particular goal. Multiple stakeholders may be associated with one goal.

Agents are responsible for the completion and/or satisfaction of goals within an organization or system. Identification of stakeholders and agents is crucial to understand the domain at hand and also to identify additional sources of information, e.g. for identifying people to be interviewed. The stakeholders for each goal are determined by asking “who or what claims a stake in this goal?” and “who or what stands to gain or lose by the completion or prevention of this goal?”

For identifying which agents are ultimately responsible for the achievement of each goal, we ask the question “who or what agent [is/should be/could be] responsible for this goal?”

In our case, we identified as stakeholders (see table 3): Requirements Engineer (RE), Project Manager (PM), Quality Assurance Manager (QAM), Software Configuration Manager (SCM), Testers, Final Users, Customer and Non-Technical Stakeholders (such as regulators, market analyst, system developers; NTS). The only agent is the Requirements Engineer. Some relationships and dependencies among stakeholders are showed in the *i** SD model in fig 3.

Auxiliary Models and Artifacts

GBTCM considers essential the generation of some artifacts and models from the information sources in order to understand, handle, formalize, and remarkably guarantee the integrity and consistency of the information respect to changes and evolution. Hence, it suggests: creating glossaries for homogenizing terms used in diverse information sources, UML class diagrams [23] for representing a conceptual model of the domain and to define the underlying ontology [24]; and as a fundamental part of the method, it requires the construction of goal-oriented models: specifically *i** as notation [25] although other options are valid. Of course all of these artifacts and models shall be synchronized (e.g. glossary terms and UML attributes should have the same name). This models and artifacts should be refined during all process.

In the RE case study, we create: *i** models (we can see an excerpt in fig. 3), glossaries and class diagrams in order to homogenize information from different sources, facilitating the communication. For example, the terms “capturing” and “extracting” coming from two different sources were unified and defined as “extracting” in our glossary.

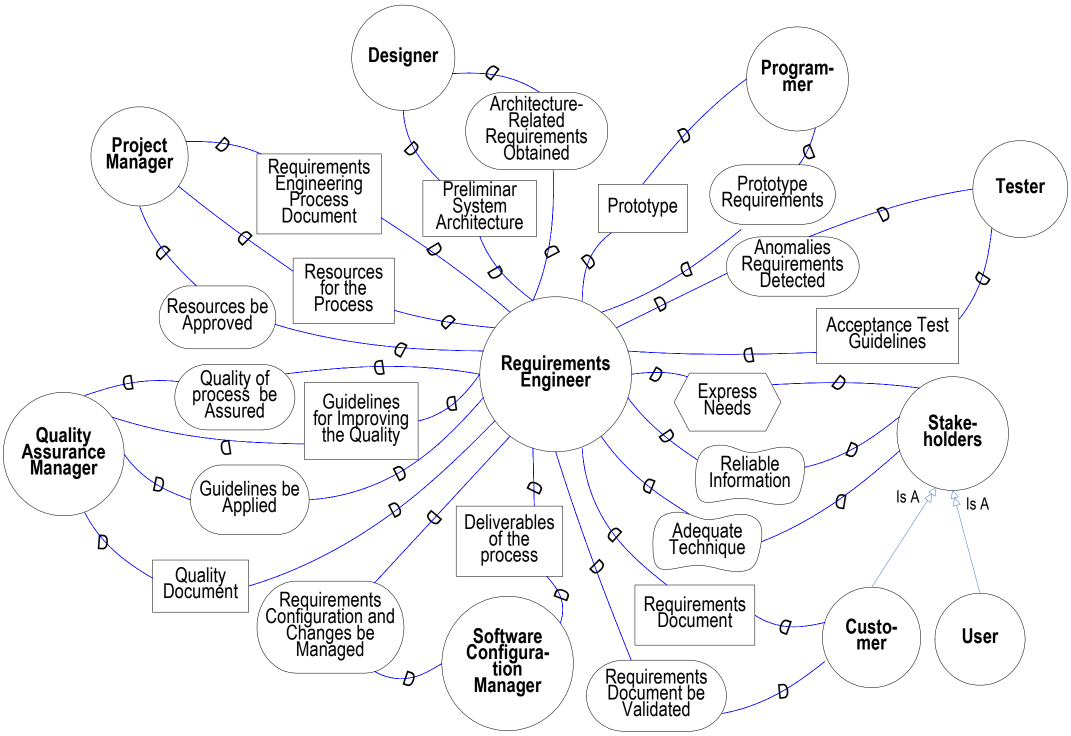


Fig. 3 Excerpt of the *i** SD model that shows relationships among system actors

Use and Creation of Heuristics

Heuristics aid us by providing prescriptive guidance for managing varying levels of detail in the information available.

Previous sections showed examples of the application of heuristics that guide the obtention of specific information by means questions. These heuristics are general and do not depend on any particular domain [6]. An important objective of GBTCM is the reuse of knowledge acquired from different case studies.

Thus, the creation of new specific heuristics of the domain is a desirable task, in order to their repeated use in other cases, achieving a high probability of success finding goals while avoiding wasted efforts.

This article focus on the general description of the application of GBTCM method to RE field; so we do not present new RE domain heuristics. More than anything else, we are aware that is necessary to perform in depth more practical cases for the foundation of new heuristics. This is part of our ongoing work.

Organization and Matching of Goals

Once goals have been analyzed and identified from all information sources, we have to organize the information firstly by means of a matching of goals from all of them, and so on, according to precedence relationships.

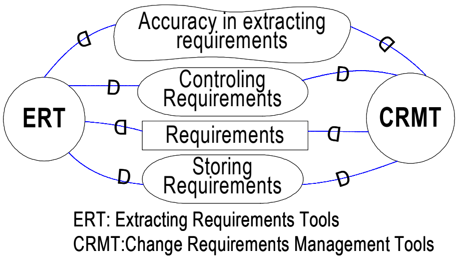
In the matching process we can observe that a goal should be taken into account in the taxonomy construction only if it exist in the market a tool that supports it (although it could be argued that discovering of goals that are not covered by any tool is a significant issue in closing the gap between tools and processes).

We perform the process of organization of goals by means of tables. Table 4 is an excerpt of this process. There we can observe the matching among some information sources: identified goals from related standards, scientific items, etc.; vendor's information of tools; and nodes of INCOSE taxonomy. At the end of the matching process we have a more complete set of goals. This resulting set of goals going to be refined in subsequent steps but firstly dependencies relationships should be specified before.

Departing from this set of goals, hierarchical dependencies are defined (by means of hierarchical tables, called goal topography). Then, refinement processes are applied concurrently with the identification and specification of dependencies among goals. This specification is done, as we mentioned before, by means of *i** models; it supports the explicit representation of potential dependencies among COTS domains. This mechanism assures the traceability of the impact of changes among domains.

Goals	Tools	Cathegory of INCOSE Taxonomy
G2:Requirements Elicitation Performed		
G2.1:Requirements Sources Defined and Analized		
G2.2:Elicitation Techniques Chosen		
G2.2.1:Extracting Requirements	Yes	RequirementsEngineering/Requirements Management/RequirementsCapture&Identification/ ToolsForElicitationOfRequirements
G2.2.1.1:Interviews	Yes	
G2.2.1.2:Scenarios	Yes	Design Domain

Table 4. An excerpt of organization of goals



Goals		
M1: Extracting Requirements		
	M1.1:Accuracy in Extracting Requirements	
	M1.2:Storing Requirements	
		M1.2.1: Controling Requirements

Fig. 4. An example of the i* SD model (left) and a hierarchical table (rigth) involving the RE tools

Refinement of the Goal Set

GBTCM is based on identification of goals and organization of knowledge from an iterative refinement process.

Once a set of goals was obtained from the matching of all information sources, and their hierarchical dependencies were specified , they are analyzed for refinement.

We used some refinement techniques proposed by GBTCM: the use of scenarios (understood as behavioral descriptions of the system and its environment) that refers to the recreation of the different situations and circumstances in which a goal is executed; the application of the Inquiry Cycle [26], a formal structure for describing discussions about requirements that consists of a series of questions and answers designed to pinpoint where information needs come from and when; and also the use of glossaries to support the reconciliation of goals and UML class diagrams.

Scenarios were used in the form of general use-case, and were represented as sequences of actions in natural language descriptions. We applied them in order to be more exhaustive and included as much as possible activities performed in the RE phase in the most projects of software development.

For instance, by means the construction of scenarios for the goal Requirements Analysis Performed, we detected (among other issues) that in many cases the process of analyzing and elaborating requirements demands (in order to be achieved) to identify the subsystems and components that will be responsible for satisfying the requirements, so we had to consider Architectural Design Done as one goal in this area.

Scenarios were very useful for uncovering and reconciling goals, checking for completeness and conflicts, and communicating with stakeholders.

We have to say that the applicability and precision in the scenarios construction depends on the criticism that the domain requires, it means that a highly critical domain (e.g., aircraft applications) requires the construction of the most detailed scenarios and special techniques.

At the end of this activity, 3 patterns of refinement were found: eliminating duplicated goals, refining goals based on system entities and consolidating synonymous goals.

Operationalizing Goals

Goal information must be ultimately operationalized (related with actions) and the general taxonomy be constructed. This is done by consolidating the goal information, and applying the Inquiry Cycle.

The Inquire Cycle was very helpful for finding the adequate question-answer pair(s) to be bound to each characterization attribute, and also in order to organize the resulting information.

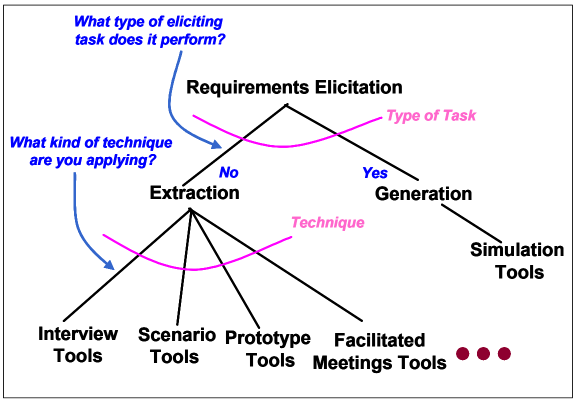


Fig. 5. Excerpt of the RE taxonomy represented as a decision tree

Level 2	Level 3	Question	Level 4
1.2 Requirements Elicitation			
	1.2.1 Generation	Do you apply simulation for generating requirements?	1.2.1.1 Simulation Tools
	1.2.1 Extraction		
		Are you using interviews?	1.2.2.1 Interview Tools
		Are you using scenarios?	1.2.2.2 Scenario Tools
		Are you using prototypes?	1.2.2.3 Prototype Tools
		Are you using facility meetings?	1.2.2.4 Facilitate Meetings Tools
		Are you using observation?	1.2.2.5 Observation Tools
		Are you using other techniques?	1.2.2.6 Other Extraction Techniques Tools

Table 5. An excerpt of the RE taxonomy

Table 5 shows and excerpt of the resulting taxonomy. Basically, each answer to the question gives place to a new level (for instance, in response to the question(s) of the column 3, level 4 arise).

Fig. 5 shows the same example as a decision tree. For lack of space we do not present all question-answer(s) related to each level, we only present a short schema in order to give a general idea of the taxonomy and resulting information.

4. CONCLUSIONS AND FUTURE WORK

From our experience we conclude that applying GBTCM supports more effective and reliable identification and organization of the knowledge about the RE domain, which are the base for the construction of goal-oriented COTS taxonomies.

In general GBTCM represents a practical approach that helps the elicitation, specification, selection and integration of COTS, based on the reuse of requirements and knowledge, and software engineering standards.

As we mentioned before, the knowledge (repository) and the taxonomy obtained help to Software Engineers -which usually carry out COTS selection and integration- to

structure and reuse better their knowledge for their repeated use during different selection process.
Above all, we claim that the taxonomy and the information and knowledge obtained (repository) could be the basis for organizations to build up the most suitable taxonomy of RE COTS according to their processes [27]. This means a better return on investment.
Our ongoing work is related with performing in depth more case studies to formalize a knowledge domain in RE with established heuristics, and complementing the RE repository with goals addressed to include modern techniques of software engineering (e.g. Agile RE, Aspect-Oriented, etc.) not only the most traditional ones shown in this article.
As future work, we are interesting in describing how the produced taxonomy can be used in a specific selection process, and how this general taxonomy can be mapped into the requirements to each organization.
Also, we will construct an exhaustive program of analysis of information sources that allows taking into account the qualities and factors of the information sources for classifing them accordingly to their relevance. In addition, we will investigate the applicability of other existent techniques for the construction of taxonomies and hierarchies as laddering or neural networks for their incorporation to GBTCM.
It is important to remark that this work is part of the PhD thesis of the first author, and it will cover the validation process of the methodology.

Acknowledgements
This work has been done in the framework of the research project UPIC, ref. TIN2004-07461-C02-01, supported by the Spanish Ministerio de Ciencia y Tecnología. Claudia Ayala's work has been partially supported by the Mexican Council for Science and Technology (CONACYT).

5. REFERENCES

[1] Carney D., Long F., "What Do You Mean by COTS? Finally a Useful Answer". *IEEE Software*, 17 (2), March/April 2000.
[2] Craig Meyers, B., Oberndorf, P., "Managing Software Acquisition", *SEI Series in Software Engineering*, 2002.
[3] Carvallo, J.P., Franch, X., Quer, C., Torchiano, M., "Characterization of a Taxonomy for Business Applications and the Relationships Among Them", *Proceedings 3rd International Conference on COTS-Based Software Systems (ICCBSS),* 2004, Lecture Notes in Computer Science Vol. 2959.
[4] Morisio, M., Torchiano, M., "Definition and Classification of COTS: A Proposal", *Proceedings 1st. International Conference on COTS-Based Software Systems (ICCBSS)*, Orlando Florida, 2002.
[5] Ayala, C.P, Botella, P., Franch, X., "Goal-Based Reasoned Construction of Taxonomies for the Selection of COTS Products", *Proceedings 8th Multi-Conference on Systemics, Cybernetics and Informatics (SCI),* Orlando Florida, July 2004. ISBN:980-6560-13-2.
[6] Antón, A.I., "Goal-Based Requirements Analysis", *Proceedings 2nd IEEE International Conference on Requirements Engineering (ICRE)*, 1996.
[7] Ayala, C.P, Botella, P., Franch, X., "On Goal-Oriented COTS Taxonomies Construction", *Proceedings 4th International Conference on COTS-Based Software Systems (ICCBSS)*, Bilbao, Spain, Febrero 2005.
[8] Mylopoulos, J., Chung, L., Yu, E., "From Object-Oriented to Goal-Oriented Requirements Análisis", *Communications ACM* 42(1), January 1999.
[9] Van Lamsweerde, A., "Goal-Oriented Requirements Engineering: A Guided Tour", *Proceedings 5th IEEE International Symposium on Requirements Engineering (ISRE)*, 2001.
[10] Standish Group Report http://www.standishgroup. *CHAOS Survey* 1994.
[11] INCOSE. "Software Engineering Tools Taxonomy" http// www.incose.org
[12] Guide to the Software Engineering Body of Knowledge, SWEBOK, www.swebok..org
[13] International Standard IEEE Std 830-1998 IEEE, "Recommended Practice for Software Requirements Specifications", June 1998.
[14] IEEE/EIA Guide. IEEE/EIA12207.1-1997, "Standard for Information Technology Software Life Cicle Processes", April 1998.
[15] International Standard ISO/IEC 12207 "Software Life Cycle Processes", 1995.
[16] ComponenSource http://www.componentsource.com
[17] eCots. Software Components Open Directory Project. http://ecots.org
[18] The Gartner Group, available on-line at http://www4.gartner.com
[19] Infrastructure Software Market Definitions for Application Development. Gartner, Dataquest Guide. 4th June, 2003. http://www.gartner.com
[20] Software Market Research Methodology and Definitions 2003-2004, Gartner Dataquest Guide, January, 2004. http://www.gartner.com
[21] CBSE Net. "Application Domain Taxonomy ". Available on-line (previous registration) at: http://www.cbsenet.org/pls/CBSEnet/eco_ricerca_documenti.concept_search_frame
[22] International Data Corporation http://www.idc.com
[23] UML 2.0 Specifications http://www.uml.org/, last accesed January 2005.
[24] Carvallo, J.P., Franch, X., Quer, C., "Defining a Quality Model for Mail Servers", *Proceedings 2nd International Conference on COTS-Based Software Systems (ICCBSS)*, 2003, Lecture Notes in Computer Science 2580
[25] Yu, E., "Towards Modeling and Reasoning Support for Early-Phase Requirements Engineering", *Proceedings 3rd IEEE International Symposium on Requirements Engineering (ISRE)*, 1997.
[26] Potts, C., Takahashi, K., Antón, A., "Inquiry-Based Requirements Analysis", *IEEE Software*, 11(2), March 1994.
[27] Basili, V.R., Caldera, G., Rombach, H.D., "The Experience Factory", J.J Marciniak (ed), *Encyclopedia of Software Engineering. vol 1.* pp 469-476. John Wiley & Sons, 1994.