

On Goal-Oriented COTS Taxonomies Construction*

Claudia P. Ayala¹, Pere Botella, and Xavier Franch

UPC-Campus Nord, c/Jordi Girona 1-3, 08034 Barcelona, Catalunya, Spain
{cayala, botella, franch}@lsi.upc.es
<http://www.lsi.upc.es/~gessi>

Abstract. This paper proposes the adoption of a goal-based method called GBRAM for facilitating the process of building taxonomies of COTS components. Since GBRAM was defined in a different setting, the main result of the paper is to adapt it to this new context obtaining the GBTCM method. We show how the different activities and artifacts of GBRAM change, and we apply the proposal to obtain a taxonomy for requirements engineering oriented tools.

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) and we applied the proposal to the family of business applications. At the leaves of the taxonomy there are *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 the proposal were satisfactory enough for our purposes, it turned out that the way to identify the discriminating characterization attributes

* This work is partially supported by the Spanish research program CICYT TIC2001-2165.

¹ Claudia P. Ayala's work has been supported by the Mexican Council for Science and Technology (CONACYT).

(which capture the relevant information for discriminating categories and domains) was not properly defined. In [3], we just took an existing taxonomy as starting point and rearranged it by observation. This was clearly a weak point of our proposal, and therefore we started to look for a better suited strategy.

In this paper, we use the notion of *goal* as introduced in the context of requirements engineering [5, 6] as 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. Our main contribution in this paper is to present a goal-based reasoning method based on the *Goal-Based Requirements Analysis Method* (GBRAM) proposed by Annie I. Antón in the field of software requirements [7] to the context of COTS taxonomies. The resulting method, *Goal-Based Taxonomy Construction Method* (GBTCM), help us to generalize, formalize, enhance and clarify the process of building taxonomies by identifying and evaluating the most suitable characterization attributes. We apply GBTCM to a particular segment of the COTS market: systems and tools for supporting the various activities embraced by the requirements engineering phase. The resulting taxonomy can be considered as another contribution of this work.

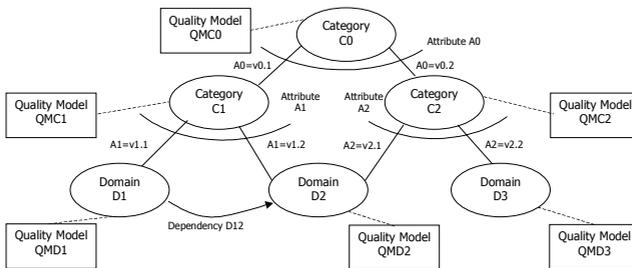


Fig. 1. The fundamental elements of COTS taxonomies

2 The Goal-Based Requirements Analysis Method (GBRAM)

GBRAM was formulated with the transformation of enterprise and system goals into requirements as primary focus, more specifically to assist analysts in gathering software and enterprise goals from many sources and to support the process of discovering, identifying, classifying, refining, and elaborating goals into operational requirements. The method's main contribution is the provision of heuristics and procedural guidance for identifying and constructing goals.

The two high level phases of GBRAM briefly explained are:

- **Goal Analysis.** Concerns the exploration of available information sources for goal identification followed by the organization and classification of goals.
- **Goal Refinement.** Concerns the evolution of goals from the moment they are first identified to the moment they are translated into operational requirements.

Fig. 2 shows the activities (ovals) and artifacts (inclined rectangles) involved in GBRAM. Its output is the Software Requirements Document (SRD) and its inputs are the diverse sources of information used in requirements elicitation. The activities are: *Explore* (entails the analysis of available information), *Identify* (aims at extracting goals applying 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 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 SRD).

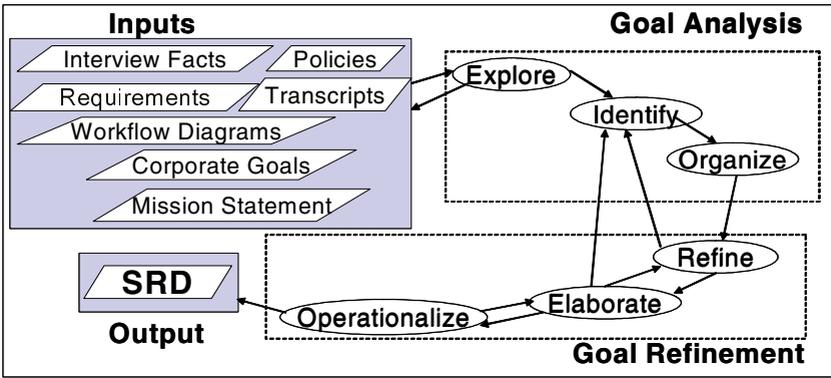


Fig. 2. Overview of the GBRAM activities

3 Customizing GBRAM for the Construction of COTS Taxonomies

We aim at applying GBRAM in a different context from which it was conceived, that is the construction of taxonomies for the COTS market using goals acquired from different information sources. The suitability of this method as a help for obtaining the characterization attributes is based on three facts:

1. It assumes the challenge of working with different sources of knowledge that are represented in different form.
2. It provides guidelines and heuristics for exploring, identifying and organizing goals (potential characterization attributes) and also allows adding new heuristics which guide us towards a high probability of success while avoiding wasted efforts.
3. It offers a guide for applying an inquiry-driven approach to goal-based analysis, that can be useful for enhancing our questions-answers mechanism linked to characterization attributes.

In the rest of the section we show how we customize GBRAM. We adjust the inputs and modify the output. We adapt and prune some activities treating the issues not as operational requirements of a system but as characterization attributes in order

to obtain the statement of main goals to be considered as potential characterization attributes in the construction of COTS taxonomies in any specific area.

3.1 Adjusting the Sources of Information

We want to emphasize the importance of the information sources (the inputs), since they are the base for obtaining goals. GBRAM's inputs correspond to usual software requirements elicitation sources in order to define the SRD for developing a software system. But the inputs for GBTCM must necessarily be different.

Table 1 summarizes the most important sources of information we consider in GBTCM. The main difference is 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.

Table 1. Information Sources to consider in our approach

Information Source		Information type	Language	Examples	
Existing Taxonomies and Ontologies		Classifications; Categories; Glossaries	Natural Language (NL); Tree-like diagrams	SWEBOK, INCOSE, Gartner, IDC	
Related Standards		Descriptions; Glossaries	NL	IEEE, EIA, ISO	
Vendors Information		Brochures; Evaluation forms; Benchmarks	NL; Values for attributes	Rational, Microsoft	
Domain Descriptions	Scientific	Academic Events, Journals; Textbooks	Precise and rigorous descriptions	NL; Models; Formulas; Schemas	ICCBSS, ICSE, TSE, [MO02]
	Divuligation	Magazines, Forums and Websites	Descriptions and tips for the general public	NL; Schemas; Tables	PCWorld, IEEE Software, COCOTS website, specialized forums
	Technical	White Papers, Surveys and Comparatives	Papers, Comparative tables	NL, Tables; Figures	Gartner, INCOSE
Oral Information	Interviews		Knowledge; Tips; Practical information	NL	ICCBSS panels, SEI courses, Business luncheons
	Talks, seminars and courses				
Test of Tools and Systems		Test results; User's manuals	Visual data; NL	Outlook, Rationale Suite	
Experiences on the field		Knowledge; Technical reports	Knowledge; NL	Past projects made	

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. A complete goal-acquisition program should take these considerations into account.

3.2 Auxiliary Models and Artifacts

Additionally to GBRAM, GBTCM considers essential the generation of some artifacts and models from the information sources cited above in order to understand, handle, formalize, and remarkably maintain the information about the domain.

As artifacts, we suggest at least to create glossaries for homogenizing terms used in the diverse information sources. As models, we suggest UML class diagrams [8] for representing a conceptual model of the domain, defining by means of classes, attributes, associations and hierarchies the underlying ontology (see [9]). As a fundamental part of our approach, we require the construction of goal-oriented models. We use i^* as notation [10] although other options are valid. Goal-oriented models align with our goal-oriented method, therefore on the one hand they aid in the process of getting goals providing a high level picture of the domain, representing and organizing its knowledge and related activities; on the other hand they are used in order to represent and record explicitly the dependencies among domains of COTS for their repeated use during different selection process, and the relationships among these domains and the actors that have an interest on them, either as users or as definers.

Of course all of these artifacts and models shall be synchronized. For instance, glossary terms and UML attributes should have the same name.

3.3 GBTCM Activities

Some activities related with goal analysis or refinement have been pruned or adapted to our approach. GBTCM finally delivers not an SRD as the original GBRAM, but a hierarchical structure of the more important goals covered in the addressed COTS market segment; the correspondence of these goals with the different sources of information (especially with existing taxonomies, standards and vendors information); the auxiliary models and artifacts; and the applicable characterization attributes.

It is worth to remark that the flow of information among activities is the same as fig 2. However, as we mentioned above, the information sources (the input of the Explore activity) in GBTCM are different, as shown in table 1. In table 2 we can realize that the output of each activity is the input of the next.

3.4 GBTCM Heuristics

One of the main contributions of GBRAM is the provision of heuristics and procedural guidance for identifying and constructing goals. Heuristics aid us by providing prescriptive guidance for managing varying levels of detail in the information available. There are four general types of heuristics used in GBRAM: identification, classification, refinement, and elaboration heuristics. Some of them are straightforward and generic, not require employing a specific inquiry technique. Others make sense only in conjunction with specific questions about the system.

Many heuristics showed in GBRAM can be mapped directly to GBTCM (section 4.2 and 4.3 show examples of applying identification heuristics –questions- that guide the obtention of specific information), but many others should be adjusted, and also some new heuristics for the specific domain can be created, which should be documented for handling the growing and evolution of the taxonomy. Applying GBTCM we can achieve a high probability of success finding the characterization attributes in a more formal way while avoiding wasted efforts.

Table 2. Activities and its inputs of GBTCM

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

4 Case Study: A Taxonomy in the Requirements Engineering Area

Our purpose in this section is to apply GBTCM to analyze the field of software Requirements Engineering (RE) and as a result propose a taxonomy in that area. We have chosen RE as case study because it is a critical area in the software development processes [11]. Therefore, to improve the efficiency of the activities performed in the area, COTS technology aid RE-related actors to simplify and facilitate their work. For keeping the description short, we focus on the most representative parts of the experience.

4.1 Sources of Information

As it can be expected for a topic such this, lots on information sources exist and many of them were gathered. Table 3 lists the sources of information more widely used for the construction of the taxonomy [12..23].

4.2 Identifying Goals and Objectives

Although we have many and diverse information sources, it should be considered as a good practice to base the process on the most solid and confident of them for extracting the main high level goals in order to assure the consistency of the set of goals, and then extracting subgoals 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 4. In subsection 4.4 we will use other information sources to decompose these high level goals.

Table 3. Main Sources of information used in the taxonomies for RE

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

4.3 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 4): 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. The relationships among these stakeholders appear in the i^* model mentioned in subsection 3.2.

Table 4. Some goals obtained from SWEBOK

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 5. An excerpt of organization of goals

Goals	Tools	Category of INCOSE Taxonomy
G2:Requirements Elicitation Performed		
G2.1:Requirements Sources Defined and Analyzed		
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

4.4 Organization and Matching of Goals

Once we had analyzed and identified goals from all information sources, we have to organize that information firstly by means of a matching of goals from all information sources, and subsequently according to precedence relationships. We represent the process of organization of goals by means of tables. Table 5 is an excerpt of this process. We can observe the matching of the goals (collected in the mentioned i^* SD model) with the existent taxonomies and vendors information.

The level of decomposition of goals is not defined in GB RAM. In GBTCM, it clearly depends on the matching of the information sources. As an example of rule of thumb in our context, one goal should be taken into account only if it exists 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). At the end of the process of matching we have a more complete set of goals. Next step we have to specify dependencies.

Specifying Goal Dependencies. In GBRAM, goals are organized only according to their temporal precedence relations. Adequate questions helps in the prerequisite findings and facilitate their organization, for example: “do any goal depend on the availability of this goal for achievement?” In GBTCM, we consider this aspect as twofold: we not only specify the temporal precedence relations of goals but also we define which are the dependencies among goals (both goals from the addressed COTS market segment or from other previously treated) for their completion, relying on another goal, agent, or resource; it means the explicit representation of potential dependencies among COTS domains by means *i** models cited in subsection 3.2.

For example, in fig. 3 we can see that besides *i** diagrams (left), we represent the precedence dependencies of goals by means of hierarchical tables (right). GBRAM refers to this last outlining mechanism as goal topography. As a result, we have a dual representation of goal models, a more technical one and a more understandable one, easy to index and read.

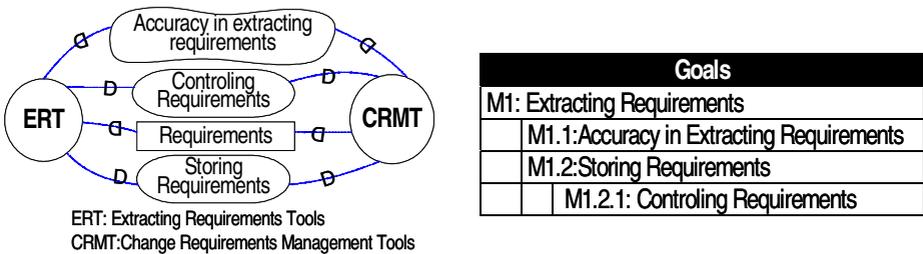


Fig. 3. An excerpt of a *i** SD model (left) and a hierarchical table (right) involving the RE tools

4.5 Reducing the Size of the Goal Set and Elaborating Scenarios

The goal topography must be refined, which implies the pruning of the goal set. Three approaches were used: eliminating duplicate goals, refining goals based on system entities, and consolidating nearly synonymous goals. The use of glossaries and class diagrams supports the reconciliation of goals. For example, the terms “capturing” and “extracting” that coming from two different sources was unified and defined as “extracting” in our glossary.

In some cases it was necessary to elaborate scenarios. Scenarios facilitate the identification of special or extraordinary circumstances which occur so that goal and requirements information should be refined. Scenarios are identified considering the goals by asking “Why?” and “Why this goal could be not achieved?” Scenarios were very useful for uncovering and reconciling goals, checking for completeness and conflicts, and communicating with stakeholders.

Table 6. An excerpt of the taxonomy for Requirements Engineering Tools

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

4.6 Operationalizing Goals

Goal information must ultimately be operationalized (related with actions). This is done by consolidating the goal information applying the Inquiry-Cycle [24] approach. The Inquiry-Cycle consists of a series of questions and answers designed to pinpoint where and when information needs arise. We apply this mechanism departing from the hierarchical set of goals in order to: get the questions-answers pair attained to each characterization attribute and appropriate organize the information resulting. Table 6 shows an excerpt.

5 Conclusions

In this paper we have described a goal-based method for the construction of COTS taxonomies called GBTCM. This approach allows the identification and elaboration of goals in a specific area and the matching and refinement of those goals into characterization attributes of the COTS field. This leads to several advantages [25]: the use of adequate information sources to obtain characterization attributes permits to browse the taxonomy in a guided way, more practical and confident; the explicit construction of i^* models as artifacts of the method representing the relationships among domains, making clearer the implications of the use a particular component as it was proposed in [3]; the organization of the information and artifacts and models resulting of the method, supports not only the reusing of knowledge gained in the specific area but also the maintenance and evolution of the COTS taxonomy. The use of GBTCM can help software engineers which usually carry out COTS selection to structure better their knowledge and may aim at a better return of investment.

Our future work will concentrate on using GBTCM in other domains inside the software development processes, and in taking into account the qualities and factors of the information sources for classify them according their relevance.

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. *Lecture Notes in Computer Science Vol. 2959*, (2004). *Proceedings 3rd International Conference on COTS-Based Software Systems (ICCBSS)*
4. Morisio, M., Torchiano, M.: Definition and Classification of COTS: A Proposal. In *Proceedings 1st. International Conference on COTS-Based Software Systems (ICCBSS)*, Orlando Florida (2002)
5. Mylopoulos, J., Chung, L., Yu, E.: From Object-Oriented to Goal-Oriented Requirements Analysis. *Communications ACM* 42(1), January 1999
6. Van Lamsweerde, A.: Goal-Oriented Requirements Engineering: A Guided Tour. In *Proceedings 5th IEEE International Symposium on Requirements Engineering (ISRE) 2001*
7. Antón, A.I.: Goal-Based Requirements Analysis. In *Proceedings 2nd IEEE International Conference on Requirements Engineering (ICRE) 1996*
8. UML 2.0 Specifications <http://www.uml.org/>, last accessed July 2004
9. Carvallo, J.P., Franch, X., Quer, C.: Defining a Quality Model for Mail Servers. In *Proceedings 2nd International Conference on COTS-Based Software Systems (ICCBSS)*. *Lecture Notes in Computer Science* 2580, (2003)
10. Yu, E.: Towards Modeling and Reasoning Support for Early-Phase Requirements Engineering. In *Proceedings 3rd IEEE International Symposium on Requirements Engineering (ISRE) 1997*
11. Standish Group Report <http://www.standishgroup.com>. CHAOS Survey 1994
12. Guide to the Software Engineering Body of Knowledge, SWEBOK, www.swebok.org
13. The Gartner Group, available on-line at <http://www4.gartner.com>
14. INCOSE. "Software Engineering Tools Taxonomy" <http://www.incose.org>
15. eCots. Software Components Open Directory Project. <http://ecots.org>
16. International Standard ISO/IEC 12207 Software Life Cycle Processes. 1995
17. IEEE/EIA Guide. IEEE/EIA12207.1-1997. Standard for Information Technology-Software Life Cycle Processes, April 1998
18. International Standard IEEE Std 830-1998 IEEE Recommended Practice for Software Requirements Specifications, June 1998
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. 16th January, 2004. <http://www.gartner.com>
21. CBSE Net. "Application Domain Taxonomy". Available on-line (previous registration) at: http://www.cbsetnet.org/pls/CBSEnet/eco_ricerca_documenti.concept_search_frame
22. International Data Corporation <http://www.idc.com>
23. ComponentSource <http://www.componentsource.com>
24. Potts, C., Takahashi, K., Antón, A.: "Inquiry-Based Requirements Analysis". *IEEE Software*, 11(2) March 1994
25. Ayala, C.P., Botella, P., Franch, X.: Goal-Based Reasoned Construction of Taxonomies for the Selection of COTS Products. In *Proceedings 8th Multi-Conference on Systemics, Cybernetics and Informatics (SCI)*. Orlando Florida (2004). ISBN:980-6560-13-2