Systematic Construction of Goal-Oriented COTS Taxonomies

Claudia Ayala

Technical University of Catalunya
Campus Nord 08034, Barcelona, Spain
cayala@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 overcoming 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. The main goal of this thesis is to provide support for a reliable and comprehensive structuring of a reuse infrastructure for the COTS components marketplace. With this aim, we propose a method for the construction and maintenance of goal-oriented COTS taxonomies based on a solid domain analysis; guiding the gathering of sources of information, modeling requirements and dependency relationship among domains, and organizing knowledge in any segment of the COTS marketplace.

1. Introduction

The use of COTS components for building systems is becoming a strategic need in a wide variety of application areas. In general, COTS are software components that provide a specific functionality, available in the market to be purchased, interfaced, and integrated into other software systems [1]. The potential benefits of this technology are mainly its reduced costs and shorter development time, while maintaining the quality. Nevertheless, many challenges must be faced for adapting the traditional software engineering activities in order to exploit these benefits [2].

One of the most critical activities in COTS-based development is the selection of the components that must be integrated therein. Selection is basically composed of two main activities, namely searching of candidates and their evaluation with respect to system requirements. However, while many approaches exist for driving evaluation [3] (e.g. CAP, CARE, CEP, CRE, OTSO, PECA, PORE, QESTA, STACE, and Storyboard), it is not the case of the searching phase, making the selection process expensive and inefficient. This lack of 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.

2. Problem Statement

One of the essential problems in reusing software components is locating and retrieving them from a large collection [4]. COTS are a special class of reusable components, this gives raise to a new problem: how to structure the knowledge about this COTS marketplace and in particular how to know which types of COTS are available and their objectives in such a way that searching the marketplace becomes a feasible task? Furthermore, COTS search must cope with some critical characteristics:

1. *The growing size of the marketplace*: New and improved products and technologies are continuously offered.
2. *The rapid changes in the marketplace*: New versions of existing products are released every few months and many times they offer services that initially were seen as belonging to different segments (e.g. current mail servers systems usually provide instant messaging facilities)
3. *Dependencies among COTS*: COTS are not designed to operate isolated [5], but in collaboration with others, which results in many dependencies among them.
4. *Type of descriptions available for COTS*: COTS suppliers do not provide the kind of structured information that would allow performing automated or at least assisted search. The situation is aggravated by the fact that they tend to highlight strengths and hide weaknesses of the licensed components.

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

In this context, 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.

3. Related Work

Due to the highly applicable nature of the subject of our research, we find related work not only concerning cientific proposals but also in the way that the COTS marketplace is really organized nowadays. Thus, profit and non-profit organizations have defined categories of services, products, and knowledge, usually structured in a hierarchical form. This type of organizations range from IT consultant companies such as Gartner [6] or Forrester [7], commercial web-based companies as ComponentSource [8], professional societies as INCOSE [9]; to the academic world in which organizations, teams and individuals have presented their own proposals that range from specific of one domain [10] to a wide range [4], [11] or even a field [12]. Moreover, it is well-known 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 [13-16]. However, these categorization proposals present some drawbacks, mainly

because 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. In [17] a survey of different approaches is given and also emphasizes some of these problems.

On the other hand, some proposals have arised for finding and identifying COTS components, such as [18], [19] and [20] which propose the use of automatic or semi-automatic search engines to identify COTS information in the web relying on component catalogues supplied by related companies. However, besides the mentioned problems of these kinds of catalogues, though these approaches achieve a high level of efficiency and low cost, they do no achieve the same level of accuracy and reliability of the acquired information as manual approaches, increasing many risks involved in COTS selection. Therefore, as other recent studies have concluded [21], we claim that there is a gap among COTS selection methods and their real application.

4. Proposed Solution and Research Method

For solving the stated problem we specifically propose a prescriptive goal-oriented method for building and maintaining a reliable reuse infrastructure that may be used in COTS search processes by arranging marketplace segments as a taxonomy. Our proposal relies on a four iteration research process supported by industrial experiences and academic cases studies we have been undertaken under action-research premises [22], as well as literature survey and grounded theory [23].

The initial iteration departed from an industrial experience 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 section 2. As a result, we studied the state of the art and evaluated the applicability of suitable methodologies for handling the problem, as related in [24].

In the second iteration, we consolidated previous results, proposing the use of goal-oriented approaches as a basement of our solution. A goal is an objective that should be achieved and may be formulated at different levels of abstraction [25]. We found that goal characteristics are helpful to build abstract, well-founded, and stable taxonomies capable of dealing with the rapid evolution of the COTS market and manage all the information related in a COTS selection process [26]. Subsequently, performing some formative case studies related with Requirements Engineering Tools, Software Application Development Tools and some sub-categories of Enterprise Applications (with emphasis with those related to Content Management), a preliminary version of our method called GBTCM (Goal-Based Taxonomy Construction Method) [27] was proposed; it was inspired by GBRAM (Goal-Based Requirements Analysis Method) [28] (a widely recognized methodology in the context of Requirements Engineering) as a framework for integrating the goal-oriented mechanisms and techniques needed for formalizing the construction of taxonomies from goals. In this stage, also the goal-modeling dependencies strategies were studied [29], adapted, and included in

the method [30]. The third iteration was focused on deeper analysis and preliminary validation of the results obtained in the previous one by means of the cases related with the Real-Time Synchronous Communication and Message-based Communication Systems, identifying and overcoming some method design flaws, resulting in an improved version of our preliminary method, called GOThIC (**G**oal-**O**riented **T**axonomy and reuse **I**nfrastructure **C**onstruction) [31].

The ongoing last iteration aims at achieving the complete validation of our approach, its tool implementation, and thesis writing.

Summarizing, our GOThIC method consist in 7 high-level activities stated as:

1. **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., by means of a knowledge acquisition program which allows extracting knowledge from the domain by reconciling the characteristics of the available sources with those of the taxonomy construction process.
2. **COTS marketplace domain analysis**. Its goal is to identify the basic elements of the domain from the information sources obtained previously, organize an understanding of relationships among these elements, and represent this understanding in a useful way.
3. **Identification, refinement, and statement of goals**. These activities are iterative and have the next objectives: *Identification* aims at extracting goals from available sources applying different goal-acquisition techniques. *Refinement* entails the goal refinement considering obstacles, scenarios to uncover hidden goals and mechanisms to discover synonymous or duplicated goals. *Statement* consists on expressing the goals in a systematic way. We use a pre/post style for specifying these goals, i.e. stating which conditions are met when others hold.
4. **Establishment of dependencies.** We have identified that a COTS component may need another for: *enabling its functionality* (e.g. in order to follow document lifecycles, 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* (e.g. a web page editing tool can complement a web browser for editing web pages); *enhancing its quality attributes* (e.g. resource utilization can be improved significantly using compression tools). Relationships are gradually identified analyzing the goal information obtained in previous activities. These relationships are declared as dependencies using goal-oriented models, specifically *i** models [32].
5. **Goal-taxonomy structuring**. The organization of goals comes from the analysis of pre and postconditions 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, thus 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.

6. **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. Using decision trees properties we defined the process of taxonomy validation as the repeated application of some stated transformation rules over the nodes to manipulate the hierarchy until reaching a stop condition. We detailed this process in [33].
7. **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 [21]. GOThIC provides an efficient mechanism to maintain a repository of the obtained knowledge. UML class diagram in Fig. 1 defines the form that this repository exhibits -see [31] for details-. 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.

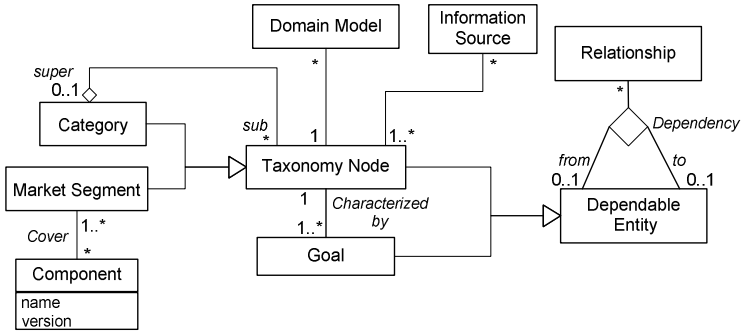

Fig. 1. Conceptual model for goal-oriented COTS taxonomies: overview.

Although presented as sequential for clarity, the method activities may in fact be intertwined and iterated as required to obtain the target infrastructure incrementally. Furthermore, GOThIC 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). Its ultimate goal is to populate a knowledge base with data according to the UML conceptual model sketched in Fig. 1. 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 and their relationships declared as dependencies. 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).

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 functionality 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 differ-

ent 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. Further method description, is available at [31].

5. Validation and Future Work

As we mentioned above, preliminary validation of each research stage was performed by means of formative case studies [26],[27],[31],[33],[34]. The results are promising from the academic point of view, therefore, our main purpose for future work is focused on industrial validation of the final proposal involving an industrial partner, by means of some action-research collaboration, based on the principles of [35] for assessing the usability and effectiveness of the GOThIC method; as well as the development of tool support starting from the UML conceptual model sketched in Fig. 1.

6. Applicability of the Proposal

Basically it 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 needs to select COTS components from these segments (i.e. communication infrastructure, ERP systems, security-related systems, etc); hence, 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. Thus, having knowledge available and classifiers to know when a market segment is of interest is a great help, 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.

7. Contributions and Target Audience

Our method is aimed at face COTS components search, based on the notion of goal for building abstract, well-founded and stable taxonomies, which may evolve as the marketplace does. It is a way to overcome the COTS marketplace characteristics:

- *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 [25]. 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 activities 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).

As a result, diverse actors may benefit from our approach: a) IT consultant companies offering assessment for business automation may structure their services better; b) Commercial web-based companies or portals may structure their offering in well-founded categories with a clear rationale behind; c) Medium- and large-size companies with their own IT department may be more confident on their own selection processes; d) Software engineers which usually carry out COTS components selection may structure better their knowledge and may aim at a better return of investment.

References

1. Meyers, C., Oberndorf, P. *Managing Software Acquisition*. Addison-Wesley 2001.
2. Hasselbring, W. "Component-Based Software Engineering" In collection of International Journal of Software Engineering and Knowledge Engineering, 2002.
3. Ruhe, G. "Intelligent Support for Selection of COTS Products" LNCS 2593, 2003.
4. Prieto-Díaz, R., Freeman, P. *Classifying Software for Reusability*. IEEE Software. 01/1987.
5. X. Franch, N., Maiden. "Modelling Component Dependencies to Inform their Selection" 2nd International Conference on COTS-Based Software Systems (ICCBSS), 2003.
6. Gartner Inc. www.gartner.com, 2006.
7. Forrester Research Inc. www.forrester.com, 2006.
8. ComponentSource www.componentsource.com, 2006.

9. INCOSE www.incose.org
10. E. Arranga. *Cobol Tools: Overview and Taxonomy*. IEEE Software, 17(2):59-61, 2000.
11. D. Carney, F. Long. *What Do You Mean by COTS? Finally a Useful Answer*. IEEE Software, 17 (2): 83-86, March/April 2000.
12. Glass, R., Vessey, I. *Contemporary Application-Domain Taxonomies*. IEEE Software. July 1995
13. eCOTS project: www.industrie.gouv.fr/rntl/FichesA/E-Cots.html
14. Jaccheri, L., Torchiano, M. "Classifying COTS Products". Proceedings 7th European Conference of Software Quality (ECSQ). Finland, 2002.
15. Morisio, M., Torchiano, M. "Definition and Classification of COTS: A Proposal". Proceedings 1st Conference on COTS-Based Software Systems (ICCBSS). 2002.
16. Erofeev, S., DiGiacomo, P. "Usage of Dynamic Decision Models as an Agile Approach to COTS Taxonomies Construction". ICCBSS 2006. IEEE Society 2006.
17. Cechich, A., Réquilé-Romanczuk, A., Aguirre, X., Luzuriaga, J. "Trends on COTS Component Identification and Retrieval". ICCBSS 2006. IEEE Society 2006.
18. C. Seacord, A. Hissam, K. Wallnau. "Agora: A Search Engine for Software Components" Internet Computing, Vol. 2 No. 6, December 1998, pp. 62-70
19. N. Yanes, et al. " MoReCOTS: a Specialized Search Engine for COTS Components on the Web". International Conference on COTS-Based Software Systems, IEEE Society 2006.
20. M. Sjachyn, L. Beus-Dukic. "Semantic Component Selection–SemaCS". International Conference on COTS-Based Software Systems, IEEE Society 2006.
21. Réquilé-Romanczuk, et al. "Towards a Knowledge-Based Framework for COTS Component Identification". Proceedings ICSE-MPEC 05. ACM 2005.
22. Baskerville R., Wood-Harper A.T. "Diversity in Information Systems Action Research Methods". European Journal on Information Systems. Vol. 7 No.2, June 1998.
23. Martin, P., Turner, B. "Grounded Theory and Organizational Research". Journal of Applied Behavioral Science.1986; 22: 141-157.
24. Ayala, C., Botella, P., Franch, X. "Goal-Based Reasoned Construction of Taxonomies for the Selection of COTS Products". SCI 2004, ISBN:980-6560-13-2. July 2004.
25. A. Lamswerdee. "Goal-Oriented Requirements Engineering: A Guided Tour". 5th IEEE International Symposium on Requirements Engineering. 2001, pp. 249-263.
26. Ayala, C.P., et al. "Goal-Based Reasoning in the Construction of Taxonomies for COTS Components" LSI-Department, Report:LSI-05-58-R.. http://www.lsi.upc.edu/dept/techreps/
27. Ayala, C., Botella, P., Franch, X. "On Goal-Oriented COTS Taxonomies Construction". 4th ICCBSS. LNCS 3412, Bilbao, Spain; February 2005.
28. Antón, A.I. "Goal Identification and Refinement in the Specification of Software-Based Information Systems". Ph.D. thesis, Georgia Institute of Technology, June 1997.
29. Ayala, C. et al. "A Comparative Analysis of i*-Based Goal-Oriented Modeling Languages". AOSDM, Taipei, Taiwán; China. July 2005.
30. Grau, G., et al. "RiSD: A Methodology for Building i* Strategic Dependency Models". SEKE 2005. Taipei, Taiwan; China. July 2005.
31. Ayala, C., Franch, X. "A Goal-Oriented Strategy for Supporting Commercial Off-The-Shelf Components Selection" 9th International Conference on Software Reuse (ICSR) 2006.
32. Yu, E. "Modelling Strategic Relationships for Process Reengineering" PhD Thesis, University of Toronto, 1995.
33. Ayala, C., Franch, X. "Transforming Software Package Classification Hierarchies into Goal-Based Taxonomies" 16th DEXA Conference. LNCS 3588, August 2005.
34. Ayala, C., et al. "Construction of a Taxonomy for Requirements Engineering COTS Components" Journal of Computer Science and Technology. Vol.5–No.2– July 2005.
35. Wohlin, C., et al. "Experimentation in Software Engineering, an Introduction" ISBN 0-7923-8682-5.