A Systematic Approach to Manage Information Quality for Supporting Software Package Selection

Claudia Ayala, Xavier Franch

Technical University of Catalunya
UPC-Campus Nord (Omega), 08034 Barcelona, Spain
{cayala,franch}@lsi.upc.edu

Abstract. Software package selection plays a crucial role in the deployment of software systems. One of its main current problems is how to deal with the vast amount of unstructured, incomplete, evolvable and widespread information that highly increases the risks of taking a wrong decision. It could be said that nowadays there are not satisfactory solutions to that problem. In this paper, we propose an approach to systematically tackle these information quality problems by stating a reference model embracing quality indicators that facilitate the collection, storage, retrieval, analysis and reuse of information in a quality assurance environment. We illustrate it with a scenario of use that shows how this reference model may be used to support software package selection decision-making.

1. Introduction

Nowadays, the amount of information available about software packages (SP) is vast and still growing. To select SP (i.e., the process of deciding the appropriate SP to be integrated in a component-based system [1]), decision-makers have to face not only the current diversity of SPs types available in the marketplace, but also the great deal of widespread, heterogeneous, and unstructured information describing each of them [2], [3]. The quality of this information largely determines the quality of the decisions made, and ultimately affects the quality of the whole software system and its development [3], [4]. Since SP selectors must rely on the information for their decision-making processes, ensuring Information Quality (IQ) is a critical success factor.

Over time, librarians and other information professionals have developed a set of criteria to be used to evaluate IQ based on careful experts' examination (e.g., authority, format, scope, etc.) [5]; however, these criteria are too general and do not provide much guidance to the particular problem of SP selection. Therefore, more specialized approaches are needed. Some recent approaches propose the use of automatic or semiautomatic search engines to identify SP, e.g., [6],[7],[8]; but, to the best of our knowledge, they have not reached a generalized consensus on their utility for the community and do not provide a systematic approach for managing IQ. Therefore, IQ is still a critical open issue from the SP selection perspective [2], [3], [4],[9].

The goal of our research is twofold. On the one hand, to develop a comprehensive framework that states those IQ aspects that are important to perform an informed SP selection, and how these aspects are feasibly gauged using a quality model schema that hierarchically describes the IQ aspects, and provides metrics to assess the value of the information used in a specific SP selection project. On the other hand, based on this framework, we also offer a tool-supported conceptual model to add capabilities for managing and reusing IQ in diverse SP selection processes.

The paper is structured as follows. In section 2 we provide a brief background of previous research that greatly justifies the need of this study. Section 3 details the processes, methods and techniques used to capture IQ needs in the context of SP selectors and their mapping to metrics. Section 4 encloses the obtained results related in the previous sections in a conceptual model that is used as a reference for systematically support SP selectors decision-making. An illustrative scenario of use is presented in section 5, whilst section 6 provides conclusions and future work.

2. Background

As a response to the SP selection needs, we formulated the GOThIC method [10]. The method is intended to guide the construction of a SP reuse infrastructure (repository) that provides well-founded and understandable goal-oriented taxonomies that describe the contents of the SP marketplace. Its main goal is to populate a knowledge base with data according to the conceptual model sketched in Fig.1 (as it is further explained in [10]). The taxonomy built with GOThIC may then be browsed during SP selection processes to locate the market segment (or segments) of interest. The DesCOTS system (Description, evaluation, and selection of COTS packages) [11] can be used to support the method.

The method is articulated by means of several activities. One of the most relevant is the activity of domain analysis of the SP marketplace segment being addressed by a SP selection project. It has the objective of capturing and representing the most important aspects of the particular SP market segment (e.g., the segment of CRM systems; they are represented by a *Taxonomy Node*) using an integrated *Domain model* based on the ISO-9126 software quality standard [12]. Once found, the *Domain model* may be used to obtain the appropriate criteria for selecting the most suitable SP. The details of the domain analysis activity were presented in [12].

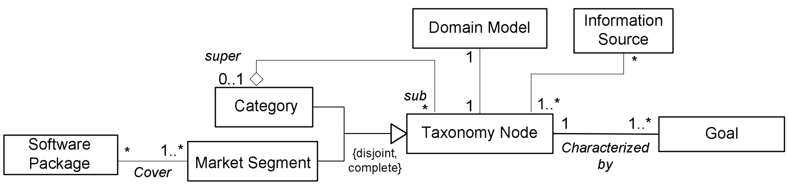

Fig. 1. Conceptual model for goal-oriented SP taxonomies in the GOThIC method: overview

It is clear that the success of domain analysis is directly related to the quality of the information used to perform this activity. The industrial validation of GOThIC shown the critical difficulty reported by software engineers to collect, process and analyze the vast amount of information sources for performing a reliable domain analysis and therefore a trustworthy decision-making. While many approaches exist to state quality characteristics and requirements for improving SP selection processes, and the quality of the SP documentation has been considered as a crucial quality aspect affecting their usability [13], the issue of where and how to get trustworthy information about them in an efficient manner has been left out [14]. Consequently, it is causing several over-costing problems or even abandoned projects because of wrong decisions based on untruthful information [15]. As a result, in order to overcome the risks associated to poor quality information [16], we realized the need of integrating an IQ management strategy to our GOThIC method to perform an efficient and proactive SP market segments domain analysis. In this sense, our current work is based on the emerging IQ research for supporting trustworthy SP selection decision-making.

3. Capturing Information Quality Dimensions for SP Selection

The work presented in this paper is based on relevant approaches from the IQ research, e.g., [5],[16],[17],[18],[19], several industrial experiences and case studies analyzing SP selection processes [20], and interviews run in software companies [21]. The industrial experiences were undertaken under action research premises and grounded theory to iteratively identify the IQ problems in the SP selection setting, trying out suitable IQ research approaches to resolve them, adapting and evaluating how successful are such strategies in practice until a satisfactory solution comes out. The interviews were used to conduct an explorative survey in 7 Norwegian software companies [21]. It consisted of a semi-structured interview conducted to managers, software architects and developers involved in SP selection projects. Our first goal was to collect information about the problems they face to get SP related information, followed by inquiring what they mean by IQ to perform an informed selection. The results are detailed below.

3.1 Identifying SP Selection Information Problems

To find out which are the most relevant dilemmas that companies face whilst processing information during SP selection, we asked interviewees about the resources they usually used to locate SP and/or information about them, as well as the perceived utility of such information for performing the different SP selection activities. Summarizing the answers, in Table 1, we provide a list of existent resources (completed with our own expertise as SP selectors we also are), their main practical use, characterization mechanisms, retrieval schema, information rendering, and some additional information they offer.

Table 1. Some existing resources to support SP selection

Information Source	Characterization	Retrieval Schema	Information Rendering	Additional Information
eCots.org	Software Categories (SC)	Browsing (B), Keyword Search (KS)	Non-Structured	Newsletter
SourceForge.net	SC	B, KS	Non-Structured	Newsletter
ComponentSource.com	SC Platforms	B, KS	Non-Structured	Newsletter, Forums, Demos
KnowledgeStorm.com	IT Solutions (ITS)	B, KS	Non-Structured	Case Studies, Forecasts Vendor white papers, Demos.
Forrester.com	ITS	B, KS	Non-Structured	IT Support, Not free
Gartner.com	ITS	B, KS	Non-Structured	IT Support, Not free
Incose.org	SC	B, KS	Semi-Structured	Research Trends
Google.com/codesearch	-	KS	-	Search public source code in the web
Tucows.com	Platforms	B, KS	Non-Structured	Freeware/Shareware Forums
Software Package Providers	Name	B, KS	Non-Structured	Requests, Forums, Documentation
CMSmatrix.org	Name	List Selection	Semi-Structured	Comparison table
Internet Mail Consortium	-	-	Non-Structured	Standards and Trends
CeBASE	-	B, KS	Non-Structured	Lessons Learned and Useful Links

Since there is a great variety of types of information sources available, we decided to group them for assessing their characteristics and actual IQ problems. Table 2 shows an excerpt of this grouping and some representative examples. Interviewees agreed that extracting SP information from these resources is a critical process because they do not have control over their availability, accessibility, heterogeneity, impartiality, incompatibilities, inconsistencies and mistakes that make difficult to guarantee IQ and lead to failures that cost dearly. This could be the reason why there was not consensus of the utility of these resources in the SP selection community. Furthermore, we found that in order to reach project constraints (mainly in terms of time and resources), actual decision-making processes for finding and/or processing SP information are rarely documented and usually based on vague factors as own experiences and intuition [22], [23]; even in the cases of organizations that periodically performed SP selection projects. This fact increases the risks of damaging the whole software development process and continuously loose tacit knowledge when more experienced people are replaced. This justifies further research on this topic.

Table 2. Information Sources Types

Types of Source	Description	Current Problem	Examples
Existing Hierarchies/ Taxonomies	SP categorizations that offer descriptions of diverse SP with different objectives.	• Proliferation of cataloguing initiatives that contain only brief and unstructured descriptions of some inventories packages. • Questionable trustworthiness of the provided information • Different descriptions of the same package and confusion in using the categorization	ComponentSource.com KnowledgeStorm.com SourceForge.net
Vendor Information	Information provided by the SP supplier as its characteristics, documentation and comparatives with previous or existent versions	• Incomplete and unstructured information. • Highlight the strengths and hide the weaknesses of their licensed packages. • Questionable veracity of the information.	Any Commercial Firm
Related Standards	Since there is no specific standardization concern to describe SP, some industrial organizations group the main vendors of particular domains maintain up-to-date information that sometimes could be considered as a reference.	• Lack of agreement among relevant manu-facturers of software packages. Therefore, is a lack of standardized concerns.	Internet Mail Consortium (IMC), Workflow Management Coalition (WfMC), Enterprise Content Management Association (AIIM)
Independent Reports	Third party organizations ranging from research to consultant often provide support for selecting SP	• Their analysis is often short-lived, ad-hoc and in some cases expensive.	Scientific: Specialized Journals Divulgation: Specialized Websites Technical: Gartner, Forrester, etc.
Experiences on the Field	Knowledge or practical information usually provided by experts or domain stakeholders that relate own experiences or lessons learned.	• It is difficult to get relevant information of this kind that really fit into our specific case, because reported experiences depend on the specific context of the author.	Forums, Talks, Luncheons, Tips, CeBASE repository,
Test of Tools and Systems	Test descriptions of tools which have been really used. They allow envisaging the real behavior of a SP in a specific environment	• Dependent to the context of the specific project and its particular requirements to test.	Test results Tucows.com CMSmatrix.org
Others	Provide some specific functionality to find expected SP functionalities. They can range from specialized searching tools to open source code detection tools	• They find wide sets of information (often not relevant) that must be also analyzed.	Koders Tool Google.com/codesearch

3.2 Determining IQ in the SP Selection Context

IQ researchers agree on the meaning of high-quality information as information which is fit for use by consumers [18]. Therefore, we considered crucial to collect information from the SP selectors about their IQ needs for performing an informed selection instead of being defined theoretically or based on researcher's experience. The analysis of the outcomes was based on the framework presented in [17] and refined in [18]. It helps to determine a basis for assessing IQ from the information consumers' perspective, and suggests to group IQ needs into 4 high-level IQ dimensions which are described by a set of quality assets that represent a single aspect or construct of IQ, as shown in Table 3. Moreover, we use the notion of information product [19] (i.e., information is treated as a product). Further evidence exist that these approaches provide comprehensive coverage of the multi-dimensional IQ construct in very diverse settings [16],[18],[19].

Table 3. Basic IQ dimensions to describe IQ in diverse settings suggested by [17].

IQ Dimension	Definition	Quality Assets
Intrinsic	Denotes that information has quality in their own right	Believability, Accuracy, Objectivity and Reputation
Representational	Includes aspects related to the format and meaning of the data	Concise representation, representational consistency, Interpretability, and Easy of understanding
Accessibility	Emphasizes the importance of the role of systems for providing access to information in a secure setting	Accessibility, and Access Security
Contextual	Highlights the requirement that IQ must be considered within the context of the task at hand	Relevancy, Timeliness, Completeness, and Appropriate Amount of Data, Value-added.

As a next step, we intended to fit the IQ needs elicited from SP selectors within the IQ dimensions stated in Table 3. Such results are shown in Table 4.

Of course, different points of view on IQ were obtained from different interviewees, but such differences were related to their specific projects requirements thus, they were successfully represented by the high-level dimensions. In addition, although in most cases decision-making processes were based on vague factors, interviewees recognized some important facts that they take into account to assess IQ, some of them are also presented in Table 4. To fit our findings into the stated dimensions, we carefully analyzed the elicited IQ needs of SP selectors to match them into the enclosed assets of each dimension; as a result, the assets were adapted, i.e. redefined, abstracted, deleted and carefully reviewed until agreement was reached. The most significant changes observed in Table 4 are:

- The *Access Security* asset belonging to *Accessibility* (Table 3) was deleted since we considered that SP selectors acted as users of the information and its security aspects were not relevant from their perspective.
- We replaced the name of the *Accessibility* asset by *Availability* in order to avoid misunderstandings with the name of the IQ dimension it belongs to.
- But more remarkably, a fifth dimension, the project dimension named *IQ Project Issues*, was added. This last relevant change appeared because our findings showed that IQ in the SP selection context is largely determined by the resources allocated to the software development project and related policies and procedures; therefore we need to take this into account.

Please note that there is a high synergy among the elicited SP selection IQ needs. Hence, many intuitive relationships become evident, for instance, some IQ needs are shared by different assets from different perspectives (e.g. the *Value-added* asset is closely related to the facts addressed by *Concise Representation* and *Representational Consistency* to denote the extent of the value added; *Reputation* enhances *Believability*; and *Accuracy* greatly depends on *Timeliness, …*).

Table 4. An excerpt of SP selection IQ needs and some facts elicited from SP selectors

IQ Dimension/Assets		SP Selection IQ Need	Some IQ Facts
Intrinsic		Own properties of the information source denoting its quality characteristics.	
1	Believability	Credible information	Due to the commercial nature of the market, one should be sure of the credibility of the information
2	Accuracy	Precise information	Due to the high volatility of the market, documents become obsolete very quickly
3	Objectivity	Impartial point of view	Commercial and sponsored resources tends to be highly partial
4	Reputation	Coming from good sources	Well-known authors/sources represent a high reliability and trustworthiness. Appropriate references, related resources and resource-dependencies are good indicators to recognize reputation.
Representational		IQ properties related to the product format	
1	Concise Representation	Highly Structured	SPs information tends to be unstructured and very difficult to be located, especially in the cases of quality of service and non-technical information.
2	Representational Consistency	Homogeneity	The lack of standards for documenting SPs results in many kinds of documentation. Such heterogeneity makes difficult to easily compare IQ.
3	Understandability	Easy to be understood	Information addressed to the general public is usually easy to understand, but the information addressed to experts requires a specific background to be understood.
4	Interpretability	Easy to be interpreted	Different kinds of representation exist (e.g., ER models, Natural Language). They should be easy to interpret by the skills and background of the involved people.
Accessibility		IQ properties describing product accessibility.	
1	Availability	Free availability preferred	Some informational resources or support are available only for a fee.
2	Easy of Operation	Easy way to retrieve the information.	The way of obtaining the data (e.g. direct download, subscription based, etc) and the resources required to process it should be compatible with the resources allocated to the project.
Contextual		Combined effect of the product quality with the project needs.	
1	Relevancy	Useful, and appropriate to the project needs	Easy for the project team to understood and process the information with the resources they have allocated)
2	Timeliness	Sufficiently current and up-to-date	The date of creation or update is the best indicator for that asset
3	Completeness	Covering all the informational project needs	Information should cover the scope of the project to assure a good understanding of the requirements.
4	Appropriate Amount of Data	Adequate volume of information to be analyzed by the resources available.	Depend on the skills of the people allocated to the project and the size of the source.
5	Value-Added	Add value to the project operations	Heterogeneous and Unstructured information is difficult to be processed. When its structure makes easier the work of the project team in any of their tasks, it is considered a value-added.
IQ Project Issues		Describes the main IQ needs of the SP selection project	
1	IQ Project needs	Describes the IQ needs of the SP selection project	The high-level project goals drive the IQ SP Selection processes, they states the criticality of the project and its IQ needs.
2	Allocated Resources	Aspects related to the set of resources allocated to the project for performing the IQ process	The resources allocated to the project play a crucial role in determining product suitability.

3.3 Determining a Measurable Framework for Assessing IQ in the SP Selection Context

Our next goal is to develop a comprehensive framework that states our IQ findings; and how they could be feasibly gauged. To manage and gauge all these different views on quality, quality models seem the most appropriate type of artefact since they provide a measurable framework which precisely defines and consolidates the different views of quality. Specifically, we propose an ISO/IEC 9126 tree-like structure [24] because: i) ISO/IEC 9126 quality standard is one of the most, if not the most, widespread quality standard available in the software engineering community, therefore, most SP selectors are familiar with it; ii) It is compatible with the *domain model* from GOThIC [12], outlining an uniform framework well-suited for the integrated evaluation of all SP selection related issues; iii) It allows considering IQ aspects as requirements from the beginning of the SP selection process in the same way that we have technical and non-technical requirements; iv) It allows optimal reusability of product quality features throughout different SP selection processes.

The ISO-IEC 9126-1 tree-like structure is based on a hierarchical model that offers quality characteristics to represent the most important quality aspects. These characteristics are further refined into multiple levels of subcharacteristics, which in turn are decomposed into attributes, yielding to a multilevel hierarchy. Intermediate hierarchies of attributes may appear making thus the model highly structured. At the bottom of the hierarchy there are the measurable attributes, whose values are computed by using some metric.

In order to elaborate the quality model from the IQ dimensions presented in Table 4, we adopted one of the most widespread approaches, the Goal-Question-Metric (GQM) approach [25], for analyzing each asset belonging to the stated dimensions. Remarkably, this study was carried out without making any assumption about the assets other than it is built on top of the validated framework presented in Table 4.

In GQM, goals of the product under measurement are identified, and then some questions are raised to characterize the way the assessment of a specific goal is going to be performed. Last, a set of metrics is associated with every question in order to answer it. The final result of the GQM approach is a hierarchical structure in graph-like form, since metrics may influence in more than one question, and questions may be related to more than one goal. Goals are composed of four elements: purpose, issue, object and viewpoint. In our framework, these elements take the following form:

- Purpose. Presence of a particular feature or characteristic in the quality model
- Issue. The GQM recommends identifying quality goals; then, we define one issue for each asset stated in Table 4. As a consequence, we have as many goals as assets.
- Object. Always the dimensions to which the asset belongs to
- View Point: Always IQ needs for SP selectors

An excerpt of this process is shown in Table 5. To define the metrics, we have used the general theory of software measurement presented in [26] as conceptual basis to define the metrics for IQ aspects. Metrics can be as simple as Integer or Boolean values or more complex as Sets or Functions. An example of a predefined function is the *mean* function used in Table 5, which has the meaning of returning the mean of all values of the function it encloses (e.g. *AuthorsBel*, *AuthorBelMarks*, *ProvBelMarks*). Moreover, we have considered very important that external marks can be computed to determine some quality attributes; for instance, to determine the believability of an author by the marks that other people have stated about him/her, and also to determine the believability of these marks by the believability of the markers.

Our GQM analysis leaded us to observe that different kinds of IQ metrics were obtained. Some of them were objective properties inherent to the product whilst others were subjective or objective but external.

In addition some metrics have applicability preconditions (e.g., Marks available for the author, which precondition is the availability of some Marker). The use of external metrics was greatly influenced by the intended reusability we aimed for the quality model.

Table 5. An excerpt of the GQM approach to refine IQ dimensions

Goal	Question	Metric	Value	Kind of metric
Purpose: *Have an appropriate* Issue: *Believability* Object: *Intrinsic IQ* View Point: *SP Selectors*	Who is/are the author(s) of the product?	Author(s) Name	AName= Set (String)	Objective, it is part of the own product properties
	What is the believability of the Author(s)?	For each author, his/her believability	AuthorsBel= Function(String→TScore) TScore: {Very High, High, Low, Very Low}	Subjective, external property of the product
	What is the overall believability of all authors?	Average of all authors believability	OveAuthorsBel= Mean(AuthorsBel)	Subjective, external property of the product
	How is the author believability obtained?	Average of all marks that he/she has received	IndAuthorBel= Mean(AuthorBelMarks)	Subjective, external property of the product
	What are the marks that the author has received?	Marks available for the author (Marker Available)	AuthorBelMarks= Function(String →TScore)	Subjective, external property of the product
	What is the believability of those marks	For each Marker, his/her believability	BelMarks= Function(String → TScore)	Subjective, external property of the product
	What is the provider Organization?	Organization Name	OrgName=String	Objective, it is part of the own product properties
	What is the believability of the Organization?	Organization Believability	OrgBel= Mean(ProvBelMarks)	Subjective, external property of the product
	What are the marks that the provider has received?	Marks available for the author (Marker Available)	ProvBelMarks= Function(String →TScore)	Subjective, external property of the product
	What is the believability of those marks	For each marker, his/her believability	BelMarksOrg= Function(String → TScore)	Subjective, external property of the product
	…			

Subsequently, following some principles stated in [27] for building a good ISO-IEC 9126 tree-like quality model, and our results of applying the GQM approach to each asset, we obtained a highly reusable quality model, an excerpt of this model is shown in Table 6. The overall model has been structured according to the following guidelines:

- **Characteristics**. The five highest-level quality factors correspond to the 5 IQ dimensions obtained in the previous section: *Intrinsic IQ*, *Representational IQ*, *Accessibility IQ*, *Contextual IQ* and *IQ Project Issues*. Due to their nature, the first three characteristics group IQ features to describe the product and can be reused in all SP selection projects, whilst *Contextual IQ* is envisaged to record the extent to which the product features meets the *IQ Project Issues*. This structure allows an optimal degree of reusability not only of the product but also the knowledge gained in every use of it.
- **First-level subcharacteristics**. The first 5 characteristics have been decomposed into 17 subcharacteristics that correspond to the assets leveraged in Table 4.
- **Intermediate subcharacteristics**. More than 30 intermediate subcharacteristics were used since most of the 17 subcharacteristics stated above were still too abstract to be measurable, and more SP IQ needs were still remaining to completely express the requirements of our interviewees; thus, whenever it was required an additional level of subcharacteristics for structuring or levelling purposes was added. It was primarily based on the GQM application results. Subcharacteristics are used mainly for classification means. This is the case of the *Intrinsic IQ/Believability* subcharacteristic which has been decomposed into other subchar-

acteristics: *Author Based Believability, Provider Based Believability* and *Product Based Believability*.

- **Attributes**. Finally, subcharacteristics were further decomposed into over 75 IQ attributes. To decide which attributes were the most suitable for evaluation and reusability purposes, we choose the most representative ones obtained from the application of GQM to the five SP IQ dimensions´ assets. In general, they are two kinds of attributes: basic attributes which stand for objectively measurable features (e.g., *Author Name* attribute categorized under the *Intrinsic IQ/Believability* subcharacteristic); and derived attributes which require to be additionally decomposed into other attributes (e.g., *Author(s) Believability*). In order to measure the attributes of our quality model, metrics were required, so we attained the metrics previously obtained by the GQM application. It is important to stand out that we try to preserve homogeneity among the attributes metrics, mainly because as we mentioned above, it is common that quality features are closely related; and such homogeneity is the basis for a successfully *Contextual IQ* attributes estimation, since it is stated as the combination of the other IQ characteristics.

Table 6. Excerpt of the Believability Subcharacteristic decomposition

/Subcharacteristics/Attributes						Description	Metric	Comments
1	Believability							
	1	Author Based Believability				Aspects that describe the believability of the product based on its authors.		
		1	Author(s) Name			Describes the name of the author(s) of the product.	AName = Set (String)	$AName \neq \emptyset$ The names are directly obtained
		2	Author(s) Believability			Describes the overall authors believability by the average of the believability of all authors.	OveAuthorsBel= Mean(AuthorsBel)	Derived Attribute
			1	Individual Author Believability		Describes the individual believability of the authors	AuthorsBel= Function(String→ TScore)	1) Dom(AuthorsBel) = AName 2) $\forall$ s $\in$ AName: AuthorsBel (s)= Mean(AuthorBelMarks)
				1	Opinion Marks about the Author	Describes the marks that markers have done about the author	AuthorBelMarks = Function(String→ TScore)	These marks are directly obtained
		…						
	2	Provider Based Believability				Aspects that describe the believability of the product based on the organization that provides it		
		1	Provider Name			Describes the name of the product provider	OrgName=String	$OrgName \neq \emptyset$
		2	Organization Believability			Describes the believability of the organization provider	OrgBel= Function(String → TScore)	1) Dom(OrgBel) = OrgName 2) $\forall$ p $\in$ OrgName: OrgBel(p)= Mean(ProvBelMarks)
			1	Opinion Marks about the Organization		Describes the marks that markers have done about the organization	ProvBelMarks = Function(String→ TScore)	These marks are directly obtained.
		…						
	3	Product Based Believability				Aspects that describe the believability of the product based on their own properties		
		…						

Finally, it is evident that some IQ attributes may influence several other attributes or subcharacteristics (as it is shown in the last column of Table 6), and thus hierarchic overlapping is also supported in the approach, by considering the hierarchy as a graph. In addition, although all quality features are involved in determining the overall IQ for a SP selection project, elaborated types of relationships among quality features and also intensities of these relationships exist, and may be depicted by means of tabular representations as done in [28].

4. A Systematic Approach for Managing and Reusing Software Packages Information Sources

In the previous sections we have achieved our primary goal, namely to understand what quality means for SP selectors, and how it can be feasibly recorded, reused, gauged and integrated into our GOThIC approach. Moreover, we enriched our approach by organizing our further findings as a set of heuristics for supporting the collection, arrangement and decision-making processes, as explained in section 4.1. However, although the obtained IQ attributes resulted satisfactory enough to describe IQ requirements in several industrial and academic selection project case studies, we realized that supporting decision-making in the face of increasing information volumes and SP information characteristics, demands a systematic management of IQ to inform SP selectors about the quality and adequacy of the information to their task without having to do a full inspection or regenerating the data anew. Besides, the reuse of the information sources and their assessment in diverse selection processes of the same domain, would improve the efficiency and effectiveness of future selection processes. As a result, we implemented our previous work as a conceptual model for systematically supporting SP selectors not only to collect, storage and assess IQ, but also to (re)use and manage it for improving their decision-making process. It is detailed in section 4.2.

4.1 Heuristics to Support IQ Assessment in the SP Selection Context

Some examples of heuristics driven the information arrangement and decision making processes are:

- “Diverse types of information sources exists, they can be grouped into: Hierarchy, Standard, Vendor Information, Independent Reports (of scientific, divulgation and/or technical nature), Oral Information, Test Of Tools Reports, Experiences, Other”,… (Descriptions and examples are provided in Table 2).
- “Information sources available can provide insights into a diverse range of software packages and/or vendor characteristics, but no requirements identified from these sources should be used without careful consideration of their confidence”.
- “Information from experts is good at quickly identifying general principles, offering explanations, validating analyses, and providing pointers that could be cross-validating with Independent Reports”.
- “Information from SP providers tends to highlight the strengths and hide the weaknesses of licensing packages”.
- “Information from standards related to the field, are the best for identifying SP high-level goals”.

– “*Test of tools and Systems* are useful for complementing the information from vendors regarded to detailed information on typical interfaces, architectures, dependencies between products for enabling or complementing their functionality”.
– Etc.,

4.2 A Conceptual Model for Systematically Supporting SP Selectors Decision-Making

Results obtained in this study, have been enclosed into a conceptual model. A sketch of this model is shown in Fig 2. The ultimate goal of this model is to incrementally build a catalogue of SP related information sources, and describe them by means of the quality features defined in the quality model explained in the previous section. This leads to inform the SP selectors about the quality of the sources to decide if they are adequate to their objectives.

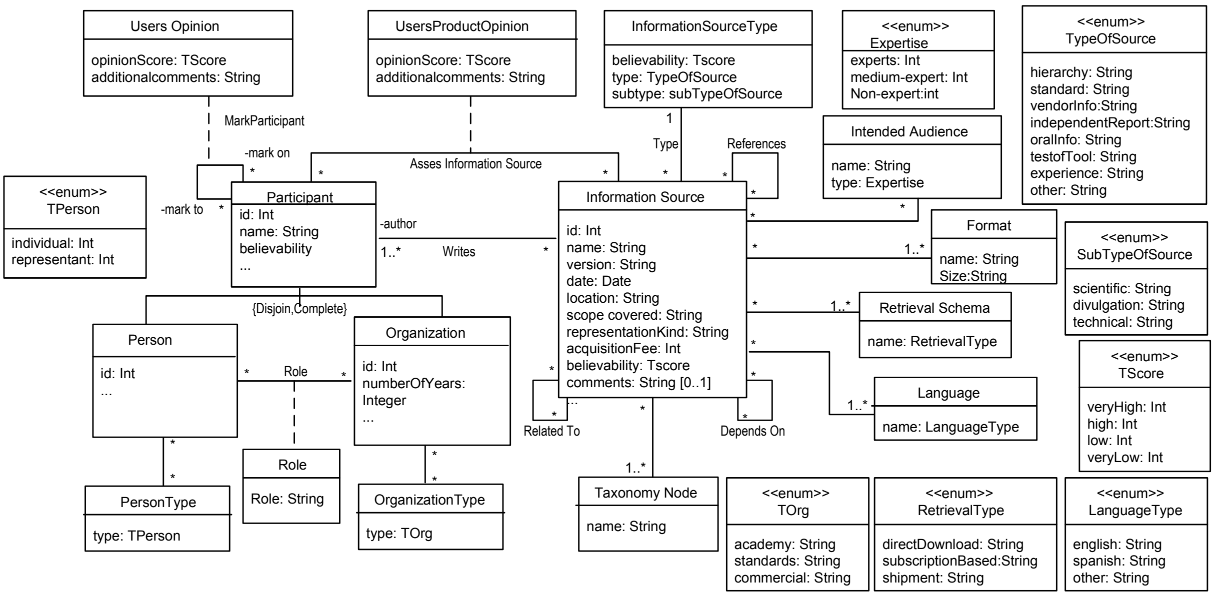

Fig. 2. An excerpt of the Reference model for recording, managing and reusing IQ in SP selection processes

At the heart of this model lies the *Information Source* class described by a set of class attributes or class relationships (e.g., the *believability* class attribute, or the *Retrieval Schema* class). All of them correspond to the IQ attributes identified in the quality model. At this respect, it is important to stand out that Fig.2 only shows an excerpt of the model, and mainly tries to denote some representative examples of the approach, i.e., in this paper we have emphasized the *Believability* subcharacteristic as a representative example of the whole quality model.

Some attributes correspond directly to metrics previously identified; for instance, the *Representation Kind* attribute categorized under the *Understandability* subcharacteristic, is directly stated by the *RepresentationKind* attribute of the *Information Source* class. On the other hand, information sources are characterized by diverse *InformationSourceType* (as detailed in Table 2) by the application of the heuristics mentioned in section 4.1.

From the reuse and management perspective, we define the *Participant* class to describe the subjects that provides a mark, create information sources, or are members of a producer organization. This class is defined by a set of objective attributes (e.g., name), and a subjective attribute named *believability*. This class can refer to a *Person* or an *Organization*. A person can play several roles in an *Organization*. On the other hand, Participants play the *Author* or *Marker* role. Author refers to the *Information Source* creator. Marker refers to who gives an opinion and/or a mark about the believability of the information source based on his/her own assessment of the source. Such opinion is denoted by the *UserProductOpinion* association class. Hence, an *Information Source* can collect more that one *UserProductOpinion*. In the same way, *Participants* can provide marks and comments about the believability of other participants.

Many other quality relationships extracted from the quality model exist in the whole conceptual model, but because of space restrictions are not mentioned here; however, some of them are easily inferred by the model. For instance Accessibility attributes of the *Information source* as *Format*, *Retrieval, Schema*, *Acquisition Cost*; Representational ones as *Language* or *Representation Kind*; and some kind of special relationships among the sources as *Related To*, *Depends On*, and *Reference* that are Intrinsic attributes that denote, the products related, dependent or referenced by the *Information Source* and have quality implications.

Of course, some integrity restrictions are defined in the model, e.g., Authors cannot issue a mark about their own *Information Sources*, or a *Participant* cannot make a mark about himself/herself.

Additionally, we have integrated this approach into our GOThIC method. The way to do that is to consider that the *Information Source* and *Taxonomy Node* classes introduced in Fig.1 are in fact the same as the stated in Fig.2. As mentioned in section 2, a GOThIC taxonomy is used to locate the taxonomy node that fulfills the needs of the user in charge of the SP selection process. Once located, the information sources related to each node can be assessed to obtain high-quality information to adapt the domain model to the specific requirements of the selection project, applying the rules we defined in [29]. Thus, our approach can be used to guide data collection, storage and use, allowing the comparison of various information sources in terms of their quality value for a specific project.

So far we have implemented the conceptual model in a software tool [30], meanwhile we have recorded over 150 information sources we have (re)used in several SP domain analysis projects. Our current work includes: On the one hand the integration of such tool into the DesCOTS system and the scenario described in section 5. On the other hand, the development of functions to systematically compute a consensual result from the match among the *IQ Project Issues´s* attributes to the *Intrinsic*, *Representational*, and *Accesibility IQ´s* attributes. It means to systematically generate *Contextual IQ* attributes results. To do this, we are basing our efforts on the criteria for information quality reasoners defined for Wang et al. [19]. Our main intention is to provide support and flexibility in dealing with the subjective, decision-analytic nature of IQ judgements.

5. An Scenario of Use: Community-Based Collaboration

One of our current projects is related to the combined application of our GOThIC approach with the creative and productive potential of the “open-source collaboration” paradigm [31]. This is jointly performed with the Norwegian University Of Science and Technology. The intended main goals of the project are:

- To provide the web-based infrastructure for enabling SPs technology users to collaborate as a community in an open-source-like environment, see Fig. 3. In such a way, they are encouraged to share knowledge (e.g., experiences, components information, and vendor comments) in a structured and friendly way based on our GOThIC metamodel. Therefore, we have designed proper templates and guidelines for editing and use in order to share the information in a structured way (as demonstrated in the Wikipedia).
- Federating actual efforts for locating and selecting SP by the open and collaborative knowledge sharing of the users (e.g., sharing a list of existing web-resources for locating SP as those cited in Table 1).
- Enabling systematic support for selecting and evaluating SP. Having structured SP information, the systematic support for evaluating and choosing components is enabled. Such information structure greatly deal with usual problems that many existing SP selection tools face (e.g., [6], [7], [8]). At this respect, we are integrating the DesCOTS system into the portal functionalities to systematically support the users of the portal in a diversity of SP selection related processes [11], as stated in Fig.3.

As a result, a web-based portal has been implemented, available at [33] –it still is at test stage-. More details about the project can be found at [31]. A snapshot of the portal is shown in Fig. 4.

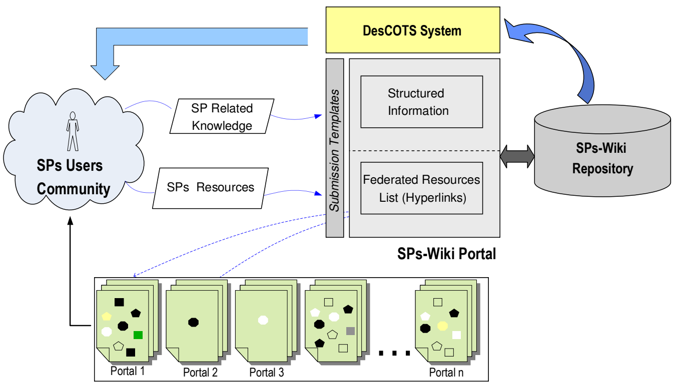

Fig. 3 SPs-Wiki Portal Main Interactions

In this context, IQ awareness plays an essential role mainly because of the increasing growing of content in the portal, its lack of centralized quality control and the diversity of the users contributions. All these make crucial to tackle the IQ quality concerns from the point of view of the SPs selectors for supporting them not only to decide the adequateness of the information to their projects but also the quality related to the contributors. At this respect, although the mechanism of providing judgments to contributors to infer their contribution reliability (as wikipedia does) is not new, offering an IQ solution specific to select SP does it. Thus, integrating our SP IQ approach, we expect not only to improve the rate of successful SP selection cases in the portal by supporting their decision-making processes under a quality assured information environment, but also improve the reuse of IQ assessments to raise the probability of success. This is especially true regarding to the selection of SP of coarse-grained granularity (i.e., ERP, CRM, and ECM are typical examples, whilst time or currency converters are not) where the criticism of the decision made tend to be very high, therefore to have this support is a great help for reducing the critical risks associated to poor quality information.

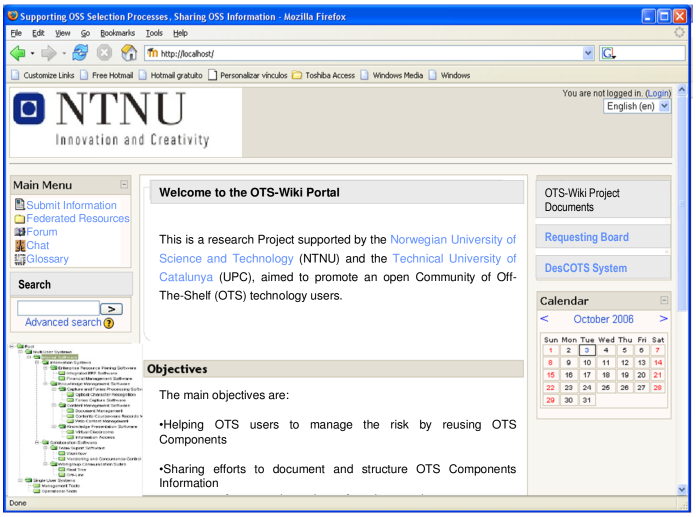

Fig. 4 A snapshot of the SP-Wiki Portal

6. Conclusions and Future Work

This research develops a framework that captures the aspects of IQ that are important to SP selectors. This framework provides a systematic approach for supporting SP selectors to decide information sources to use according to their specific quality project needs. Relevant aspects of our research are:

- We have explicitly identified the quality dimensions required to perform an informed SP selection. Moreover, they were elicited from SP selectors instead of being defined theoretically.
- We have defined the set of attributes that allows gauging, reusing, prioritizing, and managing IQ to support SP decision-making.
- These attributes were organized into an ISO/IEC 9126 tree-like structure which outlines a uniform framework for the integrated evaluation of all SP selection related issues (because it is compatible with our GOThIC approach); and IQ requirements can be considered from the beginning of the SP selection process in the same way we have other kind of requirements (e.g., functional, non-technical, …). Moreover, we have put emphasis on reuse, which allows transferring knowledge from one information quality assessing experience to another.
- This approach has been translated into a conceptual model that could be considered as a reference to systematically tackle the collection, storage, retrieval, and analysis of information sources in SP selection processes.
- Additionally, SP selectors are supported by heuristics to locate and choose information sources suitable to their quality objectives and/or resources allocated to the project.

We also remark that this proposal is part of a larger method documented elsewhere [10],[12],[20],[29],[32] and in particular can be considered as the continuation of the work presented on domain analysis at [12] in the sense that it provides a framework for establishing the properties of information to be used in that activity in order to make it more successful. Last, we stand out the existence of tool-support both as part of a larger system (the DesCOTS system [11]) and in a wiki-based environment for community-building [33].

As future work, we want to make more efficient our approach by automating the evaluation of information source objective properties, i.e., extracting properties such as authors' names and organizations, references, etc., and populating automatically the data base with these properties. Also, we aim at applying technologies such as intelligent agents, ranking algorithms, cluster analysis, web mining/data mining, personalization, recommendation, and collaborative filtering techniques, commonly integrated into internet search engines [34], to improve the construction of knowledge over these raw data.

7. Acknowledgements

This work has been partially supported by the Spanish MEC TIN2004-07461-C02-01 C. Ayala´s work has been partially supported by the Mexican Council of Science and Technology (CONACyT) and the Agència de Gestió d'Ajuts Universitaris i de Recerca (European Social Fund). The authors would like to thank to the Software Engineering Group of the Norwegian University of Science and Technology (NTNU), leaded by the professor Reidar Conradi for their support in performing the industrial study of the proposal, and Carl-Fredrik Sørensen, Simon Larsen and Kristian Aaslund of the same university for the implementation of the portal [33], and Fernando Messegue from the Technical University of Catalunya for the IQ Tool implementation [30].

References

1. Finkelstein, A., Spanoudakis, G., Ryan, M. "Software Package Requirements & Procurement" Proc. Int´l Workshop on Software Specification and Design, IEEE CS Press, 1996.
2. Cechich, A., Réquilé-Romanczuk, A., et al. "Trends on COTS Component Identification and Retrieval" Int. Conference on COTS-Based Software Systems (ICCBSS 2006).
3. Bertoa, M.F., Troya, J.M., Vallecillo, A. "A Survey on the Quality Information Provided by Software Component Vendors". In *Proceedings of the 7th ECOOP Workshop on Quantitative Approaches in Object-Oriented Software Engineering* (QAOOSE), 2003.
4. Ankolekar, A., Herbsleb, J., Sycara, K. Addressing Challenges to Open Source Collaboration with Semantic Web. 3rd Wshop on Open Source Software Engineering, (ICSE 2003).
5. R.E., Boop., L., Smith *Reference and Information Services: An Introduction*. Libraries Unlimited, 2000.
6. Seacord, R., Hissam, A., Wallnau, K. "Agora: A Search Engine for Software Components". Internet Computing, Vol. 2 No. 6, December 1998, pp. 62-70
7. Yanes, N., Sassi, S.B., Jilani, L. " MoReCOTS: a Specialized Search Engine for COTS Components on the Web". Int.Conference on COTS-Based Software Systems, IEEE, 2006.
8. Sjachyn, M., Beus-Dukic, L. "Semantic Component Selection –SemaCS". International-Conference on COTS-Based Software Systems, IEEE, 2006.
9. Simmons, G.L., Dillon, T.S. "Towards an Ontology for Open Source Software Development". In IFIP Intl. Federation for Information Processing, Volume 203, Open Source Systems, Springer, pp.65-75.

10. Ayala, C., Franch, X. “A Goal-Oriented Strategy for Supporting Commercial Off-The-Shelf Components Selection” 9th International Conference on Software Reuse (ICSR). June 2006, pp 13-24.
11. Grau, G., Carvallo, J.P., Franch, X., Quer, C., “DesCOTS: A Software System for Selecting COTS Components”. 30th EUROMICRO Conference (EUROMICRO’04). IEEE Computer Society, pp 118-126.
12. Ayala, C., Franch, X. “Domain Analysis for Supporting Commercial Off-The-Shelf Components Selection” 25th Intl. Conference on Conceptual Modeling (ER 2006). Tucson, Arizona, USA. pp 354-370.
13. Bertoa M., Troya, J.M, Vallecillo, A. ”Measuring the Usability of Software Components” The Journal of Systems and Software 79 (2006) 427–439.
14. Bobrovsky, M., Marré, M., Yankelevich. “A Software Engineering View of Data Quality”
15. CHAOS Virtual BEACON. “The Cost of ERP”. Standish Group, 2002.
16. Shankaranarayan, G., Ziad, M., Wang, R.Y. “Managing Data Quality in Dynamic Decision Environments” Journal of Database Management, 14(4), 14-32, Oct-Dec 2003.
17. Wang, R.Y., Strong, D.M. “Beyond Accuracy: What Data Quality Means to Data Consumers” *Journal of Management of Information Systems*. Vol. 12, No.4, pp5-34. 1996.
18. Lee, Y.W., Strong, D.M., Kahn, B.K., Wang, R.Y. “AIMQ: A Methodology for Information Quality Assessment” Elsevier, Information and Management. 2001.
19. Wang, R., Allen, T., Harris, W., Madnick, S. “An Information Product Approach for Total Information Awareness” IEEE, 2003.
20. Claudia Ayala. "Systematic Construction of Goal-Oriented COTS Taxonomies" 13th Doctoral Consortium at the 18th CAiSE Conference. Luxembourg, 2006.
21. Gerea, M. “Selection and Evaluation of Open Source Components”. http://www.idi.ntnu.no/grupper/su/fordypningsprosjekt-2006/gerea-fordyp06.pdf
22. Li, J. Process improvement and risk management in Off-the-Shelf Component-based development. PhD Thesis 2006. Norwegian University of Science and Technology (NTNU). http://www.idi.ntnu.no/grupper/su/publ/phd/li-phdthesis-22jun06.pdf
23. Torchiano, M., Morisio, M. Overlooked Aspects of COTS-Based Development. *IEEE Software* 21(2): 88-93 (2004).
24. International Organization for Standardization. ISO/IEC Standard 9126: Software Engineering. Product Quality 2001.
25. Basili, V., Caldiera, G., Rombach, D. Goal/question/metric paradigm. In Encyclopedia of Software Engineering. Vol. 1, 1994. J. C. Marciniak, Ed. John Wiley and Sons, New York.
26. Fenton, N.E., Pfleeger, S.L. *Software Metrics: A Rigorous and Practical Approach*. 2nd Edition, 1997.
27. Carvallo, J.P., Franch, X. “Extending the ISO/IEC 9126-1 Quality Model with Non-Technical Factors for COTS Components Selection” Proceedings WQSQ 2006. ACM
28. Chung, L., Nison, B.A., Yu, E., Mylopoulos, J. *Non-Functional Requirements in Software Engineering.* Kuwler Academic Publishers, 2000.
29. Ayala, C., Franch, X. “Transforming Software Package Classification Hierarchies into Goal-Based Taxonomies”. In Proceedings of the 16th Database and Expert Systems Applications Conference (DEXA), LNCS 3588, 2005. pp 665-675
30. Messegue, F. “Eina de suport per al anàlisi de dominis”. http://www.lsi.upc.edu/~cayala/Papers/IQToolDocumentation.pdf In catalan. Jan 2007.
31. Ayala, C., Sørensen, C.F., Conradi, R., Franch, X., Li, J. "Open Source Collaboration for Fostering Off-The-Shelf Components Selection" Third International Conference on Open Source Systems. Limerick, Ireland. June 2007.
32. Ayala, C., Franch, X. "Overcoming COTS Marketplace Evolvability and Interoperability" CAiSE Forum 2006. June 2006, Luxembourg.
33. http://cosiportal.idi.ntnu.no:8080/OTS-Wiki/ -at test stage-
34. Van Gils, B., Proper, H.A., Van Bommel, P. A Conceptual Model of Information Supply. *Data & Knowledge Engineering* 51 (2004) 189-222.