

An XML-based Approach for the Automatic Verification of Software Requirements Specifications*

Amador Durán, Beatriz Bernárdez, Antonio Ruiz, and Miguel Toro

Facultad de Informática y Estadística
Universidad de Sevilla
{amador,beat,aruiiz,mtoro}@lsi.us.es

Abstract In this paper, we present an approach for the automatic verification of software requirements specifications. This approach is based on the representation of software requirements in XML and the usage of the XSLT language not only to automatically generate requirements documents, but also to verify some desired quality properties and to automatically compute some metrics. These ideas have been implemented in REM, an experimental requirements management tool that is also described in this paper.

Keywords: requirements engineering, requirements verification, XML, XSLT

1 Introduction

Quality related activities in the requirements engineering process are basically requirements verification and validation. Paraphrasing Boehm [8], requirements validation and verification could be informally defined by the questions:

validation: "*Am I building the right requirements?*"

verification: "*Am I building the requirements right?*"

In other words, the goal of requirements validation is to ensure that requirements documents contain *actual* requirements and that these requirements are *all* the known requirements by the time the requirements documents are baselined. On the other hand, the goal of requirements verification is to ensure the quality of requirements according to desired quality properties. Some of these quality properties have to do with requirements semantics but others have to do with syntactic, structural or pragmatic aspects of requirements (see [18] for a complete classification of quality properties of requirements).

Verification of semantic properties of requirements is closely related to requirements validation and requires human participation. In fact, distinction between verification of semantic properties of requirements and requirements validation is sometimes subtle and many authors use both terms interchangeably. We think that verification of non-semantic properties should be as automated as possible.

* This work is partially funded by the Spanish CICYT project *GEOZOCO* TIC 2000-1106-CO2-01 and by the international CYTED project *WEST*.

In this paper, we present an automated approach for the verification of some quality properties of requirements and for the automatic computing of requirements metrics. Most of these properties can be classified as non-semantic, but we have also developed some heuristics to check potential problems with some semantic properties. Our approach is based on the emergent technology built around XML [5] and its companion language XSLT [4].

The rest of the paper is organized as follows. First, we briefly describe the basics of XML and XSLT needed to understand the following sections. Then, we describe REM, an experimental requirements management tool [13, 14], the XML model of requirements used by REM and how XSLT can be used to verify some quality properties of requirements expressed in XML. Finally, we discuss some related work, present some results and point out future work.

2 XML and XSLT

2.1 XML Basics

There are millions of web pages written in HTML available in Internet. In these web pages, pure information is mixed with formatting elements, making the automatic processing of information very difficult. XML [5] is a language designed for representing pure information in Internet. Information in XML is represented by *elements*. An XML element is made up of a start tag, an end tag, and other tags or data in between. For example, for representing the information about a book, we might have the following XML element named `book`:

```
<book isbn="X-XXX-XXXX-X">
  <author>Miguel de Cervantes</author>
  <title>El Quijote</title>
</book>
```

As you can see, the information about a book is between the `<book>` and `</book>` tags and it is easy to parse by a computer program. In fact, there are many XML parsers available in Internet for almost any programming language. The `author` and `title` elements are considered as children of the `book` element, thus forming a hierarchy. An XML document must always have one and only one *root element* at the top of its hierarchy.

In order to allow information interchange between two or more parties using XML, they must agree about element grammar and semantics. Element grammar is specified as regular expressions in DTDs (Document Type Definitions)¹ [5]. For example, the DTD fragment for the previous XML data might be the following:

```
<!ELEMENT book (author+,title)>
  <!ATTLIST book isbn ID #REQUIRED>
<!ELEMENT author (#PCDATA)>
<!ELEMENT title (#PCDATA)>
```

¹ There exists an alternative to DTD, called XML Schemas, which is currently in a recommendation status [6].

where it is denoted that a `book` element can contain one or more `author` elements and only one title element. Possible cardinalities are represented by the `?`, `+` and `*` symbols, meaning zero or one, one or more and zero or more respectively. An XML element can also have attributes. For example, `isbn` is defined as a required identification attribute of `book`, *i.e.* there cannot exist two books with the same value for the `isbn` attribute in the same XML document. Those elements that contain only text are said to contain `#PCDATA`, that stands for *parsed character data*.

2.2 Transforming XML

There are many situations in which XML data need to be transformed. For example, for representing the same information according to another set of tags or for presenting XML data in an HTML page. XSLT [4] is a language based on transformation patterns. An XSLT stylesheet, which is also an XML document, searches for patterns in the XML data and applies programmed transformations, thus generating some output results. For example, if we wanted to show information about books in a web browser, we could apply the following XSLT transformation rule:

```
<xsl:template match="book">
  <B><xsl:value-of select="title"/></B>
  (ISBN <xsl:value-of select="@isbn"/>)
  was written by
  <EM><xsl:value-of select="author[1]"/></EM>
</xsl:template>
```

The informal semantics of this XSLT rule are "when you find a `book` element, generate its title in boldface, then its ISBN attribute (notice the `@` prefix for attributes), and then its first author in emphasized mode". In the XSLT code, text literals like HTML tags can be mixed with element values, which are obtained by means of the `xsl:value-of` statement. If we applied this XSLT rule to the previous XML data, the result of the transformation would be something like this when rendered in a web browser:

El Quijote (ISBN X-XXX-XXXX-X) was written by *Miguel de Cervantes*

Although there are many more details about XML and XSLT, we think that this brief introduction should be enough for those readers not familiar with XML technologies in order to understand the rest of this paper.

3 REM: An XML-based Requirements Management Tool

REM (*REquirements Manager*) is an experimental requirements management tool developed by one of the authors as part of his PhD. Thesis [13, 14]. In REM, a requirements engineering (RE) project is considered to be composed by three documents:

1. a customer-oriented requirements document (the *requirements document* [19]), usually containing requirements in natural language expressed in terms of customer's vocabulary, also known as *C-requirements* [10].

2. a developer-oriented requirements document (the *specification document* [19]), usually containing requirements models and more technical information, also called *D-requirements* [10].
3. a registry for detected conflicts and negotiation support.

In REM, C-requirements and conflicts are expressed in natural language using pre-defined requirements templates and some linguistic patterns (see [14] for details). For expressing D-requirements, we have chosen a subset of the UML [9] with strong influences from [11] and [12].

3.1 REM Architecture

REM documents, *i.e.* RE projects composed by the three documents previously described, are stored in relational light-weight databases. When the user creates a new REM document, the basic structure is taken from a *REM base document* (see figure 1), that can be empty or can contain the mandatory sections of software requirements standards like [1] or [21]. Any ordinary REM document can be selected as a base document, so users can create their own base documents or reuse other REM documents.

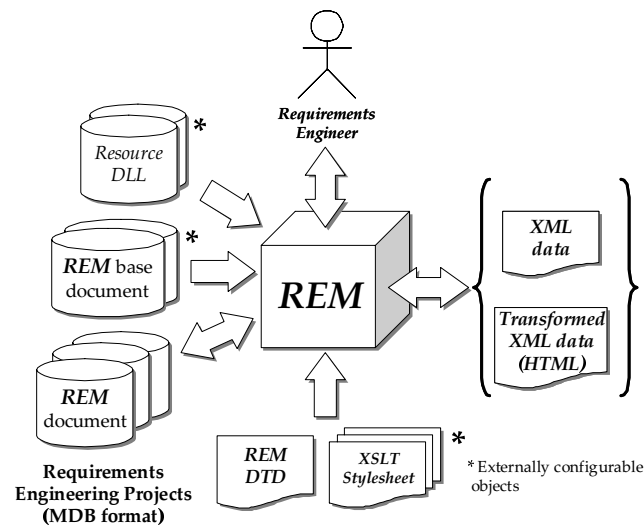


Figure1. REM Architecture

In order to provide immediate feedback on user actions, REM generates XML data corresponding to the document being edited, applies an external XSLT stylesheet that transforms XML data into HTML and shows the resulting HTML to the user. In this

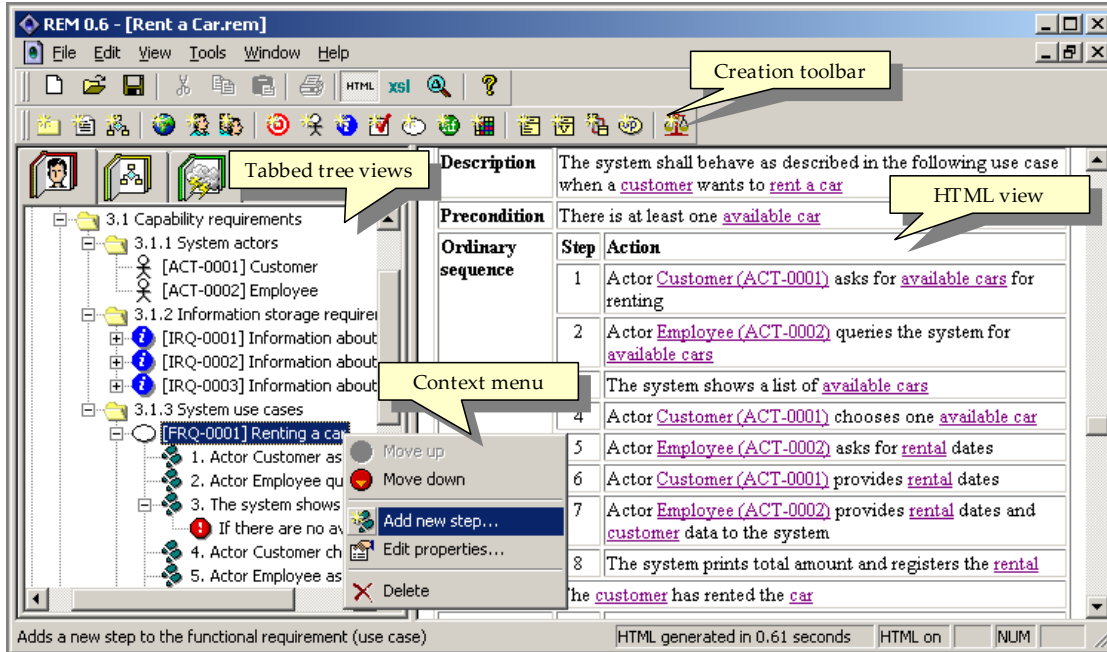


Figure2. REM User Interface

way, whenever the user changes a requirements document, he or she can see the effects immediately.²

In a similar way the REM base document can be tailored, the user can also change document appearance by selecting or creating different external XSLT stylesheets. The default XSLT stylesheet generates a highly hyperlinked document, easing navigation of requirements documents (see right side of figure 2).

Other configurable aspect of REM is the language of the user interface. The user can choose it by selecting an external resource dynamic link library (DLL). At the moment of writing, we have developed two external resource DLLs for REM, one in Spanish and other in English. Another one in Portuguese is under development.

3.2 REM User Interface

The user interface of REM presents two different views to the user (see figure 2). On the left, the user can see a tabbed view with three tree views, one for each requirements document in the RE project. On the right hand, the result of the XSLT transformation of the XML data is presented to the user in a embedded web browser.

² Actually, the performance of XSLT transformations depends on the size of the XML data, *i.e.* the requirements, and the XSLT transformations to be applied. For big documents, the transformation takes a few seconds and cannot be considered as immediate.

In any of the three tree views, the user can directly manipulate objects by drag and drop or by context menus. Only actions that have sense can be performed, following a *correct-by-construction* approach, thus increasing quality and saving verification effort. For example, actions of use case steps can be of three different classes (see figure 4): *actor action*, if the action is performed by an actor; *system action* if the action is performed by the system, or *use case action*, if the action consists of performing other use case, *i.e.* an use case *inclusion* or *extension* [9]. Actor actions and use case actions can be created only if some actor or another use case have been previously created. In general, objects can be created by means of context menus on potential parents or by means of the creation toolbar.

4 XML Model of Requirements in REM

REM is based on an object-oriented model of requirements described in [13] (a partial view of this model is shown in figures 3 and 4). The main object class of the model is the *Requirements Document*, that is composed of a sequence of REM objects (see figure 3 for a classification of REM objects).

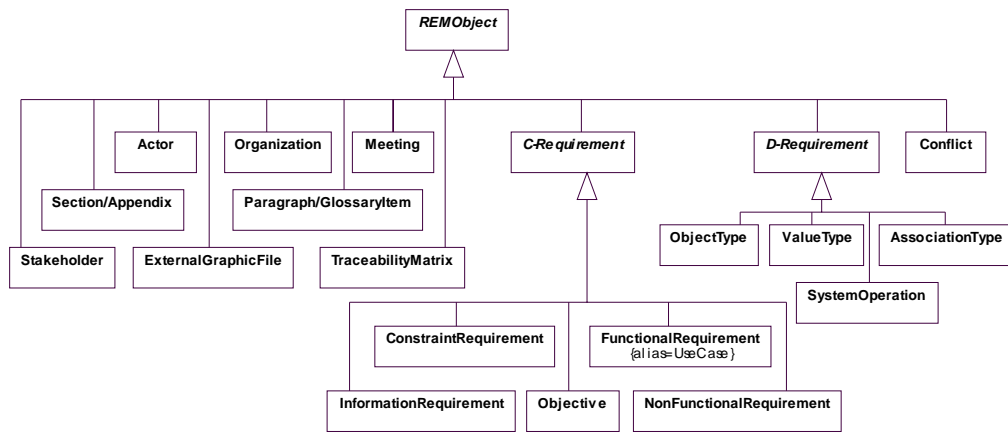


Figure3. Classification of objects in REM

We have translated our UML model of requirements into a relational schema and into a DTD following the heuristics described in [7]. As an example, the *UseCase* class in figure 4 has been translated into the following DTD element definition:

```

<!ELEMENT rem:useCase (
  rem:name, rem:version, rem:authors?, rem:sources?, rem:comments?,
  rem:importance, rem:urgency, rem:status, rem:stability,
  rem:isAbstract?, rem:triggeringEvent, rem:precondition, rem:postcondition,
  rem:frequency, rem:step* )>
<!ATTLIST rem:useCase oid ID #REQUIRED>
  
```

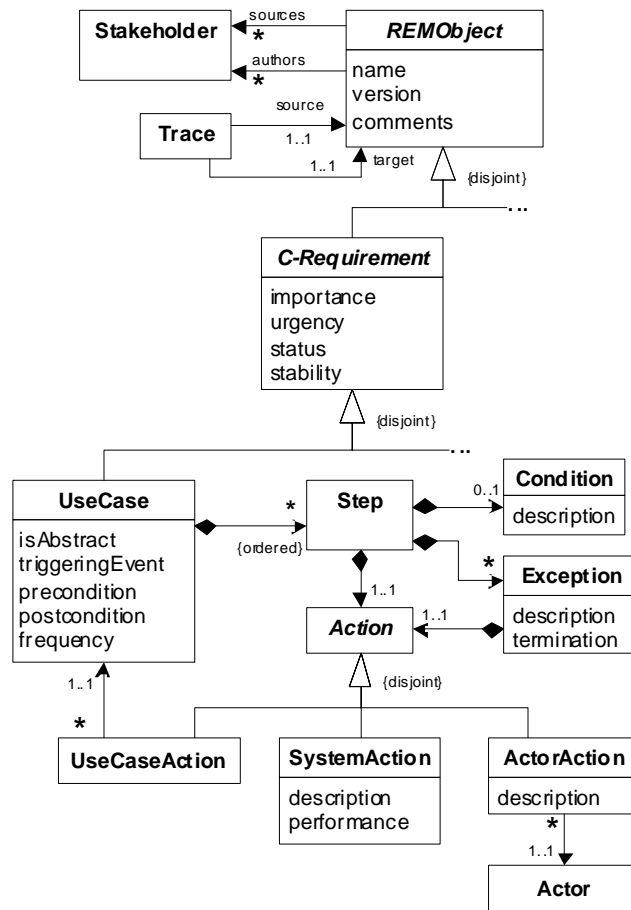


Figure4. UML model of use cases in REM

Many of the elements in the previous DTD fragment (comments, triggeringEvent, pre and postcondition), contains only text, *i.e.* natural language. In REM, text can be composed of any combination of free text, references to other objects and TBD (*To Be Determined*) marks, defined as follows:

```

<!ELEMENT rem:text (#PCDATA|rem:ref|rem:tbd) *>
<!ELEMENT rem:ref (#PCDATA)>
  <!ATTLIST rem:ref oid IDREF #REQUIRED>
<!ELEMENT rem:tbd EMPTY>
  
```

where the `rem:ref` element must have a required attribute called `oid` that it is declared as an `IDREF`, *i.e.* a reference to other element with a matching identification attribute value. The concept of an `IDREF` attribute is very similar to the *foreign key* concept in relational databases.

The `rem:tbid` element is declared as an **EMPTY** element, *i.e.* it cannot have neither subordinate elements nor data. It is simply a mark.

The DTD elements corresponding to use case steps and actions of figure 4 have been described as follows:

```
<!ELEMENT rem:step (
  rem:number, rem:condition?,
  ( rem:systemAction | rem:actorAction | rem:useCaseAction ),
  (rem:stepException*),
  rem:comments )>
<!ATTLIST rem:step oid ID #REQUIRED>

<!ELEMENT rem:systemAction (
  rem:description, rem:performance? )>

<!ELEMENT rem:actorAction (rem:description)>
  <!ATTLIST rem:actorAction actor IDREF #REQUIRED>

<!ELEMENT rem:useCaseAction EMPTY>
  <!ATTLIST rem:useCaseAction useCase IDREF #REQUIRED>

<!ELEMENT rem:stepException (
  rem:condition,
  ( rem:systemAction | rem:actorAction | rem:useCaseAction ),
  rem:termination,
  rem:comments )>
<!ATTLIST rem:stepException oid ID #REQUIRED>
```

Elements not defined in the previous DTD code (condition, description, termination, etc.) are defined as containing only text. For example:

```
<!ELEMENT rem:condition (#PCDATA|rem:ref|rem:tbid)*>
```

5 Using XSLT as a Requirements Verification Language

In the following sections we describe how some of the quality factors described in [15] can be automatically verified using XSLT when requirements are electronically stored in XML format according to the REM DTD.

5.1 Unambiguity

A requirement is unambiguous if and only if has only one possible interpretation [1]. Obviously, this is a semantic property of a requirement and cannot be automatically verified, but we can give some *hints* about potential ambiguities in a requirements document.

We agree with Leite [16] in the importance of understanding the language of the problem and in the importance of building a glossary (called *Language Extended Lexicon*, LEL, in [16]). Following Leite, the glossary should follow two principles: the

principle of circularity, (the glossary must be as self-contained as possible) and the *principle of minimal vocabulary* (use as much glossary items as possible in your requirements descriptions). Leite's principles cannot guarantee unambiguity, but they can help to build unambiguous, understandable, verifiable, consistent, concise, and cross-referenced requirements [15].

XSLT can be used to measure glossary circularity (GLC) and minimality of vocabulary (MOV). GLC can be measured as the ratio between glossary items and references to glossary items from other glossary items, according to the following formula expressed using a Z-like notation [24]:

$$GLC = \frac{\#GlossaryItem}{\#\bigcup\{g : GlossaryItem \bullet g.ref\}}$$

where the numerator is the number of glossary items and the denominator is the size of the union of all sets of references inside the text of all glossary items. The following XSLT code, where we have declared a variable for the sake of readability, can be used for computing GLC:

```
<xsl:variable name="GLC"
  select="count(//rem:glossaryItem) div
    count(//rem:glossaryItem//rem:ref)"/>
<xsl:value-of select="format-number($GLC, '#0.00')"/>
```

where the expression `//rem:glossaryItem` is an XPath expression [3] meaning "any `rem:glossaryItem` element descendant of the root", whereas the expression `//rem:glossaryItem//rem:ref` means "any `rem:ref` element descendant of any `rem:glossaryItem` descendant of the root". In XPath, the language for building navigation expressions over XML trees, an element is considered as descendant of other element if it is its child at any level of depth in the hierarchy.

A similar ratio between the number of references to glossary items in requirements and the number of requirements can be used to measure MOV:

$$MOV = \frac{\#\bigcup\{req : Requirement \bullet req.ref\}}{\#Requirement}$$

From the MOV viewpoint, it is also possible to detect those "suspicious" requirements that do not have any reference to any glossary item in their text. Since those requirements are not using the vocabulary of the customer, they should be checked for potential problems of ambiguity or understandability [15]. For example, if we want to know what use cases are "suspicious", we can use the following XSLT code:

```
<xsl:template match="rem:useCase[not(../rem:ref)]"/>
  Use case
  <xsl:value-of select="rem:name"/>
  does not use any glossary item
</xsl:template>
```

where the match expression uses brackets to select only those use cases with no descendant references. Another possibility is to determine a threshold value for the number of references per requirement and consider as suspicious all requirements with a number of references under the threshold. In that case, the match expression would be `rem:useCase[count(../rem:ref) < m]`, with m being the MOV threshold. We are currently researching proper values for the MOV threshold taking our students' projects as reference data.

5.2 Completeness

A requirements document is complete if it includes [15]:

1. Everything that the software is supposed to do, *i.e.* all the requirements
2. Responses of the software to all classes of input data in all realizable situations
3. Page numbers, figure and table names and references, a glossary, units of measure and referenced material
4. No sections marked as TBD

The two first completeness conditions have to do with semantic properties of requirements and are therefore out of the scope of our approach. On the other hand, the third completeness condition is partially satisfied by means of the *correct-by-construction* paradigm of REM: figure and table names are automatically generated, references are automatically inserted and updated, and the user can easily create a glossary. If we want to be sure about the existence of a section named **Glossary**, we can apply the following XSLT code:

```
<xsl:choose>
  <xsl:when test="//rem:section[rem:name='Glossary']"/>
    There is a glossary
  </xsl:when>
  <xsl:otherwise>
    There is no glossary
  </xsl:otherwise>
</xsl:choose>
```

where the structure formed by `xsl:choose`, `xsl:when` and `xsl:otherwise` is basically an *if-else-endif* statement with multiple `else` branches. Notice that if we want to check the existence of an element we cannot use an XSLT template. If there is no such an element, the template will never match and we will have no output.

Similar XSLT code can be used to verify if requirements documents are *organized* [15], *i.e.* if they have mandatory sections in the mandatory order with mandatory content.

The fourth condition of completeness, the absence of TBD marks, can be easily verified using XSLT. If we want to know how many TBD marks are in a requirements document we can apply the following XSLT code:

```
There are
<xsl:value-of select="count(../rem:tbd)"/>
TBD marks
```

that would generate in the output the number of occurrences of elements of type `rem:tbd` anywhere in the XML data. If we want to be more precise and we want to know, for example, what use cases have TBD marks inside their text and how many TBD marks they have, we can write the following XSLT code:

```
<xsl:template match="rem:useCase[.//rem:tbd]"/>
  Use case <xsl:value-of select="rem:name"/>
  has <xsl:value-of select="count(.//rem:tbd)"/>
  TBD marks
</xsl:template>
```

in which the select expression `"rem:useCase[.//rem:tbd]"` means "any use case with at least one descendant element of type `rem:tbd`".

5.3 Traceability

In [15], a requirements document is said to be *traceable* if and only if it is written in a manner that facilitates the referencing of each individual requirement. Since REM assigns automatically an unique identifier to every requirement (the required identifier attribute `oid`, see the DTD for use cases in section 4 as an example), this quality factor does not have to be verified explicitly.

What it must be checked is if the origin of every requirement is clear, *i.e.* if requirements are *traced* in the sense described in [15]. In our UML model of requirements, any REM object can be traced to and from other REM objects and to their human sources and authors (see figure 4). Checking if a requirement has sources and authors and if it is traced to or from other requirements is easy with XSLT. For example, the following XSLT template will match all use cases with no human sources:

```
<xsl:template match="rem:useCase[not(rem:sources)]">
  Use case <xsl:value-of select="rem:name"/>
  has no defined sources
</xsl:template>
```

And this XSLT template will match all non functional requirements not traced to other REM objects:

```
<xsl:template match="rem:nonFunctionalRequirement">
  <xsl:if test="not(//rem:trace[@source=current()/@oid])">
    Non functional requirement <xsl:value-of select="rem:name"/>
    is not traced to any object
  </xsl:if>
</xsl:template>
```

In REM, traces are defined as elements with two required attributes of type IDREF, namely `source` and `target`. The user of REM can also use traceability matrices for visual checking of non-traced requirements (see figure 5).

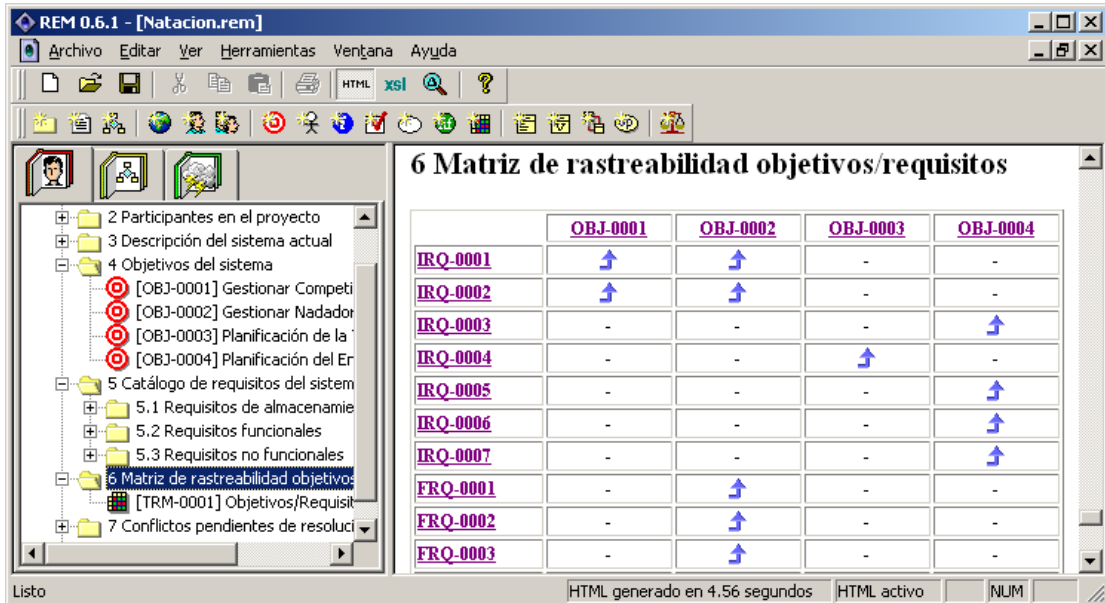


Figure5. REM Traceability matrix example

5.4 Other verifiable quality factors

Applying the same ideas, other quality factors defined in [15] can be verified using XSLT, for example:

- What requirements are not annotated with relative importance, relative stability or version.
- What requirements have potentially ambiguous words in their description, like *easy to, user-friendly*, etc. by means of XSLT string functions like `contains` [4].
- If use cases are not well structured, *i.e.* if there are too few or too many *includes* or *extends* relationships (see figure 6).
- What use cases have too few [17] or too many steps, or too much exceptions, *i.e.* too many alternative courses (see figure 6).
- What actors do not participate in any use case [17].

6 Related Work

Most work on automated requirements verification is based on Natural Language Processing (NLP), like [18] or [20]. Those approaches, focused on semantic analysis of requirements, usually make requirements engineers write requirements in a subset of natural language, demand many computer resources and have not been widely adopted in industry, probably because of their complexity.

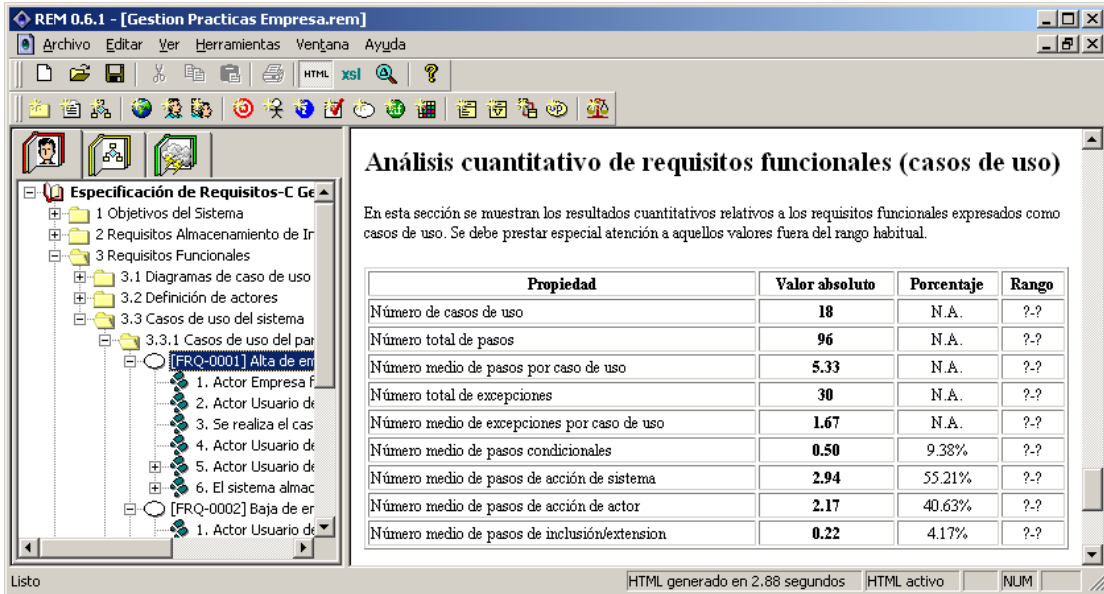


Figure6. REM Verification report example

The Automated Requirement Measurement (ARM) tool [22], is probably the most related work to the approach presented in this paper. It is a simple yet powerful tool that scans requirements documents searching for *indicators*, *i.e.* words that have been identified as indicators of good or bad quality properties.

Our approach does not use NLP but an open, simpler and lighter technology like XML/XSLT. We can offer the same functionality of ARM plus all additional verification described in previous sections. What is more, the user of REM can define his or her own XSLT verification stylesheets. From a practical point of view, we think that our results are useful for the average requirements engineer.

7 Conclusions and Future Work

In this article we have briefly presented an automated approach for the verification of software requirements. Our approach is based on a open technology like XML and XSLT. In fact, if requirements are represented in XML using a different DTD, many of the XSLT code presented in this paper should be easily adapted. Our approach does not need hard computer resources and it has proved to be useful when used with our students at the University of Seville.

Our future work is focused in developing quality metrics, so we can detect potential problems with requirements comparing quantitative values. We expect to identify some useful metrics soon by applying data mining techniques to the requirements documents generated by our students at the University of Seville and in other universities in the context of the CYTED project WEST, where the REM tool will be used soon.

Another lines for future work will be the developing of XSLT stylesheets for computing use case points [23] and for visualization of computed values in web browsers using VML (Vector Markup Language) [2].

References

- [1] IEEE Recommended Practice for Software Requirements Specifications. IEEE/ANSI Standard 830–1998, Institute of Electrical and Electronics Engineers, 1998.
- [2] Vector Markup Language (VML). W3C Note, May 1998.
- [3] XML Path Language (XPath) 1.0. W3C Recommendation, November 1999.
- [4] XSL Transformations (XSLT) 1.0. W3C Recommendation, November 1999.
- [5] Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation, October 2000.
- [6] XML Schema Part 0: Primer. W3C Recommendation, May 2001.
- [7] M. Blaha and W. Premerlani. *Object-Oriented Modeling and Design for Database Applications*. Prentice–Hall, 1998.
- [8] B. W. Boehm. Verifying and Validating Software Requirements and Design Specifications. *IEEE Software*, 1(1):75–88, 1984.
- [9] G. Booch, J. Rumbaugh, and I. Jacobson. *The Unified Modeling Language User Guide*. Addison–Wesley, 1999.
- [10] J. W. Brackett. Software Requirements. Curriculum Module SEI–CM–19–1.2, Software Engineering Institute, Carnegie Mellon University, 1990.
- [11] D. Coleman, P. Arnold, S. Bodoff, C. Dollin, H. Gilchrist, F. Hayes, and P. Jeremaes. *Object-Oriented Development. The Fusion Method*. Prentice–Hall, 1994.
- [12] D. F. D’Souza and A. C. Wills. *Objects, Components, and Frameworks with UML: The Catalysis Approach*. Addison–Wesley, 1999.
- [13] A. Durán. *A Methodological Framework for Requirements Engineering of Information Systems (in Spanish)*. PhD thesis, University of Seville, 2000.
- [14] A. Durán, B. Bernárdez, A. Ruiz, and M. Toro. A Requirements Elicitation Approach Based in Templates and Patterns. In *WER’99 Proceedings*, Buenos Aires, 1999.
- [15] A. Davis *et al.* Identifying and Measuring Quality in a Software Requirements Specification. In *Proceedings of the 1st International Software Metrics Symposium*, 1993.
- [16] J. C. S. P. Leite *et al.* Enhancing a Requirements Baseline with Scenarios. In *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE’97)*, 1997.
- [17] J. C. S. P. Leite *et al.* A Scenario Construction Process. *Requirements Engineering Journal*, 5:38–61, 2000.
- [18] F. Fabbri, M. Fusani, V. Gervasi, S. Gnesi, and S. Ruggieri. Achieving Quality in Natural Language Requirements. In *Proceedings of the 11 th International Software Quality Week*, 1998.
- [19] B. L. Kovitz. *Practical Software Requirements: A Manual of Content & Style*. Manning, 1998.
- [20] N. A. Maiden, M. Cisse, H. Perez, and D. Manuel. CREWS Validation Frames: Patterns for Validating Systems Requirements. In *Fourth International Workshop on Requirements Engineering: Foundation for Software Quality (RESFQ)*, 1998.
- [21] C. Mazza, J. Fairclough, B. Melton, D. de Pablo, A. Scheffer, and R. Stevens. *Software Engineering Standards*. Prentice–Hall, 1994.
- [22] L. Rosenberg, T. Hammer, and J. Shaw. Software Metrics and Reliability. In *9th International Symposium on Software Reliability Engineering*, 1998.
- [23] G. Schneider and J. P. Winters. *Applying Use Cases: a Practical Guide*. Addison–Wesley, 1998.
- [24] J. M. Spivey. *The Z notation: a Reference Manual*. Prentice Hall, 2^a edition, 1992.