

Automatic Procedure Building from Use Case Detection within Informal Technical Documents

Michel Lanque, Alcatel-Lucent, Villarceaux Center, 91260 NOZAY, France, Michel.Lanque@alcatel-lucent.fr
Philippe Larvet, Freelance Consultant, 91470 LIMOURS, France, phlarvet@gmail.com

1. Context of the problem

The problem we address in this paper takes place within the context of Customer Information Systems (documentation and embedded information) such as customer documentation for industrial products, i.e. User's manual, Installation or Maintenance manuals, etc. Our objective is to help technical writers (System & Development teams, Documentation writing teams) to build operational *procedures* from the contents of technical specifications. We call here "procedure" [1] the formal expression of an atomic operational function of the product. A procedure contains steps of operator actions (or end-user tasks) and data to be taken as inputs, all this being under the control of a Graphical User Interface (GUI).

As these *procedures* are fundamentally reusable (they can be used and referenced in different kinds of documents, like User's Guides, Maintenance and Installation Manuals, etc.), they are stored in a Content Management System (CMS) under a standard, easy-to-use and easy-to-assemble format: XML-DITA.

Therefore, please notice that one of the main elements of our process [5] is the idea that technical writers don't have to know neither XML nor DITA. In the contrary, we consider that, as the information already exists within informal, natural-language expressed technical documents, some dedicated and specialized tools have to be able to automatically extract and re-format this information, instead of forcing technical writers to re-express it by manipulating XML and DITA.

Before writing customer documentation, of course the product itself must be specified, designed and developed. So, technical documents exist that functionally describe the product. These informal documents (called TRS, for *Technical Requirement Specifications* in Alcatel-Lucent) are mainly based upon use cases descriptions [2]. The problem we address in this paper consists in automatically detecting the structure of use cases within technical documents and re-expressing these use cases as procedures. An application case of this process consists in detecting use cases within Alcatel-Lucent technical specifications (TRS).

2. Possible solutions

The best existing solution to this problem is fully manual: it consists in searching manually the structures of use cases within the technical documents and reorganizing manually these structures in order to write the corresponding procedures.

This solution is not good enough for many reasons:

- it is time-consuming;
- it requires special skills and competencies in technical product knowledge;
- it needs communication of complex information to Technical writers and Development teams;
- it is a local view of a whole information development (the customer information system must be processed in its E2E globality, not as a local short-term solution);
- it is not seen as part of the final product (it is not really embedded).

3. Our solution

We propose to recognize automatically the use case structure within a technical document through two main steps:

- detecting subtitles and parts of the document that describe elements of the use cases, like context, summary, actors, pre-conditions, operations, etc.,
- building automatically a procedure by reorganizing these elements according to a formal structure,

The detection of the use case structure is made by a dedicated automatic process and is helped by the use of a specific use case ontology [3]. This ontology describes the main keywords used to express a use case, the semantics of these keywords, their synonyms and their inter-relationships.

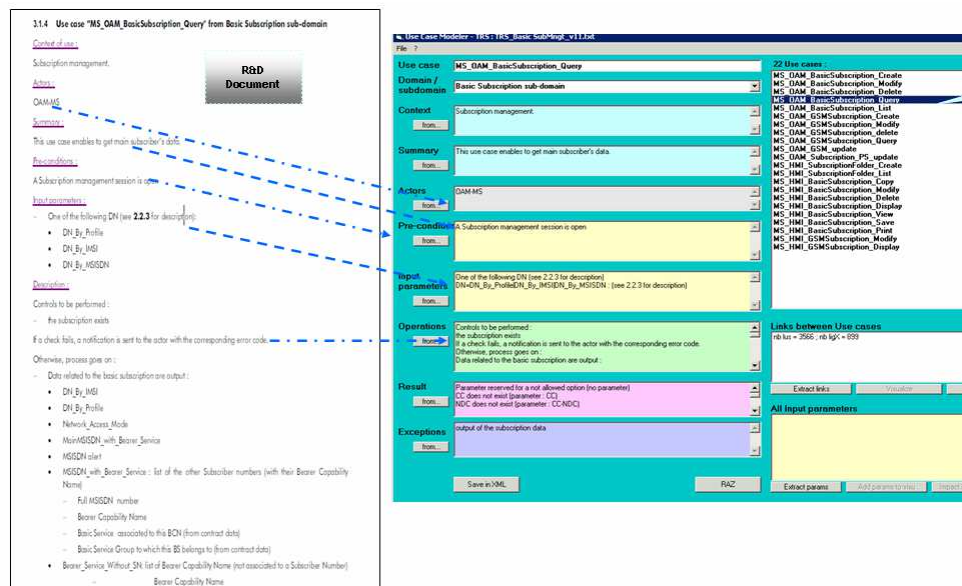


Fig. 1 Description of an embodiment of the solution: from text elements to GUI fields within ProcedureModeler, the tool that implements our solution

4. Implementing the solution

The process of the solution contains the following steps:

1. Text analysis of the document in order to extract pertinent words.
2. Word analysis to determine if a given word is a specific "keyword" depicting a specific element of a use case, for instance Actor, Summary, Pre-condition, Operation, Exception, etc. Special keywords are normally used by technical writers in order to depict use cases, but these keywords are generally written differently, according to the document templates used to write the technical document. So, the keyword detection is made in any part of the document and can be helped by a dedicated taxonomy or ontology where specialized terms, their syntactic derivations and their relationships could be described.
3. Extraction of the paragraphs of the text corresponding to this keyword detection. Each paragraph is considered as a use case element.
4. Gathering all the elements for a given use case, by measuring the semantic distance [4] between the paragraphs within the original text: taking into account this semantic distance and the position of the elements inside the text, the elements of a given use case can be assumed to be gathered together.
5. Reorganizing the elements according to a given pattern or template, in order to build automatically the structure of a procedure from the distinct paragraphs. So, for each use case (UC), the extracted paragraphs are organized in order to build the structure of the corresponding procedure. For example, the paragraphs "Actors" and "Summary" of the UC can be concatenated in the part "Context" of the procedure ; second example, the paragraph "Pre-conditions" of the UC can be used to build the part "Input parameters" of the procedure, etc.
6. Finally, an XML version of the procedure is generated, according to the DITA standard.

5. Valorizing the solution

The value and the unique benefits of this solution can be summarized as follows:

- Cost reduction for technical documentation writing
- Within the context of the Alcatel-Lucent technical documentation process, the generated XML-DITA modules are standard and can be stored and manipulated within the CMS Documentation Modular System.

The advantages of this new solution over the best existing ones can be expressed as follows:

- The definition of a given product can be made from technical modules that come from real functional documentation, which is a help for low-skilled technical writers.
- All R&D resources who have to write technical documents can improve the data capture of information without documentation skill or tool requirements.

6. Conclusion

Today, the solution proposed in this paper is part of a global process [5] and is already implemented within a prototype of a dedicated tool, whose name is ProcedureModeler (you can see an animated demonstration of the tool at www.cidm/... <http://www.infomanagementcenter.com/Demo/DemoProcModeler.html>)

Pilot projects have been launched with the use of ProcedureModeler, and results of these projects are currently gathered in order to build an assessment. The first results are encouraging and seem to prove that the way of generating XML modules from existing informal technical documents is the good one.

References

- [1] Michel Lanque, Philippe Larvet, *Procedure Analysis from Technical Documentation* Alcatel-Lucent internal document, Villarceaux Center, Nozay (France), 2008
- [2] UML use cases overview the usage requirements for a system: see for instance <http://www.agilemodeling.com/artifacts/useCaseDiagram.htm>
- [3] An ontology can be seen as an “intelligent” dictionary containing logical relationships between its terms. See for instance Tom Gruber’s definition: <http://tomgruber.org/writing/ontology-definition-2007.htm>
- [4] For semantic distance, see: Philippe Larvet, *Semantic Application Design*, Bell Labs Technical Journal, Aug.2008, Volume 13 Issue 2, Pages 75 – 91 : <http://portal.acm.org/citation.cfm?id=1405607> and <http://www3.interscience.wiley.com/journal/121376597/abstract?CRETRY=1&SRETRY=0>
- [5] Michel Lanque, *Information Development Process*, Alcatel-Lucent internal document, Villarceaux Center, Nozay (France), 2008