Requirement Mining in Safety Documents

Juyeon Kang Choi, Patrick Saint Dizier

Requirement Mining in Safety Documents

Juyeon Kang, Patrick Saint-Dizier

IRIT-CNRS,
118, route de Narbonne, 31062 Toulouse, France
kjuyeon79@yahoo.fr stdizier@irit.fr

Abstract. In this demo paper, we present a system, implemented in Dislog on <TextCoop>, that searches for requirements in safety documents (regulations, business documents). These requirements form a specific class of arguments.

1 Motivations

Requirements (Hull et al 2011) follow a number of authoring recommendations and have very different structures depending on their type and domain. Software requirements are in general short sentences describing an expected property (e.g. *the system must respond in less than 2 sec*), while, at the other extreme, security requirements are much more developed and are close to procedures: they indicate, under certain conditions the actions to carry out (e.g. *if you manipulate acid, you must wear gloves and adapted glasses, you must also make sure that the ventilation is on, never smoke, eat or drink.*).

In general, requirements related to an activity or to a product are produced by several persons with different profiles (e.g. stakeholders, technicians, engineers and business persons). Their views on a product or an activity complement each other. Requirements are also not written in isolation, they are often parts of larger documents. Besides well-known problems related to requirements such as traceability and update, a major difficulty is requirement mining i.e., given a set of documents, to identify text portions which are requirements.

In this demo paper, we show the syntactic and conceptual structure of safety requirements and how they are processed. We show that requirements are formed of a kernel associated with a number of discourse structures that define more precisely their scope and motivations. We developed the structures of these elements as found in our corpora, for French and English, on the basis of cue terms. This task is part of the LELIE project whose aim is to prevent risks of various kinds from an accurate analysis of technical documents, among which procedures and requirements.

Requirements is a type of argument that follows precise conceptual schemas. They are mainly found in specification documents and regulations which are prescriptive or regulatory documents. They remain relatively generic since their application details, which depend on the domain, are not specified. They are formed of a conclusion (the requirement kernel) possibly associated with one or

more supports, playing the role of justifications, or explanations more generally. They must not be confused with warnings and advice, which are found in procedures. In fact, warnings and advice (to a lesser extend) are implementations of requirements in the specific domain of a procedure.

2 A corpus of requirements

Our analysis of requirement structure is based on a corpus of requirements coming from 7 companies, kept anonymous at their request. Documents are in French or English. Our corpus contains about 500 pages extracted from 27 documents. The main features considered to validate our corpus are the following:

- requirements corresponding to various professional activities: product design, management, finance, and safety,
- requirements corresponding to different conceptual levels: functional, realization, etc.,
- requirements following various kinds of business style and format guidelines imposed by companies,
- requirements coming from various industrial areas: finance, telecommunications, transportation, energy, computer science, and chemistry.

Diversity of forms and contents in this corpus allows us to capture the main linguistic features of requirements, in particular safety requirements. We focus here on the linguistic and discursive structures of requirements, which parallel their logical structure, more widely addressed in the literature. Due to the complexity of the discourse structures, we carried out, in this first analysis, a manual analysis and rule elaboration. We proceeded by generalizing over closely related language cues from sets of examples, as introduced in (Marcu 1997), (Takechi et al 2003), (Saito 2006), and (Bourse et al. 2011) for procedural texts. Using learning methods would have required the annotation of very large sets of documents, which are difficult to obtain. Contrary to procedures, typography and punctuation in requirements is rather poor.

3 A conceptual organization for requirements

In general, in most types of specifications, requirements are organized by purpose or theme. The organisation follows principles given in e.g. (Rossner et al. 1992). The structure of these specifications is highly hierarchical and very modular, often following authoring and organization principles proper to a company or to an organization. Requirements may often be associated with diagrams or pictures, which will not be shown here. The higher level of these documents often starts with general considerations such as purpose, scope, or context. It can then be followed by definitions, examples, scenarios or schemas. Then follow a series of sections that address, via sets of requirements, the different facets of the problem at stake. Each section may include, for its own purpose, general elements

followed by the relevant requirements. Requirements can just be listed in an appropriate order or be preceeded by text. Each requirement can be associated with e.g. conditions or warnings and forms of explanation such as justifications, reformulations or illustrations. One of the challenges developed here is, when extracting a requirement, to also capture all the 'adjuncts' that characterize it and give its scope, purpose, limitations, priority and motivations.

While most documents share some similarities in requirement expression, there is quite a large diversity of structures which are only found in a subset of them. We categorized 20 prototypical structures for French and 14 for English which have different linguistic structures. We give below a sample of English requirement structures in BNF form (bos : beginning of structure, eos: end of structure: characterized by a punctuation or a connector, gap skips strings of words of no present interest):

- requirements composed of a modal applied to an action verb
  → `bos, gap1, Modal, {AdvP}, Verb(Action, Communication, infinive), gap2, eos`
  ex: *The Design Organisation must specify the required Performance Class for each safety barrier installation in terms of Containment Level, Impact Severity Level and the Working Width Class.*
  → `bos, gap1, Modal, {AdvP}, Aux(be), {AdvP}, Verb(Action, Loc, pastparticiple), gap2, eos`
  ex: *Road furniture and equipment must not be positioned in front (i.e. within the set-back) of a new or existing Road Restraint System (RRS).*

- Requirement with complex modals, conformity or necessity terms
  → `bos, gap1, Modal, Aux(be), expr(conform) , gap2, eos`
  ex: *All safety barriers must be compliant with the Test Acceptance Criteria.*
  → `bos, gap1, Modal, Aux, comparativeForm, gap2, eos`
  ex: *Physical entities must have at least one Ethernet interface per zone it is connected to (front, back, administrative).*
  ex: *For all other roads, the Containment Level at the ECP/MCP must be equal to or greater than that of the adjacent safety barrier e.g.*

In this demo, we present the linguistic analysis of the low-level sections of specification documents, which encapsulate requirements related to a certain feature or topic together with explanations, conditions and other discourse structures. The analysis is implemented in Dislog on the <TextCoop> platform (Saint-Dizier 2012). Here is a short illustration:
<main-requirement> <title> Monitoring safe operation of industrial trucks </title>
<warning> Working practices should be monitored by a responsible supervisor to ensure that safe systems of work are followed. </warning>
<purpose> This list is a basic guide - <circumstance> it is not exhaustive and is not intended to be a substitute for the guidance and training. </circumstance> </purpose>
<subtitle1> Operators should always: </subtitle1>
<requirement> observe floor loading limits - <restatement> find out the weight of the laden truck. </restatement> </requirement>

<requirement> make a planning of their way first <condition> if task clear </condition>. </requirement>
<requirement> ensure the load is not wider than the width of the gangways.</requirement>
<requirement> watch out before proceeding <justification> because of pedestrians and bystanders. </justification> </requirement> ....
<illustration>(see diagrams in 390-394)</illustration>. </requirement> ...
<subtitle1> They must never: </subtitle1>
<warning> lift loads that exceed the truck's rated capacity. </warning>
<warning> travel with a bulky load obscuring vision. </warning>
<warning> travel on soft ground <concession> unless the industrial truck is suitable for this purpose </concession>. </warning> ....
<subtitle1> Remember: </subtitle1>
<warning> never allow unauthorised people to operate the industrial truck. </warning>
</main-requirement>

We have carried out an indicative evaluation (showing directions for improvements). For that purpose, we considered text extracts in French, for a total of 32 pages (850 sentences). Three areas have been considered: telecommunication, energy and health. The corresponding lexical data (mainly verbs) has been created from our verb database. Our current implementation (rules and lexical data) allowed the identification and the annotation of 189 requirements, among which 178 correspond to requirements identified by experts, while the 9 remaining are erroneous. This is due in particular to the fact that some structures look like requirements but are not, or are very weak forms of requirements which are not identified as such by experts.

Acknowledgements This project is supported by the French ANR programme, LELIE project.

References

1. Bourse, S., Saint-Dizier, P., The language of explanation dedicated to technical documents, Syntagma, vol. 27, 2011.
2. Bourse, S., Saint-Dizier, P., A Repository of Rules and Lexical Resources for Discourse Structure Analysis: the Case of Explanation Structures, LREC, Istanbul, May 2012.
3. Hull, E., Jackson, K., Dick, J., *Requirements Enginnering*, Springer, 2011.
4. Mann, W., Thompson, S., *Rhetorical Structure Theory: Towards a Functional Theory of Text Organisation*, TEXT 8 (3) pp. 243-281, 1988.
5. Marcu, D., The Rhetorical Parsing of Natural Language Texts, ACL, 1997.
6. Rosner, D., Stede, M., *Customizing RST for the Automatic Production of Technical Manuals*, in R. Dale, E. Hovy, D. Rosner and O. Stock (eds.), Aspects of Automated Natural Language Generation, LNAI, Springler-Verlag, 1992.
7. Saito, M., Yamamoto, K., Sekine, S., Using Phrasal Patterns to Identify Discourse Relations, ACL, 2006.
8. Saint-Dizier, P., Processing Natural Language Arguments with the <TextCoop> Platform, Journal of Argumentation and Computation, vol 3-1, 2012.
9. Takechi, M., Tokunaga, T., Matsumoto, Y., Tanaka, H., *Feature Selection in Categorizing Procedural Expressions*, IRAL2003, 2003.