Information Systems 54 (2015) 191–208

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Formalising natural language specifications using a cognitive linguistic/configuration based approach

Matt Selway*, Georg Grossmann, Wolfgang Mayer, Markus Stumptner

Knowledge and Software Engineering Laboratory, School of Information Technology & Mathematical Sciences, University of South Australia, University Blvd, Building D Level 1, Room 19, Mawson Lakes, SA 5095, Australia

ARTICLE INFO

Available online 17 April 2015

Keywords:
Business rules
Business vocabulary
SBVR
Controlled natural language
Natural language processing

ABSTRACT

This paper addresses the problem of transforming business specifications written in natural language into formal models suitable for use in information systems development. It proposes a method for transforming controlled natural language specifications based on the Semantics of Business Vocabulary and Business Rules standard. This approach is unique in combining techniques from Model-Driven Engineering (MDE), Cognitive Linguistics, and Knowledge-based Configuration, which allows the reliable semantic processing of specifications and integration with existing MDE tools to improve productivity, quality, and time-to-market in software development. The method first learns the vocabulary of the specification from glossary-like definitions then parses the rules of the specification and outputs the resulting formal SBVR model. Both aspects of the method are tested separately, with the system correctly learning 98% of the vocabulary and correctly interpreting 98% of the rules of an SBVR SE based example. Finally, the proposed method is compared to state-of-the-art approaches for creating formal models from natural language specifications, arguing that it meets the criteria necessary to fulfil the three goals of (1) shifting control of specification to non-technical business experts, (2) reducing the manual effort involved in formalising specifications, and (3) supporting business experts in creating well-formed sets of business vocabularies and rules.

1. Introduction

The transformation from natural language business specifications into formal models is an ongoing problem in the development of information systems. While it is desirable to have domain experts develop and maintain the business vocabulary and logic themselves, there is a gap between the natural language descriptions preferred by business people (the domain experts) and the formal models required by technical experts for requirements analysis, consistency checking, compliance checking, model transformations, etc. Manual translation and verification of the natural language text, even by technical experts, is time consuming and error-prone. Furthermore, the resultant formal model requires validation by the domain experts who most likely do not know the formal notation and must be trained to use it.

To address these issues and improve the communication of specifications to technical experts, standards such as the Semantics of Business Vocabulary and Business Rules (SBVR) [1] have been proposed. SBVR attempts to reach a middle ground by providing a controlled English

* Principal corresponding author. Tel.: +61 8 8302 3943.
E-mail addresses: matt.selway@mymail.unisa.edu.au (M. Selway), georg.grossmann@unisa.edu.au (G. Grossmann), wolfgang.mayer@unisa.edu.au (W. Mayer), markus.stumptner@unisa.edu.au (M. Stumptner).

http://dx.doi.org/10.1016/j.is.2015.04.003
0306-4379/

(SBVR Structured English) representation for business people and a metamodel, with a basis in formal logic, for technical experts. SBVR represents a shift of focus from a technical perspective of business rules (see [2,3]) to a business perspective as 'it is conceptualized optimally for business people …and is designed to be used for business purposes, independent of information systems designs' [1, p. 3]. However, when using SBVR, it is still a difficult task to define complete sets of well-formed concepts and rules governing a business [4], so tools that can assist the user in formalising their concepts and rules are a necessity.

Various tools and approaches have been proposed for creating models from natural language specifications. Early work (e.g. [5,6]) focused on identifying the object-oriented elements (i.e. classes, associations, attributes, and methods) of a software specification given in unrestricted natural language for use in information systems design. More recently, SBVR-based tools (e.g. [7–9]) have been proposed for the extraction or formalisation of business rules at a higher level. Yet other approaches focus on extracting specific business aspects, such as business processes, from unrestricted texts (e.g. [10]). Each of these approaches have their pros and cons. Approaches for processing unrestricted text are predominantly automated, but produce incomplete or inaccurate models that require extensive manual revision and validation. On the other hand, more formal approaches (including many of those based on SBVR Structured English) often provide supporting tools and editors, but require manual interpretation and translation (usually by a technical expert) of the business specification into the formal notation of the tool.

To address this, our aim is to provide a method of semi-automating the translation of business specifications into formal models that fulfils three goals:

1. allow business people to develop, maintain, and validate their business specifications with minimal intervention from technical experts;
2. support business users in *formalising* their business specifications as well-formed sets of business vocabularies and rules; and
3. reduce the amount of manual effort involved in formalising business specifications.

The first goal is particularly important as, according to SBVR, a business rule is 'a rule that is under business jurisdiction' [1, p. 8]. This means that business rules can be added, removed, or modified at the discretion of the business [7]. Therefore, the focus of tools needs to shift from the creation of models by technical experts for validation by business people, to the creation and maintenance of formal models by business people themselves.

This leads naturally to the second goal. If we want business people, who are used to specifying things informally, to formalise their specifications themselves, then we need to be able to support them in creating explicit and unambiguous formal models. Consequently, high quality feedback on ambiguities, inconsistencies, and errors is important to enable domain experts to revise the models until they represent the intended meaning.

The third goal aims to minimise the manual effort of existing processes as much as possible. Therefore, in achieving the first two goals, existing business specifications should not have to be completely rewritten, new specifications should not have to be written in a highly particular or technical form, and the analysis and translation of the text into formal models should be highly automated.

In [11], we presented an approach to fulfil these goals that focused on the transformation from (controlled) natural language business specifications into their formal representation in SBVR. We used SBVR as the starting point for both the natural language business specifications and the formal representation as it is intended for use by business people and helps to reduce ambiguity, ensure strict semantics, and provide a more direct logical interpretation of the text. We then introduced a novel combination of techniques from Cognitive Linguistics, Knowledge-based Configuration, and Model-Driven Engineering (MDE) to iteratively build a formal SBVR model of the vocabulary and rules contained within the business specification. Finally, we argued that this approach combines the advantages of alternative approaches and meets specific criteria (such as naturalness, expressiveness, and parsing completeness) required to achieve the goals mentioned above.

In this paper we revise and extend the work presented in [11]. We have introduced a clearer overview of approaches to creating models from textual specifications in Section 3, including additional approaches. Furthermore, we have expanded the description of the approach (Section 5) with additional examples and incorporated multiple new heuristics that help to optimise the parsing process. Finally, a new evaluation section has been included (Section 6) that presents experimental results on the application of our prototype for SBVR to the EU-Rent case study discussed in the motivating example.

The contributions of our approach include the following:

1. Flexible syntactic analysis to allow the processing of less restricted text than many controlled natural languages.
2. Direct integration of semantic analysis into the parsing process enabling:
   (a) detailed feedback on errors and inconsistencies in the specification,
   (b) non-technical users to revise and formalise the specification themselves, and
   (c) the identification of missing or incorrect vocabulary and the suggestion of corrections during processing;
3. Specification of formal vocabulary in a glossary-like format without the need for technical notations.

The remainder of the paper is organised as follows: Section 2 introduces a motivating example that is based on the EU-Rent case study and used throughout the remainder of the paper; Section 3 provides an overview of existing techniques for generating models from natural language specifications; Section 4 briefly introduces SBVR, Knowledge-based Configuration, and Cognitive Grammar; Section 5 describes our approach to parsing natural language business specifications;