Vladimir A. Fomichov

The Advantages of Using SK–languages for Designing Semantic-Syntactic Analyzers of Recommender Systems // Kurt J. Engemann and George E. Lasker (Eds.) Advances in Decision Technology and Intelligent Information Systems, Vol. XIV. The International Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Tecumseh, Ontario, Canada, 2013, p. 37-41. - ISBN 978-1-897546-73-4

The Advantages of Using SK–languages for Designing Semantic-Syntactic Analyzers of Recommender Systems

Vladimir A. Fomichov

Department of Innovations and Business
in the Sphere of Informational Technologies
Faculty of Business Informatics
National Research University Higher School of Economics
Kirpichnaya str. 33, 105187 Moscow, Russia

vfomichov@hse.ru and vfomichov@gmail.com

Abstract

The paper describes a broadly applicable method of designing multilingual semantics-syntactic analyzers of recommender systems. The user inputs may include the questions of many kinds formed with the help of interrogative words (or without interrogative words), verbs, nouns, attributes, prepositions, the designations of the digital values of various parameters. For the queries in English and German, the developed algorithm of semantic-syntactic analysis processes the questions of many kinds, the commands, and the statements from a restricted sublanguage of NL. For the queries in Russian, the algorithm is additionally able to process the requests with participle constructions and attributive clauses. As a semantic intermediary language, the algorithm uses the SK-language determined by the considered linguistic database. The class of SK-languages is introduced by the theory of K-representations (knowledge representations), its current version is mainly stated in a monograph of the author published by Springer in 2010. The developed algorithm is implemented by means of the programming language PYTHON.

Keywords: recommender system with natural language interface; semantic representation of user query; semantic language; theory of K-representations; SK-languages; algorithm of semantic-syntactic analysis

Introduction

A branch of e-commerce called Recommender Systems (RecS) has been quickly progressing since the end of the 1990s. The software systems of this class are intended for consulting the end users of the Internet with the aim of helping them to take the decisions about the choice of products or/and services. The achievements in the field of designing natural language (NL) processing systems accumulated by the beginning of the 2000s caused the birth of the RecS with a NL-interface. An experiment carried out in USA showed that some 80 percents of the RecS users prefer to work with a natural language interface to RecS but not with numerous displayed menus (the English language has been used in this experiment) (Chai et al, 2002). This experiment gave a considerable impulse

to the development of many RecS where the key role is played by the interaction with the users by means of NL – English, Russian, German, etc.

A NL-interface of a RecS should extract the meaning of the user query in order to construct a search request in a special language (for instance, an SQL expression) and to compare the constructed search request with the current filling of the product database. There are possible two approaches to forming a search request of the kind. The first one is the straightforward transformation of a NL-query into an expression in a special language. Several variants of this approach used for the design of a NL-interface to a RecS are described in (Chai et al, 2002; Berger et al, 2004; Ludvig and Mandl, 2006).

However, it seems that broader prospects for practice are opened by the second approach, it may be reflected by the scheme

Expr => SemRepr => PragmRepr.

Here *Expr* is a NL query of a user, *SemRepr* is *a semantic representation (SR), or text meaning representation*, of *Expr*, that is, a formal structure reflecting the semantic structure (or conceptual structure) of *Expr* in an application independent form. Let's agree that the languages for constructing SRs of NL-expressions will be called *semantic languages*. The structure *PragmRepr* will be called below *a pragmatic representation* of *SemRepr*; in particular, it may be an SQL-expression or a search pattern for a Web-search engine. The approach of the kind underlay the design of RecS with NL-interface described in (Lops et al, 2010; Fomichov, 2012). Besides, this approach was used for the design of the question-answering system ORAKEL (Cimiano et al, 2007, 2008). Let's call the approach of the kind *the approach with semantic intermediary language*.

Two principal advantages of the second approach may be formulated. Firstly, it is profitable for the scientific-technical laboratories designing NL processing systems for various applications. The susbsystems realizing the transformation stage *Expr* ➔ *SemRepr* can be used in different application domains with relatively small adaptation. Secondly, the growing necessity of cross-lingual information search is an important reason for using semantic intermediary languages in order to represent the meaning of the user input and to compare this meaning with the meanings of analyzed expressions in the considered natural languages. For instance, this idea underlay the design of the RecS described in (Lops et al, 2010).

One of the central aspects of using the second approach in the design of NL-interfaces of RecS and of applied intelligent systems of other kinds is the expressive power of the employed semantic languages. Most often, the researchers use with this aim first-order logic (FOL) and the formalisms characterized by the authors as expanded versions of FOL. Discourse Representation Theory (DRT) and Episodic Logic (EL) are known formalisms from this class.

However, FOL has numerous restrictions concerning the description of semantic structure of sentences and discourses in NL. First of all, it applies to representing semantic structure of the questions of many kinds (in particular, of the questions with the answer Yes or No or with the beginning "How many") and of NL-texts including the infinitive or gerundial constructions, compound designations of sets and notions, homogeneous members of sentences, the references to the meanings of phrases and larger parts of discourse. That is why the theory of designing NL-

interfaces of RecS demands the development of more powerful and flexible semantic intermediary languages.

This paper continues the line of the work (Fomichov, 2012), where a new method of designing semantics-oriented NL-interfaces of RecS was introduced and the basic principles of semantic processing of the user queries in a very restricted sublanguage of NL were set forth. The aim of this paper is to describe the main ideas of semantic processing of much broader sublanguages of NL including the texts with the verbs and interrogative words.

2 Task Statement

In the monograph (Fomichov, 2010), a new theory of designing semantic-syntactic analyzers of NL-texts with the use of formal means for representing input, intermediary, and output data is proposed. This theory, called the theory of K-representations (knowledge representations), can be interpreted as a powerful and flexible tool of designing the NL-interfaces of RecS.

The *first constituent* of the theory of K-representations is the theory of SK-languages (standard knowledge languages), it is stated, in particular, in (Fomichov, 2010). The kernel of the theory of SK-languages is a mathematical model describing a system of such 10 partial operations on structured meanings (SMs) of natural language texts (NL-texts) that, using primitive conceptual items as "blocks", we are able to build SMs of arbitrary NL-texts (including articles, textbooks, etc.) and arbitrary pieces of knowledge about the world. The analysis of the scientific literature on artificial intelligence theory, mathematical and computational linguistics shows that today the class of SK-languages opens the broadest prospects for building semantic representations (SRs) of NL-texts (i.e., for representing meanings of NL-texts in a formal way).

The *second constituent* of the theory of K-representations is a broadly applicable mathematical model of a linguistic database, it is described in Chapter 7 of (Fomichov, 2010). The *third constituent* is formed by several complex, strongly structured algorithms carrying out semantic-syntactic analysis of texts from some practically interesting sublanguages of NL. In particular, the algorithm *SemSynt1*, described in Chapters 8 - 10 of (Fomichov, 2010), transforms a NL-text into its semantic representation being an expression of a certain SK-language (Fomichov, 2010).

The considered task is the development of a broadly applicable, multilingual semantic-syntactic analyzer of various recommender systems. The user inputs may include the questions of many kinds formed with the help of interrogative words (or without interrogative words), verbs, nouns, attributes, prepositions, the designations of the digital values of various parameters. For the queries in English and German, the algorithm of semantic-syntactic analysis is to process the same sublanguage of NL as the algorithm *SemSynt1*. For the queries in Russian, the algorithm should process the requests with participle constructions and attributive clauses. As a semantic intermediary language, the algorithm is to use the SK-language determined by the considered linguistic database.

3 Obtained Results

An expanded version *SemSynt1exp* of the algorithm *SemSynt1* has been developed, it has been implemented in the programming language PYTHON. Let's consider the examples illustrating the correspondence between the sentences in English and their semantic representations (SR) being the expressions of a certain SK-language, that is, being the K-representations of the input texts. In these examples, the SR of the input text T will be the value of the string variable *Semrepr* (Semantic representation). The considered examples illustrate the correspondence between the inputs and outputs of the developed algorithm *SemSynt1exp.*

Example 1. Let T1 = "Is the bedroom suite “Rainbow” produced in Finland?”, then

*Semrepr = Question (x1, (x1 ≡ Truth-value (Situation (e1, producing1 * (Time, certn moment * (Earlier, #now#) : t1) (Product-role, certn bedroom-suite * (Name1, “Rainbow”) : x2) (Place, certn country1 * (Name1, “Finland”) : x3))))).*

Example 2. Let T2 = ""What firm produces the medicine "Zinnat"?". Then

*Semrepr = Question (x1, Situation (e1, production1 * (Time, #now#) (Agent2, certn company1 : x1) (Product-role, certn medicine1 * (Name1, “Zinnat”) : x2))).*

Example 3. Let T3 = "What Eastern European countries have the factories where one assembles the cars of the company “Toyota”?”. Then

*Semrepr = Question (S1, (Qualitative-composition (S1, country1 * (Location, Eastern-Europe)) ∧ Description(arbitrary country1* (Element, S1) : y1, (Situation (e1, placing1 * (Agent1, y1)(Placed-object, certn factory1 : x1) (Time, #now#) ∧ Situation (e2, assembly1 * (Agent3, x1)(Time, #now#)(New-product, some car1 * (Manufacturer, certn company1 * (Name1, “Toyota”) : x2))))))).*

Example 4. Let T4 = "Who heads the company “L’Oreal”?". Then

*Semrepr = Question (x1, Situation (e1, heading1 * (Time, #now#) (Agent1, x1) (Organization-role, certn company1 * (Name1, “L’Oreal”) : x2))).*

Example 5. Let T5 = "Who and since what time has been the head of the company “L’Oreal”?". Then

*Semrepr = Question ((x1 ∧ x2), Situation (e1, heading1 * (Agent1, x1)(Time, certn moment * (Earlier, #now#) : x2)(Organization-role, certn company1 * (Name1, “L’Oreal”) : x3))).*

Example 6. Let T6 = "How many models of vacuum cleaner does the company “Miele” produce?". Then

*Semrepr = Question (x1, ((x1 ≡ Numb(S1)) ∧ Qualitative-composition (S1, model1 * (Techn-product, vacuum-cleaner) ∧ Description (arbitrary model1 * (Element, S1) : y1, Situation (e1, producing1 * (Time, #now#) (Agent2, certn company1 * (Name1, “Miele”) : x2)(Product-role, y1))))).*

4 Conclusions

The considered examples show that SK-languages are convenient for building semantic representations of the user NL queries being the questions with the answer Yes or No, the questions about the quantity of elements of a set, about the sets with compound descriptions, the questions including compound designations of sets and notions, participle constructions and attributive clauses. With respect to the analysis of the expressive power of SK-languages given in (Fomichov, 2010), it is possible to conclude that the theory of K-representations can be used as a powerful and flexible tool for designing arbitrary semantics-oriented natural language interfaces of recommender systems.

References

1. Berger, H., Dittenbach, M. and D. Merkl (2004); An Adaptive Multilingual Interface for Tourism Information; International Journal of Electronic Business, V. 2, No. 5 (pp. 531-541)
2. Chai, J., Horvath, V., Nicolov, N., Stys, M., Kambhatla, N., Zadrozny, W. and P. Melville (2002); Natural Language Assistant – A Dialog System for Online Product Recommendation; AI Magazine, V. 23, No. 2 (pp. 63-76)
3. Cimiano, P., Haase, P., Heizmann, J. and M. Mantel (2007); ORAKEL: A Portable Natural Language Interface to Knowledge Bases; Technical Report, Institute AIFB, University of Karlsruhe, Germany
4. Cimiano, P., Haase, P., Heizmann, J., Mantel, M. and P. Studer (2008); Towards Portable Natural Language Interfaces to Knowledge Bases – the Case of the ORAKEL System; Data and Knowledge Engineering (DKE), V. 65, No. 2 (pp. 325-354)
5. Fomichov, V.A. (2010); Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Series: IFSR International Series on Systems Science and Engineering, Vol. 27; Springer, New York, Dordrecht, Heidelberg, London (354 pp)
6. Fomichov, V.A. (2012); The Theory of K-representations as a Flexible Tool for Designing Natural-Language Interfaces of Recommender Systems; Advances in Decision Technology and Intelligent Information Systems, Vol. XIII (ed. K.J. Engemann and G.E. Lasker); The International Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Tecumseh, Ontario, Canada (pp. 6-10).
7. Lops, P., Musto, C., Narducci, F., De Gemmis, M., Basile, P. and G. Semeraro (2010); Cross-Language Personalization Through a Semantic Content-Based Recommender System; Artificial Intelligence: Methodology, Systems, and Applications; 14th International Conference, AIMSA 2010, Varna, Bulgaria, September 8-10, 2010, Proceedings (ed. D. Dicheva and D. Dochev), LNCS 6304; Springer, Berlin, Heidelberg (pp. 52-60)
8. Ludvig, B. and S. Mandl (2006); How to Analyze Free Text Descriptions for Recommending TV Programmes; available online at http://www8.informatik.uni-erlangen.de/IMMD8/Lectures/EINFKI/ecai_2006.pdf