



ICTE 2016, December 2016, Riga, Latvia

Service-Oriented Architecture of Intelligent Environment for Historical Records Studies

Aleksandrs Ivanovs^{a,b}, Aleksey Varfolomeyev^{c,*}

^a*Daugavpils University, Vienības iela 13, Daugavpils, LV-5401, Latvia*

^b*Institute for Regional Studies (REGI), Rezekne Academy of Technologies, Atbrīvošanas aleja 115, Rēzekne, LV-4601, Latvia*

^c*Petrozavodsk State University, Lenina pr. 33, Petrozavodsk, 185910, Russia*

Abstract

Nowadays, such socio-technical systems as virtual research environments are increasingly employed in the studies of textual heritage. The paper presents a modular platform based on a combination of Web services that provide tools for historical records publishing and research. A promising approach is the creation of semantic networks that encompass marked up texts of records (in this case, medieval documents preserved at Latvian State Historical Archives) and corresponding meta-information. The authors describe the prototype of a software system, which maps XML markup into ACE (Attempto Controlled English) statements; these statements can be automatically translated into OWL (Ontological Web Language), visualized as semantic networks, or queried by means of Web services of Attempto Project. This system is designed as an intelligent environment for collaborative research.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the scientific committee of the international conference; ICTE 2016

Keywords: Virtual research environment; Service-oriented architecture; Digital humanities; Attempto controlled English

1. Introduction

Rapid development of computer networks technologies has resulted in their introduction into almost all spheres of life, including scholarly research. Within this trend, virtual research environments¹ (VRE) are widely employed by distributed communities of researchers in order to organize collaborative processing of distributed data. At present,

* Corresponding author. Tel.: +7-953-536-0268; fax: +7-814-2711000.
E-mail address: avarf@petrsu.ru

VRE have developed into multiplex socio-technical systems that embrace common software environments, which can be adapted for different research tasks, software components, which are used as research tools, as well as users who form research teams². In this regard, the problem of interaction between the constituent parts of such systems is of utmost importance. For instance, it is desirable to ensure an exchange of research methods (algorithms) between distributed users. Moreover, the set of research tools elaborated for a definite project might be supplemented with the tools derived from other research environments.

Nowadays, many research tools are employed as autonomous software agents that interact with other programs by means of API. A common approach to the elaboration of such agents is their design in the form of Web services^{3,4} (SOAP-based or RESTful). Interfaces of Web services are described using standardized languages; it makes possible to apply to these Web services from other programs.

A number of specialized Web services have been created for VRE in the field of Digital Humanities^{5,6}. For example, in CLARIN Project⁷ that unites national language resources, dozens of Web services aimed at processing of the texts in different European languages are used. Most of the problems that can be settled by means of these Web-services are also topical in historical records studies and history research in general (e.g., named entity recognition, event extraction, sentiment analysis, etc.). However, in historical records studies, the scope of information that should be extracted from historical records is much wider; therefore, the aforementioned Web services are quite insufficient to process original historical records.

One of the noticeable trends in historical records studies is semantic representation of historical information and, in particular, the sense of the texts of historical sources, which is recorded in a formalized way suitable for computer processing^{8,9,10}. Actually, this trend reflects the topicality of numerous projects aimed at representation of cultural heritage; moreover, it corresponds fully with the Semantic Web paradigm¹¹.

In compliance with this trend, the authors have designed a modular platform, which is based on a combination of Web services that provide tools for semantic publishing and research of historical records. A promising approach is the creation of semantic networks that encompass marked up texts of records (in this case, medieval documents¹² preserved at Latvian State Historical Archives – a structural unit of the National Archives of Latvia) and corresponding meta-information. In the paper, the authors describe the prototype of a software system, which maps XML markup into ACE statements; these statements can be automatically translated into OWL, visualized as semantic networks, or queried by means of Web services of Attempto Project¹³. The system is designed as an intelligent environment for collaborative research.

The rest of the paper is organized as follows. Section 2 provides a brief overview of Attempto Controlled English and Web services specially developed for this controlled language. Section 3 presents the architecture of a virtual research environment designed by the authors for the purposes of an in-depth study of historical records, which are published in XML-format. Section 4 describes an example of application of the virtual research environment to the study of medieval charters in Old Russian. Section 5 shows how this software environment can be used to supplement and enrich semantic resources aggregated for historical e-tourism and museums. Section 6 summarizes the results of the paper.

2. Attempto controlled English

Attempto Controlled English (ACE) is one of the controlled natural languages¹⁴ that are used for knowledge representation. ACE is a subset of Standard English with a restricted syntax. It means that ACE is understandable by both humans and computers.

ACE texts can be translated into the so-called Discourse Representation Structures¹⁵ (DRS) that, in their turn, can be directly correlated with first-order logic formulas. Therefore DRS can be translated into any formal language (e.g., Rule Markup Language, Semantic Web Rule Language, OWL, etc.), which is equivalent to a subset of first-order logic. Such representation of the texts provides a basis for logic reasoning, namely: automatic question answering, theorem proving, and logical entailment.

It is very important that in the course of processing ACE texts different domain-specific lexicons can be used. Furthermore, ACE-texts can be processed using reasoner, which generate new hypotheses on the bases of the facts revealed by a researcher. Therefore, this language is quite appropriate for the purposes of semantic publication of historical records.

Download English Version:

<https://daneshyari.com/en/article/4961354>

Download Persian Version:

<https://daneshyari.com/article/4961354>

[Daneshyari.com](https://daneshyari.com)