Information Systems 69 (2017) 195–217

Contents lists available at ScienceDirect

Information Systems

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

Semantic interoperability with heterogeneous information systems on the internet through automatic tabular document exchange

Shuo Yang[a,*], Jingzhi Guo[a], Ran Wei[b]

[a] Faculty of Science and Technology, University of Macau, Macau, China
[b] Rutgers University, Newark, USA

ARTICLE INFO

Article history:
Received 15 May 2016
Revised 14 September 2016
Accepted 30 October 2016
Available online 17 November 2016

Keywords:
Information fusion
Semantic interoperability
Semantic interpretation
Semantic inference
Automatic document exchange

ABSTRACT

Internet is a common information space populated with many entities (e.g., Internet of Things) with different information system types. Each of them has its own context of how to build and process documents (e.g., form documents). This leads to heterogeneous documents in terms of syntax and semantics, which are difficult to make information fusion from one context to another. To resolve this problem, this paper uses semantic interoperability technique which consists of two automatic stages including consistent data understanding and reasonable data usage. To implement semantic interoperability, this paper proposes a novel automatic tabular document exchange (DocEx) framework comprised of a new tabular document model (TabDoc) and a semantic inference scheme to fit the two stages above respectively. In this TabDoc model, a new Tabular Document Language (DocLang) as a communication medium between users and devices is provided, which is not only an information representation language but also a rule language for semantic inference as well. Abstract sub-tree-based semantic relations constructing the logical structure of a tabular document are separated from their presentational structures, clarifying the relationship between semantic groups (e.g., a cell or a block) with the help of a common dictionary CONEX. Besides, this paper proposes a semantic inference algorithm (SIA) executing the inference procedure on received tabular documents created by a Table Designer system which integrates with SIA. Finally, the proposed framework is applied to the processing of flight ticket booking in a realistic e-business scenario. The results show that the proposed method in this paper improves the performance of information fusion among different information systems on the Internet.

1. Introduction

Currently, information fusion in Internet of Things (IoT) is a challenging issue in research [1–7]. This is because data generated by different sources with different contextual and conceptual representations are heterogeneous in terms of structures and semantics [8]. Besides, different processing logics are applied to different applications, leading to semantic conflicts [9]. Thus, information transferred among different IoT systems may suffer from semantic loss or misunderstanding. Among the techniques to tackle information heterogeneity in IoT, information fusion is still preferred in engineering applications, as it aims to integrate information from multiple sources and merge large amounts of dissimilar information into a more comprehensive and easily manageable form [1,7,10]. However, many previous methods of information fusion mainly focus on communication interoperability, but neglect interoperability on the level of the 'meaning' of the exchanged messages [11].

This paper initiates the usage of semantic interoperability technique to fulfill information fusion. Semantic interoperability is defined in [12] as the ability of two or more systems or elements to exchange information and to use the exchanged information. [13] and [14] complement this definition stating that semantic interoperability provides the means to automatically integrate and process large amounts of information without human intervention. In fact, semantic interoperability becomes an important technique in recent years. It has been used to tackle various problems such as database integration [15], information fusion among IoT devices [16] and enterprise information systems interoperability [17,18]. For example, [9] proposes a vocabulary-based method to achieve the information interoperability among different IoT devices by collaboration on different command parameters. [17,19] partition semantic interoperability procedure in three stages by defining it as the ability of a collection of communicating entities to (i) share specified information and (ii) operate on that information according to a shared operational semantics in order to (iii) achieve a specified purpose in a given context. However, with most previous approaches of semantic interoperability concentrating on data

* Corresponding author.
E-mail address: yb37416@umac.mo (S. Yang).

http://dx.doi.org/10.1016/j.is.2016.10.010
0306-4379/

understanding task, demand for improvement still remains in data usage task. For example, in [20], Sciore concludes that the central issue of semantic interoperability is maintenance of consistent meaning understanding when exchanging data among different information systems (IS). In [17], to evaluate the degree of semantic interoperability between IS_1 and IS_2, the authors focus on computing the semantic gap between their respective conceptual models without considering how to utilize the data cooperatively.

For clarity and simplicity, this paper divides semantic interoperability into two stages: (1) ***interpretation*** : automatic unambiguous information understanding [21–24] and (2) ***employment*** : automatically operate on the information according to the interpreted semantics. In this paper, we abstract an IoT system on the Internet as an information system (IS) or as a component of IS with information from outside (e.g., sensors or mobile phones or even manual input by users) as input and independent functionalities to provide services for users. To utilize semantic interoperability to handle information fusion among heterogeneous information systems (HIS), this paper models exchanged messages among IoT systems as tabular documents. This is because tabular documents have the ability to represent the meaning of most documents [21,24]. Thus, this paper studies document-based semantic interoperability to address information fusion problem among IoT-based information systems. As a result, two research problems are found, which are:

Problem 1 (*Consistent document understanding*): any tabular document (as semantic document [21]) shall be consistently understood between heterogeneous contexts (i.e., document sending party A and document receiving party B). However, in most cases, parties of document sending and receiving are often situated in different semantic communities [25]; that is, they may adopt different terms, document templates and business practices to conduct business. The outcome of inconsistent understanding of exchanged documents leads to wrong execution of documents (e.g., sales orders). Current approaches to this problem include domain-wide term association [26–28] and document standardization such as ebXML [29] and CCTS [30]. The former associates specific domains of terms with consistent semantics to facilitate consistent document understanding; however, it disregards the importance of cross-context document understanding, since document senders and receivers may have different interpretations on a same ontology due to context heterogeneity. The latter proposes common document standards for reading and writing documents. Its problem is the impracticality for all document contents in the world to be standardized.

Problem 2 (*Reasonable document usage*): any tabular document from Party A shall be executable by Party B and automatically generate more specific or previously not existed information for auxiliary decision making. The tabular document that Party B responses shall be interpretable by another user (e.g., Party A or C) without misunderstanding. Approaches in [31,32] implement this requirement by analyzing and representing documents in a logical form by using deontic and defeasible logic, transforming them into a machine readable rule notation (e.g., RuleML) as executable semantics. However, it may be too demanding for common users to represent documents in a logical form.

To resolve these two problems, this paper proposes an automatic tabular document exchange (DoxEx) framework which contains a new tabular document (TabDoc) model and a new semantic inference scheme. TabDoc model guarantees that the document writer, document writer's computer, document reader's computer and document reader of heterogeneous contexts share an identical understanding of any exchanged tabular documents. It contains basic types of semantic relations for tabular document representation and incorporates collaboratively designed common vocabulary CONEX Dictionary [21,33]. The semantic relations can be faithfully sent to the receiving parties by using types of attribute-value pairs in DocLang, an XML language designed for syntactic document representation, which has a specific processing procedure. CONEX dictionary acts as a common vocabulary that helps reduce the ambiguity of terminology used in document creation. Indeed, as an application to access and browse the CONEX dictionary that defines the data, it maps the received data elements into the vocabulary; if mappable, then it can understand them and perform appropriate processing (e.g., map to its local terms). Unlike SNOMED-CT dictionary used in clinical field [34], CONEX dictionary (served as global vocabulary) provides uniquely coded and unambiguous concepts across different domains. The tabular document created can use coded values to refer to concepts in CONEX dictionary. A code value must contain one code system (global vocabulary or local vocabulary) and a code number. Code system illustrates a concept from a particular source dictionary. Code number represents the unique identifier of a concept in a dictionary. Local dictionaries where any users can define their own specific terms must be associated to the CONEX dictionary. Thus, a tabular document is able to refer to several dictionaries to implement consistent tabular document understanding. In terms of document usage, this paper proposes a new semantic inference algorithm (*SIA*) that guides the inference engine towards a reasonable inference based on a rulebase to generate more comprehensive or previously not existed information (e.g., query result) and create new documents for further interaction. Semantic inference is an e-commerce phenomenon of reasoning from one action concept to a subsequent one between contextualized enterprises [8]. The source of rules comes from business rules and processes of document receiver parties as well as semantic relations of tabular documents. Finally, a case study shows that Table Designer system proposed based on DocEx framework is applied in the scenario of flight ticket booking. The results show that the method proposed in this paper is able to achieve information fusion among different information systems by using our semantic interoperability approach. When input information to IS comes from IoT devices, it is feasible to use CONEX dictionary as a vocabulary provider, TabDoc model as message model and semantic inference scheme as execution engine to realize information fusion among IoT-based information systems. The main contributions of this paper are:

- This approach improves the information fusion across different information systems through a new semantic interoperability approach consisting of consistent interpretation and reasonable employment of exchanged tabular documents.
- This approach contributes to the intelligence of tabular document understanding and processing by using a logic programming language and provides a base to develop intelligent applications.

The remainder of this paper is organized as follows. Section 2 discusses the related work. Section 3 presents an overview of DocEx framework. In Section 4, a methodology of semantic consistency maintenance is proposed. Section 5 describes a semantic inference scheme. Section 6 implements DocEx and conducts a case study. Section 7 compares the method used in this paper with other related ones. Finally, a conclusion is made to summarize this paper with a list of contributions and future work in Section 8.

2. Related work

2.1. Semantic interoperability

Currently, three methods are used to solve the two problems of semantic interoperability above: (1) ***Ontology mapping*** , using specific ontologies as terminology to construct information for exchange [11,18,35,36]. For example, [11] augments exchanged mes-