Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 3914 – 3923

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Using ART2 Neural Network and Bayesian Network for Automating the Ontology Constructing Process

Maryam Hourali[a], Gholam Ali Montazer[b,*]

[a,b]*IT Eng. Dept.,School of Engineering,Tarbiat Modares University, Tehran, Iran, P.O.Box:14115-179*

Abstract

Ontology is one of the fundamental cornerstones of the semantic Web. The pervasive use of ontologies in information sharing and knowledge management calls for efficient and effective approaches to ontology development. Ontology learning, which seeks to discover ontological knowledge from various forms of data automatically or semi-automatically, can overcome the bottleneck of ontology acquisition in ontology development.. In this article a novel automated method for ontology learning is proposed. First, domain-related documents were collected. Secondly, the C-value method was implemented for extracting meaningful terms from documents. Then, an ART neural network was used to cluster documents, and terms' weight was calculated by TF–IDF method in order to find candidate keyword for each cluster. Next, the Bayesian network and lexico-syntactic patterns were applied to construct the initial ontology. Finally, the proposed ontology was evaluated by expert's views and using the ontology for query expansion purpose. The primary results show that the proposed ontology learning method has higher precision than similar studies.

Keywords: Ontology; ART Neural Network; Term Frequency–Inverse Document Frequency (TF-IDF); C-value Method; Bayesian network; Lexico-Syntactic Patterns.

1. Introduction

Ontologies are defined as formal, explicit specifications of a shared conceptualization [1]. They are an essential component in many knowledge-intensive areas like the Semantic Web [2], knowledge management, and electronic commerce. The construction of domain ontologies relies on domain

* Corresponding author. Tel:+98 (21) 82883990,
E-mail address: montazer@modares.ac.ir.

1877-7058 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2012.01.594

modellers and knowledge engineers, which are typically overwhelmed by the potential size, complexity and dynamicity of a specific domain [3]. In consequence, the definition of exhaustive domain ontologies is a barrier that very few projects can overcome. Due to these reasons, nowadays, there is a need of methods that can tackle, or at least ease, the construction of domain ontologies. Automated Ontology Learning methods allow a reduction in the time. Zhou, presented a classification of ontology learning techniques which is composed of three categories: statistics-based, rule-based, and hybrid techniques. [4]. Shamsfard, classified the ontology learning approach to statistical and symbolic. The symbolic approaches are the logical, linguistic-based and template-driven approaches. Heuristic methods may be used to facilitate each approach [2].

In this paper we combined different ontology learning methods such as: statistical, linguistic, and pattern-based for proposing a novel automated ontology learning system in query expansion domain. For this purpose, natural language processing methods such as: depletion of stop words, linguistic processing, statistical processing, etc was implemented on domain related documents set. Next, by using C-value method, we extracted main terms, and construct term-document matrix respectively. Then, we clustered this matrix with ART neural network. In order to choose a candidate for each cluster, we implemented the TF-IDF method and selected the term with the highest weight as the candidate. Furthermore, the Bayesian network was implemented for constructing ontology hierarchy. Then we used lexico syntactic patterns to extract non taxonomic relations and ontology instances. Hence, the initial ontology was constructed and evaluated by domain related experts. The remainder of this paper is organized as follow: section 2, describes the architecture of ontology learning system step by step accompanied by associated empirical results. In section 3, performance evaluation of the system is described. The paper makes conclusion and suggests directions for future research in section 4.

2. Ontology Construction

The constructing ontology process comprises four steps: (1) analyzing documents, (2) clustering documents, (3) learning ontology relations, and (4) Learning ontology instances. The proposed architecture of ontology learning system is shown in Fig.1, and the process is described in detail bellow:

2.1. *Analyzing Documents*

The first step to construct ontology in a certain domain such as information technology domain, is collecting the domain related documents. We deleted the duplicated articles and reached to 3345 articles as a document set for analysis. Secondly, C-value method in linguistic and statistical parts was used to extract main terms for ontology construction. The linguistic part consists of the part of speech (POS) tagging of the corpus, the linguistic filter, and the stop list. So, (Adj|Noun)+Noun filter was used to identify those terms that contain a noun or contain an adjective that accompanied with a noun, such as: 'applied computer science', 'Internet technology'. Our stop list included some high frequency words such as: great, numerous, several, year, just, good, etc.

To some up, 3700 terms with a considerable amount of C-value in range [1,3] were extracted (about 67 percent of total terms). For increasing the accuracy and depletion the noisy inputs, we considered those terms which were existed in 3 different and separate document [13]. Finally, 1220 main terms were selected in constructing the ontology. Samples of these terms are shown in Table 1. Extracted terms were located in a term-document matrix. In this matrix, rows indicate the document sets and columns indicate the extracted terms. Matrix's cells are normalized value of TF-IDF.