An Acquisition Model of Deep Textual Semantics Based on Human Reading Cognitive Process

Jun Zhang, Shanghai University, China
Xiangfeng Luo, Shanghai University, China
Lei Lu, Shanghai University, China
Weidong Liu, Shanghai University, China

ABSTRACT

The acquisition of deep textual semantics is a key issue which significantly improves the performances of e-learning, web search and web knowledge services, etc. Though many models have been developed to acquire textual semantics, the acquisition of deep textual semantics is still a challenge issue. Herein, an acquisition model of deep textual semantics is developed to enhance the capability of text understanding, which includes two parts: 1) how to obtain and organize the domain knowledge extracted from text set and 2) how to activate the domain knowledge for obtaining the deep textual semantics. The activation process involves the Gough mode reading theory, Landscape model and memory cognitive process. The Gough mode is the main human reading model that enables the authors to acquire deep semantics in a text reading process. Generalized semantic field is proposed to store the domain knowledge in the form of Long Term Memory (LTM). Specialized semantic field, which is acquired by the interaction process between the text fragment and the domain knowledge, is introduced to describe the change process of textual semantics. By their mutual actions, the authors can get the deep textual semantics which enhances the capability of text understanding; therefore, the machine can understand the text more precisely and correctly than those models only obtaining surface textual semantics.

Keywords: Domain Knowledge, Gough Mode, Landscape Model, Reading Process, Semantic Field

INTRODUCTION

Textual semantics can be divided into two types, i.e., surface semantics and deep semantics. The former only contains textual keywords, keyword associations, sentence patterns and even the

DOI: 10.4018/jcini.2012040105
structure of the article which is from the text itself. And the latter, a type of implicit semantics not only contained in text but hidden in background knowledge, is constituted by the domain knowledge of text, and is necessary for a machine or a human being to better understand a text.

The acquisition of deep textual semantics is a key issue on the text understanding process which can significantly improve the performances of e-learning, web search and web knowledge services. Though many models have been proposed to acquire textual semantics, it is still a challenge issue because the current models only extract the surface semantics from the text itself.

There are four main types of surface textual semantic acquisition models; 1) Statistics models, such as vector space model (VSM) (Salton & Wong, 1975); 2) Cognition based models, such as element fuzzy cognitive map (EFCM) (Zhuge & Luo, 2006), concept algebra based model (Wang, 2006, 2007b) and associated linked network model (ALN) (Luo & Xu, 2010); 3) Probability topic models, such as author topic mode (ATM) (Rosen-Zvi, 2004), author-recipient-topic model (ART) (McCallum & Wang, 2004) and correlated topic models (CTM) (Blei & Lafferty, 2006); and 4) Ontology based models, such as ontology inference layer (OIL) (Fikes & McGuinness, 2001) and ontology web language (OWL) (Smith, 2002). However, the most models are based on the text itself without considering the domain knowledge. Unfortunately, there is much deep textual semantics excluded in text, such as the textual background (i.e., a part of domain knowledge). The machine or human beings would be hard to understand the text without the help of background knowledge, because the absence of the domain knowledge causes the incoherence of textual surface semantics and the bad intelligibility of text. Therefore, compared with the deep textual semantics, the surface semantics is rough, limited, narrow, and lack of coherence and cohesion for text understanding process.

There are several models having been proposed to acquire the deep textual semantics, such as “WordNet” and “HowNet” (Tang, 2007), which are stiff and inflexible in the acquisition process of deep textual semantics. From current studies, we know that there exist two key issues in the acquisition process of deep textual semantics (Luo & Lu, 2011); 1) How to obtain and organize the domain knowledge extracted from domain text set and 2) how to activate the domain knowledge for obtaining the deep textual semantics. The acquisition of domain knowledge extracted from domain text set based on human reading cognitive process has been studied by Luo and Cai (2010).

Besides, there are some other cognitive studies related with the acquisition process of domain knowledge. For instance, the brain research has focused on how the brain processes internal and external information autonomously and cognitively rather than imperatively as those of conventional computers (Wang, 2011); the latest advances and engineering applications of Cognitive Informatics (CI) have led to the emergence of cognitive computing and the development of cognitive computers that perceive, learn, and reason (Wang, 2003, 2007a, 2008, 2011; Wang et al., 2006, 2009). On the basis of the abstract syntax, a universal language processing model and the deductive grammar of English are developed toward the formalization of Chomsky’s universal grammar in linguistics (Wang, 2009).

In this paper, our work focuses on how to combine the landscape model (Broek, 1998) and the memory system model of human reading cognitive process to activate the domain knowledge in order to acquire the deep textual semantics.

This paper is organized as follows. The reading mode, landscape model and the memory system model of human reading process are introduced first. Then we present how to construct Long Term Memory (LTM) to store the domain knowledge of text. After that, the activation
Organizational Cognition in the Industrial Case Study

www.igi-global.com/chapter/organizational-cognition-industrial-case-study/27877?camid=4v1a