The Impact of Ontology on the Performance of Information Retrieval:
A Case of WordNet

Maria Indrawan, Monash University, Australia

Seng Loke, La Trobe University, Australia

ABSTRACT

The debate on the effectiveness of ontology in solving semantic problems has increased recently in many domains of information technology. One side of the debate accepts the inclusion of ontology as a suitable solution. The other side of the debate argues that ontology is far from an ideal solution to the semantic problem. This article explores this debate in the area of information retrieval. Several past approaches were explored and a new approach was investigated to test the effectiveness of a generic ontology such as WordNet in improving the performance of information retrieval systems. The test and the analysis of the experiments suggest that WordNet is far from the ideal solution in solving semantic problems in the information retrieval. However, several observations have been made and reported in this article that allow research in ontology for the information retrieval to move towards the right direction.

Keywords: ontology; semantic information retrieval; WordNet

INTRODUCTION

Semantic understanding is crucial to the success of many information technology applications. Much information technology research is still battling to solve the problem of semantic understanding for their research domain. Ontology adoption is currently the most popular approach taken by many researchers. The proliferation in the use of ontology to support semantic analysis has been found in many domains of information technology such as context awareness (Rack, Arbanowski, & Steglich, 2000; Yan & Li, 2006), service oriented computing (Bramantoro, Krishnaswamy, & Indrawan, 2005; Jingshan, Hunhns, 2006), and Semantic Web (Caliusco, Galli, & Chiotti, 2005; Dou, LePendu, Kim, & Qi, 2006). Some of the researchers adopt a specific built ontology whereas others investigate the use of a general purpose ontology, such as WordNet.

WordNet is an English lexical referencing system built in the early 1990s at Princeton University. Since its introduction, many researchers have used this lexical system for different purposes, such as multimedia retrieval (Benitez, Chang, & Smith; 2001), text summarization (Hachey & Grover, 2004), and automatic creation of domain-based ontology (Chen, Alahakoon, & Indrawan, 2005; Khan & Luo, 2002). In information retrieval research, the impact of WordNet has been investigated by a number of researchers. WordNet has been used to improve the performance of information retrieval systems by way of query expansion (Voorhees, 1993), semantic distance measure (Richrdson & Smeaton, 1995), and semantic indexing (Wang & Brookes, 2004) to name a few. The results showed by these studied are varied. Voorhees (1993) and Richrardson and Smeaton (1995) report that the recall and precision of the retrieval decreased with the inclusion of WordNet. Wang and Brookes (2004), on the other hand, report the opposite. We were encouraged by Wang and Brookes' report and decided to investigate further since we perceived a further improvement can be applied to their model. In addition, we also would like to explore the debate over the impact of WordNet in information retrieval researches. At the end of the investigation we would like to enrich the debate by reporting our experience and observations during the investigation. In order to achieve this, we organize this article as follows. In the next section, the article presents a short description of WordNet for those readers unfamiliar with this lexical system. In the third section, we lay out the current debate on the impact of WordNet in information retrieval. We introduce our improvement to Wang and Brookes' model in the forth section. The following section presents the experiment design and results. We conclude our discussion in the last section.

WordNet

The main construct of WordNet as a lexical system is the synonym set or synset. The synsets are divided into four major speech categories of noun, verb, adjective, and adverb. Within each of these categories, several semantic relations between synsets are defined. Included in the noun category are the *hypernym, hyponym, meronym,* and *holonym*.

Definition 1: *Semantic Relations of Synsets*

Let assume synsets $S=\{s_i, s_j, ..., s_n\}$ and $L=\{l_i, l_j, ..., l_n\}$ exist in the WordNet.

Hypernym: S is considered to be hypernym of L, if every L is a (kind-of) S.

Hyponym: S is considered to be a hyponym of L, if every S is a (kind-of) L.

*Meronym*: S is considered to be a meronym of L, if every S is a part-of L.

Holonym: S is considered to be a holonym of L, if every L is a part-of S.

□

In an example of taxonomy of synsets depicted in Figure 1, a *canine* is considered to be a hyponym of *carnivore* and a hypernym of *dog*.

Figure 2 shows that an *engine* is a meronym of a *car* and a holonym of a *piston*. The hypernym/hyponym relations are often referred as *hyponomic* relations. The meronym/holonym relations are referred as *part-whole* relations.

Figure 1. A hypernym/hyponym relations
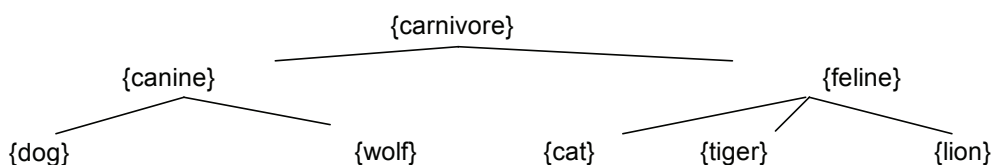
Related Content

Sports Work Strategy of College Counselors Based on MySQL Database Big Data Analysis

Xiao Zhang, Ali Yu, Xin Wangand Xue Zhang (2023). *International Journal of Information Technology and Web Engineering (pp. 1-14).*

www.irma-international.org/article/sports-work-strategy-of-college-counselors-based-on-mysql-database-big-data-analysis/332788

Query Log Analysis for Adaptive Dialogue-Driven Search

Udo Kruschwitz, Nick Webband Richard Sutcliffe (2009). *Handbook of Research on Web Log Analysis (pp. 389-414).*

www.irma-international.org/chapter/query-log-analysis-adaptive-dialogue/22012

Integrating Sustainability in Cloud Computing for Managing Sustainable Knowledge in Higher Education

Ahmed Mehrez, Ki-Hoon Leeand Fawzy Soliman (2016). *Web-Based Services: Concepts, Methodologies, Tools, and Applications (pp. 56-66).*

www.irma-international.org/chapter/integrating-sustainability-in-cloud-computing-for-managing-sustainable-knowledge-in-higher-education/140795

Security in Network Layer of IoT: Possible Measures to Preclude

B. Balamuruganand Dyutimoy Biswas (2017). *Security Breaches and Threat Prevention in the Internet of Things (pp. 46-75).*

www.irma-international.org/chapter/security-in-network-layer-of-iot/177064

The Economic Development Model and Strategy Selection of the Internet of Things Based on Big Data

Xuelin Chen, Xia Zhaoand Na Ye (2023). *International Journal of Information Technology and Web Engineering (pp. 1-16).*

www.irma-international.org/article/the-economic-development-model-and-strategy-selection-of-the-internet-of-things-based-on-big-data/330682