An Automated Learner for Extracting New Ontology Relations

Amaal Saleh Hassan Al Hashimy

Sultan Qaboos University, College of Science, Computer Science Department
Oman
amaalh@squ.edu.om

Narayanan Kulathuramaiyer

University of Malaya Sarawak UNIMAS, Sarawak, Malaysia
nara@fit.unimas.my

Abstract— **Recently, the NLP community has shown a renewed interest in automatic recognition of semantic relations between pairs of words in text which called lexical semantics. This approach to semantics is concerned with psychological facts associated with the meaning of words. Lexical semantics is an important task with many potential applications including but not limited to, Information Retrieval, Information Extraction, Text Summarization, and Language Modeling. As this task "automatic recognition of semantic relations between pairs of words in text" can be used in many NLP applications, its implementation are demanding and may include many potential methodologies. And as it includes semantic processing, the results produced still need enhancements and the outcome was always limited in terms of domain or coverage.**

In this research we developed a buffered system that handle the whole process of extracting causation relations in general domain ontologies. The main achievement of this work is the heavy analysis of statistical and semantic information of causation relation context to generate the learner. The system also builds relation resources that made it possible to learn from itself, were each time it runs the resources incremented with new relations information recording all the statistics of such relation, making its performance enhanced each time it runs. Also we present a novel approach of learning based on the best lexical patterns extracted, besides two new algorithms the CIA and PS that provide the final set of rules for mining causation to enrich ontologies.

Keywords-lexical relations, mining causation relations, ontology learning.

I. INTRODUCTION

The notion of ontology now a day is a dominant research area in the field of computer science. While this gain its main features as a "world representation scheme" from the philosophy in past, it is now gaining specific role in AI, computational linguistics, and DB theory.

Ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It is used to reason about the entities within that domain, and may be used to describe the domain. In theory, ontology is an "explicit specification of a shared conceptualization"[11].

In recent years, the acquisition of ontologies from domain texts using machine learning and text mining methods has been proposed as a means of facilitating the ontology engineering process. In this context, ontology learning has been identified as an emerging field which aims at assisting knowledge engineers as well as end-users in ontology construction (figure 1). It can be seen as a multi-disciplinary field, which integrates disciplines such as ontology engineering, machine learning, and natural language processing, among others. The use of these technologies is distributed in three main phases, lexical entry extraction, taxonomy extraction, and non-taxonomic relation extraction [13].

The next sections will explore related works and provide insights on approaches in handling lexical semantics for the purpose of ontology learning. Subsequently we propose an enhanced framework for ontology learning.
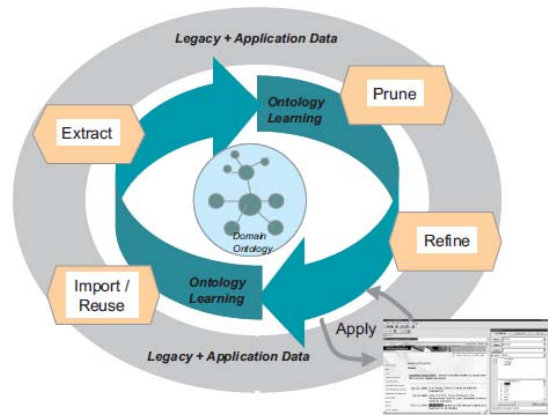

Figure1. Sample of ontology learning process
(Source: Maedche et.al. 2005)

II. RELATED WORK

Developing systems based on this kind of information is not new. Many systems have been developed Garcia in 1997 used verbs as causal indicators for causal knowledge acquisition in French. Khoo in [10] acquired causal knowledge with manually created syntactic patterns specifically for the MEDLINE text database. Works by Girju in [8] explored the acquisition of causal knowledge by using connective markers.

Ontologies wide-spread usage is still hindered by ontology acquisition being rather time consuming and, hence, expensive [3]. A number of proposals have been made to facilitate ontological acquisition through automatic discovery from