Integrating Knowledge Sources: An Ontological Approach

Hayden Wimmer, Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA

Victoria Yoon, Department of Information Systems, School of Business, Virginia Commonwealth University, Richmond, VA, USA

Roy Rada, Department Of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA

ABSTRACT

There has been a plethora of research in the area of knowledge portals, knowledge warehouses, ontologies, ontology creation and mapping, as well as the automatic creation and mapping of ontologies. While research exists in each respective area there is a lack of conceptual models that will integrate ontologies as a tool for disparate data source integration into knowledge portals or enterprise knowledge warehouses. The purpose of this work is to discuss different tools that have been developed in academic research and provide a conceptual model of how to implement these tools in relation to knowledge portals and warehouses.

Keywords: Knowledge Integration, Knowledge Portal, Knowledge Warehouse, Ontologies, Ontology Creation, Ontology Mapping

INTRODUCTION

A proliferation of Information Systems has created excessive amounts of data in organizations. Unfortunately, there are many different types of systems that are in proprietary formats or have incompatible schemas and data types, resulting in a multitude of disparate data sources in organizations. This has caused an issue that plagues researchers and organizations – integration of heterogeneous data sources (Kaza & Chen, 2008). Enterprise knowledge warehouses (EKW) and knowledge portals (KP) that fuse knowledge from multiple sources suffer from the same integration issues.

Ontologies may be used to support integration (Narock, Yoon, Merka, & Szabo, 2010). Ontologies may be employed as the backbone to assist in the development and maintenance of knowledge portals (Staab, 2001). Manually creating and mapping ontologies is burdensome and monotonous. This manual process also requires the time and experience of domain experts which is difficult to procure. Further, this manual work is prone to error due. There has been promising research in automatically

DOI: 10.4018/jkm.2013010104
creating and mapping ontologies (Doan et al., 2003; Kaza & Chen, 2008) which is the most viable direction for solving integration issues. Another viable alternative that often works with ontologies to integrate heterogeneous data and systems is multi-agent systems. Agents can be mediators between heterogeneous data sources, providing means to interoperate using ontologies for describing the data contained in the sources, and communicating with each other using a common agent communication language. Such multi-agent systems have been applied to knowledge management (Yoon, Broome, Singh, & Guimaraes, 2005) and have gained acceptance when integrating disparate data (Vizacaino, Soto, Portillo-Rodrigues, & Piattini, 2007).

Prior research exists in knowledge portals, ontologies, and multi-agent systems (Smith, 2004); however, there are few conceptual models that attempt to integrate these research areas. Also, there has been little research in employing available automated methods to assist in ontological-based data integration. A research issue is how to use ontologies and the existing automatic mapping and creation methods to integrate external sources of knowledge with internal knowledge portals. The purpose of this paper is to fill a gap in the literature by providing a conceptual framework which incorporates the two key technologies: ontology and multi-agent systems. The conceptual framework is designed to integrate internal and external knowledge sources to respond to user queries. Applying the key technologies on multiple levels, the framework will first create ontologies for external sources using automatic ontology mapping methods and second map those ontologies to an internal global ontology.

The organization of this paper is as follows. First, background information will be provided. Second, a literature review on existing methods for automatic ontology creation and mapping will be discussed. Followed is our proposed conceptual architecture for utilizing ontologies to integrate data sources. This will include a proposed research methodology and finally conclusions and future research directions will be presented.

**BACKGROUND INFORMATION**

There are various sources of knowledge in organizations. There are two knowledge sources that are common. The first is internal knowledge. Internal knowledge is in the form of organizational systems. These systems may include – but are not limited to – internal databases, legacy systems, ERP systems, knowledge warehouses, data warehouses, OLAP systems, and unstructured or semi-structured data in the form of system documentation or standard operating procedure documents. Internal databases may be Oracle or MS SQL relational databases. Legacy systems are systems that have been long engrained in organizations. ERP - or enterprise resource planning - based systems provide support for all business processes. Knowledge warehouses and data warehouses are systems that store knowledge and data respectively; however, these systems may not store data in a relational format. Knowledge warehouses (KW) have been implemented in a variety of industries with the legal profession being an avid consumer of knowledge warehouses such as Westlaw (“Westlaw,” 2012). Knowledge warehouses and data warehouses are typically in a multi-dimensional format. Many organizations have a variety of enterprise resources which they feel they are unable to leverage. Knowledge warehouses may be employed to assist in this process (Dang & Yuan, 2011). Knowledge portals are also sources of knowledge which may be internal, external, or integrate internal and external knowledge (Neubauer, 2009). OLAP systems – or online analytical application processing systems – attempt to provide business intelligence to support organizational decisions. Finally, structured data would be in the form of a database or semi-structured as a spreadsheet. An example of unstructured data is word documents or other textual data. Internal sources are easier to control and maintain than external sources. Internal sources can be controlled by the organization. The organization can utilize standards and homogeneous environments.

Arrays of external sources of knowledge are also available to an organization. External sources may be from the internet and include
Virtue-Nets
www.igi-global.com/chapter/virtue-nets/49098?camid=4v1a