

Abstract

With the popularity of Internet/Intranet, heterogeneous information integration becomes a hot IT topic in electronic business (EB) field. Heterogeneous information integration on the Web involves a number of new techniques. There have been research projects applying XML and ontology as mediated techniques to consolidate heterogeneous information. In order to manage and use information more effectively within the enterprise, a benchmark used to evaluate the mechanism of heterogeneous information integration is needed. In this research, we develop a XML and ontology benchmark workload model in heterogeneous information integration, and build a workload generation prototype. The objective of this research is to develop a workload model combines XML and ontology to test whether the heterogeneous information integration system under EB environment can overcome the diverse formats of content and derive meaning from this content. The workload model consists of XML and ontology data model and query model according to the generic data structure and query functionality. Also, a control model is created to set up the benchmark environment. In order to apply the workload model to different scenarios easier, this workload model is designed to be domain independent and generic-construct-based. Finally, we validate the research model through the prototype implementation.

Keywords: XML, Ontology, Heterogeneous Information Integration, Benchmark, Workload Model, Performance Evaluation

中文摘要

隨著網際網路和企業內部網路的盛行，異質資訊整合成為電子化企業中一個重要的議題，在網路上進行異質資訊整合涉及許多不同新的資訊技術，目前已經有些研究試圖利用延伸標記語言以及本體論當作中介技術來整合異質資訊，為了有效管理企業內的資訊，我們需要一個績效評估模型來衡量異質資訊整合的效能。在本研究中，我們提出了一個在異質資訊整合中運用延伸標記語言及本體論的績效評估工作量模型，並且建立了一個工作量產生器雛形；本研究的目標是希望發展出一個結合延伸標記語言及本體論的工作量模型，以測試在電子化企業中的異質資訊整合是否能整合不同的資訊模型，並且從這些資訊模型中衍生出語意，此工作量模型包含了延伸標記語言與本體論的資料模型與查詢模型，它們是依照延伸標記語言與本體論學名式的資料結構與查詢功能所制訂的，此外，控制模型則定義了績效評估執行環境中所需設定的變數，為了讓此工作量模型能具可攜性和延展性，以便輕易地應用在不同的領域情境中，本研究採取學名結構式且使用者定義、領域獨立的設計方法，最後，我們利用雛形實作來驗證本研究所提出的研究方法。

關鍵字：延伸標記語言，本體論，異質資訊整合，績效評估，工作量模型