數位圖書館整合之詮釋資料架構

學生：黃夙賢　　指導教授：楊維邦 博士
柯皓仁 博士

國立交通大學資訊科學與工程研究所 博士班

摘　　要

詮釋資料在數位圖書館整合中扮演重要的角色。然而傳統的詮釋資料模型並未考慮語義上的異質性以及資料結構化等因素，所以無法解決分散式環境下數位圖書館整合的問題。因此本論文從詮釋資料的角度出發，提出了一套用來整合數位圖書館的詮釋資料架構。本詮釋資料架構將數位圖書館整合分成描述、擷取以及查詢三層次，使得在數位圖書館查詢的過程中能夠保留更多的知識，讓查詢能夠節省更多時間。並且期望能解決格式、傳輸協定以及語意上不一致性的問題。

本詮釋資料架構包含三層。第一層，詮釋資料描述層。本層提出了新的詮釋資料描述語言，用來表示數位圖書館的資源。本詮釋資料描述語言擁有標準的語法，並且延伸 RDF 的描述功能。在本層當中，採用了兩種結構表示法來描述詮釋資料的結構。並且透過這兩種結構表示法，來達成詮釋資料之間的互相轉換，進而解決格式異質性的問題。

本詮釋資料第二層為資料萃取層。本層負責從分散式數位圖書館中擷取資訊，並且包裹成詮釋資料。本層採取的方式，是根據數位圖書館查詢系統產生資料的一致性，找出共同結構的部份，再根據共同結構自動抓取同一數位圖書館所產生的其他資料。第一步是將查詢系統資料編定階層號碼，接下來定義要抓取資料的階層號碼範圍，於是同一個查詢服務的資料，就可以根據定義好的階層號碼範圍抓出來。本層的特性是不需要跟

欲查詢的數位圖書館事先做溝通，即可透過網頁查詢系統自動擷取資料，因此解決通訊協定異質性的問題。

第三層，語意查詢層。本層根據數位圖書館兩個重要的組成份子：內容與服務，推導出十五種的語意關係，並將十五種語意關係應用在語意查詢當中，讓使用者在查詢數位圖書館的時候能夠獲得更多的語意資料。無論是服務查詢或者是內容查詢，都能夠透過加入語義關係函數而獲得更多正確的資料，進而解決語義異質性的問題。

實驗結果顯示採用本詮釋資料架構能夠使得數位圖書館整合獲得良好的成效。相對於傳統的關鍵字查詢系統，本論文以詮釋資料架構為基礎的查詢系統能夠在正確性以及涵蓋率都能夠獲得實質的改進。採用本詮釋資料架構能夠減輕系統管理者的負擔，當程式設計師開發數位圖書館整合性服務，例如虛擬聯合目錄或圖書館資源分配時，本架構能夠建議圖書館員如何利用現有的程式來開發新的服務。這種充分利用數位圖書館現有資源的技術正好是未來數位圖書館整合所發展的趨勢。

關鍵字：*數位圖書館整合，詮釋資料模型，資料擷取，語意查詢，語意推論*

Metadata Architecture for Digital Library Integration

Student: Su-Hsien Huang Advisor: Dr. Wei-Pang Yang

Dr. Hao-Ren Ke

Institute of Computer Science and Engineering

National Chiao Tung University

ABSTRACT

Metadata has been playing an essential role in integrating heterogeneous digital libraries (DL). However, conventional metadata architecture is insufficient to achieve interoperability among DL because of the heterogeneity in semantics and no structure consideration in metadata formats. This dissertation proposes novel metadata architecture called *M*-Architecture@DL to integrate DL seamlessly from the perspective of metadata. *M*-Architecture@DL follows Model-Extraction-Query (MEQ) model to obtain more permanent and explicit knowledge in the process of DL query. *M*-Architecture@DL contains three layers, namely metadata modeling layer, data extraction layer, and semantic query layer. The separation of *M*-Architecture@DL into three-layer achieves format, protocol and semantic interoperability in each layer.

Metadata modeling layer uses Metadata Modeling Language (MML) to describe real-world entities. MML adopts XML as its syntax and extends Resource Description Framework (RDF) by adding name hierarchy reference. MML provides two constructors, tuple and set constructors, to represent structures. With these two constructors, metadata can be translated by manipulating attributes of metadata with operations. In this layer, the format interoperability is achieved.

Data extraction layer collects data from distributed DL and encapsulates result into MML metadata. Data from DL services with similar structure can be extracted into metadata automatically by means of the common structure. In the process of extraction, the first step is to assign level ID for the sample document and determine the common part to be extracted. Then an extraction algorithm called Metadata Extractor is implemented to extract the documents according to the common structure. This layer provides a transparent way without prearrangement with distributed DLs and saves much effort to collect information through the HTTP protocol. Therefore, the protocol interoperability is achieved.

Semantic query layer retrieves metadata semantically by adding relationships in query statements. A Content and Service Inference Model (CSIM) is proposed to derive 15 relationships from two essential aspects of DL: content and services. The 15 structural relationships create operations to manipulate metadata in a query predicate and facilitate a query with as much semantics. Both content and service queries are presented to derive more semantic answers in a DL search. In this layer, the semantic interoperability is achieved.

Experiments are conducted and indicate that *M*-Architecture@DL has excellent performance in DL integration. The experiment results have shown that both accuracy and coverage are improved to a conventional keyword-based approach. Adopting *M*-Architecture@DL can alleviate the administrative load. When developing novel DL services, such as library resource planning and virtual union catalog system, librarians are recommended with alternative answers to combine existent DL components. The reuse of DL services and metadata is the future trend in DL integration.

Keywords*: digital library integration, metadata model, data extraction, semantic query, semantic inference*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES