

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ
TRUYỀN THÔNG

ĐỖ TRUNG THÀNH

TÍCH HỢP CSDL QUAN HỆ XML

LUẬN VĂN THẠC SĨ
KHOA HỌC MÁY TÍNH

THÁI NGUYÊN – 2016

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ
TRUYỀN THÔNG

ĐỒ TRUNG THÀNH

TÍCH HỢP CSDL QUAN HỆ XML

Chuyên ngành: khoa học máy tính

Mã số: 60.48.01.01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. ĐOÀN VĂN
BAN

THÁI NGUYÊN - 2016

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

MỞ ĐẦU

Chuyển đổi CSDL quan hệ sang XML và ngược lại giữ một vai trò quan trọng. Việc nghiên cứu, xây dựng thuật toán chuyển đổi CSDL quan hệ sang XML và ngược lại là việc làm cần thiết, thực hiện tốt các thuật toán này sẽ mang lại giá trị cao trong thực tiễn ngành Công nghệ Thông tin trong giai đoạn hiện nay.

XML đã và đang được sử dụng rộng rãi trong thực tế, đặc biệt với sự phát triển sôi động, mạnh mẽ của công nghệ cảm ứng trên điện thoại di động và máy tính bảng (dữ liệu XML được ứng dụng nhiều trong lĩnh vực này do lợi thế việc sử dụng, tiếp cận, xử lý và lưu trữ dữ liệu XML không cần đến các hệ quản trị CSDL). Chính vì vậy, việc xây dựng và phát triển các ứng dụng với dữ liệu XML đã và đang giữ một vai trò quan trọng. Việc khai thác các phương pháp tiếp cận, lưu trữ, xử lý tốt dữ liệu XML cũng sẽ mang lại nhiều lợi ích cho các hệ thống thông tin trong bối cảnh ngành Công nghệ Thông tin trong giai đoạn hiện nay.

Mô hình tích hợp nhiều hệ quản trị CSDL quan hệ khác nhau cùng hoạt động trên một hệ thống đang ngày càng trở lên phổ biến. Việc chuyển đổi qua lại giữa các CSDL quan hệ thông qua XML hoặc trích xuất dữ liệu theo yêu cầu của người sử dụng dưới dạng XML sẽ giữ một vai trò quan trọng của hệ thống này.

Đề tài “**Tích hợp CSDL quan hệ và XML**” gồm các nội dung nghiên cứu chính với bố cục như sau:

CHƯƠNG 1. TỔNG QUAN VỀ CSDL VÀ XML

Nghiên cứu lý thuyết cơ bản về CSDL quan hệ; phân tích, đặc tả, tiếp cận và xử lý dữ liệu XML.

CHƯƠNG 2. TÍCH HỢP CSDL VÀ XML

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

Nghiên cứu mối quan hệ giữa CSDL quan hệ với XML; các thuật toán chuyển đổi giữa CSDL quan hệ và dữ liệu XML.

CHƯƠNG 3. ỨNG DỤNG TÍCH HỢP CSDL VÀ XML

Cài đặt các thuật toán, phương pháp xử lý dữ liệu XML trên cơ sở chương 1 và chương 2 đã trình bày; cài đặt ứng dụng quản lý điểm (trên môi trường Windows Form và Asp.Net), Website tin tức (trên môi trường Asp.Net) với dữ liệu XML và một số hệ quản trị CSDL phổ biến như Sql Server, MySql, MS Access.

CHƯƠNG 1 TỔNG QUAN VỀ CSDL VÀ XML

1.1. CSDL quan hệ

1.1.1. Khái niệm về CSDL và hệ quản trị CSDL

a) Cơ sở dữ liệu (CSDL)

Cơ sở dữ liệu được xác định như là một bộ sưu tập các dữ liệu có liên quan logic với nhau; nó được tổ chức, sắp xếp theo một cách nào đó và được các hệ ứng dụng của một đơn vị/cơ quan cụ thể nào đó sử dụng[5].

b) Hệ quản trị CSDL (DBMS)

Hệ quản trị CSDL là hệ thống bao gồm nhiều phần mềm cho phép mô tả, lưu giữ, thao tác các dữ liệu trên một CSDL. Nó đảm bảo tính an toàn, bí mật của dữ liệu trong môi trường nhiều người sử dụng.

1.1.2. Kiến trúc CSDL quan hệ

Một cơ sở dữ liệu quan hệ là một tập của một hoặc nhiều quan hệ, trong đó mỗi quan hệ là một bảng hai chiều bao gồm các cột và các hàng (gọi là bảng dữ liệu hay quan hệ). Bảng dữ liệu chính là

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

hình thức thể hiện cụ thể của kiểu thực thể khi chúng ta xây dựng cơ sở dữ liệu, bảng dữ liệu được sử dụng để lưu dữ liệu về các thực thể trong lớp thực thể đó.

a) Khái niệm miền giá trị

Miền giá trị của bảng quan hệ là miền xác định các giá trị thuộc tính trong quan hệ.

b) Khung nhìn

Khung nhìn là một bảng tạm thời, có cấu trúc như một bảng, khung nhìn không lưu trữ dữ liệu mà nó được tạo ra khi sử dụng, khung nhìn là đối tượng thuộc CSDL.

1.1.3. Ràng buộc toàn vẹn CSDL quan hệ

Ràng buộc toàn vẹn và kiểm tra sự vi phạm ràng buộc toàn vẹn là một trong những vấn đề quan trọng trong quá trình phân tích, thiết kế và khai thác các hệ thống cơ sở dữ liệu.

a) Định nghĩa

Ràng buộc toàn vẹn là một điều kiện bất biến không được vi phạm trong một cơ sở dữ liệu.

b) Các yếu tố của ràng buộc toàn vẹn

Tính ràng buộc toàn vẹn dữ liệu được biểu diễn thông qua các phụ thuộc hàm và cụ thể là dựa trên ràng buộc khóa.

c) Chuẩn hóa một CSDL quan hệ

Chuẩn hóa một CSDL có nghĩa là phải biến đổi, đưa các lược đồ quan hệ về một trong các dạng chuẩn: 1NF, 2NF, 3NF, BCNF.

1.2. XML (eXtensible Markup Language)

XLM (viết tắt từ tiếng Anh eXtensible Markup Language,

Số hóa bởi Trung tâm Học liệu – ĐHTN

<http://www.lrc.tnu.edu.vn>

“Ngôn ngữ đánh dấu mở rộng”) là ngôn ngữ đánh dấu với mục đích chung cho W3C đề nghị để tạo ra các ngôn ngữ đánh dấu khác. Đây là một tập con đơn giản của SGML, có khả năng mô tả nhiều loại dữ liệu khác nhau. Mục đích chính của XML là đơn giản hóa việc chia sẻ dữ liệu giữa các hệ thống khác nhau, đặc biệt là các hệ thống được kết nối với Internet. Các ngôn ngữ dựa trên XML (chẳng hạn như: RDF, RSS, MathML, XHTML, SVG, GML và cXML) được định nghĩa theo cách thông thường, cho phép các chương trình sửa đổi và kiểm tra hợp lệ bằng các ngôn ngữ này mà không cần có hiểu biết trước về hình thức của chúng.

1.2.1. Các thành phần cơ bản của XML

- a) *Định dạng và dữ liệu ký tự*
- b) *Các khai báo trong thẻ XML*
- c) *Phần chú thích*
- d) *Chỉ thị xử lý*
- e) *Thẻ và các phần tử*
- f) *Thuộc tính*
- g) *Không gian tên gọi*

1.2.2. Đặc tả cấu trúc dữ liệu XML

Một tài liệu XML có thể được chia thành 2 thành phần chính, mỗi thành phần có các quy định khác nhau.

- Phần khởi đầu bao gồm các khai báo như: phiên bản sử dụng của XML, cách thức mã hóa dữ liệu, chỉ thị xử lý, định nghĩa kiểu dữ liệu cho tài liệu XML, các chú thích, các khoảng trắng,...

- Phần thân của tài liệu chứa nội dung của dữ liệu bao gồm một hay nhiều thành phần, tối thiểu phải có ít nhất một thành phần,

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

những thành phần này bao gồm các phần tử. Mỗi phần tử bao gồm một cặp thẻ gồm thẻ bắt đầu và thẻ kết thúc. Phần tử đầu tiên của tài liệu được coi là phần tử gốc. Tất cả các tài liệu XML được gọi là hợp khuôn dạng nếu có chứa duy nhất một phần tử gốc. Phần tử gốc chứa tất cả các phần tử và các cặp thẻ khác nhau.

Một tài liệu XML được coi là một tài liệu hợp khuôn dạng nếu tuân theo các quy tắc sau:

- Các khai báo XML phải được đặt ở dòng đầu tiên của tài liệu. Khai báo chỉ thị xử lý XML đầu tiên không bắt buộc, tuy nhiên nếu sử dụng thì phải đặt chúng ở ngay dòng đầu tiên của tài liệu. Nếu không theo yêu cầu này thì tài liệu XML được xem là không hợp lệ.

- Một tài liệu XML chỉ có duy nhất một phần tử gốc, các phần tử khác nếu có phải là con của phần tử gốc.

- Mọi phần tử XML khác rỗng phải bao gồm đầy đủ thẻ đóng và thẻ mở. Trong tài liệu XML hợp khuôn dạng, mọi thẻ mở (ngoại trừ thẻ rỗng) phải có thẻ đóng tương ứng. Nếu bộ phân tích XML phát hiện một phần tử bị thiếu thẻ đóng hay thẻ mở thì quá trình phân tích, biên dịch tài liệu sẽ bị kết thúc.

- Đóng phần tử rỗng với chuỗi />. Phần tử rỗng là phần tử chỉ có một thẻ. Dấu kết của phần tử rỗng phải là /> nếu không tài liệu sẽ được coi là không hợp khuôn dạng.

- Mọi phần tử trong tài liệu XML khác phần tử gốc đều phải nằm giữa cặp thẻ gốc.

- Các phần tử lồng nhau phải có thẻ đóng, mở tương ứng đúng vị trí. Chỉ nên sử dụng 5 tham chiếu thực thể được định nghĩa trước trong XML.

b) Đặc tả cấu trúc XML với DTD (Document Type Definition)

DTD là một tập hợp các khai báo để mô tả ràng buộc các tài liệu

Số hóa bởi Trung tâm Học liệu – ĐHTN

<http://www.lrc.tnu.edu.vn>

được viết bằng ngôn ngữ đánh dấu như SGML, XML, HTML. DTD có thể được khai báo bên trong một tài liệu XML hoặc trong một tập tin bên ngoài. Với DTD, các nhóm xây dựng hệ thống thông tin có thể thỏa thuận sử dụng một DTD tiêu chuẩn cho việc trao đổi dữ liệu. Các ứng dụng với XML có thể sử dụng một DTD tiêu chuẩn để xác định các kiểu dữ liệu được nhận từ bên ngoài vào là hợp lệ. Chúng ta có thể sử dụng DTD để kiểm tra dữ liệu XML của hệ thống.

c) Đặc tả cấu trúc XML với Schemma

XML Schema được phát triển trên nền tảng của DTD và nổi trội hơn ở một số đặc điểm:

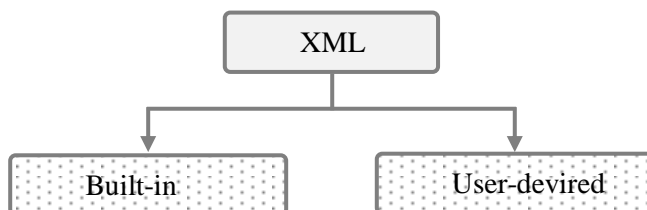
- Schema cũng là một dạng tài liệu theo chuẩn XML, do đó để mô tả cấu trúc của dữ liệu người dùng schema không cần phải làm quen với cấu trúc mới.

- Schema có thể mở rộng, người dùng có thể định nghĩa thêm các kiểu dữ liệu riêng từ các kiểu dữ liệu chuẩn.

- Schema hỗ trợ namespace.

- Schema hỗ trợ nhiều kiểu dữ liệu, định phạm vi giá trị dữ liệu, do đó có thể dễ dàng khai báo, tìm kiếm và kiểm tra tính đúng đắn của dữ liệu.

Trong Schema dữ liệu được chia thành hai kiểu: kiểu dữ liệu có sẵn và kiểu dữ liệu do người dùng tự định nghĩa:



Hình 1.3. Các kiểu dữ liệu trong Schema (Schema Datatypes)

Số hóa bởi Trung tâm Học liệu – ĐHTN

<http://www.lrc.tnu.edu.vn>

1.2.3. Phân tích tài liệu XML theo mô hình DOM

a) Mô hình đối tượng tài liệu XML

Để xử lý tài liệu XML, W3C định nghĩa tài liệu theo mô hình hướng đối tượng DOM. Theo mô hình này tài liệu XML được tổ chức theo cấu trúc hình cây bao gồm các nút đối tượng, nội dung của mỗi nút (node), trên cây có thể chứa phần tử (element), dữ liệu (text), thuộc tính (attribute),... và các nút con khác.

Mô hình DOM được W3C định nghĩa theo nhiều mức khác nhau, cụ thể là các mức sau:

Mức 0: đặc tả XML không chính thức và được áp dụng cho các trình duyệt trước đây như Netscape Navigator 3.0 và IE 3.0.

Mức 1: Tập trung vào kết hợp giữa tài liệu XML và HTML.

Mức 2: Hỗ trợ không gian tên gọi XML, khung nhìn có lọc (filtered views) và các sự kiện DOM.

Mức 3: Nâng cao các khả năng của DOM với việc cho phép nạp, xử lý, lưu lại mô hình cho nội dung tài liệu.

b) Phân tích tài liệu XML theo mô hình DOM

Với mô hình DOM, một tài liệu XML có thể được phân tích: tạo, duyệt, thêm, loại bỏ, thay đổi các phần tử thông qua các ngôn ngữ như: Java, JavaScript, Visual Studio.NET, C/C++,.... Cho dù sử dụng ngôn ngữ nào để phân tích thì ta cũng phải sử dụng tới các phương thức mà DOM cung cấp được liệt kê trong bảng dưới đây:

Phương thức	Ý nghĩa
documentElement	Lấy nút gốc của tài liệu DOM
parentNode	Lấy nút cha của nút hiện tại

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>

firstChild	Lấy nút con đầu tiên
previousChild	Lấy nút con trước đó
nextChild	Lấy nút con kế tiếp
lastChild	Lấy nút con cuối cùng
firstSibling	Lấy nút con cùng cấp đầu tiên
nextSibling	Lấy nút con cùng cấp kế tiếp
previousSibling	Lấy nút con cùng cấp trước đó
lastSibling	Lấy nút con cùng cấp cuối cùng
getElementByTagName	Lấy về danh sách các nút
Item(i)	Truy cập nút thứ i trong danh sách
childNodes	Trả về danh sách các nút trong tài liệu
doctype	Trả về kiểu của nút gốc
Length	Trả về các nút trong danh sách các nút
nodeName	Trả về tên của một nút
nodeValue	Trả về giá trị của một nút
nodeType	Xác định kiểu dữ liệu của một nút
createElement()	Khởi tạo một thành phần mới
createTextNode()	Khởi tạo dữ liệu text cho một nút

Bảng 1.4: Các phương thức trong mô hình DOM

1.2.4. Tiếp cận, lưu trữ và xử lý dữ liệu XML

XML có hai chức năng chính :

Số hóa bởi Trung tâm Học liệu – ĐHTN
<http://www.lrc.tnu.edu.vn>