

CỔNG THÔNG TIN GIÁO DỤC CỘNG ĐỒNG DỰA ONTOLOGY A COMMUNITY PORTAL FOR EDUCATION BASED ON ONTOLOGY

Phạm Huy Giang, Tạ Tuấn Anh, Đặng Văn Chuyết
Trường Đại học Bách Khoa Hà Nội

TÓM TẮT

Nội dung của bài báo đề cập đến một cách tiếp cận mới trong việc xây dựng một cổng thông tin giáo dục dựa trên nguyên tắc của một mạng xã hội. Trong đó cộng đồng người sử dụng đóng vai trò hạt nhân, cung cấp các nguồn tài nguyên cũng như tri thức được chia sẻ. Nhằm tăng khả năng suy diễn ngữ nghĩa và tính mở rộng, hệ thống được xây dựng dựa trên nền tảng của công nghệ Web ngữ nghĩa với ontology với 3 mục tiêu: (i) mô hình hóa tri thức để chia sẻ trong cộng đồng; (ii) hỗ trợ tìm kiếm theo ngữ nghĩa các tài nguyên học tập; (iii) mô hình hóa một mạng xã hội người sử dụng trong cộng đồng.

ABSTRACT

This paper presents a new approach to build an education information portal based on the principle of social networks. In such a user centric portal, users have to provide resources and knowledge for sharing. In order to increase the semantic reasoning ability, the system is built up on the basis of ontology in Semantic Web. There are three purposes to apply ontology in this system: (i) representing of shared knowledge in a community; (ii) supporting for semantic searches of learning contents; (iii) modeling a social network that represents the individuals of the community.

I. ĐẶT VẤN ĐỀ

Trong những năm qua, công nghệ thông tin, đặc biệt là Internet, đã góp phần rất lớn vào việc xây dựng và đổi mới các hệ thống giáo dục với các khóa học trực tuyến, các kho chứa tài liệu học tập, đáp ứng được những tiêu chí giáo dục mới: học mọi nơi, mọi lúc, học theo sở thích, ... Tuy nhiên, các hệ thống giáo dục đang gặp phải một số khó khăn như: (i) thiết kế, phát triển, quản lý, chia sẻ các tài nguyên giáo dục; (ii) phân phối các dịch vụ giáo dục và tài liệu học tập cá nhân [1, 2].

Web ngữ nghĩa là sự mở rộng của World Wide Web bằng cách thêm vào các mô tả ngữ nghĩa của thông tin dưới dạng mà chương trình máy tính có thể "hiểu" và do vậy cho phép xử lý thông tin hiệu quả hơn [3]. Cơ chế cho phép chia sẻ và trao đổi ngữ nghĩa của thông tin là ontology. Theo [1, 2, 4], trong các hệ thống giáo dục, ontology được sử dụng chủ yếu cho 3 mục đích: (i) xây dựng cấu trúc, diễn giải ngữ nghĩa, đánh chỉ mục, tìm kiếm tài nguyên học tập. (ii) biểu diễn và lưu trữ tri thức, và (iii) tạo các phương pháp giảng dạy, các hoạt động học tập ứng với mục tiêu, sở thích, năng lực của từng cá nhân.

Nội dung của bài báo đề cập đến một cách tiếp cận mới trong xây dựng một cổng thông tin giáo dục dựa trên nguyên tắc của một mạng xã hội. Ontology được áp dụng trong hệ thống này với 3 mục đích sau: (i) mô hình hóa tri thức chia sẻ trong cộng đồng về một lĩnh vực; (ii) tổ chức đánh chỉ mục hỗ trợ tìm kiếm theo ngữ nghĩa các tài nguyên học tập; (iii) mô hình hóa một mạng xã hội thể hiện sự đóng góp chia sẻ của các cá nhân trong cộng đồng.

II. CÁC HỆ THỐNG THÔNG TIN GIÁO DỤC DỰA TRÊN ONTOLOGY

Hiện nay đã có nhiều hệ thống giáo dục được xây dựng theo cách tiếp cận sử dụng ontology. Dựa trên các tính năng của hệ thống mà ta có thể phân loại chúng thành 3 nhóm chủ yếu sau: (i) các hệ thống chia sẻ tài nguyên giáo dục trực tuyến: GEM [5], Connexions [6]; (ii) các mạng chia sẻ ngang hàng về tài nguyên giáo dục: POOL [7], Edutella [8]; và (iii) các hệ thống Elearning dựa trên ontology: PIP [1, 9], TANGRAM [2, 10].

2.1 Hệ thống chia sẻ tài liệu GEM

GEM - The Gateway to Educational Materials (<http://thegateway.org>) là hệ thống được xây dựng bởi NLE (National Library of

Education), liên kết nhiều tổ chức giáo dục, thư viện nhằm mục đích lưu trữ và chia sẻ tài nguyên giáo dục mở. GEM hỗ trợ 2 kiểu thành viên: (1) thành viên GEM là các nhà giáo dục; (2) thành viên GEM Consortium là các tổ chức và cá nhân muốn góp phần xây dựng kho chứa tài nguyên giáo dục của Gateway. Ontology trong GEM được chia thành nhiều khối, mỗi khối là một từ vựng điều khiển (controlled vocabulary), và được ghép lại thành một lược đồ thống nhất.

2.2 Dự án Connexions

Dự án Connexions (<http://cnx.rice.edu>) thuộc trường Rice University có mục đích xây dựng kho chứa tài liệu học tập mở và các công cụ hỗ trợ tạo bài giảng cũng như chia sẻ và khai thác tri thức. Connexions khuyến khích cộng đồng gồm giáo viên, sinh viên khai thác liên kết giữa các khái niệm, khóa học và môn học, tham gia đóng góp tri thức để xây dựng và chia sẻ tài liệu học tập. Kiến trúc Connexions được thiết kế tập trung, gồm 2 thành phần chính: (1) Kho chứa nội dung phục vụ cộng đồng gồm tài liệu học tập mở và các khối tri thức; (2) các công cụ phần mềm nguồn mở hỗ trợ việc tạo, quản lý, sử dụng và chia sẻ các module thông tin.

2.3 Hệ thống lưu trữ và chia sẻ POOL

POOL - Portal for Online Objects in Learning (<http://www.edusplash.net>) là một dự án của tập đoàn TeleLearning NCE, Canada, nhằm xây dựng cơ sở hạ tầng theo mô hình peer-to-peer, kết nối nhiều kho chứa tài nguyên giáo dục không đồng nhất thành một mạng lưới [7]. POOL tập trung phát triển 2 công nghệ: (1) "POOL, POND và SPLASH" là kiến trúc phân tán cho mạng peer-to-peer gồm các kho chứa; và (2) CanCore, giao thức siêu dữ liệu mô tả tài nguyên giáo dục. Giao thức CanCore gồm một tập các thành phần mô tả tài liệu, được chọn lựa từ các thành phần của các chuẩn Dublin Core, IMS và hoàn toàn tương thích với chuẩn IEEE LOM 1484.

2.4 Mạng chia sẻ ngang hàng Edutella

Edutella(<http://www.edutella.org/edutella.shtml>) là một mạng peer-to-peer kết nối các site giáo dục không đồng nhất về hiệu năng, kích thước lưu trữ, số người sử dụng, ... với nhiều kiểu kho chứa tài nguyên, ngôn ngữ truy vấn, và kiểu lược đồ siêu dữ liệu khác nhau,

nhằm mục đích trao đổi thông tin về tài nguyên giáo dục. Các dịch vụ của Edutella gồm: (1) Dịch vụ truy vấn: truy lục vào siêu dữ liệu RDF được lưu trong các kho chứa phân tán, sử dụng các ngôn ngữ truy vấn RQL, TRIPPLE; (2) Dịch vụ nhân bản: hỗ trợ tính sẵn dùng và cân bằng tải để đảm bảo tính nhất quán và toàn vẹn; (3) Dịch vụ ánh xạ (biên dịch các từ vựng siêu dữ liệu khác nhau, cho phép kết hợp giữa nhiều bên); (4) Dịch vụ Mediation (xác định các khung nhìn kết hợp dữ liệu từ nhiều nguồn siêu dữ liệu và điều hòa xung đột và chồng chéo thông tin).

2.5 Hệ thống tạo kế hoạch giảng dạy PIP

PIP - Personalized Instruction Planner (<http://peonto.cityu.edu.hk/index.jsp>) là hệ thống hỗ trợ giảng dạy ngôn ngữ tiếng Anh trong các trường trung học ở Hồng Kông. Mục tiêu của PIP là xây dựng kế hoạch học tập cá nhân của từng học sinh dựa trên thông tin về học sinh đó như năng lực, sở thích, ... kết hợp với phương pháp giảng dạy của giáo viên. Có 5 chức năng chính được kết hợp trong PIP: (1) Chức năng quản trị: có thể thay đổi lược đồ ontology; (2) Thiết kế chương trình học cá nhân; (3) Tạo kế hoạch giảng dạy; (4) Quản lý hồ sơ người sử dụng; (5) Tìm kiếm tài nguyên học tập từ Internet (Global Search) hoặc từ PIP (Local Search). Cơ chế tìm kiếm theo từ khóa có kết hợp bộ lọc kết quả. PIP sử dụng ontology framework PEOnto, gồm 5 ontology kết hợp với nhau: (1) People Ontology biểu diễn cá nhân tham gia hệ thống; (2) Subject Domain (Language) Ontology biểu diễn tri thức lĩnh vực; (3) Curriculum Ontology biểu diễn kế hoạch giảng dạy như đường học (learning path), mục tiêu, các hoạt động học tập; (4) Pedagogy Ontology biểu diễn chiến lược/phương thức tổ chức tài nguyên giáo dục; và (5) PEA Ontology mô tả vai trò, chức năng của các Agent.

2.6 Hệ thống tạo bài giảng TANGRAM

TANGRAM(<http://iis.fon.bq.ac.yu/TANGRAM/home.html>) là ứng dụng xây dựng bài giảng và tài liệu học tập cho giảng viên và sinh viên đại học trong lĩnh vực IIS (Intelligent Information Systems). TANGRAM tập trung diễn giải chi tiết tài liệu theo cấu trúc nhằm mục đích sử dụng lại từng thành phần của tài liệu. Các module chức năng chính của TANGRAM gồm [2]: (1) module quản lý nội dung; (2)

module quản lý người sử dụng; (3) module lắp ráp động: tự động ghép các đơn vị nội dung thành tài liệu học tập cá nhân; (4) module giao diện người sử dụng.

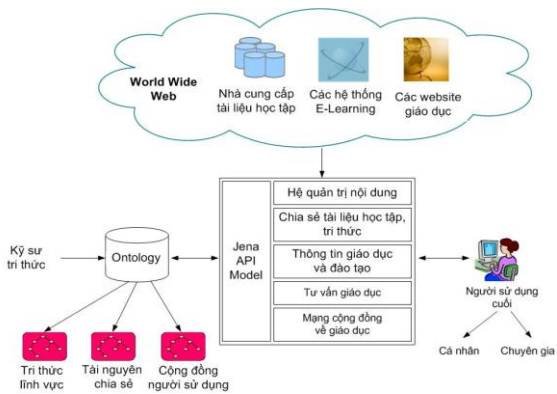
TANGRAM xây dựng nhiều ontology kết hợp để diễn giải tài liệu bao gồm: (1) ALOCoM CS (ALOCOM Content Structure) mô tả cấu trúc tài liệu; (2) ALOCoM CT (ALOCOM Content Type) mô tả thể loại tài liệu; (3) Domain IIS, mô tả lĩnh vực IIS, được phát triển dựa trên siêu lược đồ trong SKOS Core Ontology; (4) LP (Learning Path), xác định chuỗi các chủ đề học tập cho từng cá nhân thông qua Domain IIS; (5) User Model (UM) Ontology biểu diễn thông tin về người sử dụng.

III. CÔNG THÔNG TIN GIÁO DỤC CỘNG ĐỒNG BKEDUPORTAL

Từ những khảo sát đã thực hiện, chúng tôi đề xuất xây dựng một hệ thống thông tin giáo dục cộng đồng, hỗ trợ cộng đồng người sử dụng chia sẻ và tìm kiếm tri thức cũng như tài nguyên trong lĩnh vực giáo dục. Một trong những đặc điểm chủ yếu của hệ thống thông tin cộng đồng này là tất cả các nguồn tri thức lĩnh vực, tài nguyên chia sẻ và cộng đồng người sử dụng luôn có sự thay đổi thường xuyên. Người sử dụng không chỉ đóng vai trò của một người “tiêu thụ” mà còn phải đóng cả vai trò của một “nhà cung cấp”. Tùy vào vai trò của từng đối tượng, mà người sử dụng có thể tham gia cập nhật tri thức về lĩnh vực hoặc đóng góp chia sẻ các tài nguyên. Hạt nhân của hệ thống này là một công thông tin giáo dục có tên là BKeduPortal với các chức năng được minh họa trong hình 1.

Giống như các công thông tin được xây dựng trên nền tảng của các hệ quản trị nội dung (CMS), BKeduPortal cung cấp các kênh thông tin khác nhau đã được biên tập về giáo dục và đào tạo. Ví dụ, hệ thống tự động thu thập thông tin về tin tức, sự kiện, bài viết về giáo dục đã được biên tập từ nhiều website giáo dục. BKeduPortal hỗ trợ chia sẻ tài liệu theo hai phương pháp: (1) hệ thống thu thập tự động tạo tài nguyên chia sẻ (vd. sách điện tử, bài giảng từ các hệ thống đào tạo trực tuyến); (2) người sử dụng đóng góp tài liệu cho cộng đồng theo lĩnh vực chủ đề đã được mô hình hóa trong hệ thống. Ngoài ra, BKeduPortal còn có chức năng tư vấn về giáo dục đào tạo, mang sắc thái của một hệ chuyên gia. Ví dụ, có thể trả lời câu hỏi về thi tuyển đại học như quy chế, ngành nghề đào tạo,... Đặc biệt, BKeduPortal có chức năng của một mạng xã hội phục vụ cộng đồng: gắn kết những thành viên có cùng sở thích, lĩnh vực chủ đề quan tâm; cho phép các thành viên liên lạc, trao đổi, giới thiệu thông tin như kinh nghiệm làm việc, thông tin tuyển dụng, đánh giá về tài nguyên như khóa học, tài liệu.

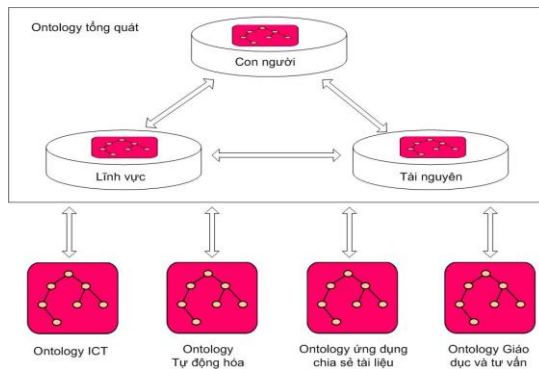
Mục đích sử dụng ontology trong hệ thống là giúp mô hình hóa dễ dàng các tri thức chung trong từng lĩnh vực (ví dụ các chủ đề học tập, danh mục ngành nghề, chuyên môn học tập,...), các nguồn tài nguyên (ví dụ tài liệu học tập, các câu hỏi thường gặp,...) được chia sẻ trong cộng đồng người sử dụng. Ngoài ra ontology còn được sử dụng để mô hình hóa các loại đối tượng sử dụng khác nhau trong cộng đồng người sử dụng để tạo thành một mạng xã hội, qua đó chúng ta có thể xác định được các chuyên gia và mối quan hệ của họ trong một lĩnh vực. Một nhóm các chuyên gia hay còn gọi là kỹ sư tri thức sẽ chịu trách nhiệm xây dựng và cải tiến lược đồ ontology. Cơ sở tri thức dựa trên ontology của hệ thống có thể được cập nhật theo hai phương thức chính: (i) Người sử dụng đăng nhập và sử dụng trực tiếp các chức năng (giao diện) của công thông tin giáo dục để tạo tri thức chia sẻ; (ii) Nguồn tri thức chia sẻ có thể được thu thập tự động từ các hệ phân hệ hỗ trợ mô tả ngữ nghĩa tài nguyên học tập dùng ontology (ví dụ hệ thống hỗ trợ xuất bản tài liệu học, thông tin khóa học sử dụng ontology, ...).



Hình 1. Kiến trúc công thông tin BKeduPortal

IV. ONTOLOGY CHO CÔNG THÔNG TIN GIÁO DỤC

Theo kiến trúc của công thông tin BKEDuPortal, mỗi chức năng hệ thống như Chia sẻ tài liệu học tập, Tư vấn giáo dục, ... là một trình ứng dụng được gắn vào công thông tin. Do đó yêu cầu của ontology là phải biểu diễn được tri thức cần thiết cho các ứng dụng. Ngoài các khái niệm tri thức chung, mỗi ứng dụng lại có đặc thù riêng. Do đó, yêu cầu thiết kế đối với ontology là: (1) phong phú và đúng đắn về tri thức; (2) khả năng sẵn sàng và dễ dàng áp dụng; (3) khả năng mở rộng và cải tiến; (4) khả năng chia sẻ tri thức giữa nhiều ứng dụng.



Hình 2. Minh họa 3 khối và các quan hệ

Từ yêu cầu thiết kế trên, ontology được thiết kế phân tầng với các khái niệm phân làm ba khối chính (hình 2).

- ❖ Một là các khái niệm mô tả tri thức về lĩnh vực dưới dạng các chủ đề và phân nhóm của chúng. Ví dụ các chủ đề về Windows XP, Linux thuộc về nhóm chủ đề Hệ điều hành, Oracle, MySQL thuộc về Cơ sở dữ liệu. Trong đó Hệ điều hành và Cơ sở dữ liệu là các nhóm con của Công nghệ thông tin. Các chủ đề được coi là một lớp các đối tượng tri thức trong chủ đề đó. Với mỗi nhóm chủ đề, có thể thêm các thuộc tính phân loại khác kèm theo chủ đề.
- ❖ Hai là các lớp mô tả nhóm các loại tài nguyên được chia sẻ trong hệ thống. Ví dụ đối với các loại tài liệu học tập, có thể có các lớp con là Sách, Bài báo, Sách điện tử, Bài giảng, ... Với mỗi nhóm tài nguyên, các thuộc tính siêu dữ liệu (meta-data) được lựa chọn sử dụng để mô tả tài liệu.

- ❖ Ba là các nhóm đối tượng người sử dụng của hệ thống. Cũng giống như tài nguyên, các thuộc tính cũng được gắn với từng nhóm người sử dụng để mô tả các cá nhân và các thông tin liên quan khác. Mỗi quan hệ giữa người sử dụng với các khối còn lại trong ontology thông qua các liên kết “là tác giả”, “là người đóng góp”, “là bạn của”, ...

Ontology được thiết kế phân tầng bao gồm ontology thành phần sau: (i) một ontology tổng quát mô tả các khái niệm chung của hệ thống; (ii) nhiều ontology lĩnh vực mô tả các miền ứng dụng khác nhau như Công nghệ thông tin, Tự động hóa, Cơ khí, ...

Để xây dựng ontology giáo dục theo nguyên tắc thiết kế ở trên, chúng tôi sử dụng mô hình biểu diễn ontology được phát triển dựa trên OWL Full với 3 mức:

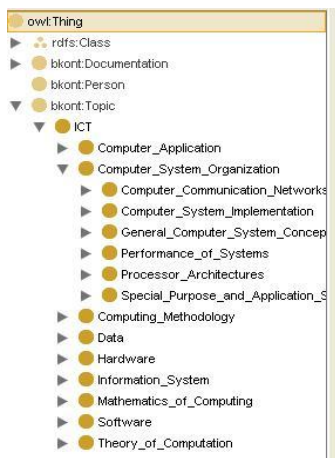
- (1) Mức siêu lược đồ (meta-schema): gồm các lớp TopicCls, DocumentationCls và PersonCls ứng với các siêu lớp dùng để tạo ra các lớp mô tả trong các khối tri thức lĩnh vực, tài nguyên và cộng đồng người sử dụng. TopicCls, DocumentationCls và PersonCls là các lớp con của owl:Class.
- (2) Mức lược đồ (schema): gồm các lớp được tạo từ các lớp ở mức meta-schema. Trong đó Topic, Documentation và Person là các lớp có kiểu tương ứng là TopicCls, DocumentationCls và PersonCls. Tùy thuộc vào mỗi loại hình ứng dụng khác nhau, mà các lớp này có thể được mở rộng với các lớp con khác nhau. Ví dụ trong ứng dụng chia sẻ tài liệu học tập về CNTT ta có ICT, Database, OS, ... là các lớp chủ đề con của Topic. Book, Paper, Ebook, ... là các lớp con của Documentation. Chú ý Topic, Documentation, Person luôn là lớp gốc cho các lĩnh vực chủ đề, tài nguyên và con người.
- (3) Mức thể hiện (Instance): bao gồm các đối tượng được tạo với các lớp định nghĩa trong mức lược đồ. Các đối tượng thể hiện chính là nguồn thông tin tri thức được chia sẻ và do cộng đồng người sử dụng cập nhật trong quá trình sử dụng. Số lượng các thể hiện trong ontology thường là rất lớn và không có sự hạn chế. Bởi cộng đồng người sử

dụng chính là nguồn cung cấp các thể hiện nên tính phong phú của hệ thống như thế này là rất lớn.

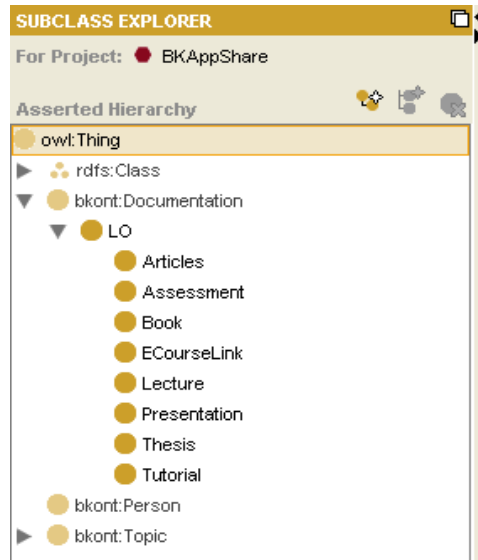
V. ỨNG DỤNG CHIA SẺ TÀI LIỆU HỌC TẬP

Để minh họa khả năng ứng dụng của ontology tổng quát, chúng tôi trình bày trong mục này việc xây dựng phân hệ chia sẻ tài liệu học tập của cổng thông tin BKeduPortal. Ứng dụng này tập trung vào việc quản lý tài liệu học tập do người sử dụng tự chia sẻ trong một lĩnh vực cộng đồng cụ thể. Hai ontology ở mức ứng dụng cần được phát triển dùng để : (i) mô tả các chủ đề học tập trong một lĩnh vực ; và (ii) tạo lược đồ mô tả các khái niệm dùng trong ứng dụng quản lý tài liệu học tập. Mặc dù chủ đề các lĩnh vực không bị hạn chế, trong nội dung bài báo này chỉ minh họa bằng một ví dụ về các chủ đề trong lĩnh vực Công nghệ Thông tin và Truyền thông (ICT).

Để xây dựng ontology chủ đề về CNTT và Truyền thông, chúng tôi sử dụng lại hệ thống phân loại lĩnh vực khoa học máy tính của ACM (ACM Computer Classification System) [4]. Hệ thống gồm 11 thành phần mức đỉnh, và mỗi thành phần được phân cấp thành 4 mức và tổng cộng có khoảng 1600 chủ đề. Mỗi chủ đề được đánh chỉ mục duy nhất theo phân cấp, ví dụ chỉ mục B.1.1.2 ứng với chủ đề ‘Microprogrammed logic array’, có thể truy cập theo đường dẫn: Hardware (B)/Control Structures and MicroProgramming (B.1)/Control Design Styles (B.1.1). Hình 3 minh họa ontology chủ đề về ICT được xây dựng bằng bộ công cụ Protege.



Hình 3. Phân cấp lĩnh vực ICT dựa trên ACM



Hình 4. Các lớp trong ontology ứng dụng

Để mô tả các khái niệm của ứng dụng, chúng tôi tạo một ontology (hình 4). Nó bao gồm lớp LO (Learning Object) được mở rộng từ lớp Documentation, có mục đích mô tả tài liệu học tập theo ngữ cảnh sử dụng với các thuộc tính đã được định nghĩa trong Dublin Core ontology như nhà xuất bản (dc:publisher), định dạng tài liệu (dc:format),... Các tài liệu học tập được phân loại theo các lớp con của lớp LO như Book, Articles, Lecture, ...

VI. KẾT LUẬN

Bài báo trình bày việc thiết kế và xây dựng ontology cho một cổng thông tin giáo dục trên cơ sở ba khối tri thức lĩnh vực, tài nguyên và con người. Để có thể đáp ứng các chức năng hệ thống như chia sẻ tài liệu, tư vấn giáo dục, ... và các yêu cầu thiết kế như có thể mở rộng, chia sẻ và sử dụng lại, ..., Ontology được thiết kế theo nguyên tắc phân module trên cơ sở của OWL trong Web ngữ nghĩa. Tuy mục đích thiết kế ontology dành cho ứng dụng cụ thể, nhưng kiến trúc ontology có thể áp dụng cho nhiều hệ thống ứng dụng khác nhau.

TÀI LIỆU THAM KHẢO

1. *A. W. P. Fok and H. H. S. Ip*; Education Ontologies Constructions for Personalized Learning on the Web; Studies in Computational Intelligence (SCI) 62, 47-82, Springer, 2007.
2. *Vladan Devedzic*; Semantic Web and Education; Springer, 2006. ISBN:0-387-35417-4.
3. *T. Berners-Lee, J. Hendler, O. Lassila*; The Semantic Web; Scientific American, May 2001
4. The ACM Computing Classification System – 1998 Version, valid in 2002, <http://www.acm.org/class/1998/>
5. <http://www.thegateway.org/documents>
6. *Geneva Henry, Richard Baraniuk, Christopher Kelty*; The Connexions Projects: Promoting Open Sharing of Knowledge for Education; Connexions Projects – <http://cnx.rice.edu>.
7. *Griff Richards, Rory McGreal, Marek Hatala and Norm Friensen*; The Evolution of Learning Object Repository Technologies: Portals for On-line Objects for Learning; Journal of Distance Education. Vol.17, No3, 2002.
8. Semantic Web Workshop 2002 Honolulu, Hawaii, May 7, 2002. Edutella: Searching and Annotating Resources within an RDF-based P2P Network.
9. *A. W. P. Fok and H. H. S. Ip*; Personalized Search of Educational Content Based on Multiply Ontologies; IEEE 2006, 1-4244-0367-7/06.
10. *Jelena Jovanovic, Dragan Gasevie, Vladan Devedzie*; Ontology-Based Automatic Annotation of Learning Content; Int'l Journal on Semantic Web & Information Systems.

Địa chỉ liên hệ: Tạ Tuấn Anh - Tel: 0983.984.775, email: anhtt@it-hut.edu.vn
Khoa Công nghệ Thông tin – Trường Đại học Bách Khoa Hà Nội
Số 1, Đại Cồ Việt, Hà Nội