

XÂY DỰNG ONTOLOGY CHO THƯ VIỆN SỐ

Nguyễn Thị Mỹ Trang

Trung tâm Học liệu, Đại học Huế

Hoàng Hữu Hạnh

Đại học Huế

TÓM TẮT

Trong bài báo này, chúng tôi trình bày tổng quan về Web Ngữ nghĩa và ứng dụng công nghệ Web Ngữ nghĩa vào thư viện số để nâng cao khả năng truy hồi, tìm kiếm thông tin. Chúng tôi đã bước đầu xây dựng SemLib Ontology, một Ontology cho thư viện số, cung cấp một khung nhìn duy nhất về tên của các đối tượng trong thư viện số hỗ trợ tìm kiếm hiệu quả trên một kho dữ liệu lớn và phân tán, đơn giản hóa hoạt động liên thông thư viện.

I. Giới thiệu về Web Ngữ nghĩa

Năm 2001, Tim Berners-Lee, trong một bài báo lần đầu tiên giới thiệu chính thức về Web Ngữ nghĩa (WebNN) đăng trên tạp chí Scientific American Online, đã đưa ra định nghĩa: “Web Ngữ nghĩa là sự mở rộng của Web hiện tại mà ở đó thông tin được định nghĩa một cách có ý nghĩa, cho phép máy tính và con người có thể hợp tác với nhau” [3].

Với WebNN, thông tin có thể được định nghĩa không chỉ cho mục đích hiển thị mà còn cho khả năng liên kết hoạt động và tích hợp giữa các ứng dụng và các hệ thống; cho phép trao đổi thông tin giữa máy – máy và xử lý thông tin một cách tự động, biểu diễn thông tin ở dạng mà máy tính có thể hiểu được. Mục tiêu ban đầu của WebNN là để hỗ trợ người sử dụng tìm kiếm thông tin trên mạng một cách nhanh chóng, chuẩn xác và thông minh hơn so với các công cụ tìm kiếm truyền thống; tức là tìm kiếm thông tin dựa vào ngữ nghĩa chứ không phải tìm kiếm theo từ khóa.

Ontology được xem như là “linh hồn” của WebNN. Ontology giúp con người và máy có thể hợp tác, cùng nhau làm việc, giúp máy có thể “hiểu” và có khả năng xử lý thông tin hiệu quả. Các Ontology được phát triển trong lĩnh vực trí tuệ nhân tạo để việc sử dụng lại và chia sẻ tri thức được thuận tiện hơn. Vào cuối thập kỷ 20 và đầu thập kỷ 21, các ontology đã trở thành một lĩnh vực nghiên cứu quan trọng trong Khoa học máy tính.

Vào năm 1998, Studer và các cộng sự đã đưa ra định nghĩa Ontology: “Ontology là một đặc tả tường minh, mang tính hình thức của sự khái niệm hóa có thể chia sẻ

được” [1]. Sự khái niệm hóa đề cập đến một mô hình trừu tượng của một số hiện tượng trong thế giới thực bằng cách xác định các khái niệm liên quan của hiện tượng đó. Tường minh có nghĩa là các khái niệm được sử dụng và các ràng buộc trên chúng được định nghĩa một cách rõ ràng. Hình thức đề cập đến việc máy có khả năng đọc ontology. Chia sẻ phản ánh quan điểm rằng một ontology nắm bắt tri thức được chấp nhận bởi một nhóm người.

Ontology cung cấp tri thức dùng chung của một miền xác định nào đó, cho phép sự giao tiếp giữa con người, các tổ chức và các hệ thống ứng dụng.

II. Ứng dụng Web Ngữ nghĩa vào thư viện số

Vấn đề đặt ra hiện nay cho các thư viện điện tử là việc liên thông hoạt động thư viện, quản lý các tài nguyên không lồ của thư viện như thế nào để hỗ trợ việc tìm kiếm, truy hồi thông tin dễ dàng hơn, chính xác hơn, tìm kiếm theo ngữ cảnh của người sử dụng. Điều này yêu cầu các thư viện phải sử dụng siêu dữ liệu chung để mô tả các bản ghi của danh mục và các từ vựng điều khiển chung để cho phép gán các định danh (ID) chủ đề cho các xuất bản phẩm.

Thư viện điện tử thường sử dụng một định dạng siêu dữ liệu nào đó để tổ chức các mô tả thư mục. Các chuẩn định dạng mô tả thư mục phổ biến là MARC¹, Dublin Core², BibTeX³,... Tuy nhiên, các chuẩn này được định nghĩa cho quá trình sử dụng của con người, chỉ có con người mới hiểu được, không định nghĩa ngữ nghĩa của các trường siêu dữ liệu theo cách máy có thể hiểu được.

Với công nghệ WebNN, sự biểu diễn của các chuẩn mô tả thư mục như các ontology là một sự lựa chọn tất yếu. Các định dạng siêu dữ liệu có ngữ nghĩa được biểu diễn trong các thuật ngữ của ontology cung cấp khả năng sử dụng các khái niệm đã được định nghĩa và suy diễn dữ liệu tiềm ẩn từ các mô tả thư mục [9].

Quá trình hoạt động liên thông giữa các thư viện khác nhau, hay thậm chí giữa các bộ sưu tập khác nhau trong cùng thư viện là một vấn đề. Việc tìm kiếm, truy hồi sẽ chính xác hơn nếu chúng ta cung cấp cho người sử dụng một khung nhìn duy nhất về tên của các đối tượng số trong thư viện. Ontology có thể cung cấp khả năng hoạt động liên thông giữa các tập dữ liệu lớn, cung cấp một khung nhìn chung cho các tập dữ liệu [8].

Bằng cách xây dựng ontology, tất cả người sử dụng có thể truy cập ontology thông qua Web, các danh mục thư viện có thể sử dụng các từ vựng giống nhau để biên

¹ <http://www.loc.gov/marc/>

² <http://dublincore.org/>

³ <http://www.bibtex.org/>

mục, đánh dấu các trường với các thuật ngữ thích hợp nhất cho lĩnh vực quan tâm. RDF [11] và OWL [12] cung cấp một hệ thống mã hóa đơn giản và nhất quán, do đó sẽ đơn giản hóa việc thực hiện hoạt động liên thông thư viện giữa hệ thống siêu dữ liệu thư viện điện tử này với các hệ thống siêu dữ liệu thư viện khác.

III. SemLIB Ontology

3.1 Giới thiệu chung

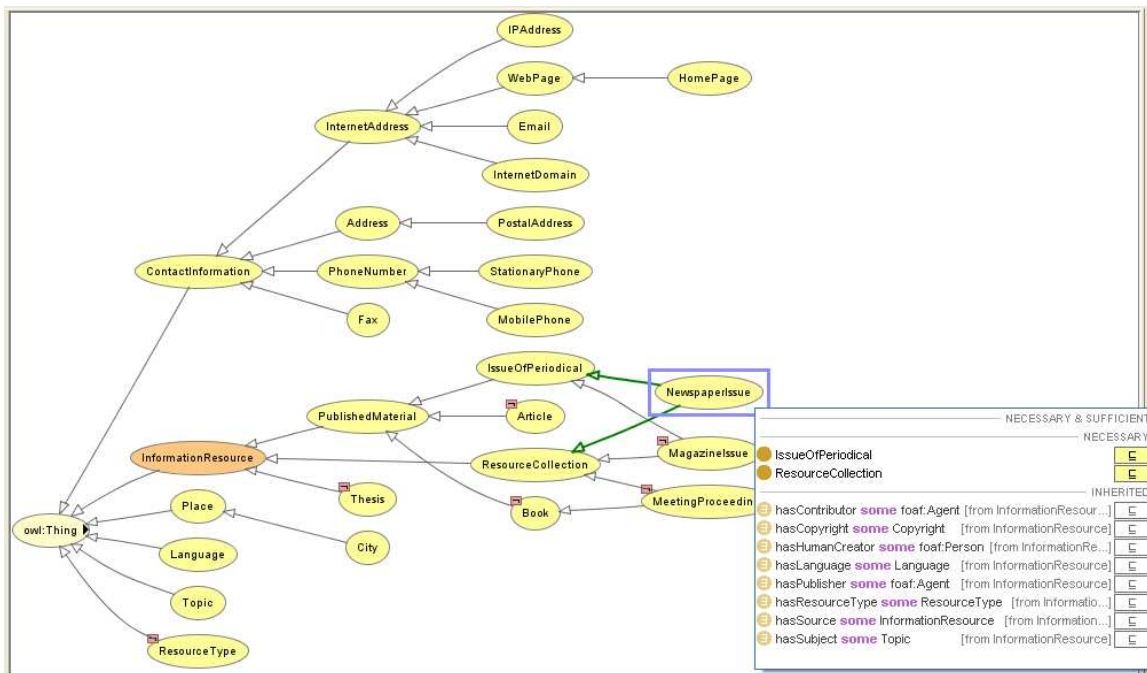
Chúng ta có thể xây dựng một tầng ngữ nghĩa trên cơ sở dữ liệu truyền thống của thư viện điện tử. Tầng ngữ nghĩa này cung cấp các khung nhìn ngữ nghĩa, các ngữ cảnh trên cơ sở dữ liệu truyền thống, hỗ trợ tìm kiếm theo ngữ cảnh cho người sử dụng, giúp tìm kiếm trên khái niệm chứ không phải tìm kiếm theo từ khóa. Chúng tôi xây dựng SemLib Ontology để mô tả các biểu ghi biên mục trong thư viện một cách có ngữ nghĩa và phát triển phương pháp để đưa SemLib Ontology vào tầng ngữ nghĩa này, phản ánh khung nhìn của người sử dụng trên cơ sở dữ liệu truyền thống, cụ thể là trên siêu dữ liệu của cơ sở dữ liệu truyền thống.

Tầng ngữ nghĩa được xây dựng trên cơ sở dữ liệu truyền thống của thư viện điện tử, tổ chức lưu trữ dữ liệu một cách có ngữ nghĩa. Các nguồn dữ liệu được chuyển sang dạng bộ ba RDF và lưu trữ với các ontology của chúng. Metastore sẽ được cập nhật một cách tự động và định kỳ với những thay đổi của các tài nguyên trong thư viện điện tử thông qua các kỹ thuật ontology tích hợp của Metatcore.

Chúng tôi đã bước đầu xây dựng SemLib Ontology (Semantic Library Ontology), để tổ chức và biểu diễn nội dung của cơ sở dữ liệu truyền thống một cách có ngữ nghĩa, cơ bản định nghĩa 31 lớp và 42 thuộc tính.

Các lớp InformationResource, PublishedMaterial, Article, Book, MeetingProceedings, IssueOfPeriodical, MagazineIssue, NewspaperIssue, ResourceCollection, Thesis, Place, City, ResourceType, Topic, ContactInformation, Address, PostalAddress, Fax, InternetAddress, Email, InternetDomain, IPAddress, WebPage, HomePage, PhoneNumber, MobilePhone, StationaryPhone, AcademicDepartment, University, Copyright, Language.

Các thuộc tính hasAbstract, hasNote, hasVolume, hasEdition, hasBeginDate, hastitle, hasNumber, hasDate, hasEndDate, hasISBN, hasISSN, hasAccessionNumber, hasPublisher, hasSubject, hasPlace, hasCity, hasContactInformation, hasInternetAddress, hasWebPage, hasEmail, hasInternetDomain, hasIPAddress, hasPhone, hasMobilePhone, hasStationaryPhone, hasFax, hasPostalAddress, hasResourceType, hasContributor, hasSource, hasLanguage, hasOwner, publishedWithin, hasOrganization, hasDepartment, hasSchool, hasCreator, hasHumanCreator, hasEditor, hasAuthor, hasCopyright, isPartOf.



Hình 1. Các phân cấp lớp trong SemLib Ontology

3.2 Mô tả các lớp trong SemLib Ontology

- Lớp **InformationResource**: lớp chứa tất cả các thư mục tài nguyên của thư viện. Lớp **InformationResource** có các thuộc tính: hasCopyright, hasHumanCreator, hasPublisher, hasSubject, hasContributor, hasLanguage, hasResourceType, hasSource, hasNote, hasDate, hasAbstract, hasTitle, hasOrganization, hasCreator.
- Lớp **PublishedMaterial**: một loại InformationResource, lớp chứa tất cả các xuất bản phẩm có trong thư viện, là lớp con của lớp **InformationResource**. Lớp **PublishedMaterial** ngoài các thuộc tính kế thừa từ **InformationResource** còn có các thuộc tính: **hasvolume**, **hasEdition**.
- Lớp **Article**: một loại **PublishedMaterial**, là một tài liệu tương đối ngắn, được xuất bản như thành phần của bộ sưu tập tài nguyên (ResourceCollection). Lớp **Article** ngoài các thuộc tính kế thừa từ **PublishedMaterial** còn có thuộc tính **publishedWithin**, **hasISSN**.
- Lớp **Book**: một loại **PublishedMaterial**, là lớp con của **PublishedMaterial**. Lớp **Book** có thuộc tính ngoài các thuộc tính kế thừa từ **PublishedMaterial** còn có thuộc tính **hasISBN**.
- Lớp **MeetingProceedings**: là lớp kỹ yếu, các bài báo được thông qua ở các hội nghị được xuất bản như một cuốn sách. Lớp **MeetingProceedings** là lớp con của lớp **Book** và **ResourceCollection**,

- Lớp **IssueOfPeriodical**: là lớp ấn phẩm định kỳ, một loại **PublishedMaterial**. Một số báo hay một tập của xuất bản định kỳ như báo, tạp chí, ... Lớp **IssueOfPeriodical** có thuộc tính **hasNumber, hasISSN**.
- Lớp **MagazineIssue**: là một loại **IssueOfPeriodical**, lớp con của lớp **IssueOfPeriodical** và **ResourceCollection**.
- Lớp **NewspaperIssue**: là một loại **IssueOfPeriodical**, lớp con của lớp **IssueOfPeriodical** và **ResourceCollection**.
- Lớp **ResourceCollection**: lớp bộ sưu tập tài nguyên, lớp con của lớp **InformationResource**. Lớp **ResourceCollection** có thuộc tính **hasVolume, hasNumber, hasEdition, hasISSN**.
- Lớp **Thesis**: lớp các luận án báo cáo, lớp con của lớp **InformationResource**. Lớp **Thesis** có các thuộc tính **hasCity, hasDepartment, hasSchool, hasAccessionNumber**.
- Lớp **Place**: thể hiện vị trí của một tổ chức, hay tài nguyên. Lớp **Place** có thuộc tính **hasPlace**.
- Lớp **City**: thể hiện tên thành phố, lớp con của lớp **Place**.
- Lớp **ResourceType**: là lớp các thể loại của nội dung tài nguyên. DCMI Type Vocabulary [35] cung cấp một danh sách các thuật ngữ có thể được sử dụng như các giá trị cho lớp **ResourceType** để xác định loại tài nguyên.
- Lớp **Topic**: lớp các chủ đề của tài nguyên thông tin. Các chủ đề xác định có thể được xây dựng như các thể hiện (cá thể) của lớp **Topic** hay các lớp con của nó. Phân cấp chủ đề được xây dựng sử dụng thuộc tính **subTopic** và thuộc tính này có tính bắc cầu.
- Lớp **ContactInformation**: chứa các thông tin của một tổ chức hay cá thể, thông tin liên lạc của nhà sản xuất, ... Lớp **ContactInformation** có thuộc tính **hasContactInformation**.
- Lớp **Address**: lớp con của lớp **ContactInformation**.
- Lớp **PostalAddress**: lớp con của lớp **Address**. Lớp **PostalAddress** có thuộc tính **hasPostalAddress**.
- Lớp **Fax**: lớp con của lớp **ContactInformation**. Lớp **Fax** có thuộc tính **hasFax**.
- Lớp **InternetAddress**: lớp con của lớp **ContactInformation**. Lớp **InternetAddress** có thuộc tính **hasInternetAddress**.
- Lớp **Email**: lớp con của lớp **InternetAddress**. Lớp **Email** có thuộc tính **hasEmail**.

- Lớp **InternetDomain**: lớp con của lớp **InternetAddress**. Lớp **InternetDomain** có thuộc tính **hasInternetDomain**.
- Lớp **IPAddress**: lớp con của lớp **InternetAddress**. Lớp **IPAddress** có thuộc tính **IPAddress**.
- Lớp **WebPage**: lớp con của lớp **InternetAddress**. Lớp **WebPage** có thuộc tính **hasWebPage**.
- Lớp **HomePage**: lớp con của lớp **WebPage**.
- Lớp **PhoneNumber**: lớp con của lớp **ContactInformation**. Lớp **PhoneNumber** có thuộc tính **hasPhoneNumber**.
- Lớp **MobilePhone**: lớp con của lớp **PhoneNumber**. Lớp **MobilePhone** có thuộc tính **hasMobilePhone**.
- Lớp **StationaryPhone**: lớp con của lớp **PhoneNumber**. Lớp **StationaryPhone** có thuộc tính **hasStationaryPhone**.
- Lớp **AcademicDepartment**: lớp con của lớp **foaf:Organization**. Lớp **AcademicDepartment** có thuộc tính **hasDepartment**.
- Lớp **University**: lớp con của lớp **foaf:Organization**. Lớp **University** có thuộc tính **hasSchool**.
- Lớp **Copyright**: có thuộc tính **hasCopyright**, **hasOwner**, **hasEndDate**, **hasBeginDate**.
- Lớp **Language**: lớp chứa các ngôn ngữ thể hiện của tài nguyên.

3.3. Mô tả các thuộc tính trong *SemLib Ontology*

- Thuộc tính **hasAbstract**: tóm tắt tài nguyên.
OWL Type: **DatatypeProperty**
Domain: **InformationResource**
Range: **xsd:string**
- Thuộc tính **hasNote**: thông tin mô tả tài nguyên.
OWL Type: **DatatypeProperty**
Domain: **InformationResource**
Range: **xsd:string**
- Thuộc tính **hasVolume**: tập của một tạp san hay một quyển sách có nhiều tập.
OWL Type: **DatatypeProperty**

Domain: PublishedMaterial ⊔ ResourceCollection

Range: xsd:string

- Thuộc tính **hasEdition:** lần xuất bản tài nguyên.

OWL Type: DatatypeProperty

Domain: PublishedMaterial ⊔ ResourceCollection

Range: xsd:Int

- Thuộc tính **hasBeginDate:** ngày bắt đầu thời hạn bản quyền.

OWL Type: DatatypeProperty

Domain: Copyright

Range: xsd:Date

- Thuộc tính **hasTitle:** tiêu đề của tài nguyên.

OWL Type: DatatypeProperty

Domain: InformationResource

Range: xsd:String

- Thuộc tính **hasNumber:** Một báo, tạp chí phát hành thường được xác định bởi tập (volume) và số (number) của nó.

OWL Type: DatatypeProperty

Domain: PublishedMaterial ⊔ ResourceCollection

Range: xsd:Int

- Thuộc tính **hasDate:** năm xuất bản tài nguyên.

OWL Type: DatatypeProperty

Domain: InformationResource

Range: xsd:string

- Thuộc tính **hasEndDate:** ngày kết thúc thời hạn bản quyền.

OWL Type: DatatypeProperty

Domain: Copyright

Range: xsd:Date

- Thuộc tính **hasISBN:** gán định danh ở dạng ISBN cho sách.

OWL Type: DatatypeProperty

Domain: Book

Range: xsd:String

- Thuộc tính **hasISSN**: gán định danh ở dạng ISSN cho báo, tạp chí.

OWL Type: DatatypeProperty

Domain: Article \sqcup IssueOfPeriodical

Range: xsd:string

- Thuộc tính **hasPublisher**: mô tả nhà xuất bản của một tài nguyên đã cho. Thuộc tính **hasPublisher** có thuộc tính nghịch đảo là **published**.

OWL Type: ObjectProperty

Domain: InformationResource

Range: foaf:Agent

- Thuộc tính **hasSubject**: gán chủ đề cho một tài nguyên.

OWL Type: ObjectProperty

Domain: InformationResource

Range: Topic

- Thuộc tính **hasPlace**: định vị vị trí của một cá nhân, tổ chức hay tài nguyên.

OWL Type: ObjectProperty

Range: Place

- Thuộc tính **hasCity**: thuộc tính con của **hasPlace**.

OWL Type: ObjectProperty

Domain: Thesis \sqcup foaf:Agent

Range: City

- Thuộc tính **hasContactInformation**:

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: ContactInformation

- Thuộc tính **hasInternetAddress**: thuộc tính con của **hasContactInformation**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: InternetAddress

- Thuộc tính **hasWebPage**: thuộc tính con của **hasInternetAddress**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: HomePage

- Thuộc tính **hasEmail**: thuộc tính con của **hasInternetAddress**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: Email

- Thuộc tính **hasInternetDomain**: thuộc tính con của **hasInternetAddress**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: InternetDomain

- Thuộc tính **hasIPAddress**: thuộc tính con của **hasInternetAddress**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: IPAddress

- Thuộc tính **hasPhone**: thuộc tính con của **hasContactInformation**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: PhoneNumber

- Thuộc tính **hasMobilePhone**: thuộc tính con của **hasPhone**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: MobilePhone

- Thuộc tính **hasStationaryPhone**: thuộc tính con của **hasPhone**.

OWL Type: ObjectProperty

Domain: foaf:Agent

Range: StationaryPhone

- Thuộc tính **hasFax**: thuộc tính con của **hasContactInformation**.

OWL Type: ObjectProperty

- Domain:** foaf:Agent

Range: Fax
- Thuộc tính **hasPostalAddress**: thuộc tính con của **hasContactInformation**.
 - OWL Type:** ObjectProperty
 - Domain:** foaf:Agent
 - Range:** PostalAddress
- Thuộc tính **hasResourceType**:
 - OWL Type:** ObjectProperty
 - Domain:** InformationResource
 - Range:** ResourceType
- Thuộc tính **hasContributor**: người chịu trách nhiệm xây dựng nội dung tài nguyên nhưng không phải là tác giả.
 - OWL Type:** ObjectProperty
 - Domain:** InformationResource
 - Range:** foaf:Agent
- Thuộc tính **hasSource**: tương đương với thuộc tính **dc:source**.
 - OWL Type:** ObjectProperty
 - Domain:** InformationResource
 - Range:** InformationResource
- Thuộc tính **hasLanguage**:
 - OWL Type:** ObjectProperty
 - Domain:** InformationResource
 - Range:** Language
- Thuộc tính **hasOwner**: mô tả cá nhân, tổ chức sở hữu bản quyền.
 - OWL Type:** ObjectProperty
 - Domain:** Copyright
 - Range:** foaf:Agent
- Thuộc tính **publishedWithin**:
 - OWL Type:** ObjectProperty
 - Domain:** Article
 - Range:** ResourceCollection
- Thuộc tính **hasOrgannization**:

OWL Type: ObjectProperty

Domain: InformationResource

Range: foaf:Organization

- Thuộc tính **hasDepartment**: thuộc tính con của **hasOrgannization**.

OWL Type: ObjectProperty

Domain: Thesis

Range: AcademicDepartment

- Thuộc tính **hasSchool**: thuộc tính con của **hasOrgannization**.

OWL Type: ObjectProperty

Domain: Thesis

Range: University

- Thuộc tính **hasCreator**: mô tả cá nhân hay tổ chức tạo tài nguyên hay bộ sưu tập tài nguyên.

OWL Type: ObjectProperty

Domain: InformationResource

Range: foaf:Agent

- Thuộc tính **hasHumanCreator**: thuộc tính con của **hasCreator**, mô tả cá nhân tạo ra tài nguyên.

OWL Type: ObjectProperty

Domain: InformationResource

Range: foaf:Person

- Thuộc tính **hasEditor**: thuộc tính con của **hasHumanCreator**.

OWL Type: ObjectProperty

- Thuộc tính **hasAuthor**: thuộc tính con của **hasHumanCreator**.

OWL Type: ObjectProperty

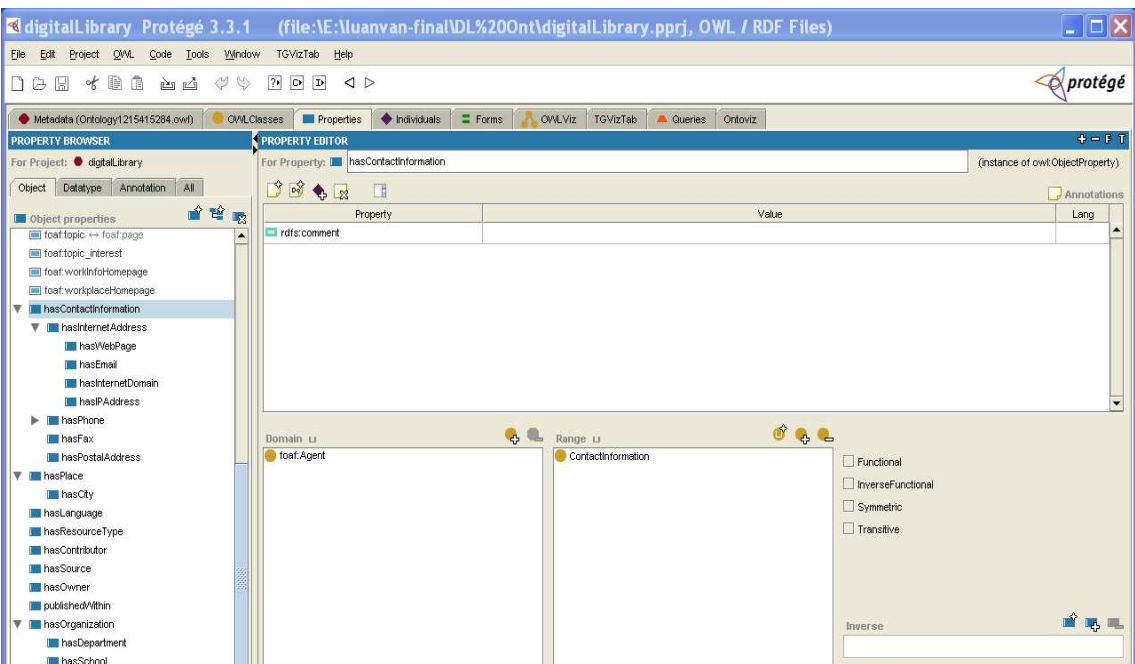
- Thuộc tính **hasCopyright**:

OWL Type: ObjectProperty

Domain: InformationResource

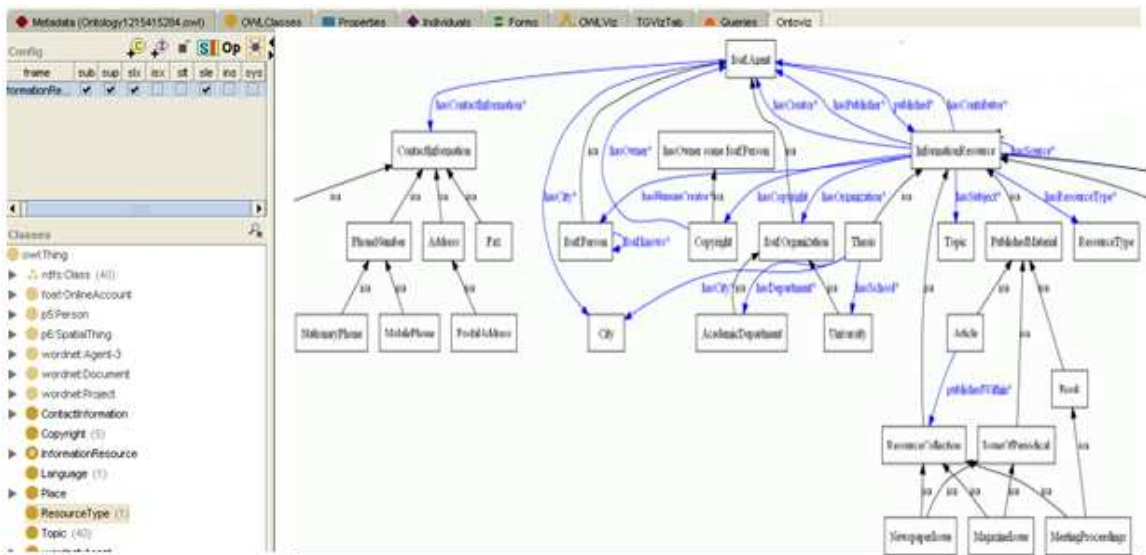
Range: Copyright

- Thuộc tính **isPartOf**: mô tả mối quan hệ giữa các thành phần của một bộ sưu tập hoặc giữa các bộ sưu tập. **OWL Type: ObjectProperty**



Hình 2. Các định nghĩa thuộc tính trong SemLib Ontology

Tất cả các thông tin về các tài nguyên thư viện, thông tin cá nhân được tổ chức một cách có ngữ nghĩa trong SemLib Ontology. SemLib Ontology đã đưa ra một số khái niệm và thuộc tính trong hệ thống thư viện, định nghĩa và mô tả mối quan hệ giữa các khái niệm đó như thế nào. Ngoài mô tả các tài nguyên trong thư viện, SemLib Ontology còn sử dụng siêu dữ liệu FOAF, một mạng liên kết các người sử dụng, để quản lý profile cá nhân, cung cấp một bộ từ vựng (ontology) cơ bản để mô tả thông tin về con người, nhóm, tổ chức, ...



Hình 3. SemLib Ontology cho Thư viện điện tử

3.4 Kho chứa ontology

Sau khi tạo ra, các ontology và phần ngữ nghĩa phải được lưu trữ vào kho chứa. Những kho chứa này thực chất là cơ sở dữ liệu cho phép lưu các mô tả trên ngôn ngữ RDFS hay OWL,... và cho phép biến đổi câu truy vấn trên ngôn ngữ này thành câu truy vấn SQL. Những kho chứa điển hình là Sesame và Jena.

Chúng tôi cũng đã mô tả một số bản ghi SemLib Ontology được biểu diễn có ngữ nghĩa ở dạng RDF.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:sl="http://www.owl-ontologies.com/semlib.owl/">
- <rdf:Description>
  <sl:hasTitle>Focus on health /</ sl:hasTitle >
  <sl:hasCreator>Hahn, Dale B.</ sl:hasCreator >
  <sl:hasCreator>Payne, Wayne A.</ sl:hasCreator >
  <sl:hasResourceType>text</ sl:ResourceType >
  <sl:hasPublisher>Boston : McGraw-Hill,</ sl:hasPublisher >
<sl:hasCity>Boston</sl:hasCity>
  <sl:hasDate>c2003.</sl:hasDate>
  <sl:hasLanguage> English.</sl:hasLanguage>
  <sl:hasNote>Includes bibliographical references and
index.</hasNote>
  <sl:hasSubject>Health.</sl:subject>
</rdf:Description>
- <rdf:Description>
  <sl:hasTitle>The right thing to do : basic readings in moral
philosophy /</sl:hasTitle>
  <sl:hasCreator>Rachels, James, 1941-2003.</sl:hasCreator>
  <sl:hasResourceType>text</sl:hasResourceType>
  <sl:hasPublisher>McGraw-Hill,</sl:hasPublisher>
  <sl:hasCity>Boston</sl:hasCity >
  <sl:hasDate>c2003.</sl:hasDate>
  <sl:hasLanguage > English.</sl:hasLanguage>
  <sl:hasNote>Includes bibliographical references.</sl:hasNote>
  <sl:hasSubject>Ethics</sl:hasSubject>
</rdf:Description>
-<rdf:Description>
  <sl:hasTitle>The nonprofit manager's resource directory /
</sl:hasTitle>
  <sl:hasCreator>Landskroner, Ronald A.</sl:hasCreator>
  <sl:hasResourceType>text</sl:hasResourceType >
  <sl:hasPublisher>Wiley, ;</sl:hasPublisher>
  <sl:hasCity>New York</sl:hasCity >
```

```

<sl:hasDate>c2002.</ sl:hasDate>
<sl:hasLanguage > English.</sl:hasLanguage>
<sl:hasSubject>Nonprofit organizations</sl:hasSubject>

<sl:hasWebPage>http://www.loc.gov/catdir/bios/wiley042/2001045384
.html
</sl:hasWebPage>
</rdf:Description>
-<rdf:Description>
  <sl:hasTitle>Designing rules for demand-driven rural investment
funds : the Latin American experience </sl:hasTitle>
  <sl:hasCreator>Wiens, Thomas B.</sl:hasCreator>
  <sl:hasCreator>Guadagni, Maurizio.</sl:hasCreator>
  <sl:hasResourceType>text</sl:hasResourceType>
  <sl:hasPublisher>World Bank,</sl:hasPublisher>
  <sl:hasCity>Washington, D.C.</sl:hasCity >
  <sl:hasDate>c1998.</sl:hasDate>
  <sl:hasLanguage > English.</sl:hasLanguage>
  <sl:hasNote>Includes bibliographical references (p. 68-
69).</sl:hasNote>
  <sl:hasSubject>Rural development projects</sl:hasSubject>
  </rdf:Description>
-<rdf:Description>
  <sl:hasTitle>A child's world : infancy through adolescence
/</sl:hasTitle>
  <sl:hasCreator>Papalia, Diane E.</sl:hasCreator>
  <sl:hasCreator>Olds, Sally Wendkos.</sl:hasCreator>
  <sl:hasCreator>Feldman, Ruth Duskin.</sl:hasCreator>
  <sl:hasResourceType>text</sl:hasResourceType>
  <sl:hasPublisher> McGraw-Hill,</sl:hasPublisher>
  <sl:hasCity>Boston</sl:hasCity>
  <sl:hasDate>c2002.</sl:hasDate>
  <sl:hasLanguage> English.</sl:hasLanguage>
  <sl:hasNote>Includes bibliographical references (B-1 to B-66)
and indexes.</sl:hasNote>
  <sl:hasSubject>Child development.</sl:hasSubject>
  <sl:hasSubject>Child psychology.</sl:hasSubject>
  <sl:hasSubject>Adolescence.</sl:hasSubject>
  </rdf:Description>
-<rdf:Description>
  <sl:hasTitle>A topical approach to life-span development /
</sl:hasTitle>
  <sl:hasCreator>Santrock, John W.</sl:hasCreator>
  <sl:hasResourceType>text</sl:hasResourceType>

```

```

<sl:hasPublisher>McGraw-Hill,</sl:hasPublisher>
<sl:hasCity>Boston</sl:hasCity>
<sl:hasDate>c2002.</sl:hasDate>
<sl:hasLanguage> English.</sl:hasLanguage>
<sl:hasNote>Includes bibliographical references (p. R1-R43) and
index.</sl:hasNote>
<sl:hasSubject>Developmental psychology</sl:hasSubject>
</rdf:Description>
-<rdf:Description>
<sl:hasTitle>Wireless mobile networking with ANSI-41 /
</sl:hasTitle>
<sl:hasCreator>Gallagher, Michael D.</sl:hasCreator>
<sl:hasCreator>Snyder, Randall A.</sl:hasCreator>
<sl:hasCreator>Gallagher, Michael D. Mobile telecommunications
networking with IS-41.</sl:hasCreator>
<sl:hasResourceType>text</sl:hasResourceType>
<sl:hasPublisher>McGraw-Hill,</sl:hasPublisher>
<sl:hasCity>New York</sl:hasCity>
<sl:hasDate>c2001.</sl:hasDate>
<sl:hasLanguage> English.</sl:hasLanguage>
<sl:Note>Includes bibliographical references and
index.</sl:Note>
<sl:hasSubject>Personal communication service
systems.</sl:hasSubject>
<sl:hasSubject>Mobile communication systems.</sl:hasSubject>
</rdf:Description>

```

IV. Kết luận và hướng phát triển

Trong bài báo này, chúng tôi nghiên cứu và xây dựng SemLib Ontology cơ bản với 31 lớp và 42 thuộc tính, bước đầu xây dựng tầng ngữ nghĩa trên cơ sở dữ liệu của thư viện điện tử để hỗ trợ người sử dụng tìm kiếm, truy hồi thông tin một cách hiệu quả hơn, đơn giản hóa hoạt động liên thông thư viện.

Trong các hướng tiếp cận tiêu biểu về ứng dụng WebNN vào thư viện số [6], [7] và [10], MarcOnt [10] là dự án được xem là lớn nhất hiện nay về nỗ lực xây dựng các chuẩn siêu dữ liệu tổng hợp cho các thư viện số. Cách tiếp cận của chúng tôi dựa trên MarcOnt, tuy nhiên, điểm khác biệt lớn đó là: chúng tôi tích hợp các tiêu đề chủ đề (subject headings), một chuẩn lớn trong công tác thư viện, vào việc mô hình hoá các khái niệm thư viện điện tử, hơn nữa, SemLib Ontology là một phần của hệ thống quản lý thư viện dựa trên WebNN mà chúng tôi đang xây dựng. Hệ thống SemLib giữ nguyên dữ liệu hiện có của các thư viện điện tử và cung cấp thêm một tầng ngữ nghĩa để có thể tạo cơ sở cho các hoạt động truy vấn, truy hồi các thông tin theo ngữ nghĩa.

Hướng phát triển trong tương lai, chúng tôi sẽ hoàn thiện SemLib Ontology, xây dựng thêm các ontology tham chiếu như ontology chủ đề hay ontology phân lớp (như DDC, LCC, CC,...) để hỗ trợ quá trình truy hồi, tìm kiếm thông tin hiệu quả hơn. Người sử dụng cũng có thể sử dụng phân cấp chủ đề (trong ontology phân lớp) để tìm các tài liệu liên quan, chọn phân cấp cha hay con để mở rộng hay lọc kết quả tìm kiếm.

TÀI LIỆU THAM KHẢO

1. Asunción M.F.L.G, Corcho O., *Ontological Engineering*, Springer-Verlag, London, 2004.
2. Berners-Lee T., Hendler J., and Lassila O., *The Semantic Web*, The Scientific American Journal, 2001.
3. Berners-Lee T., Hendler J., and Lassila O., *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American 284, 34-43.
4. Cardoso J., *Semantic Web Services: Theory, Tools, and applicationsI*, Information Science Reference, Hershey, New York, 2007.
5. Cardoso A.P.S.J., *Semantic Web Services, Processes and Applications*, Springer, 2006.
6. Quach T.T, Fong A.C.M. and Hui S.C., *A Scholarly semantic web system for advanced search functions*, School of Computer Engineering, Nanyang Technological University, Singapore, 2006.
7. Soergel D., *A Framework for Digital Library Research*, D-Lib Magazine, 2002.
8. Stuckenschmidt H., Harmelen F., *Information Sharing on the Semantic Web*, Springer, 2003.
9. Sure Y., Studer R., *Semantic Web Technologies for digital libraries*, Library Management, Vol.26, 2005.
10. Kruk S., Synak M., Zimmermann K., *MarcOnt – Integration Ontology for Bibliographic Description Formats*, International Conference on Dublin Core and Metadata Applications, Madrid, 2005.
11. Resource Description Framework, <http://www.w3.org/RDF/>.
12. Web Ontology Language, <http://www.w3.org/2004/OWL/>.

BULDING ONTOLOGY FOR DIGITAL LIBRARIES

*Nguyen Thi My Trang
Learning Resource Center, Hue University
Hoang Huu Hanh
Hue University*

SUMMARY

In this paper, we give a brief overview on the Semantic Web and related technologies applied in digital libraries in order to improve the information retrieval process. We have built an ontology for digital libraries namely SemLib Ontology which is based on standards such as Marc21, DublinCore. SemLib Ontology is the core element for intelligent digital library software systems to provide semantic unified views on library objects and this helps improving the search engine in a large data store, simplifying the inter-library integration process.