

# ỨNG DỤNG MÔ HÌNH WEB 3.0 NÂNG CAO HIỆU QUẢ DỊCH VỤ THƯ VIỆN VÀ ĐỀ XUẤT ÁP DỤNG TẠI THƯ VIỆN ĐẠI HỌC

*ThS. Trịnh Tất Đạt\**  
*ThS. Tào Ngọc Biên\*\**

**Tóm tắt:** *Thư viện số phải thường xuyên xử lý một lượng lớn thông tin từ các dạng tài liệu số. Áp dụng công nghệ Web3.0 (Web ngữ nghĩa) chúng ta có thể nghiên cứu và phát triển hệ thống thư viện số để thực hiện xử lý, lưu trữ, tìm kiếm và phân tích tất cả các kiểu thông tin số. Công nghệ ngữ nghĩa cho phép miêu tả đối tượng, thiết lập các lược đồ cần thiết trong các dạng của Ontology (bản thể luận trong logic) cho các định danh của các đối tượng số. Mục tiêu chính là làm cho thao tác giữa các phần xử lý thông minh, nhất quán, mạch lạc tương tự các lớp của đối tượng số, trong đó có các dịch vụ thông tin – thư viện.*

**Từ khóa:** *Web3.0; Thư viện số;*

## **1. Khái niệm và dịch vụ Web3.0**

### **1.1. Khái niệm Web3.0**

\* World Wide Web (WWW) đã trở thành một kho tàng thông tin khổng lồ của nhân loại và một môi trường chuyển tải thông tin không thể thiếu được trong thời đại công nghệ thông tin ngày nay. Sự phổ biến và bùng nổ thông tin trên Web cũng đặt ra một thách thức mới là làm thế nào để khai thác được thông tin trên Web một cách hiệu quả, mà cụ thể là làm sao để máy tính có thể trợ giúp xử lý tự động được chúng. Muốn vậy, trước hết máy tính phải hiểu được thông tin trên các tài liệu Web, trong khi ở thế hệ Web hiện tại thông tin được biểu diễn dưới dạng chỉ con người mới đọc hiểu được.

Web3.0 không là Web riêng biệt mà là một sự mở rộng của Web hiện tại, theo cách thông tin được xác định ý nghĩa tốt hơn, nó cho phép máy tính và người cộng tác với nhau tốt hơn. Web3.0 được hình thành từ ý tưởng của Tim Berners-Lee (1998), người phát minh ra WWW, URI (Uniform Resource Identification), HTTP, và HTML. Web3.0 là một mạng lưới các thông tin được liên kết sao cho chúng có thể được xử lý dễ dàng bởi các máy tính ở phạm vi toàn cầu. Nó được xem là cách mô tả thông tin rất hiệu quả trên WWW, và cũng được xem là một cơ sở dữ liệu có khả năng liên kết toàn cầu. WEB3.0 là một phương pháp cho phép định nghĩa và liên kết dữ liệu một cách có ngữ nghĩa hơn nhằm phục vụ cho máy tính có thể “hiểu” được. Web3.0 còn cung cấp một môi trường chia sẻ và xử lý dữ liệu tự động bằng máy tính.

Web3.0 là thế hệ Web mới, đang được phát triển và sẽ xuất hiện trong tương lai gần. Hệ thống Web mới này sẽ dần thay thế Web hiện tại song không có nghĩa là hoàn toàn khác

---

\* Trường Đại học Văn hóa, Thể thao và Du lịch Thanh Hóa

\*\* Trường Đại học Văn hóa, Thể thao và Du lịch Thanh Hóa

hệ thống Web hiện tại. Web3.0 được phát triển trên hệ thống Web hiện tại bằng cách bổ sung thêm ngữ nghĩa cho các tài nguyên Web mà máy tính có thể hiểu và tăng khả năng xử lý tự động.

\* Kiến trúc Web3.0: Web3.0 là một tập hợp/một chồng (stack) các ngôn ngữ, tất cả các lớp của Web3.0 được sử dụng để đảm bảo độ an toàn và giá trị thông tin trở nên tốt nhất [4].

- Lớp Unicode&URI: Bảo đảm việc sử dụng tập kí tự quốc tế và cung cấp phương tiện nhằm định danh các đối tượng trong Semantic Web.

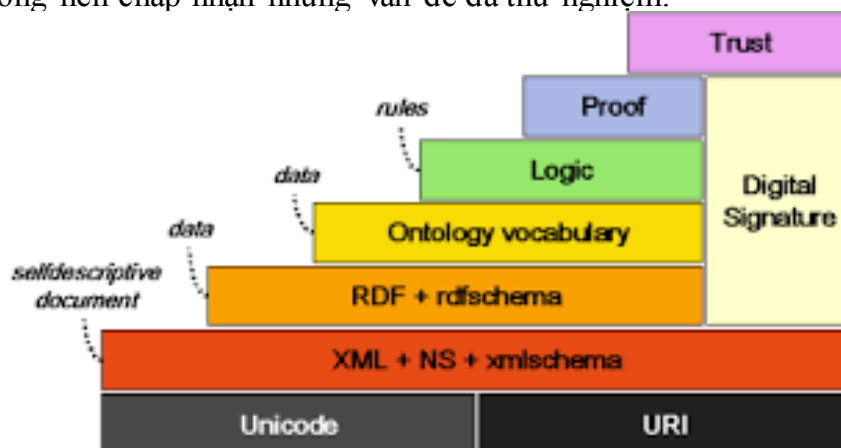
- Lớp XML cùng với các định nghĩa về namespace (vùng tên gọi/ không gian tên) và schema (lược đồ) bảo đảm rằng chúng ta có thể tích hợp các định nghĩa WEB3.0 với các chuẩn dựa trên XML khác.

- Lớp RDF [RDF] và RDFSchema [RDFS]: ta có thể tạo các câu lệnh (statement) để mô tả các đối tượng với những từ vựng và định nghĩa của URI, và các đối tượng này có thể được tham chiếu đến bởi những từ vựng và định nghĩa của URI ở trên. Đây cũng là Lớp mà chúng ta có thể gán các kiểu (type) cho các tài nguyên và liên kết. Và cũng là Lớp quan trọng nhất trong kiến trúc Web3.0.

- Lớp Ontology: hỗ trợ sự tiên hóa của từ vựng vì nó có thể định nghĩa mối liên hệ giữa các khái niệm khác nhau. Một Ontology định nghĩa một bộ từ vựng mang tính phổ biến và thông thường, nó cho phép các nhà nghiên cứu chia sẻ thông tin trong một hay nhiều lĩnh vực.

- Lớp Digital Signature: được dùng để xác định chủ thể của tài liệu (ví dụ: tác giả hay nhan đề của một loại tài liệu).

- Các lớp Logic, Proof, Trust: Lớp logic cho phép viết ra các luật (rule) trong khi lớp proof (thử nghiệm) thi hành các luật và cùng với lớp trust (chấp nhận) đánh giá nhằm quyết định nên hay không nên chấp nhận những vấn đề đã thử nghiệm.



Hình 1: Mô hình của Web3.0 được cấu thành từ nhiều tầng khác nhau

## 1.2. Dịch vụ Web 3.0

Theo định nghĩa của W3C (World Wide Web Consortium), dịch vụ Web là một hệ

thống phần mềm được thiết kế để hỗ trợ khả năng tương tác giữa các ứng dụng trên các máy tính khác nhau thông qua mạng internet, giao diện chung và sự gắn kết của nó được mô tả bằng XML.

Dịch vụ Web là tài nguyên phần mềm có thể xác định bằng địa chỉ URL, thực hiện các chức năng và đưa ra các thông tin người dùng yêu cầu. Một dịch vụ Web được tạo nên bằng cách lấy các chức năng và đóng gói chúng sao cho các ứng dụng khác dễ dàng nhìn thấy và có thể truy cập đến những dịch vụ mà nó thực hiện, đồng thời có thể yêu cầu thông tin từ dịch vụ Web khác.

Giá trị cơ bản của dịch vụ Web dựa trên việc cung cấp các phương thức theo chuẩn trong việc truy nhập đối với hệ thống đóng gói và hệ thống kế thừa. Các phần mềm được viết bởi những ngôn ngữ lập trình khác nhau và chạy trên những nền tảng khác nhau có thể sử dụng dịch vụ Web để chuyển đổi dữ liệu thông qua mạng internet theo cách giao tiếp tương tự bên trong một máy tính. Do vậy, việc phát triển và tích hợp các ứng dụng với dịch vụ Web đang được quan tâm phát triển là điều hoàn toàn dễ hiểu. Đặc biệt là việc ứng dụng dịch vụ Web ngữ nghĩa nhằm nâng cao hiệu năng các dịch vụ thư viện điện tử.

## **2. Web 3.0 trong thư viện số**

Thư viện số phải thường xuyên xử lý một lượng lớn thông tin từ các dạng tài liệu số. Phần lớn chúng được rút ra từ thư viện truyền thống, được tập trung biên tập lại thành nguồn thông tin sẵn dùng cho một nhóm người liên quan bằng cách quét bài báo, sách, tài liệu... để trở thành nguồn tài liệu số, tài nguyên đa phương tiện – là đối tượng quản lý chính của thư viện số. Bằng cách này đã làm hạn chế lợi thế của các hệ thống máy tính hiện đại và gây khó khăn cho quá trình xử lý sau này. Áp dụng công nghệ Web3.0 chúng ta có thể nghiên cứu và phát triển hệ thống thư viện số có thể thực hiện xử lý, lưu trữ, tìm kiếm và phân tích tất cả các kiểu thông tin số. Công nghệ Web3.0 cho phép miêu tả đối tượng, thiết lập các lược đồ cần thiết trong các dạng của Ontology cho các định danh của các đối tượng số. Mục tiêu chính là làm cho thao tác giữa các phần có thể xử lý thông minh, nhất quán, mạch lạc tương tự các lớp của đối tượng số và các dịch vụ.

Thành phần chính để xây dựng Web3.0 là các siêu dữ liệu, trong các thư viện số cũng vậy, các tài liệu cũng phải được biểu diễn bởi các siêu dữ liệu. Nếu tài liệu là những khối xây dựng căn bản của thư viện số, thì ngôn ngữ đánh dấu và các siêu dữ liệu là những yếu tố tổ chức. Ngôn ngữ đánh dấu được dùng để chỉ rõ cấu trúc của tài liệu riêng lẻ và kiểm soát phương thức trình bày cho người sử dụng. Siêu dữ liệu được dùng để xúc tiến việc truy cập đến những phần thích hợp của tài liệu qua việc tìm kiếm. Trong thư viện số có sự khác biệt quan trọng giữa siêu dữ liệu hiện và siêu dữ liệu ẩn. Siêu dữ liệu hiện được xác định bởi con người sau khi xem xét cẩn thận và phân tích tài liệu. Siêu dữ liệu ẩn được trích xuất tự động từ nội dung tài liệu nhờ kỹ thuật khai thác văn bản. Việc này thường khó thực hiện chính xác, được định nghĩa như một tiến trình phân tích văn bản để trích thông tin hữu ích cho mục

đích cụ thể.

Các thư viện số truyền thống cho phép người dùng tìm kiếm thông qua cấu trúc phân mục tài liệu hoặc qua tìm kiếm từ khóa nhờ kỹ thuật lập chỉ mục (Index) cho nội dung trong tài liệu. Việc định danh cho các tài liệu chứa hai hay nhiều cụm từ tương đương sẽ là vấn đề khó khăn. Do vậy, cơ chế tìm kiếm theo từ khóa và theo kiểu duyệt thư mục, chỉ mục không giải quyết được các câu truy vấn phức tạp và mang tính trừu tượng cao. Để hỗ trợ các câu truy vấn loại này, một trong các giải pháp là thêm dữ liệu ngữ nghĩa (semantec data) cho hệ thống thư viện số. Các thông tin ngữ nghĩa được biểu diễn bởi các siêu dữ liệu đi kèm với mỗi đối tượng tài liệu, cùng với một hay nhiều bản thể luận được cung cấp với ngữ cảnh ngữ nghĩa tương ứng, sẽ trả lời được các câu truy vấn mang tính trừu tượng.

Một thư viện số được tích hợp thêm ngữ nghĩa cho các tài nguyên được gọi là thư viện số ngữ nghĩa. Thư viện số ngữ nghĩa có các đặc điểm chung sau:

- Tích hợp nhiều nguồn thông tin dựa trên các siêu dữ liệu khác nhau (các tài liệu, hồ sơ người dùng, đánh dấu, phân loại,...)
- Cung cấp khả năng tương tác với các hệ thống khác (không chỉ các thư viện số với nhau) thông qua các siêu dữ liệu (RDF là một trong những tài nguyên thông dụng được dùng để trao đổi dữ liệu giữa các thư viện số với các dịch vụ khác).
- Cung cấp khả năng tìm kiếm theo ngữ nghĩa mạnh mẽ hơn so với các cách tìm kiếm thông thường và tra cứu tài liệu một cách dễ dàng.
- Nâng cao hiệu năng dịch vụ thư viện điện tử thông qua dịch vụ Web3.0

Với những thuật ngữ, câu định danh kiểu như “Vịnh Hạ Long là di sản thiên nhiên thế giới”... làm sao để máy tính có thể hiểu được? Vậy Web3.0 sẽ làm cho máy hiểu được nghĩa và tìm ra thông tin chính xác hơn Web hiện hành. Bây giờ, máy tính không phải suy luận dựa vào ngữ pháp và các ngôn ngữ đánh dấu nữa vì cấu trúc ngữ nghĩa của văn bản thực sự đã chứa nó rồi.

Thư viện số phải thường xuyên xử lý một lượng lớn thông tin từ các dạng tài liệu số. Áp dụng công nghệ Web3.0 chúng ta có thể nghiên cứu và phát triển hệ thống thư viện số, có thể thực hiện xử lý, lưu trữ, tìm kiếm và phân tích tất cả các kiểu thông tin số. Công nghệ ngữ nghĩa cho phép miêu tả đối tượng, thiết lập các lược đồ cần thiết trong các dạng của Ontology cho các định danh của các đối tượng số. Mục tiêu chính là làm cho thao tác giữa các phần có thể xử lý thông minh, nhất quán, mạch lạc tương tự các lớp của đối tượng số và các dịch vụ.

Thông thường, một thư viện số sử dụng dữ liệu mô tả có cấu trúc để mô tả hệ thống thư mục tuy nhiên các trường trong dữ liệu mô tả lại không được định nghĩa ngữ nghĩa một cách đầy đủ, việc ứng dụng Ontologies trong thư viện số không những thực hiện lưu trữ dữ

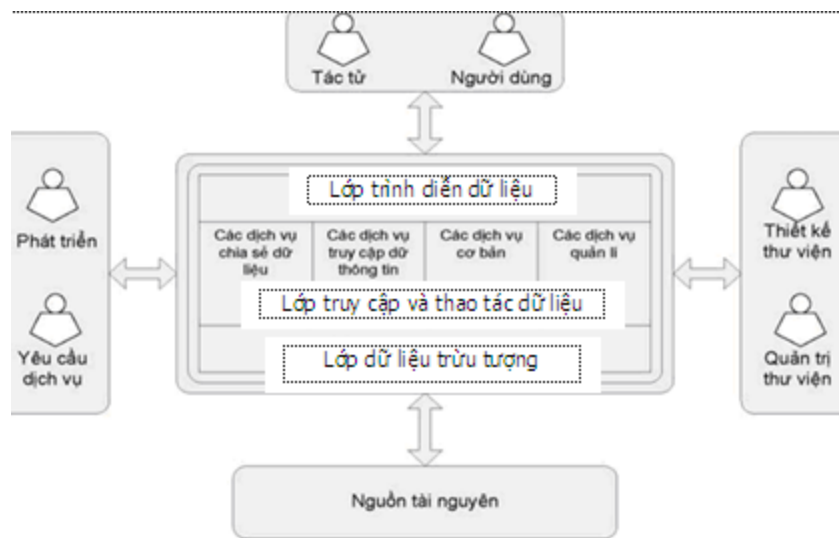
liệu để mô tả hệ thống thư mục mà còn mô tả được nội dung của nó. Thay vì trong trường hợp một quyển sách được lưu trữ trong thư viện số, chúng ta có thể tách riêng cấu trúc từng chương của nó, cung cấp mô tả cho mỗi chương và thực hiện lưu trữ mối quan hệ của các chương khác nhau. Bằng việc sử dụng tư tưởng cấu trúc của Ontologies và sử dụng tư tưởng này trong việc mô tả dữ liệu, chúng ta cung cấp một tầng tổng quát dữ liệu mô tả và nội dung.

Hệ thống dữ liệu của thư viện số rất lớn và đa dạng, nó thường phục vụ cho nhiều tổ chức, cá nhân vào nhiều mục đích khác nhau, trong khi đó dữ liệu chủ yếu thuộc vào hai dạng là dữ liệu có cấu trúc và dữ liệu phi cấu trúc. Một vấn đề đặt ra là làm thế nào để các ứng dụng sử dụng được đồng thời cả hai loại dữ liệu này, bởi vì trên thực tế mỗi ứng dụng chỉ sử dụng một loại dữ liệu có cấu trúc hoặc phi cấu trúc. Chúng ta có chuẩn chung phục vụ cho hầu hết các loại ứng dụng đó là sử dụng XML, nó được xem là nền tảng công nghệ của WEB3.0 và sẽ là cầu nối thực hiện chuẩn hóa các nguồn dữ liệu, từ đó có thể phục vụ cho mọi loại ứng dụng.

### **3. Định hướng ứng dụng Web3.0 vào thư viện đại học**

Vấn đề đặt ra hiện nay cho các thư viện số 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. Để có thể triển khai được, chúng ta cần 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 MARC, Dublin Core, BibTeX, RFID... 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.

Dựa trên các lớp kiến trúc thành Web3.0, chúng ta có thể xây dựng kiến trúc cho thư viện số dựa trên nền tảng này, đó là tạo ra các lớp Kiến trúc của thư viện số ngữ nghĩa cần phải có các thành phần tương ứng để đảm bảo cung cấp các dịch vụ trên một cách thống nhất, hình vẽ dưới đây mô tả một kiến trúc cơ bản của thư viện số ngữ nghĩa.



Hình 2: Kiến trúc cơ bản của thư viện số ngữ nghĩa

Theo mô hình, kiến trúc của thư viện số hình thành theo các lớp chính sau:

- Lớp dữ liệu trừu tượng : Cung cấp cơ chế để truy cập vào các nguồn tài nguyên.
- Lớp tầng Truy cập và thao tác dữ liệu: Hỗ trợ viện quản lý thư viện, Các dịch vụ cơ bản, Các dịch vụ truy cập thông tin, Hỗ trợ trao đổi dữ liệu.
- Lớp trình diễn dữ liệu: Cung cấp giao diện hỗ trợ người dùng.

Quá trình hoạt động liên thông, chia sẻ giữa các thư viện và giữa các bộ sưu tập khác nhau trong cùng một 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 trong thư viện số. Ontology có thể cung cấp khả năng hoạt động liên thông giữa các tập dữ liệu lớn, cung cấp một khung nhìn cho các tập dữ liệu.

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 mục, đánh dấu các trường với các thuật ngữ thích hợp nhất cho lĩnh vực quan tâm. RDF và OWL 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 số này với các hệ thống siêu dữ liệu thư viện khác.

#### 4. Kết luận

Việc phát triển và tích hợp các ứng dụng với dịch vụ Web3.0 trong thư viện số đang được quan tâm phát triển trong giai đoạn hiện nay, trong đó có thư viện trường đại học. Chúng ta hoàn toàn xây dựng được các mô hình ứng dụng dịch vụ Web3.0 tìm kiếm tài liệu, sách điện tử trong thư viện số, xây dựng mô hình ứng dụng dịch vụ hướng đối tượng nhằm nâng cao hiệu năng các dịch vụ thư viện số. Việc ứng dụng công nghệ Web3.0, theo S. Dsilva, H. Waydande, hứa hẹn sẽ "sắp xếp thông tin của nhân loại" trong thư viện số một

cách đáng kể hơn so với Google.

### TÀI LIỆU THAM KHẢO

- [1] Đặng Tiều Hùng (2004), *Phương pháp biểu diễn ngữ nghĩa lân cận siêu liên kết cho máy tìm kiếm VietSeek*, Luận văn thạc sĩ, Khoa Công nghệ, Đại học Quốc gia Hà Nội.
- [2] Đoàn Sơn (2001), *Các phương pháp biểu diễn và ứng dụng trong khai phá dữ liệu văn bản*, Luận văn thạc sĩ, Khoa Công nghệ, Đại học Quốc gia Hà Nội.
- [3] John Wiley and Sons, *Semantic Web Technologies Trends and Research in Ontology based Systems*, Jul 2006.
- [4] Liyang Yu, *Introduction to the Semantic Web and Semantic Web Services*.
- [5] S. Dsilva, H. Waydande, *Library web3.0: A virtual library - an application of web3.0 tools for library services*, Truy cập trang web <https://library.iated.org> ngày 05/12/2016.