

HƯỚNG ĐẾN MỘT HỆ THỐNG TRI THỨC DỮ LIỆU LỚN TRONG CHẨN ĐOÁN XUẤT HUYẾT NÃO

Lê Thị Hoàng Yến¹, Phan Thượng Cang² và Phan Anh Cang¹

¹Khoa Công nghệ Thông tin, Đại học Sư phạm Kỹ thuật Vĩnh Long

²Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận bài: 15/09/2017

Ngày nhận bài sửa: 10/10/2017

Ngày duyệt đăng: 20/10/2017

Title:

Towards building a large-scale knowledge system for diagnosis of cerebral hemorrhage

Từ khóa:

Dữ liệu lớn, hệ thống tri thức, hướng dịch vụ, xuất huyết não, xử lý ảnh y khoa

Keywords:

Big data, Knowledge management system, Medical image processing, SOA, Stroke

ABSTRACT

Stroke (Brain attack) is not only one of the two reasons leading human death over the world but also the most popular and dangerous cerebrovascular disease. In Vietnam, the lack of the specialized equipment as well as the force of qualified experts has become the massive problem for the accurate diagnosis as well as the efficient and well-timed treatment of stroke, especially intracerebral hemorrhage, the acute type of stroke. Based on analyzing and giving the solutions for the challenges, a service-oriented architecture for the big data-driven knowledge management system on medical intracerebral hemorrhage images was proposed. The suggested architecture provides the capability to develop the intracerebral hemorrhage knowledge systematically, consisting of the knowledge exploration and the knowledge exploitation. As a result, it can contribute to the timely and effective support in the treatment of intracerebral hemorrhage. Besides, the architecture adapts to the modern knowledge service modeling tendency. According to this trend, the knowledge management system can be expended, shared and integrated with more knowledge contributed from specialists, doctors, hospitals, and research institutes.

TÓM TẮT

Tai biến mạch máu não (TBMMN) được xem làm một trong hai nguyên nhân hàng đầu của các ca tử vong trên thế giới và là bệnh lý mạch máu não phổ biến, nguy hiểm nhất hiện nay. Ở Việt Nam, tình hình thiếu trang thiết bị và đội ngũ chuyên gia đang trở thành vấn đề nan giải trong việc chẩn đoán và điều trị hiệu quả TBMMN, đặc biệt là xuất huyết não, dạng cấp tính của TBMMN. Dựa trên việc phân tích các thách thức và đưa ra giải pháp cho các thách thức, trong bài báo này chúng tôi sẽ đề xuất một kiến trúc hướng dịch vụ cho hệ thống tri thức hướng dữ liệu lớn trên ảnh y khoa xuất huyết não. Kiến trúc được đề xuất cung cấp khả năng phát triển có hệ thống tri thức bệnh xuất huyết não bao gồm cả quá trình khai phá và khai thác tri thức; từ đó góp phần vào việc hỗ trợ kịp thời và hiệu quả trong điều trị xuất huyết não. Ngoài ra, kiến trúc cũng đáp ứng xu hướng hiện đại về mô hình dịch vụ tri thức. Theo xu hướng này, hệ thống quản lý tri thức có thể được mở rộng, chia sẻ cũng như nhận được sự đóng góp tri thức từ các chuyên gia, bác sĩ, bệnh viện và các viện nghiên cứu.

Trích dẫn: Lê Thị Hoàng Yến, Phan Thượng Cang và Phan Anh Cang, 2017. Hướng đến một hệ thống tri thức dữ liệu lớn trong chẩn đoán xuất huyết não. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Công nghệ thông tin: 112-118.

1 GIỚI THIỆU

Xếp sau nhồi máu cơ tim, TBMMN được xem là nguyên nhân thứ hai dẫn đến các ca tử vong ở người, và vị trí này đã được giữ nguyên từ năm 2000 đến năm 2015. Không những thế, số ca tử vong do TBMMN đang có xu hướng tăng khi mà số liệu được Tổ chức Y tế Thế giới (WHO) ghi nhận lần lượt là 5,41 và 6,24 triệu ca (WHO, 2017).

WHO định nghĩa TBMMN (hay còn gọi là đột quy) là sự xảy ra đột ngột các thiếu sót chức năng thần kinh, thường là khu trú hơn là lan tỏa, các triệu chứng tồn tại quá 24 giờ hoặc tử vong trong 24 giờ, loại trừ khả năng sang chấn sọ não. TBMMN là bệnh lý mạch máu não nguy hiểm và phổ biến nhất với hai dạng chính là nhồi máu não (nghe hoặc tắc mạch) chiếm 85% và xuất huyết não (vỡ mạch) chiếm 15%. Trong đó, xuất huyết não được xem là dạng cấp tính do nó xảy ra đột ngột, diễn biến của các triệu chứng diễn ra nhanh, nặng lên liên tục, khó hồi phục và thậm chí có thể gây tử vong. Xuất huyết não bao gồm xuất huyết trong não, xuất huyết dưới màng nhện, xuất huyết dưới màng cứng/ngoài màng cứng (Nguyễn Văn Đăng, 2006).

Hiện nay, việc chẩn đoán đột quy đặc biệt là dạng xuất huyết não (cấp tính) phụ thuộc vào hình ảnh thần kinh học với các phương tiện thu nhận ảnh được sử dụng phổ biến bao gồm MRI, CT scanner, DSA, siêu âm. Như một hệ quả, các bác sĩ điều trị đột quy phải làm quen với hầu hết những kỹ thuật chẩn đoán hình ảnh thông thường. Qua đó, ta có thể thấy việc chẩn đoán TBMMN vẫn còn thực hiện thủ công và mang tính chủ quan cao do phụ thuộc quá nhiều vào khả năng quan sát, phân tích và trình độ chuyên môn của các bác sĩ chuyên khoa. Một vấn đề đáng quan ngại hơn đó là lực lượng bác sĩ có đủ trình độ trong việc chẩn đoán này còn rất mỏng lại phần bố không đều giữa các tuyến, đặc biệt là bệnh viện địa phương, nơi đầu tiên tiếp nhận bệnh nhân TBMMN. Thực trạng trên cho thấy tính cấp thiết của việc xây dựng và triển khai hệ thống tri thức hỗ trợ hiệu quả trong chẩn đoán, theo dõi và điều trị kịp thời bệnh nhân TBMMN.

Theo thống kê của Hội Đột quy Mỹ, cứ mỗi 45 giây trôi qua, trên thế giới có ít nhất một người bị đột quy (Vũ Hồng Vân, 2017). Tại Việt Nam, hàng năm có một số lượng lớn bệnh nhân đến khám và điều trị tại các bệnh viện được chụp ảnh CT/MRI phục vụ cho việc chẩn đoán TBMMN với kích thước ảnh ở mỗi lần chụp từ vài trăm MB đến vài GB. Từ đó, ta có thể thấy khi số lượng bệnh nhân tăng lên, lượng dữ liệu ảnh về thần kinh học nói chung, sọ não nói riêng được lưu trữ sẽ trở nên vô cùng lớn theo thời gian. Một khía cạnh khác đáng quan tâm đó là nhu cầu cấp thiết trong việc truy xuất thời gian thực

một cách liên tục, nhanh chóng và chính xác các hình ảnh, thông tin, tri thức từ kho dữ liệu ảnh được xây dựng. Vì vậy, vấn đề tiếp theo cần xem xét đó việc lưu trữ, truy vấn và xử lý lượng dữ liệu lớn hình ảnh thần kinh học trong hệ thống quản lý tri thức hỗ trợ chẩn đoán xuất huyết não.

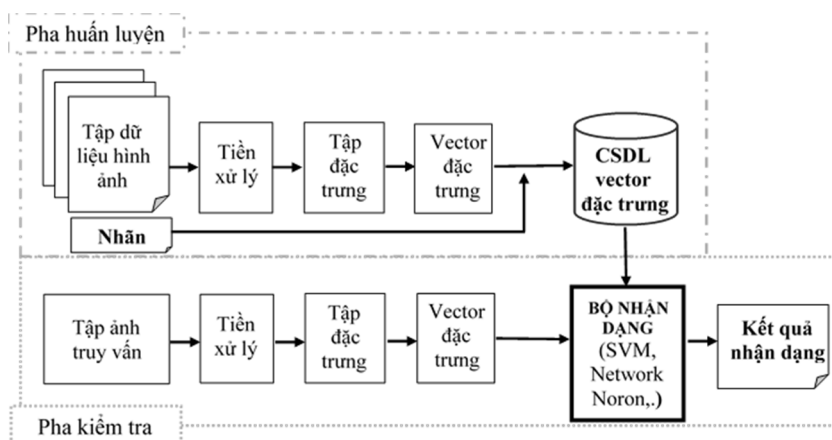
Trong bài báo này, chúng tôi tập trung vào phân tích các thách thức trong xây dựng và triển khai hệ thống quản lý tri thức hỗ trợ chẩn đoán và điều trị bệnh xuất huyết não; từ đó đưa ra các giải pháp hướng đến đề xuất kiến trúc cho hệ thống tri thức dữ liệu lớn trên ảnh y khoa xuất huyết não theo hướng dịch vụ.

2 KHO DỮ LIỆU ẢNH Y KHOA VÀ HỆ THỐNG HỖ TRỢ CHẨN ĐOÁN XUẤT HUYẾT NÃO TỪ ẢNH

Hiện nay, một lượng lớn dữ liệu ảnh CT/MRI cùng với dữ liệu bệnh án, các báo cáo y tế đính kèm (được gọi là meta-data) đang được lưu trữ tại các bệnh viện phục vụ cho công tác điều trị bệnh. Các dữ liệu này kết hợp với các nguồn thông tin khác trên Internet sẽ cung cấp cho đội ngũ y bác sĩ những tư liệu quý giá cho việc chẩn đoán lâm sàng và dự đoán kết quả. Tuy nhiên, thật không may khi các kho dữ liệu này thường ở dạng đóng, không được chia sẻ hay cho phép truy xuất giữa các bệnh viện hoặc giữa các bác sĩ với nhau khi họ có nhu cầu tham khảo các trường hợp bệnh lý tương tự. Thậm chí, không ít trường hợp bác sĩ đã từng điều trị cho bệnh nhân nhưng vẫn không nhớ tên hay các thông tin trước đây có liên quan đến bệnh do việc lưu trữ chủ yếu được thực hiện một cách thủ công trên giấy tờ. Thực trạng này làm cho việc điều trị trở nên khó khăn, thiếu kịp thời và tính hiệu quả chưa cao.

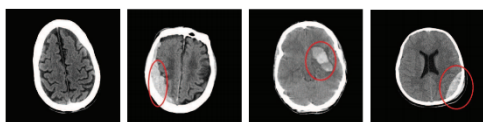
Cùng với sự phát triển của kỹ thuật ảnh y khoa, bên cạnh việc gia tăng về số lượng, độ phức tạp của dữ liệu ảnh sọ não cũng không ngừng được nâng cao. Việc lưu trữ, quản lý cũng như truy vấn nhanh chóng và hiệu quả kho dữ liệu ảnh là vấn đề cần được xem xét. Hay nói khác hơn đó là việc đảm bảo đạt độ chính xác cao trong tìm kiếm và tính toán hiệu quả.

Một trong các thách thức đối với hệ thống chẩn đoán xuất huyết não dựa trên ảnh y khoa thần kinh học đó là việc xử lý, rút trích các đặc trưng quan trọng trên các ảnh để làm cơ sở cho việc khai phá (explore) và khai thác (exploit) tri thức trong chẩn đoán bệnh. Theo một số nghiên cứu đã thực hiện, một hệ thống tự động nhận dạng và phân loại xuất huyết não tổng thể được minh họa như Hình 1 gồm các bước: tiền xử lý ảnh đầu vào, rút trích đặc trưng, lựa chọn đặc trưng, và nhận dạng (Al-Ayyoub et al., 2013; Hingene et al., 2015; Fatima et al., 2015).



Hình 1: Mô hình tổng quát của hệ thống chẩn đoán xuất huyết não qua ảnh CT/ MRI

Dựa trên nền tảng hệ thống tổng thể, trong nghiên cứu của chúng tôi, hệ thống được đề xuất sẽ hỗ trợ việc nhận dạng bốn loại xuất huyết não: xuất huyết dưới màng cứng, xuất huyết trong não và xuất huyết ngoài màng cứng Hình 2 minh họa cho ảnh CT/MRI não bình thường và ba trong bốn loại xuất huyết não.



Hình 2: Ảnh CT/MRI não bình thường và 3 loại xuất huyết não: xuất huyết dưới màng cứng, xuất huyết trong não và xuất huyết ngoài màng cứng (Al-Ayyoub *et al.*, 2013)

Các đặc trưng được nghiên cứu phục vụ cho việc phân loại và nhận dạng các loại xuất huyết não sẽ là: vị trí, kích thước, hình dạng, màu sắc và kết cấu bề mặt của vùng xuất huyết.

3 HỆ THỐNG QUẢN LÝ TRI THỨC HƯỚNG DẪU LIỆU LỚN VÀ DỊCH VỤ TRONG Y KHOA

Tri thức được hình thành một cách có hệ thống từ các đối tượng tri thức được phân loại tùy thuộc vào mức độ phát triển của chúng (dữ liệu, thông tin, và sự hiểu biết) (Bierly III *et al.*, 2000). Quản lý tri thức được định nghĩa như những hoạt động có tính tổ chức (organizational activities) liên quan đến sự tạo lập tri thức mà ở đó quá trình học xảy ra, và tri thức sẽ được tích lũy và phát triển. Trong nghiên cứu của chúng tôi, tri thức ban đầu là dữ liệu hoặc thông tin được thu thập từ các máy CT/MRI, hệ thống thông tin có sẵn, Internet, hoặc mạng viễn thông. Hệ thống quản lý tri thức (KMS - Knowledge Management System) là một dạng của hệ thống thông tin để xử lý tri thức có tổ chức. Nó bao gồm các hoạt động như nắm bắt tri thức (knowledge capture), tổ chức tri thức (knowledge organization),

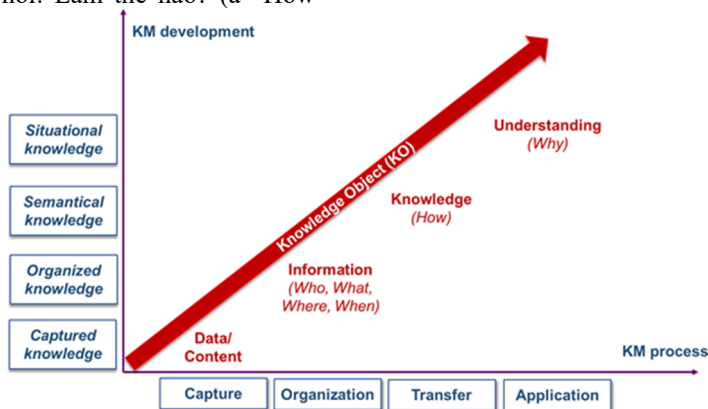
chuyển đổi tri thức (knowledge transfer) và vận dụng tri thức (knowledge application) (Le Dinh *et al.*, 2015). Vì vậy, một kiến trúc hệ thống tri thức phải hỗ trợ đầy đủ bốn hoạt động này.

Quá trình phát triển tri thức trong hệ thống quản lý tri thức được tóm tắt trong Hình 3. Nó bao gồm cả khám phá tri thức (knowledge exploration) và khai thác tri thức (knowledge exploitation). Quá trình khám phá liên quan đến việc nắm bắt (capture) và tổ chức (organization) tri thức, quá trình khai thác liên quan đến việc chuyển đổi (transfer) và ứng dụng (application) tri thức đã được tổ chức trong kho tri thức (Le Dinh *et al.*, 2015). Một đối tượng tri thức được xem là một tập các thành phần tri thức phân tầng từ thấp đến cao (Le Dinh *et al.*, 2015): nhận thức (know-what), điều kiện (know-when), tình huống (know-where), vận dụng (know-how) và cuối cùng là hiểu biết (know-why). Các thành phần tri thức của một đối tượng tri thức có thể được sử dụng và chia sẻ thông qua các quá trình chuyển đổi tri thức.

Theo quan điểm hành vi, một vòng đời của đối tượng tri thức bao gồm: tri thức được nắm bắt (captured knowledge), tri thức đã được tổ chức (organized knowledge), tri thức ngữ nghĩa (semantical knowledge), và tri thức tình huống (situational knowledge). Trước tiên, dữ liệu thô tồn tại ở bất kỳ định dạng nào. Một đối tượng tri thức (Knowledge Object - KO) ở trạng thái tri thức được nắm bắt nếu dữ liệu của nó đã được thu thập và lưu trữ trong kho tri thức. Tiếp theo đó, thông tin là dữ liệu có ý nghĩa dựa trên kết nối quan hệ. Một KO ở trạng thái tri thức tổ chức nếu dữ liệu của nó được tổ chức theo các thành phần tri thức tương ứng với cấu trúc tri thức. Một KO ở mức này có thể trả lời những câu hỏi đơn giản như: Cái gì? Ai? Ở đâu? Khi nào?. Kế đó, tri thức là sự vận dụng thích hợp các thông tin có được ở các hoạt động có tổ chức. Một KO ở trạng thái tri thức ngữ nghĩa nếu các thành

phần tri thức của nó tương ứng với cấu trúc tri thức gắn liền với các thành phần tri thức tương ứng với sự chuyển đổi tri thức (Le Dinh *et al.*, 2014). KO ở mức độ này có thể cung cấp câu trả lời hoặc hướng dẫn trả lời cho câu hỏi: Làm thế nào? (a “How”

question). Cuối cùng, sự hiểu biết là quá trình mà một tổ chức có thể lấy tri thức và tổng hợp tri thức mới từ dữ liệu thô đã được tổ chức trước đó để ra quyết định nghiệp vụ.



Hình 3: Quá trình phát triển tri thức (Le Dinh *et al.*, 2015)

Sự phát triển và những lợi ích của hệ thống quản lý tri thức đang có tác động trực tiếp đến quyết định và quy trình nghiệp vụ trong các tổ chức hiện đại. KMS làm cho các tổ chức cạnh tranh hơn trong việc nắm bắt cơ hội. Tuy nhiên, các KMS hiện tại phải đối mặt với khối lượng dữ liệu vô cùng lớn mà trước đây chưa từng gặp phải, đó là kết quả từ các hoạt động nghiệp vụ, các loại dịch vụ công nghệ thông tin khác nhau, các dạng hình ảnh và video kích thước lớn... được gọi chung là dữ liệu lớn hay “big data”. Việc tính toán xử lý trên một lượng dữ liệu rất lớn, tạo ra những rào cản đáng kể cho các tổ chức khai thác hiệu quả giá trị của dữ liệu lớn.

Mặc dù gần đây các nghiên cứu về dữ liệu lớn đã có được những chuyển biến nhanh chóng, nhưng nó vẫn thiếu các framework và kiến trúc cho phép các tổ chức có thể nắm bắt giá trị dữ liệu lớn một cách có hệ thống, đặc biệt là thúc đẩy việc nghiên cứu. Theo đó, một trong những thách thức quan trọng nhất cho KMS ngày nay là làm sao để giải quyết các tập tin dữ liệu lớn mà nó yêu cầu nội dung phải được cập nhật liên tục thường xuyên. Do đó, một thể hệ mới của KMS có thể xử lý hiệu quả nguồn dữ liệu lớn trở thành một nhu cầu thiết yếu cho các tổ chức, đặc biệt là các tổ chức hiện đại hướng dữ liệu.

Một xu hướng hiện đại cho các KMS là hệ thống hướng dịch vụ. Kiến trúc hướng dịch vụ (Service Oriented Architecture - SOA) là tập hợp các dịch vụ kết nối ‘mềm dẻo’ với nhau (nghĩa là một ứng dụng có thể ‘nói chuyện’ với một ứng dụng khác mà không cần biết các chi tiết kỹ thuật bên trong), các giao tiếp (dùng để gọi hàm dịch vụ) được định nghĩa

rõ ràng và độc lập với nền tảng hệ thống, và có thể tái sử dụng. SOA là cấp độ cao hơn của phát triển ứng dụng, chú trọng đến quy trình nghiệp vụ và dùng giao tiếp chuẩn để giúp che đi sự phức tạp kỹ thuật bên dưới. Điểm mấu chốt của SOA là thiết kế SOA tách riêng phần thực hiện dịch vụ (phần mềm) với giao tiếp gọi dịch vụ. Các chức năng của dịch vụ được công bố dưới dạng một giao diện chuẩn. Chi tiết cài đặt cụ thể các chức năng được che giấu, và người dùng dịch vụ không cần quan tâm đến; người dùng chỉ gọi sử dụng các chức năng của dịch vụ thông qua các phương thức công bố trên giao diện. KMS hướng dịch vụ cho phép các tổ chức có thể khai thác tri thức một cách có hiệu quả thông qua các dịch vụ của nó (Le Dinh *et al.*, 2016).

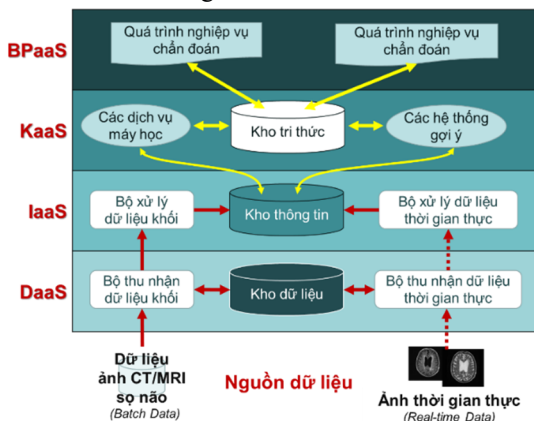
Bên cạnh đó, một khía cạnh khác cần xem xét đó là các nghiên cứu về quản lý tri thức và các KMS trong y khoa trong thời gian qua, nhìn chung đều thiếu đi phân phân tích chi tiết về việc triển khai thực tế và những bài học kinh nghiệm khi triển khai quản lý tri thức dựa trên nền tảng công nghệ thông tin trong công nghiệp chăm sóc sức khỏe (Chen *et al.*, 2013). Đa số các KMS được đề xuất chỉ hỗ trợ chẩn đoán lâm sàng, thiếu đi các chẩn đoán chuyên sâu. Việc hỗ trợ chẩn đoán y khoa qua hình ảnh thực tế vẫn dựa trên các thông tin được phân tích bởi các chuyên gia, sau đó được đánh kèm vào ảnh (Demigha and Balleyguier, 2014; Baigorri *et al.*, 2013). Đây cũng là động cơ thúc đẩy việc nghiên cứu của chúng tôi nhằm đề xuất và triển khai một kiến trúc mới cho các KMS hỗ trợ chẩn đoán tự động và giáo dục y khoa nói chung, xuất huyết não nói riêng.

4 ĐỀ XUẤT KIẾN TRÚC CHO HỆ THỐNG QUẢN LÝ TRI THỨC HƯỚNG DẪN LIỆU LỚN TRÊN ẢNH Y KHOA TRONG CHẨN ĐOÁN XUẤT HUYẾT NÃO

a. Kiến trúc tổng quan

Từ những vấn đề đã trình bày, tri thức bệnh xuất huyết não cần được phát triển một cách có hệ thống bao gồm cả khai phá (explore) và khai thác (exploit) tri thức. Dựa trên phương pháp nghiên cứu khoa học thiết kế (Hevner *et al.*, 2004), chúng tôi đề xuất một kiến trúc mới hướng dịch vụ cho các hệ thống quản lý tri thức hướng dữ liệu lớn trên ảnh y khoa trong chẩn đoán xuất huyết não. Kiến trúc sẽ bao gồm các thành phần: tập các cấu trúc (set of constructs) là các kiểu quan niệm khác nhau liên quan đến đối tượng tri thức bệnh xuất huyết não; một mô hình (model) diễn tả mối quan hệ giữa các quan niệm tri thức; một phương thức (method) là tập các hoạt động hỗ trợ quá trình quản lý tri thức trong KMS và cuối cùng là một minh họa cụ thể (instantiation) cho sự vận hành của hệ thống.

Theo kiến trúc này, dữ liệu cho các KMS hướng dữ liệu lớn dựa trên ảnh y khoa ứng dụng trong chẩn đoán và giáo dục y khoa (lĩnh vực xuất huyết não) sẽ bao gồm cả dữ liệu ảnh truyền thống và dữ liệu ảnh lớn thời gian thực. Tri thức có cấu tạo từ các đối tượng tri thức được phân lớp dựa trên mức độ phát triển của chúng, có thể là dữ liệu, thông tin, tri thức và trí tuệ. Khi đó, hệ thống được xây dựng cần có các khả năng: song song hóa các hoạt động xử lý, rút trích và lựa chọn đặc trưng quan trọng trong ảnh y khoa xuất huyết não, nhận dạng; hỗ trợ truy vấn ảnh dựa trên nội dung; xếp hạng các kết quả thu được; phân loại các dạng xuất huyết não chủ yếu; hỗ trợ ra quyết định giúp chẩn đoán tự động xuất huyết não từ ảnh y khoa dữ liệu lớn. Đặc biệt, hệ thống cũng sẽ có khả năng cung cấp tri thức ở các cấp độ khác nhau dưới dạng các dịch vụ.



Hình 4: Kiến trúc tổng quan được đề xuất

Để có thể hỗ trợ cho toàn bộ quá trình phát triển tri thức bệnh xuất huyết não, kiến trúc cho hệ thống quản lý tri thức hướng dữ liệu lớn trên ảnh y khoa trong chẩn đoán xuất huyết não (Hình 4) được đề xuất dựa trên bốn hoạt động chính của quá trình quản lý tri thức: thu nhận (capture), tổ chức (organization), chuyển đổi (transfer) và ứng dụng (application) tri thức.

– **Nguồn dữ liệu** bao gồm dữ liệu khối và dữ liệu thời gian thực liên quan đến bệnh xuất huyết não, đặc biệt là các ảnh CT/MRI sọ não.

– **Data-as-a-Service (DaaS)** là tầng thấp nhất có các thành phần thực hiện thu thập, lưu trữ hình ảnh xuất huyết não và các thông tin liên quan dưới dạng dữ liệu khối (batch data) và dữ liệu thời gian thực (real-time streaming data). DaaS có các dịch vụ cung cấp dữ liệu (hình ảnh thô).

– **Information-as-a-Service (IaaS)** là nơi dữ liệu khối và dữ liệu thời gian thực được đưa qua các bộ xử lý tương ứng, biến đổi thành thông tin bệnh xuất huyết não. Những thông tin này sẽ được cung cấp đến người dùng thông qua Dịch vụ thông tin (information services).

– **Knowledge-as-a-Service (KaaS)** là nơi sẽ sinh ra các tri thức bệnh xuất huyết não từ kho thông tin của tầng IaaS thông qua các hoạt động máy học hay hệ thống gợi ý. Tầng này sẽ gồm cơ sở tri thức phân tán, cấu trúc tri thức, chủ thích, các chức năng được đóng gói và công bố dưới dạng các dịch vụ tri thức (knowledge services).

– **Business Process-as-a-Service (BPaaS)** là tầng cao nhất, cho phép người dùng hình dung và phân tích kết quả truy vấn của họ. BPaaS cung cấp các dịch vụ đến ứng dụng bằng cách kết hợp chúng với các quá trình nghiệp vụ chẩn đoán xuất huyết não.

b. Giải pháp triển khai cụ thể các tầng

Tầng nguồn dữ liệu: Dữ liệu khối sẽ được cung cấp dưới dạng các tập tin hoặc cơ sở dữ liệu ảnh CT/MRI sọ não được lưu trữ tại các bệnh viện cùng với các meta-data đính kèm (thông tin bệnh nhân và các báo cáo y tế có liên quan). Dữ liệu thời gian thực cũng là các ảnh CT/MRI sọ não nhưng được gửi đến hệ thống liên tục trong quá trình khám và điều trị.

Tầng DaaS: HDFS (Hadoop Distributed File System) được biết đến là một hệ thống tập tin phân tán có khả năng chịu lỗi cao của Apache Hadoop. Đây cũng là một framework chiếm ưu thế trong lĩnh vực xử lý dữ liệu lớn với cơ sở hạ tầng lớn đang được triển khai và sử dụng trong nhiều lĩnh vực ứng dụng khác nhau. Thiết kế của HDFS đáng tin cậy cho việc lưu trữ các tập dữ liệu lớn cũng như truyền chúng đến các ứng dụng người dùng với băng thông

cao. Vì vậy, chúng tôi lựa chọn HDFS cho việc lưu trữ dữ liệu đã được tái cấu trúc cũng như làm sạch từ tầng nguồn dữ liệu của hệ thống.

Nạp một lượng lớn dữ liệu từ các nguồn dữ liệu thời gian thực hoặc nguồn dữ liệu khối vào trong HDFS không khả thi đối với các kỹ thuật truyền thống. Vì vậy, chúng tôi xem xét các công cụ thu nhận dữ liệu khác có tính khả thi hơn. Với dữ liệu khối, Apache Sqoop được dùng do tính hiệu quả trong việc xuất và nhập lượng lớn dữ liệu giữa HDFS và kho dữ liệu có cấu trúc như các cơ sở dữ liệu quan hệ. Bên cạnh đó, chúng tôi sử dụng các công cụ ETL (Extract, Load, Transform Tools) như các lệnh Hadoop FS shell hoặc viết các Java client API để nạp một lượng lớn dữ liệu từ tập tin văn bản hoặc NoSQL vào HDFS. Mặt khác, Apache Kafka là hệ thống hàng đợi dữ liệu (message queue) phân tán, tin cậy, thông lượng cao và độ trễ thấp. Kafka và Spark được chúng tôi kết hợp trong việc xử lý dữ liệu luồng cũng như tận dụng được những lợi thế của chúng.

Bên cạnh đó, chúng tôi cài đặt WebHDFS cho Hadoop để triển khai các dịch vụ dữ liệu. WebHDFS cho phép người dùng truy cập dữ liệu trong HDFS sử dụng cơ chế RESTful chuẩn công nghiệp. Ngoài ra, gói Hive-server2 được cài đặt cho phép người dùng thực hiện các truy vấn giống SQL và nhận về kết quả.

Tầng IaaS: Apache HBase là một cơ sở dữ liệu NoSQL hướng cột (column-oriented) được xây dựng dựa trên HDFS. Nó hỗ trợ truy xuất thời gian thực, ngẫu nhiên đến HDFS và tính năng xử lý bó. Như một kết quả, HBase được chúng tôi sử dụng để triển khai kho tri thức. HBase không chỉ cung cấp sự tiết kiệm và Restful gateway mà còn các gateway dịch vụ web cho việc tích hợp và truy xuất HBase. Vì vậy, nó thuận lợi cho việc triển khai các dịch vụ thông tin.

Không giống với các kiến trúc KMS thông thường đã được thiết kế cho dữ liệu nội bộ và có cấu trúc, kiến trúc KMS hướng dữ liệu lớn làm việc tốt với dữ liệu thô có và không có cấu trúc cũng như dữ liệu bên trong và dữ liệu bên ngoài. Thêm vào đó, nó được yêu cầu để có khả năng xử lý dữ liệu khối và dữ liệu gần thực. Như một kết quả, chúng tôi sử dụng Apache Spark cho các xử lý này. Một nguyên nhân khác cho việc chọn Spark đó là trong khi Hadoop hoạt động trên đĩa, Spark lại xử lý với dữ liệu trong bộ nhớ nhanh và tốt hơn, có khả năng hỗ trợ các tác vụ tính toán chuyên sâu. Bên cạnh đó, Spark sử dụng mô hình lưu trữ dữ liệu, được gọi là tập dữ liệu phân tán linh hoạt (RDD - Resilient distributed datasets) đảm bảo khả năng chịu lỗi và tối ưu hoạt động xuất nhập.

Với mô hình dữ liệu giống đồ thị phổ dụng linh hoạt, RDF (Resource Description Framework) là chuẩn phổ biến cho các nguồn lực miêu tả và dữ liệu ngữ nghĩa công bố trên Web. Chúng tôi sử dụng RDF như một phương thức chung cho việc mô hình hóa thông tin.

Apache Jena Elephas là một tập thư viện để truy xuất dữ liệu RDF trong HDFS. Jena là một bộ công cụ các lập trình viên Web ngữ nghĩa hàng đầu. Sự phối hợp của Jena, HBase, và Spark tạo một kho RDF phân tán, có khả năng mở rộng và hiệu quả.

Spark Core làm việc với dữ liệu khối từ kho dữ liệu (HDFS) để tổ chức nội dung theo ngữ nghĩa của chúng cũng như tạo và duy trì cơ sở dữ liệu tri thức (dùng HBase). Jena Elephas được dùng để chuyển dữ liệu không cấu trúc thành dữ liệu liên kết (RDF). Spark phân tích dữ liệu thông qua các kỹ thuật ánh xạ những đặc tả của nguồn dữ liệu vào đặc tả tri thức. Các xử lý luồng của Spark liên quan đến việc ánh xạ liên tục dữ liệu đầu vào từ Kafka thành tri thức theo thời gian thực.

Tầng KaaS: Từ quan điểm kỹ thuật, hệ thống xây dựng theo nguyên lý SOA được cấu tạo từ các dịch vụ. Chúng được định nghĩa bởi các ngôn ngữ mô tả với giao diện để hỗ trợ quá trình nghiệp vụ, và được triển khai bằng những ngôn ngữ lập trình khác nhau. Như một kết quả, chúng tôi tìm thấy một kỹ thuật lý tưởng, Web services phù hợp để triển khai tất cả các dịch vụ tri thức. Web services là kỹ thuật phổ biến và nổi tiếng nhất để triển khai SOA. Chúng tôi sử dụng kết hợp Apache Axis2/Java, Jena Elephas, SPARQL, và Spark. Spark MLlib là một thành phần chủ yếu để thực hiện các dịch vụ máy học cho phép xây dựng các mô hình dữ liệu.

Tầng BPaaS: Chúng tôi sử dụng Eclipse SOA Tools Platform để mô hình hóa quá trình nghiệp vụ. Tiếp theo, chúng tôi chuyển đổi tập tin BPMN (Business Process Modeling Notation) thành BPEL (Business Process Execution Language) với sự hỗ trợ của ngôn ngữ biến đổi ATLAS (ATL). Cuối cùng, chúng tôi triển khai nó trên Apache Orchestration Director Engine.

5 KẾT LUẬN

Có nhiều thách thức được đặt ra trong việc xây dựng hệ thống quản lý tri thức hỗ trợ công tác chẩn đoán và điều trị bệnh xuất huyết não trong thời đại big data như hiện nay: xây dựng và truy vấn kho ảnh y khoa (ảnh xuất huyết não) dữ liệu lớn từ nhiều nguồn khác nhau; chuyển hóa một cách có hệ thống dữ liệu ban đầu thành thông tin, tri thức vào kho tri thức dữ liệu lớn; phân loại, nhận dạng, hỗ trợ ra quyết định giúp chẩn đoán y khoa; xây dựng hệ

thống hướng dịch vụ quản lý tri thức dữ liệu lớn hỗ trợ chẩn đoán y khoa.

Trong nghiên cứu này, chúng tôi đề xuất một định hướng mới cho kiến trúc của KMS trong chẩn đoán xuất huyết não. Kiến trúc mới cung cấp khả năng phát triển tri thức bệnh xuất huyết não một cách có hệ thống và đầy đủ (bao gồm khai phá và khai thác tri thức). Các KMS được triển khai có thể hỗ trợ hiệu quả và kịp thời cho việc chẩn đoán cũng như điều trị, từ đó có thể nâng cao chất lượng chẩn đoán xuất huyết não nói riêng, chẩn đoán y khoa nói chung.

Ngoài ra, kiến trúc đề xuất có thể được mở rộng nhằm phát triển hệ thống tri thức toàn diện hơn phục vụ chẩn đoán và giáo dục y khoa. Dựa trên kiến trúc, hệ thống được xây dựng sẽ phát triển theo hướng mở và chia sẻ (SOA) với sự tham gia đóng góp từ các chuyên gia, bác sĩ, bệnh viện và viện nghiên cứu. Hướng đến tầm xa hơn, kiến trúc được đề xuất sẽ có tiềm năng thúc đẩy sự phát triển y tế cộng đồng tại Việt Nam. Chúng tôi hiện đang phát triển một hệ thống quản lý tri thức dựa trên kiến trúc đề xuất trên và sẽ triển khai thử nghiệm tại một bệnh viện thành phố Cần Thơ.

TÀI LIỆU THAM KHẢO

- Al-Ayyoub, M., Alawad, D., Al-Darabsah, K., Aljarrah, I., 2013. Automatic Detection and Classification of Brain Hemorrhages. WSEAS TRANSACTIONS on COMPUTERS. Vol. 12 (10): 395-405.
- Baigorri A., Villadangos, J., Astrain, J., A. Córdoba, 2013. A medical knowledge management system based on expert tagging (MKMST). Data Management and Security: Applications in Medicine, Sciences and Engineering. WIT Transactions on Information and Communication Technologies. Vol. 45: 221-231.
- Bierly III, P. E., Kessler, E. H., Christensen, E. W., 2000. Organizational learning, knowledge and wisdom. Journal of Organizational Change Management. Vol. 13(6): 595-618.
- Chen, Edward T., 2013. An Observation of Healthcare Knowledge Management. Communications of the IIMA: Vol. 13: Iss. 3, Article 7.
- Demigha, S., Balleyguier, C., 2014. KMSS: A Knowledge Management System for Senology. Proceedings of ECKM2014, Vol.1: 268-277.
- Fatima, Sridevi M, Naza, S., Anjum, K., 2015. Diagnosis and Classification of Brain Hemorrhage using CAD System. Proceeding of NCRIET-2015 & Indian J.Sci.Res. Vol 12(1):121-125.
- Hevner, A. R., March, S. T., Park, J., Ram, S., 2004. Design Science in Information Systems Research. MIS Quarterly. Vol. 28(1): 75-105.
- Hingene, M.C., Matkar, S. B., Mane, A. B., Shirsat, A. M., 2015. Classification of MRI Brain Image using SVM Classifier. LISTE - International Journal of Science Technology & Engineering. Vol. 1(9): 24-28.
- Le Dinh, T., Rickenberg, T. A., Fill, H.-G., Breitner, M. H., 2015. Enterprise Content Management Systems as a Knowledge Infrastructure: The Knowledge-Based Content Management Framework. International Journal of e-Collaboration. Vol. 11(3): 49-70, IGI Global.
- Le Dinh, T., Ho Van, T., Moreau, E., 2014. A Knowledge Management framework for Knowledge-Intensive SMEs. Proceedings of 16th International Conference on Enterprise Information Systems, S. Hammoudi, L. Maciaszek, and J. Cordeiro, Lisbon, Portugal, pp. 435-440.
- Le Dinh, T., Phan Thuong, C., Bui T., 2016. Towards an Architecture for Big Data-Driven Knowledge Management Systems. AMCIS 2016, August 11-14 2016, San Diego, CA, USA.
- Nguyễn Văn Đăng, 2006. Tai biến mạch máu não. Nhà xuất bản Y học, Hà Nội, 240 trang.
- Vũ Hồng Vân, 2017. Đột quỵ/Tai biến mạch máu não: bệnh lý mạch máu não nguy hiểm nhất, ngày truy cập 15/01/2017. Địa chỉ: <http://noitonghop.org/dot-quy-tai-bien-mach-nao-benh-ly-mach-mau-nao-nguy-hiem-nhat/>
- WHO, 2017. The top 10 causes of death, accessed on 22 March 2017. Available from <http://www.who.int/mediacentre/factsheets/fs310/en/>