

PHƯƠNG PHÁP THIẾT KẾ HỆ CƠ SỞ TRI THỨC DỰA TRÊN ONTOLOGY VÀ ỨNG DỤNG

• Đỗ Văn Nhon • Mai Trung Thành* • Hoàng Ngọc Long
Trường Đại học Quốc tế Hồng Bàng

TÓM TẮT

Hệ cơ sở tri thức là một trong các lớp ứng dụng phổ biến của trí tuệ nhân tạo. Thiết kế cơ sở tri thức và bộ suy diễn của hệ thống ứng dụng thực tế đòi hỏi việc sử dụng ontology cho biểu diễn tri thức. Trong các ứng dụng như hệ giải vấn đề thông minh trong các miền tri thức hình học phẳng, điện học, hay hóa học thì tri thức có cấu trúc rất trừu tượng và phức tạp bao gồm các khái niệm, các quan hệ, các toán tử, các hàm, các luật. Giải pháp biểu diễn tri thức theo ontology là tất yếu để biểu diễn các miền tri thức như trên và tổ chức cơ sở tri thức cho hệ thống trên máy tính. Trong bài báo này sẽ trình bày phương pháp thiết kế cơ sở tri thức và bộ suy diễn của hệ cơ sở tri thức dựa trên môn ontology cho miền tri thức, gọi là ontology về cơ sở tri thức các đối tượng tính toán, và các ứng dụng của nó trong thực tế như hệ giải toán trong miền tri thức hình học, hệ chẩn đoán bệnh.

Từ khóa: hệ cơ sở tri thức, biểu diễn tri thức, suy diễn, ontology

AN ONTOLOGY BASED METHODS FOR DESIGNING KNOWLEDGE-BASED SYSTEMS AND APPLICATIONS

• Do Van Nhon • Mai Trung Thanh • Hoang Ngoc Long

ABSTRACT

Designing of the knowledge bases and the inference engines of knowledge-based systems in practice requires representations in the form of ontologies. For applications such as the intelligent problem solver in plane geometry, the knowledge contains a complicated system of concepts, relations, operators, functions, and rules. This situation motivates an ontology-based solution with such components. In this article, an ontology called Ontology of Computational Object Knowledge Base (COKB), will be presented in details. We also present a model for representing problems together reasoning algorithms for solving them, and design methods to construct applications. The above methodology has been used in designing many applications such as systems for solving analytic geometry problems, solving problems in plane geometry, and the expert system for diabetic micro vascular complication diagnosis.

Keywords: knowledge-based system, ontology, knowledge representation

1. TỔNG QUAN

Trong việc ứng dụng CNTT, đặc biệt là các hệ cơ sở tri thức, hiện nay là một trong những đòi hỏi trong nhiều lĩnh vực của thời kỳ cách mạng công nghiệp 4.0. Các ứng dụng dựa trên hệ cơ sở tri thức đã nhận được rất nhiều sự quan tâm của các doanh nghiệp, tổ chức, cũng như chính phủ. Hệ cơ sở tri thức là những hệ thống thông tin dựa trên tri thức, cho phép mô hình hóa các tri thức của

* Tác giả liên hệ: ThS. Mai Trung Thành, Email: thanhmt@hiu.vn

(Ngày nhận bài: 10/10/2022; Ngày nhận bản sửa: 11/11/2022; Ngày duyệt đăng: 16/11/2022)

chuyên gia, dùng tri thức này để giải quyết vấn đề phức tạp thuộc cùng lĩnh vực. Một hệ cơ sở tri thức có 2 thành phần quan trọng đó là cơ sở tri thức (knowledge base) và bộ suy diễn (inference engine). Cơ sở tri thức bao gồm các sự kiện, các luật, các khái niệm và các quan hệ. Bộ suy diễn là bộ xử lý tri thức theo mô hình hóa được lập luận bởi chuyên gia. Bộ suy diễn hoạt động trên thông tin về vấn đề đang xét, so sánh với tri thức lưu trong cơ sở tri thức rồi rút ra kết luận. Do đó, có thể nói rằng tính hiệu quả của hệ cơ sở tri thức phần lớn phụ thuộc vào cơ sở tri thức của hệ thống. Để xây dựng các hệ thống trên, nhà thiết kế phải xây dựng được một cơ sở tri thức và một bộ suy diễn phù hợp để giải quyết các vấn đề dựa trên tri thức. Chất lượng hoạt động của các hệ thống thông minh này phụ thuộc rất lớn vào cách biểu diễn, tổ chức cơ sở tri thức đã có và phương pháp suy luận giải quyết vấn đề trên tri thức. Tuy nhiên, cơ sở khoa học kỹ thuật hiện nay liên quan đến thiết kế cơ sở tri thức vẫn còn chưa hoàn chỉnh và còn nhiều vấn đề phải nghiên cứu phát triển.

Hiện nay có nhiều phương pháp biểu diễn tri thức trong thiết kế các ứng dụng thông minh, theo [1-3] ta có thể phân thành các nhóm sau đây:

Các phương pháp biểu diễn dựa trên logic hình thức: đây là phương pháp cổ điển nhất để biểu diễn tri thức trên máy tính, dựa trên hai dạng phổ biến đó là logic mệnh đề và logic vị từ. Điểm mạnh của phương pháp này đó là có tính hình thức cao, được nghiên cứu khá đầy đủ về cơ sở lý thuyết, vấn đề, thuật giải và có các ngôn ngữ lập trình dành riêng cho loại biểu diễn tri thức này, có thể kể đến như PROLOG, LISP. Điểm khó khăn của phương pháp này đó là không can thiệp được sâu hơn cấu trúc của mệnh đề, cũng như không biểu diễn được các loại tri thức có cấu trúc phức tạp, nhiều loại thành phần và quan hệ.

Các phương pháp biểu diễn tri thức dạng thủ tục: phương pháp này bao gồm một tập các chỉ thị hay các bước, hành động cụ thể, mỗi thủ tục thường được xây dựng để giải quyết một bài toán, vấn đề theo ngữ cảnh cụ thể. Phương pháp này khá linh động trong việc thiết kế để giải một số vấn đề nhỏ, mang tính ngữ cảnh (context), cung cấp đủ phương thức để giải quyết đúng, đủ các vấn đề cụ thể. Điểm hạn chế của phương pháp này đó là quan tâm nhiều về các thao tác tính toán, hành động, xử lý để giải quyết vấn đề, chưa quan tâm nhiều đến việc biểu diễn cấu trúc phức tạp của tri thức, cũng như là tri thức nhiều thành phần, nhiều mối quan hệ.

Các phương pháp biểu diễn tri thức dạng luật dẫn: phương pháp này cung cấp giải pháp biểu diễn là tập các yếu tố và tập các luật dẫn có tính chất kéo theo. Mỗi luật dẫn có dạng $h(r) \# g(r)$, trong đó $h(r)$ là tập các các sự kiện giả thiết của luật r , $g(r)$ là tập các sự kiện kết luận của luật r . Phương pháp này khá phổ biến trong tri thức thực tế, tỏ ra hiệu quả với các tri thức có dạng luật dẫn. Điểm khó khăn của giải pháp này đó là sự kiện được đề cập chưa có sự định nghĩa rõ ràng là gì? Và chúng hướng nhiều đến các dạng tri thức luật dẫn (hỗ trợ sự suy luận trên luật), chưa quan tâm nhiều đến biểu diễn cấu trúc, phân loại tri thức vì vậy tính biểu diễn tri thức chưa được cao. Hơn thế nữa, với những tri thức có cả hàng ngàn luật, hoặc là luật có sự phân nhóm thì về mặt lý thuyết của giải pháp này vẫn chưa được đề cập đến và kỹ thuật xử lý cho vấn đề này.

Các phương pháp biểu diễn dạng mạng: là phương pháp có dạng như đồ thị có cấu trúc gồm tập nút và tập các cung. Trong đó mỗi nút biểu diễn hay đại diện cho các đối tượng, mỗi cung biểu diễn mối quan hệ giữa các đối tượng. Các phương pháp dạng này có tính trực quan cao, dễ hiểu, cách biểu diễn khá tự nhiên, và rất linh động khi có thể thêm nút mới hoặc các quan hệ giữa các nút một cách thuận tiện. Tuy nhiên, hiện nay chưa có một quy chuẩn nào cụ thể nói về các nút và các cung. Vì vậy khi sử dụng, nhà thiết kế phải tự mình định nghĩa lại về cấu trúc của nút và cung này. Hơn thế nữa, việc linh động trong quá trình thêm nút và cung có thể dẫn đến các mối quan hệ mâu thuẫn trong tri thức và rất khó kiểm soát.

Các phương pháp biểu diễn cấu trúc: Phương pháp này cho phép sử dụng các cấu trúc dữ liệu phức tạp và các cấu trúc dữ liệu trừu tượng trong biểu diễn, có thể kể đến một số cấu trúc phổ biến như: Frames, Class. Ý tưởng chính của các phương pháp này, thường tập trung vào biểu diễn tri thức của một đối tượng hay khái niệm, vì vậy mọi tri thức liên quan đến khái niệm sẽ được đóng gói lại thành một khung khép kín trong nội tại của đối tượng. Điểm mạnh của phương pháp này đó là cho phép biểu diễn được nhiều tri thức liên quan đến một đối tượng gồm các thuộc tính, hành vi, ...vv, và thể hiện được tính tự nhiên của tri thức. Tuy nhiên cũng như các phương pháp ở trên, Frame, Class chưa đặt nhiều trọng tâm trong việc biểu diễn tri thức có nhiều thành phần khác và các mối quan hệ giữa các khái niệm hay đối tượng.

Nghiên cứu và phát triển mô hình cho dạng cơ sở tri thức theo tiếp cận Ontology là một cách tiếp cận hiện đại để thiết kế các thành phần tri thức của hệ cơ sở tri thức. Ontology ngày nay càng trở nên phổ biến trong khoa học máy tính. Ontology là một khái niệm bắt nguồn từ triết học, trong lĩnh vực khoa học máy tính theo Tom Gruber [4] thì Ontology là một “*đặc tả của hình thức cho hệ thống khái niệm trong tri thức của con người để có thể chia sẻ được và xử lý được trên máy tính*”. Phương pháp tiếp cận này tỏ ra mạnh mẽ và phù hợp với hiện đại, vì có thể dễ dàng thấy rằng mọi miền tri thức của con người hiện nay đều được hình thành từ các khái niệm, thuộc tính, tính chất và các quan hệ giữa các khái niệm này. Điều này cho phép Ontology là công cụ đắc lực cho nhà thiết kế, để thiết kế được các cơ sở tri thức cho xây dựng các ứng dụng thông minh trong khoa học về trí tuệ nhân tạo. Tuy nhiên, dù là một giải pháp hiện đại và mạnh mẽ trong biểu diễn tri thức, Ontology hiện nay vẫn là một phương pháp khá tổng quát và mang nhiều tính chất về định hướng hơn là một phương pháp cụ thể có sự hỗ trợ về kỹ thuật, phương pháp xây dựng. Điều này dẫn đến nhà thiết kế sẽ tốn rất nhiều công sức và thời gian để có thể đưa ra được các định nghĩa phù hợp theo phạm vi ngữ cảnh hay phạm vi tri thức cụ thể. Trong bài báo này, nhóm nghiên cứu sẽ trình bày phương pháp thiết kế cơ sở tri thức và bộ suy diễn của hệ cơ sở tri thức dựa trên một Ontology cho miền tri thức, gọi là Ontology về cơ sở tri thức các đối tượng tính toán và các ứng dụng của nó trong thực tế như hệ giải toán trong miền tri thức hình học, hệ chẩn đoán bệnh.

2. HỆ CƠ SỞ TRI THỨC – KIẾN TRÚC VÀ QUY TRÌNH THIẾT KẾ

2.1. Hệ cơ sở tri thức và kiến trúc hệ thống

2.1.1. Hệ cơ sở tri thức

Tri thức được hiểu là kiến thức hay sự hiểu biết của con người trong các lĩnh vực hay miền tri thức nhất định, như kiến thức về toán học nói chung, hay các miền tri thức toán học với phạm vi thu hẹp hơn như kiến thức về hình học phẳng, hình học không gian, hình học giải tích; kiến thức về vật lý nói chung hay kiến thức về điện một chiều, điện xoay chiều, cơ học; kiến thức về các loại bệnh trong y khoa, v.v. Tri thức thường bao gồm nhiều thành phần trừu tượng, cùng với những mối liên hệ rất đa dạng và phức tạp giữa các thành phần [3]. Các thành phần tri thức thường gặp bao gồm thành phần tri thức khái niệm, thành phần tri thức quan hệ, thành phần tri thức toán tử, thành phần tri thức hàm, các sự kiện và các luật; trong đó thành phần tri thức khái niệm cùng với các quan hệ cơ bản trên khái niệm là phần nền tảng của tri thức. Tri thức còn bao gồm cả kinh nghiệm và các quy tắc hỗ trợ cho quá trình suy luận giải quyết các vấn đề liên quan đến tri thức trong một lĩnh vực hay phạm vi tri thức nhất định.

Một hệ cơ sở tri thức sẽ giúp chúng ta quản lý tri thức của các miền tri thức khác nhau trên máy tính, giúp giải quyết những vấn đề liên quan đến tri thức; cao cấp hơn là hệ thống có khả năng suy luận hay suy diễn giải quyết những vấn đề trừu tượng và phức tạp. Hệ thống thường có một tổ chức lưu trữ tri thức phục vụ cho việc giải quyết các yêu cầu hay vấn đề ứng dụng được đặt ra; tổ chức tri thức này được gọi là cơ sở tri thức của hệ thống [5].

Các hệ cơ sở tri thức có thể được phân loại dựa vào các khía cạnh với những tiêu chuẩn đặc trưng nhất định như dựa vào tính đóng mở, dựa vào phương pháp biểu diễn tri thức, dựa vào mục tiêu ứng dụng và kiến trúc cũng như sự vận hành của hệ thống, hay phân loại theo lĩnh vực ứng dụng.

Một hệ cơ sở tri thức đóng có cơ sở tri thức được xây dựng với một số “tri thức lĩnh vực” ban đầu, và chỉ những tri thức đó mà thôi trong suốt quá trình hoạt động hay suốt thời gian sống của nó. Hệ cơ sở tri thức mở có khả năng bổ sung tri thức trong quá trình hoạt động, thường là có mô-đun khám phá tri thức dựa trên các phương pháp máy học. Các hệ cơ sở tri thức kết hợp có những đặc tính từ sự kết hợp giữa hệ đóng và hệ mở, kết hợp giữa cơ sở tri thức và cơ sở dữ liệu, kết hợp giữa hệ cơ sở tri thức này với một hệ cơ sở tri thức khác, v.v.

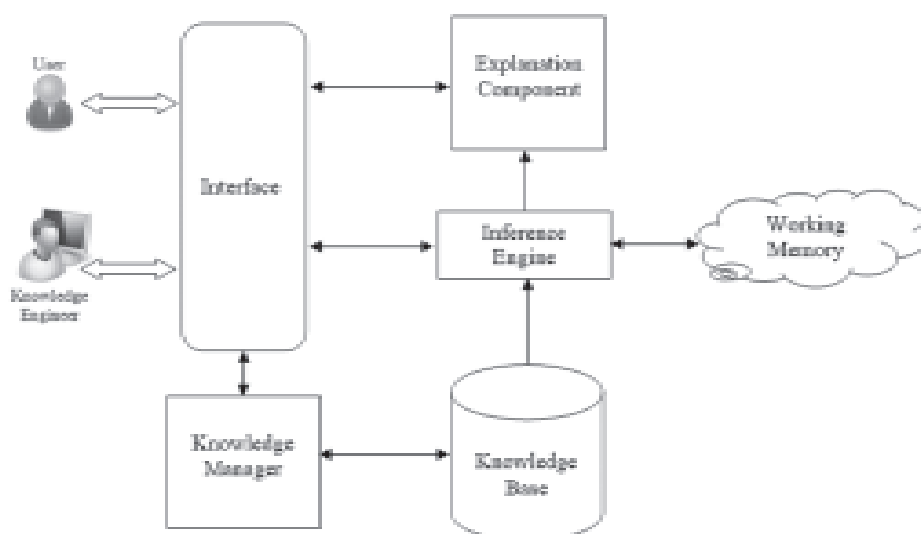
2.1.2. Kiến trúc hệ thống

Kiến trúc một hệ cơ sở tri thức gồm các thành phần như: cơ sở tri thức (knowledge base), bộ suy diễn (inference engine), giao diện (interface), mô-đun giải thích (explanation component), mô-đun quản lý tri thức (knowledge manager), vùng nhớ làm việc (working memory); người sử dụng (user) chức năng của hệ cơ sở tri thức mang lại; kỹ sư tri thức (knowledge engineer) vai trò quản trị tri thức.

Trong các thành phần của hệ thống theo sơ đồ trên, CSTT (Knowledge Base) và Bộ suy diễn (Inference Engine) là hai thành phần trung tâm của hệ thống.

Cơ sở tri thức là kiến thức của miền tri thức ứng dụng, được mô hình hóa và đặc tả trên máy tính. Làm cơ sở cho thiết kế các hệ thống, đòi hỏi xử lý và giải quyết các vấn đề dựa trên tri thức. Việc lựa chọn một mô hình biểu diễn tri thức phù hợp có vai trò rất quan trọng trong việc thiết kế cơ sở tri thức cũng như thiết kế bộ suy diễn.

Bộ suy diễn hay động cơ suy diễn chính là thành phần suy luận giải quyết các vấn đề dựa trên cơ sở tri thức, bộ suy diễn phải mô phỏng lại khả năng cách “tư duy” của con người trong việc giải quyết vấn đề. Hiện nay để thiết kế bộ suy diễn trong các hệ cơ sở tri thức, nhà thiết kế cũng đã có những chiến lược trong thiết kế các kiểu suy luận như: suy diễn tiến, suy diễn lùi, suy diễn dựa tiến lùi kết hợp, suy diễn dựa trên mẫu bài toán (pattern simple), suy diễn dựa trên bài toán mẫu (sample problems), heuristics,...



2.2. Quy trình thiết kế hệ thống

Để thiết kế các hệ cơ sở tri thức, có các chức năng giải quyết các vấn đề dựa trên tri thức ta có thể qua các giai đoạn sau đây:

● **Giai đoạn 1:** Thu thập tri thức và các bài toán cụ thể (vấn đề, yêu cầu, ...). Giai đoạn này có thể bao gồm các việc cụ thể như sau:

+ Xác định miền tri thức cùng với phạm vi tri thức ứng dụng theo nhu cầu thực tế đặt ra.

+ Xác định nguồn thu thập tri thức. Nguồn thu thập thường là tài liệu, sách vở và đặc biệt là từ những nhà chuyên môn hay người chuyên gia. Tri thức được thu thập thường bao gồm các khái niệm, các mối liên hệ, quan hệ giữa các khái niệm, luật, sự kiện, ...vv.

+ Thu thập các bài toán hay vấn đề cụ thể, các yêu cầu, cùng với các bảng hay biểu mẫu thực tế. Từ đó các bài toán sẽ được phân lớp để làm căn cứ cho việc thiết kế bộ suy diễn của hệ thống. Trước hết là các bài toán có khuôn mẫu, rồi đến các dạng bài toán tổng quát.

● **Giai đoạn 2:** Thiết kế cơ sở tri thức. Giai đoạn này lấy kết quả thu thập tri thức và các vấn đề (yêu cầu) của giai đoạn 1 làm căn cứ để xây dựng được cơ sở tri thức cho hệ thống. Để có được cơ sở tri thức ta có thể thực hiện các bước sau đây:

+ Lựa chọn mô hình biểu diễn tri thức phù hợp (đã có) với miền tri thức đã thu thập hoặc xây dựng mô hình riêng cho tri thức được thu thập.

+ Trên cơ sở mô hình biểu diễn tri thức, xác lập tổ chức cơ sở tri thức cụ thể trên máy tính. Nhà thiết kế có thể sử dụng các công cụ phù hợp cho việc tổ chức tri thức trên máy tính.

● **Giai đoạn 3:** Thiết kế bộ suy diễn. Giai đoạn này dựa trên cơ sở tri thức đã được xây dựng trong giai đoạn 2, và các bài toán hay vấn đề đã thu thập được trong giai đoạn 1. Để có được bộ suy diễn ta có thể thực hiện các bước sau đây:

+ Dựa trên các bài toán cụ thể đã thu thập được, ta có thể phân loại các bài toán nhằm xác định các dạng bài toán; từ đó xây dựng mô hình cho các dạng bài toán làm căn cứ cho việc thiết kế các thuật giải suy diễn cho các lớp vấn đề.

+ Ứng với các lớp bài toán đã được mô hình hóa, có thể chọn lựa các chiến lược suy diễn và xây dựng các thuật giải suy diễn nhằm giải quyết các bài toán hay vấn đề. Việc này đòi hỏi nhà thiết kế vận dụng các chiến lược và phương pháp suy diễn, cũng như nghiên cứu phát triển các phương pháp này.

+ Kiểm tra, đánh giá và nâng cao năng lực suy diễn để tìm lời giải cho vấn đề hay kết quả mong muốn.

● **Giai đoạn 4:** Thiết kế giao diện hệ thống, giao diện phải đảm bảo được sự thuận tiện, dễ dàng sử dụng. Trong đó gồm:

+ Nhóm giao diện dành cho đối tượng sử dụng phổ thông (người dùng chính).

+ Nhóm giao diện dành cho người quản trị tri thức.

3. THIẾT KẾ HỆ CƠ SỞ TRI THỨC DỰA TRÊN ONTOLOGY COKB

3.1. Ontology COKB cho biểu diễn tri thức

Mô hình tri thức về các đối tượng tính toán (Computational Objects Knowledge Base - COKB) [7] gồm có 6 thành phần (**C, H, R, Ops, Funcs, Rules**), trong đó **C** là tập các khái niệm, mỗi khái niệm là một đối tượng tính toán, có cấu trúc theo phân cấp:

$$C(0) = \{c \mid \text{dom}(c) \subseteq \square \mid c \text{ là khái niệm tiền đề trong miền tri thức}\}$$

$$C(k) = \{c \mid (\forall x \in \text{attrs}(c), \exists i < k, \exists c' \in C(0), x = c') \wedge (\exists x \in \text{attrs}(c), \exists c' \in C(k-1), x = c')\}, k > 0$$

Ký hiệu: $\text{dom}(c)$ và $\text{attrs}(c)$ là tập các thuộc tính của khái niệm.

Mô hình đối tượng tính toán gồm các thành phần nội tại, được trang bị bên trong nó các lớp bài

toán nội tại cùng các thuật giải cho các lớp bài toán gồm:

- + Bài toán về xác định bao đóng tập sự kiện
- + Bài toán xem xét tính giải được của bài toán $H \rightarrow G$
- + Bài toán tìm lời giải cho bài toán $H \rightarrow G$
- + Bài toán sự xác định đối tượng

H là tập các quan hệ đặc biệt trên các thành phần **C**; **R** là tập các quan hệ khác trên **C**; **Ops** là tập các toán tử trên **C**; **Funcs** là tập các thành phần tri thức về hàm; **Rules** là tập các thành phần tri thức luật. Mỗi thành phần luật, có cấu trúc gồm giả thiết và kết luận. Trong đó, các sự kiện giả thiết, kết luận theo mô hình tri thức **COKB** được phân loại thành 12 nhóm sự kiện và thuật giải hợp nhất sự kiện gồm:

- + Sự kiện loại 1: Sự kiện thông tin về loại của một đối tượng.
- + Sự kiện loại 2: Sự kiện về tính xác định của một đối tượng thay thuộc tính của đối tượng (đã được xác định).
- + Sự kiện loại 3: Sự xác định của một thuộc tính hay một đối tượng thông qua một hằng hay biểu thức hằng.
- + Sự kiện loại 4: Sự kiện về sự bằng nhau giữa một đối tượng hay một thuộc tính với một đối tượng hay một thuộc tính khác. Sự bằng nhau giữa hai đối tượng cùng loại sẽ được hiểu theo nghĩa là các thuộc tính của chúng tương ứng bằng nhau.
- + Sự kiện loại 5: Sự kiện về sự phụ thuộc của một đối tượng hay của một thuộc tính của các đối tượng thông qua một công thức tính toán hay một đẳng thức theo các đối tượng hoặc các thuộc tính
- + Sự kiện loại 6: Sự kiện về một quan hệ trên các đối tượng hay các thuộc tính của đối tượng.
- + Sự kiện loại 7: Sự kiện về sự xác định của một hàm.
- + Sự kiện loại 8: Sự xác định của một hàm thông qua một hằng hay biểu thức hằng.
- + Sự kiện loại 9: Sự bằng nhau giữa một đối tượng | một thuộc tính của đối tượng và một hàm
- + Sự kiện loại 10: Sự bằng nhau giữa một hàm và một hàm khác.
- + Sự kiện loại 11: Sự phụ thuộc của một hàm vào một hàm khác hay một đối tượng hay một thuộc tính của đối tượng khác thông qua một biểu thức.
- + Sự kiện loại 12: Sự kiện về mối quan hệ giữa hàm và một đối tượng khác.

3.2. Quy trình thiết kế cơ sở tri thức

3.2.1. Biểu diễn tri thức

Biểu diễn tri thức là khâu bắt buộc và có vai trò rất quan trọng trong quá trình thiết kế cơ sở tri thức. Quá trình này đòi hỏi nhà thiết kế phải có sự am hiểu về các phương pháp biểu diễn tri thức đã có, từ đó làm cơ sở lựa chọn hay vận dụng một cách linh hoạt các phương pháp biểu diễn tri thức đã có, để thực hiện mô hình hóa tri thức, đặc tả và tổ chức tri thức lên máy tính.

Hiện nay có nhiều phương pháp biểu diễn tri thức khác nhau từ các phương pháp biểu diễn tri thức cơ bản truyền thống đến các phương pháp hiện đại [1]. Việc lựa chọn tri thức phù hợp phụ thuộc chính vào quá trình thu thập tri thức, phân loại tri thức, yêu cầu của hệ cơ sở tri thức được thiết kế. Để có thể xây dựng mô hình cho tri thức, ta thực hiện các bước sau:

- Bước 1: Thu thập tri thức. Tri thức được thu thập cần phải xác định rõ miền tri thức gì? nguồn thu thập từ đâu? Thông thường nguồn tri thức sẽ được chọn lọc từ các tài liệu chính thống như sách, vở đặc biệt là các chuyên gia về miền tri thức cần thu thập. Việc thu thập này sẽ bao gồm cả các lớp vấn đề, các yêu cầu hệ thống.

● Bước 2: Phân loại tri thức. Tri thức sau khi thu thập cần có sự phân loại, như khái niệm, quan hệ, luật, hàm, toán tử, hay các dạng bài tập, các phương pháp giải bài tập, các kỹ thuật, mẹo giải quyết bài tập.

● Bước 3: Chọn lựa phương pháp biểu diễn tri thức phù hợp với miền tri thức, các yêu cầu chức năng được thu thập và phân loại.

● Bước 4: Bắt đầu thiết lập các danh mục tri thức theo mô hình tri thức được chọn từ các tri thức được thu thập, phân loại.

3.2.2. Tổ chức cơ sở tri thức

Việc lựa chọn mô hình biểu diễn tri thức phù hợp và tập tri thức được thu thập, ta cần phải biểu diễn đặc tả chúng trên máy tính, để từ đó có thể thiết kế được các động cơ suy diễn, giải quyết các vấn đề trên tri thức. Việc tổ chức, đặc tả, lưu trữ cơ sở tri thức phải đảm bảo các tiêu chuẩn sau đây:

+ Phải biểu diễn được đầy đủ tri thức.

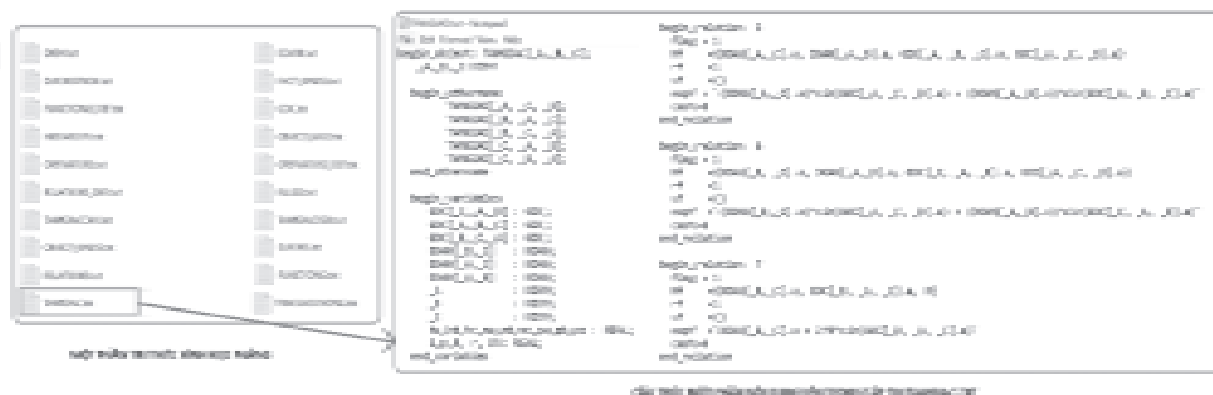
+ Tri thức phải đảm bảo sự nhất quán, không gây mâu thuẫn.

Tổ chức lưu trữ tri thức trên máy tính phụ thuộc vào kỹ năng sử dụng các công nghệ, kỹ thuật của nhà phát triển, ta có thể dụng các công cụ hỗ trợ lưu trữ tri thức sau đây:

+ Sử dụng các tập tin (*.TXT).

+ Sử dụng các công cụ lưu trữ như: MySQL, SQL server, MS Access, MS Excel, ...

Ví dụ: ta có cấu trúc tổ chức cơ sở tri thức về kiến thức hình học phẳng, được biểu diễn theo mô hình các đối tượng tính toán gồm tập các tập tin *.TXT, và một phần cấu trúc bên trong của tập tin TAMGIAC.txt



4. THIẾT KẾ BỘ SUY DIỄN GIẢI QUYẾT VẤN ĐỀ DỰA TRÊN CƠ SỞ TRI THỨC

4.1. Vấn đề tổng quát hóa và mô hình hóa vấn đề trong thiết kế bộ suy diễn

Để thiết kế được động cơ suy diễn giải quyết các lớp vấn đề tự động thông minh trên cơ sở tri thức, đòi hỏi các lớp vấn đề này cần phải được mô hình hóa và đặc tả được trên máy tính. Từ cơ sở đó, nền tảng căn cứ, để giúp nhà thiết kế có thể xác định các chiến lược suy luận giải quyết các vấn đề bên trong thuật giải. Việc tổng quát hóa và mô hình hóa phụ thuộc vào từng phương pháp biểu diễn tri thức, tùy vào mỗi phương pháp biểu diễn tri thức mà ta sẽ có cấu trúc hay mô hình bài toán, đặc tả khác nhau.

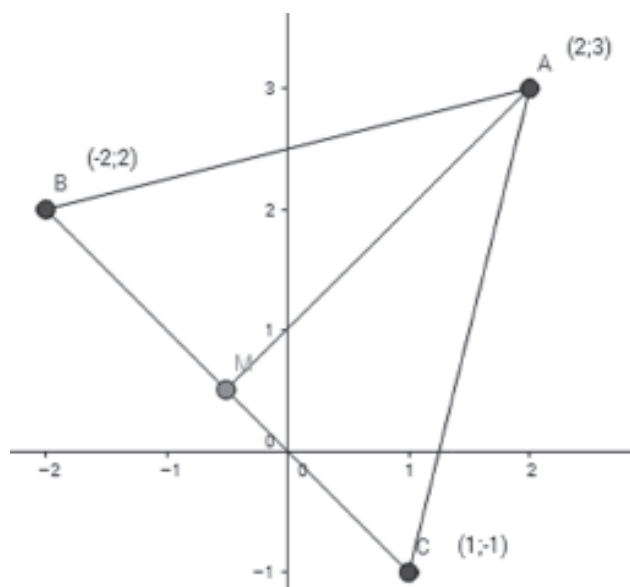
Chẳng hạn như: trên mô hình tri thức COKB, các lớp bài toán sẽ được mô hình hóa dưới dạng mạng các đối tượng tính toán [7], có cấu trúc gồm: $(O, F) \rightarrow G$. Trong đó, (O, F) là giả thiết của bài toán; G là mục tiêu của bài toán.

Ví dụ: Trong miền tri thức hình học giải tích 2 chiều, ta cho tam giác ABC, với các điểm $A(2;3)$,

$B(-2;2)$, $C(1; -1)$, và điểm M là trung điểm của BC . Yêu cầu hãy chứng minh rằng tam giác ABC là tam giác cân, chứng minh rằng $MA \perp BC$, và tính diện tích tam giác ABM .

Giả sử tri thức về hình học giải tích hai chiều, được mô hình hóa theo tri thức COKB, thì ta có thể mô hình hóa lớp bài toán này như sau:

$O = \{[TAMGIAC[A,B,C], "TAMGIAC"], [A, "DIEM"], [B, "DIEM"], [C, "DIEM"], [M, "DIEM"]\};$
 $F = \{A.x = 2, A.y = 3, B.x = -2, B.y = 2, C.x = 1, C.y = -1, M = TRUNGDIEM(B, C)\};$
 $G = \{["CHUNG MINH"], [TAMGIAC[A,B,C], "TAMGIACCAN"]\},$
 $["CHUNG MINH"], ["VUONG", DOAN[M, A], DOAN[B, C]], ["TINH", TAMGIA[A,B,M].S]\};$



4.2. Kỹ thuật suy diễn giải quyết vấn đề một cách thông minh

Để thiết kế bộ suy diễn giải quyết vấn đề thông minh dựa trên tri thức, việc lựa chọn một chiến lược phù hợp cho suy luận là rất quan trọng, phản ảnh khả năng tư duy thông minh của con người. Vì vậy, việc thiết kế kỹ thuật suy diễn để đảm bảo được sự thông minh của con người, đòi hỏi phải đưa vào đó là các kinh nghiệm, kỹ thuật suy luận thông minh của con người.

Trong khoa học máy tính các nhà thiết kế về biểu diễn tri thức và suy luận trên máy tính cũng đã có một số phương pháp suy luận cụ thể như [2 - 3]: suy diễn tiến, suy diễn lùi, suy diễn tiến-lùi kết hợp, suy diễn theo trường hợp (case based reasoning) [8 - 9], suy diễn theo mẫu bài toán (sample problems) [10], suy diễn theo bài toán mẫu (pattern problems), suy diễn theo các quy tắc heuristics.

Kỹ thuật suy diễn giải quyết vấn đề thông minh trên mô hình tri thức các đối tượng tính toán chủ yếu dưới dạng suy diễn tiến. Trong đó, mỗi bước suy diễn sinh ra các sự kiện mới từ các sự kiện đã có, theo dạng sau đây:

+ Suy diễn dựa trên các “*quy tắc suy luận*”, ta có cá quy tắc suy luận sau đây:

- Tự động sinh ra các sự kiện mặc nhiên từ tập sự kiện đã biết trước.
- Sinh ra sự kiện mới dựa trên việc áp dụng một luật dẫn.
- Sinh ra sự kiện mới dựa trên hành vi của một đối tượng.
- Sinh ra sự kiện mới dựa trên việc sử dụng các tính chất, hay thực hiện các thủ tục của một hàm.
- Sinh ra sự kiện mới dựa trên việc giải một hệ phương trình được thành lập bằng cách kết hợp

nhiều dạng quan hệ tính toán.

- Sinh ra đối tượng mới dựa trên việc áp dụng một luật dẫn có phát sinh đối tượng.

+ Suy diễn dựa trên mẫu bài toán (sample problems)

+ Suy diễn dựa trên bài toán mẫu (pattern problems)

Dựa trên các suy luận sinh ra sự kiện mới từ các sự kiện đã có, việc lựa chọn một quy tắc suy luận hay một bước suy luận để sinh ra sự kiện mới cũng là một vấn đề rất quan trọng, ảnh hưởng rất lớn đến tốc độ giải quyết bài toán, cũng như phản ánh sự tư duy của con người. Dựa trên đó các heuristics được sử dụng trong thiết kế bộ suy diễn gồm:

+ Ưu tiên sử dụng các luật có mục tiêu luật cùng tên đối tượng/hàm với mục tiêu của bài toán.

+ Ưu tiên sử dụng các quan hệ tính toán có liên quan đến sự kiện mục tiêu của bài toán.

+ Ưu tiên sử dụng các hành vi bên trong của một đối tượng tính toán có chứa đối tượng cùng đối tượng với đối tượng trong mục tiêu của bài toán.

+ Ưu tiên sử dụng các sự kiện hàm.

+ Ưu tiên sử dụng các bài toán mẫu (pattern problem) trong trường hợp có thể sinh ra được các sự kiện cùng loại, hoặc cùng tên với đối tượng trong mục tiêu bài toán.

+ Ưu tiên sử dụng mẫu bài toán trong trường hợp có thể sinh ra được các sự kiện cùng loại, hoặc cùng tên với đối tượng trong mục tiêu bài toán.

+ Với các trường hợp khác, ưu tiên sử dụng các quy tắc suy luận, có khả năng sinh ra được các sự kiện cùng loại, hoặc cùng tên với đối tượng trong mục tiêu bài toán.

5. ỨNG DỤNG

Ontology là một giải pháp tiếp cận hiệu quả, để thiết kế các hệ cơ sở tri thức hỗ trợ giải quyết các vấn đề dựa trên tri thức. Mô hình về các đối tượng tính toán là một trong những giải pháp cụ thể theo hướng tiếp cận ontology, mô hình đã tỏ ra mạnh mẽ và hiệu quả trong việc thiết kế các cơ sở tri thức có cấu trúc phức tạp, nhiều thành phần. Với các kiến trúc, quy trình thiết kế được trình bày và ontology về các đối tượng tính toán (COKB) cho biểu diễn tri thức, đã giúp các nhà phát triển (knowledge engine) xây dựng được một số hệ cơ sở tri thức với các vấn đề như suy diễn giải quyết các vấn đề tự động [12-14], hệ chuyên gia chẩn đoán bệnh trong y khoa [15], tư vấn kiến trúc đã được xây dựng, [15]. Bên dưới đây là một số minh họa thực tế của một số ứng dụng:

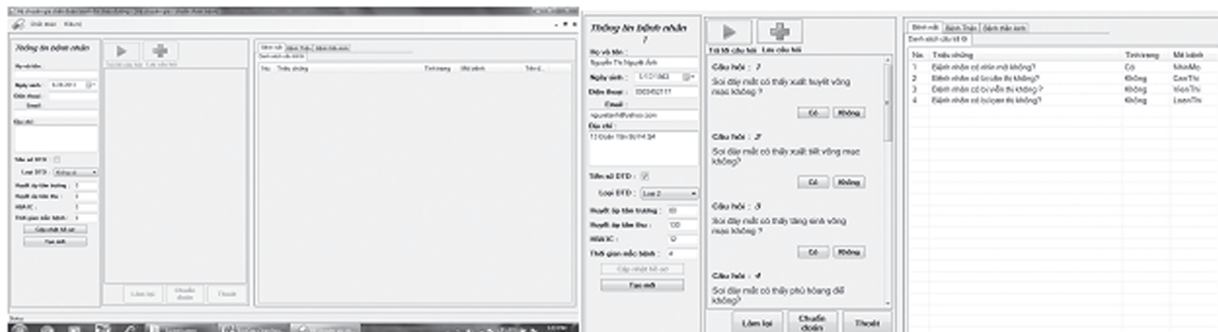
5.1. Ứng dụng giải quyết vấn đề tự động trên miền tri thức hình học không gian

The screenshot displays a software interface for solving a geometry problem. It is divided into three main sections:

- Solution - Lời giải:** This panel shows five steps of the solution process:
 - Bước (1):** M nằm trong tam giác ABC. H_0 là điểm $H_0 = AM \cap CB$, H_0 nằm giữa CB, M nằm giữa AH_0 .
 - Bước (2):** $H_0 = AM \cap CB$ (1). $H_0 \in AM$, $H_0 \in CB$.
 - Bước (3):** N nằm trong tam giác DAB. H_3 là điểm $H_3 = AN \cap DB$, H_3 nằm giữa DB, N nằm giữa AH_3 .
 - Bước (4):** $H_3 = AN \cap DB$ (3). $H_3 \in AN$, $H_3 \in DB$.
 - Bước (5):** $H_0 \in AM$ (2), $MA \subset mp(MNA)$. $H_0 \in mp(MNA)$.
- Graph:** A diagram showing a triangle with medians intersecting at a point. Labels include M , N , H_0 , H_3 , and various planes like $mp(MNA)$.
- List of Facts:** A list of 20 facts:
 - $H_0 \in mp(H_1 H_0 H_3)$
 - $M \notin CH_3$
 - D phân biệt với H_0
 - C phân biệt với H_3
 - N nằm trong tam giác DAB
 - M nằm trong tam giác ABC
 - $ND \subset mp(CDN)$
 - $MN \subset mp(MNA)$
 - $MA \subset mp(MNA)$
 - $DB \subset mp(DBC)$
 - $CB \subset mp(DBC)$
 - $BM \subset mp(BDM)$
 - $AN \subset mp(MNA)$
 - $H_3 \in mp(H_1 H_0 H_3)$
 - $MN \notin mp(BCD)$
 - $D \in mp(CDN)$
 - $D \in mp(BDM)$
 - $C \in mp(CDN)$
 - MN cắt $mp(DBC)$
 - H_3 phân biệt với H_0

Hệ giải toán Hình học không gian chương trình lớp 11 đã được thử nghiệm trên các dạng toán của các miền kiến thức được phân loại ở trên. Chương trình đưa ra lời giải thực hiện theo từng bước, diễn giải một cách tường minh, tự nhiên và phù hợp với lối tư duy của con người.

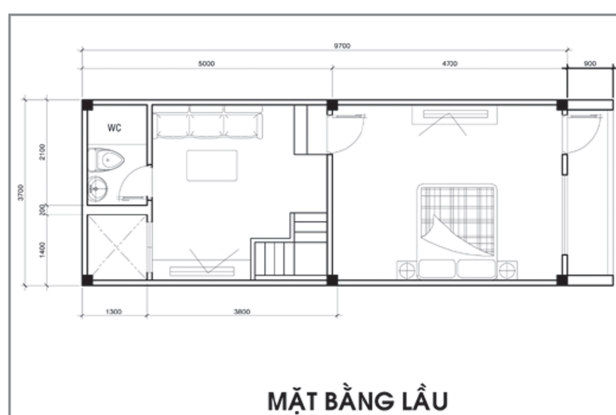
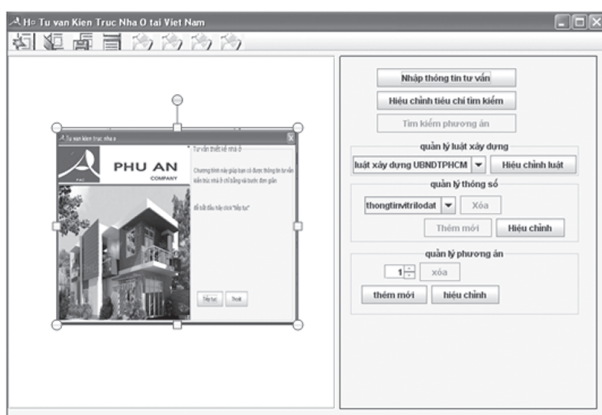
5.2. Hệ thống chẩn đoán các biến chứng trên mạch máu nhỏ của bệnh đái tháo đường



Kết quả thực nghiệm của hệ chuyên gia ESDMCD được đánh giá bằng cách so sánh với kết quả chẩn đoán của chương trình với kết quả chẩn đoán của bác sĩ chuyên gia trên 106 bệnh nhân. Dữ liệu được thu thập từ Bệnh viện Đa khoa quận 4 từ ngày 16/7/2012 - 12/11/2012. Trong tổng số 106 trường hợp có 4 trường hợp kết quả chẩn đoán khác biệt, 102 trường hợp cho kết quả tương đương chiếm tỷ lệ 96.2%.

5.3 Hệ tư vấn kiến trúc nhà ở tại Việt Nam

Hệ thống cung cấp các thông tin tư vấn liên quan đến quy định xây dựng nhà ở của địa phương, có đủ thông tin tư vấn liên quan đến từng không gian các phòng, thông tin tư vấn của các khu vực phụ như ban công, sân vườn, nhà xe, cầu thang,... thông tin về hướng bố trí các phòng trong nhà thế nào cho phù hợp với hướng chính của căn nhà. Ngoài ra, hệ thống còn kiểm nhanh các phương án đã được thiết kế trước đây.



6. KẾT LUẬN

Trong bài báo đã trình kiến trúc và phương pháp, quy trình thiết kế hệ cơ sở tri thức trong thực tế. Đặc biệt là quy trình thiết kế thành phần KB, IE trong một hệ cơ sở tri thức. Bên cạnh đó bài báo đã trình bày phương pháp thiết kế cơ sở tri thức và bộ suy diễn của hệ cơ sở tri thức dựa trên ontology về cơ sở tri thức các đối tượng tính toán và các ứng dụng của nó trong thực tế như hệ giải toán trong miền tri thức hình học, hệ chẩn đoán bệnh. Các ứng dụng đã được triển khai thử nghiệm và đã được đưa vào sử dụng trong thực tế. Hướng nghiên cứu phát triển của các hệ cơ sở tri thức nói

chung, biểu diễn tri thức và suy luận nói riêng cần phải nghiên cứu và đưa ra được các hướng xây dựng các hệ cơ sở tri thức dựa trên sự tích hợp.

TÀI LIỆU THAM KHẢO

- [1] Frank van Harmelen, Vladimir, Bruce, “Handbook of Knowledge Representation”, Elsevier, 2008.
- [2] Ronald J. Brachman, Hector J. Levesque, “Knowledge Representation and Reasoning”, Elsevier Inc, 2004.
- [3] Stuart Russell, Peter Norvig, “Artificial Intelligent – A modern approach (third edition)”, Prentice Hall, 2020.
- [4] Gruber, T. R., “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, International Journal Human-Computer Studies, 43, 907-928, 1995.
- [5] Rajendra Arvind Akerkar and Priti Srinivas Sajja, “Knowledge-Based Systems”, Jones and Bartlett Publisher, LLC, 2010.
- [6] Đỗ Văn Nhơn, Nguyễn Đình Hiến, “Các hệ cơ sở tri thức”, NXB ĐHQG-HCM, 2017.
- [7] Nhơn Do, Thanh T. Mai, “Perfect COKB Model and Reasoning Methods for the design of Intelligent Problem Solvers”, Proceedings of 17th International Conference on Intelligent Software Methodologies, Tools, and Techniques, Japan, 2017.
- [8] Janet L. Kolodner, “An Introduction to Case-based Reasoning”, Artificial Intelligent Review 6, pp. 3-34, 1992.
- [9] Michael M. Richter, Rosina O. Weber (2013), “Case-based Reasoning – A Text book”, Springer.
- [10] Nhơn V. Do, Hien D. Nguyen, Thanh T. Mai, “Designing an Intelligent Problems Solving System based on Knowledge about Sample Problems”, Proceeding of 5th Asian conference on Intelligent Information and Database Systems (ACIIDS 2013), Kuala Lumpur, Malaysia, March 2013, LNAI 7802, pp. 465-475, Springer, 2013.
- [11] Nhơn V. Do, Hien D. Nguyen, Vương T. Phạm, “A methodology for designing knowledge-based systems and applications”, 2022.
- [12] Đỗ Văn Nhơn, Nguyễn Đình Hiến, Phạm Thị Vương, “Xây dựng hệ hỗ trợ giải toán đại số tuyến tính trên cơ sở tri thức gồm các miền tri thức phối hợp”, *Tạp chí Khoa học Trường Đại học Cần Thơ*, 2017.
- [13] Hien D. Nguyen, Nhơn V. Do, “Intelligent Problem Solver In Educations For Discrete Mathematics”, New Trends in Intelligent Software Methodologies, Tools and Techniques, Volume 297, 2017.
- [14] Hien D. Nguyen, Nhơn V. Do, Thanh T. Mai, Vương T. Phạm, “A method for designing the Intelligent system in learning of Algorithms”, Proceedings of 18th International Conference on Intelligent Software Methodologies, Tools, and Techniques (SOMET 2019), Kuching, Malaysia, Frontiers in Artificial Intelligence and Applications, vol. 318, pp. 658 - 671, Sept., 2019.
- [15] Nhơn Do, Diem Nguyen, “Intelligent Problem Solving about Functional Component of COKB Model and Application”, The 6th International Conference on Computational Collective Intelligence Technologies and Applications, 2014.