

Các hệ thống dựa trên tri thức

TS. NGUYỄN ĐÌNH HÓA

KHOA CNTT1 – PTIT

DINHHOA@GMAIL.COM ĐT: 0942807711

Giới thiệu chung

- Các vấn đề cơ bản về các hệ thống dựa trên tri thức: khái niệm, ứng dụng, sự phát triển của các hệ thống dựa trên tri thức, các ưu điểm của chúng.
- Một số hệ thống dựa trên tri thức cơ bản
- Một số công cụ để xây dựng các hệ thống dựa trên tri thức.

Kế hoạch giảng dạy

- + Nghe giảng lý thuyết: 36 tiết (18 kíp)
- + Thảo luận và Hoạt động nhóm: 08 tiết (4 kíp)
- + Tự học: 01 tiết

Đánh giá môn học

Điểm chuyên cần: 10%

Điểm bài tập lớn: 90%

Tài liệu môn học

- [1] Nguyễn Quang Hoan, “Các Hệ thống dựa trên tri thức”, giáo trình. Học viện Công nghệ Bưu chính Viễn thông.
- [2] Joseph Giarratano, Gary Riley. “Expert systems: Principles and Programming”, third edition, PWS Publishing Company, 2002.
- [3] Rajendra Arvind Akerkar, Priti Srinivas Sajja. “Knowledge-Based Systems”. Jones and Bartlett Publishers, 2010.

Các yêu cầu cơ bản

Tham dự bài trên lớp đầy đủ

Không mất trật tự trong giờ học lý thuyết

Làm đủ bài tập lớn và bài tập về nhà (nếu có)

Hăng hái tham gia xây dựng bài học trên lớp

Giới thiệu chung

Một **hệ thống dựa trên tri thức (KBS)**, hay còn gọi là **hệ chuyên gia**, “là một chương trình máy tính thông minh sử dụng tri thức và các thủ tục suy luận để giải những bài toán tương đối khó khăn trong một lĩnh vực hẹp của đời sống, đòi hỏi những chuyên gia mới giải được”

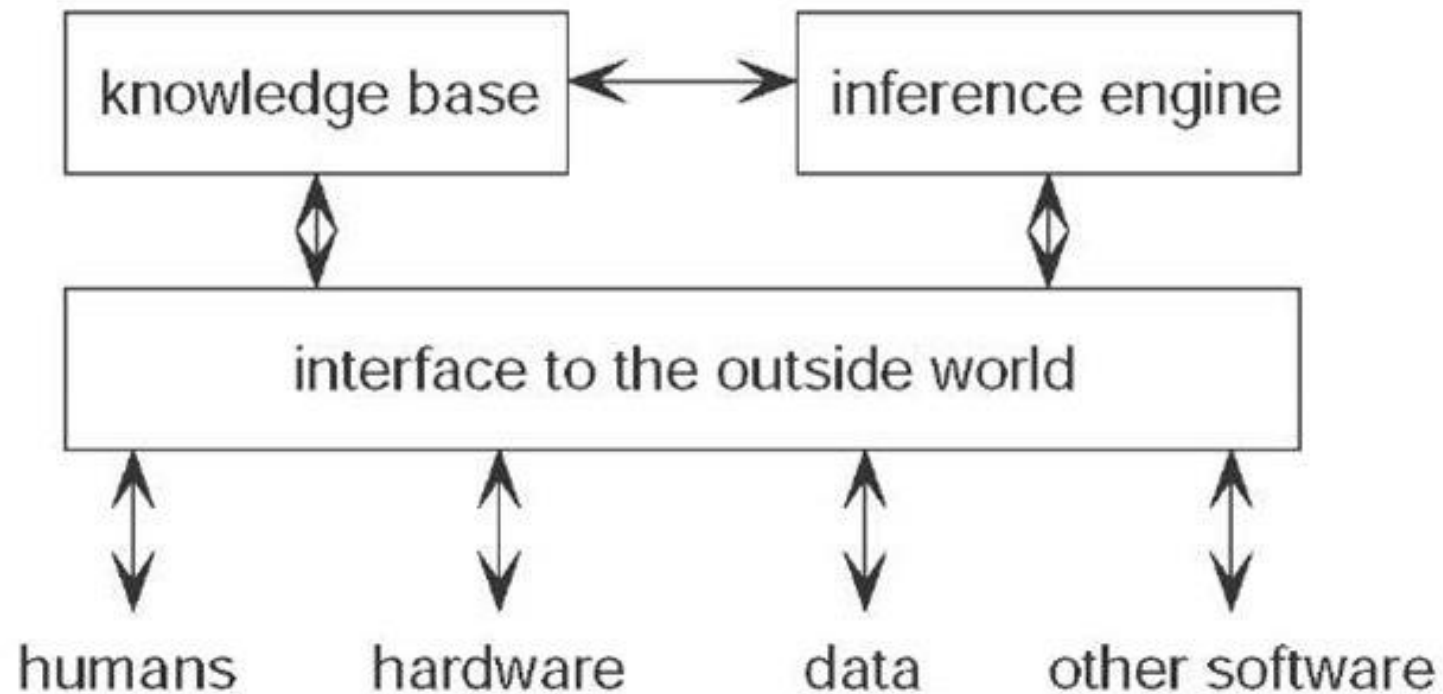
- Là một trong những lĩnh vực ứng dụng của trí tuệ nhân tạo.
- Là một hệ thống tin tin học có thể mô phỏng năng lực quyết định và hành động của một chuyên gia.
- Là hệ thống sử dụng tri thức trong một lĩnh vực cụ thể để đưa ra giải pháp cho một vấn đề với “chất lượng chuyên gia”.

Các thành phần của một KBS

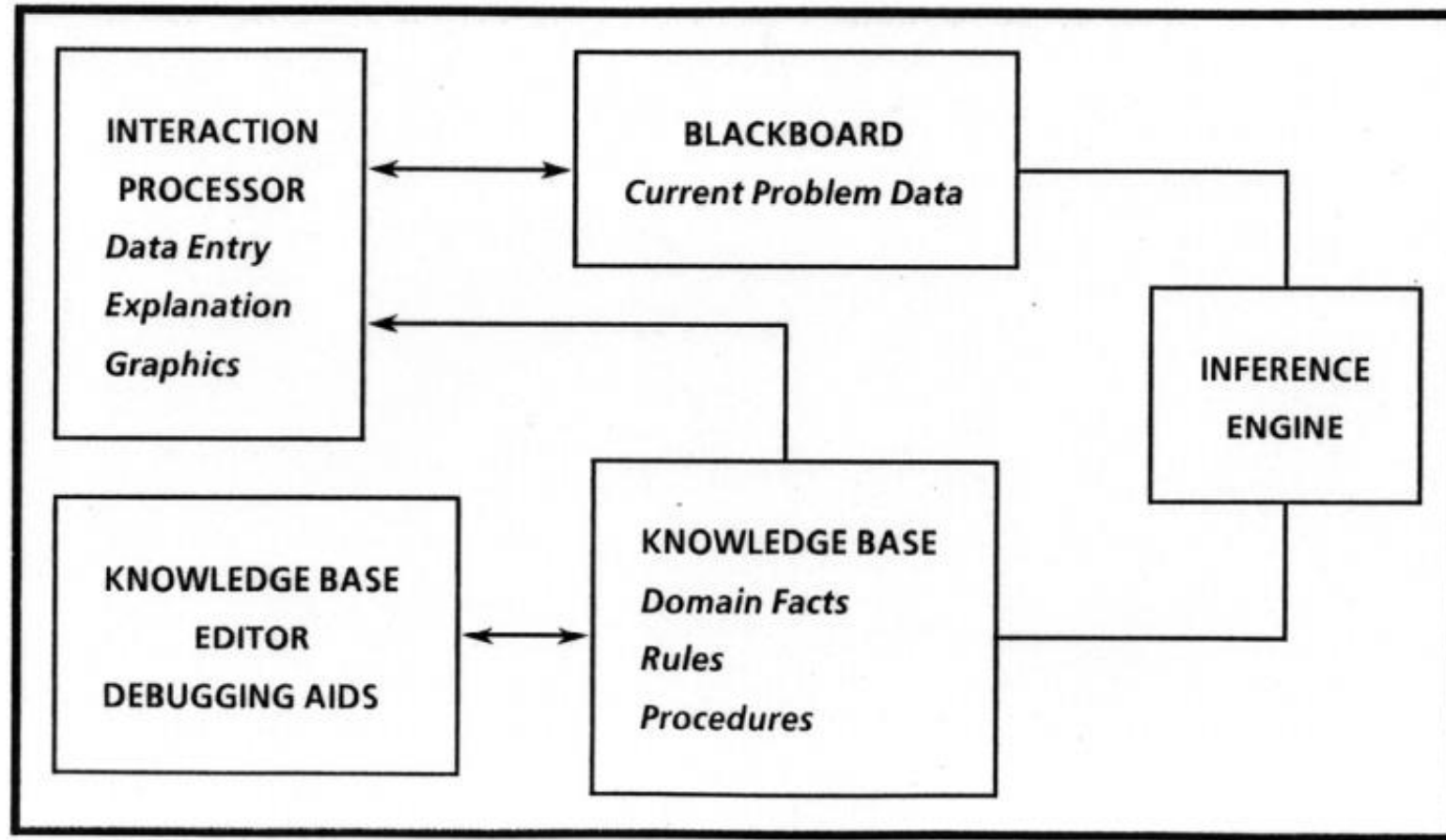
Một hệ thống dựa trên tri thức gồm ba thành phần chính:

- Cơ sở tri thức (knowledge base)
- Máy suy diễn hay cơ chế suy diễn (inference engine)
- Hệ thống giao tiếp với người sử dụng (user interface)

Các thành phần của một KBS



Kiến trúc của một KBS



Ưu điểm của KBS

Giảm chi phí, tăng chất lượng sản phẩm và dịch vụ

- Tăng khả năng phục vụ của chuyên gia
 - Cung cấp kiến thức chuyên gia cho người bình thường
 - Giảm thời gian chờ đợi dịch vụ
 - Tăng phạm vi phục vụ
- Tổng hợp kiến thức chuyên gia từ nhiều nguồn
- Tạo ra các dịch vụ với chất lượng ổn định
- Tăng khả năng tự động hóa các dịch vụ
- Xây dựng bộ dữ liệu về quá trình cung cấp dịch vụ

Khi nào cần xây dựng KBS?

Phụ thuộc vào chi phí xây dựng và triển khai thuật toán

Phụ thuộc vào lĩnh vực áp dụng: đơn giản hay cao siêu, lĩnh vực rộng hay hẹp,...

Đã có sẵn các hệ thống được triển khai thực tế trong cùng lĩnh vực chưa?

Có sẵn kiến thức chuyên gia, hoặc sự giúp sức của chuyên gia trong lĩnh vực này chưa?

Khả năng đóng góp cho xã hội của hệ thống đến mức nào?

Hệ thống có khả năng mở rộng kiến thức không?

Có sẵn các ví dụ điển hình trong thực tế thuộc lĩnh vực cần quan tâm chưa?

Tính hiệu quả của hệ thống ra sao khi triển khai thực tế?

Tốc độ xử lý và tương tác thông tin của hệ thống?

Tri thức

Tri thức là **sự hiểu biết** về một đối tượng, sự việc, hoàn cảnh, sự kiện trong một lĩnh vực nhất định.

Tri thức là một khái niệm trừu tượng, có thể được cụ thể hóa thành ba thành phần:

- Các sự kiện (events, facts)
- Các mối quan hệ, quy luật liên quan (Rules) giữa các sự kiện
- Tri thức có tính heuristic (rút ra từ kinh nghiệm, từ suy diễn mang tính may rủi; tìm ra, phát hiện ra)

Tri thức do con người (chuyên gia) diễn giải và chuyển giao trong một lĩnh vực cụ thể nào đó.

Ví dụ:

- Tri thức về bệnh Covid-19, do các bác sĩ chuyên khoa hoặc nhà nghiên cứu y học mô tả và chuyển giao
- Tri thức về chiến lược đầu tư do các nhà cố vấn đầu tư diễn giải và chuyển giao

Ví dụ về tri thức

Sự kiện: “trời mưa”, “đường ướt” là các hiện tượng/sự việc có thể được nhận thức bởi con người.

Mối quan hệ giữa các sự kiện: “trời mưa” -> “đường ướt”, gọi là logic mệnh đề

Một trong số các cách để biểu diễn tri thức là sử dụng các luật, ví dụ:

Luật 1: IF ... THEN...

IF “trời mưa” THEN “đường ướt”

IF “xe không khởi động được” THEN “lỗi trong hệ thống điện”

Luật 2: IF... AND... THEN ...

IF “lỗi trong hệ thống điện” AND “Điện thế ắc quy nhỏ hơn 10V” THEN “lỗi tại bộ ắc quy”

Phân loại tri thức

Tri thức thực tế: là các thuật ngữ, chi tiết cụ thể và các yếu tố cơ bản trong bất kỳ lĩnh vực nào. Đây là các thông tin cần được học thông qua tiếp xúc, lặp đi lặp lại, và đọng trong trí nhớ.

Tri thức về khái niệm: thông tin được nắm bắt về các mối quan hệ và chức năng giữa các thành phần/chi tiết để tạo nên khối kiến thức lớn, có cấu trúc. Loại tri thức này liên quan đến 3 yếu tố:

- Nắm được về loại thông tin
- Nắm được nguyên lý và tổng quát hóa
- Nắm được lý thuyết, mô hình, cấu trúc

Phân loại tri thức

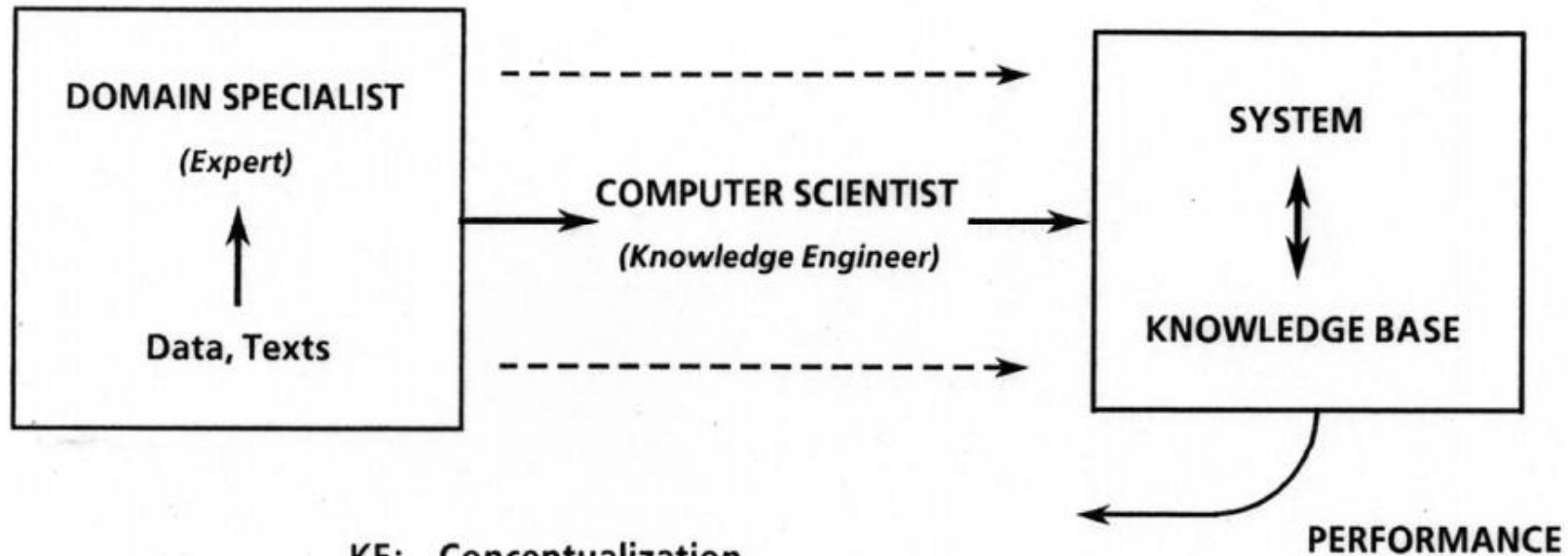
Tri thức được mô tả theo thủ tục: là dạng tri thức liên quan đến kỹ năng thực tế trong một công việc cụ thể nào đó (trả lời câu hỏi “làm thế nào..”), bao gồm:

- Các kỹ năng, thuật toán cho từng đối tượng cụ thể
- Các kỹ thuật, phương pháp cho từng đối tượng cụ thể
- Các tiêu chí lựa chọn thủ tục nào phù hợp trong từng trường hợp cụ thể

Tri thức siêu nhận thức: bao gồm

- Tri thức mang tính định hướng
- Tri thức về các nhiệm vụ tư duy (theo ngữ cảnh, theo điều kiện)
- Tự nhận thức

Thu thập tri thức



KE: Conceptualization
Construction
Refinement

KE: Using the framework
vs
Designing the framework

Hệ cơ sở tri thức

Hệ CSTT là hệ thống dựa trên tri thức, cho phép mô hình hóa các tri thức của chuyên gia, dùng tri thức này để giải quyết vấn đề phức tạp cùng lĩnh vực.

Hệ CSTT gồm: các khái niệm, các sự kiện, các luật và quan hệ giữa chúng.

Sự kiện	Lập luận (suy diễn)
Sự kiện 1	Lập luận 1
Sự kiện 2	Lập luận 2
.....
Sự kiện n	Lập luận m

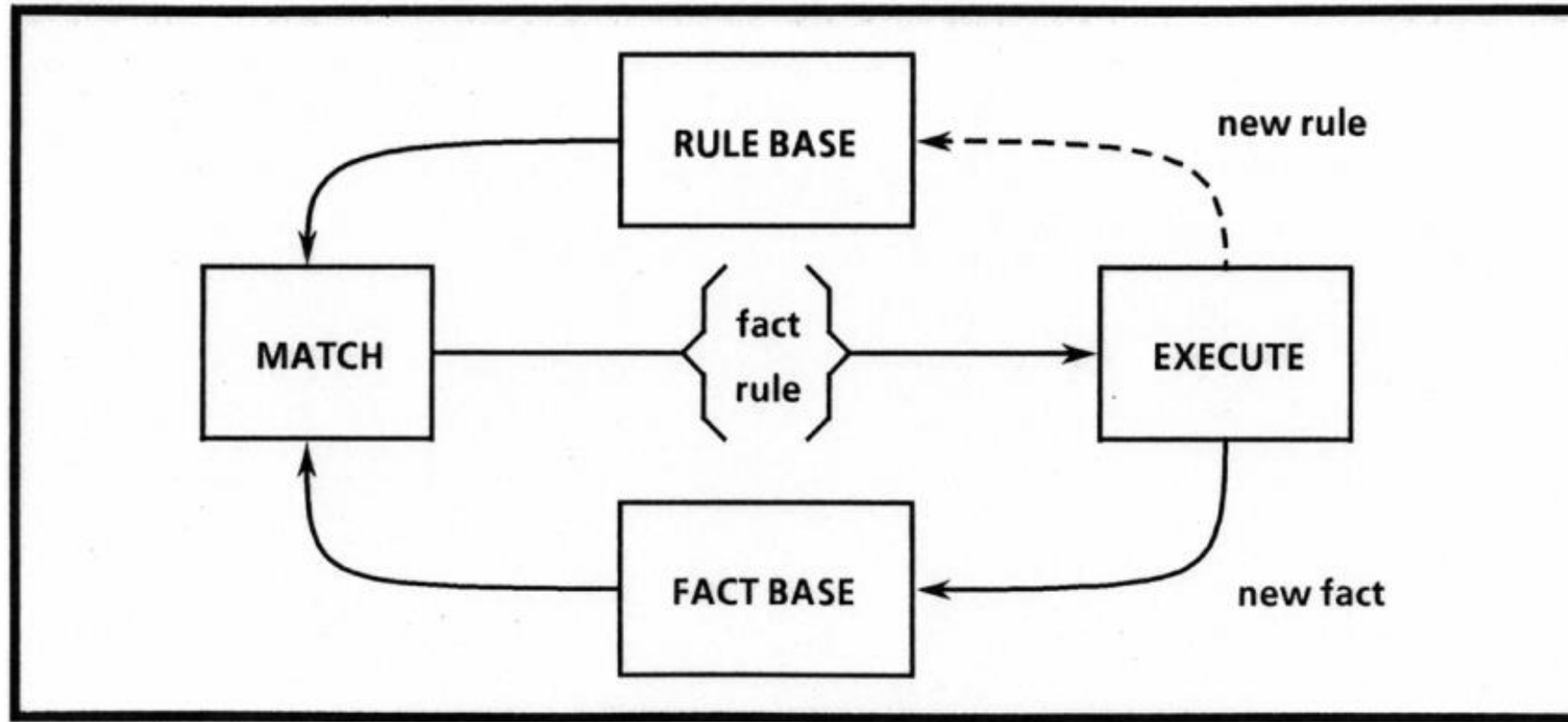
Máy suy diễn

Là bộ xử lý tạo ra sự suy luận bằng cách quyết định xem những luật nào sẽ làm thỏa mãn các sự kiện, các đối tượng, chọn ưu tiên các luật thỏa mãn, thực hiện các luật có tính ưu tiên cao nhất.

Có nhiệm vụ so trùng các sự kiện được chứa trong bộ nhớ làm việc với tri thức được chứa trong CSTT nhằm dẫn ra kết luận cho vấn đề.

Nếu CSTT có chứa luật, KBS sẽ tìm ra luật mà các tiên đề của luật so trùng với các sự kiện được chứa trong bộ nhớ làm việc, lúc đó KBS sẽ thêm các kết luận của luật đó vào bộ nhớ làm việc, rồi tiếp tục tìm ra sự so trùng khác.

Quy tắc hoạt động suy diễn



Ví dụ về máy suy diễn

Giả sử hệ CSTT chỉ có 2 luật:

Luật 1: **IF** “xe không khởi động được” **THEN** “lỗi trong hệ thống điện”

Luật 2: **IF** “lỗi trong hệ thống điện” **AND** “Điện thế ắc quy nhỏ hơn 10V” **THEN** “lỗi tại bộ ắc quy”

Bước 1:

- KBS: “có phải xe không khởi động được?”
- Người: “Đúng”

Chú thích: người dùng trả lời “đúng” nên KBS thêm vào một sự kiện “Xe không khởi động được”

Máy suy diễn thực hiện so trùng, và thấy Luật 1 có thể sử dụng được.

Hệ thống bổ sung thêm kết luận của Luật 1, đó là:

“lỗi trong hệ thống điện”

Ví dụ về máy suy diễn

Bước 2:

- KBS: “có phải điện thế của ắc quy dưới 10V?”
- Người: “Đúng”

Chú thích: vì người dùng trả lời “Đúng” nên KBS thêm vào một sự kiện “Điện thế ắc quy dưới 10V”

Động cơ suy diễn của KBS thực hiện so trùng, thấy Luật 2 có thể sử dụng, nên nó bổ sung thêm phần kết luận của Luật 2, đó là

“Lỗi tại bộ Ắc quy”

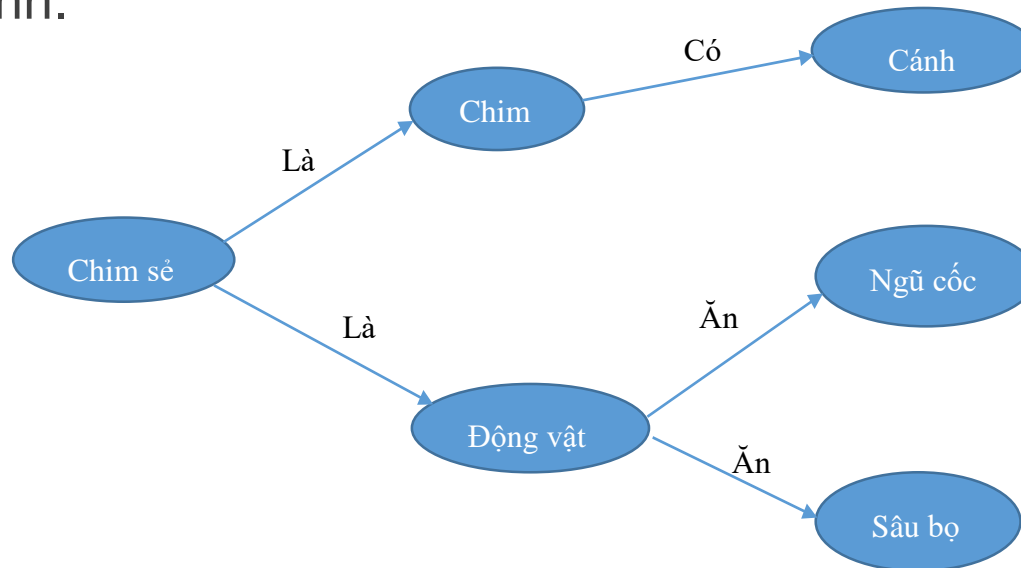
Phiên làm việc kết thúc vì hệ thống chỉ có 2 luật nói trên.

Biểu diễn tri thức

Mô tả tri thức bằng mạng ngữ nghĩa (semantic network)

Định nghĩa 1: Mạng ngữ nghĩa là sự mở rộng và phát triển từ mô tả bộ ba Object – Attribute – Value (OAV).

Mạng ngữ nghĩa là mạng (gồm các nút và cung $G=\{V, U\}$, trong đó nút được gán một ngữ nghĩa nhất định.



Biểu diễn tri thức

Mạng tính toán (Computational network):

Định nghĩa 1: mạng tính toán là một trường hợp riêng của mạng ngữ nghĩa, gồm các nút được mô tả bởi:

- Hàm: Ký hiệu nút dạng bằng một dạng (ví dụ dạng hình chữ nhật);
- Biến: ký hiệu nút dạng khác (ví dụ dạng hình tròn);

Và các cung mô tả các mối liên hệ giữa các nút hàm và các nút biến

Định nghĩa 2: mạng tính toán là một cấu trúc (M, F) với $M = \{x_1, x_2, \dots, x_m\}$ là tập các biến đơn trong các miền xác định tương ứng D_1, D_2, \dots, D_m , và $F = \{f_1, f_2, \dots, f_m\}$ tập các quan hệ giữa các tập biến số trong M .

$f_i: u_i \rightarrow v_i$ trong đó u_i, v_i là các tập con khác rỗng của M thỏa mãn: $u_i \cap v_i = \emptyset$

Ví dụ về mạng tính toán

Cho tam giác ABC với tập các biến $M =$

$\{a, b, c\}$: các cạnh của tam giác

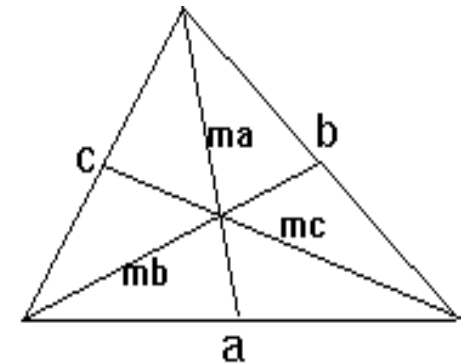
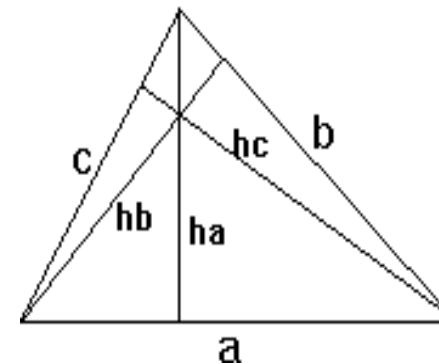
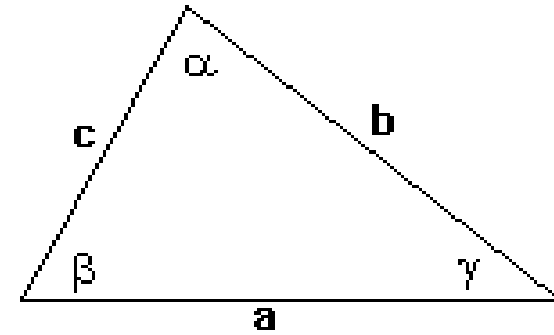
α, β, γ : các góc trong tam giác

h_a, h_b, h_c : ba đường cao của tam giác

S : diện tích tam giác

p : nửa chu vi tam giác, ...}

Tập các hàm $F = \{f_1, f_2, \dots, f_m\}$ được xác định như sau



Ví dụ về mạng tính toán

Quan hệ giữa ba góc:

- $f_1: \alpha + \beta + \gamma = \pi \text{ (rad)}$

Định lý Cosine:

- $f_2: a^2 = b^2 + c^2 - 2bc \cdot \cos(\alpha)$

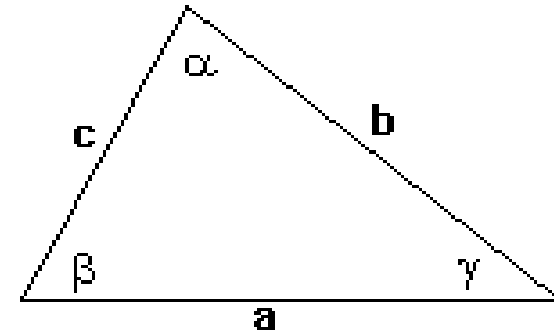
- $f_3: b^2 = a^2 + c^2 - 2ac \cdot \cos(\beta)$

- $f_4: c^2 = a^2 + b^2 - 2ab \cdot \cos(\gamma)$

Định lý sine:

- $f_5: \frac{a}{\sin \alpha} = \frac{b}{\sin \beta}$

- $f_6: \frac{c}{\sin \gamma} = \frac{b}{\sin \beta}$



Ví dụ về mạng tính toán

Các công thức tính diện tích:

- $f_7: S = \frac{1}{2}(ab \cdot \sin \gamma)$
- $f_8: S = \frac{1}{2}(bc \cdot \sin \alpha)$
- $f_9: S = \frac{1}{2}(ac \cdot \sin \beta)$
- $f_{10}: S = \frac{1}{2}a \cdot h_a$
- $f_{11}: S = \sqrt{p(p-a)(p-b)(p-c)}$

Các công thức tính chu vi:

- $f_{12}: p = (a + b + c)/2$

Các vấn đề trong mạng tính toán

Ứng dụng của mạng tính toán:

Với tam giác ABC, giả sử đã biết các cạnh cùng các góc, tính các thông số còn lại

- Giả thiết: đã biết $A = \{a, \beta, \gamma\}$
- Tính các biến: $B = \{b, c\}$

Vấn đề 1: $A \rightarrow B$ có giải được không?

Định nghĩa 3: Bài toán $A \rightarrow B$ được gọi là giải được khi có thể tính được giá trị các biến thuộc B xuất phát từ giả thiết về giá trị của các biến có trong A . Ta nói rằng một dãy quan hệ $\{f_1, f_2, \dots, f_m\} \subseteq F$ là một lời giải của bài toán $A \rightarrow B$.

$\{f_1, f_2, \dots, f_m\}$ được gọi là lời giải tốt nếu không thể bỏ bất kỳ thành phần nào trong đó

Lời giải được gọi là tối ưu nếu nó có số bước giải ít nhất trong tất cả các lời giải tốt.

Các vấn đề trong mạng tính toán

Lời giải có thể như sau:

$$A = \{a, \beta, \gamma\} \xrightarrow{f_1} A_1 = \{a, \beta, \gamma, \alpha\} \xrightarrow{f_5} A_2 = \{a, b, \beta, \gamma, \alpha\} \xrightarrow{f_6} A_3 = \{a, b, c, \beta, \gamma, \alpha\}$$

Lời giải này sẽ là $\{f_1, f_5, f_6\}$

Một bài toán có thể có nhiều lời giải khác nhau.

Vấn đề 2: nếu $A \rightarrow B$ không giải được, cần thêm các yếu tố nào?

Một số vấn đề trong mạng tính toán

Trong một bài toán có thể xuất hiện nhiều đối tượng khác nhau. Bài toán chỉ tập trung vào một đối tượng, nhưng tri thức trên đối tượng này không đủ để giải quyết bài toán

- Ví dụ: cho tứ giác với thông tin về 4 cạnh và 1 góc, cần phải tính diện tích tứ giác đó.

Tri thức về các thành phần cơ bản của một mạng tính toán (các biến đơn) có thể thay đổi theo thời gian: cần chỉnh sửa cấu trúc tập biến (ví dụ thêm khái niệm biến hàm)

- Ví dụ: cường độ dòng điện xoay chiều thay đổi theo thời gian: $i(t) = I \cdot \cos(\omega t + \varphi)$

Tập quan hệ tính toán và cơ chế suy luận còn đơn giản: cần phải mở rộng

Vai trò của KBS

Cung cấp các hệ thống với mức thông minh cao

Hỗ trợ con người trong khám phá và phát triển các lĩnh vực chưa được biết tới

Cung cấp lượng lớn tri thức trong các lĩnh vực khác nhau

Hỗ trợ quản lý tri thức trong các cơ sở tri thức

Giải quyết các vấn đề một cách tốt hơn so với các hệ thống thông tin truyền thống

Thu thập các nhận thức mới bằng mô phỏng các tình huống chưa được biết tới

Hỗ trợ, cải thiện đáng kể hiệu suất phần mềm

Giảm đáng kể thời gian và chi phí phát triển các hệ thống điện toán

Phân loại KBS

Hệ chuyên gia

Các hệ thống liên kết

Các hệ CSDL liên kết, tương tác người dùng thông minh

Các hệ thống dựa trên tri thức cho Công nghệ phần mềm

Các hệ thống dựa trên tri thức cho đào tạo thông minh

Các khó khăn khi xây dựng KBS

Phụ thuộc vào tri thức: giới hạn bởi năng lực chuyên gia

Ví dụ: IF “xe không khởi động được” THEN “Lỗi trong hệ thống điện”

Mối quan hệ giữa hai khái niệm ở trên không thể diễn đạt bằng một hàm số cụ thể, mà chỉ có thể sử dụng hàm heuristic (kinh nghiệm) được đưa ra bởi chuyên gia.

Phụ thuộc vào đặc tính (động hoặc tĩnh) của tri thức: một số tri thức luôn thay đổi theo thời gian, dẫn đến hệ thống KBS cần phải luôn phát triển.

Phụ thuộc vào độ lớn của tri thức: để giải quyết một vấn đề nào đó, cần phải có một lượng tri thức đủ lớn. Đôi khi, các tri thức khác nhau có các cấu trúc khác nhau, cần có các quy tắc để kết nối thông tin giữa chúng.

Các khó khăn khi xây dựng KBS

Phụ thuộc quá trình thu thập tri thức: việc diễn đạt tri thức về cùng một sự vật hiện tượng của các chuyên gia khác nhau sẽ khác nhau

- Các kỹ sư tri thức trình bày yêu cầu với các chuyên gia để lấy thông tin về cách thức giải quyết yêu cầu của chuyên gia
- Chuyển đổi các tri thức từ chuyên gia sang tri thức của hệ thống
- Hiện nay chưa có một thủ tục chung được định trước cho việc thu thập và mô tả tri thức

Quá trình học của hệ thống diễn ra chậm: để giải quyết một vấn đề, có thể phải áp dụng nhiều tri thức, kỹ thuật và công cụ. Cần đủ thời gian để hình thành các tiến trình của KBS và môi trường áp dụng, tạo sự liên kết giữa KBS và cơ sở dữ liệu

Lập trình thông minh

Theo truyền thống: Program = data + algorithm

Với KBS: Intelligence = Knowledge + Inference

Các ngôn ngữ, công cụ để xây dựng KBS:

- Các công cụ truyền thống: PROLOG, LISP (List Processing)
- Các công cụ tiên tiến: AIML (Artificial Intelligence Modeling Language), Matlab, JavaNNS (Java Neural Network Simulator), CLIPS (C Language Integrated Production System).