

**Đại Học Quốc gia TP HCM**

**Chương trình đào tạo thạc sĩ CNTT qua mạng**

---

**Bài giảng môn học:**

# **Công nghệ tri thức và ứng dụng**

**Giảng viên : GS.TSKH. Hoàng Kiếm**

# Nội dung môn học

---

Mở đầu: Giới thiệu tổng quan

## **Phần I: Quản lý tri thức (knowledge management)**

Chương 1: Tiếp nhận, biểu diễn tri thức

Chương 2: Tối ưu hóa CSTT

## **Phần II: Các hệ CSTT (knowledge-based systems)**

Chương 3: Bên trong một hệ CSTT

Chương 4: Phân loại các hệ CSTT

Chương 5: Một số hệ điển hình

## **Phần III: Khai mỏ dữ liệu và khám phá tri thức (Data mining and Knowledge Discovery)**

Chương 6: Máy học & khám phá tri thức.

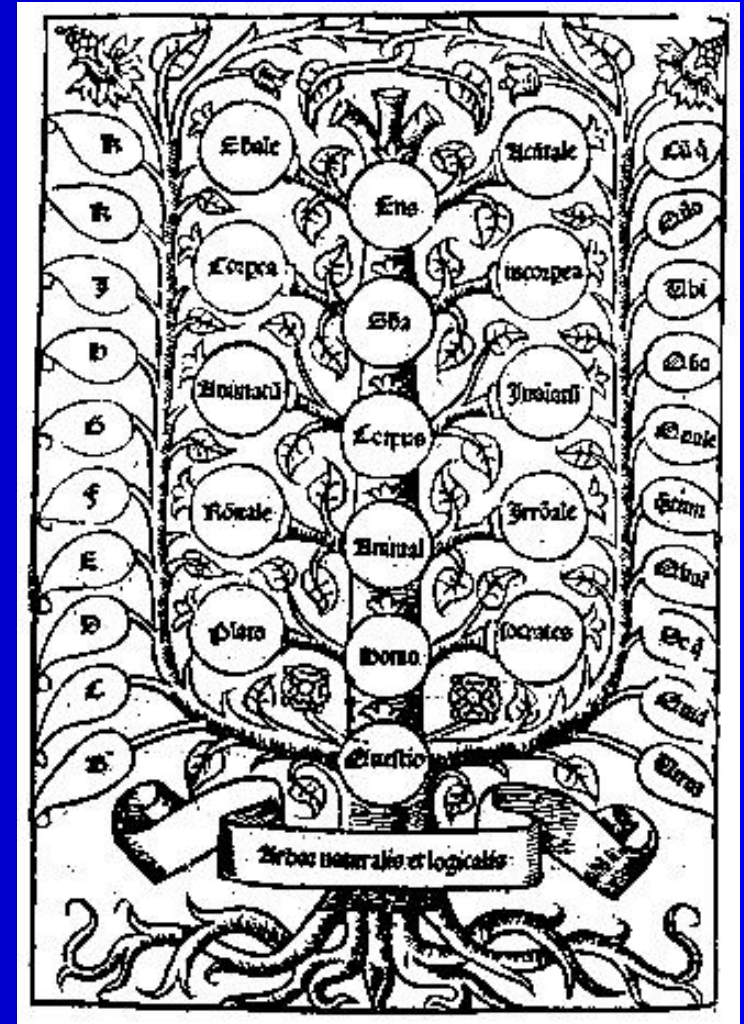
Chương 7: Khai mỏ dữ liệu.

Tổng kết: Tóm tắt, giới thiệu một số công trình nổi bật

## Mở đầu: Giới thiệu tổng quan

# Công nghệ tri thức là gì ?

❖ Công nghệ tri thức (Knowledge Engineering): có thể xem là một nhánh nghiên cứu của trí tuệ nhân tạo, phân tích tri thức lĩnh vực và chuyển nó thành những mô hình tính toán đưa vào máy tính để phục vụ những nhu cầu cần thiết. (*John F.Sowa. Knowledge representation: Logical, philosophical, and Computational Foundations. Copyright @2000 by Brooks/Cole. A division of Thomson Learning*)



# Công nghệ tri thức là gì ? (tt)

---

❖ Công nghệ tri thức (Knowledge Engineering): là các phương pháp, kỹ thuật được những kỹ sư tri thức (knowledge engineers) dùng để xây dựng những hệ thống thông minh như: hệ chuyên gia, hệ cơ sở tri thức, hệ hỗ trợ quyết định, etc. *(Dr Dickson Lukose. Department of Mathematics, Statistics and Computer Science - The University of New England. Dr Rob Kremer Department of Computer Science The University of Calgary Calgary, Alberta, T2N 1N4 Canada. Courses: KNOWLEDGE ENGINEERING, PART A: Knowledge Representation. July 1996)*

- ❖ Công nghệ tri thức là những phương pháp, kỹ thuật dùng để:
  - ✓ Tiếp nhận, biểu diễn tri thức.
  - ✓ Xây dựng các hệ cơ sở tri thức
  - ✓ Khám phá tri thức



# Khoa học tri thức (knowledge science)

JAIST: National graduate institution for advancement of the frontiers of science and technology

**Khoa học về sáng tạo, quản lý, khai thác, sử dụng tri thức**

**Dựa trên sự kết hợp của Khoa học thông tin (TTNT) + Khoa học hệ thống + Khoa học xã hội (kinh tế)**

- school of information science (1992)
- school of materials science (1993)
- school of **knowledge science** (1998)



# Vai trò của công nghệ tri thức

---

❖ Cùng với sự phát triển nhanh chóng, vượt bậc của ngành công nghiệp máy tính, nhu cầu của người dùng đối với máy tính ngày một cao hơn: không chỉ giải quyết những công việc lưu trữ, tính toán bình thường, người dùng còn mong đợi máy tính có khả năng thông minh hơn, có thể giải quyết vấn đề như con người. Và từ đó trí tuệ nhân tạo nói chung và đặc biệt là công nghệ tri thức ra đời và phát triển

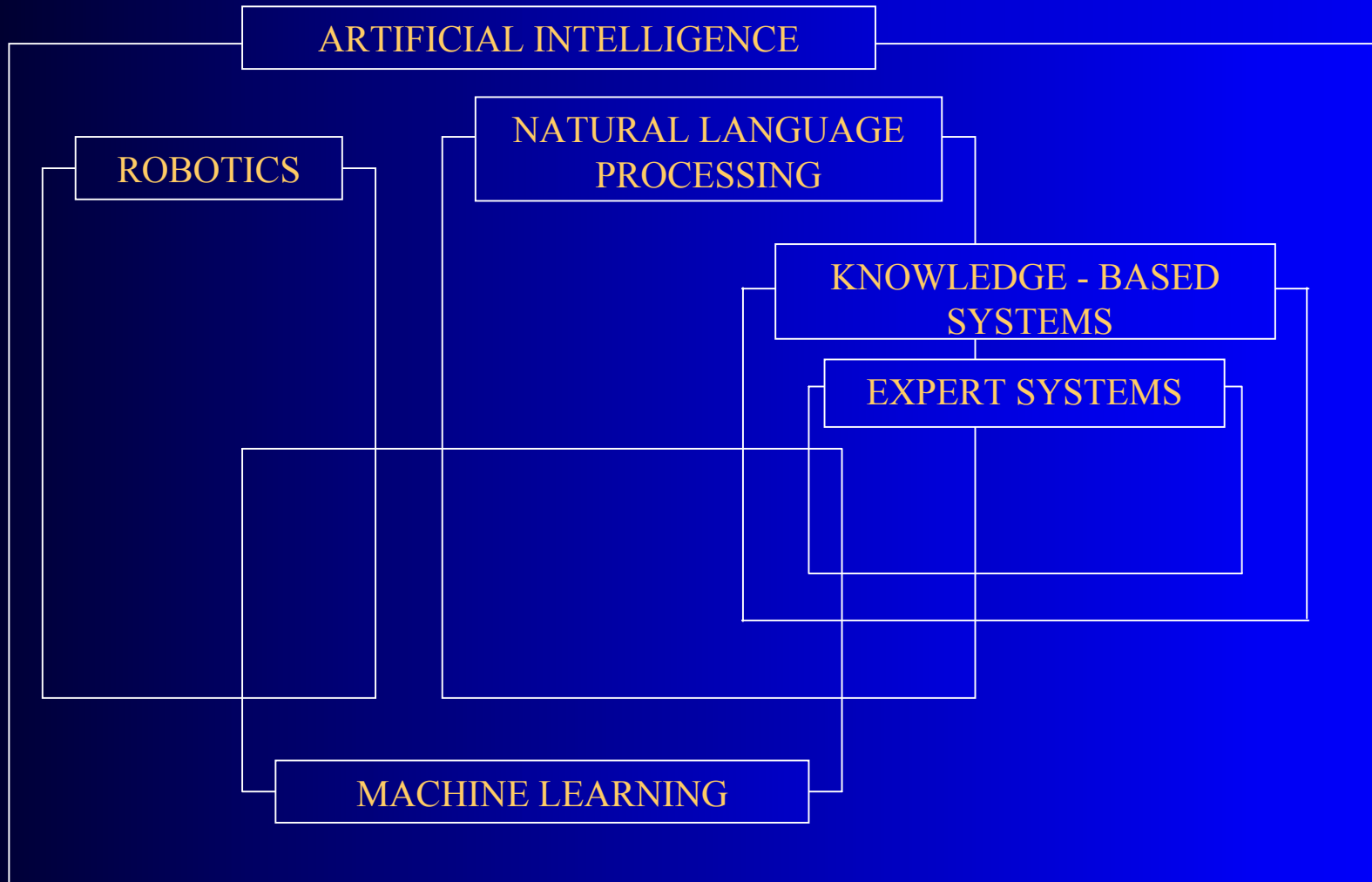
❖ Công nghệ tri thức đóng vai trò hết sức quan trọng trong việc phát triển Công nghệ thông tin, nâng cao sự hữu dụng của máy tính, giúp con người gần gũi với máy tính hơn.

❖ Công nghệ tri thức còn góp phần thúc đẩy nhiều ngành khoa học khác phát triển, khả năng phát triển khoa học dựa trên tri thức liên ngành

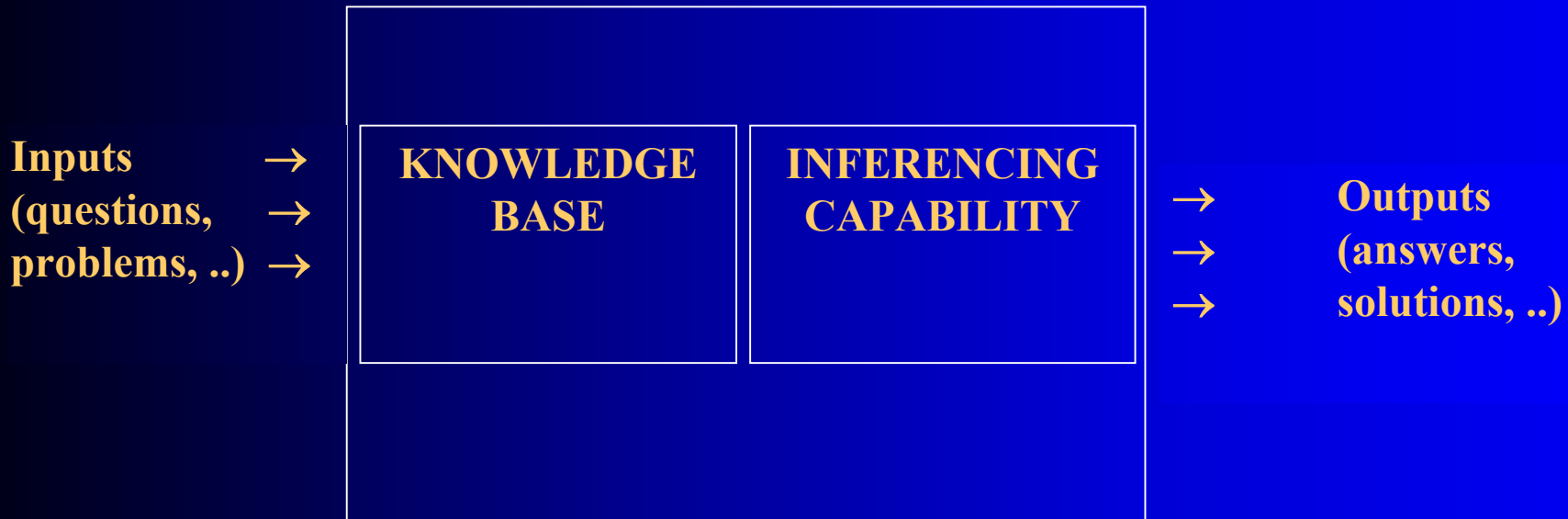
❖ ...

# Các lĩnh vực trong thông minh nhân tạo (AI)

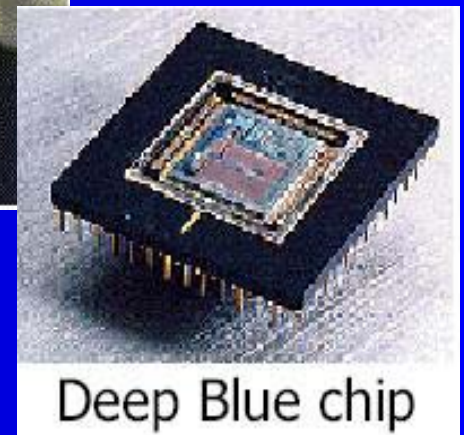
---



# Áp dụng các khái niệm của AI vào máy tính



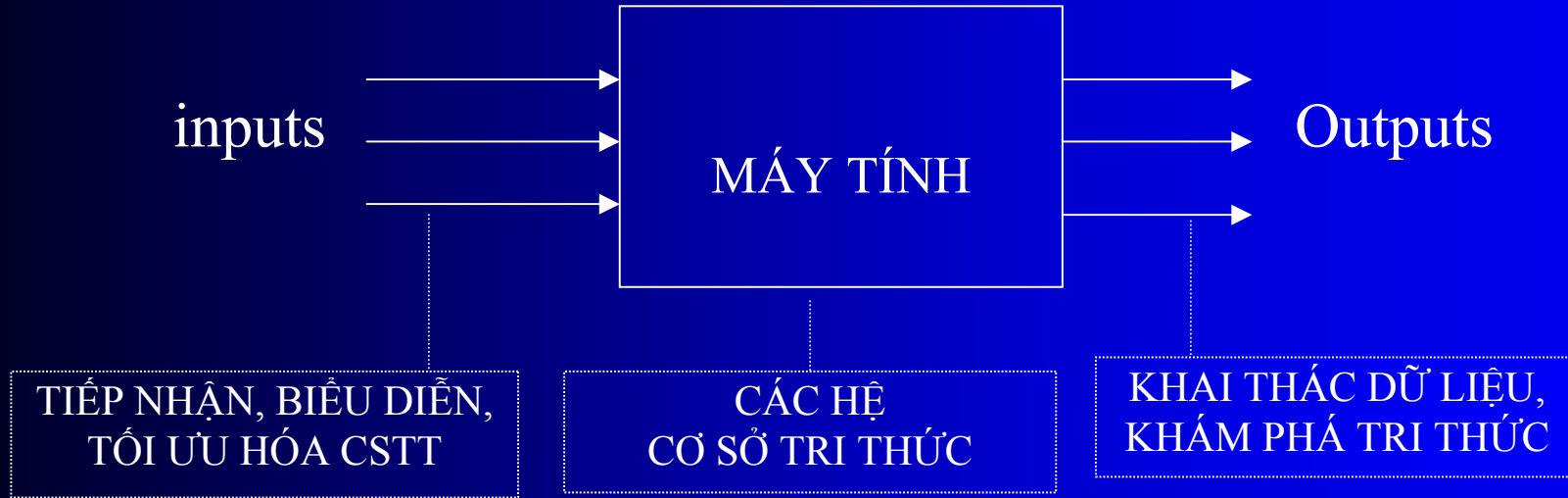
❖ Một trong những mục tiêu quan trọng của lĩnh vực nghiên cứu này là làm cho máy tính có khả năng tiếp nhận, giải quyết vấn đề giống như con người, thậm chí hơn cả con người (máy tính IBM Deep Blue đã chiến thắng vua cờ Kasparov).



Deep Blue chip

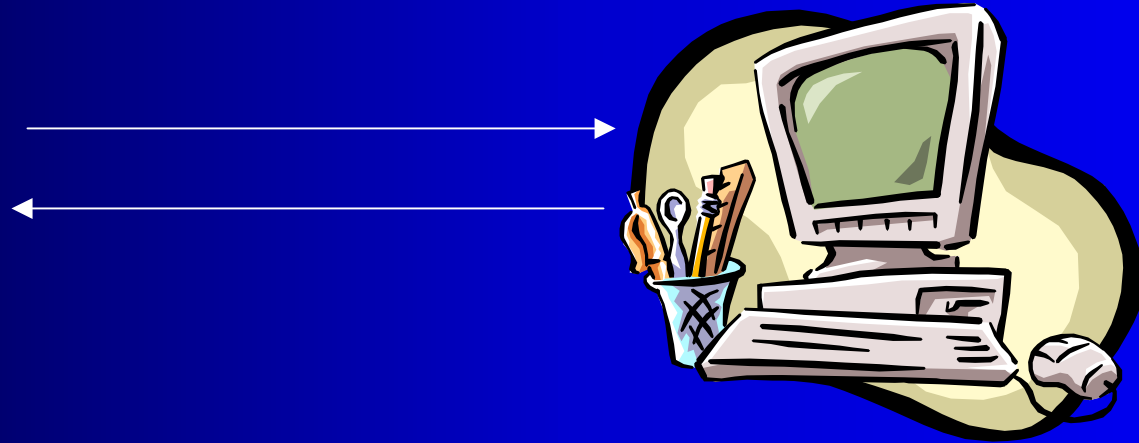
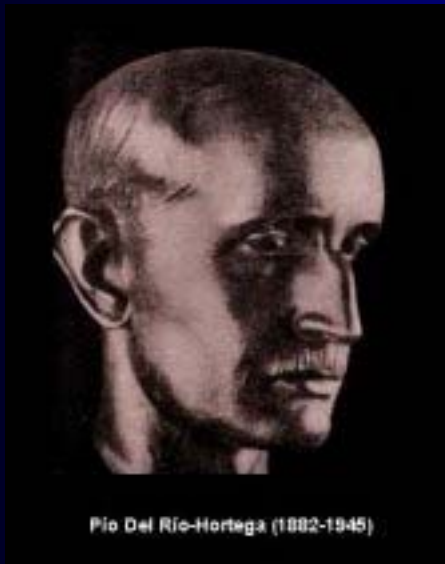


# Hướng nghiên cứu, phát triển công nghệ tri thức



- ❖ **Quản lý tri thức (knowledge management):** bao gồm tiếp nhận, biểu diễn và tối ưu hóa cơ sở tri thức...
- ❖ **Các hệ cơ sở tri thức (knowledge-based systems):** tìm hiểu cấu trúc bên trong của một hệ cơ sở tri thức, phân loại các hệ cơ sở tri thức, và một số hệ cơ sở tri thức điển hình.
- ❖ **khai mở dữ liệu, khám phá tri thức (Data mining, knowledge discovery):** nghiên cứu về phương pháp, kỹ thuật để khai mở dữ liệu và khám phá tri thức.

# Đưa tri thức vào máy tính



- Nhận thức  $\Leftrightarrow$  Tiếp nhận, biểu diễn và tối ưu hóa cơ sở tri thức
- Suy luận  $\Leftrightarrow$  Động cơ suy diễn
- Phản ứng  $\Leftrightarrow$  Phản ứng, trả lời
- Tình cảm  $\Leftrightarrow$  “Bộ xử lý tình cảm” ?

# Quản lý tri thức: Tiếp nhận tri thức

---

Có thể chia thành 2 cách để tiếp nhận tri thức như sau:

## ❖ Thụ động

- Gián tiếp: những tri thức kinh điển.
- Trực tiếp: những tri thức kinh nghiệm (không kinh điển) do “chuyên gia lĩnh vực” đưa ra.

## ❖ Chủ động

- Đối với những tri thức tiềm ẩn, không rõ ràng hệ thống phải tự phân tích, suy diễn, khám phá để có thêm tri thức mới

# Quản lý tri thức: Tiếp nhận tri thức

## ❖ Giao tiếp người-máy

In: Keyboard, Mouse, sensors, touch-pad, touchable screen, speech-recognition, ...

Out: text, graphics, voice, ...



Con người



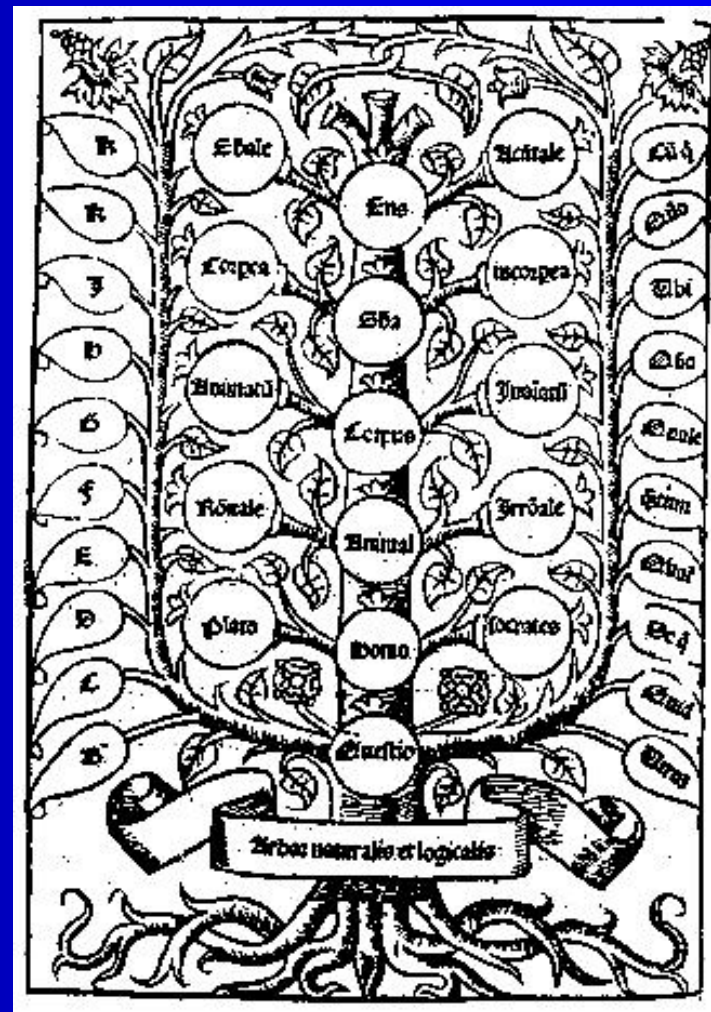
Bộ xử lý ngôn ngữ tự nhiên



# Quản lý tri thức: Biểu diễn tri thức

## ❖ Phương pháp biểu diễn tri thức

- Logic mệnh đề & logic vị từ
- Hệ luật dẫn
- Đối tượng-thuộc tính-giá trị
- Mạng ngữ nghĩa
- Frame
- Script
- ...





# Quản lý tri thức: Tối ưu cơ sở tri thức

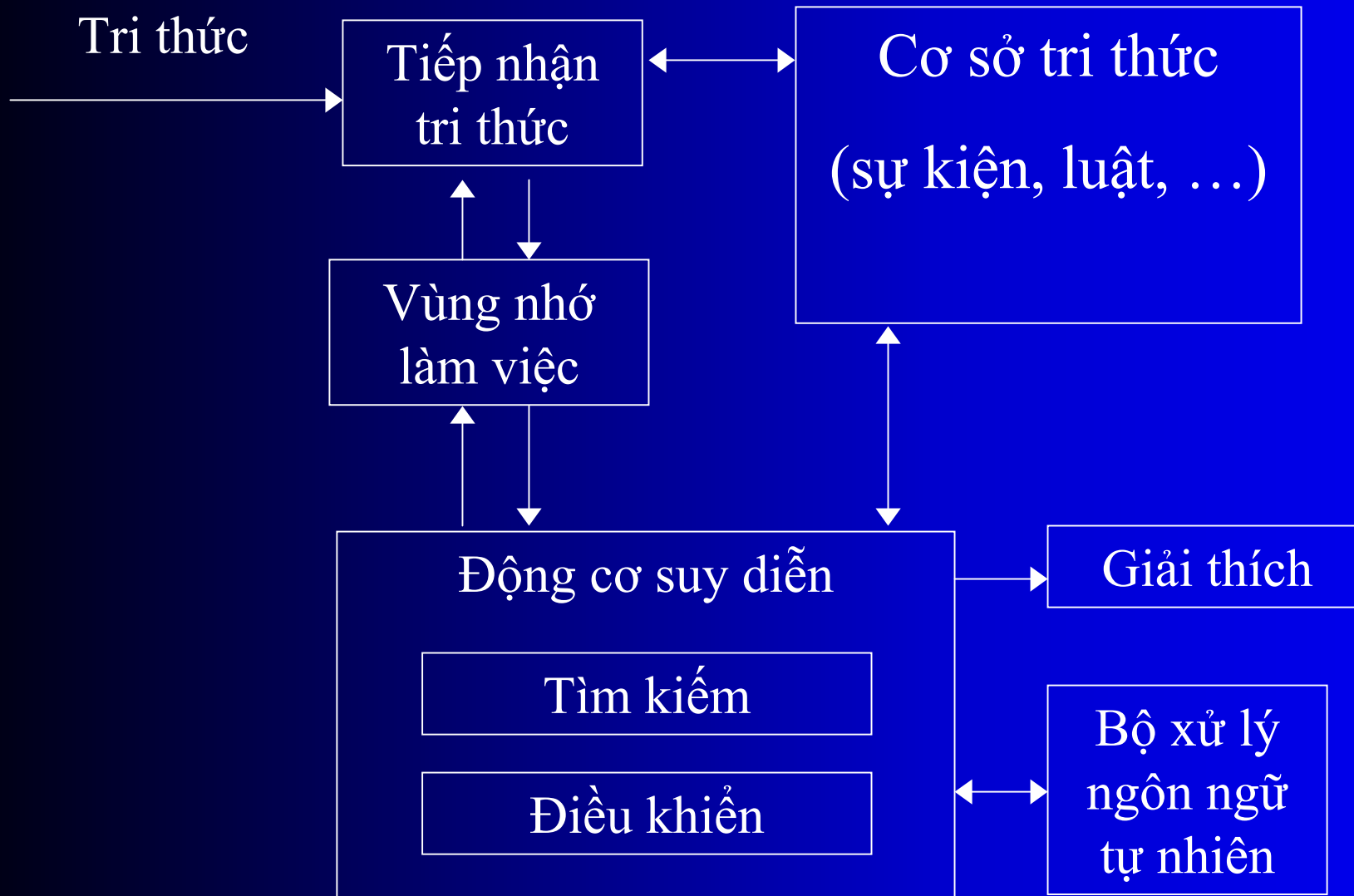
---

## Tại sao tối ưu cơ sở tri thức ?

- ❖ Vấn đề mâu thuẫn, trùng lặp, dư thừa nảy sinh khi tri thức được tiếp nhận và biểu diễn trong cơ sở tri thức. Vì vậy đòi hỏi chúng ta phải có phương pháp để tối ưu cơ sở tri thức.
- ❖ Tùy thuộc vào cách biểu diễn tri thức, chúng ta sẽ có phương pháp thích hợp để tối ưu cơ sở tri thức.

Ví dụ: diễn hình cho vấn đề này là bài toán loại bỏ luật thừa trong cơ sở tri thức luật.

# Tổng quan hệ cơ sở tri thức



# Các hệ cơ sở tri thức: đóng, mở, kết hợp

---

❖ **Hệ cơ sở tri thức đóng:** là những hệ 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.

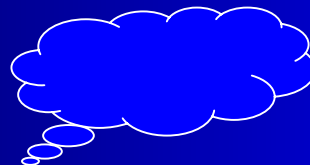
Ví dụ: những hệ cơ sở tri thức về kinh dịch, những hệ giải toán, thường là những hệ cơ sở tri thức giải quyết vấn đề...



# Các hệ cơ sở tri thức: đóng, mở, kết hợp (tt)

❖ **Hệ cơ sở tri thức mở:** là những hệ cơ sở tri thức tiên tiến hơn, nó có khả năng bổ sung tri thức trong quá trình hoạt động, khám phá.

Ví dụ: Những hệ giải toán cho phép bổ sung tri thức trong quá trình suy luận (tri thức ban đầu là những tiên đề và một số định lý, tri thức bổ sung là những định lý mới, những tri thức heuristics, ...); những hệ cơ sở tri thức chẩn đoán, dự báo chẳng hạn: hệ chẩn đoán y khoa MYCIN và EMYCIN, những hệ dự báo thời tiết, khí hậu, động đất, ...



# Các hệ cơ sở tri thức: đóng, mở, kết hợp (tt)

---

❖ **Hệ cơ sở tri thức kết hợp:** bao gồm sự kết hợp giữa hệ đóng và hệ mở, hệ kết hợp giữa CSTT và CSDL, hệ kết hợp giữa hệ CSTT này với một hệ CSTT khác, ... Những hệ cơ sở tri thức kết hợp thường phát triển mạnh dựa trên tri thức liên ngành.

Ví dụ: những hệ hỗ trợ ra quyết định trong đời sống, kinh tế và khoa học; (kinh dịch, tử vi áp dụng với đời sống; kinh dịch, tử vi áp dụng với y học; ...); những hệ chẩn đoán, dự báo đòi hỏi tri thức liên ngành; ...



# Các hệ cơ sở tri thức: phân loại theo phương pháp biểu diễn tri thức

---

Tùy thuộc vào phương pháp biểu diễn tri thức mà chúng ta có thể phân loại các hệ cơ sở tri thức

- ❖ Hệ cơ sở tri thức dựa trên logic mệnh đề và logic vị từ
- ❖ Hệ cơ sở tri thức dựa trên luật dẫn
- ❖ Hệ cơ sở tri thức dựa trên đối tượng
- ❖ Hệ cơ sở tri thức dựa trên Frame
- ❖ Hệ cơ sở tri thức dựa trên mạng ngữ nghĩa
- ❖ ...

Hệ CSTT kết hợp một số phương pháp biểu diễn đã nêu trên.

# Các hệ cơ sở tri thức: phân loại theo ứng dụng

---

❖ **Hệ giải quyết vấn đề:** thường là hệ có tính chất đóng, nhưng đôi khi cũng có hệ mang tính mở.

Ví dụ: Những hệ giải toán, thuật giải Vương Hạo, thuật giải Robinson, ...

❖ **Hệ hỗ trợ quyết định:** thường là các hệ mang tính kết hợp (CSDL + tri thức ngành + hàm toán học + ..), đối tượng sử dụng là các nhà lãnh đạo.

Ví dụ: những hệ thống đánh giá doanh nghiệp (tình hình tài chính, kết quả kinh doanh, qui trình nghiệp vụ, qui trình sản xuất, tính chuyên nghiệp trong quản lý, ...), những hệ thống lập kế hoạch (planning), ...

# Các hệ cơ sở tri thức: phân loại theo ứng dụng (tt)

❖ **Hệ dự báo, chẩn đoán:** thường cũng giống như những hệ hỗ trợ ra quyết định với tính ngoại suy cao hơn.

Ví dụ: Bài toán chẩn đoán hỏng hóc xe, chẩn đoán y khoa, dự báo thị trường chứng khoán, thời tiết ...

❖ **Hệ điều khiển:** là những hệ điều khiển có gắn với CSTT. Những hệ thống này thường ứng dụng trong công nghiệp, trong điều khiển tự động hóa, thường là những hệ thống thời gian thực (real-time systems). Một số hệ thống này có sử dụng kết hợp lý thuyết mờ để xử lý.

Ví dụ: Máy giặt, Máy bơm nước với bộ điều khiển mờ, ...

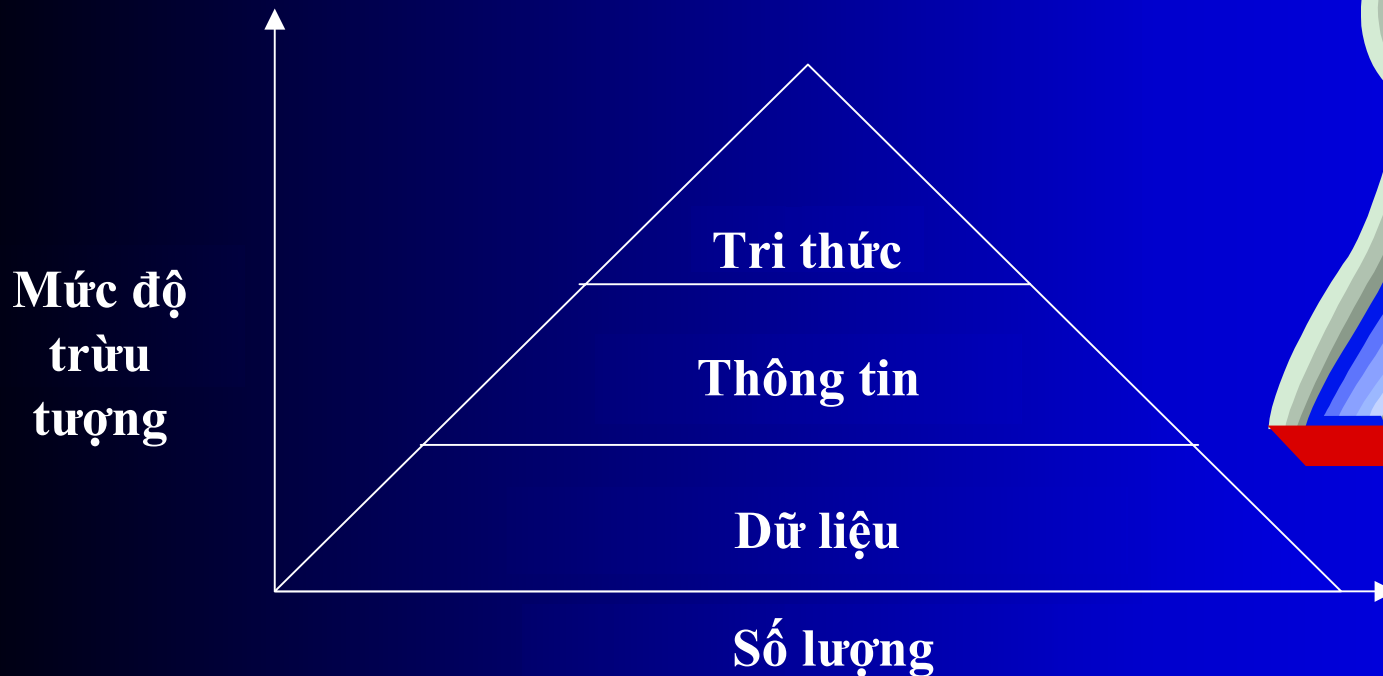


# Máy học và khám phá tri thức

Thế nào là khám phá tri thức (knowledge discovery) ?

❖ Khám phá tri thức là tìm ra những tri thức tiềm ẩn, những tri thức mới (không phải là những tri thức kinh điển, kinh nghiệm, ...)

Thừa dữ liệu, thông tin nhưng thiếu tri thức.



# Máy học và khám phá tri thức (tt)

---

Ví dụ: Trong toán học

❖ Dữ liệu: 1, 1, 2, 3, 5, 8, 13, 21, 34, ...

Mối liên hệ này có thể được biểu diễn bằng công thức sau:  $U_n = U_{n-1} + U_{n-2}$

⇒ Công thức tìm ra ở trên chính là tri thức

Ví dụ: Trong vật lý

Công thức:  $U = I \times R$  là tri thức rút ra từ thực nghiệm

Ví dụ: Chuồn chuồn bay thấp thì mưa, bay cao thì nắng, bay vừa thì râm

⇒ Lời nhận xét trên là tri thức rút ra từ kinh nghiệm đời sống.



# Máy học và khám phá tri thức (tt)

---

Thế nào là máy học (Learning Machine) ?

❖ Máy tính hay chương trình máy tính có khả năng *tự hoàn thiện* từ “*kinh nghiệm*”.

❖ Máy học còn có nghĩa là việc mô hình hóa môi trường xung quanh hay khả năng một chương trình máy tính sinh ra một cấu trúc dữ liệu mới khác với cấu trúc hiện có. Chẳng hạn việc tìm ra những luật *If...then...* từ tập dữ liệu đầu vào.

(Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski. *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers, 1998)

# Máy học và khám phá tri thức (tt)

---

**Phân loại các phương pháp máy học:** có nhiều quan điểm phân loại khác nhau

Phân loại thô:

- ❖ Học giám sát (supervised learning)
- ❖ Học không giám sát (unsupervised learning)

Phân loại theo 2 tiêu chuẩn cùng lúc: “cấp độ học” & “cách tiếp cận”

Cấp độ học:

- ❖ Học vẹt (Rote learning)
- ❖ Học theo giải thích (by explanation)
- ❖ Học theo ví dụ, trường hợp (by examples, cases)
- ❖ Học khám phá (by discovering)

# Máy học và khám phá tri thức (tt)

---

Cách tiếp cận:

- ❖ Tiếp cận thống kê
- ❖ Tiếp cận toán tử logic
- ❖ Tiếp cận hình học (phân hoạch không gian, xây dựng cây định danh, ...)
- ❖ Tiếp cận mạng Neural
- ❖ Tiếp cận khai mở dữ liệu
- ❖ ...

# Nhà kho dữ liệu và khai mỏ dữ liệu

---

Tạo DATA WAREHOUSE = Biến đổi dữ liệu thành tri thức yểm trợ  
tiến trình ra quyết định.

## Dữ liệu

Khoa học

Giáo dục

Công nghệ

Kinh doanh

Thị trường

Thời tiết

...

Tạo Data Warehouse



Tri thức yểm trợ ra quyết  
định

# Nhà kho dữ liệu và khai mỏ dữ liệu (tt)

---

DatawareHouse = Business Information + Decision Making

*(IBM BPEC'96 Conference, San Diego, USA)*

Sự bùng nổ của các CSDL lớn vượt quá khả năng diễn dịch và lĩnh hội của con người, phát sinh yêu cầu sáng tạo các công cụ kỹ thuật mới để phân tích dữ liệu một cách thông minh và tự động nhằm tạo ra tri thức hữu dụng hỗ trợ tốt cho tiến trình ra quyết định.

*(Usama, Data Mining and Knowledge Discovery, 1995)*



# Nhà kho dữ liệu và khai mỏ dữ liệu (tt)

## Một số bài toán điển hình về data mining

Bài toán khám phá luật kết hợp

Bài toán nhận dạng mẫu

Bài toán phân loại dữ liệu

Bài toán gom nhóm dữ liệu

Bài toán lập mô hình

Bài toán dự báo

...



# Tài liệu tham khảo

---

- [1] GS.TSKH Hoàng Kiếm. Bài giảng cao học môn học cơ sở tri thức và ứng dụng. ĐHKHTN-TPHCM.
- [2] GS.TSKH Hoàng Kiếm. Thư viện những báo cáo khoa học, bài thu hoạch môn cơ sở tri thức và ứng dụng. Các lớp cao học thuộc khoa CNTT- ĐHKHTN. TPHCM.
- [3] GS.TSKH Hoàng Kiếm, TS. Đỗ Văn Nhơn, Th.sĩ Đỗ Phúc. Giáo trình Các hệ cơ sở tri thức. Đại Học Quốc Gia TPHCM – 2002.
- [4] GS.TSKH Hoàng Kiếm, Th.sĩ Đinh Nguyễn Anh Dũng. Giáo trình Trí tuệ nhân tạo. Đại Học Quốc Gia TPHCM – 2002.
- [5] John F.Sowa. Knowledge representation: Logical, Philosophical, and Computational Foundations. Copyright @ 2000 by Brooks/Cole. A division of Thomson Learning.
- [6] Adrian A.Hopgood. Knowledge-based systems for Engineers and Scientists. The Open University – CRC Press. Boca-Raton Ann-Arbor London Tokyo 1998.
- [7] Sharon Wood. Planning and decision making in dynamic domains.Ellis Horwood Series in Artificial Intelligence - 1998.
- [8] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski. Data Mining Methods for Knowledge Discovery. Kluwer Academic Publishers, 1998
- [9] Citeseer - Scientific Literature Digital Library. Artificial Intelligence-  
<http://citeseer.nj.nec.com/ArtificialIntelligence/> - 2003

# Thư viện số CiteSeer

Một số nghiên cứu đang được quan tâm về AI

<http://citeseer.nj.nec.com/ArtificialIntelligence/>

- Expert Systems :

<http://citeseer.nj.nec.com/ArtificialIntelligence/ExpertSystems/>

- Knowledge Representation:

<http://citeseer.nj.nec.com/ArtificialIntelligence/KnowledgeRepresentation/>

- Natural Language Processing:

<http://citeseer.nj.nec.com/ArtificialIntelligence/NaturalLanguageProcessing/>

- Optimization:

<http://citeseer.nj.nec.com/ArtificialIntelligence/Optimization/>

- Planning:

<http://citeseer.nj.nec.com/ArtificialIntelligence/Planning/>

- Robotics:

<http://citeseer.nj.nec.com/ArtificialIntelligence/Robotics/>

