



Xây dựng, phát triển và đánh giá thuật toán

Nhập môn Công nghệ thông tin 1

Nội dung

- Nghiên cứu khoa học
- Nghiên cứu thuật toán
- Vai trò và chức danh trong nghiên cứu khoa học



Nghiên cứu khoa học



Khái niệm

- Nghiên cứu khoa học thường được mô tả là một quy trình tìm hiểu tích cực, cần cù và có hệ thống nhằm khám phá, lý giải tri thức hay thậm chí tạo ra những tri thức mới.



Phân loại nghiên cứu khoa học

- Nghiên cứu thường được chia làm hai loại:
 - Nghiên cứu cơ bản: phát triển các lý thuyết hiện có nhằm làm cho nó càng gần giống với thế giới tự nhiên.
 - Nghiên cứu ứng dụng: cách thức đưa các lý thuyết vào sản xuất các sản phẩm phục vụ đời sống.

"Điện đã có thể không bao giờ được phát minh nếu người ta chỉ lo việc cải tiến những ngọn nến"



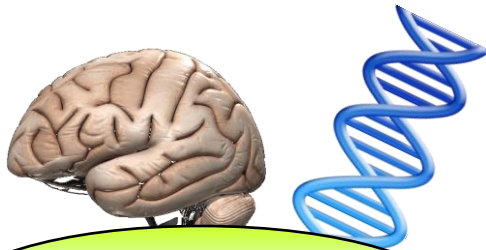
Tại sao nghiên cứu khoa học

- Mức độ tổng quát:
 - Giúp tri thức nhân loại ngày càng mở rộng và phát triển.
 - Đáp ứng được nhu cầu và thỏa mãn của con người nhiều hơn.
 - ...
- Mức độ cá nhân:
 - Để kiếm sống.
 - Để thỏa **đam mê** khám phá.
 - ...

(Thảo luận)



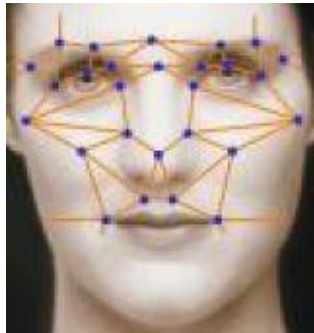
Một số hướng nghiên cứu



**Tính toán
mềm**



**Khai thác
dữ liệu**



Nhận dạng



Thị giác máy tính

Một số hướng nghiên cứu



Search engine



Bài toán Trí tuệ nhân tạo



Chẩn đoán y khoa



Bioinformatics

Một số hướng nghiên cứu



**Thị trường chứng khoán
Tài chính, ngân hàng**

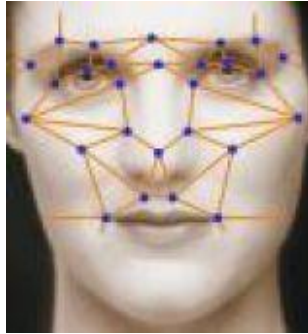


Hệ thống siêu thị



Tổng hợp, phân loại, gom cụm văn bản

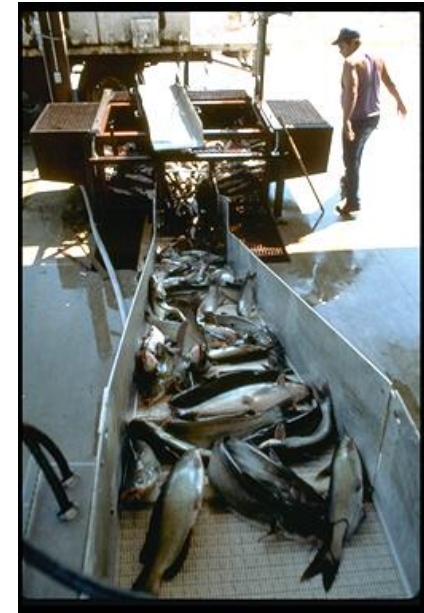
Một số hướng nghiên cứu



Nhận dạng mặt người



Nhận dạng chữ viết



**Phân loại cá
trong công nghiệp**

Nghiên cứu thuật toán



Khái niệm

- Thuật toán hay giải thuật nói chung là một tập hợp hữu hạn của các chỉ thị hay phương cách được định nghĩa cho việc hoàn tất một số việc từ một trạng thái ban đầu cho trước dẫn đến kết quả mong muốn.
- Một bài toán có thể được giải quyết bởi các thuật toán khác nhau.



Ví dụ

- Thuật toán để giải phương trình bậc nhất $P(x): ax + b = c$ (với a, b, c là các số thực) có thể thực hiện qua một số bước sau:

Nếu $a = 0$

$b = c$ thì $P(x)$ có nghiệm bất kì

$b \neq c$ thì $P(x)$ vô nghiệm

Nếu $a \neq 0$

$P(x)$ có duy nhất một nghiệm $x = (c - b)/a$



Tâm quan trọng của thuật toán

- Boolos & Jeffrey (1974, 1999) đã đưa ra nhận xét sau:
 - *Không có con người nào có thể viết đủ nhanh, đủ dài, đủ nhỏ để liệt kê tất cả các thành phần của một tập rất lớn gần như vô hạn mà chỉ bằng cách lần lượt viết ra tên của chúng theo một số quy ước.*
 - *Tuy nhiên, con người có thể đưa ra cách thức để xác định phần tử thứ n bất kì. Từ đó, cách thức này sẽ được thực hiện bởi các máy điện toán.*



Câu hỏi thuật toán

- Các nhà phát triển thuật toán thường tự đặt 4 câu hỏi phản biện (critical) khi họ đánh giá các thuật toán:
 - Có phải thuật toán giải quyết bài toán đã được nêu ra?
 - Có phải thuật toán rõ ràng, rành mạch?
 - Thuật toán có đưa ra một kết xuất?
 - Thuật toán có kết thúc trong một khoảng thời gian hợp lý?



Quá trình phát triển thuật toán

- Xác định đầu vào
- Xác định tiến trình thực hiện
- Xác định đầu ra
- Phát triển lược đồ HIPO
- Xác định các module liên quan



Xác định đầu vào

- Thuật toán cần dữ liệu gì?
- Như thế nào để có dữ liệu đó?
- Định dạng dữ liệu thế nào?



Xác định tiến trình thực hiện

- Làm cách nào để thao tác với dữ liệu để sinh ra những kết quả có ý nghĩa?



Xác định đầu ra

- Dữ liệu nào cần được trả ra?
- Định dạng dữ liệu trả ra?

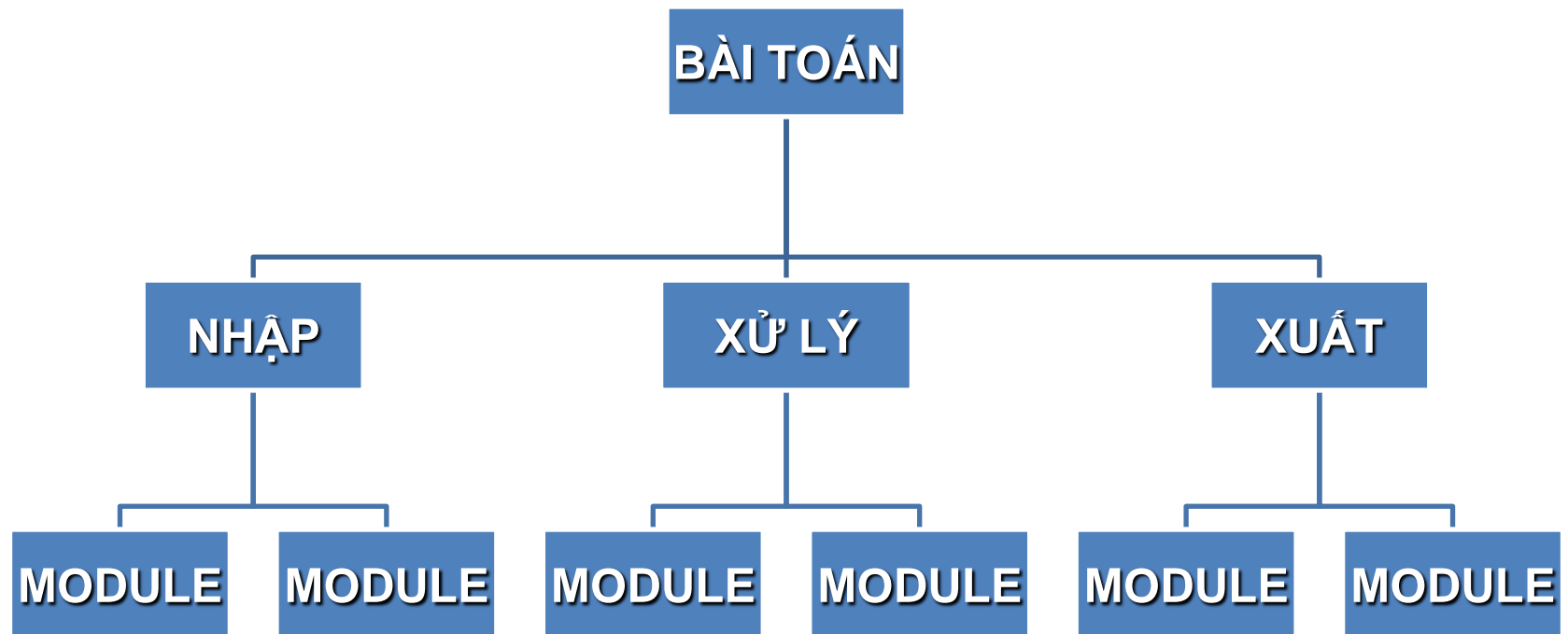


Phát triển lược đồ HIPO

- HIPO (Hierarchy of Inter-Processes-Outputs) là một kĩ thuật phục vụ cho việc lên kế hoạch và ghi tài liệu cho thuật toán.
- HIPO là một biểu đồ phân tầng thể hiện cấu trúc điều khiển và một bộ nhập-xử lý-xuất để mô tả dữ liệu đến, dữ liệu xuất từ đâu và những xử lý được thực thi bởi các module trên lược đồ phân tầng này.



Phát triển lược đồ HIPO



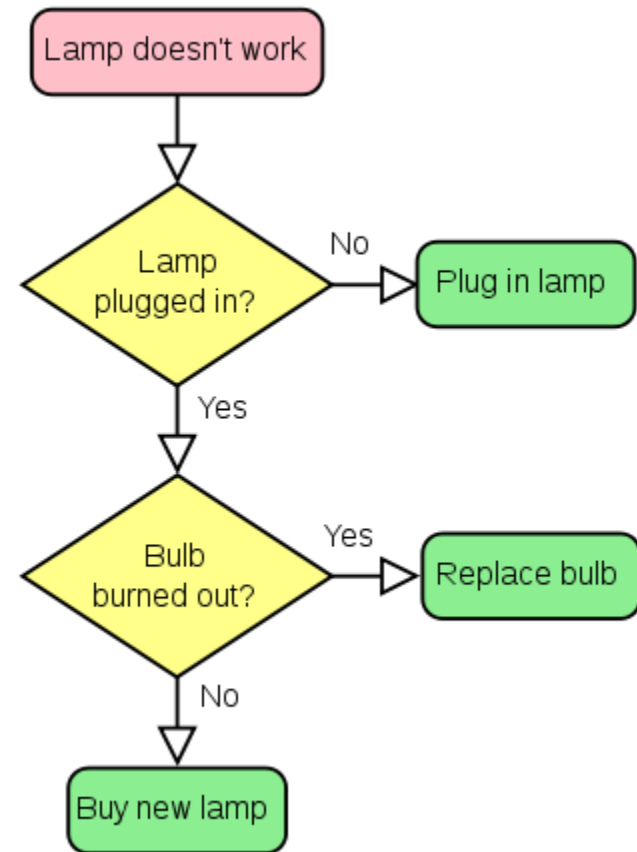
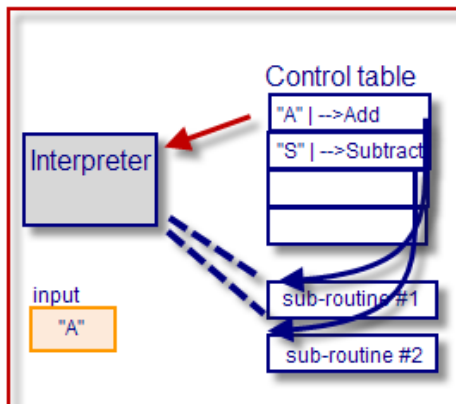
Xác định những module liên quan

- Như thế nào để tách những bài toán lớn thành những mảnh nhỏ hơn và có thể quản lý được?
- Các module cần dữ liệu đầu vào nào?
- Những xử lý cần được thực hiện trong mỗi module?
- Dữ liệu kết xuất của từng module?



Thể hiện thuật toán

- Thuật toán có thể được thể hiện trong:
 - Ngôn ngữ tự nhiên
 - Mã giả
 - Lược đồ flowchart
 - Ngôn ngữ lập trình
 - Bảng điều khiển



Đánh giá thuật toán

- Thuật toán được đánh giá dựa trên khối lượng tài nguyên (thời gian và bộ nhớ) cần để thực thi nó.
 - Độ phức tạp về mặt không gian.
 - Độ phức tạp về mặt thời gian.



Đánh giá thuật toán

- Đánh giá thuật toán quan trọng vì:
 - Việc sử dụng vô ý một thuật toán không hiệu quả có thể ảnh hưởng đến hiệu năng hệ thống.
 - Trong các ứng dụng thời gian thực, một thuật toán chạy quá lâu có thể làm cho kết quả của nó đã lỗi thời hoặc vô dụng.
 - Một thuật toán không hiệu quả cũng có thể tiêu tốn một khối lượng tính toán hay vùng nhớ một cách không kinh tế để chạy.



Phân loại thuật toán

- Có nhiều cách để phân loại thuật toán:
 - Theo cách thực thi: tuần tự, song song,...
 - Theo phương pháp thiết kế: vét cạn, chia để trị, ...
 - Theo lĩnh vực nghiên cứu: tìm kiếm, sắp xếp, ...
 - Theo độ phức tạp: khối lượng thời gian cần để hoàn thành so với kích thước dữ liệu nhập.



Vai trò, chức danh trong nghiên cứu khoa học



Các chức danh khoa học Việt Nam

- Các chức danh trong nghiên cứu khoa học ở Việt Nam:
 - Học vị:
 - Cử nhân: người hoàn thành một chương trình đào tạo các môn khoa học ở cấp đại học.
 - Thạc sĩ: người nắm vững một lĩnh vực trong nghiên cứu khoa học.
 - Tiến Sĩ: người có thể đưa ra các phát kiến mới.
 - Học hàm:
 - Phó Giáo Sư
 - Giáo Sư



