**1. Định nghĩa Chương trình dịch? Các giai đoạn của Chương trình dịch? Cho**

**ví dụ?**

* Định nghĩa: chương trình dịch là một chương trình thực hiện việc chuyển đổi một chương trình hay một đoạn chương trình con từ ngôn ngữ nguồn một cách tương đương sang ngôn ngữ đích
* Ngôn ngữ nguồn là NNLT bậc cao (Pascal, Java, C…)
* Ngôn ngữ đích là NNLT bậc thấp như Assembly hoặc ngôn ngữ máy

Gồm 2 pha, 6 giai đoạn:

* Giai đoạn phân tích:
* Phân tích từ vựng
* Phân tích cú pháp
* Phân tích ngữ nghĩa
* Giai đoạn tổng hợp
* Sinh mã trung gian
* Tối ưu mã trung gian
* Sinh mã đích

Ví dụ: Chương trình GCC (GNU Compiler Collection) là một trình dịch phổ biến được sử dụng để dịch mã nguồn từ ngôn ngữ lập trình C sang mã máy tính có thể thực thi được trên nhiều nền tảng khác nhau.

Ví dụ, khi viết mã nguồn Java, bạn có thể sử dụng trình biên dịch (compiler) của Java để biên dịch mã nguồn thành bytecode. Trình biên dịch của Java sẽ phân tích mã nguồn, dịch nó sang bytecode và tạo ra file .class chứa bytecode. Sau đó, các file .class này có thể được đưa vào môi trường thực thi Java để thực thi chương trình.

Ví dụ: CPython dịch mã nguồn Python thành mã byte code, một mã trung gian giữa mã nguồn và mã máy tính. Mã byte code sau đó được thực thi bởi một trình thông dịch Python. Quá trình này giúp cho mã nguồn Python có thể chạy trên nhiều nền tảng khác nhau mà không cần phải biên dịch lại cho từng nền tảng một cách riêng biệt.

**2. Các thế hệ ngôn ngữ lập trình? Vai trò của Chương trình dịch đối với ngôn**

**ngữ lập trình?**

* 1GL, thế hệ thứ nhất, là dạng ngôn ngữ ở trình độ thuần túy máy móc, nghĩa là các mệnh lệnh và dữ kiện được biểu tượng dưới dạng những con số 0 và 1. Gọi là ngôn ngữ máy.
* 2GL, thế hệ thứ hai, là dạng ngôn ngữ mệnh đề hay "Assembler" (thường được gọi là "Assembly Language"),có nhiệm vụ hoán chuyển (convert) những mệnh đề (statements) thành ngôn ngữ máy móc. Ví dụ điển hình các mệnh lệnh của ngôn ngữ nầy: ADD 12,8 (cộng 12 và 8)
* 3GL, thế hệ thứ ba, là ngôn ngữ lập trình bậc cao (high level), như PL/I, C, hay JAVA, xữ dụng COMPILER để hoán chuyển những mệnh lệnh của loại ngôn ngữ cao cấp sang ngôn ngữ máy móc. Đối với Java, chẳng hạn, mệnh lệnh được hoán chuyển được gọi là BYTECODE,và mệnh lệnh nầy lại được hoán chuyển một lần nữa sang một loại ngôn ngữ máy móc thích ứng qua dạng máy giả định (Virtual machine), một bộ phận trong hệ thống vận hành của máy tính (operating system). Dạng ngôn ngữ thế hệ thứ ba nầy đòi hỏi rất nhiều kiến thức về chương trình điện toán.
* 4GL, thế hệ thứ tư, được phác họa gần giống ngôn ngữ tự nhiên hơn thế hệ thứ ba. Một trong những ví dụ của các mệnh lệnh xữ dụng trong database đại để như sau. EXTRACT ALL CUSTOMERS WHERE "PREVIOUS PURCHASES" TOTAL MORE THAN $1000 (Chọn tất cả những khách hàng nào có tổng số tiền mua hàng trước đây vươt quá $1000).
* 5GL, thế hệ thứ 5, Sử dụng phương pháp tượng hình để tạo mệnh lệnh và những mệnh lệnh nầy được hoán chuyển bằng những compiler của thế hệ thứ ba hay thứ tư. Phương pháp lập chương trình nầy giúp dể nhận diện hệ thống đơn tữ hoá trong lập trình (Object-oreiented Programming) và giúp kéo các hình con để ráp nối càc bộ phận của chương trình.

Về vai trò của chương trình dịch đối với ngôn ngữ lập trình, chương trình dịch là công cụ giúp biên dịch mã nguồn của ngôn ngữ lập trình thành mã máy để máy tính có thể thực thi. Quá trình biên dịch bao gồm các giai đoạn như phân tích cú pháp (parser), dịch mã nguồn thành mã trung gian (compiler), tối ưu hóa (optimizer), và phát hành (linker). Chương trình dịch đóng vai trò quan trọng trong quá trình phát triển phần mềm, giúp cho người lập trình có thể viết mã nguồn ở mức độ cao hơn và tối ưu hơn, và đồng thời tạo ra được các ứng dụng có hiệu suất tốt hơn.

**3. Định nghĩa văn phạm? Bài tập về văn phạm sinh ngôn ngữ? Biểu thức chính quy đặc tả ngôn ngữ?**

Văn phạm (Grammar) trong khoa học máy tính là một tập hợp các quy tắc định nghĩa cách một chuỗi các ký tự có thể được sắp xếp để tạo thành một ngôn ngữ. Văn phạm bao gồm các ký hiệu (Symbols) và các quy tắc để xây dựng các chuỗi ký tự hợp lệ trong ngôn ngữ. Các ký hiệu có thể là các ký tự, từ, hoặc các biểu thức.

Văn phạm G = {VT, VN, P, S}

trong đó:

* VT : tập các kí hiệu kết thúc của bộ chữ (a, b, …)
* VN: tập các kí hiệu không kết thúc (A, B, …)
* S ∈ VN: kí hiệu bắt đầu
* P: tập các sản xuất có dạng α → β (α ∈ V\*VV\*, β ∈ V\*)

Biểu thức chính quy (Regular expression) là một chuỗi ký tự đại diện cho một mẫu ký tự nào đó. Biểu thức chính quy thường được sử dụng để đặc tả các chuỗi ký tự trong một ngôn ngữ cụ thể. Ví dụ, biểu thức chính quy "a\*b" có thể đặc tả các chuỗi ký tự bắt đầu bằng 0 hoặc nhiều ký tự "a", và kết thúc bằng một ký tự "b". Biểu thức chính quy cũng có thể được sử dụng để tạo ra các trình phân tích cú pháp (parser) để xử lý các chuỗi ký tự trong ngôn ngữ.

**4. Định nghĩa otomat? Các thành phần của otomat? Bài tập về otomat đoán**

nhận ngôn ngữ và biểu thức chính quy? Vẽ đồ thị chuyển?

Otomat (Automaton) là một mô hình tính toán được sử dụng để mô tả hành vi của một hệ thống, bao gồm các hệ thống số học, hệ thống điều khiển và các ứng dụng khác. Otomat có thể được xem như một bảng điều khiển tự động, nơi mà các đầu vào được đưa vào và các đầu ra được tạo ra dựa trên trạng thái hiện tại của hệ thống.

Các thành phần của otomat bao gồm:

1. Tập trạng thái (State set): một tập hữu hạn các trạng thái mà otomat có thể ở.
2. Bộ chuyển trạng thái (Transition function): một hàm để xác định trạng thái tiếp theo của otomat dựa trên trạng thái hiện tại và đầu vào.
3. Trạng thái ban đầu (Initial state): trạng thái mà otomat bắt đầu ở.
4. Tập trạng thái kết thúc (Accepting state set): một tập con của tập trạng thái, xác định các trạng thái mà otomat dừng lại và đưa ra đầu ra cho một chuỗi đầu vào.
5. **Biểu thức chính quy?**

Biểu thức chính quy (Regular expression) là một chuỗi ký tự đại diện cho một mẫu ký tự nào đó. Biểu thức chính quy thường được sử dụng để đặc tả các chuỗi ký tự trong một ngôn ngữ cụ thể. Ví dụ, biểu thức chính quy "a\*b" có thể đặc tả các chuỗi ký tự bắt đầu bằng 0 hoặc nhiều ký tự "a", và kết thúc bằng một ký tự "b". Biểu thức chính quy cũng có thể được sử dụng để tạo ra các trình phân tích cú pháp (parser) để xử lý các chuỗi ký tự trong ngôn ngữ.

**Bài tập**













