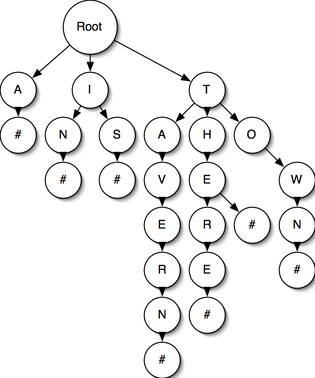
**WEEK 1**

**Bài 1: Cho 2 số nguyên a và b, thực hiện phép chia 2 số này mà không sử dụng phép toán nhân , chia , mod.**

* Độ phức tạp : O(n), O(logn)

**Bài 2: Tìm hiểu cấu trúc dữ liệu Trie, xây dựng ứng dụng auto complete sử dụng cấu trúc Trie.**

1. Cấu trúc dữ liệu Trie ( Cây tiền tố)



* Nút gốc ( root) là rỗng , không lưu kí tự nào cả
* Các nút con của root đại diện cho kí tự đầu tiên của các từ lưu trong Trie
* Các nút con tiếp theo là kí tự thứ 2 của các từ , …cho đến khi từ kết thúc bởi node #

1. Các hàm chính của cấu trúc Trie

* Insert : chèn 1 từ khóa vào trong Trie
* Search: tìm kiếm một từ đã có trong Trie chưa

1. Ứng dụng auto complete

Khi người dùng nhập vào ô tìm kiếm một chuỗi thì sẽ gợi ý tất cả các từ có trong Trie bằng chuỗi đó .

**Bài 3: Viết câu lệnh SQL tìm người có lương cao thứ 2 từ bảng lương của nhân viên ?**

1. Cách 1: Sử dụng với mệnh đề IN

SELECT MAX(salary) FROM employees WHERE salary NOT IN (SELECT MAX(salary) FROM employees);

1. Cách 2: Sử dụng toán tử “<”

SELECT MAX(salary) FROM employees WHERE salary < (SELECT MAX(salary) FROM employees);

1. Cách 3: Sử dụng từ khóa LIMIT trong Mysql

SELECT salary FROM ( SELECT salary FROM employees ORDER BY salary DESC LIMIT 2) AS emp ORDER BY salary LIMIT 1;

**Câu 4: Học có giám sát là gì ? Học không có giám sát là gì ? Lấy ví dụ. Implement mô hình Linear Regression.**

1. Học có giám sát ( Supervise Learning)

* Là thuật toán dự đoán đầu ra (outcome) của một dữ liệu mới dựa trên các cặp (input, outcome) đã biết từ trước
* Cần học một hàm y = f(x) từ tập học {(x1, y1), (x2, y2 ), …(xN, yN)} sao cho yi

= f(xi)

* Học có giám sát chia làm 2 loại chính:

+ Phân loại ( classification): nếu y chỉ nhận giá trị từ một tập rời rạc

+ Hồi quy ( Regression): nếu y nhận giá trị là một số thực

* Ví dụ:

Lọc thư rác

Phân loại trang web

Nhận dạng chữ số viết tay

1. Học không giám sát ( Unsupervise learning)

* Thuật toán chỉ có dữ liệu đầu vào mà không biết được outcome( nhãn)
* Dựa vào cấu trúc dữ liệu để phân nhóm hoặc giảm số chiều của dữ liệu
* Cần học một hàm y = f(x) từ tập học cho trước { x1, x2, …, xN}
* Ví dụ:

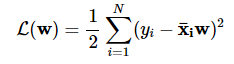
Phân cụm ( Kmeans) : Phân nhóm khách hàng dựa trên hành vi mua hàng

Phát hiện các cộng đồng trong mạng xã hội

Phát hiện xu hướng

1. Implement mô hình Linear Regression

* Linear Regression là học có giám sát
* Sử dụng hàm lỗi ( loss function) :



* Tìm nghiệm w\* bằng cách lấy đạo hàm của hàm lỗi cho bằng 0 và giải phương trình thu được:

w\* = ( AT A)-1 ATy

* Trong đó A là ma trận dữ liệu cỡ Nx(m+1) : N là số dữ liệu qua sát được, m là số thuộc tính của một input.