Information retrieval

1. Bài toán truy tìm thông tin

* Cho trước:
* 1 tập tài liệu lớn, ổn định
* Một nhu cầu thông tin
* Bài toán: tìm tất cả tài liệu có liên quan đến nhu cầu thông tin

1. Những vấn đề cần giải quyết:

* Biểu diễn tập tài liệu như thế nào (xử lý đánh chỉ mục)
* Biểu diễn nhu cầu thông tin như thế nào (xử lý công thức truy vấn)
* Bằng cách nào hệ thống có thể trả về những tài liệu liên quan đến nhu cầu thông tin một cách hiệu quả
* Kết quả trả về dk trình bày như thế nào?(Mô hình truy tìm thông tin)

1. Đánh chỉ mục

Một số hình thức đánh chỉ mục phổ biến

* Cây hậu tố
* Chỉ mục ngược (được dùng trong đa số bộ máy tìm kiếm hiện nay)
* Chỉ mục trích dẫn
* Chỉ mục Ngram

1. Chỉ mục ngược

Ý tưởng: với mỗi từ t, lưu một danh sánh những tài liệu chứa t.