

BÁO CÁO SƠ BỘ TIỂU LUẬN CUỐI KỲ

Tên đề tài: Phân tích và So sánh các thuật toán khớp chuỗi

Nhóm thực hiện: Nhóm 3

Môn học: Thiết kế và đánh giá thuật toán

Giảng viên hướng dẫn: Nguyễn Thị hồng Minh, Trần Bá Tuấn

1. Mục tiêu của đề tài

- Tìm hiểu, phân tích nguyên lý hoạt động của các thuật toán khớp chuỗi: Naive, Rabin-Karp, Boyer-Moore.
- Cài đặt các thuật toán bằng Python.
- Đánh giá, so sánh hiệu suất các thuật toán qua thực nghiệm trên các văn bản có độ dài khác nhau, vị trí khác nhau.

2. Các công việc đã thực hiện

- Đọc tài liệu tham khảo và giáo trình liên quan về các thuật toán.
- Phân tích nguyên lý hoạt động và độ phức tạp thời gian, không gian.
- Cài đặt thành công thuật toán Naive và Rabin-Karp bằng Python.
- Thực hiện kiểm thử với các văn bản mẫu, đo thời gian thực thi, số lần so khớp.
- Viết phần mô tả lý thuyết và triển khai của hai thuật toán trên.

3. Minh chứng công việc đã làm

- Minh chứng được đính kèm với file
- Có báo cáo, slide, code đã làm trong thời gian qua

4. Các công việc sẽ thực hiện trong thời gian tới

- So sánh hiệu năng của ba thuật toán trên tập dữ liệu lớn hơn (văn bản Unicode, tiếng Việt).

- Test nhiều trường hợp văn bản hơn, tìm dữ liệu text không dùng dữ liệu tự sinh nữa
- Viết và hoàn thiện các phần còn thiếu của báo cáo chính thức
- Hoàn thành slide thuyết trình và phần trình bày cho buổi báo cáo cuối kỳ.

5. Phân công công việc nhóm

- Nguyễn Thùy Dương: Tìm hiểu lý thuyết phần Rabin - Karp, viết báo cáo , code thuật toán, giao diện cho các thuật toán
- Lê Thị Hạnh: Tìm hiểu lý thuyết phần Naive, Cài đặt mã nguồn và kiểm thử các thuật toán, test.
- Đặng Thùy Dương: Tìm hiểu lý thuyết phần Boyer Moore, code thuật toán, làm slide.