

CHALLENGE 01: SO KHỚP CHUỖI

1 Nội dung

1.1 Tìm hiểu

Đối sánh chuỗi (string matching) là một nhánh các thuật toán tìm kiếm được ứng dụng rộng rãi trong đời sống cũng như nhiều lĩnh vực khoa học khác nhau. Hãy tham khảo **chương 32**, sách **Introduction to Algorithms** (*3rd edition*) và thực hiện các yêu cầu sau:

1. Định nghĩa bài toán *Đối sánh chuỗi*, bao gồm các điều kiện đầu vào và yêu cầu bài toán. Nêu ra 2 ứng dụng của so khớp chuỗi trong thực tế, chỉ rõ các **input** và **output** tương ứng.
2. Giải thích thuật toán, cho ví dụ (chạy từng bước), chỉ ra độ phức tạp (giải thích) và lập bảng so sánh các thuật toán đối sánh chuỗi sau:
 - Brute-force (Naive String-matching)
 - Rabin-Karp
 - Knuth-Morris-Pratt

1.2 Lập trình

- **Crossword game:** Cho một bảng kí tự kích thước $W \times H$ như **Hình 1**. Viết một chương trình minh họa cho phép đưa vào chuỗi kí tự (là một từ có nghĩa), hãy cho biết từng chuỗi đó xuất hiện tại tọa độ nào trong bảng kí tự đã cho theo hàng ngang (từ trái sang phải) hoặc hàng dọc (từ trên xuống dưới).
- **Input:** File *input.txt* có nội dung như sau.
 - Dòng thứ nhất: 2 số nguyên dương W và H lần lượt tượng trưng cho chiều ngang và chiều cao của bảng kí tự, cách nhau bởi khoảng trắng ().
 - H dòng tiếp theo, mỗi dòng W kí tự, cách nhau bởi khoảng trắng.
 - Từ dòng thứ $(H + 2)$: Mỗi dòng chứa một chuỗi ký tự (là một từ có nghĩa), cần được tìm kiếm trong bảng kí tự.
 - Kết thúc khi gặp dòng chứa kí tự #

	1	2	3	4	5	6	7	8	9	10
1	U	C	M	A	R	V	E	L	O	L
2	S	F	D	Q	U	E	U	E	G	I
3	T	R	A	V	E	R	S	A	L	N
4	E	F	G	S	T	A	X	F	N	K
5	K	D	P	U	Z	U	V	U	C	E
6	I	S	I	B	L	I	N	G	S	D
7	E	N	N	T	S	L	Y	C	L	L
8	D	E	G	R	E	E	A	B	I	I
9	G	X	Z	E	M	O	Q	R	F	S
10	E	R	T	E	B	G	R	O	O	T

Hình 1. Bảng kí tự

• Output:

- Dòng đầu tiên: Số từ tìm được.
- Các dòng tiếp theo: Theo cấu trúc [**Word**_**Pos**_**Dir**] (_ là khoảng trắng). Trong đó:
 - * **Word** là từ được cho.
 - * **Pos** là toạ độ của kí tự bắt đầu, có dạng (x, y) ($x, y \geq 1$). Trả về (0, 0) nếu không tìm được từ.
 - * **Dir** là chiều của từ: **LR** nếu chiều của từ từ trái sang phải, **TD** nếu chiều từ trên xuống dưới. Trả về **NF** nếu không tìm được từ.

• Ví dụ:

Input	Output
9 10 U C ... L ... E R ... T MARVEL LIST XXX #	2 MARVEL (1,3) LR LIST (7,10) TD XXX (0, 0) NF

2 Yêu cầu và chấm điểm

2.1 Yêu cầu

- Sinh viên làm bài theo nhóm 4 người. Đăng kí và nhận mã nhóm tại [link](#) Google Drive đi kèm. Chỉ chấm tối đa 10 nhóm.

Lưu ý: Chỉ đăng ký nhóm sau khi đã hoàn thành bài làm. Thời gian nộp bài và đăng kí nhóm không được chênh lệch quá 15pht.

- Nhóm sinh viên nộp bài trên Moodle dưới dạng [Mã_nhóm.rar] hoặc [Mã_nhóm.zip], là file nén của thư mục [Mã_nhóm]. Thư mục này bao gồm:
 - File báo cáo được trình bày dưới dạng văn bản [Mã_nhóm.pdf] hoặc dạng slide [Mã_nhóm.pptx]. File này chứa nội dung tìm hiểu ở phần 1.1 và trình bày hướng giải quyết vấn đề lập trình ở phần 1.2.
 - * Nếu bài nộp file báo cáo là dạng slide, cần giải thích cặn kẽ từng slide ở phần *Note* của slide đó.
 - * Trang đầu (hoặc slide đầu) của File báo cáo cần ghi rõ thông tin (Họ và tên, MSSV) của từng thành viên trong nhóm.
 - * File báo cáo cần trình bày có cấu trúc, logic, rõ ràng, mạch lạc, trả lời thẳng vào trọng tâm của câu hỏi. Độ dài của bài nộp không vượt quá 15 trang đối với file văn bản, và 30 trang đối với slide.
 - File lập trình [Mã_nhóm.cpp]. Yêu cầu mã nguồn cần trình bày logic, có chú thích rõ ràng.

2.2 Chấm điểm

- Điểm bài Challenge 01 sẽ chiếm 2% điểm tổng kết.
- Phần file nộp: Tối đa 10 điểm.
- Các GV sẽ chọn nhóm có file bài nộp tốt nhất, và một nhóm ngẫu nhiên **có nộp bài** để trình bày. Điểm thuyết trình: Cộng tối đa 5 điểm vào phần file nộp.
- Các bài nộp sai quy định sẽ không được chấm điểm.
- Các bài làm giống nhau sẽ bị **0** điểm môn học.