

Đề thi lập trình C nâng cao

(Thời gian làm bài 120 phút - Được phép dùng tài liệu tham khảo)

Trong lĩnh vực xuất bản, mỗi ấn phẩm có thể trích dẫn những ấn phẩm khác dưới dạng tài liệu tham khảo. Tuy có nhiều cách xác định một ấn phẩm (bằng tên ấn phẩm, tên tác giả, số xuất bản v.v), nhưng để đơn giản hóa vấn đề chúng ta coi mỗi ấn phẩm có một tên duy nhất, là một chuỗi ANSCII không chứa khoảng trắng và ít hơn 100 ký tự. Ngoài ra, để quá trình lưu trữ được tối ưu, chúng ta sẽ cấp cho mỗi ấn phẩm một mã số duy nhất (số nguyên, trị tuyệt đối nhỏ hơn 10^6), gọi là mã ấn phẩm. Như vậy có thể coi vấn đề trích dẫn trong lĩnh vực xuất bản như vấn đề đồ thị, trong đó mỗi ấn phẩm là một đỉnh của đồ thị, mỗi trích dẫn là một cạnh có hướng của đồ thị. Chúng ta sẽ gọi ngắn gọn những đồ thị như vậy là đồ thị trích dẫn.

Định dạng tệp dữ liệu đồ thị: Dữ liệu đồ thị trích dẫn được cho trong tệp văn bản theo cấu trúc sau:

- Dòng đầu tiên chứa 2 số nguyên: N (số lượng ấn phẩm) và M (số lượng trích dẫn), được ngăn cách bởi một dấu cách.
- Mỗi dòng trong N dòng tiếp theo chứa thông tin về một ấn phẩm gồm: Tên ấn phẩm và mã số ấn phẩm được ngăn cách bởi một dấu cách. Dữ liệu vào đảm bảo tên và mã số cho mỗi ấn phẩm là duy nhất, không trùng lặp.
- Mỗi dòng trong M dòng tiếp theo chứa thông tin về một trích dẫn dưới dạng một cặp mã số u v, được ngăn cách nhau bởi một dấu cách, thể hiện ấn phẩm u trích dẫn ấn phẩm v.

Để nâng cao hiệu quả xử lý đồ thị trích dẫn cần viết chương trình đọc dữ liệu từ tệp văn bản và tự động thực hiện các chức năng theo yêu cầu.

Yêu cầu: Chương trình phải hoạt động theo tham số dòng lệnh tương tự các công cụ trong môi trường Linux. Các chức năng xử lý và cấu trúc tham số dòng lệnh được liệt kê trong *Danh sách chức năng*. Một số ví dụ câu lệnh chạy chương trình và kết quả mong đợi được cho trong bảng 1 và bảng 2.

Chú ý: Lưu đồ thị theo danh sách kê sử dụng cây đồ đen (thư viện JRB) là bắt buộc. In và chỉ in ra màn hình những thông tin theo yêu cầu, không in ra bất kỳ thông báo phụ nào khác.

Gợi ý: Có thể đọc file chỉ sử dụng hàm `fscanf` vì tên ấn phẩm không chứa khoảng trắng.

Danh sách chức năng:

Trong danh sách chức năng này sử dụng các ký hiệu sau:

pr: Tên chương trình, sinh viên tự dịch code của mình thành chương trình với tên là pr;

file_name: Tên tệp văn bản chứa dữ liệu đồ thị theo cấu trúc đã mô tả, v.d., g.txt;

id, i, j, u, v: Là các mã số ấn phẩm, v.d., 101;

item_name: Là tên ấn phẩm, v.d., TruyenKieu.

1. *Trợ giúp & giới thiệu (help & about)*. Xử lý tham số phù hợp với bảng 1 để khi chạy lệnh trong cột “Lệnh”, chương trình in ra thông báo như trong cột “Thông báo”. **[2 điểm]**

Bảng 1. Chức năng trợ giúp & giới thiệu

Lệnh	Thông báo
<code>./pr help</code>	Supported command: help, i2n, n2i, cc, cr, odg, rel.
<code>./pr about</code>	C advanced 20142 final exam.

2. *Tra cứu tên ấn phẩm theo mã số (i2n: id to name)*. Khi chạy lệnh theo cú pháp

`./pr i2n file_name id`

Chương trình cần in ra tên ấn phẩm có mã số là id trong đồ thị cho trong file_name. **[1 điểm]**

3. *Tra cứu mã theo tên ấn phẩm (n2i: name to id)*. Ngược với i2n, khi chạy lệnh theo cú pháp

`./pr n2i file_name item_name`

Chương trình cần in ra mã số của ấn phẩm có tên là item_name trong đồ thị cho trong file_name. **[1 điểm]**

4. *Cùng trích dẫn (cc: co-citation)*. Khi chạy lệnh theo cú pháp

`./pr cc file_name i j`

Chương trình cần in ra tập hợp tên của tất cả các ấn phẩm được trích dẫn đồng thời bởi ấn phẩm i và j trong đồ thị được cho trong file_name, $\{u | i \rightarrow u; j \rightarrow u\}$. Mỗi ấn phẩm một dòng. **[2 điểm]**

5. Cùng được trích dẫn (cr: co-reference). Khi chạy lệnh theo cú pháp

`./pr cr file_name i j`

Chương trình cần in ra tập hợp tên của tất cả các ấn phẩm trích dẫn đồng thời i và j trong đồ thị được cho trong file_name. $\{u | u \rightarrow i; u \rightarrow j\}$. *Mỗi ấn phẩm một dòng.* [2 điểm]

6. Bậc ra lớn hơn (odg: out-degree greater). Khi chạy lệnh theo cú pháp

`./pr odg file_name k`

Chương trình cần in ra tập hợp tên của tất cả các ấn phẩm có bậc ra của đỉnh tương ứng lớn hơn k trong đồ thị cho trong file_name. *Mỗi ấn phẩm một dòng.* [1 điểm]

7. Ấn phẩm liên quan (rel: related). Khi chạy lệnh theo cú pháp

`./pr rel file_name u`

Chương trình cần in ra tập hợp tên các ấn phẩm được trích dẫn bởi u hoặc được trích dẫn bởi ít nhất một ấn phẩm được trích dẫn bởi u trong đồ thị file_name. *Mỗi ấn phẩm một dòng.* [1 điểm]

Ví dụ:

Dữ liệu sử dụng trong ví dụ được cho trong tệp *g.txt* như hình 1, gồm 12 dòng đánh số từ 1 tới 12. Các số thứ tự dòng trong hình 1 sẽ không có trong *g.txt*.

Hình 1. Ví dụ tệp *g.txt* chứa đồ thị trích dẫn hợp lệ

g.txt:					
1	5 6	5	Omega 8	9	1 8
2	Alpha 1	6	Theta 100	10	3 8
3	Beta 3	7	1 5	11	5 100
4	Gamma 5	8	3 5	12	100 1

Bảng 2. Ví dụ chạy chương trình

STT	Câu lệnh	Kết quả	Ghi chú
1	./pr i2n g.txt 5	Gamma	Đúng
		Tên đỉnh la: Gamma	Sai (quy cách output)
2	./pr n2i g.txt Omega	8	Đúng
	./pr n2i g.txt omega	8	Sai (phân biệt chữ hoa/thường).
3	./pr cc g.txt 1 3	Gamma Omega	Đúng (có thể in theo thứ tự bất kỳ)
4	./pr cr g.txt 5 8	Alpha Beta	Đúng (có thể in theo thứ tự bất kỳ)
5	./pr odg g.txt 0	Alpha Beta Gamma Theta	Đúng (có thể in theo thứ tự bất kỳ)
6	./pr rel g.txt 5	Theta Alpha	Đúng (có thể in theo thứ tự bất kỳ)

Xác nhận của bộ môn

G.V. Ra đề

TS. Nguyễn Bá Ngọc