PERSONAL PROJECT

reported by

20112096

Joonyoung Choi

**Design of data-sturcture**

# 0 프로젝트 개요

실생활에서 찾을 수 있는 dynamic graph로 표현될 수 있는 자료를 선택하고 graph 구조로 표현하는 방법과 실질적 데이터를 취합한다. 그렇게 구축한 주제의 graph가 나타내는 바를 파악하고 평가할 수 있도록 한다.

* 진행에 대한 내용은 아래에 링크된 Github를 통해 업데이트한다.
* [20112096 GitHub](https://github.com/mydream757/Design-of-DataStructure)
* 프로젝트 진행 내역은 지속적으로 첨가/수정될 수 있다.
* 다음에 진행 계획을 첨부한다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 9월 | | | | | | |
| S | M | T | W | T | F | S |
|  |  |  |  |  |  | 1 |
| 2 | 3 | 4 | 5 | 6 | 7  Problem definition & modeling | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20  Problem solving | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10월 | | | | | | |
| S | M | T | W | T | F | S |
| 30 | 1 | 2 | 3 | 4  Implementation & Data collection | 5 | 6 |
| 7 | 8 | 9 | 10 | 11  Testing & Evaluation | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 28 | 29 | 30 | 31 |  |  |  |

1. Problem definition & modeling (9/5~912)

주제 조건 확인

후보군 선정

적합성 평가 및 주제 선정

모델링

# 1 Problem definition & modeling

1. 주제 조건 확인

주제(문제) 선정을 위해 다음과 같은 조건을 확인하였다.

* 데이터 간의 관계를 일정한 기준으로 비교할 수 있는가
* 실질적인 데이터를 수집할 수 있는가
* 그래프로 구성할 수 있는가

<주제 조건 표 예시>

|  |  |  |  |
| --- | --- | --- | --- |
|  | 후보1 | 후보2 | 후보3 |
| 데이터 간의 관계를 일정한 기준으로 비교할 수 있는가 |  |  |  |
| 실질적인 데이터를 수집할 수 있는가 |  |  |  |
| 그래프로 구성할 수 있는가 |  |  |  |

1. 후보군 선정
2. 노래

노래들의 ‘사용자 리뷰, 장르, 작곡가 등’을 이용.

흥미를 느낀 음악과 연관성이 높은 음악을 찾을 가능성을 더 높일 수 있을 것.

1. 도서

도서들의 ‘제목의 키워드, 장르, 사용자 리뷰, 저자 등’을 이용

아주 많은 도서들 중 원하는 내용의 책을 찾을 가능성을 더 높일 수 있을 것.

1. 역사

사건의 ‘발생 년도, 대상, 키워드, 성격 등’을 이용

다양한 사건 간의 관계를 파악하는 데에 용이할 것.

1. 범죄

범죄 분류 중 하나를 정하여 ‘범행 수법, 피해자 유형, 범행시기 등’의 데이터를 이용

C와 동일한 이유로 후보 선정

1. 연예인 간 관계

연예인들의 출연작, 출연 프로, 데뷔 시기 등의 데이터 이용하여 구성

어떤 프로그램에서 출연시킬 게스트를 구상하거나, 새로운 프로그램의 방송 패널 구상을 고려할 때 사용될 수 있을 것이라 판단.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A 노래 | B 도서 | C 역사 | D 범죄 | E 연예인 간 관계 |
| 데이터 간의 관계를 일정한 기준으로 비교할 수 있는가 | 가장 중요한 ‘리듬’ 을 비교함이 어려움 | 키워드, 장르, 내용 등 | 발생 년도, 키워드, 유형 외에 파악 중 | 범행 수법, 피해자 유형, 범행시기 | 출연작, 출연 프로, 데뷔 시기, 연애 관계 등 |
| 실질적인 데이터를 수집할 수 있는가 | 멜론 등 음악사이트 | 인터파크 등 도서 판매 사이트 | 한국사 연표 및 기타 자료 | (대검찰청 사이트 통계 자료 외 상세열람 불가)  뉴스 기사 정보 활용 | 네이버 등 포털 사이트 정보 이용 |
| 그래프로 구성할 수 있는가  (구성한다면?) | Undirected graph  Vertex – songs  Arc – similarity | Undirected graph  Vertex – Books  Arc – similarity | Directed graph  Vertex – events  Arc – relation | Undirected graph  Vertex – events  Arc - similarity | Undirected graph  Vertex – entertainer  Arc - relation |

1. 적합성 평가 및 주제선정

* A, B, D의 경우 그래프 구조의 필요성에 대한 의문이 제기되었고, C는 사건 간의 연관성 파악에 어려움이 크다는 것 때문에 주제 선정에서 배제하였다.
* E의 경우, 소규모 집단을 그래프로 표현해 볼 수 있다는 점과 이것이 SNS 등의 네트워크 구조와 유사할 수 있다는 점을 생각해봤을 때 그래프로 구성하기에 적절하다고 판단하였다.
* 또한 2) 후보군 선정에서 제시한 바와 같이 그래프로 얻을 수 있는 정보를 활용할 방안이 충분히 실용적이라 판단하였다.

1. 모델링

* 주제: 연예인들 간 관계 그래프 모델링
* 설계:
* 그래프 형태: 무방향 그래프
* Vertex: 인물
* Arc: 연관성(무방향)
* Weight: 인물 간 공통되는 출연작의 수에 따른 가중치
* 데이터 수집: 포털 사이트의 인물 검색과 위키 자료를 활용할 것.
* 자료 구조

\*모종의 이유로 은퇴를 하거나(정점 삭제) 신규 활동하는 연예인(정점 삽입)이 발생한다.

\*\*정점의 개수에 비해 간선의 개수가 그리 많지 않을 것이다.  
\*\*\*출연작에 따라 간선이 추가된다(간선 삽입).

위의 조건에 의해 다음과 같이 구성할 것이다.

|  |  |
| --- | --- |
| 사용 언어 | C++ |
| 정점 타입 | Struct PERSON{  이름  나이  출연작 리스트  } |
| 정점들에 대한  컨테이너 | Vector<PERSON> p; |
| 간선의 리스트 | Pair<int , int> adj[]  (first: Vertex의 index, second: weight) |

정점 타입: 구조체 타입으로 정점(연예인)에 대한 정보를 포함

PERSON

이름

나이

출연작

…

정점들에 대한 컨테이너: 리스트 형태로 구축하기 위해 vector 타입을 사용

P

P + 1

P +2

간선의 리스트: pair 타입을 사용해 정점의 index와 가중치를 저장할 것

adj[1]

adj[2]

adj[3]

adj[4]

3

3

4

1

3

2

1

3

4

1

1

3

1

1

2

1

\*자료 구조 시각화

3

2

1

1