



# 상 경 계 열

성명		지원 학부·학과		수험 번호										
----	--	----------	--	-------	--	--	--	--	--	--	--	--	--	--

## 유의 사항

1. 90분 안에 답안을 작성하십시오.
2. 답안지는 검정색 펜(샤프, 볼펜, 연필)으로 작성하십시오.
3. 답안지와 문제지, 연습지를 함께 제출하십시오.
4. 다음 경우는 0점 처리됩니다.
  - 1) 답안지를 검정색 펜(샤프, 볼펜, 연필)으로 작성하지 않은 경우
  - 2) 자신의 신원을 드러내는 표기나 표현을 한 경우
  - 3) 답안을 해당 답란에 작성하지 않은 경우

※ 감독의 지시가 있을 때까지 다음 장으로 넘기지 마시오.

[문제1] (가)를 참조하여 (나)의 ㉠가 어떤 것일지 설명하고, A사의 ‘추천 시스템’과 구글의 ‘페이지랭크’의 방법적 특징을 비교한 후, (다)의 ㉡가 페이지랭크에서는 어떤 조건들 때문에 발생 가능성이 적은지 기술하시오.(600자, 50점)

(가)

인터넷 주문형 비디오 공급사인 A사의 영화 추천 시스템이 호평을 받고 있다. A사는 내가 좋아할 영화를 어떻게 아는 걸까? A사는 나를 모른다. 하지만 내가 과거에 무슨 영화를 재미있게 보았는지는 안다. 보았던 영화들에 내가 평점을 주었기 때문이다. A사의 회원 수가 충분히 많다면, A사는 이 정보와 함께 다른 회원들이 준 평점을 이용하여 나의 영화 취향을 꽤 정확하게 예측할 수 있다. A사의 영화 추천 시스템은 상관관계에 바탕을 두고 있다. 우선 내가 몇 편의 영화에 평점을 준다. A사는 내가 준 평점을 다른 사람들의 평점과 비교하여 내 평점과 상관관계가 높은 평점을 준 사람들을 찾아낸다. 그 사람들은 내가 좋아하는 영화를 좋아하는 경향이 있다. 이 작업을 토대로 A사는 나와 취향이 비슷한 회원들에게 높은 평점을 받았지만 내가 아직 보지 않은 영화를 추천할 수 있다.

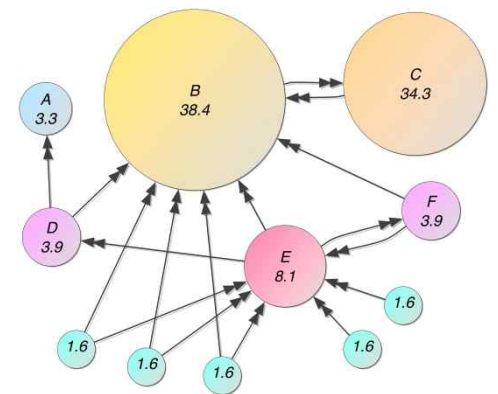
(나)

지난 세기말 등장한 구글의 검색 엔진도 사용자에게 만족도 높은 탐색 결과를 제시함으로써 단번에 인터넷 검색 시장을 석권하였다. 웹 페이지의 ‘유용성’ 또는 그것이 사용자에게 주는 ‘만족도’ 같은 주관적 특성을 어떻게 측정하고 예측할 수 있을까? ㉠ 사용자가 입력하는 검색어 각각에 대하여 (가)의 A사의 방법을 사용하는 것도 생각할 수 있다. 그러나 하루에 수억 개의 검색어를 처리해야 하는 구글로서 이런 방법을 사용하는 것은 현실적으로 불가능하다.

당시 월드 와이드 웹(WWW)의 정보 축적량은 폭발적으로 증가하고 있었는데, 이 증가 속도가 효과적인 탐색 엔진의 개발에 대한 비관적 전망을 키우고 있었다. 그것은 웹이 정보들을 연결하고 있는 특정한 방식 때문이었는데, 웹에서는 하이퍼링크(hyperlinks)라는 방식으로 페이지 제작자의 결정에 따라 임의의 정보가 다른 임의의 정보와 연결될 수 있다. 하이퍼링크는 컴퓨터의 속도와 연산 능력을 이용하여 막대한 양의 정보를 연결할 수 있는 자연스러운 방법이였기 때문에 도입된 것이었다. 이로 인해 내용 관련성과 관계없이 모든 것이 다른 모든 것과 연결될 수 있게 되자 웹의 방대한 크기와 연결의 임의성이라는 이 두 요소는 사람들이 찾고자 하는 정보를 찾아내는 일을 난제로 만들었다.

구글의 창업자인 래리 페이지는 ‘페이지랭크(PageRank)’라는 알고리즘을 개발하여 웹 탐색의 새로운 영역을 열었다. 페이지랭크는 웹 페이지들의 연결 구조로부터 각 페이지의 중요도 값을 얻는 한 가지 방법을 준다. 웹 페이지의 중요성이란 본질적으로 주관적인 문제로서 독자들의 관심, 지식, 태도에 달려 있는 문제이지만, 그럼에도 여전히 웹 페이지의 상대적 중요성에 관하여 객관적으로 말할 수 있는 부분이 꽤 있다. 웹의 양적 팽창이 어느 단계를 넘어서면 일부 사용자들의 특이성에도 불구하고 사람들이 각각의 웹 페이지에 주는 관심과 집중의 정도, 즉 웹 페이지의 중요도를 객관적이고 기계적으로 측정할 수 있다.

페이지랭크는 모든 웹 페이지가 기본적으로 동등하다는 웹의 독특한 민주적 본성에 의 존하면서도 웹의 방대한 연결 구조 자체가 각 페이지의 가치를 나타내는 하나의 지표가 될 수 있다고 본다. 요컨대, 구글은 어떤 한 페이지 P에서 다른 페이지 Q로 가는 링크를 (몇 개의 링크를 걸었든) P가 Q에게 주는 한 표의 투표로 해석한다. 그러나 구글은 페이지가 받는 링크의 숫자만을 고려하는 것이 아니라, 투표하는 페이지도 분석한다. ‘중요도’가 높은 페이지가 주는 표에는 가중치를 부여하여 그 표를 받는 페이지의 중요도는 더 큰 폭으로 높아진다. <그림>에서, 페이지 C는 단 한 개의 페이지로부터 표를 받고 있을 뿐이고 페이지 E는 여섯 개의 페이지로부터 표를 받고 있다. 그러나 C에게 투표한, 그리고 C에게만 투표한 페이지 B가 중요도가 높기 때문에 C가 E보다 중요도가 더 높다.



<그림>

(다)

페이지랭크로 계산되는 링크의 수가 페이지의 중요도를 나타낸다면, 마찬가지로 인터넷의 언론 보도와 논평 기사들에 대한 댓글의 수는 그 기사의 중요도를, 또는 적어도 그 기사가 화제가 되는 정도를 나타낸다고 볼 수 있을 것이다. 최근 우리 사회에서는 기사에 대한 긍정적인 또는 부정적인 댓글의 수를 늘리려 하거나, 댓글에 대한 추천 수를 늘려 댓글의 노출 순서를 바꾸려 하는 등, 댓글 조작에 의해 여론의 향배를 오도하려 한 사람들의 존재가 확인되고 있고, 이러한 여론 ㉡ 왜곡을 우려하는 목소리가 점점 커지고 있다.

[문제 2] 다음 물음에 답하시오. (50점)

1. 이차함수  $y = x^2$ 의 그래프 위에  $x$ 좌표가 자연수이며 그 곱이 2000인 서로 다른 두 점이 있다.  
이 두 점과 원점  $(0, 0)$ 을 지나는 원의 중심  $(x_0, y_0)$ 에 대하여  $2x_0 + 2y_0$ 의 최솟값을 구하시오.
2. 이차함수  $y = x^2$ 의 그래프 위에 있고  $x$ 좌표가 1부터 10까지의 자연수인 10개의 점들 중 서로 다른 두 점을 임의로 선택할 때, 이 두 점과 원점  $(0, 0)$ 을 지나는 원의 중심의  $y$ 좌표를 확률변수  $Y$ 라 하자.  $Y$ 의 기댓값을 구하시오.
3. 이차함수  $y = x^2$ 의 그래프 위에  $x$ 좌표가 1부터 20까지의 자연수인 20개의 점들이 있다.  
이들 중 임의로 선택한 서로 다른 두 점과 원점  $(0, 0)$ 을 지나는 원의 중심  $(x_0, y_0)$ 에 대하여  $|x_0|$ 의 값이 120일 때,  $y_0$ 의 값을 확률변수  $Z$ 라 하자.  $Z$ 의 기댓값을 구하시오.

