

[1~2] 다음 글을 읽고 물음에 답하시오. 2017년 입법고시 PSAT 언어논리 가책형 19~20번

(가) 수십억 개의 웹 페이지가 존재하지만, 대부분의 페이지는 그다지 흥미롭지 않은 것들이다. 많은 페이지들이 유용한 정보를 전혀 담고 있지 않은 스팸 페이지이다. 다른 페이지들은 개인 블로그, 결혼 공지, 가족사진 앨범 등이다. 엄청난 웹 페이지의 수가 검색엔진이 유용한 페이지를 찾는 것을 어렵게 만든다.

검색엔진에 특정한 단어를 입력했을 때, 웹상에는 해당 단어를 포함한 수백만 개의 페이지가 존재한다. 검색엔진은 어떻게 가장 유명한 페이지를 선택할 수 있을까? 한 가지 효과적인 방법은 웹 페이지 간의 링크를 인기 척도로 사용하는 것이다. 가장 분명한 척도는 각 페이지에 대한 진입링크 개수를 세서 이것을 순위 부여 알고리즘의 자질로 사용하는 것이다. 이들 측정 방법들 중에, 페이지랭크가 가장 자주 언급된다.

페이지랭크는 랜덤 서퍼(random surfer) 아이디어에 기반하고 있다. 앨리스라는 사람이 지금 웹 브라우저를 사용하고 있다고 상상해보자. 앨리스는 매우 따분해서, 시간을 때우기 위해 웹 페이지를 이리저리 돌아다니고 있다. 그녀의 브라우저는 “surprise me”라는 특별한 버튼이 있는데, 이것을 클릭하면 임의의 웹 페이지로 이동한다. 웹 페이지가 로딩될 때마다, 그녀는 “surprise me”를 누를 것인지, 아니면 페이지에 있는 링크를 누를지를 결정한다. 그녀가 페이지 상의 링크를 누를 때도, 그녀는 특별히 그 링크에 관심이 있는 것은 아니고 단지 랜덤하게 하나를 선택한 것이다. 앨리스는 매우 따분해서 이러한 웹 브라우저를 영원히 계속한다. 이를 좀 더 정교한 형태로 표현하자면, 앨리스는 다음 알고리즘에 따라 웹을 돌아다닌다.

1. 0과 1사이의 임의의 숫자 r 를 고른다.
2. 만약 $r < K$ 이면 : “surprise me” 버튼을 누른다. (단, $0 \leq K \leq 1$)

3. 만약 $r \geq K$ 이면 : 현재 페이지에 있는 링크 중에 하나를 무작위로 누른다.

4. 처음부터 다시 시작한다.

보통 K 값은 충분히 작아서 앨리스가 “surprise me” 버튼보다는 링크를 더 많이 누른다고 가정한다. 비록 웹 페이지들을 랜덤하게 돌아다닐지라도 앨리스는 유명한 페이지를 그렇지 않은 페이지들보다 더 자주 보게 될 것이다. 왜냐하면, 앨리스는 주로 링크를 따라가는데 링크는 유명한 페이지를 가리키는 경향이 있기 때문이다.

(나) 웹은 A, B, C의 3개의 페이지로만 구성되어 있다. 페이지 A가 페이지 B와 C를 가리키고 있고, 페이지 B는 페이지 C를, 페이지 C는 페이지 A를 가리키고 있다고 하자. 주어진 페이지에 있는 링크는 무작위로 선택된다. 따라서 페이지 A에서 페이지 C로 갈 확률은 진출링크가 두 개이므로 50%이다. “surprise me” 버튼을 무시한다면, 페이지 C의 페이지랭크 $PR(C)$ 는 식 $PR(C) = \frac{PR(A)}{2} + \frac{PR(B)}{1}$ 로 계산된다.

여기에는 한 가지 명백한 문제가 있는데, . 따라서 모든 페이지의 페이지랭크 값이 동일하다고 가정하여 $PR(A) = PR(B) = PR(C) = 0.33$ 으로 놓고, 이를 근거로 반복 계산을 수행하여 값을 계산한다. 반복의 각 단계마다 이전 반복 때의 페이지랭크 값을 사용하여 새로운 페이지랭크 값을 계산한다. 예를 들어, 첫 번째 반복에서 $PR(C) = 0.33/2 + 0.33 = 0.5$ 이다.

※ 진입링크 : 특정한 페이지를 가리키는 링크

1. 윗글에 대한 추론으로 옳지 않은 것은?

- ① (1-K)는 한 단계에서 앨리스가 현재 페이지에 있는 링크 중에 하나를 무작위로 누를 확률을 나타낸다.
- ② 특정한 페이지의 페이지랭크 값은 검색엔진 사용자가 입력한 키워드에 따라 달라진다.
- ③ 많은 페이지로부터 링크를 받는 페이지는 그렇지 못한 페이지에 비해 상대적으로 더 높은 페이지랭크 값을 가질 것이다.
- ④ 웹 페이지 수가 더 이상 증가하지 않더라도 페이지랭크 값은 변화할 수 있다.
- ⑤ 막대한 웹 페이지의 수는 검색엔진을 통해 필요한 정보를 찾는 것을 어렵게 한다.

2. 윗글에 대한 추론으로 옳은 것을 <보기>에서 모두 고르면?

<보 기>

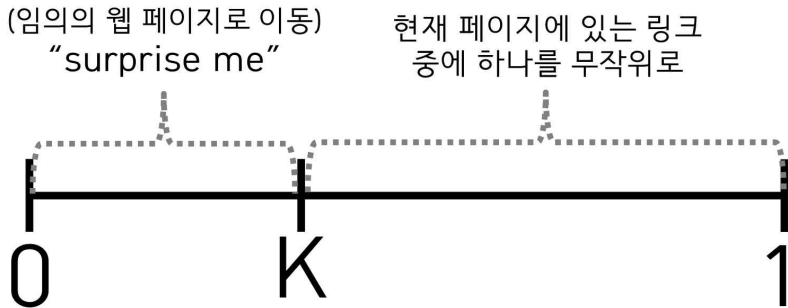
| |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>ㄱ. 특정한 페이지의 페이지랭크는 그 페이지를 링크하고 있는 페이지들의 페이지랭크 값에 의존적이다.</p> <p>ㄴ. 빈 칸에 “바로 우리가 각 페이지의 페이지랭크 값을 모른다는 것이다. 우리가 구하고자 하는 것이 바로 페이지랭크 값이기 때문이다”라는 말이 들어갈 수 있다.</p> <p>ㄷ. (나)의 첫 번째 반복에서 $PR(A) = 0.33$이다.</p> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- ① ㄱ ② ㄱ, ㄴ ③ ㄱ, ㄷ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

해설

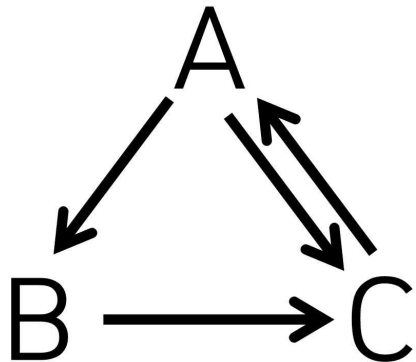
(가)를 이해하는 것이 어렵지는 않을 겁니다. **문제** 엄청난 웹 페이지의 수가 검색엔진이 유용한 페이지를 찾는 것을 어렵게 만든다. **해결** 페이지랭크 구조의 글이죠.

앨리스가 따르는 알고리즘은 결국 주사위를 던져서 K보다 작으면 아무데나 뽕 가고, K 이상이면 페이지 내 링크 중 랜덤하게 하나를 골라 뽕 간다는 거죠?



(나)가 처음에 좀 어려웠을 수 있습니다. 일단 빈칸 앞을 보면 $PR(C)$ 가 $PR(A)$, $PR(B)$ 에 의존하는 수식이 등장합니다. 그리고 뒤에는 빈칸의 문제를 해결하기 위해 "모든 페이지의 페이지랭크 값이 동일하다고 가정"했구요. 따라서 빈칸에는 $PR(A)$, $PR(B)$ 를 몰라서 $PR(C)$ 를 구할 수 없다는 식의 문제점이 들어가야 합니다. 모르기 때문에 일단 동일하다고 '가정'했겠죠!

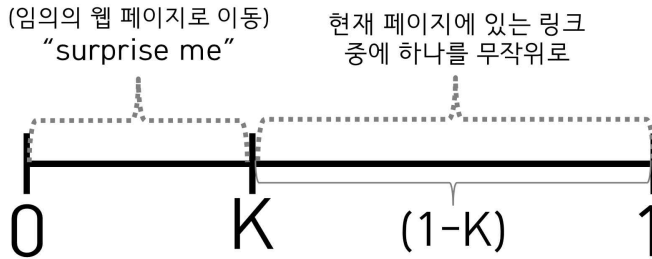
수식의 경우 초등학교 산수 수준이라 별로 어렵지 않습니다. A, B, C 사이의 링크 관계도 단순하고요. 이미지로 나타내면 다음과 같습니다.



문1 해설

① (1-K)는 한 단계에서 앨리스가 현재 페이지에 있는 링크 중에 하나를 무작위로 누를 확률을 나타낸다. O

: (가)에 소개된 알고리즘을 그대로 묘사했으니 적절합니다.



② 특정한 페이지의 페이지랭크 값은 검색엔진 사용자가 입력한 키워드에 따라 달라진다. X

: 지문에 따르면, 페이지랭크 값은 “웹 페이지 간의 링크”에 의해 결정됩니다. (나)에 구체적인 식도 나왔지만 ‘키워드’가 변수로 들어가지 않죠! 정답~

③ 많은 페이지로부터 링크를 받는 페이지는 그렇지 못한 페이지에 비해 상대적으로 더 높은 페이지랭크 값을 가질 것이다. O

: 페이지랭크는 “각 페이지에 대한 진입링크 개수를 세서 이것을 순위 부여 알고리즘의 자료로 사용”하는 것이므로, 선지가 적절하다는 느낌이 확 오죠? (물론 시험장에서는 전혀 그럴 필요 없지만, (나)에 제시된 식을 이용해서 계산해봐도 알 수 있습니다.)

④ 웹 페이지 수가 더 이상 증가하지 않더라도 페이지랭크 값은 변화할 수 있다. O

: 지문에 따르면, 페이지랭크 값은 “웹 페이지 간의 링크”에 의해 결정됩니다. 따라서 페이지 수가 증가하지 않아도, 링크가 바뀌면 페이지랭크 값도 바뀔 수 있습니다.

⑤ 막대한 웹 페이지의 수는 검색엔진을 통해 필요한 정보를 찾는 것을 어렵게 한다. O

: 1문단 마지막 줄을 그대로 옮겼네요. ㅎ (㉔는 문제-해결 구조의 글에서 문제를 선지화했고, 나머지 선지는 전부 해결을 선지화했다는 건 제가 말 안 해도 느꼈겠죠?)

문2 해설

ㄱ. 특정한 페이지의 페이지랭크는 그 페이지를 링크하고 있는 페이지들의 페이지랭크 값에 의존적이다. ○

: $PR(C) = \frac{PR(A)}{2} + \frac{PR(B)}{1}$ 를 보면, 특정한 페이지 C의 페이지랭크는 그 페이지를 링크하고 있는 페이지 A, B의 페이지랭크 값에 의존하고 있다는 게 바로 보이네요.

ㄴ. 빈 칸에 “바로 우리가 각 페이지의 페이지랭크 값을 모른다는 것이다. 우리가 구하고자 하는 것이 바로 페이지랭크 값이기 때문이다”라는 말이 들어갈 수 있다. ○

: 글을 읽으면서 추론해봤던 내용과 맥락이 같네요! (이 정도 빈칸 문제는 지문을 읽으면서 정답에 관한 이미지를 잡은 후, 비교해보는 정도로 풀 수 있어야 합니다.)

ㄷ. (나)의 첫 번째 반복에서 $PR(A) = 0.33$ 이다. ○

: (나)에 제시된 수식을 이해했는지 묻는 문제입니다. 지문의 설명을 그대로 적용하면 됩니다.

[페이지 C에서 페이지 A로 갈 확률은 진출링크가 한 개이므로 100%이다.

“surprise me” 버튼을 무시한다면, 페이지 C의 페이지랭크 $PR(C)$ 는 식

$PR(A) = \frac{PR(C)}{1} + 0$ 로 계산된다. (페이지 B에서 페이지 A로 가는 링크는 없으므로 0으로 나타냄)]

낯선 소재지만 그렇게 어렵지는 않았죠? 이 지문이 변형되어 출제된다면 두 가지 지점을 물어볼 수 있습니다.

생각1. 램덤 서퍼 기능은 왜 필요할까요?

생각2. 반복 계산을 수행하면 PR(A), PR(B), PR(C)는 어떤 값에 수렴할까요?

덧 : 고1, 2라면 시간 있을 때 『미래를 바꾼 아홉 가지 알고리즘』(에이콘출판사) 등 양질의 도서를 읽어보기 바랍니다. 배경지식이 없어도 문제를 풀 수 있지만, 있어서 나쁠 건 없습니다. 2017학년도에서 수능 만점을 받은 이영래 씨가 “사실 문제보다는 한국교육과정평가원 기출 문제를 중심으로 공부했고, 독서를 많이 한 것이 수능 만점에 도움이 된 것 같다”에서 보듯, 독서는 교양과 지적만족을 넘어 시험에도 도움이 될 수 있습니다.

『미래를 바꾼 아홉 가지 알고리즘』과 관련된 수능/모의평가 기출문제

「3장 페이지랭크」

X 페이지랭크(2017년 입법고시 PSAT)

「4장. 공개키 암호화」

X 공개키 암호화 방식(2005학년도 9월 모의평가)

「5장 오류 정정 코드」

X 해시 함수(2016학년도 9월 모의평가)

「6장 패턴 인식과 인공지능」

X 우편번호 자동분류기(2010학년도 9월 모의평가)

X 인공신경망 퍼셉트론(2017학년도 6월 모의평가)

「7장 데이터 압축」

X 동영상 압축 기술(2009학년도 수능)