

<문제 1>

Page1:

<https://www.goodreads.com/search?page=1&qid=cZTYj1phOX&query=harry+potter&tab=books&utf8=%E2%9C%93>

Page2:

<https://www.goodreads.com/search?page=2&qid=cZTYj1phOX&query=harry+potter&tab=books&utf8=%E2%9C%93>

Page3:

<https://www.goodreads.com/search?page=3&qid=cZTYj1phOX&query=harry+potter&tab=books&utf8=%E2%9C%93>

⇒ 페이지를 이동할 경우 위 url 에서 page=숫자 부분이 페이지에 맞게 변화한다.

<문제 2>

수집된 URL

[1]

/book/show/3.Harry_Potter_and_the_Sorcerer_s_Stone?from_search=true&from_srp=true&qid=POWOMrNsaG&rank=1

클릭하여 이동한 URL

[1]

https://www.goodreads.com/book/show/3.Harry_Potter_and_the_Sorcerer_s_Stone?from_search=true&from_srp=true&qid=POWOMrNsaG&rank=1

⇒ 수집된 URL에서는 클릭하여 이동한 URL 과 다르게 앞부분에 <https://www.goodreads.com> 가 생략되어있다.

<문제 3>

Harry potter

<https://www.goodreads.com/search?utf8=%E2%9C%93&query=harry+potter>

The chronicles of Narnia

<https://www.goodreads.com/search?utf8=%E2%9C%93&query=the+chronicles+of+narnia>

The lord of the rings

<https://www.goodreads.com/search?utf8=%E2%9C%93&query=the+lord+of+the+rings>

⇒ “https://www.goodreads.com/search?utf8=%E2%9C%93&query=” 뒤에 키워드가
적혀지는데, 키워드 안에 있는 띄어쓰기는 +로 표시된다.