

Ergodic Markov Model을 이용한 정보 검색 모델

강인호⁰ 이어진 한영석^{*} 김길창

한국과학기술원 전산학과
{ihkang, yjlee, gckim}@csone.kaist.ac.kr
수원대학교 전산학과^{*}
yshan@mail.suwon.ac.kr

An Information Retrieval Model based on an Ergodic Markov Model

In-Ho Kang⁰ Yeo-Jin Lee Young S. Han^{*} Gil Chang Kim

Dept. of Computer Science, KAIST
Dept. of Computer Science, Suwon Univ.^{*}

요 약

인터넷의 급속한 양적 증가로 인해 색인어 기반의 검색 방식만으로는 원하는 정보를 찾아 내기가 쉽지 않다. 색인어 기반의 검색 방식에서는 색인어로 나타나지 않는 특징을 이용할 수 없으며, 질적으로 균등한 검색 결과를 제시하지 못하기 때문이다. 따라서 사이트의 여러 가지 특성에 따라 계층적으로 분류해놓은 웹 디렉토리를 이용하거나, 관련 전문가들의 추천 리스트를 이용하여 검색하기도 한다. 본 연구에서는 기존의 색인어 기반의 검색 모델에 웹 디렉토리나 추천 문서 같은 문서 간의 링크 정보를 결합할 수 있는 정보 검색 모델을 제시한다. 특정 질의어의 검색 결과로 얻어낸 문서와 그 문서와 연결된 문서 집합을 이용하여 네트워크를 구성한다. 이 네트워크에 검색기가 제시하는 순위와 유사도, 그리고 문서 간의 링크 정도를 이용해서 확률값을 정해준다. 그리고 Ergodic Markov Model의 특성을 이용하여 색인어 정보와 링크 정보를 결합한다. 본 연구에서는 특정 문서가 질의어에 부합되는 정도를 사용자가 그 문서로 이동할 확률값으로 계산하는 방식을 보인다.

1. 서론

웹을 통한 정보 검색은 사용자들에게 가장 친숙한 정보 검색 방법이 되었다. 사용자는 원하는 정보를 얻기 위하여 질의어와 유사한 문서들을 제공하는 정보 검색 시스템을 사용한다. 현재 대부분의 정보 검색 시스템은 색인어 기반의 방식을 사용하고 있다. 그러나 색인어 기반의 정보 검색 시스템은 제공하는 정보의 양이 증가함에 따라 사용자의 정보 요구 만족도를 충족 시켜주지 못하고 있다.

구글¹ 검색 사이트를 보자. 구글 사이트의 메인 페이지에는 문헌 검색이나 정보 검색 전문 사이트임을 알 수 있는 용어가 희박하다. 즉 색인어 기반

의 검색 방식만으로는 구글을 검색 사이트로 추천할 수 없다. 사이트의 내용이나 정보와 달리 회사의 이미지를 고려하여 페이지를 구성할 경우, 글로 나타나지 않은 특성은 색인어 기반 방식으로는 찾아내기 어렵다. 게다가 대부분의 사용자는 정보 검색에 충분하지 못한 한 단어에서 세 단어 정도 길이의 질의어를 이용한다[5]. 따라서 근거로 사용할 자료조차 부족하다.

우리는 다른 많은 웹페이지에서 구글을 검색 사이트로 추천하는 것을 볼 수 있다. 웹 상의 추천, 분류 리스트 같은 하이퍼 링크는 검색에 중요한 정보로 사용할 수 있다. Yahoo² 나 Infoseek³에서는 주제별, 용도별에 따라 그 분야의 전문가가 구분하고 관리하는 대규모의 웹디렉토리를 제공하기

² www.yahoo.com

³ www.infoseek.com

¹ www.google.com

도 한다. 여기서 링크는 두 문서가 내용적으로 관련이 있다는 링크 생성자의 주관적인 판단에 의해서 생성되므로, 링크로 연결된 문서도 추가 정보로서의 가치가 있다. 많은 사람들이 추천하는 문서가 좋은 시작점이 될 수 있으며 어느 정도의 질을 보장할 수 있다.

본 연구에서는 링크 정보를 이용하는 정보 검색 모델을 제시한다. 기존의 색인어 기반의 정보와 링크 정보를 Ergodic Markov Model을 이용하여 결합하는 방법을 보인다.

2. 관련연구

정보 검색 시스템에서 하이퍼링크에 대한 가정은 일반적으로 다음과 같다.

- 하이퍼링크는 관련 있는 문서들에 대한 연결이어야 한다.
- 링크를 통해 도달한 문서가 다른 저자에 의해 지어진 것이라면 그 문서는 추가 정보로서의 가치가 있다.

즉, 하이퍼링크로 연결되어 있다면 두 문서는 서로 어느 정도의 연관성을 가지고 있다고 생각할 수 있다[5]. 이 두 가지 가정으로부터 링크 된 문서 역시 추출할 필요가 있다는 결론이 나온다. 링크 정보를 이용하는 방법으로 크게 세 가지를 들 수 있다. Hub와 Authority를 결합하는 방법[7]과 PageRank[4]를 이용하는 방법 그리고 Belief Network[9]을 이용한 링크 정보의 사용을 들 수 있다.

2.1 Hub와 Authority

검색기를 이용하여 사용자 질의어에 대한 검색 결과 S 를 구한다(그림 1). 그리고 S 가 링크하는 문서, 그리고 S 를 링크하는 문서를 이용하여 확장된 문서 집합 T 를 구한다.

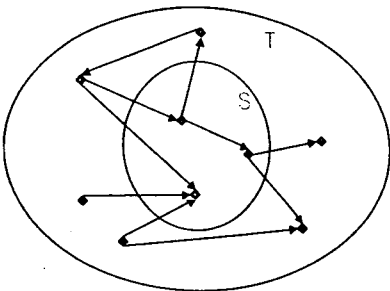


그림 1 검색 결과의 확장

이 확장된 결과 집합 T 에서 링크를 기반으로 Hub와 Authority를 계산한다. 좋은 Hub 문서는 높은 Authority를 가지는 문서를 많이 가리키고 있는 것이며, 높은 Authority를 가지는 문서는 좋은 Hub 문서로부터 많은 가리킴을 당하는 문서를 말한다[7].

$$hub(p) = \sum_{\forall q, p \text{ points } q} authority(q) \quad (1)$$

$$authority(p) = \sum_{\forall q, q \text{ points } p} hub(q) \quad (2)$$

수식 (1),(2)를 이용하여 Hub와 Authority 값을 계산하고, 그 합한 값으로 문서를 순위화 한다. 특정 사이트의 Hub와 Authority는 질의어에 따라 추출 문서 집합이 달라지기 때문에 그 값이 고정되지 않고 달라지는 형태이다. 따라서 단순한 문서의 인기도를 나타내는 것이 아니라, 질의어에 어느 정도 관련된 인기도를 얻을 수 있다. 그러나 색인 작업시 미리 계산할 수 없다는 단점과 문서가 질의어와 얼마나 일치하는지를 나타내는 것이 어렵다는 단점을 가진다.

2.2 PageRank

링크를 통해서 문서의 순위가 전파되어 나가는 개념이다[4]. T_1, \dots, T_n 이 페이지 A 를 가리키고 있을 때, 페이지 A 는 수식 (3)과 같은 페이지 순위(PR)를 가진다.

$$PR(A) = (1-d) + d(PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n)) \quad (3)$$

수식 (3)에서 d 는 사용자가 특정페이지에서 만족하지 못하고 다른 페이지로 이동할 확률을 얘기하며 $C(T_i)$ 는 T_i 가 가리키는 문서 개수를 나타낸다. 구글에서는 사용자가 질의어로 입력하는 단어를 모두 가지는 문서에 대해서 미리 계산해 놓은 PageRank를 가지고 순위를 결정한다. 즉 색인어 정보에서는 동일하다고 생각되는 문서에 대해서 링크 정보를 이용해서 우선 순위를 결정하는 방식이다. 따라서 색인 작업시 PageRank를 미리 계산할 수 있다는 장점을 가진다. 그러나 링크의 종류나 특성에 따라서 차별하기가 쉽지 않다는 단점을 가진다.

2.3 Belief Network

색인어와 문서를 노드로 표현하고 영향을 주는 여부에 따라 링크를 주어 Belief Network을 구성한다. [2]에서는 색인어 층, 문서 층, 그리고 질의어

층으로 구성하여 각 색인어가 문서에 미치는 정도를 계산할 수 있게 했다. [9]에서는 기존의 색인어 정보 외에 다른 종류의 정보를 추가하기 위하여 한 층을 그림 2와 같이 추가하였다. 이 층은 링크를 이용한 Hub와 Authority 그리고 색인어 정보를 문서와 연결시킬 수 있게 한다. 그림 2에서 K 는 색인어 층을 나타내며, C 는 색인어 정보, A 는 Authority, H 는 Hub를 나타낸다. [9]에서는 미리 Belief Network을 만드는 것이 아니라 검색기가 제시하는 결과 중 일부만을 이용해서 질의마다 동적으로 만든다.

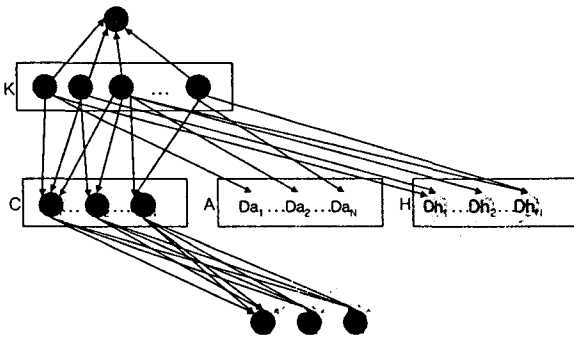


그림 2 Belief Network을 이용한 정보의 결합

이와 같이 링크 정보를 이용하는 정보 검색 방식은 기존의 검색 방식으로 검색 결과를 얻어낸 뒤, 링크를 이용하여 결과 집합을 확장한다. 그리고 링크 정보를 이용하여 결과 집합을 재순위하는 형태를 보인다. 본 연구에서는 기존의 검색 모델이 사용하는 정보를 기준으로 검색 모델을 설명하는 것과 달리 사용자가 링크 정보를 이용하여 상태를 이동해나가는 것으로 표현한다. 기존의 색인어 기반의 검색기가 제시하는 정보와 링크를 이용하여 이동하는 것을 모델링하여 두 정보를 결합할 수 있는 모델을 제시한다.

3. Ergodic Markov Model을 이용한 정보 검색

본 절에서는 색인어 정보와 링크 정보를 합치기 위해 사용하는 Ergodic Markov Model을 설명하고 이를 정보 검색에 이용하는 방법을 보인다.

3.1 Ergodic Markov Model

특정 시간 n 에 상태 i 에 있는 것을 $X_n = i$ 로 나타낼 경우, 상태 i 에서 j 로 한 단계만에 이동할 확률은 수식(4)와 같이 나타낼 수 있다.

$$P_{ij} = P[X_{n+1} = j | X_n = i] \quad (4)$$

그리고 k 단계만에 i 에서 j 로 이동할 확률은 수식(5)와 같이 나타낼 수 있다.

$$P_{ij}^k = P[X_{n+k} = j | X_n = i] \quad (5)$$

Ergodic Markov Model은 확률분포이며, 상태(State)와 상태간의 이동(Transition) 그리고 이동이 일어날 확률(Transition Probability)로 표현한다. 확률분포이기 위해서는 수식(6)을 만족해야 한다.

$$\sum_{ij} P_{ij} = 1 \quad (6)$$

Ergodic Markov Model이기 위해서는 Markov 특성을 가지는 확률 분포 외에 다음과 같은 조건이 더해진다. 모든 상태가 다른 상태들로부터 찾아올 수 있는 경로가 존재해야 한다. 그리고 비주기적(aperiodic)이면서 positive recurrent 해야 한다 [3].

자신의 상태로 k 단계 만에 되돌아오는 경로가 있을 때 $k+1, k+2, \dots, \infty$ 단계의 이동으로 되돌아올 수 있는 경로 또한 존재할 경우 이를 비주기적인 상태(aperiodic state)라고 얘기한다. 이때 상태에 다시 되돌아오는데까지 예상 이동 횟수가 유한한 값을 가질 때 positive recurrent라고 한다.

Ergodic Markov Model은 다음과 같은 특성을 가진다.

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j \quad (7)$$

$$\sum_{j=0}^{\infty} \pi_j = 1 \quad (8)$$

$$\pi_j = \sum_{i=0}^{\infty} \pi_i \cdot P_{ij} \quad (9)$$

즉 시간을 무한대로 고려할 경우 상태 i 에서 상태 j 로 이동할 확률(stationary probability)은 π_j 로 수렴하며, 이 값은 시작하는 상태에 상관없이 동일한 값을 가진다[3]. 예를 들어 그림 3과 같은 Ergodic Markov Model이 있을 경우, 상태 0, 1, 2의 stationary probability는 각각 0.2, 0.6, 0.2의 값을 가진다. 즉 상태 0, 1, 2로 이동할 확률값은 각각 0.2, 0.6, 0.2가 된다. Stationary Probability는 수식(9)를 이용하여 행렬식으로 표현하여 계산할 수 있다.

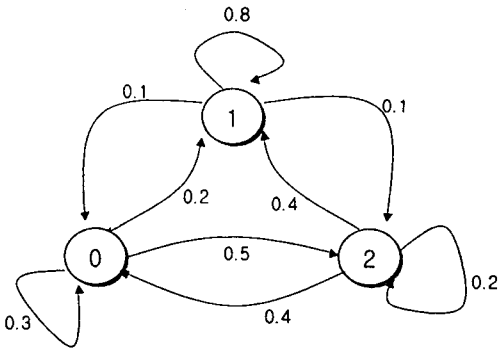


그림 3 Ergodic Markov Model의 예

$$[0.2 \ 0.6 \ 0.2] \times \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.8 & 0.1 \\ 0.4 & 0.4 & 0.2 \end{bmatrix} = [0.2 \ 0.6 \ 0.2]$$

그림 4 Stationary Probability 계산 예

3.2 정보 검색의 Ergodic Markov화

사용자가 정보 검색을 위해서 검색기를 실행하고 문서들을 이동하며 정보를 찾아가는 과정을 Ergodic Markov Model로 표현한다. 사용자는 검색기를 이용하여 검색된 문서와 그 문서의 링크를 이용해서 확장된 결과 집합에서 정보를 찾는다.

문서를 상태로 간주하고, 두 문서간에 링크가 존재하는 경우 상태간의 이동이 가능하다고 표현하면 그래프를 얻을 수 있다. 이 그래프에 정보 검색기를 이용하여 얻어낸 검색 결과 리스트를 가지는 페이지를 상태로 추가할 경우, 모든 상태가 연결된 그래프를 얻을 수 있다. 즉 사용자는 결과 리스트 페이지에서 특정 문서를 하나 선택해서 이동을 하고 그 문서에서 정보를 얻거나 문서가 가리키는 다른 문서로 이동 혹은 결과 리스트 페이지로 되돌아온다는 가정을 한다.

이 그래프에 각각의 상태에 머물 확률과 다른 상태로 이동할 확률 값을 어떻게 결정하는가에 따라서 다양한 검색 모델을 얻을 수 있다. 문서 간의 유사도를 이용해서 이동 확률을 구할 수도 있을 것이며, 실제 웹 이동 기록을 분석해서 얻을 수도 있다. 본 연구에서는 우선 링크에 동일한 비중을 두고 이동한다고 가정하고 모델을 제시한다.

본 연구에서는 검색 결과로 나온 결과 리스트 페이지에서 해당 문서로 이동할 확률을 검색 순위 (r)에 비례하여 할당한다. 사용자들이 검색 순위가 높은 문서를 선호하고 우선적으로 채택한다는 가정을 가진다. 그리고 그 문서로 이동해서 정보를 찾을 수도 있고, 링크를 이용하여 다른 페이지로 이동을 하거나 원래의 검색 결과 페이지로 이동할 수 있다. 그 문서에서 만족할 확률은 검색기가 제시하는 유사도($sim.$)를 이용한다. 그리고 나머지 다른 페이지로 이동할 확률값은 동일하게 분배한다.

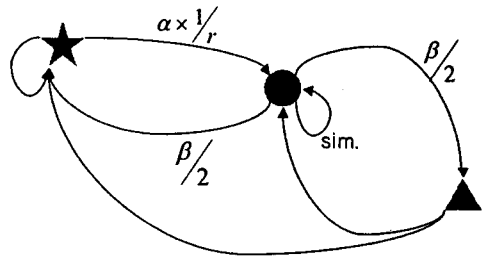


그림 5 상태 이동 확률 결정의 예

그림 5에서 별표는 결과 리스트 페이지를 나타내며 원은 검색 페이지 그리고 삼각형은 링크된 페이지를 나타낸다. 여기서 $sim. + \beta = 1$ 이 된다. α 는 수식 (6)을 만족시키기 위한 상수이다.

이와 같은 방식으로 정보 검색을 Ergodic Markov 모델로 표현할 수 있다. 여기서 구해지는 stationary probability는 그 문서로 이동할 확률값 즉 그 문서를 선택할 확률값이 된다. 이 확률값을 기준으로 정렬하여 순위를 결정한다.

4. 실험 및 평가

4.1 실험

제안한 모델을 평가하기 위해 한국어 정보 검색 실험 집합인 KEMONG-SET을 이용한다. KEMONG-SET 실험 집합은 (주) 계몽사에서 출판한 학생 대백과 사전으로, 각 표제어와 그에 대한 설명을 하나의 문서 단위로 하고 있다. 총 23,113개로 구성되어 있는데, 각 문서에 포함된 용어의 수는 20개에서 1,000개 정도이고, 평균 문서의 크기는 56어절을 포함하고 있다. 질의 개수는 45개이고, 평균 질의에 포함된 단어는 6어절이다. 전체 문서에서 각 질의 당 적합한 평균 문서 개수는 9개이다.

KEMONG-SET은 문서들 사이의 참조 정보를 나타내기 위해 <see>와 <seealso> 태그를 사용한다.

<see> 태그는 다른 문서에 관련어에 대한 설명이 있기 때문에, 그 문서에서 단어에 대한 설명을 찾아보라는 것이고 <seealso> 태그는 설명하는 단어와 밀접한 관련이 있는 문서를 가리킨다. 예를 들면, 그림 6에서 1.4 후퇴에 대한 설명은 6.25전쟁을 설명한 문서를 보면 알 수 있는 것이다. 그리고 '2극 진공관'에 대해서 좀 더 알고 싶은 경우 '진공관'에 대한 설명도 살펴보라는 것이다. 본 연구에서는 <see>와 <seealso> 정보를 이용하여 문서들을 연결시킨다.

```
<id> 00004
<title> 1·4 후퇴 -四後退
<see> 6·25 전쟁
```

```
<id> 00013
<title> 2극 진공관 二極真空管
<seealso> 진공관
<contents>
음극과 양극의 두 극으로 이루어진 진공관.
열전자를 방출하는 음극(필라멘트)과 열전자를 ...
```

그림 6 see와 seealso 예

그러나 <see>와 <seealso> 정보가 약 4,000여개로 전체 문서수에 비해 너무 부족하기 때문에 인위적으로 링크를 부여했다. 각 문서의 첫번째 문장을 타이틀에 대한 직접적인 정의문이라고 간주하고 정의문이 관련있는 문서를 링크로 연결하였다. 예를 들면, 그림 6에서는 “음극과 양극의 두 극으로 이루어진 진공관.”을 ‘2극 진공관’의 정의문으로 간주한다. 관련이 있는 두 문서는 다음과 같은 방식으로 얻어낸다. 먼저 동일한 정의문을 가지는 경우, 두 문서에 링크를 부여하였다. 그리고 정의문을 이용해서 상위어를 추출하고 상위어를 설명하는 문서와 그 상위어를 타이틀로 가지는 문서에 링크를 부여하였다. 상위어를 추출하기 위해서 사용하는 패턴은 다음과 같다.

- ()의 한 가지.
- ()의 하나.
- ()에 딸린 ().

각 문서의 정의문에 위와 같은 패턴이 나타날 경우 괄호에 해당하는 부분을 타이틀로 가지는 문서와 연결하였다. 예를 들어 ‘피아노 3중주’, ‘피아노 4중주’, ‘피아노 5중주’는 “실내악 연주 형태의 한 가지”라는 동일한 정의문을 가졌다. 따라서 ‘피아노 3중주’, ‘피아노 4중주’ 그리고 ‘피아노 5중주’는 서

로 링크를 가진다. 그리고 ‘사채권’의 정의문은 “유가 증권의 한 가지”로 ‘유가 증권’을 상위어로 얻을 수 있다. 따라서 ‘유가 증권’을 타이틀로 가지는 문서와 ‘사채권’을 타이틀로 가지는 문서는 서로 링크를 가진다. 이와 같이 정의문을 이용하여 12,000여개의 링크를 만들 수 있었다. 위와 같은 see, seealso 정보와 정의문을 이용한 정보를 이용하여서 문서 간의 이동 확률 즉 상태 이동 확률값을 얻었다.

색인어 정보로는 최대 엔트로피 품사 태거[6]를 이용하여 문서를 품사 태깅한 뒤 명사만을 그 대상으로 벡터 방식[1][8]으로 순위를 계산하였다.

4.2 실험 결과 분석

색인어 정보만을 이용하여 얻은 문서 결과에서 상위 150개의 문서를 가지고 확장한 문서집합을 대상으로 링크 정보를 이용하여 확률값을 결정했다. 그 실험 결과는 표 1과 같다.

표 1 실험 결과

Recall	Precision		
	Content ⁴	C+ See ⁵	C+ S+ Def ⁶
0.0	0.6319	0.6724	0.6740
0.1	0.6301	0.6724	0.6740
0.2	0.5956	0.6366	0.6466
0.3	0.5331	0.5749	0.5753
0.4	0.4724	0.5121	0.5040
0.5	0.4430	0.4823	0.4963
0.6	0.3806	0.3931	0.4116
0.7	0.2405	0.2757	0.3027
0.8	0.1970	0.2290	0.2634
0.9	0.1682	0.1991	0.2322
1.0	0.1482	0.1787	0.2151
Average	0.4037	0.4385	0.4541

본 연구에서 사용한 색인어 정보만을 이용하는 방식으로는 6개의 정답 문서를 찾지 못했다. 그리고 실험 결과 문서의 링크 정보를 이용하여 정답 문서가 추가 되는 경우는 없었다. 따라서 상위 150개의 문서를 대상으로 했을 경우, 제시하지 못한 정답 문서는 링크 방식에서 각각 19개, 18개로 늘어났다. 그럼에도 불구하고 성능은 향상됨을 알 수 있다. 이는 색인어 정보로 순위가 높은 문서가 링

⁴ 색인어 정보만을 가지고 벡터방식으로 얻은 결과
⁵ 색인어 정보와 see, seealso 태그를 이용한 결과
⁶ 색인어 정보와 see, seealso 태그 그리고 정의문을 이용한 결과

크를 이용하여 관련 문서에 영향을 주어 순위를 변동시켰기 때문이다.

예를 들어 22번 질문인 “작용 반작용에 대해 그림으로 설명해주세요”에 대해서 색인어 정보만을 이용할 경우 타이틀이 ‘반작용’인 문서를 1순위로 정답을 제시하였다. 이는 타이틀인 ‘반작용’이 희귀한 단어이면서 문서의 유일한 단어이기 때문이다 (그림 7). 그리고 작용·반작용을 설명하는 문서인 ‘작용·반작용’과 ‘작용·반작용 법칙’ 두 문서는 각각 2순위, 3순위로 제시하였다. 반면 내용면에서 작용 반작용 법칙과 밀접한 관련이 있는 ‘운동의 법칙’ 문서는 다른 색인어도 많아서 22순위로 제시되었다. 그러나 이 문서들이 서로 링크를 가짐으로써 22순위였던 ‘운동의 법칙’이 1순위로 올라가고 2순위가 ‘반작용’, 3순위가 ‘작용·반작용’, 그리고 4순위가 ‘작용·반작용의 법칙’이 되었다.

```
<id> 08091
<title> 반작용 反作用
<see> 작용·반작용
```

그림 7 ‘반작용’을 설명하는 문서

그림 8은 벡터 방식의 검색기에서 상위 n개의 문서를 채택했을 때의 11-point average를 보인다.

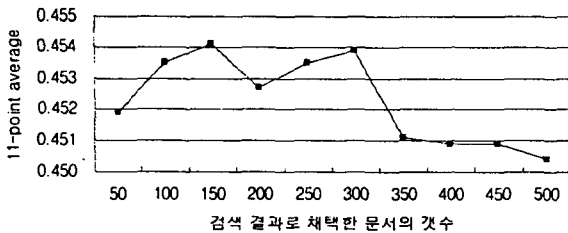


그림 8 기본 문서 집합 크기와 정확률의 관계

상위 350개 이상을 사용할 경우 질의와 관련없이 링크를 많이 받는 문서가 우선시 되어 성능이 낮아지는 것을 볼 수 있다.

5. 결론 및 향후 연구

본 연구에서는 색인어 기반의 정보 검색 방식에 링크 정보를 결합하여 정보 검색 과정을 표현할 수 있는 모델을 제시하였다. 다른 문서에서의 링크를 이용하여 현재 문서에 색인어로 나타나지 않은 정보를 얻을 수 있을 뿐만 아니라 문서의 질적 보장도 얻을 수 있다. 이를 위해서 웹 문서를 상태로 간주하고 문서간의 링크를 이용하여 문서를 이동하는 것을 상태간의 이동으로 표현하여 Ergodic

Markov로 표현하고 이에 따른 특정 상태에 이동할 확률값을 계산할 수 있는 정보 검색 모델을 제시하였다.

본 연구에서는 상태간의 이동 확률을 상호 참조(see, seealso)와 정의문의 유사성을 이용하여 이동 확률을 계산하는 방식을 보였다. 이 외에도 카테고리 정보, 실제 사용자들의 이동 통계 정보 등도 이용할 수 있을 것으로 보인다.

6. 참고 문헌

- [1] G. Salton and M. McGill. "Introduction to Modern Information Retrieval", McGraw-Hill, New York, NY, 1983
- [2] H. Turtle and W. Croft. "Evaluation of an inference network-based retrieval model." ACM Transactions on Information Systems, 9(3): 187-222, 1991
- [3] Sheldon M. Ross "An Introduction to Probability Models", Academic Press, Incorporated, 1997
- [4] S. Brin and L. Page. "The anatomy of a large-scale hypertextual web search engine", In Proc. Of the 7th International World Wide Web Conference(WWW7), pages 107-117, Brisbane, Australia, 1998
- [5] Krishna Bharat and Monika R. Henzinger, "Improved Algorithms for Topic Distillation in a Hyperlinked Environment", 21st ACM SIGIR, 1998
- [6] 강인호, 김재훈, 김길창 "최대엔트로피 모델을 이용한 한국어 품사 태깅" 제 10회 한글 및 한국어 정보처리 학술 대회, 1998
- [7] Kleinberg, J. "Authoritative sources in a hyperlinked environment." Proc. Of 9th ACM-SIAM Symposium on Discrete Algorithms, 1998
- [8] R. Baeza-Yates and B. Ribeiro-Neto. "Modern Information Retrieval", Addison Wesley, Essex, England, 1999.
- [9] Ilmerio Silva, et al. "Link-Based and Content-Based Evidential Information in a Belief Network Model", SIGIR 2000