이용자 중심 요약문 생성에 관한 실험적 연구

An Experimental Study on Generation of User-focused Summaries

김정하, 정영미 (연세대학교 대학원 문헌정보학과)

Jung-Ha Kim, Young-Mee Chung
Dept. of Library and Information Science, Graduate School of Yonsei University

본 연구에서는 단락검색 기법을 응용하여 이용자의 질의에 적합한 최적의 요약문을 자동 생성하는 방안을 모색하고자 하였다. 이를 위해 먼저 실험문헌집단을 구축한 후, 실험을 통해 이용자 중심 요약문을 생성하는 정적 단락검색 기법과 동적 단락추출 기법의 최적의 모형을 찾고 이들의 성능을 비교하였다.

1 서론

자동요약이란 본래 문헌이 지닌 의미는 그대로 둔 채 중복되거나 가치없는 정보는 제거함으로서 정보의 복잡도를 줄이는 작업이다. 검색 가능한 정보가 증가하여 이용자 개개인이 처리할 수 있는 용량의 범위를 벗어나기 시작하자 '요약시스템'과 같은 도구의 필요성이 절실해졌다.

본 연구에서는 단락검색 기법을 응용하여 이용자의 질의에 적합한 최적의 요약문을 자동 생성하는 두 가지 기법을 비교하였다. 첫째는 정적 단락검색' 기법으로 이는 문헌을 의미단위 단락으로 미리 구분해 두었다가 질의가 입력되면 질의와 가장 유사한 단락을 요약문으로 제시하는 방법이다. 두 번째 기법은 본 연구에서 새롭게 제안하는 요약문 생성 방법으로 확장활성화 알고리즘을 문장단위에 적용한 것이다. 새로운 기법은 '동적 단락추출'이라 명하였다. 정적 단락검색에서는 단락을 미리 구분하여 두고 검색하는 데 반해 동적 단락추출에서는 우선 주요한 문장을 검색하고 이 검색된 문장을 중심으로 확장하여 단락을 추출한다.

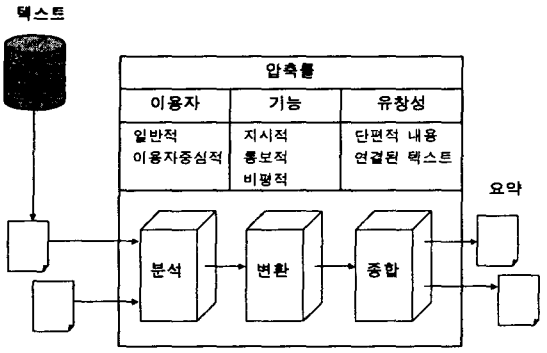
사전실험을 통해 두 기법을 최적화한 후, 각 기법이 제시하는 요약문의 성능을 내재적 평가와 외재적 평가를 통해 알아보았다. 내재적 평가에서는 요약문 자체의 품질을 측정하고, 외재적 평가에서는 요약문이 문헌의 적합성 판정에 얼마나 유용한지 측정하였다.

2 이론적 배경

2.1 자동요약

그림 1은 텍스트 자동요약시스템의 구조를 도식화한 것이다. 자동요약의 과정은 크게 세 단계로 나뉘어진다(Mani and Maybury 1999). 분석단계에서 입력 텍스트를 분석한 후, 변환단계에서 분석 결과를 바탕으로 요약에 포함될 텍스트의 부분을 선정하며, 종합단계에서 선정된 것을 종합하여 출력물로 제시한다.

생성되는 요약문의 특성은 그림 1과 같이 압축률, 이용자, 기능, 유창성의 측면으로 나누어볼 수 있다. 특히, 이용자 중심 요약은 90년대 중반 이후 정보검색의 맥락에서 많은 연구자들이 관심을 가져왔다(Sanderson 1998). 이



<그림 1> 텍스트 요약시스템의 구조

용자 중심 요약에서는 이용자의 질의와 요약 대상물의 단락수준, 문장수준의 정보를 이용한다. 실험을 통해 생성할 요약문은 '고정 길이', '이용자 중심적', '지시적', '연결된 텍스트'의 성격을 갖는다.

요약시스템의 성능을 평가할 때에는 요약문 대 원문, 시스템이 작성한 요약문 대 수작업으로 작성한 요약문, 시스템 대 시스템 등을 비교한다. 일반적으로 텍스트 요약을 평가하는 방법은 내재적 평가와 외재적 평가의 두 가지 범주로 구분할 수 있다. 내재적 평가(intrinsic evaluation)는 요약문의 분석을 통하여 인간이 요약문의 품질을 직접적으로 판단한다. 요약문의 가독성(readability), 요약문 내 주요개념이 포함되어 있는지의 여부, 이상적인 요약과의 유사성 정도 등을 통해 측정한다.

외재적 평가(extrinsic evaluation)는 요약문의 품질을 요약문이 '다른 작업'을 수행하는데 얼마나 영향을 미치는지로 측정한다. 즉, 검색 문헌의 적합성 평가에 요약문이 얼마나 도움이 되는지, 분류에서 요약문이 활용되는 정도는 어떠한지로 요약문의 품질을 평가한다.

2.2 단락검색

단락검색(passage retrieval)은 문헌검색과 유사하지만 검색 이전 단계에 문헌으로부터 단락을 추출하는 과정이 선행되어야 한다. 즉, 단락검색의 과정은 문헌검색과 동일하지만 검색의 대상물이 문헌이 아닌 문헌으로부터 추출한 단락이 된다. 이때 '단락'의 개념을 어떻게 정의하느냐에 따라 단락검색의 효율성이 결정된다. 단락검색과 관련된 이전 연구들에서 내린 단락의 의미는 담화 단락(discourse passages), 고정크기 단락(window passages), 의미 단락(semantic passages)의 세 가지로 구분할 수 있으며, 그 중 단락검색의 성능이 가장 우수한 것으로 알려진 것은 의미 단락이다. 문헌을 의미 단락으로 구분하는 대표적인 알고리즘으로는 Heart(1994)의 텍스트타일링(TextTiling) 알고리즘이 있다.

3 요약문 생성 실험

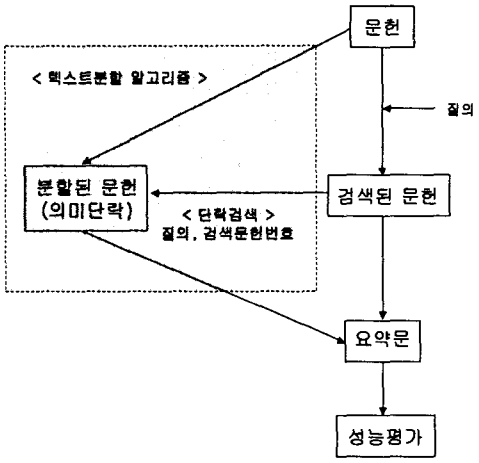
3.1 실험문헌집단

대개의 단락검색 실험에서는 문헌검색에서 사용하는 실험문헌집단을 그대로 사용한다. 이때 발생하는 문제점으로 다음의 세 가지를 들 수 있다: ①적합성 판정이 문헌 수준에서만 이루어져 있다. ②문헌의 길이가 다소 짧은 경우가 대부분이다. ③문헌검색을 위한 실험문헌집단은 실제 환경에서 사용되는 데이터집합과는 거리가 멀다.

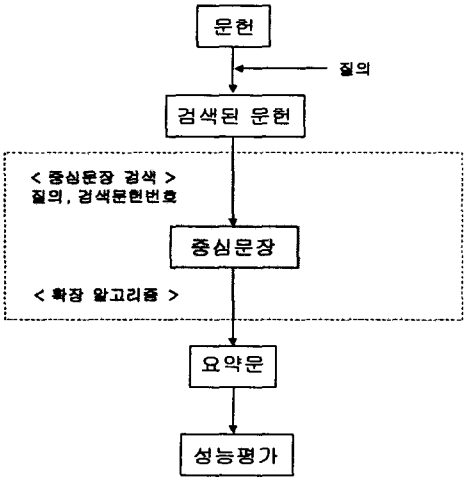
본 연구에서는 기존의 실험문헌집단 중에서 연구에 맞는 것을 발견할 수 없었으므로 주간조선, 주간동아, 한겨레21, 뉴스위크(한국판)의 경제 및 국제면 2000년 1월 ~ 12월 기사 중 길이가 5,000 바이트 이상인 기사 총 458건으로 실험문헌집단을 구성하였다.

실험에 사용할 질의는 실험용 문헌을 분석하여 경제부분 15개, 국제부분 15개로 모두 30개를 추출하였다. 실험용 질의는 질의번호, 질의제목, 질의설명, 기본질의어, 수작업 확장질의어로 구성된다. 문헌을 검색할 때에는 기본질의어를 사용해도 무방하나 검색된 문헌에서 질의와 적합한 문장을 검색하기 위해서는 질의가 구체적이어야 할 필요가 있으므로 본 실험에서는 문헌검색이나 단락검색, 문장검색에 수작업 확장질의어를 사용하였다.

요약문의 성능 평가를 위해서는 적합문헌의 문장에 대한 적합성 평가가 마련되어 있어야 하므로 각 질의에 적합한 문헌의 모든 문장에 대한 적합성 평가를 수행하였다.



<그림 2> 정적 단락검색 기법을 이용한 요약문 생성



<그림 3> 동적 단락추출 기법을 이용한 요약문 생성

3.2 정적 단락검색 기법

정적 단락검색 기법은 문헌을 의미단위 단락으로 미리 구분해 두었다가 질의가 입력되면 질의와 가장 유사한 단락을 요약문으로 제시하는 방법이다. 정적 단락검색 기법은 그림 2와 같이 텍스트 분할과 단락검색의 두 단계로 구성된다. 텍스트 분할 알고리즘은 문장간의 유사도를 측정하여 문헌을 의미 단락으로 자동 구분하는 과정으로 Hearst (1994)의 텍스트타일링을 바탕으로 하여 구현하였다. 이 단계에서는 문헌을 단락으로 구분할 때 고려되는 이웃문장의 수와 유사계수를 변수로 하여 실험하였다. 단락검색에서는 분할된 단락을 대상으로 질의에 적합한 단락을 선정할 때 사전실험에서 확인된 최적의 용어 가중치를 사용하였다.

3.3 동적 단락추출 기법

동적 단락추출 기법은 확장활성화(spreading activation) 기법을 문장단위에 적용한 요약문 생성 방법이다. 문헌을 '선형 의미망'으로 간주하고, 질의에 적합한 중심문장을 찾은 후 중심문장과 이웃문장의 유사도를 측정하여 이웃문장을 중심문장과 함께 요약문에 포함시킬지 여부를 결정한다. 즉, 중심문장을 기준으로 앞뒤로 확장하여 확장된 문장들을 요약문에 포함시킨다. 동적 단락추출 과정은 그림 3과 같이 중심문장 검색 부분과 확장 알고리즘 부분으로 구성된다. 동적 단락추출을 위한 실험에서는 사전실험 결과 가장 높은 성능을 보이는 용어 가중치를 적용하여 중심문장을 검색한 다음 확장활성화 알고리즘 중 병렬적 bnb 탐색을 이용한 알고리즘을 적용하여 요약문을 생성하였다.

3.4 정적 단락검색과 동적 단락추출 기법의 성능 평가

(1) 내재적 평가

실험을 통해 최적화한 정적 단락검색 기법과 동적 단락추출 기법을 전체 실험질의 30건에 대한 적합문헌 195건에 적용한 결과는 표 1과 같다. F 값은 수작업으로 선정한 요약후보 문장과 시스템이 선정한 요약후보 문장을 비교하여 측정하였으며, 연결성은 요약의 가독성을 측정하는 척도이다.

F 척도에서 동적 단락추출 기법이 정적 단락검색 기법에 비해 4.1% 높게 나타났다. 요약문 자체의 품질은 동적 단락추출이 우수한 것으로 보이며, 연결성은 동적 단락추출이 단일 단락을 요약문으로 제시하는 정적 단락검색보다 떨어지는 것으로 나타났다.

<표 1> 요약 성능 평가 결과

	정적 단락검색	동적 단락추출
F 값	0.394	0.435
연결성	1.000	0.680

(2) 외재적 평가

6명의 이용자를 대상으로 평가를 수행하였으며, 각 이용자에게는 기법별로 5개씩 총 10개의 질의를 제시하였다. 각 질의에 대해 검색된 상위 15개 문헌의 문헌번호와 요약문을 제시한 후, 문헌의 적합성 여부를 적합, 부적합으로 평가하여 실험문헌집단에 미리 판정되어 있는 적합문헌과 비교하였으며, 질의별로 적합성 판정에 걸리는 시간을 측정하게 하였다.

문헌검색기법으로는 벡터공간모델을 선택하였으며 단어빈도(1+ log(tf)), 코사인 정규화값, 역문헌빈도(1+log(N/df))를 용어 가중치로 사용하였다. 질의와 문헌의 내적을 구하여 가장 점수가 높은 상위 15개 문헌에 대해 요약문을 생성하여 이용자에게 제시하였다.

평가결과 정확률, 재현율, F 값에서 모두 동적 단락추출 기법이 정적 단락검색 기법보다 우수하게 나타났으며, 정확률에서 2.1%, 재현율에서 4.3%, F 값에서 3.1%의 성능 차이를 보였다(표 2 참조). 이러한 결과는 이용자가 동적 단락추출 기법을 통해 생성한 요약문을 보았을 때 정적 단락검색 기법을 이용한 요약문을 보았을 때보다 많은 문헌의 적합성을 정확하게 판정했음을 의미한다. 평가결과로 미루어보아 문헌의 적합성 판정에는 연결성보다 요약문의 품질이 많은 영향을 미치는 것으로 판단된다.

<표 2> 요약문에 대한 이용자의 문헌 적합성 판정 결과

	정적 단락검색	동적 단락추출
평균 정확률	0.699	0.720
평균 재현율	0.751	0.794
F 값	0.724	0.755

적합성 판정 외에 질의별로 적합성 판정에 소요되는 시간을 측정하도록 하였다. 표 3과 같이 동적 단락추출이 정적 단락검색을 이용하였을 때보다 우수한 성능을 보였다. 동적 단락추출이 정적 단락검색에 비해 문헌 당 적합성 판정에 걸리는 시간을 평균 0.51초 앞당길 수 있었던 것은 동적 기법의 경우 질의와 유사한 문장을 3개 선정하고 확장을 하여 6개의 문장을 추출하였으므로 정적 기법에 비해 적합성 판정에 단서가 될 수 있는 문장이 요약의 앞쪽에 출현할 가능성이 크기 때문이다. 따라서, 동적 단락추출 기법을 이용하였을 때 이용자들은 적합성 판정을 좀 더 빠르고 정확하게 할 수 있었다.

<표 3> 질의별 적합문헌 판정에 소요된 시간

	정적 단락검색	동적 단락추출
질의당 소요시간	4분 31초	4분 16초

4 결론

내재적 평가 중 요약문의 품질 면에서 우수한 성능을 보인 동적 단락추출이 외재적 평가에서도 좋은 성능을 보였다. 이용자에게 각 기법을 통해 생성한 요약문을 보여주고 문헌의 적합성을 판정하도록 하였을 때 동적 단락추출을 적용한 경우가 정적 단락검색을 적용한 경우보다 정확률, 재현율에서 모두 우수하게 나타났다. 또한, 동적 단락추출 기법을 통해 생성한 요약문이 정적 단락검색 기법을 통해 생성한 요약문에 비해 적합문헌 판정에 걸리는 시간을 단축할 수 있었다.

결론적으로 이용자 중심 요약문을 생성하는 방법으로 본 연구에서 제안하는 동적 단락추출 기법이 정적 단락검색 기법에 비해 효율적인 방법임을 실험을 통해 증명하였다.

참고문헌

Hearst, M.A. 1994. "Multi-paragraph segmentation of expository text". ACL '94. Las Cruces, NM.

Mani, I., and M.T. Maybury. 1999. Advances in automatic text summarization. The MIT Press.

Sanderson, M. 1998. "Accurate user directed summarization from existing tools". CIKM '98 45-51.