

# 탐색결과에 근거한 자연어질의 자동확장 및 응용에 관한 연구 고찰\*

## The Pragmatics of Automatic Query Expansion Based on Search Results of Natural Language Queries

노정순(Jung-Soon Ro)\*\*

### 목 차

- |                                   |                  |
|-----------------------------------|------------------|
| 1 서론                              | 4.1.2 추론네트워크모델   |
| 2 검색결과에 기반을 둔 자연어 질의확장<br>시스템의 속성 | 4.2 확장용어의 선정     |
| 3 벡터공간모델                          | 5 다중질의 결합        |
| 3.1 용어와 문헌의 가중치                   | 6 의사적합성피드백탐색     |
| 3.2 적합성피드백을 사용한 질의확장              | 7 적합성판정          |
| 4 확률모델                            | 7.1 배움표본의 크기와 내용 |
| 4.1 문헌의 순위화                       | 7.2 적합성판정척도      |
| 4.1.1 이진독립모델                      | 8 탐색용어의 수정 및 확장  |
|                                   | 9 결론             |

### 초 록

본 연구는 자연어 비볼리언 탐색에서 탐색결과에 근거하여 질의를 수정, 확장, 결합하여 검색효과를 향상시키는 시스템들을 개념모델별로 성능을 고찰하고, 성능에 영향을 끼치는 요소들을 분석하여, 이론적인 개념의 틀을 제시하였다. 용어의 가중치기법, 문헌의 순위화방법, 용어선정알고리즘, 질의확장에 사용된 문헌수와 용어수, 적합성판정정보의 출처 및 척도, 배움표본의 크기, 부적합문헌정보의 사용여부, 용어확장방법, 질의의 크기, DB의 종류와 크기 등에 의해 영향을 받는 것으로 분석되었다.

### ABSTRACT

This study analyses the researches on automatic query modification, expansion and combination based on search results of natural language queries and gives a conceptual framework for the factors affecting the effectiveness of the relevance feedback. The operating and experimental systems based on the vector space model, the binary independence model and the inference net model are reviewed, and it is found that the effectiveness of query expansion is affected by conceptual models, algorithms for weighting terms and documents and selecting query terms to be added, size of relevant and non-relevant documents to be used and size of terms to be added in relevance feedback, query length, type and size of DBs, etc.

\* 이 논문은 1997년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

\*\* 한남대학교 문헌정보학과 교수

■ 논문 접수일 : 1999년 5월 12일

## 1 서 론

정보를 필요로 하는 이용자(End-User)가 직접 탐색을 수행하는 온라인 정보검색시스템에서 탐색자들은 얼마나 성공적으로 탐색을 수행하는가? 온라인 검색시스템의 주종을 이루는 불논리검색시스템에서 최종이용자들이 겪는 탐색상의 어려움, 특히 주제탐색의 어려움은 여러 문헌에 보고되어 있다.

OPAC에서 대부분의 이용자는 관련있는 몇개의 문헌으로 만족하나, 서지 정보검색시스템에서 많은 이용자는 관련있는 모든 문헌을 검색하기를 원하는 것으로 보인다(Khoo & Poo 1994). 그러나 OPAC에서 주제탐색(주제명탐색+서명키워드탐색)으로 문헌을 하나도 검색하지 못한 실패율은 대략 30%(Lynch 1989)에서 심한 경우는 62%(Hunter 1991)나 되는 것으로 보고되었다. IR시스템에서도 CD-ROM 이용자(최종이용자)는 스스로 CD-ROM을 쉽게 잘 탐색한다고 믿고 있지만 실제 그들이 검색한 문헌은 중요문헌의 1/3뿐이었다(Lancaster et al. 1994).

최종이용자가 겪는 이러한 탐색의 어려움은 사용하는 특정 DB나 시스템에서 사용되는 시소러스나 주제명표목, 용어사전화일, 시스템언어 등에 대한 이해부족은 물론, 불논리와 검색결과를 축소 혹은 확대하는데 필요한 탐색전략(strategy)과 기법(tactic) 등에 대한 이해가 부족함에서 기인한다(Borgman 1996). 특히 최종이용자들이 정보요구를 불논리탐색식으로 표현하는데 겪는 어려움과 불논리검색이 갖는 단점을 해결하기 위해 불연산자를 사용하지 않은 자연어로 정보를 요구하고 자동으로 질의를 생성하여 질의에 대한 유사도값으로 문헌의 순위를 부여하는 비블리언 탐색시스템이 수많은 텍스트 기반 실험시스템에

서 연구되고 있으며, Dialog나 Lexis/Nexis, Westlaw와 같은 상용 벤더시스템과 CITE, Okapi, Cheshire와 같은 OPAC에서 불논리탐색과 함께 제공되고 있다.

비블리언 탐색에서 이용자는 불연산자를 사용하지 않고 탐색용어를 입력하거나 자연어 문장형식으로 정보요구를 입력한다. 문장형식의 정보요구는 파싱과 스테밍을 통해 불용어를 제외한 개념어가 추출되어 초기질의백터가 자동생성된다. 이 초기질의백터는 시소러스나 의미네트워킹(WordNet, Nearest Neighbours, Maximum Scanning Tree 등) 등과 같은 정보원을 사용하여 확장되거나(지식 기반 확장), 초기질의백터의 탐색으로 검색된 문헌 중 적합문헌에 출현한 용어들을 사용하여 확장된다(탐색결과기반 확장). 다양한 방법으로 검색된 결과를 결합하여 검색효과를 향상시키기도 한다. 비블리언 탐색에서 자연어 질의탐색은 탐색전문가가 수행한 불논리탐색에 비해 차이가 없거나(Borlund & Ingwersen 1997, Hersh & Hickam 1995, Hersh et al. 1994, Lu et al. 1995, Paris & Tibbo 1998, Robertson & Thompson 1990, Turtle 1994 등), 불논리탐색보다 우수한 것으로 보고되었다(Hersh & Day 1997, Losee & Bookstein 1988, Fox 1988, 1990, Fox et al. 1993, Turtle & Croft 1992).

본 연구는 자연어질의 비불논리검색에서 지금까지 이루어진 다양한 질의 생성 및 수정, 확장에 관한 기법 중 탐색결과에 근거하여 자동으로 질의를 수정 혹은 확장하는 시스템에 대한 연구를 고찰하였다. 텍스트검색 시스템을 연구하고 설계하는데 구성요소별로 올바른 선택결정을 할 수 있도록 다음과 같은 질문을 중심으로 분석하였다.

1. 어떤 개념모델을 사용할 것인가?

2. 어떤 용어가중치기법과 문헌순위부여방법을 사용할 것인가?
3. 이용자의 적합성 판정이 꼭 필요한가?
4. 적합성판정정보(배움표본)의 크기와 내용은?
5. 피드백에서 부적합문헌의 사용 효과는?
6. 질의를 자동 수정할 것인가? 확장할 것인가? 확장한다면 그 크기는? 어떤 용어를 선택할 것인가?
7. 용어선택 알고리즘은?
8. 용어의 자동선택이 좋은가? 대화식 수동선택이 좋은가?

본 연구에서 질의(query)란 불논리연산자나 인접연산자로 표현된 탐색식이나 벡터로 표현된 질의를 의미하고, 연산자로 연결되는 개개용어나 벡터를 이루는 용어는 질의용어(query term)로 표현된다. 질의확장(Query Expansion: QE)이란 적합문헌에 출현하는 용어 중 초기질의에 사용되지 않은 용어를 추가로 피드백질의에 포함시키거나, 시소러스의 동의어나 하위개념, 용어네트워크에서 가장 유사도가 큰 단어(Nearest Neighbours) 등을 추가하는 것을 의미한다. 그러나 질의수정(Query Modification)이란 적합성정보에 의해 초기질의용어의 가중치만 수정하고 새로운 용어는 질의에 추가하지 않는 것을 의미한다.

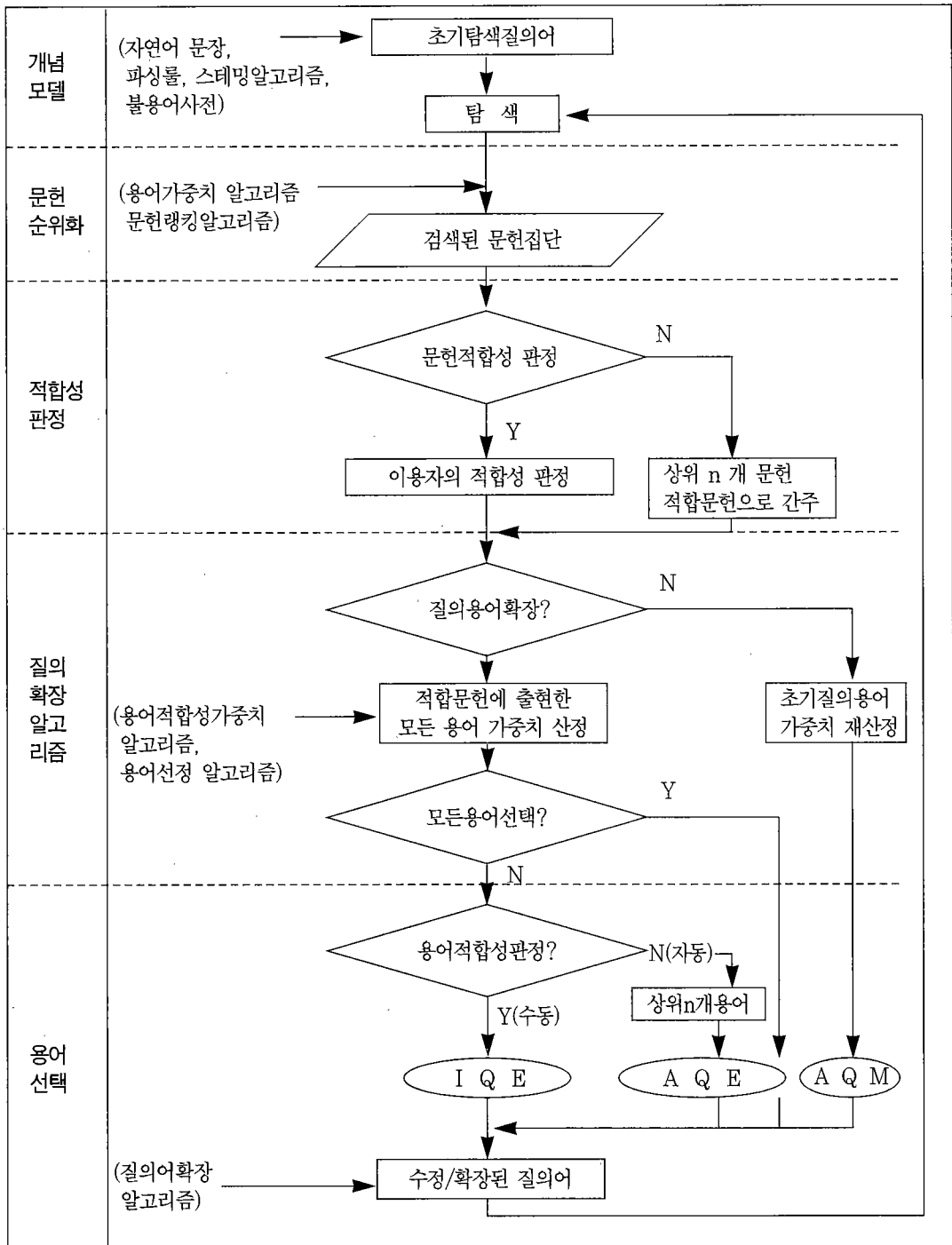
## 2 검색결과에 기반을 둔 자연어 질의확장시스템의 속성

〈그림 1〉은 QE시스템의 5가지 속성 즉 개념모델, 문헌순위부여방법, 적합성 판정, 적합성피드백 질의확장알고리즘, 및 용어선택방법과 QE시스템에서 이들의 위치와 역할을 보여준다. 개념모델은 탐색접근방식으로 다른 속성을 결정한다.

문헌의 순위부여방법과 용어의 적합성가중치부여방법 및 질의확장알고리즘은 개념모델에 따라 결정된다. 지금까지 비블리언환경에서 QE시스템에 대한 연구는 벡터공간모델과 확률모델 환경내에서 주로 수행되어 왔다. 블리언탐색과 달리 비블리언탐색은 문헌용어와 질의용어에 가중치를 부여하고 문헌을 순위화한다. 벡터모델에서 문헌과 질의는 각각  $D=(d_1, d_2 \dots d_k)$ 와  $Q=(q_1, q_2 \dots q_k)$ 의 형태로 이루어진 k차원의 벡터로 이루어지고, 두벡터의 유사도값 순으로 문헌에 순위가 부여된다. 그러나 확률모델에서는 일반적으로 문헌에 출현하는 모든 질의용어의 가중치의 합으로 문헌이 순위화된다.

검색된 문헌에 대한 적합성정보는 이용자에 의해 화면에서 입력되거나 이용자의 적합성 판정없이 검색된 상위 n개의 문헌이 적합문헌으로 간주된다.

벡터모델에서 적합성피드백탐색은 적합성판정정보에 의거하여 초기질의벡터에 적합문헌벡터를 추가하고 부적합문헌벡터를 마이너스하여 새로 확장된 질의벡터가 생성되고, 문헌은 다시 생성된 새질의벡터와의 유사도값으로 순위화된다. 확률모델에서 초기질의용어의 가중치는 적합성정보에 근거하여 재산정되거나(Automatic query modification), 적합문헌에 출현한 비질의용어가 적합성가중치에 의해 질의용어로 선택되어 추가되고, 문헌은 문헌에 출현하는 수정된 질의용어의 가중치의 합으로 다시 순위화된다. 적합성가중치순으로 순위가 부여된 적합문헌에 출현하는 비질의용어는 전시된 후보용어중 이용자가 선택하거나(Interactive query expansion), 컴퓨터시스템이 탐색자와 상호작용없이 상위 n개의 용어를 선택하여(Automatic query expansion) 질의를 확장한다.



〈그림1〉 질의확장시스템의 속성

### 3 벡터공간모델

#### 3.1 용어와 문헌의 가중치

〈표 1〉은 벡터모델에서 초기탐색에 주로 사용되는 문헌용어와 질의용어의 가중치알고리즘을 정리한 것이다. 문헌용어의 가중치는 문헌에서 용어  $t$ 의 빈도수( $tf$ ), 문헌의 길이( $dl$ ), DB내 용어  $t$ 가 출현하는 문헌수( $n$ ), DB내 문헌수( $N$ ), DB에서 용어  $t$ 의 빈도수( $tfc$ ), 정규화(normalization) 등에 의해 영향을 받는다. 정규화란 문헌(또는 질의)의 길이가 길수록 짧은 문헌(또는 질의)에 비해 주제를 나타내는 질의용어의 출현빈도수는 높아지므로 문헌(혹은 질의)의 길이에 의해 용어의 가중치가 영향을 받는 것을 최소화하기 위한 것이다. 질의용어나 문헌용어의 가중치를 계산하는 방법은 여러 연구에서 제안되었고 몇몇 연구에서는 다양한 용어가중치기법 비교 평가되었다(Noreault et al. 1981, Ro 1988, Salton & Buckley 1988).

벡터모델에 근거한 적합성피드백시스템에서 용어가중치의 효과는 주로 SMART시스템에 기반을 둔 연구에서 수행되었다. 6개의 적합성피드백의 성능을 테스트한 Salton과 Buckley(1990)의 연구에서 문헌용어와 질의용어를 모두 atc로 코싸인정규화시킨 가중치는 5개 데이터set에서 이진가중치보다 탁월하게 우수하였다. 그러나 용어의 평균출현빈도수가 가장 낮은(질의용어1.00, 문헌용어 1.21) NPL 실험set에서는 이진가중치가 우수하였다.

자연어 질의문으로부터 질의벡터를 생성하고 대규모의 본문텍스트로부터 문헌벡터를 생성하는 TREC<sup>1)</sup> 실험에서 SMART는 문헌용어와 질의용어의 가중치로 코싸인정규화된  $ntc, ntc$ 와

(TREC1에서)  $lnc, ltc$ (TREC2와 TREC3에서)를 사용하였다. 그러나 SMART의 코싸인정규화는 Inquiry의 최대단어빈도정규화와 Okapi의 바이트길이정규화에 비하여 낮은 검색효과를 보였으며, 코싸인정규화를 수정한 Pivot 코싸인정규화로 그 성능을 향상시킬 수 있음이 보고되었다(Singhal et al. 1996). SMART는 TREC4에서 Pivot 코싸인정규화(Lnu, ltu)가 코싸인정규화(lnc, ltc)보다 우수함을 증명하고(Buckley et al. 1995), TREC4 이후 모든 TREC실험에서 용어의 가중치공식으로 사용하였다. Lnu, ltu는 SMART를 기반으로한 AT&T의 TREC6 실험에서도 사용되었으나(Singhal et al. 1997), AT&T의 TREC7에서는 Pivot 바이트길이정규화(dnb, dtn)가 사용되었다(Singhal et al. 1998).

2개의 단일용어로 이루어진 구(phrase)나 복합명사는 각각의 단일용어로 취급하여 가중치를 계산하는 대신 하나의 구로써의 가중치를 부여할 때 성능이 보다 우수하였다. 구의 자동인식 및 가중치부여방법은 여러 연구에서 비교테스트되었으나(Fagan 1989, Croft et al. 1991, Keen

1) 대규모의 텍스트검색에서 연구를 격려하고 아이디어를 교환하여 탁월한 검색성능을 보인 검색기술을 빠르게 상품화하기 위한 목적으로, 미국립표준기술연구소(National Institute of Standards & Technology)에서 1992년부터 매년 주관하는 Text REtrieval Conference. 각기 다른 검색특성을 지닌 검색시스템이 동일환경에서 평가될 수 있도록 매년 표준데이터(표준 DB, 표준질의문 및 적합성판정정보)를 제공한다(Routing run). 표준질의문(topic statement)은 title과 description, narrative, concepts(TREC1과 2에서만)으로 이루어지며, 실험에 참여하는 시스템은 각 요소 혹은 전체 질의문으로 부터 질의를 생성한다. 특별탐색(Ad-hoc run) 목적으로 제공되는 질의문에 대해서는 공식적합성판정정보를 제공하지 않고 실험에 참여하는 시스템이 보고한 탐색결과를 취합하여 적합성을 판정하여 다음해 Routing run을 위한 데이터로 제공한다.

〈표 1〉 용어가중치(wt) 알고리즘

$tf$  : 문헌이나 질의내 용어  $t$ 의 출현빈도수  
 $N$  : DB내 레코드(문헌)총수  
 $n$  : DB내 용어  $t$ 가 출현하는 문헌수  
 $tf_c$  : DB내 용어  $t$ 의 총출현 빈도수  
 $dl$  : 문헌에 출현하는 단어수(문헌길이) ( $avdl$  은 평균문헌길이)

문헌(혹은 질의)내 용어빈도수

- |     |                                |                             |
|-----|--------------------------------|-----------------------------|
| (1) | 1                              | Binary(Singhal et al. 1996) |
| (2) | $tf$                           | nln(Salton & Buckley 1988)  |
| (3) | $\log tf + 1.0$                | lnn(Salton & Buckley 1988)  |
| (4) | $c + (1-c) \frac{tf}{\max tf}$ | ann( $c=0.5$ )              |

역문헌 빈도수

- |     |  |                            |
|-----|--|----------------------------|
| (5) | $\log \frac{N}{n}$                               | Idf(Salton & Buckley 1988) |
| (6) | $-\log \frac{n}{\max n} = \log \frac{\max n}{n}$ | (Sparck Johns 1979b)       |
| (7) | $\log \frac{N-n}{n}$                             | icf(Croft & Harper 1979)   |
| (8) | $\log \frac{N-n+0.5}{n+0.5}$                     | (Robertson 1986)           |

문헌(혹은 질의) 내 용어빈도수×역문헌 빈도수

- |      |  |                                   |
|------|--|-----------------------------------|
| (9)  | $tf \cdot \log \frac{N}{n}$                            | ntn                               |
| (10) | $\log(tf+1.0) \times \log \frac{N}{n}$                 | ltn                               |
| (11) | $\frac{\log(tf+1) \times idf}{\log dl}$                | PRISE(Harman 1986, 1992)          |
| (12) | $\frac{(tf/n)^a}{\log \max(cutoff, n)}$                | avtf(Kwok 1996, Kwok & Chen 1998) |
| (13) | $0.5 + 0.5 \frac{tf}{\max tf} \times \log \frac{N}{n}$ | atn                               |
| (14) | $1 + \log(1 + \log tf) \cdot \frac{N+1}{n}$            | dtn(AT&T TREC7)                   |

정규화

- |      |   |                                      |
|------|---|--------------------------------------|
| (15) | $\frac{\log(tf)+1.0}{\sqrt{\sum (\log(tf)+1.0)^2}}$   | lnc(SMART TREC2)                     |
| (16) | $\frac{tf \cdot \log \frac{N}{n}}{\sqrt{\sum (tf \cdot \log \frac{N}{n})^2}}$   | ntc(SMART TREC1, Aalbersberg 1992)   |
| (17) | $\frac{(\log tf+1.0) \times \log(\frac{N}{n})}{\sqrt{\sum [(\log tf+1.0) \times \log(\frac{N}{n})]^2}}$                     | ltc(SMART TREC2, Buckley et al 1994) |
| (18) | $\frac{0.5 + 0.5 \frac{tf}{\max tf}}{\sqrt{\sum (0.5 + 0.5 \frac{tf}{\max tf})^2}}$   | anc(Salton & Buckley 1998)           |
| (19) | $\frac{0.5 + 0.5 \frac{tf}{\max tf} \times \log \frac{N}{n}}{\sqrt{(0.5 + 0.5 \frac{tf}{\max tf})^2 (\log \frac{N}{n})^2}}$ | atc(Salton & Buckley 1990, 1991)     |
| (20) | $\frac{1.0 + \log tf}{(1.0 - slope) \times pivot + slope \cdot unique \ terms}$   | Lnu(SMART TREC4, Pivot 문헌길이 정규화)     |
| (21) | $\frac{(1.0 + \log tf) \cdot (\log \frac{N}{n})}{(1.0 - slope) \times pivot + slope \cdot Unique \ terms}$                  | ltu(SMART TREC4)                     |
| (22) | $\frac{1 + \log(1 + \log tf)}{0.8 + 0.2 \cdot \frac{dl \text{ in byte}}{avd \text{ in byte}}}$                              | dnb(AT&T, Pivoted 바이트길이 정규화)         |

1991, 1992 등), 적합성피드백에서 비교평가는 거의 수행되지 못하였다. TREC실험에서 SMART(TREC1에서 TREC7까지)는 TREC1의 실험문헌set에 15번 이상 출현하는 구를 대상으로 복합어리스트를 만들어, 복합어리스트에 존재하는 복합어만을 벡터에 포함시키고 복합어의 출현빈도에 근거하여 단어의 가중치부여방법과 동일한 기법으로 가중치를 부여하였다. Okapi는 질의문에 출현하는 인접용어(단어s와 t로 이루어진)의 가중치를 인접용어st의 출현빈도 대 단어 s와 t의 출현빈도를 곱한 값의 비율  $f(st) \cdot c / (f(s) \cdot f(t))$ 로 부여하였다.

벡터공간모델을 기반으로 하는 시스템에서 문헌벡터와 질의벡터의 유사도는 단순유사도 혹은 내적유사도(Inner Similarity) 공식  $sim(D, Q) = \sum (d_k \cdot q_k)$ 로 표현된다. 단순유사도에서 문헌과 질의벡터를 이루는 용어에 이진수의 가중치를 부여한다면, 문헌의 유사도는 문헌벡터에 들어있는 질의용어수로 표현된다(Coordination match). 문헌벡터를 2진벡터로하고 질의용어에만 가중치를 준다면 문헌의 유사도는 문헌벡터에 들어있는 질의용어의 가중치의 합이 된다(Weighted coordination match). Noreault et al.(1981)은 질의어와의 유사도에 따라 문헌의 순위를 부여하는데 사용될 수 있는 67개의 유사도공식을 소개하고, 그중 24개의 공식을 테스트 하였다.

### 3.2 적합성피드백을 사용한 질의확장

벡터모델을 기반으로한 피드백 질의확장시스템에서 질문벡터 Q는 적합·부적합문헌에 출현하는 용어에 근거하여 새로운 질문벡터 Q'로 수정된다. 적합성피드백에 의한 질문벡터 수정은

1960년대 Rocchio(1971)에 의해 제안되었고, Ide(1971)는 Rocchio의 공식을 확장하여 두가지의 다른 피드백 전략을 조사하였다. 확장된 질문벡터에서 질의용어의 가중치는 초기질의용어의 가중치에 적합문헌집단에 출현하는 질의용어의 가중치를 합하고 부적합문헌 집단에 출현하는 질의용어의 가중치는 마이너스하여 새로운 가중치를 갖는다. 적합문헌에 출현하는 비질의용어는 새로운 질문벡터에 추가된다. 그러나 부적합문헌에만 출현한 비질의용어는 새 질문벡터에 추가되지 않고 적합문헌에서 추출된 비질의용어의 가중치를 계산하는데 사용된다.

〈표 2〉 (벡터 적합성피드백 확장모델)

	$R_i$ : 적합문헌 i의 벡터
	$S_i$ : 부적합문헌의 i의 벡터
	$n1$ : 적합문헌의 수
	$n2$ : 부적합문헌수
Rocchio	$Q = \alpha Q + \beta \sum_{i=1}^{n1} \frac{R_i}{n1} - \gamma \sum_{i=1}^{n2} \frac{S_i}{n2}$
Ide Regular	$Q = Q + \sum_{i=1}^{n1} R_i - \sum_{i=1}^{n2} S_i$
Ide dec-hi	$Q = Q + \sum_{i=1}^{n1} R_i - S_i$
Incremental	$Q = Q + \beta R_i$ (Di가 적합문헌일 때) $Q = Q - \gamma S_i$ (Di가 부적합문헌일 때)

Salton과 Buckley(1990)는 6개의 표준실험 set에서 이 세방법을 다른 세 확률모델과 비교하였다. 전체적으로 Ide dec-hi가 가장 우수하였다. 문헌용어에 가중치를 주었을 때는 세가지 벡터기법이 모두 확률기법보다 우수하였고, 벡터기법 중에서는 용어의 가중치나 확장용어의 크기와 상관없이 Ide dec-hi > Ide regular > Rocchio순의

효과를 보였다. Rocchio에서는  $\beta=0.75$ ,  $\gamma=0.25$  일 때 가장 우수하였다.

Aalbersberg(1992)는 본문(Fulltext) DB를 사용하여 검색된 문헌을 하나씩 전시하여 적합문헌이면 확장을 반복하는 피드백시스템을 제안하고 3가지 벡터기법과 비교하여 평가하였다. 검색된 문헌수를 15개로 제한하고  $\beta$ 와  $\gamma$ 를 변화시켰을 때, 가장 높은 정확률은 Ide 45.06%( $\beta=1.0$ ,  $\gamma=0.1$ 에서), Ide dec-hi 44.90%( $\beta=0.6$ ,  $\gamma=0.3$ 과  $\beta=0.8$ ,  $\gamma=0.5$ 에서), Rocchio 44.89%( $\beta=1.0$ ,  $\gamma=0.6$ 에서)의 순이었다. 세기법 모두에서 피드백으로 탐색을 확장하기 전( $\beta=\gamma=0$ )과 비교해 약 18%의 향상을 보였다. 하나씩 전시하여 그 문헌이 적합문헌일 경우에만 탐색을 재수행하는 방식에서는  $\beta=0.4$ ,  $\gamma=0.1$ 에서 50.85%의 정확률을 얻었다. Rocchio 방식으로 부적합문헌은 무시하고 적합문헌에 출현한 용어의 가중치만을 합할 경우( $\beta=1.0$ ,  $\gamma=0.0$ )가 부적합문헌의 비율을 40% 이상으로 했을 때( $\beta \leq 0.6$ ,  $\gamma \geq 0.4$ )보다 우수하였다.

TREC 연구에서 SMART시스템(코넬대학)은 적합문헌에 출현하는 모든 용어로 질의를 확장하는 대신 적합문헌에 출현하는 용어 중 상위  $n$ 개의 용어를 선택하여 질의를 확장하였다. TREC1에서는 Ide dec-hi공식이 사용되었으나 TREC2 이후에는 Rocchio공식이 사용되었다. 로치오의  $\alpha:\beta:\gamma$ 값은 8:16:4(TREC2와 3), 64:64:2(TREC4), 8:64:64(TREC5), 8:8:8(TREC 6과 7)이 사용되었다.

TREC1에서부터 TREC3까지 초기 실험에서는 비교적 많은 용어(300~500개)가 선택되었다. TREC2에서 질의용어를 확장하지 않고 초기 용어의 가중치만 수정하여 피드백탐색을 수행한 결과 +7%의 향상을 보았으나, 20개 용어와 6개

의 구를 추가하여서는 +17%, 300개 용어와 50개 구를 추가하여서는 +38%의 향상을 보였다. TREC3에서는 질의용어는 확장하지 않고 가중치만 수정하여 +6%의 향상을, 300개의 용어와 30개의 구로 질의를 확장하여서는 +24%의 향상을 보였다.

TREC4 이후는 Pivot 정규화가중치기법과 동적피드백최적화(Dynamic Feedback Optimization), 소규모의 용어(50개 용어와 10개의 구)를 선택하여 SMART의 성능을 크게 향상시켰다. Pivot 정규화가중치와 동적피드백최적화와 함께 TREC5에서 적용된 QZ(Query Zone)과 TREC6에서 적용된 SC(Superconcept)는 TREC7까지 사용되었다. TREC5와 TREC6에서는 100개의 용어와 10개의 구 외에 적합문헌과 적합문헌 두배의 부적합문헌에 동시 출현하는 한 쌍의 용어(pair) 50개를 선택하여 피드백질의벡터에 추가하였다.

TREC1에서부터 TREC7까지 전체적으로 SMART의 성능은 가중치알고리즘을 Pivot 정규화로 변경( $ntc, ntc \rightarrow Lnc, ltu$ )하고 QZ와 DFO를 사용하여 크게 향상되었으나 SC는 큰 도움이 되지 못하였다. 한편 SMART의 TREC 공식보고는 아니지만 Buckley 등(1994)은 TREC2에서와 같은 용어가중치( $lnc, ltc$ )와 로치오계수(8:16:4)를 사용하여 적합문헌수가 증가할수록(500개까지), 또한 확장에 사용할 용어를 많이 선택하면 할수록(4000개까지) 검색효과는 향상되었음을 보고하였다.

동적피드백최적화(DFO)는 질의확장에 사용된 용어를 선정 후 가중치를 재조정하는 것으로, 그 효과는 TREC3 데이터를 사용한 Buckley와Salton(1995)의 연구에 소개되었다. 상위 200개 문헌중 적합문헌집단에 출현하는



용어를 용어가 출현하는 적합문헌수( $r$ )로 가중치를 부여하여 상위 50개의 용어를 선택한 후 Rocchio 공식으로( $\alpha:\beta:\gamma=2:4:1$ ) 질의벡터를 수정하면 정확률이 10% 향상되었다. 그러나 여기에 확장된 질의용어의 가중치를 50% 증가시켜 (pass 비율=50%) 성능이 향상되면 변경된 가중치를 유지하나 성능이 향상되지 않으면 이전의 가중치로 돌아가 패스비율을 25%로 줄여 같은 작업을 반복하고 그래도 성능이 향상되지 않으면 패스비율을 12.5%로 다시 줄여 3회 반복한 가중치로 질의를 확장했을 때는 정확률이 초기검색에 비해 20% 향상되었다. 이 DFO방식에서 가장 높은 정확률은  $\alpha, \beta, \gamma$ 값을 (2, 64, 8)로 주고 패스비율을 [1, 0.5, 0.3, 0.2, 0.1, 0.05]로 주었을 때 얻을 수 있었다(30% 향상). 이때 확장하는 용어수를 단일어 50개에서 단일어 200개와 구 10개로 늘리면 더 좋은 결과를 얻을 수 있었으나, 단일어 300개와 구 30개로 늘리면 단일어 200개와 구 10개만큼 우수하지는 못하였다. 이 DFO방식은 QZ의 효과를 실험한 Singhal et al.(1997)의 연구와 SMART를 사용하여 TREC에 참여한 AT&T의 연구에서도 확인되었다.

QZ(query zone)란 문헌공간벡터에 있는 모든 문헌 중에서 질의의 주제영역에 있는 문헌만을 모은 작은 벡터공간으로, 초기질의를 QZ로 제한 탐색하여 질의를 확장하기 위한 것이다. Singhal et al.(1997)의 연구에서는 3개의 QZ가 사용되었다. 초기질의로 검색된 상위  $n$ 개의 문헌과 기타 알고있는 적합문헌으로 이루어진 QZ-1에서는  $n = 2000$ 일 때 가장 효과적이었다(QZ를 사용하지 않았을 때 보다 +11.4%(TREC3), +6.4%(TREC4)향상). 유사도값이 임계치값 이상인 문헌으로 이루어진 QZ-2에서는 임계치

=0.25일 때가 가장 효과적이었다(+11.3%(TREC3), +5.1%(TREC4)향상). DFO를 사용한 QZ-3에서는 +12.1%(TREC3), +8.9%(TREC4) 향상되었다.

SMART시스템을 기반으로 TREC6에 참여한 AT&T는 기본적으로는 코넬에서 사용된 것과 같이 Lnu, ltu에 의한 상위 5000개 문헌의 QZ, Rocchio를 사용한 용어와 구의 선택, DFO(3 pass)에 의한 피드백질의가중치의 재산출 등을 사용하였다. 코넬은 10% 이상의 적합문헌에 출현하는 단어와 5% 이상의 적합문헌에 출현한 구를 대상으로 QZ에 있는 부적합문헌정보만 사용하여(문헌용어가중치는 Lnu 사용) 100개의 단어와 10개의 구를 선정하였지만, AT&T는 로치오(8, 64, 64)를 사용하여(문헌용어가중치는 Ltu 사용) 100개의 단어와 20개의 구를 선정하였다. QZ를 사용하여 +8.0%, DFO로 확장용어의 가중치를 재산정하여 +15.9%, 100개의 pair-word를 추가하여(QZ + pair-word + FDO) +25.4%의 정확률 향상을 보였다.

한글텍스트를 대상으로 피드백질의확장을 연구한 김주연 등(1997)은 1053개 문헌과 30개 질의로 구성된 한글 KT 테스트set를 사용하여 모든 적합문헌에 출현하는 용어  $t$ 와 질의와의 유사도를 계산하고 모든 부적합문헌에 출현하는 용어  $t$ 와 질의와의 유사도를 계산한 후, 적합문헌에서의 유사도값에서 부적합문헌에서의 유사도값을 뺀 값(최종 관련도)으로 용어의 순위를 부여하였다. 7~12개 용어로 확장할 때가 성능을 가장 향상시켰으며, 10~15개 이상의 용어로 확장할 경우에는 확장하지 않는 방법보다 오히려 성능을 감소시키기도 하였다.

## 4 확률모델

### 4.1 문헌의 순위화

#### 4.1.1 이진독립모델

확률모델은 원래 Maron과 Kuhns(1960)가 자동색인을 위해 제안한 확률색인모델에서 출발하였으나 오늘날 질의확장에 사용되는 대부분의 확률모델은 Robertson과 Sparck Jones(1976)의 확률모델을 기반으로 한다. 이 모델은 특정질문으로 검색된 적합문헌에 나타나는 용어는 검색된 부적합문헌에 나타나는 용어보다 높은 적합성 가중치를 가진다는 전제를 기반으로 한다. 이진독립모델(Binary Independence model)에서 문헌은 질의용어의 출현여부에만 관심을 갖고 "1"이나 "0"으로 구성된 문헌벡터  $D=(d_1, d_2 \dots d_k)$ 로 가정되며, 색인어는 적합문헌과 부적합문헌에서 독립적으로 발생한다는 가정 아래 Bayes의 확률 공식을 사용하여 질의어에 대한 문헌의 적합성 정도를 문헌이 적합문헌일 확률 대 부적합문헌일 확률로 표시한다. 용어가 적합문헌에 출현하는 확률  $p$ 값을  $r/R$ 로, 부적합문헌에 출현할 확률  $q$ 값을  $(n-r)/(N-R)$ 로 대입한  $F_4$ 는 용어가 적합문헌에 한번도 출현하지 않을 경우를 대비하여 불확실성(Uncertainty)을 허용한  $F_4(.5)$ 로 수정되었다(Sparck Jones 1979a).

이진독립모델과 투포아슨독립모델(TPI)을 불논리적 정규형(Boolean Conjunctive Normal Form) 탐색순위화와 비교한 연구(Losee & Bookstein 1988)에서  $F_4$ 는 TPI보다 우수하였다 ( $F_4/TPI$ )CNF). 또한 1개 문헌의 적합성정보만 사용한 실험에서도  $F_4$ 는 TPI보다 우수하였다(Losee 1988). 그러나 이 결론은 초록이나 제목 탐색에 제한되며, 본문(Fulltext) 탐색에서는 오

히려 TPI가  $F_4$ 보다 검색효율이 우수한 것으로 소개되었다(문성빈 1993).

TREC실험에서 확률모델을 기반으로한 시스템 중 가장 우수한 성능을 보인 시스템 중의 하나인 Okapi는  $F_4(.5)$ 를 기반으로한 대표적인 시스템이다. Okapi는 TREC1에서  $F_4(.5)$ 로 문헌을 순위화하였으나, TREC2에서는 문헌용어  $tf$ 나 질의 용어  $qtf$ 의 분포를 각각 투포아슨분포의 혼합물로 가정하고,  $tf$ 와  $qtf$ 를 정규화시켜  $F_4(0.5)$ 와 곱한 BM15와 BM11를 사용하였다(Robertson & Walker 1994). TREC3 이후는 BM11과 BM15를 결합한 BM25로 순위화하였다.

BM11, BM15, BM25에서 문헌의 가중치는 개개 질의용어의 가중치를 모두 합한 후 문헌길이 수정요인  $k_2 - ng \frac{avdl-dl}{avdl+dl}$  을 추가로 합한다. BM11과 BM15를 사용한 TREC2에서  $k_1$ 은 0.0에서 2.0까지는 값이 증가할수록 정확률을 증가시켰으나 2.0 이후는 값이 증가할수록 감소하였다( $k_1=2.0$ 이 최고). 문헌길이 수정요인 상수  $k_2$ 는 0.0보다는 0.5가 성능을 향상시켰고  $k_2 \geq 1$ 은 성능을 향상시키지 못하고  $k_2$ 값이 증가할수록 성능을 저하시켰다. TREC4 이후에는 BM25의 공식에서  $k_2=0$ 을 대입함으로써 문헌길이 수정요인을 무시하였다.  $k_3$ 는 무한대로 증가시킬수록 성능을 향상시키며, 긴 질문에서 더 큰 효과를 발휘하였다. BM11이 BM15보다 우수하였고, BM25를 사용하기 시작한 TREC3에서 성능이 크게 향상되었으며 TREC5-7에서 Okapi는 성능이 가장 좋은 시스템 중의 하나였다. Okapi의 정규화된  $tf$ 공식은 Inquiry(TREC4 이후), AT&T의 dnb(TREC7), IRIT, IBM, 오스트리아국립대학(ANU)의 ACSys, Waterloo대학 등의 시스템에 채택되었다.

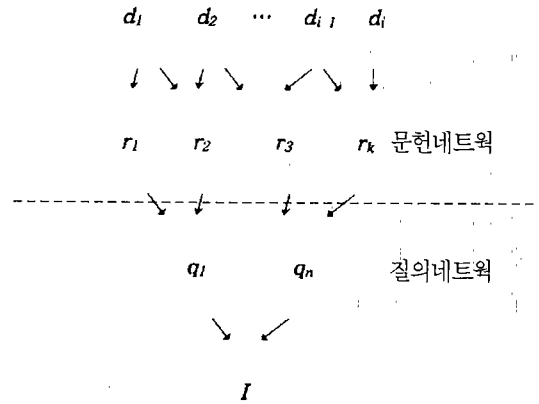
적합정보가 없을 때( $r=R=0$ )는 용어가 적합문

현에 출현할 확률  $p$ 를  $1/2$ 로 주고  $q$ 를  $n/N$ 으로 주면  $F4$ 는  $\log((N-n)/n)$ 이 됨으로써  $idf$ 와 동일한 기능을 수행하며, 이 공식은 초기질의용어의 가중치를 부여하는데 사용되었다(Croft & Harper 1979, Losee 1988 등). Okapi를 비롯하여  $F4(.5)$ 를 사용하는 대부분의시스템에서는  $\log(N-n+0.5)/(n+0.5)$ 가 사용된다. Greiff (1998)는 추론네트워크모델에서 증거가중치 (Weight of Evidence)이론으로  $idf = -\log 0(occ) = \log((N-n)/n)$ 의 공식을 이론적으로 규명하고 Inquiry에서 TREC1- TREC6 데이터를 가지고 테스트하여 그 효과를 증명하였다.

#### 4.1.2 추론네트워크 모델

Pearl(1988)의 불확실성을 기반으로한 추론네트워크모델을 정보검색에 응용한 Turtle & Croft(1991)의 추론네트워크모델은 문헌네트워크와 질의네트워크로 구성된다. <그림 2>는 기본 네트워크모델에서 텍스트노드와 질의의 개념노드를 생략한 간단한 추론네트워크이다. 문헌노드( $d_i$ )는 하나의 문헌을 나타내며, 문헌을 표현하는 하나 이상의 개념노드(concept representation node) ( $r_k$ )를 갖는다. 질의네트워크에서 정보요구( $I$ )를 표현하는데 다중질의( $q_n$ )가 사용되고, 각 질의노드( $q_n$ )는 하나 이상의 개념( $r_k$ )를 사용하여 정보요구를 표현한다.

추론네트워크에서 문헌은 정보요구( $I$ )가 문헌노드( $d_i$ )에 의해 만족될 확률(belief)로 순위화된다. 문헌노드의 확률은 네트워크를 통해 증식되어  $I$ 노드의 확률을 추론한다. 자식노드의 확률은 부모노드의 확률과 연결(link matrix)에 의해 결정된다.  $d_i$ 노드와  $r_k$ 노드 사이의 연결은 이 개념( $r_k$ )이  $d_i$ 에 출현한다는 명제에 대한 증거이고,  $r_k$ 와  $q_n$  사이의 연결은 개념  $r_k$ 가 최종확률을 계산하



<그림 2> 추론네트워크 모델

는데 결합됨을 나타낸다.

추론네트워크모델에서는 두 피드백기법이 사용된다. 첫째는 증거(evidence)를 추가하는 것으로, 네트워크에 나타난 확률분포는 변화시키지 않고 이 분포와 일치하도록 확률(belief)만 변경하는 방법으로 네트워크의 분포가 옳을 때 적절한 방법이다. 두번째 방법은 네트워크의 분포상태(Topology)와 링크를 변경하는 것으로, 초기 네트워크가 진짜분포가 아니라 의사분포로 만들어졌고 진짜분포에 관한 좋은 정보를 얻을 수 있을 때 사용하기 적절한 방법이다. 이용자의 적합성판정정보를 사용하여 질의네트워크를 변경하는 두번째 방법에서  $I$ 노드와 추가되는  $r$ 노드 사이에 링크를 추가하는 것으로 질의에 용어가 추가되고, 질의용어의 가중치는  $I$ 노드에서  $\#wsum(\text{weighed sum})$  형식을 사용하여 표현된다.

노드 $Q$ 에 부모노드 개념  $A, B, C$ 가 연결된 경우 용어  $A, B, C$ 가 각각 출현하는 문헌이 질의( $I$ )를 만족시킬 확률(belief)을 각각  $bel(A), bel(B), bel(C)$ 라고 할 때, 용어  $A, B, C$ 가 모두 출현하는 문헌이 질의에 적합할 확률(belief)은

다음과 같이 산출된다.

$$bel_{usum}(W_a \cdot bel(A), W_b \cdot bel(B), W_c \cdot bel(C)) = \frac{(W_a \cdot bel(A) + W_b \cdot bel(B) + W_c \cdot bel(C)) W_d}{W_a + W_b + W_c}$$

대표적인 추론네트워크 모델 기반시스템인 Massachusetts대학의 Inquiry에서 용어 A가 출현하는 문헌이 질의를 만족시킬 확률 bel(A)는 TREC6 이후 다음과 같이 산출된다.

$$bel(A) = 0.4 + 0.6 \times \frac{tf}{tf + 0.5 + 1.5 \cdot \frac{dl}{avdl}} \times \frac{\log\left(\frac{N+0.5}{n}\right)}{\log(N+1.0)}$$

문헌내 출현빈도는 TREC3까지 최대값정규화 방법  $\log(tf+0.5)/\log(\max tf + 1.0)$ 이 사용되었으나, TREC4에서는 Okapi의 문헌길이정규화 방법  $tf/(tf+0.5+1.5 \cdot (dl/avdl))$ 이 함께 사용되었고, TREC5 이후에는 Okapi의 방법으로 완전히 대체되었다.

TREC Routing실험에서 Inquiry는 R개의 적합문헌과 동수의 부적합문헌(NR)을 대상으로 용어가 출현하는 적합문헌수  $r \geq 5$ 인 용어만  $(r/R) - (nr/NR)$ 순으로 순위화하여 상위 n개 용어를 선택한 후 Belief를 로치오공식으로 다시 산출하였다. 로치오가중치는 DFO에 의해 다시 조정되었다. TREC4와 TREC5에서는 50개 용어와 150개 개념, TREC6에서는 20개 용어와 40개의 구가 선정됐으며, 로치오공식은  $\beta:\gamma=2:0.5$  (TREC4),  $\beta:\gamma=8:2$  (TREC5),  $\alpha:\beta:\gamma=1:4:0.5$  (TREC6)가, DFO에서는 5 패스(TREC4와 5)와 3 패스(TREC6)가 사용되었다.

#### 4.2 확장용어의 선정

확률모델에서 문헌을 순위화하는데 질의용어의

가중치공식으로 사용된 F4와 F4(.5)는 적합문헌에 출현하는 비질의용어를 선택하여 질의를 확장하기 위한 목적에서 용어선정알고리즘으로도 사용되었다. 용어선정알고리즘으로 F4(.5)는 F4(n/N)으로 수정되어 효과가 증명되었으며 (Robertson 1986), 이것은 Shaw(1995)에 의해 확인되었다. 그러나  $r=0$ 일 때는  $p=1/N^2$ 로,  $r=R$ 일 때는  $p=1-(1/N)^2$ 로 F4의 p값을 수정하는 것이 보다 우수한 것으로 보고되었다(Shaw 1995). Salton과 Buckley(1990)의 연구에서도 6개의 DB 모두에서 적합문헌에 출현하는 모든 용어를 대상으로 질의를 확장했을 때 F4(n/N)이 F4(0.5)보다 더 높은 정확률을 가져왔다. 그러나 적합문헌에 가장 많이 출현하는 상위 n개의 용어만을 대상으로 질의를 자동확장했을 때는 모든 DB에서 F4(0.5)가 정확률을 더 많이 향상시켰다.

적합문헌에 출현하는 비질의용어 중 상위 20의 용어만 선정하여 질의를 확장하는 목적에서 Harman(1988)은 Cranfield 표준실험set를 사용하여 6개 용어선정공식을 비교하였다.  $tf \cdot idf$ 와  $tf \cdot idf \cdot r$ 이  $\log(rt) \cdot idf$ 나  $r$  혹은  $idf$ 만을 사용하는 것보다 좋은 용어를 선정하였다. 1992년 연구에서 Harman(1992)은 적합문헌과 부적합문헌에 용어가 출현할 비율과 관련된 4가지 공식을 같은 목적아래 평가하였다. 20개 용어를 선정 확장하여 모두 100% 이상 정확률이 향상되었는데, 가장 좋은 공식은 F4(.5)와  $w(p-q)$ 였고, AID와 Porter공식은 상대적으로 좋지 못하였다. F4(.5)를 제외한 나머지 공식들은 1988년 연구의  $tf \cdot idf$ 보다 좋은 결과를 가져오지 못하였다. 두 실험에 사용된 용어선정알고리즘으로 20개의 용어를 선택하여 확장된 탐색은 알고리즘에 상관없이 모두 100% 이상의 정확률 향상을 보였지만, 용어는 확장하지 않고 초기질의용어의 가중치만 다시

산정하여 문헌의 순위를 다시 부여했을 때는 정확률이 23.9% 향상되었다.

Efthimiadis(1993)는 CIRT시스템에서 INSPEC 표준실험set를 사용하여 용어확장에 사용될 후보용어 25개를 선정하는데 6개의 공식을 비교하였다.  $w(p-q) \succ EMIN = Porter \succ F4(.5) = F4(n/N) \succ Zoom$  순으로 우수했으며, 상위25개의 후보용어중 이용자가 5개만을 선정했을 때 5개 용어의 평균순위는  $EMIN \succ w(p-q) \succ Porter \succ F4(.5) = F4(n/N) \succ Zoom$  순으로 우수했다. 유의수준 0.05에서 상위 다섯 개의 공식은 Zoom보다 우수하였다. r-hilo와 r-lohi 두 공식을 추가로 비교한 1995년 연구(Efthimiadis 1995)에서는 65등까지의 후보용어리스트 중에서 이용자가 선택한 5개 적합용어의 평균순위로 평가하였다.  $w(p-q) = EMIN \succ r-lohi = Porter \succ F4(n/N) \succ F4(0.5) \succ Zoom \succ r-hilo$  순으로 우수하였다. 유의수준  $\alpha=0.05$ 에서 상위 여섯 알고리즘 사이에는 차이가 없었다.

Efthimiadis와 Biron(1994)은 TREC2에서 Okapi시스템을 기반으로 5개의 용어선정 알고리즘을 테스트하였다. 확장질의로 문헌을 순위화하는데는 BM15를 사용하였다. 10개의 적합문헌정보로 질의를 확장했을 때 r-lohi와 r-hilo가 다른 알고리즘보다 우수하였다( $r-lohi \succ r-hilo \succ EMIN \succ w(p-q) \succ Porter$ ). 그러나 BM15대신 BM25를 사용한 City대학의 Okapi TREC3 실험에서 r은  $w(p-q)$ 보다 우수하지 못하였다.

City대학의 Okapi는 TREC실험에서 용어선정 알고리즘으로  $w(p-q)$ 를 사용하였다( $w=F4(.5)$ ). 그러나 TREC1에서 q값은 무시해도 좋은 것으로 간주되었고, p값 r/R은 하나의 질의에 사용되는 모든 질의용어에 대해 적합문헌수(R)가 같으므로 R변수를 무시하여  $w(p-q)=w \cdot r$ 로 수정하여 사

용하였다. TREC2에서는  $w \cdot p$ 를 사용하였다. TREC3에서는 p값을 모든 적합문헌의  $tf/R$ 를 합한 값 혹은  $rtf/R$ 로 바꾸었다. 확장에 사용되는 용어수는 30개가 20개보다 좋은 결과를 가져왔다. TREC4와 5에서는 피드백탐색을 반복하여 수행할 때 한번 질의용어로 추가된 용어는 계속 질의용어로 사용하는 대신 성능을 상승시키지 못하면(적합문헌을 검색하지 못하면) 질의용어에서 제거하는 방법을 연구하였다. TREC6에서는  $w(p-q)=w \cdot r$ 이 다시 사용되었다.

Haines와 Croft(1993)는 추론네트워크모델에서 용어선정알고리즘 EMIM, PMIM, P-4, idf, rdfidf, rtf, rtfidf의 효과와, 선정된 용어에 다시 가중치를 부여하는 방법 rtf와 rtfidf의 효과를 초기질의용어와 추가된 용어의 상대가중치를 50 : 50에서 90 : 10까지 변화를 주어 비교 테스트하였다. 용어선정알고리즘은 초록문(CACM) DB에서는 idf, EMIM, rdfidf, rtfidf가 rtf나 p-4보다 우수하였으나, 본문(Full text)탐색에서는 p-4, rdfidf, EMIM, PMIM이 우수하였다. 가중치를 다시 부여하는 것은 본문보다는(rtfidf는 +30.6%, rtf는 +34.4% 향상) 초록문에서(rtfidf는 +105.4%, rtf는 +83.3% 향상) 매우 효과적이었다. 초기질의용어와 추가된 용어에 부여하는 상대가중치는 50 : 50(본문) 또는 35 : 65(초록)로 하였을 때 가장 우수하였다.

## 5 다중질의 결합

다중질의 결합이란 같은 정보요구에 대해 다른 탐색자에 의해 표현되는 탐색식의 차이에 따라, 혹은 문헌을 대표하는 표현방법(서명, 초록, 통제어)에 따라 상이한 적합문헌이 검색된다는 연구

〈표 3〉 확률모델에서 질의응어의 적합성 가중치

R : 질의에 대한 적합문헌 총수  
 r : 질의용어 t가 출현하는 적합문헌수  
 N : DB내 문헌총수  
 n : 질의용어 t가 출현하는 문헌수

문헌순위화

- (1)  $\log \frac{r/R}{n/N}$  F1(Robertson & Sparck Jones 1976)
- (2)  $\log \frac{(r(R-r))}{(n-r)(N-n-R+r)} = \frac{r(N-n-R+r)}{(n-r)(R-r)}$  F4( " )
- (3)  $\log \frac{(r+0.5)(N-n-R-r+0.5)}{(n-r+0.5)(R-r+0.5)}$  F4(.5) (Okapi TREC1)
- (4)  $\log \frac{(r+\frac{n}{N})(N-n-R+r-\frac{n}{N}+1)}{(n-r+\frac{n}{N})(R-r+\frac{n}{N}+1)}$  F4(n/N) (Robertson 1986)
- (5)  $t \cdot F4$  (Salton & Waldstein 1979)
- (6)  $t^f \cdot \log \frac{(1+f)(N+r)(1+R)}{(f-r)(N-R)}$  TPI(Losee 1988, 문성빈 1993)
- (7)  $F4(.5) \cdot S_1 S_3 \cdot \frac{tf}{k_1 \cdot \frac{dl}{avdl} + t^f} \cdot \frac{atf}{k_3 + atf}$  BM11(Okapi TREC2)
- (8)  $F4(.5) \cdot S_1 S_3 \cdot \frac{tf}{(k_1 + t^f)} \cdot \frac{atf}{k_3 + atf}$  BM15(Okapi TREC2)
- (9)  $F4(.5) \cdot S_1 S_3 \cdot \frac{(k_1+1)t^f}{k_1((1-b) + b \frac{dl}{avdl}) + t^f} \cdot \frac{(k_3+1)atf}{k_3 + atf}$  BM25(Okapi TREC3 이후)
- (10)  $\frac{tf}{2 \times (0.25 + 0.75 \times \frac{dl}{avdl}) + t^f} \times \log \frac{N-n+0.5}{n+0.5}$  ANU/ACSys(Hawking et al. 1998)
- (11)  $bel(A) = 0.4 + 0.6 \times \frac{tf}{tf + 0.5 + 1.5 \frac{dl}{avdl}} \times \frac{\log(\frac{N+0.5}{n})}{\log(N+1.0)}$  Inquiry

용어선정

- (12)  $\log \frac{(r+0.5)(N-n-R+r+0.5)}{(n-r+0.5)(R-r+0.5)} \cdot (\frac{r}{R} - \frac{n-r}{N-R})$  w(p-q) (Robertson 1990)
- (13)  $\frac{r}{R} - \frac{n}{N}$  Porter(Porter & Galpin 1988)
- (14)  $\log \frac{rN}{Rn} \cdot r - \log \frac{(n-r)N}{(N-R)n} \cdot (n-r) - \log \frac{(R-r)N}{(N-n)R} \cdot (R-r)$   
 $+ \log \frac{(N-n-R+r)N}{(N-n)(N-R)} \cdot (N-n-R+r)$  EMIN(van Rijsbergen 1977)
- (15) r ( r이 같은 경우엔 용어빈도수가 낮은 것부터 높은 순으로)
- (16) r ( r이 같은 경우엔 용어빈도수가 높은 것부터 낮은 순으로)
- (17)  $\frac{RT \times n}{N}$  (RT: 질의어로 검색된 문헌수) AID(Doszkocs 1978)
- (18)  $rt^f$  (적합문헌 집단에서 용어 k의 빈도수) ESA/IRS Zoom(Haines & Croft 1993)
- (19)  $t^f \cdot id^f \cdot r$  (Harman 1992)
- (20)  $\log(rt^f) \cdot t^f$  ( " )
- (21)  $r \cdot id^f$  rdfidf(Haines & Croft 1993)
- (22)  $r \cdot \frac{\log(N+0.5)/n}{\log(N+1)}$  rdfidf(수정)(Allan 1995)
- (23)  $rt^f \cdot id^f$  rtfidf(Haines & Croft 1993)

결과에 근거하여 하나의 표현으로 검색되는 문헌의 결과를 다른 표현으로 검색되는 문헌결과와 결합하여, 문헌의 순위를 재배열하는 것이다.

추론네트워크모델 Inquiry을 기반으로 Turtle & Croft(1991)는 자연어질의 표현과 불질의 표현의 검색효과를 결합하여, Belkin et al.(1993)은 TREC2에서 5개의 불논리질의 표현 탐색을 연속적으로 결합하여, 검색효과를 향상시킬 수 있음을 보고하였다. 그러나 Belkin 등의 연구에서 5개의 불논리질의 결합은 자연어 Inquiry탐색보다 우수하지 못하였고, Inquiry자연어질의탐색에 불논리질을 결합하여도 Inquiry자연어탐색을 향상시키지 못하였다(Rutgers의 TREC1). TREC2에서 5개의 불논리질을 Inquiry의 자연어질의로 변환한 후 결합할 경우엔 2개 혹은 3개 질의결합이 5개 모두를 결합하는 것보다 우수하였다(Belkin et al. 1995).

Inquiry TREC5에서는 Description필드만 사용하여 기본 처리한 질의와 핵심(Core)처리질의, LCA처리질을 결합하여, Narrative필드만 사용한 경우에는 기본처리질의와 LCA처리질의만을 결합하여 성능을 향상시켰다.

벡터공간모델 SMART를 기반으로한 질의결합 실험(Virginia Tech.)에서 3개의 P-norm탐색은 2개의 질의벡터(긴 질의, 짧은 질의)보다 우수하였다(P-norm (1.5=2.0)>1.0)>긴 질의벡터>짧은 질의벡터). 두 벡터를 결합하는 것은 다른 어떤 두 질의의 결과를 결합하는 것보다 좋지 못하였고, 다섯 질의의 결과를 모두 결합하는 것이 결합하지 않은 다섯 결과보다 우수하였다. 6가지 결합방법 중에는 각각의 가중치를 합하는 Combsum방식이 가장 좋았다(Fox & Shaw 1994). TREC3에서는 다섯 탐색결과를 모두 결합하는 것보다는 하나의 질의벡터와 하나의 P-norm을 결합하는 것이 더욱

좋았다(Shaw & Fox 1995).

이준호(1995)는 SMART를 사용하여 6가지 가중치의 벡터매칭결과를 2개씩 결합하여 검색효과를 향상시켰다. 코사인정규화를 사용한 매칭의 결과를 코사인정규화를 사용하지 않은 가중치의 매칭결과와 결합할 때 가장 성능이 향상되었다.

벡터모델과 확률모델의 검색결과를 결합한 연구(Savoy et al. 1996)에서 Okapi의 BM25탐색은 6종류의 벡터매칭기법 어느 것보다 우수하였으며, 6개의 벡터모델의 탐색결과를 결합한 결과를 다시 Okapi와 결합시킬 경우 Okapi의 성능보다 낮은 결과를 가져왔다. 벡터매칭기법 중에서는 Inc.ltc가 가장 우수하였다.

RMIT의 TREC4에서는 이진조합매칭(CDM)과, idf로 가중치를 준 조합매칭(WCM), 코사인 유사도(COS), Okapi 변형유사도로 검색한 결과 긴 질의에서는 COS>Okapi>WCM>CDM, 짧은 질의에서는 Okapi>COS>WCM>CDM의 성능을 얻었다. 긴 질의에서 Okapi 와 COS를 결합하여 가장 좋은 향상을 보았으나, 짧은 질의에서는 어떤 결합도 Okapi보다 좋지 못하였다(Wilkinson et al. 1996).

Waterloo대학은 질의용어 몇 개만 사용하여 자동생성한 질의(cover density ranking기법)과 Okapi 초기탐색과 피드백탐색을 수정한 각각의 탐색결과를 결합하여 성능을 향상시켰다(TREC6).

다중질의결합기법을 적합성피드백에 응용한 이기호 등의 연구(1997)에서는 SMART 환경에서 하나의 질의에 대해 초기탐색으로 검색된 상위 30개의 문헌을 적합문헌으로 간주하여 2개의 벡터질의확장기법과 3개의 확률모델질의확장기법으로 5개의 확장된 피드백질의를 생성한 후, 각 질의벡터의 가중치를 합하는 방법으로 결합된 질의 벡터를 만들어 검색을 수행하였다. 5개 피드백방

법을 각각 독립적으로 검색하여 검색된 200개 문헌을 대상으로 평균정확률을 산출했을 때 Ide dec-hi > Rocchio > F4(n/N) > F4(.5) > Fuhr 확률공식(Fuhr & Buckley 1991) 순으로 정확률이 향상되었다. 5개의 피드백벡터를 2단계에서는 두개씩, 3단계에서는 3개씩, 4단계에서는 4개씩, 5단계에서는 5개 모두를 결합하여 질의벡터를 생성한 후 검색을 실행한 결과, 평균검색효과는 단계가 증가함에 따라 약간씩 증가하였으나, Rocchio를 F4(.5)나 F4(n/N)과 결합했을 때 가장 높은 정확률을 얻었다(+26.1% 향상).

## 6 의사적합성피드백탐색

Alta Vista 탐색엔진(의 refine)과 같은 시스템은 이용자의 적합성판정을 요구하지 않고 추가할 후보용어를 가중치와 함께 보여주고, 이용자가 추가할 탐색어를 선택하도록 하는 대화형질의확장(IQE) 시스템이다. 이용자의 적합성판정을 요구하지 않는 대신 초기질문으로 검색된 상위 n개의 문헌을 적합문헌으로 간주한다. 순위화된 검색결과에서 상위 n개의 문헌을 적합문헌으로 간주하는 의사적합성피드백(Pseudo relevance feedback 혹은 blind query expansion)은 TREC3의 특별실험(ad-hoc run)에서 공식 사용된 이후 TREC7에 이르기까지 적합성판정정보가 없는 상황(ad-hoc run)에서 대부분의 시스템이 보편적으로 사용한 방법이다. 이 방법은 작은 DB에서보다는 TREC과 같이 큰 DB를 사용한 환경에서 매우 효과적이었다.

Croft와 Harper(1979)의 연구에서 초기질의에 사용된 용어만을 대상으로 가중치를 다시 계산할 경우( $\log(N-n)/n$  사용), 상위 5개 문헌을 적합문

헌으로 간주했을 때 가장 탐색이 향상되었다.

벡터공간모델에서 SMART는 TREC3에서는 상위 30개의 문헌에서 500개의 단어와 10개의 구를, TREC4에서는 상위 20개의 문헌에서 50개의 단어와 10개의 구를 선정하였으나, TREC5 이후는 보다 정교한 QZ(Query Zone)를 사용하여 질의를 확장하였다. 상위 1,000개의 문헌에 출현하는 질의용어의 동시출현빈도를 이용하여 상위 50개의 문헌을 재순위화한 후 상위 20개의 문헌을 적합문헌으로 간주하였다. TREC6과 TREC7에서는 상위 30개의 문헌을 클러스터링하여 초기질의와 가장 적합한 두 개의 클러스터를 선택하고, 로치오(8, 8, 8)로 가중치를 산정한 후, 100개의 단어와 20개의 구(TREC6), 25개의 단어와 5개의 구(TREC7)를 선택하였다. TREC4까지는 적합문헌정보만으로( $\gamma=0$ ) 적합문헌에 출현하는 용어의 순위를 결정하였으나, TREC5이후에는 501~1000등의 문헌을 부적합문헌으로 간주하여 적합문헌에 출현하는 용어의 적합성가중치를 계산하는데 사용하였다.

Mitra et al.(1998)은 벡터모델에서 초기탐색(Lnu,ltu)으로 1,000개의 문헌을 검색하여 상위 20개 문헌은 적합문헌으로, 501-1,000번째 문헌은 부적합문헌으로 간주하여, 로치오공식으로 적합문헌에 출현한 용어 중 25개의 단일어와 5개의 복합어만 추가하여 질의를 확장했을 때, TREC6만 제외하고 TREC3-5 데이터에서 정확률이 21.2%-39.1% 향상되었음을 보고하였다. 이 질의확장검색은 초기탐색으로 검색된 t개의 문헌을 문헌에 출현하는 질의용어 가중치(IDF)의 합으로 다시 순위를 부여하는 필터링을 거친 후 상위 20개의 문헌을 적합문헌으로 간주하여 Rocchio 적합성피드백탐색을 수행하는 필터링 피드백탐색과 비교되었다(w단어의 문단(passage)제한 때



칭). 필터링으로 피드백탐색은 3.2%(TREC4)~20.3%(TREC6) 향상되었다. 가장 성능을 향상시키는 w값과 t값은 TREC 데이터에 따라 달랐다. 이 필터링피드백탐색은 퍼지블논리필터링을 이용한 피드백탐색과 겨룰만한 성능을 보였다.

SMART를 기반으로한 AT&T(TREC6) 또한 Lnu, ltu로 문헌을 1000등까지 검색하여 검색된 문헌 내에서 질의용어의 중요도를 다시 계산하여 상위 50개의 문헌을 다시 순위화한 후(필터링) 상위 20개의 문헌을 적합문헌으로 간주하였다. 501~1000등의 문헌을 부적합문헌으로 간주하여 로치오(8, 8, 8)로 25개의 단어와 5개의 구를 선택하여 질의를 확장하였다.

TREC7에서 AT&T는 짧은 질의에서 tf의 영향을 줄이기 위해 용어가중치를 dnb, dtn으로 수정하여 1000개의 문헌을 순위화하였으며, 상위 10개의 문헌을 적합문헌으로, 501~1000등의 문헌을 부적합문헌으로 간주하여 로치오(3, 2, 2) 공식으로 상위 20개의 단어와 5개의 구를 추가하였다. 확장용어를 선택하기 위해 로치오로 적합, 부적합문헌의 가중치를 계산할 때는 dtb가 사용되었고, 확장된 질의로 문헌을 최종 검색할 때는 dnb를 사용하여 1000개의 문헌을 재순위화하였다. 코넬(TREC5)에서 부적합문헌정보는 검색효과에 영향을 미치지 못했으나, AT&T(TREC6)에서는 상위 20개의 적합문헌정보만으로 질의를 확장하여 정확율이 6.3% 향상되었으나 부적합문헌정보까지 사용하였을 때는 14.1% 향상되었다. TREC7에서 AT&T의 특별실험은 코넬의 것보다 우수하였다.

확률모델에서 Okapi는 TREC3에서는 상위30개 문헌에서 상위40개 용어를, TREC4에서는 50개 문헌의 첫 끝문단에서 20개 용어를, TREC5에서는 15개 문헌에서 30개 용어(긴 질의에서는 50개 문헌에서 55개 용어)를 선정하였으나, TREC6과 TREC7에서는 비교적 작은수의 문헌과 용어를 사용하였다(짧은 질의는 7개 문헌 20개 용어, 긴 질의는 15개 문헌 30개 용어): 문헌수를 20-100으로, 용어수를 2-50으로 변화를 주어 실험한 TREC3에서 상위30개 문헌에서 40개 용어를 선택하여 질의를 확장할 때 검색이 가장 향상되었다.

용어선정알고리즘 w(p-q)는 특별실험 TREC6과 TREC7에서는 TSV로 수정되어 사용되었다. w(p-q)에서 q를 무시하는 대신 알고있는(검색된) 부적합문헌수S와 알고 있는 부적합문헌 중 질의용어가 출현하는 문헌수 s를 사용한 q값이 사용되었다(TSV=w · (r/R - α · s/S), (0<α<1)). TSV에서 w값은 F4(5) 대신 S와 s를 사용하는 아래의 공식을 사용한다. 상위R개의 적합문헌 다음 500개의 문헌을 건너뛰어 그 다음 500개의 문헌을 부적합문헌(S)로 간주하였다. 피드백탐색에서 문헌의 가중치를 계산할 때는 초기질의용어에 보너스(×2.5)를 주었다.

Inquery는 TREC4까지는 Phrase Finder를 사용하여 질의용어와 가장 밀접한 개념을 검색하여 질의를 확장하였으나, TREC5 이후에는 LCA(Local Context Analysis) 기법으로 상위 n개의 문헌에 출현하는 상위 n개의 구문에서 m개의 개념을 선택하여 질의를 확장하였다.

$$w = \frac{K_5}{K_5 + \sqrt{R}} \left( K_4 + \log \frac{N}{N-n} \right) + \frac{\sqrt{R}}{K_5 + \sqrt{R}} \cdot \log \frac{r+0.5}{R-r+0.5} - \left( \frac{K_6}{K_6 + \sqrt{S}} \cdot \log \frac{n}{N-n} + \frac{\sqrt{S}}{K_6 + \sqrt{S}} \cdot \log \frac{s+0.5}{S-s+0.5} \right)$$

$bel_{wsum}(Wa\ bel(A), Wa\ bel(B), \dots Wm\ bel(m))$ 을 계산하는데 추가되는 각각의 부모노드 A, B,  $\dots M$ 이  $wsum$ 에 공헌하는 각각의 상대 공헌도  $W_a, W_b, \dots W_{m-1}, W_m$ 은 다음과 같이 산출된다.

$$W_m = 1.0 - 0.9 \times (m - 1)/m$$

Inquery는 TREC5에서 LCA기법으로 300단 어크기의 문단 1000개에서 70개의 개념을, TREC6과 TREC7에서는 문단 30개에서 50개의 개념을 선택하였다. 이 방법은 상위  $n$ 개 문헌에 출현하는 용어의 가중치를 Rocchio(1,1,0) 공식으로 수정하여 질의확장하는 방법이나 Phrase Finder로 확장하는 방법보다 우수하였다(Xu & Croft 1996). Xu와 Croft의 연구(1996)는 코넬 연구에서와 같이 로치오기법으로 상위50개의 단어와 10개의 구로 자동확장할 경우, TREC3데이터에서는 10개의 문헌을 사용했을 때(+20.5%), TREC4데이터에서는 5개 문헌을 사용했을 때(+14.0%) 가장 좋은 결과를 얻었다. Inquery의 문단을 사용한 LCA자동확장은 TREC3데이터에서는 상위100개 문단을 사용하여 +23.5%, TREC4데이터에서는 상위200개 문단을 사용하여 +24.4% 향상시킴으로써 SMART의 문헌을 사용한 자동확장보다 우수하였다.

Kwok(1996)은 PIRCS(Probabilistic Indexing & Retrieval Component System)에서 TREC2와 TREC4의 짧은 질의(평균 6-8개 질의용어)를 탐색할 때 RSV(Retrieval Status Value)에 의해 순위화된 초기탐색의 결과에서 상위 30개의 문헌을 적합문헌으로 간주하여 질의를 자동확장한 검색이 적합성판단정보로 수동 확장하는 것과 차이가 없음을 보고하고 있다.  $avtf = (tfc/n) / \log \max[cutoff, n]$ 의 공식으로 초기질의용어의 가중치만 수정했을 때와 초기질

의용어에 보다 높은 가중치를 주고 적합문헌에 출현한 용어 중 가장 높은 가중치를 갖는 비질의 용어를 최고 3개까지만 추가하여 질의를 확장했을 때가 테스트되었다. avtf공식은  $\alpha=1.5$ 일 때 가장 검색효과가 우수하였다.

또한 1998년 연구(Kwok & Chan 1998)에서는 이 avtf를 응용하여 검색효율을 더욱 향상시키는 방법을 연구하였다. TREC5와 TREC6 데이터를 사용하여 고빈도 단어는 제외시키고, 적합문헌으로 간주된 상위  $n$ 개의 문헌내의 용어빈도수와 역문헌빈도를 기반으로한 단어분리  $\log(1-qk)/qk$  ( $qk$ 는  $tfc/DB$ 내 용어총수)를 avtf와 결합시켰을 때 avtf보다 정확률이 29%(TREC5), 12.5%(TREC6) 향상되었다.

의사적합성피드백을 사용한 자동질의확장과 이용자의 적합성판정을 통한 질의확장의 성능에 대해서는 코넬에서 관심을 갖고 연구하였다. SMART TREC4와 TREC5에서 이용자의 적합성판정으로 질의확장하는 것이 의사적합성피드백보다 +7.7%(TREC4), +15%(TREC5) 우수하였다. 검색된 문헌을 10개와 15개로 각각 제한한 TREC6과 TREC7에서 이용자의 적합성판정으로 질의를 확장하는 것이 의사적합성피드백에 비해 3명의 이용자 평균 +34.7%(TREC6)와 +22.3%(TREC7) 더 우수하였다.

## 7 적합성판정정보

### 7.1 배움표본(learning Sample) 크기와 내용

질문1 : 질의에 대한 DB내 전체 적합정보 대신 초기질의로 검색된  $n$ 개의 문헌에 대한 적합성 정보를 사용한 적합성피드백 질의확장 탐색은 탐

색성능을 향상시키는가?

Sparck Jones(1979a)는 검색되지 않은 적합 문헌에 대한 정보를 알 수 없는 실제검색시스템에서 확률공식 F1과 F4에  $c=0.5$ 를 추가함으로써 초기검색의 결과로 얻는 적합문헌만으로도 검색효과를 향상시킬 수 있음을 보고하였다. 또 Sparck Jones(1979b)는 1-3개의 적합문헌정보만으로 피드백 탐색을 수행하였을 경우와 DB내 모든 적합문헌정보를 가지고 피드백탐색을 수행하였을 경우를 비교하였다. Cranfield 표준set에서는 하나의 적합정보만 사용했을 경우( $R=1$ )와 2개의 적합정보를 사용했을 경우( $R=2$ ), 그리고 모든 적합정보를 사용하여 탐색을 확장시켰을 때 향상된 효과사이에는 차이가 없었다. UKCIS 표준set에서는  $R=1$ 과  $R=2$ ,  $R=3$  간에는 차이가 없었으나  $R=$ 모든 적합문헌 만큼 우수하지는 못하였다.

Okapi의 TREC3 특별실험에서 상위n개의 문헌을 적합문헌으로 간주하는 의사적합성피드백탐색(ad-hoc)이 공식적합성정보를 모두 사용하는 피드백확장(Routing)보다 우수하지는 못하였지만, TREC특별실험 전반에 걸쳐 의사적합성피드백은 매우 효과적이었다.

질문2 : 질의를 확장하여 문헌의 순위를 다시 부여하는데 적합문헌만 필요한가 부적합정보도 필요한가?

먼저 벡터공간모델을 기반으로 한 시스템에서 상위 15개 문헌만을 대상으로 적합성을 판정한 Salton과 Buckley(1990)의 연구는 3개의 벡터질의확장기법중 적합문헌은 모두 포함하나 부적합문헌은 하나로 제한하여 질의어를 확장하는 Ide dec-hi가, 적합문헌과 부적합문헌 모두를 사용하는 Ide regular나 적합문헌과 부적합문헌의 비율을 3:1로 사용한 Rocchio방식보다 우수함을

보고하고 있다. Rocchio공식에서는  $\beta=0.75$ ,  $\gamma=0.25$ 가  $\beta=\gamma=0.5$ 나  $\beta=1$ ,  $\gamma=0$ 보다 우수했다.

Aalbersberg(1992) 또한 15개의 배움표본으로  $\beta$ 와  $\gamma$ 값을 0.1~1.0로 변화를 주면서 3개의 벡터질의확장기법을 비교하였다. 세가지 벡터기법 모두에서  $\beta=0$ ,  $\gamma=1$ 일 때 정확률은 가장 낮았다. 그러나  $\beta=1$ ,  $\gamma=0$ 일 때 가장 좋은 결과가 나오지는 않았다. 즉 부적합문헌만 가지고 피드백확장을 했을 때 가장 나쁜 결과가 나왔으나, 적합문헌만을 대상으로 탐색을 확장했을 때 가장 좋은 결과를 얻지 못하였다.

SMART(코넬)은 TREC 특별탐색에서 초기에는 적합문헌만 사용하였으나(TREC3과 TREC4에서  $\gamma=0$ ), TREC5부터는 501~1000 등의 문헌을 부적합문헌으로 간주하여 확장용어의 가중치(로치오)를 산출하였다. 이 두 방법의 우열은 AT&T의 TREC보고서에 보고되었다. SMART를 기반으로한 AT&T의 TREC3-6 의사적합성피드백탐색에서 적합문헌(상위 20개)만 사용하는 것보다는 부적합문헌(501~1000등 문헌)을 추가하는 것이 검색효과를 향상시켰다.

확률모델을 사용한 Okapi와 Inquiry에서도 부적합문헌정보를 적합문헌정보와 함께 사용하는 것이 검색효과를 더 향상시켰다. Okapi는 TREC6 이후 Robertson의 용어선정알고리즘  $w(p-q)$ 를 TSV로 바꾸어 500+R등의 문헌에서부터 500개의 문헌을 부적합문헌으로 간주하여 사용하고 있다.

Inquiry에서도 TREC3에서 로치오공식으로 용어의 Belief값을 계산할 때 검색된 적합문헌 수와 동수의 부적합문헌을 사용하였다. Inquiry의 TREC4에서는 용어선정공식으로 부적합정보 S와 s를 사용한 선정공식을 사용하였다. Inquiry를 기반으로한 Rutgers대학의 TREC6과

TREC7 실험에서는 로치오공식을 사용하여 용어의 Belief값을 재산정할 때 TREC6에서는 부적합 문헌까지 사용하는 것이, TREC7에서는 적합 문헌만 사용하는 것이 조금 우수하였으나 통계적으로 유의한 수준은 아니었다.

질문3 : 대화형피드백시스템에서 검색된 문헌을 하나씩 화면에 전시하여 적합성을 판정하게 하는 것이 좋은가 아니면 여러 문헌을 한꺼번에 전시하여 여러 개의 적합성판정정보로 적합성피드백 탐색을 수행하는 것이 좋은가? 복수의 배움표본이 필요하다면 몇 개의 적합문헌과 부적합문헌을 사용했을 때 가장 좋은 검색효과를 볼 수 있는가?

탐색자가 적합성을 판정하지 않고 상위 n개의 문헌을 적합문헌으로 간주하는 시스템에서, Croft & Harper(1979)는 적합문헌에 출현하는 비질의 용어를 질의어에 추가할 목적에서는 상위 3개 문헌을 배움표본으로 사용하는 것이 상위 5개를 사용하는 것보다 좋았다고 밝히고 있다.

Robertson(1986)은 검색된 R개의 적합문헌만을 배움표본으로 사용하여 검색된 문헌에 출현하는 모든 용어로 확장된 질의어를 사용할 때, 적합문헌수가 1개 일 때보다는 2개 일 때 더 높은 검색효과를 가져옴을 보고하였다.

Ide의 벡터적합성피드백 용어확장을 사용한 SMART시스템에서 적합문헌에 출현하는 상위 30개 용어를 용어확장에 사용하는 환경에서는 5개의 적합문헌수 보다는 10개의 적합문헌수가 피드백탐색 성능을 향상시켰다(Hersh et al. 1994). TREC 데이터와 Rocchio의 벡터확장을 사용한 SMART 테스트에서 배움표본에 사용되는 적합문헌수가 많으면 많을수록 검색효율은 증가하였다(Buckley et al. 1994). 같은 시스템을 사용한 Allan(1996)의 연구에서는 모든 적합문

헌을 사용하여 질의를 확장했을 경우 정확률이 80.9% 증가하였으나, 10%의 적합문헌만 사용하여도 76.1% 증가하였다. 10%의 적합문헌 정보를 n개씩(2, 4, 6, 8, 16, 32, 62개씩) 추가하면서 탐색을 확장할 경우에 반복할수록 정확률은 증가하였다.

Aalbersberg(1992)는 검색된 문헌을 하나씩 전시하여 적합문헌이면 확장을 반복하는 시스템과 15개의 검색된 문헌에 대해 3가지 벡터기법으로 탐색을 확장하는 시스템을 비교하였다. 15개의 검색된 문헌에 대한 적합성 정보로 질의를 확장하는 3가지 벡터기법으로는 확장하기 이전보다 검색효율이 약 18% 증가되었으나, 이전벡터에 추가되는 새문헌의 벡터를 결합하여 1개씩 확장을 반복하는 기법으로는 32% 증가하였다.

Allan(1995)은 검색된 상위 n개의 문헌 중 적합문헌에 출현한 용어를  $t(=\min(3+2r, 300))$ 개 선택하여 질의를 확장(확장된 용어의 가중치는  $S_k \times tf \times \frac{\log((N+0.5)/n)}{\log(N+1)}$ ,  $S_k$ 는 k가 초기질의 용어이면 1, 추가된 용어이면 0.3)하여 검색된 문헌이 너무 많으면 짧은 문헌은 전문(full text)을 제공하나 긴 문헌은 문단(passage)으로 제공하는 시스템을 제안하고, 검색된 문헌수는 40개, 문단의 크기는 300일 때가 가장 효과적이었다고 보고하고 있다.

## 7.2 적합성판정척도

이용자가 적합성을 판정하는 시스템에서 적합성판정에는 2진척도(Okapi) 또는 3진척도(SMART, Waterloo)가 사용되고 있다. SMART는 TREC7의 IHP실험(Interactive High Precision Treck)에서 3진척도를 사용하지만 "possibly relevant"로 판정된 문헌은 피드백에

사용하지 않았다. 적합성판정척도를 연구변수로 삼아 판정척도가 피드백탐색에 끼친 영향을 분석한 연구는 거의 없다. 다만 Shaw(1995)의 연구에서 3개의 수정된 BI모델(F4)에서 적합성판정척도가 질의확장탐색에 끼친 영향이 분석되었다. 3간척도(매우 적합, 약간 적합, 부적합)에서 “매우 적합”과 “약간 적합”을 적합으로 간주했을 때와 “매우 적합”만을 적합문헌으로 간주했을 때를 비교하였다. 초기질의어에 사용된 질의용어만을 대상으로 가중치만 다시 계산하고 용어는 확장하지 않을 경우엔 3개의 수정된 F4 모두에서 “매우 적합”만을 적합문헌으로 간주했을 때 더 좋은 결과를 얻었다. 그러나 문헌에 출현하는 모든 비질의용어까지 포함하여 탐색을 확장했을 때는 F4(C=0.5)를 제외한 다른 두 F4모델에서 “매우 적합”만을 적합문헌으로 간주했을 때 더 좋은 결과를 얻었다.

## 8 탐색용어의 수정 및 확장 (AQM : AQE : IQE)

지금까지 수행된 적합성피드백에 의해 탐색 질의를 수정 및 확장하는 방법은 다음 4가지로 구분될 수 있다.

1. 초기질의어에 사용된 용어만 그대로 사용(용어의 적합성가중치만 재산출하고 질의어 확장은 안함) (Automatic Query Modification : AQM)
2. 적합문헌에 출현하는 모든 용어를 사용하여 자동질의확장(AQE-A)
3. 적합문헌에 출현한 용어를 가중치순으로 순위부여하여 상위 n개의 용어만 사용하여 자동질의확장(AQE-S)
4. 가중치순으로 화면에 전시된 적합문헌에 출

현한 용어 중 이용자가 선택하여 질의확장(Interactive Query Expansion : IQE) :

질문1 : 질의수정(AQM)이 좋은가, 질의확장(AQE-A와 AQE-S)이 좋은가, 확장하는 용어의 크기는 어느 정도가 좋은가?

Wu와 Salton(1981)은 AQM과 AQE-A를 다시 세분하여 다음 4가지 방법을 단어의 가중치 부여방법(2가지)과 DB(2가지)를 달리하여 비교하였다. 첫번째 AQM은 초기탐색에서 사용된 가중치는 무시하고 적합성피드백으로 재산출된 가중치만 사용하는 것이고(AQM-1), 두번째 AQM은 초기탐색때의 가중치에 적합성피드백으로 재산출된 가중치를 합한 것이다(AQM-2). 첫번째 AQE-A는 적합문헌에 출현하는 모든 단어에 대해 재산출된 가중치만 부여하는 것이고(AQE-A1), 두번째 AQE-A에서는 모든 단어에 초기탐색의 가중치와 재산출된 가중치를 합하였다(AQE-A2). Cranfield 표준set로 초기탐색에서 IDF가중치를 사용했을 때 검색효과는 AQE-A2) AQM-1) AQM-2) AQE-A1 순으로 우수했고, IRW가중치를 사용했을 때는 AQM-1) AQE-A2) AQM-2) AQE-A1 순이었다. Medlars 표준set에서는 IDF와 IRW모두에서 AQE-A2) AQE-A1) AQM-1) AQM-2 순이었다(AQE-A)AQM). 두개의 AQE-A 방법은 모두 적합성판정을 위해 검색된 문헌(배움표본)을 10개에서 15개로 증가시켰을 때 다른 두 방법보다 검색효과를 증가시켰다.

MEDLINE의 CF DB에서 실험된 Shaw(1995)의 연구에서도 AQE-A가 AQM보다 우수하였다.

Harman(1992)은 Cranfield 표준set를 사용하여 AQM을 상위 20개의 용어로 탐색을 확장하는 AQE-S와 비교하였다. AQM으로는 초기탐색

에 비해 정확률이 23.9% 향상되었으나, AQE-S로는 112.2% 향상되었다. AQE-S에서 용어수를 10, 30, 40으로 변화시켰을 때는  $n=30 > 10 > 40$  순으로 저하되었다. 그러나 가장 나쁜  $n=40$ 일 때도 초기탐색에 비해 102.3%의 높은 향상을 보였다.

Haines와 Croft(1993)는 적합문헌정보만 사용하여( $\beta=1, \gamma=0$ ) Rocchio의 벡터공식으로 질의를 확장할 때(AQE-S) 추가되는 용어수와 그 비율이 초록과 본문(fulltext) 검색효과에 끼친 영향을 AQM과 비교 연구하였다. 용어선정알고리즘에 따라 차이는 있지만 초록문(CACM)에서는 확장된 용어수들에서  $n=1$ 에서 100으로 증가시켜도 AQM에 비해 여전히 평균 100%의 향상을 보였다. 7개 알고리즘의 평균 피드백질의확장은 본문보다는 초록문 탐색에서 그 효과가 월등하였다. 초록에서는 초기질의 용어와 추가용어를 65 : 35의 비율로 했을 때 효과가 가장 우수하였으며(+99.3%), 본문(WEST)에서는 50 : 50의 비율로 새 질문벡터를 작성할 때 가장 효과가 우수하였다(+33%).

TREC2에서 코넬의 SMART는(Buckley et al. 1995) Rocchio의 알고리즘으로 질의용어의 가중치만 재계산했을 경우(AQM) 초기탐색에 비해 7%의 향상을 보였으나, 상위 20개의 단어와 10개의 구로 질의어를 확장했을 경우엔(AQE-S) 17%의 향상을 보였고, 300개의 단어와 50개의 구로 확장했을 경우엔(AQE-S) 38%의 향상을 보였다(AQE-S>AQM). 같은 연구환경에서(Rocchio  $\alpha, \beta, \gamma=8, 16, 4$ ) 확장하는 용어를 4000개까지 증가시켰을 때 확장되는 질의용어가 많으면 많을수록 검색효과는 증가한다는 피드백 탐색의 효과와 확장되는 용어수와의 선형관

et al. 1994).

Salton과 Buckley(1990) 연구에서 AQE-A가 AQE-S보다  $F4(0.5)$ 를 제외한 모든 질의확장방법(Ide, Ide dec-hi, Rocchio,  $F4(n/N)$ )에서 좋은 탐색효과를 보였으나, 그 차이는 근소한 것이므로 축적용량과 처리시간을 고려한다면 AQE-S를 사용하는 것이 타당하다고 권고하고 있다.

확률모델에서는 Salton과 Buckley(1990) 연구에서 언급한 바와 같이 AQE-S가 AQE-A보다 우수하였고( $F4(.5)$ 사용), BM11를 사용한 경우 Okapi TREC2에서는 EQE-S가 AQM보다 우수하였다.

질문2 : 추가할 용어는 자동으로 선정하는 것이 좋은가, 탐색자가 선정하는 것이 좋은가?

이 주제는 주로 확률모델을 기반으로한 시스템에서 연구되었다. City 대학은 AQE(Okapi)와 IQE(XOkapi), 그리고 지식기반(시소러스) 대화식 질의확장시스템인 ENQUIRE를 사용하여 INSPEC DB와 London Business School의 OPAC에서 비교하였다(Beaulieu 1997). EQ로 더 많은 적합문헌을 얻었다고 응답한 이용자의 비율은 ENQUIRE Inspec(56%)>AQE(Okapi 50%)>ENQUIRE LSB(36%)>IQE(XOkapi. 31%)순으로 시소러스를 기반으로 질의를 확장하는 ENQUIRE가 가장 좋은 기능을 보였고 적합성피드백 검색은 AQE가 IQE보다 우수하였다. 그러나 OPAC만을 가지고 비교했을 때는 AQE(적합성피드백기반)>IQE(지식기반)>IQE(적합성피드백기반) 순의 성능을 볼 수 있다.

초기의 TREC(TREC2와 TREC3)에서 Okapi는 IQE의 탐색에서 자동질의확장(AQE-S) 만큼 좋은 결과를 얻지 못하였다. 그러나 이것은 두 시스템을 통제된 환경에서 비교한 연구결과는 아니

경에서 실험한 연구(Beaulieu et al. 1996)에서는 4개의 시스템이 평가되었다. 이용자의 적합성 판정정보로부터 상위 20개의 용어를 자동선택한 시스템(DIAG2)이 가장 우수하였고, 그 다음은 의사적합성피드백(상위 40개 문헌)으로 상위 20개의 용어를 자동선택한 시스템(AUTO), 이용자의 적합성판정정보로부터 선정된 후보용어리스트에서 이용자가 추가한 구를 포함하여 20개의 용어를 선택한 시스템(INT), 이용자의 적합성판정정보로부터 선정된 후보용어리스트에서 20개의 용어를 선택만 하고 구는 추가하지 않은 시스템(DIAG1) 순서로 우수하였다. AQE-S(DIAG2와 AUTO)가 IQE(INT와 DIAG1)보다 우수하였다. 그러나 TREC4에서 이용자의 적합성판정에 의거한 IQE는 의사적합성피드백을 이용한 자동용어 확장(Ad hoc)과 공식적합정보를 사용한 자동확장(Routing)보다 우수하였다.

Okapi에서와 비슷한 BM25변형공식과 용어선정공식  $w(p-q)$ 를 사용한 오스트리아국립대학의 ACSys에서 수동으로 작성한 초기질의로 검색된 문헌 중 상위  $n$ 개의 문헌을 적합문헌으로 간주할 때 이용자가 용어를 추가하거나 가중치를 수정하는 IQE가 상위  $m$ 개의 용어로 질의를 자동확장하는 AQE-S보다 우수하였다(Hawking et al. 1998).

TREC3 데이터 중 Wall Street Journal 신문 기사 173,252개를 대상으로 AQE-S와 IQE를 비교한 실험은(Magennis & van Rijsbergen 1997) 적합문헌에 출현하는 상위  $n$ 개의 비질의 용어를 확장할 때(F4 사용)  $n=1$ 에서  $n=20$ 까지 변화를 주면서 AQE-S의 효과를 분석하였다. AQE-S는  $n=6$ 일 때 가장 효과가 좋았고 6개를 넘으면 감소하였다. 그러나 6개(+29.85%)와 20개(+27.35%)사이의 효과의 차이는 유의한

수준의 것이 아니었다. 이 결과는 검색된 문헌수(배움표본)를 100개로 했을 때의 결과이고, 배움표본수를 40으로 했을 때는  $n=5$ 가 가장 좋았다. 검색된 문헌수를 60과 80으로 했을 때  $n=6$ 이 가장 좋은 검색결과를 가져왔다. 이 연구에서는 IQE가 AQE-S보다 좋은 결과를 가져올려면 탐색자가 좋은 탐색어를 선정할 능력이 있어야 한다는 가정 아래 탐색경험이 많은 탐색자와 경험이 없는 비경험자를 대상으로 IQE를 실시하였다. 그 결과는 경험자가 수행하는 IQE는 AQE-S보다 좋았으나, 비경험자가 수행한 IQE는 AQE-S보다 나쁜 결과를 가져왔다 ( $IQE(\text{경험자}) > AQE-S > IQE(\text{비경험자})$ ).

Inquery에서는 IQE가 AQE-S보다 우수함을 보고하고 있다. TREC2 Routing테스트에서 IQE는 AQE보다 +2.4% 우수하였다. TREC3의 공식적합정보를 사용한 Routing테스트에서 IQE가 AQE보다 +1.1% 우수하였고, 의사적합정보를 사용한 특별테스트에서 IQE는 AQE보다는 +15.5% 우수하였다.

## 9 결 론

본 연구에서는 비블리언탐색환경에서 탐색결과에 근거하여 질의를 수정 또는 확장, 결합하여 검색효과를 향상시키는 시스템들의 성능을 문헌을 통하여 고찰하고 성능에 영향을 끼치는 요소들을 분석하였다. 시스템의 성능은 사용된 개념모델, 용어의 가중치기법, 문헌의 순위화방법, 용어선정 알고리즘, 질의확장에 사용된 문헌수와 용어수, 적합성판정정보의 출처, 배움표본의 크기, 부적합 문헌정보의 사용여부, 용어확장방법, 질의의 크기, DB의 종류와 크기에 의해 영향을 받는 것으로 분석되었다.

DB의 종류와 크기, 질의의 크기, 적합성정보 등을 통일한 TREC 환경에서 벡터모델을 사용한 SMART와 이진독립확률모델에 기반한 Okapi, 추론모델에 기반한 Inquiry가 성능이 우수한 대표적인 시스템이었고, 다른 많은 시스템들이 이 셋 특히 SMART와 Okapi를 기반으로 탐색기법을 응용하고 있다(AT&T, IRIT, ANU/ACSys, Mercure, Waterloo 등).

용어의 가중치는 문헌이나 질의의 길이에 따라 영향을 받는 것을 최소화하기 위한 정규화(SMART의 Pivot 코사인정규화, Okapi의 문헌 길이정규화)기법이 성능을 향상시켰다.

적합성판정정보는 이용자의 판정정보나 사전에 알고 있는 공식적합정보를 사용하는 것이 의사적합문헌을 사용하는 것보다 좋은 결과를 가져오나, 공식적합정보가 없는 환경에서는 의사적합성 피드백은 매우 효과적이었다. 특히 논리적사고력이 약한 탐색자는 부적합문헌을 유용한 문헌으로 판정하는 경향이 높기 때문에(Allen 1993) 이용자의 논리적사고능력에 따라 이용자의 적합성판정활동 대신 의사적합성정보를 사용하는 것도 고려해 볼 수 있다. 질의확장에 사용되는 적합성정보는 부적합문헌도 포함시키는 것이 성능을 향상시켰다.

질의수정보다는 질의확장이 우수하였고, 적합 문헌에 출현하는 모든 용어로 확장하는 것은 서지 DB에서는 유용하나, 본문 DB에서는 적합문

헌에 출현하는 용어 중 선택하여 사용하는 것이 현실적으로(성능과 시간 면으로) 타당해 보인다. 탐색자가 주제영역의 전문지식이 많은 사람일 경우에는 전시된 상위n개의 용어 중 이용자가 선택하거나 구(복합명사)로 조합, 새로운 용어를 추가할 수 있게하는 IQE가 타당해 보인다. 또한 본문 DB에서는 문헌검색보다는 문단검색이 유용하였고, 문단검색은 윈도우크기의 문단(passage)검색이 문법단위의 문단(paragraph)검색보다 우수하였다(Callan 1994). 피드백탐색의 결과를 결합하면 보다 좋은 검색효과가 가능하였다.

그러나 이러한 연구의 결론은 대부분 영문텍스트를 대상으로 수행된 결과이기 때문에 한글텍스트에서 이들 변수에 대한 테스트없이 실제 온라인시스템의 성능을 향상시키기 위해 분석된 기법들을 응용하는 것은 무리라고 생각된다. 본 연구에서는 자연어탐색의 성능향상에 영향을 미치는 요소들을 규명하고, 현재까지 진행된 연구들을 소개하고 분석함으로써, 한글텍스트에서 이에 대한 연구를 수행하는데 방향을 제시하려 하였다. 탐색결과의 통계적인 데이터에 의한 성능향상기법 외에도 지식기반 용어확장시스템이나 구문과 문법에 기반을 둔 자연어연구가 수행되고, 불논리탐색에서 응용 가능한 기법들이 연구되어 한글 OPAC이나 서지DB, 본문DB 검색시스템이 발전할 수 있기를 기대한다.



## 참 고 문 헌

- 김주연, 김병만, 고희일. 1997. 연관문헌내 용어 발생 분포를 이용한 질의용어확장. 『1997년도 한국정보과학회 봄 학술발표논문집』, 24(1): 487-490.
- 문성빈. 1993. 적합성 피드백을 이용한 전문검색시스템의 검색효율성 증진을 위한 연구. 『정보관리학회지』, 10(2): 43-67.
- 이준호. 1995. 다중 가중치기법을 이용한 검색효과의 개선. 『정보관리학회지』, 12(2): 213-223.
- 이기호, 이준호, 이규철. 1997. 다중 질의 결합을 통한 검색 효과의 개선. 『한국문헌정보학회지』, 31(3) : 135-146.
- Aalbersberg, I. 1992. "Incremental relevance feedback." *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 11-22.
- Allan, B. 1993. "Logical reasoning and retrieval performance." *Library & Information Science Research*, 15: 95-105.
- Allan, J. 1995. "Relevance feedback with too much data." *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 337-343.
- Allan, J. 1996. "Incremental relevance feedback for information filtering." *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 270-278.
- Allen, J. et al. 1996. "Recent experiments with Inquiry." <<http://trec.nist.gov/pubs/trec4/>>.
- Allen, J. et al. 1997. "Inquiry at TREC-5." <<http://trec.nist.gov/pubs/trec5/>>.
- Allan, J. et al. 1998. "Inquiry does battle with TREC-6." <<http://trec.nist.gov/pubs/trec6/>>.
- Allan, J. et al. 1999. "Inquiry and TREC-7." <<http://trec.nist.gov/pubs/trec7/>>.
- Beaulieu, M. 1997. "Experiments on interfaces to support query expansion." *Journal of Documentation*, 53(1): 8-19.
- Beaulieu, M., S. Robertson, and E. Rasmussen. 1996. "Evaluating interactive systems in TREC." *Journal of the American Society for Information Science*, 47(1): 85-94.
- Beaulieu, M. and M. J. Gatford. 1998. "Interactive Okapi at TREC-6." <<http://trec.nist.gov/pubs/trec6/>>.
- Belkin, N. J. et al. 1993. "The effect of multiple query representation on information retrieval system performance." *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 339-346.
- Belkin, N. J. et al. 1995. "Combining the evidence of multiple query representations for information retrieval."

- Information Processing & Management*, 31(3): 431-448.
- Belkin, N. J. et al. 1998. "Rutgers' TREC-6 interactive track experience."  
<<http://trec.nist.gov/pubs/trec6/>>.
- Belkin, N. J. et al. 1999. "Rutgers' TREC-7 interactive track experience."  
<<http://trec.nist.gov/pubs/trec7/>>.
- Borgman, C. L. 1996. "Why are online catalogs still hard to use?" *Journal of the American Society for Information Science*, 47(7): 493-503
- Borlund, P. and P. Ingwersen. 1997. "The development of a method of the evaluation of interactive information retrieval systems." *Journal of Documentation*, 53(3) : 225-250.
- Broglio, J. et al. 1995. "Document retrieval and routing using the Inquiry system."  
<<http://trec.nist.gov/pubs/trec3/>>.
- Burkley, C. and G. Salton. 1995. "Optimization of relevance feedback weights." *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 351-357.
- Burkley, C., G. Salton, and J. Allan. 1994. "The effect of adding relevance information in a relevance feedback environment." *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 192-300.
- Burkley, C., J. Allan, and G. Salton. 1995. "Automatic routing and retrieval using SMART: TREC-2." *Information Processing & Management*, 31(3): 315-326.
- Burkley, C., A. Singhal, and M. Mitra. 1997. "Using query zoning and correlation within SMART: TREC5."  
<<http://trec.nist.gov/pubs/trec5/>>.
- Burkley, C. et al. 1995. "Automatic query expansion using SMART: TREC3."  
<<http://trec.nist.gov/pubs/trec3/>>.
- Burkley, C. et al. 1996. "New retrieval approaches using SMART: TREC4."  
<<http://trec.nist.gov/pubs/trec4/>>.
- Burkley, C. et al. 1998. "Using clustering and superconcepts within SMART: TREC6."  
<<http://trec.vist.gov/pubs/trec6/>>.
- Buckley, C. et al. 1999. "SMART high precision: TREC7."  
<<http://trec.nist.gov/pubs/trec7/>>.
- Callan, J. P. 1994. "Passage-level evidence in document retrieval." *Proceedings of 17th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 302-310.
- Callan, J., W. B. Croft, and J. Broglio. 1995. "TREC and Tipster experiments with Inquiry." *Information Processing & Management*, 31(3):327-343.
- Croft, W. and D. Harper. 1979. "Using probabilistic models of information retrieval without relevance information." *Journal of Documentation*, 35:

- 285-295.
- Croft, B. et al. 1994. "TREC-2 routing and ad-hoc retrieval evaluation using the Inquiry system."  
 <<http://trec.nist.gov/pubs/trec2/>>.
- Croft, B., H. R. Turtle, and D. D. Lewis. 1991. "The use of phrases and structured queries in information retrieval." *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 21-30.
- Efthimiadis, E. 1993. "A user-centered evaluation of ranking algorithms for interactive query expansion." *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 146-159.
- Efthimiadis, E. 1995. "User choices : A new yardstick for the evaluation of ranking algorithms for interactive query expansion." *Information Processing & Management*, 31(4): 605-620.
- Efthimiadis, E. 1996. "Query expansion." *Annual Review of Information Science and Technology*, 31: 121-187.
- Efthimiadis, E. and P. V. Biron. 1994. "UCLA-Okapi at TREC2: Query Expansion experiments."  
 <<http://trec.nist.gov/pubs/trec2/>>.
- Fagan, J. L. 1989. "The effectiveness of a nonsyntactic approach to automatic phrase indexing for document retrieval." *Journal of the American Society for Information Science*, 40: 115-132.
- Fox, E. A. 1998. "Testing the applicability of intelligent methods for information retrieval." *Information Services and Use*, 7(4-5): 119-138.
- Fox, E. A. 1990. "Advanced retrieval methods for online catalogs." *Annual Review OCLC Research*, July 1989 to June 1990. (Dublin, OH: OCLC): 32-34.
- Fox, E. A. and J. A. Shaw. 1994. "Combination of multiple searches."  
 <<http://trec.nist.gov/pubs/trec2/>>.
- Fox, E. A. et al. 1993. "Development of a Modern OPAC: From REVTOLC to MARIAN." *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 248-259.
- Fuhr, N. and C. Buckley. 1991. "A probabilistic learning approach for document indexing." *ACM Transactions on Information systems*, 9(3): 223-248.
- Greiff, W. R. 1998. "A theory of term weighting based on exploratory data analysis." *Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 11-19.
- Haines, D. and W. B. Croft. 1993. "Relevance feedback and inference networks."

- Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 2-11.
- Harman, D. 1992. "Relevance feedback revisited." *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 1-10.
- Harman, D. 1993. "Overview of the first Text Retrieval Conference (TREC-1)." <http://trec.nist.gov/pubs/trec1/>.
- Harman, D. 1994. "Overview of the second Text Retrieval Conference (TREC-2)." <http://trec.nist.gov/pubs/trec2/>.
- Harman, D. 1995. "Overview of the third Text Retrieval Conference (TREC-3)." <http://trec.nist.gov/pubs/trec3/>.
- Harman, D. 1996. "Overview of the fourth Text Retrieval Conference (TREC-4)." <http://trec.nist.gov/pubs/trec4/>.
- Hawking, D., P. Thistlewaite, and Nick Craswell. 1998. "ANU/ACSys TREC-6 experiments." <http://trec.nist.gov/pubs/trec6/>
- Hersh, W. and B. Day. 1997. "A Comparison of Boolean and Natural Language Searching for the TREC-6 Interactive Task." <http://trec.nist.gov/pubs/trec6/>.
- Hersh, W. and D. Hickam. 1995. "An Evaluation of interactive Boolean and natural language searching with an online medical textbook." *Journal of the American Society for Information Science*, 46(7): 478-489.
- Hersh, W. et al. 1994. "OHSUMED: An interactive retrieval evaluation and new large test collection for research." *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 192-201.
- Hunter, R. N. 1991. "Success and failures of patrons searching the online catalog at a large academic library: a transaction log analysis." *RQ*, 30: 395-402.
- Keen, E. M. 1991. "The use of term position devices in ranked output experiments." *Journal of Documentation*, 47: 1-22.
- Keen, E. M. 1992. "Term position ranking: some new test results." *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 66-76.
- Khoo, C. and D. C. C. Poo. 1994. "An expert system approach to online catalog subject searching." *Information Processing & Management*, 30(2): 223-238.
- Kwok, K. L. 1996. "A new method of weighting query terms for Ad-Hoc retrieval." *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 187-195.

- Kwok, K. L. and M. Chan. 1998. "Improving two-stage Ad-Hoc retrieval for short queries." *Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 250-256.
- Lancaster, F. W. et al. 1994. "Searching databases on CD-ROM : Comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries." *RQ*, 33(3): 370-386.
- Losee, R. M. 1988. "Parameter estimation for probabilistic document retrieval models." *Journal of the American Society for Information Science*, 39: 8-16.
- Losee, R. M. and A. Boostein. 1988. "Integrating Boolean queries in Conjunctive Normal Form with probabilistic retrieval models." *Information Processing & Management*, 24(3): 315-321.
- Lu, X., J. D. Holt, and D. J. Miller. 1995. "Boolean system revisited : Its performance and its behavior." (<http://trec.nist.gov/pubs/trec4/>).
- Lynch, C.A. 1989. "Large database and multiple database problems in online catalogs." *OPACS and beyond: Proceedings of a joint meeting of the British Library, DBMIST, and OCLC*. Dublin, ohio: OCLC.
- Magennis, M. and C. J. van Rijsbergen. 1997. "The potential and actual effectiveness of interactive query expansion." *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 324-332.
- Mitra, M., A. Signhal, and C. Buckley. 1998. "Improving automatic query expansion." *Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 206-214.
- Noreault, T., M. McGill, and M. B. Koll. 1981. "A Performance evaluation of similarity measures, document term weighting schemes & representations in a Boolean environment." In: R. N. Oddy ed. *Information Retrieval Research* (London, Butterworths, 1981): 57-76.
- Paris, L. A. and H. R. Tibbo. 1998. "Freestyle vs. Boolean: a comparison of partial and exact match retrieval systems." *Information Processing & Management*, 34(2/3): 175-190.
- Porter, M. and V. Galpin. 1988. "Relevance feedback in a public access catalogue for a research library: Muscat at the Scoatt Polar Research Institute." *program*, 22(1): 1-20.
- Ro, J. S. 1988. "An evaluation of the applicability of ranking algorithms to improve the effectiveness of full-text retrieval. II. On the effectiveness of ranking Algorithms on Full-text

- retrieval." *Journal of the American Society for Information Science*, 39: 147-160.
- Robertson, S. 1986. "On relevance weight estimation of query expansion." *Journal of Documentation*, 42: 182-188.
- Robertson, S. 1990. "On term selection for query expansion." *Journal of Documentation*, 46(4): 359-364.
- Robertson, S. and K. Sparck Jones. 1976. "Relevance weighting of search terms." *Journal of the American Society for Information Science*, 27:129-146.
- Robertson, S. and C. L. Thompson. 1990. "Weighted searching: The CIRT experiment." *K.P. Jones ed., Informatics 10-prospects for intelligent retrieval*. (London: Aslib): 153-165.
- Robertson, S. and S. Walker. 1994. "Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval." *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 232-241.
- Robertson, S., S. Walker, and M. Hancock-Beaulieu. 1995. "Large test collection experiments on an operational, interactive system: Okapi at TREC." *Information Processing & Management*, 31(3): 345-360.
- Robertson, S., S. Walker, and M. Beaulieu. 1997. "Laboratory experiments with Okapi: Participation in the TREC programme." *Journal of Documentation*, 53(1): 20-34.
- Robertson, S., S. Walker, and M. Beaulieu. 1999. "Okapi at TREC-7: automatic ad hoc, filtering, VLC and interactive." <http://trec.nist.gov/pubs/trec7/>.
- Robertson, S. et al. 1994. "Okapi at TREC-2." <http://trec.nist.gov/pubs/trec2/>.
- Robertson, S. et al. 1995. "Okapi at TREC-3." <http://trec.nist.gov/pubs/trec3/>.
- Robertson, S. et al. 1996. "Okapi at TREC-4." <http://trec.nist.gov/pubs/trec4/>.
- Salton, G. and C. Buckley. 1988. "Term-weighting approaches in automatic text retrieval." *Information Processing & Management*, 24(5): 513-523.
- Salton, G. and C. Buckley. 1990. "Improving retrieval performance by relevance feedback." *Journal of the American Society for Information Science*, 41(4): 288-297.
- Salton, G. and R. K. Waldstein. 1978. "Term relevance weights in on-line information retrieval." *Information Processing & Management*, 14:29-35.
- Savoy, J., M. Ndarugendamwo, and D. Vrajitoru. 1996. "Report on the TREC4 experiment: combining probabilistic and vector-space schemes." <http://trec.nist.gov/pubs/trec4/>.
- Shaw, W. M., Jr. 1995. "Term-relevance computations and perfect retrieval

- performance." *Information Processing & Management*, 31(4): 491-498.
- Shaw, J. A. and E. A. Fox. 1995. "Combination of Multiple Search." <http://trec.nist.gov/pubs/trec3/>.
- Singhal, A. 1997. "AT & T at TREC-6." <http://trec.nist.gov/pubs/trec6/>.
- Singhal, A. et al. 1998. "AT & T at TREC-7." <http://trec.nist.gov/pubs/trec7/>.
- Singhal, A., C. Buckley, and M. Mitra. 1996. "Pivoted document length normalization." *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 21-29.
- Singhal, A., M. Mitra, and C. Buckley. 1997. "Learning routing queries in a query zone." *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 25-32.
- Singhal, A. et al. 1996. "Document length normalization." *Information Processing & Management*, 32(5): 619-633.
- Sparck Jones, K. 1979a. "Experiments in relevance weighting of search terms." *Information Processing & Management*, 15: 133-144.
- Sparck Jones, K. 1979b. "Search term relevance weighting given little relevance information." *Journal of Documentation*, 35(1): 30-48.
- Sparck Jones, K. 1995. "Reflections on TREC." *Information Processing & Management*, 31(3): 291-314.
- Sparck Jones, K. 1998. "Summary performance comparisons : TREC2-6." <http://trec.nist.gov/pubs/trec6> .
- Spink, A. and R. Losee. 1996. "Feedback in information retrieval." *Annual Review of Information Science & Technology*, 31: 33-78.
- Turtle, H. 1994. "Natural language vs. Boolean query evaluation : A comparison of retrieval performance." *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 212-220.
- Turtle, H. and W. B. Croft. 1991. "Evaluation of an inference network-based retrieval model." *ACM Transactions on Information Systems*, 9(3):187-222.
- Turtle, H. and W. B. Croft. 1992. "A Comparison of text retrieval models." *The Computer Journal*, 35(3): 279-290.
- Van Rijsbergen, C. J. 1977. "A theoretical basis for the use of co-occurrence data in information retrieval." *Journal of Documentation*, 33: 106-119.
- Vorhees, E. M. and D. Harman. 1997. "Overview of the fifth Text REtrieval Conference(TREC-5)." <http://trec.nist.gov/pubs/trec5/>.
- Vorhees, E. M. and D. Harman. 1998. "Overview of the sixth Text REtrieval

- Conference(TREC-6)."  
<<http://trec.nist.gov/pubs/trec6/>>.
- Walker, S. et al. 1998. "Okapi at TREC-6 : Automatic ad hoc, VLC, routing, filtering and QSDR."  
<<http://trec.nist.gov/pubs/trec6/>>.
- Wilkinson, R., J. Zobel, and R. Sacks-Davis. 1996. "Similarity Measures for short queries."  
<<http://trec.nist.gov/pubs/trec4/>>.
- Wu, H. and G. Salton. 1981. "The Estimation of term relevance weights using relevance feedback." *Journal of Documentation*, 37: 194-214.
- Xu, J. and W. B. Croft. 1996. "Query expansion using local and global document analysis." *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Science*, 4-11.