

문단 기반 멀티에이전트 메타 검색 엔진 설계

강남오*, 박사준**, 황수철***

요약

인터넷에 공개되는 문서가 증가함에 따라, 사용자들이 정보를 정확하고 효과적으로 찾아내는 작업은 더욱 어려워지고 있다. 일반적으로, 사용자들은 질의어에 대해 검색 엔진에서 산출한 결과 리스트를 이용하여 웹 페이지를 방문하고, 원하는 정보를 획득하기 위해 각 페이지 내의 자료를 검색한다. 이러한 작업들은 추가의 시간과 노력을 필요로 한다. 또한 검색의 질을 향상하기 위해 이용되는 검색 엔진의 수가 늘어나면 엔진마다 각기 다른 결과 리스트가 산출됨으로 인해, 이들 사이의 합리적인 통합을 통한 검색은 더욱 어려워진다. 이의 개선을 위해 우리는 멀티에이전트와 결과 문서들내 지역적 정보 이용을 도입하였다. 본 논문에서 우리는 문단 기반 멀티에이전트 메타 검색 엔진을 소개한다. 본 시스템은 사용자 질의어에 대해 다수의 검색 엔진을 이용하여 검색을 수행하고, 산출된 웹 문서로부터 문단 단위로 유사도를 측정함으로써 지역 중심적인 결과 리스트를 생성한다.

Design of Paragraph-based Multiagent Metasearch Engine

Nam-Oh Kang*, Sa-Joon Park**, Su-Chul Hwang***

ABSTRACT

As the number of documents published on the Internet increases, users are gradually difficult to retrieve information accurately and efficiently. In general, users visit web pages by using the result list generated from a search engine and search each web page of the browsed list to discover information. And these processes are necessary for an additional time and work. In addition, for improving the retrieved result, increasing the number of search engines used in the retrieval process produces many different result lists, and it makes much more difficult for users to search information based on an integrated searching list. For solving these problems, we introduce using the multiagent and the localized information in the result documents. In this paper, we propose a Paragraph-based Multiagent metasearch engine which produces local-centric result list based on the calculated paragraph-unit similarity in web documents, which is browsed by numerous search engines for a user query.

Key Words : Multiagent systems, Metasearch engine, Passage retrieval, Ontology, Information retrieval systems

* (주)리써치링크 (✉namohkang@gmail.com)

** 대구한의대학교 모바일콘텐츠학부

*** 인하공업전문대학

· 제1저자(First Author) : 강남오 · 교신저자(Correspondent Author) : 황수철

· 접수일(2012년 2월 5일), 수정일(1차 : 2012년 3월 5일), 게재확정일(2012년 3월 7일)

1. 서론

인터넷에서의 신속한 정보공유를 위한 서비스로서 월드 와이드 웹(World Wide Web)이 소개된 이후, 웹을 통하여 제공되는 정보는 기하급수적으로 증가하고 있다. 이에 따라 방대한 웹상의 데이터로부터 사용자가 원하는 정보를 정확하고, 효과적으로 찾아내는 분산 정보 검색(Distributed information retrieval-DIR)은 더욱 어려워지고 있다. 분산 정보 검색은 웹상의 다양한 정보 데이터베이스(웹 사이트)를 검색, 사용자가 원하는 정보를 포함하는 웹페이지의 검색 결과 리스트를 제공하는 서비스로써, 통합 검색(federated search)으로도 알려져 있다. 이의 응용으로는 웹 검색 엔진(Web search engine), 메타 검색 엔진(meta-search engines), 분산 유전정보 검색, 그리고 뉴스 검색 엔진 등이 있다[1-3].

분산 정보 검색은 정보원 기술(description), 정보원 선택, 질의어 변환, 그리고 결과 통합이라는 4개의 하위 작업으로 구성된다[2]. 정보원 기술은 각각의 데이터베이스가 다루는 주제들을 파악하고 기술하며, 정보원 선택은 주어진 질의어를 다루는 정보원을 선택하는 방법을 다룬다. 질의어 변환은 질의어를 선택된 정보원 데이터베이스에서 사용하는 질의어 양식으로 변환하는 방법을 다루며, 마지막으로 결과 통합은 선택된 정보 데이터베이스들로부터 획득된 결과들을 하나의 통합된 결과 리스트로 통합하는 과정을 다룬다[1]. 이처럼 분산 정보 검색 서비스를 구축하는 것은 많은 노력과 시간이 드는 작업으로, 웹의 경우 구글, 야후, Bing, 네이버 등의 검색엔진이 그 예이다.

웹 검색 엔진들은 서로 다른 정보원 기술, 정보원 선택, 질의어 변환, 그리고 결과 통합 방법을 취한다. 따라서 동일한 사용자 질의에 대하여서도 검색엔진에 따라 서로 다른 검색 결과를 도출한다. 메타 검색 엔진은 사용자의 질의어를 기존의 검색엔진에 제공하고, 서로 다른 중요도 측정에 의해 산출된 결과 리스트들

을 바탕으로 하나의 결과 리스트를 산출해낸다. 메타 검색 엔진의 주요 논의 점은 사용자 질의어를 검색엔진에 어떻게 전달할 것인가, 산출된 결과 리스트로부터 정확한 검색 결과를 어떻게 판단할 것인가, 그리고 서로 다른 검색 결과들을 어떻게 하나로 통합 할 것인가 등이다[4]. 이러한 문제를 해결하기 위해 메타검색 엔진 구축에 멀티 에이전트 시스템을 활용하려는 시도가 제기 되고 있다.

메타 검색엔진 구축에 멀티 에이전트의 활용은 다음과 같은 장점을 가진다. 첫째 사용자의 질의어를 에이전트마다 각기 다른 형태로 변환 가능함으로써, 다양한 검색이 가능하다. 둘째 기존의 검색 엔진들에서 도출된 결과에 다양한 추가의 처리를 통하여 새로운 결과 리스트 도출이 가능하다. 또한 추가의 처리를 통하여 사용자가 요구하는 부가 정보를 제공하는 것이 가능하다.

현재 대부분의 웹 검색엔진은 사용자의 질의어를 바탕으로 웹 페이지의 내용을 분석, 질의어가 포함된 문서 리스트를 산출한다. 이를 바탕으로 해당 사이트를 방문한 경우, 사용자는 자신이 원하는 정보를 다루는 부분을 재검색해야할 경우가 많다. 웹 사이트가 방대하게 늘어나고 등록된 문서 크기가 점차 커지고 있어서, 방문한 웹 페이지의 내용을 다시 검색하는 것은 추가의 노력과 시간을 요하는 작업이다. 그리고 문서 단위로의 검색은 지역적 유사도를 제대로 반영하지 못하고 있다. 이러한 단점을 해결하기 위해 본 논문에서는 다수의 검색엔진으로부터 사용자 질의에 대한 결과를 산출하고 이를 문단 단위로 유사도를 측정, 지역 중심적인 유사도 중심의 결과를 산출할 수 있는 문단 기반 멀티 에이전트 메타 검색 시스템을 소개한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 통하여 주요 기술적인 면을 고찰하고 3장에서는 문단 기반 웹 검색을 위한 멀티 에이전트 협력 기반의 검색 구조를 제시한다. 그리고 4장에서는 결론 및 향후 연구 방향을 언급 한다.

II. 관련 연구

메타 검색 엔진의 적용 동기는 1) 참여하는 검색 엔진의 구성을 확장함으로써 효과적으로 검색 영역을 확장, 2) 사용자로부터 입력된 하나의 질의어를 개별 검색 엔진에 특화된 질의어로 적응시킴으로써 사용자에게 대한 편리성 증대, 3) 검색 엔진마다 서로 상이한 중요도로 산출된 결과 리스트를 하나의 최종 결과 리스트로 통합하는 검색의 효과성 증대에 있다. 지난 20년 동안 많은 메타 검색 엔진의 연구 및 개발들이 이들의 문제에 맞추어져 있다. 특히 특정 영역에 특화된 검색 엔진들로부터 검색 결과의 추출과 이의 통합은 가장 중요한 주제로 다루어 졌다[4].

m개의 검색 시스템 $S_i, i = 1 \dots m$ 이 있다고 할 때, 주어진 사용자 질의어 q에 대하여 각 검색 시스템은 유사도 측정에 따라 검색된 문서 리스트 l_i 를 산출한다. 사용자에게 하나의 관련 문서 리스트를 산출하기 위해, 각각의 검색된 문서 리스트는 통합이 되어야 한다[5-7]. 각 정보 검색 시스템은 서로 다른 유사도 측정 기법을 바탕으로 검색 리스트를 산출하였기 때문에 이를 조절하고 합리적으로 통합하는 것은 어려운 일이다[1].

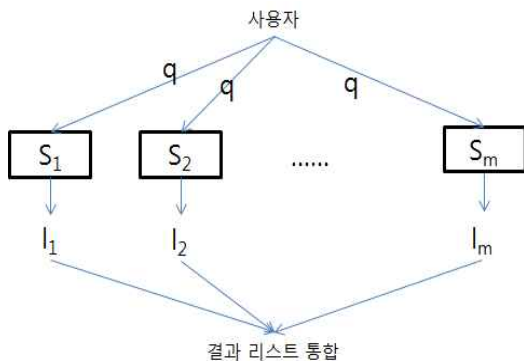


그림 1. 분산 정보 검색의 처리 절차

Fig.1 The process of distributed information retrieval

초창기 검색 엔진들은 수치적 유사도 값을 검색 결과에 내포시키고 있었다. 따라서 메타 검색 엔진은 이 값들을 정규화 시켜서 통합된 결과 리스트를 산출했다[4]. [8]에서는 유사도 값이 없을 경우에 결과 리스트의 랭킹을 활용하여 통합 결과 리스트를 산출하는 방법을 제시하였다. [9]에서는 각 검색 엔진이 산출한 결과 리스트의 문서들을 메타 검색 엔진이 모두 다운로드 하고서 공통의 유사도 함수를 활용하여 통합된 결과 리스트를 산출하는 방법을 제시하였다.

이 방법의 장점은 하나의 유사도 측정 함수를 활용함으로써 순위를 산출하는데 통일성을 높인다는 점이다. 하지만 사용자의 요청에 대한 응답속도가 떨어진다는 단점이 있다. 최근의 많은 검색 엔진들은 문서의 제목과 문서의 요약본(snippet)을 제공한다. [10]에서는 제목과 문서의 요약본을 활용하여 순위를 산출하는 방법을 제시하였다.

특정 영역에 특화된 검색 엔진들에 대한 메타 검색 엔진의 구축 시에는 검색 엔진의 선택 또한 중요한 고려 사항으로 다루어진다[4]. [11-13]에서는 주어진 사용자 질의어에 따라 검색 엔진의 선택 및 중요도에 대한 가중치를 부여하는 방법을 제시하였다. 즉 각 검색 엔진은 자신의 주요 대표성(representatives)을 가지고 있으며 메타 검색 엔진은 이 정보를 사용자 질의어와 비교, 검색 적합도를 평가하여 산출된 결과들에 가중치를 적용, 최종 결과 리스트를 산출한다. [14]에서는 질의어와 그 결과로부터 해당 검색 엔진의 주요 대표성을 파악하는 방법을 제시하였다.

사용자의 질의어 확장에 있어서 [15-17]에서는 연관 피드백(relevance feedback)을 통하여 사용자 질의어를 자동 확장하는 방법을 제시하였다. [18]에서는 사용되는 문서 집합의 개선을 통해 질의어의 자동 확장 방법을 제시하였다.

III. 문단 기반 멀티에이전트 메타 검색

2장에서 살펴본바와 같이 대부분의 분산 정보 검색이 특화된 정보 검색 시스템 혹은 데이터베이스를 바탕으로 한 질의어 변환, 유사도 측정 그리고 결과 리스트의 통합을 다루고 있다. 하지만 이들 시스템을 분산 웹 검색 시스템에 바로 적용하기에는 다음과 같은 점들을 고려하여야 한다.

첫째 사용자 질의어에 대한 통일된 변환 및 각 검색 엔진으로부터 산출된 검색 리스트를 바탕으로 통일된 유사도 측정 방안을 고려하여야 한다. 특정 분야의 내용에 정보 검색 시스템이 특화된 경우 이에 대한 정보를 추출하고 이를 활용하여 사용자의 질의어 변환 및 특화된 유사도 측정에 활용할 수 있다. 하지만 웹 검색 엔진은 웹상의 문서를 저장하고 있으며 특정의 분야에 특화되어 있지 않다. 따라서 사용자 질의어의 변경에 있어 일반적인 변경 혹은 사용자의 의도가 파악된 변경이 필요하다.

둘째 사용자가 원하는 정보를 지역적으로 검색 혹은 검색된 웹 페이지의 요약 방안이 필요하다. 이는 현재의 웹 정보 검색에서도 요구 되는 기능이다. 웹 검색 엔진은 사용자의 질의어를 바탕으로 웹 페이지의 내용을 분석하여 질의어가 포함된 문서 리스트를 산출한다. 이를 바탕으로 해당 사이트를 방문한 경우, 사용자는 원하는 정보를 다루는 부분을 재검색해야 하는 단점이 있다. 현재도 웹상의 등록 사이트가 늘어나고 등록된 문서 크기가 점차 커지고 있는 실정이다. 따라서 사용자가 검색 결과에 따라 방문한 웹 페이지로부터 자신이 원하는 정보를 찾기 위해 모든 페이지를 검색하는 것은 많은 노력과 시간을 소비하게 한다. 몇몇의 사이트는 검색을 위한 기능이 있지만 대부분 웹 페이지를 데이터베이스에 저장해두고서 사용자가 찾고자 하는 단어나 구를 입력하면 이를 이용하여 해당 부분을 검출하는 방식이다.

본 장에서는 멀티에이전트를 활용하여 다수의 웹 검색 엔진으로부터 사용자 질의어를 검색하고, 산출된 검색 리스트를 바탕으로 사용자 질의어와 검색 문

서간의 지역적 유사도를 측정, 검색 리스트를 통합하는 문단 기반 멀티에이전트 웹 검색 시스템을 소개한다. 시스템의 전체적 구조는 다음과 같다.

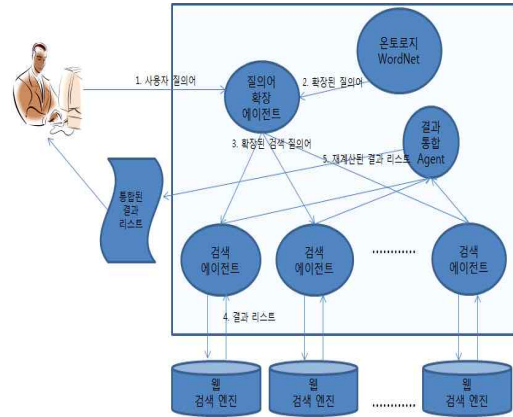


그림 2. 문단 기반 멀티에이전트 웹 검색 시스템
Fig. 2 Multiagent retrieval system based on paragraph

3.1 질의어 확장 에이전트(Query Expansion Agent)

질의어 확장 에이전트는 사용자가 입력한 질의어를 확장함으로써 웹 검색 엔진으로부터 보다 향상된 결과를 산출하도록 돕는다. 질의어 확장 에이전트는 온토로지 혹은 단어 사전을 활용하여 사용자 질의어를 변환하거나 혹은 유사단어를 첨가함으로써 웹 검색 결과를 향상 시킨다.

본 시스템에서는 WordNet을 활용하여 사용자 질의어를 확장 하였다. 질의어 확장 에이전트는 사용자의 질의어를 단어 단위로 분할하고, 분할된 단어는 WordNet을 통하여 유사어를 추출하고 기존 사용자 질의어에 첨가함으로써 사용자 질의어를 확장하였다. 예를 들어서 사용자 질의어가 “programming language”인 경우 질의어 확장 에이전트는 사용자 질의어를 단어 단위로 분할, 이들을 각각 WordNet에 질

의하여 유사어를 추출한다. 단어 "programming"의 경우 "scheduling"과 "programing"이 유사단어로 추출되었고, "language"의 경우 "linguistic communication"이 추출되었다. 질의어 확장 에이전트는 이를 각 웹 검색 엔진의 질의어 연산 규칙에 고려하여 확장한다. 구글 검색 엔진의 경우 OR 연산은 연결된 단어중 적어도 하나를 포함을 의미하고 공백은 AND 연산을 의미한다. 확장된 사용자 질의어는 사용자 질의어의 단어는 반드시 포함하면서 추가로 유사어를 확장하는 구조를 가지도록 구성한다. 따라서 사용자 질의어 "programming language"에 대해 확장된 질의어는 "programming language scheduling OR programing OR linguistic communication" 이 산출된다.

3.2 검색 에이전트(Retrieval Agent)

검색 에이전트는 웹 검색 엔진으로 부터 도출된 결과리스트의 사이트를 방문하고 해당 페이지를 분석하여 사용자 질의어와 문단 단위로의 유사도를 측정함으로써 재평가된 결과리스트를 결과 통합 에이전트에 전달하는 역할을 담당한다.

검색 에이전트는 질의어 확장 에이전트로부터 확장된 사용자 질의어를 입력 받아 등록된 웹 검색 엔진에 이를 전송하고 이의 검색 결과 리스트를 되돌려 받는다. 보통 웹 검색 엔진이 산출하는 결과 리스트에 포함된 웹 페이지 문서의 수는 방대하다. 따라서 검색 에이전트의 방문이 필요한 사이트의 문서 페이지를 걸러내는 작업이 필요하다. 본 시스템에서는 검색 결과에서 상위 20개의 사이트를 방문 한다.

검색 에이전트는 방문한 사이트의 웹 페이지로부터 문단을 추출한다. 문단의 추출은 HTML에서 <P> 태그를 활용한다. - <P> 태그가 발견되지 않을 경우, 문장의 개행을 참조하였다. 추출된 문단에서 불용어(Stopword)를 제거한 후 남겨진 단어들에 대하여 스

테밍(Stemming)을 수행한다. 스템밍은 Port 알고리즘을 활용하였다. 문단화, 불용어 제거 그리고 스템밍이 수행된 문단은 동일한 작업을 거친 사용자 질의어와 유사도 검사를 실시한다.

보통 문서 검색에서 단어의 중요도를 산출하기 위해 빈도수를 바탕으로 한 tf-idf(Term frequency - inverse document frequency)를 이용한다. 하지만, 문서 단위의 유사도 측정과는 달리, 방문 페이지에서 문단 단위의 유사도 측정은 idf 값을 산출하기 어렵다. 그리고 한 문단 내에서 쓰인 단어라 할지라도 다음 문단에서는 동일한 단어가 아니라 비슷한 용어로 대응이 되는 경우가 많다. 따라서 이러한 상황을 고려한 단어의 중요도 산출 방법이 필요하다. 본 시스템에서는 [19]에서 사용한 문단 내 단어 중요도 산출 방식을 도입하였다.

$$w_{k,i} = \log\left(\frac{1}{n_i} + \sum_k e^{-a\left(\frac{d_k}{D}\right)^2} \frac{1}{n_{k,i}}\right) \quad (2)$$

여기서, d_k 는 웹 페이지 내에서 주어진 문단 k 의 거리를 나타낸다. D 는 해당 웹 페이지의 문단 개수를 나타낸다. n_i 는 해당 문서 내에서 단어 i 의 발생 빈도수를 나타내며, $n_{k,i}$ 는 문단 k 에서 단어 i 의 발생 빈도를 나타낸다. a 는 실험을 통해 결정되는 계수 값으로써 여기서는 1의 값을 사용하였다[19].

검색 에이전트는 방문한 각 웹 문서의 문단 내에서 사용자 질의어의 단어 중요도를 산출하고 이의 정보를 결과 통합 에이전트에 넘긴다.

3.3 결과 통합 에이전트(Result Aggregation Agent)

결과 통합 에이전트는 검색 에이전트가 산출한 결과 리스트를 하나로 통합하여 사용자에게 제공하는 역할을 수행한다. 결과 통합 에이전트는 확장된 사용

자 질의어 q 와 각 검색 엔진에서 산출한 웹 페이지의 문단 p 에 공통으로 존재하는 단어 중요도 w_i 를 유사도 계산에 활용한다. 유사도 계산은 벡터 유사도 측정 (vector similarity)[20]을 활용한다. 즉 질의어 q 와 문단 p 사이의 유사도는 두 벡터 사이 각도의 코사인 값으로 측정 한다.

$$S(p, q) = \frac{\sum_i F_{p,i} F_{q,i}}{\sqrt{\sum_i F_{p,i}^2} \sqrt{\sum_i F_{q,i}^2}} \quad (3)$$

여기서 $F_{p,i}$ 와 $F_{q,i}$ 는 문단과 질의어에서 단어 i 에 대해 가중치가 적용된 빈도수 이다. 즉 $F_{p,i} = w_i * f_{p,i}$ 로서 w_i 는 문단에서 단어 i 에 대한 가중치이다. 그리고 $f_{p,i}$ 는 해당 문단에서 발생한 단어 i 의 발생 빈도이다.

본 시스템에서는 사용자 질의어와 문서의 지역적 유사도를 중시 여긴다. 결과 통합 에이전트는 검색 에이전트가 산출한 사용자 질의어와 문서 내 문단의 유사도 값을 내림차순으로 정렬한다. 그리고 문단을 포함한 문서의 URL을 이용하여 결과 리스트를 작성하고 사용자에게 전송한다.

3.4 PMA(Paragraph-based Multiagent) 검색 시스템

제안된 문단 기반 멀티에이전트 검색 시스템은 C#언어를 사용하여 .Net 환경에서 구현하였고, 사용자 질의어의 확장을 위하여 WordNet을 활용하였다.

PMA 검색 시스템은 사용자의 질의어를 입력받아 질의어 확장을 거친 후 선택된 검색 엔진에 전송한다. 시스템은 각 검색 엔진의 결과 리스트 중 HTML 파일에 대하여 문단 기반 검색을 수행 후 이를 바탕으로 재정렬된 결과 리스트를 산출한다[그림 3 참조].

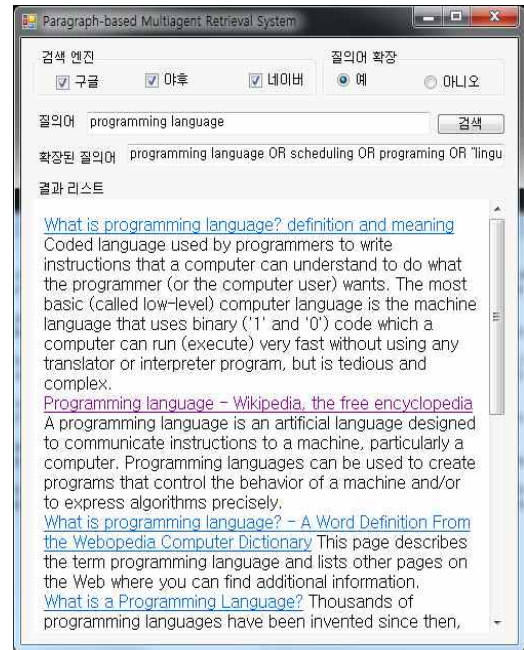


그림 3. 문단 기반 멀티 에이전트 검색의 결과
Fig. 3 The result of PMA retrieval system

IV. 결론 및 향후 연구 방향

인터넷의 정보는 폭발적으로 증가하고 있으며, 대부분의 경우 이러한 정보는 웹을 통하여 제공된다. 사용자들은 웹에서 정보를 얻기 위해 다양한 웹 검색 엔진을 사용한다. 사용자는 결과 리스트를 바탕으로 자신이 원하는 정보를 획득하기 위해 웹 페이지를 방문하고 페이지 내의 자료를 재검색하는 절차를 거친다. 본 논문에서는 멀티에이전트를 활용하여 다수의 웹 검색 엔진을 통해 사용자의 질의어를 검색하고, 검색된 웹 페이지의 문단을 기반으로 한 지역적 유사도를 통하여 결과 리스트를 도출하는 문단 기반 멀티에이전트 메타 검색 시스템을 설계하였다.

본 시스템의 성능 향상을 위해 향후 다음의 분야에 대한 연구를 수행할 예정이다. 본 시스템에서는 사용자 질의어의 확장에 있어 특정 분야의 온토로지를 사

용하지 않고 WordNet을 활용하여 확장을 수행했다. 보다 정확한 검색을 위하여 사용자의 질의어를 분석하고 특정 분야의 온토로지를 적용하는 방안에 대한 연구가 필요하다. 또한 HTML로 작성된 웹페이지의 경우 형식화가 미약하다. 즉 문서 내에 정확한 문단에 대한 정보가 없는 경우가 많다. 이런 경우 본 시스템에서는 문장의 줄 바꿈을 참조하도록 하였으나, 문서 내에서의 의미 흐름을 파악하여 문단의 영역을 구분 지을 수 있는 방법의 도출이 요구된다. 그리고 현재는 HTML 문서에만 국한된 검색이 이루어지나 향후 PDF 및 다른 서식 파일에 대한 검색 기능도 추가가 요구된다. 뿐만 아니라 본 시스템에서는 사용자의 질의어와 문단의 유사도 값을 내림차순으로 정렬함으로써 최종의 결과 리스트를 통합 하였다. 하지만 문서 내 문단들의 유사도 값을 통합하여 문서 전체의 유사도를 측정하는 연구도 필요하다. 이는 지역적 유사도뿐만 아니라 전체 유사도를 동시에 반영함으로써 보다 많은 정보가 포함된 결과 리스트의 산출이 가능하게 한다. 본 논문에서는 웹상의 문서에 대한 문단 기반 검색 시스템 설계 방법을 제시하였다. 향후 TREC WEB 데이터를 통한 기존의 기법과 비교를 수행할 예정이다. 이 경우 기존의 웹 검색 엔진을 활용할 수 없으므로 이에 대한 합리적 기준이 요구된다.

참고문헌

- [1] A. Frad, M. Kahani, R. Ghaemi, and H. Tabatabaee, "Multi-agent Data Fusion Architecture for Intelligent Web Information Retrieval," World Academy of Science, Engineering and Technology Vol. 28, pp.73-78, 2007.
- [2] J. Callan, "Distributed information retrieval," In Advances in Information Retrieval, W.B. Croft, Ed. Kluwer Academic Publishers, pp.127-150, 2000.
- [3] L. Si, J. Callan, "A Semisupervised Learning Method to Merge Search Engine Results," ACM Transactions on Information Systems, Vol. 21, No. 4, pp.457-491, October 2003.
- [4] W. Meng, "Metasearch Engines", Encyclopedia of Database Systems 2009. pp.1730-1734, 2009.
- [5] E. M. Voorhees, N. K. Gupta, and B. Johnson-Laird, "The collection fusion problem", Proceedings of the third Text Retrieval Conference (TREC-3), pp.95-104, 1994.
- [6] E. M. Voorhees, N. K. Gupta, and B. Johnson-Laird, "Learning collection fusion strategies", Proceedings of the 18th ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, pp.172-179, 1995.
- [7] R. R. Yager, "On Ordered Weighted Averaging aggregation Operators in Multi Criteria Decision Making", IEEE Trans. on Systems, Man and Cybernetics. 18(1), pp.183-190, 1998.
- [8] Aslam J. and Montague M. "Models for metasearch." In Proceedings of the ACM SIGIR Conference, New Orleans, LA, pp.276-284, 2001.
- [9] Lawrence S. and Lee Giles C. "Inquirus, the NECi meta search engine." In Seventh International World Wide Web conference, Brisbane, Australia, pp.95-105, 1998.
- [10] Lu Y. Meng W., Shu L. Yu C., and Liu K. "Evaluation of result merging strategies for metasearch engines." WISE Conference, New York, NY, pp.53-66, 2005.
- [11] Yuwono B. and Lee D. "Server ranking for distributed text resource systems on the internet." DASFAA Conference, Melbourne, Australia, pp.391-400, 1997.
- [12] Callan J., Lu Z. and Croft W.B. "Searching distributed collections with inference networks." In Proceedings of the ACM SIGIR Conference, Seattle, WA, pp.21-28, 1995.
- [13] Meng W. Wu Z. Yu C. and Li Z. "A highly scalable and effective method for metasearch." ACM TOIS, 19(3), pp.310-335, 2001.
- [14] Dreilinger D. and Howe A. "Experiences with selecting search engines using metasearch." ACM Trans. Inf.

- Syst., 15(3), pp.195-222, 1997.
- [15] J. Rocchio. "Relevance feedback in information retrieval. In the SMART Retrieval System Experiments in Automatic Document Processing." pp. 313-323, 1971.
- [16] G. Salton, C. Buckley, and E.A. Fox. "Automatic query formulations in information retrieval." Journal of the American Society for Information Science, 34(4), 262-280, 1983.
- [17] G. Salton and C. Buckley. "Improving retrieval performance by relevance feedback." Journal of the American Society for Information Science 41(4), pp.288-297, 1990.
- [18] Mandar Mitra, Amit Singhal, Chris Buckley, "Improving Automatic Query Expansion." SIGIR 1998, pp.206-214, 1998.
- [19] A. Gelbukh, N. Kang, S. Han, "Combining Sources of Evidence for Recognition of Relevant Passages in Texts", LNCS 3563, pp.283-290, 2005.
- [20] G. Salton, A. Wong, C. S. Yang, "A Vector Space Model for Automatic Indexing." Commun. ACM 18(11): pp. 613-620, 1975.

박사준(Sa-Joon Park)



1990년 중앙대학교 전자계산학과(학사)
1994년 중앙대학교 대학원 컴퓨터공학과
(공학석사)
2004년 중앙대학교 대학원 컴퓨터공학과
(공학박사)
2005년~현재 대구한의대학교 모바일콘텐츠학부 부교수
※ 관심분야: 인공지능, 시맨틱 웹, 모바일콘텐츠

황수철(Su-Chul Hwang)



1986년 중앙대학교 전자계산학과(학사)
1988년 중앙대학교 대학원 전자계산학과
(이학석사)
1993년 중앙대학교 대학원 전자계산학과
(공학박사)
1991~현재 인하공업전문대학 교수
※ 관심분야: 인공지능, 지능형시스템, 인터넷/모바일 응용

저자소개



강남오(Nam-Oh Kang)

2000년 중앙대학교 대학원 컴퓨터공학과
(공학석사)
2006년 중앙대학교 대학원 컴퓨터공학과
(공학박사)

2011년~현재 리써치링크
※ 관심분야: 시맨틱웹, 모바일콘텐츠