

지능형 E-mail 문서 관리기 시스템 설계

최승혁 김용성 김영천

전북대학교 컴퓨터공학과

shchoi@mail.chonbuk.ac.kr, yskim@moak.chonbuk.ac.kr, yckim@moak.chonbuk.ac.kr

Designed of Intelligent E-mail Document Management System

Seung-Hyuck Choi Young-Seong Kim Young-Chun Kim

Dept. of Computer Technology, Chonbuk University

요 약

정보의 다양화와 급속한 정보량의 증가로 정보검색 시스템의 필요성이 증대되고 있다. 그러나 현재 시스템 자체로는 사용자들의 요구나 선호도를 만족시킬 수 없고 그만큼 사용자에게 편의성을 제공하지 못한다. E-mail의 정보를 보다 효율적으로 관리하기 위하여 사용자가 원하는 문서를 그룹핑하고 지식라인 생성이 가능한 문서 관리기 시스템을 생성해야한다. 따라서 본 논문에서는 LSA기법을 이용하여 문서를 순차적으로 관리하고 유사한 지식을 그룹화 시키는 문서관리 알고리즘을 제안하고 사용자의 선호도에 맞는 알고리즘을 생성하기 위한 문서관리시스템을 설계한다.

1. 서 론

현대 정보화 사회에서는 정보를 유효적절하게 처리하여 가치가 있는 형태로 저장한 후, 이를 신속, 정확하게 처리하는 것이 매우 중요하게 되었다. 기하급수적으로 증가하고 있는 많은 정보들중에서 유용한 정보들을 이끌어 내는 것이란 무척 어려워지고 있다. 이러한 적합한 정보를 얻기 위한 시간과 노력이 증가하고 있는 반면에 정보검색 시스템을 이용하여 적합한 정보를 찾는 것에 여전히 많은 시간을 허비해야 한다. 따라서 사용자의 관심분야나 선호도를 고려하여 수많은 검색결과 중에서 보다 적합한 정보만을 선택, 제공할 수 있는 방법이 연구되어야 하고 특정 사용자의 관심분야나 선호도를 고려하여 문서의 순위를 결정하기 위해서는 사용자의 요구가 몇 개의 용어 조합이 아닌 상세하면서 선호도를 반영 할 수 있는 형태로 표현되어야 한다. 사용자의 요구를 표현하기 위해서 사용자 위주의 정보검색 기법의 중요성이 증가하면서 정보필터링 분야에서 사용자 프로파일로 연구가 되고 있다. 본 논문에서는 검색된 수많은 결과 중에서 특정사용자의 선호도를 고려한 최적의 문서만을 제공하기 위하여 사용자 프로파일과 잠재적구조분석 (LSA; Latent Structure Analysis)을 이용한 지능형 E-mail문서 관리 시스템을 제안한다. 본 논문은 사용자 프로파일을 구축하여 사용자의 선호도를 표현하고 검색결과 문서들을 대상으로 사용자의 선호도를 분석한 다음 사용자 프로파일과 분석결과로 표현된 문서들과의 유사성을 비교하여 적합성 정도에 따라 사용자에게 최적의 문서를 제공하는데 목적이 있다.

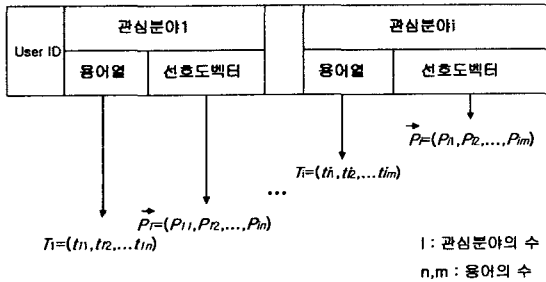
2. 관련 연구

본 논문에서는 특정 사용자의 선호도를 반영하여 용어열과 선호도 벡터로 구성된 사용자 프로파일을 설계하고 효과적으로 E-mail을 관리 할 수 있는 방법을 제안하였다. 본 논문에서 설명하는 사용자 관심을 표현하는 사용자 프로파일은 사용자가 원하는 정보의 성향을 효과적으로 표현하기 위한 수단이다. 따라서 사용자 프로파일은 사용자의 취향이나 정보검색의 성향을 파악하여 사용자의 정보를 표현함으로써 정보검색 시스템이 효과적인 검색을 할 수 있도록 도와준다. 이러한 사용자 프로파일은 정보필터링의 핵심이 되고 있다. 본 논문에서는 지식관리 시스템으로서 E-mail을 사용하였고 사용자 위주의 지식관리 시스템과 사용자의 선호도를 나타내기 위해 프로파일을 설계하였고 E-mail에서 추출한 정보를 이용해서 프로파일을 갱신한다. 이 시스템을 이용하면 이미 구축되어 있는 E-mail시스템을 사용하기 때문에 별도의 문서를 공유,관리할 수 있는 구조화를 쉽게 할 수 있다. 지금까지 사용되어 온 방식은 구체적인 검색이 매우 어려울 뿐만 아니라 E-mail을 자체적으로 관리를 해왔을 뿐 추가적인 파일까지는 관리가 제대로 되지 않았다. 본 논문에서는 추가적인 파일까지도 체계적으로 관리 할 수가 있고 사용자가 관심도에 따라 효율적으로 E-mail을 관리 할 수 있는 시스템을 설계하고자 한다. 또한 문서순위결정을 고려하여 사용자의 질의어와 검색된 문서들이 얼마나 유사한가에 의해 문서의 순위를 결정하고 사용자가 가장 적합한 문서를 참조하는 방법을 적용하고 사용자의 선호도에 의해 데이터를 입력해 놓고 필터링 함으로써 검색된 문서의

수를 줄일 수가 있고 사용자가 원하는 정보를 쉽게 얻을 수 있도록 한다. 본 논문에서는 LSA를 이용한 필터링기법과 문서들간의 연관성을 이용하여 사용자 프로파일과 분석결과로 표현된 문서들과의 유사성을 비교하여 최적의 문서를 제공하는데 목적이 있다.

3.1. 사용자 프로파일의 구조

사회가 점점 복잡해지고 정보의 가치와 양이 많아짐에 따라 적합한 정보를 얻는 시간과 노력이 매우 증대되고 있다. 또한 막대한 E-mail중에 적합한 결과를 찾아내기 위해 많은 노력을 기울여야 할 것이다. 따라서 개인의 관심분야나 선호도를 고려하여 일반 정보검색 시스템이 제공하는 수많은 검색결과 중에서 보다 적합한 결과만을 선택,제공 할 수 있는 방법이 요구되고 있다. 따라서 E-mail을 이용하여 지능형 문서관리기 시스템을 구축할 경우 정보의 구축이 쉽고 데이터 관리가 쉬워진다. 이러한 요구에 따라 사용자의 질의와 문서의 유사성 비교를 통해 순위를 결정하여 사용자에게 제시하는 문서순위결정에 관한 연구가 되어야 하고 이 논문에서 제안하는 사용자 프로파일의 구조는 다음과 같다.



3.2. 프로파일과 LSA

일반적으로 사용되고 있는 용어기반 정보검색 시스템은 사용자가 입력한 질의어의 정합매칭으로 검색을 수행하기 때문에 사용자가 입력한 질의어가 비적합 문서의 키워드와 일치하더라도 그문서까지 결과로 제공하는 경우가 발생한다. 그리고 사용자가 질의어를 적절히 입력하지 못한 경우에는 적합한 문서일지라도 질의어를 포함하지 않기 때문에 검색결과로 얻을 수 없는 경우도 발생한다. 현재 용어기반 정보검색 시스템의 문제점을 해결하기 위해 사용되는 방법으로는 동의어와 다의어, 문서에서 추출된 색인들을 부울 조합 방법으로 조합해서 색인을 만드는 방법이 있다.

4. 문서관리 시스템 설계

4.1 사용자 프로파일의 갱신

(1) 사용자 접근에 의한 갱신

사용자가 특정어에 더 중요성을 부여하고자 할 때 선호도를

반영하는 방법이다.

(2) 확장된 질의의 확인을 통한 갱신

확장된 질의어에 대한 사용자의 확인을 통해 최종 선택된 질의어의 가중치를 사용자 프로파일의 선호도에 반영하는 방법이다.

(3) 사용자 프로파일을 참조한 질의확장

사용자 프로파일의 선호도가 0.9이상인 용어를 추가하여 질의를 확장한다.

(4) 시소러스를 참조한 질의확장

확장된 질의가 사용자 프로파일의 용어열에 있는 경우에는 선호도 벡터의 선호도 질의의 가중치로 한다.

4.2 E-mail의 정보구조

E-mail을 이용한 지능형 문서관리 시스템은 다른 사람이 생성한 정보를 추출한다는 것이다. 그러므로 이러한 정보를 관리하기 하기 위해서 생성일, 생성자, 정보형태 등의 정보를 E-mail의 내용에서 추출해야 한다. 다음 그림은 E-mail의 원본 메시지 구조를 head 부분과 body부분으로 구분한 것으로 지능형 문서관리 시스템을 구축하기 위한 여러 정보를 포함하고 있는 것을 알 수 있다.

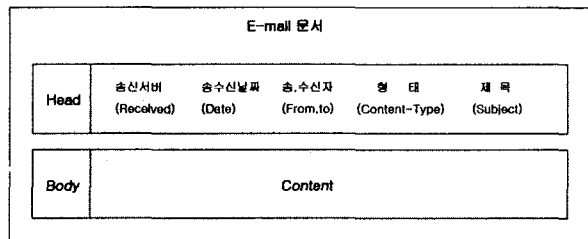


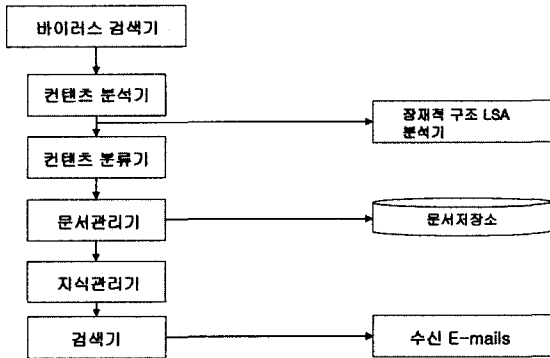
표1은 Data, Content-Type, Received, From, To 와 같은 정형화 정보와 Subject, Content와 같은 비 정형화 정보로 구분 할 수 있다. 정형화된 정보는 관리나 분류가 단순한 질의에 가능하고 비 정형화 정보는 정보를 파악하기 위해서 정형화 정보로의 변환이 필요하다.

정형화 정보			비정형화 정보	
날짜	형태	문자열	문자열	
Date	Content-Type	Received From To	Subject	Content

4.3 시스템 구조

일반적인 E-mail 시스템은 단순히 메일을 송수신 하는 기능과 메일 분배기능, 개인 스케줄링 기능을 가지고 있지만 지능형 E-mail 문서관리시스템은 해킹방지, E-mail 분류 및 공유, 사용자 위주의 주제별 검색 등 다양한 기능을 제공할 수 있게 구성된다.

즉 지능형 E-mail 문서관리시스템은 E-mail을 지식관리시스템을 기반으로 E-mail 송수신은 물론, 문서관리기틀 이용하여 정보를 체계적으로 구축하여 업무 처리의 효율성을 높이고자 한다. 또한 업무처리의 능률을 향상시키기 위하여 E-mail에 대한 관리, 운영할 수 있는 기반을 조성함으로써 E-mail에 첨부되어 오는 파일에 대한 검색, 개인 스케줄링, 메일 지식공유 및 관리를 보다 효율적으로 할 수 있다.



4.4 문서순위결정 알고리즘

문서들을 사용자 요구를 기준으로 분석할 때 분석기준과 문서들간의 관계를 파악할 수 있는데 분석된 결과를 가지고 사용자 프로파일과 비교하여 적합한 순서대로 문서의 순위를 정할 수 있다. 본 논문에서 제안하는 사용자 위주의 문서순위결정 알고리즘은 다음과 같다.

- 입력 : 1. 사용자 프로파일 2. 검색결과 문서
 출력 : 1. 순위가 결정된 검색결과 문서
 2. 갱신된 사용자 프로파일

User_centered_document_ranking

begin

1. 관심분야의 프로파일 벡터를 선택.
2. 질의확장
 IF (갱신 회수 > 5)
 then 사용자 프로파일을 이용하여 질의를확장
 else 사용자 프로파일과 시소러스를 이용하여 질의를 확장
3. 확장된 질의의 확인을 통한 갱신을 수행하여 사용자 프로파일 갱신
4. 검색결과 문서와 사용자 프로파일비교
5. 검색결과로 얻은 문서들을 분석
6. 분석결과와 사용자 프로파일을 비교하여 문서의 순위를 결정한 후 사용자가 원하는 범위내에서 결과를 출력
7. 사용자 적합성 피드백에 의한 갱신을 수행하여 사용자 프로파일 갱신

4.5 성능평가

본 논문에서 제안한 지능형 E-mail문서관리 시스템의 목적에 따라 다음과 같이 평가한다. 먼저 사용자 프로파일 갱신방법을 이용한 효율성 면과 문서순위결정 기법을 이용하여 사용자의 요구를 얼마나 충족시키는지 검증한다.

첨부파일문서	첨부명
D1	정보 필터링
D2	적합성 피드백을 이용한 필터링 에이전트
D3	향상된 인터넷 정보검색을 위한 사용자 모델과 필터링 에이전트
D4	정보 필터링 에이전트
D5	특정 순위결정을 위한 문서 필터링
D6	사용자 적응성 에이전트를 이용한 정보검색과 필터링
D7	정보 필터링과 검색

5. 결론

본 논문에서는 제안한 지능형 E-mail문서관리 시스템을 이용하여 사용자의 세부 관심분야에 해당되는 문서를 사용자가 원하는 형태로 제공함으로써 수많은 검색결과 문서중에서 최적의 문서를 선택하는데 소요되는 노력을 최소화 할 수 있다. 따라서 이 기술을 웹 기반 정보검색 에이전트에 결합하면 최상위 순위부터 검색결과를 제시하여 사용자에게 최적의 검색결과와 함께 검색 편의성을 제공 할 수 있다. 본 논문에서 제안한 지능형 문서관리시스템을 이용하면 문서의 관리나 처리를 훨씬 수월 하게 할 수 있을 뿐만 아니라 Attach 파일들을 종류별로 효과적으로 관리 할 수 가 있다. 앞으로 연구에는 E-mail정보를 XML구조를 이용하여 시스템을 설계하고 제안한 알고리즘으로 구현하고자 한다.

6. 참고 문헌

- [1].우선미 “사용자프로파일과 잠재적 구조분석을 이용한 검색된 문서의 순위결정기법”, 전북대학교 박사학위논문, 2001.8
- [2].정영미, “정보검색론”, 구미무역(주) 출판부, pp.1-325, 1993
- [3].이상진, “효과적 지식공유를 위한 동적 폴더의 구현”, 서울대학교 석사학위논문, 2002.2
- [4].유춘식,우선미, “자연어처리,통계적기법,적합성 검증을 이용한 자동색인 시스템에 관한연구”, 한국정보처리학회 논문지, 제5권, 제6호, pp.1552-1562, 1998
- [5].Bracha Shapira,et al. "information Filtering: A New Two-Phase Model using Stereotypic User Porfiling," Journal of Intelligent Information systems, Vol.8, 1997.