

개념 계층과 사용자 정보를 이용한 지능적 이미지 검색 시스템 설계

홍성용*, 나연복
단국대학교 컴퓨터공학과

Design of Intelligent Image Retrieval System utilizing Conceptual Hierarchy and User Information

Seongyong Hong*, Yunmook Nah
Department of Computer Engineering, Dan-Kook University

요 약

본 논문은 개념 계층과 마이닝 기법을 적용하여 이미지 데이터베이스에서 이미지 데이터에 대한 검색 기법을 지능화 하고 시스템을 효율적으로 관리하는데 있다. 개념 계층을 이용하여 이미지 정보를 지능적인 형태로 저장, 검색하고, 효율적인 웹서비스를 하기위해 사용자의 정보와 사용자가 이용한 질의 정보 그리고 이미지 정보를 통합하여 사용자에게 좀더 많은 부가 정보들 제공한다 또한, 사용자의 의도를 추측하여 사용자의 의도를 반영한 지능적인 이미지 시스템을 개발하기 위한 방법을 제시한다.

1. 서 론

최근 인터넷 분야가 급성장함에 따라 웹상에 있는 수많은 정보들 가운데에서도 멀티미디어 정보의 양이 급격히 늘어나고 있다. 따라서 멀티미디어 정보를 검색할 수 있는 여러 검색 방법이 시도 되어지고 있다. 그 중에서도 이미지 데이터의 양이 급증하고 있으며, 이미지 데이터를 검색하기 위한 가장 기본적인 방법으로 이미지 정보의 내용을 텍스트로 기술하고 그 기술된 내용에 의해서 검색을 하는 방식이 보편적으로 사용되고 있다. 한편 내용 기반 검색 시스템들은 이미지의 색상, 질감, 모양등 속성을 기반으로 하는 검색 방식이다. 그러나 이러한 검색 방법들은 사용자의 의도와 만족도와는 상관 없이 검색을 수행하고 있다. 본 논문에서는 이미지 정보를 개념 계층을 이용하여 표현하고 사용자의 정보를 추출하여 분석하고, 분석된 결과를 사용자에게 피드백(feed back)함으로써, 지능적인 이미지 검색을 수행하는 시스템을 설계 하였다. 본 시스템에서 이미지 검색은 사용자로부터의 지식 추출과 추출된 데이터를 데이터베이스에 저장하고, 지식탐사를 위해 주어진 데이터베이스로부터 대상이 되는 데이터를 선별하는 과정과 선별된 대상 데이터를 마이닝(mining)하여 얻어진 결과를 이해하기 쉽게 표현하는 과정으로 이루어 진다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에

대해서 살펴보고, 3장에서는 시스템 설계에 따른 시스템 구조와 개념 계층에 의한 이미지 정보의 표현 방법과 사용자의 정보 추출 방법에 대해 설명하고, 추출된 정보를 지능적으로 질의 처리 하기 위한 방법에 대하여 기술한다. 마지막으로 결론과 향후 연구에 대해 기술한다

2. 관련 연구

기존의 이미지 검색 시스템으로는 IBM사의 QBIC 이나 Columbia 대학의 Safe, VisualSEEK, WebSEEK, CBVQ와 같은 내용 기반과 텍스트 기반의 검색 시스템이 있다. 그러나 이러한 시스템들이 공통적으로 가지고 있는 단점은 사용자의 의견이나 만족도를 고려하지 못하고 있다. 그래서 지식 탐사나 데이터 마이닝(data mining)과 같은 연구가 활발히 이루어지고 있는데, 이러한 것은 사용자에게 좀더 많은 정보를 제공하고, 필요한 정보에 대하여 탐색 공간을 줄이고 탐색의 결과로 보다 흥미로운 지식을 얻을 수 있다[1]. 관계형 데이터베이스에서의 지식 탐사에 있어서 메경지식은 중요한 역할을 한다.

2.1 개념 계층

개념 계층(concept hierarchy)이란 개념 공간상에 존재하는 개념들을 레벨(level)에 따라 가장 일반적인 개념에서부터 그에 대응되는 구체적인 개념순으로 나열한 것이다. 개념 계층은 주로 트리(tree), 격자(lattice), DAG(directed acyclic graph)등의 형태를 가진다. Han

* 이 논문은 한국 과학 재단의 특정 기초 연구비 지원에 의한 것임.

과 Fu는 배경 지식으로 사용되는 개념 계층이 데이터베이스 상에서 나타나는 실제 데이터들의 분포에 따라서 특정 마이닝 작업에 적합한 형태로 동적으로 정제될 수 있으며, 또한 영역이 수치인 속성에 대해서는 자동으로 생성될 수 있음을 보였다[2].

2.2 지능 질의 처리

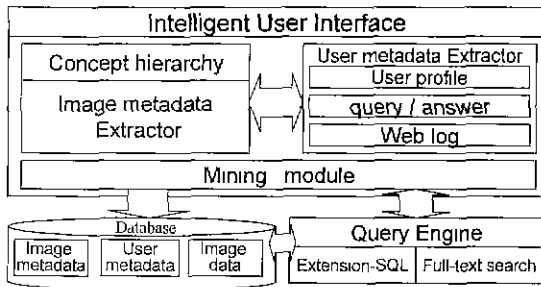
지능 질의 처리란 시스템이 사용자의 질의에 대해 구체적인 질의의 답만을 제공하는 것이 아니라 사용자의 분석 결과나 좀더 많은 정보, 또는 사용자의 의도를 추측하여 사용자의 의도를 반영한 정보를 제공하는 것이다[3]

2.3 사용자 접근 패턴 분석

웹 서버의 로그(log)파일은 웹 서버에 접속한 사용자, 액세스 시간, 접근한 웹 페이지와 같은 정보를 담고 있다. 이러한 웹 로그파일을 분석하여 얻은 결과는 웹 데이터를 관리하거나 사용자의 성향을 분석하는 등의 여러 가지로 이용될 수 있다[4].

3. 시스템 설계

본 논문에서는 [그림1]과 같은 지능적인 이미지 검색 시스템 구조를 제안한다

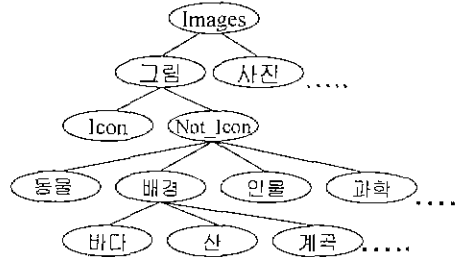


[그림1] 지능적 이미지 정보 검색 시스템

본 시스템의 사용자는 특정 분야에 국한되지 않은 사용자로 가정한다. 사용자 정보 인식 시스템은 사용자의 신상 정보나 사용자의 IP, 사용시간, URL 등의 정보를 자동으로 추출하여 데이터베이스에 메타데이터(metadata) 형태로 기록한다. 즉, 웹 로그를 데이터베이스에 기록 저장 하는 것이다. 이러한 사용자 정보를 수집하는 단계는 사용자가 처음으로 시스템에 접속하였을 때 이루어지며, 사용자의 마지막 요청후 일정시간 후에 종료가 된다 이미지 인식 시스템은 이미지에 대해서 그 이미지에 관한 정보를 개념 계층에 기반하여 표현되고, 이미지 자체와 그 이미지에 대한 메타데이터를 구분하여 데이터베이스에 저장 한다 새로운 이미지가 등록될 경우에는 이미지 데이터와 이미지 메타데이터를 개념 계층에 의한 분류를 하여 웹상에서 지능적으로 등록하기 위한 사용자 인터페이스가 제공된다. 마지막으로 사용자가 원하는 이미지를 지능적으로 검색하기 위해서 질의 하는 내용과 질의에 대한 답변 내용을 데이터베이스에 기록한다. 만일, 다른 사용자가 같은 질의를 하거나 비슷한 질의를 하였을 때, 사용자의 정보와 질의 정보를 다양하게 분석하여 부가적인 정보를 제시한다

3.1 개념 계층을 기반으로한 이미지 메타데이터

지식 탐사는 주어진 데이터베이스로부터 대상이 되는 데이터를 선별하는 과정, 선별된 대상 데이터를 적절한 형태로 변환하는 과정, 변환된 데이터에 대해 마이닝 알고리즘을 적용하여 얻어진 결과를 이해하기 쉽게 표현하는 과정으로 이루어 진다. 이러한 지식 탐사 과정에 있어서 배경 지식(background knowledge)은 중요한 역할을 한다. 개념 계층에서 부모와 자식간의 관계는 일반화된 개념과 그 개념에 대응되는 상세화 된 개념과의 관계이다. [그림2]는 이미지에 대해 정의된 개념 계층의 한 예를 보인다



CH on Attributes

- {그림, 사진} ⊂ Images
- {Icon, Not_Icon} ⊂ 그림
- {동물, 배경, 인물, 과학...} ⊂ Not_Icon
- {바다, 산, 계곡...} ⊂ 배경

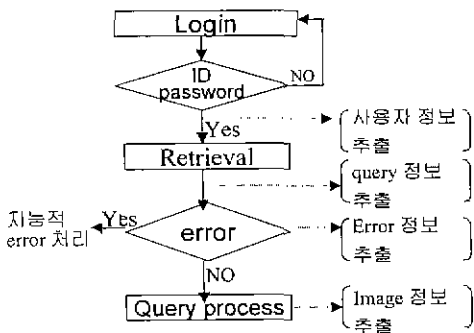
[그림2] 이미지 묘사 개념 계층의 예

3.2 사용자 모델링

사용자의 행동 양식을 관리하여 시스템을 지능적이고 효율적으로 만들 수 있다. 사용자의 질의 패턴에 따른 사용자들의 분류와 클러스터링(clustering)을 수행함으로써 사용자들에게 관련된 정보를 제공한다. [그림3]은 사용자의 질의 정보를 자동으로 추출하여 데이터베이스에 저장 하기 위한 과정을 나타내고 있다. 사용자의 질의에 대해 마치 사람이 대답하는 것과 같이 좀 더 의미 있는 결과와 부가적인 정보를 제공하는 협조적 응답(cooperative answering)방식이다 예로 "1999년 8월에 남자들이 가장 많이 검색한 이미지는 무엇인가?" 라는 질의에 대해서 단순한 형태의 답변 보다는 "검색 횟수는 255회 이고 주로 컴퓨터직종을 가진 남자들이 찾았던 이미지는 dbmm.jpg입니다." 라는 형식에 답변을 유추해 낼 수 있다. 또한 질의의 내의 가정 처리에 의해 잘못된 형태의 질의문, 예로 "이미지중에 전체가 모두 빨간색을 가지고 있는 이미지는?" 이란 질의에 대해서 단순히 에러나 아무런 정보도 표시하지 않는것 보다 "이미지 중에는 100% 빨간색을 가진 이미지는 없습니다" 라고 표현 할 수 있다

3.3 지능적 질의 처리

질의는 다양한 지능적인 인터페이스로 구성이 되며 사용자가 원하는 형태의 질의를 할 수 있고, 이미지의 경우 미니어췌(miniature), 즉 축약이미지(thumbnail)형식으로 보여지게 된다 이러한 이미지를 통해 직접 질의를 할 수 있고, 개념 계층 구조를 이용한 질의를 함으로써 직접적으로 원하는 이미지를 바로 찾을 수 있다. 시스템 내에는 각 속성의 도메인에 대해 이미 정의된 개념 계층이 존재한다. 지식 탐사의 대상이 되는 테이블과 속성들을 선택한다. 각 속성의 실제 값을 개념



[그림3] 사용자 질의 정보 추출 과정

계층의 일반화된 개념으로 치환하고 치환된 결과 같은 값을 가지게된 튜플들을 합치고 개수를 알아낸다. [그림4]는 이러한 과정으로 사용자의 접속 통계를 하기 위한 알고리즘을 나타낸다. 이러한 지능적인 질의를 하기 위해 확장-SQL문을 사용하고, 다양한 질의 형태를 제공하기 위해 전문 질의 기법을 사용한다[5]. 지능적 질의 처리 과정은 다음의 과정을 거쳐서 처리 된다.

[1단계] 데이터 추출 단계

- ① 트랜잭션에 따른 사용자 정보, 사용자 질의 정보, 이미지 정보를 데이터베이스에 기록.

[2단계] 질의 형태 분석 단계

- ① 사용자의 질의를 일반화된 형태로 가공
- ② 확장-SQL 로의 변환.
- ③ 질의 정보의 mining.
- ④ 지능적 Error 처리.

[3단계] 데이터 분석 단계

- ① 사용자 접속 횟수에 따른 분석.
- ② 사용자 그룹 클러스터 (직장별, 나이별...).
- ③ 검색된 이미지 분류

[4단계] 시스템 관리 단계

- ① 데이터베이스 중 자주 사용되는 부분의 변화 추이 분석.
- ② 사용자 그룹의 관심분야 추이 변동 분석

4. 결론

본 논문에서는 이미지에 대한 정보를 개념 계층을 기반으로 하여 표현하였으며, 사용자의 정보를 이용하여 이미지 정보 검색이 지능적으로 이루어 질 수 있도록 설계 하였다. 사용자의 정보는 접속 횟수, 직업, 시간, 국가, 나이별 등으로 그룹화 되어 분석 처리되고, 사용자에게 부가적인 정보로 제공되어지며, 검색시에 유용한 정보로 쓰이게 된다 이 시스템의 장점은 다음과 같다. 첫째, 기존의 특정한 질의방식만을 지원하는 이미지 검색 시스템들에 비해서 사용자에게 다양한 질의 형태를 제공한다. 둘째, 사용자의 정보를 자동으로 데이터베이스에 기록하고, 사용자의 질의나 답변에 대하여 다양하게 분석하여 그 결과를 사용자에게 제공하므로써 질의 효과와 사용자 만족도를 높일 수 있다. 셋째, 개념 계층을 기반으로한 이미지데이터의 표현으로 이미지 마이닝(분류 classification)을 할 수 있고, 검색의 폭을 줄이여 시간을 줄일 수 있다

향후 연구로 이러한 시스템이 좀더 지능적인 형태를 갖출수 있도록 자동으로 추출된 사용자의 정보에 대한

Procedure Count_Users

```
//전체 사용자의 개수를 구한다.
BEGIN row count
  If Result = EOF
    Then Error
  Else num=0,
  Do While Result ≠ EOF
    Then num=num+1
    Result,next
  Loop
END
```

Procedure Count_ID

```
//사용자 접속 통계
BEGIN
  For row count Do
    BEGIN
      item=0,
      row=0;
      //row 의 개수가 num 보다 작을때까지
      For row less then num Do
        BEGIN
          If (a row ≠ ID) Then
            row=row+1
          Else
            item=item+1
            row=row+1
          END
        END
      percent = (item/num)*100
    END
  END
```

num: Table record number
 item UserID
 row: Select record
 a : Table columne

[그림4] 사용자 접속 통계 알고리즘

다양한 분석과 마이닝 기술을 적용해 많은 정보를 사용자에게 제공할 수 있는 방법을 연구 개발하고, 이러한 과정을 자동으로 처리할 수 있도록 Transact-SQL을 이용해 개발할 예정이다.

[참고문헌]

- [1] 김성민, 이동하, 남도원, 이진영, "관계형 데이터베이스에서의 지식탐사를 위한 개념 계층의 재어기법", '98 한국전문가시스템 추계학술대회는논문집, 1998년 12월, pp. 83-90.
- [2] J.Han and Y.Fu, "Dynamic Generation and Refinement of concept Hierarchies for Knowledge Discovery in Database," AAAI'94 Workshop on Knowledge in Databases (KDD'94), Seattle, WA, July 1994, pp. 157-168.
- [3] J.Han, Y.Huang, N.Cercone, and Y.Fu, "Intelligent Query Answering by Knowledge Discovery Techniques," IEEE Transactions on Knowledge and Data Engineering, 8(3): 373-390, 1996.
- [4] D-H.Lee, D.-Y.Seo, N.-H.Kim, J.-Y.Lee, "Discovery and Application of User Access Patterns in the World Wide Web," Proceedings of the 4th World Congress on Expert Systems, March 16-20, 1998. pp. 321-327.
- [5] 권병희 지, SQL Server Bible Ver.7 영진출판사, 1999년