

개념 그래프를 이용한 개념 기반 검색시 검색 질의어의 문맥 확장

*
배환국, 전성진, 김기태
중앙대학교 컴퓨터공학과

{bae, sjjhun, ktkim}@ailab.cse.cau.ac.kr

Context Extension In Concept-based Searching Using the Conceptual Graph

*
Hwan-Kook Bae, Sung-Jin Jhun, Ki-Tae Kim
Dept. of Computer Science & Engineering, Chung-Ang University

요 약

웹그래프는 웹문서 간의 하이퍼링크를 각 웹문서의 핵심어간의 링크관계로 추상화하며, 이 관계를 이용하여 핵심어의 개념 그래프를 구축하고 질의의 확장이나 영역지식을 제공하는 개념 기반 검색이 가능한 검색 시스템이다. 본 논문에서는 웹그래프에서 가능했던 질의어에 대한 한 단계의 질의 확장에 그치지 않고, 최초의 질의어와 이후 확장어들에 대한 문맥을 유지하며 추가적인 다단계의 확장이 가능하도록 하는 방법을 제시한다. 검색 시스템 사용자는 최초 질의어가 속하는 해당 분야에 대한 지식이 부족한 경우, 문맥을 유지한 확장을 통하여 자신이 찾고자 하는 바를 명확히 해 나가며 세부 질의를 구축할 수 있고 해당 분야에 대한 지식을 얻는 것이 가능하다.

1. 서론

과거에는 서적에서 주로 정보를 획득하였으나, 정보의 바다라 불리는 인터넷이 널리 퍼지면서 새로운 정보원(Information Source)으로 점점 더 부각되고 있다. 새로운 지식 베이스인 인터넷에서 정보를 찾기 위해서 검색엔진은 필수적이다. 인터넷 상에서의 검색 방법 중 하나로 개념 기반 검색이 있다. 이 때 개념이란 특정 단어를 유사어 혹은 관계어로 확대시켜 나아가는 과정을 말하며, 개념 기반 검색은 단어의 의미를 분석, 단어의 개념 관계를 이용하여 검색을 확장하여 계층 표현을 가능하게 하는 검색 방법이다. 웹그래프는 개념 기반 검색 방법을 이용한 검색엔진이다[최 98]. 웹그래프는 웹 문서의 키워드를 추출하기 위하여 하이퍼링크의 앵커 텍스트(Anchor Text)를 이용하였다. 웹그래프는 하이퍼링크 정보 중 하나인 앵커 텍스트를 이용하여 웹 문서의 핵심어를 간단하고 빠르게 추출하고, 웹 문서 간의 하이퍼링크를 각 웹 문서의 핵심어간의 링크관계로 추상화하며, 이 관계를 이용하여 핵심어의 개념 그래프를 구축하고 질의의 확장이나 영역지식을 제공하는 개념 기반 검색이 가능한 시스템이다[최 98]. 검색하고자 하는 분야에 대하여 전문적인 지식이 있는 질의자는 해당 분야에 대한 정확한 지식을 가지고 세부 질의를 할 수 있으나, 전문적인 지식이 없는 질의자는 찾고자 하는 바를 정확히 알지 못한 채 질의하는 경우가 많다. 이런 경우 사용자는 최종적으로 웹 문서의 내용을 확인하기 전에 그래프의 확장을 이용한 네비게이션을 통해서 점차 세부적인 질의를 해 나갈 수 있으며, 이 과정을 통하여 자신이 잘 알지 못하는 분야에 대한 지식을 획득할 수 있다. 웹그래프는 검색엔진의 사용자가 입력한 최초의 질의어로부터 다음 단계의 관련어로의 확장을 하여 그리프로 보여 줌으로써 위와 같은 기능을 어느 정도 제공하였으나, 다시 그 관련어들을 계속해서 다음 단계의 관련어들로 확장할 때에는, 최초로 질의한 질의어로부터 최종 관련어까지의 관계를 유지하지 못한 채 확장함으로써, 추가적인 확장시 문맥을 유지하지 못한다는 단점이 있다. 본 논문에서는 웹그래프에서의 개념 그래프를 생성하는 방식을 확대하여, 최초의 질의어로 나온 개념 그래프를 추가적으로 확장할 때 문맥을 유지하며 확장할 수 있도록 하는 방법을 제시한다. 문맥을 유지한 그래프의 확장으로 질의자의 의도를 정확히 표현할 수 있는 상세 질의를 구축하도록 하여 검색에 도움을 줄 수 있도록 하는 것이 목적이다. 본 논문의 구성은 다음과 같다. 2 장에서는 본 논문에 대한 관련 연구, 3 장에서는 본 논문에서 제안한 문맥 유지 확장의 방법에 대하여 설명한다. 4 장에서는 설계 및 구현 환경을 소개한다. 5 장에서는 인공지능 분야에 대한 웹 문서들을 모은 후, 문맥을 유지 하지 않고 단순한 확장을 하는 경우와 각각의 문맥 유지 방법을 적용시켰을 경우에 문맥을 유지하는 정도를 해당 분야에 대한 전문지식을 가진

그룹이 평가하도록 하였다. 6 장에서는 5 장의 실험 결과를 기반으로 본 논문에 대한 결론을 맺으며 향후 연구 과제에 대해 소개한다.

2. 관련연구

2.1. 웹그래프

웹그래프란 웹문서의 링크 정보를 이용하여, 사용자의 질의에 도식화된 개념 그래프를 보여줌으로써 원하는 정보를 쉽게 검색할 수 있도록 하는 검색엔진을 말한다[최 98][박 00][이 00]. 이 검색엔진은 개념 기반 검색 기법을 도입하였다. 개념 기반 검색은 단어의 의미를 분석하여 단어의 개념 관계를 이용해서 검색을 확장하게 해줄 수 있는 검색 방법이다. 따라서 사람의 사고 방식과 유사하며, 효과적이라 할 수 있다. 또한 이 시스템은 키워드를 추출하는데 있어서 하이퍼링크 정보를 이용하였다. 이러한 하이퍼링크의 특징과 정보를 바탕으로, 요약 정보인 Anchor Text 나 Alt Text 에서 추출된 핵심어와 관련 웹 문서서 저장하는 특성을 이용하여 구축된 개념그래프를 사용하여 구축된 지식을 웹의 개념지식이라 한다[최 98][조 99].

2.2 개념 추출 방법

2.2.1 핵심어 추출

웹그래프는 웹문서의 핵심어 추출을 위하여 하이퍼링크 정보중 하나인 Anchor Text 와 웹문서의 제목(title) 태그를 이용한다[최 98]. 구체적인 핵심어 추출 방법은 다음과 같다.

웹문서 B 의 핵심어 = A1 + A2 + A3 + ... + Title

2.2.2 하이퍼링크 정보를 이용한 개념화

핵심어를 추출한 후 웹문서마다 하나 이상의 핵심어를 가지게 된다. 각 문서의 핵심어와 하이퍼링크인 링크를 이용하여 개념을 생성한다. 링크는 계층적 구조 혹은 내용의 참조를 나타내는데, 이를 이용하여 웹문서간의 계층적 구조와 참조가 핵심어간의 계층적 구조와 참조로 추상화되며, 이것이 개념으로 나타난다[최 98]. 특정 핵심어를 가지는 웹문서는 하나 이상이고, 웹문서는 하이퍼링크 정보를 이용하여 서로 연결하고 있다. 특정 핵심어를 가진 웹문서가 연결하고 있는 웹문서들의 핵심어 리스트중 동일한 핵심어를 가지는 것끼리 분류를 한다. 그리고 핵심어 별로 분류된 문서들은 링크로 추상화하여 웹문서간의 관계를 핵심어간의 관계로 바꾸어 개념을 생성한다.

3. 문맥 유지 확장

3.1 문맥 유지 확장

웹문서들은 단어 간의 의미관계를 정확하게는 표현하고 있지 못하다. 이러한 한계는 서로간에 잘 조직화되고 특정 영역 밖으로 넘어가지 않는 범위 내에서의 웹문서들을 정보원으로 이용할 때 상당 부분 극복되어질 수 있다. 본 논문에서는 '문맥'을 '개념 기반

검색엔진에서 사용자가 최초 입력한 질의어로부터 최종 질의 확장되어 나타난 관련어까지의 단어들 사이의 관계'라고 정의하기로 한다. 그리고, 이 정의에서의 '관계'란 단지 용어간의 의미상에서의 계층적이거나 동등한 관계만을 말하는 것이 아닌, 웹문서 저작자의 의도가 반영되어 나타날 수 있는 관계를 포함하는 것을 말하기로 한다. 우선, 그래프 상에서 최초 질의어로부터 다음 관련어로 확장하는 것을 1 단계의 확장이라고 하자. 관련어들 중 하나에 대해서 그래프 상에서 다시 다음의 관련어로 확장하는 것을 계속해서 2 단계의 확장이라고 하겠다. 2 단계의 확장으로부터 나온 관련어들 중 하나에 대해서 그래프 상에서 다시 다음의 관련어로 확장하는 것을 3 단계의 확장이라고 하겠다. 이렇게 계속되는 확장을 1, 2, 3, ..., n 단계의 확장이라고 지칭하기로 한다. 이러한 방식으로 그래프 상에서 관련어들을 계속 다음 단계의 관련어들로 확장해 나감으로써 질의 확장을 하게 된다. 각 단계의 확장을 위해 선택되는 관련어를 해당 단계의 확장 선택어라고 부르기로 한다. 예를 들어, n 단계에서의 확장을 위해 선택되는 관련어를 n 단계의 확장 선택어라고 지칭한다. 각 단계의 확장시마다 추출되는 관련어들을 구하기 위한 핵심어를 가지고 있는 웹문서들을 각 단계마다의 관련어 추출 웹문서로 부르기로 하겠다. 예를 들어, n 단계에서의 관련어를 추출하기 위한 핵심어를 가지고 있는 웹문서를 n 단계의 관련어 추출 웹문서로 지칭한다. 기존의 웹그래프에서는 n 단계의 확장을 위해서 단지 해당 단계의 확장 선택어를 핵심어로 가지고 있는 웹문서들을 그 단계의 관련어 추출 웹문서로 이용했다. 하지만 이렇게 해서 구해진 관련어들은 문맥을 유지하지 못하게 된다. 본 논문에서는 문맥을 유지하며 질의를 확장하기 위하여 방법을 제시한다. 3.2 에서 각각의 방법에 대한 설명을 한 후, 3.3 에서 질의, 확장처리 시의 구체적인 알고리즘을 설명하도록 하겠다.

[표 1] 용어 정의

n 단계의 확장	이전 단계에서 n 단계로의 확장
n 단계의 확장 선택어	n 단계의 확장을 위해 선택되는 관련어
n 단계의 관련어 추출 웹문서	n 단계의 관련어 추출을 위한 핵심어를 가지고 있는 웹문서
n 단계의 관련어	n 단계 확장시 나온 관련어

3.2 문맥 유지 확장의 방법

본 논문에서 제안하는 문맥 유지 확장의 방법을 분류하면 다음과 같다.

첫번째로, 핵심어간의 링크관계를 이용한 확장이다.

두 번째로, 핵심어들의 and 연산을 이용한 확장이다.

3.2.1 핵심어간의 링크관계를 이용한 확장

이것은 최초의 질의어를 핵심어로 가진 웹문서로부터 계속된 하이퍼링크의 연결된 상에 존재하는 웹문서들만을 관련어 추출에 이용하여 문맥을 유지하는 방법이다. 1 단계의 확장시에는 최초 질의어를 핵심어로 가지고 있는 문서들로부터 하이퍼링크로 연결된 문서들을 1 단계의 관련어 추출 웹문서로 한다. 이는 기존의 웹그래프와 같은 방식이다. 2 단계 확장시에는 기존의 웹그래프에서는 2 단계의 확장 선택어를 핵심어로 가지고 있는 웹문서들을 2 단계의 관련어 추출 웹문서로 하였다. 이 방법은 앞서 기술한 바와 같이 문맥을 유지할 수 없다. 링크관계를 이용한 방법에서는 1 단계의 관련어 추출 웹문서이면서 2 단계의 확장 선택어를 핵심어로 가지고 있는 웹문서로부터 하이퍼링크로 연결된 문서들만을 2 단계의 관련어 추출 웹문서로 선택한다. 다음 3 단계 확장에서도 마찬가지로 2 단계의 관련어 추출 웹문서이면서 3 단계의 확장 선택어를 핵심어로 가지고 있는 웹문서로부터 하이퍼링크로 연결된 문서들만을 2 단계의 관련어 추출 웹문서로 선택한다. 계속 확장되는 경우도 마찬가지로 n 단계의 확장 시에는 이전 단계의 관련어 추출 웹문서이면서 n 단계의 확장 선택어를 핵심어로 가지고 있는 웹문서로부터 하이퍼링크로 연결된 문서들만을 n 단계의 관련어 추출 웹문서로 선택한다.

3.2.2 핵심어들의 and 연산을 이용한 확장

최초 질의어와 이후 단계의 확장 선택어들을 모두 핵심어로 가지고 있는 웹문서들을 해당 단계의 관련어 추출 웹문서로 하여 관련어들을 구하는 방법이다.

3.3 질의, 확장 처리

질의 확장에 대한 내용은 정적으로 저장되어 있지 않고, 질의어와 각 단계의 확장 선택어에 따라 수행시에 동적으로 생성된다. 이러한

이유는, 모든 질의어와 확장에 대비한 그래프의 구축은 많은 시간적, 공간적 제약을 받기 때문이며, 모든 경우의 확장에 대비한 그래프를 미리 구축한다는 것은 사실상 불가능하다. 실제 질의, 확장이 처리되는 방법은 다음과 같다. 클라이언트는 사용자로부터 질의나 확장에 대한 요구가 있을 경우 JAVA Applet을 통하여 이를 입력받고, 최초 질의어에 해당하는 노드부터 확장이 요구된 노드까지의 단어들의 리스트를 서버로 전송한다. 그러면 서버에서는 단어 리스트를 이용하여 다음 단계의 관련어들을 추출해내고, 클라이언트로 전송하여 이 관련어들을 이용하여 갱신된 그래프가 사용자에게 보여지게 된다. 아래에 최초 질의처리, 핵심어간의 링크관계를 이용한 확장, 핵심어들의 and 연산을 이용한 확장 각각에 대한 알고리즘을 설명한다.

3.3.1 최초 질의처리 알고리즘

1. 질의어를 입력받음
2. If 질의어가 색인에 있으면
 질의어를 갖고있는 URL 들을 모음
 URL들의 모음을 m_URLKey 에 저장
 Else Return
3. m_URL_Key 에 저장된 각각의 URL 에 대하여
 3-1 URL 이 가지고 있는 하이퍼링크 정보를 구함
 3-2 하이퍼링크 정보를 m_DestURLKey 에 저장
4. m_DestURLKey 에 저장된 각각의 URL 에 대하여
 4-1 URL 이 가지고 있는 핵심어들을 구함
 4-2 핵심어들을 m_listKeyword(하이퍼 핵심어 리스트)에 저장
5. 핵심어 빈도에 따른 리스트 정렬
6. 질의어와 개념어를 개념 그래프 표현을 위한 일정한 포맷의 문자열 로 만들
7. 자바 애플릿에 파라미터로 전송

3.3.2 핵심어간의 링크 관계를 이용한 확장 알고리즘

1. Cnt 와 WordList 를 입력받음
2. WordList 를 노드별로 분리하여 순서대로 saWordList(스트링 배열) 에 저장한다.
3. saWordList[0]의 핵심어를 가지고 있는 URL 들에 대하여
 3-1 이 URL 들을 m_URLKey 에 저장한다.
 3-2 m_URLKey의 연결 URL 들을 m_DestURLKey 에 저장한다.
4. for i = 1 to Cnt-1
 4-1 m_DestURLKey 에 저장된 URL 들 중 saWordList[i] 의 키워드를 가지고 있는 URL 들을 m_URLKey 에 저장한다.
 4-2 m_URLKey의 연결 URL 들을 m_DestURLKey 에 저장한다.
5. m_DestURLKey 에 저장된 각각의 URL 에 대하여
 5-1 URL 들이 가지고 있는 핵심어들을 구한다.
 5-2 핵심어들을 m_listKeyword 에 저장한다.
6. 핵심어 빈도에 따른 리스트 정렬
7. 그래프 표현을 위한 형식의 문자열 구축
8. 자바 애플릿에 파라미터로 전송

3.3.3 핵심어들의 and 연산을 이용한 확장 알고리즘

1. Cnt 와 WordList 를 입력받음
2. WordList 를 노드별로 분리하여 순서대로 saWordList(스트링배열)에 저장한다.
3. saWordList 의 핵심어들을 모두 포함하는 URL 들을 m_URLKey 에 저장한다.
4. m_URLKey의 연결 URL 들을 m_DestURLKey 에 저장한다.
5. m_DestURLKey 에 저장된 각각의 URL 에 대하여
 5-1 URL 들이 가지고 있는 핵심어들을 구한다.
 5-2 핵심어들을 m_listKeyword 에 저장한다.
6. 핵심어 빈도에 따른 리스트 정렬
7. 그래프 표현을 위한 형식의 문자열 구축
8. 자바 애플릿에 파라미터로 전송

4. 설계 및 구현

본 시스템은 웹문서를 모아서 데이터베이스에 저장하는 수집기(Spider)와 모아진 데이터베이스의 내용을 토대로 핵심어를 추출하고 이를 인덱스화 하는 색인기(Indexer), 그리고 질의를 처리하는 질의 처리기(Query Processor), 끝으로 질의의 결과를 시각화하여 보여주는 시각화 인터페이스 부분으로 구성되어 있다. 전체 시스템은 크게 Off-line Batch Job 과 On-line Processing 부분으로 구분한다. Off-line Batch Job 부분은 웹문서를 모아서

인덱스를 구축하는 부분으로 최종적으로 인덱스 파일 시스템의 구축을 목표로 한다. On-line Processing 부분은 사용자의 질의를 받아 인덱스 파일 시스템을 이용하여 그래프의 결과를 보여주는 기능을 담당한다. 자료를 수집하는 과정은 지속적으로 필요하며 자료의 양이 방대하기 때문에 엄청난 시간이 소요된다. 반대로 질의 처리는 시스템의 반응시간이 신속하여야 한다. 이런 이유로 수집과 질의 처리 부분을 Off-line 과 On-Line 으로 분리하였다. 전체적인 동작은 Off-line 으로 웹문서를 수집하여 데이터베이스에 저장한 후, 색인기를 통하여 인덱스 파일을 만든다. 검색 시스템이 웹 서버를 통하여 질의나 질의 확장에 대한 요구를 받으면, 질의처리가 인덱스 엔진을 통하여 인덱스 파일을 접근하여 그래프를 작성한 후 클라이언트로 전송한다. 클라이언트는 최초 질의시에는 전송받은 그래프를 보여주며, 추가 질의 확장 시에는 전송 받은 그래프를 기존에 가지고 있던 그래프에 추가하여 갱신된 그래프를 보여준다.

5. 실험 및 평가

5.1 실험 환경

인공지능관련 초기 URL 들을 시작점으로 하여 수집기를 돌려 각 URL 내에 존재하는 하이퍼링크를 따라가며 웹문서들을 수집하였다. URL 을 저장하기위한 큐로 DB 를 사용하였으며, 수집된 URL 은 총 293481 개, URL 에 존재하는 하이퍼링크는 740570 개, 실제 방문한 URL 은 39849 개이다. 검색 엔진 서버는 WindowsNT 상의 Internet Information Server 에서 동작하는 ISAPI Server Extension 을 사용하여 구현하였다. 실험평가는 [표 2]의 최초 질의어들에 대하여 인공지능에 관한 전문지식이 있는 5 명의 전문가 그룹이 문맥유지를 하지 않고 확장하는 경우, 핵심어간의 링크관계를 이용하는 경우, 핵심어들의 and 연산을 이용한 확장의 경우, 링크관계와 and 연산을 모두 이용하는 경우로 나누어 각각 5 단계까지 확장을 해 나가며 각각의 단계마다 문맥이 유지되는 정도를 평가자 각각이 자신이 가지고 있는 인공지능에 대한 지식을 바탕으로 문맥유지에 대한 만족도를 0%에서 100%사이에서 평가하도록 하였다[표 1]. 본 논문에서는 확장시의 관련어를 24 개로 제한하였는데, 한 번의 확장시마다 관련어 수의 제곱으로 가능한 확장의 수가 늘어나므로 모든 경우의 가능한 확장에 대하여 수치적으로 정확하게 평가하는 것은 힘든 일이다. 따라서 각 평가자는 자신이 가지고 있는 인공지능에 대한 기존 지식을 바탕으로 여러 관련어들 중 문맥을 유지하지 못하는 것으로 판단되는 단어에 대하여는 이후에 확장선택어로 선택하지 않고, 이전 단계의 단어들에 대하여 문맥을 유지한다고 판단되는 단어들에 대하여는 각각에 해당하는 웹문서들을 확인하도록 하여 해당 관련어와 이에 해당하는 웹문서의 실제 내용이 최초의 질의어로부터 해당 단계의 관련어까지의 문맥유지도를 평가하고 추가적인 확장을 통하여 각 단계별로 평가하도록 하였다.

5.2 실험결과

[표 2] 전문가 집단의 문맥유지도 평가

문맥유지방법 \ 확장단계	확장단계				
	1	2	3	4	5
And	83%	80%	75%	38%	15%
링크	83%	71%	57%	44%	27%
and, 링크	83%	78%	69%	56%	42%

[표 2]은 앞서 설명한 평가방법으로 각각의 평가자들이 3 가지의 서로 다른 문맥유지방법 각각에 대하여 단계별로 평가한 문맥유지도를 평균을 내어 계산한 결과이다. 표의 내용을 보면, 우선 모든 경우에서 1 단계는 같은 문맥유지도를 나타냈다. 핵심어들의 and 연산을 이용하여 확장을 하는 경우에는 약 3 단계까지는 높은 문맥유지도를 나타내다가 4 단계 이후에는 급격히 떨어졌다. 이것은 많은 핵심어들을 한꺼번에 모두 포함하는 웹문서의 희소성으로 인해 핵심어들의 and 연산만을 이용한 계속적인 확장은 한계에 부딪히기 때문이다. 핵심어간의 링크관계를 이용하여 확장을 하는 경우에는 2 단계 확장까지는 70%이상의 높은 문맥유지도를 나타냈지만 3 단계 이후에는 비교적 낮은 문맥유지도를 나타냈다. 마지막으로 핵심어들의 and 연산과 핵심어간의 링크관계를 함께 적용하여 확장을

한 경우에는 1 단계에서 3 단계까지는 and 연산만을 이용한 경우와 비교하여 문맥유지도가 비슷하거나 약간 떨어지는 경향을 보였다. 하지만, 4-5 단계에서는 and 연산이나 링크관계 중 어느 하나만을 이용하는 경우보다 더 나은 문맥유지도를 보였다.

6. 결론 및 향후 연구과제

본 논문에서는 개념 기반 검색 엔진인 웹그래프의 개념 그래프에 기반하여 핵심어들의 and 연산과 핵심어간의 링크관계를 이용한 문맥 유지 확장 방법을 제시하였다. 핵심어들의 and 연산을 이용한 확장은 둘 이상의 단어가 함께 복합어를 이루는 경우에 좋은 결과를 나타내며, 핵심어간의 링크관계를 이용한 확장은 관련어에 대한 내용을 한 웹 문서에서 모두 설명하지 않고 링크로 가리키고 있는 웹 문서에서 설명하는 경우에 좋은 결과를 나타낸다. 핵심어간에 이러한 확장 방법을 통하여 사용자는 최초의 질의어와의 관계를 유지하며 그래프 상에서 질의를 확장해 나갈 수 있고, 질의 확장 과정 중에 해당 영역에 대한 지식을 획득할 수 있으며, 구축된 상세 질의를 반영하는 결과를 얻을 수 있다. 검색하고자 하는 분야에 대하여 전문적인 지식이 있는 질의자는 해당 분야에 대한 정확한 지식을 가지고 상세 질의를 할 수 있으나, 전문적인 지식이 없는 질의자는 찾고자 하는 바를 정확히 알지 못한 채 질의하는 경우가 많다. 이런 경우 사용자는 최종적으로 웹 문서의 내용을 확인하기 전에 그래프의 확장을 이용한 네비게이션을 통해서 점차 세부적인 질의를 해 나갈 수 있으며, 이 과정을 통하여 자신이 잘 알지 못하는 분야에 대한 지식을 획득할 수 있다. 향후 연구 과제로는 다단계 확장시 문맥의 유지도를 최대한 유지하는 방법에 대한 연구가 필요하며, 특정 분야의 웹 문서를 수집함으로써 질의 확장 시 정확도를 높이기 위하여 웹 문서를 분야별로 분류하는 방법에 대한 연구가 필요하다.

참고문헌

[김 92] 김명철, 이운재, 최기선 김길창, "시소러스 작성을 위한 개념 획득 도구", 4 회 한글 및 한국어정보처리 학술발표논문집, pp39~49, 1992
 [박 00] 박사준, 김상경, 황수현, 김기태, "전문가 검색 엔진에서 개념그래프를 이용한 Web 정보 획득", 한국정보과학회 학술발표논문집 제 27 권 1 호 2000.4
 [이 98] 이재운, 김태수, "WordNet 과 시소러스", 제 11 회 언어정보연찬회 발표논문집, 언어 정보의 탐구 1, pp232~269, 1998
 [이 00] 이권국, 신일수, 이상준, 김기태, "전문가 검색엔진에서 데이터 마이닝을 이용한 개념관계 추출", 한국정보과학회 논문집 제 27 권 1 호 2000.4.
 [조 99] 조민재, "웹의 개념 지식을 이용한 자동 시소러스 생성법의 설계 및 구현", 중앙대학교 92 회 석사학위 논문, 1999
 [최 98] 최준영, "인터넷상의 하이퍼링크를 이용한 개념 그래프 기반 검색 시스템", 중앙대학교 90 회 석사학위 논문, 1998
 [Berry98] Michael J. A. Berry, Gordon Linoff, "Data Mining Techniques: For marketing, Sales, and Customer Support", Wiley Computer Publishing, pp 216~242, 1998
 [Brin98] Sergey Brin, Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Proceeding of 7th World Wide Web Conference, 1998
 [Chen95] Hsinchun Chen, Chris Schuffels, Rich Orwig, "Internet Categorization and Search: A Self-Organizing Approach", .http://ai.bpa.arizona.edu/papers/som95/som95.html
 [Jiawei2001] Jiawei Han, Micheline Kamber, "Data Mining - Concepts f. and Techniques", Morgan Kaufmann 2001
 [Klienberg98] Jon M. Klienberg, "Authoritative Sources in a Hyperlinked Environment", Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, 1998
 [Kosala2000] R. Kosala and H. Blockeel, "Web Mining Research: A Survey", SIGKDD Exploration, June 2000, Volume 2, Issue 1