

논문 2012-49CI-2-7

효율적 검색의도 파악을 위한 쿼리 단어 가시화에 관한 연구

(Revealing Hidden Relations between Query-Words for an Efficient
Inducing User's Intention of an Information Search)

권순진*, 홍철의**, 김원일***

(Soonjin Kwon, Chuleui Hong, and Wonil Kim)

요약

본 논문은 검색 질의어 단어 입력 시에 드러나지 않은 쿼리 단어들을 가시화함으로써, 검색 주체의 의도 구조를 선택할 수 있게 하며, 탐색 효율을 제고하도록 제안한다. 검색 질의어를 입력할 때 검색 의도를 파악한다면 효과적인 검색 서비스가 가능할 것이다. 이렇게 하기 위하여, 의도 구조와 요소를 설정함과 아울러 쿼리 단어의 생략된 관계에 해당하는 단어를 복원하여 가시화하는 과정이 필요하다. 관련된 연구들을 검토하며, 검색 의도 구조를 정의하고, 쿼리 단어의 가시화를 위한 방법과 의도 구조에 적합한 쿼리 단어를 확장 생성하는 과정을 보인다. 이 과정에서 의도 구조의 여러 계층 중 하나를 쿼리 단어 범위로 할당하는 예제와 실험을 수행하였다. 탐색 효율 상승의 검색결과와 탐색효율 하락의 검색결과를 분석하였다. 향후 연구로는 의도 결절을 확장하여 구성 요소를 학습할 수 있도록 자동화하는 연구가 필요하다.

Abstract

This paper proposes to increase an efficiency of somebody searching information by a visualization of an unseen query words with well-selected user's intent structures. If a search engine identifies user's intent to pursue information, it would be an effective search engine. To do so, it is needed that relationships between query-words are to be visible after recovering words lost during formulated, and that an intention structure/elements is to be established. This paper will review previous studies, after then, define a simple structure of the search intent, and show a process to expand and to generate the query words appropriate to the intent structure with a method for the visualization of the query words. In this process, some examples and tests are necessary that one of the multiple intent structured layers is to assign to a range of query-words. Increasing/Decreasing an efficiency are analyzed to find. Future research is needed how to automate a process to extend structural nodules of user's intent.

Keywords : Visualizing queries' relationship, information search intent, query relation type, Intention structure, Intention identifier

I. 서론

본 논문은 어떤 주체가 정보를 검색(Information

Retrieval)할 때 드러나지 않은 쿼리 단어들을 가시화함으로써, 미리 정의된 사용자 의도구조를 선택할 수 있게 하며, 탐색 효율을 제고하도록 제안한다. 검색 질의어를 입력할 때 검색 주체의 의도를 파악한다면 효과적인 검색 서비스가 가능할 것이다. 이렇게 하기 위하여, 의도 구조와 요소를 설정함과 아울러 쿼리 단어의 생략된 관계에 해당하는 단어를 복원하여 가시화하는 과정이 있어야 한다.

검색 주체의 정보 요구에 적절한 정보를 가져다주는 것은 정보 탐색의 중요한 과제이다. 그러나 현행 검색

* 학생회원, *** 정회원-교신저자, 세종대학교 디지털콘텐츠학과

(Department of Digital Contents, Sejong University)

** 정회원, 상명대학교

(Department of Computer science, Sangmyung University)

접수일자: 2011년2월15일, 수정완료일: 2012년3월3일

결과로서 검색엔진을 사용하는 사람에게 많은 참고 자료를 주지만, 이들은 보조 자료에 불과하고 이마저도 대부분이 정보 요구에의 부합성이 떨어진다. 이의 원인은 정보 요구와 검색 대상 자료의 구조가 불일치한 결과에서 초래되며, 무엇보다도 정보 요구에 대한 검색결과와의 적합 여부를 사람이 판별하기 때문이다. 따라서 사용자가 사물인 경우에도 이 둘을 비교시킬 수 있도록 공유할 수 있는 새로운 중간구조가 필요함을 의미한다.

본 논문은 다음과 같이 구성하였다. 복층의 의도 구조 및 쿼리 단어에 관해 관련연구들을 검토하였으며, 쿼리 단어들의 형성 흐름을 분석하여 관계양상을 도출하였다. 이렇게 함으로써, 사용자가 쿼리를 입력하기 직전 단계에서의 의도를 회상하는 실마리를 제공할 수 있다. 이를 숨겨진 쿼리 단어의 관계를 가시화하는데 활용한다. 두 질의어의 관계에 잠재된 특정 관계 유형을 정의하고 드러내는 가시화 과정을 거침으로써, 질의 표현 시에 탈락되는 의도표상에 관한 단서를 드러내고자 하였다. 가시화 과정에는 첫째, 관계의 형태와 내용(관계 항목의 프레임과 구조 및 계층, 요소, 값 등), 둘째, 관계 유형 식별자의 결정 및 Tagger 표현과 부착 방법, 셋째, 관계 유형의 식별과정 규칙 저장, 넷째, 관계 온톨로지 모형 및 저장 기준 등이 포함된다. 구문의 의미 관계를 표현한 데이터 배열로 실험 모델을 구성하는 예제와 검색 엔진을 이용한 실험이 수행될 것이다.

II. 관련 연구

관련 연구는 크게 두 부류로 나누어 검토한다. 의도 및 의도 구성 요인에 관한 연구, 그리고 질의어 및 질의어들의 관계 가시화에 관한 연구이다.

1. 의도 모델

의도모델에 관련된 연구로는 다양한 연구가 있다.

Carberry는 정보추구자의 질의를 추동하는 과업관련 계획에서의 시스템의 신념(system's beliefs) 표현을 다루었다.^[1] 후보계획과 맥락모델 사이의 관계성들을 예로 들어 적용하고 있으며 의도된 계획을 일부 다루고 있다.

Lee and Liu는 목표기반 의도 추출에 관해 다루었다.^[2] 목표구조(goal structure)를 정의하고, 목표의 속성/방향/행선지의 계층구조를 보였고, 사용자 의도들을 추출하는 과정, 단순화시킨 개체 클래스 온톨로지, 그리

고 목표선택 과정과 아키텍처를 다루었으며 물품구입시의 실제 예를 보였다. 배경지식의 중요성을 주장하였다.

또 다른 연구에서, Cantador 등은 특정 도메인 내의 사용자 선호들을 다층 온톨로지 기반의 복합 추천 모델로 제안하였다.^[3] 협력적 추천을 실제 사용자 프로파일을 통해 수동과 자동으로 실험하였다. 여러 가지 추천 방법들을 혼합하여 제안하고 있으며 개인화 온톨로지 기반 콘텐츠 검색(Retrieval) 모델에서 가중치를 부여한 의미적 선호성을 사용하고 있다. 추천 기법을 단순화하여 의도어 추천에 응용가능성이 있다. 그러나 과정상의 변수가 많아서 복잡하고 개인프로파일을 의도자원으로 전환하는 문제는 정보를 제한하게 되는 윤리적 문제가 있어 더 연구가 필요하다.

이 외에도 Schmitz의 연구, Ponnada와 Sharda의 연구, Rose와 Levinson의 연구는 웹에서의 목표와 온톨로지 유추를 다루었다.^[4,6,8] Rao와 Geogeff의 Multi-Agent Systems에 의한 결정과정, Chulef와 Read and Walsh의 목표의 계층구조에 관해 연구하였다.^[5,7]

이들의 연구를 정보 검색 프로세스와 대비하여 구조화해보면, 의도 기반 및 복층 구조의 온톨로지 모델을 형상화할 수 있다. <그림 1>에서는 'Denny's burger'라는 질의어 평면 위의 상위 구조를 표현하였다.

<그림 1>에서 정보를 추구하는 질의어가 어떤 의도를 가지고 출발하였는지를 유추한 경우의 의도 구조와

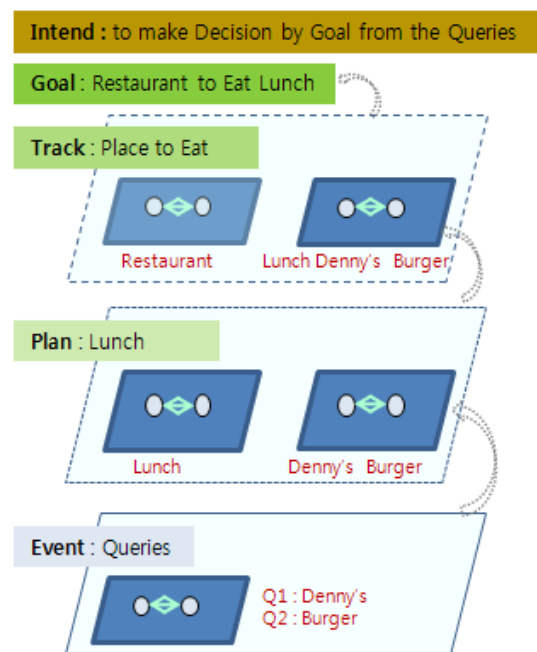


그림 1. 의도와 질의 과정
Fig. 1. Flow from Intention to Query.

속성을 나타내었다.

정보 검색 분야 연구들의 대부분이 사용자 의도를 파악하는 방법으로서 프로파일이나 웹상의 스니펫을 활용하는 것을 볼 수 있다. 그러나 실세계의 실체와 달리 사고의 흐름인 의도는 구조파악이 어려운 면이 많아 검색에 적용하는 검색 의도 연구는 모험적인 연구이다. 주로 개인화 검색에서 프로파일을 형성하거나 웹에서 검색 주체가 서핑한 흔적을 목록화한 스니펫을 활용하는 연구 등이 있다. 웹에서 들면 프로파일링 방법이 있다. 이 방법은 의도를 구조적으로 파악하기에 미흡한 데다 프라이버시를 침해하는 정보 윤리적 법률적인 분쟁의 소지가 있다. 개인 프로파일을 의도자원으로 전환시키는 것은 또 다른 차원의 별개의 연구 문제로 볼 수 있다. 국내의 연구로 스니펫에 의한 방법이 있다.^[9] 이 방법은 동일한 성공적 검색 흐름이 재구성될 가능성이 적다는 한계가 있다.

따라서 검색 의도를 명시적으로 밝혀져 최소한의 구조화로 사용자 스스로 검색 흐름 도중에 경로를 이탈하지 않도록 하는 정도에서 검색 성능 향상에 활용하는 연구가 바람직하고 합리적이다. 본 연구의 결과를 응용한다면, 검색 주체가 사물인 경우에도 유용할 뿐만 아니라, 프라이버시 침해의 소지가 없으므로 여러 면에서 더욱 유용할 것이다.

다음으로 의도의 구성 요소에 관한 관련 연구를 검토한다.

앞서 검토하였던 의도 모델 연구에서는 목표와 경로와 출발점이 나타난다. 구체적으로 의도에 영향을 미치는 선행 요인에 관한 대표적 연구를 보자면 Spyrou 외, Cohen 외, Kroy, Taylor의 연구를 들 수 있다.

Spyrou 등은 의도의 이론과 모델을 연구하고 침해탐지에 적용하도록 연구하였다.^[10] Cohen 등의 연구에서는 에이전트의 신념/행동/의도를 다루면서 의도의 속성과 논리에 관한 원리를 제시하였다.^[11] Kroy는 정보 수집 과정에서 의도 생성과정과 언어로 수행의도가 변환되는 과정을 보였다.^[12] Taylor는 정보욕구가 내재적/의식적/공식적/절충적 니즈로 구체화 됨을 연구하였다.^[13]

검색 주체의 의도는 검색 목표가 수행될 경우에 달성될 수 있으며 목표를 달성할 수 있도록 계획이 전제되어야 한다. 왜냐하면 질의 행위는 정보 추구 목표나 의도가 전제되어야 의미가 있기 때문이다.

목표 모형은 계획과 시작점 그리고 도달점 방향과 경로 등으로 구성할 수 있다. 실험 시스템의 복잡도를 고



그림 2. 의도 모델 선택
Fig. 2. Selecting Intention model.

려하고 연구 목적 달성을 위해 최소한의 의도 구성 요소로서 [목표-경로-출발점]으로 정의한다.

이 모델을 정보 검색과 관련하여 간단히 표현하면, 시스템 구성요소와 처리방법으로는, 목표의 내용과 방향 계획에, 관계의 식별, 목표 모델의 탐색, 사용자 의도 표현, 후보 목표 모델 선택의 단계가 필요할 것이다. <그림 2 참조>

의도기반 정보검색에는, 질의어들의 관계를 가시화한 후에 식별하고, 유추할 목표 모델을 의도 기반의 목표 모델 중에서 탐색하여, 사용자 의도를 포함한 질의어로 재생성하는 과정들이 있어야 한다. 즉, 질의어들의 관계 유형과 구조를 미리 정의하여 저장해 두고, 질의어 입력 시에 관계를 동적으로 시각화하여 제시함으로써, 입력자가 관계를 선택하게 할 수 있게 하면, 다양한 의도에 대처하며 선택적으로 개입하게 할 수 있는 의도 자원이 형성될 것이다.

2. 쿼리 단어 생성 흐름과 질의어 관계 양상

검색 주체의 질의 의도를 파악하기 위해서는 우선 질의어의 구체화 과정을 보아야 한다. 즉 색인어 집합을 보는 것이 아니라, 질의어가 나오게 된 역방향을 추적해야 하는 것이다. 이전의 정보 검색(Information Retrieval)에 관한 연구들은 색인어 방향으로 많은 연구가 이뤄져 왔으나, 정보검색 주체의 의도를 파악하는데 기여하지 못하였으며, 검색 주체에 관해 연구하더라도 색인어에 연관된 맥락(Context)을 파악하는 정도에 그칠 수밖에 없었다.

질의자의 의도가 구체화되는 과정은, 무의식적/암묵적/함묵적 요구로부터, 의식적 요구 과정을 거쳐 언어 표현으로 형식화 변화하며 수행의도의 결과가 질의어로 구체화되거나 절충되고 있다.^[12,13] 따라서 질의 요구의 구체화 단계에서 질의자가 어떤 문제에 봉착하여 정보요구가 발생된 것인가를 알고, 거기에 맞는 질의어를 구체화하는 것인가의 관점에서 문제를 인식하여야 할 것이다. 이러한 정보 추구 행동에 있어서 단계마다 요구되는 정보 또는 매체가 달라진다. 이들 단계마다 검색 의도를 파악하기 위해서는 각 단계별로 정보(요구)

주체에게 대화와 질문을 해야 하나 이로 인해 시스템도 사람도 모두 번거로워진다. 따라서 사용자의 불편을 최소화하는 수준에서 정보 요구 만족의 해법을 찾기 위해서, 검색창에 입력되는 질의어와 그 관계 모델에서 함수와 변수를 추적 및 설정하는 방법이 최선이라고 판단하게 된다.

본 논문에서는 그 방법을 찾기 위하여 질의어의 관계를 드러내고 식별자를 설정할 수 있도록 한다. 대부분의 검색 질의어는 평균 2.35 개로 구성되므로 질의어를 특정 2개로 한정하여 질의어 관계를 가시화할 것이다.^[14,15] 용어 개체들의 관계 종류를 분석한 연구(백지원의 2006)를 보면, 계층 관계나 포함 관계 등의 여러 가지 관계 양상이 있다.^[16] 이들 연구를 참조하여 특정 관계에 대한 형태를 분석하고 관계 유형 등을 검토하며, 특정 관계를 선정한 후, 관계를 내포하는 형태와 단서를 부착할 수 있을 것이다. 이 과정을 중심으로 관계를 표현한 데이터 배열을 구성할 수 있게 된다.

질의어 '종로 비빔밥'과 'Denny's Burger'의 예를 든다. '종로'와 'Denny's'는 장소, 지역 및 식당 등의 속성을 가지고 있으며, '비빔밥'이나 'Burger'는 음식명, 맛, 조리법, 식당 등의 속성을 가진다. 따라서 속성들의 연결 관계를 확장해 나가면, 관계 의도에 해당하는 맛있는 식당에 관한 정보의 연결 관계를 드러낼 수 있을 것이다. 관계 확장 과정에서 매칭되는 관계 항목의 프레임과 구조 및 계층, 요소, 값 등을 구성할 수 있게 된다. 이들 관계 유형의 구체적 항목을 식별할 수 있도록 식별자를 결정하고 Tagger를 부착하며 그 방법을 기술하면 관계 유형의 식별 과정이 되며, 이들 규칙을 저장하면 관계 온톨로지 형태로 가시화 될 수 있다. 관계 온톨로지 구축에 관한 방법론적인 것은 다른 연구에서 연구성과를 기대한다.

III. 검색 의도 정의

검색 주체의 질의 의도는 정보 요구를 유발하는 과정이며, 정보를 검색하기로 판단한 합리적 이유들이 포함된다. 의도 구조에 질의어 관계 집합을 매칭할 수 있게 하려면 몇 가지 제한점이 따른다.

가장 큰 제한점은 의도의 공간 범위를 축소하거나 구조를 단순화해야 한다는 것이다. 왜냐하면, "의도"가 갖는 의미는 포괄적이며 추상적이기 때문이다. 포괄적인 집합 요소의 많은 구성 요소 모두에다가 질의어 관계를

비교할 경우에, 합리적인 시스템 구성이 어려워 질 수 있다. 또한 추상적인 개념은 구성 요소의 추출에 많은 시간이 소요되기도 하고 자동화하는 데 많은 시간과 비용이 투입되어야 할 것으로 예상되기 때문이기도 하다.

두 번째 제한 점은 해 공간을 단순화해야 한다는 점을 들 수 있다. 질의 의도는 질의어와 질의어 관계 집합을 모두 포함하는 한 계층 위의 개념인 의도(Intent, Intention, Intending to do)가 나올 수 있다. 질의어 차원의 한 계층 위의 개념이 나타나도록 추론하되 정보요구가 발생하는 공간을 가정하여야 한다. 즉, 정보 검색 주체가 어떤 문제 극복의 과정에 놓여 정보 추구 행동을 검색 질의어로 표현하게 된 것인가를 가정하게 된다. 문제 극복 과정을 도입 및 정형화하고, 추론할 수 있는 복합적인 요소를 가정하되, 질의와 상위 구조의 개념이 동일 차원의 공간에서 다루어질 수 있도록 단순화한 공간을 가정하여야 할 것이다. 왜냐하면 차원이 다른 공간에 존재하는 요소일지라도 직선 노드를 가정하게 되면 동일 차원의 해 공간이 구성될 수 있기 때문이다.

이렇게 가정하게 되면, 여러 가지 복합적인 요소가 제거되면서 정보 탐색 패턴의 최소 범주 이내에서 해 공간을 구성할 수 있게 될 것이며, 간단한 실험으로 증명할 수 있게 된다. 따라서 의도를 {의도의 시작}과 {의도의 중간 과정, 전략} 및 {의도의 끝}의 세 요소 집합으로 한정하고 이 집합 요소에 질의어 관계를 매칭하는 관계로 추론하도록 하는 것이 합리적일 것이다. 매칭 방법은 메타데이터에 의한 자동 매핑이나, 또는 전문가에 의한 직접 매칭 등이 있을 수 있다. 매핑시스템에는 매핑데이터 생성코드와 개념정의 및 규칙이 필수적이다. 질의의도의 연관성을 효율적으로 탐사할 규칙 생성에 관한 연구는, 포괄적 집합군 탐색 방안과 의도구성 및 구성요소의 속성상속에 의존이 예상되므로 난이도가 높아 후속 연구로 별도로 진행되고 있다.

1. 질의어 Q_1 , Q_2 의 관계와 의도 공간 설정

관계 R 은, Q_1 과 Q_2 의 곱집합 ($Q_1 \times Q_2$) 의 부분 집합 $G(R | R=q_1, q_2, \dots, q_n)$ 로 가정하고, $q_1 R q_2$ 로서 표기한다. 질의어 관계 집합인 $q_1 _R q_2$ 는 질의어 Q_1 과 질의어 Q_2 의 함수이다.

$$R = f(Q_1, Q_2)$$

질의어 관계(R)의 공간 범위는 질의 의도(I)의 집합

범위 내에서 평면상의 확장 범위로 가정한다.

$$R = f(Q1, Q2)$$

$$I \supseteq (R , Q1 , Q2)$$

2. 질의 의도(I) 과정

의도의 과정을 $I = [Is , Ip , Ie]$ 로 나타낼 수 있다.

Is 는 {의도, 시작} : 정보획득이 필요하게 된 이유와 배경에 해당, 즉 Why에 해당하며 ‘점심’, ‘Lunch’ 등이다. 질의가 발생한 이벤트의 상위 구조이다.

Ip 는 {의도, 중간} : 정보 대상물의 획득 과정에 해당, 즉 How, When, Where에 해당하며 ‘식당’, ‘Restaurant’이며, 검색 매체 상에서 Track의 역할이지만 장소 의미로 한정하여 구분한다.

Ie - {의도, 끝} : 질의 결과 취할 목표물에 해당, 즉, What, Who에 해당하며 ‘비빔밥’, ‘Burger’이다. 이는 현재 검색창에 발생한 이벤트로 표현되어 있으므로 이로부터 취득된다.

3. 질의어 관계 모델 유형의 범위와 매칭

위에서 나타난 질의 의도의 자원별로 관계(R)을 할당하여 모델의 유형을 설정할 수 있으며, 그 결과 다음 세 가지 유형을 기준으로 분류할 수 있다.

- 1) $Rm1 = (R, Is)$ 의 경우 데이터
- 2) $Rm2 = (R, Ip)$ 의 경우 데이터
- 3) $Rm3 = (R, Ie)$ 의 경우 데이터

이들 3 가지 데이터 군으로부터 질의어 관계를 식별할 수 있는 3가지의 상위 범위 항목 부류로 정한다. 검색 주체의 의도 흐름과 질의어 관계를 회상하는 흐름은 서로 역방향으로 전개된다.

IV. 쿼리 단어의 관계 가시화 모델링

쿼리 단어로 ‘종로 비빔밥’이나 ‘Denny’s Burger’가 입력될 경우에 현행 구글 검색 결과로는 수 백 만 건에서부터 수 천 만 건의 검색 결과가 나온다. 여타 검색 엔진의 검색 결과도 대동소이하다. 하지만 이런 결과

표 1. 의도구조를 활용하여 의미를 확장한 연결 관계
Table 1. Results on Modelling Visualization of Relation.

*‘종로 비빔밥’의 맥락과 유사하며, 한글/영문 어순의 역순을 적용하되, 영문 인스턴스의 구성예제는 bonanza’s burger에 대해 구문 의미 모델을 구성하고 실험을 실시함

건수	Q1	Q2	의미 관계 확장 예시
13871	종로	비빔밥	종로 비빔밥 맛집
*nnnn	Denny	Burger	Lunch Denny’s burger
*3456	Bonanza	Burger	Lunch Bonanza’s burger
6003	Chicago	seminar	IT Tech Chicago Seminar

모두가 검색자에게 활용되거나 의미가 있는 것은 아니다. 오히려 검색자의 주의를 산만하게 하여 데이터를 확인하는 과정에서 검색 경로를 이탈하는 현상이 발생한다.

따라서 의미 관계를 확장하되 방향성을 의도 구성 요소 방향으로 제시할 필요가 있다. 현행 정보 검색의 상위 구조인 의도의 구조와 속성 요소들에 의한 데이터를 검색자에게 제시함으로써 검색자의 검색 목표를 상기하게 하고 검색 도중에 경로이탈 현상도 예방할 수 있다. <표1>에서는 의미 관계를 의도 방향으로 확장한 결과를 나타내었다. 한글의 ‘종로 비빔밥’은 영어로 보통 미국인의 어순으로 표현하면 ‘Denny’s Burger’정도가 된다.

<표 1>에서 데이터 건수는 ‘종로 비빔밥’이 질의어로 입력될 때 의미적으로 가능한 연관 의미가 발생하는 자연어 문장 집합이다. 예컨대 ‘종로 비빔밥을 맛있게 하는 점심 식당’ 등이다.

‘점심’이나 ‘Lunch’는 의도 구조에 의한 데이터 집합이다. 이 집합은 ‘종로 비빔밥’을 질의한 질의어 데이터 집합과 다르다. ‘종로 비빔밥’은 의도 구조인 ‘점심’으로부터 파생된 데이터의 상위 구조의 데이터 집합 군에 속한다. 또한 질의어 데이터 집합 군은 색인어와 웹 데이터를 저장한 데이터와도 다르다. 의도 구조에서 파생된 데이터 집합 군은 질의어 데이터 집합과 색인어(웹 데이터)데이터를 연결할 수 있는 새로운 공유 데이터 집합이다. 새로운 집합인 이유는 자신의 데이터와 타인의 데이터를 연결하는 제3의 데이터이기도 하거나, 과거 데이터와 현재 데이터를 연결하는 미디어 역할 성격을 지니는 데이터라 할 수 있기 때문이다.

1. 인스턴스 데이터 및 질의어 관계의 확장

질의어 관계를 확장하여 의미적으로 나타나는 경우

의 데이터 수는 '종로 비빔밥'은 총 13,871 건을 의미체계 모델로 구성하였다. '남원 추어탕'의 경우는 11,891 건이었으며, 'Denny's burger' 또는 'Bonanza burger' 등의 영어 데이터 경우는 종로비빔밥과 비슷할 것으로 추정하였다. 왜냐하면 한글과 영어의 어순만 다를 뿐이지 맥락은 같기 때문이다. 'Chicago Seminar'에 관련된 의미있는 데이터는 6,003 건을 구성하였다.

'Bonanza burger'는 3,456 건을 구성하였다.

Bonanza's burger의 의미 체계 구성에 관한 경우의 최대치 산출 수식은 <수식1>과 같다.

$$\forall q \in Q, (R, Q_1, Q_2) \in I, q \text{는 } Q \text{의 속성 집합 개수}$$

$$a \quad b \quad \ell \quad c$$

$$\{\sum_{i=1}^{a} (q_1)_i \cdot \sum_{j=1}^{b} (q_1)_j\} \cdot \sum_{m=1}^{\ell} R_m \cdot \{\sum_{k=1}^c (q_2)_k\} \quad (\text{식1})$$

i 는 Q1의 전치 수식어 경우 수, a 는 의미 수식 최대치
j 는 Q1의 후치 수식어 경우 수, b 는 의미 수식 최대치
k 는 Q2의 전치 수식어 경우 수, c 는 의미 수식 최대치
m 는 이항관계의 변화값(ℓ=1,2,3 :의도의 속성별 시작과 중간, 끝)

한편 '종로 비빔밥'을 의미적으로 확장하여 나타난 의도 데이터 분류는 3부류 6건으로 한정할 수 있었다. 맛(맛집, 맛있는 곳), 식당(맛집 식당, 한식집), 점심(한정식)이었다.

2. 상위 관계로 확장했을 경우 의도 선택

두 질의어 관계를 기준으로 의미적으로 확장하면 '종로 비빔밥'의 경우에 총 13,871 건이 의미가 있고, 의도를 선택하는 유형은 3부류 6개의 상위 구조 분류 기준 중에서 선택할 수 있다. 이 경우 상위 기준은 장소 또는 음식점 의도로 속성 식별자를 부착할 수 있을 것이다.

Denny's Burger의 경우는 Lunch가 의미 확장 후의 선택에 효과적으로 나타났다. 선택의 규칙은 사용자 선택에 의하므로 질의어의 선형 변환에 따르는 복잡성과 불확실성을 줄일 수 있다. 이 방법은 사용자에게 추가적인 자료를 요구하지도 않으면서 사용자 부하를 줄일 수 있도록 가상 창으로 인터페이스를 구성하여 해결할

수 있다. 의도 상자와 질의 상자를 구분한다면, 의도를 선택하는 사용자에게 유용할 뿐만 아니라, 검색 중간 과정에서도 검색 목표에 대한 주의 환기 효과가 클 것이다. 또한 의도창을 검색창과 분리할 수 있게 되면, 검색 주체가 사람이 아닐 경우에도 검색이 가능하도록 할 수 있게 되는 이점이 있다.

3. 의도에 적합한 단어의 생성 실험

질의어와 질의어 사이의 관계를 의미적으로 확장하여 발생가능한 경우의 관계 유형 군을 분류하여 실험하였다. 실험의 목적과 목표는 다음과 같다.

검색 질의어 사이의 관계를 의미적으로 확장하여, 관계유형 군을 상위 의미 계층인 의도 낱말 군 중에서 선택한다. 선택한 결과가 검색엔진에서 검색대상 범위의 적합성 기준으로 기능할 수 있는지 여부를 판별하고, 성능 향상시의 효율이 상승하는지를 측정한다. 실험의 목표는 아래 두 가지로 대별된다.

- 1) 질의어 간의 의미적 관계로 성립이 가능한 어휘 개체를 나열하고, 의도 단어 군을 형성
- 2) 의도단어와 질의어를 결합하기 전과 후의 검색결과를 수치로 비교하여 효율을 산출

한편 의도어는 맛집, 음식점, 식당을 가정하였다.

- 1) 맛집을 찾는 질의어(Q1,Q2)를 선정하여 의미적 관계(Rm)를 확장

$$Q_1 \ R_m \ Q_2 \rightarrow R_m \ Q_2 \ A_j$$

- (1) Q₁ 속성 : 지명, 먹거리촌 이름 (전주, 남원, 양평, 마산, 춘천, 함흥, 평양, 수원, 대구, 포항, 제주, 등촌동)
- (2) Q₂ 속성 : 한식, 식사, 메뉴, 맛집 (비빔밥, 추어탕, 해장국, 야구찜, 막국수, 냉면, 아바이순대, 갈비찜, 두루치기, 과메기, 옥돌구이, 샤브샤브)
- (3) R_m 속성 : 장소를 수식하는 후치 형용사 +메뉴를 수식하는 전치 형용사(Q1 에 있는 유명한 돌솥 Q2)

- 2) 지명과 메뉴에 관한 질의어 Q1 Q2와 관계 R_m일 경우 후속 형용사 및 의도적 관계 의미를 확장하여 식당을 찾는 의도로 추정

$$R_m \ Q_2 \ A_j \rightarrow I_n$$

(1) Q₂ 속성 : 한식, 식사, 메뉴 (비빔밥, 추어탕, 해장국, 갈비탕)

(2) Aj 속성 : 한정식 메뉴 뒤에 오게 될 식사 장소를 수식하는 전치 형용사 (Q2 맛있게 잘하는)

(3) In 속성 : 맛집류의 장소에 관한 정보를 탐색하는 의도 (Aj 맛집, 식당, 한정식)

“종로 비빔밥”의 실제 예에서 3 부류 6 가지의 상위 군을 구성할 수 있었다. 이러한 상위 군의 단어를 검색 질의결과를 의미적으로 한정할 수 있는 범위로 설정한다. 이 경우에 유형 군을 선택하는 과정에서 무한 반복의 위험이 있을 수 있다. 상위구조의 의도 낱말 군을 제한하는 방법으로 속성범위를 제한하는 방법과 검색주체가 개입하도록 하는 방법이 있다. 본 논문에서는 무한 반복을 피할 수 있도록 질의자를 개입시켜서 선택하도록 한다. 즉, 종로 비빔밥의 경우는 맛집 또는 점심이나 한정식 등 6개의 상위 구조 중 한 개를 선택할 수 있고, ‘Denny’s Burger’의 경우는 마찬가지로 ‘Lunch’ 또는 ‘Restaurant’를 클릭하여 선택한다.

3) 맛집을 찾는 의도를 추정하는 경우 인스턴스 예시



구글을 통해 검색한 결과(표2)로서, ‘종로 비빔밥’의 경우는 의도어를 반영하여 검색할 경우에 검색결과가 줄거나 늘어났다. ‘Denny’s Burger’의 경우는 85,300,000

표 2. 질의어와 의도어 및 검색결과 (단위: 천건, %) Table 2. Results on queries and intention.

검색 (Q1 Q2)	의도어 추가후 검색 (Q1 Q2 + I) I=맛집,음식점,식당				효율향상(%) =(A-I)/A×100, I=B,C,D		
	결과 (A)	맛집 (B)	음식점(C)	식당 (D)	맛집	음식점	식당
종로 비빔밥	150	95.1	2580	2330	36.6	-1,620.0	-1,453.3
남원 추어탕	344	463	557	197	-34.6	-61.9	42.7
양평 해장국	54.5	59.1	432	364	-8.4	-632.7	-567.9
마산 아구찜	854	588	1460	57.1	31.1	-71.0	93.3
춘천 막국수	129	76.7	1400	1110	40.5	-985.3	-760.5
함흥 냉면	116	57	614	506	50.9	-429.3	-336.2
수원 갈비찜	900	38.8	647	517	95.7	28.1	42.6
대구 두루치기	48.6	23.1	159	134	52.5	-227.2	-175.7
포항 파메기	109	39.8	656	564	63.5	-501.8	-417.4
제주 옥돔구이	58.2	28.9	398	352	50.3	-583.8	-504.8
등촌동 샤브샤브	26.8	13.9	254	211	48.1	-847.8	-687.3

(검색일자: 구글,2011.12.23.오전)

건에서 1,850,000으로 검색 결과가 대폭 줄었다. 맛집의 경우(표2) ‘종로 비빔밥’은 36.6%, ‘Denny’s Burger’의 경우는 97.8%가 각각 향상되었다.

맛집의 경우, 효율이 하락한 경우는 “마산 아구찜”과 “양평 해장국” 경우이다.

1) 마산 아구찜의 경우

마산과 아구찜을 결합한 “마산아구찜”으로 검색할 경우에 효율이 상승하였다. 이는 마산아구찜이 하나의 복합 고유명사로 굳어진 결과로 보인다. 예를 들어 “종로 마산아구찜”으로 검색하면 854,000건이며, “종로 마산아구찜 맛집”으로 검색할 경우 11,000 건으로 효율이 98.7% 상승한다.

2) 양평 해장국의 경우

“양평해장국”을 복합고유명사로 결합한 경우에는 효율이 상승하였다. (54,500 건에서, 23,900 건으로 효율 56.1% 상승)

의도어를 맛집으로 가정하면 위의 <표 2>와 같이 대체적으로 검색결과를 제한하는 결과를 볼 수 있다. 그러나 의도어를 같은 의미인 음식점이나 식당으로 가정할 경우는 일관성이 없는 결과가 나온다. <표 2>에서의 음식점을 의도어로 검색한 12건 모두가 오히려 검색 결과가 늘어났다. 즉 효율이 감소된 것이다. 맛집의 경우는 대체적으로 의도어로서 기능한 반면에 음식점이나 식당은 의도어로서 부적절함을 나타내게 된다. 검색하는 주체의 의도를 반영하지 못하는 경우와 타인의 의도를 검색주체의 의도에 반영한다는 것은 정보탐색엔진의 결합이거나 검색질의창에서 의도를 식별 할 수 없는 것이라 할 수 있겠다. ‘Bonanza’의 경우 ‘Delicious food house’를 입력하게 되면 향상율이 더욱 커진다. bonanza(228,000,000건), Bonanza’s burger(6,380,000건), Lunch bonanza’s burger(1,530,000건), 그리고 Delicious food house bonanza’s burger(291,000건)으로 검색결과가 축소되었다.

한편 사용자 의도에 대한 적중률은 사용자가 클릭하는 방법을 채택하였으므로 논의하지 않았다. 이는 다양한 검색 엔진을 통한 실험에서 유효성을 비교 및 검증하여야 할 과제로 판단한다.

VI. 결론 및 향후 연구

본 연구에서는 두 질의어 사이의 관계로부터 검색의 상위구조인 의도를 유추함으로써, 검색 주체의 의도에 적합한 질의어를 구성하도록 하였으며, 그 결과로 검색의 효율성을 판단하는 과정을 보이게 하였다. 두 질의어의 의미적 관계를 확장하는 ‘종로 비빔밥’의 경우 13,871 건의 인스턴스를 구성했고, 의도 선택 유형은 6 가지로 식별할 수 있었다. 또한 영어인 ‘Denny’s Burger’의 경우에도 같은 맥락으로 추정하여 실험하였다. 이 모델링의 의미는 두 개의 질의어로 의도를 식별하기 위해서는 많은 관계 중에서 의도 한 개를 식별한다는 것을 의미하게 된다.

본 연구에서 제안하는 모델은 검색 주체의 의도를 유추함으로써 기존 정보검색(IR) 엔진과 분리되어, 실용 영역에서는 자료 구조의 범주 변화에 동적으로 대응할 수 있는 기반이자 방향이 된다. 질의어를 형성(Formulation)하는 과정에 있어 의도의 힌트를 추천할 때에도 유용할 것이다. 본 연구 모델링의 질적인 효과로는 정보검색의 상호 작용 척도를 늘리는 의의가 있다. 질의어 수준의 상호 작용 뿐만 아니라 검색 대상이 적합한지를 판단하는 기준과의 상호 작용 차원을 하나 더 늘리는 효과가 있다. 정보 과부하 속에서 사람의 검색 경로 이탈 현상을 크게 개선할 수 있는 가능성을 확인하였다. 이 과정에서 식별자는 질의자의 의도 모형을 선택하는 중요한 역할을 할 것이다.

본 연구 결과는 Jansen 등의 연구 결과에서 정보 의도 통계 결과 범위 내의 정보 요구 중 한 가지를 톱다운(Top-Down) 방식으로 역추적하여 실험했다는 의미가 있다 할 것이다.^[14~15] 향후 연구로는 연관성 탐사규칙 생성에 관한 연구가 진행 중이고, 남은 연구로는 의도 구조를 한 계층 더 추가할 경우에 데이터베이스의 자동 구성과 시간적 효율성을 다른 방법으로 측정하는 연구가 이루어져야 할 것이다.

참 고 문 헌

- [1] Sandra Carberry, “Modelling the User’s Plans and Goals,” *Computational Linguistics*, V.14, No.3, 1988.9
- [2] C. L. Lee and A. Liu, “Modeling the Query Intention with Goals,” Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA’05), 2005.
- [3] Iván Cantador, Alejandro Bellogín and Pablo Castells, “A Multilayer Ontology-based Hybrid Recommendation Model,” *AI Communications*, Vol.21 No.2-3, IOS Press, p203-210, 2008.
- [4] Patrick Schmitz, “Inducing Ontology from Flickr Tags”, WWW 2006, May 22 - 26, Edinburgh, UK, 2006.
- [5] Anand S. Rao and Michael P. Geogeff, “Formal Models and Decision Procedures for Multi-Agent Systems,” *Technical Note* 61, 1995, in www.cs.ucf.edu
- [6] Mohan Ponnada and Nalin Sharda, “Model of a Semantic Web Search Engine for Multimedia Content Retrieval,” In the proceeding of 6th IEEE/ACIS International Conference on Computer and Information Science, Melbourne, 2007.7.11-13
- [7] A. Chulef, S. J. Read and D.A. Walsh, “A Hierarchical taxonomy of human goals,” *Motivation and Emotion*, Vol.25, No.3, p191-232, 2001.
- [8] D.E. Rose and D. Levinson, “Understanding user goals in web search,” Proceedings of the 13 th international conference on World Wide Web, p13-19, 2004.
- [9] 장재형 김태연 김재광 이지형, 개체명을 활용한 스타니펫 생성 방법, Proceedings of KIIS Spring Conference 2010, Vol. 20, No. 1, p178-179, 2010
- [10] T. Spyrou and J. Darzentas, “Intention Modelling: Approximating,” *Information Systems Security*, 1996, p321~336
- [11] Philip R. Cohen and Hector J. Levesque, “Intention Is Choice with Commitment,” *Artificial Intelligence* 42, p213-261, 1990.
- [12] 맥 크로이 저 심영보 옮김, *마음의 지도 I*, 넥서스, p80,1997.
- [13] R. S. Taylor, “Question Negotiation and Information Seeking in Libraries,” *College and Libraries* 28, 178-194, 1968.
- [14] B. J. Jansen, A. Spink, J. Bateman & T. Saracevic, “Real life information retrieval: a study of user queries on the web,” ACM SIGIR Forum, 1998.
- [15] B. J. Jansen, D. L. Booth and A. Spink, “Determining the informational, navigational, and transactional intent of Web queries,” 2007, Available online at www.sciencedirect.com, 2011.3.1
- [16] 백지원 정연경, 용어관계의 분류모형 개발에 관한 연구, *정보관리학회지* 제23권 제1호, 2006.

— 저 자 소 개 —



권 순 진(학생회원)
 2010년 세종대학교 전자정보공학
 대학 디지털콘텐츠학과
 박사 수료
 2007년~2009년 방송통신국가표
 준심의위원회 IT분과 전
 문위원

<주관심분야 : 의도기반 정보검색, 검색모델, 질
 의어관계>



홍 철 의(정회원)
 1985년 한양 대학교 학사
 1989년 New Jersey Institute of
 Technology 석사
 1992년 University of Missouri -
 Rolla
 1992년~1997년 한국전자통신
 연구소

1997년 상명대학교 컴퓨터 공학과 교수 임명
 <주관심분야 : 수평 분산 시스템, 최적화 알고리
 즘, 멀티미디어 어플리케이션, 인공 에이전트>



김 원 일(정회원)-교신저자
 1982년 한양대학교 공과대학 졸업
 1981년~1985년 시스템 디자이너
 와 프로그래머로 대한항
 공 전산실 근무
 1988년 Southern Illinois 대학교
 컴퓨터 공학과 학사 졸업

1990년 Southern Illinois 대학교 컴퓨터 공학과
 석사 졸업
 2000년 Syracuse 대학교 컴퓨터 정보학과 박사
 졸업
 2000년~2001년 Bhasha INC 기술 연구원 근무
 2002년~2003년 아주대학교 근무
 2003년~현재 세종대학교 전자정보공학대학 디지
 털콘텐츠학과 부교수

<주관심분야 : 인공지능, 정보보안, 멀티미디어
 콘텐츠>