

## 시맨틱 검색 시스템의 구현과 평가에 관한 연구\*

한동일\*\* · 권혁인\*\*\* · 최호준\*\*\*\*

### A Study on the Implementation and Evaluation of a Semantic Search System\*

Dong-IL Han\*\* · Hyeong-In Kwon\*\*\* · Ho-Joon Choi\*\*\*\*

#### ■ Abstract ■

In this paper, we present an application called Semantic Search which is built on different supporting technologies and is designed to improve traditional web searching. The Semantic Search is becoming crucial challenges on semantic web. The assessment and the implementation of the research on Semantic Search is not full-fledged whereas its research is highly interested. Also there exists only little research that offers a commercial use Semantic Search System that should be taken into the account in measuring the effectiveness of a Semantic Search System.

This paper proposes an implementation and evaluation for the Semantic Search System. Firstly, we built Semantic Search System which includes a case of development and its procedure. Secondly, We presented the measurement of our Semantic Search System's effectiveness. Finally, the evaluation offers useful implications to the researchers and practitioners to improve the research level to the commercial use.

Keyword : Semantic Search, Semantic Web, Measurement

논문투고일 : 2007년 10월 15일      논문수정완료일 : 2008년 09월 07일      논문게재확정일 : 2008년 09월 18일

\* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장 동력 핵심기술 개발사업(2005-S-083-02, 차세대 웹을 위한 시맨틱 서비스 에이전트 기술 개발)의 일환으로 수행하였음.

\*\* KT 미래기술연구소 책임연구원

\*\*\* 중앙대학교 상경학부 교수

\*\*\*\* KT 미래기술연구소 수석연구원

## 1. 서론

일반적으로 웹 이용자들이 새로운 정보(Information)가 필요한 이유는 어떤 행위를 수행하기 위해 지식이 필요하거나 불완전한(Incomplete) 문제를 해결하기 위해서라고 한다. 이와 같은 경우, 검색 이용자는 새로운 정보를 검색하여 그의 불완전한 지식을 완전한(Complete) 상태로 유지하고자 한다[12, 19]. 그럼에도 불구하고, 현재의 웹 검색 이용자들은 대부분 검색 행위 자체에 게으르고(Lazy), 한 두 단어의 키워드만 입력하는 질의 행위, 극소수 만인 고급 검색 기능을 이용하는 행위 등을 보여 주고 있다[29, 30]. 이는 인지적 구두쇠(Cognitive Miser) 현상으로, 정보 과부하(Information overload) 속에서 원하는 정보를 찾기 위해 부담해야 할 인지적 노력(Cognitive load)을 최소화하고자 하는 검색 이용자들의 행동으로 판단된다.

이러한 인간 행위(Human behavior)를 반영하듯 현재 웹 검색은 검색 이용자에게 키워드와 적중(Hit)됨을 기준으로 대량의 검색결과를 제공하여, 검색 이용자의 검색 요구(requirement)에 좀 더 근접한 결과를 찾기 위해서는 서로 상이한 검색 결과를 어쩔 수 없이 볼 수 밖에 없는 것이 현실이다. 최소의 질의어 입력으로 최대의 검색 결과 속에서 헤매고 있는 현상을 초래하고 있다. 다시 말하면, 현재의 웹 검색 방식은 검색 이용자의 니즈(Needs)를 충족하기 위해 정확한 검색을 할 수 있을 정도로 충분히 지능적이지는 못하다[25].

그러나 시맨틱 웹[13] 환경에서는 찾고자 하는 용어(term)의 복잡한 의미(Meaning)와 용어간의 관계(Relationship)를 온톨로지를 매개로 기계가 이해(Machine-understandable)할 수 있어 인간의 개입을 최소화 한다. 그와 동시에 능동화된 에이전트(Autonomous Agent)의 복잡한 질의 해석을 통해 만족스러운 검색결과를 제공하게 된다. 이와 같이 시맨틱 검색은 정보 검색 이론의 핵심적인 개념인 검색 이용자의 정보 니즈(Information Needs)를 가장 잘 충족시킬 수 있다. 그 이유는, 시맨틱 기술

을 활용한 의미기반 검색 엔진은 검색 이용자가 정확한 의미를 몰라도 사람의 말을 이해하는 것처럼 검색 문장의 의미를 파악해 의미에 맞는 대상을 찾아주는 방식이기 때문이다[23]. 검색에 투입되는 수고와 인지/심리적 부담감을 최소화하면서도 검색 결과는 정확하고, 검색 의도와 의미적 연관된 정보를 검색해주기 때문이다. 현재까지는 시맨틱 웹 기술이 잠재력을 완전하게 실현시키기 위해서는 좀더 많은 연구가 필요하지만, 그 기술에 대한 미래는 매우 밝게 전망되고 있다[28].

따라서 본 연구에서는 기존 웹 검색과 시맨틱 웹 관련 이론적인 연구를 토대로 시맨틱 웹 기술을 적용하고, 기존 검색을 개선시킬 수 있도록 시맨틱 검색 시스템을 설계하고, 베타 서비스 형태로 개발한 사례를 제시하고자 한다. 아울러 개발한 시맨틱 검색 시스템의 효과성 측정결과도 제시하고자 한다. 제 2장에서는 관련 연구를 제시하고, 이를 토대로 제 3장에서는 시맨틱 검색 시스템의 설계 및 구현 사례를 제시하고자 한다. 제 4장에서 개발된 시맨틱 검색 시스템의 효과성을 측정하고, 제 5장에서는 평가 결과에 대한 토의를 하고자 한다. 마지막으로 제 6장에서는 결론 순으로 기술하였다.

## 2. 관련 연구

다양한 시맨틱 웹 응용 중에 하나인 시맨틱 검색의 기술에 대해 알아보려 한다. 또한 아직 연구 초기 단계인 시맨틱 검색 시스템의 효과성 측정에 대한 관련 연구도 살펴보고자 한다.

### 2.1 시맨틱 검색 기술 연구

현재 인터넷의 보편화와 방대한 웹 데이터로 인해 더 이상 기존 웹 검색 방식에 의존하지 않고 정보의 의미를 검색할 수 있는 시맨틱 검색에 대한 필요성이 증대되고 있다. 기존 웹 검색과 시맨틱 검색의 차이점은 첫째, 기존 웹 검색이 텍스트

(Texts)키워드매칭 방식이라면, 시맨틱 검색에서는 객체(Objects) 검색 방식이며 둘째, 검색 결과도 기존 웹 검색에서는 단순 URL이라면 시맨틱 검색에서는 객체의 개념(Concepts), 속성(Properties), 값(Instances)을 포함한다[8]. 이러한 시맨틱 검색은 기존 웹 검색보다 해당 키워드가 의미하는 하나 또는 그 이상의 개념(Concept)을 찾고자 할 때 유용하다[14, 15]. 또 다른 차별성으로는 두 사람이 같은 사물을 설명하는데 같은 용어(Term)를 이용하는 확률이 20% 이하라는 지적[20]에서도 알 수 있듯이, 검색 이용자의 질의어를 직접 데이터 집합에 매칭(Matching)하는 기존의 방법과는 다르게 시맨틱 검색은 의미기반의 검색결과를 제공할 수 있다.

이러한 시맨틱 검색은 복잡한 검색 질의를 토대로 만족스러운 검색 결과를 제공하기 위해 웹 콘텐츠에 대한 의미 이해뿐만 아니라 논리적인 추론도 할 수 있다. 따라서 아래와 같은 다양한 시맨틱 웹 기술이 활용된다.

첫 번째로, 다양한 소스(Source)의 콘텐츠(텍스트, 이미지, 멀티미디어)로부터 시맨틱 메타데이터를 생성 및 저장하는 영역인 지식획득(Knowledge Acquisition) 분야는 크게 2가지의 기술적 이슈가 있다. 우선, 다양한 콘텐츠로부터 시맨틱 검색 대상이 되는 시맨틱 메타데이터 생성이 매우 어렵다. 시맨틱 메타데이터 생성에서 생성되는 양과 질에는 상충관계(Trade-off)[5]가 있기는 하지만, 대규모 메타데이터 생성과 시맨틱 어노테이션(annotation)이 가능[6]하다. 둘째, 다수의 생성된 메타데이터를 저장하고 관리할 수 있는 대용량 인덱스 기법이다. 시맨틱 웹의 본질적인 특성상 콘텐츠를 RDF와 RDF Schema(RDF/RDFS), OWL로 표현해야 하고, 이러한 대용량의 콘텐츠를 효율적으로 구조화하여 저장, 검색할 수 있는 대용량 인덱스 처리의 필요성이 높아지고 있기 때문이다[7].

두 번째로, 온톨로지의 스키마와 인스턴스를 구축하고, 이러한 온톨로지 기반 질의 확장 등을 통해 시맨틱 검색을 처리하는 지식의 표현(Kn-

owledge Representation) 영역은 크게 세 가지를 고려해야 한다[5]. 우선, 실제 응용 가능한 온톨로지는 반정형화(Semi-formal)되게 표현되어야 한다. 왜냐하면 다양한 소스로부터 지식을 추출하고 통합하며, 다수의 작업자에 의해 온톨로지가 구축되므로 일관성의 부재, 제약조건 위반 등과 같은 반정형화된 표현이 불가피하다[21]. 둘째, 구축될 온톨로지의 표현이 지나치게 구체화되고 상세화 된다면(Heavy Ontology), 온톨로지에 표현된 지식을 포착(capture)하기가 매우 어렵기 때문에 실제 애플리케이션 적용될 때 별다른 가치를 제공하지 못한다. 또한 추론의 성능 저하도 우려된다. 셋째, 시맨틱 웹 응용 분야가 지식의 발견 및 지능적/분석적 애플리케이션에 적용할 때에는 도메인과 태스크 특화된 심도 깊은 시맨틱 메타데이터와 객체/개념(Entity/concept) 보다는 관계(Relation)의 로직 처리가 요구된다.

그 외에도 멀티미디어 콘텐츠를 시맨틱 검색에 적용하기 위한 위에서는 추가적인 고려가 요구된다[27]. 멀티미디어에서 고품질의 검색결과를 제공하기 위해서는 기술 기반(Description-based)과 콘텐츠 기반(content-based) 표현방식을 모두 포함하여 멀티미디어를 표현해야 한다.

세 번째로, 검색 이용자와 검색 대상이 되는 정보와의 상화작용 영역(Human Information Interaction)인 지식의 이용(Knowledge Utilization) 분야에서는 정보 검색 이용자 요구(Requirement)를 만족시키는데 초점이 있다.

현재까지 시맨틱 검색의 상용화 사례는 미흡하며, 시맨틱 웹 검색 인터페이스에 대한 연구는 연구자의 실험 결과를 확인해 보는 실험실 수준에 머물러 있다. 따라서 개념(concept)과 관계(relation) 중심으로 입력 폼을 구성하는 등 기존 방식과 너무 상이한 검색 입력 창으로 구성하거나, 검색 방식이 너무 복잡하여 익숙하지 못한 이용자로부터의 거부감을 초래할 수 있다는 문제점이 있다.

## 2.2 기존 시맨틱 검색 연구의 제약점

기존 시맨틱 검색 문헌 연구를 토대로 시맨틱 검색을 설계 및 구현하기 위한 제약점을 요약하면 아래와 같다.

첫째, 지식획득 계층에서의 고품질 시맨틱 메타데이터 생성 (반)자동화와 생성된 메타데이터의 대용량 처리구조를 위한 인덱스 개발이 요구된다. 현재까지 시맨틱 메타데이터 생성의 자동화 기법으로 래퍼(Wrapper), 지도 학습(Supervised Learning), 비지도 학습(Unsupervised Learning) 등을 시도하고 있으나 구체적으로 적용된 사례가 부족하며 고품질의 의미 있는 메타데이터 생성은 매우 어려운 실정이다. 또한 대용량 처리를 위한 상용화 수준의 인덱싱 방식에 대한 구현사례가 미흡하다.

둘째, 지식표현 계층 온톨로지의 형태와 범위는 응용 분야에 따라 다르게 구축해야 한다. 시맨틱 검색 분야는 복잡한 응용 분야이므로 온톨로지의 개념뿐만 아니라 관계를 처리해야 할 경우가 있으므로 적절한 크기와 범위(5~8 depth)를 설정하여 온톨로지를 구축(Right Ontology)하여야 한다.

마지막으로 지식이용 계층에서는 세 가지 관점에서 살펴볼 수 있다. 우선 시맨틱 웹 검색 이용자 관점에서 검색의 니즈(Needs)를 반영할 수 있도록 시맨틱 웹 검색 시스템의 상호작용성을 고려해야 한다[1]. 다음으로, 검색 이용자가 그들의 정보 니즈를 질의(Query) 형태로 형식화할 때, 질의 생성 시 정보 니즈를 명쾌하게 명시적으로 표현할 수 있도록 시맨틱 인터페이스를 설계해야 한다. 마지막으로 시맨틱 검색 인터페이스가 너무 복잡하고 검색 과정을 인식해야 하므로 조작이 어렵거나, 학습에 많은 시간이 소요될 것이므로 유용성에 대해서도 신중한 고려가 요구된다.

## 2.3 시맨틱 검색 시스템의 평가(Evaluation)

### 연구

시맨틱 검색은 검색을 위한 하나의 시맨틱 웹

애플리케이션으로 검색 패러다임의 변화를 추구하고 있는데도 불구하고, 대부분 관련연구는 기존 웹 검색 시스템 품질의 한 평가기준인 정확률과 재현율과 관련된 이슈[9, 10]에 집중되어 있을 뿐, 시맨틱 검색 시스템의 성공측정에 관한 포괄적인 연구가 부족한 실정이다.

전통적으로 정보시스템 성공의 측정에 관한 연구[34]에서는 성공을 측정하기 위한 6가지 요소로서 시스템 품질, 정보 품질, 시스템 이용, 사용자 만족도, 개인 성과, 그리고 조직 성과를 제시하고 있다. 근래에는 웹 기반 정보시스템에의 응용에 한계점이 있어서 정보시스템의 품질요소로서 정보 품질, 시스템품질 이외에 서비스 품질[36]을 추가하고, ‘시스템 이용’은 조직 내에서 강제화됨으로 인해 어쩔 수 없이 이용하는 경우도 존재하기 때문에 제외하였다. 이후 제시된 정보시스템 성공의 수정모델[35]에서는 ‘시스템 이용’에 ‘이용 의도’를 추가하고, 또 개인성과 및 조직성과는 ‘도입 효과’로 통합하였다. 웹 기반 정보시스템 효과성의 측정을 위한 독립변수 및 종속변수를 제시하는 이 수정 모델은 시맨틱 검색 시스템의 성공측정을 위해서도 유용하게 응용 될 수 있다.

최근 시맨틱 검색 시스템에 대한 연구 결과들을 종합해 보면, 검색 과정 측면에서 시맨틱 검색 시스템 자체의 사용 용이성, 시스템과의 상호작용성, 신속한 반응 시간 등이 중요한 요인으로 강조되고 있다. 또한 시맨틱 검색 시스템의 검색 결과 측면에서는 정보의 정확율 및 재현률이 시스템의 주된 품질 요인으로 간주되고 있다[8-10, 12, 16].

또 다른 측면의 연구에서는 시맨틱 검색 시스템이 좋은 품질을 유지한다면 사용자들에게는 어떠한 효익이 있는지에 관한 연구가 있다. [11, 18, 19]의 연구에서는 정보 니즈의 개념이 모든 검색 시스템의 성공에 핵심적임에도 불구하고 대부분의 검색에서 이러한 개념을 효과적으로 다루지 못하고 있다고 지적하였다. [8]의 연구에서도 사용자의 지식개선을 위해 검색을 수행한다고 하였다. 즉, 검색을 통해 기대할 수 있는 주요 개인적 효익은

불완전한 지식의 개선 혹은 완성을 통한 개인의 정보 니즈 충족도의 향상이라고 할 수 있다. 이와 더불어, 검색 이용자가 이러한 검색 시스템을 이용한 결과로 시스템 효과에 대한 인식을 하게 되는데, 필요로 하는 정보와 실제 검색 결과로 받은 정보간의 차이에서 정보의 만족 또는 불만족을 측정할 수 있다[34].

한편 [1]의 연구에서는 시맨틱 검색 시스템 품질의 개념을 이론적으로 정립하고 검색의 효과성에 영향을 미치는 시맨틱 검색 시스템 품질평가 요인을 포괄적으로 밝혀내는데 초점을 두었다. 이러한 목적을 달성하기 위한 실증 연구 결과, 시스템 품질 요인에서는 상호작용성이 정보 니즈 충족도에, 정보 품질 요인에서는 정보의 재현성과 정보의 최신성이 각각 정보 니즈 충족도와 사용자 만족도에, 그리고 서비스 품질 요인에서는 신뢰성이 사용자 만족도에 유의적인 영향을 보여주었다. 또한 정보 니즈 충족도가 사용자 만족도에 유의적인 영향을 보여주었다. 시스템, 정보, 그리고 서비스의 세 가지 품질요인과 정보시스템 결과요인 간의 관계를 조명하고자 했던 기존 연구(가령, [34, 37])들에서는 주로 이들 세 가지 품질요인을 기본 분석단위로 하여 변수들 간의 관계를 분석한 반면, [1]연구에서는 이들 품질단위를 더 세분화하여 세부 척도들을 분석단위로 한 연구를 수행함으로써 관련변수들 간의 관계를 더 깊이 있게 이해하는데 기여하였다.

최근 검색 이용자의 니즈를 충족시키는 시맨틱 검색 시스템의 설계에 많은 어려움이 존재하고 있다. 아직 시맨틱 검색 시스템들이 베타 버전 단계 [3, 22]에 있으므로, 시맨틱 검색 시스템의 성공 가능성을 높이기 위해서는 사용자의 니즈를 충족시킬 수 있는 시맨틱 검색 서비스 웹 사이트 품질요소에 관해 보다 더 깊이 있는 이해가 필요하다는 점에서 [1]연구의 중요성을 발견할 수 있다. 따라서 본 연구에서는 [1]의 평가 프레임워크를 기반으로 시맨틱 검색 시스템의 효과성을 측정하고자 한다.

### 3. 시맨틱 검색 시스템 설계와 구현

본 장에서는 앞서 제시한 요구사항을 반영하는 시맨틱 검색 시스템을 설계와 구현을 하였다. 설계된 시맨틱 검색 시스템은 시맨틱 웹의 다양한 기술과 기능을 포함한 구조이다.

#### 3.1 아키텍처 설계

제안하는 시맨틱 검색 시스템의 아키텍처는 크게 지식획득 계층, 지식표현 계층, 지식이용 계층으로 구성된다.

우선 지식획득 계층은 시맨틱 Crawler, 시맨틱 메타데이터 생성기, 시맨틱 레파지토리로 구성되어 있다. 시맨틱 Crawler에서 수집된 콘텐츠(텍스트, 이미지)는 시맨틱 메타데이터 생성기를 통해 메타데이터가 생성되고, 시맨틱 레파지토리에 트리플 인덱스 구조로 저장된다. 시맨틱 Crawler는 온톨로지를 토대로 내용기반의 콘텐츠를 수집하는 기능이다. 특히 시맨틱 메타데이터 생성기에서는 텍스트에 대해서는 온톨로지 기반 추출, 규칙 기반 추출, 자연어 처리 기반 추출 등의 기법을 통합하는 시맨틱 텍스트 분석 방식을 활용하여 고품질의 시맨틱 메타데이터를 구체적이고 실질적으로 생성하였다. 반면 이미지/동영상의 경우는 객체인식 이미지/동영상 처리 기술을 활용하여 하위 수준(Low Level)의 메타데이터(예: 색상, 질감 등)를 생성은 물론이고 상위 수준(High Level)의 메타데이터(예: 건물, 바다, 야경 등)도 생성하고, 이미지/동영상이 속한 컨텍스트(예: 제목, 태그 등)까지 활용하여 기술 기반과 콘텐츠 기반의 메타데이터 생성 기법을 통합하여 시맨틱 메타데이터를 생성하였다.

다음은 지식표현 계층으로 지식획득 계층에서 생성된 메타데이터를 대상으로 온톨로지가 추가/갱신되고, 이러한 온톨로지를 기반으로 시맨틱 질의 처리기, 결과정보 처리기가 운영될 수 있다. 시맨틱 질의 처리기는 지식획득 계층의 시맨틱 인덱스

를 대상으로 검색과 관리를 수행한다. 특히 시맨틱 질의 처리기는 검색 이용자가 입력한 질의문을 토대로 온톨로지를 참조하여 온톨로지에서 사용하는 용어(Term)로 정규화(Normalization)하고 온톨로지의 구조에 따른 확장된 질의어를 생성하여 검색을 수행한다. 지식표현 계층의 또 다른 기능인 결과 정보 처리기에서는 시맨틱 질의 처리기를 통해 가져온 검색결과를 검색결과 제공부에 표시하기 전에 온톨로지의 가중치 기반 랭킹 또는 온톨로지 개념/속성에 따른 분류 등의 과정을 통해 시맨틱 랭킹과 클러스터링을 수행한다. 결과적으로 지식표현 계층에서는 서비스에 적절한(Right) 온톨로지를 기반으로 다양한 시맨틱 질의 처리기를 통해 관계(Relation) 중심의 로직 처리를 수행한다.

마지막으로 지식이용 계층은 검색 이용자와 검색 시스템간의 상호작용이 활발하게 진행되는 영역이다. 크게 질의어 입력부와 질의어 생성부, 검색결과 제공부로 구성되어 있다. 우선 질의어 입력부에서는 Multi-Facet 인터페이스로 구성한다. 이러한 이유는,

첫째, 정보 니즈(Information Needs) 파악은 단 하나의 질의(Query)로는 명시적으로 표현될 수 없으므로 다중 인터페이스(Multi-interface)를 통해 해결하고자 한다[11, 18, 19].

둘째, 시맨틱 웹이 다양한 질의들을 그래프 형태(Graph Patterns)로 명세화(Formalize) 하기에는 용이하나 검색 이용자에게는 질의를 명세화하기 어려우므로, 복잡한 질의 패턴을 직관적으로 해결하도록 시맨틱 인터페이스 지원이 요구된다[12, 16, 31, 32].

셋째, 대부분의 경우 검색 이용자는 검색 대상 분야의 전문가도 아니고 찾고자 하는 대상을 정확하게 알 수 없으므로, 검색 이용자의 검색 과정(Searching Process)을 지원해야 한다[12, 16, 33].

한편 질의어 생성부에서는 입력된 질의문에 대해 지식표현 계층의 시맨틱 질의처리가 처리할 수 있도록 입력된 질의어를 트리플 질의어로 생성하는 역할을 수행하도록 하여 질의 생성의 어려움

을 해소할 수 있다.

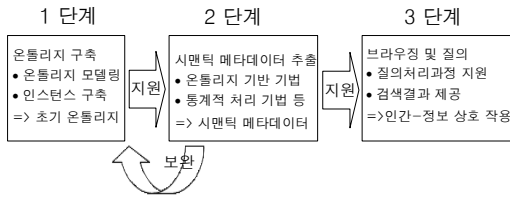
또한 검색 결과 제공부는 검색 결과의 해석이 용이하도록 속성(Property) 또는 관계(Relation)를 중심으로 시맨틱 클러스터링 기법을 적용한 결과와 시맨틱 랭킹 기법을 적용하여 의미기반의 우선 순위에 따른 검색 결과를 제공하도록 구성하였다.

### 3.2 시맨틱 검색 시스템 구현

본 연구를 통해 시맨틱 기술을 활용한 실질적인 시맨틱 검색 시스템이 STARS(Semantic Technology bAsed Retrieval System)이며, 현재 베타 서비스[3]를 진행 중인 상태이다. 전체적으로 3계층으로 구성되어 있으며 각 계층별로 독특한 특징들을 포함하고 있다.

우선 지식획득 계층은 온톨로지, 연관 규칙, 자연어 처리, 객체 인식 기술을 병합하여 고품질의 의미 있는 메타데이터 (반)자동 추출이 가능하고, 트리플 기반 인덱싱으로 시맨틱 메타데이터의 저장과 관리가 용이하다. 지식표현 계층은 기존검색과 추론기능을 혼합한(Hybrid) 검색을 통해 온톨로지 기반 검색과 온톨로지 가중치 적용으로 연관 검색(Associate Search)이 가능하여 구축된 서비스 온톨로지를 참조하여 관계 중심의 질의 확장이 쉽다. 또한 이미지에 대한 객체인식 결과를 포함하는 이미지 온톨로지와 이미지의 주변 텍스트를 기반으로 구성된 서비스 온톨로지를 모두 활용하여 좀 더 정확한 이미지 검색도 가능하다. 마지막으로 지식이용 계층은 검색 이용자의 니즈를 파악할 수 있도록 입력 방식을 기존 검색어 입력 창만 제공하는 방식에서 탈피하여, 검색어 입력창, 질의어 추천창, 그리고 큐브 인터페이스의 동기화를 통해 구성하였고, 검색 결과도 의미 기반의 클러스터링 방식으로 제공한다.

이러한 특징을 가지고 있는 STARS는 시맨틱 검색 서비스를 제공하기 위해서는 [그림 1]과 같은 과정으로 통해 구축된다.

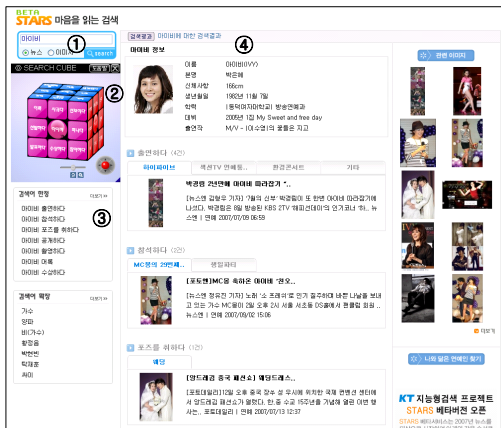


[그림 1] 시맨틱 검색 시스템의 구축과정

1단계인 온톨로지 구축 단계에서는 프로테제라는 개발도구를 이용하여 온톨로지를 디자인, 인스턴스 생성 등을 통해 초기 온톨로지(Bootstrap Ontology)를 구축하였다.

2단계인 시맨틱 메타데이터 생성 과정은 온톨로지 매핑 알고리즘과 온톨로지, 연관규칙, 자연어처리, 객체인식 기반 시맨틱 정보 추출 알고리즘을 이용하였다.

3단계인 브라우징 및 질의 과정에서는 트리플 인터페이스(Triple Interface : ①②③)를 통해 검색 이용자와의 상호작용 과정과 검색 이용자의 정보 니즈를 파악하는 질의 처리과정을 지원하도록 구성하고 검색결과(④)를 제공하였다([그림 2] 참조).



[그림 2] 시맨틱 검색 화면

전체적으로 초기 온톨로지(Bootstrap Ontology)를 구축하고, 다양한 소스의 정보로부터 추가적인 메타데이터를 생성하여 온톨로지를 추가/보완 한 후, 시맨틱 검색 서비스 제공을 수행하는 단계로

요약된다. 단, 본 연구에서는 베타 서비스를 목표로 개발하였으므로 검색의 대상, 온톨로지와 메타데이터 범위를 일부 뉴스 기사로 한정하였다.

## 4. 시맨틱 검색 시스템의 평가

앞서 살펴보았듯이 본 연구에서는 [1]의 평가모델 프레임워크를 활용하여 시맨틱 검색에 필요한 구성요소를 중심으로 위 시맨틱 검색 시스템의 효과성을 측정하였다.

### 4.1 평가 모델

기존 연구문헌을 통해 알 수 있듯이 시맨틱 검색 관련 연구가 아직 부족한 실정이며, 또한 시맨틱 검색 시스템 품질이 검색 이용자의 정보 니즈 충족도와 전체적인 만족도에 어떻게 영향을 미치는지에 대한 이론적 연구는 아직 부족한 실정이다. 이에 본 논문에서는 기존 연구의 한계를 극복할 수 있도록 관련된 선행연구를 토대로 개발된 새로운 평가모형[1]을 토대로 제한한 시맨틱 검색 시스템을 평가하고자 한다. 따라서 본 연구에서의 평가모형은 [35]의 정보시스템 성공모형을 시맨틱 검색 시스템 환경에 맞게 수정하고, 웹사이트 기반 고객 지원 시스템 평가모델[37]을 결합하였다. 또한 시맨틱 검색 시스템 관련 성과 연구가 미흡한 실정이고 시맨틱 검색 시스템이 웹 사이트를 통해 서비스를 제공하기 때문에 [36-38]의 연구에서와 같이 시맨틱 검색 시스템의 품질을 세 가지 영역, 즉 시스템 품질 영역, 정보 품질 영역, 서비스 품질 영역으로 나누어 고려하였으며, 이러한 시맨틱 검색 시스템 품질요소가 검색 이용자의 개인적 효익과 사용자 만족에 미치는 영향을 실증적으로 알아볼 수 있도록 평가모형을 구성하였다. 다만 기존 평가모형[1]에서 누락된 정보품질 요소인 정보의 정확성을 추가하여 시맨틱 검색 시스템을 평가하고자 한다. 왜냐하면, 정보의 정확성(precision)은 대부분의 검색에서 중요한 속성이다. 웹 기반 정

보의 정확성은 소수의 질 좋은 검색 결과를 제공 하되 부적합한 정보를 제공하지 않음을 의미하기 때문이다[17]. 즉, 정보의 정확성은 시맨틱 검색 시스템이 사용자 니즈와 관련된 문서만을 얼마나 잘 찾아내느냐를 의미하는 질적인 개념이며, 이러한 정보 결과는 사용자의 정보니즈 충족도에 긍정적 영향을 미칠 것이기 때문이다.

## 4.2 분석 방법

본 연구를 위해 수집된 자료는 SPSS 12와 AMOS 4.01 통계프로그램을 이용하여 다음과 같이 분석하고자 한다.

첫째, 연구대상자들의 인구사회학적 특성 및 시맨틱 검색 이용행태를 살펴보기 위해 빈도분석을 실시하였다.

둘째, 본 연구에서 사용된 시스템 품질의 상호작용성, 정보품질의 정확성, 재현성, 최신성, 그리고 서비스품질의 신뢰성 구성변수들의 개념타당성을 검증하기 위하여 확인요인분석(Confirmatory Factor Analysis CFA)을 실시하였다.

셋째, 연구대상자들의 인구사회학적 특성과 이용행태에 따라 시맨틱 검색 시스템의 품질과 정보 품질, 서비스 품질, 정보니즈 충족도, 사용자 만족도 등의 차이 검증을 위해 t 검정과 일원변량분석(One-way ANOVA)을 실시하였다.

넷째, 시스템 품질과 정보 품질, 서비스 품질, 정보니즈 충족도, 사용자 만족도간의 인과관계 살펴보기 위해 상관분석과 경로분석을 실시하였다.

위 분석을 위해 설문 조사 방법을 선택하였다. 설문은 관련 연구에서 제시된 개념적 설명들을 토대로 초기항목을 개발하고 내용타당성 테스트와 항목 분류과정과 예비설문과정을 통해 완성하였다. 본 조사의 모집단은 전시회에 참석하여 시범 적용하고 있는 시맨틱 검색 서비스 베타 버전 사이트(stars.paran.com)에 접속하여, 시맨틱 검색의 기능을 충분히 경험해 본 후 설문에 응하도록 하였다. 자료 수집은 2007년 10월 3일 간에 걸쳐

총 140부를 설문지를 배포하여 129부를 회수하였다. 그 결과 시맨틱 검색시스템의 잠재적 또는 실제 이용자들이 대다수를 차지하였다.

## 4.3 실증분석 결과

### 4.3.1 조사대상자 분석

#### 4.3.1.1 인구사회학적 표본의 특성

조사대상자의 인구사회학적인 특성은 우선 성별은 남성이 81.4%, 여성이 18.6%로 남성이 대부분이었으며, 연령층은 20대가 30.2%, 30대가 29.5%, 40대가 15.5%, 50세 이상이 4.7%로 분포하였다.

학력은 대(제)졸이 45.7%, 고(제)졸이 27.1%, 대학원 이상(22.5%), 전문대졸이 4.7%로 분포하였으며, 직업은 학생이 42.6%로 가장 많았고, 다음으로 전문직이 15.5%, 연구직이 14.0%, 사무직이 8.5%, 공무원/교직이 6.2%, 자영업이 5.4% 등으로 분포하였다.

#### 4.3.1.2 시맨틱 검색 이용행태

시맨틱 검색 이용행태가 어떠한지를 살펴보았다. 시맨틱 검색에 대한 인지도는 잘 알고 있는 편이다(17.1%)와 매우 잘 알고 있다(2.3%) 등 비교적 잘 알고 있다는 응답이 19.4%였고, 전혀 알지 못한다(20.9%)와 잘 모르는 편이다(30.2%) 등 잘 모른다는 응답이 51.1%로 잘 모르는 응답자가 많은 것으로 나타났다.

다음으로 시맨틱 검색 시스템 이용경험은 어느 정도 경험이 있다(19.4%)와 매우 많은 경험이 있다(2.3%) 등 경험이 비교적 많다는 응답이 21.7%, 전혀 경험이 없다(51.9%)와 조금 경험이 있다(15.5%) 등 경험이 비교적 적다는 응답이 67.4%로 나타났다.

하루 평균 인터넷 검색시간은 1~2시간 미만인 32.6%로 가장 많았으며, 3시간 이상이 30.2%, 2~3시간 미만인 19.4%, 30분~1시간 미만인 10.9%, 30분 미만이 7.0%로 각각 분포하였으며, 네이버나 구글 등 기존검색 시스템에 대한 만족도는 조금 만족(39.5%), 매우 만족(8.2%)로 만족한다는 응답



이47.7%로 나타났으며, 매우 불만족(2.3%)와 약간 불만족(10.1%) 등 불만족하다는 응답이 12.4%로 만족한다는 응답이 많은 편이었다.

마지막으로 1회 검색시 검색키워드 입력수는 두 단어가 29.5%로 가장 많았으며, 다음으로 네 단어 이상이 28.7%, 세단어가 25.6%, 한 단어가 15.5%로 각각 분포하였다.

4.3.2 변수의 타당성과 신뢰성 분석

본 연구에서 사용된 시스템 품질의 상호작용성, 정보품질의 정확성, 재현성, 최신성, 그리고 서비스 품질의 신뢰성 구성변수들의 개념타당성을 검증하기 위하여 확인요인분석(Confirmatory Factor Analysis ; CFA)을 실시하였다.

각 단계별로 항목 구성의 최적상태를 도출하기 위한 적합도를 평가하기 위하여 기초부합치인  $\chi^2$  통계량( $p > 0.05$  바람직함), GFI(Goodness of Fit Index : 0.9이상 적합), 조정부합치인 AGFI(Adjusted Goodness of Fit Index : 0.9이상 적합), 원소한 평균차이인RMR(Root Mean Square Residual : 0.05이하 적합), 표준부합지수인 NFI(Normed Fit Index : 0.9이상 적합), 비교부합치인 CFI (Co-

mparative Fit Index : 0.9이상 적합) 등을 이용하였다.

한편 표본의  $\chi^2$ 값은 검증력이 표본의 크기에 민감한 지수이기 때문에 표본의 크기가 클 경우  $\chi^2$  값을 기준도 고려해야하지만 다른 적합도 지수를 더욱 우선적으로 고려해서 모형의 적합도를 평가해야 한다.

시스템과 정보, 그리고 서비스 품질의 확인적 요인분석의 모형의 적합도와 분석결과는 다음 <표 1>과 <표 2>와 같다. 먼저 모형 적합도는  $\chi^2 = 62.933(p = .015)$ , GFI = .926 AGFI = .869, RMR = .047, NFI = .914, CFI = .967로 나타나 모형은 비교적 적합한 것으로 나타났다.

확인적 요인분석 결과 시스템 품질의 상호작용성, 정보품질의 정확성, 재현성, 최신성, 그리고 서비스품질의 신뢰성 요인의 구성변수 모두 유의한 것으로 나타났다.

<표 1> 시스템 품질 요인의 확인적 요인분석 모형 적합도

$\chi^2(p)$	GFI	AGFI	RMR	NFI	CFI
62.933 (p = .015)	.926	.869	.047	.914	.967

<표 2> 시스템, 정보, 서비스 품질의 요인의 확인적 요인분석 결과

구 분	경 로	경로 계수	S.E	t	
시스템 품질	B1 <- 질의어 추천	1.000			
	B2 <- 추천질의어 수정	1.090	0.174	6.259***	
	B3 <- 질의어 선택	1.086	0.191	5.672***	
정보 품질	B4 <- 상위랭킹	1.000			
	B5 <- 부적합결과 배제	0.847	0.131	6.464***	
	B10 <- 검색결과 정확	1.253	0.156	8.051***	
	정보의 재현성	B6 <- 대상의 누락없음	1.000		
		B7 <- 폭넓은 검색 결과	0.795	0.123	6.484***
	정보의 최신성	B8 <- 검색결과가 최신	1.000		
B9 <- 최근정보우선 제공		0.916	0.120	7.608***	
서비스 품질	신뢰성	B11 <- 성실한 수행	1.000		
		B12 <- 기대에 부응	1.058	10.03	10.034***

주) \*\*\* p < .001.

확인적 요인분석을 통해 검증된 요인별 구성항목의 내적일관성을 나타내는 신뢰도 검증결과는 <표 3>에서 보이듯, 모든 요인에서 Cronbach's  $\alpha$  계수가 0.7이상으로 나타나 요인별 구성변수들은 내적일관성이 높음을 알 수 있다.

<표 3> 신뢰도 검증 결과

구 분		항목수	Cronbach's $\alpha$
시스템 품질	상호작용성	3	.73
	정보의정확성	3	.74
정보 품질	정보의재현성	2	.74
	정보의최신성	2	.75
서비스 품질	신뢰성	2	.79

### 4.3.3 이용행태에 따른 차이 분석

#### 4.3.3.1 시맨틱 검색 시스템 품질

시맨틱 검색 이용행태와 시맨틱 검색 시스템 품질과의 관련성을 살펴보기 위해 상관분석을 실시한 결과는 <표 4>와 같다. 분석결과 시맨틱 검색에 대한 인지도, 시맨틱 검색 시스템 이용경험, 하루평균 인터넷 검색시간, 1회 검색시 검색키워드 입력수 등과 시맨틱 검색 시스템 품질과는 의미 있는 상관관계는 나타나지 않았으나, 기존 검색 시스템

만족도와 시맨틱 검색 시스템 품질과는 부(-)의 유의미한 상관관계를 보이는 것으로 나타났다.

구체적으로 정보의 신뢰성( $r = -.268, p < .01$ ), 정보의 재현성( $r = -.265, p < .01$ ), 정보의 정확성( $r = -.239, p < .05$ ), 검색시스템의 상호작용성( $r = -.197, p < .05$ ) 등과 유의미한 부(-)적 상관관계를 보이는 것으로 나타나 기존 검색 시스템 만족도가 낮을수록 시맨틱 검색 시스템에 대한 서비스의 신뢰성, 정보의 재현성, 정보의 정확성, 검색시스템의 상호작용성 등에 대해 긍정적으로 인식하고 있음을 알 수 있다.

#### 4.3.3.2 정보니즈 충족도와 만족도

시맨틱 검색 이용행태와 정보니즈 충족도, 사용자 만족도와의 관련성을 살펴보기 위해 상관분석을 실시한 결과는 <표 5>와 같다. 분석결과 시맨틱 검색에 대한 인지도, 시맨틱 검색 시스템 이용경험, 하루 평균 인터넷 검색시간, 1회 검색시 검색키워드 입력수 등과 시맨틱 검색 시스템의 정보니즈 충족도 및 사용자 만족도와 의미 있는 상관관계는 나타나지 않았으나, 기존 검색 시스템 만족도와 시맨틱 검색 시스템의 정보니즈 충족도와 사용자 만족과는 부(-)의 유의미한 상관관계를 보이는 것으로 나타났다.

<표 4> 시맨틱 검색 이용행태와 시맨틱 검색 시스템 품질과의 관련성

구 분	시스템 품질	정보 품질				서비스 품질	
	상호 작용성	정보의 정확성	정보의 재현성	정보의 최신성	전 체	신뢰성	
시맨틱 검색에 대한 인지도	.139	-.178*	-.043	-.132	-.139	-.066	
시맨틱 검색시스템 이용 경험	.048	-.183*	-.013	-.172	-.146	-.173	
하루 평균 인터넷 검색시간	.018	.024	.065	.039	.051	.055	
기존 검색 시스템 만족도	-.197*	-.239**	-.265**	-.140	-.251**	-.268**	
1회 검색시 검색어 입력수	-.143	.091	.050	-.064	.027	-.027	

주) \*  $p < .05$ , \*\*  $p < .01$ .

〈표 5〉 시맨틱 검색 이용행태와 시맨틱 검색 시스템의 정보니즈 충족도와 사용자 만족과의 상관관계

구 분	정보니즈 충족도	전반적 이용 만족도
시맨틱 검색에 대한 인지도	.044	.009
시맨틱 검색 시스템 이용경험	-.059	-.006
하루 평균 인터넷 검색시간	.091	.101
기존 검색 시스템 만족도	-.158*	-.315**
1회 검색시 검색어 입력수	-.014	-.015

주) \* p < .05, \*\* p < .01.

구체적으로 시맨틱 검색 시스템 전반적인 이용 만족도( $r = -.315, p < .01$ ), 시스템의 정보니즈 충족도( $r = -.158, p < .05$ ) 등과 유의미한 부(-)적 상관관계를 보이는 것으로 나타나 기존 검색 시스템 만족도가 낮을수록 시맨틱 검색 시스템에 대한 전반적인 이용 만족도와 정보니즈 충족도 등에 대해 긍정적인 반응을 보이고 있음을 알 수 있다.

#### 4.3.4 인구사회학적 특성에 따른 차이 분석

##### 4.3.4.1 시스템 품질, 정보 품질, 서비스 품질 인식 차이

인구사회학적 특성에 따른 시맨틱 검색 시스템 품질과 정보 품질, 서비스 품질에 대한 인식 차이가 있는지를 살펴보기 위해 t 검정과 일원변량분석(ANOVA)을 실시한 결과는 다음과 같다. 먼저 전체 응답자의 시맨틱 검색 시스템 품질에 대한 인식 수준이 어떠한 지를 살펴본 결과는 다음 <표 6>에서 보는 바와 같이, 시스템 품질의 상호작용성(3.67)은 전반적으로 긍정적인 반응을 보였다. 다음으로 정보 품질은 3.55로 비교적 높게 나타났으며, 요인별로는 정보의 재현성(3.61)이 가장 높고, 정보의 최신성(3.55), 정보의 정확성(3.42) 순으로 긍정적인 반응을 보였다.

마지막으로 정보의 신뢰성은 3.49로 역시 비교적 긍정적인 반응을 보였다.

〈표 6〉 시맨틱 검색 시스템 품질에 대한 인식

구 분		N	M	SD
시스템 품질	상호 작용성	129	3.67	.64
	정보의 정확성	129	3.42	.77
정보 품질	정보의 재현성	129	3.61	.79
	정보의 최신성	129	3.55	.85
	전 체	129	3.53	.68
	서비스 품질	신뢰성	129	3.49

상세한 분석에서는 성별과 학력에 따른 인식차이는 없었으나, 연구대상자의 연령층에 따라 시맨틱 검색 시스템 품질에 대한 인식 차이가 있는지를 살펴본 결과는 시스템 품질의 상호작용성과 정보 품질의 정보의 정확성 요인에 대한 인식에서 통계적으로 유의미한 차이를 보이는 것으로 나타났다. 먼저 검색 시스템의 상호작용성의 경우 30대 연령층(3.92)이 10대(3.56)와 20대(3.56), 40세 이상(3.57)의 연령층에 비해 더욱 긍정적인 반응을 보였다( $p < .05$ ).

다음으로 정보의 정확성의 경우 10대(3.71)가 가장 긍정적인 반응을 보였고, 20대(3.15)에도 대체적으로 긍정적 반응이었다( $p < .05$ ).

##### 4.3.4.2 정보니즈 충족도와 사용자 만족도 인식 차이

인구사회학적 특성에 따라 시맨틱 검색 시스템의 순효익과 사용자 만족에 대한 인식 차이가 있는지를 살펴보기 위해 t 검정과 일원변량분석(ANOVA)을 실시한 결과는 다음과 같다. 먼저 전체 응답자의 시맨틱 검색 시스템의 순효익과 사용자 만족 수준이 어떠한 지를 살펴본 결과는 다음 <표 7>에서 보는 바와 같이 정보니즈 충족도는 3.78점으

로 비교적 높았고, 사용자 만족도 역시 3.71점으로 높게 나타났다.

〈표 7〉 시맨틱 검색 시스템의 정보니즈 충족도와 사용자 만족 수준

구 분	N	M	SD
정보니즈 충족도	129	3.78	.80
사용자 만족도	129	3.71	.72

상세한 분석에서는 성별과 연령에 따른 인식차이는 없었으나, 연구대상자의 학력에 따라 시맨틱 검색 시스템의 정보니즈 충족도와 사용자 만족도 차이가 있는지를 살펴본 결과는, 시맨틱 검색 시스템을 통한 정보니즈 충족도는 학력층에 통계적으로 유의미한 차이를 보이는 것으로 나타났다. 구체적으로 대학원 이상의 학력 집단(4.14)이 대(재)졸(3.71)과 고(재)졸 집단(3.63)에 비해 높은 정보니즈 충족도를 보이는 것으로 나타났으며, 전반적으로 학력이 높을수록 시맨틱 검색 시스템에 대한 정보니즈 충족도가 높은 것을 알 수 있다.

4.3.5 시맨틱 검색 시스템의 시스템의 품질, 정보 품질, 서비스 품질, 정보니즈 충족도, 사용자 만족과의 관계

시맨틱 검색 시스템의 시스템 품질과 정보 품질, 서비스 품질, 정보니즈 충족도, 사용자 만족도 등

인과관계를 살펴보기에 앞서 연구 개념들 간에 어떠한 관련성이 있는 상관분석 결과는 <표 8>과 같다. <표 8>에서 알 수 있듯이 시맨틱 검색 시스템의 시스템 품질과 정보 품질, 그리고 서비스 품질 요인들과 정보니즈 충족도, 사용자 만족도간에 모두 유의한 정(+적 상관관계를 보이고 있는 것으로 나타났으며, 정보니즈 충족도와 사용자 만족도간에도 유의한 정(+적 상관관계를 보이고 있는 것으로 나타났다.

다음으로 시스템 품질과 정보 품질, 서비스 품질 요인들과 정보니즈 충족도, 사용자만족도 등 연구 개념들 간의 인과관계를 검증하기 위하여 경로분석을 실시하였으며, 그 결과는 <표 9>와 같다.

먼저 시스템 품질 요인인 상호작용성이 정보니즈 충족도와 사용자 만족도에 미치는 영향에 대한 분석결과 상호작용성은 정보니즈 충족도에 유의한 정(+영향을 미치는 것으로 나타나(경로계수 = 0.250, t = 3.186(p < .01)), 시맨틱 검색 시스템의 상호작용성을 긍정적으로 인식할수록 정보니즈 충족도가 높아짐을 알 수 있으나, 사용자 만족도에는 유의미한 영향을 미치지 않는 것으로 나타났다.

다음으로 정보 품질 요인이 정보니즈 충족도와 사용자 만족도에 미치는 영향을 살펴보면, 먼저 정보니즈 충족도의 경우 정보의 정확성(경로계수 = 0.263, t = 2.390(p < .05))과 재현성(경로계수 = 0.198, t = 2.060(p < .05))은 정보니즈 충족도에 유의한 정

〈표 8〉 연구개념들간의 상관관계

구 분	시스템 품질	정보 품질				서비스 품질	정보니즈 충족도	사용자 만족도
	상호 작용성	정확성	재현성	최신성	전체	신뢰성		
상호 작용성	1							
정확성	.388*							
재현성	.416**	.579**	1					
최신성	.360**	.564**	.572**	1				
전체	.458**	.838**	.845**	.852**	1			
신뢰성	.396**	.752**	.620**	.602**	.776**	1	.462**	
정보니즈 충족도	.462**	.516**	.499**	.435**	.570**	.462**	1	
사용자 만족도	.406**	.459**	.563**	.451**	.580**	.559**	.499**	1

주) \* p < .05, \*\* p < .01.

〈표 9〉 경로분석 결과

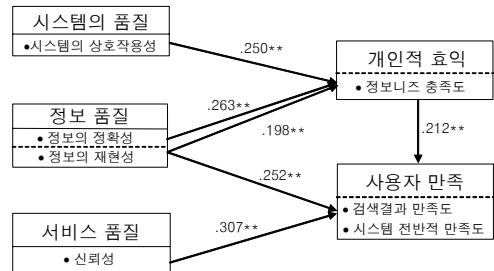
경로	표준화 경로계수	SE	t
상호작용성 → 정보니즈 충족도	.250	.099	3.186**
상호작용성 → 사용자 만족도	.100	.088	1.292
정보의 정확성 → 정보니즈 충족도	.263	.114	2.390*
정보의 정확성 → 사용자 만족도	-.092	.100	-0.862
정보의 재현성 → 정보니즈 충족도	.198	.098	2.060*
정보의 재현성 → 사용자 만족도	.252	.085	2.705**
정보의 최신성 → 정보니즈 충족도	.091	.088	0.982
정보의 최신성 → 사용자 만족도	.045	.076	0.510
신뢰성 → 정보니즈 충족도	-.012	.116	-0.102
신뢰성 → 사용자 만족도	.307	.100	2.789**
정보니즈 충족도 → 사용자 만족도	.212	.076	2.517*

주) \* p < .05, \*\* p < .01.

(+)적 영향을 미치는 것으로 나타났다. 따라서 시맨틱 검색 시스템의 검색결과에 대한 정확성과 재현성이 높을수록 정보니즈 충족도는 높아짐을 알 수 있다. 한편, 사용자 만족도에는 정보 품질 요인 중 정보의 재현성 요인만이 유의한 정(+)적 영향(경로계수 0.252, t = 2.705(p < .01))을 미치는 것으로 나타나, 정보의 재현성이 사용자 만족도에 특히 중요한 요인임을 알 수 있다.

서비스 품질 요인인 신뢰성이 정보니즈 충족도와 사용자 만족도에 미치는 영향을 살펴보면, 먼저 신뢰성은 정보니즈 충족도에 유의한 영향을 미치지 않았으며, 사용자 만족도에는 유의한 정(+)의 영향(경로계수 = 0.302, t = 2.789(p < .01))을 미치는 것으로 나타났다. 따라서 시맨틱 검색 시스템의 신뢰성이 높을수록 사용자 만족도가 높아짐을 알 수 있다.

마지막으로 정보니즈 충족도가 사용자 만족도에 미치는 영향을 살펴본 결과 유의한 정(+)적 영향(경로계수 = 0.212, t = 2.802(p < .05))을 미치는 것으로 나타나, 사용자의 정보니즈 충족도가 높을수록 검색 시스템에 대한 만족도 높은 것을 알 수 있다([그림 3] 참조).



[그림 3] 경로분석 결과

## 5. 토 의

### 5.1 시맨틱 검색 시스템 구현 관점

제안한 시맨틱 검색 시스템 베타 서비스인 STARS는 아래와 같은 특징을 가지고 있다.

첫째, 지식획득 관점에서 시맨틱 메타데이터 생성의 자동화를 위해 온톨로지 기반 추출, 규칙 기반 추출, 자연어 처리 기반 추출, 객체인식 추출 등의 기법을 사용하였고, 생성된 메타데이터를 구조화하기 위해 시맨틱 레퍼지토리를 개발하였다.

둘째, 지식표현 관점에서는 온톨로지의 형태와 범위는 시범 서비스 제공을 목표로 기존 하향식 방식을 탈피하여 상향식으로 구축하여 도메인의 적합성, 서비스의 유용성을 강조하였다.

셋째, 지식이용 관점에서도 시맨틱 웹 인터페이스를 일반 웹 검색과의 차이점을 극복할 수 있도록 기존 검색 입력 창을 포함하는 인터페이스의 유용성과 조작의 용이성을 충분히 반영하였다. 더욱이 시맨틱 검색 관점에서 검색의 니즈를 반영할 수 있도록 시맨틱 검색 시스템 설계시 검색 이용자와 검색 시스템간 상호작용성을 트리플 인터페

이스(검색어 입력창, 질의어 추천창, 큐브 인터페이스)를 통해 최대한 고려하였다.

## 5.2 시맨틱 검색 시스템 평가 관점

위 평가결과를 토대로 제안한 시맨틱 검색 시스템의 효과성을 살펴보면 아래와 같은 특성이 있다. 첫째, 인터넷 검색 이용행태에 따른 제안한 시맨틱 검색 시스템의 품질과 만족도에서는 기존 검색에 대한 만족도가 낮은(만족도가 보통 이하인 64.3%) 이용자가 제안한 시맨틱 검색 시스템의 시스템 품질, 정보 품질, 서비스 품질에 대해 긍정적으로 인식하였다. 또한 이러한 이용자는 정보니즈 충족도, 검색결과 만족도, 전반적 이용 만족도 측면에서도 제안한 시맨틱 검색 시스템을 긍정적으로 인식하였다.

둘째 인구사회학적 특성에 따른 시맨틱 검색 시스템의 품질과 만족도의 인식차이에 대해서는, 제안한 시맨틱 검색 시스템의 시스템 상호작용성에 대해서는 30대 > 10대 > 40대 순으로 긍정적인 반응을 보인 반면, 정보의 정확성에 대해서는 10대 > 30대 > 40대 > 20대 순으로 긍정적인 반응을 보였다. 또한 제안한 시맨틱 검색 시스템은 고학력자에게 정보니즈 충족도를 더욱더 충족시켜줄 수 있었다.

셋째, 시맨틱 목표를 극대화시키기 위해 6가지 관점에서 집중적으로 제안한 시맨틱 검색 시스템을 분석한 결과 상호작용적인 측면을 발견하였다. 우선 제안한 Tripple Interfaces(검색어 입력창, 질의어 추천창, 큐브 인터페이스)는 사용자가 시스템과 상호작용할 수 있도록 검색 질의어의 추천, 수정, 선택을 통해 검색과정을 지원하고 있으며 정보니즈 충족도를 높여주었다. 제안한 시맨틱 검색 시스템은 검색 결과 제공 측면에서도, 사용자의 기대에 부합할 수 있는 수준으로 검색 결과의 재현성과 정확성을 유지하였고 정보니즈 충족도에 긍정적인 영향을 미쳤다. 특히 제공한 시맨틱 검색 시스템의 검색 결과는 정보의 재현성 측면에서 사용자 만족도에도 긍정적으로 기여했다. 또한 제안한 시맨틱 검색 시

스템의 서비스 제공 측면에서는 믿음직하고 성실하게 서비스를 수행할 수 있는 능력을 확보하여 사용자 만족도에 유의한 영향을 미쳤다. 마지막으로 사용자의 정보 니즈 충족도가 높을수록 사용자 만족도도 높아지므로, 사용자 만족도를 높여야 하는 상황에서는 시맨틱 검색 시스템의 품질(시스템 품질, 정보 품질)을 높여 사용자의 정보 니즈 충족도를 제고하는 노력도 병행하였다고 판단된다.

## 6. 결 론

현재까지 시맨틱 검색은 비교적 새로운 개념이고 시맨틱 검색 시스템 관련 연구도 아직 초기단계에 있으므로, 시맨틱 검색 시스템의 설계 및 구현은 물론, 포괄적인 측면에서의 시맨틱 검색 시스템에 대한 성공 측정에 관한 연구는 소수에 지나지 않는다. 따라서 본 논문은 시맨틱 검색 시스템 개념을 이론적으로 정립하고, 이를 토대로 설계와 구현을 하였으며, 검색의 효과성에 영향을 미치는 시맨틱 검색 시스템 품질평가 요인을 토대로 제안한 시맨틱 검색 시스템을 평가하여 객관성을 확보한 시맨틱 검색의 실질적이고도 포괄적인 연구에 기여하였다고 판단된다.

결과적으로 시맨틱 검색 시스템 'STARS'를 평가모형을 통해 평가해본 결과 전반적으로 긍정적인 평가결과를 확보하였다.

향후 제안한 시맨틱 검색 시스템을 평가모형에서 제시한 6개의 평가요인은 물론 추가적인 관점에서도 수정/보완하여 시맨틱 검색 시스템의 완성도 향상에 노력할 계획이다.

## 참 고 문 헌

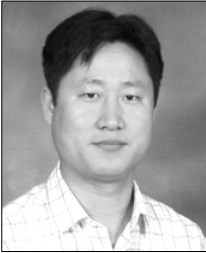
- [1] 한동일, 홍일유, "검색 효과성에 영향을 미치는 시맨틱 검색 시스템 평가요인에 관한 실증적 연구", 『KMIS International Conference』, (2007), pp.52-57.
- [2] Clusty, <http://clusty.com>.

- [3] Stars, <http://stars.paran.com>.
- [4] OntoWeb, <http://www.ontoweb.org>.
- [5] Sheth, A., "From Semantic Search and Integration to Analytics", *Dagstuhl Seminar on Semantic Interoperability and Integration*, 2004.
- [6] Dill et al., "SemTag and SemSeeker : Bootstrapping the Semantic Web via automated semantic annotation", *Proceedings of the 12th International WWW Conference(WWW 2003)*, Budapest, Hungary, 2003.
- [7] Harth, A. and S. Decker, "Optimized Index Structures for Querying RDF from the Web", *Proceedings of the 3rd Latin American Web Congress (LA-WEB)*, (2005), pp.71-80.
- [8] Guha, R., R. McCool, and E. Miller, "Semantic Search", *WWW 2003 Conference*, ACM Press, Budapest, Hungary, 2003.
- [9] Bonino, D. et al., "Ontology Driven Semantic Search", *WSEAS Transaction on Information Science and Application*, Vol.1 (2004).
- [10] Sure, Y. and V. Iosif., "First Results of a Semantic Web Technologies Evaluation", *DOA '2002 Conference*, 2002.
- [11] Wissbrock, F., "Information Need Assessment in Information Retrieval ; Beyond Lists and Queries", *27th German Conference on Artificial Intelligence*, KI2004, University of Ulm, Germany, 2004.
- [12] Albertoni, R. et al., "Semantic Web and Information Visualization", *1st Italian Semantic Web Workshop*, Ancona, Italy, 2004.
- [13] TimBerners-Lee, James Hendler and Ora Lassila, "The Semantic Web", *Scientific American*, 2001.
- [14] Bangyong, L. et al., "Association Search in Semantic Web : Search+Inference", *WWW 2005 Conference*, Chiba, Japan, 2005.
- [15] Richa, C. et al., "A Hybrid Approach for Searching in the Semantic Web", *WWW 2004 Conference*, New York, USA, 2004.
- [16] Makela, E. et al., "Ontogator A Semantic View-Based Search Engine Services for Web Applications", *5th International Semantic Web Conference 2006*, ISWC 2006, Athens, GA, USA, 2006.
- [17] Peter Morville., "Ambient Findability : What We Find Changes Who We become", *O'REILLY*, 2005.
- [18] Oddy, R., "Information retrieval through man-machine dialogue", *Journal of Documentations*, Vol.33(1977), pp.1-14.
- [19] Belkin, N., "Anomalous states of knowledge as a basis for information retrieval", *Canadian Journal of Information Science*, Vol.5(1980), pp.133-143.
- [20] Bate, M., "Subject Access in Online Catalogs : A Design Model", *Journal of the American Society for Information Science*, Vol.11(1986), pp.357-376.
- [21] Gruber, T., "It Is What It Does : The Pragmatics of Ontology, invited talk at Sharing the Knowledge", *International CIDOC CRM Symposium*, Washington, DC., 2003.
- [22] Hakia, <http://www.hakia.com>.
- [23] Jiming, Liu, Ning Zhong, Yiyu Yao, and Zbigniew W. Ras, "The Wisdom Web : New Challenges for Web Intelligence (WI)", *Journal of Intelligent Information Systems*, Vol.20, No.1(2003).
- [24] Han, et al., "Fox Service : An Implementation Case of Ontology-based Search Agent in Mobile Environments", *The 7th International Conference on Mobile Data Management*, 2005.

- [25] Ponnada, M. and N. Sharda, "Model of a Semantic Web Search Engine for Multi-media Content Retrieval", *In the proceeding of 6<sup>th</sup> IEEE/ACIS International Conference on Computer and Information Science*, Melbourne, 2007.
- [26] Parasuraman, A., V. A. Zeithaml and L. L. Berry, "SERVQUAL : A Multiple-Item Scale for Measuring effectiveness", *MIS Quarterly*, Vol.19, No.2(1988), pp.173-187.
- [27] Chu, Heting, "Research in image indexing and retrieval as reflected in the literature", *Journal of the American Society for Information Science and Technology*, Vol.52, No.12(2001), pp.1011-1018.
- [28] Thomas B. Passin, "Explorer's Guide to the semantic web", *Manning publications*, Canada(2004), pp.2-6.
- [29] Spink, A. et al., "Searching the Web : the public and their queries", *J.Am.Soc.Inf.Sci. Technol.*, Vol.52(2001), pp.226-234.
- [30] Belkin, N. J. and W. B. Croft, "Information filtering and information retrieval : two sides of the same coin?", *Commun.ACM*, vol.35(1992), pp.29-38.
- [31] Athanasis, N. et al., "Generation on the fly queries for the semantic web : The CIS-FORTH graphical RQL interface(CRQL)", *Proceedings of the Third International Semantic Web Conference*, (2004), pp.486-501.
- [32] Catarci, T. et al., "An ontology based visual tool for query formulation support", *Proceedings of the 16<sup>th</sup> European Conference on Artificial Intelligence*, IOS Press, (2004), pp.308-312.
- [33] Colucci, S. et al., "A semantic-based fully visual application for matchmaking and query refinement in B2C e-marketplaces", *ICEC '2006*, 2006.
- [34] DeLone, W. H. and E. R. McLean, "Information systems success : the quest for the dependent variable", *Information Systems Research*, Vol.3, No.1(1992), pp.60-92.
- [35] DeLone, W. H. and E. R. McLean, "The Delone and McLean Model of Information Systems Success : A Ten-Year Update", *Journal of Management Information Systems*, Vol.19, No.4(2003), pp.9-30.
- [36] Seddon, P. B. and M. Y. Kiew, "A partial test and development of DeLone and McLean model of IS success", *Australian Journal of Information Systems*, Vol.4, No.1 (1996), pp.90-104.
- [37] Negash, S., T. Ryan, and M. Igbaria, "Quality and effectiveness in Web-based customer support systems", *Information and Management*, Vol.2029(2002), pp.1-12.
- [38] Liu, C. and K. P. Arnett, "Exploring the factors associated with Web site success in the context of electronic commerce", *Information and Management*, Vol.38(2000), pp.23-33.
- [39] D'Ambra, J., R. E. Rice, "Emerging factors in user evaluation of the World Wide Web", *Information and Management*, Vol.38, No.6 (2001), pp.373-384.
- [40] Molla, A. and P. S. Licker, "E-commerce systems success : An attempt to extend and respecify the DeLone and McLean model of IS success", *Journal of Electronic Commerce Research*, Vol.2, No.4(2001), pp. 1-11.
- [41] Flavian, C. and R. Gurrea, "The choice of digital newspapers : influence of reader goals and user experience", *Internet Research*, Vol.16, No.3(2006), pp.231-247.



## ◆ 저 자 소 개 ◆

**한 동 일 (dihan@kt.com)**

중앙대학교에서 경영학 석사를 하고, 동 대학에서 경영정보전공 박사과정 수료 상태이다. 현재 KT 미래기술연구소에서 시맨틱 웹 분야를 연구 중이다. 주요 관심분야는 시맨틱 웹, 웹 2.0, 시맨틱 웹서비스 분야 등이다.

**권 혁 인 (hikwon@cau.ac.kr)**

중앙대학교에서 컴퓨터공학 석사를 하고, 파리 제6대학에서 전자계산학 박사를 받았다. 현재 중앙대학교 상경학부 교수로 재직 중이며 주요 연구관심 분야로 서비스 사이언스, 비즈니스 모델, 게임 콘텐츠 등이 있다. 또한 한국데이터베이스학회, IT서비스학회 등에 논문을 게재하였으며, 디지털콘텐츠 생산 및 유통 기반 구축사업, 게임 산업 인력양성사업 등 다수의 프로젝트 연구책임자로 연구를 수행하였다.

**최 호 준 (chopchop@kt.com)**

한양대학교에서 전산학 석사를 하고, 현재 KT 미래기술연구소에서 시맨틱 웹 분야에서 연구 중이며, 관련 분야의 연구책임자로 프로젝트를 수행 중에 있다. 주요 관심분야는 시맨틱 웹, 시맨틱 웹서비스 분야 등이다.