

온톨로지 기술과 스피어만 상관계수를 적용한 시맨틱 정보 검색 향상

이병욱*

가천대학교 컴퓨터미디어융합학과

Improvement of the Semantic Information Retrieval using Ontology and Spearman Correlation Coefficients

Byungwook Lee*

Dept. of Computer Media Convergence, Gachon University

요약 질의 키워드의 정보 검색은 키워드의 의미가 다양하여 검색된 문서들이 사용자의 요구에 부합되지 않는 문제점을 가지며, 사용자의 상황과 특성이 사용자 마다 달라 정보가 매우 적고 연관성을 찾기 어렵다. 또한, 일반 상관계수의 사용은 정보에 대한 연관성을 나타내지 못하는 문제가 있다. 본 연구에서는 시맨틱 웹 기술을 기반으로 인선에 필요한 다양한 개념들과 지식으로 구성된 인선 온톨로지와 인선 규칙들을 구축하고 규칙들을 지원하는 인선 절차와 인선 결과의 적합성을 확인할 수 있는 지식검색 시스템을 제안한다. 제안된 시스템에서는 스피어만 상관계수를 이용하여 사용자의 상황과 특성에 적합한 정보를 제공하여 제한적인 정보 추천의 단점을 해결하였다. 키워드 기반 검색과 시맨틱 기반 검색 실험 결과 시맨틱 기반 검색이 키워드 검색에 비하여 정확도는 90.3%, 재현율은 71.8%의 성능을 보였다.

주제어 : 시맨틱 웹, 온톨로지, 정보 검색, 키워드, 스피어만 상관계수

Abstract Information retrieval by query keywords have some mismatching problems to fit user's requirement for the retrieved documents due to the varieties of users. These problems are originated from the different situations and characteristics of user's requirement. Also, it has a problem that general correlation coefficients did not display the information relations. In this thesis, it is to suggest knowledge retrieval system to verify feasibility of personnel selection procedure and results supporting selection rules after construction of personnel selection ontologies and rules composed of various concept and knowledge based on the semantic web technology. In the suggested system, it is to clear disadvantages of limited information retrieval providing the suitable information to satisfy user's different situations and characteristics using Spearman's coefficients. Experimental results by this semantic-based information retrieval show 90.3% of accuracy and 71.8% of recall compared with legacy keyword information retrieval.

Key Words : semantic web, ontology, Information retrieval, keyword, Spearman coefficient

* 이 논문은 2013년도 가천대학교 교내연구비 지원에 의한 결과임.(GCU-2013-R189)

Received 4 September 2013, Revised 27 September 2013

Accepted 20 November 2013

Corresponding Author: Byungwook Lee(The University of Gachon)

Email: leebw@gachon.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

유비쿼터스 기술의 발전과 멀티미디어 콘텐츠의 성장과 무선 인터넷의 확장으로 인하여 실시간으로 정보를 교환하고 시간과 공간적으로 사용자의 상황을 인지하여 사용자의 상황에 맞는 서비스를 제공하며 환경적 상황에 따라 필요한 정보를 제공하고 있다[1]. 따라서 사용자의 개별적인 취향과 선호에 따른 콘텐츠를 제공하는 인식 기술과 사용자의 선호 정보를 입력 받아 취향과 선호에 맞는 콘텐츠를 추천하는 역할이 검색 시스템에서 중요한 역할을 차지하고 있다. 검색에 의한 결과를 확인하기 위해서는 사용자로부터 정보를 입력 받아야 하며 정확한 정보의 제공을 위해 사용자로부터 많은 정보를 입력 받게 되어 입력에 할애되는 시간이 상당히 많이 소요된다. 효율적인 콘텐츠 제공을 위한 콘텐츠 검색 기법은 사용자의 개별적인 취향과 선호에 따른 콘텐츠를 추천하기보다 개념적인 정보를 제한적으로 추천하게 되며 이 또한 검색을 위한 시간이 필요하게 된다[2, 9].

전통적인 인선 시스템들은 인사 데이터베이스를 대상으로 인선 조건에 맞는 인물을 키워드 기반으로 검색하는 방식이다. 키워드 기반의 질의 방식으로는 정보 규모가 방대할 때 너무 많은 검색 결과가 제공되는 단점이 있다. 시맨틱 웹은 컴퓨터 프로그램이 웹상의 정보를 이해하고 인간 대신 자동으로 정보를 검색하고 추론할 수 있는 차세대 웹을 의미한다. 시맨틱 웹 기술을 이용하면 컴퓨터가 인사 정보를 이해하여 직원의 경력과 전공과 업무 능력에 관한 다양한 지식들을 활용하여 지식 검색을 수행하고 결과로 얻은 인선 자료를 인사 담당자에게 제공할 수 있다. 시맨틱 웹 기술에는 XML, RDF, 온톨로지, OWL, SWRL 등이 있으며, 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수단이다. 지식검색에서는 각 정보에 대한 온톨로지가 구축되어야 하고 이를 기반으로 추론 검색할 수 있는 논리가 설정되어야 한다. 또한, 피어슨(Pearson) 상관계수 방법은 선형적 관계만을 나타내고, 특정 규칙에 대한 연관성을 표현하기 어렵고, 수량변수만 표현하는 특징 때문에 본 연구에서는 인선을 위한 인선 온톨로지를 설계하고 인사업무에 활용할 수 있는 인선규칙과 보직규칙을 정의하고 인선 시스템을 구축하였으며, 본 논문에서는 스피어만 상관계수를 이용하여 사

용자 상황과 특성에 적합한 정보를 추천하는 시스템을 제안하였다. 적은 정보와 연관성을 스피어만 상관계수를 이용하여 서열화하고 사용자의 상황과 특성에 적합한 정보를 서열화한 목록으로 제공하므로 존재하는 제한적인 개념적 정보 추천의 단점을 해결하였다.

제안된 방법의 수행 과정에서는 상황별 객체 선호도 처리를 수행하고, 스피어만 상관계수를 이용하여 상황별 선호도를 예측하여 처리하고, 스피어만 상관계수를 이용한 선호도 예측으로부터 서열화된 최종 추천 객체 목록을 생성한다. 시맨틱 기반 검색이 키워드 검색에 비하여 정확도는 90.3%, 재현율은 71.8%로 나타났다.

본 논문의 구성으로는 2장에서 선호도 선정 상관계수, 선호도 추천 시스템, 시맨틱 웹 기술에 대해서 살펴보고, 3장에서는 본 논문에서 제안한 인선 온톨로지 및 시맨틱 웹 기술, 상황별 객체 선호도 처리, 스피어만 상관계수를 이용한 선호도 예측에 대하여 설명한다. 그리고 4장에서는 실험 환경과 실험 결과, 5장에서는 결론을 다룬다.

2. 관련 연구

2.1 선호도 선정과 상관계수

상황 정보는 입력된 정보의 상호 작용을 통하여 사용자의 상황을 특성화한다. 상황 인식 서비스 시스템에서 서비스의 가능 여부와 서비스의 종류를 결정하는 방법은 사용자로부터 입력된 정보와 사용자의 상황 정보를 고려하여 선호도 선정 기법을 이용한다. 선호도 선정 기법을 사용하기 위해 데이터를 모델링하는 방법은 상관계수를 이용한다[3]. 상관계수는 연속된 자료 값을 갖는 두 개의 특성 사이의 관계를 객관적인 수치로 표현하기 위한 방법이다. 두 특성의 모집단을 x 와 y 변수로 나타내면 다음 식 (1)과 같다.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (1)$$

\bar{x} 는 x 의 평균값을 나타내고 \bar{y} 는 y 의 평균값을 나타낸다. 식 (1)로부터 x , y 의 모집단 전체를 표현하여 분산 값을 구하기 위한 형태로 표현하면 식 (2)와 같이 나타난다.

$$S_{xx}^2 = \sum_{i=1}^n (x_i - \bar{x})^2, S_{yy}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2)$$

S_{xx}^2 은 x 로부터 평균값의 차이에 대한 제곱 값의 평균을 낸 분산을 나타내고 S_{yy}^2 은 y 로부터 평균값의 차이에 대한 제곱 값의 평균을 계산한 분산을 나타낸다. 식 (2)로부터 특성의 분산 값을 구하는 식은 다음 식 (3)과 같이 나타낸다.

$$S_{xy}^2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3)$$

식 (3)에서 구해진 분산 값을 이용하여 상관계수를 계산하며 식 (4)와 같다.

$$r = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} \quad (4)$$

r 은 상관계수를 나타내고 상관계수 r 은 두 특성의 관련성을 나타내는 척도로 사용되며 연관성이 양으로 나타나면 상관계수의 값도 양의 값을 갖는다. 연관성의 표현은 -1과 1사이의 값으로 표현되며 -1인 경우에는 음의 직선상에 나타나고 1인 경우에는 양의 직선상에 나타나게 된다.

상관계수를 이용한 선호도 측정 방법은 비슷한 선호도를 가지는 사용자들의 객체에 대한 평가를 바탕으로 추천하는 방법이기 때문에 높은 예측 추천 능력을 가지는 장점이 있다.

2.2 선호도 추천 시스템

선호도 추천 서비스 기반의 시스템은 데이터를 이용하는 방법에 따라 분류하며 시스템에 저장되어 있는 기존 데이터와 사용자에게 의해 입력되는 입력 데이터, 기존 데이터와 입력 데이터를 조합하여 형성된다[4].

사용자 선호도 추천 서비스 시스템에는 단일 협업 기법과 내용 기반 기법, 지식 기반 기법, 유용성 기반 기법 등이 있다.

단일 협업 기법은 사용자의 기초 정보와 고객들의 선호도를 선별하여 비슷한 패턴을 보이는 고객들을 하나의 그룹으로 묶어 추천 서비스를 결정하며 기존 데이터와

입력 데이터를 각 항목에 대한 점수로 표현하고 k -최대 근접 방법을 이용하여 점수가 유사한 고객을 찾아 처리하므로 기존 데이터의 의존도가 높으며 사용자 상황을 고려하기가 어려운 단점이 있다. 내용 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징으로부터의 사용도와 입력 데이터의 점수에 따라 클러스터링을 이용하여 분류하므로 클러스터링 방법에 의존해야 한다.

지식 기반 기법은 고객 요구에 충족하는 지식의 기존 데이터의 필요와 흥미에 대한 입력 데이터를 이용하여 욕구를 충족시키는 항목을 추정하므로 확률에 의존하게 되고 기존 데이터가 많이 필요한 단점이 있다. 유용성 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징과 입력되는 데이터로부터 선호 항목의 유용성을 클러스터링 방법을 이용하여 분류한 후 선택하므로 클러스터링 방법에 의존해야 한다[5].

2.3 시맨틱 웹 기술

시맨틱 웹 기술에는 XML, RDF, 온톨로지, OWL, SWRL 등이 있다. XML은 웹 문서 자료를 구조화하기 위하여 만든 표식 언어로 기계가 정보의 내용을 조금 이해할 수 있게 하였고, RDF는 웹상의 자원들을 3진구조로 표현함으로써 표현의 명확성을 향상하였다. 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수단이다. OWL은 클래스와 속성을 이용하여 온톨로지를 정의할 수 있다. SWRL은 온톨로지 규칙을 이용하여 추론할 수 있는 언어이다[8].

OWL에서 클래스는 동일한 특성을 가진 개체들의 그룹을 나타내며, 속성은 클래스 간의 관계를 표현하는 객체 속성과 속성 값이 자료형의 종류임을 표현하는 자료형 속성으로 구분된다. SWRL을 사용하면 인선 의사결정을 도와주는 인선 지식을 규칙 형태로 정의할 수 있다. SWRL은 OWL을 확장시킨 언어로서 OWL의 클래스와 속성을 이용하여 규칙을 정의한다.

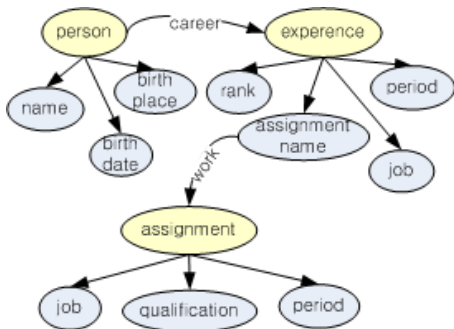
온톨로지 기반의 검색 기법은 온톨로지를 구성하는 개념과 개념간의 관계를 이용하여 검색하는 기법이다. 이 기법에서는 온톨로지의 계층 구조간의 관계를 정의하고 관계를 이용한 의미적인 검색을 지원한다. 온톨로지 기반의 검색 기법에 관한 연구는 TAP과 SEWISE 시스템 등에서 진행되었다. TAP는 미국의 스탠포드 대학에서 SUO(Standard Upper Ontology) 온톨로지와 Cyeupper

온톨로지를 이용하여 콘텐츠에 대한 검색 영역을 확장하기 위해 제안된 시스템이다. TAP에서는 RDF 기반의 RDQL 형태로 표현된 질의문을 처리해야 하는 문제점이 있다. SEWISE는 웹상의 다양한 형태의 자원을 XML 형식으로 구조화하고 도메인 온톨로지를 기반으로 검색하기 위한 시스템이다. 하지만 이 시스템은 웹상의 다양한 자원들을 Wrapper를 이용하여 XML 형식으로 전환해야 하는 문제점이 있다.

3. 시스템 모델

3.1 인선 온톨로지 및 시맨틱 웹 기술

인선을 위한 시스템을 구축하기 위하여 데이터베이스를 추론 검색할 수 있는 온톨로지 논리 체계를 설계한다. 인선 온톨로지는 인선관련 지식들을 온톨로지로 구축한 것으로 인물 지식, 교육 지식, 경력 지식, 보직 지식 등에 관한 개념과 서로의 관계가 정의되어 있다. 인선 규칙은 인사 규정을 반영해야 하는 인선규칙과 보직을 위한 최소한의 자격 기준을 정의하는 규칙이다. 인사규칙은 조직의 각종 보직에 대한 필요조건과 인선에 관한 인선 지식이 규칙의 형태로 정의되어 있다. 인선/보직 지식은 데이터베이스 안에 있는 RDF 형식의 인선 지식과 보직 지식에 대한 검색 기능을 제공한다. 추론 엔진은 인사 규칙과 인선 온톨로지와 지식들을 대상으로 데이터베이스에서 보직 대상자를 선발하는 모듈이다. OWL과 SWRL로 작성된 지식 정보를 이용하여 실시간으로 입력된 인선 정보에 대한 추론을 실행하고, 새로운 정보가 추론될 경우 추론된 결과를 메모리에 저장한다.



[Fig. 1] Ontology for personnel selection

인선 온톨로지는 [Fig. 1]에서와 같이 인선 의사결정에 사용되는 다양한 개념들과 그들의 관계를 정의한 온톨로지이다.

시맨틱 웹 기술로 운용되는 인선 의사결정지원 시스템에서는 인선 온톨로지를 기반으로 표현된 RDF 형태의 직원과 보직 정보가 필요하다. RDF의 객체에서 시간에 따라 변화하는 정보는 타원으로 변화하지 않는 정보는 사각형으로 표현하며, 등록 시스템을 통해 입력된 정보는 정보 등록 모듈에 보내고 다시 미리 정의된 변환 규칙에 따라 RDF 형태로 변환된 뒤에 데이터베이스의 인물(직원) 테이블과 보직 테이블에 저장된다.

3.2 상황별 객체 선호도 처리

다양하게 변화하는 사용자의 상황에 대해 단말기로부터 사용자의 정보와 단말기 위치를 전송받아 주변의 정보와 상황 및 특성을 사용자 데이터로 분류하여 상황에 대한 모든 추천 객체로 이용한다. 사용자의 상황을 분석하고 특성을 인식하여 사용자의 상황에 맞는 추천 가능 범위와 사용자의 특성에 따라 추천 형태를 결정하여 사용자의 위치에서 가까운 범위 내의 사용자의 특성에 적합한 추천 객체를 생성한다.

사후확률은 Y 의 관측값 y 에 대한 θ 의 조건부 확률임을 확인할 수 있으며, y 의 값은 주어진 값이기 때문에 사후확률은 θ 에 대한 함수로 나타나고 식 (5)와 같이 표현된다.

$$f(\theta) = k(\theta|y) \tag{5}$$

사후 확률 $p(\omega_i|x)$ 이 가장 큰 값을 가진 객체를 결정하기 위해 사용하였으며 식 (6)과 같이 표현된다.

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \tag{6}$$

x 는 특징 벡터를 나타내고 ω_i 는 특징 벡터들이 포함된 클래스를 나타낸다. $p(\omega_i|x)$ 는 특징 벡터가 포함된 클래스의 확률을 나타낸다. 식 (6)에 대하여 베이저안 정리를 적용하여 정리하면 식 (7)처럼 나타낼 수 있다.

$$p(w_i|x) = \frac{p(x|w_i)p(w_i)}{\sum_{k=1}^N p(x|w_k)p(w_k)} \quad (7)$$

식 (7)에서 $p(x)$ 특징 벡터 x 의 확률 값을 나타내며 결정 규칙에 영향을 미치지 않아 제거하여 특징 벡터를 포함한 항과 클래스 항으로 정리하면 식 (8)과 같이 표현된다.

$$\Lambda(x) = \prod_{k=1}^n p(x_k|\theta) \quad (8)$$

특징 벡터를 포함한 클래스 w_i 의 확률 값이 특징 벡터를 포함한 클래스의 확률 값이 선택된다. 식 (8)에서 $\Lambda(x)$ 는 우도비(likelihood)의 계산을 위한 수식으로 표현되며 확률 밀도 함수의 예측이 가능하고 선호도 예측 처리를 위해 사용하였다.

스피어만 상관계수를 이용한 선호도 예측은 스피어만 상관계수가 데이터가 순서에 의한 척도인 경우에 사용된다. 측정형 변수나 순서형 분류형 변수들의 상관관계를 자료의 순위 값에 의하여 계산하는 방법으로 순서형 분류형 변수들의 상관관계를 계산한다. 데이터의 값 대신 순위를 이용하는 상관계수이며 데이터를 작은 것부터 차례로 순위를 정하여 서열 순서로 바꾸어 순위를 이용해 상관계수를 구하는 방법이다. 두 클러스터 간의 연관 관계가 있는지 없는지를 밝혀주며 데이터에 이상점이 있거나 표본크기가 작을 때 유용하게 사용된다[6].

사용자간의 유사도를 계산하기 위하여 사용자의 객체에 대한 선호도를 테이블로 형성한 후 식 (8)과 같이 스피어만 상관계수를 이용하여 선호도 점수를 변환한다.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x}) - (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (8)$$

변수 x 의 i 번째 관측치의 순위와 변수 y 의 i 번째 관측치를 계산하여 전체 서열 순위는 ρ 에 의해 순위로 결정된다.

4. 실험 결과 및 분석

시맨틱 기반 검색 시스템에서 사용자는 추천된 검색어 중에서 자신이 원하는 검색어를 선택하며 선택된 검색어는 SPARQL로 변환되어 데이터베이스에 전달된다. SPARQL은 사용자가 찾고자 하는 직원 및 보직 정보에 관한 조건들을 트리플 구조로 질의할 수 있는 언어이다. 데이터베이스에 저장된 정보가 RDF 형식이기 때문에 SPARQL을 이용할 수 있다. [그림 2]는 인선 담당자가 ‘엔진 설계’를 선택한 결과 화면을 보여준다. 인선 담당자는 자신이 선택한 업무가 어떤 업무인지 ‘설명’ 항목을 통하여 확인할 수 있다.



[Fig. 2] Semantic-based retrieval interface form

[Fig. 2]는 인선 담당자가 ‘엔진 설계’를 선택한 결과 화면을 보여준다. 인선 담당자는 자신이 선택한 업무가 어떤 업무인지 ‘설명’ 항목을 통하여 확인할 수 있다.

본 논문에서 제안한 시스템의 성능을 평가하기 위해 효과성을 측정하여 정확도와 재현율로 나타낸다. 정확도 (precision)는 검색된 데이터들 가운데 연관된 데이터의 비율을 계산하여 연관된 데이터만을 선택하였을 경우 높은 결과를 보인다. 실제 선택한 데이터와 추천한 데이터의 교집합을 추천한 데이터로 나누어 표현한다. 재현율 (recall)은 검색된 데이터를 관련된 모든 데이터의 수로 나눈 수치이다. 선택한 데이터와 추천한 데이터의 교집합을 선택한 데이터로 나누어 표현한다[7].

성능 평가의 또 다른 방법은 질의를 기반으로 평가하는 방법이다. 질의 리스트를 평가하기 위한 방법으로 정

확도와 재현율을 사용하고 있으며 정확도는 질의 리스트 중 몇 개를 실제로 선택했는지를 평가하고 재현율은 사용자가 선호하는 질의를 사용하여 얼마나 많은 검색 결과가 추천되었는지를 평가한다. 질의 리스트를 평가하기 위해 정확도와 재현율을 이용하여 기존 연구와 비교 실험을 수행하였다.

<Table 1> Compare experiment result of effectiveness

category	precision(%)	recall(%)
single cooperation	86.5	72.2
content based	87.9	72.7
knowledge based	84.2	65.4
usefulness based	88.5	74.1
suggestion	90.3	71.8

<Table 1>의 실험 결과 기존 추천 서비스 시스템들과 비교하여 정확도는 90.3%, 재현율은 71.8%의 성능을 보였다.

k-means를 이용한 방법, Pearson 상관계수를 이용한 방법과의 사용자가 선호하는 대상의 처리율과 선호대상의 정확성, 그리고 추천 목록 생성의 정확성을 각각 선호도 처리(preference process), 선호도 예측(preference estimate), 추천 목록 생성(suggest list generate)으로 표현하여 비교 실험을 수행하여 <Table 2>에 나타내었으며 평가하기 위한 수식은 식 (9)와 같이 표현한다.

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

<Table 2> Compare experiment result of recommendation service system

category	k-means	Pearson	Spearman
preference process(%)	93.3	95.6	95.2
preference estimate(%)	94.5	96.7	97.8
suggest list generate(%)	96.2	97.1	96.8

<Table 2>와 같이 실험 결과 선호도 처리에서 Spearman 상관계수를 이용한 방법이 k-means를 이용한 방법보다 성능이 높게 나타났으나 Pearson 상관계수를 이용한 방법과는 비슷한 결과를 보였으며 선호도 예측에서는 성능이 가장 높게 나타났다. 추천 목록 생성 비교 실험에서는 k-means를 이용한 방법보다 성능이 높게 나

타났으나 Pearson 상관계수를 이용한 방법과는 비슷한 결과를 보였다. 따라서 Spearman 상관계수를 이용하여 선호도 예측이 가능함을 실험을 통하여 확인하였다.

5. 결론

전통적인 인선 시스템들은 인사 데이터베이스를 대상으로 인선 조건에 맞는 인물을 키워드 기반으로 검색하는 방식이다. 키워드 기반의 질의 방식으로는 정보 규모가 방대할 때 너무 많은 검색 결과가 제공되는 단점이 있다. 시맨틱 웹 기술을 이용하면 컴퓨터가 인사 정보를 이해하여 직원의 경력과 전공과 업무 능력에 관한 다양한 지식들을 활용하여 지식 검색을 수행하고 결과로 얻은 인선 자료를 인사 담당자에게 제공할 수 있다. 본 연구에서는 인선을 위한 인선 온톨로지를 설계하고 인사업무에 활용할 수 있는 인선규칙과 보직규칙을 정의하고 인선 시스템을 구축하였으며, 본 논문에서는 스피어만 상관계수를 이용하여 사용자 상황과 특성에 적합한 정보를 추천하는 시스템을 제안하였다. 적은 정보와 연관성을 스피어만 상관계수를 이용하여 서열화하고 사용자의 상황과 특성에 적합한 정보를 서열화한 목록으로 제공하므로 존재하는 제한적인 개념적 정보 추천의 단점을 해결하였다.

제안된 방법의 수행 과정에서는 상황별 객체 선호도 처리를 수행하고, 스피어만 상관계수를 이용하여 상황별 선호도를 예측하여 처리하고, 스피어만 상관계수를 이용한 선호도 예측으로부터 서열화된 최종 추천 객체 목록을 생성한다. 시맨틱 기반 검색이 키워드 검색에 비하여 정확도는 90.3%, 재현율은 71.8%로 나타났다.

ACKNOWLEDGMENTS

This work was supported by the Gachon University research fund of 2013.”(GCU-2013-R189)

REFERENCES

- [1] Jung Heon Man, Lee Jung Hyun, “Probability-

- annotated Ontology Model for Context Awareness in Ubiquitous Computing Environment”, Journal of the Korea society of computer and information, Vol.11, No. 3, pp239-248, 2006.
- [2] Chan-Shik Ahn, Sang-Yeob Oh. Efficient Continuous Vocabulary Clustering Modeling for Tying Model Recognition Performance Improvement. Journal of the Korea Society of Computer and Information. Vol. 15, No. 1, pp. 177-183, 2010.
- [3] Latifur Khan, Dennis McLeod, Eduard H. Hovy, "Retrieval effectiveness of an ontology-based model for information selection", VLDB J. 13(1): pp.71-85, 2004.
- [4] Harry Chen, Sovrin Tolia, Craig Sayers, Tim Finin, and Anupam Joshi, "Creating Context-Aware Software Agents", Proc. of the first GSFC/JPL Workshop on Radical Agent Concepts, pp.186-200, 2001.
- [5] Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedle, "Application of Dimensionality Reduction in Recommender System-A Case Study", ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.
- [6] Jonathan L. Herlocker, Joseph A. Kosran, Al Borchers, and John Riedl, "An Algorithm Framework for Performing Collaborative Filtering", Proceedings of the 1999 Conference on Research and Development in Information Retrieval, 1999.
- [7] Latifur Khan, Dennis McLeod, Eduard H. Hovy, "Retrieval effectiveness of an ontology-based model for information selection", VLDB J. 13(1): pp.71-85, 2004.
- [8] Byungwook Lee, Semantic Ontology Speech Information Extraction using Non-parametric Correlation Coefficient, The Journal of Digital Policy and Management. Vol. 11, No. 09, pp. 147-151, 2013.
- [9] Chan-Shik Ahn, Sang-Yeob Oh. Echo Noise Robust HMM Learning Model using Average Estimator LMS Algorithm. The Journal of Digital Policy and Management. Vol. 10, No. 10, pp. 277-282, 2012.

이 병 욱(Lee, Byungwook)



- 1973년 2월 : 연세대학교 (공학사)
- 1984년 2월 : George Washington Univ. Dept. of Computer Science (공학석사)
- 1994년 2월 : 중앙대학교 전자계산학과(공학박사)
- 1985년 3월 ~ 현재 : 가천대학교 컴퓨터미디어융합학과 교수
- 관심분야 : 데이터베이스, 분산 시스템, 디지털 융합
- E-Mail : leebw@gachon.ac.kr