

지식검색 서비스에서의 소셜 네트워크 기반 영향력 지수 알고리즘

최창현*, 박건우**, 이상훈***

An Influence Value Algorithm based on Social Network in Knowledge Retrieval Service

Chang Hyun Choi *, Gun Woo Park **, Sang Hoon Lee ***

요약

집단지성을 이용한 지식검색 서비스는 개방적 구조와 축적된 자료를 공유할 수 있다는 커뮤니티적인 특성으로 큰 인기를 얻고 있다. 하지만 방대한 지식공유 속에서 사용자가 진정으로 원하는 답변 획득은 점점 더 어려워지고 있다. 최근 알고리즘에서 가장 정교하다고 평가 받는 구글을 통해 상위에 랭크된 검색 결과들 중에는 집단지성을 통해 구축된 위키피디아, 야후 Q/A와 같은 소셜 검색엔진의 검색 결과들이 상당수 존재한다. 본 논문은 대부분의 질문은 인간으로부터 문제해결의 실마리를 얻을 수 있다는 점과 온라인상의 사용자에게 대한 연구를 통해 지식검색 서비스 사용자 중 영향력 자를 찾는 것에 목적을 둔다. 이에 국내 소셜 검색엔진의 대표인 네이버 지식iN을 중심으로 지식검색내의 사용자 활동성과 신뢰성을 소셜 네트워크 기반으로 정의하고, 사용자간의 관계를 중앙성으로 분석하는 영향력 지수 알고리즘을 제안한다. 제안된 알고리즘을 통한 영향력 지수는 지식검색 서비스에서 문제 해결의 실마리를 가진 사용자를 랭킹화 함으로써 질문에 적합하고 신뢰성 있는 답변을 하는 사용자를 분별하는 지표가 되며 이를 바탕으로 지식검색 서비스내의 영향력 자를 식별 가능하게 된다. 이는 지식검색 서비스 사용자의 최대 목적인 사용자가 필요로 하는 정보와 지식을 보다 용이하게 획득 가능케 함으로써 검색 만족도 향상에 큰 기여를 할 것이다.

Abstract

Knowledge retrieval service that uses collective intelligence which has special quality of open structure and can share the accumulative data is gaining popularity. However, acquiring the right needs for users from massive public knowledge is getting harder. Recently, search results from Google which is known for its exquisite algorithm, shows results for collective intelligence such as Wikipedia, Yahoo Q/A at the highest rank. Objective of this paper is to show that most answers come from human and to find the most influential people in on-line knowledge retrieval service. Hereupon, this paper suggest the influence value calculation algorithm by analyzing user relation as centrality which social network is based on user activeness and reliance in Naver 지식iN. The influence value calculated by the suggested algorithm will be an important index in distinguishing

• 제1저자 : 최창현

• 투고일 : 2009. 09. 16, 심사일 : 2009. 10. 19, 게재확정일 : 2009. 10. 22.

* 국방대학교 전산정보학과 석사과정 ** 국방대학교 전산정보학과 박사과정 *** 국방대학교 전산정보학과 교수

reliable and the right user for the question by ranking users with troubleshooting solutions in the knowledge retrieval service. This will contribute in search satisfaction by acquiring the right information and knowledge for the users which is the most important objective for knowledge retrieval service.

▶ Keyword : 소셜 네트워크(Social Network), 영향력 지수(Influence Value), 중앙성(Centrality)

I. 서론

지식검색 서비스는 2002년 10월 네이버의 지식iN을 시작으로 다음, 야후 등과 같은 포털들의 참여로 국내 검색의 대표적 서비스로 성장하였다. 비즈니스 위크지가 구글이 네이버로부터 배워야 할 점으로 지식iN 서비스를 주목했을 정도로(Moon & Woyke, 2006) 지식iN 서비스는 시작 직후부터 지금까지 국내 이용자들에게 폭발적인 인기를 끌고 있다. 이러한 인기는 누구나 어떠한 주제에 대해서도 질문과 답변을 할 수 있다는 개방적 구조와, 이렇게 축적된 자료를 공유할 수 있다는 커뮤니티적인 특성에 기인한다. 하지만 불특정 다수의 사용자에 의해 구축된 방대한 자료 속에는 검증되지 않은 답변이나 추측성 답변들로 인하여 답변의 신뢰성, 정확성, 전문성이 저하되고 지식의 질적 하락이 초래될 수 있는 문제점도 있다. 이러한 문제점은 사용자가 진정으로 원하는 답변 획득을 점점 더 어렵게 만들고 있다. 이러한 문제를 해결하기 위해 지식 검색 서비스의 결과물로 대변되는 답변 문서의 특성을 평가하는 기준 제시 연구가 수행되었다(1). 또한 연결어, 추정어, 개인의견 단어, 광고성 단어, 가치판단어, 이모티콘 등의 텍스트 요소로 문서의 신뢰도를 평가하는 연구(2)와 문서의 조회 수나 추천 수 등의 비텍스트 정보를 이용하는 연구도 진행되었다(3). 하지만 답변 결과물의 성격이 문서를 포함하여 동영상, 이미지, 음성 등으로 다양해짐에 따라 문제 해결의 어려움을 겪고 있다. 본 논문은 지식 검색 서비스의 이러한 문제점을 집단지성의 특성과 소셜 네트워크 관점에서 풀어보고자 한다. 이는 객관적 수치를 바탕으로 높은 영향력 지수를 갖는 사용자라면 어떠한 형태(문서, 동영상, 이미지, 음성 등)든 양질의 질문/답변을 통해 지식공유의 근본적 목적에 부합될 수 있을 것이라는 가정에서 시작되며, 이를 위해 대표적인 국내 지식검색 서비스인 네이버 지식iN에서의 사용자 활동성, 신뢰성을 바탕으로 카테고리별 영향력 지수 알고리즘을 제안한다. 산정된 영향력 지수는 문제 해결의 실마리를 가진 사용자를 찾음에 있어 훌륭한 지표가 될 것이다. 논문은 2장에서는 관련연구, 3장에서는 소셜 네트워크 기반 영

향력 요소, 속성 추출 및 이를 통한 영향력 지수 알고리즘을 제안한다. 4장에서는 실험을 위한 데이터 셋과 제안한 알고리즘을 통한 영향력 지수 산정 및 평가 결과를 제시하고 마지막으로 결론 및 향후 연구 과제를 제시한다.

II. 관련 연구

1. 집단지성과 지식검색

지식과 정보의 구분이 모호해지면서 등장한 지식의 여러 유형 가운데 사회적 형식지¹⁾는 온라인이란 환경에 힘입어 큰 영향력을 미치고 있다. 이러한 지식은 광의의 개념으로서 일상생활과 관련된 지식, 다양한 지식생산자가 제공하는 지식, 상대적으로 불안정하고 유동적인 지식, 집합적으로 구성되는 지식이다(4). 즉, 온라인에서의 지식은 우리의 일상생활과 관련된 정보, 상식, 조언까지도 포함하는 보다 확장된 개념으로서, 사용자를 포함한 다양한 지식생산자들이 직접 제공하는 상대적으로 불안정하고 유동적인 지식이다. 또한 이를 기반으로 한 지식검색 시스템에서 제공하는 지식은 현재 나의 목적에 어떠한 의미가 있는가에 따라 현재 시점에서 창출되는 지식으로 인터넷의 장속에서 집합적으로 공유되고 끊임없이 구성되는 특성을 지닌다. 이러한 점은 집단지성의 발현과 관련이 깊다. 집단지성은 다수의 사용자가 개개인의 작업 및 지식을 공유하고 취합하여 일반적 사실을 도출해 낸다는 원리를 가지고 있다. 이 원리는 다수 사용자의 참여에 의해 어떤 사실에 대한 해결의 실마리를 얻는다는 것이 그 핵심이다(5).

2. 소셜 네트워크 분석

소셜 네트워크는 최근 온라인을 중심으로 하여 하나 이상의 상호 의존적인 관계에 의해 구성된 개인 또는 집단의 사회적 구

1) 형식지 : 스펠더(Spender)는 지식을 ① 의식하고 있는 지식(개인적 + 형식지), ② 습관화된 지식(개인적 + 암묵지), ③ 집단화된 지식(사회적 + 암묵지), ④ 객관화된 지식(사회적 + 형식지)으로 구분하였다. (Nahapiet & Ghoshal, 1998)

조체(Social Structure)로 정의할 수 있다. 대표적인 소셜 네트워크 서비스(SNS : Social Network Service)인 위키피디아, 지식검색 서비스, 프렌드스터(Friendster), 오르kut (Orkut)과 한국의 싸이월드 등에서 볼 수 있듯이 정보과학 분야에서는 기 구성된 소셜 네트워크의 현상을 웹 환경에 응용하는 연구가 활발히 진행 중이다. 이러한 연구는 첫째, 소셜 네트워크상에 존재하는 웹 사용자 간의 연결성(Connectivity) 확장을 통한 검색 효율의 향상 방법과 둘째, 실제 사회 현상과 소셜 네트워크상의 현상에 대한 비교 분석, 마지막으로 네트워크 구성의 효율성 및 보안 등 소셜 네트워크 자체에 대한 연구로 나눌 수 있다[6], [7], [8], [9], [10]. 사회학, 통신공학, 경제학 등에서 폭넓게 연구 중인 소셜 네트워크 분석은 소셜 네트워크의 형태와 특성을 알고리즘 적으로 연구하는 것으로 전체 관계망에서의 위치와 그 효과를 측정하는 위치적 접근법(Positional approach)과 연결망의 직접적인 관계에 초점을 둔 관계적 접근법(Relational Approach)으로 분류된다 [11]. 위치적 접근법에선 사람들과의 사회적 관계에서 각자가 차지하는 위치 하나하나를 가리켜 사회적 지위라고 부르며, 각각의 사회적 지위에 따라 기대되는 행위를 가리켜 사회적 역할이라고 한다. 관계적 접근법은 연결망 내 구성원들의 상호작용에 의한 전염효과, 즉 직접적인 관계 유무에 초점을 두어 '결속 집단' 이라 한다.

소셜 네트워크의 분석은 노드간의 관계 구조를 찾아내기 위해 그래프 이론을 이용한 소서메트리(Sociometry)와 수학적 방법인 계량적 방법을 이용한다. 수학적 방법의 기본은 행렬과 그래프의 이해이다. 구성된 (i, j) 사이의 관계가 있고 없음을 1과 0으로 나타내는 행렬을 인접 행렬(Adjacency Matrix)이라고 부르며 행렬의 항(Cell)은 i로부터 j에 이르는 관계로 표현된다. 그림 1의 예에서 A, G항이 1인데, 이는 A로부터 G에 이르는 관계가 있다는 뜻으로(예: A가 G를 친구로 선택한 경우) 그래프에서는 A와 G간 연결된 화살표로 표현된다. 인접행렬의 대각선은 자기 선택 유무를 나타내는 것으로 일반적으로 0으로 처리 한다. 한 노드가 다른 노드와 연결될 때 연결된 노드의 수를 연결정도(Degree)라 한다. 한 노드에서 다른 노드로 나가는 수의 합을 외향 연결정도(Out-Degree)라 하고 다른 노드로부터 들어오는 수의 합을 내향 연결정도(In-Degree)라 한다. 전체 연결망에서 특정한 노드의 내향 연결정도가 큰 경우 이 점은 중요한 역할을 하는 경우가 많다

[12]. 소셜 네트워크 분석 기법은 결속, 중앙성, 구조적 등 위성등의 주제로 구별할 수 있다. 이중 중앙성(Centrality)이란 한 노드가 맺은 관계의 정도를 통해 권력과 영향력이라는 개념과 연결되어 가장 많이 쓰이는 지표 가운데 하나이다. 프리만

(Freeman)은 중앙성을 지역 중앙성(Local Centrality)과 전체 중앙성(Global Centrality)으로 구분하였다[11]. 한 노드가 지역적으로 중앙성을 갖는다는 것은 그 노드가 속한 관계망 환경에서 다수의 다른 노드와 연결 관계를 갖는 것, 즉 그 노드가 다수의 이웃 노드를 직접적인 연결로서 가지고 있다는 것을 의미한다. 반면 한 노드의 전체 중앙성은 그 노드가 전체적인 관계망 구조에서 전략적으로 중요한 위치를 가진다는 것을 뜻한다[13]. 중앙성은 관계의 연결 정도를 통해서 계산되는 수치로 내향 연결정도를 이용한 내향 중앙성과 외향 연결정도를 이용한 외향 중앙성의 합에 전체 연결정도의 합을 통해 얻은 중앙성을 나누어 구한다[11].

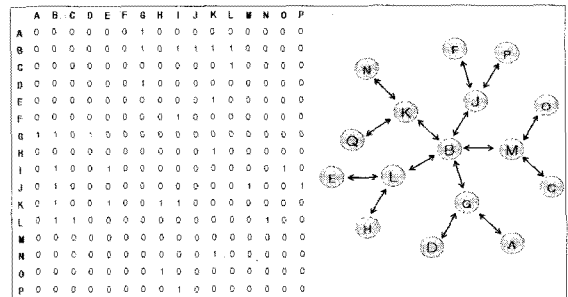


그림 1. 인접행렬과 그래프
Fig. 1. Adjacency Matrix & Graph

$$C_i = \frac{\sum_{j=1}^n (Z_{ij} + Z_{ji})}{\sum_{i=1}^n \sum_{j=1}^n (Z_{ij})}, 0 \leq C_i \leq 1 \tag{1}$$

(C : Centrality
Z_{ijk} : k 연결망에서 i 구성원으로부터 j 구성원까지의 관계)

III. 본 론

1. 영향력 지수(Influence Value)

1.1 영향력 요소와 속성 산정

영향력의 사전적 의미는 '어떠한 사물의 효과나 작용이 다른 것에 미치는 힘'이다. 크리스 와이드너²⁾는 영향력을 주어진 공간에서 구성원의 잠재력을 끌어내고 긍정적인 변화의 기

2) 크리스 와이드너 : 동기부여 및 리더십 전문가, 『메이 드 포 석세스』, 『크리스 와이드너 쇼』 진행자, 미 주간지 E-zine 집필자

운을 만들어 내는 인간의 능력이라 하였다[14]. 다양한 목적으로 지식 검색 서비스를 이용하는 사용자는 집단지성의 발현을 통해 문제 해결의 실마리를 찾고자 한다. 이것은 또한 지식 검색 서비스의 최대 목적이 된다. 온라인 소셜 네트워크에서의 영향력 자는 이러한 목적의 중심에 있는 사용자이다. 국의 대표적 지식공유 검색 사이트인 야후 Q/A의 질문/답변에 대한 특성 분석은 흥미 있는 결과를 제시한다[15]. 총 25개의 카테고리에 대한 질문/답변의 길이, 답변 실마리의 내용, 질문자/답변자 오버랩 등을 통한 야후 Q/A의 소셜 네트워크 구조는 영향력 자를 중심으로 구성되고, 그림 2와 같이 3가지 클러스터로 분류된다.

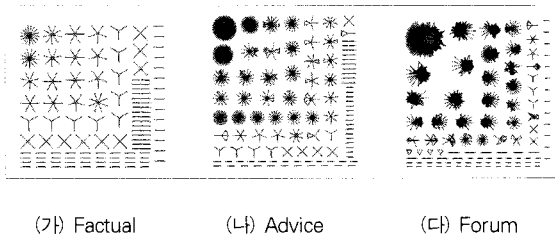


그림 2. 야후 Q/A 소셜 네트워크 구조
Fig. 2. Yahoo Q/A Social Network Structure

첫째는 어떠한 사실 여부에 대한 질문/답변으로 과학적 사실, 프로그래밍, 학문 등이 해당되는 'Factual' 클러스터이다. 두 번째는 조언을 찾고 제공하며, 일상적이면서도 다소 전문적일 수 있는 'Advice' 클러스터로 패션, 결혼/이혼 상담 등이 질문/답변으로 등장한다. 마지막으로 'Forum' 클러스터는 "스포츠에서 누가 이길 것 같은가?", "가장 감명 깊게 봤던 영화는 무엇인가?" 등과 같이 개개인의 의견, 감정 등을 묻고 답하는 특성을 보인다. 이처럼 상이한 질문/답변 특성을 갖는 클러스터내 사용자들은 유사한 질문/답변 행동양식과 지식의 신뢰기준을 가지고 영향력 자를 중심으로 관계망을 형성한다 [16]. 또 다른 연구에서는 가상공간내의 핵심 멤버 찾기를 사용자의 관심과 태그(Tag) 수로 보았으며[17], 구글의 페이지랭크(Page Rank)는 각 사이트의 주목도를 통해 영향력 있는 사이트를 순위화 한다[18]. 이러한 영향력은 온라인이라는 가상 세계에서 어떠한 형태로든 각각의 개체 사이에서 발생하는 관계 속에서 맺어진다는 공통점이 있다. 본 논문에서는 지식검색 서비스에서의 관계를 사용자간 이루어지는 질문과 답변 행위로 본다. 이를 통해 영향력의 구성 요소를 활동성과 신뢰성으로 정의하고 국내 지식검색 서비스로부터 공통적으로 추출 가능한 사용자의 질문수, 답변수, 답변채택수 및 질문확정수를 속성으로 정의한다.

1.2 활동성

지식검색 서비스는 사용자들이 지식을 검색, 공유 및 활용하며, 직접 만들어가는 서비스이다. 이러한 지식검색 서비스의 시스템은 사용자가 궁금한 것을 '질문'하고, 질문기간 동안 다른 사용자가 이것에 대해 '답변'을 하는 형식으로 이루어진다. 질문자는 질문기간이 지나면 수많은 답변들중 하나를 '답변채택'한다. 이것이 하나의 공유 가능한 지식으로 완결되어 저장 된다[4]. 이렇게 저장된 지식들은 사용자들이 함께 공유하고 검색하며 활용할 수 있다.

본 논문에서는 지식 서비스 내에서 사용자들이 지식, 정보의 생산자와 소비자로서 동시에 활동하는 빈도를 활동성이라 정의한다. 이러한 행동양식은 다수의 사용자가 개개인의 작업 및 지식을 공유하고 취합하여 일반적 사실을 도출해 낸다는 집단지성의 원리를 가지고 있다. 이 때 잘못된 사실이 전체 집합에 포함되기도 하는데 잘못된 사실은 전체집합에서 아주 작은 부분이므로 참과 거짓의 구분이 가능하다. 이를 수학적으로 해석하면 다음과 같다. 사용자들마다 알고 있는 지식을 $X_k = a + nk$ 로 표현할 때 a 는 옳은 지식이고 nk 는 잘 모르는 지식을 의미한다. 여기서 k 는 사용자의 번호를 의미한다. 지식 X 에 대해 1부터 k 까지의 사용자가 알고 있는 지식을 모두 합하면 식(2)와 같이 표현된다[5].

$$x_{CI} = \sum_{k=1}^K x_k = Ka + \sum_{k=1}^K nk \dots\dots\dots (2)$$

(CI : Collective Intelligence)

식 (2)에서 옳은 지식은 그 양이 증가할수록 선형적으로 증가하지만 잘 모르는 지식은 모두 같은 지식이 아니며 상호 간섭이 적기 때문에 서로 독립적이라고 볼 수 있다. 지식검색에서의 사용자 질문, 답변 행위도 서로 독립적이다. 사용자 i 의 활동성을 질문과 답변에 대한 회수로 계산시 아래와 같은 수식(3)이 유도 가능하다.

$$U_{i_Activity} = \sum_{n=1}^p (u_{i_question})_n + \sum_{m=1}^q (u_{i_answer})_m \dots\dots\dots (3)$$

- (사용자 $i = \{1, 2, 3, \dots, k\}$ 일 때,
- p, q : 개인별 질문수, 응답수에 대한 변화값,
- $u_{i_question}$: 사용자 i 의 질문,
- u_{i_answer} : 사용자 i 의 답변)

1.3 신뢰성

지식검색 서비스 내에서의 신뢰성이란 질문에 대해 답변이 얼마나 지식·정보로써 믿음만한 것인가에 대한 문제이다. 이에 본 연구에서는 질문자의 답변자 채택을 신뢰성의 핵심 요소로 고려한다. 지식검색에서 지식은 관련 연구에서 언급하였듯이 사용자의 목적에 어떠한 의미가 있는가에 따라 현재 시점에서 창출되는 지식이다. 이러한 점을 고려시 질문에 대한 수많은 답변중 질문자가 답변을 선택하는 행위는 질문자에게 있어 답변 채택자와의 신뢰성 관계가 성립한 것이라 볼 수 있다. 본 논문에서는 카테고리별 사용자의 질문/답변 행동양식과 신뢰성 기준이 상이함을 고려한다. 따라서 카테고리별로 질문자와 답변채택자와의 관계를 통한 소셜 네트워크를 구축하고 이를 통해 각 노드의 중앙성 지수를 산출한다. 한 노드가 네트워크 내에서 중앙성을 갖는다는 것은 그 점이 속한 환경에서 다수의 다른 노드와 연결 관계를 갖는다는 것이다. 이를 지식검색내 사용자간의 지식공유 측면으로 이해한다면 노드를 향해 오는 내향 중앙성(In-Centrality)은 신뢰성의 요소로, 밖으로 나가는 외향 중앙성(Out-Centrality)은 사용자의 지식공유 구축에 대한 활동성 요소로 볼 수 있다. 따라서 소셜 네트워크 분석의 내/외향 연결정도 수식(11)을 토대로 아래와 같은 식(4)를 유도 가능하다.

$$\begin{aligned}
 indegree_{ik} &= \sum_{j=1}^N Z_{ijk} = Z_{jk} \\
 outdegree_{ik} &= \sum_{j=1}^N Z_{ijk} = Z_{ik} \\
 \dots\dots\dots & \dots\dots\dots (4)
 \end{aligned}$$

Where, Z_{ijk} : k 연결망에서 i 사용자로부터 j 사용자와의 관계

$indegree$: 사용자 i 가 다른 모든 사용자들 j 로부터 받는 답변채택수

$outdegree$: 사용자 i 로부터 다른 모든 사용자 j 에게 가는 질문수

1.4 영향력 지수 알고리즘

지식검색 서비스의 가장 큰 장점은 다수의 지식제공자들이 집합적으로 구성하는 지식의 생산이라는데 있다. 하지만 수많은 사용자의 참여는 지식의 전문성 및 신뢰성을 떨어뜨린다. 이를 해결하기 위해 네이버 지식iN에서는 전문가 지수³⁾ 알고리즘(이후, 전문가 지수로 표기)을 제공하나 활동성에만 중

심을 둔 구조로 (가령, 몇몇 사용자간의 의도적인 질문/답변 채택은 전문가 지수 수치의 상승을 가져옴) 문제해결에 한계가 있다[19]. 본 논문에서는 이러한 단점을 보완할 수 있도록 사용자의 활동성과 신뢰성을 기반으로한 알고리즘을 제시하고자 한다. 네이버 지식iN은 질문/답변의 범주를 총 11개의 카테고리와 144개의 세부 카테고리로 분류한다. 각 카테고리에서의 질문/답변 행위는 연구[15]에서와 같이 카테고리별 상이한 활동성과 신뢰성 기준을 가진다. 이에 사용자 k 로 구성된 연결망 내에서 사용자 i 의 영향력 지수 알고리즘을 수식(5)와 같이 제안한다.

$$\begin{aligned}
 U_{i-IV} &= \alpha \sum_{n=1}^p (u_{i-question})_n C_{i-outdegree} \\
 &+ \\
 (1-\alpha) U_{i-ASR} &\sum_{m=1}^q (u_{i-answer})_m C_{i-indegree} \\
 \dots\dots\dots & \dots\dots\dots (5)
 \end{aligned}$$

(α : 카테고리별 질문/응답수의 특성을 고려한 가중치
 C : 카테고리내 사용자 노드의 중앙성 지수

$$C_{i-indegree} = \frac{indegree_{ik}}{k-1}, \quad C_{i-outdegree} = \frac{outdegree_{ik}}{k-1}$$

k : 네트워크에 존재하는 총 노드수, $0 \leq C_i \leq 1$

ASR : 답변 채택률 % = 채택 답변수 / 전체 답변수

사용자의 질문/답변수 증가는 활동성 수치 증가를 가져온다. 이와는 상반되게 질문/답변채택수의 증가로 추가되는 노드가 많아질수록 전체 소셜 네트워크의 중앙성 지수는 감소 효과를 받게 된다. 신뢰성이 결여된 답변수의 증가일지라도 활동성의 증가를 가져오며, 이로 인해 영향력 지수 증가를 가져올 수 있으므로 답변에 대한 산술적 신뢰성을 답변 채택률 (ASR : Answer Selection Ratio)을 적용하여 정규화 한다. 가중치 산정은 인공지능망의 역전파 알고리즘을 이용한다. 이를 통해 카테고리별 질문/응답수의 특성이 가질 최적의 가중치를 산정한다. 인공지능망에선 은닉노드의 활성화 함수로서 시그모이드⁴⁾ 함수를 이용하며, 수식(6)과 같다.

3) 전문가 지수 : {Q&A의 (채택된 답변수*10) + 채택 되지 못한 답변수} + {오픈백과의 작성글수*15 + 기타 글수*3 + 릴레이 MVP 글수*10 + 릴레이 글수}

4) 신경망에서 선형적으로 분리되지 않는 입력 벡터 집합

$$f(\text{sum}_j) = \frac{1}{1 + e^{-\text{sum}_j}} \dots\dots\dots (6)$$

출력 층의 신경세포들은 은닉 층으로부터 입력을 받아 목표 값과 활성 값의 차인 오차를 구한다. 이후 오차신호를 계산하고 신경세포 i 와 j 간의 연결강도들(wij)을 변화시키면서 최적의 가중치를 구한다.

2. 실험 및 평가

본 장에서는 제안한 영향력 지수 알고리즘의 실험 및 평가를 위해 네이버 지식iN 서비스에서의 스포츠(구기 종목) 카테고리 내에서 오가는 질문/답변을 기반으로 사용자간 소셜 네트워크를 형성한다. 이후 영향력 지수 알고리즘과 전문가 지수 알고리즘을 통한 사용자 랭킹화를 거친 후, 랭킹화된 사용자가 제시한 실제 질문/답변 분석을 통해 알고리즘의 랭킹 정확도를 NDCG로 평가한다.

2.1 데이터 셋 구성

네이버 지식검색에서는 하루에만 수천~수만 개의 질문/답변이 이루어진다. 표본 선정의 용이성과 전체 표본 프레임 구성이 어려운 연구 사정상 체계적 표본추출⁵⁾ 방식으로 데이터를 수집하였다. 체계적 표집은 뽑은 사람의 주관이 완전히 배제된 상태에서 동등한 확률로 뽑히도록 표본을 추출하므로 객관적이며 체계적인 방법이다. 본 연구에서는 두 단계에 걸쳐 제안한 알고리즘의 실험을 위한 데이터 셋을 구성한다. 첫째, 카테고리별 질문/답변의 활동성, 신뢰성 특성을 확인하기 위해 네이버 지식iN의 카테고리를 표 1과 같이 Factual, Advice, Forum 클러스터로 분류하였으며, Factual 클러스터내 여행/레저, Advice 클러스터내 교육/학문, Forum 클러스터내 컴퓨터 카테고리에 대해 각각 사용자 50명을 샘플링 하였다. 이후 네트워크 분석도구인 넷마이너(NetMiner)⁶⁾를 통해 네이버 지식iN의 소셜 네트워크 구조 및 질문/답변 특성 분석을 실시하였다. 그림 3은 네이버 지식iN의 소셜 네트워크 구조이다.

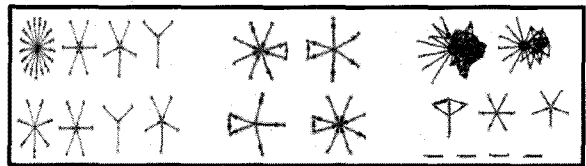
의 에러수정 절차는 종료되지 않으므로(연속미분 불가) 이를 해결하기 위해 비선형 값을 선형적으로 변환해 주어야 함. 시그모이드 함수는 0과 1에 무한히 가까워지는 형태를 가진 함수로 어떤 값이 시그모이드 함수를 거치게 되면 미분 가능한 형태로 변환됨.

5) 체계적 표집(systematic sampling) : 확률적 표집 방식의 하나로, 전체 모집단의 크기를 N이라 하고 일정한 질서에 따라서 n크기의 표본을 추출하는 방식.

6) <http://www.netminer.com>

표 1. 사용자 활동성, 신뢰성에 따른 카테고리 분류
Table 1. Classification Category based on User Activity & Trust

[검색]		구분	카테고리 재분류
카테고리 :	컴퓨터, 통신	Forum	엔터테인먼트, 사회, 문화 스포츠
게임	엔터테인먼트		여행, 레저
비즈니스,경제	쇼핑	Advice	쇼핑, 의학, 건강, 의학
사회,문화	건강, 의학		가정, 생활, 교육, 학문
가정,생활	여행, 레저	Factual	컴퓨터, 통신, 게임
스포츠	교육, 학문		비즈니스, 경제



(가) Factual (나) Advice (다) Forum

그림 3. 네이버 지식iN의 소셜 네트워크 구조
Fig. 3. Naver 지식iN's Social Network Structure

이는 야후의 질문/답변에 대한 특성 분석^[15] 및 유사한 질문/답변 행동양식에 따른 지식의 신뢰기준^[16]에 부합되는 소셜 네트워크 구조를 보여준다. 둘째, 질문/답변을 통한 실험용 소셜 네트워크를 구축하기 위해 구기 종목 카테고리 내에서 활동 중인 사용자중 매 30분 기준 최초답변자를 선정하는 방법으로 표본인원을 추출하였다. 총 표본인원은 997명이었으며, 지식iN의 활동 기간 중 장기 미사용자로 선정하는 기준인 3개월을 데이터 수집 기간으로 정하고 '09년 3~5월간 그들의 질문/답변 관계를 수집하였다. 수집된 질문/답변 셋은 총 3,777개로 실험 및 평가를 목적으로한 3,256개의 훈련 데이터 셋과 521개의 테스트 데이터 셋으로 분류 하였다. 분류된 테스트 데이터 셋의 답변수 분포율은 표 2와 같다.

표 2. 테스트 데이터 셋의 질문/답변 구성
Table 2. Q/A Ratio of Test Data set

답변수	질문/답변 셋 (EA)	Total %
3	260	47.8 %
4	126	24.1 %
5	69	13.2 %
6	34	6.5 %
7	12	2.3 %
8	10	1.9 %
9	12	2.3 %
10	10	1.9 %

답변수가 3~5인 질문/답변 셋의 비중이 85.1 %로 높은 비중을 차지했으며, 제한한 알고리즘의 NDCG 랭킹 정확도 평가를 위해 답변수 1, 2인 데이터 셋에 대해선 전처리 작업을 거쳐 제거하였다. 그림 4는 넷마이너를 통해 구축된 997명 사용자의 질문/답변 관계로 구성된 소셜 네트워크이다.

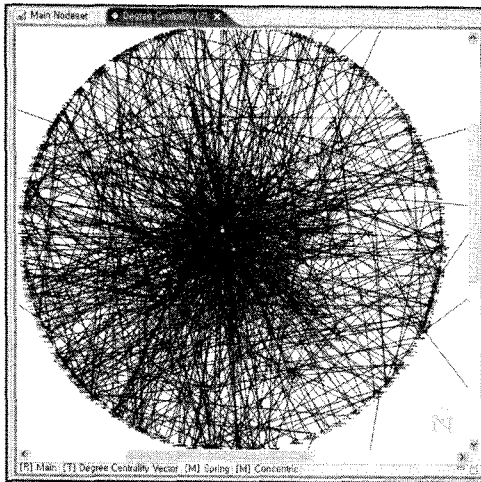


그림 4. 구기 종목 지식N 사용자의 소셜 네트워크
Fig. 4. Ball Games 지식N User's Social Network

2.2 영향력 지수 산정

표 3은 표본인원 997명을 통해 구성된 결속 망에 제한한 영향력 지수 알고리즘을 적용한 결과이다. '구기 종목' 카테고리 내 영향력 지수는 'O'열에 해당하며, 영향력 지수에 따른 랭킹은 'A'열, 전문가 지수에 따른 랭킹은 'B'열에 해당한다. 사용자 997, 165, 4번의 순위로 랭크되었으며, 카테고리별 질문/답변 수에 따른 가중치 값은 질문 : 0.12, 답변 : 0.88 이었다. 전반적인 랭킹 순위는 채택된 답변의 영향을 크게 받았으며, 광고성 글이나 부정확한 방법(동일 사용자의 반복적인 질문/답변)등을 통한 답변수의 증가는 답변 채택률 과 중앙성을 통해 상쇄됨을 확인하였다.

표 3. 영향력 지수 산정 결과
Table 3. I.V Computation Result

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
IR	RIK	NA	RA	Q	NC	D-DEG	IN	OUT	DEG	답변수	질문수	채택률	중심성	중심성	중심성
1	1	897	0.097492	0.003009	183	136	4	0.979	77	0.013038	18.912862	13.524988	0.078270	1117	
2	2	185	0.097202	0.01003	162	88	10	0.912	79	0.1008	5.57766	9.670066	0.91244626	894	
3	4	888	0.061196	0.003009	83	74	3	0.482	9	0.009027	4.886726	4.807753	0.9112496	154	
4	5	170	0.051187	0.003009	115	70	3	0.409	45	0.009027	4.35309	4.362117	1.93100028	154	
5	5	4	0.040113	0.003009	113	74	2	0.402	49	0.004162	2.96898	2.972982	2.61277489	758	
6	6	213	0.022048	0.004212	80	21	5	0.289	57	0.0206	0.907818	0.527979	0.44741618	157	
7	7	211	0.021083	0.003006	31	22	2	0.71	9	0.004212	0.465386	0.467968	0.40704216	126	
8	8	136	0.015845	0.003009	21	16	3	0.762	5	0.009027	0.24072	0.249747	0.20194668	155	
9	9	242	0.014442	0.003009	28	15	4	0.536	13	0.012038	0.21069	0.222666	0.19593594	123	
10	10	139	0.013938	0.001003	14	13	1	0.609	1	0.01003	0.166507	0.17051	0.14903131	151	
11	11	238	0.012038	0.002006	16	14	2	0.875	2	0.004212	0.168804	0.172216	0.148444	142	
12	12	98	0.011938	0.003009	19	14	3	0.737	5	0.009027	0.154462	0.163469	0.13623761	143	
13	13	59	0.012038	0.001003	27	12	1	0.444	15	0.021002	0.144402	0.145845	0.122714328	135	
14	14	182	0.011039	0.001003	14	12	1	0.487	2	0.001003	0.132196	0.133389	0.1169468	123	
15	17	380	0.011039	0.004212	18	11	4	0.978	8	0.018048	0.121365	0.137413	0.11744135	133	

평균 질문수, 답변수, 답변채택수는 표 4와 같으며 답변채택률은 45.6 %로 지식검색의 사용자는 높은 활동성에 비해 낮은 신뢰성을 가지는 것으로 확인되었다.

표 4. 평균 질문/답변수, 답변채택수, 채택률
Table 4. Average counts of Q/A & Selection Answer, ASR

구분	질문수	답변수	답변 채택수	답변 채택률
평균	1.52	2.90	1.66	45.6%

2.3 알고리즘 평가

2.3.1 NDCG

본 논문에서는 알고리즘의 평가를 위해 NDCG [20] 와 3 점 척도 법[1], [3]을 사용한다. NDCG@k는 정보검색 분야에서 검색결과에 랭크 정확도를 측정하는 계량법이다. 이는 50% 이상의 검색 사용자가 검색결과에 1.2 페이지 정도만 참고하는 것에 주목하여 Precision, Recall, F1 과 다르게 rel(Relevance)의 상이한 수준을 분간하고 실제적인 rel에 따른 랭킹을 더 좋게 여긴다. 본 논문에서는 영향력 지수 산정 알고리즘과 전문가 지수 알고리즘을 통해 랭킹화된 사용자가 제시한 답변을 3점 척도 법으로 rel 점수를 부여한 후, NDCG 적용을 통해 알고리즘의 정확도를 비교 평가한다.

2.3.1.1 CG

CG(Cumulative Gain)는 n개의 랭킹 결과물의 등급을 모두 합한 값이다. 관련성에 따라 0~4 사이의 값을 가진다고 할시 Ideal CG 와 현재 CG를 기반으로 절대적인 점수를 도출 가능하다. 예로, 첫 번째 결과에서 여섯 번째 결과까지의 rel 등급이 3, 3, 2, 0, 1, 2 라 한다면 CG는 11의 값을 가진다.

$$CG_P = \sum_{i=1}^p rel_i \dots\dots\dots (7)$$

(P : 랭킹위치, rel : 에디터가 부여한 relevance 점수)

2.3.1.2 DCG

상기 CG 는 실제 랭킹의 위치에 따른 영향을 전혀 받지 않는 방법으로 랭킹에 따른 페널티 적용을 통해 DCG(Discounted Cumulative Gain)를 도출한다. 즉, 에디터가 산정한 rel 점수에는 비례하지만 낮은 랭킹의 결과에 대해서 점진적으로인 페널티를 주기 위해 LOG 함수를 사용한다.

$$DCG_P = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2 i} \dots\dots\dots (8)$$

2.3.1.3 IDCG

수식 (8)을 통한 DCG 점수는 현재 랭킹 알고리즘 결과의 상태를 보여준다. 이를 가장 이상적인 랭킹 점수로 나누어 정규화된 NDCG를 산정한다.

$$NDCG_P = \frac{DCG_P}{IDCG_P} \dots\dots\dots (9)$$

(NDCG : Normalized Discounted Cumulative Gain
IDCG : Ideal Discounted Cumulative Gain)

2.3.2 3점 척도법

NDCG 정확도 적용을 위해선 에디터의 rel 점수 부여가 필요하다. rel 점수 부여는 지식검색 결과물의 평가 기준을 제시한 기준 연구[1], [2]에서 사용한 3점 척도 법을 사용한다. 에디터는 문헌정보학 전공자로 구성되었으며, 평가전 각각의 질문/답변 셋에 대해 이용자가 조회한 문서를 기록한 클릭 로그를 참고하여 이용자의 정보 요구를 파악하고, 질의의 주제 범위와 유의어 등을 확인하였다. 답변의 적합성은 "적합", "보통", "부적합"의 3점 척도를 사용하였으며, 구체적인 평가기준은 표 5와 같다.

표 5. 지식N 답변의 적합성 평가 기준
Table 5. Criteria of 지식N Answer's Suitability Evaluation

적합성	단일 질문에 대한 기준	복수 질문시 답변 충족률
적합	- 질문 내용이 답변에 모두 등장하며, 답변이 질문 내용에 적합한 정보를 제공 - 질문 내용이 답변에 등장하지 않지만, 답변이 질문 내용에 적합한 정보를 제공	60 % 이상

보통	- 질문 내용의 일부에만 해당하는 답변을 제공	30 ~ 60 %
부 적 합	- 질문 내용이 답변에 등장하지 않으며, 답변이 질문과 전혀 관련이 없음 - 질문 내용의 일부에 대한 정보를 제공 하나 답변이 질문 내용의 초점에서 벗 어날 때	30 % 미만

신뢰성은 "높음", "보통", "낮음"의 3점 척도를 사용하였으며, 질문의 유형에 따라 요구되는 답변의 수준이 다르기 때문에, 지식검색 문서에 포함된 질문에 따라 지식형 질문(전문적인 지식을 필요로 하는 범주)과 생활형 질문(생활 상식이나 신변 잡기적인 내용을 묻는 범주)으로 구분하여 rel 점수를 부여하였다. 표 6은 답변의 신뢰성 평가 기준이다.

표 6. 지식N 답변의 신뢰성 평가 기준
Table 6. Criteria of 지식N Answer's Trust Evaluation

신뢰성	지식형	생활형
높음	- 공신력 있는 정확한 출처 - 객관적으로 확실한 근거 (이론/학문적 예시 등) - 논리적 설명 - 자료 첨부 (표, 그림, 사진) - 질문에 대한 핵심 답변	- 논리적인 개인 의견 - 속담, 격언, 생활지식 등 학문적 근거는 없지만 상식적 답변 - 질문에 대한 핵심 답변
보통	- 답변을 하였으나 근거부족	- 정확한 출처가 나오지는 않고 답변의 의견에 의존하나 어느 정도 논리적임
낮음	- 비방, 욕설, 음란한 글 - 명예훼손 글 - 추측성 답변 - 근거가 없는 개인 의견 - 질문과는 전혀 관련 없는 답변 - 광고성 글	

상기 3점 척도 법을 통해 NDCG의 실제 rel 을 산정하며, rel 에 대한 CG는 답변당 최대 4점에서 0점의 분포를 가진다.

질문에 대한 답변은 에디터의 3점 척도 법에 의해 rel 점수가 부여되고, 이를 기반으로 IDCG가 산정된다. 영향력 지수 알고리즘 및 전문가 지수 알고리즘에 따른 랭킹으로 DCG가

계산되며, $NDCG_P = \frac{DCG_P}{IDCG_P}$ 공식에 따라 NDCG 정

확도가 산정된다.

2.4 평가 결과 분석

표 7은 테스트 데이터 셋에 영향력 지수 알고리즘을 적용

한 답변수 3 ~ 10까지의 각 영역별 NDCG@k 의 정확도를 나타낸다. NDCG@3에서 NDCG@10으로 답변수가 증가할수록 각 시도별 정확도 편차가 큰 것을 볼 수 있는데 이는 명백한 답변을 할 수 없는 질문(의견을 묻거나, 개인감정을 묻는 질문 등)에 주로 답변이 많이 달리는 지식검색 시스템의 특성을 보여주는 결과라 할 수 있다.

표 7. 영향력 지수 알고리즘을 통한 랭킹의 NDCG 정확도
Table 7. NDCG of I_V Algorithm

시도	NDCG@3	NDCG@4	NDCG@5	NDCG@6	NDCG@7	NDCG@8	NDCG@9	NDCG@10
4	0.67897959	0.70881275	0.48780771	0.77895569	0.56497044	0.59953388	0.7525048	0.79150331
5	0.799999	0.70285596	0.78485175	0.55018524	0.70690608	0.50726987	0.6642662	0.4894075
6	0.799999	0.7184317	0.74387493	0.73475895	0.70278837	0.89342425	0.7598649	0.62334273
7	0.799999	0.71533689	0.799999	0.71933465	0.5149274	0.70342527	0.6472634	0.61334256
8	0.6523179	0.79127069	0.77527863	0.67845808	0.79453822	0.70342734	0.6034261	0.5632411
9	0.79434371	0.799999	0.74200466	0.65551579	0.77029023	0.70254634	0.5842167	0.59432123
10	0.61839401	0.71740377	0.700894	0.76837793	0.72375877	0.71354352	0.4925789	0.63135678
11	0.799999	0.799999	0.78225426	0.61573124	0.58232342		0.6623954	
12	0.799999	0.7810861	0.70142388	0.73271307	0.69422421		0.669714	
13	0.799999	0.9458133	0.74588897	0.77258814				
14	0.799999	0.7728002	0.996574	0.70711565				
15	0.799999	0.78265435	0.78995489	0.61248539				
16	0.799999	0.61846718	0.799999	0.79028426				
17	0.799999	0.7028671	0.78433144	0.79439054				
18	0.799999	0.799999	0.60352511	0.58228666				
19	0.79117069	0.74544459	0.63120138	0.78814930				
20	0.74631251	0.73793207	0.7587734	0.79439054				
34	0.79127069	0.61646718	0.70141368	0.70711565				
69	0.6523179	0.75338481	0.799999					
126	0.61839401	0.79127069						
260	0.61839401							
avg	0.76201241	0.75398136	0.70260775	0.71271398	0.68895517	0.68035021	0.6615468	0.64963006

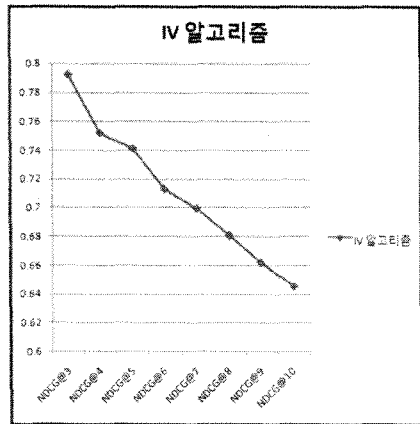


그림 5. 영향력 지수 알고리즘의 NDCG@k 변화그래프
Fig. 5. NDCG@k's Graph of I_V Algorithm

그림 5는 각 영역별 NDCG@k 의 평균 정확도를 나타내는 그래프이다. NDCG@3 은 79.20%의 평균 정확도를 보였으며, NDCG@4 는 평균 75.19%, 이후 NDCG@5~10 에 걸쳐 평균 정확도가 감소하는 결과를 보였다. 특히, NDCG@10 의 경우 64.56% 평균 정확도로 NDCG@1 대비 14.64%의 평균 정확도 하락을 보였으며, 답변수의 증가에 따라 랭킹의 평균 정확도 그래프는 상용로그 함수와 같은

그래프 경향을 보였다. 이는 상위 랭커의 정확도에 큰 비중을 두는 NDCG 평가 방법의 특성이 반영된 것으로 질문에 대한 답변수가 많을 때 사용자는 상위 답변내의 결과만을 주로 참고한다는 연구[21]에 부합되는 결과이다. 질문/답변 특성을 분석한 결과 NDCG@3 ~ 5의 경우 명확한 정보·지식에 관한 공유가 주를 이루었고, 질문과 답변의 실마리가 대체로 짧았으며, 이전 답변자의 정확한 답변이후엔 답변이 추가되지 않는 특징을 보였다. NDCG@6 ~ 10의 경우 질문과 답변의 실마리가 대체로 길었으며, 사용자간의 개인 의견, 감정을 공유하거나 조언을 구하는 질문/답변이 주로 이루어졌다. 또한 질문자가 답변채택을 완료하기 전까지 지속적인 답변이 이루어지는 특징을 보였다. 이는 기존 지식검색 서비스의 질문/답변 특성분석을 실시한 연구[15]와 유사한 결과이다.

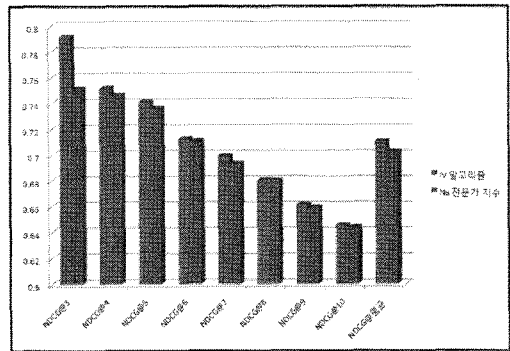


그림 6. 영향력 지수 / 전문가 지수 알고리즘의 랭킹 정확도 비교 그래프
Fig. 6. Comparison Ranking Precision Graph for I_V & 전문가지수 Algorithm

그림 6은 제안한 알고리즘과 전문가 지수 알고리즘의 평균 정확도를 비교한 그래프이다. 테스트 데이터 셋의 47.8%에 해당하는 답변수 3에서의 랭킹 평균 정확도는 영향력 지수 알고리즘이 79.20%로 전문가지수 알고리즘의 75.16% 대비 4.04% 높은 평균 정확도를 보였으며, 테스트 데이터 셋의 85.1%에 해당하는 NDCG@3 ~ NDCG@5의 평균에 대해선 영향력 지수 알고리즘이 76.18%로 전문가지수 알고리즘의 74.49% 대비 1.69% 높은 평균 정확도를 보였다. 사용자 랭킹의 정확도가 높다는 것은 제안한 알고리즘이 지식검색 서비스 내에서 영향력이 높은 사용자를 찾는데 성능이 우수하다는 것을 의미한다. 결과적으로 어떠한 형태(문서, 동영상, 이미지, 음성 등)든 양질의 질문/답변을 통해 지식공유의 근본적 목적에 부합되도록 노력하는 온라인 서비스내의 중심이 되는 사용자를 영향력 지수라는 객관적 수치로 선별 가능함을

보여주는 것이라 할 수 있다. 또한 산정된 영향력 지수는 문제 해결의 실마리를 가진 사용자를 찾음에 있어 훌륭한 지표가 될 수 있으며, 이를 통해 영향력 있는 사용자로부터 사용자가 진정으로 원하는 지식·정보의 획득이 보다 용이해질 수 있을 것으로 판단한다.

IV. 결 론

최근 인터넷 환경은 분산 컴퓨팅 및 서비스 기반 인터넷으로 빠르게 변화하고 있다. 이러한 변화를 가속화시키는 중요 기술로 집단지성은 다양한 분야에서 활용되고 있다. 집단지성을 이용한 지식검색 서비스는 다양한 사용자들의 지식 및 경험이 축적되어 대중적으로 활용 가능한 형태의 정보를 만들어 낼 수 있다는 것이 그 특징이다. 본 논문에서는 지식검색 사용자의 사용 목적과 가장 부합된다고 할 수 있는 영향력자 찾기를 사용자의 활동성과 신뢰성을 기반으로 소셜 네트워크 분석기법을 적용하여 해결하고자 하였다. 본 논문에서 제안한 알고리즘을 통한 영향력 지수는 지식검색 서비스에서 문제 해결의 실마리를 가진 사용자를 찾는 지표가 된다. 또한 어떤 한 가지 이상의 목적을 가지고 소셜 네트워크를 이용하는 사용자의 행동양식을 분석 함에도 유용하게 적용 가능하다. 향후 데이터 셋을 추가하여 카테고리별 사용자 행동양식을 보다 세밀하게 분석하고, 하나의 지식으로 저장된 질문/답변 셋에 대한 다른 사용자의 반응을 추가 요소로 반영한다면 지식검색 서비스의 만족도를 한층 더 향상시킬 수 있을 것으로 본다.

참고문헌

- [1] 박소연, 이준호, 전지운, "지식검색 서비스 개선을 위한 문서의 적합도 및 신뢰도 분석," 한국문헌정보학회지, 제 40권, 제 2호, 300쪽, 2006년.
- [2] 이정태, 송영인, 임해창, "신뢰도 자질을 이용한 지식검색 문서의 품질 평가," 한국정보과학회 학술발표 논문집, 63-65쪽, 2007년 10월.
- [3] Jiwoon Jeon, W.Bruce Croft, Joon Ho Lee, Soyeon Park, "A Framework to Predict the Quality of Answers with Non-Textual Features," In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 228-235, 2006.
- [4] 김희연, "정보사회에서의 지식과 지식검색에 대한 고찰," 정보통신정책, 제18권, 14호, 통권398호, 6-8쪽, 2006년 8월.
- [5] Szuba T, "Computational Collective Intelligence," Wiley and Sons NY, 2001.
- [6] M. V. Vieira, B. M. Fonseca, R. Damazio, P. B. Golgher, D. de Castro Reis and B. Ribeiro-Neto, "Efficient Search Ranking in Social Networks," In Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM), pp. 563-572, 2007.
- [7] E. Spertus, M. Sahami and O. Buyukkokten, "Evaluating Similarity Measures : A Large-Scale Study in the Orkut Social Network," In Proceedings of the 11th ACM SIGKDD international conference on Knowledge Discovery in Data mining, pp. 678-684, 2005.
- [8] L.A. Adamic, O. Buyukkokten and E. Adar, "A Social Network caught in the Web," First Monday, pp. 6-8, 2003.
- [9] D. Fono and K. Raynes-Goldie, "Hyperfriendship and Beyond : Friends and Social Norms on LiveJournal," In Internet Research Annual Volume 4 : AIOR conference, pp. 91-103, 2006.
- [10] C. Dwyer, "Digital Relationships in the 'My Space' Generation : Results From a Qualitative Study," In Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS), pp. 19-28, 2007.
- [11] Freeman L, "Centrality in Social Networks : A Conceptual Classification," Social Networks, No.1, 1979.
- [12] 김용학, "사회연결망 분석," 박영사, 7-36쪽, 82-122쪽, 2003년.
- [13] John P. Scott, "Social Network Analysis : A Handbook," SAGE Publications, London, pp. 7-16, 27-65, 2000.
- [14] 크리스 와이드너, "영향력," 리더스북, 12-25쪽, 2008년.
- [15] Lada A.Adamic, Jun Zhang, Eytan Bakshy, Mark S.Ackerman, "Knowledge Sharing and Yahoo Answers : Everyone Knows Something," WWW 2008, pp. 667-670, April 2008.

- [16] Soojung Kim, Jung Sun Oh, Sanghee Oh, "Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective," ASIST 2007, pp. 5-8, 2007.
- [17] Haiqiang Chen, Xueqi Cheng, Yue Liu, "Finding Core Members in Virtual Communities," WWW 2008, pp. 1233-1234, April 2008.
- [18] Sergey Brin and Larry Page, "The anatomy of a large-scale hypertextual web search engine," In Proceedings of the Seventh International World Wide Web Conference, 1998.
- [19] 김희연, "웹에서 유통되는 정보·지식의 신뢰연구," 정보통신정책, 제19권, 8호, 통권 415호, 19-25쪽, 2007년.
- [20] Kalervo Jarvelin, Jaana Kekalainen, "Cumulated Gain-based Evaluation of IR Techniques," ACM, 2002.
- [21] 장혜란, 이은태, "인터넷포털 지식검색의 질문응답서비스 성능 평가," 한국과학기술정보연구원, 40-57쪽, 2006년.

저자 소개



최창현

2000: 해군사관학교 전자과 학사
 2008: 국방대학교 전산정보학과 석사과정
 현재: 국방대학교 전산정보학과 석사과정
 관심분야: 정보검색, 소셜 네트워크



박건우

1997: 충남대학교 컴퓨터과학과 학사
 2007: 연세대학교 컴퓨터과학과 석사
 2008: 국방대학교 전산정보학과 박사과정
 현재: 국방대학교 전산정보학과 박사과정
 관심분야: 정보검색, 소셜 네트워크, 네트워크 보안



이상훈

1978: 상원대학교 정보통신공학과 학사
 1989: 연세대학교 산업대학원 전산학과 석사
 1997: 일본 교토대학교 정보공학 박사
 서일대학 겸임교수, 충남산업대학교 교수, 일본 교토대학교 교환교수
 현재: 국방대학교 전산정보학과 교수
 관심분야: 정보검색, 데이터베이스, 미디어 융합