

# 소셜 네트워크 기반 사용자 유사성 발견을 통한 개인화 및 소셜 검색

박 건 우<sup>†</sup> · 오 정 운<sup>\*\*</sup> · 이 상 훈<sup>\*\*\*</sup>

## 요 약

소셜 네트워크(Social Network)는 웹 환경에서 개인 중심의 네트워크로 구성되어 웹 사용자별 프로파일을 탐색하고 새로운 연결을 형성함으로써 정보의 소통을 지원한다. 따라서 유사한 내재적 정보를 가진 웹 사용자들로 구성된 소셜 네트워크를 찾아서 검색에 적용한다면 검색의 효율성과 검색 결과에 대한 웹 사용자의 만족도를 향상시킬 수 있다.

본 논문에서는 첫째, 웹 사용자간 직접 또는 간접적인 연결로 구성된 소셜 네트워크를 구성 한다. 둘째, 사용자들의 속성(Feature)에 내재된 정보를 이용하여 주제(topic)별 웹 사용자 간 유사성(Similarity)을 산정한 후, 주제(Topic)별 변화되는 유사성에 따라 소셜 네트워크를 재구성 한다. 마지막으로 산정된 유사성과 웹 사용자들의 검색결과에 대한 만족도, 즉 검색 패턴(Search Pattern)을 비교 실험 한다. 실험 결과 주제별 유사성이 높은 웹 사용자 간에는 검색 패턴 또한 유사함을 확인 하였다. 이와 같은 사실을 검색에 적용한다면 개인화 검색(Personalized Search) 및 소셜 검색(Social Search)의 효율성 및 신뢰성 향상에 기여 할 수 있다.

키워드 : 소셜 네트워크, 개인화 검색, 소셜 검색, 사용자 유사성, 검색 패턴

## Personalized and Social Search by Finding User Similarity based on Social Networks

GunWoo Park<sup>†</sup> · JungWoon Oh<sup>\*\*</sup> · SangHoon Lee<sup>\*\*\*</sup>

## ABSTRACT

Social Networks which is composed of network with an individual in the center in a web support mutual-understanding of information by searching user profile and forming new link. Therefore, if we apply the Social Network which consists of web users who have similar immanent information to web search, we can improve efficiency of web search and satisfaction of web user about search results.

In this paper, first, we make a Social Network using web users linked directly or indirectly. Next, we calculate Similarity among web users using their immanent information according to topics, and then reconstruct Social Network based on varying Similarity according to topics. Last, we compare Similarity with Search Pattern. As a result of this test, we can confirm a result that among users who have high relationship index, that is, who have strong link strength according to personal attributes have similar search pattern. If such fact is applied to search algorithm, it can be possible to improve search efficiency and reliability in personalized and social search.

Keywords : Social Networks, Personalized Search, Social Search, User Similarity, Search Pattern

## 1. 서 론

일반적으로 웹 사용자들은 검색 결과 중 단지 5순위에서 20순위 정도를 살펴볼 뿐, 이때 원하는 정보가 존재하지 않으면 바로 다음 검색 페이지로 넘어가는 경향을 갖고 있다

[1, 2]. 이러한 검색 패턴은 웹 사용자의 검색 의도(Search Intent)라 할 수 있으며, 이는 질의 기록(Query Log)나 웹 사용자의 반응(User Interaction) 등을 이용하여 간접적으로 파악 할 수 있다. 그러나 이러한 접근방법으로는 웹 사용자가 질의를 입력하는 의도를 정확히 파악하기 어려우며 개인 프라이버시(Privacy) 문제로 인해 제한을 받는다. 즉 웹 사용자의 질의는 그 길이가 매우 짧고, 구체적이지 못하며, 동일한 질의를 입력한 경우에도 웹 사용자 간의 검색 의도는 서로 다르다[2, 3]. 그러나 이러한 모호함에도 불구하고 웹 사용자의 검색 의도를 가장 잘 나타내는 것은 기본적으로

<sup>†</sup> 준 회 원 : 국방대학교 전산정보학과 박사과정  
<sup>\*\*</sup> 정 회 원 : 해군 이종무함 작전관  
<sup>\*\*\*</sup> 종신회원 : 국방대학교 전산정보학과 교수  
논문접수: 2009년 2월 23일  
수 정 일 : 1차 2009년 6월 1일  
심사완료: 2009년 6월 5일

질의 그 자체이다[4]. 본 논문에서는 이러한 질의의 모호함과 검색 의도 파악의 어려움을 질의를 입력하는 주제인 웹 사용자로부터 해결하고자 하였다. 이는 소셜 네트워크(Social Network)에 존재하는 웹 사용자의 유사성(Similarity)에 따른 검색의 공통적인 경향을 이용하면 보다 효율적인 검색을 할 수 있다는 생각에서 비롯되었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대하여 기술하고, 3장에서는 주제별 웹 사용자의 유사성을 산정하기 위한 방법을 제시하고, 4장에서는 웹 사용자의 유사성과 검색 패턴사이의 연관성 및 검색 선호도에 대한 실험 결과를 기술한다. 마지막 5장에서는 결론 및 향후 계획에 대하여 기술한다.

## 2. 관련 연구

### 2.1 웹 검색의 효율성 향상

최근 웹 검색 효율의 향상에 대한 많은 연구가 진행되어 왔으며, 이는 크게 개인화 검색(Personalized Search)과 문맥 판단 검색(Contextual Search)으로 구분할 수 있다.

개인화에 대한 연구는 사용자들이 특정한 주제에 더 많은 관심을 가지고 있으며 이러한 경향이 검색 결과에 영향을 미친다는 관찰에서 기인하였다. 이는 일정한 알고리즘 기반의 검색 결과에 사용자의 주제에 대한 선호도를 반영하거나 사용자의 질의 기록 및 반응상의 선호도를 반영하는 방법 등으로 연구가 진행되어 왔다[5, 6]. 그러나 개인정보의 활용은 많은 시간과 저장 공간을 필요로 하며, 사용자의 주제에 대한 선호도를 안다고 할지라도 각각의 주제에 대한 정확한 가중치(Weight)를 반영하는 것은 여전히 어려운 과제이다[6]. 이는 사용자의 질의가 주제와 검색 의도 간의 복잡한 관계로 얽혀 있기 때문이며, 이를 문맥을 통하여 알아내고자 하는 것이 문맥 판단의 방법이다. 이러한 연구는 질의 문맥 파악, 질의 문맥과 이전의 질의 기록, 사용자 반응을 종합적으로 분석하는 방법과 질의와의 일치, 인지도, 문서의 경향(document vectors)에 따른 문맥의 학습 효과를 이용하는 방법 등이 있다[7-9].

본 논문에서는 소셜 네트워크 상에 존재하는 웹 사용자들을 대상으로 주제별 유사성이 큰 웹 사용자 간에는 검색 패턴 또한 유사함을 발견하여 검색에 적용함으로써 개인화 및 소셜 검색에 기여하고자 하였다.

### 2.2 소셜 네트워크(Social Network)

소셜 네트워크는 하나 이상의 상호 의존적인 관계에 의해 구성된 개인 또는 집단으로 구성된 사회적 구조체(Social Structure)로 정의할 수 있다. 대표적인 SNS(Social Network Service)에는 프렌드스터(Friendster), 오컷(Orkut)와 한국의 싸이월드(Cyworld) 등이 있다.

사회학, 통신공학, 경제학 등에서 폭넓게 연구 중인 SNA(Social Network Analysis)는 소셜 네트워크의 형태와 특성을 알고리즘 적으로 연구하고 있으며, 정보과학 분야에

서는 기 구성된 소셜 네트워크의 현상을 웹 환경에 응용하는 연구가 진행되고 있다. 이러한 연구에는 첫째, 소셜 네트워크 상에 존재하는 웹 사용자 간의 연결성(Connectivity)의 확장을 위한 구성원 또는 동호회 등에 대한 검색 효율의 향상 방법과 둘째, 실제 사회 현상과 소셜 네트워크 상의 현상에 대한 비교 분석, 마지막으로 구성의 효율성 및 보안 등 소셜 네트워크 자체에 대한 연구 등 3가지 분야로 나눌 수 있다[10-13].

소셜 네트워크 분석에 대한 접근법은 전체 관계망에서의 위치와 그 효과를 측정하는 위치적 접근법(Positional Approach)과 연결망의 직접적인 관계에 초점을 둔 관계적 접근법(Relational Approach)이 있다. 기본적인 출발점은 행렬과 그래프를 이해하는 것이다. 즉 구성원 (i, j) 사이의 관계가 있고 없음을 '1'과 '0'으로 나타내는 행렬을 인접 행렬(Adjacency Matrix)이라 하며, 완전 연결망의 기본 형태이다[14]. 이러한 소셜 네트워크는 사용자의 명확한 참여 의사 표시에 의하여 형성되고, 구성원 각각의 활동 정도가 서로 연관되어 연결성의 증강이 도모되는 것을 알 수 있다. 또한, 실제 사회에서와 유사한 형태(사람과 사람 간의 관계 설정)로 구성되고 변화되는 것을 관찰할 수 있다. 이러한 특성 때문에 최근의 연구들은 소셜 네트워크의 구성원에 대한 조사를 토대로 그 성격을 규명하는 데 집중되어 있다.

본 논문에서는 소셜 네트워크 상에 존재하는 웹 사용자 간의 연결성이 특정한 주제별로 차이가 있음에 착안하고, 이를 수치화하여 웹 검색에 적용하기 위한 방법론을 제시한다.

## 3. 소셜 네트워크 기반 사용자 유사성(Similarity)

소셜 네트워크의 특징은 구성원간의 연결이 존재한다는 것이며, 사용자는 적어도 한 가지 이상의 목적을 가지고 소셜 네트워크를 이용한다. 그러나 단순한 연결 관계인 사용자 간에는 모든 주제, 즉 질의에 대한 같은 결과를 만족할 수는 없다. 이는 사용자 마다 같은 질의라도 전혀 다른 맥락이나 관심사에서 검색에 대한 결과를 기대 할 수 있기 때문이다. 따라서 본 논문에서는 주제별 사용자 간 유사성이 변화하는 것을 사용자 속성과 주제와의 상호관계를 이용하여 검색 효율성과 신뢰성을 향상 시키고자 하였다.

### 3.1 사용자 유사성의 구성 요소

#### 3.1.1 사용자 프로파일

사용자 유사성을 산출하기 위해 소셜 네트워크에 존재하는 웹 사용자가 갖고 있는 속성(Attributes)과 주제(Topic) 사이의 연관성을 부여하는 것은 매우 중요한 문제이다. 본 연구에서는 웹 사용자 간의 유사성을 산정하기 위해 SNS의 대표적인 사이트인 '오컷'의 일반적인 속성을 이용하였다. '오컷'의 속성 요소인 프로파일은 대분류 항목으로 일반, 친목, 연락처, 직업, 개인정보 등 5개 분야로 분류되어 있다. 이름과 성별, 국가를 기본 입력 요소로 하여 이메일 주소를 통해 사용자를 식별 한다. 친구 찾기 등 다양한 서비스를

제공받기 위해서는 기타 정보 항목에 대해 프로파일을 입력 하도록 구성되어 있다. 중요한 정보는 모두 일반 분야에서 제공하고 있으며, 나머지 4개 분야(친목, 연락처, 직업, 개인 정보)는 일반 분야와 연관되어 자동적으로 함께 수정된다.

사용자 유사성을 산정하기 위한 첫 번째 주된 요소인 사용자 속성은 필수 입력 항목인 일반 분야에 해당되며 다음과 같다.

- 성별 : 여, 남
- 관계현황 : 솔로, 기혼, 연애 중
- 생년 : 1920 ~ 1990년 까지
- 도시 : 직접 입력
- 언어 : 기본 1개 언어
- 고등학교, 대학, 회사/조직 : 직접 입력
- 관심분야 : 친구, 취미활동 파트너, 비즈니스, 데이트 중 선택
- 친구 등록 : 상호 친구 등록, 비 등록으로 구분
- 가입한 동호회의 공통 여부 : 동일한 동호회 가입 여부

3.1.2 ODP(Open Directory Project)

사용자 유사성을 산정하기 위한 두 번째 요소인 주제에 대한 정의이다. 모든 사람이 불특정의 모든 분야에서 유사한 경우를 산정하는 것은 매우 어렵다. 본 연구에서는 모든 주제를 포함하면서 15개의 커뮤니티로 정제된 ‘오컷’의 ODP(Open Directory Project)를 주제로 선정 하였다. ODP는 최적의 검색결과를 얻기 위한 가장 포괄적인 웹 디렉토리이며 구성은 (그림 1)과 같다. 따라서 모든 분야의 주제는 ODP의 15개 디렉토리 중 어느 한 분야에 포함된다고 간주 할 수 있다.

이상과 같이 사용자 간 유사성을 산정하기 위한 구성 요소로 첫째, SNS를 이용하기 위한 입력 항목인 개인 프로파일과, 다음으로 모든 주제를 포함할 수 있도록 정제되어져 있는 ODP에 해당하는 15개의 주제를 정의한다.



(그림 1) ODP의 15개 디렉토리

3.2 사용자 속성과 주제의 연관성

본 연구에서는 사용자의 속성(Attributes)에 내재된 정보를 이용하여 주제 별 사용자 간 유사성을 산정하고자 하였다. 유사성 산정을 위해 3.1절에서 언급한 사용자의 속성인

개인 프로파일, 즉 성별, 관계현황, 생년 등 11개의 항목과 Arts, Business 등 15개의 주제에 대한 상호 연관성 정의가 우선시 되어야 한다.

여러 사회학자, 심리학자들은 사이버공간인 가상 공동체에서의 사람과 사회 현상에 대한 연관성 및 상호작용을 연구 하였다[15, 16]. 물론 이러한 정의는 시간과 장소에 따라 유동적이지만, 본 연구에서는 현재까지 연구 되어졌던 학술을 근거로 <표 1>과 같이 사용자의 속성과 주제 사이의 연관성을 정의 한다. ‘1’은 값을 부여하는(Positive Count) 항목이며, ‘0’은 값을 부여하지 않는(Negative Count) 항목이다. 이에 따라 속성 ‘성별’은 주제 ‘Arts’에 영향을 미치는, 즉 연관성이 있는 것을 의미하며, ‘Business’에는 연관성이 없는 것을 의미한다. 즉, <표 1>의 (a), (b)에서 ‘1’에 해당하는 항목은 사용자 간 관계에 영향을 미치는 값을 의미한다.

유사성(Similarity)은 웹 사용자 간의 속성에 대한 유사점(Resemblance)과 접근성(Access)을 주된 요소로 하고 있다. 속성에 대한 유사점(Resemblance)은 사용자 간 기본적인 정보의 유사성을 나타낸다. 접근성(Access)은 사용자 간 연결의 존재 여부에서 측정 될 수 있으며, Borgatti 는 소셜 네

<표 1> 주제와 속성 사이의 연관성 매트릭스

유사성(S) 속성(F) 주제(T)	유사점 (Resemblance)						접근성 (Access)				
	r-feature 1	r-feature 2	r-feature 3	...	m	Sum (A <sub>fj</sub> )	a-feature 1	a-feature 1	...	m	Sum (R <sub>ij</sub> )
topic 1	1	0	1	-	1	Σ <sub>f</sub> r <sub>f1</sub>	0	1	-	1	Σ <sub>f</sub> r <sub>f1</sub>
topic 2	0	1	1	-	0	Σ <sub>f</sub> r <sub>f2</sub>	1	0	-	1	Σ <sub>f</sub> r <sub>f2</sub>
topic 3	1	1	1	-	0	Σ <sub>f</sub> r <sub>f3</sub>	0	0	-	1	Σ <sub>f</sub> r <sub>f3</sub>
⋮	-	-	-	-	-	-	-	-	-	-	-
topic n	0	1	1	-	1	Σ <sub>f</sub> r <sub>fn</sub>	1	1	-	1	Σ <sub>f</sub> r <sub>fn</sub>

(a) 주제와 속성 사이의 연관성 매트릭스 일반화

유사성(S) 요소 주제	유사점 (Resemblance)										접근성 (Access)		
	성별	관계	생년	도시	언어	학교	대학	회사	관심분야	합계 (R <sub>ij</sub> )	친구 등록	동호 가입	합계 (A <sub>ij</sub> )
Arts	1	0	0	1	1	1	1	1	1	7	1	1	2
Business	0	0	1	1	0	1	1	1	1	6	1	1	2
Computer	1	1	0	1	1	0	0	0	1	5	1	1	2
Game	1	0	1	1	0	0	0	0	1	4	1	1	2
Health	1	0	1	1	0	0	0	1	1	5	1	1	2
Home	1	1	1	1	1	0	0	1	1	7	1	1	2
Kids & Teens	1	1	1	1	1	0	0	0	1	6	1	1	2
News	0	0	0	1	1	1	1	1	1	6	1	1	2
Recreation	0	1	1	1	0	1	1	1	1	7	1	1	2
Reference	0	1	1	1	1	1	1	1	1	8	1	1	2
Regional	1	0	0	1	1	0	0	1	1	5	1	1	2
Science	1	0	1	1	1	1	1	1	1	8	1	1	2
Shopping	1	1	1	1	0	0	0	1	1	6	1	1	2
Society	1	0	1	1	1	0	0	1	1	6	1	1	2
Sports	1	1	1	1	0	0	0	0	1	5	1	1	2

(b) 15 Topics X 11 Attributes 연관성 매트릭스

트위크 상에 존재하는 사용자 간의 친구등록, 동호회 가입 등 사용자 행동(User Behavior)이 접근성에 해당된다고 하였다 [17]. 따라서, 사용자 간 유사성을 산정하기 위한 주된 요소인 속성에 대한 유사점(R)은 <표 1>의 (b)와 같이 ‘성별’에서 ‘관심분야’ 까지 9개 요소에 대해 선정 하며, 접근성(A)은 ‘친구등록’과 ‘동호회 가입’ 등 2개 요소에 대해 선정 한다.

### 3.3 주제별 사용자 유사성 산정 알고리즘

연결망 자료는 대부분 관계의 유무를 0과 1호 나타내는 이분형(Binary) 자료로 표현되므로 이러한 자료를 위해 고안된 유사성과 상이성 지표로 구조적 등위성<sup>1)</sup>을 측정하기도 한다. 대표적인 지표로 단순 합치(Simple Matching), 재카드(Jaccard), 러셀과 라오(Russell and Rao) 등을 들 수 있다 [14]. 본 논문에서 제안하는 유사성 산정 알고리즘은 단순 합치 측도를 기반으로 하였으며, 15개 주제별 사용자 간 유사성 산정 방법은 <식 1>과 같다.

$$S_{ij} = k \cdot \frac{\sum tr_{ij}}{R_{ij}} + (1-k) \cdot \frac{\sum a_{ij}}{A_{ij}} \quad (1)$$

- Where,  $S_{ij}$  : 사용자 i, j 간의 유사성(Similarity)
- R : 주제(topic)와 유사점이(Resemblance) 있는 항목의 전체 집합
- A : 주제(topic)와 접근성이(Access) 있는 항목의 전체 집합
- t : 주제(topic)
- $r_{ij}$  : 사용자 i, j 간 유사점(Resemblance) 일치 항목의 집합
- $a_{ij}$  : 사용자 i, j 간 접근성(Access) 일치 항목의 집합
- k : 주제(topic)의 영향 요소인 R, A 간의 가중치

즉 어떠한 주제  $T_i$  에 대한 사용자 i 와 j 쌍의 유사성은 속성의 유사점(R)과 접근성(A)의 평균화(Normalize)된 합으로 계산한다. 이에 따라 우선 유사점과 접근성에 해당하는 사용자의 속성이 어떤 주제  $T_i$  에 대해 영향 요소로 작용하는지 정의하여야 한다. 영향요소가 있으면 ‘1’의 값을 부여하고, 영향요소가 없을 시에는 ‘0’의 값을 부여한다. 다음으로 영향요소가 있는 항목에 대해서 사용자 i 와 j 간 속성이 일치하거나 유사한 경우를 산정한다. 이때 일치하거나 유사한 사용자 속성 값은 각각 유사점과 접근성 별로 구분하여 해당 분포에 위치한다. 이때 상수 k는 속성의 유사점과 접근성의 균형 조정 요소(Balancing Factor)로 0.82(유사점의 속성 수 / 전체 속성 수)를 부여하였다. 이는 유사점과 접근성이 갖는 세부 속성의 수에 대한 비율이다. 즉 유사점이 전체 11개의 속성 중 9개로 접근성에 대한 속성보다 더 많은 세부 속성을 갖고 있기 때문에 유사점에 더 많은 가중치

1) 연결망 기법은 결속(포괄성, 연결정도 및 밀도 등), 중앙성(중앙에 위치한 정도로 권력과 영향력이라는 개념과 연결되어 가장 많이 쓰이는 지표), 구조적 등위성(연결망 내의 행위자들이 서로 어느 정도 같은 유형의 관계를 맺고 있는지를 나타내는 개념) 등의 주제로 구별 할 수 있다.

를 부여하기 위해 설정한 값이다.

## 4. 실험 및 평가

주제별 사용자 유사성 산정에 대한 효율성을 검증하기 위해 질의 입력에 따른 사용자의 검색 결과 만족도를 비교 실험 한다. 즉 어떠한 주제  $T_k$ 에 대해 개인과 유사성이 높은 사용자 간에는 검색패턴 또한 유사하다는 사실을 검증한다.

### 4.1 실험 데이터 구성

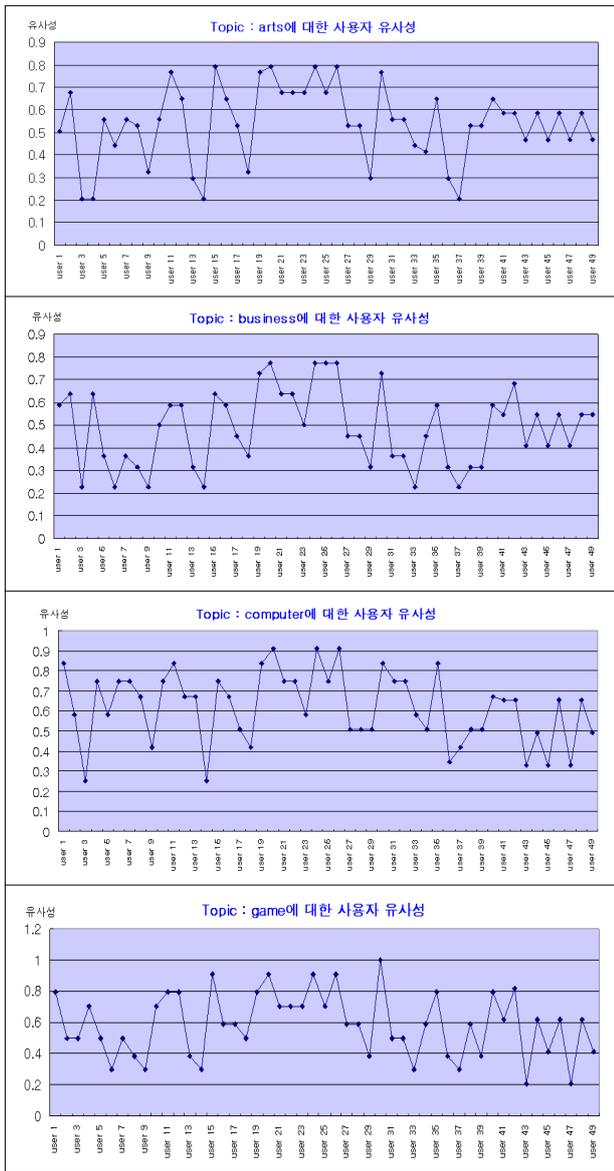
소셜 네트워크는 사용자인 노드(Node)와 노드 간의 연결성을 표현하는 에지(Edge)로 구성된다. 본 연구에서의 실험 데이터 구성은 저자와 직접적인 연결성(1-홉)을 갖는 사용자 40명(40 노드)과 이를 통해 연된(2-홉) 사용자 9명(9 노드), 즉 저자를 포함하여 총 50명을 대상으로 실험을 실시하였다. 또한, 실험의 신뢰성을 위하여 실험 대상자의 개인 프로파일(성별, 연령, 직업, 주소 등)은 다양하게 구성 하였다.

### 4.2 유사성 산정

저자 중심의 소셜 네트워크를 구성 한 후 제안한 알고리즘에 따라 유사성을 산정하였다. <표 2>는 주제별 사용자 유사성을 산정한 결과를 나타낸다. <표 2>와 같이 주제 ‘Arts’에 대한 사용자 간 유사성은 user2가 저자와 연결강도가 가장 강하며, 다음으로 user12 > user1 = user12 > user49 > user33 > user34 > user11 순이다. 또한, 주제 ‘Business’에 대해서는 user2 > user1 > user49 > user1 = user11 > user49 > user34 > user13 > user33 순임을 알 수 있다. 한편, 주제 ‘Arts’와 ‘Business’ 분야에서는 저자와 user2간의 유사성이 가장 큰 것을 확인 할 수 있지만, 주제 ‘Computer’ 및 ‘Game’ 분야에 대해서는 저자와의 유사성이 user1과 user12가 user2 보다 더 큰 것을 확인 할 수 있다.

<표 2> 주제별 사용자 유사성 산정

구분	user 1	user 2	user 00	user 12	user 13	user 00	user 33	user 34	user 00	user 49
Arts	0.50	0.68	...	0.65	0.30	...	0.44	0.41	...	0.47
Business	0.59	0.64	...	0.59	0.32	...	0.23	0.45	...	0.55
Computer	0.84	0.58	...	0.67	0.67	...	0.58	0.50	...	0.49
Game	0.80	0.50	...	0.80	0.39	...	0.30	0.59	...	0.41
Health	0.67	0.75	...	0.67	0.44	...	0.25	0.51	...	0.49
Home	0.77	0.68	...	0.65	0.53	...	0.44	0.53	...	0.59
Kids&teens	0.86	0.64	...	0.73	0.59	...	0.50	0.59	...	0.55
News	0.59	0.64	...	0.59	0.32	...	0.36	0.45	...	0.55
Recreation	0.65	0.56	...	0.53	0.41	...	0.32	0.41	...	0.59
Reference	0.69	0.60	...	0.59	0.39	...	0.40	0.49	...	0.62
Regional	0.67	0.75	...	0.67	0.34	...	0.42	0.51	...	0.49
Science	0.59	0.71	...	0.69	0.39	...	0.40	0.49	...	0.51
Shopping	0.73	0.64	...	0.59	0.43	...	0.36	0.45	...	0.55
Society	0.86	0.77	...	0.73	0.45	...	0.36	0.59	...	0.55
Sports	0.84	0.52	...	0.67	0.51	...	0.42	0.51	...	0.49



(그림 2) 웹 사용자 사이의 관계지수

(그림 2)는 주제 ‘Arts’ 등 4개의 주제에 대해 <표 2>와 같이 산정된 사용자 간 유사성을 그래프로 표현 한 것이다. (그림 2)에서 볼 수 있듯이 주제에 따라 사용자 간 유사성이 달라지는 것을 확인 할 수 있다.

4.3 주제별 질의 선정

다음 단계인 주제별 질의 선정은 ODP의 15개 디렉토리에서 사용자의 선호도가 높은 항목을 아래와 같이 각 주제별 5개씩 선정 하였다.

- Arts : 가요, 극장, 캐릭터, 문학, 디자인
- Business : 통신, 관광, 은행, 경매, 아르바이트
- Computer : 인터넷, 온라인게임, 채팅, 컴퓨터, 친구 찾기
- Game : 비디오, 시뮬레이션, 인터넷게임, 게임, 놀이
- Health : 병원, 한의학, 정신건강, 다이어트, 식품

- Home : 결혼정보, 육아, 요리법, 이사, 임신
  - Kids : 육아건강, 어린이영화, 동화, 어린이잡지, 학습
  - News : 지역신문, 방송, 인터넷 방송, 방송국, 날씨
  - Recreation : 경품, 연예인, 영화, 온라인 서점, 뉴스
  - Reference : 유학원, 외국어, 온라인 학습, 고시, 유치원
  - Regional : 여행정보, 낚시, 자동차, 애완동물, 독서
  - Science : 과학, 사회과학, 과학 잡지, 교육, 환경
  - Shopping : 꽃, 패션, 식음료, 쇼핑, 생활용품
  - Society : 종교, 법률정보, 정치, 이슈, 사회운동
  - Sports : 스포츠, 운동, 구기 종목, 스포츠경기, 건강운동
- 또한, 질의에 대한 검색결과는 구글(google) 검색엔진을 이용 하였으며, 결과에 대한 페이지 수는 3페이지로 제한하였다. 이는 사용자의 검색 결과에 대한 탐색 범위를 설문한 연구 보고서 [18]에 따르면 상위 3페이지까지만 확인하는 사용자가 응답자의 88%에 달한다고 하였기 때문이다.

4.4 유사성에 따른 검색 결과 확인

주제별 사용자 간 유사성이 높은 사용자 그룹은 검색 패

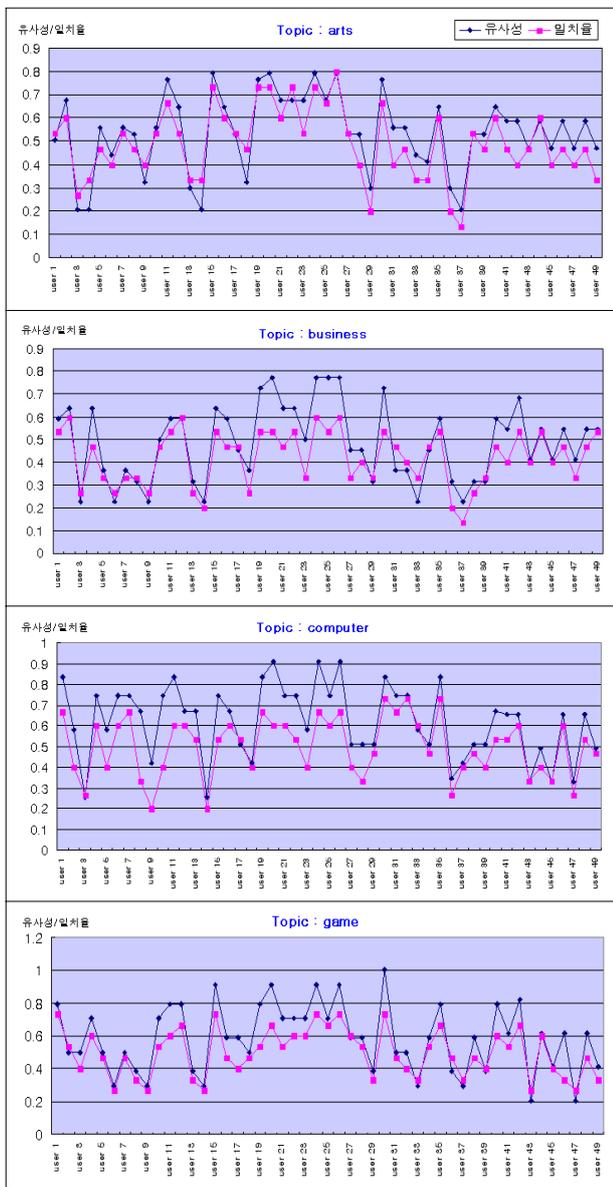
구분	사용자별 검색 정보 선택 항목
주제	질의어, 저자, user1, user2, user3, ..., user49
Arts	가요: 1,3,9, 1,4,10, 1,5,9, 1,2,5, ... 1,2,3
	극장: 1,4,13, 2,5,13, 1,13,26, 2,5,10, ... 3,5,7
	캐릭터: 1,13,16, 1,13,20, 13,16,25, 1,7,9, ... 1,12,13
	문학: 3,6,8, 3,6,9, 1,3,10, 1,3,7, ... 3,5,7
	디자인: 2,4,7, 2,4,7, 2,4,5, 1,2,5, ... 1,5,8
Business	통신: 4,7,16, 4,5,9, 1,4,7, 1,2,4, ... 3,4,7
	관광: 2,5,9, 2,5,7, 5,6,9, 1,3,5, ... 2,5,7
	은행: 10,11,13, 2,10,11, 1,10,11, 1,2,3, ... 2,4,6
	경매: 1,2,7, 1,3,4, 1,2,3, 1,4,5, ... 2,5,7
	아르바이트: 1,4,7, 1,5,7, 1,3,9, 1,2,3, ... 1,2,4

(그림 3) 사용자별 검색결과 선택

턴 또한 유사함을 검증하고자 유사성과 검색 결과의 일치율을 비교 실험 하였다. 앞서 기술한 바와 같이 각 주제별 5개 항목에 대한 질의 선정에 따른 검색 결과에 사용자가 만족하는 검색 결과를 3개씩 선택하도록 했다. (그림 3)은 주제 ‘Arts’ 및 ‘Business’에 해당하는 각 질의에 대한 검색 결과를 각 사용자 선택한 결과이다. (그림 3)과 같은 방법으로 15개 주제별 5개 질의, 즉 총 75개의 질의에 대한 검색 결과를 사용자별로 선택하도록 하여 검색 일치율을 확인 하였다.

4.5 사용자 유사성에 따른 검색 결과 비교 분석

사용자의 속성에 내재된 정보를 이용하여 주제별 사용자 간 유사성을 산정 한 후 질의에 대한 사용자의 검색 결과 만족도를 비교 실험한 결과를 제시한다. (그림 4)는 주제

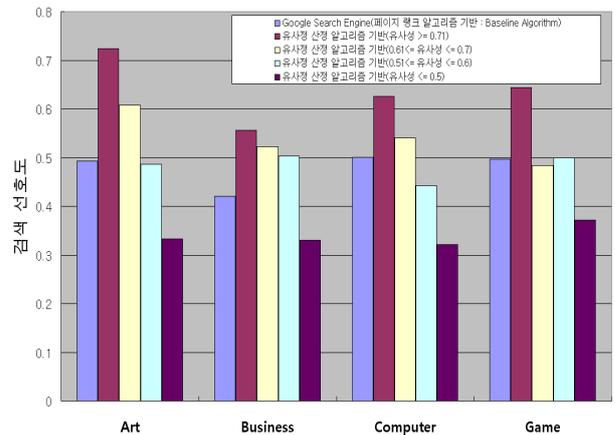


(그림 4) 주제별 유사성에 따른 검색 일치율 비교

‘Arts’, ‘Business’, ‘Computer’, ‘Game’ 등 4개의 주제에 대한 사용자의 유사성과 검색 패턴의 추이를 나타낸다. (그림 4)를 통해 알 수 있듯이 주제별 사용자의 유사성이 높고 낮음에 따라 검색 결과의 선호도 또한 유사한 형태로 변화하는 것을 확인 할 수 있다. 이는 주제별 연결 강도가 강한, 즉 유사성이 높은 사용자 간에는 검색 패턴 또한 유사하다는 사실을 보여주는 것이다. 본 연구에서는 (그림 4)에서 표현 한 4개의 주제 이외의 나머지 11개 주제에 대해서도 유사성과 검색 패턴이 유사한 형태를 나타내는 것을 확인 하였다.

4.6 결과 비교 분석 비교

(그림 5)는 기준이 되는 페이지 랭크 알고리즘과 사용자 유사성 알고리즘을 기반으로 웹 사용자들의 검색 선호도를 비교 분석한 결과이다. 그림에 나타났듯이 소셜 관계 지수가 매우 높은(0.71 이상) 웹 사용자들 간 공유되는 웹 페이지에 검색 선호도가 페이지 랭크 알고리즘에 의해 제공되는 검색 결과에 대한 검색 선호도보다 높게 나타나는 것을 확인 할 수 있다. 즉 소셜 관계 지수가 높은 웹 사용자들이 제공하는 웹 페이지들을 상위에 재순위화 하여 웹 사용자들에게 제공함으로써 검색 선호도가 향상되는 것을 확인 할 수 있다. 따라서 개인의 속성에 따른 주제별 유사성이 높은 사용자 간에는 검색 패턴 또한 유사하다는 사실을 검색에 적용한다면 개인에게 보다 적합하고 연관성이 높은 검색 결과 제공을 통해 개인화 및 소셜 검색의 효율성을 향상 시킬 수 있다.



(그림 5) 검색 선호도 비교 분석 결과

5. 결론 및 향후 연구

본 연구에서는 사용자를 중심으로 형성된 소셜 네트워크를 이용하여 사용자 간의 유사성을 산정하고, 이들 사이의 검색 결과를 분석 하였다. 이를 위해 직접 또는 간접적으로 연결된 사용자 간의 단순한 연결망인 소셜 네트워크를 주제별 변화되는 연결 강도, 즉 유사성에 따라 소셜 네트워크

를 구성 하였다. 이를 기반으로 질의에 대한 사용자의 검색 결과 만족도를 유사성과 비교 실험함으로써 주제별 개인과 유사성이 높은 사용자 간에는 검색 패턴 또한 유사하다는 결과를 확인하였다. 이와 같이 개인의 속성에 따른 주제별 유사성이 높은 사용자 간에는 검색 패턴이 유사하다는 사실을 개인화 및 소셜 검색에 적용한다면 효율성을 향상 시킬 수 있을 것이다.

정보의 존재와 같은 질적 요소의 결합은 여전히 어려운 문제로서 사용자가 가진 내재적 속성을 밝히는 데 중요한 역할을 할 것으로 기대되며, 주제별 영향 요소에 대한 분석 능력의 향상이 필요하다. 따라서 향후에는 사용자 속성과 검색에 있어서의 행동 패턴을 실제 응용프로그램 개발을 통해 검색 엔진에 적용함으로써 검색의 효율성과 신뢰성 향상 정도를 파악해야 할 것이다.

### 참 고 문 헌

[1] K. Tanaka. : Usage of Web Knowledge and User Behaviors for Web Search. In Proceedings of KJDB 2007, 2-19, 2007.

[2] B. J. Jansen, A. Spink and T Saracevic. : Real life, real users, real needs : A study and analysis of user queries on the Web. Information Processing and Management, 36(2) : 207-227, 2000.

[3] J. R. Wen, J. Y. Nie and H. J. Zhang. : Clustering user queries of a Search Engine. In Proceedings of the 10th International World Wide Web Conference, 162-168, 2001.

[4] R. Kraft and C. C. Chang : Search with Context. In Proceedings of the 11th International World Wide Web conference, 477-486, 2002.

[5] T. H. Haveliwala.: Topic-Sensitive PageRank., In Proceedings of the 11th International World Wide Web Conference, 517-526, 2002.

[6] F. Qiu and J. Cho. : Automatic Identification of User Interest For Personalized Search. In Proceedings of the 15th International World Wide Web Conference, 727-736, 2006.

[7] A. Border.: A Taxonomy of Web Search. SIGIR Forum. 36(2), 3-10, 2002.

[8] U. Lee, Z. Liu and J. Cho.: Automatic Identification of User Goals in Web Search. In Proceedings of the 14th International World Wide Web Conference, 391-400, 2005.

[9] D. E. Rose and D. Levinson.: Understanding User Goals in Web Search. In Proceedings of the 13th International World Wide Web Conference, 13-19, 2004.

[10] 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) '07, 563-572, 2007.

[11] D. Fono and K. Raynes-Goldie.: Hyperfriendship and Beyond: Friends and Social Norms on LiveJournal. In Internet Research Annual Volume 4: AIOR Conference, 91-103, 2006.

[12] C. Dwyer: Digital Relationships in the 'MySpace' Generation: Results From a Qualitative Study. In Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS), 19-28, 2007.

[13] C. Dwyer, S. R. Hiltz and K. Passerini.: Trust and A comparison of Facebook and MySpace. In Proceedings of the 13th Americas Conference on Information Systems (AMCIS), 2007.

[14] 김용학, "사회 연결망 분석", 박영사, 2003.

[15] Hagel, Jhon and Arthur G. Armstrong. : Net Gain ; Expending Markets through Virtual Communities, HBS Press, 1997.

[16] 마크스미스 외, "사이버공간과 공동체", 조동기 역, 나남출판, 186, 2001.

[17] S. P. Borgatti and R. Cross.: A Relational View of Information Seeking and Learning in Social Networks. Management Science, 49(4): 432-445, 2003.

[18] IProspect, Search Engine User Behavior, 2006. [http://iprospect.com/WhitePaper\\_2006\\_Search\\_EngineUser\\_Behavior.pdf](http://iprospect.com/WhitePaper_2006_Search_EngineUser_Behavior.pdf)



### 박 건 우

e-mail : pgw4050@hotmail.com

1997년 충남대학교 컴퓨터과학과(학사)

2007년 연세대학교 컴퓨터과학과(석사)

2007년~현 재 국방대학교 전산정보학과  
박사과정

관심분야: 정보검색, 소셜 네트워크, 네트워크, 네트워크 보안



### 오 정 운

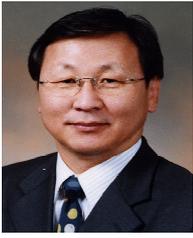
e-mail : influential@hanmail.net

1998년 부경대학교 해양시스템공학(학사)

2009년 국방대학교 전산정보학과(석사)

현 재 해군 이종무함 작전관

관심분야: 정보검색, 소셜 네트워크, 데이터마이닝



## 이 상 훈

e-mail : hoony@kndu.ac.kr

1978년 성균관대학교 정보통신공학과  
(학사)

1989년 연세대학교 산업대학원 전산학과  
(석사)

1997년 일본 교토대학교 정보공학(박사), 서일대학 겸임교수,  
충남산업대학교 교수, 일본 교토대학교 교환교수

현 재 국방대학교 전산정보학과 교수

관심분야: 정보검색, 데이터베이스, 미디어 융합