사용자 프로파일을 이용한 개인화된 토픽맵 랭킹 알고리즘

(Personalized Topic map Ranking Algorithm using the User Profile)

박 정 우 † 이 상 훈 ††
(Jungwoo Park) (Sanghoon Lee)

요 약 토픽맵에서 사용자의 토픽 선택에 따라 제공되는 정보는 개별 사용자의 관심과 배경지식이 고려되지 않고 최초 도메인 전문가에 의해 구축된 토픽맵 상의 토픽(Topic)과 연관되는 관계(Association), 자원(Occurrence)만을 이용하여 사용자에게 토픽맵 정보를 제공하고 있다. 이에 토픽맵은 개인화된 정보 제공 측면의 단점을 보완하고자 개별 사용자를 위한 개인화 기능으로 개인 선호항목 설정, 필터링(Filtering), 범위제한(Scope) 등 사용자가 직접 관심정보를 사전에 설정하는 기능을 제공하고 있으나 토픽맵 사용자를 위한 개인화 측면에서 만족스럽지 못하다.

따라서 본 논문에서는 특정 도메인 토픽맵에서 사용자가 원하는 개인화된 정보를 제공 하기 위해 사용자 클릭정보 수집을 통한 프로파일 정보와 이를 이용한 토픽 선호도 백터(Topic Preference Vector), 토픽맵 지식층의 기본요소인 토픽(Topic)과 관계(Association)를 이용한 개인화된 토픽맵 랭킹 알고리즘(PTR)을 제안한다. 사용자는 PTR 알고리즘을 이용하여 개인 선호도가 고려되어 랭킹된 토픽맵 정보를 제공받을 수 있게 됨으로써 개인화된 정보 제공 측면에서의 성능 향상을 가져올 수 있는 장점을 가진다.

키워드 : 토픽맵, 개인화, 사용자 선호도, 토픽-관계

Abstract Topic map typically provide information to user through the selection of topics, that is using only topic, association, occurrence on the first topicmap which is made by domain expert without regard to individual interests or context. For the purpose of supplementation for the weakness which is providing personalized topic map information. personalization has been studied for supporting user preference through preseting of customize, filtering, scope, etc in topic map. Nevertheless, personalization in current topicmap is not enough to user so far.

In this paper, we propose a design of PTRS(personalized topicmap ranking system) & algorithm, using both user profile(click through data) and basic element of topic map(topic, association) on knowledge layer in specific domain topicmap. therefore User has strong point that is improvement of personal facilities to user through representation of ranked topicmap information in consideration of user preference using PTRS.

Key words : Topicmap, Personalization, User Preference, Topic-Association

† 비 회 원 : 국방대학교 전산정보학과
pjw2236@naver.com
†† 종신회원 : 국방대학교 전산정보학과 교수
hoony@kndu.ac.kr
논문접수 : 2007년 12월 6일
심사완료 : 2008년 7월 29일

정보과학회논문지: 소프트웨어 및 응용 제35권 제8호(2008.8)

1. 서 론

오늘날 우리는 인터넷이나 각종 IT 단말기의 보급과 정보기술의 발달로 정보가 넘치고 있는 환경에 살고 있다. 이처럼 폭발적으로 증가하는 정보들로부터 사용자가 필요한 정보를 얻기 위해 많은 검색 시스템을 이용하는데, 이의 주된 목적은 하나의 특정 토픽에 관한 지식을 확인하는 것이다[1].

지금도 전 세계 수많은 사람들이 매일 자신의 관심영역에 대한 웹서핑을 하면서 개인의 관심주제를 찾고 있으며, 정보 검색시스템을 사용하는 경우에도 그 결과가

방대하기 때문에 사용자가 원하는 정보를 정확하게 얻는 것이 어려운 것이 현실이다. 즉, 오늘날의 검색엔진은 사용자의 관심(Interest)과 배경지식(Context)을 고려하지 않고 검색결과를 제공하는 단점을 가지고 있다. 그래서 이러한 문제점을 해결하기 위해 컴퓨터가 정보의 의미를 분석 가능케 하는 시멘틱웹(Semantic Web)이 등장하였다.

일종의 개념 네트워크라 할 수 있는 시멘틱웹에서는 RDF, OWL, Topic map 등의 온톨로지(Ontology) 언어들이 있다. RDF나 OWL과 같은 기존의 온톨로지 언어들은 기존의 정보 데이타들을 변환하여야 하며 온톨로지 구축에 시간과 노력이 많이 요구되는 단점이 있다. 반면 ISO 표준인 토픽맵(Topic maps)은 지식층(Knowledge Layer)과 정보층(Information Layer)을 쉽게 연결하기 위해 만들어졌다. 이에 토픽맵은 기존 정보의 형태를 변환하지 않고도 정보자원을 통합하여 서로 다른 토픽(Topics)들의 병합이 가능하다. 또한, 관계(Associations) 기능으로 무한한 정보제공과 보다 광범위한 연계가 가능하다는 장점을 가지고 있다. 하지만 이러한 장점을 가진 토픽맵 또한 개별 사용자의 선호도에 맞는 개인화된 정보를 제공하지 못하고 있는 실정이다[2,3].

웹검색 분야에서도 개인화를 위한 선행연구가 여러 가지 있었다. 하지만 보통 사람들이 많은 관심을 가지는 내용을 높은 순위에 올려 개별 사용자에게 제공하는 방식이 오늘날 웹검색의 일반적인 형태가 되었다. 예를 들면, 구글 개인화 검색(Google Personalized Search) 기능은 사용자가 직접 관심을 가지는 카테고리를 선정하여 개인 프로파일(Profile)을 만들어 사용하도록 요구하고 있다. 이러한 프로파일은 검색결과를 개인화 하는데 유용하게 사용될 수 있다. 그리고 상업적 정보 필터링 시스템 역시 이런 방식을 사용하고 있다. 또한, 개인 프로파일은 페이지랭크의 개인화된 결과를 만들어 내기 위한 웹검색에서 배경자료(Context)로 사용되어져 왔다. 이외에도 사용자의 검색 히스토리를 기반으로 하는 카테고리에 사용자 질의를 맵핑시켜 개인화된 검색결과를 끌어내는 기술이 사용되고 있다.

개인화된 검색결과를 제공하기 위해서는 사용자가 명백히 드러내는 사용자 질의 히스토리, 브라우징 히스토리, 웹 커뮤니티, 그리고 클라이언트 측에서의 상호작용이나 클릭 정보 분석 등이 필요하다. 그러나 현재의 검색시스템 보다 더 높은 질적 수준을 가진 개인화를 이루기 위해서는 사용자의 숨겨진 의도를 추론해내는 알고리즘이 개발되어야 실질적인 개인화가 가능하다고 볼 수 있다. 따라서 향후 미래에는 사람의 마음까지 읽어내어 정보를 제공할 수 있는 웹으로의 진화가 진행될 것이며, 이와 더불어 이를 구현하기 위한 수단으로 토픽맵 온톨로지의 중요성은 보다 부각되고 있다.

본 논문에서는 특정 도메인 토픽맵 사용자에게 개인화된 정보를 제공하기 위해 사용자의 토픽 선호도 벡터(Topic Preference Vector)와 토픽맵 지식층(Knowledge Layer)의 핵심요소인 토픽(Topics)과 관계(Associations)를 이용한 개인화된 토픽맵 랭킹(PTR: Personalized Topic map Ranking) 알고리즘과 더불어 이를 이용한 개인화된 토픽맵 랭킹 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로써 토픽맵의 구조와 개인화 기능에 대해 Ontopia에서 제공하는 토픽맵을 기준으로 살펴보고, 토픽맵 랭킹 알고리즘 구현을 위한 토픽 선호도 벡터(Topic Preference Vector)와 BM(Best Matching)에 대해 살펴본다. 3장에서는 제안하는 개인화된 토픽맵 랭킹 알고리즘과 PTRS을 설명한다. 그리고 사례연구를 통해 랭킹 알고리즘의 적용 사례를 보여준다. 4장에서는 기존 토픽맵과 비교를 통해 제안된 시스템의 효율성을 평가하며, 5장에서는 결론을 내린다.

2. 관련연구

2.1 토픽맵(Topic maps)

토픽맵(Topic maps)은 차세대 웹인 시멘틱웹 구현을 위한 지식 표현 방법론으로서, 분산 환경 하에서 지식 구조를 정의하고, 정의된 구조와 지식 자원을 연계하는데 쓰이는 기술 표준이며, 정보자원의 구성, 추출, 네비게이션에 관련된 새로운 패러다임이라 할 수 있다. 현재 시멘틱웹에 대한 기술 연구는 그림 1과 같이 크게 웹의 표준을 담당하고 있는 W3C를 중심으로 한 RDF(Resource Description Framework) 기술과 ISO를 중심으로 한 Topic maps 기술로 나눌 수 있다[2-4].

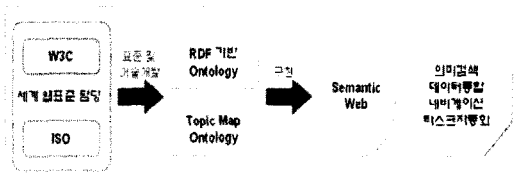

그림 1 온톨로지 종류

2.1.1 토픽맵 구조

토픽맵의 구조는 그림 2와 같이 표현될 수 있다. 토픽맵은 지식층과 정보층의 이중 구조를 나타내는데, 지식층(Knowledge Layer)은 상위 계층으로 토픽(Topics)과 토픽간의 관계(Associations)로 구성된다. 토픽은 특정 주제를 나타내는 표현이고, 관계는 주제들 간의 연관된 관계를 나타낸다. 정보층은 디지털 컨텐츠나 텍스트,

URL 등을 나타내며, 이들 지식층과 정보층은 어커런스(Occurrence)를 통해 상호 연결이 이루어진다.

토픽맵은 토픽(Topics), 관계(Associations), 어커런스(Occurrences)로 구성되어 있다.

토픽(Topics)은 실세계의 주제를 기술한다. 주제는 "기계에 의해 주소화가 가능한 어떤 것"이나, "사람"과 같이 주소화 할 수 없는 것 또는 "음악"과 같은 추상적 개념일 수도 있다.

관계(Associations)는 토픽 집합간의 관계를 기술한다. 관계 구성은 추가적 형태의 정보를 포함할 수 있는데 토픽들 간의 "관계원형"을 명시하거나 또는 각 관계들 간에 작용하는 "역할"이다. 그래서 관계는 계층적인 일반적 개념을 표현하는데 사용될 수 있거나, "회의", "미팅"과 같은 더 복잡한 관계의 그룹을 표현할 수 있다[3].

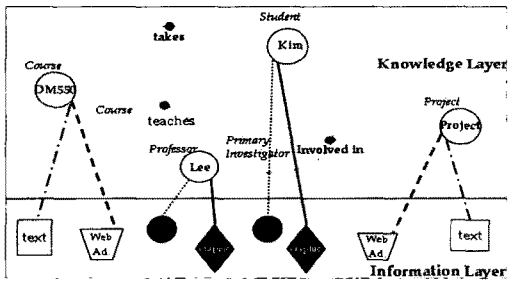

그림 2 토픽맵 구조

어커런스(Occurrences)는 토픽과 관계되는 세부 정보자원이다. 자원은 정보자원과 토픽 또는 정보자원간의 관계원형 둘 다 명시하는데 사용될 수 있다. 예를 들어, 토픽을 어떤 사람으로 표현할 수 있고, 그 사람의 홈페이지 URL인 자원을 가질 수도 있다.

현재의 토픽맵에서 사용자에게 정보를 제공하는 방식은 평소 사용자가 선호하는 토픽이나 관심을 고려하지 않은 채 단지 도메인 전문가가 구축한 토픽맵 속의 토픽(Topics)과 관계(Associations), 어커런스(Occurrences)의 상호 연결을 통해 사용자에게 정보를 제공하고 있는 실정이다.

2.1.2 토픽맵 개인화(Topic maps Personalization)

토픽맵에서 개인화를 위한 일반적 기능은 그림 3과 표 1에서 처럼 Customize와 Filter 기능, 그리고 Scope, 클릭 히스토리 로그정보 표현 등이 있다. Customize에서는 개인 선호도에 맞게 Model, View, Skin에 대한 설정 기능이 있으며, Filter에서는 개인이 선호하는 토픽(Topics), 연관관계(Associations), 어커런스(Occurrences)의 세부항목을 사전에 선정하여 개인화된 정보를 제공받을 수 있다[3]. 그러나 현재의 토픽맵에서 사용자가 관심을 가지는 토픽을 선택했을 때 이와 연관된

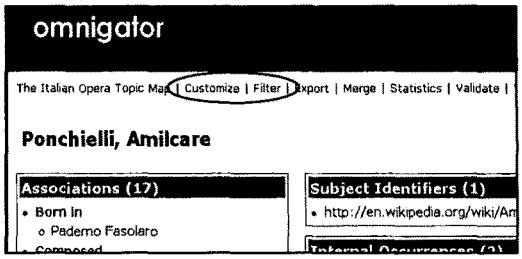

그림 3 토픽맵의 개인화

표 1 토픽맵의 개인화 지원 기능

구 분	Preset	세 부 기 능
Customize	Model	· Nontopoly · Complete
	View	· number of columns
	Skin	· Colors
Filter	Base name context	· Code type, language, name type/form
	Variant name context	· Name form, etc.
	Association context	· Name, publication, publisher, etc.
	Occurrence context	· language, network location, etc.

정보(Association Roles/Topics)들은 토픽맵에서는 단순히 알파벳 순서대로 나열되어 사용자에게 제공되며, Vizigate를 통한 토픽 정보 제공 역시 토픽 Type별 순서로 보여 지게 되어 사용자의 토픽 정보 개인화 측면에서 만족스럽지 못하다.

2.2 사용자 선호도(User Preference)

오늘날 웹상에는 수십억 개의 웹페이지가 넘쳐나고 있으며, 웹페이지의 영역 또한 매우 다양하다. 그러나 실제 사용자가 관심을 가지고 이용하는 웹페이지의 수는 매우 제한되어 있다. 즉, 사용자가 관심을 가지는 토픽(Topics)의 숫자 또한 매우 한정되어 있다. 그리고 웹페이지에 대한 선호도(Preference)는 웹페이지 속에서 사용자가 관심 있어 하는 토픽에 의해 영향을 받는다. 예를 들면, 과학에 관심 있는 물리학자에게 비디오 게임 관련 토픽이 있는 페이지의 내용을 제공한다면 그것이 아무리 질(Quality)이 높고 인기 있는 내용이라 할지라도 관심 없어 할 것이다[5-9].

2.2.1 토픽 선호도 벡터(Topic Preference Vector)

토픽 선호도 벡터는 다음과 같이 정의된다. 토픽의 집합 T=[T(1),....,T(m)]가 m개의 토픽을 가지고 있으며, 사용자가 i번째 토픽에 관심을 가지고 클릭하면, 토픽은 T(i)로 표현되고 선호도 벡터가 부여된다. 따라서 토픽의 집합 T의 벡터 값은 정규화 되어 아래와 같이 표현

될 수 있다.

$$\sum_{i=1}^{m} T(i) = 1 \quad (1)$$

예를 들면, 단 2개의 토픽 "Computers"와 "News"가 있다고 가정한다. 그리고 사용자는 "Computers"에 3회, "News"에 1회 관심을 가졌다면 사용자의 토픽 선호도 벡터는 [0.75, 0.25] 로 나타낼 수 있다[10].

2.3 BM(Best Matching)

BM은 Robertson과 Sparck Jones등이 Okapi에서 사용한 확률기반 용어 가중치 계산함수이다. BM은 최적 매칭(Best-matching)의 약어로, 사용한 가중치 함수의 종류에 따라, BM11, BM15, BM25등으로 나누어진다. BM 함수에 의해 질의어의 용어들의 가중치가 계산되며, 이를 이용하여 문서의 적합도를 계산할 수 있다.

Best-Matching에 의한 확률 모델은 문서가 질의어와 연관이 있는지 없는지 뿐만 아니라 얼마나 연관이 있는지를 나타낼 수 있다. BM을 통해 계산된 질의어에 포함된 용어의 가중치를 통해 질의어의 확장이나 제거에 이용하거나, 검색된 문서의 연관도를 통해 문서의 순위화 등에 사용하여 정보 검색의 효율을 높이는데 사용할 수 있다.

검색시 용어의 빈도수와 문헌의 길이가 고려되지 않을 경우 길이가 긴 문헌에 대해서는 검색의 성능이 떨어지게 되는데, BM에서는 이를 해결하기 위해, BM15, BM11, BM25등에서 용어의 가중치에 의한 2 Poisson 모델을 통해 용어의 빈도수와 문헌의 길이를 반영하여 이를 해결하고 있다. BM은 문서의 검색 뿐만 아니라 Site-Finding과 같은 작업의 평가에서도 유용하게 쓰이는 함수이다[11-13].

2.4 프로파일을 이용한 가중치 부여

Eric hovitz는 확률적 가중치 부여함수인 BM25를 기반으로 아래 그림 4와 식 (2)에서처럼 프로파일을 이용하여 질의 용어(Query Terms)에 대한 문서들의 적절성(Relevance)을 나타내고, 이를 기반으로 문서에 가중치를 부여한 후 순위화 하였다[14].

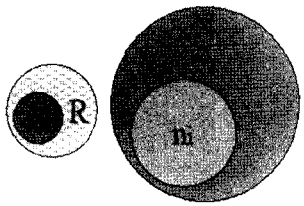

그림 4 Weighting using Profile

$$W_i = \log \frac{(r_i + 0.5)(N - n_i + 0.5)}{(n_i + 0.5)(R - r_i + 0.5)} \quad (2)$$

Where, N: size of corpus
n_i: number of docs contain term i
R: number of docs in profile
r_i: number of docs in R, contain term i

3. 개인화된 토픽맵 랭킹(PTR) 시스템

3.1 개인화된 토픽맵 랭킹 시스템

본 논문에서 제안하는 개인화된 토픽맵 랭킹 시스템은 클라이언트 측에서 작동 되도록 되어있다.

개인화된 토픽맵 랭킹 절차는 다음과 같다. 첫 번째, 사용자가 토픽맵 웹 인터페이스에서 로그인후 선호하는 정보들에 대한 클릭 정보를 사용자 프로파일 데이타베이스(User Profile DB)에 수집한다. 두 번째, 사용자가 실시간 토픽(Topic)에 대한 클릭 이벤트를 발생시킬 때마다 사용자 프로파일 정보를 바탕으로 사용자의 토픽 선호도 벡터(Topic Preference Vector)가 계산되어 DB에서 실시간 유지된다. 세 번째, 사용자 선호도 벡터와 토픽맵 지식층(Knowledge Layer)의 기본 요소인 토픽(Topics)과 관계(Associations)를 이용한 토픽맵 랭킹 알고리즘에 의해 사용자 선호도에 맞는 개인화된 토픽(Topics)정보를 순위화 하여 사용자에게 제공되어 진다.

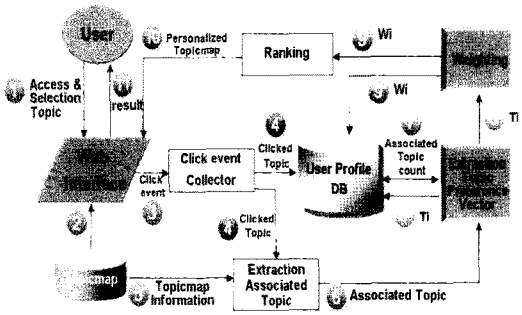

그림 5 개인화된 토픽맵 랭킹 시스템 구성도

표 2 PTR Procedure

PTRS Procedure
User action: click topic output: Personalized topicmap information begin 1. Access to topic map through Web interface 2. User choose a preference topic 3. Count for click through topic and store in user profile 4. Update for user preference vector 5. Ranking associated topics with user click topic by PTR Algorithm 6. Provide user to ranked topics

3.2 개인화된 토픽맵 랭킹 알고리즘

최적의 개인화 알고리즘은 양적으로 풍부하고 높은

질을 가진 프로파일과 웹 코퍼스(corpus)에 달려있다.

본 논문에서 제안하는 시스템을 위한 개인화된 토픽맵 랭킹 기법은 사용자의 Click 정보 수집을 통한 프로파일을 이용 개인 선호도 정보를 반영한 토픽 선호도 벡터를 이용하고, 토픽맵 지식층의 기본 구성요소인 토픽들과 그들의 최적의 연관관계를 이용한다.

3.2.1 개인화된 토픽맵 랭킹 함수

N개의 토픽수를 가지는 토픽맵과 사용자 프로파일(U)은 그림 6과 같이 표현될 수 있다. 토픽맵에서 사용자가 선호하는 토픽을 랭킹하는 함수는 사용자 프로파일과 사용자 선호도 벡터를 이용하여 식 (3)과 같이 표현된다.

$$W_i = T_i \times \log[\frac{(U_i+0.5)(N-n_i+0.5)}{(n_i+0.5)(U-u_i+0.5)}+k] \quad (3)$$

Where, T_i: topic preference vector of topic i

N: size of topics in a specific topic map

n_i: number of topics associated topic i

U: number of topics in user profile

u_i: number of topics in U, associated topic i

k: normalization constants

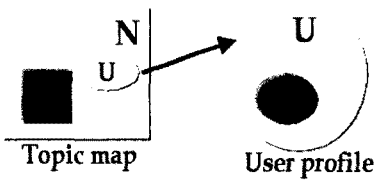

그림 6 토픽맵과 사용자 프로파일

3.2.2 개인화된 토픽맵 랭킹 알고리즘

식 (3)에서 표현된 함수의 개인화된 토픽맵 랭킹 알고리즘은 표 3과 같이 표현된다.

표 3 PTR 알고리즘

PTR Algorithm

1. Let be T = [T_1,.....Tn)]
2. Let T_i be user *preference vector considering* user click counts
3. Let N be the size of topics in a specific topic map
4. Let n_i be the number of topics in user profile
5. Let U be the number of topics associated topic i
6. Let u_i be the number of topics in U, associated topic i
7. Let W_i be the weight of topic i
8. for each T_i, 0< i <N
9. $W_i = T_i \times \log[\frac{(U_i+0.5)(N-n_i+0.5)}{(n_i+0.5)(U-u_i+0.5)}+k]$
10. end
11. for each W_i, 0< i <N
12. T_i.sort() as Wi
13. end
14. Return T

3.3 사례 연구

총 토픽수(N) 20개를 가지는 토픽맵을 구축하면, 특정 도메인 토픽맵은 그림 7과 같이 표현된다. 또한 사용자 프로파일의 총 토픽 수(U)가 5개이고, 사용자별로 토픽 선호도 벡터를 가진다고 가정한다.

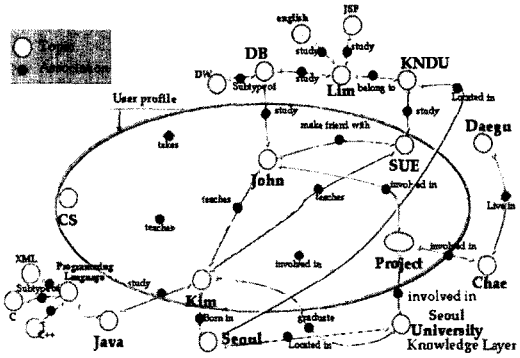

그림 7 20개의 토픽을 가지는 토픽맵

그림 7의 토픽맵에서 사용자 프로파일내의 5개 토픽 모두에 대해 T_i=0.2의 값을 가지고 있는 상태에서 사용자가 선호하는 토픽 "John"을 클릭 하였다고 가정할 때, 전체 토픽수 N=20, 연관되는 토픽의 수는 n_{john}=5, n_{cs}=2, n_{kim}=7, n_{pro}=4, n_{sue}=3이고, 사용자 프로파일 총 토픽수 U=5, 프로파일 내에서 연관되는 토픽의 수는 u_{john}=4, u_{cs}=2, u_{kim}=4, u_{pro}=2, u_{sue}=2이다. 따라서 토픽맵 랭킹 함수(3)에 의해 식 (4)~식 (7)과 같이 선택된 토픽과 연관관계에 있는 토픽별로 가중치가 부여된다. 따라서 표 2와 같이 사용자 선호도와 토픽들 간의 연관관계를 고려해 랭킹된 토픽 정보를 사용자에게 제공할 수 있게 된다.

$$W_{cs} = 0.2 \cdot \log[\frac{(2+0.5)(20-2+0.5)}{(2+0.5)(5-2+0.5)}+0.5] \quad (4)$$
$$= 0.144$$

$$W_{kim} = 0.2 \cdot \log[\frac{(4+0.5)(20-7+0.5)}{(7+0.5)(5-4+0.5)}+0.5] \quad (5)$$
$$= 0.146$$

$$W_{project} = 0.2 \cdot \log[\frac{(2+0.5)(20-3+0.5)}{(4+0.5)(5-2+0.5)}+0.5] \quad (6)$$
$$= 0.088$$

$$W_{sue} = 0.2 \cdot \log[\frac{(2+0.5)(20-3+0.5)}{(3+0.5)(5-2+0.5)}+0.5] \quad (7)$$
$$= 0.110$$

표 4 랭킹된 토픽

순 위	토 픽	가중치(W_i)
1	Kim	0.146
2	Cs	0.144
3	Sue	0.110
4	Project	0.088

4. 성능평가

토픽맵 랭킹 알고리즘을 통해 제공된 토픽 정보의 적절성을 비교하기 위해 기존 토픽맵에서 제공된 정보와 개인화된 토픽맵 랭킹 시스템을 통해 제공된 정보를 사용자 입장에서 비교하여 평가 하였다.

4.1 Recall & Precision

Recall과 Precision은 정보검색(IR)에서 중요한 성능 측정 기준으로 사용하는 지표이다. Precision은 검색 결과 중에 실제로 관계되는 문서가 몇 개인가를 의미한다. 즉, 결과의 "정확도"이다. Recall은 검색어와 관계되는 문서 전체 중에 몇 개를 찾아내느냐를 의미한다. 즉, 결과의 "재현율"이다[15].

4.2 평가 방법

제안한 PTRS을 통해 제공된 정보의 적절성 평가를 위해 20명의 대학원생을 선정하여 온토피아에서 제공하는 Italian Opera 토픽맵을 각자 10일간 사용하여 개인별 프로파일을 유지하고, 제안된 시스템을 통한 토픽정보와 기존 토픽맵에서 제공되는 토픽정보가 얼마나 개인 선호도에 부합되는지에 관한 평가를 한다. 여기서 평가 대상 토픽은 관계역할(Association Role) 20개 이상 가지는 토픽들로 한정하여 사용자들이 직접 적절성을 평가하였다. 그리고 상위 10개의 토픽에 대해서만 Precision과 Recall이 조합된 F-Score를 사용하여 비교평가 하였다.

평가 절차는 다음과 같다. 먼저 사용자 자신이 선호하는 토픽을 선택하게 되면, 사용자는 PTRS로부터 토픽 정보를 제공받게 되고, 기존 토픽맵과 PTRS로부터 제공되는 토픽 정보를 사용자가 직접 적절성을 0(부적절) 또는 1(적절)로 평가하였다.

4.3 평가 결과

평가 참가자 20명이 직접 평가한 데이타를 상위 10개의 토픽에 대한 평균값을 Precision, Recall, F-Score로 구분하여 비교하면, 그림 8, 9, 10과 같이 표현된다.

Precision은 사용자의 선호도가 반영된 토픽이 얼마나 정확히 제공되었는지를 의미한다. 그림 8의 Precision

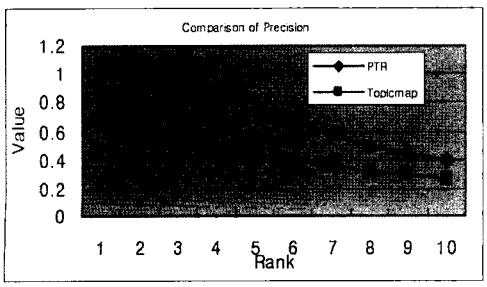

그림 8 Precision 비교

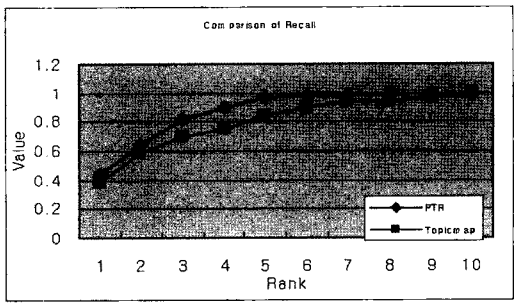

그림 9 Recall 비교

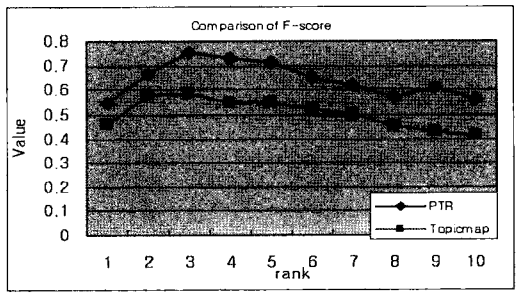

그림 10 F-Score 비교

비교에서는 기존 토픽맵에 비해 32.5%의 효율성 향상을 가져왔다.

그림 9의 Recall 비교에서 또한 11.3% 효율성 향상을 가져왔다. 이는 제한된 범위의 도메인 토픽맵의 특성상 비교적 작은 수의 토픽들 간의 Recall 비교에서는 기존 토픽맵과 큰 차이를 보이지 않음을 알 수 있다. 하지만 Recall 역시 제안된 시스템의 효율성을 보여 주었다. 그림 10에서의 Precision과 Recall을 조합한 F-Score의 비교에서도 역시 23%의 효율성 향상을 보여주어 제안한 시스템의 성능을 입증하였다. 위에서 보여준 결과는 연관 토픽(Association Roles)의 수가 20개 정도인 토픽들을 대상으로 평가 하였다. 하지만 토픽의 수가 방대해지고, 토픽간의 관계(Associations)가 더욱 복잡해진다면 본 시스템에 의한 효율성은 보다 증가할 것이다.

5. 결론 및 향후 연구

본 논문에서는 특정 도메인 토픽맵을 이용하는 사용자가 원하는 정보를 찾을 때 개인의 관심도에 맞고 보다 개인화된 토픽맵 정보를 제공할 목적으로 사용자 선호도와 토픽들 간 최적의 연관관계를 반영한 개인화된 토픽맵 랭킹 알고리즘을 제안하였다.

본 알고리즘을 통해 토픽맵 사용자는 실시간 토픽 선택 행위에 따라 개인 선호도에 맞게 랭킹된 정보를 제공 받을 수 있게 됨으로써 토픽맵의 개인화 성능 향상

을 가져올 수 있는 장점을 가진다. 또한 기존 토픽맵에서 제공하는 정보와의 비교를 통해 제안한 시스템의 효율성을 확인하였다.

향후 연구는 본 연구에서 토픽간의 1차적 관계만을 이용하여 제시한 개인화된 토픽맵 랭킹 알고리즘을 발전시키기 위해 연관되는 토픽 노드의 확장을 통해 보다 개인화되고 최적화된 연관관계에 있는 토픽을 찾아낼 수 있는 방향으로 연구할 예정이다.

참 고 문 헌

[1] Dandan Wang, Darina Dicheva, Christo Dichev, Jerry Akouala, "Retrieving information in topic maps: the case of TM4L," Proceedings of the 45th annual southeast regional conference, pp.88-93, 2007.

[2] 한국전자거래진흥원, TopicMaps 응용 표준 및 활용 가이드라인 개발, 2003.

[3] Steve Pepper, "The TAO of Topic Maps," http://www.ontopia.net

[4] http://www.frotoma.com

[5] Liu F, Yu C and Meng W. "Personalized Web search by mapping user queries to categories," In Proceedings of CIKM'02, pp.558-565, 2002.

[6] Yabo Xu, Ke Wang, Benyu Zhang, Zheng Chen, "Privacy-enhancing personalized web search," Proceedings of the 16th international conference on World Wide Web, pp.591-600. 2007.

[7] Google personalized search, http://www.google.com/psearch

[8] Zhongming Ma, Gautam Pant, Olivia R. Liu Sheng, "Interest-based personalized search," ACM Transactions on Information Systems, Volume 25 Issue 1 Article NO5, 2007.

[9] Kelly D and Teevan J., "Implicit feedback for inferring user preference," SIGIR Forum, 37(2), pp.18-28, 2003.

[10] Feng Qiu, Junghoo Cho, "Automatic Identification of User Interest For Personalized Search," Proceedings of the 15th international conference on WWW, Session: Improved search ranking, pp.727-736, 2006.

[11] S. E. Robertson, S. Walker, "Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval," Proceedings of the 17th annual international ACM SIGIR conference, pp.232-241, 1994.

[12] Karen Spärck Jones, Steve Walker, and Stephen E. Robertson. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments (parts 1 and 2). Information Processing and Management, 36(6):779-840. 2000.

[13] Stephen E. Robertson, Steve Walker, and Micheline Hancock-Beaulieu. Okapi at TREC-7. In Proceedings of the Seventh Text REtrieval Conference. Gaithersburg, USA, November 1998.

[14] Jaime Teevan, Susan T. Dumais, Eric Horvitz, "Personalizing search via automated analysis of interests and activities," Proceedings of the 28th annual international ACM SIGIR conference, Session: User studies, pp.449-456, 2005.

[15] Stefan Buttcher, Charles L. A. Calrke, Brad Lushman, "Term proximity scoring for ad-hoc retrieval on very large text collections," Proceedings of the 29th annual international ACM SIGIR conference, Poster Session, pp.621-622, 2006.

[16] Andrea Ernst-Gerlach, Norbert Fuhr, "Retrieval in text collections with historic spelling using linguistic and spelling variants," Proceedings of the 2007 conference on Digital libraries, pp.333-341, 2007.

박 정 우

1997년 해군사관학교 해양학과(학사). 2005년 아주대학교 경영대학원(석사). 2006년~2008년 국방대학교 전산정보학과 석사과정. 2008~현재 해군 제3함대 목포훈련대장. 관심분야는 정보검색, 시멘틱웹, 데이타마이닝 etc.

이 상 훈

성균관대학교 전자공학과 졸업(학사, 석사, 박사). 1978년~1980년 한국전자통신연구원 전임연구원. 1985년~1986년 일본동경공업대 객원연구원. 1988년~1999년 성균관대학교 교학처장, 전기전자 및 컴퓨터공학부장, 정보통신대학원장, 정보통신기술연구소장. 1996년~1998년 국무총리실 정보화추진위원회 자문위원. 2002년~2003년 한국정보보호학회장. 2003년~2004년 성균관대학교 연구처장. 1982년~현재 성균관대학교 정보통신공학부 교수. 2000년~현재 정보통신부 지정 정보보호 인증기술연구센터장. 관심분야는 정보검색, 데이타베이스 등