

관심 커뮤니티 추천을 위한 분석 질의를 지원하는 지능형 시맨틱 블로그 모델링

양경아[○] 양재동¹ 이마검¹

케이테크[○], 전북대학교 전자정보공학부¹

kayang[○]@knowledgetech.co.kr, {jdyang¹, limajian¹}@chonbuk.ac.kr

Intelligent Semantic Blog Modelling for Supporting Analysis Queries to Recommend Interest Communities

Kyung-Ah Yang[○], Jae-dong Yang¹ and Limajian¹

[○]KnowledgeTech

¹Division of Electronics and Information Engineering, Chonbuk National University

최근 괄목할만한 성장을 이루고 있는 블로그는 오픈 소스로 공개된 소프트웨어를 통해 사용자들이 작성한 글들을 웹 상에 손쉽게 유포하는 새로운 온라인 커뮤니티 방식으로 각광받고 있다. 하지만 블로그는 웹에 분산되어 존재하며 그 자체적으로 독립적인 속성을 지니므로 현재 제공되는 키워드 검색만으로는 블로그 사용자의 관심 분야에 대한 적절한 블로그 자료를 검색하는데 한계가 있다. 블로그 검색에 대한 이러한 문제를 해결하기 위해 태그 접근방식과 시맨틱 웹 접근 방식으로 블로그 공간을 조직화하고 카테고리화하려는 연구가 수행되었다[1][2].

태그를 이용한 블로그 분류 방식은 별도의 훈련 과정없이 사용자가 손쉽게 처리할 수 있다는 장점이 때문에 널리 사용되고 있지만 같은 사물을 지칭하는 두 개의 다른 태그들을 처리하는 매커니즘을 제시하지 못하며 태그 간의 의미적인 정보가 부재하다는 단점을 지닌다. 또한 시맨틱 웹 접근방식은 전문가가 아닌 일반 사용자들이 사용하였을 때 시맨틱 웹 언어를 학습하여 이를 이해하고 처리할 수 있어야 한다는 단점이 있다. 기존의 두 연구방식 모두 사용자들이 동적으로 상호작용하는 블로그 공간 안에서 일어나는 커뮤니티 활동을 모니터링하여 유동적으로 변화하는 상황을 사용자에게 반영하는 시스템은 지원하지 않고 있다. 따라서, 다양한 형식으로 다양한 장소에 분산되어 있는 블로그들을 효율적으로 관리하고 블로그들 사이에 발생하는 상호작용 활동을 분석하여 블로그 사용자의 성향에 맞는 서비스를 제공하기 위해 구조화된 블로그 환경을 지원해야 한다. 블로그 환경을 외부에서 모니터링하고 사용자의 성향을 동적 혹은 정적으로 분류하여 컨텍스트(context)에 맞는 서비스를 제공하기 위한 지능적인 블로그 모델링의 설립은 필수적이라 할 수 있다.

본 논문에서는 계층화된 개념 정보인 온톨로지를 블로그 환경 내에 지식정보로 활용함으로써 블로그 공간을 체계적으로 분석, 관리하여 블로그 사용자 간의 관심 커뮤니티 간의 효율적인 접근 방식 제공하는 지능형 시맨틱 블로그 모델링을 제안한다. 본 논문의 연구내용은 다음과 같이 크게 세 가지 영역으로 구분할 수 있다. 첫째, 지능형 시맨틱 블로그 모델링에서는 온톨로지 개념을 나타내는 개념 객체, 포스트 객체, 블로그 객체와 객체 간 설정 가능한 다양한 관계성을 정의한다. 또한 객체와 관계성에 대해 구조적 연산을 지원하는 연산자를 제공함으로써 다양한 관점에 따라 지능형 시맨틱 블로그를 효과적으로 브라우징 할 수 있도록 한다. 둘째, 클러스터링 연산을 기반으로 관심사가 비슷한 커뮤니티를 생성하고 커뮤니티의 관계성을 분석하여 블로그 사용자의 관심도가 높은 다른 커뮤니티 그룹을 추천할 수 있도록 지원한다. 셋째, 지능형 시맨틱 블로그 모델링에서 시스템 관리자가 의도하는 객체를 참조할 수 있는 시맨틱 질의 기법을 제안한다. 시스템 관리자는 블로그 환경에 대한 시맨틱 질의의 결과를 분석하여 블로그 시스템의 상태를 파악할 수 있다. 지능형 시맨틱 블로그에서는 정적인 분류체계로서의 온톨로

지가 아니라 커뮤니티의 동적인 상호작용 활동에 따라 온톨로지의 개념과 관계성이 생성되어 동적으로 변화함으로써 커뮤니티의 요구에 따라 유연성 있게 관리될 수 있다.

지능형 시맨틱 블로그 모델링에서는 개념과 포스트, 블로그를 기본 객체로 간주한다. 하나의 포스트에 여러 개의 태그를 첨부하듯 하나의 포스트에 여러 개의 개념을 할당할 수 있으며 하나의 개념 객체에 포함된 포스트 객체들은 공통된 관심사를 공유하는 커뮤니티로 간주된다. 특정 포스트에 할당된 여러 개의 개념은 그 포스트를 나타내는 키워드 기능을 하며 다양한 관점을 반영한다. 시맨틱 블로그 모델링에서 모든 객체는 수평적, 수직적으로 다른 객체와 관계성을 가질 수 있다. 시맨틱 블로그 모델링에서는 객체와 관계성에 구조적 연산을 지원하는 연산자를 제공함으로써 다양한 관점에 따라 블로그 객체 간에 설정되어 있는 관계성 링크를 참조하여 효과적으로 브라우징 할 수 있도록 한다. 또한 관계성 링크 참조 연산자와 객체 집합에 대한 AND/OR 연산자 및 클러스터링 연산자 등 다양한 연산자를 지원하여 시스템 관리자가 의도하는 결과를 브라우징할 수 있도록 한다.

또한 클러스터링 연산을 이용하여 관심 커뮤니티를 추출하고 다른 블로그 커뮤니티에 추천할 수 있도록 한다. 객체와 관계성 그리고 이들 데이터 모델에 적용할 수 있는 클러스터 연산자를 이용하여 커뮤니티 그룹을 대상으로 동적으로 세부적인 커뮤니티 즉 클러스터를 생성할 수 있으며 이들에 설정되어 있는 관계성 정보를 분석하여 연관된 다른 커뮤니티를 추천할 수 있다. 커뮤니티 내부의 포스트 객체 집합을 대상으로 [3]를 응용한 클러스터링 연산을 수행하여 클러스터 집단을 추출한 후 각각이 맺고 있는 외부 객체와의 관계성 링크를 분석하여 각 클러스터가 밀접한 상호작용하는 외부의 커뮤니티를 추출할 수 있다. 특정 클러스터와 외부 커뮤니티와의 정보 교환 경로를 클러스터가 속한 같은 커뮤니티의 멤버들에게 제시하여 정보 유통경로를 보다 일반화시킴으로써 커뮤니티에서 유용하게 사용가능한 정보 채널을 공유할 수 있게 된다.

지능형 시맨틱 블로그 모델링에서는 [4]에서 제시하는 질의기법을 참조하여 시스템 관리자가 의도하는 객체를 추출할 수 있는 시맨틱 질의를 제안한다. 지능형 시맨틱 블로그 모델링에서 제공되는 질의는 객체의 참조 범위를 특정 일부분으로 한정하는 영역질의와 참조범위를 동시 또는 선택적으로 제한하는 제약질의로 표현되며 질의기법을 통해 의미적으로 일정한 관계를 지니는 모든 객체들을 참조할 수 있다. 앞서 정의된 연산자는 시맨틱 질의에서도 사용될 수 있다.

본 논문에서 제안하는 지능형 블로그 모델링에서는 블로그 환경을 구조화하고 블로그 커뮤니티의 관계성을 분석하여 사용자의 관심도가 높은 다른 커뮤니티 그룹을 추천할 수 있도록 하여 블로그 사용자의 상황에 맞는 적합한 커뮤니티를 참조할 수 있도록 지원한다. 또한 시스템 관리자가 시맨틱 질의 기법을 이용하여 세부적이고 다양한 표현을 통해 관리자의 목적에 맞는 적절한 결과를 도출하여 블로그 시스템에 결과를 반영할 수 있도록 한다.

참고문헌

- [1] Conor Hayes, Paolo Avesani Sriharsha Veeramachaneni, *An Analysis of Bloggers and Topics for a Blog Recommender System*, 7th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2006.
- [2] Steve Cayzer. *Semantic Blogging and decentralized knowledge management*. Commun. ACM, 47(12):47-52, 2004.
- [3] Chricota, Fabien Jourdan and Guy Melancon. *Software Components capture using Graph Clustering*, Program Comprehension. 11th IEEE International Workshop, 2003.
- [4] 최재훈, 김기현, 양재동, 이동길, "재사용 부품 검색 시스템에서 객체기반 시소러스를 이용한 패킷 질의 확장", 정보과학회논문지:소프트웨어 및 응용, Vol 27, No.2, 2002.