

Non-First Normal Form에 입각한 eBook Annotation 온톨로지의 설계와 구현

신성욱[○] 김종석^{*} 임순범^{**} 최윤철^{*}
[○] 연세대학교 컴퓨터학과
^{**} 숙명여자대학교 멀티미디어과
uk1216[○]@dreamwiz.com

Design and Implementation of eBook Annotation Ontology Based on Non-First Normal Form

Sung-Wook Shin[○], Jong-Suk Kim^{*}, Soon-Bum Lim^{**}, Yoon-Chul Choy^{*}
[○] Dept. of Computer Science, Yonsei University
^{**} Dept. of Multimedia Science, Sookmyung Women's University

요 약

본 연구에서는 온라인 다중 사용자 환경의 eBook 어노테이션 시스템 개발에서 데이터를 의미 기반으로 관리하고, 데이터에 대하여 상호 공통적인 이해를 표현하며, 그리고 데이터에 대한 무결성 검사 등을 지원하기 위해서 eBook 어노테이션 온톨로지를 구축하였다. eBook 어노테이션 데이터에 대한 상호 공통적인 이해의 표현을 위해서 한국 전자책 문서 표준인 EBKS(Electronic Book of Korea Standard)를 기반으로 구축 하였으며 구축된 온톨로지는 Conceptual Graph(CG)를 사용하여 표현하였다. 의미 기반의 처리를 위해서 본 온톨로지에서는 다국어(Multilingua) 관계를 고려하였으며 또한 어노테이션 데이터 생성 시 중요도를 표현하기 위해서 중요성 axiom을 고려했고, NF²(Non-First Normal Form)에 입각하여 온톨로지를 설계함으로써 어노테이션 데이터의 검색에 활용도를 높였다. 제안된 온톨로지는 어노테이션 데이터의 재사용성을 높일 수 있고 의미 정보활 활용함으로써 eLearning, cyberclass과 같은 다중 사용자 환경에서 효과적인 협업을 가능하게 한다. 본 연구에서 구현한 eBook annotation 시스템은 구축한 온톨로지를 사용함으로써 의미 기반의 데이터 관리가 가능하다. 또한 어노테이션 생성 시 온톨로지 구조를 모르더라도 어노테이션을 생성할 수 있는 인터페이스를 구현하였다.

1. 서 론

정보화 시대가 도래함으로써 카메라, 카세트, 텔레비전과 같은 기존 아날로그 미디어들이 디지털화하고 있다. 이러한 디지털화의 변화 속에서 기존의 정보 전달 매체인 책도 디지털화하고 있다. 우리는 이러한 전통적인 책의 디지털화한 형태를 전자책(eBook)이라고 한다. 기존의 책이 디지털화 함으로써 인쇄나 유통 등의 중간 과정들의 단순화, 멀티미디어 기술의 활용으로 보다 실감 있는 정보 전달, 영구적인 수명, 그리고 저장 및 관리의 용이성과 같은 장점을 갖는다[1].

기존의 종이 책에서 독자의 생각을 표현하기 위해서 사용되었던 어노테이션은 웹스터 사전에서 "의견이나 설명을 위해 추가되는 노트"라고 정의되어 있다. 본 논문에서는 전자책이라는 환경을 고려하여서 어노테이션을 "원본 문서에 추가되는 모든 정보"로 정의한다.

전자책에서는 기존의 종이책에서 사용하였던 어노테이션을 적용함으로써 다중 사용자 환경에서 사용자의 생각이나 의견을 추가할 수 있다. 전자책 어노테이션이 기존의 종이책 어노테이션과 구별되는 특징은 어노테이션에 대한 키워드 검색이 가능하고 생성한 어노테이션에 대한 공유가 가능하다는 점이다[2].

본 연구는 어노테이션 데이터의 관리에 관한 연구로써 어노테이션 데이터를 관리하는데 있어서 상호 공통적인 이해를 표현 하며, 의미 기반의 데이터 처리와 저장된 데이터 간의 관계에서 데이터 무결성 검사, 중요성 등의 추론이 가능하도록 온톨로지를 설계한다. 이를 위해 전자책에 대한 상호

공통적인 이해를 표현하는 국내 전자책 표준인 EBKS을 기반으로 온톨로지를 설계하였다[3]. 또한 설계 된 온톨로지는 CG를 사용하여 표현된 후 eBook 어노테이션 시스템 구현 시 사용된다. 구현된 eBook 어노테이션 시스템에서는 다국어 관계, 동의어 관계를 사용하여 어노테이션 데이터를 검색 할 경우 의미 기반으로 수행하도록 하였다. 또한 어노테이션 데이터 입력시 무결성 검사와 중요성 추론을 통해서 사용자는 의미 기반의 데이터 관리 및 처리를 하게 된다. 본 시스템에서는 사용자가 온톨로지의 구조에 대하여 몰라도 의미 기반의 annotation을 수행할 수 있도록 인터페이스를 설계하였다.

2. 관련 연구

본 절에서는 온톨로지의 의미에 대한 정의와 설계 방법을 살펴 본다. 또한 대표적인 온톨로지 표현 언어인 CG에 대하여 소개한다. 그리고 eBook annotation 온톨로지를 설계하기에 앞서 기존의 개발된 온톨로지를 비교, 분석하였고, NF²(Non-First Normal Form)에 관한 정의와 장점, 활용분야를 소개한다.

2.1 기존의 개발된 온톨로지 비교 분석

정보 시스템에서의 온톨로지는 인공지능의 지식 표현, 지식 베이스, 자연어 처리 등의 분야에서 1990년대 초부터 연구되어 왔다. 인공지능에서 온톨로지는 특정 도메인의 지식에 대한 개념화의 명시적인 명세라고 말한다.

온톨로지가 최근에 활발히 연구되고 있는 이유는 지식의 재사용을 가능하게 하고, 사람들이나 시스템간에 서로 다르게 이해할 수 있는 정보의 구조에 대하여 공통적인 이해를 갖게 하며, 특정 도메인의 가정이나 사실들을 가시화하여 보

여주기 때문이다[4].

eBook 어노테이션 온톨로지를 설계하기에 앞서 이미 개발된 온톨로지들을 비교 분석하였다. 비교 분석 대상 온톨로지로는 상식적 지식에 대한 온톨로지인 CYC, 기업 및 상거래를 위한 온톨로지인 TOVE, 그리고 의학 분야의 온톨로지인 UMLS이다. 이들 온톨로지를 비교 분석한 차원은 온톨로지 구성의 일반적인 요소들로서 다음과 같다[5].

- 영역/일반 온톨로지인가?
- 개념의 분류법(Taxonomy)이 존재하는가?
- 개념 구조 및 개념간의 관계를 제공하는가?
- Axiom을 사용하였는가?
- 어떤 표현 언어를 사용하였는가?

[표 1]은 위의 5가지 차원에 따라 CYC, TOVE, UMLS를 분석한 결과를 보여주고 있다.

온톨로지	영역/일반	개념 분류법	개념 구조	Axiom	표현 언어
CYC	일반	○	○	명시적 사용	F.O.L
TOVE	기업	○	○	명시적 사용	F.O.L
UMLS	의학	○	×	사용 없음	ASN.1

[표 1] CYC, TOVE, UMLS 비교 분석

위의 차원에 따라 본 연구에서 설계한 eBook 어노테이션 온톨로지를 평가하면 eBook 어노테이션에 대한 영역 온톨로지이며 EBKS를 중심으로 개념 분류법이 존재하고 개념 구조 및 개념간의 관계가 존재하며 axiom이 제공되고 표현 언어로 CG를 사용하였다.

온톨로지를 표현하기 위한 언어는 웹에서의 표현을 목적으로 하는 언어와 전통적인 언어로 나눌 수 있다. 웹에서의 표현을 목적으로 하는 온톨로지 언어로는 RDF(S), OIL, DAML+OIL 등이 있고 전통적인 온톨로지 언어는 KIF, F-Logic, CG와 같은 것들이 있다. 본 연구에서는 웹에서의 온톨로지 사용이 아니기 때문에 전통적인 온톨로지 언어를 사용하였으며 그 중에서 온톨로지를 다이어그램 형태로 표현이 가능한 CG를 사용하였다.

CG(Conceptual Graph)는 Sowa에 의해 1984년 개발되었으며 Chen의 ER 다이어그램과 비슷하지만 철학적, 심리학적, 언어학적, 객체 지향의 원리에 기반을 두고 비주요한 더욱 진보된 지식 기반 표현 방법이다[6]. CG는 F.O.L(First Order Logic)에 기반을 두고 있으나 다이어그램을 사용하여서 로직을 보여주고 있다.

2.2 NF²(Non-First Normal Form)

NF²는 전혀 정규화되지 않은 관계(Unnormalized)로 정의할 수 있다. 이는 한 속성의 값으로 집합값(Set Value)도 허용된 관계를 나타낸다.

NF²가 갖는 장점으로는 다중값(multi-valued)의 지원이 가능하고, 여러 개의 keyword를 한 속성에 포함 가능하다는 것이다.

활용분야로는 멀티미디어정보검색(Multimedia Information Retrieval)과 웹 검색(Web Retrieval)이다.[7]

3. eBook 어노테이션 온톨로지

지금까지 개발된 eBook annotation 시스템에서는 의미

기반의 정보 저장 및 관리, annotation 데이터에 대한 상호 공통적인 이해의 표현 등을 지원하지 않았다. 이러한 문제를 해결하고자 본 연구에서 개발한 시스템에서는 온톨로지를 사용하여 데이터를 의미 기반으로 표현하고 저장하도록 하였다.

3.1 NF²에 입각하여 eBook annotation 온톨로지 생성

- 온톨로지 캡처

EBKS에서 사용하는 용어들은 eBook 분야에서 공통적으로 사용하는 용어들이며 서로간의 이해를 공통적으로 표현해 주기에 EBKS를 중심으로 concept과 relation을 추출한다. Annotation에 관한 concept과 relation은 온라인 다중 사용자 환경을 고려하여서 추출하였다.

- 온톨로지 코딩

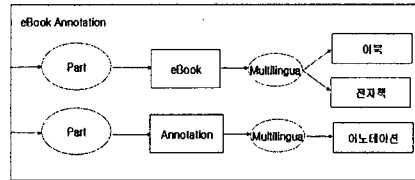
비정규형인 NF²에 기반하여 온톨로지를 코딩하였다. 이를 통해 얻게 되는 장점은 다중 키워드 검색이 가능하다는 것이다[8]. Formal한 언어인 CG를 사용한다. Formal한 언어를 사용함으로써 생기는 장점은 서로 다른 온톨로지 간의 비교가 가능하며 잠재적인 가정들을 명시적으로 드러내고 설계된 온톨로지가 목적과 범위에 적합한지를 평가하는 기준을 제공하는 것이다[4].

- 기존의 온톨로지 통합

아직 eBook 관련 온톨로지가 없다. 상위 레벨의 온톨로지와의 통합은 본 연구에서는 고려하지 않는다.

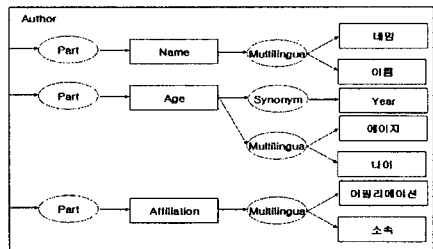
3.2 CG를 이용한 eBook annotation 온톨로지 표현

[그림 1]은 eBook 어노테이션 온톨로지의 상위 구조이다. 'eBook'이라는 것에 대한 한국어 표현으로써 '이북', '전자책'이 쓰이고 있다. 영어에 대한 한국어 표현의 관계를 사용함으로써 어노테이션 데이터 검색의 경우 용어의 같은 의미에 대한 언어적 표현의 차이로 인해서 의미 해석의 결과가 다르게 나오는 것을 막을 수 있다.



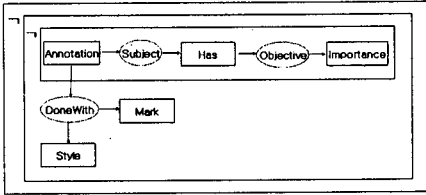
[그림 1] eBook annotation 온톨로지의 상위 구조

[그림 2]는 author의 구성에 관한 온톨로지이다. 여기서 우리는 author에 대한 동의어 관계를 볼 수 있다. 동의어 관계를 통해서 다국어(multilingua) 관계와 같이 동의어에 대한 의미 해석의 차이가 발생하는 것을 막을 수 있다. author의 하위 구성 및 다른 온톨로지의 구성도 구성, 다국어, 동의어 관계를 통해서 이루어진다.



[그림 2] author의 구성 요소

[그림 3]은 온톨로지의 axiom 중 중요성을 보여주고 있다. Annotation이 될 때 mark와 style이 같이 되었을 경우에 그렇지 않은 경우에 비해서 중요성을 갖기에 중요성이라는 axiom이 필요하다.

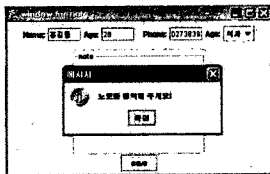


[그림 3] 중요성 axiom

4. eBook 어노테이션 시스템의 구현 및 고찰

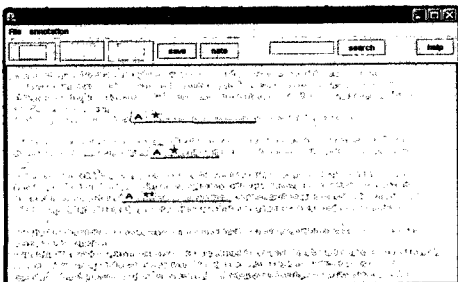
eBook 어노테이션 시스템이 온톨로지를 사용하고 있으나 온톨로지의 구조에 대해서 사용자가 알아야 어노테이션을 수행하는 것이 아니기에 온톨로지를 보여 주고 있지 않다. eBook 온톨로지의 동의어, 다국어, 무결성 검사, 그리고 중요성 관계는 시스템 내부에서 구현되어 있다.

[그림 4]는 무결성 검사를 구현한 화면이다. 노트를 입력하기 위해서 노트 창을 불러 왔으나 사용자가 노트를 입력하지 않고 저장을 하려고 하는 경우에 경고 메시지가 화면에 나타나게 된다. 이를 통해서 노트 입력하는 것을 선택한 경우에 노트가 입력되지 않는 것을 막고 데이터의 무결성을 유지할 수 있도록 한다.



[그림 4] 무결성 검사를 구현한 화면

[그림 5]은 중요성 axiom을 구현한 화면이다. 본 시스템에서는 마크만 수행 된 경우보다 마크와 노트가 같이 수행 된 경우 사용자가 데이터에 대하여 중요성을 표현한 것이라 생각하여서 중요성 관계를 시스템에 구현하였다. 책의 일부분을 선택하여 마크를 수행하면 수행 후 그 부분에 대하여 중요성을 검사 하여서 '별'로써 중요도를 화면에 보여주게 된다.



[그림 5] 중요성 axiom을 구현한 화면

5. 결론 및 향후 연구방향

본 연구에서는 eBook 어노테이션 시스템 개발 시 eBook 어노테이션 데이터에 대한 상호 공통적인 이해의 표현, 데이터의 상호 운용, 의미 기반의 데이터 저장 및 관리, 추론을 지원하기 위해서 eBook 어노테이션 온톨로지를 설계하였다. 데이터 처리 시 동의어와 다국어 사용으로 인해 의미가 다르게 해석 되는 것을 방지하기 위해 동의어와 다국어 (multilingua) 관계를 추가하였다. 그리고 데이터 무결성 검사와 어노테이션입력 시 특정 데이터에 대한 중요성을 반영하기 위해서 axiom을 추가하였다. 설계 된 온톨로지는 의미 기반으로 데이터를 처리하도록 하기 때문에 eLearning, Cyberclass과 같은 다중 사용자 환경에서 협업을 가능하게 하며 어노테이션 데이터의 재사용성을 높일 수 있다.

본 eBook 어노테이션 시스템에서는 온톨로지에서도 의미 기반의 처리를 위해서 고려한 관계를 구현하고 있음을 알 수 있었다. 온톨로지의 구조를 사용자가 모른다 하여도 시스템 내부적으로 구현된 온톨로지는 사용자로 하여금 의미 기반의 데이터 관리를 하도록 한다.

본 연구에서는 고려하지 않은 어노테이션 데이터를 어떻게 사용자 인터페이스에서 보여 줄 것인가에 대한 것도 연구 되어야 할 것이다. 데이터를 의미 기반으로 효과적으로 관리 하는 것도 중요하지만 처리 된 데이터를 eBook의 컨텐츠를 가리지 않으면서 사용자에게 보여 주는 것도 중요하기 때문이다. 또한 eBook 어노테이션 온톨로지의 활용을 위해서는 이미 개발된 다른 온톨로지와의 통합에 대한 연구가 필요하다. 이는 eBook annotation 시스템이 온톨로지를 사용하는 다른 시스템이나 Semantic Web과의 상호 데이터 교환에 필요 하기 때문이다. 특히 Semantic Web과의 연동을 위해서는 RDF(S)나 웹 온톨로지 표현 언어인 WOL(Web Ontology Language)로 본 연구에서 설계된 온톨로지를 표현하는 것이 필요하다.

참고문헌

- [1] 문화관광부, "디지털시대의 전자책(e-Book) 발전방향", 2000
- [2] Ilia A. Ovsiannikov, (1999) Annotation Technology, International Journal of Human-Computer Studies v.50 n.4
- [3] EBK(e-Book of Korea) Consortium. (2001) A Study of Korean Standardization of eBook documents", Technical Report
- [4] Noy, Natalya Fridman and Deborah L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology
- [5] Sowa, John F., ed. (1998) Conceptual Graphs, draft proposed American National Standard, NCITS.T2/98-003.
- [6] Noy, Natalya Fridman and Carole D. Hafner. (1997) The State of the Art in Ontology Design: A Survey and Comparative Review. AI Magazine. Fall 1997.
- [7] Kalervo J., and Timo N. (1995) AN NF2 RELATIONAL INTERFACE FOR DOCUMENT RETRIEVAL, RESTRUCTURING AND AGGREGATION, SIGIR '95. Seattle WA USA.
- [8] FUHR, N., AND ROLLEKE, T. (1996) A probabilistic NF2 relational algebra for integrated information retrieval and database systems. In Proceedings of the 2nd World Conference on Integrated Design and Process Technology. Society for Design and Process Science, Austin, Tex. -17-30.