

# 온톨로지를 이용한 eBook Annotation 시스템의 설계 및 구현

신성욱\*, 김종석\*\*, 고승규\*\*, 임순범\*\*\*, 최윤철\*\*  
\*성결대학교 컴퓨터공학과  
\*\*연세대학교 컴퓨터학과  
\*\*\*숙명여자대학교 멀티미디어학과  
e-mail: uk1216@dreamwiz.com

## Design and Implementation of eBook Annotation System using Ontology

Sung-Wook Shin\*, Jong-Suk Kim\*\*, Seung-Kyu Ko\*\*, Soon-Bum Lim\*\*\*, Yoon-Chul Choy\*\*  
\*Dept. of Computer Engineering, Sungkyul University  
\*\*Dept. of Computer Science, Yonsei University  
\*\*\*Dept. of Multimedia Science, Sookmyung Women's University

### 요 약

본 연구에서는 온라인 다중 사용자 환경의 eBook annotation 시스템 개발에서 데이터를 의미 기반으로 관리하고, 데이터에 대하여 상호 공통적인 이해를 표현하며, 그리고 데이터에 대한 무결성 검사 등을 지원하기 위해서 eBook annotation 온톨로지를 설계하였다. eBook annotation 데이터에 대한 상호 공통적인 이해의 표현을 위해서 한국 전자책 문서 표준인 EBKS(Electronic Book of Korea Standard)를 기반으로 설계 하였으며 설계 된 온톨로지는 Conceptual Graph(CG)를 사용하여 표현하였다. 의미 기반의 처리를 위해서 본 온톨로지에서는 동의어(synonym) 관계와 다국어(multilingua) 관계를 고려하였으며 또한 annotation 데이터 생성시 오류 방지와 중요도를 표현 하기 위해서 무결성 검사, 중요성 axiom을 고려 했다. 제안된 온톨로지는 annotation 데이터의 재사용성을 높일 수 있고 의미 정보를 활용함으로써 eLearning, cyberclass과 같은 다중 사용자 환경에서 효과적인 협업을 가능하게 한다. 본 연구에서 구현한 eBook annotation 시스템은 설계한 온톨로지를 이용함으로써 의미 기반의 데이터 관리가 가능하다. 또한 annotation 생성 시 온톨로지 구조를 모르더라도 annotation을 생성할 수 있는 인터페이스를 구현하였다.

### 1. 서론

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 사용 되었던 annotation은 웹스터 사전에서 “의견이나 설명을 위해 추가되는 노트”라고 정의 되어 있다. 본 논문에서는 전자책이라는 환경에서 annotation을 “원본 문서에 추가되는 모든 정보”로 정의한다.

전자책에서는 기존의 종이책에서 사용하였던 annotation을 적용함으로써 다중 사용자 환경에서 사

용자의 생각이나 의견을 추가할 수 있다. 전자책 annotation이 기존의 종이책 annotation과 구별되는 특징은 annotation에 대한 키워드 검색이 가능하고 생성한 annotation에 대한 공유가 가능하다는 점이다.[2]

전자책에서의 annotation 시스템 구성은 인터페이스 측면과 원본 문서와 annotation 관리 측면으로 나눌 수 있다. Annotation 인터페이스 연구는 다중 사용자 환경에서 효과적으로 annotation을 생성하고 출력하는 것에 관한 것이다. 원본 문서와 annotation 데이터 관리에 관한 연구는 입력된 annotation 데이터를 어떻게 원본 문서와의 관계를 고려하여 저장하고 관리하는지에 대한 연구이다.

본 연구는 annotation 데이터의 관리에 관한 연구로써 annotation 데이터를 관리하는데 있어서 상호 공통적인 이해를 표현 하며, 의미 기반의 데이터 처리와 저장된 데이터 간의 관계에서 데이터 무결성 검사, 중요성 등의 추론이 가능 하도록 온톨로지를 설계한다. 이를 위해 전자책에 대한 상호 공통적인 이해를 표현하는 국내 전자책 표준인 EBKS을 기반으로

로 온톨로지를 설계하였다.[3] 또한 설계 된 온톨로지는 CG를 사용하여 표현된 후 eBook annotation 시스템 구현 시 사용된다. 구현된 eBook annotation 시스템에서는 다국어 관계, 동의어 관계를 사용하여 annotation 데이터를 검색 할 경우 의미 기반으로 수행하도록 하였다. 또한 annotation 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 온톨로지와 온톨로지 기반의 annotation 시스템을 비교, 분석하였다.

### 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 시스템 간에 서로 다르게 이해할 수 있는 정보의 구조에 대하여 공통적인 이해를 갖게 하며, 특정 도메인의 가정이나 사실들을 가시화하여 보여주기 때문이다.[4]

eBook annotation 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.[6]

- 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	o	o	명시적 사용	F.O.L
TOVE	기업	o	o	명시적 사용	F.O.L
UMLS	의학	o	x	사용 없음	ASN 1

[표1] CYC, TOVE, UMLS 비교 분석

위의 차원에 따라 본 연구에서 설계 한 eBook annotation 온톨로지를 평가하면 eBook annotation에 대한 영역 온톨로지이며 EBKS를 중심으로 개념 분류법이 존재하고 개념구조 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, CG 와 같은 것들이 있다. 본 연구에서는 웹에서의 온톨로지 사용이 아니기 때문에 전통적인 온톨로지 언어를 사용하였으며 그 중에서 온톨로지를 다이어그램 형태로 표현이 가능한 CG 를 사용하였다.

CG(Conceptual Graph)는 Sowa 에 의해 1984 년 개발 되었으며 Chen 의 ER 다이어그램(Chen, 1981)과 비슷하지만 철학적, 심리학적, 언어학적, 객체 지향의 원리에 기반을 두고 비유얼한 더욱 진보된 지식 기반 표현 방법이다.[5] CG 는 F.O.L(First Order Logic)에 기반을 두고 있으나 다이어그램을 사용하여서 로직을 보여주고 있다.

### 2.2 온톨로지 기반 annotation 시스템

웹 문서에 대한 annotation을 지원 하는 기존의 시스템들도 온톨로지의 사용 목적은 의미 기반으로 데이터를 처리하도록 하기 위함이다.

#### • CREAM

CREAM 에서는 HTML 문서에 대한 annotation 을 하는 데 있어서 온톨로지 기반으로 할 수 있도록 하는 시스템이다. CREAM 에서 생성 된 annotation 데이터는 온톨로지를 기반으로 기존의 데이터와 통합을 하기 때문에 구조화된 형태를 갖게 되며 지식의 재사용성을 높이는 결과를 갖고 온다. 사용자 인터페이스에서 온톨로지를 그래픽하게 보여 줌으로써 사용자가 쉽게 온톨로지 기반으로 annotation 을 하도록 하고 있다. 웹 기반의 annotation 시스템이기에 온톨로지를 표현하기 위해서 DAML+OIL 을 사용하였다. CREAM 을 시스템으로 구현한 것을 Ont-O-Mat 이라고 한다.[7]

#### • Ontology-based Text Annotation Tools (in KA<sup>2</sup>-initiative)

Web 문서에 대하여 온톨로지 기반으로 annotation 을 할 수 있도록 하는 KA<sup>2</sup>-initiative 에서는 의미 기반의 annotation 을 수행하는 것을 문서에 대하여 annotation 을 수행하는 것과 개발 된 영역 온톨로지의 개발과 진화간의 순환 작업이라고 말하고 있다. Annotation 생성 시 구문 오류 방지와 annotation 이 수행 되는 목적 대상에 대하여 정확한 개념을 선택할 수 있도록 하기 위해서 annotation 다이어로그 인터페이스를 개발하였다.[8]

## 3. eBook annotation 온톨로지

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

3.1 eBook annotation 온톨로지의 목적과 범위

• 목적

-eBook annotation 시스템에서 사용자와 개발자, 개발자와 개발자, 시스템과 시스템간의 상호 공통적인 이해를 표현하도록 한다.

-eBook annotation 시스템에서 데이터가 다른 언어로 표현 될 때 의미 해석의 차이가 발생하는 것에 대하여 일관성 있는 해석이 가능하도록 한다.

-각각의 언어들은 각 단어와 같은 의미를 갖는 동의어가 있다. 정보 시스템에서도 동의어에 대한 의미적 해석을 지원해야 한다.

-생성되는 annotation 데이터에 대하여 데이터간의 관계 속에서 중요도를 체크할 수 있게 한다.

-생성되는 annotation 데이터에 대하여 미리 설정해 놓은 검사 규칙으로써 데이터 무결성을 검증한다. 이것으로써 불필요한 데이터의 오류가 발생하는 것을 막을 수 있다.

• 범위

-온라인 다중 사용자 환경의 eBook annotation 시스템을 대상으로 한다.

3.2 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

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.

3.3 eBook annotation 온톨로지 평가

설계한 eBook annotation 온톨로지가 목적에 맞는지를 체크하기 위해서 다음과 같은 질문들을 사용하였다.

Q1: eBook annotation 시스템에서 사용되며 생성되는 데이터에 대한 상호 공통적인 이해의 표현을 제공하는가?

Q2: eBook annotation 데이터에 대하여 다국어 사용으로 인한 의미 해석의 차이 문제를 해결해 주는가?

Q3: 정보 시스템이 동의어를 의미적으로 해석하도록 하는가?

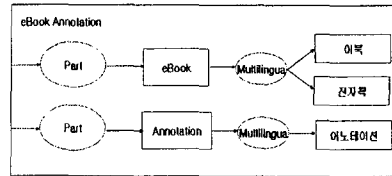
Q4: 생성되는 annotation 데이터에 대하여 중요도를 줄 수 있는가?

Q5: 데이터가 잘못 입력되는 경우에 대하여 경고하

고 사전에 막을 수 있는가?

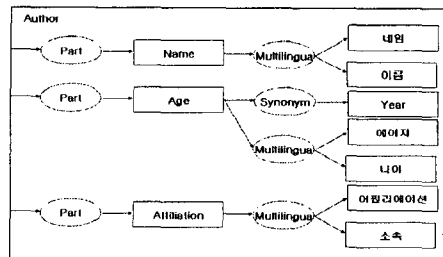
3.4 CG를 이용한 eBook annotation 온톨로지 표현

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



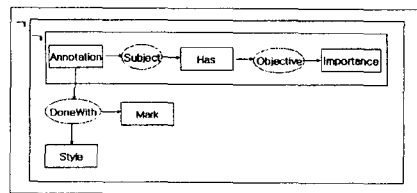
[그림 1] eBook annotation 온톨로지의 상위 구조

[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 이루어 진다.



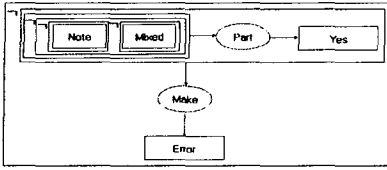
[그림 2] author의 구성 요소

[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

[그림 4]는 데이터 무결성 검사에 대한 axiom을 표현한 것이다. Annotation type이 note나 mixed로 선택되었는데 입력된 텍스트가 없을 때 에러를 발생하게 된다. 이러한 데이터 무결성 검사 axiom은 annotation 인터페이스와의 관계에서 의미 있게 사용 될 수 있다.

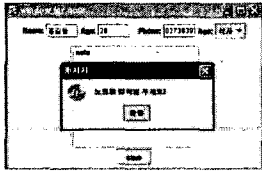


[그림 4] 데이터 무결성 검사 axiom

#### 4. eBook annotation 시스템의 구현 및 고찰

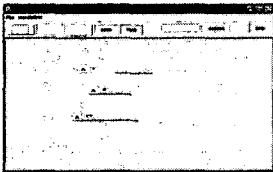
eBook annotation 시스템이 온톨로지를 사용하고 있으나 온톨로지의 구조에 대해서 사용자가 알아야 annotation을 수행하는 것이 아니기에 온톨로지를 보여 주고 있지 않다. eBook 온톨로지의 동의어, 다국어, 무결성 검사, 그리고 중요성 관계는 시스템 내부에서 구현되어 있다.

아래의 [그림 5]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 5] 무결성 검사를 구현한 화면

아래의 [그림 6]은 중요성 axiom을 구현한 화면이다. 본 시스템에서는 마크만 수행된 경우보다 마크와 노트가 같이 수행된 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 6] 중요성 axiom을 구현한 화면

#### 5. 결론 및 향후 연구방향

본 연구에서는 eBook annotation 시스템 개발 시 eBook annotation 데이터에 대한 상호 공통적인 이해의

표현, 데이터의 상호 운용, 의미 기반의 데이터 저장 및 관리, 추론을 지원하기 위해서 eBook annotation 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 annotation 입력 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 annotation 데이터의 재사용성을 높일 수 있다.

본 eBook annotation 시스템에서는 온톨로지에서 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 annotation 데이터를 어떻게 사용자 인터페이스에서 보여 줄 것인가에 대한 것도 연구 되어야 할 것이다. 데이터를 의미 기반으로 효과적으로 관리하는 것도 중요하지만 처리된 데이터를 eBook의 콘텐츠를 가리지 않으면서 사용자에게 보여 주는 것도 중요하기 때문이다. 또한 eBook annotation 온톨로지의 활용을 위해서는 이미 개발된 다른 온톨로지와의 통합에 대한 연구가 필요하다. 이는 eBook annotation 시스템이 온톨로지를 사용하는 다른 시스템이나 Semantic Web과의 상호 데이터 교환에 필요하기 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) 발전 방향”, 2000
- [2] Ilia A. Ovsiannikov, “Annotation Technology”, International Journal of Human-Computer Studies v.50 n.4, 1999
- [3] EBK(e-Book of Korea) Consortium, “A Study of Korean Standardization of eBook documents”, Technical Report, 2001
- [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] S. Handschuh, S. Staab, and A. Maedche. CREAM □ Creating relational metadata with a component based, ontology driven framework. In *Proceedings of K-Cap 2001*, Victoria, BC, Canada, October 2001.

[8] M. Erdmann, A. Maedche, H.-P. Schnurr, and Steffen Staab. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In *P. Buitelaar & K. Hasida (eds). Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, Luxembourg, August 2000.