

# 메타데이터 정의를 통한 위키 시스템의 데이터 재사용 방안

박재휘<sup>0</sup> 공기현 명재석 이동주 이상구

서울대학교

{yjhan<sup>0</sup>, rawtilt, justake, therocks, sglee}@europa.snu.ac.kr

## Metadata Formatting Scheme on Data in Wiki-System for Reusability

Jaehui Park<sup>0</sup>, Gihyun Gong, Jaeseok Myung, Dongjoo Lee, Sang-goo Lee  
Seoul National University

### 요 약

위키는 사용자 누구나 문서를 생성하고 수정이 가능하여 지식관리 도구로 활용이 되고 있다. 하지만 컴퓨터가 위키 시스템에 축적된 지식을 재사용하기 위해서는 문서의 구조를 인식하고 처리하는 과정을 거쳐야 한다. 본 연구에서는 문서의 메타데이터를 정의하여 문서를 구성하고 해당 메타데이터에 따른 입력 방식을 제안하여 데이터에 대한 유효성을 검증할 수 있게 하였다. 그리고 문서를 메타데이터에 따라 분해하여 저장하는 방식을 통하여 타 응용에서 의미 있는 데이터의 재사용이 가능하도록 하였다. 제안된 방식은 음악을 주제로 한 위키 시스템인 WikiMusic을 통해 구현하고 관련 연구와 비교하여 평가하였다.

### 1. 서 론

최근 웹 2.0이라는 새로운 웹의 경향이 제시되면서, 새로운 방식의 서비스들이 생겨나고 있다. 그 중에서 위키피디아[1]는 현재 세계에서 규모가 가장 큰 온라인 백과사전으로 알려져 있다. 위키피디아는 공개된 저작 방식을 통해서 웹 2.0의 이념인 공유와 참여를 잘 활용하고 있다고 알려져 있다. 위키피디아가 지금까지 축적해온 지식은 다양한 분야로 활용되는 것이 가능하다.[2]

이렇게 사용자들의 집단 지성을 활용한 위키피디아와 같은 위키 시스템이 각광 받고 있지만 축적된 데이터의 재사용이 어렵기 때문에[5] 그 효용성의 이면에는 새로운 문제점들이 드러나고 있다.

현재의 위키피디아는 단순히 사람들이 읽고 쓰는 백과사전에 불과하다. 다시 말하면, 컴퓨터가 자동으로 축적된 정보와 지식을 다양한 분야로 활용할 수는 없는 형태로 구성되어있다는 것이다. 그래서 데이터를 다시 활용하기 위해서는 텍스트를 추출하고 해당하는 의미 정보를 처리하는 과정을 거쳐야 한다.

또, 위키피디아와 같이 다수 사용자의 참여를 통한 데이터 수집 방식은 수집된 데이터의 유효성을 보장할 수 없다. 또한 잘못된 정보, 허용 범위 외의 정보, 반달리즘 문제는 여전히 해결해야 할 문제로 남아있는 상태이다.

본 연구에서는 위키에 저장할 데이터의 범위를 축소하여 해당 데이터가 가질 수 있는 메타데이터를 정의하고, 메타데이터에 따른 데이터를 입력 받을 수 있도록 하였다. 입력된 데이터는 관계형 데이터베이스의 구조화된 테이블에 메타데이터에 따라 분해되어 저장

된다.

이와 같은 방식으로 사용자 입력 시점에서의 데이터 유효성 검증이 가능하다. 그리고 위키 문법을 사용해서 작성하는 기존 문서 창 대신 메타데이터에 따른 사용자 인터페이스를 제공하여 사용자 편의성도 증가한다. 또한, 정의된 메타데이터에 부여된 의미정보를 파악하여 다른 응용에 적용이 가능하다.

이어지는 2장에서는 위키의 문제점을 파악하고 해결법을 제시한 관련연구 사례를 소개한다. 3장에서는 메타데이터 정의를 통한 위키 문서의 작성 방식을 제안하고, 4장에서는 제안된 문서를 재사용 하기 위한 데이터베이스 저장 방식을 제안하였다. 5장에서는 3장과 4장의 제안에 따라 구현한 사례를 서술하고, 마지막으로 6장에서 결론과 추후 연구를 서술한다.

### 2. 관련 연구

위키는 위키 위키 웹의 줄임말로써, 사용자 누구나 쉽게 문서를 생성하고 수정 가능하도록 하는 웹 또는 그런 일을 하는 소프트웨어를 말한다.[3] 이런 상호작용이 편리한 특성으로 인해 위키는 웹 상의 집단 지성을 구축하는 효율적인 도구로 활용된다. 기본적인 특성은 간단한 마크업 언어를 사용하고, 각 문서가 하이퍼링크로 연결되어 있는 것이다. 또한 다양한 구조의 문서를 생성하고 수정하기 쉽도록 데이터베이스 테이블의 한 필드에 문서단위로 저장되는 방식을 따른다. 그래서 일반 웹 문서의 정적인 구조에 반해 위키 문서는 내용과 내부의 구조가 동적으로 자유롭게 변할 수 있다.

하지만 이렇게 쉽고 편하게 사용할 수 있는 위키 문서

관리 구조는 문서를 사람이 보고 활용하기엔 효율적이지만, 컴퓨터가 활용하기에는 어려움이 따른다.

그래서 위키 문서에 컴퓨터가 인식할 수 있는 의미를 부여하고 풍부하게 하기 위한 연구사례들이 있었는데 Platypus 와 Semantic Wikipedia 를 그 예로 들 수 있다.

### 2.1 Platypus

Platypus[4] 는 위키 데이터의 공유와 재사용을 위해 시맨틱 웹 기술을 적용한 위키 문서 작성방식을 제안하였다. 모든 위키 문서를 RDF문서와 결합해서 표현하여 기존의 시맨틱 웹 기술로 데이터를 재사용이 가능하도록 했다. 하지만 일반 위키 문서 편집 창과 RDF 문서 편집 창이 독립적으로 분할되어 존재하기 때문에 동일한 내용인지 일관성을 수동으로 확인해야 한다. 또, 사용자가 RDF 문법을 파악해야 하기 때문에 일반 위키 사용자가 RDF문서를 저작하기에는 어려움이 따른다. 그리고 기존에 축적되어 왔던 위키 데이터들과는 호환이 되지 않아서 문서들을 다시 작성해야 하는 문제점이 있다.

### 2.2 Semantic Wikipedia

Semantic Wikipedia[5] 는 위키피디아의 축적된 지식을 활용하기 위해 미디어위키 익스텐션을 개발하였다. 기존의 링크 방식과 문서 구조에 typed link와 attribute 를 도입함으로써 문서간 링크에 의미를 부여하고 문서내부의 데이터 값을 활용할 수 있도록 하였다. 하지만 작성된 문서를 RDF export 기능을 사용하여 다시 RDF 문서로 저장하는 과정을 거쳐야 활용할 수 있는 문제점이 있다.

본 논문에서는 실제 위키피디아가 사용되는 방식에 기반하여 관련 연구들에서 찾은 문제점들을 해결 할 수 있는 방법을 제안한다.

## 3. 메타데이터

### 3.1 메타데이터란

메타데이터란 데이터에 관한 데이터를 말하는 것으로, 데이터의 사용과 관리, 이해에 활용되는 데이터를 뜻한다. 예를 들어 도서관에 있는 책에 관한 메타데이터는 책 제목, 저자, 출판 연도, 책의 위치 등이 될 수 있고 해당하는 메타데이터 값들은 각각 'A discipline of programming', 'Edsger W. Dijkstra', '1976', '005.1 D569d' 가 된다.

본 연구에서는 하나의 위키 문서를 표현할 수 있는 관련 데이터를 메타데이터라고 하였다. 메타데이터는 위키 문서의 소 단락 제목, 문서 분류 등이 될 수도 있다.

### 3.2 메타데이터의 활용

최근 위키피디아는 사용자 규모가 증가하고 정보의 양이 급격히 늘어나게 되면서 일부 문서에 대해 정형화된 메타데이터를 마련하였다. 예를 들어, 'John Lennon' 이라는 제목의 문서에는 출생일, 출신지,

사망일, 직업 등 사람에 관련된 일반적인 정보를 정해진 테이블로 명시하였다. 이것은 이 문서에서 서술하는 정보의 주제가 '사람'이기 때문이다.

이와 같이 위키피디아는 대부분의 문서에 대한 자유로운 편집을 허용하고 있는 반면 일부 문서에 관해서는 정해진 메타데이터가 존재하고, 사용자가 메타데이터에 따라 입력을 할 수 있도록 유도하고 있다.

어떤 메타데이터는 정해진 주제에 대해 필수적이며 또한, 메타데이터에 대응하는 값의 타입이나 범위가 정형화 되어 있다. 예를 들어 'John Lennon' 의 출생일은 '9 October 1940'이다. 인물을 수록한 백과사전인 경우 출생일은 인물을 표현하는데 필수적인 정보이며 해당하는 값의 타입과 범위는 유일한 날짜라는 것을 알 수 있다.

October 1940 – 8 December 1980), was a 20th-century English songwriter, singer and instrumentalist & founder of the Beatles. Lennon and Paul McCartney formed a critically acclaimed and top range for the Beatles and other artists. Lennon, with his cynical edge and knack for toylike optimism and gift for melody, complemented one another uniquely.<sup>[c]</sup> In his solo career, as "Imagine" and "One Peace a Chance".

<sup>[d]</sup> Lennon sat on television, in films such as *A Hard Day's Night* (1964), in books such as *In His Own Words*. He channeled his fame and penchant for controversy into his work as a peace activist.

<sup>[e]</sup> Cynthia and Sean, with his second wife, avant-garde artist Yoko Ono. Lennon was murdered in 1980 and Ono returned home from a recording session.

<sup>[f]</sup> 100 Greatest Britons voted Lennon into eighth place. In 2004, *Rolling Stone* ranked Lennon as the 50th Greatest Artist of All Time<sup>[g]</sup> and ranked the Beatles at number 1.

Background information	
<b>Birth name</b>	John Winston Lennon
<b>Born</b>	9 October 1940
<b>Origin</b>	Liverpool, England
<b>Died</b>	8 December 1980 (age 40) New York City, New York, USA
<b>Genre(s)</b>	Pop-rock Sitar-rock Rock and roll Psychedelic rock Neo-progressive rock
<b>Occupation(s)</b>	Singer-songwriter, guitarist, poet, artist, activist
<b>Instrument(s)</b>	Guitar, Harmonica, Piano, Bass, Melodica, Sitar
<b>Years active</b>	1957 – 1979, 1980
<b>Label(s)</b>	Parlophone, Capitol, Apple, Vee-Jay, EMI, Geffen
<b>Associated acts</b>	The Beatles Plastic Ono Band The City Mac
<b>Website</b>	JohnLennon.com <sup>[h]</sup>

그림 1 위키피디아의 'John Lennon' 문서의 Born 메타 데이터와 그 값

위키피디아의 정형화된 메타데이터와 같이 특정 주제의 문서에 대해 주제를 표현하는 메타데이터를 정의하는 일은 위키 프로젝트 [6]에서 해당 분야의 전문가 그룹을 통해 이루어지고 있다.

### 3.3 메타데이터 정의의 효과

문서에 대한 메타데이터를 정의함에 따라, 메타데이터 타입을 확인 하고 값도 일정 범위와 조건 내에서 다루는 것이 가능해진다. 기존의 위키 문서 편집창 대신 메타데이터의 범위와 조건에 따라 값을 입력 하는 사용자 인터페이스를 마련하면, 메타데이터 값에 대한 데이터 유효성 검증이 가능해져서, 데이터 정확도가 높아진다. 예를 들어 '출생일' 이라는 메타데이터의 값은 날짜와 연도로 표시된다는 것을 미리 예상할 수 있으므로 해당 데이터에 대한 유효성 검증을 사용자 인터페이스를 통해 프로그램에서 해 줄 수 있다.

또, 문서의 주제를 표현 할 기본적인 요소를 메타데이터로 정의해 놓음으로써 문서 편집 시 사용자 편의성 증가와 더불어 문서 주제의 모호함도 감소한다.

## 4. 문서 저장

### 4.1 데이터 무결성

위키피디아와 같은 집단 지성 소프트웨어는 사용자가

많을수록 데이터의 정확도가 높아진다고 하는 가정을 가지고 있다. 하지만 데이터가 정확도와는 별도로 데이터 유효성이 보장되어야 데이터베이스 내의 무결성을 보장할 수 있다.

3장에서 제안한 메타데이터의 정의를 통해 메타데이터 값의 유효성 검증이 가능해지면 문서 저장 시 메타데이터 값의 무결성은 보장된다. 따라서 문서를 저장할 때 관계형 데이터베이스의 구조화된 테이블에 메타데이터를 저장이 가능해진다.

정해진 메타데이터에 따라 문서를 분해하여 저장이 가능하게 되면, 문서의 주제에 대한 전문가가 메타데이터를 미리 정의하고 해당 의미를 부여한 데이터베이스 스키마를 디자인 할 수 있다. 이렇게 데이터베이스에서 위키 문서의 메타데이터를 따로 관리하면 문서 관리 시 효율성이 증가한다. 이런 방식은 기존 위키피디아의 문서 저장 방식으로는 할 수 없었던 메타데이터의 활용이 이루어진다.

4.2 데이터 모델

기존 위키 문서의 데이터 모델은 다음과 같다.

$$\text{System} = \{\text{Article}_1, \text{Article}_2, \dots, \text{Article}_n\}$$

$$\text{Article}_i = \{(v,l) \mid v \in \text{Value}, l \in \text{Link}(\text{Article}_j) \text{ or } l \text{ is null}\}, i \neq j$$

System 은 위키 시스템을 의미하고 문서를 의미하는 Article의 집합이다. 그리고 Article은 value와 link의 쌍으로 구성된다. Value는 텍스트로 표현된 값이며 Link는 다른 Article과의 연결을 의미한다. 이와 같이 하나의 Article에는 다른 Article과의 링크를 여러 개 포함할 수 있어서 System은 문서들이 노드가 되고 링크가 연결선이 된 그래프 구조로 표현된다.

각 문서의 내용은 제한 없이 자유로운 값으로 구성되기 때문에 기존의 위키 시스템에서는 문서간의 링크 외에는 컴퓨터가 수집하여 바로 이해할 수 있는 정보가 없다.

하지만 본 연구에서 제시하는 데이터 모델은 문서에 포함되는 모든 값이 특정 메타데이터에 종속됨으로써 메타데이터에 관련한 의미 정보를 포함하게 된다. 이런 의미 정보는 메타데이터를 정의할 때 부여되고, 그에 대한 도메인 제약 조건(D<sub>i</sub>)또한 적용이 가능하다. 그리고 문서는 네임스페이스(ns)[7]로 분류되어 각 문서의 메타데이터 집합을 다르게 가질 수 있도록 한다.

$$\text{System} = \{ns_i:\text{Article}_j \mid i, j > 0\}$$

$$ns_i:\text{Article}_j = \{T_1, T_2, \dots, T_n\}$$

$$T_i = \{(v,l) \mid v \in \text{Value}, v \in D_i, l \in \text{Link}(\text{Article}_k) \text{ or } l \text{ is null}\}, i, j > 0, j \neq k$$

$$D_i = \{C_1, C_2, \dots, C_n\}$$

그리고 제안한 모델은 기존의 위키 시스템이 가진 문서의 그래프 구조를 그대로 사용하면서, 메타데이터와 네임스페이스의 조합으로 더욱 풍부한 의미정보를 가진다. 예를 들어 'John Lennon'이라는 사람의

메타데이터인 '발매 음반'에 대해서 'Stand by me'라는 문서가 링크 되었을 때, 네임스페이스가 Artist인 'Artist:John Lennon' 문서에서 네임스페이스가 Album인 'Album:Stand by me'의 문서로 '발매 음반'이라는 메타데이터 값으로써 링크를 형성한다. 그래서 문서 내에 포함되는 의미정보는 "아티스트 John Lennon 이 발매한 음반 Stand by me"가 된다. 이렇게 링크 자체도 풍부한 의미 정보를 지니게 되면서 기존 위키 문서보다 풍부한 표현력을 가지게 된다.

5. 구현

본 연구는 All Music Guide (AMG)[7]의 음악 메타데이터 정의를 기반으로 음악 주제 위키 시스템인 WikiMusic을 구현하였다.

5.1 구현 환경

위키문서를 관리할 기본 엔진은 위키피디아 엔진인 미디어 위키[8]를 사용하였고 메타데이터 값은 Oracle에 저장한다. 메타데이터와 그 값은 All Music Guide에 정의된 내용을 사용하였다. 그리고 에디터 익스텐션은 자바 스크립트와 php를 이용하여 구현하였다.

5.2 구현 내용

위키에 저장할 문서는 음악, 앨범, 아티스트로 네임스페이스를 제한하였고, 각 네임스페이스 별 메타데이터는 제시한 데이터 모델에 따라 정의하였다. 그림 2는 Music 네임스페이스에 대한 메타데이터 정의 도식이다.

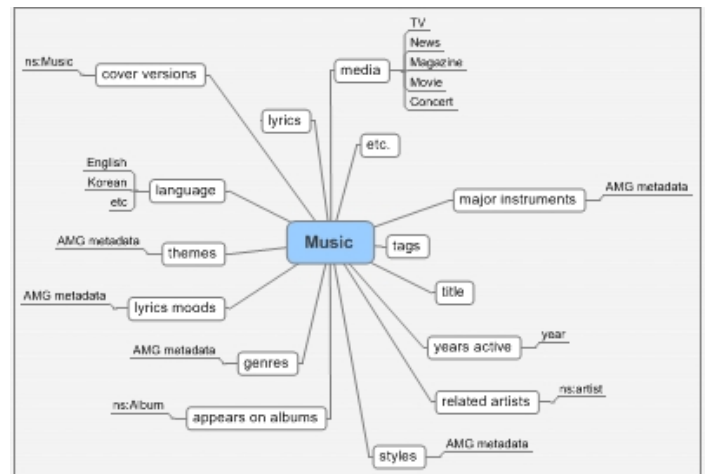


그림 2 네임스페이스가 Music 인 문서의 메타데이터 정의

각 문서의 에디터는 사용자가 위키 문법을 알지 못해도 사용할 수 있도록 기본적인 HTML 컨트롤을 통해 입력을 하도록 하였다. 이 방식으로 정해진 메타데이터에 검증된 값들이 입력될 수 있다.

그리고 위키 문법의 그래프 구조를 활용하는 링크는 네임스페이스를 사용자가 쉽게 선택할 수 있게 버튼으로 표현하였다. 그리고 편집된 문서가 보여지는 화면은 네임스페이스에 종속된 템플릿[8]으로 구성하였다.

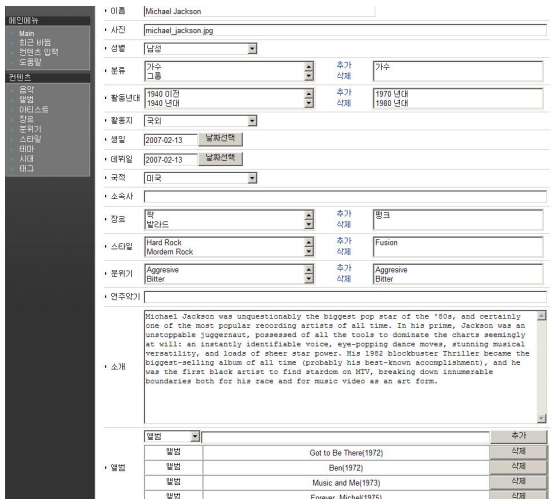


그림 3 정의된 메타데이터에 따른 사용자 입력 에디터  
마지막으로, 사용자 편의를 위해 왼쪽 사이드 바에 네임스페이스 별, 주요 메타데이터 별 네비게이션 문서를 링크해 두고 해당 문서에 쉽게 접근할 수 있도록 하였다.

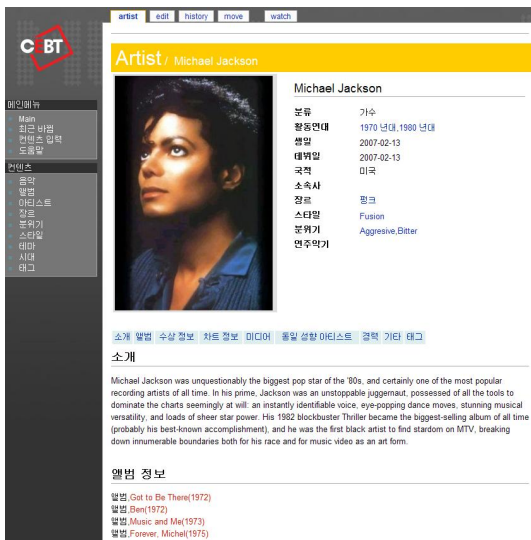


그림 4 네임스페이스가 아티스트인 문서의 예  
5.3 평가  
관련 연구와 비교하여 본 연구에서 구현한 시스템은 다음과 같은 기준에 따라 평가하였다.

사용자 편의성: WikiMusic은 위키 문법이나 RDF 문법을 완전히 배제하고 기존 위키 시스템보다 구체적인 의미를 지니는 문서를 사용자가 쉽게 작성할 수 있게 하였다.

표현력: 관련연구의 위키 시스템은 시맨틱 언어의 Schema를 구성하는 과정이 존재하고, WikiMusic 에서 메타데이터를 구성하는 과정과 맞물린다. 현재 구현된 시스템은 문서의 정해진 주제를 표현하는 것에 구체화되어 있으므로 일반적인 문서 관계를 표현하는 데는 한계가 있다. 하지만 정해진 메타데이터로 구성된 문서에 대해서는 동일한 표현력을 가진다.

유연성: WikiMusic은 메타데이터를 정의해 문서의

구조를 고정시켜 놓았으므로 기존의 위키 시스템이나 관련 연구의 시스템보다는 유연성은 떨어진다. 하지만 데이터의 유효성을 검증하고 컴퓨터가 이해할 수 있는 문서를 구성하기 위해서는 유연성은 일부 제한되게 된다.

호환성: WikiMusic은 미디어 위키 엔진의 기능을 전부 수용하고 미디어 위키 익스텐션[8]으로 개발 되었으므로 기존 위키피디아와의 호환이 가능하다.

WikiMusic은 기존 위키 시스템보다 유연성을 제한하면서 표현력을 증가시켜서 컴퓨터에 대한 문서의 이해도를 높였고, 시맨틱 웹 기술이 적용된 위키 시스템보다 일반적인 문서에 대한 표현력은 낮지만 사용자 편의성을 높여서 집단 지성을 활용할 수 있는 기회를 증가시켰다.

### 6. 결론

본 연구에서는 위키 문서에 대해 메타데이터를 정의한 저작 방법을 제안함으로써 위키 시스템으로부터 유효성이 검증된 데이터를 얻을 수 있게 하였다. 그리고 문서 내용을 메타데이터에 따라 분해한 값을 데이터베이스에 저장하는 방식으로 다른 응용 프로그램 등에서 활용이 가능하도록 하였다.

WikiMusic은 현재 진행 중에 있는 음악 컨텐츠 프로젝트의 일부로써 지능형 음악 추천 엔진의 기반 데이터 수집에 활용되고 있다.

추후에는 사용자 참여를 고려한 시맨틱 웹 언어 적용 방식 등의 연구 과제가 남아있다.

### 7. 참고문헌

[1] Wikipedia, [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)  
 [2] J. Wales. Wikipedia and the free culture revolution, OOPSLA/ WikiSym Invited Talk, 2005.  
 [3] Bo Leuf, and Ward Cunningham, The Wiki Way: Collaboration and Sharing on the Internet, Addison-Wesley, 1st edition, 2001.  
 [4] S. E. Campanini, P. Castagna, and R. Tazzoli. Platypus wiki: A semantic wiki wiki web. In Semantic Web Applications and Perspectives, Proceedings of 1st Italian Semantic Web Workshop, 2004.  
 [5] Max Völkel, Markus Krötzsch, Denny Vrandečić, Heiko Haller, Rudi Studer, Semantic Wikipedia, In Proceedings of the 15th International Conference on World Wide Web, 585-594, 2006  
 [6] WikiProject, <http://en.wikipedia.org/wiki/Wikipedia:WikiProject>  
 [7] All Music Guide, <http://www.allmusic.com>  
 [8] MediaWiki, <http://www.mediawiki.org>