효율적인 기술 문서화를 위한 콘텐트 재사용성 접근방법

(Approaching Content Reuse for Efficient Technical Documentation)

구 흥 서*
(Koo, Heung Seo)

요 약 기업에서 수백 또는 수천 개의 문서를 작업해야 하는 다수의 프로젝트를 수행할 때, 여러 개의 문서에 포함된 동일한 콘텐트나 본질적으로 유사한 콘텐트를 변경하지 않아도 되기 때문에 콘텐트의 싱글소싱은 커다란 이득이 있다. OASIS의 DITA (Darwin Information Typing Architecture)는 기술문서들을 저작하고, 생성하고, 배포하기 위한 XML기반의 아키텍처이다. 이 기술은 정보타입에 따라 분류한 토픽 모듈들을 생성한 다음, 다양한 방식으로 이런 콘텐트를 이용하는 설계 규칙들의 집합이다. 본 논문에서는 서로 다른 정보제품을 제작할 때 기존의 콘텐트 콤포넌트의 재사용성을 향상시키기 위해, DITA기반의 기술문서 개발을 위한 접근방법을 연구하였다.

핵심주제어 : 콘텐트 재사용, 싱글소싱, DITA, 정보제품, 기술문서

Abstract The single-sourcing of content is extremely beneficial because when we are managing several projects with hundreds or thousands of documentation, we don't want to be changing the same content, or substantially similar content in multiple locations. The Darwin Information Typing Architecture (DITA) is an XML-based architecture for authoring, producing, and delivering technical documents. It consists of a set of design principles for creating information-typed topic modules and for using that content in various ways. In this paper, we examine the approach of using The Darwin Information Typing Architecture for technical documents development to enhance the reuse of existing content components for difference information products.

Key Words : Content reuse, Single-sourcing, DITA, Information product, Technical Document

1. 서 론

기업에서는 제품 개발의 일부 작업으로 다양한 기술문서들을 제작하고 있으며, 이러한 문서는 제품 사용설명서, 홍보자료, 제품사용 설명서, 기술 매뉴얼, 유지보수 설명서 등 다양한 형태의 콘텐트로 존재한다. 과거에는 기업들이 기술문서를 제품을 판매하기 위해 형식적으로 제공하는 액세서리로 인식하기도 했지만, 최근 일부 선진기업들이 차별화된 경쟁 전략의 하나로 사용자편의성 제공 및 원가절감 차원에서 인식하기 시작하였다. 또한 제조물책임법(PL법, Product Liability)의 발효로 기술문서의 중요성이 더욱 높아지고 있는 상황이다.

* 청주대학교 컴퓨터정보공학과

기술문서의 목적은 제품에 관한 정확하고 관련 있는 정보를 고객들에게 제공하여 고객이 제품 사용 할 때 편의성을 향상시키고 제품을 정확하게 작동시킬 수 있도록 지원하는 것이다. 기업에서 생산하는 제품의 기능 또는 구조가 복잡해지고 유사 제품의 종류가 다양해지면서 이러한 기술문서를 제작하는 비용도 급격하게 증가하여 제품 원가에 압박을 가하고, 기술문서의 내용오류도 증가하고 있어 선진기업들은 이에 대한 해결책을 모색하고 있다.

이러한 해결책으로 보다 정확하고 관련성이 높은 정보를 신속하게 사용자에게 제공하면서 기술문서 제작, 출판(publishing), 유지관리 작업의 효율을 향상시키기 위해 지난 15년간 여러 기술을 적용해 왔다. 이러한 기술로는 nroff/troff, TeX/TeXinfo, SGML, HTML, XML, DocBook, DITA가 있다[1]. 이 중에서 가장 최근에 개발된 기술로 2005년에 OASIS (Organization for the Advancement of Structured Information Standards)에서 정식 규격[2]으로 승인한 DITA 기술이 크게 활용될 것으로 예측된다.

DITA의 활용에 대한 산업계의 관심이 국내에서도 높아지고 있으나 DITA의 효과적인 활용방법, DITA를 활용한 기술문서 저작 프로세스에 대한 연구가 미흡한 실정이다. 그러므로 DITA의 효과적인 활용방법 및 저작 프로세스를 정립하여 기업에서 활용하면 기술문서 제작시 발생하는 제작기간 초과, 비용 상승, 내용오류 발생 등 문제점들을 적절하게 해결할 수 있을 것이다.

본 논문은 이러한 문제를 해결하기 위해 연구로서 DITA 기술을 활용하여 기술문서를 효율적으로 제작할 수 있도록 콘텐트 재사용성을 향상시킬 수 있는 접근방법을 연구하였다. 이 접근방법으로 네 가지 설계요소를 제시하였으며, 이 설계요소들은 DITA 이전부터 많이 참조되었던 콘텐트 정보모델링 방법론과 DITA 기술의 특성을 반영한 것이다.

본 논문의 구성은 먼저 2장에서 본 논문의 주제와 관련된 배경 지식으로 기술문서 출판 시스템의 요구사항, 콘텐트 재사용 기법, 그래뉼러리티(granularity)에 대해 살펴본다. 3장에서는 DITA의 기능 중에서 콘텐트 재사용성과 관련된 주요 기능을 조사 분석한다. 그리고 4장에서는 콘텐트 재사용성 향상을 위한 DITA의 주요 설계 요소를 제시하고, 5장에서 결론을 맺는다.

2. 배경 지식

2.1 기술문서 출판 시스템의 요구사항

기술문서 출판 시스템을 효율적으로 구축하기 위해서는 다음 네 가지 요소를 충족시켜야 한다[3].

(1) 높은 정보 공유

하나의 정보가 조직 내부·외부의 여러 그룹의 프로세스에 자동으로 통합되어 공유될 수 있어야 생성된 정보를 더욱 유용하게 활용할 수 있게 된다.

(2) 높은 재사용성

정보의 재사용 및 속성 변경이 가능해야 동일한 정보를 중복 제작 하거나 복사본을 유지하지 않아도 되며, 정보 불일치성(information inconsistency)을 피할 수 있다. 제작자 입장에서는 동일한 정보소스를 하나만 저작하여 여러 문서에 적용하고, 원본 정보소스가 변경되면 이 정보가 포함된 모든 문서에 자동적으로 반영될 수 있어야 기술문서 제작 및 갱신 비용을 낮출 수 있다.

(3) 정확한 정보 연관성 표현

정보를 작은 단위의 모듈형식으로 저작하여 관리를 할 수 있어야 문서 제작자의 필요나 제품 개발자의 요구사항에 따라 쉽게 통합하거나 재구성할 수 있다.

(4) 콘텐트의 자동화 처리

기술문서 제작 측면에서 중요한 항목이 제작시간 단축과 비용 절감이다. Aberdeen그룹의 보고서[4]에 따르면 기술문서 출판기간의 충족율이 업계평균 87% (일류기업 100%, 후진기업 68%)이므로 개선이 필요한 부분이다. 그리고 이 항목은 시간단축과 비용절감이 계속 요구받기 때문에 제작자 관점에서는 문서제작 자동화의 필요성이 높아지고 있다. 최근 문서제작과 출판 시스템 구현에 있어 XML 기술이 주목 받고 있

는 것은 자동화된 프로세스를 적용할 수 있기 때문이다.

2.2 콘텐트 재사용

콘텐트 재사용은 기술 문서화 분야에서 최근 각광을 받고 있는 콘텐트의 싱글-소싱에 있어 중요한 요소이다. 콘텐트 재사용 기법은 수백 또는 수천 페이지에 이르는 기술문서들을 제작하는 여러 프로젝트를 관리하는 경우, 문서의 여러 위치에 내장된 동일한 콘텐트를 개별적으로 변경하지 않고 한 번에 변경하거나 반영할 수 있어서 비용적인 면이나 시간적인 면에서 큰 장점을 가진다.

콘텐트 재사용 방법을 적용하기 위해서 일반적으로 정보모델링 설계단계에서 다음 두 가지 재사용 기법[5] 중 하나를 사용한다.

(1) 편의적인 재사용 기법

전통적인 효율적 재사용 기법은 어느 콘텐트 엘리먼트가 재사용 가능하고, 그 콘텐트가 어디에 속하고, 어떤 문서에 그것을 삽입해야 하는지에 관한 저작자의 지식에 전적으로 의존한다. 이 접근방법을 편의적인 재사용 기법(opportunistic reuse)이라 한다. 편의적인 재사용 기법은 저작자에게 어느 콘텐트가 재사용 가능한지 결정과 콘텐트를 재사용하는 선택을 저작자에게 부여하기 때문에 저작자들에게 높은 유연성을 제공한다는 장점을 가진다. 그러나 편의적인 재사용 기법은 잠재적으로 재사용 가능한 콘텐트의 식별, 콘텐트의 재사용 결정, 재사용하기 원하는 콘텐트의 포함 위치를 일일이 탐색해야 하는 부담을 저작자가 모든 져야 하기 때문에 재사용 빈도수가 낮아진다는 단점이 있다. 만일 콘텐트 재사용에 대한 동기가 부족하거나 적합하게 재사용 가능한 엘리먼트에 대한 인식이 부족하거나 저작자가 적당한 엘리먼트를 탐색하기 어려운 경우, 콘텐트 재사용 빈도수는 더욱 낮아질 것이다. 또한 저작자들이 콘텐트를 부적절하게 재사용할 수도 있으며 이러한 경우를 방지할 수 있는 안전장치가 존재하지 않는다.

(2) 체계적인 재사용 기법

체계적인 재사용 기법(systematic reuse)은 계획된 재사용 방법을 사용한다. 특정 콘텐트가 특정 위치에 재사용 가능한 것으로 결정되면, 그 재사용 가능한 콘텐트는 해당 위치에 자동적으로 삽입된다. 체계적인 재사용 기법은 콘텐트의 재사용을 지원하므로, 재사용 가능한 콘텐트가 존재하는지를 파악하고, 재사용 가능한 콘텐트를 탐색하여 적합한 위치에 삽입해야 하는 저작자의 부담을 감소시킨다.

체계적인 재사용 기법을 사용하려면 이 기법을 지원하는 콘텐트관리시스템(CMS)이 필요하다. CMS는 어느 콘텐트가 재사용될 수 있는지를 식별하기 위해 사용자의 정보모델을 이용한다. 콘텐트 엘리먼트가 존재하고 저작자의 특정 요구사항에 해당되면 그 재사용 가능한 엘리먼트는 문서의 적당한 지점에 자동적으로 삽입된다.

체계적인 재사용 기법은 재사용 계획 수립, 상세한 모델과 재사용맵 작성, 그리고 동적 콘텐트 엔진같은 적합한 기술이 필요하기 때문에 구현하는데 많은 비용이 소요된다. 그러나 이 기법은 최고의 투자효과(ROI: Return Of Investment)를 생성할 수 있다. 투자효과는 보장된 품질의 재사용을 통해 실현되면 이것은 저작자의 동기부재나 재사용 가능한 콘텐트에 대한 지식부재와 관계없이 자동적으로 이루어진다.

2.3 그래뉼러리티

싱글소스 정보는 포함되는 모든 콘텐트를 식별하여 파일이나 문서 단위가 아닌 엘리먼트 단위로 구성한다. 이 작업에서 중요한 설계 요소가 정보를 어느 수준 또는 크기까지 분할할지를 결정하는 것이며, 이러한 정보조각을 그래뉼러리티(granularity)라고 한다. 그러므로 그래뉼러리티는 콘텐트로서 의미를 가지는 가장 작은 정보조각이 된다.

그래뉼러리티는 크기가 커지면 관리하기 쉽지만 콘텐트 재사용성이 낮아질 가능성이 높아진다는 특성을 가진다. 그러므로 모델을 얼마나 작게 분할할 것인지 여부가 그래뉼러리티의 논쟁점(issue)이다. 그래뉼러리티는 정보의 의미를 유지하면서 크기를 가장 작게 설계하는 것이 중요하다.

3. DITA의 주요 기능

DITA는 IBM사가 자사의 솔루션인 Eclipse 개발환경에 관한 도움말 문서를 제작하기 위해 2001년에 초기 버전을 개발하였다. 그리고 IBM의 소스코드 기증으로 2005년에 OASIS(Organization for the Advancement of Structured Information Standards) 정식 규격으로 승인되었다.

DITA는 제품에 대한 도움말, 사용자 매뉴얼, 기술문서 등을 다양한 디스플레이와 배포방법에서 재사용이 쉬운 모듈방식 기술문서를 저작·생산·배포하기 위한 XML기반 정보구조이다[1,6].

3.1 DITA의 저작 원리

DITA 기술이 지향하는 핵심적인 저작 원리는 다음과 같다[6].

(1) 토픽지향 내용 작성

DITA는 재사용 가능한 토픽을 지원하여, 저작자가 내용상 독립적으로 유지 가능한 최소 단위로 저작할 수 있다. 이 기능으로 저작자는 결과물이 최종적으로 포함될 문서를 고려하지 않고 특정 주제를 완전히 커버하도록 토픽 내용 작성에만 집중할 수 있다. 그러므로 작성된 토픽의 공유율 향상을 기대할 수 있다.

(2) 정보분류

DITA는 정보분류 기능을 지원하는데, 이것은 콘텐트가 배포되는 방법과는 독립적이며 콘텐트에 대해 묘사하는 방법과 관련 있다. 이 기능으로 저작자가 기술문서에서 장 또는 부록을 작성하는 것 대신에 작성해야 하는 정보의 구조(structure)와 의미(semantics)를 사용하여 개념(concept), 작업(task), 참조 토픽(reference topic) 작성에 집중할 수 있다.

(3) 세분화

DITA는 이 원리를 통해 저작자가 더 세분화된 정보 타입을 생성할 수 있도록 지원한다. 그래서 이미 생성된 정보의 구조와 의미를 활용하여 특정 독자들에게 적합한 새로운 정보 타입을 생성할 수 있다. 그러므로 생성된 토픽·도메인의 재사용율 향상을 기대할 수 있고, 문서제작 프로세스의 상호연동을 통한 자동화 구축 가능성이 높아진다.

(4) 상속 가능 프로세스

출판·번역 같은 DITA 인식 프로세스가 더 세분화된 타입들을 자동적으로 처리하여 결과물을 특정 독자에 요구에 적합하게 생성할 수 있다.

3.2 DITA의 특성 분석

이번 절에서는 DITA 기술에서 적극적으로 활용해야 하는 기능들을 설명한다.

(1) 모듈 방식으로 콘텐트 저작

기술문서를 제작할 때 독립적인 단위로 존재할 수 있는 모듈 방식으로 정보를 생성할 수 있도록 토픽 DTD(Document Type Definition)의 정의를 지원한다. 토픽이란 정보를 구성하는 하나의 구성요소로 완전한 문서가 아니며, 소주제의 한 가지 측면을 표현한 작은 단위의 콘텐트이다. 모듈 방식으로 작성된 이러한 토픽들을 계층적으로 조립하여 기술문서 하나를 제작할 수 있다. 정보를 모듈 방식으로 저작한 정보 결과물이 얼마나 잘 설계가 되어 있는지가 정보 재사용율에 영향을 끼치며, 결과적으로 기술문서 출판 시스템의 유지보수 비용을 낮춰줄 수 있다.

토픽들을 계층적으로 구성하는 요소를 맵(map)이라고 하며, 이 메커니즘을 사용하여 기술문서를 계층구조로 정의할 수 있다. 예를 들어 기술문서 맵은 장, 절, 소절, 부록으로 이루어진 계층구조를 가질 수 있다.

(2) 정보 세분화 기능으로 유연성 구현

토픽 DTD은 제목, 단락, 목록 등 서로 다른 서식을 지정하여 고유한 요소집합을 지정할 수 있다. 그리고 정보 세분화 메커니즘을 사용하여 이미 정의된 토픽 DTD를 제작자의 필요나 사용자의 요구에 따라 적절하게 속성을 재정의 할 수 있다. 정보 세분화 구조를 얼마나 잘 설계하고 유지·관리하는지가 문서제작 프로세스의 자동화 처리에 영향을 준다.

세분화 메커니즘은 상속받은 토픽 DTD에 대해 새로운 태그의 추가 또는 기존 태그의 속성 변경을 통해 이미 저작된 정보의 재사용율을 향상시킬 수 있다.

4. 콘텐트 재사용을 위한 설계 요소

이번 장에서는 DITA 기반으로 기술문서 저작시스템을 구축할 때, 문서화 효율을 향상시키기 위한 주요 방안 중 하나인 콘텐트 재사용을 위한 접근방안을 제안한다. 콘텐트 재사용이 DITA 기반 저작시스템 활용의 효율성에 가장 큰 영향을 미치는 요소 중 하나이므로, 적합한 콘텐트 재사용 전략 수립이 효율성 높은 기술문서 저작시스템의 구축에 큰 영향을 미친다.

(1) 체계적인 재사용 기법 사용

DITA 기반 기술문서 저작시스템을 구축하려면, 기술의 복잡도와 관련 소프트웨어 도입 등으로 초기 비용과 유지보수 비용이 크게 증가하므로, 시스템의 활용성을 극대화할 수 있는 체계적인 재사용 기법(systematic reuse)을 적극적으로 적용해야 한다.

(2) 정보모델링 프로세스에 기초한 싱글소스에 대한 세밀한 분석 작업 수행

정보모델은 콘텐트관리시스템(CMS: Content Management System)과 동적콘텐트관리(dynamic content management) 시스템에서 싱글소싱을 유지하기 하는데 중요한 요소이다. 정보모델은 재사용 가능한 콘텐트를 효과적으로 작성·구조화·배포하기 위한 청사진 역할을 한다. 따라서 콘텐트 제작 프로젝트를 수행할 때 프로젝트의 정보모델을 설계하는 작업은 싱글소싱을 위해 가장 중요한 과정 중 하나이다[5].

정보모델링 프로세스는 콘텐트 분석단계(개괄 분석 단계, 상세 분석 단계)와 모델링 단계(정보제품 모델, 엘리먼트 모델, 시멘틱 모델)로 크게 나눌 수 있다. 콘텐트 분석단계는 다시 개관분석 단계(top-level analysis)와 상세분석 단계(in-depth analysis)로 나눌 수 있으며, 개관분석 단계에서는 싱글소스에 포함된 모든 콘텐트에 대한 분류 목차를 작성하여, 재사용 가능한 콘텐트를 파악한다. 그런 다음, 상세분석 단계에서 유사한 콘텐트들에 대한 표준화 여부를 결정한다. 이 표준화 작업을 통해 콘텐트 관리의 용이함을 높이고 재사용성을 향상시킬 수 있다.

(3) 싱글소스에 대해 독립된 주제별로 DITA의 토픽 맵 작성

싱글소스 분석 산출물을 기반으로 DITA의 특정 주제 중심으로 콘텐트 맵을 작성한다. 이 맵을 토픽 맵이라고 하는데, 토픽 맵은 내용적으로 독립적으로 유지될 수 있도록 하면서 그래뉼러리티를 가능한 작게 작성하는 것이 중요하다. 토픽 맵의 장점을 충분히 활용하려면 토픽들을 재사용 가능한 단위로 구성해야 하며, 그래뉼러리티의 수준이 고객 요구, 저작자의 요구사항, 정보집합(information set)을 유지관리 하는데 요구되는 유지관리 수준에 적합하도록 정보모델을 반영해야 한다.

(4) CMS를 이용하여 싱글소스의 콘텐트 관리

CMS를 이용하여 싱글소스의 콘텐트를 관리한다. 온라인을 기반으로 한 동적문서 출판 시스템은 콘텐트를 동적으로 출판할 수 있어야 한다. 따라서 콘텐트가 동적으로 변경되면, 관련된 저작자들에게 콘텐트 변경을 통보하여 저작자들이 쉽게 인지할 수 있도록 지원해야 한다. 일반적으로 CMS는 특정 콘텐트가 변경되면 저작자들에게 통보할 수 있는 워크플로우 기능 지원한다.

5. 결 론

최근 제품이 복잡해지고 제조물책임법(PL법)이 발효되면서, 선진기업들이 정확한 제품 문서화가 고객 만족과 밀접한 관련이 있는 중요 작업으로 인식하고 많은 투자를 하고 있다. 국내의 많은 기업도 제품 사용설명서가 더 이상 제품 개발과정의 부산물이 아니라는 인식이 확산되고 있지만, 제품 문서화를 위한 자동화 프로세스의 개발 필요성에 대한 인식이 낮고 중요한 자동화 기술 중 하나인 DITA를 적용하기 위한 접근방안을 찾지 못하고 있는 실정이다.

본 논문은 기업에서 제품 문서화 작업의 효율성을

향상시키기 위해 OASIS의 DITA 기술을 활용하려는 경우에 중점 고려사항 중 하나인 콘텐트 재사용성 향상을 위한 접근방법을 연구하였다.

4장에서 제시한 네 가지 설계요소인 체계적인 재사용 기법 사용, 정보모델링 프로세스에 기초한 싱글소스에 대한 세밀한 분석 작업 수행, 싱글소스에 대해 독립된 주제별로 DITA의 토픽 맵 작성, CMS를 이용하여 싱글소스의 콘텐트 관리는 콘텐트 재사용성을 향상시켜 기술문서 제작 효율성에 영향을 미칠 것이다.

콘텐트 재사용성은 제품 문서화 작업의 효율을 향상시킬 뿐아니라 콘텐트의 일관성을 유지함으로써 정보제품(information product)에 대한 신뢰도를 높일 수 있다. 그러나 DITA의 적용은 도입 및 운영비용이 추가적으로 소요되고 문서화 작업의 복잡도를 증가시킬 수 있는 위험요소도 존재하므로 주의 깊게 접근해야 한다.

이러한 위험요소가 존재하지만 투자효과(ROI)를 높일 수 있는 방안 중 하나가 콘텐트 재사용성을 최대한 활용하는 것이다. DITA에 대한 국내 연구가 미흡한 현실에서 이 연구 결과는 DITA 기술을 적용하고 있는 기업에서 참고 자료로 활용할 수 있을 것이다.

참 고 문 헌

[1] Deborah Pickett, "An Introduction to DITA", Open Source Developers' Conference 2005, 2005.

[2] http://xml.coverpages.org/DITA-OASIS-CFP.html.

[3] DITA: Four Letters You Need to Know, PTC, 2006.

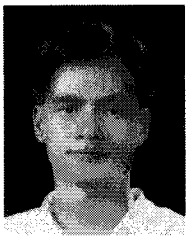
[4] Chad Jacson, Mehul Shah, Next-Generation Product Documentation, Aberdeen Group, 2006.12.

[5] Ann Rockley, "Information Modeling: A Practical Approach", STC's 50th Annual Conference Proceedings, 2003.

[6] Namahn, Darwin Information Typing Architecture, A research note by Namahn, 2000.

[7] Michael Priestley, "DITA XML: A Reuse by Reference Architecture for Technical Documentation", SIGDOC'01, ACM, 2001.10.

[8] Sarah O'Keefe, Accessing DITA as a Foundation for XML Implementation, Scriptorium Publishing Services, 2009.



구 흥 서 (Koo, Heung Seo)

- 1985년 : 인하대학교 전산학 (이학사)
- 1989년 : 인하대학원 전산학 (이학석사)
- 1993년 : 인하대학원 전산학(이학박사)
- 1994년 3월 ~ 현재 : 청주대학교 컴퓨터정보공학과 교수
- 관심분야 : 데이터베이스, ECM, u-헬스케어, Technical Documentation