

콘텐츠 세계화 관리 시스템 (GMS:Globalization Management System)

이 영 미* 이 경 일**

목 차

1. 서 론
2. 웹사이트의 세계화
3. GMS 모델 및 기능
4. GMS의 기술 요소
5. 결 론

1. 서 론

1.1 GMS의 개요

GMS(Globalization Management System)는 기업의 문서, 데이터베이스, 기업용 프로그램, 웹 사이트, 매뉴얼 등에 있는 다국어 정보를 관리하는 기업용 프로그램이다. GMS는 기업 정보의 다국어, 다지역, 다문화별 관리를 가능하게 한다. 따라서 기본적으로 전세계에서 동시에 발생하는 수백개의 지역화(localization, L10N) 프로젝트를 관리할 수 있어야 하며, 더 나아가 세계화(Globalization, G11N) 관리 기능을 제공해야 한다. 세계화 프로세스라는 면에서 볼 때, 소프트웨어 프로그램의 세계화 프로세스와 웹사이트의 세계화 프로세스는 성격이 달라서 접근 방법이나 요구되는 기술이 다르기에 본 논문에서는 웹사이트 세계화(이하, WG)중심의 GMS 이해를 목표로 내용을 전개하고자 한다.

1.2 GMS의 필요성

세계화 및 지역화 대상이 되는 기업의 콘텐츠는

이메일에서부터 정책 자료, 계약서, 마케팅 자료, 소프트웨어, 웹사이트 등 그 종류가 매우 다양하고 방대하며, 복잡한 형태를 가지고 있다. 조금 더 상세히 살펴보면, (1)각 기업에서 다루어야 할 분야가 법, 공학, 마케팅, 고객응대 등의 매우 넓은 영역을 포괄하고 있으며, (2)워드, HTML, PDF, XML 등 문서의 포맷도 매우 다양하고, (3)파일 시스템, 데이터베이스, CMS 등의 저장 관리되는 형태도 상이하다. 또한 (4)각 콘텐츠의 소유자(소유 부서)가 다양하며, (5)테스크 탭에서부터 기업용 서버까지 그 운용 환경도 매우 다양하다.

이러한 기업 콘텐츠들이 지속적으로 여러 언어로 번역되며, 각 언어 쌍들이 상호 연동되어 유기적으로 관리되어야 할 경우, 기업은 번역 혹은 지역화 프로세스를 관리하는 것이 얼마나 어려운지 절감하게 된다. 심지어 사내 공통으로 사용할 용어 정리조차 어렵다는 것을 알게 되며, 많은 경우에 사내 번역 수요가 어디에서 발생할지, 그 전체 비용은 얼마나 될지도 알기 어렵다.

WG의 경우 콘텐츠의 발생 빈도가 매우 높을 뿐만 아니라, 그 생명 주기가 상당히 짧아서 세계화를 위한 프로세스 관리 및 다국어 콘텐츠 관리

* 모비코엔시스메타(주) 연구소장

** 모비코엔시스메타(주) 부사장

가 더욱 중요해진다. 대다수의 사람들이 처음에는 WG를 번역과 동일한 것으로 생각한다. 그래서 콘텐츠 세계화를 위해 수정 대상 콘텐츠를 분석, 분류해 내고 사람의 손으로 번역 작업을 수행한 후, 목표 콘텐츠를 재구성하는 과정을 거치게 된다. 특히, HTML 및 XML로 구성 완료되어 있는 page나 CGI형태의 동적인 콘텐츠 세계화에는 더 많은 노력이 필요했는데, 모든 웹 콘텐츠로부터 텍스트를 추출하고, 번역을 수행한 후 다시 HTML 등으로 재구성해야 하며, 특히 복잡하게 연결되어 있는 링크들을 다시 연결해 주어야 하는 어려움이 있었다. 사실 실질적인 번역 비용은 세계화 노력의 10% 내지 15% 정도만을 차지한다. 세계화란 언어와 문화, 국가를 초월해 제품 전개와 마케팅 콘텐츠 관리를 포함한 복잡한 비즈니스 프로세스이기 때문이다.

지역과 국가, 지리적인 경계를 넘기 위해서는 통화(currency), 세금 관련 내용, 검열 등의 사회적 관습, 배송 등의 물류 체계, 해당 지역의 제품 지원 사항 등 폭 넓은 문제를 다뤄야 한다. 또한 콘텐츠 지역화(L10N) 활동은 회사 내외부의 많은 사람들의 협업을 필요로 한다. 게다가 이 협업은 언어와 시간대를 초월해야 할뿐 아니라, 세계의 여러 조직에 흩어져 있는 여러 그룹의 목표를 조율해야만 한다. 지역화는 대개 번역사, 감사자, 지역의 마케팅 관리자, 법률 검토자, 품질 검토자, 개발자 등등 여러 사람으로부터 입력을 받게 되는데, 검토 프로세스의 경우 조직과 부서에 따라, 또 제품 라인에 따라 콘텐츠 종류 및 목표 시장에 따라 달라진다. 가장 어려운 것은 항상 변화하는 웹 사이트, 마케팅 정보 및 제품 관련 문서의 다중 버전이 동기화되어 여러 언어로 최신 정보를 유지하는 프로세스를 관리하는 것이다. 이 문제는 목표 시장 수가 증가하고 특정 시장에 대한 맞춤 정보 요구가 증가할수록 기하급수적으로 복잡해진다.

1.3 GMS 도입 효과

대부분의 GMS는 다양한 소스의 콘텐츠들을 관리할 수 있는 단일 지역화 관리부를 제공한다. 따라서 서로 다른 CMS가 여러 개 있는 경우에도, GMS는 한 개만 사용하게 된다. 한 시스템에서 모든 지역화 관리가 이루어질 수 있으므로 GMS는 업무 복잡도를 낮추어 주고 중앙집중식의 용어 관리가 가능하므로 일관된 브랜딩에 도움을 주며, 지역화 프로세스 관리 일원화, 지역화 프로세스와 비용의 가시화 등의 성과를 이루게 된다. 즉, 기업은 GMS를 통해 글로벌 비즈니스 관리 비용 감소, 시장 진입 시간 감소, 브랜드 메시지 일관성 확보, 효과적인 기업 정보 전달을 이루는 것이다. 다국어 사이트 관리 비용 감소 및 번역 비용 감소로 인해 기업은 더 많은 콘텐츠를 더 많은 직원 및 파트너, 고객과 함께 나눌 수 있게 되는 것이다.

2. 웹사이트 세계화

웹사이트 세계화는 웹사이트가 점점 확대되고 콘텐츠가 수시로 변경되는 특성 때문에 GMS에서 일반적으로 요구되는 기능 외에도 콘텐츠 일관성 확보와 시기 적절성을 유지하는 것이 관건이다. 따라서 콘텐츠 관리는 WG의 아주 중요한 목표다. 특히 E-commerce는 실시간 거래를 기본으로 하기에 E-commerce 콘텐츠는 비즈니스와 고객의 요구사항 변화에 대해 지속적으로 응답해야 한다. 즉, 관련 정보와 트랜잭션, 메시지 모두가 다국어 사이트 전체에 걸쳐서 정확하고, 일관적이며 최신의 것이 되어야만 하는 것이다.

이를 위해 GMS와 연동된 CMS(이하 Global CMS)는 웹사이트 콘텐츠 상의 변화를 감지하고 즉각 지역화 프로세스가 시작되도록 돕는다. 워크플로우는 웹사이트 콘텐츠의 개발, 저작, 리뷰, 승인, 시험, 저장과 사용 전체에 걸쳐 업무를 조정하

는 역할을 한다. 일반적으로 GMS는 새로운 콘텐츠를 기존의 콘텐츠와 비교하기 위해 TM (Translation Memory) 기술을 사용하며 전에는 번역된 적이 없었던 콘텐츠만을 추출하여 번역을 수행한다. 최근에는 기계번역(MT:Machine Translation)기술의 발전으로 TM과 함께 MT 기술이 결합되어 높은 비용 효과를 내고 있다. 지역화된 또는 번역된 내용은 관련 웹 페이지 또는 데이터베이스에 자동으로 올라간다.

Global CMS에는 여러 종류가 있으며 지원하는 웹 애플리케이션이 얼마나 복잡한가에 따라 그 특징이나 기술이 다양하게 나타난다. 특히 몇몇 해외 선진 회사의 e-Biz용 Global CMS의 경우 완벽한 세계화 솔루션을 제공하기도 한다. 기능 비교에 대한 자세한 내용은 3장에서 다루기로 한다.

3. GMS 모델 및 기능

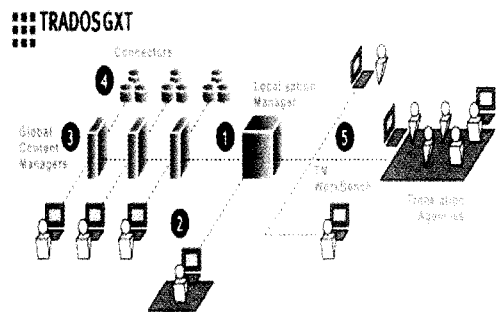
3.1 GMS 시스템 모델

일반적으로 GMS는 세 개의 구성요소를 가진다. 첫째는 콘텐츠 관리부이며 소스 로케일(locale)을 모델로 하여 소스상의 변화가 다른 로케일에 반영되도록 관리한다. 따라서 변화를 감지하고 포맷을 변경하여 지역화 관리부에 전달해주며, 지역화가 된 내용은 본래의 포맷으로 바꾸어준다. 둘째는 지역화 관리부인데 보통 TM과 MT기술을 활용하여 지역화가 필요한 텍스트만을 추출하고 이를 자동 번역하고, 그 내용을 번역사 및 감수자에게 전달한다. 지역화 프로세스 관리용 사용자 인터페이스는 미리 보기 기능과 TMX(Translation Memory exchange)지원 기능을 제공한다. 셋째는 워크플로우로 전체적인 세계화 프로세스 관리를 위해 사용된다.

인터넷 시대에는 가능한 한 빠르고 정확하게 정보를 습득하고 분배하는 능력이 요구되는데, 이는

언어장벽이 없는 경우이나 가능하다. 정확성이라는 면에서는 사람을 통한 번역이 바람직하지만 번역시간이 많이 걸리므로, MT는 실시간 솔루션이 요구되는 경우 초벌 번역을 제공하기 위하여 통합 사용되는 것이다.

GMS사례로서 본 모델에 충실하면서도 대표적인 업체 중 하나인 TRADOS사의 GMS, TRADOS GXT를 소개한다.



(그림 1) TRADOS GXT

- 1) Localisation Manager : 지역화 프로세스 관리를 위한 중앙 워크플로우 자동화 시스템
- 2) Workers : 지역화 프로젝트를 발주하고 추적하는 부서
- 3) Global Content Managers : 업무 규정, 워크플로우 자동화, 사이트 관리, 변화 감지를 수행하는 CMS 인터페이스
- 4) Connectors : 기업용 애플리케이션 연동을 위한 인터페이스
- 5) Localisation Vendors : 지역화를 위해 TM workbench를 사용하는 그룹

세계화를 위해서는 기본적인 구성요소 외에도 다양한 툴을 활용해야 한다. 이러한 툴에는 Code Analyzer, Internationalization checker와 같은 국제화(Internationalization, I18N) 툴, TM과 용어관리

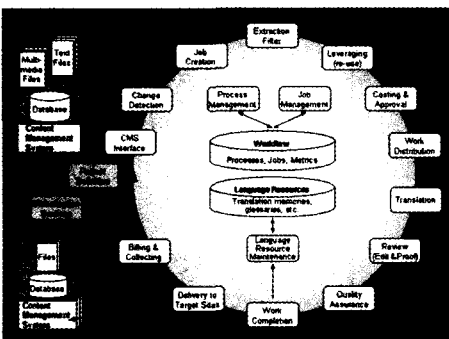
(Terminology management) 기능을 포함하는 CAT (Computer Aided Translation) 툴이 있다. 그 밖에도 S/W 지역화 툴, QA 툴, 전자 출판용 (Desktop Publishing) 툴 등이 활용된다.

3.2 GMS Workflow 모델

대표적인 모델로는 Cadieux의 모델[2]이 있는데, 이는 다국어 웹사이트의 자동 관리를 위한 단계와 구성 요소를 열거한 생성 모델(그림 2)로 인정을 받고 있다. Cadieux는 이미 GlobalSight, Uniscape, Tridion, WolrdPoint사의 제품에 대한 비교 평가를 위해 이 모델을 성공적으로 사용한 바 있으며, 국내에서는 모비코앤시스메타(주)에서 이에 준한 시스템을 구현한 바 있다[3].

Cadieux 모델은 워크플로우 엔진과 언어 자원, 통합 관리 기능을 중심으로, 콘텐츠 세계화의 전체 프로세스를 자동화하고자 하는데 목표를 두고 있다. Cadieux 모델은 다음의 세 가지 면에서 특히 유용한 모델이다.

- 콘텐츠가 한 로케일에서 다른 로케일로 변환되는 워크플로우를 논리적으로 전개한다.
- 콘텐츠 세계화를 위해 반드시 수행해야 하는 활동을 명확하게 정의한다.
- 객관적인 제 3자 관점을 제공한다



(그림 2) Generic model of the automated process provided by Web site globalization technologies

3.3 GMS에서 필요한 기능

다음 <표 1>에서는 Global CMS에서 요구되는 기능을 GMS와 CMS 측면에서 나누어 보았다. 이를 통해 GMS의 보다 세부적인 기능 파악이 가능하며 CMS와의 관계를 좀 더 명확히 이해할 수 있다.

<표 1> GMS 기능 분류표

기능	GMS	통합 요소	CMS
Leverage Previous Translations	●		
Filters to Extract Natural Language Content form Native File Formats	●		
Segment Content by Phrase and Sentence and Match it with Localized Content	●		
Terminology Management Across Languages	●		
Translation Tools	●		
Pricing of Localization Costs by Project	●		
Localization of Graphic	●		
Business Rules for Managing Relationships Between Multiple Multilingual Site	●		
Tracking Localization Schedules	●		
Workflow	●	●	●
Change Detection	●	●	●
Content Creation Workflow			●
Collaboration in Content Creation			●
Project Management			●
Staging			●
Publishing			●
Archiving & Versioning			●
Input Methods			●
Developer Access			●

4. GMS의 기술 요소

4.1 자동 번역, 번역 자동화 기술

최근 10년간 자동 번역 기술은 비약적으로 발전을 해 왔다. 자동 번역 기술은 크게 (1)기계번역(MT)기술, (2)번역메모리(TM)기술, (3)휴먼번역지원(CAT)기술, (4)응용 어플리케이션 개발 기술 등으로 나누어 볼 수 있다.

4.1.1 기계번역 기술

기계번역 기술은 컴퓨터가 발명된 초기부터 많은 연구와 투자가 진행된 깊은 역사를 가지고 있다. 최근 10년 동안 놀라운 컴퓨터 성능 개선 및 언어 자원 처리 기술의 발전 덕분에 번역률 부분에서도 상당한 진전이 이루어진 것이 사실이다. 그러나, 현재 유사 언어권(한국어/일어, 영어/독어 등)이 아닌, 상이한 구조를 가진 언어 간 번역(한국어/영어)의 경우 GMS와 연동되어 상업적으로 쓰기에는 부족한 성능을 보이고 있다. 최근에는 MT가 인간 대신 번역을 수행하는 역할보다는 빠르게 해외 정보를 확인하거나, 실시간 다국어 의사소통(다국어 채팅), 다국어 CRM에서의 다국어 게시판 기능, 다국어 정보 검색 등에 적용되어 Global CMS와 연동되는 사례들이 늘고 있으며, TM 기술과의 결합 및 통계적 방법을 사용한 새로운 패러다임의 MT기술 개발이 진행되고 있다. GMS에의 실용적 적용을 위해서는 보다 높은 수준의 MT 기술 개발이 필요하며, 특히 엔진 차원에서의 TM과의 결합을 모색하는 등의 추가 연구가 필요할 것으로 보인다.

4.1.2 번역메모리 기술

사람이 번역한 문장을 기억하고 있다가 이와 비슷한 문장을 재번역하고자 할 때, 과거에 사람이 번역한 내용을 재활용하여 번역에 소요되는 비용과 시간을 효과적으로 줄이는 기술이다. 번역메모

리 기술을 위해서는 원문과 번역문을 상호 연동하는 텍스트 얼라인먼트 기술과, 단위 번역 문장(Translation Unit)을 관리 운용하기 위한 DB 및 언어자원 관리 기술이 응용되어야 한다. 최근에는 TMX 표준 규약을 따른 이기종 TM들을 통합하여 대규모 통합 TM 시스템을 구축하고 이를 상업적으로 사용하기 위한 연구가 활발히 진행 중에 있다. 보다 실용적인 GMS 및 Global CMS 시스템의 공급을 위해서는 형태소 분석 및 구문 분석 등의 자연언어처리(NLP) 기술이 깊게 활용되고, 구문 패턴에 의한 지능적 구문 matching 및 부분적 대역 문장 생성을 수행하는 차세대 TM system에 대한 연구가 필요할 것이다.

4.1.3 휴먼번역 지원 기술

CAT(Computer Aided Translation) tool 관련 기술로서 사람이 번역을 할 때, 컴퓨터가 여러 측면에서 업무 효율을 높여줄 수 있도록 편리한 기능을 지원해 주는 기술이다. CAT tool은 일종의 워크벤치로서 TM 혹은 MT와의 연동을 통한 자동번역을 지원하며, terminology 시스템과의 연동을 통해 통일되고 규격화된 번역이 가능하도록 지원하고, 편집 및 부분적 번역문 갱신 등의 번역가에게 필요한 제반의 기능을 제공하고 있다. 최근에는 다수의 사람들이 온라인상에서 CAT 시스템을 같이 사용할 수 있는 Web 기반의 CAT 기술 개발이 완료단계에 있으며, 기업의 인트라넷에 연동되어 표준 문서 작성 및 공문 수발신 시스템과 결합되는 양상을 보이고 있다.

4.1.4 응용 어플리케이션 개발 기술

지난 몇년간 자동 번역 시스템과 연동되어 사용자 편의를 증대시킬 수 있는 다양한 자동번역 어플리케이션 소프트웨어 기술 개발이 수행되었다. 초창기의 인터넷 자동 번역 기술을 비롯해, 현재는 실시간 다국어 채팅, 다국어 게시판, 다국어 e-mail 번역 시스템 등의 e-biz의 모든 부분에 자동 번역

시스템을 결합해 활용되고 있다. 웹 세계화의 산업적 성공을 위해서는 이 부분의 추가적 연구가 필요할 것이다.

4.2 세계화 프로세스 관리 및 협업 기술

앞에서 이미 논의되었던 것처럼 세계화 프로세스는 매우 복잡하고 다양한 기술이 결합되어 있다. 워크플로우 시스템은 이렇게 복잡한 다단계 업무를 효과적으로 관리할 수 있도록 설계되어야 하며, task 정의, 생성, 업무 할당, 비용 배분, 일정관리, 고객에게의 결과물 전달까지의 모든 업무를 일괄적으로 관리하여야만 한다. 특히, 대규모 프로젝트 또는 적은 양의 콘텐츠라도 빠른 시간 안에 번역 및 세계화하여 반영해야 하는 e-Biz 환경에서는 온라인 상에서의 협업이 필수적이다. 수십명의 협력자들이 상호 연동되고, 업무를 교류, 조정할 수 있는 협업 시스템의 지원이 없이는 이러한 형태의 프로젝트를 수행하는 것이 불가능하게 된다. 국내에서는 최근 몇 년 동안 워크플로우 시스템과 협업 시스템에 대한 연구가 활발히 진행되고 있으나, GMS 영역에의 도입은 아직 시도되고 있지 못한 실정이다.

4.3 콘텐츠 동기화 및 분석/생성 기술

WG 시스템은 콘텐츠들이 고정되어 있지 않고, 지속적으로 생성, 갱신되며 매우 빠른 속도로 전파되는 인터넷의 특성을 반영해야만 한다. 특히, 다국어 콘텐츠들은 상호 연동이 되어 있어서, 한국어 page의 일부가 변경된다면 변경된 부분을 자동 감지(auto-detection)하고, 영어, 일본어 등의 다른 언어 page의 해당 부분을 빠르게 수정할 수 있도록 워크플로우에 task를 생성할 수 있어야만 한다.

이러한 것이 실제로 가능하려면, 다국어로 되어 있는 같은 내용의 콘텐츠들이 상호 연동되고 동기화 될 수 있어야 하며, 동기화를 유지하기 위해서

는 변경을 자동 감지하고 처리할 수 있는 기능을 지원하여야만 한다. 특히, HTML 형식의 웹 page, 게시판에 첨부된 문서들을 자동 동기화하기 위해서는 이들의 문서 형식을 분석하여 실제 번역 대상 콘텐츠만 분리하여 동기화 및 번역을 수행하고, 변경된 부분만 다시 반영하는 콘텐츠 분석 및 생성 기술을 지원해야만 한다.

5. 결 론

지난 십년간 세계 경제는 세계화의 바람 속에서 인터넷의 빠른 성장에 힘입어 급속히 하나가 되어가고 있다. 이는 선진 번역업체에게 S/W와 웹 애플리케이션의 지역화, 엔지니어링, 시험, 관리뿐만 아니라 제품이나 콘텐츠 자체를 세계화하기 위한 기술을 요구하게 된다.

세계적 관련 기업들은 이에 부응하여 종합적인 세계화 솔루션에 막대한 투자를 하고 있다. 특히, 최근에는 다국어 콘텐츠 관리, 번역관리 시스템, TM, MT 등이 워크플로우 상에서 통합되어 실시간 지역화가 가능해지고 나아가 세계화 프로세스 관리 및 마케팅, 세일즈를 위한 컨설팅을 제공하고 있다.

반면 국내에서는 CMS, CAT 등 IT 시스템 관점의 지협적 연구 개발과 상업화는 진행되었으나, 통합 시스템으로서의 GMS 혹은 Global CMS에 관한 연구는 미미한 실정이다. 이 분야의 국내 선도 기업인 모비코엔시스메타는 2001년도부터 GMS 부문에 대해 지속적인 연구 개발을 수행하고 있으며, 상용 제품으로서 iGlobiz를 소개한 바 있다.

본 논문에서는 GMS에 대한 소개와 시장 상황 및 기술적 과제들을 소개함으로 GMS에 대한 이해를 돕고자 했다. 최근 들어 국내에 공격적으로 진출하고 있는 다국적 기업들에 대응하기 위해서는 해당 업계와 학계가 협력하여 관련 기술 개발에 힘을 기울여야 할 것으로 생각되며, 특히 동북아 허

브 국가를 표방하고 있는 대한민국 정부에서도 세계화 관련 기술에 보다 폭넓은 지원을 해야 할 것으로 판단된다.

참고문헌

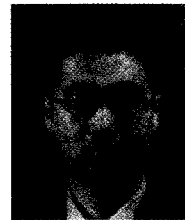
- [1] Howard Schwartz, "Seven Deadly Sins of eBusiness Globalization", Uniscap, 2001.
- [2] Pierre Cadieux, "Evaluating Technologies for Web Site Globalization", Multilingual Computing issue #35 Oct/Nov, 2000.
- [3] 이영미, 서형국, 이경일, "Cadieux 모델에 기반한 Web Globalization 시스템 구현", 제 15회 한글 및 한국어 정보처리 학술대회, 2003.
- [4] Jordan Woods, "Maximizing ROI, Best Practices For Content Globalization", GlobalSight White paper, 2002.
- [5] Barry Parr and Maureen McManus, "Web-Site Globalization, The Next Imperative for the Internet 2.0 Era", IDC White Paper, 2002.
- [6] LISA, "Where Enterprise Globalization Becomes a Reality!", LISA Global Strategies Summit Asia, 2001.

저자약력



이 영 미

1991년 호서대학교 전자계산학과(학사)
 1994년 광운대학교 전자계산학과(석사)
 1994년~1998년 KAIST정보시스템연구소 연구원
 2000년~2001년 모비코앤시스메타(주) 기술연구소
 선임연구원
 2002년~현재 모비코앤시스메타(주) 연구소장
 관심분야 : CAT, TM, GMS, CMS, KMS, Collaboration
 이 메 일 : zerome@mobico.com



이 경 일

1996년 인하대학원 전자재료공학과(공학석사)
 1997년~1999년 LG중앙연구소 연구원
 1999년~2001년 현대전자연구소 연구원
 2001년~2003년 (주)시스메타 대표이사
 2003년~현재 모비코&시스메타(주) 부사장
 관심분야 : NLP, Information Retrieval, Text mining,
 Machine Translation
 이 메 일 : tony@sysmeta.com