

# 게임전문용어전자사전 구현에 관한 연구

임충재\*, 윤태수\*

\*동서대학교 게임&멀티미디어공학전공

e-mail : dooly@dongseo.ac.kr

## Implementation of the Game Terminology Electronics Dictionary

Choongjae Im\*, Taesoo Yun\*

\*Dept. of Game & Multimedia Engineering, Dong-Seo University

### 요 약

세계는 지식정보의 창출 및 활용정도가 국가경쟁력을 좌우하는 지식기반사회로 급격히 전환하고 있다. 이러한 상황에서 주요 지식정보의 집중적인 디지털화를 통해서 국가 지식정보 공유체계를 구축하는 것이 무엇보다 시급한 상황이다. 특히 급격하게 성장하고 있는 디지털 콘텐츠 분야의 체계적인 정보화 사업은 관련 분야의 선진국과 경쟁하기 위해서 절실하게 요구된다. 본 논문은 게임 산업에 대한 체계적인 정보화를 추진하기 위해서 의미를 기반으로 한 게임전문용어전자사전 구축을 위한 연구 결과를 소개한다. 의미를 기반으로 한 게임전문용어전자사전 구축하기 위해서 다양한 형태의 게임 관련 자료를 통해서 망뭉치를 추출하였다. 추출한 망뭉치로부터 빈도수 등을 이용해서 중요한 의미를 갖는 표제어를 추출하였으며, 추출한 표제어를 이용해서 게임전문용어 시소러스를 구축하였다. 또한 웹 기반으로 게임전문용어전자사전을 구축하기 위해서 SDML 을 이용하여 전문용어를 표기했으며 다양한 검색 기능을 추가해서 게임전문용어전자사전의 프로토타입 시스템을 구축하였다.

### 1. 서론

세계는 지식정보의 창출 및 활용정도가 국가경쟁력을 좌우하는 지식기반사회로 급격히 전환하고 있다. 이러한 상황에서 주요 지식정보의 집중적인 디지털화를 통해서 국가 지식정보 공유체계를 구축하는 것이 무엇보다 시급한 상황이다.

최근 국내게임산업은 얼마되지 않는 기간에 세계 게임업계가 주목할만한 성과를 얻고 있다. 하지만 게임 기반기술, 게임산업동향, 게임 콘텐츠 기획 등과 같은 분야는 게임 선진국인 미국이나 일본에 비해서 상당한 격차가 있는 것이 사실이다. 짧은 게임산업의 역사를 가진 국내의 영세 게임업체가 전통과 대규모 자본을 갖춘 미국과 일본의 게임업체와 경쟁하기 위해서는 게임에 대한 양질의 정보 확보가 무엇보다도 시급하다.

본 논문은 게임산업에 대한 체계적인 정보화를 추

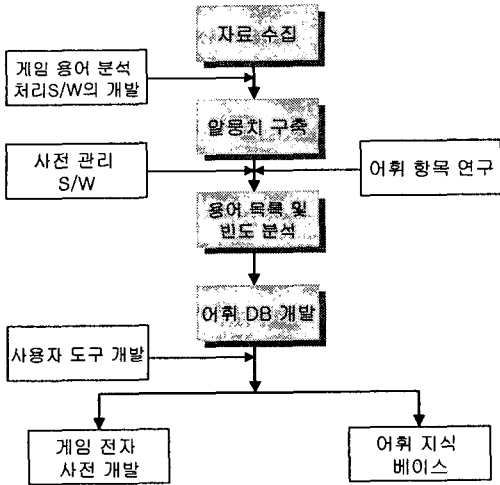
진하기 위해서 의미를 기반으로 한 게임전문용어전자사전 구축을 위한 연구 결과를 소개한다. 2 장에서는 게임전문용어전자사전 구축을 위한 방법론을 소개하고, 3 장에서는 이 방법론을 활용한 실제 구축 과정에 대해서 자세히 설명한다.

### 2. 게임전문용어전자사전 구축 방법론

본 장에서는 게임산업에 필요한 정보를 다양한 관점에서 접근하여 원하는 항목을 쉽게 찾을 수 있는 게임산업에 특화된 게임전문용어전자사전의 구축 방법론을 설명한다.

[그림 1]은 게임전문용어전자사전 구축을 위한 전체적인 방법론을 그림으로 나타낸다.

게임전문용어전자사전을 구축하기 위해서는 먼저 게임 분야의 전문용어에 대한 의미계층체계를 구축한다. 그리고, 의미계층에 기반하여 말뭉치로부터 전문



[그림 1] 게임전문용어전자사전 개발 과정

용어를 추출하여 시소러스를 개발해야 한다. 의미계층 체계에서 가지는 의미간의 관계는 주로 등가관계, 계층관계, 연관관계가 대표적이다. 이러한 관계는 게임 산업에 적합한 말뭉치(corpus)를 구축하여 이 말뭉치로부터 전문용어를 추출하여 그 용어의 관계들을 선택함으로써 가능하다. 본 연구에서는 대표적인 온톨러지의 일종인 WordNet 을 기반으로 데이터베이스진흥원의 “시소러스 개발지침”을 따랐다.

## 2.1 말뭉치 구축

### 2.1.1 말뭉치의 정의 및 필요성

말뭉치란 정의는 사람에 따라 다르나 대략 다음과 같은 의미로 쓰이고 있다.

- 대규모 언어 데이터베이스
- 인간의 음성언어(문어, 구어)를 대용량 컴퓨터에 저장하고 이를 필요에 따라 가공하여 언어연구에 사용하는 것
- 컴퓨터가 판독할 수 있는 형태로 저장된 자연어의 용례들과 이들 용례에 대한 부가적인 정보

언어 정보화 단계에서 가장 먼저 시도되어야 할 일이 바로 말뭉치 구축이다. 말뭉치는 다용도의 언어공학 처리를 위해 필요한 기초 자원이다. 말뭉치는 용어 추출, 용어의 사용빈도 파악, 해당분야 언어모델링 등에 필요한데, 해당 분야의 용어학적 대표성을 가질 만큼의 양과 소분야별 포괄성을 지닐 때 의미있는 말뭉치가 된다. 말뭉치는 과학기술의 발전에 따라 새로운 정보가 출현하고 아울러 새로운 용어가 끊임없이 생성된다는 시사성을 고려하여 다양한 수단과 매체로 구축되어야 한다. 기존에는 인쇄화된 문서가 대부분이었으나, 현재에는 전자파일로 혹은 웹으로부터 추출을 통해 말뭉치를 수집한다.

### 2.1.2 말뭉치 구축의 기본원칙

말뭉치가 신뢰성을 가지기 위해서는 모집단을 연구하기 위해 구성된 표본집단처럼 구성되어야 한다. 모집단으로부터 표본을 구성할 때 가장 중요한 문제는 대표성이다. 따라서 어떤 전문분야를 연구하기 위한 말뭉치는 그 전문분야에 대한 대표성을 지녀야 하고, 구축된 말뭉치는 다양한 언어연구에 활용될 수 있는 것이어야 한다. 즉, 균형성과 정보성을 가져야 한다.

### 2.1.3 말뭉치의 구축의 단계

일정한 규모의 말뭉치를 구축할 때 그 구축 단계는 다음과 같다.

1. 적절한 규모와 말뭉치의 범위 결정
2. 텍스트의 우선 순위 결정
3. 문헌 목록 검색을 통한 텍스트 선정
4. 기계 가독형으로 작성
5. 텍스트를 이용할 수 있는 형태로 가공
6. 오류 수정 및 유지 보수

## 2.2 시소러스 구축

시소러스는 원래 자연언어 처리분야에서 전문용어의 개념(의미)간 관계를 계층적으로 표현한 데이터베이스로서 인터넷의 발달과 함께 정보검색 시스템에서 효율적인 정보의 검색 및 활용을 위해 전자사전구축을 위한 필수적인 요소이다. 즉, 전문적인 용어를 모르더라도 전자사전이 용어의 의미(개념)별 계층관계를 이용한 시소러스를 기반으로 구축되었다면 해당 용어에 대한 분류어휘 목록과 계층적 구조를 탐색하여 쉽게 원하는 결과를 얻을 수 있다.

시소러스에서 사용하는 용어관계는 크게 두 가지로 구분할 수 있다. 첫 번째는 거시적 레벨로 각 용어간의 의미적 관계인 등가관계, 계층관계, 연관관계라는 기본적인 관계이며, 두 번째는 미시적 레벨로, 등가관계에 속한 용어의 집합과 계층관계에 있는 용어, 연관관계에 있는 용어 범주의 상호관계 및 전체와의 관계이다. 시소러스는 이러한 관계성을 이용, 탐색 시 키워드로 사용될 색인어나 탐색어를 선택하기 위한 도구로 사용된다.

## 3. 게임전문용어전자사전 구축

이 장에서는 2 장에서 설명한 구축 방법론을 이용해서 실제로 게임전문용어전자사전 구축에 대해서 설명하고자 한다.

### 3.1 자료 수집

효율적인 말뭉치 구축을 위해서는 게임관련 말뭉치 구성 텍스트 입력자료를 효율적으로 확보하는 것이 관건이다. 본 연구에서는 게임전문용어의 신뢰성을 확

보하기 위해서 다음과 같은 다양한 소스로부터 말뭉치를 구축하였다.

- PC 통신 정보검색 : 천리안이나 나우누리 등 PC 통신상의 게임, 정보통신 관련 동호회로부터 수집
- 인터넷 정보검색 : 인터넷의 보급과 정보통신의 발달로 인터넷을 이용한 정보 습득은 가장 보편적이며 수집 자료의 양적 면에서 가장 큰 부분이다.
- 관련문서 : 학술단체 및 관련기관에서 제작하여 판매하는 자료로서 각종 전자사전과 특허청자료, 전자책, 게임관련잡지 등이 이에 해당된다.

### 3.2 말뭉치(corpus) 구축

본 연구에서는 게임전문용어전자사전 구축을 위한 말뭉치 구축은 게임관련 DB 또는 원문자료를 작성 또는 수집하여 원시말뭉치를 만드는 1 단계와 수집 작성된 자료를 가공하고 표준화하는 2 단계로 나누어 진행하였다.

제 1 단계 작업은 원시 말뭉치 구축 방법론을 개발하고 게임관련 기초자료를 수집, 작성하는 단계이다. <표 1>에 구축한 게임전문용어 말뭉치 구성 텍스트 목록 및 어절수를 정리하였다.

<표 1> 게임전문용어 말뭉치 구성 텍스트 목록

자 료	말뭉치
1999 년도 국내 게임산업 동향조사, 게임종합지원센터, 1999	10,000
온라인게임 산업 육성방안, 한국전자통신연구원, 2000	5,000
게임백서, (재)게임종합지원센터, 2001	10,000
인터넷정보검색	12,500
PC 통신 정보검색	5,000
신문, 게임관련잡지	7,500

제 2 단계는 범용성이 높은 말뭉치의 표준화 규격을 연구, 개발하고 이에 기반한 말뭉치를 만드는 단계이다. 제 1 단계에서 수집 작성된 게임용어 말뭉치는 전자사전 구축을 위한 중요한 기반이 되지만 정형화되지 않은 자료이기 때문에 자료의 중복과 질적인 면에서 많은 문제점을 안고 있다. 이러한 문제점을 해결하기 위해서는 말뭉치 자료구축의 표준화 및 규격화가 이루어져야 하며 여기에는 말뭉치의 문서형식과 마크업(markup) 방식 등의 각종 정보의 표기방식을 정해야 한다.

본 연구에서는 50,000 여 어절로 구성된 게임전문용어 말뭉치로부터 추출된 약 1,000 여개의 게임전문용어를 대상으로 전자사전의 프로토타입을 설계하였다. 말뭉치로부터 추출된 게임전문용어 목록을 국내는 물론 국제적으로 상호 데이터 교류 및 공유를 위해서 국제규격에 맞는 마크업 언어로 정보화하여 게임전문용어전자사전을 구축하려고 한다. 이를 위해서 본 연

구에서의 게임전문용어전자사전의 작성은 국어정보베이스에서 정의한 TDMS(Text and Dictionary Management System)의 규격을 만족하면서 국제적인 표준 인코딩 방법인 SGML(Standard Generalized Markup Language)을 따르는 TEI(Text Encoding Initiative) 지침을 적용하여 작성하였다. <표 2>는 SDML 을 이용하여 정의한 게임전문용어전자사전의 예이다.

<표 2> TEI 지침 기반의 전문용어 정보 (예)

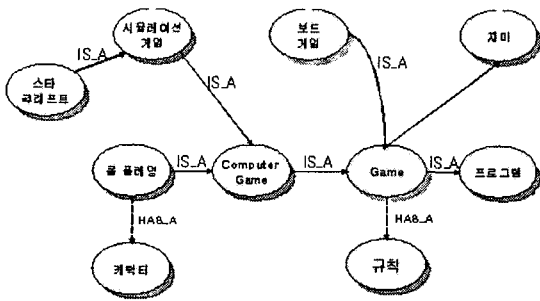
```
<termEntry id = MS1>
<descrip type = 'subjectLevel1'>장르</descrip>

<ntig lang = 'kor'>
<termGrp><term> 롤플레이팅게임 </term>
<termNote type = 'partOfSpeech'></termNote></termGrp>
<descripGrp><descrip type = 'definition'>
컴퓨터 게임의 한 장르로서, 플레이어가 게임속에서 어떤 인물의 역할을 맡아 수행하는 게임의 종류. 원래 RPG 라는 장르는 컴퓨터 게임 장르가 아니라 예전에 미국에서 개발되어 큰 인기를 얻었던 TRPG(Table Talk RPG)에서 시작되었던 것이다. RPG에서 빼놓을 수 없는 요소는 플레이어의 능력치와 성장이라 할 수 있다.플레이어가 게임을 진행시켜 나감에 따라 게임 속의 인물은 능력치가 점점 높아져 그 게임 속 세계에 더 잘 적응하게 된다. 결국 플레이어는 시나리오에 설정된 그 어떤 목적을 이룰 수 있을 때까지 인물을 성장시키는게 게임의 주요 관건이다.
</descrip>
<ptr type = 'sourcIdentifier' target = '...'></descripGrp>
<adminGrp><admin type = 'responsibility'></admin>
</adminGrp>
</ntig>
<ntig lang = 'en'>
<termGrp><term> role playing game </term>
<termNote type = 'partOfSpeech'></termNote></termGrp>
<descripGrp><descrip type = 'definition'>
</descrip>
<ptr type = 'sourcIdentifier' target = '...'></descripGrp>
.....
```

### 3.3 시소러스

본 연구를 위해 표제 대상에 오른 1000 여개의 게임전문용어들을 살펴본 결과, 다음과 같은 특징을 얻을 수 있었다. 단어들 사이에 반대어 관계가 나타나는 경우는 드물거나 크게 의미가 없었다. 부분-전체 관계는 상당히 유용한 것으로 보이며, 상-하위어 관계는 필수적이다. 따라서, 기본적인 동의어 표시 및 외국어(원문 및 영문표시) 및 사례관계들을 포함하는 시소러스를 고려하게 되었다. 이러한 관점에서 본 연구에서는, 게임용어를 위한 시소러스에서는 상하위어 관계와 부분-전체 관계의 정보를 중심으로 하여야 한다고 보고, 이를 위한 시소러스의 형식과 구조를 전개하였다. 이러한 방법을 이용해서 추출한 게임전문용어간의 의미망을 구축하였다.

[그림 2]는 제안한 관계를 표제어 “롤플레이팅게임”을 예로 의미망으로 표현한 것이다. 각각의 단어는 그래프를 통해서 다른 단어와 특정 관계로 이어져 있다. 여기에서는 Is\_a (A Is a B; A는 B의 한 종류이다) 관계와 Have\_a (A Have a B, A는 B를 부분으로 가지고 있다) 관계를 상징한 의미망이다.

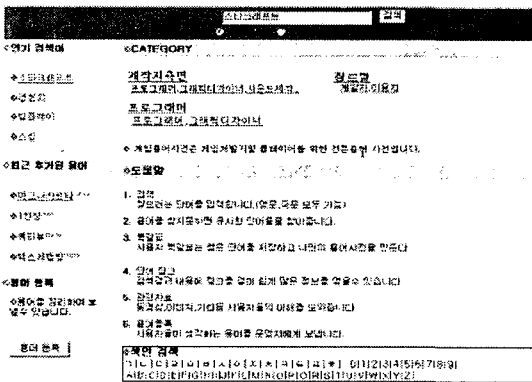


[그림 2] 게임전문용어 시소러스 구축 (예)

3.4 웹 기반의 게임전문용어전자사전 구축

의미 기반의 게임전문용어전자사전을 사용하기 편리하기 위해서 웹 기반으로 구현하였다. 게임전문용어 전자사전 서버는 ASP 를 이용하여 구축하였으며, 데이터베이스는 MS-SQL 을 이용했다.

[그림 3]은 구축한 게임전문용어전자사전의 초기 GUI 화면을 보여 준다.



2002.2 Copyright © Dongsoo, U. All Rights Reserved.

[그림 3] 게임전문용어전자사전 초기 GUI 화면

<표 3>은 웹 기반의 게임전문용어전자사전의 전체 기능을 나타낸다.

<표 3> 웹 기반의 게임전문용어전자사전의 기능

기능	설명
키워드 검색	키워드를 입력하면 키워드에 해당하는 자료를 검색하여 결과를 보여주는 기능으로서 찾고자 하는 용어를 게임용어와 개발용어의 범주로 나누어 검색 가능
카테고리 검색	분류된 카테고리를 따라가면서 검색을 수행할 수 있다. 사용방법은 일반검색 엔진의 디렉토리방식과 유사하게 찾고자 하는 게임용어의 분류를 선택하고 서브 카테고리로 이동해서 다시 찾아가는 방식

도움말	검색에 익숙하지 않은 초보자에게 정보검색 방법을 소개하는 기능으로서 관심있는 도움말을 클릭하면 데이터 서비스가 가능
색인	검색에 익숙하지 않은 초보자에게 한글 첫머리글자와 색인어를 이용해서 정보검색을 할 수 있도록 하는 기능
인기 검색어	이용자들이 가장 많이 검색하는 용어를 보여주는 기능
최근 추가 용어	가장 최근에 등록된 게임 용어를 보여주는 기능
용어 등록 신청	자신이 용어가 궁금할 때 신청하거나 이용자가 용어 등록을 할 수 있도록 하는 기능으로서 용어등록 메뉴를 클릭해서 등록 신청할 용어를 기록
히스토리 기능	그 동안 자신이 검색했던 용어들을 보여주는 기능

4. 결론

본 논문은 게임산업에 대한 체계적인 정보화를 추진하기 위해서 의미를 기반으로 한 게임전문용어 전자사전 구축을 위한 연구 결과를 소개하였다. 2 장에서는 게임전문용어전자사전 구축을 위한 방법론을 소개하였고, 3 장에서는 이 방법론을 활용한 실제 구축 과정에 대해서 설명하였다.

앞으로 게임산업 현장에서 필요로 하는 전문용어 전자사전을 구축하기 위해서는 전문용어 추출을 위한 신뢰성있는 말뭉치의 자동구축에 관한 연구, 사전정보를 효율적으로 저장, 검색, 수정할 수 있는 사전관리 시스템에 대한 연구, 타 전문 분야 국가 데이터베이스와의 연동을 위한 데이터베이스 표준화에 대한 연구 등이 선행되어야 한다.

참고문헌

[1] Juntae Yoon, Key-Sun Choi, Mansuk Song "Corpus-Based Approach for Nominal Compound Analysis for Korean Based on Linguistic and Statistical Information", Natural Language Engineering vol 7/No 3, 251-270, 2001.

[2] Young-Soog Chae, Young-Bin Song, Yong-Suk Choi, Key-Sun Choi "Building Knowledge Base from KIBS Corpus and MRD:A Case Study of Verbs", Proceedings of the 1th Workshop on Multi-lingual Information Processing and Asian Language Processing(MAL'99), pp. 96-101, Beijing, China, 1999,11.

[3] Young C. Park and Key-Sun Choi, "Automatic Thesaurus Construction Using Bayesian Networks", Information Processing and Management, Vol. 32, No. 5, pp. 543-553, 1996.

[4] Foskett, Douglas J. 1980. "Thesaurus" in A. Kent, H. Lancour, and J.E. Daily (eds.) Encyclopedia of Library and Information Science, 30: pp. 416-463. (New York: Marcel Dekker, Inc.)