

시맨틱 웹 환경에서 적합한 문장을 제공하는 이야기 쓰기 도우미에 관한 연구*

A Study of Retrieval Model Providing Relevant Sentences in Storytelling on Semantic Web

이태영(Tae-Young Lee)**

초 록

이야기 쓰기를 돕는 본문 및 문장 검색시스템의 구축을 위해서 (1) 이야기와 단락 및 문장의 구조를 분석하고 (2) 색인작성과 탐색 질문에 적용되는 언어 추론을 연구하였다. 이야기 쓰기에 필요한 이야기, 단락, 그리고 문장으로 구성된 사항 데이터베이스와 필요한 추론규칙으로 이루어진 지식베이스와 온톨로지가 고안되었다. 추론의 기초인 실례(實例) 파일들은 시맨틱 웹 환경에서 작동될 마크업 언어 형식으로 만들어졌다. 시맨틱 웹 환경에서 실용적인 시스템이 되려면 단락과 문장을 정확히 대변하는 색인 방법론과 이를 정밀하게 지식베이스화 할 수 있는 마크업 언어의 창조가 필수적이라 사료된다.

ABSTRACT

Structures of stories, paragraphs, and sentences and inferences applied to indexing and searching were studied to construct the full-text and sentence retrieval system for storytelling. The system designed the database of stories, paragraphs, and sentences and the knowledge-base of inference rules to aid to write the story. The Knowledge-base comprised the files of story frames, paragraph scripts, and sentence logics made by mark-up languages like SWRL etc. able to operate in semantic web. It is necessary to establish more precise indexing language represented the sentences and to create a mark-up languages able to construct more accurate inference rules.

키워드: 이야기 쓰기, 시맨틱, 웹, 사항검색, 지식베이스, 문장검색, 단락검색, 추론규칙
storytelling, semantic web, fact retrieval, knowledge-base, sentence retrieval,
paragraph retrieval, inference rule

* 이 연구는 전북대학교 발전재단 연구비에 의해 이루어졌음.

** 전북대학교 인문대학 문헌정보학과 교수(taehyun@chonbuk.ac.kr)

■ 논문접수일자: 2009년 10월 1일 ■ 최초심사일자: 2009년 10월 8일 ■ 게재확정일자: 2009년 10월 15일
■ 정보관리학회지, 26(4): 7-33, 2009. [DOI:10.3743/KOSIM.2009.26.4.007]

1. 서론

1.1 연구의 목적

인간의 암묵지를 형식지로 표현하고 체계적인 관리를 통해 그 지식을 공유하고 응용하는 것은 인류 발전을 위해 반드시 있어야 할 필수적인 활동이다. 그런데 사람의 생각을 원고지나 컴퓨터 워드 화면에 마음먹은 대로 기술하는 것은 상당히 어려운 작업으로 글 쓰는 직업을 가진 이들도 작문의 고통을 흔히 토로하고 있다. 이와 관련된 조사에서 80% 이상의 응답자가 글쓰기를 어렵다고 응답하였으며, 글을 쓸 때 어려운 문제로 (1) 글 구조 파악, (2) 단락 구성, (3) 문장 작성, (4) 어휘 구사 등 크게 4가지 요인이 거론되었다(이태영 1999).

글쓰기 초보자의 경우에는 작문의 어려움이 더 커질 수 있는데, 그것은 유사한 과거의 명문들을 참조/활용하면 많은 도움을 받을 수 있을 것이다. 예컨대 짜임새 있는 단편소설은 이야기의 구조/체계와 형식을 전수할 수 있으며, 유려한 문장들로 쓰여진 수필은 작문 시 문장의 전형을 보여줄 수 있는 것이다. 다시 말해서 스크립트나 프레임 형태로 작성된 소설의 줄거리, 단락/문장의 뼈대를 알려 주는 단락/문장의 구조 형식, 그리고 줄거리와 구조 형식의 예로 실제 글, 단락, 문장들을 제공하는 도우미 시스템이 있으면 작문 초보자들에게 참조와 모방 및 창작의 동력으로 작용하여 글쓰기는 훨씬 수월해질 수 있다.

실제 문장, 단락, 글들을 많이 확보하고 있는 시스템을 구축하는 것은 오늘날 많은 수의 글들이 디지털 자료로 저장되고 있어 그리 어려

운 일은 아니다. 문제는 저장된 글, 단락, 문장, 어구들이 이용자가 필요로 할 때 이용자의 요구에 정확히 부합하는 데이터를 제공할 수 있는 능력이다. 이용자가 원하는 글, 단락, 문장과 같은 사항을 검색하기 위해서는 기존의 문헌 검색시스템에서 사용하는 키워드 위주의 검색 방법으로는 해결되지 않는다.

“과충류가 분비하는 독물들은 많이 있습니까?”라는 문장을 검색할 때, {과충류-독물}과 같은 주제명 검색이나 {과충류, 독물}과 같은 단순한 명사 중심의 키워드 검색, 또는 {과충류(객체), 독물(주체)}과 같은 명사 키워드에 역할기호를 추가한 ‘키워드+역할’ 검색 방법으로는 위 문장에 있는 “많이, 있습니까”란 단어를 대변할 수 없다. 우선 용언 ‘있습니까’에서는 ‘존칭’, ‘의문형’이란 어미의 처리가 대두되고 부사 ‘많이’에서는 “많다/적다, 높다/낮다, 두껍다/얇다”와 같은 수량적 의미가 글을 분석하기 위한 언어 지식에 추가되어야 한다.

웹 환경에서의 이러한 문제는 RDF, OWL, SWRL 등 XML을 기반으로 하는 마크업 언어들이 위와 같은 언어 구조 관계를 푸는 열쇠 알고리즘을 많이 적재할 수 있으면 보다 많은 문제가 해결 되고 더 발전하여 버너스리(2001)가 추구하는 시맨틱 웹 환경을 구축할 수 있다. 현재 XML로 작성되는 웹문서는 중첩되는 여러 경로와 접두어를 포함한 노드 이름을 필연적으로 갖는 구조적 검색(structured retrieval)을 이루게 되므로(Manning, Raghavan and Schutze 2008, 178) 우선적으로 경로(path)와 노드(node)의 운영에서 언어 구조 관계들을 보다 단순하게 정리하고 실제 문장 예를 제시하여 보다 구체적으로 판단하게 하면 어느 정도 만족될 것이다.

만족한 문장/단락/글 검색은 이용자가 찾는 알맞은 문장/단락/글의 구조와 어휘를 제공하여야 한다. 이를 위해서 문헌 검색에 시소러스가 있듯이, 이야기 구조를 정의하고 웹 문서의 노드와 경로 및 술어를 중심으로 한 문맥 관계를 정확히 표현할 수 있도록 어휘/문장/단락/글에 대한 온톨로지가 있어야 한다.

구조적 검색에 사용되는 색인은 오늘날 한 탐색어에 수백만 내지 수억 개씩 재현되는 문서와 사이트들을 여과하는 기능(Manning, Raghavan and Schutze 2008, 179)도 있지만, 경로가 틀려 알맞은 문장이 검색되지 않는 경우도 발생할 수 있다. 따라서 본고는 이야기와 단락 및 문장의 구조와 관계를 정의하고 그 지식을 담은 온톨로지를 작성한 후 이를 토대로 시맨틱 웹 환경에서 이야기 쓰기를 도와 줄 수 있는 글쓰기 도우미 검색시스템의 한 유형을 제시하고자 한다.

1.2 연구의 범위와 방법

본 연구를 수행하기 위하여 어린이 동화책에 수록된 가족에 대한 짧은 이야기(short story)들 중에서 무작위로 50개의 표본을 추출하였다. 이야기의 기술 분량은 보통 책 크기의 2-3페이지 정도로 제한하였고 문체 종류는 설명문, 서사문, 묘사문, 대화문으로 한정하였으며 단락의 형식도 이삼형(1994)이 분류한 13가지 유형으로 구별하였다.

검색용 지식베이스 종류로 이야기 전체인 글 프레임, 이야기의 부분 부분인 단락 스크립트, 단락을 구성하는 개개 문장들을 나타내는 문장 논리를 설정하였다. 이야기 프레임과 단락의 스크립트 및 문장의 술어논리를 작성하기 위하

여 표본을 분석하였고 자동 이야기 쓰기 시스템들 중, 민스트렐(minstrel)과 브루투스(brutus)의 줄거리(Pérez, R. P. and Sharples, M 2004)를 비롯하여 여타 글쓰기 자료들에서 사용하고 있는 방법과 용어들을 그대로 차용하였다. 표본 분석 결과는 만들어진 시스템의 구성소에 부기하였다. 웹 문서는 XML에 기초한 RDF, OWL, SWRL의 형식을 응용하여 XML 문서로 구현하였고, 이 문서의 특성상 검색시스템에는 구조적 검색 방법(Manning, Raghavan and Schutze 2008, 184)을 도입하였다. 그리고 언어의 논리관계를 표시하기 위하여 형식언어를 응용하였으며 색인작성과 탐색 시에 사용되는 접근점으로 위의 색인언어에 덧붙여 그림을 활용하였다.

본 구조 검색의 색인은 이야기 전체를 탐색하는 프레임 형식의 1차 접근점과 이어서 이야기의 부분을 탐색하는 단락별 접근점 및 세부적 단위로 문장을 탐색하는 3차 접근점으로 구성하였다. 첫째인 이야기 프레임은 형식과 주제로 구성하였는데 형식은 배경, 테마, 줄거리, 해법으로 전개하였고 주제는 한명숙의 8가지 구조를 그대로 사용하였다. 둘째의 단락 스크립트는 단락의 소 주제문을 중심으로 뒷받침 문장들의 정보를 통합하여 단락이 표출하고자 하는 바를 표현하였다. 변수로 “유형(Type), 주체(Actor), 객체(Object), To, From, At, 상태(Status)”를 설정하였으며 ‘유형’에는 단락의 소 주제 항목들을 배정하였고 ‘상태’에는 단락이 추구하였던 바의 달성도를 표시하였다. 셋째, 문장 논리는 술어논리(Predicate Logic) 형식을 따르며, 변수는 “유형, 주체, 객체, 가중치”로 결정하였고 문장 간의 관계를 표시하기 위하여 생크의

‘개념의존문법’을, 한국어의 절의 관계를 표시하기 위하여 용언어미 범주를 사용하였다. 이 3가지 접근은 다시 ‘주제’, ‘형식’, ‘환경조건’과 같은 3가지 표목들로 세분되었다.

2. 시맨틱 웹 환경

2.1 온톨로지

시맨틱 환경하에서는 사람이 문서를 인식하고 서로 소통하듯이, 기계도 문서를 인식하고 사람과 또는 기계들 끼리 상호 대화할 수 있게 된다. 사람이 문서에 쓰여진 자연언어들을 이해하듯이 기계가 문서의 자연언어 시스템을 이해하려면 현재의 웹 문서와 달리 상당한 언어적 지식을 정형적으로 갖고 있는 형식언어가 장착되어야 한다. 그리고 사람들 사이에서 이

종 언어와 방언에 따라 발생하는 의사 불소통과 같은 문제점을 제거해 주는 표준을 담당할 온톨로지가 있어야 한다.

온톨로지 안의 용어들은 ‘개체’, ‘개념(클래스)’, ‘속성’, ‘관계’로 이루어져 있는데, 시소러스와 마찬가지로 흔히 is_a와 part_of의 두 관계로 연결된다. 그러나 시소러스와는 달리 <표 1>처럼 여러 가지 규칙 기능들이 포함된다. 여기서 AGROVOC과 비교되는 가설적 온톨로지는 단순히 BT와 NT만으로 구성된 선언만 있는 것이 아니라 추론을 할 수 있는 규칙을 갖고 있다(<http://www.fao.org/docrep/008/af234e/af234e01.htm>).

다음은 위키피디아에서 정의한 온톨로지의 통상적인 구성소들이다.

- (1) 개체: 예 또는 객체(기본 수준)
- (2) 클래스: 집합, 모음, 개념, 객체 유형, 사물 종류

<표 1> AGROVOC의 시소러스와 가설적 온톨로지와의 비교

	AGROVOC 시소러스	가설적 온톨로지
선언	AGROVOC의 비차별 계층 구조 milk NT cow milk NT milk fat cow NT cow milk Cheddar cheese BT cow milk	온톨로지의 차별 구조 milk <includesSpecific> cow milk <containsSubstance> milk fat cow <hasComponent> cow milk* Cheddar cheese <madeFrom> cow milk
규칙	없음	Rule 1 Part X <mayContainSubstance> Substance Y IF Animal W <hasComponent> Part X AND Animal W <ingests> Substance Y Rule 2 Food Z <containsSubstance> Substance Y IF Food Z <madeFrom> Part X AND Part X <containsSubstance> Substance Y

- (3) 속성(에트리뷰트): 객체가 갖고 있는 관점, 성질, 자질, 특성, 계수
- (4) 연관: 클래스와 개체들이 상호 관련될 수 있는 방법
- (5) 복합관계(Function terms): 한 서술에서 개별 용어 대신에 사용될 수 있는 관계들로부터 형성되어지는 복합 구조
- (6) 제한: 입력되는 임의의 주장에 공식적으로 언급된 진실 사항
- (7) 규칙: 특정 형식의 주장으로부터 유도될 수 있는 논리적 추론을 기술하는 IF-THEN 형식문
- (8) 공리: 적용 영역 안에서 포괄적 의미의 이론을 말하는 논리 형식의 주장들(규칙 포함이고 공리적 문으로부터 이끌어져 나온 이론도 포함)
- (9) 사건: 속성과 관계의 변화 과정

2.2 마크업 언어

W3C에서 채택하고 있는 RDF/RDF-S는 레코드를 하나의 기술 단위로 취급해온 기존의 방식과는 달리, 자원(주어), 성질(술어), 성질

값(목적어)을 하나의 단위로 취급하는, 이른바 "Triple" 개념으로 사람과 기계가 동시에 이해할 수 있는 형태로 'rdfs:Class, rdfs:subClassOf'와 같이 클래스와 클래스 간의 관계, 'rdf:Property, rdfs:subPropertyOf'와 같이 속성과 속성간의 관계, 'rdfs:range, rdfs:domain, rdfs:Bag, rdfs:Seq'과 같이 메타데이터의 속성과 클래스 간의 관계 등을 표현할 수 있게 되어 있다.

RDF-S를 좀 더 확장하여 강화한 것이 OWL (Web Ontology Language)이다. OWL을 이용하면 임의의 어휘를 구성하는 용어의 의미와 용어들 간의 관계를 명시적으로 표현할 수 있다. OWL은 RDF, RDF-S보다 더 많은 의미 표현 수단을 제공하게 되는데, 시소러스의 동의, 계층, 연관 관계를 말하는 성질들은 <표 2>와 같이 바꾸어 표시될 수 있다(<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>).

SWRL(Semantic Web Rule Language)은 시맨틱 웹 사회가 규칙과 OWL의 통합을 위하여 최근에 제안한 DLP(Description Logic Programs)의 대항마로 W3C가 2004년 5월에 출시한 것으로, DLP가 혼 논리(Horn Logic)와 OWL의 교집합이라면 SWRL은 그것들의

<표 2> 온톨로지에서의 BT/NT, USE/UF, RT 성질

성 질	성질 또는 속성 명(코드)
USE/UF	owl:equivalentClass, owl:sameAs
BT/NT: is	owl:Class, rdfs:SubClassOf, owl:Thing, rdf:ID, rdf:type
BT/NT+RT: has	owl:ObjectProperty, DatatypeProperty, rdfs:subPropertyOf, owl:DatatypeProperty, rdfs:domain, rdfs:range, owl:Restriction, owl:onProperty, rdf:datatype, rdf:resource(주황색은 공통임)
RT	owl:allValueFrom, owl:someValueFrom, owl:Cardinality, owl:hasValue, owl:differentFrom, owl:distinctMembers, owl:intersectionOf, rdf:parseType, owl:unionOf, owl:complementOf, owl:oneOf, owl:disjointWith

합집합이다(wikipedia. <http://en.wikipedia.org/wiki/SWRL>).

〈그림 1〉과 같이 SWRL 파일 구조는 크게 변수 선언과 한번만 출현하는 헤드(규칙에서의 if part), 여러번 출현할 수 있는 바디(규칙에서의 then part)의 세 가지 부분으로 이해할 수 있다.

XML을 바탕으로 한 이러한 마크업 언어는 결과적으로 구조적 색인 및 탐색을 수행하게 되어 문서를 검색할 때에는 유리하지만 대신에 경로가 맞지 않으면 탐색 실패를 초래한다. 〈그림 1〉에서 〈swrl : ClassAtom〉은 〈swrl : Imp〉와 〈swrl : body〉 및 〈rdf : first〉를 거쳐야 만날 수 있다. 따라서 A 문서의 색인이 이와 같은 경로로 되어 있다면 탐색식도 〈swrl : Imp〉 → 〈swrl : body〉 → 〈rdf : first〉 → 〈swrl : ClassAtom〉의 순서로 작성되어야 A 문서가 탐색된다. 구조적 탐색에 맞는 매칭함수는 아래의 유사도 측정 〈공식 1〉과 같다(Manning, Raghavan and Schutze 2008, 190).

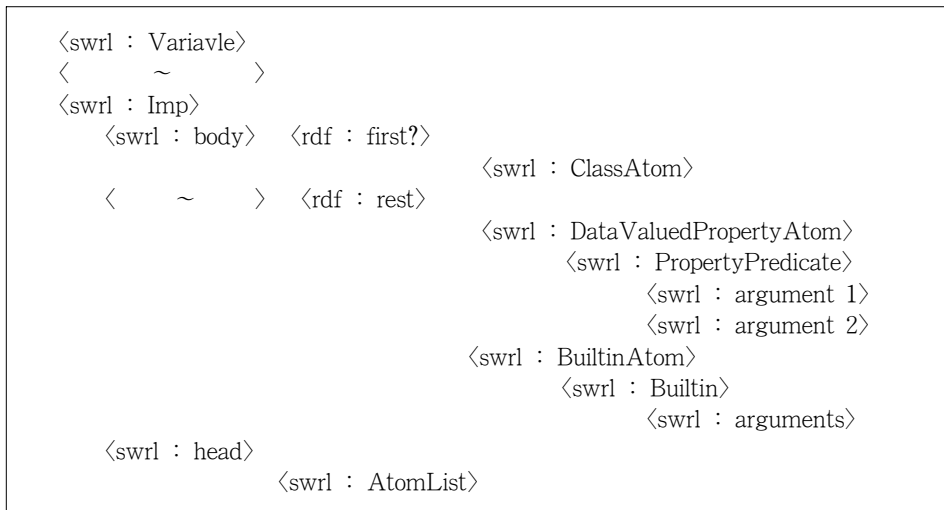
〈공식 1〉 유사합성도(q, d) =

$$\sum_{c_k \in B} \sum_{c_l \in B} C_R(c_k, c_l) \sum_{t \in V} \frac{weight(q, t, c_k) \cdot weight(d, t, c_l)}{\sqrt{\sum_{c \in B, t \in V} weight^2(d, t, c)}}$$

2.3 유한 상태기와 미분

문장은 단어로 이루어지는데 문장을 이루는 단어들은 품사, 구문적 역할, 의미 등 여러 가지로 분석되어 왔다. 입력되어지는 문자열에서 어떤 언어의 문장인가 아닌가를 결정하고자 할 때 유한 상태기(finite state machine) 모형을 사용한다. 결정론적(deterministic) 유한 상태기는 5종의 $\langle \Sigma, S, s_0, \delta, F \rangle$ 로 정의된다.

- (1) Σ 는 입력 문자열이고 유한하며 공집합을 배제한 부호 집합이다.
- (2) S 는 공집합을 배제한 유한 상태 집합이다.
- (3) s_0 는 S 의 요소이며 초기 상태를 나타낸다.
- (4) δ 는 상태 전이 함수를 말한다.
- (5) F 는 최종 상태 집합을 말하며 $F \subset S$ 이다.



〈그림 1〉 SWRL의 기본 구조(송용욱 외 2006)

이와 같은 유한 오토마톤(상태기)을

$$M = \langle \Sigma, S, s_0, \delta, F \rangle \dots(1)$$

로 나타낸다. M이 어느 상태 s_i 에 있을 때, 입력 a 가 주어지면 상태 s_j 로 이행했을 경우, 이를

$$\delta(s_i, a) = s_j$$

라 쓰고 상태전이함수라 한다. S 는 이 같은 함수집합이다. 한편 결정론적 유한 상태기에 비해 유한 상태 변환기(transducer)는 6중의 $\langle \Sigma, \Gamma, S, s_0, \delta, \omega \rangle$ 로 정의된다(http://en.wikipedia.org/wiki/Finite_state_machine 2009).

(1) Γ 는 출력 문자열이고 유한하며 공집합을 배제한 부호 집합이다.

(2) ω 는 출력 함수이다.

다음으로 언어의 미분(differentiation)이라는 개념을 생각하여 보자. 문자열 x, y, z 가 주어지고 $x = yz$ 이면 $D_y x = z$ 가 아래의 조건으로 성립하며 이 식은 x 를 y 로 미분한 결과가 z 임을 말한다.

$$D_y x = \begin{cases} z : x = yz \text{ 일 때} \\ \emptyset : x = yz \text{ 가 되는 } y, z \text{ 가 존재하지 않을 때} \end{cases}$$

$$\text{또 } D_\lambda x = x$$

$$D_x x = \lambda \text{ 이다.}$$

예를 들어

$$\left. \begin{aligned} D_a(abb) &= bb \\ D_{ab}(abb) &= b \\ D_c(ab) &= \emptyset \end{aligned} \right\}$$

이다(황도삼, 최기선, 김태석 2007, 54-55).

문자열 중, 의미가 있는 문자열(단어), 즉 언어 L 에 속하는 것을 a 로 규정하면 $a \in L$ 의 관계가

성립된다. 당연히 $L \subset S^*$ (알파벳으로 이루어진 모든 문자 집합)이다.

$$\alpha \gamma_1 \beta \in L$$

$$\alpha \gamma_2 \beta \in L$$

이 때 $\alpha \sim \beta$ 를 문맥이라고 부르자. \sim 의 부분에는 문자열 $\gamma_1, \gamma_2, \dots$ 가 들어간다. 이 경우 $\{\gamma_1, \gamma_2, \gamma_3, \dots\}$ 는 문맥 $\alpha \sim \beta$ 에서 같은 역할을 하는 것으로 보고, 이를 관계 $R_{\alpha\beta}$ 로 표현한다.

$$\gamma_1 R_{\alpha\beta} \gamma \Rightarrow \gamma_2 R_{\alpha\beta} \gamma_1 \dots(2)$$

$$\gamma_1 R_{\alpha\beta} \gamma_2, \gamma_2 R_{\alpha\beta} \gamma_3 \Rightarrow \gamma_1 R_{\alpha\beta} \gamma_3 \dots(3)$$

로 동치관계이다. 여기에서 $\alpha\beta$ 는 λ 이어도 좋다. $\alpha \sim \beta$ 의 그룹으로서, $\alpha\gamma_1\beta \in L$ 일수 있는 $\alpha \sim \beta$ 를 생각하고, 이 모든 것에 대해 $\alpha\gamma_2\beta \in L$ 이면, γ_1, γ_2 는 강한 의미에 있어 동의어라고 정의할 수 있다(황도삼, 최기선, 김태석 2007, 39).

3. 이야기 지식베이스 요소

3.1 이야기 구조 프레임

3.1.1 내용 구조

이야기들의 내용과 스타일은 시스템에 의해 저장되는 수사적 지식, 이야기 세계 지식, 그리고 상식적 지식을 포함한 지식의 유형에 의존한다. 페레즈와 샤펴(2004, 16)은 이야기 전개 과정에서 필수적으로 정립하여야 할 사항으로 이야기의 진행 순서, 연결성, 흥미, 신기함, 자연어 생성을 들었다. 그리고 김효영과 이순

형(2005)은 아동들에게 “문자적 이해(사실, 줄거리), 추론적 이해(감정추론, 내용추론, 경험연결), 공감적 이해(감정 공감)”에 대하여 조사하였으며, 한명숙(2003)은 <그림 2>와 같은 이야기 구조 구성 유형을 소개하였다.

자동 이야기 쓰기 시스템 중의 하나인 브루투스에서는 아래의 핵심과정으로 이야기를 전개하였다(Pérez and Sharples 2004, 17).

충족 자아	행복세계	기쁜관계 → 협력구조
		슬픈관계 → 배반구조
불행세계	행복세계	기쁜관계 → 해결구조
		슬픈관계 → 긴장구조
결핍 자아	행복세계	기쁜관계 → 극복구조
		슬픈관계 → 갈등구조
불행세계	행복세계	기쁜관계 → 화해구조
		슬픈관계 → 불화구조

<그림 2> 이야기 구조 구성 유형

- 주제 프레임의 예시화(실증화)
- 등장인물들이 미리 정의된 목표 집합(줄거리 전개)을 달성하려 시도하는 시뮬레이션 과정
- 최종 출력을 생산하는 이야기 문법의 확장 역시 자동 이야기 쓰기 시스템 중의 하나인 민스트렐은 PAT라고 알려진 아래의 6가지 미리 정의된 스키마 주제들로 이야기들을 전개하였다(Pérez and Sharples 2004, 19).

- 소개 장면:
 - 테마 장면: 세상-사실(전제조건)
 - 결정(목표를 성취할 계획)
 - 관계
 - 결말(계획실행의 결과)
- 대단원 장면:

3.1.2 전개 스키마

민스트렐의 창시자인 터너(1993, 45)는 테마 스키마를 “가치, 결정, 결과, 연결, 계획가, 계획 및 실제적 목표, 계획, 실세계 사건”으로 분석하였다. 그리고 저자수준(저자입장)에서 ‘목표(goal) 스키마에 포함시킬 하부유형(subtype)으로 “Tell-Story → Check-Story-For-Suspense → Check-Scene-For-Suspense → Add-Suspense-to-Scene → Check-Story-For-Tragedy → Check-Scene-For-Tragedy → Add-Tragedy-to-Scene → Check-Story-For-Foreshadowing → Add-Foreshadowing-to-Scene → Check-Story-For-Characterization → Add-Characterization → Check-New-Scene → Check-Consistency → Check-Preconds → Make-Goal-Consistent → Check-Affects → Add-Story-Intros → Add-Denouement → Connect → Instantiate”을 설정하였다.

손다이크(1977, 224-239)는 이야기 문법을 <그림 3>과 같이 만들었다. 그리고 브링스조르드와 페루치(2000, 153)는 브루투스에서 저자가 스토리를 만들어 가는 과정에 대해 “① 떠오른

번호	규칙
(1)	Story → Setting + Theme + Plot + Resolution
(2)	Setting → Characters + Location + Time
(3)	Theme → (Event)* + Goal
(4)	Plot → Episode*
(5)	Episode → Subgoal + Attempt* + Outcome
(6)	Attempt → {Event*, Episode}
(7)	Outcome → {Event*, State}
(8)	Resolution → {Event*, State}
(9)	{Subgoal, Goal} → Desired State
(10)	{Characters, Location, Time} → State

<그림 3> 손다이크의 이야기 문법

창조적 이야기를 발전 → ② 독자의 마음에 형상을 생성 → ③ 의식적 조망으로 이야기 꾸밈 → ④ 순수문학적 픽션의 중심에서 개념들을 계산 → ⑤ 진정으로 흥미있는 이야기를 생산 → ⑥ 깊고 영구적인 이야기 구조로 두드려 만들 → ⑦ 기계적인 산문 회피”와 같이 7단계로 요약하였다.

3.1.3 프레임 구조

본 논문의 프레임은 위의 한명숙과 <그림 3>을 중심축으로 하여 <표 3>과 같이 슬롯(slot)과 패싯(facet)으로 작성하였고, 개념 표현이 잘 안되거나 본인의 표현 능력 밖에 있는 사안들을 위해 비슷한 상황의 그림을 택하도록 하였다.

3.2 단락 스크립트

3.2.1 단락의 성질

단락이란 단일한 논제를 취급하는 관련된 문장들의 총합이다. 가능한 한 효과적이기 위해서는 단락은 단일성(통일성), 연결성, 주제문, 그리고 적절한 전개를 포함하여야 한다. <표 4>는 퍼듀대학의 OWL(Online Writing Lab)에서 권장한 사항들이다(<http://owl.english.purdue.edu/owl/resource/606/01/>).

이삼형(1994, 63-90)은 단락의 문장들 간의 관계의 의미를 ① 수집, ② 부가, ③ 공제, ④ 인과, ⑤ 이유(과인), ⑥ 비교/대조, ⑦ 상세화, ⑧ 문제/해결, ⑨ 초담화와 같이 설정하였다.

<표 3> 이야기 검색을 위한 이야기 프레임

a) 구조 형식

제1슬롯	제2슬롯	패싯1	패싯2		그림&소리(0)
배경	등장인물	남/녀(유아, 아동, 노인)	그림&소리(011)	그림&소리(01)	
	장소	산, 바다, 강, 도시, 집 등	그림&소리(012)		
	시간	1990s, 21c 등			
테마	사건	연애, 사고, 공부, 살인 등	그림&소리(021)	그림&소리(02)	
	목표	권선징악, 훈육, 담술 등	그림&소리(022)		
줄거리/에피소드들	하부목표	편지 등	그림&소리(03)		
	시도	배달 등			
	결과	답장 등			
해법	사건	연애, 사고, 공부, 살인 등	그림&소리(04)		
	상태	결혼, 우정, 합격, 사형 등			

b) 주제

제1슬롯	제2슬롯	제3슬롯	제1슬롯	제2슬롯	제3슬롯
충족자아	행복-기쁜	협력구조	결핍자아	행복-기쁜	극복구조
	행복-슬픈	배반구조		행복-슬픈	갈등구조
	불행-기쁜	해결구조		불행-기쁜	화해구조
	불행-슬픈	긴장구조		불행-슬픈	불화구조

3.2.2 단락의 하부 유형
이야기를 구성하는 단락들에는 저마다 이야기
가 부여한 역할들이 있다. 예를 들면 시대적 배경
을 설명하거나, 장소를 묘사하고, 심리 상태를 묘

사하거나, 감정을 읽어내는 부분들이 같은 목표
하에 응집성과 연결성을 고려하면서 단락으로
묶여 표현되곤 한다. 이러한 목표들과 활동 하부
유형(subtype)을 보면 <표 5, 6, 7>과 같다.

<표 4> 단락이 갖추어야 할 성질

		내역
단일성		전 단락에 걸쳐 단일한 초점이 모아져야 한다.
연결성	논리의 다리	① 한 논제의 유사 아이디어가 문장과 문장 사이를 누비게 한다. ② 뒷 문장(이어지는 문장)이 앞 문장과 같은 문장 형식으로 만들어질 수 있다.
	말의 다리	① 키워드가 여러 개의 문장에서 반복하여 쓰일 수 있다. ② 이음동어가 여러 문장에 걸쳐 반복될 수 있다. ③ 대명사는 이전에 출현한 문장들을 알아보도록(조회) 할 수 있다. ④ 전이 단어는 아이디어를 다른 문장에 링크시킬 수 있다.
주제 문장		주제 문장은 단락이 취급할 아이디어나 논점을 일반적인 방법으로 알리는 문장이다.
적정 전개		잘 전개되는지의 여부를 다음 사항으로 판단한다. ① 예와 삽도를 이용한다. ② 자료를 인용한다(사실, 통계, 증거, 세목 등). ③ 증언을 실험한다(인용, 의역과 같이 다른 이의 말). ④ 일화 또는 이야기를 사용한다. ⑤ 단락 안에서 용어를 정의한다. ⑥ 비교와 대조를 활용한다. ⑦ 원인과 이유를 평가한다. ⑧ 효과와 결과를 검토한다. ⑨ 논제를 분석한다. ⑩ 논제를 기술한다. ⑪ 사건의 연대기(시간 단위)를 제공한다.

<표 5> 생크와 애벌슨의 목표 분류(Turner 1993, 46)

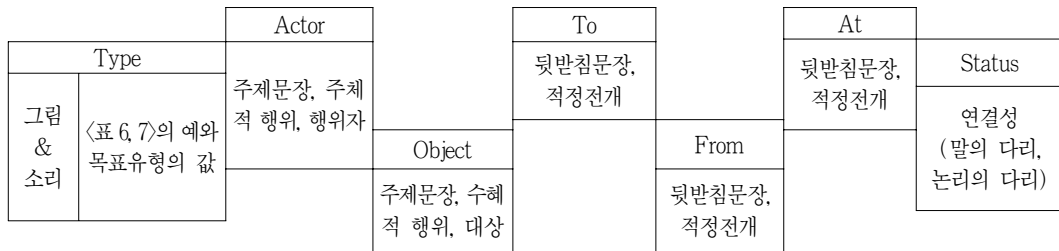
목표유형	내용기술	예
Satisfaction Goal(S-Goal)	Recurring bodily desires	S-Hunger, S-Sleep
Delta Goal(D-Goal)	Desired State Change	D-Location, D-Control
Enjoyment Goal(E-Goal)	Pleasurable activities	E-Travel, E-Entertainment
Achievement Goal(A-Goal)	Long-term attainment of social status	A-Skill, A-Status
Preservation Goal(P-Goal)	Activated when status is threatened	P-Health, P-Possessions
Crisis Goal(C-Goal)	Active preservation goals	C-Health, C-Possessions

<표 6> MINSTREL 등장인물 수준의 목표 유형(Turner 1993, 47)

클래스	목표유형	내용기술
S-Goal	S-Hunge	Satisfy one's hunger.
D-Goal	D-Control	Control something.
	D-Loc	Change location.
	Destroy	Destroy some object of person.
	Scare	Cause fear in someone.
A-Goal	A-Status	Achieve status in society.
	A-Love	Find romantic love.
	A-Affection	Find brotherly affection.
C-Goa	C-Health	Health crisis.
Meta Goals	Favor	Do a favor for someone.
	Anti-Favor	Cause the failure of someone's goals.
	Retract	Retract a goal from being active.
	Deception	Cause someone to be deceived about something.

〈표 7〉 활동 스키마(Turner 1993, 48)

슬롯	의미
Type	Identifies the particular sub-type of action
Actor	Identifies the actor who performs the action
Object	The thing or person being acted upon
To	The direction the action is taking
From	The origin of the action
At	Where the action occurs
Status	Whether the actor succeeded in performing the action



〈그림 4〉 단락 검색을 위한 단락 스크립트

3.2.3 스크립트 구조

본고에서 설정한 단락의 스크립트는 상기한 〈표 5, 6, 7〉을 중심으로 하고, 앞의 3.2.1절에 서술한 내용을 참조하여 〈그림 4〉와 같이 작성하였다. ‘Type(유형)’에는 〈표 6, 7〉의 ‘예’와 ‘목표 유형’의 값이 그대로 항목 값으로 적용되었다. ‘Actor(행위자)’에는 단락의 주제문장과 주제문장의 주어가 오며 주제문장이 모호할 때는 단락 전체를 이끄는 주체가 기록된다. ‘Object(대상)’에도 마찬가지로 주제문장의 목적어가 오며 주제문장이 모호할 경우 단락 전체에 작용하는 객체를 적는다. ‘행위자’와 ‘대상’은 단락의 단일성을 해치지 않아야 한다. 시간과 장소 및 대상 등을 나타내는 ‘To, From, At’은 주제문장, 또는 주제문장을 뒷받침하는 문장들에서 〈표 4〉의 ‘적정전개’에 나열된 항목들을 최대한 수용

하는 해당 문장이나 구절을 선택하였다. ‘Status(현상)’는 “행위자, 대상, To, From, At”으로 선정된 문장과 구절들이 유지하는 연결성(다리 역할) 정도를 수치로 환산한 값을 찾아 대입하였다. 그리고 이야기 프레임과 마찬가지로 그림과 소리를 수록하였다.

3.3 문장 논리

3.3.1 문장 정제 과정

퍼듀대학 홈페이지(<http://owl.english.purdue.edu/>)에 있는 글쓰기 콘텐츠를 살펴보면 문장 작성할 때 유의해야 할 사항으로, (1) 연결성(Coherence)(http://owl.english.purdue.edu/handouts/general/gl_cohere.html), (2) 간결성(Conciseness)(<http://owl.english.purdue.edu/>)

e.edu/handouts/general/gl_concise.html), (3) 문장 명쾌성(http://owl.english.purdue.edu/handouts/general/gl_sentclar.html), (4) 은유를 사용한 창조(http://owl.english.purdue.edu/handouts/general/gl_metaphor.html)를 들고 있다. 터너(1993, 42)는 저자수준의 목표스키마(Goal Schema)의 슬롯으로 “유형(Type), 행위자(Actor), 배경/무대(Object), 순위(Priority)”를 설정하였다. 여기서 ‘유형’은 목표의 특정한 하부유형을 아우르고 ‘행위자’는 목표를 주체하는 행위자를 말한다. ‘배경/무대’는 목표가 적용되는 이야기의 장면이나 개념이며, ‘순위’는 목표의 중요성을 지적하는 0-100까지의 숫자이다. 문장의 정제는 퍼듀대학의 네 가지

유의사항과 터너의 네 가지 슬롯을 바탕으로 진행되었다.

문장들을 정제하는 과정은 2.3장에서 살펴본 형식언어를 사용하여 <그림 5>에 일곱 항목으로 기록하였다. 그것을 간추리면 다음과 같다. (1) 문장에서 핵심의미 파악에 별로 도움이 되지 않는 품사와 어구들을 구별하여 미분화시킨다. (2) 반복 출현하여 독자적 단어를 형성하고 그 의미를 강화시키는 복합어들을 정규언어로 표현하여 색인 형식언어를 만든다. (3) 대명사의 유무에 상관없이 앞뒤 문장의 의미 연결을 정규언어화한다. (4) 동의어와 은유적 표현을 표준화한다.

(1) 기능어(감탄사, 부사, 관형사, 대명사, 접속사 등)를 제거하는 문제 → 단어 미분화	
(2) 같은 문자의 반복 또는 몇몇 문자로 여러 단어를 운율적으로 조합하는 문제 → 정규언어화 $\{ a^n \ n=0,1,2,\dots \} : (\lambda, a, aa, aaa, \dots)$ 예) 정*당*하게 → 정규문법 $\alpha \subset L \ (L \subset S^*)$ 에서 “정당하게, 정정당당하게”가 생성 $\{ (ab)^n \} : (\lambda, ab, abab, ababab, \dots)$ 예) 다사* → 정규문법 $\alpha \subset L \ (L \subset S^*)$ 에서 “다사, 다사다사”가 생성 $\{ (a+b)^* \} : (\lambda, (a+b), (aa+ab+ba+bb), (aaa+aab+\dots), \dots)$ 예) 다사* → 정규문법 $\alpha \subset L \ (L \subset S^*)$ 에서 “다사, 다사다사”가 생성 $\{ (a+b)c(b(a+b)c)^* a \} \rightarrow (ac\lambda a, bc\lambda a, acbaca, acbbca, \dots)$ 예) (가+오)려(오(가+오)려)*가 → 정규문법 $\alpha \subset L \ (L \subset S^*)$ 에서 “가려가, 오려가, 가려오 가려가, 가려오오려가”가 생성	
(3) <표준, 최신, 현대>국어와 같이 저의미 수식어구들의 처리 문제 → ‘국어’로 의미 미분화	
(4) ① 이음동의어(약어, 두자어 포함), ② 동음이의어, ③ 연관어들을 포함하여 문장의 뜻이 표준성을 나타내게 하는 문제 → 의미 정규언어화 ↔ 색인어 확장, 상위 색인어, 상위 분류기호 적용	
(5) 대명사가 가리키는 전위 문장의 뜻을 받는 문제 → 정규언어화	
(6) 문맥 상에서 앞 뒤 문장의 뜻을 포함하여야 하는 문제 → 정규언어화 정보 -- 시스템 → 색인, 검색, 탐색, 목록/자료 야생 -- 보호대책 → 동물, 포유류, 사자 (1) $\gamma =$ 검색, $\gamma_1 =$ 색인, $\gamma_2 =$ 탐색 (1) $\gamma =$ 검색, $\gamma_1 =$ 색인, $\gamma_2 =$ 탐색 (2) $\gamma_1 =$ 색인, $\gamma_2 =$ 탐색, $\gamma_3 =$ 목록/자료 (2) $\gamma_1 =$ 동물, $\gamma_2 =$ 포유류, $\gamma_3 =$ 사자 $\gamma_1 R_{\alpha\beta} \gamma \Rightarrow \gamma_2 R_{\alpha\beta} \gamma_1 \dots(1)$ $\gamma_1 R_{\alpha\beta} \gamma_2, \gamma_2 R_{\alpha\beta} \gamma_3 \Rightarrow \gamma_1 R_{\alpha\beta} \gamma_3 \dots(2)$	
(7) 은유적 표현에 대한 표준화 → 정규언어화 ← 온톨로지(이야기 쓰기용)	

<그림 5> 문장 정제 과정

3.3.2 논리 구조

상기한 내용을 참조하여 설정된 문장 논리는 <그림 6>과 같다. 단락에서는 ‘행위자, 대상 등’이 ‘유형’과 동격이었지만 문장에서는 “행위자, 배경/무대 등”은 ‘유형’에 하부속성으로 종속시켰다. 여기서 ‘유형’의 패킷은 생크(1975: 1977)의 용언범주(1)를 모체로 하여 “ATRANS(생각/사고/연상하다, 이전하다), ATTEND(집중하다, 주목하다), GRASP(소유하다, 보유하다, 갖다), EXPEL(출력하다, 배출하다, 분실하다), INGEST(입력하다, 먹다, 습득하다), MBUILD(작성하다), MTRANS(전송하다, 강의하다, 공부하다, 미치다, 정화하다), MOVE(움직이다, 오다/가다), PROPEL(추진하다), PTRANS(운송하다, 옮기다, 이사하다), M-Fight(싸우다, 전쟁하다), SPEAK(말하다), PBUILD(세우다, 꽃다, 짓다, 만들다)”와 같이 13개 범주를 마련하였다.

‘행위자’는 문장에서 주체 역할을 하는 어구, ‘배경/무대’는 객체 역할을 하는 어구를 대입한다. ‘순위’는 연결성(“이음동어의, 용어 반복, 대명사, 전이어”를 감안), 간결성(“한정사, 수식어구, 불필요 절/부정사구, 허사, 완곡한 표현, 과도한 세부사항”의 배제), 명쾌성(“명사 용어열, 동사의 명사형 단어”의 회피), 은유적 표현을 분석한 후 해당되는 값을 부여한다. 그리고 문장 논리에도 이야기와 단락에서처럼 그림과 소리를 채택하

고 특징 있는 용언구와 단어길이 및 연결어미(복문의 경우)를 등재하였다.

4. 이야기 지식베이스 설계

4.1 이야기 온톨로지

이야기 쓰기에 관련한 이야기 구조, 단락, 문장을 제공하는 검색 시스템에는 이야기 쓰기 전반을 말하여 주는 편람적 지식이 필요하다. 왜냐하면 어떤 지식이 어떤 경우에 일반적으로 적용되는지 구체적으로 알 필요가 있기 때문이다. 이러한 지식을 담고 있는 것을 온톨로지라고 하자. 따라서 이야기 쓰기 도우미용 온톨로지는 이야기 쓰기에 필요한 제반 도구들을 포함하고 있어야 한다. 본 시스템에서는 (1) 주제어들 사이의 동등, 계층, 연관을 밝히는 어휘관계, (2) 명사와 용언 범주와 문장 생성을 위한 범주 간의 결합 정보, (3) 용언어미와 문장에서 절과 절 사이의 연결정보, (4) 주제어와 결연하는 이야기 글, 단락, 문장들에 대한 링크 정보, (5) 주제어의 가중치와 최신성을 나타내는 정보, (6) 그림, 소리 및 글/단락/문장 모형을 장치하였다.

Predicate Type: (ATRANS, ATTEND, GRASP, EXPEL, INGEST, MBUILD, MTRANS, MOVE, PROPEL, PTRANS, SPEAK)(그림&소리)		
Actor	Predicate Type	Object
연결어미, 단어길이, 특징적 용언구, 순위 (연결성, 간결성, 명쾌성, 은유표현)		

<그림 6> 문장 검색을 위한 문장 논리

1) 생크는 개념의준문법에서 용언을 “ATRANS, ATTEND, GRASP, EXPEL, INGEST, MBUILD, MTRANS, MOVE, PROPEL, PTRANS, SPEAK”의 11가지 개념 범주로 분리하였다.

4.1.1 어휘와 문맥 코드

(1) 계층과 연관 관계 정보: 명사의 '종-속'과 부분-전체를 알려주는 상위어와 하위어의 계보를 인식할 수 있게 각각의 'Class'나 'ObjectProperty'의 관계 규칙 정보를 다음의 XML 코드와 같이 작성하였다.

```
<owl:ObjectProperty rdf:ID="BT">
  <rdf:type rdf:resource="&owl:TransitiveProperty" />
  <rdfs:domain rdf:resource="&owl:Thing" />
  <rdfs:range rdf:resource="#Region" />
</owl:ObjectProperty>

<Region rdf:ID="SantaCruzMountainsRegion">
  <BT rdf:resource="#CaliforniaRegion" />
</Region>

<Region rdf:ID="CaliforniaRegion">
  <BT rdf:resource="#USRegion" />
</Region>
```

(2) 동등관계 정보: 명사의 'Class'나 'ObjectProperty'의 동등관계 정보를 역시 XML 코드로 다음과 같이 작성하였으며 마지막 예는 비유되는 단어들의 관계를 'swrl:reference' 속명을 개설하여 표시하였다.

```
<owl:Class rdf:ID="Wine">
  <owl:equivalentClass rdf:resource="&vin:Wine"/>
</owl:Class>

<owl:Class rdf:ID="TexasThings">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#locatedIn" />
      <owl:someValuesFrom rdf:resource="#TexasRegion" />
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

<Wine rdf:ID="MikesFavoriteWine">
  <owl:sameAs rdf:resource="#StGenevieveTexasWhite" />
</Wine>
```

```

<owl:Class rdf:ID="검은 머리">
  <swrl:reference rdf:resource="http://www.mirohan.com/compare/view.xml">
    (owl:sameAs rdf:resource="#검은 구름" />
  </swrl:reference>
</Wine>
    
```

(3) 명사 최고범주 정보: 동종의 자질 정보가 작용되는 명사들의 집합을 나타내는 용어로서 <그림 7>의 [명사]에 그 범주가 있으며 각 명사들에게 최고의 상위어로 작용한다. 그것들을 XML 코드로 “종속, 부분전체” 관계를 표현하면 다음과 같이 작성되었다.

[명사]: 동작, 동물, 인공물, 속성, 신체부위, 인지, 커뮤니케이션, 사건, 감정, 음식, 집합, 위치, 동기, 자연물, 자연현상, 인물, 식물, 소유, 과정, 수량, 관계, 형상, 상태, 물질, 시간
 [용언]: 신체 기능과 치료, 변화, 커뮤니케이션, 경쟁, 소비, 접촉, 인지, 창조, 동작, 감정/심리, 상태, 지각, 소유, 사회 상호작용, 날씨

<그림 7> WordNet 범주어 예(김영택 2001, 146-147)

```

<owl:Class picture="Pic01.htm" Sound="Sou01.htm" rdf:ID="물질">
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing picture="Pic011.htm" Sound="Sou011.htm" rdf:about="#식물"/>
    <owl:Thing picture="Pic012.htm" Sound="Sou012.htm" rdf:about="#동물"/>
    <owl:Thing picture="Pic013.htm" Sound="Sou013.htm" rdf:about="#매체"/>
    <owl:Thing picture="Pic014.htm" Sound="Sou014.htm" rdf:about="#인공물"/>
    <owl:Thing picture="Pic015.htm" Sound="Sou015.htm" rdf:about="#자연물"/>
    <owl:Thing picture="Pic016.htm" Sound="Sou016.htm" rdf:about="#음식"/>
    <owl:Thing picture="Pic017.htm" Sound="Sou017.htm" rdf:about="#인물"/>
  </owl:oneOf>
</owl:Class>
    
```

(4) 용언 최고범주 정보: 동종의 자질 정보가 작용되는 용언들의 집합을 나타내는 범주 정보로 <그림 7>의 [용언]에 그 범주가 있으며 각 용언들의 최고 상위어로 작용한다. 또한 각 용언들에는 그것들과 어울리는 그림과 소리들이 단어 회상에 대비하여 참조되었으며 XML 코드로 나타내면 다음과 같다.

```

<owl:Class picture="Pic02.htm" Sound="Sou02.htm" rdf:ID="동작">
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing picture="Pic021.htm" Sound="Sou021.htm" rdf:about="#이동"/>
    <owl:Thing picture="Pic022.htm" Sound="Sou022.htm" rdf:about="#작용"/>
  </owl:oneOf>
</owl:Class>
    
```

```

<owl:Thing picture="Pic023.htm" Sound="Sou023.htm" rdf:about="#지각"/>
<owl:Thing picture="Pic024.htm" Sound="Sou024.htm" rdf:about="#통신"/>
<owl:Thing picture="Pic025.htm" Sound="Sou025.htm" rdf:about="#창조"/>

</owl:oneOf>
</owl:Class>

```

(5) 단어 결합 정보: “명-명, 명-용, 주-목적어, 주-보어, 목적어-보어”의 결합을 규명하는 명사와 용언의 범주들을 찾아 작성하였다. 명사 범주를 ‘NounTopCat’, 용언 범주를 ‘VerbTopCat’, 명사가 문장에서 주체적인 구실을 할 때, ‘Subjectives’, 목적어나 보어의 구실을 할 때, ‘Objectives’로 명명하고, ‘개가 책을 읽다’와 같이 결합의 오류를 발생하는 경우를 바로잡는 <그림 5>의 ‘문장 정제 과정’들을 다음과 같이 연결하였다.

```

(가) <owl:ObjectProperty rdf:ID="VerbTopCat">
    <rdfs:domain rdf:resource="#NounTopCat">
        <rdfs:restriction rdf:resource="#NounTopCat"/>
    </rdfs:domain>
    <rdfs:rule rdf:resource="#문장정제규칙01"/>
</owl:ObjectProperty>
(나) <owl:ObjectProperty rdf:ID="VerbTopCat">
    <rdfs:range rdf:resource="#NounTopCat">
        <rdfs:restriction rdf:resource="#NounTopCat"/>
    </rdfs:range>
    <rdfs:rule rdf:resource="#문장정제규칙02"/>
</owl:ObjectProperty>
(다) <owl:ObjectProperty rdf:ID="VerbTopCat">
    <rdfs:domain rdf:resource="#Subjectives"/>
    <rdfs:range rdf:resource="#Objectives">
        <rdfs:restriction rdf:resource="#NounTopCat"/>
    </rdfs:range>
    <rdfs:rule rdf:resource="#문장정제규칙03"/>
</owl:ObjectProperty>

```

(6) 절 결합 정보: 절과 절 사이의 용언어미들의 결합 관계를 표현하기 위하여 ‘Class’나 ‘Object Property’에 버금가는 ‘Synthesis’를 개설하고 속성에 ‘subject’ 주어와 ‘object’ 목적어를 포함하여 절과 절들의 결합 정보를 처리하였다. 실제로 연결 정보는 <표 8>의 연결코드를 사용하여 다음의 XML 코드처럼 인물을 주어로 한 부절과 또 인물을 목적어로 한 주절은 합쳐져 하나의 문장을 만들 수도 있음을 알려 주고 있다.

〈표 8〉 용언어미의 종류

종류	연결 어미	종류	연결 어미
+가정	-면, -거든, -(ㄷ)아도, -(ㄱ)어도, -라도	+과상	들고 있었다(과진)/들려 있었다(과상)
+과정1	-여, -여서, -해, -해서, <동사의 경우>: -ㄱ(서), -ㅏ(서) -어(서) -아(서)	+과정2	-ㄴ(는)데, -다가, -(ㄴ)는 중에
+과정3	-게, -진 채, -데다	+나열1	-고
+나열2	'고'로 대체될 수 있는 것: -며, -고서, -하는가 하면, -면서	+대립	-지만, -거늘, -되, -(으)나, -는데 불구하고, -했으나, -뿐만 아니라, -ㄴ 반면
+목적	-려고, -르, -러, -도록, -라, -도록, -려면	+반반	-지~지, -던(든)지~던(든)지
+불문	-(이)든(지) 간에, -하더라도	+비교	-보다, -ㄴ(는) 것 보다, -ㄴ 만큼, -라기 보다는
+삽입	-ㄱ(이, 을), -음(이, 을), -기(가, 는, 를, 로), -느냐가, -는 것(이, 은 등), -(을)듯, -기 만큼, -(하, 라)고(는), -라는	+상태	-ㄴ 채(로, 로는 등), -는 데(에, 에 있어 등), -ㄴ(는) 거(것이)-, -ㄴ 결과, -ㄴ(이, 리), -듯
+선택	-거나, -든지, -는가	+수식	-ㄴ, -는, -던
+시간	-기 전, -ㄴ(는) 후, -ㄴ 때(까지), -에 앞서	+염려	-까봐
+원인1	-(으)니까, -기 때문에, -(으)므로, -기에, 있는 만큼, -함으로써, -ㄴ 까닭에	+원인2	-(다)니, -자, <형용사의 경우>: -어(서)(는), -아(서)(는), -ㄱ(서)(는), -ㅏ(서)(는), -더니
+의문	-면서도, -고서도, -ㄴ 뿐더러, 있거나 한지	+전제	-아야, -어야, -(는)다고, -해야(만), -건만, -ㄴ 망정, -는 한
+종결	-다, -까, -오, -세, -게나	+현상	들고 있다(현진), 들려 있다(현상)
+동의	-가 까지	+동조	-하기도, -도
+반박	있지 않은, -하지 않은	+누가	-ㄴ수록
+추측	-할(수, 것, 지), -르, -ㄴ 리	+수사	-던 것

```

<owl:Synthesis rdf:ID="Sentence">
  <owl:intersectionOf>
    <owl:Class subject="human" object="" rdf:resource="#beforeClause[나열1]" />
    <owl:Class subject="" object="human" rdf:resource="#afterClause[종결]" />
  </owl:intersectionOf>
</owl:Synthesis>
    
```

4.1.2 결연과 가중치 코드

‘결연’은 관용어, 사자성어 또는 그림/소리를 대변하는 장르로 미리 기록되어 있는 단락이나 문장들을 발췌하여 이용하는 방법을 말한다. 예를 들면 아래의 ASP.NET 언어 코드처럼 그림이나 소리를 연결하여 마음에 들면 그리로 클릭하여 글 정보를 얻는다. 그 글 정보의 가중치는 0.8로 정한다.

```

public void pictureButton_Click(Object Sender, EventArgs E) {
    string picTure = Request.Form["imapicture"];
    if (picTure == "PIC1000") {
        Response.Write("Thank You" + "<IMG SRC='./pictureDB/ball106.gif'>");
        Response.Write("<html>");
        Response.Write("<body>");
        Response.Write("<A HREF='./storyDB/story001.htm'><IMG SRC='./pictureDB/D.gif'>
            0,8</A>");
        Response.Write("</body>");
        Response.Write("</html>");
    }
}

```

4.2 색인화 지식베이스

4.2.1 글 지식베이스

이야기 글 전체를 나타내는 지식베이스를 만들기 위하여 앞 장에서 분석된 글 지식을 토대로 <표 9>와 같은 색인언어를 마련하였다. 여기에 그림도 색인 표목에 오르도록 하였다. 색인 그림은 그림으로 주제 어휘들을 대신하는 것으로 예를 들면 “가을 단풍이 든 산”을 나타낸 이미지는 ‘가을 산’이라는 색인어와 같은 역할을 할 수 있으며 나아가 ‘가을 산’보다 더 많은 내포 의미를 이용자에게 줄 수 있는 것이다. 따라서 원 글에 단어, 그림, 소리로 이루어진 색인언어를 부가시켜 다음의 <글 지식베이스 예>처럼 XML 파일로 지식베이스를 구축하였다. 이 글 지식베이스에 대한 xsd:schema 부분은 <부록 1>에 첨부하였다.

<표 9> 글 색인 표목

접근점	내역
글 장르	역사, 소년/소녀, 애정, SF, 환타지, 무협 등
소재	<그림 7>의 키워드로
비언어	그림과 소리, 그림과 소리의 주제명은 <표 3>의 b)의 제3슬롯, <표 5>와 <표 6>의 ‘목표유형’, <그림 6>과 <그림 7>의 키워드로
주제	<표 3>의 b)에서 ‘제3슬롯’을 사용
문체	정서, 묘사, 논술, 대화 등으로
의미량	1p 이하, 2p 이하, 3p 이하로
독서 수준	대학생, 중고등학생, 초등학생, 유아용, 성인용, 노인용
정보 질	상, 중, 하로

<글 지식베이스 예-XML 문서 부분>

```

<?xml version="1.0" encoding="UTF-8"?>
<stories xmlns="x-schema:Mirohan4.4.xdr">

```

```

<story author="구연희" title="검은구름" goalsubject="협력" subjectmatter="인물"
      id="0001">
  <background:character> 이모 와 조카 </background:character>
  <background:location> 집 </background:location>
  <background:time> 봄 과 여름 </background:time>
  <theme:event>검은 머리 와 검은 구름</theme:event>
  <theme:goal>검은 머리가 검은 구름 </theme:goal>
  <plot literarystyle="공트" sentencestyle="설명문">
    <episode quantity="2" quality="중" level="전체" subgoal="협력"
      attempt="연상" result="동심 세계">
      </episode>
  </plot>
  <resolution:event>이모 머리감는 것에 대한 조카 생각</resolution:event>
  <resolution:state>동심의 세계</resolution:state>
  <picture> </picture>
  <sound> </sound>
</story>
</stories>

```

글 지식베이스에서 채택한 접근점들은 표본분석 결과, 모든 표본 글에서 95% 이상 출현한 구성소들을 선정하였다.

4.2.2 단락 지식베이스

단락을 대표하는 단락 지식베이스는 <표 10>과 같은 색인언어로 정립하고 여기에 단락 원문을 부가하여 쉽게 단락 문장들을 참조할 수 있게 만들었다. 단락에서 설정한 3.2.3절의 접근점 중 ‘행위자’와 ‘대상’은 서로 관련이 깊은 점을, ‘from’, ‘to’, ‘at’은 출현빈도가 95%에 못 미치나 서로 합하면 100%에 이르고, 또 서로 관련이 있는 점을 들어 다음의 <단락 지식베이스 예>처럼 합성된 접근점을 만들었다. ‘현상’은 수량화된 색인언어이므로 대표격인 ‘유형’에 합쳐 표시한다.

<표 10> 단락 색인 표목

접근점	내역
단락위치	<표 3>의 a)에서 ‘제1슬롯’과 ‘제2슬롯’을 사용
소주제	<표 9>의 소재와 동일
비언어	<표 9>의 비언어와 동일
하부목표	‘주어, 목적어’의 명사와 ‘보어’의 ‘부터-로-에’에 있는 명사
단락형식	이삼형의 분류대로
문체	정서, 묘사, 논술, 대화 등으로
의미량	50words 이하, 100words 이하, 150words 이하/이상으로
독서수준	<표 9>의 독서수준과 동일
정보질	<표 9>의 정보질과 동일

<단락 지식베이스 예(1, 2차 탐색 부분)>

```
<?xml version="1.0" encoding="UTF-8"?>
<paragraphes>
  <paragraph subjectmatter="신체부위" paragraphform="과인" sentencestyle="설명"
    quantity="100"location="배경">
    <type-status>머리감기 </type-status> 90
    <actor-object>조카-이모 </actor-object>
    <from-to-at>검은구름-비-까만 머리 </from-to-at>
    <paratext>잔뜩 비누칠을 한 머리를 문질러 거품을 내고 있는데 .....</paratext>
  </paragraph>
</paragraphes>
```

4.2.3 문장 지식베이스

문장을 검색하기 위한 색인언어는 <표 11>과 같이 설정되었고 이것을 토대로 문장 검색용 문장 지식베이스를 다음의 예와 같이 구축하였다.

<표 11> 문장 색인 표목

접근점	내역
술어범주	<그림 6>의 술어 유형을 사용
소재영역	이동, 병원, 여행(호텔, 기차 등)
주체/객체	사람, 기계, 동물
비언어	<표 9>의 비언어와 동일
절정보	<표 8>의 연결어미 코드 사용
문체	정서, 묘사, 논술, 대화 등으로
의미량	① 23-27단어, ② 23-27 이상, ③ 23-27 이하
독서수준	<표 9>의 독서수준과 동일
정보질	<표 9>의 정보 질과 동일

<문장 지식베이스 예(1, 2차 탐색 부분)>

```
<?xml version="1.0" encoding="UTF-8"?>
<sentences>
  <sentence predicatetype="ATRANS" actor="머리" object="검은구름"
    senstyle="묘사">
    비늘 끝만 살짝 갖다 대도 짝하고 금이 갈 것만 같은 하늘의 푸름이.....
    <clause connectivecode="나열1" connectivesuffix="고" />
  </sentence>
</sentences>
```

5. 검색과 평가

5.1 검색엔진과 인터페이스

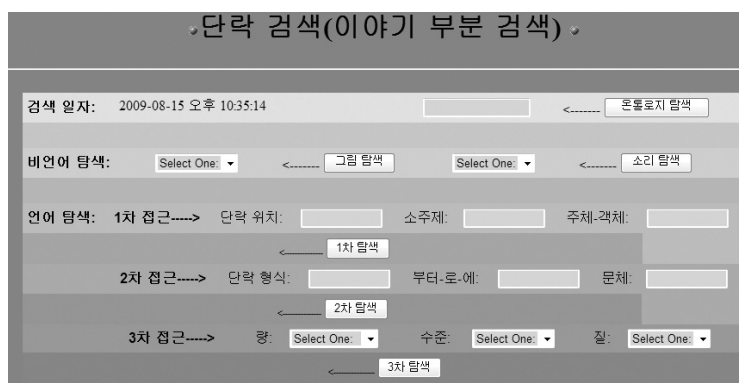
이용자들이 XML 파일로 작성된 색인과 원문을 수록하고 있는 이야기 지식베이스에 접근할 수 있도록 ASP.NET 프로그램 언어로 글, 단락, 문장 검색엔진과 인터페이스를 구축하였다. 검색엔진과 인터페이스는 4단계의 경로를 거치는 노드(nodes)와 애트리뷰트(attributes) 및 엘리먼트(elements)들로 작성되었다(그림

8, 9, 10 참조).

글, 단락, 문장을 검색하는 각각의 엔진에는 공통적으로 온톨로지와 그림 및 소리를 탐색할 수 있도록 해당 입력창과 버튼을 설치하였다. OWL과 SWRL의 형식으로 작성된 온톨로지는 이야기 탐색에 연계된 주제명/키워드를 결정하고 명문과 미사여구 및 사자성어 등을 추론 또는 추적할 수 있다. 그림을 검색시스템에 도입한 이유는 시스템이나 이용자들이 색인과 탐색식을 만드는 과정에서 사물이나 감정을 묘사하는데 언어의 한계를 느낄 때 어휘 대신 사용하도록 장



〈그림 8〉 글 검색 인터페이스



〈그림 9〉 단락 검색 인터페이스



〈그림 10〉 문장 검색 인터페이스

치한 것이다. 그림이나 소리 버튼을 클릭하면 순서에 입각하여 〈그림 7〉에 나열된 wordnet의 명사 범주어를 색인어로 삼은 사진들이 출력되고 마음에 드는 사진을 클릭하면 그 것에 대응한 이야기, 단락, 문장들이 모두 출력된다.

글(그림 8 참조), 단락(그림 9 참조), 그리고 문장(그림 10 참조) 검색 엔진에서 '1차 탐색'은 XML로 작성된 글/단락/문장 지식베이스 파일의 루트 노드 다음에 오는 2단계 노드와 애트리뷰트 그리고 엘리먼트들을 매개로한 값들이 검색되고, '2차 탐색'은 그 다음 경로인 3단계 노드와 애트리뷰트 그리고 엘리먼트들의 값들이 검색 대상이며 '3차 탐색'은 그 다음 경로인 4단계 노드 수준에 있는 값들이 검색된다.

5.2 탐색문 작성/탐색 전략

기본적인 탐색 전략은 첫째, 검색 대상에 대한 키워드나 주제명이 확실치 않을 경우, 글, 단락, 문장 검색 시스템 모두 앞의 〈그림 8, 9, 10〉 인터페이스에 있는 1차, 2차, 3차 탐색을 만족될 때까지 순차적으로 실행한다. 둘째로 대상 어휘

가 확실치 않을 경우에는 각 인터페이스에 있는 '온톨로지 탐색'란에 해당 단어를 입력하여 확인하거나, 〈그림 11〉 인터페이스에 있는 '가나다' 순 이야기 온톨로지서 확인한 후 첫째와 같이 한다. 셋째, 대상 어휘가 아주 모호할 때는 〈그림 11〉 인터페이스에 있는 '가나다' 순 이야기 온톨로지를 열람하거나 '그림 탐색'이나 '소리 탐색'으로 접근하여 찾고자 하는 대상과 비슷한 그림과 소리를 클릭하여 출력되는 글, 단락, 문장을 살펴 보면서 불만족일 때는 알맞은 단어를 골라 첫째와 같이 한다.

특히 단락 검색에서 '부터-로-에' 색인어는 주절과 주절에서 가까운 부절 순서로 주어, 목적어, 보어들을 뽑아내어 '주격 단어-목적격 단어-치소격 단어'와 같이 작성하여 입력한다. 그리고 문장 검색의 '절정보' 색인어를 입력할 때 한 문장이 여러 절을 가지고 있을 시에는 사다오의 방식(Makoto 1996, 172)을 원용하여 부절의 중요도 순서를 "(1) 원인, 목적, 대립, 전제, 불문, (2) 반반, 비교, 삽입, 의문, 반박, 가정, 상태, (3) 선택, 염려, 시간, 동조, 누가, (4) 추측, 수사, 과정, 나열, 수식, 현상, 과상"의 순서로 정하였다.



〈그림 11〉 온톨로지 탐색 인터페이스

〈표 12〉 온톨로지 무사용(A)과 사용(B) 비교 검증

	대응차		t	자유도	유의확률(양쪽)
	평균	표준편차			
A-B	-12.52941	6.94728	-7.436	19	0.001

5.3 평가

본고에서 구축한 문장 검색 시스템이 유용한가에 대한 질문에 대한 이용자의 답변은 재론의 여지 없이 긍정으로 결론을 내릴 수 있었다. 이 시스템이 택한 온톨로지의 효용과 구조 색인의 경로 단계별 탐색에 대한 검색효율을 F 척도($2PR/(P+R)$)로 20회 측정하고 t검정을 SPSS 12.0 K로 실시하였다(표 12 참조).

첫째, 온톨로지 사용과 무사용의 결과는 유의 확률(유의도 수준)이 0.001로 0.05보다 적으므로 의미있는 차이가 있는 것으로 나타났다. 둘째, 1차 탐색과 2차 탐색의 차이 검증(대응표본 t검증(오택섭, 최현철 2003, 110))에서는 유의확률 0.005, 셋째, 2차 탐색과 3차 탐색의 차이 검증(대응표본 t검증)에서도 유의확률 0.016으로 1차 탐색과 2차 탐색, 2차 탐색과 3차 탐색이 모두 차이가 있어 구조 색인의 각 경로 단계는 검색효율 향상에 이용될 수 있다.

6. 결론

이용자들에게 알맞은 이야기 쓰기 재료를 제공할 수 있는 색인 또는 탐색언어 개발을 위하여 먼저 이야기 구조와 단락을 분석하였고, 다음으로 이야기 문장들의 구성을 구문/의미적 역할로 분석하였다. 이야기 글 색인은 (1) 한명숙의 “①협력구조, ②배반구조, ③해결구조, ④긴장구조, ⑤극복구조, ⑥갈등구조, ⑦화해구조, ⑧ 불화구조” 8가지와 (2) 민스트렐과 브루투스의 “①배경, ②테마, ③줄거리, ④해법”, 그리고 “①등장인물, ②장소, ③시간, ④사건, ⑤목표, ⑥에피소드(하부목표, 시도, 결과), ⑦사건, ⑧상태”를 (1)은 이야기 주제, (2)는 단락 위치로 설정하였다. 또한 WordNet의 명사와 용언 범주를 이야기 글의 소재로 삼았다.

이야기 단락 색인은 Precis 역할 연산자와 민스트렐을 참조하여 “유형, 행위자, 대상, -로, -부터, -에, 상태”의 변수를 색인 시스템에 도입하

는 형식을 취하였고, 단락의 문장들간의 관계의 미는 연결 유형을 고려한 결과 ①수집, ②부가, ③공제, ④인과, ⑤이유, ⑥비교/대조, ⑦상세화, ⑧문제/해결, ⑨초담화로 나누었다. 이야기 문장 색인은 ①생크의 11가지 술어 범주, ②문장의 주어와 목적어, ③부절의 연결어미를 중심으로 작성하였다.

색인파일과 탐색문 작성의 일관성을 위하여 '이야기 온톨로지'를 구축하였으며, 이 온톨로지에는 단어별 관계, 그림과 소리 색인 관계, 문장 생성규칙, 가중치 등이 포함되었다. 구조화된 이야기, 단락, 문장의 색인과 원문을 웹 상에

서 표현하기 위해서 OWL과 SWRL의 형식을 참조하여 XML로 만들었다.

구조 색인의 경로와 중심 접근점인 소재와 소주제 및 량, 수준, 질 등의 효용척도를 산출하기 위하여 SPSS 12.0 K를 사용하였다. 실험결과로 경로와 소재 주제 등이 색인의 역할인 여과 기능을 충분히 수행한다고 판단할 수 있었으며 나아가 앞으로, (1) 이야기 구조 틀과 단락 구조 틀의 주제 추출, (2) 문장의 대표어 작성, (3) 각종 규칙의 정제, (4) 이를 기초로 하여 보다 정밀한 지식베이스 구축의 노력이 필요하다고 사료된다.

참 고 문 헌

- 김영택 외. 2001. 『자연언어처리』. 서울: 생능출판사.
- 김효영, 이순형. 2005. 이야기 구조에 따른 만 2, 3세 유아의 이야기 이해. 『아동학회지』, 26(1): 15-29.
- 송용욱, 홍준석, 김우주, 윤숙희, 이성규. 2006. 차세대 웹을 위한 SWRL 기반 역방향 추론엔진 SMART-B의 개발. 『한국지능정보시스템학회논문지』, 12(2): 67-79.
- 오택섭, 최현철. 2003. 『사회과학 데이터 분석법 1』. 서울: 나남출판.
- 이삼형. 1994. 『설명적 텍스트의 내용 구조 분석 방법과 교육적 적용 연구』. 박사학위논문, 서울대학교 대학원.
- 이태영. 1999. PC통신과 웹에서 지역알림정보의 작성을 돕는 전문가적인 서비스 모형에 관한 연구-지역주민의 견문을 중심으로-. 『정보관리학회지』, 16(1): 89-118.
- 한명숙. 2003. 이야기 구조의 구성 모형에 관한 연구. 『국어교육학 연구』: 559-583.
- Berners-Lee, T., J. Hendler, and O. Lassila. 2001. "The semantic web." *Scientific American*, may.
- Bringsjord, S., and D. A. Ferrucci. 2000. *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine*, Mahwan, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Cooper, M. D. 1998. "Design consideration and retrieval using empirical facts: Evaluation retrieval systems." *JASIS*,

- 49(10): 903-919.
- FAO Corporate Document Repository. [online]. [cited 2009.3.3]. <<http://www.fao.org/docrep/008/af234e/af234e01.htm>>.
- Giarratano, J. and G. Riley. 1998. *Expert Systems: Principles and Programming*. Boston: PWS Publishing Company.
- Nagao, Makoto. 1996. 『차세대 웹과 온톨로지의 핵심기술 자연언어처리』. 황도삼, 최기선, 김태석 공역. 서울: 홍릉과학사.
- Manning, C. D., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- On-line Writing Center of Purdue University. [online]. [cited 2008.10.10]. <<http://owl.english.purdue.edu/>>.
- On-line Writing Center of Purdue University. [online]. [cited 2009.1.21]. <http://owl.english.purdue.edu/writers/by_topic.htm>.
- Pe´rez, R. P. and M. Sharples. 2004. "Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA." *Knowledge-Based Systems*, 17: 15-29.
- Pretorius, A. J., "Ontologies-introduction and overview." [online]. [cited 2009.1.20]. <http://www.starlab.vub.ac.be/teaching/Ontologies_Intr_Overv.pdf>.
- Schank, R. C. and R. P. Abelson. 1975. "Scripts, plans and knowledge." *IJCAI*, 75: 151-157.
- Schank, R. C. and M. Selfridge. 1977. "How to learn/What to learn." *IJCAI*, 77: 8-14.
- Thorndyke, P. W. 1977. *Cognitive Structure in Comprehension and Memory of Narrative Discourse*, New York: Academic Press, 224-239. Quoted in S. Bringsjord, and D.A. Ferrucci. 2000. *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine*, Mahwan, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Turner, S. R. 1993. *MINSTREL: A computer model of creativity and storytelling*. PhD Dissertation, University of California LA.
- Wikipedia, the free encyclopedia. [online]. [cited 2009.3.1]. <http://en.wikipedia.org/wiki/Semantic_Web>.
- World Wide Web Consortium. [online]. [cited 2009.2.12]. <<http://www.w3.org/Submission/SWRL/>>.
- World Wide Web Consortium. [online]. [cited 2009.2.12]. <<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>>.
- World Wide Web Consortium. [online]. [cited 2009.2.12]. <<http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>>.

〈부록 1〉 글 지식베이스-XML 스키마 부분

```

<?xml version="1.0" encoding="euc-kr" ?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
targetNamespace="http://www.mirohan.com" xmlns="http://www.mirohan.com">
  <xsd:complexType name="storyType">
    <xsd:attribute name="author" type="xsd:string" use="required" />
    <xsd:attribute name="title" type="xsd:string" use="required" />
    <xsd:attribute name="goalsubject" type="xsd:string" use="required" />
    <xsd:attribute name="subjectmatter" type="xsd:string" use="required" />
    <xsd:attribute name="id" type="xsd:int" use="required" />
    <xsd:sequence>
      <xsd:element name="background:character" type="background:characterType" />
      <xsd:element name="background:location" type="background:locationType" />
      <xsd:element name="background:time" type="background:timeType" />
      <xsd:element name="theme:event" type="theme:eventType" />
      <xsd:element name="theme:goal" type="theme:goalType" />
      <xsd:element name="plot" type="plotType" />
      <xsd:element name="resolution:event" type="resolution:eventType" />
      <xsd:element name="resolution:state" type="resolution:stateType" />
      <xsd:element name="picture" type="pictureType" />
      <xsd:element name="sound" type="soundType" />
      <xsd:element name="paragraphes" type="paragraphesType" />
      <xsd:element name="sentences" type="sentencesType" />
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="paragraphesType">
    <xsd:sequence>
      <xsd:element name="paragraph" type="paragraphType" maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="sentencesType">
    <xsd:sequence>
      <xsd:element name="sentence" type="sentenceType" maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="plotType">
    <xsd:attribute name="literarystyle" type="xsd:string" use="required" />
    <xsd:attribute name="sentencestyle" type="xsd:string" use="required" />
    <xsd:sequence>
      <xsd:element name="episode" type="episodeType" maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>

```

```

</xsd:complexType>
<xsd:complexType name="episodeType">
  <xsd:attribute name="quantity" type="xsd:int" use="required" />
  <xsd:attribute name="quality" type="xsd:string" use="required" />
  <xsd:attribute name="level" type="xsd:string" use="required" />
  <xsd:attribute name="subgoal" type="xsd:string" use="required" />
  <xsd:attribute name="attempt" type="xsd:int" use="required" />
  <xsd:attribute name="result" type="xsd:string" use="required" />
</xsd:complexType>
<xsd:complexType name="paragraphType">
  <xsd:attribute name="subjectmatter" type="xsd:string" use="required" />
  <xsd:attribute name="paragraphform" type="xsd:string" use="required" />
  <xsd:attribute name="sentencestyle" type="xsd:string" use="required" />
  <xsd:attribute name="quantity" type="xsd:int" use="required" />
  <xsd:attribute name="place" type="xsd:string" use="required" />
  <xsd:sequence>
    <xsd:element name="type-status" type="type-statusType" />
    <xsd:element name="actor-object" type="actor-objectType" />
    <xsd:element name="from-to-at" type="from-to-atType" />
    <xsd:element name="paratext" type="paratextType" />
  </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="sentenceType">
  <xsd:attribute name="predicatetype" type="xsd:string" use="required" />
  <xsd:attribute name="actor" type="xsd:string" use="required" />
  <xsd:attribute name="object" type="xsd:string" use="required" />
  <xsd:sequence>
    <xsd:element name="clause" type="clauseType" maxOccurs="unbounded" />
  </xsd:sequence>
</xsd:complexType>
<xsd:element name="stories">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="story" type="storyType" maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
</xsd:schema>

```

