

# Factotum SemNet을 활용한 개연규칙 검증

양재균\*, 배재학\*, 유해영\*\*, 이종혁\*\*\*

\*울산대학교 컴퓨터·정보통신공학부

\*\*단국대학교 정보컴퓨터학부

\*\*\*포항공과대학교 전자컴퓨터공학부 컴퓨터공학과

e-mail: {jgyang, jhbae}@ulsan.ac.kr, yoohy@dankook.ac.kr, jhlee@postech.ac.kr

## Verification of Abductive Rules with Factotum SemNet

Jae-Gun Yang\*, Jae-Hak J. Bae\*, Hae-Young Yoo\*\*, Jong-Hyeok Lee\*\*\*

\*School of Computer Engineering and Information Technology,  
University of Ulsan

\*\*Division of Information and Computer Science,  
DanKook University

\*\*\*Dept. of Computer Science & Engineering,  
Pohang University of Science and Technology

### 요 약

독해 과정에서 문장간 개연성 파악은 저자의 지출의도 파악에 필수적인 요소로 작용한다. 이런 중요성은 온라인 문서의 처리에도 적용된다. 이 점은 기계를 이용한 자연어 처리과정에서 문서의 개연규칙 추출이 중요함을 의미한다. 본 논문에서는 개연규칙 추출 자동화의 전단계로서 미리 준비한 개연규칙을 검증해 보았다. 검증과정에서는 Factotum SemNet을 활용하였다. 검증과정을 통해, 준비된 개연규칙이 글 이해에 필요한 개연성을 잘 반영하고 있음을 확인하였고, 개연규칙 검증에 Factotum SemNet이 유용함을 보였다.

### 1. 서론

인간은 정보전달을 위한 수단으로 대화나 글을 이용한다. 대화나 글은 단일 문장으로 종결될 수도 있지만, 보통은 여러 문장이 연속된 형태로 표현된다. 그 결과 보다 용이하게 주제에 대한 정보를 전달할 수 있게 된다. 연속된 각 문장 사이에서는 서로 연관된 단어들 이 필연적으로 등장하게 마련이다. 최종적으로 청취자나 독자는 이 "연관된 단어들"로 표상되는 의미적 개연성을 파악함으로써 글이나 대화의 전체적인 의미를 이해하게 된다. 이렇듯 개연성 파악은 대화 속에 산재된 정보를 추출하는 문제와 관련이 있다[1].

인간은 누구나 글이나 대화로부터 올바른 개연성을 추출하는 능력을 가지고 있다. 이러한 능력은 인간이 광범위한 지식체계를 보유하고 있으며, 또한 관심이 없는 요소를 무시하고 서로 관계가 있는 사실들을 걸러낼 수 있기 때문에 가능하다. 따라서, 개연성은 문맥이나 독자의 배경, 지적수준 혹은 다른 요소들에 의해서 사람마다 다르게 결정된다[2].

Mike wanted Paul out of town for a few days.  
He suggested to Paul that he get away for a weekend.

[그림 1] 개연성 파악을 위한 예문

[그림 1]에 예시한 문장에서 "want"는 12가지, "suggest"는 7가지의 상호 독립적인 사전적 의미를 가진다[3]. 그러나 두 문장을 읽는 독자는 "want"와 "suggest" 사이에서 모종의 강한 연관성을 감지할 것이다. 설령 인식의 구체적인 과정에 대한 인지여부에 관계없이 두 어휘 사이의 개연성은 파악된다. 개연성 파악을 통해서, 두 문장에서 "Mike가 Paul에게 원하는 어떤 것이 있기 때문에 Paul에게 무언가를 제안한다."라고 이해하는 것이다.

한편 정보기술의 발달로 온라인 문서량은 지속적으로 증가하고 있으며, 개인이 처리 가능한 범위를 상회하고 있다. 이러한 상황은 온라인 문서처리를 도와줄 자연어처리 도구의 필요성을 증가시키고 있다. 자연어처리에 관한 기술 중에서 개연성 파악은 대상이 되는 글이나 대화의 거시적 이해에 간여한다. 또한, 정교한 원문 요약이 가능하도록 도와준다 [5, 6]. 이에 본 논문에서는 Factotum SemNet[4]을 활용하여 미리 준비한 개연규칙을 검증해 보았다.

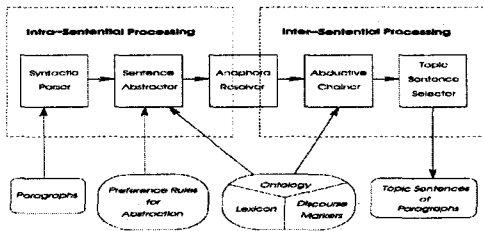
### 2. 원문이해

원문이해는 원문 이해심도에 따라서 표층·중층·심층 등의 3단계로 구분할 수 있다. 이중 표층적 원문이해는, 단어간의 직·간접적인 의미결속을 토대

로 성취하는 표층적인 원문이해이다[5]. 현시점에서 가능한 자연언어 처리도구나 자원을 활용하면 표층적 원문이해가 가능하다. 원문이해 시스템을 활용하여 주어진 원문에 대한 문장추상화를 통해 검증할 개연규칙을 마련하였다.

2.1 원문이해 시스템

원문이해 시스템[6]은 대용어 해소기를 중심으로 문장내 처리부와 문장간 처리부로 나뉜다. 구성요소는 다음과 같다: 온톨로지 및 심층사전, 구문분석기, 문장추상기, 대용어 해소기, 화제문 선정기. 이 중에서 심층사전은 Prolog 데이터베이스로 구현하였다. 구문분석기는 Link Grammar Parser[7]의 원시 프로그램을 개작한 것이다. 마지막으로, 대용어 해소기는 선형사 구조[8]에 관련된 알고리즘의 구현이다.



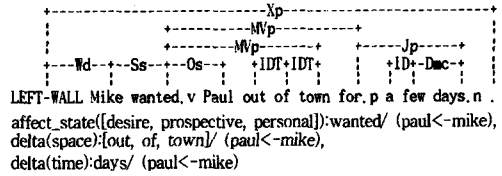
[그림 2] 원문이해 시스템 구성도

2.2 문장추상화

문장추상화(Sentence Abstraction)란 개념추상화를 도입한 문장교열작업이다. 이를 통해서, (1) 문장의 주요의미를 성형적으로 표현함과 아울러 (2) 개연사슬의 고리가 될 수 있는 후보어나 줄거리 단위의 심상이 될 수 있는 문장 구성성분을 분리해낼 수 있다.

문단기반의 원문이해 방법론을 구현하는 과정에서 OfN(Ontology for Narratives)[9]이라는 온톨로지를 설정하였다. OfN 구현은 Roget 시소러스[10]를 심층사전으로 삼아 이를 재구성하여 얻었다. 문장추상화를 위해서는 구문분석 결과에서 OfN 범주가 확인된 문장추상화에 참가할 요점을 선택하여야 한다. 요점이란 문장추상화의 적절한 대상이 될 수 있는 후보단어이다.

다음 예문을 보자: Mike wanted Paul out of town for a few days. [그림 3]의 하단에 이 예문에 대한 문장추상기 SABOT의 출력이 있다. Prolog로 구현한 문장추상기 SABOT가 요점어 mike, paul, wanted, out of town, 그리고 days 등을 선별해내었다. 술어 affect\_state와 delta는 각각 OfN의 심상과(시간, 공간)의 변화에 대응한다.



[그림 3] 추상화된 예문

2.3 개연규칙

개연규칙은 문장간 구성성분들의 개연적인 결속성을 나타내며 문장내 구성성분들이 가지는 OfN 정보로서 표현된다. 2항 또는 3항인 개연규칙의 일반적인 모습은 다음과 같다.

Ante <= Post [{=+}, =->, =\*>} Cons]

여기에서 (1) Ante, Post, Cons는 pred(args)의 형태로 표현된다, (2) pred(args)는 OfN에 명시된 개념으로 7가지의 범주로 표현된다, (3) ==> and ==->는 Post와 Ante에 제한사항이 있음을 나타내고, (4) ==>는 담화표지가 있음을 나타낸다. [표 1]은 Mike와 Paul의 이야기[5]를 처리하여 선정한 개연규칙의 예이다[6].

[표 1] 개연규칙

1	% 마음이 통하면 주고 싶어진다 affection(sympathetic) <= affection(offer)
2	% 풀이 죽으면 소극적이 된다 event(inactivity) <= event(descend).
3	% 장소를 바꾸고 싶을 때 여행을 한다 delta(space) <= event(journey) ==-> affection(prospective).

개연규칙은 크게 2항 규칙과 3항 규칙으로 나눌 수 있다. [표 1]에서 규칙 (1)과 (2)는 2항 규칙, (3)은 3항 규칙이다.

3. 개연규칙 검증

Mike와 Paul의 이야기에서 추출한 개연규칙은 인간의 견지에서 보편적인 수긍이 가능하다. 인간의 경우는 개연규칙을 직관적으로 추출, 파악하지만 그 과정을 자동화하는 것은 용이하지 않다. 이런 이유로 개연규칙 추출의 자동화에 앞서 선정한 개연규칙을 검증해 보았다. 개연규칙의 검증에는 SemNet의 기능관계(functional relation)를 적용하였다.

3.1 Factotum SemNet

Factotum SemNet은 MICRA에서 Roget 시소러스와 Webster 사전을 이용해서 구축한 semantic network이다. 이 의미망은 Roget 시소러스의 분류체계를 보완, 재구성하였다. 또한 is-a등의 의미관계표식(semantic relational marker)을 이용하여 어휘의 의미계승을 가능하게 하였다[11].

[그림 4]와 같이 Factotum SemNet은 약간의 계층이 추가되었지만 전체적으로는 Roget 시소러스의 분류체계를 이용하고 있다. relation은 “(” 기호 내부에 표기하고 해당 범주의 동의어 목록은 같은 행에 기술하고 있다. 한 예로 {{result\_of}}가 나타나는 행의 의미는 “conversion”은 “transforming”의 결과임을 뜻한다. [그림 4]의 하단에 Factotum SemNet의 이해를 돕고자 서술적인 형태로 기술하였다.

A6.1.4 CONVERSION (R144)  
 #144. Conversion.  
 N. {{has\_subtype(change, R140)}} conversion, transformation.  
 {{has\_case: @R7, initial state, final state}}.  
 {{has\_patient: @R3a, object, entity}}.  
 {{result\_of}} {{has\_subtype(process, A7.7)}} converting, transforming.  
 {{has\_subtype}} processing.  
 transition.  
 =>  
 <change, has-subtype, conversion> <change, has-subtype, transformation>  
 <conversion, has-case, initial state> <conversion, has-case, final state>  
 <conversion, has-patient, object> <conversion, has-patient, entity>  
 <conversion, is-result-of, converting> <conversion, is-result-of, transforming>  
 <process, has-subtype, converting> <process, has-subtype, transforming>  
 <conversion, has-subtype, processing>

[그림 4] Factotum SemNet의 구조(일부)

Factotum SemNet에는 400개 이상의 relation이 존재한다. [표 2]에 relation의 일부를 예시하였다. 대부분의 relation은 대칭적인 성격이 아니기 때문에 역방향으로도 기능하는 relation은 드물다. 현재로서는 의미망을 서술하는 어떤 특징은 relation을 추출하는 과정에서 무시되었다. 예를 들어, relation의 세 목록에서 “&”기호가 앞에 나타날 수 있는데, 이 기호의 의미는 해당 relation이 항상 적용되지 않고 어떤 경우에만 유효하다고 한정시킨다.

[표 2] Factotum SemNet의 relation(일부)

Relation	의미
has-subtype	is-a와 반대의 관계
is-property-of	종속된 특성
is-caused-by	대상의 원인을 지시
has-property	대상의 두드러진 특성
has-part	물리적인 요소
has-high-intensity	특성에 대한 강화자
has-high-level	행위에 대한 함의
is-antonym-of	반대의 의미
is-conceptual-part-of	개념적인 요소
causes <sub>mental</sub>	동기

### 3.2 Factotum SemNet을 활용한 개연규칙 검증

Mike와 Paul의 이야기에서 추출한 개연규칙을 Factotum SemNet의 functional relation을 이용하여 검증해 보았다. 이를 통해 relation을 이용하여 개연규칙의 추출 자동화가 가능함을 보이고자 한다.

[표 3] Factotum SemNet을 이용한 개연규칙 검증

개연규칙 1		
관련어휘	friend	want
개연사슬	• Friend {{has_subtype: favorite}} • Favorite {{has_patient(affection)}} • Affection {{similar to: sympathetic affections}} • Sympathetic affection {{has_subtype(feeling)}} • Feeling {{affects(human condition)}} {{caused_by(desire)}} • Desire {{has_tobiet: want}}	
어휘사슬	friend - favorite - affection - feeling - desire - want	
의미	마음이 좋하면 바램이 생긴다.	
개연규칙 2		
관련어휘	friend	present
개연사슬	• Offer ≡ present • Offer {{of(physical object)}} {{has_result: @R784, offering, gift}}. • Giving {{has_subtype(gift)}} {{causes(respect, R928)}} tribute. • Respect {{conceptual_part_of(esteeem: love &c. R897)}}. • Love {{has_high_intensity(friendship)}} {{has_subtype(passion, @R821)}} love, affection, tenderness, fondness. • Friendship {{has_subtype(affection)}} {{has_property(friend)}} friendship, amity, friendliness.	
어휘사슬	present - offer - gift - giving - respect - love - friendship - friend	
의미	마음이 좋하면 주고 싶어진다.	
개연규칙 3		
관련어휘	want	suggest
개연사슬	• Advice {{has_subtype: @R616a, suggest, prompt, recommend, prescribe, advocate, exhort}}. • Advice {{&causes(get advice)}} {{has_subtype(consent, R762)}} • Consent {{&caused_by: turn a willing ear &c. (willingness)}}. • Willingness {{has_subtype(desire, R866)}} • Desire {{has_tobiet: want &c. (requirement) R630}}.	
어휘사슬	suggest - advice - consent - willingness - desire - want	
의미	충고할 마음은 바램이 있을 때 생긴다.	
개연규칙 4		
관련어휘	puttering	settled
개연사슬	• Quiescence ≡ settle &c. (be located) R184; • Quiescence {{of(person/beat): sleep, repose, inactivity}}. • Inactivity ≡ dilatory, listless, lackadaisical, puttering, pottering.	
어휘사슬	settle - quiescence - inactivity - puttering	
의미	풀이 죽으면 소극적이 된다.	
개연규칙 5		
관련어휘	asked	hear
개연사슬	• Request ≡ asked • Request {{has_subtype(request[for=favor/donation[2], @R784]]}} solicitation, appeal. • Court {{has_subtype}} {{alternative_to(trial court)}} {{performed_by(hear[obj=appeal, @R969], @R969)}} appellate court, higher court, court of appeal. • Communication {{&causes(tell): @R418, hear, overhear}}.	
어휘사슬	asked - request - appeal - court - hear	
의미	요청을 듣다.	
개연규칙 6		
관련어휘	hear	thought
개연사슬	• Thought ≡ thought • Thought {{conceptual_part_of: @M5.4.1, discuss, hammer at}}. • Conversation {{(case_of(issue, @M3.1.1.R1))}} confer, discuss. • Conversation {{(participant_in(person[num>1]))}} ... talk, discourse, verbal intercourse; oral communication, commerce. • Communication {{&causes(tell): @R418, hear, overhear}}.	
어휘사슬	thought - discuss - conversation - communication - hear	
의미	듣고 생각하다.	
개연규칙 7		
관련어휘	asked	thought
개연사슬	• Request ≡ asked • Request {{has_subtype(suggestion)}} proposal, proposition. • Meaning {{conceptual_part_of: suggestion &c. (information) R527b; figure of speech &c. R521}}. • Meaning {{has_conceptual_part: concept, R453}}. • Concept {{has_conceptual_part(topic)}} {{has_conceptual_part(thought, R451)}} concept, idea, conception, notion, conception, thought. • Thought ≡ thought	
어휘사슬	asked - request - suggestion - meaning - concept - thought	
의미	요청에 대처할 방법을 강구하다.	

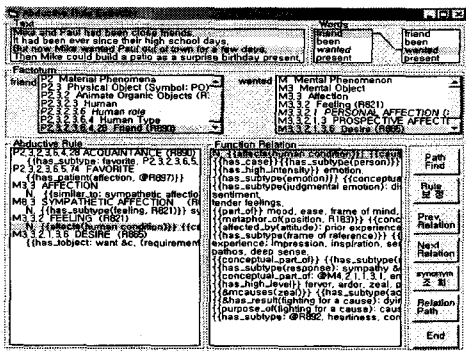
\* “≡” 기호는 Factotum SemNet에 없는 표기로 Synonym을 뜻함.

[표 4] Factotum SemNet을 이용한 개연규칙 검증(계속)

개연규칙 8	
관련어휘	thought idea
개연사슬	· Thought ⇒ thought · Thought ((has_conceptual_part: concept, R453; reasoning, R476)) · Concept ((has_conceptual_part(topic)) (has_conceptual_part(thought, R451))) concept, idea, conceation, notion, conceation, thought.
어휘사슬	thought - concept - idea
의미	착상이 떠오른다.
개연규칙 9	
관련어휘	went here
개연사슬	· Communication ((causes(tell): @R418, hear, overhear)). · Communication ((has_subtype(assertion)), M8.1.1.5) {(property_of(possible))} {(not_property_of(certain))} {(topic_of(possibility[2]))} hint, suggestion, proposal ... · Request ((has_subtype(proposal)) (in_frame(parliamentary procedure))) motion. · Motion ((caused_by(propel)) (intransitive) move[1], translocate, go, travel, be in motion, be moving, be in transit.
어휘사슬	hear - communication - proposal - motion - go
의미	뜻을 전하러 가다.
개연규칙 10	
관련어휘	emergency bus
개연사슬	· Journey ⇒ bus · JOURNEY ((has_result(travelling)) ((location_of(land, R342)) {(has_subtype(event, R151))} journey, travel by land. · Event ((has_subtype(event))) crisis, emergency, adventure, pass
어휘사슬	bus - journey - event - emergency
의미	우발사건이 발생하여 이동하다.

[표 3]의 개연규칙 3 “충고할 마음은 바람이 있을 때 생긴다.”를 예로 들어 개연규칙 검증 과정을 설명하고자 한다. 이 개연규칙은 [그림 1]의 문장에 나타나는 어휘 “want”와 “suggest” 사이의 개연성을 규칙화한 것이다. 우선, Factotum SemNet에서 “suggest”를 검색한 결과 중에서 어휘적 밀접도가 가장 높은 “Advice”범주를 “suggest”의 모범주로 선정한다. 이때의 relation은 “suggest”가 “advice”의 subtype임을 밝히고 있다. 다음 단계에서 “advice”는 “consent”를 subtype으로 가짐을 알 수 있다. 같은 방법으로 계속 relation을 검색하면 “want”에 닿을 수 있다. 그리고 이때의 개연 어휘사슬은 “suggest - advice - consent - willingness - desire - want”와 같이 연결된다. 결국, Factotum SemNet의 relation과 synonym 목록을 이용한 개연규칙 검증이 가능함을 알 수 있다.

3.3 개연규칙 추출 인터페이스



[그림 5] 개연규칙 추출 인터페이스

개연규칙을 검증하는 과정을 통해 개연규칙 추출 자동화에 필요한 요소와 기능을 선정하였다. 이를 바탕으로 [그림 5]와 같이 개연규칙 추출 인터페이스를 설계하였다.

4. 결론

본 논문에서는 온라인 문서내의 개연규칙을 자동 추출할 목적으로, Factotum SemNet[4]의 Functional Relation 탐색을 통해 미리 준비한 개연규칙을 검증해 보았다.

검증에 사용된 개연규칙은 Mike와 Paul의 이야기 [5]에서 추출하였으며, 대부분 수동이 가능한 것이었다. 이 개연규칙은 의미망(Semantic Network)의 Relation과 유의어 탐색을 통해서 검증이 가능함을 보였다. 또한 개연규칙 검증과정을 개관함으로써, 개연규칙 추출 자동화에 적용할 수 있는 개괄적인 방법을 찾을 수 있었다. 검증의 결과로 개연규칙 어휘사슬을 얻음으로서 개연규칙의 타당성 결정을 보다 용이하게 보일 수 있었다. 끝으로 개연규칙 추출 자동화에 필요한 요소와 기능을 정리하여 개연규칙 추출 인터페이스를 설계하였다.

참고문헌

[1] Sanda M. Harabagiu. "WordNet-Based Inference of Textual Cohesion and Coherence." FLAIRS-98, May 1998, Sanibel Island, FL.  
 [2] Sanda M. Harabagiu, and Dan Moldovan. "A Parallel Algorithm for Text Inference." International Parallel Processing Symposium, IPPS-96, April 1996, Honolulu, Hawaii, pages 441-445.  
 [3] Merriam-Webster Online.  
<http://www.webster.com>  
 [4] MICRA Inc., Factotum SemNet  
<http://www.micra.com/>  
 [5] Bae, J.-H. J. and Lee, J.-H. "Topic Sentence Selection with Mid-Depth Understanding." Proc. of ICCPOL, pp. 199-204, 2001.  
 [6] 김곤, 배재학. 개연성 규칙과 문장추상화를 활용한 문서요약. 한국정보처리학회 제18회 춘계학술발표대회 논문집, 제9권, 제2호, pp. 359-362, 선문대학교.  
 [7] Sleator, D. and Temperley, D. "Parsing English with a Link Grammar." Third International Workshop on Parsing Technologies, August 1993.  
<http://www.link.cs.cmu.edu/link/>  
 [8] Grefenstette, G., "Producing Intelligent Telegraphic Text Reduction to Provide an Audio Scanning Service for the blind", In *Working Notes of the Workshop on Intelligent Text Summarization*, pp. 111-117, 1998.  
 [9] 양재근, 배재학. "온톨로지 정보를 이용한 범주 재편성: Roget 시소러스의 경우." 한국정보처리학회, 제 9권, 제 1호, pp.515-518, 2002.  
 [10] Roget's Thesaurus.  
<http://promo.net/cgi-promo/pg/t9.cgi?entry=22&full=yes&ftp>  
[site=ftp://biblio.org/pub/docs/books/gutenberg/](http://biblio.org/pub/docs/books/gutenberg/)  
 [11] O'Hara, Tom, and Janyce Wiebe (forthcoming), "Classifying functional relations in Factotum via WordNet hypernym associations", to appear in Proc. Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2003).