

# 전자장비 고장진단 질의응답을 위한 인과관계 정의 및 추출

(Definition and Extraction of Causal Relations for  
Question-Answering on Fault-Diagnosis of Electronic  
Devices)

이 신 목<sup>†</sup>      신 지 애<sup>\*\*</sup>  
(Sheen-Mok Lee)      (Ji-Ae Shin)

**요 약** 온톨로지의 인과관계는 특정 응용을 위한 추론에서 중요한 역할을 하므로, 인과관계는 응용에서 쓰이는 추론의 형태에 근거하여 정의되어야 한다. 본 논문에서는, 전자장비의 고장진단 질의응답을 위한 온톨로지에서의 인과관계를 정의하고 추출하는 모델을 제시한다. 질의응답의 패턴을 분석하여 인과범주를 정의하고, 질의응답에서 나타나는 개념들 사이의 관계들 중 인과범주에 속하는 경우를 인과관계로 정의한다. 인과관계 인스턴스는 응용분야의 정의문으로부터 어휘 패턴을 이용하여 추출되고 시소러스 정보를 이용하여 점진적으로 확장된다. 분야 전문가들의 평가 결과, 본 모델은 관계분류에 있어서 92.3%의 평균 정확률과 추출 단계의 인과관계 인식에 있어서 80.7%의 정확률을 보인다.

**키워드** : 인과관계, 응용 온톨로지, 고장진단 질의응답, 관계추출, 관계정의, 인과추론

**Abstract** Causal relations in ontology should be defined based on the inference types necessary to solve problems specific to application as well as domain. In this paper, we present a model to define and extract causal relations for application ontology for Question-Answering (QA) on fault-diagnosis of electronic devices. Causal categories are defined by analyzing generic patterns of QA application; the relations between concepts in the corpus belonging to the causal categories are defined as causal relations. Instances of casual relations are extracted using lexical patterns in the concept definitions of domain, and extended incrementally with information from thesaurus. On the evaluation by domain specialists, our model shows precision of 92.3% in classification of relations and precision of 80.7% in identifying causal relations at the extraction phase.

**Key words** : causal relation, application ontology, diagnostic QA, relation extraction, relation definition, causal inference

## 1. 서론

인과관계의 사전적 의미는 일반적으로 협의적 의미와

광의적 의미로 나눌 수 있다. 협의적 의미는 “원인과 결과 사이의 관계”이며, 광의적 의미는 “정규적으로 상호연관성을 지닌 사건 사이의 관계”이다[1]. 예를 들어, “전자장비를 수리하는” 사건과 “전자장비의 고장 문제가 해결된” 사건 사이에는 두 가지 의미의 인과관계가 모두 나타나지만, “화상 카메라용 디바이스 드라이버”와 “화상 카메라” 사이에는 협의적 의미의 인과관계는 나타나지 않고, 광의적 의미의 인과관계만이 나타난다. 일반적으로, “인과관계”라는 용어는 다양한 분야에서 다양한 용도로 나타나며, 대개의 경우, 협의적 의미 범위를 포함하고 광의적 의미 범위에 속하는 영역에 위치한다.

온톨로지에서의 인과관계 또한, 협의적 의미와 광의적

<sup>†</sup> 학생회원 : 한국과학기술원 전산학과 박사과정 연구원  
smlee@world.kaist.ac.kr

<sup>\*\*</sup> 정 회원 : 한국정보통신대학교 공학부 교수  
jiae@icu.ac.kr

논문접수 : 2008년 1월 18일

심사완료 : 2008년 4월 22일

Copyright©2008 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제35권 제5호(2008.5)

의미 사이에 존재하며, 온톨로지의 분야와 용도를 고려하여 정의하여야 한다. 왜냐하면, 분야와 응용에 따라서, 필요로 하는 인과관계의 형태는 매우 다르게 나타날 수 있기 때문이다. 예를 들어, “감염시키는(infect)” 관계의 경우, 의학 분야 온톨로지에서는 생화학적 반응을 표현하는 유용한 인과관계이지만, 정보기술 분야에서는 비교적 유용성이 두드러지지 않는다. 또한, 같은 분야 내에서도 응용에 따라 인과관계의 형태가 매우 다르게 나타날 수 있다. 예를 들어, 정보기술 분야 내에서 전자장비의 고장진단에 대한 질의응답을 수행하는 응용과 제품간 비교를 위한 응용에서 나타나는 인과관계를 비교할 수 있다. 고장진단을 위한 응용에서는, 전자장비가 어떠한 원인과 과정을 거쳐서 고장 났는지를 분석하여야 한다. 이와 같은 고장의 전달 경로를 알기 위하여서는, 고장을 전달시킬 수 있는 기능적 연관성을 알아야 한다. 따라서, “화상 카메라용 디바이스 드라이버”와 “화상 카메라” 사이의 기능적 연관성을 나타내는 의미관계 (“drive”)는 고장진단을 위한 응용에서 유용한 인과관계라 할 수 있다. 하지만, 제품비교 응용에서는 주변 기기간의 기능적인 연관성이 인과관계로서의 유용성을 가지는 경우가 많지 않다. 즉, 하나의 관계가 인과관계인지의 여부를 판단하는 데에는 애매성이 존재하며, 분야와 응용에 따라서, 필요로 하는 인과관계의 형태가 다르다. 따라서, 분야와 응용에 맞는 인과관계를 정의하고 추출함으로써, 보다 효율적인 추론이 가능하다.

본 연구에서는, 정보기술이라는 분야의 특성과 고장진단 질의응답이라는 응용의 특성을 함께 고려한, 정보기술 분야에서 사용되며 전자장비 고장진단을 위한 질의응답을 응용의 목표로 하는 온톨로지의 인과관계를 정의하고 추출하는 모델을 제시한다. 첫째로 특정응용에서의 추론의 역할에 따라 어떠한 범주의 인과관계가 필요한지를 결정하여야 한다. 둘째로, 각 범주별로 정의된 역할을 수행하는 의미관계를 인과관계로 정의하여야 한다. 마지막으로, 인과관계로 정의된 의미관계를 말뚝치로부터 추출하여야 한다. 이와 같은 과정이 본 연구에서 어떻게 다루어졌는가를 요약하면 다음과 같다.

첫째, 고장진단 질의응답에서 이루어지는 인과추론 과정을 분석하여, 추론에 필요한 인과관계 범주<sup>1)</sup>를 설정한다. 이를 위하여, 본 연구에서 다루고자 하는 전자장비의 고장진단을 위한 질의응답 말뚝치를 분석한다. 예를 들어, 고장의 원인을 분석하려면 하나의 구성요소에서 일어난 고장이 다른 구성요소의 고장을 야기시키는, 즉, 고장상태를 전파시키는 기능을 수행하는 인과관계를

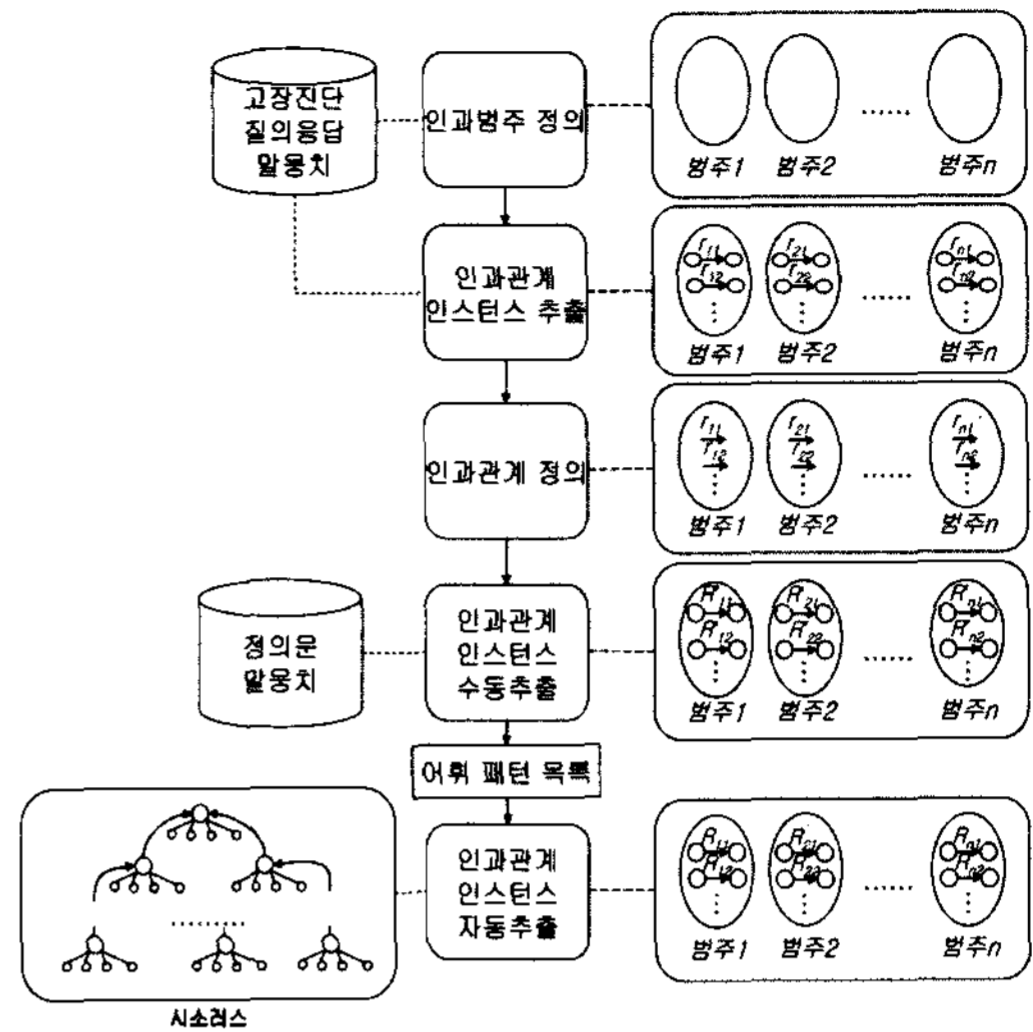


그림 1 인과관계 정의 및 인과관계 인스턴스 추출 모델

찾아야 한다. 이와 같은 관계를 본 논문에서는 “전달” 범주로 정의한다.

둘째, 각 인과범주에 포함되는 인과관계를 정의한다. 이를 위하여, 질의응답 말뚝치로부터 인과관계 인스턴스<sup>2)</sup>를 추출한다. 말뚝치에서 특정 개념쌍 사이에 인과관계가 있다고 판정한 경우, 인과관계 인스턴스를 추출하고, 추출한 인과관계 인스턴스의 정의역과 치역 사이의 의미관계를 부여한다. 개념쌍 사이의 관계가 기존에 정의된 의미관계로 표현되지 않는 경우, 새로운 의미관계를 정의하여 정의된 의미관계를 부여한다. 예를 들어, 고장진단 질의응답을 위한 추론과정에서, “화상 카메라용 디바이스 드라이버”에서 난 고장이 “화상 카메라”에서의 고장을 야기시키는 “전달”범주의 관계가 포함될 경우, 두 개념쌍을 정의역과 치역으로 하는 인과관계 인스턴스를 추출하고, 두 개념 간의 의미관계로서 “drive”를 부여한다. 의미관계 “drive”가 “전달” 범주에 이미 정의되어 있지 않은 경우, “drive” 관계를 “전달” 범주의 관계에 포함시킨다.

셋째, 정의한 인과관계에 해당하는 인과관계 인스턴스를 정보통신 분야 정의문<sup>3)</sup>으로부터 추출한다. 이를 위하여, 인과관계를 나타내는 어휘 패턴을 추출하고, 이를 이용하여 초기 인과관계 인스턴스를 추출한다. 초기 인과관계 인스턴스는 시소러스 정보를 이용하여 확장한다.

1) 인과범주: 인과관계의 분류체계로서, 응용 추론에서 수행하는 역할이 같은 인과관계는 같은 인과범주에 속한다.

2) 두 개념 사이의 의미관계가 특정한 인과범주에 속하는 경우, 인과관계라 한다. 특정한 인과범주에 속하는 하나의 의미관계를 지칭하는 경우에도 사용한다. 인과관계의 정의에서 전자의 경우를 “인과관계 인스턴스”라고도 한다.

3) 본 논문의 4장에서 인과관계 인스턴스 추출의 대상 말뚝치로 정보통신 분야 정의문을 사용한 이유를 설명하고 있다.

확장된 관계는 어휘 패턴을 확장하는 데 다시 사용한다. 본 논문에서는, 이와 같은 어휘 패턴과 인과관계의 점진적 확장 모델을 제안한다.

이 논문은 다음과 같이 구성되어 있다. 2장에서는 온톨로지 및 인과관계에 대한 기존 연구를 살펴본다. 3장에서는, 전자장비의 고장진단 질의응답으로부터 인과관계를 정의하는 모델을 제시하고, 4장에서는 정보통신분야 정의문으로부터 인과관계 인스턴스를 추출하는 모델을 제시한다. 5장에서는 인과관계 정의 및 추출 모델에 관한 실험 및 평가 결과를 보이고, 6장에서 결론과 향후 연구 방향을 고찰한다.

## 2. 관련 연구

본 장에서는 일반적인 인과관계의 정의 및 추출에 대한 연구에 대하여 논하고자 한다.

### 2.1 인과관계 정의

어떠한 관계를 인과관계로 판단할 것인지에 대한 문제는 현재까지의 인과관계 추출 연구들에서 꾸준히 논의하여 왔지만, 온톨로지의 분야와 응용을 동시에 고려한 객관성 있는 인과관계 정의에 대한 연구는 아직까지 없는 것으로 알고 있다.

기존의 인과관계 정의 방법들을 분류하면, 언어학적 이론에 기반한 방법, 사전정보를 이용한 방법, 해당 분야 말뭉치의 관찰을 통한 방법으로 구분할 수 있다. 표 1은 각각의 방법들의 특성을 나타낸다. 세 방법 모두 온톨로지 응용을 고려하지는 않았으며, 온톨로지 분야나 객관성에 대하여서도 고려하지 않은 경우가 많다.

언어학적 이론에 기반한 방법은 인과관계에 대한 기존의 언어학적 분석 결과에 기반하여 기준을 제시하는 방법이다. 예를 들어, 한국어 문서에서 나타나는 인과관계를 인과/결과/조건 의 세 가지 타입으로 정의한 언어학적 분석[2]에 의거하여, 인과관계 초기 데이터를 수동으로 태깅할 수 있다[3].

사전에서 제공하는 인과관계에 대한 정보로부터 인과관계를 정의하는 방법도 자주 사용한다. 영어에서의 인과관계 추출을 위한 많은 방법론들이 워드넷 시소러스의 사전 정의문에 있는 일반적인 인과관계 정보를 이용

하여 초기 패턴을 추출하였다[3-5]. 이 방법은 명확하고 객관적인 정의 기준을 제시한다는 장점이 있다.

한편, 분야 말뭉치의 관찰을 통한 방법에서는 대상으로 하는 분야에서 실제로 나타나는 문장들을 이용하여 인과관계 판단 기준을 설명한다. 예를 들어, 한국어 특허문서에서 나타나는 인과관계의 기준을 설명하기 위하여, 특허문서에 실제로 나타나는 문장들 가운데, 인과관계 판단이 어려운 예제들을 분류하고, 각 경우에 대하여 인과관계의 판단기준을 제시할 수 있다[6]. 이 경우는, 분야 말뭉치를 관찰하여 기준을 설정하므로, 분야의 특성을 자연스럽게 고려할 수 있다는 장점이 있다.

하지만, 현재까지의 인과관계 연구들에서는 “화상 카메라용 디바이스 드라이버”와 “화상 카메라” 사이의 관계와 같이 특정한 형태의 응용에서 필요로 하는 인과관계를 정의할 수 없다. 따라서, 응용을 고려한 새로운 형태의 인과관계 정의 방법론이 필요하다.

### 2.2 인과관계 추출

인과관계 추출에 관한 기존의 연구는 사용한 정보에 따라 다음과 같은 두 가지로 분류할 수 있다.

첫째, 사건을 연결하는 어휘 및 구문 정보에 기반한 방법이다. 일반적으로, 인과관계를 나타내는 동사나 접속사 등이 원인과 결과 사건을 연결시키는 경향이 있다. 인과관계를 나타내는 패턴은 어휘나 구문 형태로 나타나며, 이들을 말뭉치로부터 학습한 결과를 추출에 이용한다[5,7].

둘째, 사건어휘 및 사건어휘의 시소러스 정보를 함께 고려하는 방법이다. 특정한 어휘/구문 패턴이 연결시킬 수 있는 사건의 종류에는 제약이 있을 수 있다. 사건 표현의 최상위 시소러스 정보를 이용하여 이와 같은 제약조건을 나타낼 수 있다[3,4]. 어휘/구문 패턴과 어휘쌍 및 시소러스의 상위어 정보쌍 등에 대한 확률값을 계산하여 인과관계 분류를 위한 자질로 사용할 수도 있다[3].

특히, 사건 표현이 명사구로 나타나는 경우, 이와 같은 정보는 인과관계 추출에 더욱 유용하다. 본 연구에서도, 명사구 간 인과관계를 대상으로 하므로, 연결 패턴과 더불어 어휘 및 시소러스 정보를 함께 활용하고자 한다.

## 3. 고장진단 질의응답을 위한 인과관계 정의

본 장에서는 본 연구에서 응용으로 선택한 전자장비의 고장진단에 대한 질의응답을 위한 인과관계의 범주를 정의하고, 각 범주별로 인과관계를 정의하는 과정을 설명한다.

### 3.1 고장진단 질의응답에서의 인과범주 정의

고장진단 질의응답을 위한 인과관계의 범주는 표 2와

표 1 인과관계 정의 방법들의 특징

기준 방법	객관성	분야의 고려	응용의 고려
A	X	X	X
B	O	X	X
C	X	O	X

- A: 언어학적 이론에 기반한 방법
- B: 사전정보를 이용한 방법
- C: 해당분야 말뭉치의 관찰을 통한 방법

표 2 고장진단 질의응답을 위한 인과범주

인과범주	역 할	예제
원인	외부로부터의 고장의 최초 원인이 특정 개념에 고장/이상을 일으키는 관계	harm(웬바이러스, 컴퓨터시스템)
전달	특정 개념의 고장/이상이나 고장 해결 효과가 다른 개념의 고장/이상이나 효과로 옮겨가는 관계	drive(프린터드라이버, 프린터), hasFunction(웹캠, 화면디스플레이)
증상	특정 개념의 고장/이상이 외부적으로 특정 증상을 보이는 관계	hasState(3d 게임, 락 현상)
처리	특정 개념(일반적으로 "사용자")이 다른 개념에 작용하여 문제해결의 최초 실마리를 제공하는 관계	execute(사용자, 패치작업) install(사용자, 패치프로그램)
효과	특정 개념이 작용하여, 다른 개념의 고장 /이상을 해결하거나 가능성을 감소시키는 관계	solveProblemOf (패치작업, 보안 모듈프로그램)

System

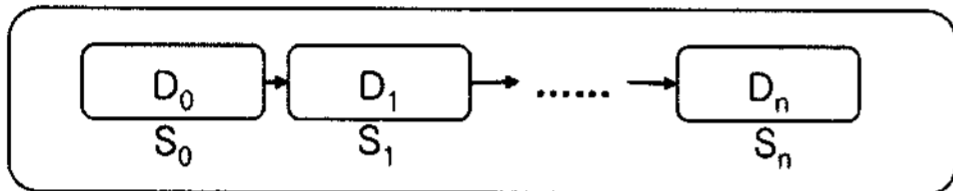


그림 2 전자장비 시스템의 개략도

같이 분석한다. 표 2에서는 5가지의 인과범주를 정의한다. 이에 대한 예를 들기 위하여, 그림 2의 전자장비 시스템에 대한 고장 상황을 가정한다.

그림 2에는  $n$ 개의 구성요소가 있으며, 각각의 내부 부품 사이에는 기능적 의존성이 존재한다. 그림 1에서,  $D_i$ 는  $i$ 번째 구성요소를 의미하며,  $S_i$ 는  $i$ 번째 구성요소의 상태를 의미한다.  $S_i$ 가 0이면 구성요소가 고장 상태임을, 1이면 정상 상태임을 의미한다. 구성요소  $D_i$ 와  $D_j$  사이의 화살표는  $S_j$ 가  $S_i$ 에 의존적임을 의미한다. 그림 3은 그림 2의 시스템에 고장이 발발하여, 이를 처리하는 과정을 Petri-net의 형태로 나타낸다.

그림 3의 타원형 노드에는 전자장비의 상태를 각 구성요소별 상태로 표현하며, 실선은 각 상태 사이에 발생하는 사건을 나타낸다. 그림 3은 고장의 원인이 다른 구성요소에 전파되어 증상이 나타나는 과정과, 해결책을 적용하였을 때, 다른 구성요소에도 그 영향력이 미쳐서, 문제가 해결되는 과정을 나타낸다.

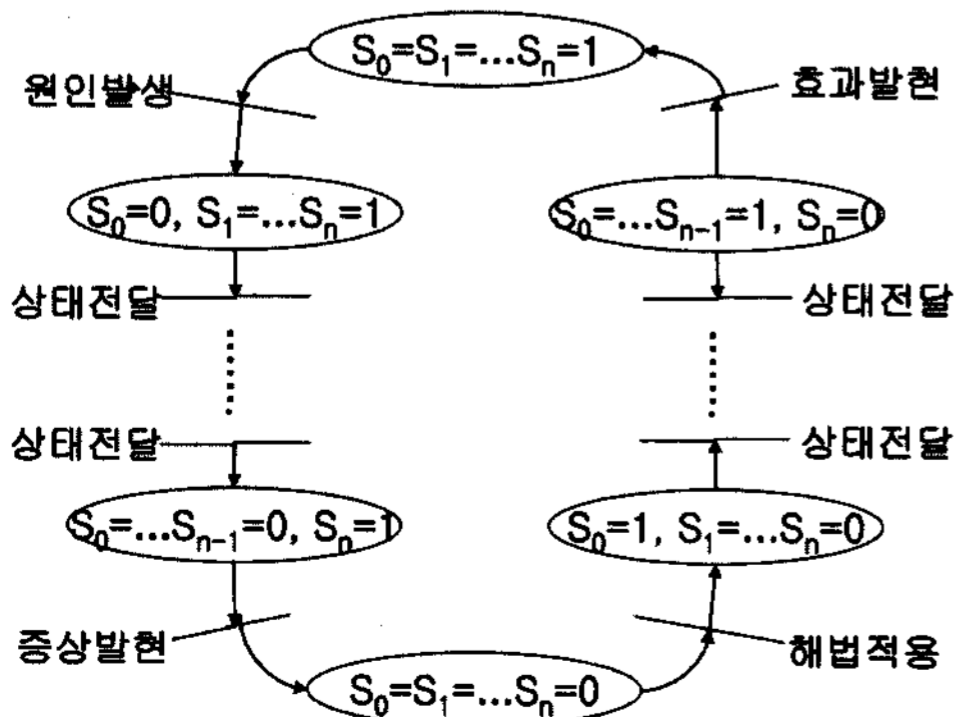


그림 3 전자장비 시스템의 고장발생 및 해결 과정

그림 3에서 보듯이, 이 과정에서, 원인발생, 상태전달, 증상발현, 해법적용, 효과발현이라는 5가지 유형의 사건이 발생한다. 각 사건은 온톨로지 상에서 표 2의 인과범주로 표현된다. 원인발생 사건은 원인 범주로, 상태전달 사건은 전달 범주로, 증상발현 사건은 증상 범주로, 해법적용 사건은 처리 범주로, 효과발현 사건은 효과 범주로 각각 나타난다. 다음에, 표 3의 화상대화 시스템의 고장을 예로 들어 설명하고자 한다.

표 3 화상대화에서의 고장진단에 대한 질의응답 예제

Question : 넷미팅의 화면이 보이지 않습니다.  
Answer : 화상 카메라의 디바이스 드라이버에서 발생한 문제일 수 있습니다. 디바이스 드라이버를 교체해 보세요

표 3의 질의응답에서 나타나는 주요 사건을 그림 3의 형태로 표현하면 그림 4와 같다.

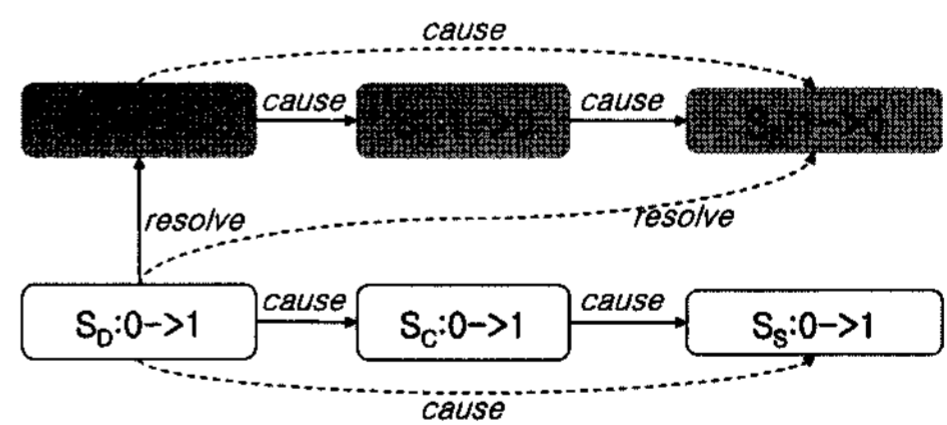


그림 4 화상대화 시스템의 고장 및 해결 과정

그림 4에서,  $S_s$ 는 화면디스플레이의 상태를,  $S_c$ 는 화상카메라의 상태를,  $S_D$ 는 디바이스 드라이버의 상태를 나타낸다. " $S_x:1 \rightarrow 0$ "은 정상부품  $x$ 가 고장나는 사건을 나타내며, " $S_x:0 \rightarrow 1$ "은 고장난 부품  $x$ 가 정상으로 돌아오는 사건을 나타낸다. 즉, 그림 4는, 디바이스 드라이버에 이상이 생겨서, 화상 카메라와 화면 디스플레이에도 이상이 생기는 고장의 전파과정과, 디바이스 드라이버를 교체하여 각 부품이 정상으로 돌아오는 해결과정을 나타낸다. 그림 4에 나타난 사건들 간의 인과관계를 나타내면 그림 5와 같다.

표 4 인과관계로부터 사건을 유도하는 과정

인과관계	인과추론/추정 규칙	추론으로 얻은 관계	사건	범주
changeInto (화면디스플레이, 고장상태)	changeInto(A, 고장상태) → hasState(A, 고장상태)	hasState (화면디스플레이, 고장상태)	Ss:1→0	증상
hasFunction (화상카메라, 화면디스플레이)	hasFunction(A, B) and hasState(B, 고장상태) → hasState(A, 고장상태)	hasState (화상카메라, 고장상태)	Sc:1→0	전달
drive (디바이스드라이버, 화상카메라)	drive(A, B) and hasState(B, 고장상태) → hasState(A, 고장상태)	hasState (디바이스드라이버, 고장상태)	Sd:1→0	전달
replace (사용자, 디바이스드라이버)	replace(A, B) → hasState(B, 정상상태)	hasState (디바이스드라이버, 정상상태)	Sd:0→1	처리
drive (디바이스드라이버, 화상카메라)	hasState(A, 정상상태) and drive(A, B) → hasState(B, 정상상태)	hasState (화상카메라, 정상상태)	Sc:0→1	전달
hasFunction (화상카메라, 화면디스플레이)	hasState(A, 정상상태) and hasFunction(A, B) → (B hasState 정상상태)	hasState (화면디스플레이, 정상상태)	Ss:0→1	효과

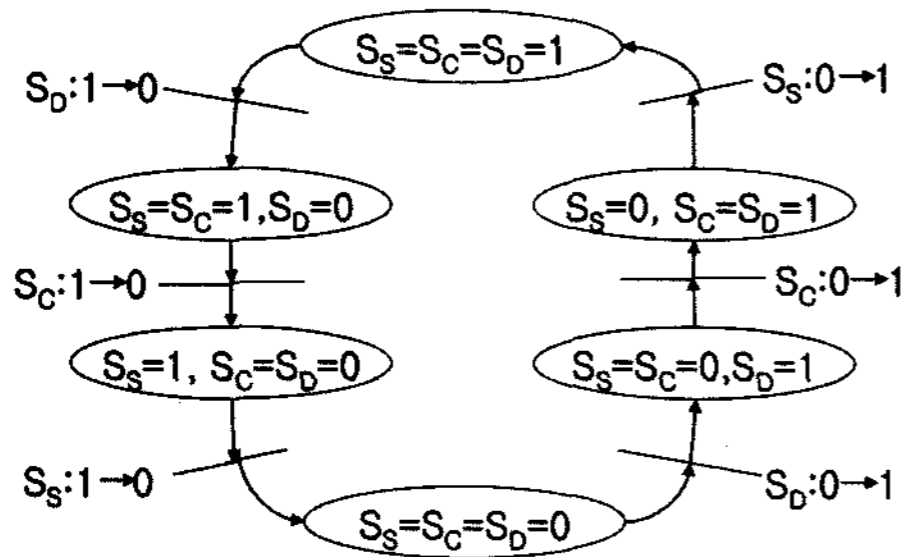


그림 5 화상대화 시스템에서의 사건 간 인과관계

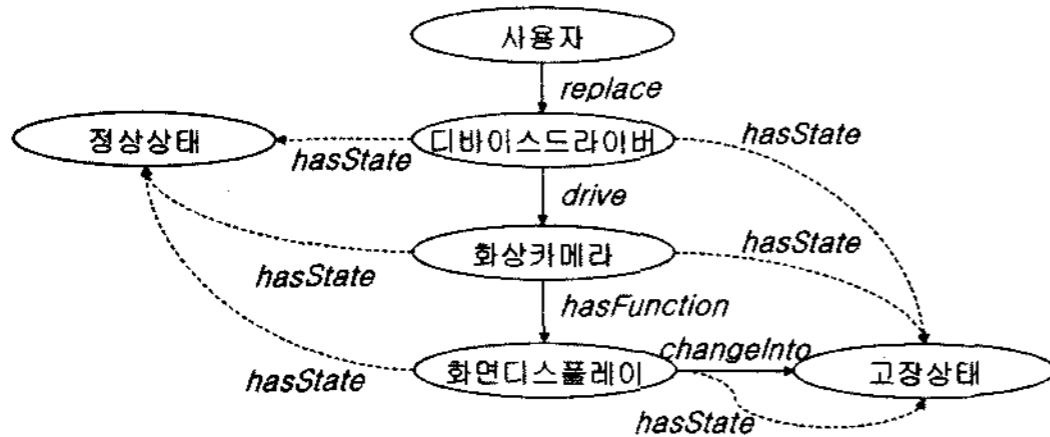


그림 6 고장진단 및 해결을 위한 인과관계 인스턴스

그림 5에서 실선은 표 3의 예제에 직접 드러나는 인과관계를, 파선은 추론 과정에서 나타나는 인과관계를 나타낸다. 회색 노드는 원인을 분석하는 경로를 나타내며, 흰색 노드는 문제를 해결하는 과정을 나타낸다. 그림 6에는 그림 5의 인과사건들을 나타내기 위한 인과관계 인스턴스를 표현한다.

그림 6에서 실선은 온톨로지에 존재하는 관계들을, 파선은 추론을 통하여 얻은 관계를 나타낸다. 표 4는 그림 6의 온톨로지 관계에 인과추정/추론 규칙을 적용하여 그림 5의 인과사건을 도출하는 과정을 표로 나타낸다. 첫번째 열의 관계에 두번째 열의 규칙을 적용하여, 세번째 열의 새로운 관계를 도출한다. 이 과정을 통하여 네번째 열에 나타난 사건을 표현할 수 있다.

### 3.2 고장진단 질의응답에서 인과관계의 정의

본 장에서는 전자장비의 고장진단영역의 FAQ 웹사이트<sup>4)</sup>에서 인과추론에 참여하는 인과관계 인스턴스를 추출하고, 추출한 인과관계 인스턴스에 의미관계를 부여하며 인과범주를 설정함으로써, 각 인과범주별 인과관계를 정의한다.

#### Algorithm: DefineCausalRel(QA)

**Input:**

QA={{(q,a)|q is a question and a is an answer}

C={{c|c is a causal category}

**Output:**

RC={{(r,c)|r :a relation, c:a causal category}

- 1 RC={}
- 2 **foreach** qa in QA:
- 3 N={n|n is a node whose label is a term in qa}
- 4 G=N
- 5 **foreach** n1 in N:
- 6 **foreach** n2 in N:
- 7 G<sub>new</sub>(n1,n2)={p|p is a path of semantic relations between n1 and n2}
- 8 G=GUG<sub>new</sub>(n1,n2)
- 9 **end**
- 10 **end**
- 11 R={r|r=label(e), where e is an edge in G}
- 12 **foreach** r in R:
- 13 **foreach** c in C:
- 14 **if** category(r,qa)==c
- 15 RC=RC U {(r,c)}

4) 삼성전자 데스크탑 컴퓨터의 고장진단을 위한 FAQ(<http://www.sec.co.kr>)와 휴렛팩커드 레이저젯 5000 프린터의 FAQ(<http://h50177.www5.hp.com/support/C4110A/solve/>) 사이트



16 end  
17 end  
18 end

위의 알고리즘에서 N은 질의응답에 나타난 용어들을 노드로 표현한 집합이다. label(n), label(e)는 노드의 용어 레이블과 에지의 관계 레이블을 의미하며, category(r, qa)는 특정한 관계 r의 특정 질의응답 qa상에서의 인과범주를 의미한다.

표 3의 질의응답 예제를 위의 분류 알고리즘에 적용하는 과정을 예시하면 다음과 같다. 먼저, 질의응답 상의 용어들을 이용하여 그림 7과 같이 그래프를 초기화한다.

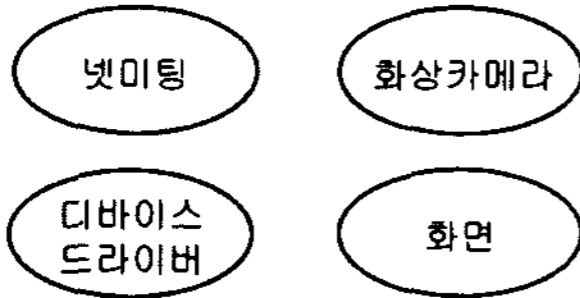


그림 7 질의응답 내의 용어를 이용한 그래프 초기화

현재의 그래프에 속한 모든 노드 사이에 경로를 생성한다. 각각의 경로는 노드가 나타내는 개념 사이의 의미관계로 연결시킨다. 현재의 노드만으로 두 노드가 연결되지 않는 경우는, 한 개의 노드와 관계를 추가한다. 이와 같은 방법으로 모든 노드 사이에 경로가 있는 그래프를 생성한다. 그림 8은 그림 7의 그래프를 알고리즘에 따라, 점진적으로 확장하여 약한 연결 그래프를 만든 결과이다.

그림 8과 같은 그래프에서 에지에 해당하는 각 의미관계별로 추론에서 필요한지를 검토하여 그 역할에 의하여 인과범주를 할당한다. 각 의미관계를 할당된 인과범주에 속하는 인과관계로 정의한다.

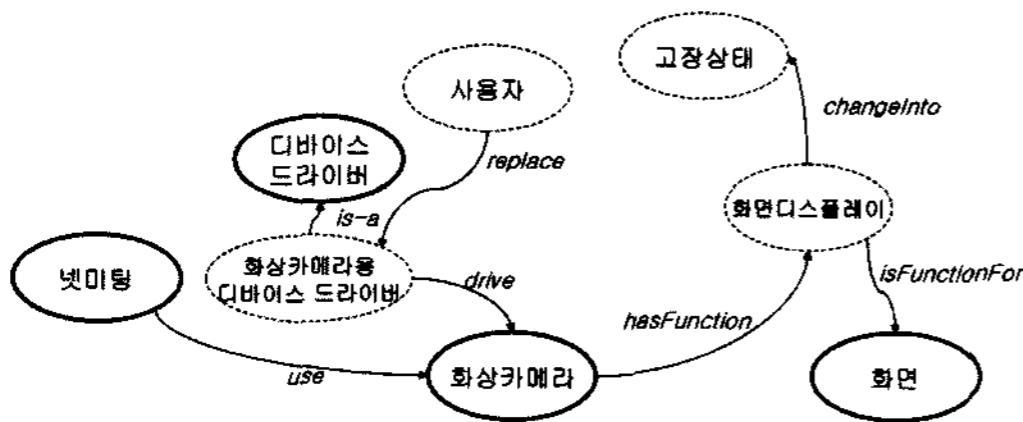


그림 8 의미관계에 의한 약한 연결 그래프



그림 9 질의응답 추론에 필요한 의미관계 연쇄

그림 8의 개념 및 관계들로부터, 질의응답의 추론에 필요한 관계를 추적하면 그림 9와 같다.

그림 9에 나타난 관계들을 인과범주별로 분류하면, 표 5와 같이 정리할 수 있다. 즉 질의응답으로부터 추출한 의미관계 인스턴스를 고장진단 질의응답을 위한 인과범주를 이용하여 분류한다.

표 5 각 의미관계를 인과범주에 대응시킨 예제

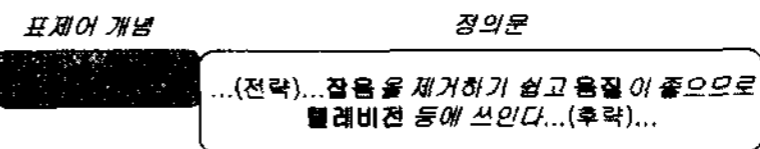
changeInto	: 증상
hasFunction, drive	: 전달
replace	: 처리

#### 4. 고장진단 질의응답을 위한 인과관계 인스턴스 추출

본 절에서는 3.2에서 인과범주별로 정의한 각 인과관계의 인스턴스를 정보통신분야 용어 정의문<sup>5)</sup>에서 추출하는 과정을 설명한다.

본 연구에서 질의응답 말뭉치가 아닌 용어 정의문을 추출 대상으로 삼는 이유는, (1) 정의문에는 분야에서 자주 쓰이는 간단하고 유용한 어휘 패턴들("에 쓰이다",

입력 정의문:



출력 인과관계 인스턴스 집합:



그림 10 인과관계 인스턴스 추출의 입출력 예제

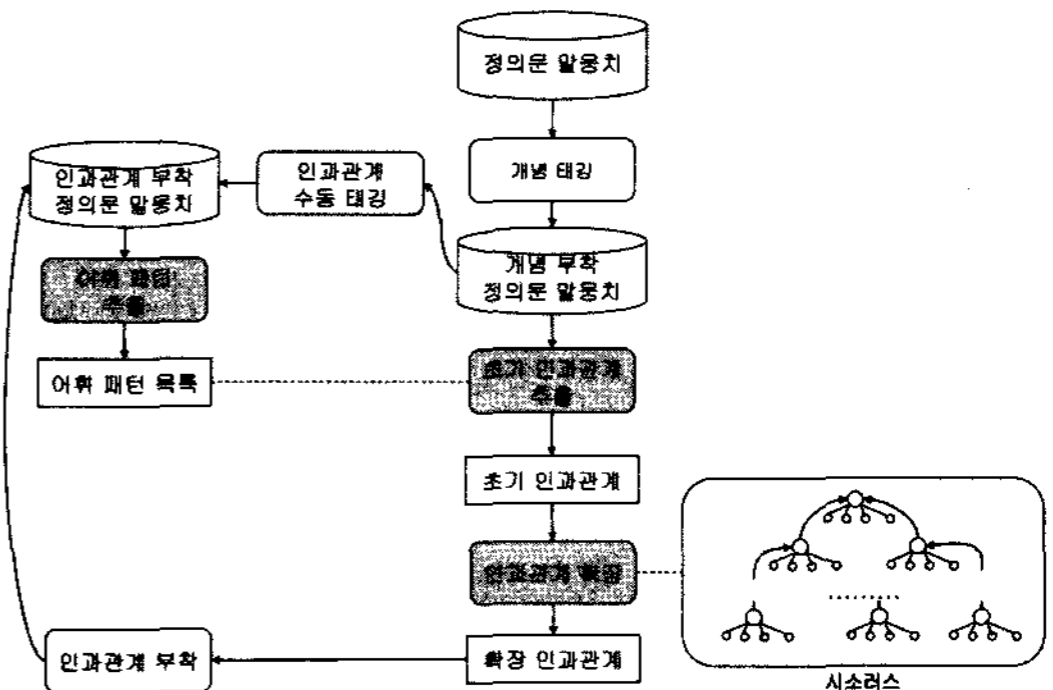


그림 11 어휘 패턴과 시소러스 정보를 이용한 인과관계 인스턴스 추출 모델

5) 한국정보통신기술협회에서 제공하는 정보통신 용어 사전 (<http://word.tta.or.kr/>)을 사용한다.

“를 제어하다”)이 많이 나타나는 반면, (2) 질의응답 말뭉치의 패턴 구조(예: “A 중 B의 실행이 안 됩니다.”)는 상대적으로 다양한 형태로 나타나기 때문에, 다량의 질의응답 말뭉치가 필요한데, 이를 획득하기 위한 소스를 구하기 쉽지 않으며, (3) 구성이 가능하다고 하더라도 말뭉치 상의 하나의 패턴이 두 개 이상의 의미관계로 구성된 네트워크로 표현되어야 하는 경우가 자주 나타나서, 패턴 규칙의 정규화가 어렵다.

분야 정의문에는 특히 표제어 용어의 개념과 정의문 상의 용어의 개념 사이의 관계가 분명하게 나타난다. 따라서, 본 연구에서는 표제어 용어와 정의문 상의 용어 간에 나타나는 인과관계를 추출 대상으로 삼는다. 그림 10에 본 논문에서 수행하는 인과관계 인스턴스 추출의 입출력 예제를 보인다.

그림 11은 본 논문에서 제안하는 인과관계 인스턴스 추출 모델을 나타낸다. 그림에서 보듯이, 본 논문에서는 어휘 패턴을 추출하고, 이에 기반하여 초기 인과관계 인스턴스를 추출하며, 시소러스 정보를 이용하여 인과관계 인스턴스를 확장한다. 확장한 인과관계 인스턴스는 다시 어휘 패턴을 추출하는 데 사용하며, 이와 같은 과정이 반복되면서, 인과관계의 집합과 인과패턴의 집합이 점진적으로 확장된다. 각각의 과정을 각 절 별로 설명하고자 한다.

4.1 어휘 패턴의 추출

본 절에서는 초기 인과관계 추출을 위한 어휘 패턴의 추출에 대하여 설명한다. 어휘 패턴의 추출을 위하여, 20개의 정의문 집합으로부터 표제어 용어의 개념과 정의문 상용어의 개념 간의 3장에서 정의한 인과관계를 수동으로 추출한다. 본 논문에서는, 표제 개념이 tw인 정의문에 속한 한 문장에서, 임의의 개념 w와 w의 오른쪽에 위치하며 가장 가까운 개념 w' 사이에 나타나는 부분을 tw와 w 개념 간의 인과관계 인스턴스의 어휘 패턴이라 정의한다.

그림 10의 예에서, 표제 개념인 “주파수 변조”와 정의

문 상의 개념 “음질” 사이에 “makeBetter”이라는 “효과” 범주의 인과관계가 있다고 판단한 경우, “음질”과 “음질”의 바로 뒤에 위치한 “텔레비전” 사이의 “이 좋으므로”가 “주파수 변조”와 “음질” 사이의 인과관계 인스턴스에 대한 어휘 패턴이 된다. 표 6은 각 인과범주와 인과관계별로 할당한 어휘 패턴들 가운데 자주 나타나는 경우들을 보이고 있다. 표 6의 모든 관계에서 정의역은 <TERM>이 발생한 정의문의 표제 용어이고, 치역은 <TERM>이 된다.

4.2 어휘 패턴을 이용한 인과관계의 자동추출

어휘 패턴을 이용하여 인과관계를 자동으로 추출하는 방법은 다음과 같다.

Algorithm: CausalExtract(LP, TT)

Input:

LP\_TRIPLE={ (lp,r,c) | lp is a lexical pattern, r is a relation represented by lp, and c is a causal category containing r }

TT={ tt | tt is a title term }

Output:

CR={ (cau, rel, cat, res) | cau is a title term, res is a term in definition sentence of cau, rel is a causal relation between cau and res, cat is a causal category of rel }

```

1  foreach lp_triple in LP_TRIPLE:
2    lp=the first element of lp_triple
3    r=the second element of lp_triple
4    c=the third element of lp_triple
5    foreach tt in TT:
6      def(tt)=definition sentence of tt
7      T={ t | t is a term in definition sentence of tt }
8      foreach t in T:
9        if ( t is the rightest term in def(tt) ):
10         lp(t)=the right context of t in def(tt)
11       else:
12         t'=the leftmost term in the right context
           of t in def(tt)
13         lp(t)=lexical part between t and t'
14       endif
15       if (lp(t)==lp):
16         insert (tt, r, c, t) into CR
17       endif
18     end
19   end
    
```

표 6 자주 발생하는 인과관계의 어휘 패턴

인과범주	인과관계	어휘 패턴
원인	Generate	<TERM>가 많이 생기므로
전달	Enable	<TERM>를 가능하게 하는
전달	Transmit	<TERM>를 전송하는
전달	Provide	<TERM>를 제공할 수 있게 된다
전달	Control	<TERM>를 제어하는
전달	isUsedIn	<TERM>에서는 많이 쓰인다
처리	Complete	<TERM>가 완전히 이루어지도록 하는 것
처리	Remove	<TERM>을 제거하기 쉽고
효과	Increase	<TERM> 등을 늘리고
효과	isAppropriateFor	<TERM>에 적합하다

표 7 인과관계 자동추출 예제

인과범주	개념1	관계명	개념2	관계패턴
원인	주파수 변조	Generate	측파대	<TERM>가 많이 생기므로
전달	처리 제어	Control	방송기기	<TERM>를 제어하는
전달	크로미넌스 채널	Transmit	크로미넌스 신호	<TERM>를 전송하는
처리	고선명 텔레비전	Complete	대형화	<TERM>를 실현시킨
효과	고선명 텔레비전	Increase	가로세로비	<TERM>를 늘리고
효과	FM 방송	isAppropriateFor	스테레오 방송	<TERM>에 적합하다

어휘 패턴과 그것이 나타내는 인과관계, 그 인과관계가 속하는 인과범주의 정보를 나타내는 집합 LP-TRIPLE이 입력으로 들어와서, 추출된 인과관계 인스턴스와 그것이 속하는 인과범주를 나타내는 CR이 출력으로 나온다. 정의문 상의 모든 용어에 대하여, 오른쪽 문맥으로부터 어휘 패턴 후보를 찾아서, 기존의 어휘 패턴과 매칭한 후, 매칭이 되면, 인과관계 인스턴스로 판단하여 추출한다. 이와 같은 방법으로 자동 추출한 인과관계 인스턴스의 예는 표 7과 같다.

4.3 시소러스 정보를 이용한 관계의 확장

4.2에서 구한 초기 인과관계의 정의역과 치역에 해당하는 개념들에 속한 어휘들의 시소러스 관련어 정보를 참조하여, 정의문 상에서 더 많은 인과관계 인스턴스를 추출한다. 이 때, 시소러스는 CoreNet을 사용하며, CoreNet의 동일한 개념 노드에 속한 어휘들을 시소러스 관련어로 정의한다. 확장 알고리즘은 다음과 같다.

Algorithm: CausalExpand(CR)

Input:

CR={ (cau,rel,cat,res) | cau is a title term, res is a term in definition sentence of cau, rel is a relation, cat is a causal category }

Output: updated set of CR

```

1  foreach cr=(t1,r,c,t2) in CR:
2    CAU={t1}U{t|relWord(t,t1)=true}
3    RES={t2}U{t|relWord(t,t2)=true}
4    foreach cau in CAU:
5      foreach res in RES:
6        if cau is a title term and res is a term in
           the definition sentence of cau
7          CR=CR U {(cau,r,c,res)}
8    end
9  end
10 end
    
```

입출력으로 쓰이는 CR은 인과관계 인스턴스의 집합이다. 즉, 4.2에서 추출한 인과관계 인스턴스의 집합이

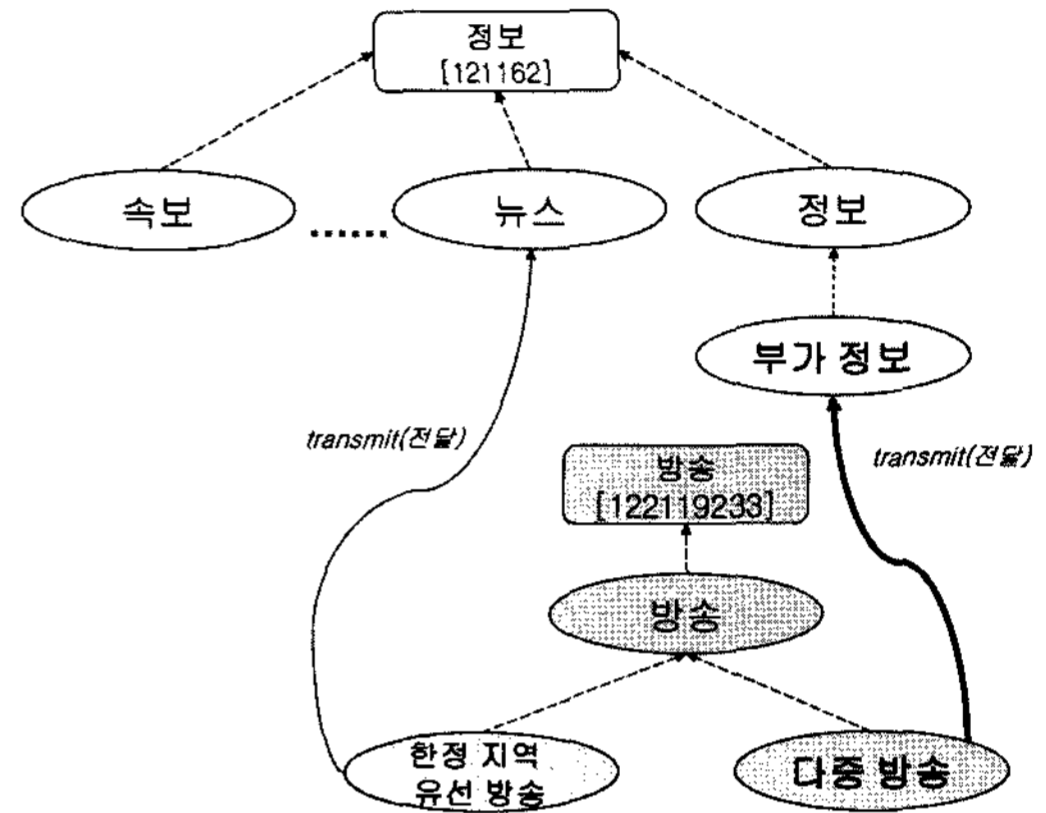


그림 12 시소러스 정보를 이용한 인과관계 확장

입력으로 들어가서, 새로운 인과관계 인스턴스가 추가된 집합이 출력으로 나온다. relWord(w1,w2)는 (1) w1과 w2가 동일한 중심어를 포함하거나, (2) w1에 포함된 어휘와 w2에 포함된 어휘가 시소러스 상에서 같은 개념노드에 속하는 관련어인 경우 true의 값을 반환한다.

인과관계의 확장을 위하여 CoreNet을 이용한다. CoreNet의 각 개념 노드는 그 개념을 나타내는 어휘들을 포함한다.

그림 12에서, 둥근 사각형은 개념 노드를, 원형 노드는 CoreNet의 각 개념 노드에 속하는 어휘 노드를 의미하며, 굵은 화살표는 초기 인과관계를, 가는 화살표는 확장된 인과관계를, 파선은 CoreNet 상의 개념과 어휘 간 관계를 의미한다. 그림 12는 4.2에서 추출한 “다중 방송”과 “부가 정보” 사이의 초기 인과관계 인스턴스를 이용하여, “한정 지역 유선 방송”과 “뉴스” 사이의 인과관계 인스턴스 후보를 추출하는 과정이다. CoreNet 체계에서, “뉴스”와 “정보”는 관련어이므로, “뉴스”와 “부가 정보” 역시 관련어로 간주한다. 또한, “한정 지역 유선 방송”과 “다중 방송”은 동일한 중심어를 가진 관련어이다. 따라서, “한정 지역 유선 방송”과 “뉴스” 사이의 인과관계 인스턴스를 “다중 방송”과 “부가 정보” 사이의 인과관계 인스턴스로부터 도출한다.

5. 실험



본 장에서는 질의응답으로부터의 인과관계 정의 및 분야 정의문으로부터의 인과관계 인스턴스 추출을 위한 실험 및 평가결과를 설명하고, 추출한 인과관계의 온톨로지 구축에의 유용성을 평가한다.

**5.1 인과관계 정의**

**5.1.1 인과관계 정의 실험 결과**

본 절에서는 고장진단영역 질의응답 응용을 위한 인과관계를 정의한 결과를 제시하고 분석한다. 전자장비 고장진단 사이트<sup>6)</sup>와 프린터 고장진단 사이트<sup>7)</sup>로부터 총 26개의 질의응답을 분석하여, 질의응답에 필요한 137개의 인과관계 인스턴스를 수동으로 추출하였다. 추출한 인과관계는 총 84가지 종류이며, 인과범주별로 추출 결과를 제시하면 표 8과 같다.

표 8에서 보듯이, “전달”과 “처리” 범주가 가장 다양한 인과관계로 나타난다. “전달” 범주는, 고장상태의 전달과 처리상태의 전달에 공통적으로 나타날 수 있고, 표 3의 예제에서와 같이 연쇄 형태로 나타나는 경우가 많으므로, 추출된 인과관계 인스턴스 수 역시 가장 많다.

빈도수가 비교적 높은 인과관계의 유형은 표 9와 같다. 표 9는 각 관계들의 인과범주와 정의역, 치역을 명시한다. 정의역과 치역에는 Object, Function, Attribute의 세 가지를 할당한다. “전달” 범주(use(그래픽메모리, 시스템 메모리))와 “처리” 범주(use(사용자, 가상화면 유틸리티))에 속하는 “use”와 같이 다양한 인과범주에 속

표 8 인과범주별 인과관계 정의 및 인과관계 인스턴스 추출 결과

인과범주	인과관계 수	인과관계 인스턴스 수
처리	23	28
효과	13	13
전달	34	50
원인	8	10
증상	6	36

표 9 인과관계 정의 결과의 일부

인과관계	인과범주	정의역	치역
use	전달	Object	Object
hasFunction	전달	Object	Function
isFunctionOf	전달	Function	Object
isAttributeOf	전달	Attribute	Function
setTo	처리	Object	Function
drive	전달	Object	Object
harm	원인	Object	Object
needCompatabilityWith	원인	Object	Object
isPartOf	전달	Object	Object

6) <http://www.sec.co.kr>  
7) <http://www.hp.co.kr>

하는 경우도 있다.

**5.1.2 인과관계 정의에 대한 평가**

인과관계 정의에 대하여서는 자동으로 평가하기 위한 지침이 아직 없으므로, 본 연구에서는, 분야 전문가에게 평가를 의뢰한다. 즉, 인과관계 정의에 참여하지 않은 분야 전문가 3명에게 정의 및 분류 결과에 대하여 얼마나 동의하는지의 여부를 문의한다.

**인과범주 정의에 대한 평가 :**

응용을 위한 인과범주를 정의한 작업의 평가를 위하여, 평가자들이 정의 결과에 대하여 동의하는지의 여부를 문의한 결과, 대체적으로 동의한다는 대답을 얻었다. 단, “전달” 관계의 경우, 텍스트상에 명확하게 드러나지 않는 경우가 많고 다양한 유형이 존재해서, 판단이 힘들므로, 세부유형분류를 하자는 의견이 있었다.

**관계 분류에 대한 평가 :**

설정된 인과범주에 대하여 각 질의응답을 표현하기 위한 관계들을 추출하고 분류한 결과는 다음의 두 가지의 기준에 의하여 평가한다.

- (1) 질의응답을 수행하는 데에 필요한 관계를 적절하게 생성하였는지의 여부  
이에 대한 판단을 위하여, 평가자에게 다음의 세 가지 사항을 물었다.
  - 각 질의응답으로부터 추출한 관계가 추론에 필요한 관계인지의 여부
  - 관계와 개념의 할당과 명명이 적절한지의 여부
  - 각 질의응답에서의 추론에 더 필요한 관계들이 있는지의 여부
 위의 세 가지를 모두 만족한 경우를 정확하다고 판단하여 평가한 결과, 92.3%의 정확률을 얻었다.
- (2) 올바른 인과범주를 각 관계에 할당하였는지의 여부  
(1)의 기준을 모두 통과한 경우에 한하여 인과범주가 적절한지를 문의하였다. 이에 대하여, 98.1%의 정확률을 보였다.

인과관계의 분류에 대한 평가에 있어서, 세 사람의 평가자가 지적한 오류의 개수에는 차이가 있었다. 총 138개의 관계에 대하여, 평가자 A는 29개의 오류를, 평가자 B는 2개의 오류를, 평가자 C는 1개의 오류를 지적하였다. 평가자 A가 지적한 오류의 상당수 (79.3%)와 평가자 B, C가 지적한 모든 오류가 (1)번의 두번째 사항에 대한 것이었다. 예를 들어, “harm(웬바이러스, 컴퓨터 시스템)”과 같이, 말뭉치의 표현을 그대로 옮긴 듯한 경우, 관계의 체계에 의하여 정제시킨 형태로 수정하여야 한다는 의견이 있었다.

둘째 기준에 대한 판단에서는 평가자 A가 7개의 오

류를, 평가자 B가 1개의 오류를 지적하였다. 지적한 오류 가운데 75%가 “전달” 범주 판단의 애매성에 기인한다. 나머지 경우는 “효과”와 “처리” 범주의 애매성이 원인인 오류이다. 이는 인과범주의 정의 결과에 대한 평가와 동일한 맥락을 지닌다.

**5.2 인과관계 인스턴스 추출**

**5.2.1 인과관계 인스턴스 추출 실험 결과**

3.2에서 정의한 인과관계를 IT 방송 분야 20개의 정의문에서 수동 추출하였다. 그 결과, 19가지 종류의 59개 인과관계 인스턴스를 추출하였다. 정의문에서는 주로 “효과”와 “전달” 범주가 많이 나타난다.

수동 추출한 인과관계 인스턴스로부터 표 6과 같은 어휘 패턴을 추출한다. 어휘 패턴을 732개의 용어에 대한 정의문에 적용하고 시소러스 정보를 이용하여 확장한 결과, 126개의 자동 추출 결과를 얻었다. 자동 추출한 인과관계 인스턴스를 범주별로 정리하면 표 10과 같다.

표 10 인과관계 인스턴스 추출 결과

인과범주	인과관계 종류	인과관계 인스턴스의 수	어휘 패턴의 수
처리	3	26	14
효과	5	11	8
전달	9	81	45
원인	1	1	1

**5.2.2 인과관계 인스턴스 추출 성능 평가**

추출 결과에 대한 성능평가는 세 가지 기준에 의한다. 첫째, 고장진단 질의응답을 위한 인과관계 인스턴스 인지의 여부이다. 둘째, 올바른 의미관계를 부여하였는지의 여부이다. 셋째, 올바른 인과범주를 할당하였는지의 여부이다. 의미관계 및 인과범주를 할당한 말뭉치가 존재하지 않으므로, 분야 전문가가 수동으로 검증한다. 세 가지 기준에 대하여 정확률(P)과 평가자들 간의 의견일치율(A)을 (1)과 같이 계산한다.

$$P = \frac{\sum_{i=1}^3 \sum_{j=1}^N e_{ij}}{3N} (e_{ij} = 0, 1)$$

$$A = \frac{\sum_{j=1}^N s_j}{N} (s_j = \begin{cases} 1, & \text{if } e_{0j} = e_{1j} = e_{2j} \\ 0, & \text{otherwise} \end{cases}) \quad (1)$$

N은 인과관계나 관계명의 판단에서는 전체 추출결과 수이며, 인과범주를 판단할 때에는 인과관계와 관계명이 맞다고 판단한 추출결과 수이다.  $e_{ij}$ 는 i번째 평가자가 j번째 추출결과에 대하여 평가한 결과이며, 동의하는 경우 1을, 그렇지 않은 경우 0을 부여한다.  $s_j$ 는 하나의 추출결과에 대하여 모두 동일한 평가를 한 경우 1을, 그렇지 않은 경우 0을 부여한다. 표 11은 설명한 계산 방식에 의한 평가 결과이다. 표 11을 분석하면, 어휘 정보만을 이용한 결과에서는, 일반적으로 정확률(P)이 의견일치율(A)보다 높다. 즉, 올바른 추출에 대한 판단이 불명확한 경우가 비교적 많다. 반면, 어휘정보와 시소러스를 함께 사용한 결과에서는 정확률(P)보다 의견일치율(A)이 높다. 즉, 대체로 맞는 경우와 틀린 경우의 구분이 분명하다.

어휘 정보에 시소러스 정보를 추가한 경우, 정확률이 감소하는 경향이 나타난다. 특히, 관계명 부착의 경우, 30% 이상 감소하는데, 이는 관련어 정보를 이용하여 인과관계를 확장한 경우, 구체적인 관계까지 보존하지는 않기 때문이다. 추출 개수에 있어서는, 어휘 정보만을 이용한 결과는 33개인 반면, 어휘정보와 시소러스 정보를 함께 이용한 결과는 126개이다. 추출 개수가 3.8배 가량 증가한다. 즉, 어휘 정보는 정확성에 있어서, 시소러스 정보는 확장성에 있어서 각기 장점을 가진다.

**5.3 고장진단 질의응답을 위한 온톨로지**

본 절에서는 3.2절과 4.1절에서 얻은 인과관계 인스턴스를 이용하여, 고장진단 질의응답을 위한 온톨로지를 구축한다. 즉, (1) 수동으로 질의응답으로부터 추론 가능한 형태로 추출한 인과관계와 (2) 정의문으로부터 반자동으로 추출한 인과관계를 통합하여, 하나의 온톨로지를 만든다. 그 결과, 263개의 클래스와 56개의 의미관계, 85개의 의미관계 인스턴스로 이루어진 온톨로지를 구축하였다. 이 온톨로지로부터 다음을 관찰하였다.

**새로운 질의응답 추론 가능성의 검토 :**

구축한 온톨로지의 인과관계 가운데  $relation_1(A_1, A_2)$ ,  $relation_2(A_2, A_3), \dots, relation_{n-1}(A_{n-1}, A_n)$ 와 같은 형태의 인과관계 연쇄를 추출한다. 인과관계 연쇄 가운데, 질의응답 추론 예제 구성이 가능한 예는 다음과 같다. isUsedIn(MP3 코덱, 원앰프)와 isUsedFor(원앰프, 인터넷 라디오)와 같은 디바이스 간의 전이 범주의 경우,

표 11 인과관계 인스턴스 자동추출 성능평가

	어휘 정보만을 이용한 결과(33개)			어휘정보와 시소러스를 이용한 결과(126개)		
	인과관계 판단	관계명 부착	인과범주 할당	인과관계 판단	관계명 부착	인과범주 할당
정확률	97.0%	95.0%	96.9%	80.7%	64.3%	83.3%
의견일치율	93.9%	87.9%	93.9%	83.3%	88.9%	84.1%

“인터넷 라디오”의 고장의 원인 혹은 전달 경로가 “MP3 코덱”임을 알 수 있다.

#### 분야 온톨로지와의 비교 :

마지막으로, 새로이 구축한 온톨로지와 기존의 분야 온톨로지의 관계 집합을 비교한다. 이를 통하여, 응용 온톨로지를 위한 인과관계 정의 및 추출 작업이 기존의 분야 온톨로지와 어떠한 차이점을 만들어 내는지를 알아보하고자 한다. 공통적인 의미를 가진 의미관계의 비율을 추정하기 위하여, 관계명에 사용된 단어들을 비교한다. 온톨로지의 관계명을 단어 단위로 분할하여 원형을 구한다. 이 때, 전치사나 be동사 등은 제외한다. 남은 단어들의 온톨로지별 빈도수를 구하고, 그 결과를 서로 비교한다.

97개의 의미관계를 가진 분야 온톨로지에는 113개(중복제거 75개)의 단어가, 응용 온톨로지의 의미관계에는 81개의 단어가 사용된다. 이 가운데 공통적으로 사용되는 단어는 function, use, error, control, property의 5개이다. 즉, 공통적인 의미를 가진 의미관계가 나타날 확률이 매우 적다.

단어의 품사별로 비교한 결과, 분야 온톨로지는 명사가 66개로 가장 많이 나타나며, 응용 온톨로지는 동사가 34개로 가장 많이 나타난다. 즉, 응용 온톨로지는 동적인 의미관계가 주를 이루는 반면, 분야 온톨로지는 정적인 의미관계가 주를 이룬다.

## 6. 결론

온톨로지에서의 인과관계는 응용의 추론에서 주된 역할을 하므로, 관계 정의와 추출시 응용에서 필요로 하는 추론의 형태를 고려하여야 한다.

본 연구에서는 고장진단 질의응답을 위한 온톨로지에서의 인과관계를 정의하고 추출함에 있어서, 질의응답 말뭉치를 이용하여 인과관계를 정의하고, 분야 정의문으로부터 인과관계 인스턴스를 점진적으로 추출한다. 어휘 패턴과 시소러스 정보가 인과관계 인스턴스 추출에 이용된다. 어휘 패턴은 용어 사이의 관계에 특정한 의미관계를 부여하고, 인과범주에 속하는지의 여부를 판단하기 위한 명확하고 신뢰도 높은 단서를 제공한다. 정확 매칭에 의하여서만 결과를 산출하므로, 제한된 경우에만 적용되는 어휘 패턴의 단점을 보완하기 위하여, 시소러스의 관련어 정보를 사용하여 적용률을 향상시킨다.

분야 전문가들의 평가 결과, 인과관계의 인과범주별 분류 결과는 92.3%의 평균 정확률을 보이면서, 본 논문의 인과범주 정의가 사람의 직관과 유사함을 보였다. 추출 결과는 인과관계 인식에 있어서 80.7%의 정확률을 보이면서, 어휘 정보와 시소러스 정보가 정의문으로부터 인과관계 인스턴스를 추출하는데 좋은 단서를 제공함을

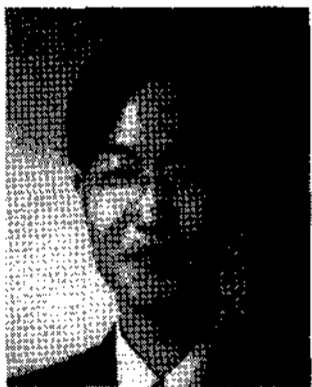
보였다.

본 논문에서 제시된 모델을 다음과 같은 방향으로 확장하는 연구를 진행 중이다. 첫째, 인과관계의 정의 단계에서 분야 정보를 고려할 수 있다. 본 연구에서는 인과관계 인스턴스를 추출할 경우에만 분야 정의문의 분야 정보를 사용하였다. 하지만, 질의응답에서 발견된 특정 관계에 의미관계를 부여하는 경우와 같이, 분야 정보가 필요한 경우가 있으므로, 인과관계를 정의할 때도, 질의응답에서 발견된 인과관계 인스턴스에 분야 정의문으로부터의 분야 정보를 고려하여 의미관계를 자동으로부터 부여하는 모델을 제안하고자 한다. 둘째, 인과관계 인스턴스의 추출에 있어서, 정의문 상의 용어 간의 인과관계 인스턴스 추출로 범위를 확장하고자 한다. 표제 용어와의 관계와는 달리, 한 문장에서의 인과관계 인스턴스 추출은 어휘 정보 이외의 다양한 정보를 필요로 한다. 따라서, 구문 정보, 의미 정보 등을 패턴화시키는 방법을 고려하는 중이다. 셋째, 고장진단 질의응답으로부터 인과관계의 정의 과정과 분야 정의문으로부터 인과관계 인스턴스의 추출 과정의 상호 연관성을 증대시키는 방법을 제시한다. 구체적으로, 인과관계의 정의 과정에서 인과관계 인스턴스 추출의 대상이 되는 개념 어휘와 인과관계 인스턴스 추출의 단서가 되는 패턴을 질의응답 말뭉치로부터 가져와서, 인과관계 인스턴스 추출의 초기 개념 어휘 및 초기 패턴으로 사용하고자 한다.

## 참고 문헌

- [1] Merriam-Webster Online Dictionary. <http://www.m-w.com>. 2005.
- [2] 윤평현, 국어의 접속어미 연구. 한신문화사. 1989.
- [3] D. Chang and K. Choi, Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities, Information Processing & Management, Volume 42, Issue 3, Pages 662-678, 2006.
- [4] R. Girju, Automatic Detection of Causal Relations for Question Answering, In Proceedings of the 41st ACL, Workshop on Multilingual Summarization and Question Answering, 2003.
- [5] R. Girju and D. Moldovan. Mining Answers for Causation Questions, AAAI Symposium on Mining Answers from Texts and Knowledge Bases. 2002.
- [6] 이신목, 김현수, 황금하, 최기선, “한국어 특허문서상에서의 인과관계 관찰 및 추출,” 2006년도 한국인지과학회 춘계학술대회, 2006.
- [7] C. S. G. Khoo, J. Kornfilt, R. N. Oddy, and S. H. Myaeng. Automatic Extraction of Cause-Effect Information from Newspaper Text without Knowledge-based Inferencing. Literary and Linguistic Computing. Volume 13, Number 4, pp. 177-186. 1998.

- [8] Y. Kitamura and R. Mizoguchi, Functional Ontology for Functional Understanding, Twelfth International Workshop on Qualitative Reasoning, pp. 77-87, 1998.
- [9] P. Pantel and M. Pennacchiotti, Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations, joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics, 2006.
- [10] V. Nastase and S. Szpakowicz. Exploring Noun-modifier Semantic Relations. In Fifth International Workshop on Computational Semantics(IWCS-5), pages 285-301, Tilburg, the Netherlands. 2003.
- [11] J. Kim, Causes and Events: Mackie on Causation, Journal of Philosophy, Vol. 68, 1971, pp. 426-41. Reprinted in E. Sosa, ed., Causation and Conditionals, Oxford University Press, 1975.
- [12] R. Mizoguchi, *オントロジー工学, 人工知能學會 編集. オーム社. 2005.*
- [13] J. F. Sowa, Processes and Causality. Available at: <http://www.jfsowa.com/ontology/causal.htm>. 2002.
- [14] N. Guarino, Formal Ontology and Information Systems. In. Proceedings of the First International Conference on Formal Ontologies in. Information Systems (FOIS), pp. 3-15. 1998.



#### 이 신 목

1999년 한국과학기술원 전자전산학과(학사). 2001년 한국과학기술원 전자전산학과(석사). 2001년~현재 한국과학기술원 전자전산학과 박사과정. 관심분야는 온톨로지, 관계추출, 질의응답

#### 신 지 애

정보과학회논문지 : 소프트웨어 및 응용  
제 35 권 제 2 호 참조