

## 지식베이스를 이용한 작업자 증상 기반 화학물질 추정 시스템 설계

주용택·이동훈\*·신은지\*\*·유상우\*\*\*·†신동일\*\*\*\*

명지대학교 재난안전학과 박사과정, \*명지대학교 재난안전학과 박사, \*\*명지대학교 재난안전학과 석사, \*\*\*명지대학교 재난안전학과 석사과정, \*\*\*\*명지대학교 재난안전학과 교수  
(2021년 4월 23일 접수, 2021년 6월 18일 수정, 2021년 6월 19일 채택)

### Worker Symptom-based Chemical Substance Estimation System Design Using Knowledge Base

Yongtaek Ju · Donghoon Lee · Eunji Shin · Sangwoo Yoo · †Dongil Shin

Dept. of Disaster and Safety, Myongji University, Yongin 17058, Korea

(Received April 23, 2021; Revised June 18, 2021; Accepted June 19, 2021)

#### 요약

본 논문에서는 산업현장 화학물질 인체 접촉 증상 기반 지식베이스 구축 및 화학물질 추정 시스템 설계에 대한 연구이다. 미국NIH에서 제공하는 WISER 프로그램의 499개의 화학물질 접촉 증상 정보로 활용하였다. 지식베이스 구축을 위해 AllegroGraph 7.0.1 프로그램을 이용하였으며 입력된 Chemical structure로 Triple 값인 Cas No., Synonyms, Symptom, SMILES, InChI, Formula를 사용 하였다. 또한 작업자의 증상을 안내하는 방법은 AI 스피커를 활용한 방식이 가능하며 지식베이스 구축 결과 암모니아(CAS No: 7664-41-7)를 기준으로 39개의 증상이 WISER 프로그램과 동일함을 확인 하였다. 이를 통해 화학물질 추정 시스템의 증상 추출 과정에 지식베이스 구축이 가능하였다.

**Abstract** - In this paper, a study on the construction of a knowledge base based on natural language processing and the design of a chemical substance estimation system for the development of a knowledge service for a real-time sensor information fusion detection system and symptoms of contact with chemical substances in industrial sites. The information on 499 chemical substances contact symptoms from the Wireless Information System for Emergency Responders(WISER) program provided by the National Institutes of Health(NIH) in the United States was used as a reference. AllegroGraph 7.0.1 was used, input triples are Cas No., Synonyms, Symptom, SMILES, InChI, and Formula. As a result of establishing the knowledge base, it was confirmed that 39 symptoms based on ammonia (CAS No: 7664-41-7) were the same as those of the WISER program. Through this, a method of establishing was proposed knowledge base for the symptom extraction process of the chemical substance estimation system.

**Key words** : knowledge base, symptoms, chemical substance estimation, WISER, AllegroGraph

#### I. 서론

산업안전보건법에서 명시하고 있는 사고유형으로는 중대재해(사망자 1명 이상 발생, 3개월 이상의 요양을 요하는 부상자가 동시에 2명 이상 발생,

부상자 또는 직업성 질병자가 동시에 10명 이상 발생), 중요사고(사망자 2명 이상 발생, 사상자 3명 이상으로 사망 반드시 포함하여 발생), 중대사고(사망자 3명 이상 발생, 사상자 5명 이상으로 사망 반드시 포함하여 발생)로 구분하고 있다. 이 중 중요사고와 중대사고는 중대산업사고로 산업안전보건법 제 49조의 2항에서 정한 위험물질의 누출·화재·폭발 등으로 인하여 사업장 내의 근로자에게 피해를 주

†Corresponding author:dongil@mju.ac.kr  
Copyright © 2021 by The Korean Institute of Gas

**Table 1.** Number of major accidents(2009~2018)

구분		계	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
발생 건수	중대재해	8,929	911	970	958	974	945	790	831	875	862	813
	중대산업 사고	361	30	34	41	58	40	28	28	36	34	32
합계		9,290	941	1,004	999	1,024	985	818	851	911	896	845
사망 자수	중대재해	9,128	928	987	994	1,029	978	792	840	882	873	825
	중대산업 사고	708	57	58	84	120	84	50	54	69	68	64
합계		9,836	985	1,045	1,078	1,194	1,062	842	894	951	941	889

**Table 2.** Chemical accident law

소관부처	관련 법령	관리 대상	관리제도
환경부	화학물질 관리법	유해화학 물질	위해관리계획서, 장외영향평가서
고용 노동부	산업안전 보건법	유해, 위험물질	공정안전보고서
산업통상 자원부	고압가스 안전관리법	고압가스	안전성향상 계획
소방청	위험물 안전관리법	위험물	예방규정

거나 사업장 인근지역에 피해를 줄 수 있는 사고로 명시하고 있다.

2018년을 기준으로 과거 10년간 안전보건공단에서 조사한 중대재해는 8,929건 중 사망 9,128명, 중대산업사고는 361건 중 사망 708명으로 확인되고 있다. 매년 사고와 인명피해가 지속적으로 발생하고 있는 실정이다. [1]

이 중 중대산업사고의 위험물질에 해당하는 화학물질은 누출, 폭발, 화재 및 복합사고로 이어진다. 또한 작업장 근로자 및 인근 주민의 인명피해로 이어질 수 있어 안전관리 조치가 필요하다. 화학사고의 정의는(화학물질관리법 제2조 13항) 시설의 교체 등 작업시 작업자의 과실, 시설 결함·노후화, 자연재해, 운송 사고 등으로 인하여 화학물질이 사람이나 환경에 유출·누출되어 발생하는 일체의 상황으로 언급되어 있으며 Table 1과 같이 화학사고 관련법의 제도를 통해

관리되고 있다. [2]

현행법에서는 「산업안전보건법」 제27조에 따라 가연성 또는 독성물질의 가스나 증기의 누출을 감지하기 위한 가스누출감지 경보설비 설치, 「화학물질관리법」 제24조제1항, 같은 법 시행규칙 제21조제2항 관련 [별표 5]에 따라 유해화학물질(폭발성, 인화성, 급성독성 유해화학물질) 소량 기준 이상을 취급하는 취급시설에 설치를 의무화하고 있다. 하지만 현행 화학물질 감지 시스템은 알람 형태가 대부분으로 특정 위치에 고정되어 감지되고 있으며 작업자의 화학물질 접촉에 의한 안전관리는 미흡하다. 특히 중대산업 사고 원인 중 하나인 질식사고 예방과 작업자의 떨어짐, 이상온도, 화학물질 중독 등을 감지하기 위해서는 화학물질 접촉 증상과 실시간 센서 정보 융합 감지시스템의 지식서비스 활용이 필요하다. 이에 본 연구에서는 지식베이스를 이용한 작업자 증상 기반 화학물질 추정 시스템을 설계 하였다.

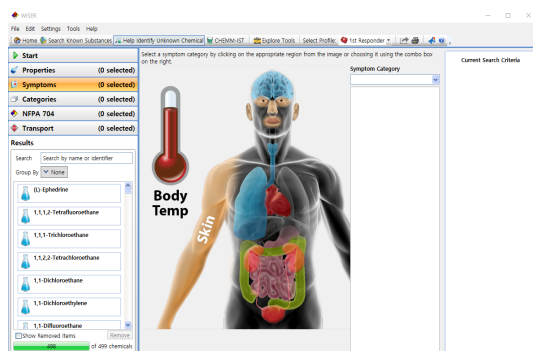
## II. 증상 기반 화학물질 추정 기술

미국 NIH(National Institutes of Health)에서 제공하는 WISER(Wireless Information System for Emergency Responders) 프로그램은 화학물질 사고 시 작업자 및 구조대원을 지원하도록 설계된 시스템으로 물질 식별 지원 (Including substance identification support), 물리적 특성(physical characteristics), 인체 건강 정보(human health information), 격리 및 억제 지침(containment and suppression guidance)을 포함하여 유해 물질에 대한 광범위한 정보를 제공하고 있다. 특히 작업자의 화학물질 접촉 증상 정보를 빠르게 접근할 수 있는 특징을 가지고 있다.

WISER의 증상 분류는 Table 3와 같이 증상 카테고리

**Table 3.** Symptoms of WISER(wireless information system for emergency responders)

Symptom Category	Symptoms	ea
Temperature	chills, high body temp, low body temp, shivering	4
Neurological	agitation, dizziness, drowsiness, fatigue/weakness, headache, lack of coordination, lowered mental state, numbness/tingling, paralysis, slurred speech, spasms/seizures, unresponsive	12
Ears	hearing loss, tinnitus	2
Eyes	constricted pupils, dilated pupils, eye irritation/redness, eye swelling, impaired vision, light sensitivity, tearing, vision loss	8
Nose	bloody nose, nasal irritation, runny nose, sneezing	4
Mouth/Throat	coughing/choking, drooling/salivation, dry mouth/throat, mouth irritation, throat irritation	5
Cardiovascular	arrhythmia, bradycardia, chest pain, hypertension, hypotension/shock, hypoxia/cyanosis, tachycardia	7
Respiratory	chest discomfort, congestion, coughing/choking, coughing blood, hypoxia/cyanosis, irregular breathing, pulmonary edema, rapid breathing, resp burning/irritation, respiratory arrest, shortness of breath, slow breathing, sneezing, wheezing	14
Gastro/Urinary	abdom.discomfort, abdom.distention, diarrhea, nausea, urinary incontinence, urinary pain/burning, urination. bloody, vomiting, vomiting blood	9
Skin	blistering, cyanosis/blue, dry skin, frostbite, itching, jaundice/yellow, numbness/tingling, pale, peeling/exfoliation, rash, skin burns/burning, skin redness, skin swelling, sweating	14



**Fig. 1.** WISER's Worker symptoms program.

리로는 온도, 신경, 귀, 눈, 코, 입/목, 심혈관, 호흡기, 위장/비뇨, 피부 10개로 구분하였으며 세부적으로 총 79개의 증상으로 분류되어 Fig. 1과 같이 인체 이미지에 해당부분을 선택하여 세부 증상을 선택할 수 있다. 또한, 세부 증상을 선택 할 경우 후보 화학물질 결과가 확인된다. 그리고 여러 세부 증상을 중복하여 선택하면 화학물질 범위가 줄어 특정 화학물질에 대한 정확도를 높일 수 있다.[3]

### III. 지식베이스 구축

#### 3.1. 지식베이스

지식베이스는 시맨틱 웹, 질의 응답 등 자연어 처리 분야의 다양한 응용 시스템에 사용되는 중요한 요소이며, 지식베이스 구축에 대한 다양한 연구들이 진행되고 있다. 지식베이스에서 지식은 주로 ‘주어(subject)-서술어(predicate)-목적어(object)’의 트리플 형태로 표현되는 것으로서 2개의 개체와 그 관계로 구성된다. 영어에 비해 한국어를 포함한 다른 언어들은 상대적으로 데이터가 적기 때문에 지식베이스를 효과적인 방법으로 구축하는 것이 중요하다.[4, 5, 6] 영어권에서는 WordNet, YAGO, Cyc, BabelNet과 같은 지식베이스들이 주로 사용되고 있다. 그 중에서도 Wikipedia, WordNet 및 GeoName에서 정보를 추출하여 구축된 YAGO(Yer Another Great Ontology)는 IBM 왓슨(Watson)의 인공지능 시스템 등에 이용된 고성능의 지식베이스로서[7] 2019년 천만 개 이상의 엔티티(인간이 생각하는 개념 또는 정보의 세계에서 의미있는 정보의 단위)를 갖고 있으며 이 엔티티에 대한 1억 2천만개 이상의 사실관계를 포함한다.

### 3.2. AllegroGraph

본 연구에 사용된 AllegroGraph는 시맨틱 웹 애플리케이션을 구축하기 위한 데이터베이스 및 애플리케이션 프레임 워크이다. 정보를 저장하는 방법은 여러 가지가 있지만 W<sub>3</sub>C는 Resource Description Framework (RDF)에서 표준화하였다. RDF는 지식을 assertions(단정)이라 하는 subject(누가), predicate(한다), object(무엇을)로 세 문장과 같이 분해한다. 이러한 분명한 이유로 assertion(단정)을 triple(트리플)이라고 한다. 데이터와 메타 데이터를 Triple로 저장할 수 있다. SPARQL (표준 W<sub>3</sub>C 쿼리 언어) 및 Prolog와 같은 다양한 쿼리 API를 통해 이러한 Triple을 구성한다. 내장된 추론과 함께 RDFS ++ 추론을 적용 할 수 있다.

### 3.3. Chemical Structure

본 연구에서 구축한 Database 중 Chemical Structure는 첫째, SMILES (Simplified Molecular Input Line Entry System)는 ASCII (American Standard Code for Information Interchange) 문자열을 사용하여 화학물질의 구조를 설명하기 위한 표기 형식이다. 일반적으로 분자에 대한 여러 SMILES로 작성 가능하며 canonicalization SMILES를 사용하였다. 둘째, InChI (International Chemical Identifier)는 원자 및 결합 연결성 나타내며 이성질체 정보 등 화학물질에 대하여 설명한다. SMILES 보다 더 많은 정보 표현 이 가능하고 모든 화학물질구조가 고유한 InChI를 가지고 있다.

### 3.4. Transform chemical structure to fingerprints

본 연구에서 구축한 Database 중 Transform chemical Structure to fingerprints는 분자의 구조를 이진법인 bit 문자열(0과1)로 인코딩가능하며 각 bit는 미리 정의 된 구조적 특징을 가지고 있다. 분자에 미리 정의 된 특징이 있는 경우 기능에 해당하면 1, 특징이 없으면 0으로 표시하는 방식이다. 이와 같은 방법으로 첫째, MACCS Keys (Molecular ACCess System Keys) 가장 일반적으로 사용되는 구조의 키 중 하나로 960 bits의 키와 166 bits의 키 중 166 bits 만 공개적으로 사용 가능하다. 둘째, PubChem Keys 881개의 구조 키로 다양한 하위 구조와 특징들을 포괄 PubChem의 화학물질 유사성 탐색에 사용된다.

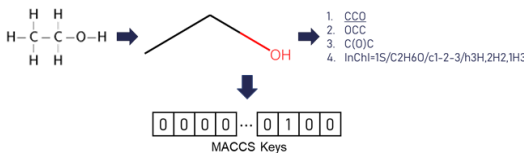


Fig. 2. MACCS Keys.

## IV. 화학물질 추정 시스템 설계

### 4.1. 화학물질 추정 시스템

Fig. 3은 지식베이스 구축 및 Web기반 지식서비스 개발을 위한 시스템의 설계를 나타낸다. 이 시스템은 산업현장에서 예상치 못한 증상이 발생하였을 경우에 접촉된 화학물질을 빠르게 찾아내어 대처할 수 있도록 한다.

센서 정보 및 음성대화 정보를 바탕으로 능동적인 대응에 필요한 화학물질 누출 정보 및 작업자의 증상 정보를 수집한다. 수집한 정보를 통해 누출된 화학물

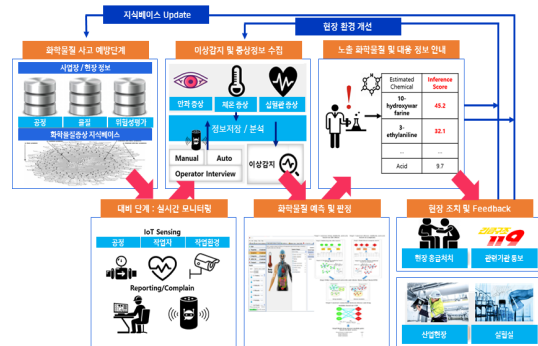


Fig. 3. System design for knowledge service.

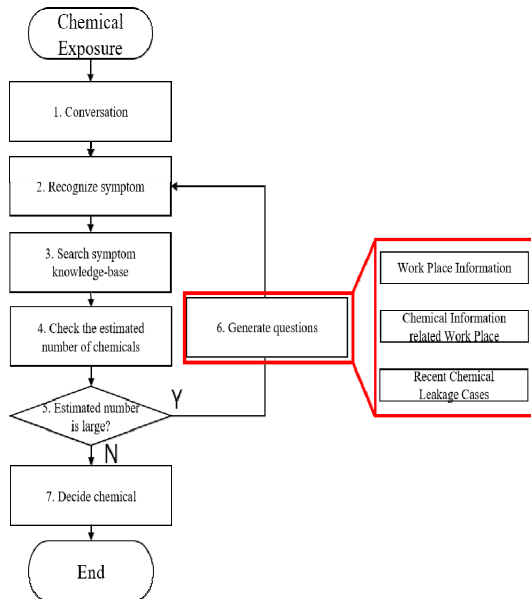


Fig. 4. Flow chart of chemical estimation.

질을 인지하고, 현장 대응정보를 제공하며, 더불어 유관기관에 통보하는 시스템이다.

본 논문에서 제안하는 시스템은 작업자와의 직접적인 의사소통을 통해 정보를 수집하고 대응정보를 전달하는 자연어 인식을 이용한 증상기반 화학물질 추정이다. 작업자에게 증상을 안내할 수 있으며, 시스템 서버로 정보를 전달하고 서버는 정보를 수집할 수 있다.

#### 4.2. 화학물질 추정 방법

Fig. 4는 AI 스피커 등과의 음성 대화를 통해 인체에 발현된 증상을 입력하고, 증상에 기반을 두어 노출된 화학물질을 추정하는 과정을 나타낸다. AI 스피커와 대화를 통해 증상을 기반으로 하여 화학물질을 추정하는 방법은 크게 두 가지가 있다. 첫째, 화학물질이 누출되었을 경우, 증상이 발현된 사용자가 AI 스피커와 대화를 시도하여 증상을 입력한 후 화학물질을 추정하는 방법이 있다. 둘째, 센서 정보를 통해 사용자보다 먼저 화학물질의 누출을 인지하고 AI 스피커가 사용자와 대화를 시도하여 사용자가 발현된 증상을 인식하게 된다. 그리고 대화를 통해 증상을 입력받아 화학물질을 추정하는 방법이 있다.

두 방법 모두, 입력받은 증상정보를 구축한 증상 데이터베이스 서버에 전송하여 Chemical Structure 를 통한 추론으로 화학물질을 추정하며, 이에 따른 대응정보를 제공한다.

추정된 화학물질의 개수가 많은 경우, AI 스피커의 능동적인 질문 생성을 통해 추정 화학물질의 개수를 줄일 수 있다. 질문은 화학물질 누출이 발생한 장소의 화학물질 목록, 최근 누출사고 기록 등의 환경정보를 통해 생성한다. 추정된 화학물질의 개수가 적을 경우, 추정된 화학물질에 대한 대응정보를 제공한다. 대화를 통해 추정된 누출 화학물질 및 증상정보에 대한 기록을 기존 데이터베이스에 저장하여, 추후 비슷한 사고가 발생할 경우 화학물질의 추정 속도를 높일 수 있도록 한다.

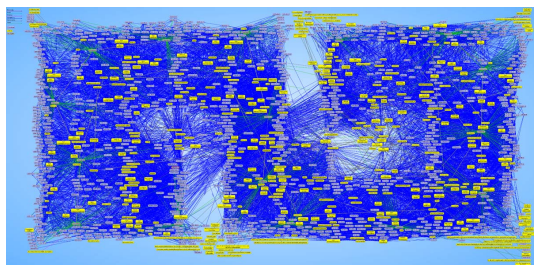


Fig. 5. Knowledge base of symptoms.

AI 스피커와의 대화를 통해 증상을 입력받기 때문에, 언어를 통해 입력받을 수 있으며, 동의어를 입력받아도 증상정보를 입력할 수 있다. 이는 단일 언어를 제공하고, 동의어 입력하기 어려운 WISER와 구분되는 차별점이다.

## V. 결 과

### 5.1. 증상 지식베이스 구축

화학물질을 추정하기 위해서, 화학물질 물성과 증상정보가 담긴 데이터베이스가 절대적으로 필요하다. 본 연구에서는 화학물질 물성과 증상이 노드(node)로, 둘 사이의 관계를 edge로 나타낸 그래프 데이터베이스(graph database)인 지식그래프를 생성하였다. Fig.5는 증상을 기반으로 한 화학물질 추정 방법에서 이용하는 증상 데이터베이스의 지식그래프를 나타낸다. 화학물질의 CAS 번호와 관련된 증상이 edge와 같은 관계(relation)으로 연결되어있으며, 화학물질의 CAS 번호와 화학물질명 또한, edge로 연결되어 있다. 하나의 화학물질은 한 가지 이상 즉, 복수의 증상을 나타낼 수 있으므로, 하나의 화학물질은 복수의 증상과 edge를 통해 연결된다. Wiser의 499개 화학물질을 대상으로 입력된 Triple은 Cas No., Synonyms, Symptom, SMILES, InChI, Formula 이다.

### 5.2. AllegroGraph의 SPARQL

조건문 Where을 사용하여 증상 정보 입력 예시로 headache(두통), pale(창백한), bloody nose(코피), chest

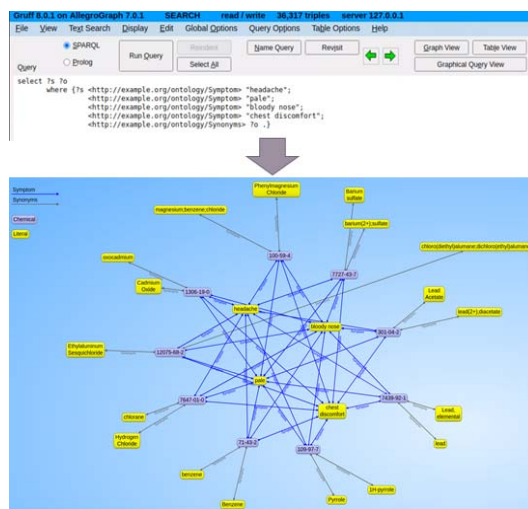
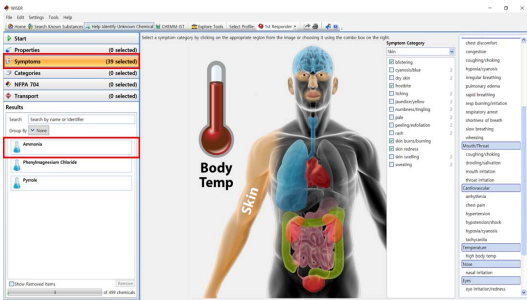
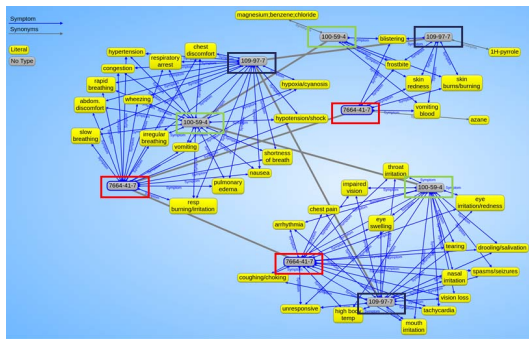


Fig. 6. AllegroGraph's SPARQL.





(a) WISER



(b) AllegroGraph

**Fig. 7.** Ammonia symptoms of (a)WISER, (b)AllegroGraph.

discomfort(흉통) 4개의 증상을 갖는 화학물질 검색 시 select ?s ?o 조건을 만족하는 s, o 출력으로 9개의 화학물질이 Cas No.와 synonyms로 출력되었다.

### 5.3. WISER 프로그램과 비교결과

WISER의 Ammonia 증상은 총 39개로 Ammonia(CAS No: 7664-41-7), Phenylmagnesium Chloride(CAS No: 100-59-4), Pyrrole(CAS No: 109-97-7)의 3개 물질이 확인되었으며 AllegroGraph에 동일한 증상 39개를 입력한 결과 동일하게 Ammonia, Phenylmagnesium Chloride, Pyrrole의 3개 물질이 추출되어 지식베이스 구축 완성도를 확인 하였다.

## VI. 결론

본 연구에서는 산업현장 화학물질 인체 접촉 증상과 실시간 센서정보 융합 감지시스템 지식서비스 개발을 위한 자연어처리 기반 지식베이스 구축 및 화학

물질 추정 시스템 설계에 대해 연구이다. 미국 NIH(National Institutes of Health)에서 제공하는 WISER(Wireless Information System for Emergency Responders) 프로그램의 499개의 화학물질 접촉 증상 정보를 reference로 활용하였다. AllegroGraph 7.0.1를 이용하였으며 입력된 Triple은 Cas No., Synonyms, Symptom, SMILES, InChI, Formula 등 화학물질 정보를 입력하였다.

지식베이스 구축 결과 WISER의 Ammonia 증상은 총 39개로 Ammonia(CAS No: 7664-41-7), Phenylmagnesium Chloride(CAS No: 100-59-4), Pyrrole(CAS No: 109-97-7)의 3개 물질이 확인되었으며 AllegroGraph에 동일한 증상 39개를 입력한 결과 동일하게 Ammonia, Phenylmagnesium Chloride, Pyrrole의 3개 물질이 추출되어 지식베이스 구축 완성도를 확인 하였다. 이를 통해 화학물질 추정 시스템의 증상 추출 과정에 지식베이스 구축 방안을 제시하였다.

향 후 구축된 지식베이스를 활용하여 산업현장 화학물질 인체 접촉 증상과 실시간 센서정보 융합 감지시스템 지식서비스의 인공지능 학습 모델에 필요한 온톨로지로 활용 할 것이다.

## 감사의 글

본 연구는 산업통상자원부 지식서비스산업 핵심 기술개발사업(K\_G012000926501)의 연구비지원에 의해 수행되었습니다. 이에 감사드립니다.

## REFERENCES

- [1] KOSHA(Korea Occupational Safety & Health Agency), <http://www.kosha.or.kr/>
- [2] Jeong-Min Cha, Seong-Ho Hyun, “A Study on the Analysis of Domestic Hazardous Material Accidents in Recent 10 Years”, *Korean Journal of Hazardous Materials*, 7(1), 54-64, (2019)
- [3] NIH (National Institutes of Health), <https://wiser.nlm.nih.gov/>
- [4] Kim, J.R., Ro, K.H., “Construction of Knowledge Base Based on Graph Database for College Student Career Advice Using Public Data”, *Journal of The Institute of Electronics and Information Engineers*, 56(10), 41-48, (2019)
- [5] Hong, G.W., Myaeng, S.H., “Generation of open relation embeddings from natural language sentences using BERT and knowledge base relation embeddings”, *Journal of The Korean Institute of Infor-*

- mation Scientists and Engineers*, 533-535, ICC JEJU, Korea, (2019)
- [6] Kim, J.H., Lee, M.J., "Knowledge Extraction Methodology and Framework from Wikipedia Articles for Construction of Knowledge-Base", *Journal of Intelligence and Information Systems*, 25(1), 43-61, (2019)
- [7] Jung, S.W., Choi, M.S., Kim, H.S., "Construction of Korean Knowledge Base Based on Machine Learning from Wikipedia", *Journal of KIISE*, 42(8), 1065-1070, (2015)
- [8] Kingma, D. and J. Ba, *Adam: A method for stochastic optimization*, Proceedings of the 3rd International Conference for Learning Representations, (2015)
- [9] B. Saha and K. Goebel, *Battery data set*, "NASA AMES Prognostics Data Repository, (2007)
- [10] A. Jain, K. Nandakumar and A. Ross, "Score normalization in multimodal biometric systems", *Pattern Recognit.*, 38(12), 2270-2285, (2005)