



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

유해화학물질의 시각적 안전관리를
위한 MSDS 지도 개발



2020년 2월

부 경 대 학 교 대 학 원

안 전 공 학 과

신 명 우

공학석사학위논문

유해화학물질의 시각적 안전관리를 위한 MSDS 지도 개발

지도교수 : 서 용 윤

이 논문을 공학석사 학위논문으로 인준함

2020년 2월

부 경 대 학 교 대 학 원

안 전 공 학 과

신 명 우

신명우의 공학석사 학위논문을 인준함.

2020 년 2 월 21 일



위 원 장 공학박사 이 창 준 (인)

위 원 공학박사 이 의 주 (인)

위 원 공학박사 서 용 윤 (인)

목 차

제 1 장 서 론	1
제 2 장 연구 배경	4
2.1 유해화학물질 관리제도 및 방법	4
2.2 안전관리를 위한 데이터 시각화 연구	7
제 3 장 연구 방법론	8
3.1 연구 절차	8
3.2 텍스트마이닝	9
3.3 자기조직화지도(SOM: Self-Organizing Map)	11
제 4 장 분석결과	13
4.1 STEP 1 : 데이터 수집	13
4.2 STEP 2 : 전처리 및 키워드 추출	15
4.3 STEP 3 : MSDS 지도 작성	17
4.4 STEP 4 : MSDS 지도 응용	19
제 5 장 토의	22
5.1 물질이 금지물질과 허가물질 셀에 포함된 경우	24
5.2 물질이 금지물질 셀에만 포함된 경우	26
5.3 물질이 두 가지 셀 모두에 포함되지 않는 경우	28
5.4 새로운 물질이 추가되는 경우	28
제 6 장 결론	30
참고문헌	31

표 목 차

Table 1 Chemical substance representative case	5
Table 2 Chemical control low act.33	6
Table 3 Summary of hazardous and chemical materials in MSDS map	23
Table 4 Guide text identical to prohibited material	27



그 립 목 차

Fig. 1 Research procedure	8
Fig. 2 Text mining method	10
Fig. 3 SOM method	12
Fig. 4 Data collection : MSDS data list publised by OSHA	14
Fig. 5 Pakcage code	15
Fig. 6 Result of document-term matrix	16
Fig. 7 MSDS map clustering results	18
Fig. 8 Prohibited substances	20
Fig. 9 Permitted hazardous substances	21
Fig. 10 Prohibited and Permitted substances	25
Fig. 11 General substance	26
Fig. 12 Management method for new substances	28

Development of MSDS Map for Visual Safety Management of Hazardous and Chemical Materials

Myungwoo Shin

Department of Safety Engineering, Graduate School,
Pukyong National University

Abstract

For preventing the accidents generated from the chemical materials, thus far, MSDS (Material Safety Data Sheet) data have been made to notify how to use and manage the hazardous and chemical materials in safety, However, it is difficult for users who handle these materials to understand the MSDS data because they are only listed based on the alphabetical order, not based on the specific factors such as similarity of characteristics. It is limited in representing the types of chemical materials with respect to their characteristics. Thus, in this study, a lots of MSDS data are visualized based on relationships of the characteristics among the chemical materials for supporting safety

managers. For this, we used the textmining algorithm which extracts text keywords contained in documents and the Self-Organizing Map(SOM) algorithm which visually addresses textual data information. In the case of Occupational Safety and Health Administration (OSHA) in the United States, the guide texts contained in MSDS documents, which include use information such as reactivity and and potential risks of materials, are gathered as the target data. First, using the textmining algorithm, the information of chemicals is extracted from these guide texts. Next, the MSDS map is developed using SOM in terms of similarity of text information of chemical materials. The MSDS map is helpful for effecttively classifying chemical materials by mapping prohibited and hazardous substances on the developed the SOM map. As a result, suing the MSDS map, it is easy for safety managers to detect prohibited and hazardous substances with respect to the Industrial Safety and Health Act standards.

제 1 장 서 론

산업이 발전함에 따라 삶의 편의성을 제공하기 위한 화학물질들이 다양하게 사용되고 있으며, 새로운 화학물질이 지속적으로 개발되고 증가하고 있다. 미국화학회(American Chemical Society)에서 운영하는 Chemical Abstract Service에 따르면 2015년 6월 1일까지 1억 종이상의 화학물질이 등록되어 있으며, 그 중 170만 종이상이 유통되고 있다. 현재 국내에서 유통되는 화학물질만도 4만3천여 종이며, 매년 200~400종의 신규화학물질들이 기계·재료·의약품 등을 개발하기 위해 유용하게 사용되고 있다¹⁾.

그러나 이와 같은 화학물질의 유용성에도 불구하고, 화학물질의 보관과 사용과정에서, 가연성·반응성·부식성·독성 등 위험성과 관련한 문제점이 나타나고 있다. 화학물질 사고는 다양한 화학물질에 의하여 꾸준히 발생하고 있으며 ‘폭발’, ‘누출’, ‘중독’ 등 발생형태 또한 다양하게 나타난다. 최근 국내에서 발생했던 구미 불산 누출사고에서는 작업자 5명이 사망했으며, 사고현장에 출동한 소방관 및 경찰, 인근주민 등 1,100여명이 누출사고로 인하여 검사와 치료를 받았다. 그리고 농작물도 말라죽는 물적손실 피해도 있었다. 이처럼 화학물질이 대량으로 사용될수록, 사고 발생 시 인명피해와 물적피해를 동반하는 중대재해로 이어질 가능성이 높아지고 있다.

이에 따라 구체적으로 물질안전보건자료(MSDS : Material Safety Data Sheet)를 작성하여 올바른 보관과 사용방법을 제시 및 배포하고 있다. 국내에서도 사고대비물질을 추가하고 정부에서 사용방법을 관리하는 등 화학물질관리법과 산업안전보건법이 변화하고 있다. 이는 화학물질을 유통하고 사용하는 기관 및 기업에 대해 안전한 사용의무를 촉구함에 따라 사고와 재해를 감소하는 효과가 있을 수 있다²⁾.

그러나 현재 국내·외에서 작성되고 있는 MSDS는 물질의 특성이나 유형 또는 관리·사용방법의 유사성과는 관련없이 일괄적으로 알파벳 또는 숫자 형태로 목록화(list)되어 있기 때문에, 기존 물질의 위험성 및 취급방법의 유사성을 파악하기도 어렵다. 또한 화학물질은 전문가뿐만 아니라 일반적 인 근로자들에게도 쉽게 노출된다는 점에서, 사용물질에 대한 정보를 쉽게 파악할 수 있는 관리방법을 제시할 필요가 있다³⁾. 따라서 기존 MSDS를 물질의 특성이나 관리 및 사용방법의 유사성을 파악하고, 그 연관성을 시각적으로 표현하여 일반 근로자도 쉽게 확인할 수 있는 방법을 개발할 필요가 있다.

이를 위해, 본 연구에서는 유해화학물질의 관리를 위한 물질 간 연관성 정보를 시각화하는 방법을 개발한다. 구체적으로, OSHA(Occupational Safety and Health Administration)의 MSDS 자료에 포함된 위험성, 취급 방법, 응급처치방법 등을 제시하고 있는 가이드 데이터(guide text)를 이용하여 물질의 정성적 정보 키워드를 도출하고, 키워드 간의 관계를 분석 및 시각화하여 MSDS 지도(MSDS map)를 개발한다. MSDS 지도는 유해화학 물질의 취급방법, 위험성, 응급처치방법 등의 연관성을 보여주는 도구로 정의한다. 방법론적으로, 우선, 텍스트마이닝(textmining)을 활용하여 MSDS에 작성된 guide text 데이터의 키워드를 도출한 후, 자기조직화지도(SOM : Self-Organizing Map)를 이용하여 키워드의 유사도에 따라 물질을 클러스터링하고, 지도 형태로 시각화한다. 작성된 MSDS 지도를 활용하여, 산업안전보건법 기준에 따라 허가대상 유해물질과 금지물질을 매핑 및 분류함으로써, 기존물질은 물론 새로이 추가된 물질을 MSDS 지도에 표시된 화학물질들과 시각적으로 비교하는 관리방법을 제안한다.

MSDS 지도는 실무 관리자에게 MSDS 관리법을 제시하며, 금지물질과 허가물질에 대한 분류에 따른 화학물질의 위험성 검토와 새로운 물질이 추

가될 경우 기존에 사용하고 있는 유사물질과의 관계를 바탕으로 MSDS를 새로 작성하고 지속적으로 관리할 수 있는 장점을 제시한다. 그리고 시각적으로 표현된 화학물질 분류를 통하여 현장근로자도 취급방법, 응급처치 방법등을 쉽게 파악하는데 도움을 줄 것으로 기대된다.



제 2 장 연구 배경

2.1 유해화학물질 관리제도 및 방법

화학물질사고는 2014년도 130회, 2015년도 117회, 2016년도 84회, 2017년도 92회, 2018년도 68회가 발생하였다⁴⁾. 발생빈도는 매년 비슷한 수준으로 발생하였으며, 화학누출사고는 Table 1에서 볼 수 있듯이 인명피해와 물적피해를 동반하는 중대재해로 이어지는 경우가 많다. 따라서, 화학물질을 취급하는 사업장은 Table 2와 같이 국내에서 화학물질관리법의 제 33조 “유해화학물질 안전교육”에 따라 유해화학물질 취급시설의 기술인력, 유해화학물질관리자 그리고 유해화학물질 취급 담당자는 환경부령으로 정하는 교육기관이 실시하는 유해화학물질 안전교육을 받도록 되어 있으며, 해당 사업장의 모든 종사자에 대하여 정기적으로 유해화학물질 안전교육을 실시하도록 되어있다.

그러나 대부분의 사업장에서는 시청각 교육이나 강의를 통한 교육으로 해당 물질에 대하여 위험성을 인지시키기 보단 법적기준을 충족시키기 위한 형식적 또는 의무적인 교육만 실시하고 있다.⁵⁾ 즉, 근로자가 참여하지 않는 교육은 근로자로 하여금 유해물질의 정확한 정보를 얻기 힘들기 때문에 유해물질의 정보를 얻기 위해서는 목록형태로 되어 있는 MSDS 데이터를 통해 얻어야 한다. 목록형태의 MSDS 데이터는 물질특성이나 취급방법과는 상관없이, 숫자, 알파벳 순으로 나열되어 있어 그 특징의 유사성을 파악하기 힘들다. 이는 화학물질관리법 제48조에 따른 화학물질종합정보시스템에서도 마찬가지로 문제점을 가지고 있다.

따라서 본 연구에서는, 일반적인 근로자들도 화학물질에 대한 정보를 용이하게 얻을 수 있도록 화학물질을 특성 또는 관리와 사용방법에 따라 분류하여, 관리되고 있는 화학물질들의 관계를 시각적으로 확인할 수 있는 MSDS 지도를 개발하고자 한다.

Table 1. Chemical substance representative case

Year	Location	Accident	Damage
2012	Gumi	Hydrofluoric acid leakage	• 5 deaths • 398 slight injury
2013	Hwaseong	Hydrofluoric acid leakage	• 1 deaths • 4 serious injury
2013	Yeosu	Polyethylene storage tank explosion	• 6 deaths • 11 slight injury
2018	Pohang	Oil tank oil vapour explosion	• 1 deaths • 4 slight injury

Table 2. Chemical control law Act.33

Article 33 (Safety education of hazardous chemicals) ① The technical personnel of hazardous chemical processing facilities pursuant to article 28 paragraph 2 of Harmful Chemicals, the managers of hazardous chemicals under Article 32 and other personnel in charge of handling hazardous chemicals pursuant to the Presidential Decree shall receive safety education of hazardous chemicals conducted by educational institutions (hereinafter referred to as "Safety education of hazardous chemicals.") ② When a person who is required to receive safety training for hazardous chemicals is hired, the operator of hazardous chemicals shall receive safety training for harmful chemicals from the person concerned. In this case, the operator of the hazardous chemicals shall bear the costs of training. ③ The operator of hazardous chemicals shall provide regular safety training for all employees of the business establishment as provided under the Environment Ministry Ordinance.
--

2.2 안전관리를 위한 데이터 시각화 연구

안전관리를 위하여 SOM, 네트워크 분석 등을 활용하여 다양한 산업에 데이터 시각화를 통한 안전관리방안 연구가 활발히 진행되고 있다^{6,7,8)}. 특히, SOM을 활용한 기존연구에서는 국내에서 건설업에 대하여 사고사망의 재해개요를 분석하여 위험요인지도를 작성하는 연구가 진행되었으며, 사고사망에 대한 주요 요인들을 파악하여 효과적인 안전관리 방법을 제시하였다^{7,8)}. 또한, 감시장비 시스템이 내재하고 있는 다양한 환경 및 정적 이상 감지에서 낮은 강인성 문제를 보완하고 비용 효율성 문제를 해결하기 위한 지능형 이상 감지 시스템을 제안하는 연구를 진행하였다⁹⁾. 국외에서는 이탈리아 목재 가공업에 대하여 재해개요를 분석하여 사고유발 원인과 매개체를 찾아내는 연구가 진행되었고, 브라질에서는 복합적 사고에 대한 위험성을 파악하여 사고로 이어지는 상황을 판단하여 시각화하는 연구가 진행되었다^{8,10)}. 따라서 본 연구에서는, 유해화학물질의 유사물질과 주요요인을 도출하기 위해, SOM을 활용한 MSDS 지도를 개발하여 유해화학물질을 분류하고, 금지물질과 허가물질에 대한 위험성에 따라 MSDS를 검토하는 방안을 제안한다.

제 3 장 연구 방법론

3.1 연구절차

본 연구의 절차는 Fig. 1과 같이 진행된다. 데이터수집이 수월하고 신뢰도가 높으며, 다양한 설명이 적혀있는 미국 OSHA의 MSDS 데이터를 수집한다. 다음으로 수집된 데이터에서 Guide text 데이터에 대하여 텍스트마이닝을 통하여 키워드를 추출한 후, MSDS 지도를 작성한다. MSDS 지도는 SOM을 통하여 작성하며, SOM은 비지도 학습에 의한 클러스터링 방법으로, 추출한 키워드를 유사도에 따라 시각화한다. 마지막으로 MSDS 지도에 산업안전보건법 기준에 따른 금지물질과 허가대상 유해물질로 분류하여 MSDS 지도를 활용하는 관리방법을 제시한다.

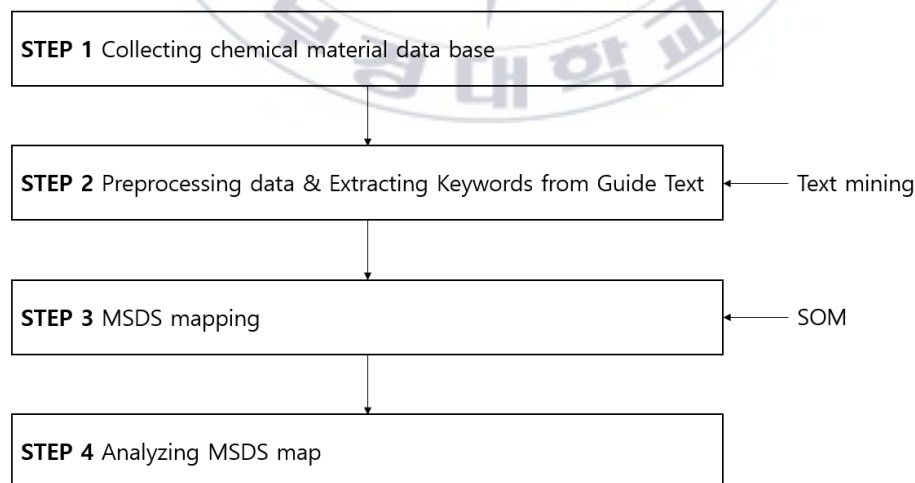


Fig. 1. Research procedure

3.2 텍스트 마이닝

텍스트 마이닝은 텍스트로 구성된 비정형데이터에서 자연어 처리 기술을 기반으로 유용한 정보를 얻는 방법이며, 문서의 연계성 파악, 분류 혹은 군집화, 요약 등 빅데이터에 숨겨진 의미있는 정보를 발견하는 것이다. 현재 텍스트 마이닝은 특허분석, 고객요구분석, 감성 분석 같은 사회과학분야에서 다양하게 사용 되고 있다^{6,7)}. 텍스트 마이닝의 과정은 첫 번째로 텍스트데이터를 수집하고, 두 번째로 데이터의 전처리 과정을 거친다, 전처리 과정에서는 조사나 불용어(stopwords) 등 의미를 가지지 않는 단어를 제거하거나 데이터가 포함되어 있지 않은 데이터(missing data)들을 제거한다. 마지막으로 정형데이터인 문서-키워드 매트릭스(DTM : Document-Term Matrix)를 생성하여, 다양한 데이터마이닝 알고리즘을 처리할 수 있는 input data로 활용한다. 텍스트 마이닝과 같은 자연어 처리 기술은 우리가 사용하는 언어를 컴퓨터가 인식하여, 이를 처리할 수 있도록 하는 기술이다. 자연어 처리는 크게 두 가지 작업으로 나눌 수 있다. 첫째, 필요한 정보뿐만 아니라 언어에 있어서의 어휘와 구문, 의미에 관한 지식을 사용해서 문어를 처리하는 것이다. 둘째, 음성에서 발생하는 애매함을 비롯한 음성학에 대한 부가적인 지식을 필요로 하는 구어를 처리하는 것이다. 구체적으로 문서가 입력되면 전처리 과정에서 언어를 탐지하고, 탐지된 언어를 기반으로 문서, 문단, 문장의 기본단위로 구분하여 품사를 매칭하고 불용어를 제거하는 작업을 한다. 이후 모델링 작업을 수행하여 원하는 결과를 도출하는 과정이 자연어 처리 과정이다.

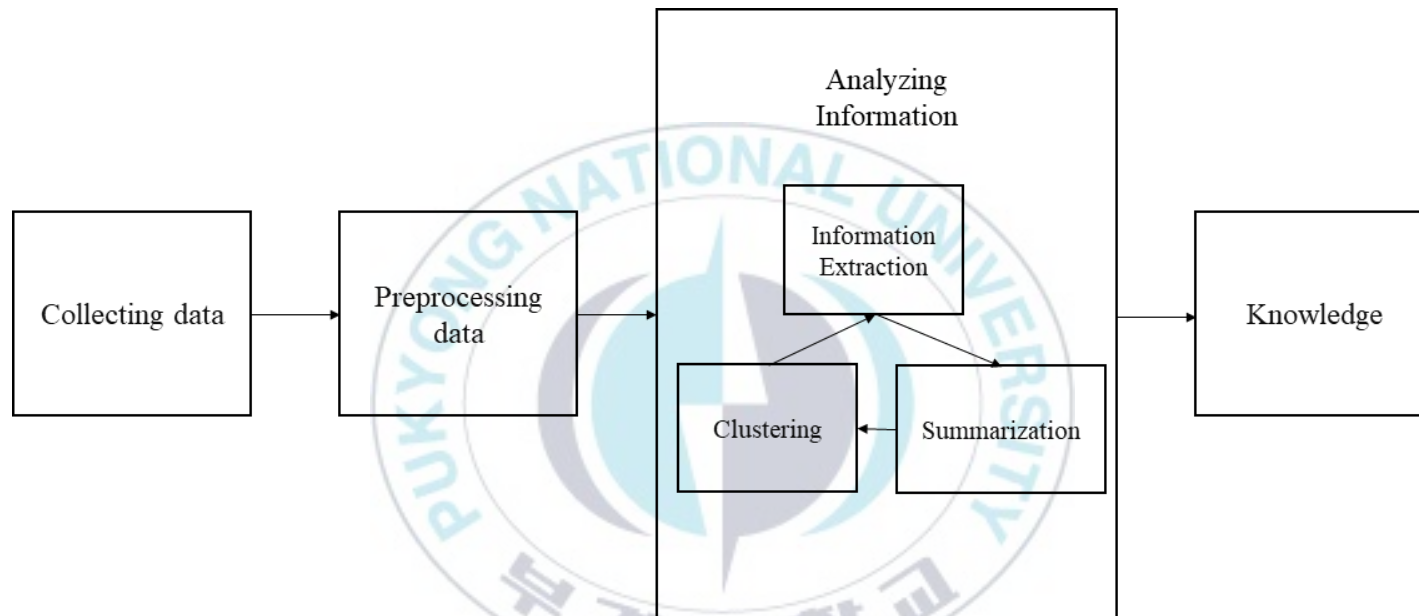


Fig. 2. Text mining method

3.3 자기조직화지도(SOM : Self-Organizing Map)

SOM은 입력 패턴에 대하여 정확한 해답을 주지 않고 스스로 학습할 수 있는 능력이 있는 비지도 학습에 의한 클러스터링 방법이며 Fig. 3.와 같이 input 데이터에 대한 유사 데이터에 가중치를 지속적으로 업데이트하여 학습시키는 클러스터링 방법이다^{9,10}. SOM은 뉴런의 거리에 따라 승자 뉴런을 정하며 승자 뉴런과 반경 안에 드는 모든 뉴런 등도 유사 조정하는 과정을 거치며 반경을 줄이는 과정을 거친다^{11,12}. 즉, 가중치가 변화가 없을 때 까지 앞에 과정을 반복하여 클러스터링 결과를 얻는다. 학습 단계에서 입력 벡터 X 와 각 뉴런의 가중치 벡터 m_i 의 위상적인 거리를 구하게 된다. 이때 입력 벡터 X 와 가장 인접한 거리의 뉴런을 승자 뉴런으로 정의하게 되며 식 (1)과 같이 경쟁 학습 과정을 거친다.

$$(1) \text{winner}(X) = \min_{i \in 1, \dots, M} \|X - m_i(t)\|$$

이후 승자 뉴런을 중심으로 전체 가중치 벡터는 식 (2)와 같이 갱신된다.

$$(2) m_i(t+1) = m_i(t) + \epsilon(t) \Lambda_{\text{winner}(x), i} (X - m_{\text{winner}(X)}(t))$$

여기서 $\epsilon(t)$ 는 학습 속도를 결정하는 학습률이며, 이웃 함수로 정의되는 $\Lambda_{\text{winner}(X), i}$ 는 승자 뉴런을 중심으로 이웃한 뉴런과의 연결 강도를 식 (3)과 같은 오목한 형태의 가우시안 필터 스칼라 곱으로 조정한다.

$$(3) \Lambda_{\text{winner}(X), i} = \exp\left(-\frac{\|r_i - r_{\text{winner}(x)}\|^2}{2\sigma^2(t)}\right)$$

여기서 r_i 와 $r_{\text{winner}(X)}$ 는 SOM의 격자 구조 내 뉴런의 위치이며, $\sigma^2(t)$ 는 이웃 함수의 반경을 나타내는 분산이다. $\sigma^2(t)$ 과 $\epsilon(t)$ 는 0~1의 범위에서 학습이 진행됨에 따라 점진적으로 감소하는 선형 함수 형태를 갖는다. 결과적으로 SOM의 학습 기법의 가우시안은 방사형으로 거리가 멀어짐에 따라 값이 작

아지는 분포이기 때문에 입력 벡터와 뉴런과의 경쟁 학습 과정을 통해 승자가 된 뉴런을 중심으로 연결 강도를 조정하며 최적화된다¹³⁾. SOM을 활용하면 다차원 데이터를 2차원으로 시각화 할 수 있다. 실제로 SOM을 활용하여, 사고문서의 개요 등을 텍스트마이닝을 거친 후 시각화하여 사고유형을 구분하고, 관리방안을 마련하는 연구가 수행되어 왔다^{7,8,14)}. 본 연구에서는 화학물질의 유사성이나 유형에 관계없이 목록형태로 되어 있어 쉽게 파악하기 힘든 MSDS자료를 효과적으로 시각화하여 나타내기 위하여 SOM을 활용한다.

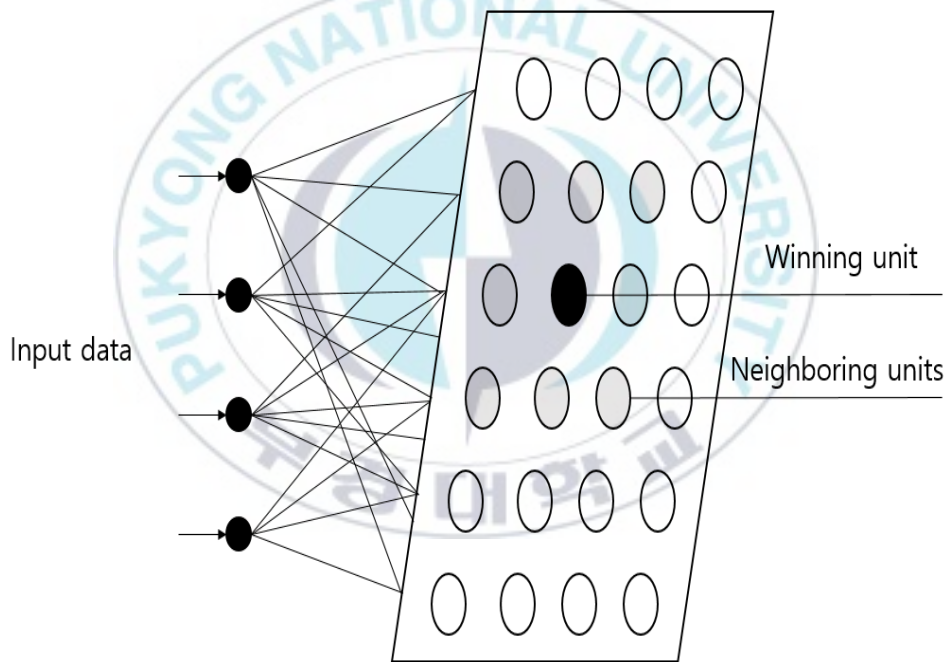


Fig. 3. SOM method

제 4 장 분석결과

4.1 STEP 1 : 데이터 수집

화학물질 데이터 수집을 위하여 OSHA Chemical database (<https://www.osha.gov/chemicaldata/>, 2018년 6월까지 접속)에서 제공하는 800개의 화학물질 데이터를 수집하였다. OSHA에서 제공하는 데이터도 Fig. 3과 같이 특별한 기준없이 물질을 목록 형태로 정리하여 제공하고 있으며, MSDS 문서에는 Chemical name, Formula, Synonyms, Physical description, Reactivity, Firsy aid, Guide text가 메타 데이터로 포함되어 있다. 본 연구에서는 물리화화적인 물성특정보다는 MSDS의 위험성, 취급방법, 응급처치방법 등을 중심으로 분석하기 위해 guide text를 사용하여, 화학물질의 사용과정의 안전성을 개선하고자 한다. 이는 OSHA에서 자체적으로 제공하는 데이터로, guide text에는 MSDS의 기본항목 16가지를 모두 포함하기 보다는, 유해위험성, 응급조치요령, 폭발화재 또는 누출사고 시 대처방법, 취급 및 저장방법, 안전성 및 반응성등에 대한 정보를 포함하고 있다. 데이터 전처리 과정에서는 모든 문서에 대하여 guide text가 제공되어 있지 않아, guide text가 없는 화학물질자료는 제거하여 800개 중 539개의 문서가 추출되었다. 최종적으로, 539개 문서에 대하여 이름순서대로 labeling을 실시하는 과정을 거쳤다.

1	Chemical name	Formula	Synonyms	Physical description	Reactivity	First aid
2	1,1,1,2-TETRACHLORO-2,2	CCl ₃ CClF ₂	2,2-Difluoro-1,1,1,2-tetra	Colorless solid with a slight	Chemically-	Eyes: Irrigat
3	1,1,1,2-TETRACHLOROETHANE	C ₂ H ₂ Cl ₄		Yellowish-red liquid.	STABILITY:	EYES: First c
4	1,1,1-TRICHLOROETHANE	C ₂ H ₃ Cl ₃	Chloroethene; 1,1,1-Trich	Colorless liquid with a mi	STABILITY:	EYES: First c
5	1,1,2,2-TETRACHLORO-1,2-DI	CCl ₂ CFCl ₂ F	1,2-Difluoro-1,1,2,2-tetra	Colorless solid or liquid (	Chemically-	Eyes: Irrigat
6	1,1,2-TRICHLORO-1,2,2-TRIF	C ₂ Cl ₃ F ₃	Chlorofluorocarbon-113	Colorless to water-white	REACTIVITY	EYES: First c
7	1,1,2-TRICHLOROETHANE	C ₂ H ₃ Cl ₃	Ethane trichloride; beta	Colorless liquid with a sw	STABILITY:	EYES: First c
8	1,1-DICHLORO-1-NITROETH	C ₂ H ₃ Cl ₂ NO ₂	Dichloronitroethane	Colorless liquid with an	STABILITY:	EYES: First c
9	1,1-DICHLOROETHANE	C ₂ H ₄ Cl ₂	Asymmetrical dichloroe	Colorless, oily liquid with	STABILITY:	EYES: First c
10	1,1-DIFLUOROETHYLENE	C ₂ H ₂ F ₂	Difluoro-1,1-ethylene; 1	Colorless gas with a faint	STABILITY:	EYES: First c
11	1,1-DIMETHYLHYDRAZINE	C ₂ H ₈ N ₂	Dimazine; DMH; UDMH	Colorless liquid with an	AIR AND W	Warning: Ef
12	1,1-DIMETHYLPROPYLACET	C ₇ H ₁₄ O ₂	tert-Amyl acetate; Pentylacetate			Call a physic
13	1,1-METHYLENE BIS(4-CYCL	C ₁₅ H ₂₂ N ₂ O ₂	Dicyclohexylmethane 4,	Clear, colorless to light-y	STABILITY:	EYES: First c
14	1,2,3-TRICHLOROPROPANE	C ₃ H ₅ Cl ₃	Allyl trichloride; Glycero	Colorless liquid with a ch	STABILITY:	EYES: First c
15	1,2,4-TRICHLOROBENZENE	C ₆ H ₃ Cl ₃	unsym-Trichlorobenzen	Colorless liquid or crystal	STABILITY:	EYES: First c
16	1,2-DIBROMO-3-CHLOR	C ₃ H ₅ Br ₂ Cl	1-Chloro-2,3-dibromop	Dense yellow or amber li	STABILITY:	EYES: First c
17	1,2-DICHLORO-1,1,2,2-TETR	C ₂ Cl ₂ F ₄	1,2-Dichlorotetrafluoro	Colorless gas with a faint	Chemically-	INHALATIO
18	1,2-DICHLOROETHYLENE, A	C ₂ H ₂ Cl ₂	Acetylene dichloride; ci	Colorless liquid (usually a	STABILITY:	EYES: First c
19	1,2-DICHLOROPROPANE	C ₃ H ₆ Cl ₂	Dichloro-1,2-propane; 1	Colorless liquid with a ch	STABILITY:	EYES: First c
20	1,2-EPOXYETHYLBENZENE	C ₈ H ₈ O				
21	1,2-PHENYLENEDIAMINE	C ₆ H ₈ N ₂			STABILITY:	EYES: First c
22	1,3-BUTADIENE	C ₄ H ₆	Biethylene; Biviny; Buta	Colorless gas with a mild	AIR AND W	EYES: First c
23	1,3-DICHLORO-5,5-DIMETH	C ₅ H ₆ Cl ₂ N ₂ O	Dactin; DDH; Halane	White powder with a chlo	STABILITY:	EYES: First c
24	1,3-DICHLOROPROPENE	C ₃ H ₄ Cl ₂	3-Chloroallyl chloride; 1	Colorless to straw-colore	STABILITY:	EYES: First c
25	1,3-DIOXOLANE	C ₃ H ₆ O ₂				Move victim
26	1,3-PHENYLENEDIAMINE	C ₆ H ₈ N ₂			STABILITY:	EYES: First c
27	1,4-DICHLORO-2-BUTENE	C ₄ H ₆ Cl ₂			AIR AND W	INHALATIO
28	1,5-NAPHTHENE DIISOCYAN	C ₁₀ H ₆ (NCO) ₂	1,5-Diisocyanatonaphth	White to light-yellow, crys	This compo	Eye: If this c
29	1,6-HEXANE DIAMINE	C ₆ H ₁₆ N ₂			STABILITY:	EYES: First c
30	1-CHLORO-1-NITROPROPAN	C ₃ H ₆ ClNO ₂	Korax; Lanstan	Colorless liquid with an	STABILITY:	EYES: First c
31	1-CHLORO-2,3-EPOXYPROP	C ₃ H ₅ ClO	2-Chloropropylene oxide	Colorless liquid with a sli	CHEMICAL	Warning: Ef
32	1-CHLORO-2-PROPANOL	C ₃ H ₇ ClO	sec-Propylene chlorohy	Clear colorless or light ar	REACTIVITY	EYES: First c
33	1-HEXADECANETHIOL	CH ₃ (CH ₂) ₁₅ S	Cetyl mercaptan; Hexad	Colorless liquid or solid (	Oxidizers, s	Eyes: Irrigat
34	1-HEXENE	C ₆ H ₁₂				SKIN OR EV
35	1-HYDROXYANTHRA-QUIN	C ₁₄ H ₈ O ₃				
36	1-METHOXY-2-PROPANOL	C ₄ H ₁₀ O ₂	Dowtherm209; 1-Metho	Clear, colorless liquid with	CHEMICAL	EYES: First c
37	1-NAPHTHYLAMINE	C ₁₀ H ₉ N	1-Aminonaphthalene; 1	Colorless crystals with an	STABILITY:	EYES: First c
38	1-NITRONAPHTHALENE	C ₁₀ H ₇ NO ₂				
39	1-NITROPROPANE	CH ₃ CH ₂ CH ₂ N	Nitropropane; 1-NP	Colorless liquid with a so	Amines; str	Eyes: Irrigat
40	1-PROPYLENE	C ₃ H ₄	Allylene; Propine; Propy	Colorless gas with a swe	CHEMICAL	Eyes: Frostbi
41	1-UNDECANETHIOL	CH ₃ (CH ₂) ₁₀ S	Undecyl mercaptan	Liquid.	Oxidizers, r	Eyes: Irrigat
42	2,2-BIS-(BROMOMETHYL)-1	C ₅ H ₁₀ Br ₂ O ₂	Dibromoneopentyl glyco	Off-white powder	REACTIVITY	EYES: First c

Fig. 4. Data collection : MSDS data list publised by OSHA

4.2 STEP 2 : 전처리 및 키워드 추출

키워드 추출은 R프로그램을 이용하여 Fig. 5와 같이 tm package와 NLP package를 통하여 실시하였다. 전처리 과정에서 접속사·지시어 등 불용어(stopwords)를 제거하고 데이터가 포함되지 않은 데이터를 제거한 후 텍스트마이닝을 수행하였다. 또한, 글자 수가 3~12개의 의미가 확보되는 단어들을 추출하여, 최종적으로 문서와 키워드의 DTM을 구축하였다. Fig. 6의 행(row)은 문서번호를 의미하는 case들이며, 열(column)은 추출된 키워드 하나의 record를 구성하게 된다. 해당 셀의 숫자는 문서에 키워드가 언급된 횟수를 말한다. 이 DTM은 SOM을 구축하기 위한 기본 input으로 사용되며, 각 MSDS문서에 포함된 키워드들의 유사성을 SOM에서 분석할 수 있도록 한다.

```
library(rJava)
library(NLP)
library(tm)
dtm <- DocumentTermMatrix(doc,
                           control=list (tokenize=words,
                                         removeNumbers=T,
                                         removePunctuation=T,
                                         wordLengths=c(3,12),
```

Fig. 5. Package code

1		administer agent	aluminum answer	apparatus	appropriate area	areas	artificial	available	aware	away	back	basement: breathing	burn	calcium	call
2	1	1	1	1	1	2	1	2	1	1	1	1	3	1	2
3	2	1	1	1	1	2	1	2	1	1	1	1	3	1	2
4	3	1	1	1	1	2	1	2	1	1	1	1	3	1	2
5	4	1	0	0	1	1	3	1	1	1	2	1	3	1	0
6	5	1	0	0	1	1	3	1	1	1	2	1	3	1	0
7	6	1	0	0	1	1	3	1	1	1	2	1	3	1	0
8	7	1	0	0	1	1	3	1	1	1	2	1	3	1	0
9	8	1	0	0	1	1	3	1	1	1	2	1	3	1	0
10	9	1	0	0	1	1	3	1	1	1	2	1	3	1	0
11	10	1	0	0	1	1	3	1	1	1	2	1	3	1	0
12	11	1	0	0	1	1	3	1	1	1	2	1	3	1	0
13	12	1	0	0	1	1	3	1	1	1	2	1	3	1	0
14	13	1	0	0	1	1	3	1	1	1	2	1	3	1	0
15	14	1	0	0	1	1	3	1	1	1	2	1	3	1	0
16	15	1	0	0	1	1	3	1	1	1	2	1	3	1	0
17	16	1	0	0	1	1	3	1	1	1	2	1	3	1	0
18	17	1	0	0	1	1	3	1	1	1	2	1	3	1	0
19	18	1	0	0	1	1	3	1	1	1	2	1	3	1	0
20	19	1	0	0	1	1	3	1	1	1	2	1	3	1	0
21	20	1	0	0	1	1	3	1	1	1	2	1	3	1	0
22	21	1	0	0	1	1	3	1	1	1	2	1	3	1	0
23	22	1	0	0	1	1	3	1	1	1	2	1	3	1	0
24	23	1	0	0	1	1	3	1	1	1	2	1	3	1	0
25	24	1	0	0	1	1	3	1	1	1	2	1	3	1	0
26	25	1	0	0	1	1	3	1	1	1	2	1	3	2	0
27	26	1	0	0	1	1	3	1	1	1	2	1	3	2	0
28	27	1	0	0	1	1	3	1	1	1	2	1	3	2	0
29	28	1	0	0	1	1	3	1	1	1	2	1	3	2	0
30	29	1	0	0	1	1	2	0	1	2	1	1	0	3	1
31	30	1	0	0	1	1	2	0	1	2	1	1	0	3	1
32	31	1	0	0	1	1	4	3	1	1	2	1	2	3	2
33	32	1	0	0	1	1	4	3	1	1	2	1	2	3	2
34	33	1	0	0	1	1	4	3	1	1	2	1	2	3	2
35	34	1	0	0	1	1	4	3	1	1	2	1	2	3	2
36	35	1	0	0	1	1	4	3	1	1	2	1	2	3	2

Fig. 6. Result of document-term matrix

4.3 STEP 3 : MSDS지도 작성

MSDS 지도 작성은 SOM을 활용하여 키워드의 빈도수에 따라 시각화하였다. SOM은 MATLAB의 Kohonen SOM toolbox로 구현하였으며, 10×15 매트릭스에 전체문서를 유사도에 따라 배치하였다. 매트릭스의 크기 결정은 10×8 매트릭스부터 시작하여, 점차적으로 크기를 조정함으로써 Quantization error와 Topographic error가 가장 낮은 오류를 갖는 매트릭스 크기를 선정하였다. 이는 Sample size나 결과에 따라서 유연하게 조정 가능하다. Fig. 7에 표시된 문서 번호는 각 셀(cell)의 대표문서(vote)를 나타내며, 이는 같은 문서 내에 변수 값과 학습 가중치 값의 함수 결과값이 가장 큰 문서를 나타낸다. MSDS 지도에 매핑된 물질들은, 물질 특성별로 독성물질, 물 반응성 물질, 가연성 고체, 가연성 액체, 기타물질로 클러스터링하여, 물질을 용이하게 관리할 수 있는 유형 정보를 나타낸다.

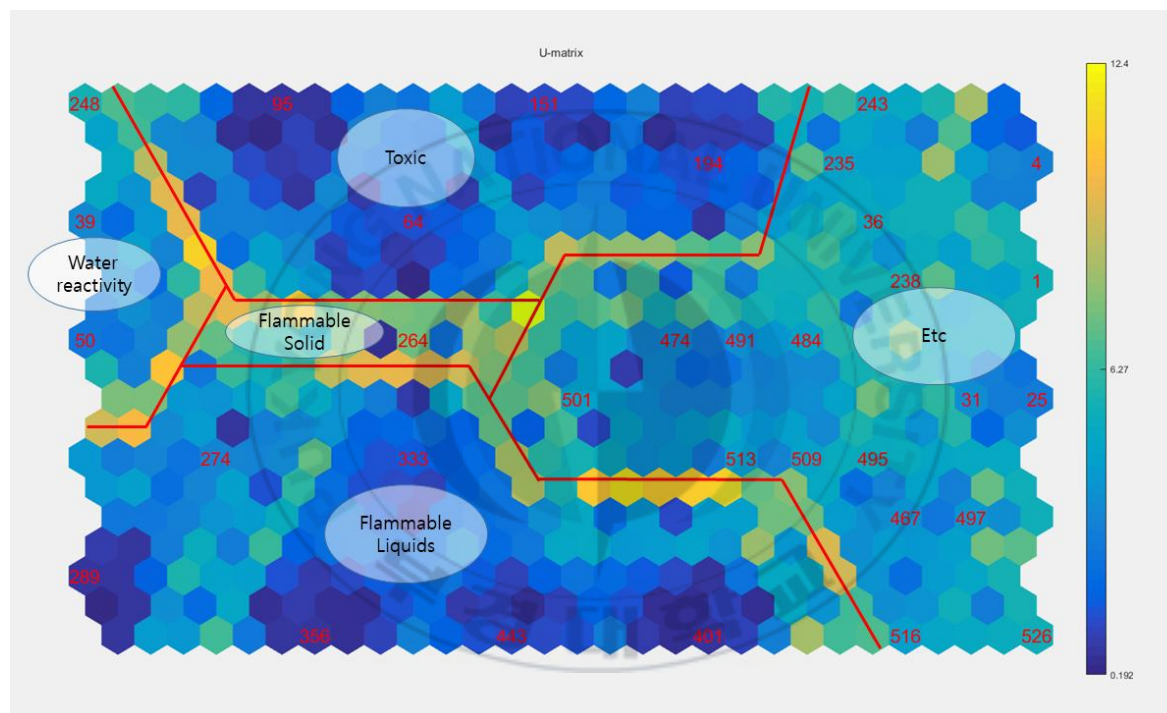


Fig. 7. MSDS map clustering results

4.4 STEP 4 : MSDS 지도 응용

MSDS 지도에 화학물질 취급소에서 따르는 산업안전보건법 기준에 따라 금지물질(prohibited substances), 허가대상 유해물질(permitted hazardous substances)을 표시하여 유해화학물질이 지도상에 위치하는 클러스터를 살펴해보았다. 결과적으로, 금지물질들은 Fig. 8 의 그림과 같이 나왔으며(마름모 표시), 대부분 물질이 독성물질에 포함되며, 그 외에 가연성 액체, 기타물질로 일부 분류되었다. 허가대상 유해물질의 경우 Fig. 9의 그림과 같이 나왔으며(네모 표시), 역시 독성물질이 대부분을 차지하며, 물 반응성·가연성 고체·기타물질로 나머지가 분류되었다.



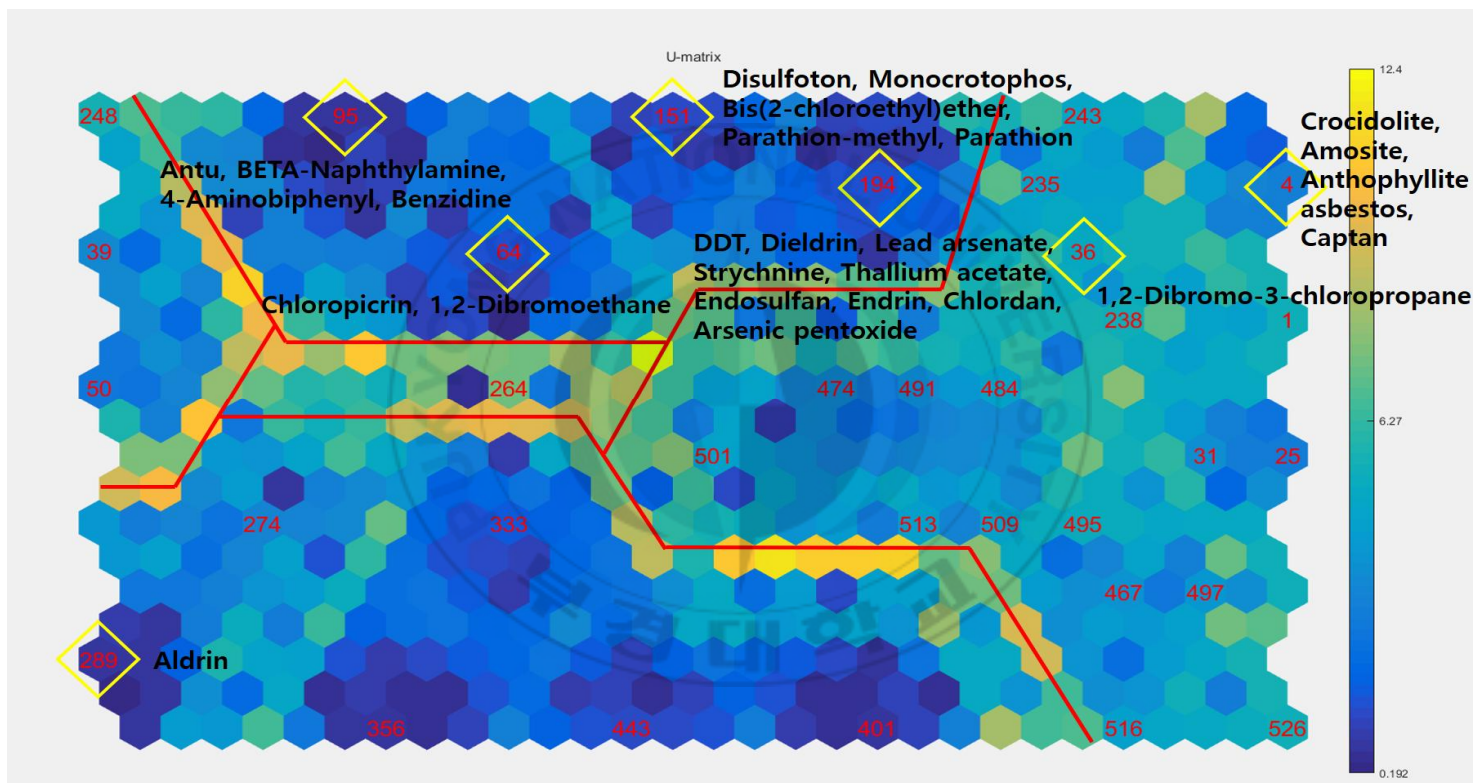


Fig. 8. Prohibited substances

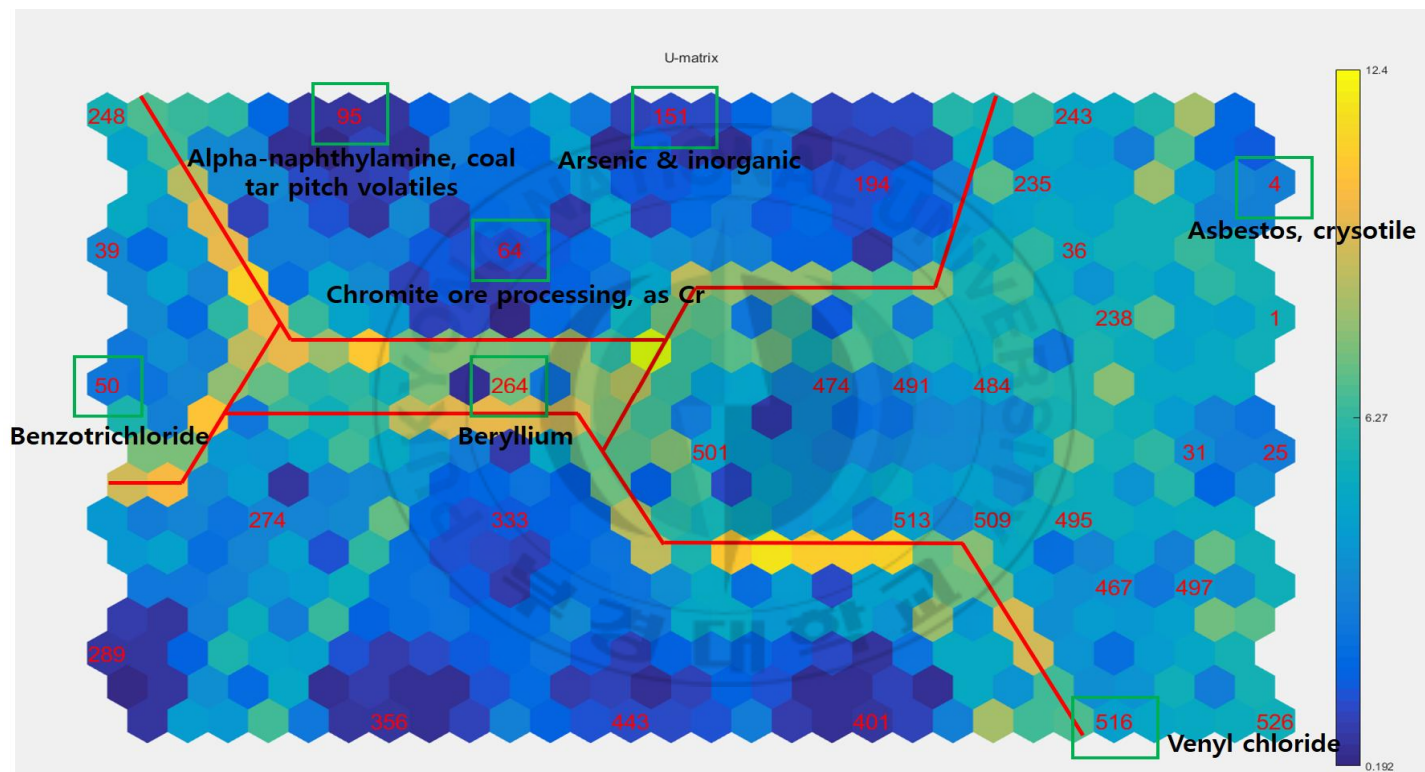


Fig. 9. Permitted hazardous substances

제 5 장 토의

MSDS 지도를 시각화하고, 셀의 화학물질 정보를 확인한 결과, Table 3의 해당 셀에서 확인할 수 있듯이, MSDS지도상에 “금지물질”과 같은 셀에 묶여있는 MSDS 화학물질문서는 95번의 경우 56가지, 151번의 경우 43가지, 64번의 경우 31가지, 194번의 경우 41가지, 289번의 경우 44가지, 36번의 경우 3가지, 4번의 경우 21가지가 포함되어 있었다. 이 중에서 95번, 151번, 64번, 4번의 경우는 허가대상물질도 같이 포함하는 셀로 나타났다.

즉, guide text가 없는 문서를 제외한 후 539가지 화학물질에 대하여 클러스터링 결과 금지물질과 같은 셀에 포함된 문서가 전체문서의 45%인 239가지가 존재하는 것을 나타내며, 금지물질 혹은 허가물질 셀에 포함된 화학물질들에 대하여 guide text 상에서는 같은 취급방법을 사용하기 때문에 해당 물질들에 대하여 새로운 검토가 필요할 것이다. 또한 새로운 물질이 추가될 경우, MSDS 지도를 갱신하여, 새로운 물질이 추가되는 셀과 클러스터를 비교분석하여 다음과 같은 물질관리 방법을 참조할 수 있다.

Table 3. Hazardous and chemical materials in MSDS map

Properties	Vote	Total # of materials	Prohibited substances	Hazardous substances subject to permission	General substances
Toxic	95	56	• Antu(110) • Beta-Naphthylamine(113)	• Alpha-naphthylamine(101) • Coal tar pitch volatiles(116)	• Cresols(117) • Benzidine(111) • Etc
	151	43	• Disulfoton(168) • Monocrotophos(180) • Bis(2-chloroethyl)ether(151) • Parathion-methyl(177) • Parathion(185)	• Arsenic&inorganic(160)	• Ethylene(520) • Dioxathion(167) • etc
	64	31	• Chloropicrin(75) • 1,2-Dibromoethane(78)	• Chromite ore processing, as Cr(76)	• Barium sulfate(244) • Sodium hydroxide(88) • etc
	194	41	• DDT(205) • Dieldrin(206) • Lead arsenate(216) • Strychnine(225) • Thallium acetate(228) • Endosulfan(208) • Endrin(209) • Chlordan(203) • Arsenic pentoxide	–	• Carbon tetrabromide(201) • Methyl iodide(219) • etc
Flammable Liquids	289	44	• Aldrin(297)	–	• Chloroacetone(301) • Methanol(315) • etc
Etc	36	3	• 1,2-Dibromo-3-chloropropane(36)	–	• Acetylene tetrabromide(37) • Bromoform(38)
	4	21	• Crocidolite(10) • Amosite(5) • Anthophyllite asbestos(9) • Captan(12)	• Asbestos, chrysotile(6)	• Diuron(18) • Pyrethrum(22) • etc

5.1 물질이 금지물질과 허가물질 셀에 포함된 경우

금지물질과 허가대상 유해물질이 같이 포함되는 경우는 Fig. 10의 그림과 같이 대표문서번호 95, 151, 64, 4번의 문서가 같은 셀에 포함되었다. 금지물질은 제조·수입·양도·제공 또는 사용하여서는 안 되는 물질을 말하며, 허가대상 유해물질은 사용 전 고용노동부장관의 허가를 받은 후 사용이 가능한 물질을 말한다. 따라서, 물질을 재평가한 뒤 금지물질인지 허가대상인지를 위주로 확인할 필요가 있다. 95번 셀에서 확인할 수 있듯이 Cresols의 경우 같은 셀에 포함된 Benzidine과 같이 독성물질로 분류되어 위험성은 존재하지만, 법적규제가 존재하지 않아 위험성이 낮다고 판단되고 있다. 또한, 허가대상물질들도 금지대상물질과의 비교를 거쳐, 보다 엄밀한 기준에 따라 두 대상물질의 차이를 구별할 필요가 있을 것이다. 따라서 해당 물질과 같이 다른 물질들에 대해서도 재검토가 필요할 것이다.

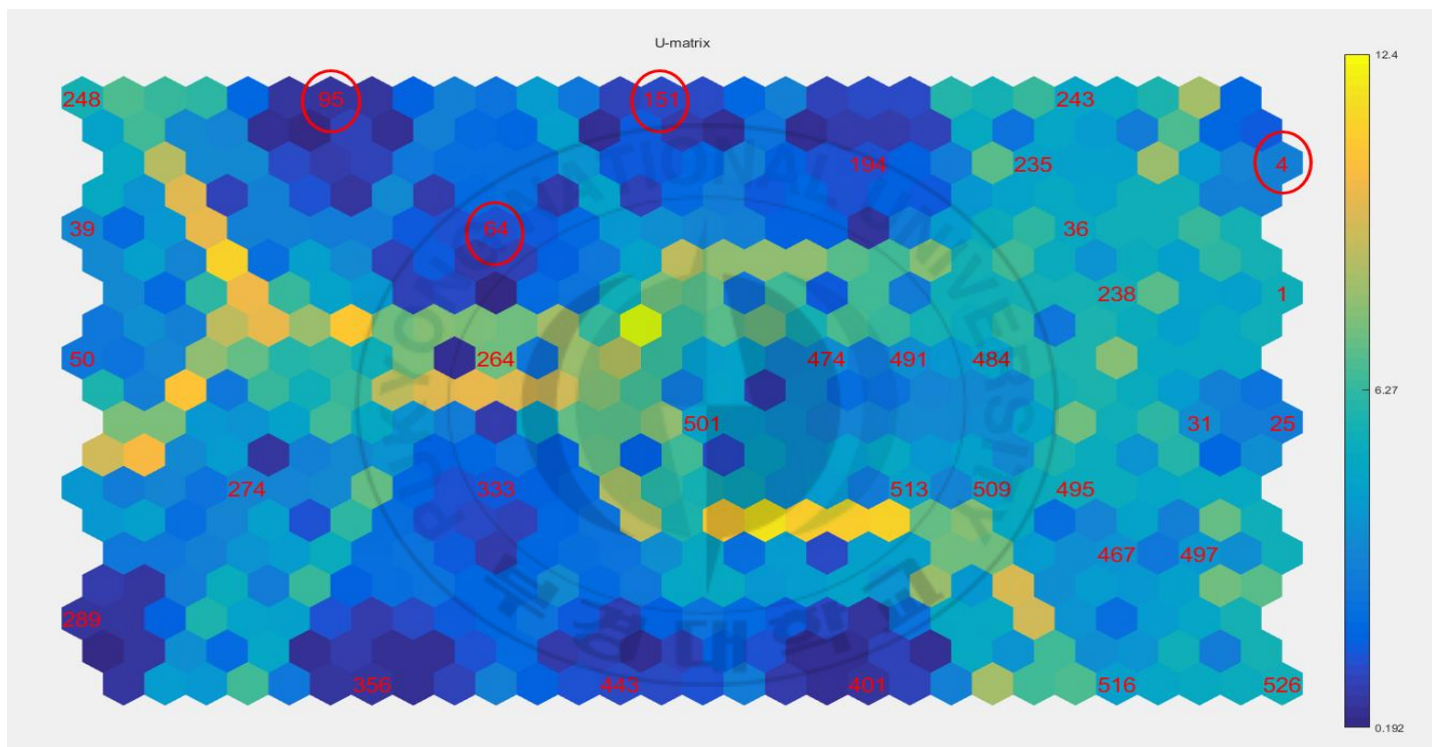


Fig. 10. Prohibited and Permitted substances

5.2 물질이 금지물질 셀에만 포함된 경우

금지물질에는 포함되지만 허가대상 유해물질에는 포함되어 있지 않은 경우는 보다 엄격히 관리대상으로 삼아 재평가할 필요가 있다. Fig. 11의 그림과 같이 194, 289, 36번 셀의 포함된 물질들의 경우 guide text상에서는 금지물질들과 같은 취급방법을 나타내고 있기 때문에 해당 물질들의 위험성을 재평가 할 필요가 있다. 따라서, Carbon tetrabromide, Methyl iodide, Chloroacetone, Acetylene tetrabromide, Bromoform 등을 대상으로 금지물질과 비교하여 평가를 할 필요가 높은 물질이라 볼 수 있다.

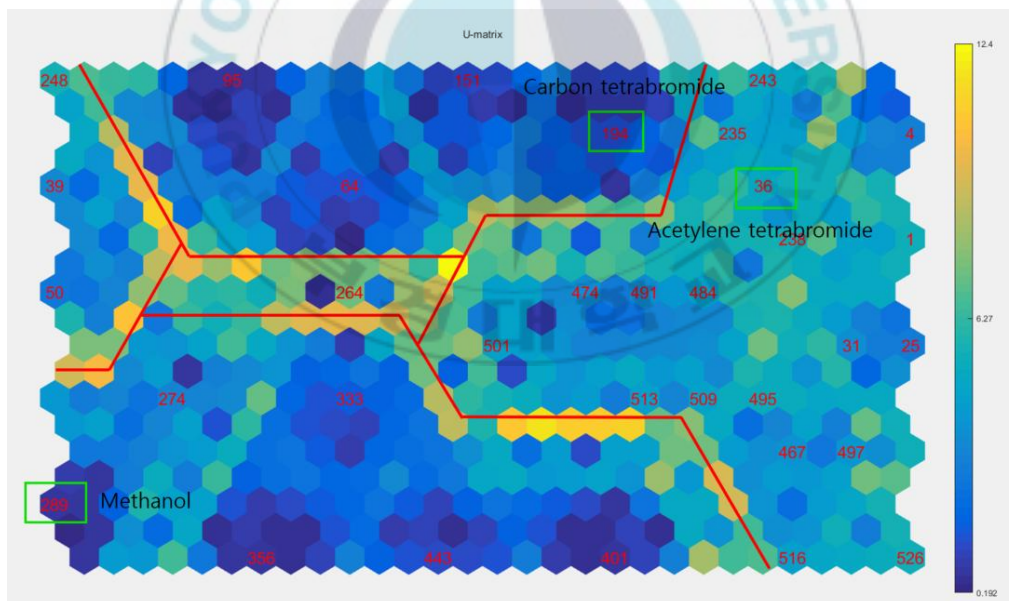


Fig. 11. General substance

Table 4. Guide text identical to prohibited material

Cell number	Substance name	Guide text
36	Acetylene terabromide	Substances (irritating) potential hazards health
194	Carbon terabromide	Toxic(non-combustible) potential hazards health
289	Methanol	Flammable liquids-Toxic potential hazards health

5.3 물질이 두 가지 셀 모두에 포함되지 않는 경우

MSDS 지도를 보면, 금지물질과 허가대상물질에 모두 포함되어 있지는 않지만, 그 물질들과 가까운 셀들도 보이게 된다. 이는 물질특징이 유사하나, 반응성이나 유통방법, 보관방법 등에 있어 그 차이가 어느 정도 안전한 물질이라고 볼 수 있다. 그러나 보다 안전한 물질자료 관리를 위해서는 금지물질과 가까운 위치에 있는 물질의 위험성을 검토할 필요가 있다. 예를 들어, Fig. 8의 36번 물질과 가까이 위치한 235번(Bromine pentafluoride etc.)과 238번(Ammonium persulfate etc.)의 물질들은 36번 물질과 guide text 내용이 유사하지만, 금지물질과 허가대상물질에 포함되지 않았다는 점에 비추어보아, 안전성을 재검토해야 할 가능성이 높다.

5.4 새로운 물질이 추가되는 경우

새로운 물질의 guide text를 기존의 MSDS 데이터베이스에 추가하여 새로운 MSDS 지도를 작성하여, 위치하는 셀이나 클러스터의 분석을 수행하면, 이 물질의 위험성을 예비적으로 파악할 수 있다. 예를 들어, Fig. 12와 같이 95번 주변 셀에 새로운 물질이 포함되었다면, 95번의 관리방법을 새로운 물질의 관리방법으로 참조할 수 있다.

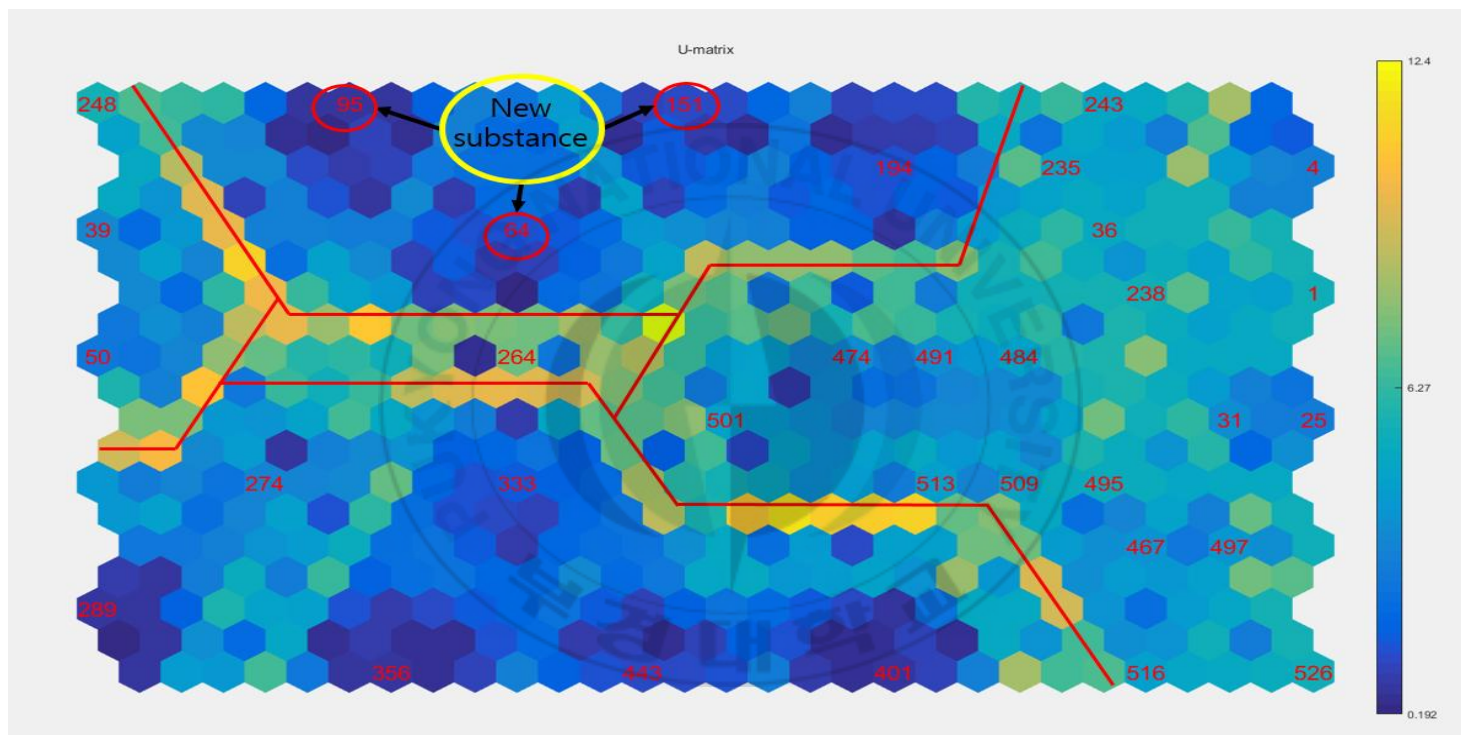


Fig. 12. Management method for new substances

제 6 장 결론

본 연구에서는 OSHA의 MSDS 자료에 포함되어 있는 guide text를 사용하여, 유해물질의 정보를 시각적으로 확인할 수 있는 MSDS 지도를 개발하였다. 이를 위해 텍스트마이닝을 활용하여 키워드를 도출하였고 SOM을 통하여 군집화 하였다. 또한, 화학물질의 특성별로 클러스터링 된 지도를 통하여 산업안전보건법 기준에 따라 분류하여 분석하였다. 그 결과 위험도가 높은 금지물질과 같은 셀의 분류된 물질 239가지에 대하여 재검토가 필요한 것을 확인하였고, 기존의 물질별 특성과 상관없이 알파벳순으로 제공되는 목록 형태의 MSDS에 비해, 유사성이 높은 MSDS 물질들을 시각적으로 표현하여 관리자와 근로자 모두에게 유해화학물질을 관리하는데 도움을 줄 것으로 기대된다.

그러나 본 연구에서는 전처리 과정에서 불용어를 제거하지 않고 실시하여 지시어, 상황 등에 대한 키워드 중복이 많았으며, OSHA의 800가지 화학물질 중에서도 guide text가 없는 문서를 제외하면 539가지 화학물질에 대한 자료로 수행하여 텍스트마이닝을 거쳐 키워드를 추출하였지만, 데이터가 부족하여 키워드 분석으로는 어려워 문서를 재검토하여 클러스터링 하는 과정을 거쳤다. 또한, 한국산업안전보건공단에서 제시하는 국내 MSDS의 각 유해물질문서들은 OSHA와는 다른 구성으로 제시되어 있다는 점도 국내 실무에 적용하기 위해서 추가적으로 고려해야 할 부분이다. 무엇보다도, 본 연구는 물성정보와는 관계없이 MSDS에서 제시하고 있는 관리 및 사용방법, 위험성 위주로 분석한 한계점이 있다. 따라서 향후, 전문가 자문을 통해 구축된 MSDS 지도에 물성정보도 같이 반영할 수 있는 형태로는 개발 프로세스도 추가적으로 수행할 필요가 있다.

참고문헌

1. J. Yang, C. Lim and S. Park, “ A Study on the Priority for the Hazard and Risk Evalution of Chemicals(HREC)according to the Industrial Safety and Health Act(ISHA)”, Journal of Korean Society of Occupational and Environmental Hygiene, Vol. 22, No. 1, pp. 73-81, 2012.
2. J.Park, S. Ham, S. Kim, K. Lee, K. Ha, D. Park and C. Yoon, “Study on the Chemical Management - 1. Chemical Characteristics and Occupational Exposure Limits under Occupational Safety and Health Act of Korea”, Journal of Korean Society of Occupational and Enviromental Hygiene, Vol. 25, No. 1, pp. 45-57, 2015.
3. B. Song, S. Choi, Y. Kang, G. Lyu, Y. Jo, H. Im, J. Kwon and K. Park, “Safety Management System for Emergency Handling of Environmental Toxic Gas Release”, Korean Journal of Hazardous Materials, Vol. 2, No. 2, pp. 27-31, 2014.
4. Ministry of Environment, Chemistry Safety Clearing-hOUSE, <https://csc.me.go.kr/>
5. C. Yoon, S. Ham, J. Park, S Kim, S. Lee, K. Lee and D. Park, “Comparison between the Chemical Management Contents of Laws Pertaining to the Ministry of Environment and the Ministry of the Employment and Labor”, Journal of Environmental Healyh Sciences, Vol. 40, No. 5, pp. 331-345, 2014.
6. Y. Suh, “Data Analytics for Social Risk Forecasting and Assessment of

New Technology”, J. Korean Soc. Saf., Vol. 32, No. 3, pp. 83-89, 2017.

7. S. Kang and Y. Suh, “On the Development of Risk Factor Map for Accident Analysis using Textmining and Self-Organizing Map(SOM) Algorithms”, J. Korean Soc. Saf., Vol.33, No. 6, pp. 77-84, 2018.

8. R. Moura, M. Beer, E. Patelli and J. Lewis, “Learning from Major Accidents : Graphical Representation and Analysis of Multi-attribute Events to Enhance Risk Communication”, Safety Science, Vol. 99, pp. 58-70, 2017.

9. T. Kohonen, “The Self-Organizing Map”, Neurocomputing, Vol. 21, No. 1-3, pp. 1-6, 1998.

10. T. Kohonen, “Self-Organization and Associative Memory”, Springer-Verlag, Berlin, Heidelberg, 2012.

11. T. Kohonen, “Essentials of the Self-Organizing Map”, Neural Networks, Vol. 37, pp. 52-65, 2013.

12. J. Vesanto and E. Alhoniemi, “Clustering of the Self-Organizing Map”, IEEE Transactions on Neural Network, Vol. 11, No. 3, pp. 586-600, 2000.

13. J. Kim and J. Cho, “Real-Time Change Detection Architecture Based on SOM for Video Surveillance Systems”, J. KIIT, Vol. 17, No. 4, pp. 109-117, 2019.

14. F. Palamara, F. Piglione and N. Piccinini. “Self-Organizing Map and Clustering Algorithms for the Analysis of Occupational Accident Database”, Safety Science, Vol. 49, pp. 1215-1230, 2011.