

# MEDLINE 검색을 통한 산업안전보건 분야에서의 인간공학 연구동향 : 워드임베딩을 활용한 초록 단어 모델링을 중심으로

김준희\* · 황의재\*\* · 안선희\* · 곽경태\*\*\* · 정성훈\*\*\*\*†

## Research Trends of Ergonomics in Occupational Safety and Health through MEDLINE Search: Focus on Abstract Word Modeling using Word Embedding

Jun Hee Kim\* · Ui Jae Hwang\*\* · Sun Hee Ahn\* · Gyeong Tae Gwak\*\*\* · Sung Hoon Jung\*\*\*\*†

### †Corresponding Author

Sung Hoon Jung

Tel : +82-33-760-2971

E-mail : movt\_architect@daum.net

Received : June 28, 2021

Revised : September 2, 2021

Accepted : September 24, 2021

**Abstract** : This study aimed to analyze the research trends of the abstract data of ergonomic studies registered in MEDLINE, a medical bibliographic database, using word embedding. Medical-related ergonomic studies mainly focus on work-related musculoskeletal disorders, and there are no studies on the analysis of words as data using natural language processing techniques, such as word embedding. In this study, the abstract data of ergonomic studies were extracted with a program written with selenium and BeautifulSoup modules using python. The word embedding of the abstract data was performed using the word2vec model, after which the data found in the abstract were vectorized. The vectorized data were visualized in two dimensions using t-Distributed Stochastic Neighbor Embedding (t-SNE). The word “ergonomics” and ten of the most frequently used words in the abstract were selected as keywords. The results revealed that the most frequently used words in the abstract of ergonomics studies include “use”, “work”, and “task”. In addition, the t-SNE technique revealed that words, such as “workplace”, “design”, and “engineering,” exhibited the highest relevance to ergonomics. The keywords observed in the abstract of ergonomic studies using t-SNE were classified into four groups. Ergonomics studies registered with MEDLINE have investigated the risk factors associated with workers performing an operation or task using tools, and in this study, ergonomics studies were identified by the relationship between keywords using word embedding. The results of this study will provide useful and diverse insights on future research direction on ergonomic studies.

Copyright©2021 by The Korean Society of Safety All right reserved.

**Key Words** : ergonomics, occupational safety and health, research trends, word embedding, word2vec, t-SNE

## 1. 서론

인간공학 연구는 1911년 Taylor의 과학적 관리법 (Scientific Management)이라는 이론을 통해 시작되었다<sup>1)</sup>. 이후 1920년대 Frank와 Gilbreth가 동작에 관한 연구와 공장 경영에 관한 연구를 수행하였다<sup>2)</sup>. 이들은 주로 숙

련된 작업과 이에 관련된 피로감 혹은 장애인들을 위한 작업장과 장비 디자인 등에 관하여 연구하였다<sup>2)</sup>. 하지만 이러한 연구들을 통한 공장 장비나 절차의 개선이 계획 및 실행되지 못하였다. 그 후 제2차 세계대전이 일어나게 되고, 본격적으로 군수 산업을 중심으로 인간공학 인력들을 고용하여 생산품에 대한 인간의 신체적,

\*코리아테크 R&D 책임연구원 (Research and Development, KOREATECH Corporation)

\*\*연세대학교 물리치료학과 연구교수 (Department of Physical Therapy, Yonsei University)

\*\*\*쿠팡 EHS Sr. Health & Ergonomics (EHS, Coupang Corporation)

\*\*\*\*연세대학교 물리치료학과 박사후과정 (Department of Physical Therapy, Yonsei University)

인지적, 감성적, 사회문화적 요소들을 반영하고자 하였다<sup>3)</sup>. 이후 1980년대까지 군수 산업 혹은 우주 비행 분야에 집중되었던 인간공학이 의학, 자동차, 컴퓨터 등 점차 여러 분야에 확대되기 시작하였다<sup>4,5)</sup>. 1990년대 이르러 미국 산업안전보건청 (OSHA)에서는 안전보건을 저해하는 근골격계 질환의 발생을 줄이기 위한 목적으로 인간의 신체와 심리적 능력을 고려하여 작업 환경을 갖추는 인간공학 프로그램을 마련하였다<sup>6)</sup>. 이를 통해, 각 산업 분야들은 제품, 작업, 환경 등을 설계하는 데 있어 인간공학의 중요성이 점차 증가하였다. 우리나라는 1931년 동아일보에 게재된 칼럼인 ‘산업발달과 인간공학’을 통해서 인간공학이 처음 소개되었으며, 1967년 군수상보에 ‘인간-기계 계통’의 게재를 통하여 인간공학의 기본 개념이 소개되었다<sup>7)</sup>. 이후, 우리나라의 인간공학은 1980년대부터 본격적으로 수행되었다고 볼 수 있다. 1982년에 대한인간공학회가 설립되었으며, 한국표준과학연구소와 한국원자력연구소 등 국가 출연연구소에 인간공학을 연구하는 팀이 구성되기도 하였다<sup>7)</sup>. 이와 더불어 인간공학을 접목한 제품 개발을 위한 산학연구가 본격적으로 진행되면서 다양한 기업들이 인간공학과 관련된 연구에 참여하기 시작하였다<sup>7)</sup>.

우리나라는 2000년대부터 산업보건 분야에 있어 근골격계 질환이 중요한 문제로 제기되기 시작하였다<sup>8)</sup>. 산업이 점차 고도화되면서 노동의 강도가 강화되고 인체에 부담을 주는 작업이 증가하였고 이는 곧 근골격계 질환의 증가로 이어졌다. 하지만 이 시기에 개인의 건강과 복지에 관한 관심이 높아지면서 근골격계 질환 문제는 양대 노총의 임·단협의 핵심쟁점으로 대두될 정도로 노사 간 갈등 문제의 중심이 되기도 하였다<sup>9)</sup>. 이에 따라 2003년부터 정부는 근골격계 질환을 예방하기 위한 사업주의 의무를 산업안전보건법에 명시하였고, 유해요인조사를 의무적으로 실시하도록 하였다<sup>7)</sup>. 이후 대기업을 중심으로 근골격계 질환에 관한 특성들을 이해하기 시작하면서 작업장 안전보건 강화를 위해 인간공학적 작업 환경 개선이나 인간공학 프로그램을 도입하는 등 대처 방안들을 마련하였고 근골격계 질환들을 예방 및 개선하기 시작하였다<sup>10,11)</sup>. 그 결과, 점점 많은 기업이 근골격계 질환을 경제적인 비용이 발생하는 문제로 인식하게 되었다. 위와 같은 이유로 근골격계 질환이 발생하지 않도록 인간공학적 작업장 평가와 작업환경개선, 의학적 조치, 교육 및 훈련 등을 통하여 근골격계 질환에 대한 유해요인들을 제거하고 사업장 안전보건 체계를 강화하려는 방법들이 다방면으로 수행되고 있다<sup>12,13)</sup>. 이처럼 인간공학은 인간의 신체와 심리적 능력을 고려하여 안전과 능률의 향상을 목적으로

발전해왔다.

인간공학은 크게 신체 특성과 관련된 영역, 인지 및 관련된 영역, 기계와 관련된 영역 등으로 나누어 볼 수 있는데, 인간공학에서 신체 특성과 관련된 영역은 의학 분야 연구들과 밀접한 연관이 있다<sup>9)</sup>. 인간공학에서 대표적으로 다루지는 주제 중 하나인 근골격계 질환은 반복적이고 누적되는 특정한 일이나 동작과 연관되어 신체 일부를 무리하게 사용했을 때 나타나는 질환으로 신경, 근육, 인대, 관절 등에 문제가 발생하여 통증이나 이상 감각, 마비 등의 증상이 일어나는 질환들을 총칭하는 용어이다<sup>9)</sup>. 특히 오랜 시간 동안 반복적인 작업으로 인한 외부의 스트레스가 누적되어 질병이 나타나기 때문에 누적 외상성 질환이라고 부르기도 한다<sup>9)</sup>. 이외에도, 공학 심리학, 인지 공학 등 인간의 심리와 인지 영역을 공학과 연결하여 인간과 기계의 효과적인 상호작용을 설명하고자 하는 노력을 지속해서 수행하고 있다<sup>14-16)</sup>. 이처럼 인간은 인간공학의 중심이 되는 구성요소이기에 의학 분야는 인간공학과 관련성이 높은 기초 학문 가운데 하나로 여겨지고 있다.

Word2vec은 비정형 데이터인 문자의 의미를 다차원 벡터 공간에 표현하여 이를 통해 벡터 간의 유사도를 계산할 수 있도록 해주는 워드임베딩 방법 중 하나이다<sup>17)</sup>. 특히, 주변 단어를 고려하여 임베딩을 하므로 단어들의 문맥상 의미도 정량화하여 문법적 해석과 단어들의 거리를 통한 의미론적 추론으로 새로운 정보를 습득할 수 있게 해준다<sup>17)</sup>. 기존 인간공학 연구들은 작업자들이 가지고 있는 특성이나 작업 환경을 기반으로 그에 대한 근골격계 질환의 위험성이나 발생률을 파악하거나 이에 대한 예방과 개선의 효과를 검증하기 위해 주로 연산이 가능한 정형 데이터들을 사용하고 있다. 하지만, 방대하게 존재하는 인간공학 연구들 가운데 의학 분야와 관련성이 높은 인간공학 연구들의 비정형 데이터인 문자 자체를 데이터로 이용해 인간공학 연구들에 대한 동향과 특징을 분석해본 연구는 아직 이루어지지 않았다.

그래서, 본 연구는 미국 국립 의학도서관에서 제공하는 대표적인 의학 분야 데이터베이스인 MEDLINE에 등록된 인간공학 연구들의 초록 데이터를 이용하여 Word2vec 기반의 워드임베딩 방법으로 전반적인 연구 동향과 특징들을 살펴보았다. 또한 연구들의 초록에서 사용되는 주요 키워드들을 선정하고 이들의 관계를 분석하고자 하였다. 즉, Word2vec을 활용하여 초록을 구성하는 단어들을 벡터화하고, 문맥을 통해 단어를 학습시켰다. 의미를 보존하기 위해 벡터 공간에 존재하는 논문들의 주요 단어들과 관련성이 높은 단어들을

통해 이들의 관계를 분석해보았다. 이처럼, 본 연구는 의학 서지 데이터베이스인 MEDLINE에 등록되어 있는 ‘인간공학’ (Ergonomics)을 주제로 하는 논문들의 초록 데이터를 워드임베딩 방식으로 모델링하여 연구 동향을 분석하고자 하였다.

## 2. 이론적 배경

### 2.1 워드임베딩

컴퓨터가 인간의 언어를 이해하도록 언어를 숫자로 변환하는 자연어 처리 기법을 이용하면 문자로 구성된 데이터로부터 의미 있는 정보를 획득할 수 있다. 이러한 방법으로는 신경망 기반 학습 방식을 이용하는 워드임베딩이 있다. 워드임베딩은 컴퓨터가 자연어를 이해할 수 있도록 자연어를 벡터로 변환하는 과정을 의미하며 분산 표현이라고도 한다<sup>17)</sup>. 즉, 워드임베딩 과정을 통해 언어를 모델링 하면 단어 간의 유사도를 반영할 수 있도록 단어들의 의미를 숫자 형태인 벡터로 표현할 수 있다. 이러한 방식은 단어에 대한 개념을 쉽게 이해할 수 있는 효과를 줄 수 있으며 특정 단어들에 대한 의미적인 분석을 통해 다양한 주제에 대한 해석이 가능해진다. 이처럼 워드임베딩 방법을 통한 자연어처리기법들은 현재까지 활발하게 연구되고 있으며, 컴퓨터 공학, 산업, 경제학, 건설 등 다양한 분야에서 활용되고 있다<sup>18-20)</sup>. 이렇듯, 워드임베딩 방법을 이용하면 기존의 방식으로 다루어진 주제들로부터 점점 더 다양하고 넓은 주제에 대한 분석과 해석이 가능해진다.

### 2.2 Word2vec

구글 (Google)의 Mikolov 등<sup>21)</sup>이 제안한 Word2vec 모델은 단어를 수학적 형태로 표현하여 벡터의 형태로 의미를 나타내는 워드임베딩 방법의 대표적인 모델이다. 이러한 방법은 단어들의 의미 관계가 기계에 보존되도록 머신러닝 알고리즘을 이용하는 것이다. Word2vec 모델은 특정 임베딩 공간상에서 같은 유사도를 갖는 단어들은 가까운 거리를 가진다는 가정을 바탕으로 단어들을 벡터의 형태로 보존한다. 이처럼, 워드임베딩 방식으로 Word2vec 모델을 이용하면 주어진 문장에 대한 단어들이 벡터의 형태로 보존되어 이에 대한 문법적 해석과 단어들의 거리를 통해 의미론적 추론이 가능하다. 가장 대표적인 예로, Word2vec을 통해 학습시킨 단어들 가운데 ‘한국’, ‘서울’, ‘일본’, ‘도쿄’가 존재한다고 할 때, ‘한국’ - ‘서울’ + ‘일본’= ‘도쿄’와 같은 의미론적 추론이 가능한 형태로 단어들이 보존되는 것이다<sup>22)</sup>. Word2vec 모델은 학습하고 예측하는 데 있

어 CBOW (Continuous Bag-Of-Words)나 Skip-gram 방법을 이용한다. CBOW 방법은 Word2vec 모델의 하이퍼 파라미터인 문맥 윈도우  $k$  값을 이용하여 주변에  $2k$  개의 단어들이 존재할 때 그 중심에 특정 단어가 나타날 확률을 계산하는 방식이고, Skip-gram 방법은 중심에 단어가 제시되었을 때  $2k$  개의 특정 조합을 가진 주변 단어가 나타날 확률을 계산하는 방식으로 워드임베딩을 수행한다. 두 가지 방법 중에서 Skip-gram 방법이 CBOW 방법에 비교하여 연산을 수행하는 시간이 더 오래 걸리기 때문에 속도는 더 느리지만 대신 정확도가 더 높다고 보고되고 있다<sup>22)</sup>. Cosine similarity는 내적 공간의 두 벡터 간 각도의 코사인값을 이용하여 측정된 벡터 간의 유사한 정도를 의미한다. 데이터가 1에 가까울수록 두 단어가 양의 상관관계를 가지고 있다는 것을 의미한다. 0에 가까울수록 관련성이 없다고 볼 수 있으며 -1에 가까울수록 음의 상관관계를 가지고 있다는 것을 의미한다.

### 2.3 t-SNE (t-Distributed Stochastic Neighbor Embedding)

Word2vec 모델을 통해 워드임베딩을 수행한 데이터들은 다차원으로 구성되어 있어 이를 시각화하기 위해서는 더 적은 차원으로 데이터들을 축소해야 한다. 이를 위해, 비선형적 차원 축소 방법으로 데이터상의 군집들을 다른 방법들보다 더 잘 보여주는 경향이 있다고 보고된 Maaten와 Hinton<sup>23)</sup>이 제안한 t-SNE 방법이 주로 사용되고 있다.

## 3. 연구 방법론

### 3.1 자료 수집

본 연구에서는 인간공학 관련 해외 임상 의학 연구를 수행한 논문들을 선정하기 위해 펌메드 사이트를

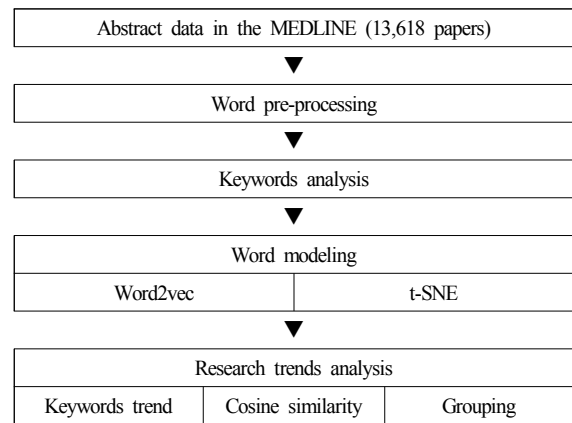


Fig. 1. Flowchart of the study.

(<https://pubmed.ncbi.nlm.nih.gov/>)에서 검색어를 ‘인간공학’ (“Ergonomics”)으로 수행하였다. 펍메드에 초록이 등록되어 있고 의학 서지 데이터베이스인 MEDLINE에 해당하는 논문은 2021년 3월 30일 기준 13,693건이 등록되어 있었다. Python을 이용하여 Selenium 모듈과 BeautifulSoup 모듈로 작성한 프로그램으로 논문들의 제목, 저자, 저널명, 초록 데이터, 출판일 등을 수집하였다. 수집된 자료들 가운데 중복에 해당하는 논문들을 제외한 결과 13,618건의 논문 데이터들이 수집되었다.

### 3.2 단어 전처리

키워드 분석과 워드임베딩 프로그램을 수행하는 데 있어 불필요한 단어들은 전처리로 제거작업을 수행하였다. 초록 형식에서 각 문단의 제목으로 사용되는 단어 (Background, Purpose, Method, Result, Conclusion 등), 연구 방법론적으로 사용되는 단어 (Study, Group, Control 등)들을 분석 단어에서 제외되도록 처리하였다. 또한, 복수형 및 단수형 명사들이 하나의 단어로 통합되고 단수형 동사 및 복수형 동사의 통합이 이루어지도록 NLTK (Natural Language Toolkit) 모듈의 Lemmatization을 활용하였다. 이와 함께, 분석에 중요하지 않은 관사 (a, an, the), 전치사 (of, on, for), 대명사 (It, He, Her), 특수 문자 및 숫자 데이터와 같은 단어들을 제거되도록 프로그램을 수행하였다.

### 3.2 키워드 빈도 분석

Python의 Pandas 모듈과 카운터 함수로 키워드 분석 프로그램을 작성하여 “인간공학” 관련 연구 논문들의 초록을 구성하는 단어들의 빈도를 분석하였다. 이때, 같은 의미가 있는 단어들은 같은 단어로 병합하도록 처리하여 분석결과의 정확성이 향상되도록 하였다. 분석결과에 따라 빈도수가 높은 단어 중에서 검색어인 ‘인간공학’을 제외한 10개의 단어를 키워드로 선정하여 분석하였다.

### 3.3 Word2vec 모델링

‘인간공학’을 주제로 하는 논문들의 초록 데이터를 Python의 Gensim 모듈의 Word2Vec 함수를 이용하여 정확도가 더 높은 Skip-Gram 방식을 이용하였다. 해당 모델을 개발한 구글이 뉴스 데이터셋을 학습시키는 데 사용하였던 하이퍼 파라미터값을 이용하여 벡터 크기를 300, 문맥 윈도우를 10으로 설정하고 초록을 구성하는 단어들을 학습시키는 모델링을 진행하였다<sup>24)</sup>. 그 다음 키워드들과 연관성이 높은 단어들은 Cosine similarity를 이용하여 그 유사한 정도를 수치로 나타내었다.

### 3.4 t-SNE

t-SNE 방법을 이용하여 Word2vec 모델링으로 300차원 공간에 배열된 단어들을 2차원 공간으로 시각화하였다. 그다음 ‘인간공학’과 높은 빈도를 나타낸 키워드 10개 단어와 거리가 가까운 단어들을 분석하였다. 이때 나타내는 값이 작을수록 데이터 간 거리가 가깝고 데이터 간의 유사성 혹은 관련성이 높은 것으로 분석하였다.

## 4. 결과 분석

### 4.1 자료 수집 결과

미국 국립 의학도서관에서 운영하는 의학 서지 데이터베이스인 MEDLINE에 ‘인간공학’을 검색어로 수행하였을 때 검색되는 논문 가운데, 중복을 제외한 13,618편의 논문 초록 데이터들이 수집되었다. 이를 10년 단위로 분류하였을 때 1989년 이전에 출판되어 의학 서지 데이터베이스에 등록된 ‘인간공학’ 관련 연구들은 460편에 불과하였다. 하지만 1990년대에 출판된 논문은 2,125편의 논문이 의학 서지 데이터베이스에 등록되었으며, 2000년대는 4,029편, 2010년대에는 6,408편이 의학 서지 데이터베이스에 등록되었다 (Fig. 1). 이를 통해, 의학 서지 데이터베이스에 등록되는 ‘인간공학’을 주제로 하는 연구들이 큰 폭으로 증가하는 것으로 나타났다. 또한, 최근 10년간의 논문 숫자를 살펴보면 급격한 증가를 나타냈던 2012년을 제외하고는 매해 꾸준히 연구들이 증가하는 것으로 나타났다 (Fig. 2).

### 4.2 키워드 빈도 분석결과

의학 서지 데이터베이스에 ‘인간공학’을 주제로 하는 논문들의 초록에서 가장 많이 사용된 단어 10개를 연도별 키워드 단어로 선정하였다 (Table 1). 전체 연구

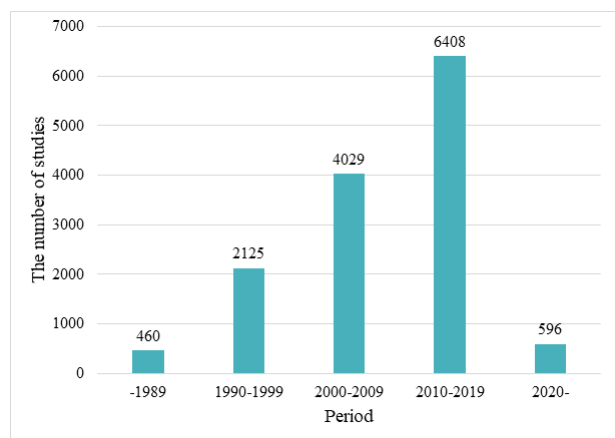


Fig. 2. Number of studies in 10 years.

Table 1. 10 top keywords and frequency

-1989		1990-1999		2000-2009		2010-2019		2020-2021		Total (13,618)	
Word	Count	Word	Count	Word	Count	Word	Count	Word	Count	Word	Count
Work	502	Use	2091	Use	4913	Use	8755	Use	863	Use	16907
Use	285	Work	1763	Work	3429	Work	5681	Work	531	Work	11906
System	197	Task	1400	Task	2479	Task	3910	System	387	Task	8227
Time	176	Time	984	System	2174	Design	3494	Factor	382	System	7173
Worker	176	System	976	Factor	2023	System	3439	Time	357	Factor	6772
Subject	174	Factor	831	Design	1945	Factor	3394	Risk	334	Design	6633
Muscle	163	Design	760	Time	1784	Risk	3308	Health	305	Time	6283
Load	161	Condition	760	Worker	1714	Time	2982	Worker	304	Risk	5784
Design	147	Muscle	742	Effect	1632	Worker	2784	Task	296	Worker	5624
Task	142	Effect	738	Risk	1572	Effect	2731	Design	287	Effect	5434

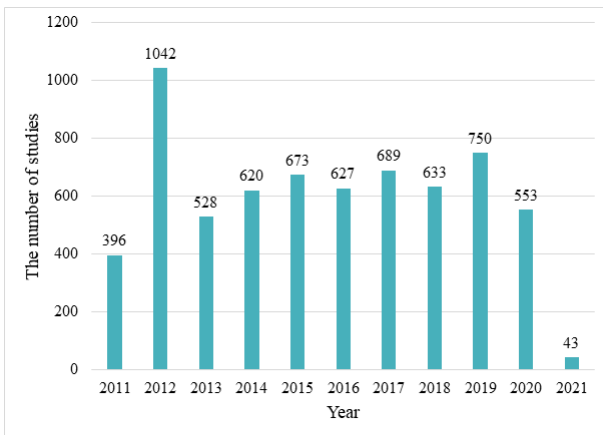


Fig. 3. Number of studies per year within the last 11 years.

에서 공통으로 가장 많이 사용된 단어는 ‘사용하다’ (Use), ‘일’ (Work), ‘과제’ (Task)로 나타났다. 이와 같은 단어들을 제외하면, 1989년 이전 시기에는 ‘시스템’ (System), ‘시간’ (Time), ‘근로자’ (Worker)의 단어들이 높은 빈도로 사용되었다. 1990년대에는 이와 함께 ‘요인’ (Factor), ‘디자인’ (Design), ‘상태’ (Condition), ‘근육’ (Muscle) 등의 단어들이 높은 빈도로 사용되었다. 2000년대에는 이전에 높은 빈도로 나타나지 않았던 ‘위험’ (Risk)이라는 단어가 추가적으로 높은 빈도로 사용되었으며, 2010년대에는 이전과 유사한 단어들이 지속적으로 높은 빈도를 나타냈다. 2020년대에 이르러서는 ‘건강’ (Health)이라는 단어가 새롭게 높은 빈도로 사용되기 시작했다.

### 4.3 word2vec 모델링 결과

‘인간공학’과 빈도수가 가장 높은 상위 10개의 키워드를 Word2vec으로 모델링을 수행한 결과는 Table 2에 나타났다. 키워드들은 전체 논문들의 초록 데이터에서

가장 자주 사용되었던 단어들과 이들과 높은 관련성과 유사성을 나타내는 단어의 경우 1에 가까운 숫자로 표현된다. ‘인간공학’과 가장 높은 관련성을 나타낸 단어는 ‘작업장’ (Workplace)이었다. 그다음으로 ‘디자인’, ‘공학’ (Engineering), ‘작업’ (Occupational), ‘안전’ (Safety), ‘건강’, ‘교육’ (Education)이 뒤를 이어 높은 연관성을 나타냈다. ‘일’ (Work)과 관련성이 높은 단어는 ‘작업장’, ‘근로자’, ‘직업’ 등이 있었고, 특징적으로 ‘활동’ (Activity), ‘근골격계’ (Musculoskeletal)가 있어서 높은 관련성을 나타냈다. ‘과제’와 연관성이 높은 단어들은 ‘모의’ (Simulated), ‘수행력’ (Performance), ‘활동’, ‘실험’ (Experiment)이 있었다. ‘시스템’과 관련성이 높은 단어에는 ‘기술’ (Technology), ‘장비’ (Device), ‘과정’ (Process), ‘인간’ (Human), ‘접근’ (Approach), ‘소프트웨어’ (Software) 등이 나타났다. ‘요인’은 ‘심리사회적’ (Psychosocial), ‘이슈’ (Issue), ‘위험’ 등의 단어와 높은 연관성을 나타냈다. ‘디자인’과 연관성이 높은 단어에는 ‘평가’ (Evaluation), ‘사용자’ (User), ‘발전’ (Development), ‘새로운’ (New), ‘연구’ (Research) 등이 나타났다. ‘위험’ (Risk)과 연관성이 높은 단어는 ‘장애’ (Disorder), ‘위험’ (Hazard), ‘근골격계’, ‘손상’ (Injury) 등이 존재했다. ‘효과’ (Effect)와 관련성이 높은 단어에는 ‘영향’ (Influence), ‘충격’ (Impact), ‘차이’ (Difference) 등이 있었다.

### 4.4 t-SNE 결과

Word2vec으로 벡터화된 단어들의 데이터를 t-SNE 방법을 이용하여 다시 2차원으로 시각화하였다 (Fig. 3). 이때, 초록을 구성하는 단어들 가운데 빈도가 10회 이상인 모든 단어를 2차원 배열에 위치하도록 하였다. 이 가운데 ‘인간공학’과 빈도 상위 10개의 키워드 단어들과 가장 가까운 거리에 있는 단어들을 Table 3에 나

Table 2. Word2vec results of 11 keywords

	Ergonomics	Use	Work	Task	System	Factor	Design	Time	Risk	Worker	Effect
1	Workplace	Based	Workplace	Simulated	Technology	Psychosocial	Ergonomics	Minute	Disorder	Employee	Influence
	0.463	0.465	0.458	0.373	0.401	0.385	0.445	0.396	0.446	0.556	0.497
2	Design	Developed	Worker	Performance	Device	Issue	Designer	Longer	Hazard	Job	Impact
	0.445	0.412	0.456	0.360	0.385	0.377	0.408	0.386	0.422	0.478	0.402
3	Engineering	Designed	Job	Activity	Process	Risk	Evaluation	Duration	Musculoskeletal	Work	Difference
	0.405	0.383	0.449	0.340	0.384	0.367	0.396	0.385	0.413	0.456	0.351
4	Occupational	Applied	Occupational	Experiment	Human	Ergonomics	User	Day	Injury	Workplace	Relationship
	0.396	0.378	0.392	0.321	0.379	0.363	0.394	0.361	0.412	0.450	0.346
5	Safety	Performed	Activity	Manual	Approach	Related	Development	Period	Associated	Industry	Association
	0.387	0.344	0.360	0.308	0.364	0.361	0.380	0.353	0.402	0.403	0.311
6	Health	Design	Musculoskeletal	Tracking	Software	Aspect	New	Average	Developing	Occupational	Affect
	0.387	0.341	0.360	0.299	0.363	0.339	0.374	0.340	0.383	0.373	0.287
7	Education	Technique	Organization	Demand	Interface	Characteristic	Research	Interval	Occupational	Health	Interaction
	0.384	0.340	0.352	0.288	0.341	0.329	0.365	0.325	0.380	0.368	0.282
8	Principle	Utilized	Ergonomics	Job	Framework	Work	Engineering	Spent	Incidence	Musculoskeletal	Investigate
	0.379	0.339	0.352	0.280	0.334	0.328	0.360	0.316	0.374	0.362	0.282
9	Research	New	Health	Complex	Application	Organizational	Usability	Hour	Prevalence	Compensation	Experiment
	0.375	0.337	0.341	0.277	0.325	0.314	0.352	0.315	0.368	0.355	0.279
10	Technical	Analysis	Physical	Pushing	Use	Disorder	Process	Shorter	Factor	Workforce	Whether
	0.372	0.336	0.335	0.273	0.324	0.309	0.345	0.311	0.367	0.353	0.268

타냈다. 해당 키워드와 높은 연관성을 지닌 단어들은 가까운 거리에 위치하게 된다. 이차원으로 시각화된 데이터들을 바탕으로 ‘인간공학’ 연구들의 키워드들은 네 개의 그룹으로 구분될 수 있었다. 각각의 그룹은 A, B, C, D그룹으로 나타났다. A그룹은 ‘인간공학’을 중심으로 ‘사용’, ‘디자인’, ‘시스템’의 키워드들, 그리고 이들과 가까이 위치한 단어들이 그룹을 형성하고 있었다. B그룹은 ‘효과’, ‘요인’, ‘위험’ 키워드들을 중심으로 이들과 가까운 거리에 있는 단어들이 그룹을 이루었다. C그룹은 ‘일’, ‘근로자’ 키워드들을 중심으로 이들과 가까운 거리에 있는 단어들이 그룹을 형성하였다. D그룹은 ‘과제’, ‘시간’ 키워드들을 중심으로 이들과 가까운 거리에 있는 단어들이 그룹을 이루었다.

### 5. 고찰

현대사회는 많은 양의 텍스트 데이터들이 굉장히 빠른 속도로 생산 및 축적되고 있다. 이러한 텍스트 데이터를 컴퓨터가 이해하도록 처리하고 이 가운데 의미 있고 필요한 정보만을 효율적으로 추출하여 활용하는 것은 다양한 영역에서 인간이 발견하지 못할 수도 있는 정보를 발견해 줄 가능성을 제시해준다. 본 연구는 미국 국립 의학도서관의 의학 서지 데이터베이스에 ‘인간공학’을 키워드로 포함하고 있는 연구들의 초록

데이터들을 텍스트 데이터 처리 방법의 하나인 워드임베딩 방법을 이용하여 데이터들을 모델링하고 이를 분석하였다. 의학 서지 데이터베이스에서 ‘인간공학’을 키워드로 검색하여 분석에 이용된 논문은 1959년부터 2021년 3월 30일까지 총 13,618편이었다. 초록 데이터를 추출하고 모델링 하여 분석한 결과 ‘인간공학’ 연구와 관련하여 다음과 같은 특징과 연구 동향을 살펴볼 수 있었다.

#### 5.1 인간공학 연구 건수

의학 서지 데이터베이스에 등록된 ‘인간공학’ 연구는 1959년을 시작으로 현재에 이르기까지 꾸준히 증가하는 추세를 나타냈다. 출판된 논문들은 10년 단위로 살펴보았을 때, 그 수가 거의 1.5 ~ 2배에 가까운 숫자로 증가하였다. 이 중 2011년도부터 매해 출판되는 인간공학 관련 연구들은 약 400편에서 700편 정도가 꾸준히 의학 서지 데이터베이스에 등록되고 있다. 특히 2012년은 의학 관련 인간공학 연구가 굉장히 활발히 진행되어 1,042편의 논문이 의학 서지 데이터베이스에 등록되어 있다.

#### 5.2 키워드 빈도

전체 논문들의 초록에서 가장 많이 사용되는 단어들을 통해 의학 관련 ‘인간공학’은 ‘일’이나 ‘과제’를 수

Table 3. Distance of 11 keywords in t-SNE

	Ergonomics	Use	Work	Task	System	Factor	Design	Time	Risk	Worker	Effect
1	Workplace	Based	Worker	Simulated	Technology	Risk	Development	Longer	Disorder	Employee	Influence
	0.599	0.556	0.597	0.680	0.663	0.677	0.653	0.650	0.605	0.477	0.483
2	Design	Developed	Job	Performance	Device	Psychosocial	User	Duration	Hazard	Job	Impact
	0.620	0.598	0.622	0.684	0.689	0.682	0.663	0.659	0.616	0.573	0.616
3	Engineering	Applied	Workplace	Activity	Interface	Related	Evaluation	Minute	Associated	Work	Difference
	0.681	0.683	0.622	0.713	0.699	0.684	0.690	0.668	0.628	0.597	0.620
4	Occupational	Designed	Occupational	Manual	Approach	Characteristic	New	Average	Musculoskeletal	Workplace	Relationship
	0.691	0.684	0.679	0.739	0.699	0.724	0.711	0.689	0.648	0.615	0.671
5	Health	New	Activity	Experiment	Human	Issue	Designer	Interval	Injury	Employer	Association
	0.698	0.705	0.684	0.745	0.708	0.729	0.711	0.697	0.649	0.705	0.713
6	Safety	Performed	Musculoskeletal	Complex	Process	Stressor	Research	Day	Stress	Occupational	Interaction
	0.703	0.710	0.703	0.764	0.721	0.735	0.713	0.714	0.675	0.708	0.732
7	Factor	Order	Physical	Time	Software	Aspect	Engineering	Rate	Factor	Operator	Experiment
	0.711	0.715	0.709	0.772	0.723	0.739	0.721	0.719	0.677	0.723	0.751
8	Research	Provided	Factor	Motion	Use	Work	Approach	Period	Developing	Compensation	Change
	0.712	0.726	0.741	0.779	0.727	0.741	0.731	0.727	0.710	0.723	0.755
9	Education	System	Organization	Use	Design	Role	Usability	Increased	Occupational	Health	Investigate
	0.719	0.727	0.742	0.783	0.742	0.745	0.736	0.733	0.711	0.723	0.762
10	Principle	Data	Stress	Situation	New	Problem	Use	Mean	Prevalence	Musculoskeletal	Objective
	0.719	0.731	0.746	0.784	0.744	0.748	0.741	0.741	0.711	0.732	0.764

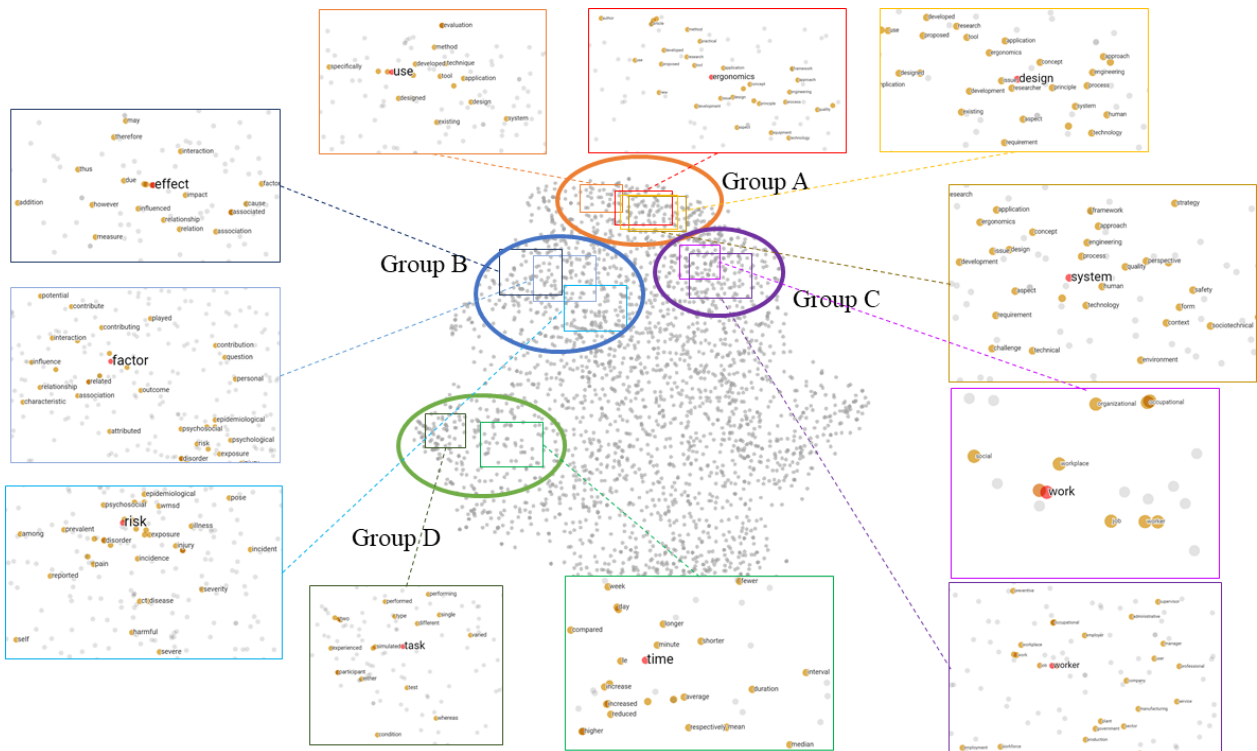


Fig. 3. t-SNE 2D visualization of 11 keywords.

행하기 위해 무엇인가를 ‘사용’하여 수행하는 일련의 과정에 관한 연구들이 주를 이루었다는 사실을 유추해 볼 수 있다. 즉, 논문들은 인간이 도구를 사용하는 영

장류로써 주어진 작업 환경 내에서 도구를 사용하여 작업을 수행하는 것을 가장 기본으로 연구했던 것으로 보인다.

1989년도 이전까지는 ‘시스템’, ‘시간’, ‘근로자’의 단어들이 높은 빈도로 사용되었는데, ‘근로자’들에게 ‘시간’ 효율적이고 인체의 신체적 특성에 적합한 작업장 환경을 제공하기 위하여 ‘시스템’을 구축하기 위한 연구들이 진행됐음을 알 수 있다.

1990년대에는 작업장에서의 근로 환경이 인체에 미칠 수 있는 ‘요인’들에 대한 분석과 함께 근로자의 ‘상태’와 ‘근육’이 연관된 신체적 특성에 관한 연구들이 증가한 것으로 볼 수 있다. 즉, 근골격계 질환을 예방하기 위한 인간공학적 연구들이 본격적으로 수행되기 시작했다는 것을 의미한다. 또한, 이 시기에는 사용자 중심의 인간공학적 ‘디자인’에 대한 연구들도 활발하게 진행된 것으로 볼 수 있다.

2000년대에 이르러서는 ‘위험’에 관한 요소들이 더욱 주목받아 연구가 진행되었다. 이는 인체에 악영향을 미치는 위험 요소에 대해 미리 예방하고 대비하기 위함이라 유추해 볼 수 있다. 이 시기에 맞추어 우리나라도 2000년대부터 급격하게 증가하기 시작한 작업성 근골격계 질환을 예방하기 위해 인체에 가해지는 위험 요소 관한 연구들이 활발하게 진행되었다<sup>8,11,25)</sup>.

2020년대에 이르러서는 ‘건강’ 단어가 논문에서 높은 빈도를 나타내기 시작했다. 이는 생활 수준의 향상으로 장기간 관리가 필요하고 일상생활에 밀접한 관계가 있는 근골격계 질환에 관한 인간공학적 연구가 더욱더 개인의 건강에 초점을 맞춰 수행되고 있음을 유추해볼 수 있다. 또한, 2019년 말 발생한 코로나 사태 등으로 증가한 건강에 관한 관심이 ‘인간공학’ 연구에 반영된 것으로 볼 수 있다<sup>26-28)</sup>.

### 5.3 단어 간 의미적 유사도

Word2vec 모델을 이용한 워드임베딩 방식을 통해 살펴본 ‘인간공학’ 단어와 가장 높은 관련성을 나타내는 단어는 ‘작업장’이었다. 작업장은 인간공학 연구와 관련된 가장 기본적인 환경 요인이다. 환자를 대상으로 하는 임상 연구들이 병원이라는 환경에서 이루어진다면 인간공학 분야에서 근로자를 대상으로 하는 연구들은 작업장이라는 환경에서 이루어지는 것이다. 다음으로 ‘디자인’, ‘공학’을 통해 인간공학의 의학적 분야 연구에서 인간의 신체적 특성에 적합한 디자인과 공학의 요소가 기본적으로 적용되고 있다는 사실을 유추해 볼 수 있다. 즉 이와 함께, ‘작업’, ‘안전’, ‘건강’, ‘교육’ 등의 단어들을 통하여 인간공학 연구들이 근로자의 신체적 건강을 위한 안전 지식과 대처능력 습득, 작업의 잠재적 위험 요소 제거 등과 관련된 교육에 관련성이 높다는 것을 추측해 볼 수 있다.

‘일’은 ‘작업장’, ‘근로자’, ‘작업’ 등이 관련성이 높았으며, 생산 활동을 위해 사용되는 ‘근골격계’와 관련된 ‘활동’ 단어들이 높은 연관성을 나타냈다. 특히 의학 관련 인간공학 연구에서는 인체의 ‘근골격계’를 사용하는 ‘일’에 대한 면밀한 관찰과 분석이 주를 이루었을 것이다. ‘근골격계’는 인간이 움직임을 만들어내어 ‘일’을 수행하기 위한 가장 기본적이고 필수적인 시스템으로 이와 관련된 연구들이 활발히 이루어졌을 것이다. ‘과제’는 ‘모의’, ‘수행력’, ‘활동’, ‘실험’ 등의 단어들과 관련성이 높은 것으로 나타났는데, 이는 인간공학 연구들이 모의과제를 통해 실제 작업이나 활동을 통해 근골격계와 같은 신체적 특성에 미치는 영향을 평가하는 방법으로 수행되었다는 사실을 말해준다.

‘시스템’과 관련성이 높은 단어에는 ‘기술’, ‘장비’, ‘과정’과 같은 단어들이 높은 관련성을 나타냈고, 이를 통해 기술이나 장비의 개선 혹은 작업 과정에 대한 분석 등으로 인간의 신체적 특성에 적합한 시스템으로 개선하는 방안의 인간공학적 연구들이 수행되었음을 알 수 있다. 또한, ‘디자인’과 관련성이 높은 단어에는 ‘평가’, ‘사용자’, ‘발전’, ‘새로운’, ‘연구’ 등이 있다. 이 단어들을 통해 새로운 연구를 바탕으로 사용자의 평가를 통한 디자인의 발전 방안을 인간공학 분야에서 연구했음을 알 수 있다.

‘위험’ 단어와 관련성이 높은 단어들을 바탕으로 ‘근골격계’에 대한 ‘손상’ 혹은 이로 인한 ‘장애’에 대한 위험성을 검증하고 분석하는 연구들이 수행되었다고 볼 수 있다.

### 5.4 t-SNE 및 그룹화

t-SNE를 이용하여 워드임베딩 방법으로 벡터화된 데이터를 2차원으로 시각화하면 다음과 같은 형태의 모습이 나타난다(Fig. 3). 이렇게 시각화된 데이터들 가운데 ‘인간공학’과 높은 사용 빈도를 나타낸 상위 10개의 단어를 키워드로 살펴보았을 때, 본 연구에서는 이들의 관계를 4개의 그룹으로 나누어 볼 수가 있었다.

A그룹은 ‘인간공학’을 중심으로 ‘사용’, ‘디자인’, ‘시스템’의 키워드가 가까이 위치하며 이들과 거리가 가까운 단어들이 그룹을 형성하고 있었다. 인간공학에서 다루는 인간공학적 디자인이 도구의 사용, 기계의 조작, 시스템 제어 등을 수행할 때 인간의 형태적, 생리적 특성을 고려하여 도구를 사용하기 쉽게 하거나 작업자의 근골격계에 부담을 줄이는 것을 목적으로 한다는 점을 바탕으로 이들 단어가 인간공학 연구에서 가까운 거리에 위치하여 높은 연관성을 나타내는 것으로 볼 수 있다.



B그룹은 ‘효과’, ‘요인’, ‘위험’ 키워드들과 거리가 가까운 단어들로 구성된다. 이는 인간공학적 관점에서 신체에 대한 위험성을 가지고 있는 요인들에 대한 분석을 바탕으로 인간공학적 개선을 통해 효과를 입증한 연구들이 수행되었기 때문으로 볼 수 있다.

C그룹은 ‘일’, ‘근로자’ 키워드들과 거리가 가까운 단어들로 구성된다. 이를 통해 근로자들의 특성이나 고용 형태에 따른 작업 형태 등이 연구에서 다루어진 것으로 볼 수 있다.

D그룹은 ‘과제’, ‘시간’ 키워드들과 거리가 가까운 단어들로 구성된다. 이를 통해 실제 환경과 유사한 과제를 수행하는 데 있어 걸리는 시간이나 효율적인 작업 방식을 조사하고 이를 분석하여 인간공학적으로 개선하는 연구들이 수행되었을 것으로 볼 수 있을 것이다.

## 6. 한계점

이 연구의 한계점은 다음과 같다. 첫 번째, 의학 서지 데이터베이스인 Pubmed의 MEDLINE에서 ‘인간공학’을 검색어로 하는 논문을 분석하였기에 인간공학 분야의 모든 연구를 대표한다고 보긴 어려우며 인간공학 분야 중 의학 영역과 관련이 깊은 연구들을 대상으로 수행하였다. 둘째, 본 연구에 사용된 논문들의 복합 명사들은 각각의 단어로 분리되어 처리되었기 때문에 이들에 대한 의미 해석에 있어 한계점을 가지고 있다. 셋째, 해당 모델의 성능을 최적화하기 위한 하이퍼 파라미터의 조율은 수행되지 않아 모델 성능에 관한 평가가 이루어지지 않았다. 넷째, 2021년 3월까지 의학 서지 데이터베이스에 등록된 논문들의 초록 데이터를 기반으로 단어 데이터들의 관계를 순서나 시간과 관계 없이 모델링 하였다. 그러므로 시간의 흐름에 따른 주요 단어들의 관계성 변화나 동향 분석은 부족하다.

## 7. 결론

본 연구는 의학 서지 데이터베이스에 ‘인간공학’을 검색어로 등록된 연구들의 초록 데이터의 단어들을 수집하여 자연어 처리 기법의 하나인 워드임베딩 방법을 통해 단어들의 관계성을 모델링하고 이를 분석한 첫 연구이다. 키워드 분석을 통해 인간공학 연구들은 근로자가 도구를 사용하여 작업이나 일을 수행하는 것과 관련된 다양한 위험 요인들을 자료화하고 이에 대한 분석을 수행해 왔음을 파악할 수 있었다. 그리고, 워드 임베딩을 통해 ‘인간공학’ 연구에서 주로 많이 사용되는 단어 간의 관계성을 파악할 수 있었으며 이들의 관

계성을 통하여 수행되었던 연구들에 관한 분석이 가능하였다. 다차원으로 구성된 데이터들을 2차원으로 축소한 결과 키워드를 4개 그룹으로 구분할 수 있었으며, 그룹에 따른 연구 동향들을 파악해볼 수 있었다. 이와 같은 결과들을 통하여 이전에 진행되었던 ‘인간공학’ 연구들에 대한 동향 분석이 가능하였으며 앞으로 수행될 ‘인간공학’ 연구들의 주제가 더욱 다양화될 수 있을 것이다.

## References

- 1) F. W. Taylor, “Scientific Management, Routledge; 2004.
- 2) B. Price, “Frank and Lillian Gilbreth and the Manufacture and Marketing of Motion Study, 1908-1924”, Bus. Econ. Hist., Published online, pp. 88-98, 1989.
- 3) W. F. Moroney and C. M. Adams, “Placement Opportunities for Human Factors Engineering and Ergonomics Professionals: Part I: Industry, Government/Military and Consulting Positions”, Proc. Hum. Factors. Ergon. Soc. Annu. Meet., Vol. 40, pp. 436-440, 1996.
- 4) M. Zare, M. Croq, F. Hossein-Arabi, R. Brunet and Y. Roquelaure, “Does ergonomics improve product quality and reduce costs? A review article”, Hum. Factors. Ergon. Manuf. Serv. Ind., Vol. 26, No. 2, pp. 205-223, 2016.
- 5) H. W. Hendrick, “Ergonomics in Organizational Design and Management”, Ergonomics, Vol. 34, No. 6, pp. 743-756, 1991.
- 6) National Institute for Occupational Safety and Health, “Participatory Ergonomic Interventions in Meatpacking Plants”, Published online, 1994.
- 7) M. K. Chung, M. H. Yun, J. H. Park, I. S. Lee and J. H. Lim, “40 Years of Ergonomics in Korea: Accomplishments, Challenges and 40 More Years Ahead”, J. Korean. Inst. Ind. Eng., Vol. 40, No. 6, pp. 568-579, 2014.
- 8) D. K. Lee and J. H. Kim, “A Study on the Prevention System of Musculoskeletal Disorders in Korea and Other Countries”, J. Ergon. Soc. Korea, Vol. 29, No. 4, pp. 423-433, 2010.
- 9) B. Y. Jeong, “Ergonomics' Role for Preventing Musculoskeletal Disorders”, J. Ergon. Soc. Korea, Vol. 29, No. 4, pp. 393-404, 2010.
- 10) Y. C. Kim and S. R. Chang, “The Effect and Management of the Ergonomics Program for the Prevention of WMSDs in a Shipbuilding Industry”, J. Korean Soc. Saf., Vol. 21, No. 6, pp. 101-105, 2006.
- 11) D. C. Bae and S. R. Chang, “Ergonomic Evaluation

- and Improvement of the Manufacturing Lines of Compressors”, *J. Korean Soc. Saf.*, Vol. 20, No. 1, pp. 148-152, 2005.
- 12) K. S. Kim, J. K. Park and D. S. Kim, “Status and characteristics of occurrence of work-related musculoskeletal disorders”, *J. Erg. Soc. Korea*, Vol. 29, No. 4, pp. 405-422, 2010.
  - 13) C. H. Kim, M. K. Moon, J. E. Lee and Y. H. Jung, “A Study of Musculoskeletal Disorders in a Chemical Industry”, *Proc. 2007. Spring. Conf. Ergon. Soc. Korea*, Published Online, pp. 151-155, 2007.
  - 14) J. Y. Kim, S. Y. Yoon, S. W. Yoo, S. J. Lee, W. Y. Kim, K. Y. Seon and Y. S. Kim, “Cognitive Approach to Develop A Safety Management System to Control Human Error (SMaSHE) among Construction Workers”, *Proc. 2008. Spring. Conf. Ergon. Soc. Korea.*, Published online, pp. 2-7, 2008.
  - 15) D. W. Han, “A Psychological Approach to the Safety Problems in Korean Society”, *Korean. Psychol. J. Cult. Soc. Issues.*, Vol. 9, pp. 35-55, 2003.
  - 16) K. U. Kyung and J. S. Park, “The state of the art and R&D Perceptives on Haptics”, *Electron. Telecommun. trends.*, Vol. 21, No. 5, pp. 93-108, 2006.
  - 17) S. H. Choi, J. S. Seol and S. G. Lee, “On Word Embedding Models and Parameters Optimized for Korean”, *Korean Lang. Inf. Sci. Soc.*, Published Online, pp. 252-256, 2016.
  - 18) S. G. Cho and S. B. Kim, “Finding Meaningful Pattern of Key Words in IIE Transactions Using Text Mining”, *J. Korean. Inst. Ind. Eng.*, Vol. 38, No. 1, pp. 67-73, 2012.
  - 19) C. W. Jeong and J. J. Kim, “Analysis of Trend in Construction using Textmining Method”, *J. Korean Digit. Archit. Inter. Assoc.*, Vol. 12, No. 2, pp. 53-60, 2012.
  - 20) S. H. Kim, Y. J. Lee, J. Y. Shin and K. Y. Park, “Text Mining for Economic Analysis”, *J. Korean. Econ. Anal.*, Vol. 26, No. 1, pp. 1-85, 2020.
  - 21) T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, “Distributed Representations of Words and Phrases and Their Compositionality”, *Neural. Inf. Process. Syst.*, Published Online, pp. 3111-3119, 2013.
  - 22) J. U. Heu, “Korean Language Clustering using Word2Vec”, *J. Inst. Internet. Broadcast. Commun.*, Vol. 18, No. 5, pp. 25-30, 2018.
  - 23) L. van der Maaten and G. Hinton, “Visualizing data using t-SNE. *journal of Machine Learning Research* 9”, Published online, 2008.
  - 24) Google Code Archive, “<https://code.google.com/archive/p/word2vec/>”, Retrieved on 03.30.2021.
  - 25) D. C. Bae and S. R. Chang, J. H. Jung and S. E. Jin, “Ergonomic Evaluation of a Forearm Supporter for a Mouse”, *J. Korean Soc. Saf.*, Vol. 20, No. 2, pp. 169-174, 2005.
  - 26) N. Panjaitan, A. Y. B. Ali and H. A. Samat, “Ergonomic Research Trends in the Health”, *IOP. Conf. Series. Mater. Sci. Eng.*, Vol. 1003, 2020.
  - 27) F. Sasangohar, J. Moats, R. Mehta, S. C. Peres, “Disaster Ergonomics: Human Factors in COVID-19 Pandemic Emergency Management”, *Hum. Factors*, Vol. 62, No. 7, pp. 1061-1068, 2020.
  - 28) P. Carayon and S. Perry, “Human Factors and Ergonomics Systems Approach to the COVID-19 Healthcare Crisis”, *Int. J. Qual. Heal. Care.*, Vol. 33, No. Supplement 1, pp. 1-3, 2021.