



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0092937
(43) 공개일자 2024년06월24일

(51) 국제특허분류(Int. Cl.)

G06Q 50/10 (2012.01) G06F 16/93 (2019.01)
G06F 16/951 (2019.01) G06F 17/18 (2006.01)
G06F 18/23 (2023.01) G06F 18/27 (2023.01)
G06F 40/279 (2020.01) G06N 20/00 (2019.01)
G06Q 10/06 (2012.01)

(52) CPC특허분류

G06Q 50/10 (2015.01)
G06F 16/93 (2019.01)

(21) 출원번호 10-2022-0175758

(22) 출원일자 2022년12월15일

심사청구일자 2022년12월15일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대학교)

(72) 발명자

한승우

인천광역시 연수구 신송로82번길 6, 402동 803호
(송도동, 송도풍림아이원4단지아파트)

김유진

경기도 용인시 수지구 수풍로 47, 102동 1803호(
풍덕천동, 동문아파트)

(뒷면에 계속)

(74) 대리인

특허법인태백

전체 청구항 수 : 총 10 항

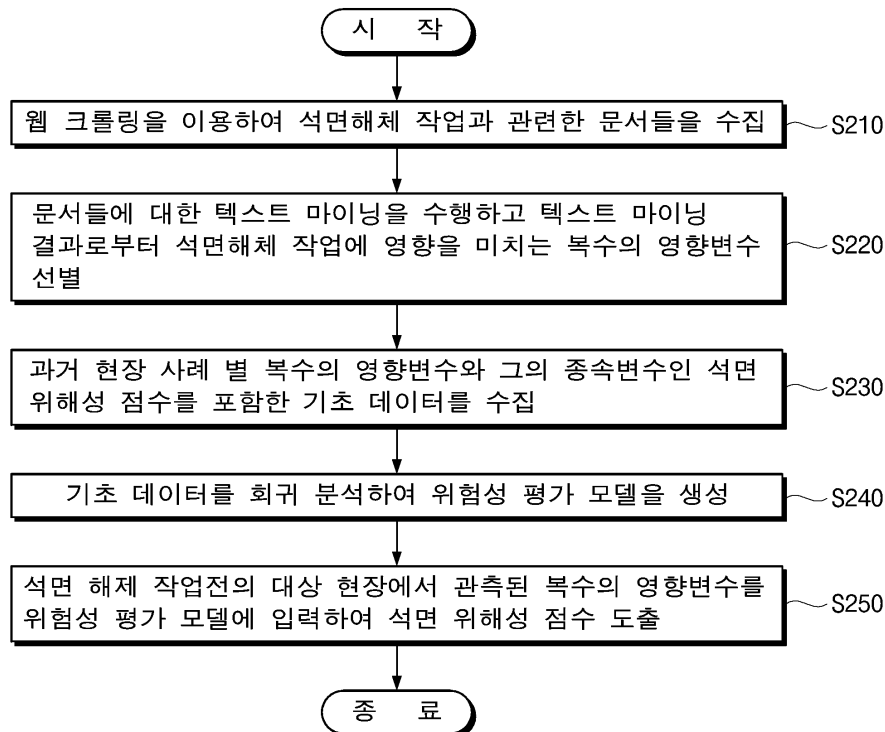
(54) 발명의 명칭 텍스트마이닝을 활용한 석면 해체 및 제거작업의 위험성 평가 장치 및 그 방법

(57) 요약

본 발명은 텍스트마이닝을 활용한 석면 해체 및 제거작업의 위험성 평가 장치 및 그 방법에 관한 것이다. 본 발명에 따르면, 석면 해체 및 제거작업의 위험성 평가 장치에 의해 수행되는 위험성 평가 방법에 있어서, 웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집하는 단계; 상기 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과로부터 석면해체 작업에 영향을 미치는 복수의 영향변수 선별

(뒷면에 계속)

대표도 - 도2



스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별하는 단계; 및 과거의 현장 사례 별로 현장에서 기 관측된 상기 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성하는 단계를 포함한다.

본 발명에 따르면, 석면 해체작업 관련 문서들에 대한 텍스트 마이닝 결과로부터 도출한 영향 변수들로 구성된 위험성 평가 모델을 활용하여 석면 해체작업 전에 위험성 평가가 가능하므로, 기존의 작업자의 안전이 고려되지 않은 위해성 평가 방식에서 벗어나 정량화된 위험성 평가에 도움을 줄 수 있다.

(52) CPC특허분류

G06F 16/951 (2019.01)

G06F 17/18 (2013.01)

G06F 18/23 (2023.01)

G06F 18/27 (2023.01)

G06F 40/279 (2020.01)

G06N 20/00 (2021.08)

G06Q 10/0635 (2023.01)

(72) 발명자

노재윤

경기도 안산시 상록구 감골로 83, 602동 1002호(사동, 신우아파트)

김동규

서울특별시 강서구 화곡로13길 107, 112동 201호(화곡동, 화곡푸르지오)

김민승

경기도 평택시 만세로 1738-25(죽백동, 나라소리국악기제작소)

이호현

인천광역시 미추홀구 인하로120번길 3, 403호(용현동, 주영주택)

이재우

서울특별시 용산구 이촌로87길 14, 105동 202호(이촌동, 강촌아파트)

김도현

인천광역시 미추홀구 인하로105번길 16, 405호(용현동, 동림주택)

조형준

대전광역시 유성구 은구비남로 34, 805동 802호(노은동, 열매마을아파트 8단지 새미래)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711163251

과제번호 2021R1A2C1007467

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 개인기초연구(과기정통부)

연구과제명 석면해체작업을 위한 빅데이터 기반 친환경 및 최적화 공정프로세스 모델 개발

기 여 율 1/1

과제수행기관명 인하대학교

연구기간 2022.03.01 ~ 2023.02.28

명세서

청구범위

청구항 1

석면 해체 및 제거작업의 위험성 평가 장치에 의해 수행되는 위험성 평가 방법에 있어서,

웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집하는 단계;

상기 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별하는 단계; 및

과거의 현장 사례 별로 현장에서 기 관측된 상기 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성하는 단계를 포함하는 석면 해체 및 제거작업의 위험성 평가 방법.

청구항 2

청구항 1에 있어서,

상기 문서들은 논문, 보고서, 기사, 법령자료 중 적어도 하나를 포함하는 석면 해체 및 제거작업의 위험성 평가 방법.

청구항 3

청구항 1에 있어서,

상기 복수의 영향변수를 선별하는 단계는,

상기 문서들을 텍스트 마이닝하여 추출한 텍스트 데이터들의 빈도 분석을 통하여 빈도수가 높은 순으로 정렬 후 상위 순위에 해당하는 복수의 중심 단어를 선별하는 단계;

N-gram 기법을 통해 획득한 각 중심 단어와 그 주변 단어의 셋과, 워드투백의 Skip-gram을 이용하여 획득한 각 중심 단어와 그 주변 단어의 셋을 조합하여, 각 중심 단어 별로 상기 중심 단어와 유사성이 높은 상위 복수 개의 주변 단어를 영향 단어 군집으로 추출하는 단계; 및

석면위해성평가자료에 존재한 복수의 평가항목 중에서 상기 추출된 영향 단어 군집과 연관된 평가항목들을 상기 복수의 영향변수로 선별하는 단계를 포함하는 석면 해체 및 제거작업의 위험성 평가 방법.

청구항 4

청구항 1에 있어서,

모델 학습 이전에 상기 과거 사례별 상기 복수의 영향변수와 상기 위해성 점수를 정량화하여 전처리하는 단계를 더 포함하는 석면 해체 및 제거작업의 위험성 평가 방법.

청구항 5

청구항 1에 있어서,

석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 상기 위험성 평가 모델에 입력하여 상기 대상 현장에 대한 석면 위해성 점수를 예측 결과로 도출하는 단계를 더 포함하는 석면 해체 및 제거작업의 위험성 평가 방법.

청구항 6

텍스트마이닝을 활용한 석면 해체 및 제거작업의 위험성 평가 장치에 있어서,

웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집하는 문서 수집부;

상기 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별하는 영향변수 선별부; 및

과거의 현장 사례 별로 현장에서 기 관측된 상기 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성하는 모델 생성부를 포함하는 석면 해체 및 제거작업의 위험성 평가 장치.

청구항 7

청구항 6에 있어서,

상기 문서들은 논문, 보고서, 기사, 법령자료 중 적어도 하나를 포함하는 석면 해체 및 제거작업의 위험성 평가 장치.

청구항 8

청구항 6에 있어서,

상기 영향변수 선별부는,

상기 문서들을 텍스트 마이닝하여 수집된 텍스트 데이터들의 빈도 분석을 통하여 빈도수가 높은 순으로 정렬 후 상위 순위에 해당하는 복수의 중심 단어를 선별하고,

N-gram 기법을 통해 획득한 각 중심 단어와 그 주변 단어의 셋과, 워드투백의 Skip-gram을 이용하여 획득한 각 중심 단어와 그 주변 단어의 셋을 조합하여, 각 중심 단어 별로 중심 단어와 유사성이 높은 상위 M개의 주변 단어를 영향 단어 군집으로 추출하며,

석면위해성평가자료에 존재한 복수의 평가항목 중에서 상기 추출된 영향 단어 군집과 연관된 평가항목들을 상기 복수의 영향변수로 선별하는 석면 해체 및 제거작업의 위험성 평가 장치.

청구항 9

청구항 6에 있어서,

모델 학습 이전에 상기 과거 사례별 상기 복수의 영향변수와 상기 위해성 점수를 정량화하여 전처리하는 전처리부를 더 포함하는 석면 해체 및 제거작업의 위험성 평가 장치.

청구항 10

청구항 6에 있어서,

석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 상기 위험성 평가 모델에 입력하여 상기 대상 현장에 대한 석면 위해성 점수를 예측 결과로 도출하는 예측부를 더 포함하는 석면 해체 및 제거작업의 위험성 평가 장치.

발명의 설명

기술 분야

[0001] 본 발명은 텍스트마이닝 기법을 활용한 석면 해체 및 제거작업의 위험성 평가 장치 및 그 방법에 관한 것으로서, 보다 상세하게는 석면해체작업과 관련된 문서들에 대한 텍스트 마이닝 결과로부터 도출한 영향 변수들로 구성된 위험성 평가 모델을 기반으로 석면 해체작업 전에 위해성 평가를 진행할 수 있도록 하는 석면 해체 및 제거작업의 위험성 평가 장치 및 그 방법에 관한 것이다.

배경 기술

[0002] 석면해체 및 제거작업은 석면의 위해성으로 인하여 작업 환경이 일반 건축물해체보다 작업자의 건강에 악영향을 끼치며, 이에 따라 해체공사비용도 상승하는 것으로 나타난다. 특히 석면이 포함된 내장재의 경우 보호뿐만 아니라 내부 밀폐시스템 조건하에 작업을 실시함에 따라 작업효율이 급격하게 낮아지는 것으로 나타난다.

[0003] 석면 해체 과정에서 발생하는 비산먼지 및 석면 농도는 해체 제거 작업 6시간 후에 감소하지만, 이후에도 석면

에 충분히 노출될 가능성이 있다. 또한, 슬레이트 제거사업의 경우 여러 기관에서 다발적으로 진행되며, 프로세스의 통일성이 없어 효율성이 떨어진다. 따라서 석면 해체공사 이전에 작업자를 위한 공정상의 위험성 평가가 필요하다.

[0004] 아울러, 기존의 석면해체 및 제거작업 표준품셈의 경우 해체공사에서 나타나는 다양한 공종 및 재료 항목을 반영하지 못하고, 해당 항목에 기초하여 모델링 프로세스 및 해체공사 품셈으로 한정되었다.

[0005] 또한 현재 의무적으로 시행되는 석면해체공사의 감리 중, '석면위해성평가'의 경우, 단순히 해당 건축물의 석면해체가 필요한지, 불필요한지의 여부를 판단하는 지표로서, 작업자의 안전은 고려되지 않는 상황이다. 이에 따라 작업자의 안전 혹은 위해도를 평가할 수 있는 모델 작성이 필요하다.

[0006] 본 발명의 배경이 되는 기술은 한국등록특허 제10-2207287호(2021.01.19 공고)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 석면해체작업과 관련된 문서들에 대한 텍스트 마이닝 결과로부터 도출한 영향 변수들로 구성된 위험성 평가 모델을 기반으로 석면 해체작업 전에 위해성 평가를 진행할 수 있도록 하는 텍스트마이닝 기법을 활용한 석면 해체 및 제거작업의 위험성 평가 장치 및 그 방법을 제공하는데 목적이 있다.

과제의 해결 수단

[0008] 본 발명은, 석면 해체 및 제거작업의 위험성 평가 장치에 의해 수행되는 위험성 평가 방법에 있어서, 웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집하는 단계; 상기 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별하는 단계; 및 과거의 현장 사례 별로 현장에서 기 관측된 상기 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성하는 단계를 포함한다.

[0009] 또한, 상기 문서들은 논문, 보고서, 기사, 법령자료 중 적어도 하나를 포함할 수 있다.

[0010] 또한, 상기 복수의 영향변수를 선별하는 단계는, 상기 문서들을 텍스트 마이닝하여 추출한 텍스트 데이터들의 빈도 분석을 통하여 빈도수가 높은 순으로 정렬 후 상위 순위에 해당하는 복수의 중심 단어를 선별하는 단계; N-gram 기법을 통해 획득한 각 중심 단어와 그 주변 단어의 셋과, 워드투백의 Skip-gram을 이용하여 획득한 각 중심 단어와 그 주변 단어의 셋을 조합하여, 각 중심 단어 별로 상기 중심 단어와 유사성이 높은 상위 복수 개의 주변 단어를 영향 단어 군집으로 추출하는 단계; 및 석면위해성평가자료에 존재한 복수의 평가항목 중에서 상기 추출된 영향 단어 군집과 연관된 평가항목들을 상기 복수의 영향변수로 선별하는 단계를 포함한다.

[0011] 또한, 상기 위험성 평가 방법은, 모델 학습 이전에 상기 과거 사례별 상기 복수의 영향변수와 상기 위해성 점수를 정량화하여 전처리하는 단계를 더 포함할 수 있다.

[0012] 또한, 상기 위험성 평가 방법은, 석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 상기 위험성 평가 모델에 입력하여 상기 대상 현장에 대한 석면 위해성 점수를 예측 결과로 도출하는 단계를 더 포함할 수 있다.

[0013] 그리고, 본 발명은, 텍스트마이닝을 활용한 석면 해체 및 제거작업의 위험성 평가 장치에 있어서, 웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집하는 문서 수집부; 상기 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별하는 영향변수 선별부; 및 과거의 현장 사례 별로 현장에서 기 관측된 상기 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성하는 모델 생성부를 포함한다.

[0014] 또한, 상기 위험성 평가 장치는, 모델 학습 이전에 상기 과거 사례별 상기 복수의 영향변수와 상기 위해성 점수를 정량화하여 전처리하는 전처리부를 더 포함할 수 있다.

[0015] 또한, 상기 위험성 평가 장치는, 석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 상기 위험성 평가 모델에 입력하여 상기 대상 현장에 대한 석면 위해성 점수를 예측 결과로 도출하는 예측부를 더 포함할 수

있다.

발명의 효과

[0016] 본 발명에 따르면, 석면 해체작업 관련 문서들에 대한 텍스트 마이닝 결과로부터 도출한 영향 변수들로 구성된 위험성 평가 모델을 활용하여 석면 해체작업 전에 위험성 평가가 가능하므로, 기존의 작업자의 안전이 고려되지 않은 위험성 평가 방식에서 벗어나 정량화된 위험성 평가에 도움을 줄 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 실시예에 따른 석면 해체 및 제거작업의 위험성 평가 장치의 구성을 나타낸 도면이다.

도 2는 도 1의 장치를 이용한 석면 해체 및 제거작업의 위험성 평가 방법을 설명하는 도면이다.

도 3은 도 2의 S220 단계를 설명한 도면이다.

도 4는 본 발명의 실시예에서 S221 단계의 결과를 예시한 도면이다.

도 5는 기존 문헌에 대한 N-gram 결과 예시를 나타낸 도면이다.

도 6은 본 발명의 실시예에서 중심 단어 대상으로 영향 변수 군집을 추출한 결과를 예시한 도면이다.

도 7은 본 발명의 실시예에 따라 생성한 위험성 평가 모델의 변수들과 계수를 요약한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0018] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시 예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0019] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0020] 도 1은 본 발명의 실시예에 따른 석면 해체 및 제거작업의 위험성 평가 장치의 구성을 나타낸 도면이다.

[0021] 도 1에 나타난 것과 같이, 본 발명의 실시예에 따른 석면 해체 및 제거작업의 위험성 평가 장치(100)는 문서 수집부(110), 영향변수 선별부(120), 모델 생성부(130)를 포함하며, 예측부(140), 데이터 수집부(150) 및 전처리부(160)를 더 포함할 수 있다. 여기서, 각 부(110~160)의 동작 및 각 부 간의 데이터 흐름은 제어부(미도시)에 의해 제어될 수 있다.

[0022] 문서 수집부(110)는 웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집한다. 문서 수집부(110)는 논문, 보고서, 기사 및 법령 전반에 걸쳐 석면해체작업과 관련된 기존 문헌들을 수집할 수 있다.

[0023] 영향변수 선별부(120)는 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝을 수행하고 텍스트 마이닝 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별한다.

[0024] 영향변수 선별부(120)는 텍스트 마이닝 기법을 이용하여 석면해체 작업 관련 문서들에 포함된 텍스트 데이터를 분석하고 이를 기반으로 석면해체공사에서 작업자 및 실 사용자를 대상으로 안전에 영향을 끼치는 요소들을 도출할 수 있다.

[0025] 모델 생성부(130)는 과거의 현장 사례 별로 현장에서 기 관측된 복수의 영향변수(독립변수)와 그에 대한 종속변수인 현장의 기 평가된 석면 위험성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성한다.

[0026] 이에 따라, 본 발명의 실시예는 현장에서 관측 가능한 복수의 영향 변수로부터 석면 위험성 점수를 예측해 주는 위험성 평가 모델을 구현할 수 있다.

[0027] 여기서, 모델 학습 이전에 과거 사례별 수집된 복수의 영향변수와 위험성 점수는 전처리부(160)에서 사전에 정

량화되어 전처리될 수 있다. 이에 따라, 각 변수들 중에서 수치화되지 않은 비정형 데이터, 정성적이거나 모호한 변수에 대해서는 퍼지(fuzzy) 기법 등을 통해서 전처리부(160)에서 수치화된 변수 값으로 가공하여 모델 학습이 가능한 형태로 제공할 수 있다. 이에 따라, 모델 생성부(130)는 전처리를 통해 정량화된 데이터셋을 바탕으로 회귀분석을 사용하여 위험성 평가 모델을 작성할 수 있다.

[0028] 여기서, 모델 생성부(130)는 과거 현장 사례 별 복수의 영향변수와 석면 위해성 점수를 포함한 기초 데이터를 데이터 수집부(150)로부터 제공받아 모델 학습을 진행할 수 있다. 데이터 수집부(150)는 과거의 각 사례별 건축물 석면해체 위해성평가자료를 기반으로 기초 데이터에 해당하는 데이터셋을 수집하고 이를 모델 생성부(130)로 제공할 수 있다. 석면 위해성평가자료는 해당 건축물의 석면노출 위험도를 평가한 자료로, 정부, 관계 기관 등에서 제공하는 공공데이터로부터 확보 가능하다.

[0029] 예측부(140)는 석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 기 학습된 위험성 평가 모델에 입력하여 대상 현장에 대한 석면 위해성 점수를 예측 결과로 도출하여 제공한다.

[0030] 이에 따라, 본 발명의 실시예는 석면 해체 작업전에 해당 임의 현장에 대한 각 영향변수를 사전 학습된 위험성 평가 모델에 입력하는 것만으로도 해당 현장에서 예상되는 석면 위해성 점수를 예측 결과로 출력해줄 수 있다.

[0031] 본 발명의 실시예에 따른 장치(100)는 석면 해체공사 이전에 석면 위해성 점수를 결정하는 서버 그 자체일 수도 있고, 사용자 단말(미도시) 상에 구현되는 어플리케이션에 해당할 수도 있다. 사용자 단말은 관련 어플리케이션(응용 프로그램)이 실행된 상태에서 본 발명의 장치(100)와 네트워크 접속되어 분석 대상이 되는 석면해체 작업전의 현장에 대한 변수 데이터 입력에 대응하여 예측되는 해당 현장에 대한 석면 위해성 점수를 화면을 통해 출력하여 제공할 수 있다.

[0032] 도 2는 도 1의 장치를 이용한 석면 해체 및 제거작업의 위험성 평가 방법을 설명하는 도면이고, 도 3은 도 2의 S220 단계를 설명한 도면이다.

[0033] 먼저, 문서 수집부(110)는 웹 크롤링을 이용하여 석면해체 작업과 관련한 문서들을 수집한다(S210). 이때, 수집되는 문서들은 논문, 보고서, 기사, 법령자료 등을 포함할 수 있다.

[0034] 문서 수집부(110)는 웹크롤링 코드(python)를 통해 석면 해체 및 제거작업의 텍스트 데이터를 수집할 수 있으며, python 함수 urllib.request와 BeautifulSoup를 사용하여 코드를 구현할 수 있다. 예를 들어, '석면해체'를 검색어로 하여, 논문, 보고서 등의 문헌은 RISS(Research Information Sharing Service)에서, 기사는 네이버 뉴스에서 데이터를 크롤링하여 수집할 수 있다.

[0035] 웹 크롤링을 통해 수집된 문서들은 데이터 마이닝 분석을 위하여 영향변수 선별부(120)로 전달된다.

[0036] 다음으로, 영향변수 선별부(120)는 석면해체 작업과 관련한 문서들에 대한 텍스트 마이닝 분석을 수행하고 텍스트 마이닝 분석 결과를 이용하여 석면해체 작업에 영향을 미치는 복수의 영향변수를 선별한다(S220).

[0037] 이러한 S220 단계는 도 3과 같이 빈도 분석을 통한 중심 단어 선별 과정(S221), 중심 단어를 기초로 N-gram과 워드투백을 이용한 영향 단어 군집 추출 과정(S222), 그리고 영향 단어 군집을 이용한 영향변수 선별 과정(S223)을 포함할 수 있다.

[0038] 우선, 영향변수 선별부(120)는 문서 수집부(110)에서 수집된 문서들을 텍스트 마이닝 분석하여 추출한 텍스트 데이터들의 빈도 분석을 통하여 빈도수가 높은 순으로 정렬 후 상위 순위에 해당하는 복수의 중심 단어를 선별한다(S221).

[0039] 이러한 중심 단어 선별 과정은 텍스트 데이터의 빈도 분석을 수행하여 문헌들에서 중심 내용이 되는 중심단어를 선별하는 과정에 해당한다. 해당 과정에서 연구 동향 파악 및 영향요인 추출에 무의미한 단어들을 제거하는 전처리 과정을 진행할 수 있다. 이후, 단어 수를 카운트하여 빈도수가 높은 순으로 나열하여 최다 빈도 단어부터 중심단어로 선별한다.

[0040] 도 4는 본 발명의 실시예에서 S221 단계의 결과를 예시한 도면이다. 도 4는 연도 별로 분석된 워드 클라우드와 각 단어 별 출현 빈도수를 이미지 및 테이블로 도시한 것이다. 단어별 출현 빈도로부터 일부 무의미한 단어를 제거하여 전처리하고 출현 빈도가 높은 순으로 중심 단어들을 선별할 수 있다.

[0041] 다음으로, N-gram 기법을 통해 획득한 각 중심 단어와 그 주변 단어의 셋과, 워드투백의 Skip-gram을 이용하여 획득한 각 중심 단어와 그 주변 단어의 셋을 조합하여, 각 중심 단어 별로 상기 중심 단어와 유사성이 높은 상

위 복수 개의 주변 단어를 영향 단어 군집으로 추출한다(S222).

- [0042] 본 발명의 실시예는 중심단어 대상 동시 출현 빈도 분석을 통해서 영향 요인을 도출하는 과정을 거치는데, 이를 통해 영향 요인의 관계성 및 문맥을 파악한다.
- [0043] 도 5는 기존 문헌에 대한 N-gram 결과 예시를 나타낸 도면이다. N-gram은 문자열 등에서 n개의 연속된 (어휘, 음절 등의) 구성요소를 표현한 것이다. N-gram은 횟수만 고려한 것이 아닌 단어가 연결될 때 생기는 의미를 고려하여 연이어 사용된 n개의 단어의 단어쌍의 관계를 표현한 그림으로, 단어 각각의 의미뿐만 아니라 연결되어 생기는 의미도 파악할 수 있어 영향요인 간 연관성과 고려가 필요한 사항을 문맥적으로 파악할 수 있다. 그런데 단어의 의미와 문맥만 파악 가능하므로 워드투백을 통한 영향 요인 추출 과정을 거치게 된다.
- [0044] 도 6은 본 발명의 실시예에서 중심 단어 대상으로 영향 변수 군집을 추출한 결과를 예시한 도면이다. 도 6은 5가지의 중심 단어에 대한 영향 변수 군집 데이터를 보여주며, 진동, 평가, 빈도, 종류, 크기를 중심단어로 하고 그에 대한 상위 연관 단어 10개를 추출하여 나타낸 것이다. 중심 단어에 대해 유사성 높은 상위 10개를 군집으로 추출하며, 이때 해당 단계에서 영향요인 추출에 무의미한 조사, 접속사와 같은 단어들은 제거한다.
- [0045] 본 발명의 실시예는 각 중심 단어에 대해 N-gram을 통해 얻은 데이터 셋과 워드투백의 Skip-gram을 통해 얻은 데이터 셋을 조합하여, 각 중심 단어 별로 중심 단어와 유사성이 높은 상위 10개의 주변 단어를 영향 단어 군집으로 추출할 수 있다.
- [0046] 이러한 단계에서는 중심 단어와 관련성 높은 단어를 통하여, 위험 요인으로 고려가 필요한 사항들을 정립할 수 있다. 여기서 석면위해성 평가항목 중에서 물리적 평가, 건물 사용으로 인한 손상 가능성 평가, 건축물 유지 보수 활동에 기반한 손상 가능성 평가, 작업자의 인체노출 가능성 평가들이 위해 요인으로 고려해야 할 사항과 관련될 수 있다.
- [0047] 다음으로, 석면위해성평가자료에 존재한 복수의 평가항목 중 추출된 영향 단어 군집과 연관된 것으로 판단되는 평가항목들을 석면해체 작업에 영향을 미치는 복수의 영향변수로 선별한다(S223).
- [0048] 이때, 각 영향 단어 군집에서 실제 현장에서 수집 가능한 정보를 선정할 수 있으며, 중심 단어에 해당하는 데이터가 유효하지 않은 경우, 대체적으로 수집 가능한 정보인 유사성이 높은 단어를 변수로 취할 수 있다.
- [0049] 예를 들어, '진동'을 중심 단어로 부위, 자체, 원인, 상태 등의 단어가 출현되는데, 이 단어들은 '석면위해성평가점수' 기준 중 '진동에 의한 손상가능성' 항목의 평가기준에 명시되어 있다. 또한, '크기'를 중심단어로 입자, 유량 등의 단어가 출현되는데, 이는 '석면위해성평가점수' 기준 중 '유량'항목의 평가 기준에 명시되어 있다.
- [0050] 그리고, '평가'가 중심 단어로 선정됨에 따라 '석면위해성평가점수'항목에 포함되는 손상상태, 비산성, 진동, 유량, 누수, 유지보수형태, 유지보수빈도가 차용될 수 있다.
- [0051] 본 발명의 실시예에서 석면위해성평가자료를 데이터로 사용한 이유는 해당 자료의 석면위해성평가점수가 석면해체공사 이전에 해당 건물을 평가하는 평가 지표에 해당하고 중심 단어 및 관련단어를 평가할 수 있는 정성적 지표로 제시되기 때문이다.
- [0052] 그 외에도 석면해체작업 전에 수집 가능한 데이터는 건강진단개인표를 통한 '나이', '노출기간', 석면 해체/제거작업계획서를 통한 '음압기 사용계획', 석면농도측정결과보고서를 통한 석면 '측정농도' 등이 존재한다. 이 중에서 작업자 개개인의 석면 해체 및 제거작업의 위험성을 확인하기 위해 파악 가능한 건강진단개인표에서 '나이', '노출기간'을 사용할 수 있다.
- [0053] 본 발명의 실시예의 경우, 텍스트 마이닝 분석을 통해 도출한 영향 단어 군집과 연관되는 것으로 판단되는 항목들로, 석면위해성평가자료에 포함된 여러가지 평가항목 중에서도 '손상상태', '비산성', '진동', '유량', '누수', '유지보수형태', '유지보수빈도', '최초 노출된 나이', '노출기간'을 포함한 9가지 변수를 모델 구축을 위한 영향변수로 선정하였다.
- [0054] 여기서 물론, 선정된 영향변수들이 반드시 상술한 바로 한정되는 것은 아니며 일부 항목이 삭제될 수도 있고 석면위해성평가자료에 존재한 다른 평가항목들이 더 추가 및 보강될 수도 있다.
- [0055] 그리고, 위험성 평가 모델을 작성하기 위해, 이렇게 선정된 영향변수들에 대한 과거 현장 사례별 데이터셋을 수집하도록 한다.

[0056] 이를 위해, 데이터 수집부(150)는 과거 현장 사례 별 복수의 영향변수와 그의 종속변수인 석면 위해성 점수를 포함한 기초 데이터를 수집한다(S230). 데이터 수집부는 공공데이터로 제공되는 석면해체 위해성평가자료에서 과거 각 현장 사례별로 9가지 영향변수 데이터와 해당 현장의 석면 위해성 점수를 데이터셋으로 수집하고 이를 회귀 모델 학습에 필요한 기초 데이터로 제공할 수 있다.

[0057] 이와 같이 데이터 수집부(150)는 과거의 각 사례별 수집된 건축물 석면해체 위해성평가자료를 기반으로 기초 데이터를 수집하고 이를 모델 생성부(130)로 전달할 수 있다.

[0058] 여기서, 모델 학습 이전에 과거 사례별 수집된 복수의 영향변수와 위해성 점수는 전처리부(160)에서 정량화되어 전처리될 수 있다. 전처리부(160)는 각 변수들 중 수치화되지 않은 비정형 데이터를 모델 학습이 가능한 수치화된 변수 값으로 가공할 수 있다.

[0059] 데이터 정량화 방법으로, 예를 들어 Fuzzy C-means(FCM) Clustering Algorithm을 사용하여 명목형 데이터를 연속형으로 변환시킬 수 있다. 그 과정에서 FCM 알고리즘을 통해 초기 가중치를 설정하고, 가중치가 수렴할 때까지 fuzzy 군집화를 반복할 수 있다.

[0060] 본 발명의 실시예에서 사용되는 9가지 영향 변수는 위험성 평가 모델의 독립 변수에 해당하며 아래의 표 1과 같이 정리될 수 있다. 각 변수에 대한 값의 범위도 나타내었다.

표 1

[0061]

변수명	손상상태	비산성	진동	유량	누수
범위	0,1,2,3	0,1,2,3	0,1,2	0,1,2	0,1,2
변수명	유지보수형태	유지보수빈도	최초 노출된 나이	노출기간[년]	
범위	0,1,2,3	0,1,2,3	26~61	0~20	

[0062] 모델 학습 이전에 필요한 변수 데이터를 정량화하고 전처리하는 방법은 다양하게 존재하며 기 공지된 다양한 기법이 적용될 수 있다.

[0063] 모델 생성부(130)는 전처리를 통해 정량화된 데이터셋을 바탕으로 회귀분석을 사용하여 위험성 평가 모델을 작성할 수 있다.

[0064] 모델 생성부(130)는 과거의 현장 사례 별로 현장에서 기 관측된 복수의 영향변수와 그에 대한 종속변수인 현장의 기 평가된 석면 위해성 점수 간의 관계를 회귀 분석을 통해 학습하여 위험성 평가 모델을 생성한다(S240).

[0065] 모델 생성부(130)는 과거 현장에서의 복수의 영향변수(손상상태, 비산성, 진동, 유량, 누수, 유지보수형태, 유지보수빈도, 최초 노출된 나이, 노출기간)를 입력 값(독립 변수)으로 하고 그에 대응하여 예상되는 석면 위해성 점수를 출력 값(종속 변수)으로 하여, 복수의 독립 변수와 그 종속 변수 간의 관계를 머신 러닝 분석하여 각 독립 변수에 대한 가중치를 추정할 수 있다.

[0066] 회귀 모델은 일반적으로 아래 수학적 1의 형태를 가진다.

수학적 1

[0067]

$$Y = \omega_0 + \omega_1 X_1 + \dots + \omega_p X_p + \epsilon$$

[0068] 여기서, X_i 는 i 번째 독립변수, ω_i 는 i 번째 독립변수에 적용되는 i 번째 가중치(계수), ϵ 는 회귀 모델을 통해 예측된 위해성 점수(Y)와 각 독립변수(X_i) 간의 오차 보정 계수를 나타낸다.

[0069] 본 발명의 실시예의 경우 최종적으로 아래와 같은 회귀 모델을 생성할 수 있다.

수학식 2

$$y = -0.011(\text{비산상}) + 0.005(\text{손상상태}) - 0.004(\text{진동}) - 0.013(\text{기류}) + 0.009(\text{누수}) \\ - 0.001(\text{유지보수형태}) + 0.001(\text{유지보수빈도}) - 0.076(\text{최조노출된나이}) \\ + 0.015(\text{노출기간}) + 5.225$$

[0070]

[0071]

도 7은 본 발명의 실시예에 따른 위험성 평가 모델의 변수들과 계수를 요약한 도면이다. 수학식 2의 각 독립변수에 곱해진 계수는 도 7의 결과로부터 결정된 것에 해당한다.

[0072]

이러한 본 발명은 회귀 모델에 대한 종속변수 데이터의 정규성을 확인하고, VIF(분산팽창요인) 검증을 통해 독립변수간 독립성을 판단하는 과정을 통해 모델의 적합성을 판단할 수 있다. 그리고 도출된 회귀모델의 정확성을 판단하기 위해 수정된 결정계수($adj-R^2$)를 사용한다. 결정계수가 0.8 이상일 경우 해당 모델을 석면해체공사의 위험성 평가 모델로 지정할 수 있다.

[0073]

위험성 평가 모델이 구축된 이후부터는 분석 대상이 되는 작업장에 대해 관측되는 9가지 영향변수들을 수학식 2의 회귀 모델에 입력하는 것만으로도 해당 작업장에서 예상되는 석면 위해성 점수를 제공할 수 있다.

[0074]

즉, 예측부(140)는 석면 해체 작업전의 대상 현장에서 관측된 복수의 영향변수를 앞서 구축된 위험성 평가 모델에 입력하여 석면 위해성 점수를 예측 결과로 도출한다(S250).

[0075]

이후, 출력부(미도시)는 예측 결과를 디스플레이, 사용자 단말 등을 통해 사용자에게 제공할 수 있다.

[0076]

이와 같이, 본 발명에 따르면, 석면해체 작업의 위해영향 요소를 추출하기 위해 사용한 자료들(석면해체공사와 관련한 논문, 보고서, 기사, 법령 등)을 이용하여 객관화된 위해영향 요소를 추출할 수 있으며, 추출한 요소들을 입력변수로 하는 회귀모델을 사용하여 현장에서 상용화 가능한 간단한 수식 형태로 위험성 평가 모델을 제공할 수 있다.

[0077]

본 발명에 따르면, 작업자 개개인의 석면 해체 및 제거작업의 위험성을 확인할 수 있으며, 본 모델이 상용화됨에 따라 석면해체공사에서 정량화된 위험성 평가가 가능하다.

[0078]

또한, 본 발명은 석면해체작업과 관련된 논문, 보고서, 기사 및 법령 자료들로부터 위험성 평가 모델을 위한 영향 변수들을 추출하였기에, 석면해체 공정 혹은 제도 및 그에 따른 이슈가 발전 또는 변화하더라도 시대에 따른 안전성의 기준을 반영하기 용이한 장점을 갖는다.

[0079]

아울러, 본 발명의 기술은 기존의 작업자의 안전이 고려되지 않은 위해성 평가방식을 보완하여 추후 ESG적 경영 측면에서 해당 자료를 지속 가능한 사회적 비용 산출에 사용될 수 있다.

[0080]

본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

[0081]

100: 석면 해체 및 제거작업의 위험성 평가 장치

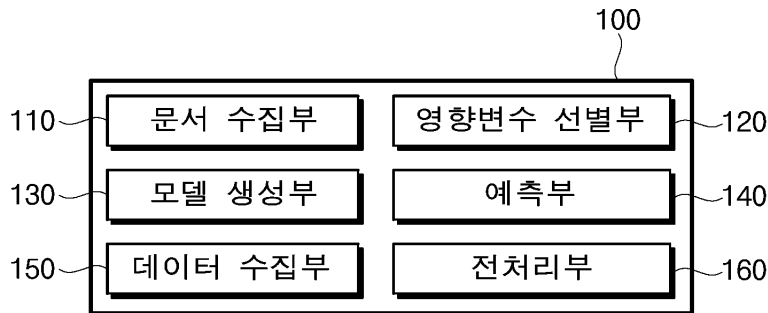
110: 문서 수집부 120: 영향변수 선별부

130: 모델 생성부 140: 예측부

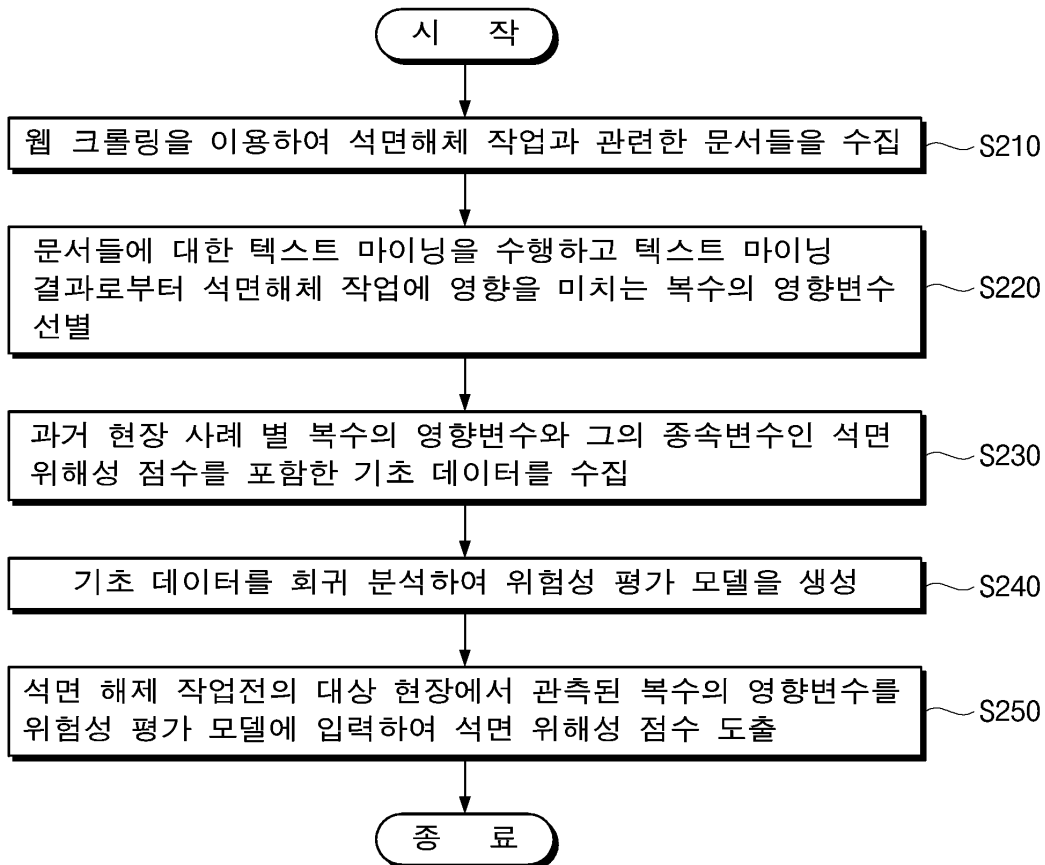
150: 데이터 수집부 160: 전처리부

도면

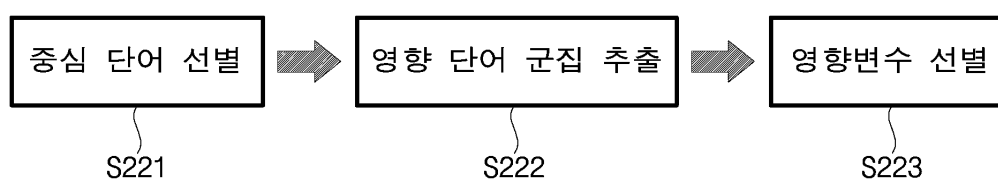
도면1



도면2



도면3



도면6

단어	분석값	단어	분석값	단어	분석값	단어	분석값	단어	분석값
진동	1.000000	평가	1.000000	빈도	1.000000	종류	1.000000	크기	1.000000
부위	0.998891	등급	0.999685	변화	0.999179	위치	0.999281	광물	0.999342
자체	0.998782	안전성	0.999037	중요도	0.998993	습윤제	0.998821	길이	0.999041
원인	0.998749	업체	0.998938	값	0.998977	물	0.998797	최대	0.998983
폴리	0.998719	받다	0.998533	성분	0.998892	도포	0.998725	입자	0.9989
요한	0.998716	미흡	0.99835	검정	0.998826	행위	0.998718	변화	0.998816
규산나트륨	0.998708	낮다	0.998121	요인	0.998745	함유량	0.998702	관찰	0.99878
방식	0.998702	시공	0.998008	최소	0.998718	시나리오	0.99866	유량	0.998658
분산	0.998683	우수	0.997981	예측	0.998643	거리	0.998641	크다	0.99861
상태	0.99868	안정	0.997959	낮다	0.998606	발열반응	0.99864	유리	0.998551
기능	0.998664	성	0.997912	유의	0.998604	아크릴	0.998534	현미경	0.998497

도면7

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	통계량 변화량				
					R 제곱 변화량	F 변화량	df1	df2	유의확률 F 변화량
1	.935 ^a	.873	.873	.295	.873	2822.017	9	3683	.000

a. 예측값: (상수), 노출기간, 누수, 손상상태, 유지보수빈도, 유량, 진동, 최초노출된나이, 유지보수형태, 비산성

계수^a

모형		비표준화 계수		표준화 계수		t	유의확률	상관계수			공선성 통계량	
		B	표준오차	베타				0차	편상관	부분상관	공차	VIF
1	(상수)	5.225	.026			200.539	.000					
	비산성	-.011	.009	-.009		-1.267	.205	-.019	-.021	-.007	.692	1.445
	손상상태	.005	.008	.004		.579	.562	.000	.010	.003	.881	1.135
	유지보수형태	-.001	.005	-.001		-.141	.888	.006	-.002	-.001	.804	1.244
	유지보수빈도	.001	.012	.000		-.017	.986	-.006	.000	.000	.853	1.173
	진동	-.004	.009	-.002		-.396	.692	.021	-.007	-.002	.887	1.128
	유량	-.013	.017	-.005		-.772	.440	.013	-.013	-.005	.919	1.089
	누수	.009	.010	.005		.863	.388	.016	.014	.005	.863	1.159
	최초노출된나이	-.076	.001	-.906		-149.362	.000	-.930	-.926	-.876	.936	1.069
	노출기간	.015	.001	.096		15.901	.000	.325	.253	.093	.935	1.069

a. 종속변수: 평가점수