



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0165515
(43) 공개일자 2022년12월15일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2022.01) G06K 9/62 (2022.01)
G06N 20/00 (2019.01) G06T 7/60 (2017.01)
(52) CPC특허분류
G06V 30/40 (2022.01)
G06K 9/6267 (2022.01)
(21) 출원번호 10-2021-0074240
(22) 출원일자 2021년06월08일
심사청구일자 2021년06월08일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
엘아이지넥스원 주식회사
경기도 용인시 기흥구 마북로 207 (마북동)
(72) 발명자
문두환
세종특별자치시 도움3로 75, 1101동 503호 (중촌동, 가재마을11단지)
이진원
서울특별시 서대문구 연희로16길 36-10, 302호
(뒷면에 계속)
(74) 대리인
특허법인우인

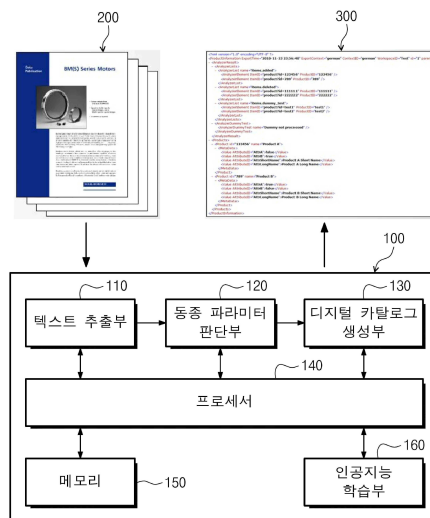
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 카탈로그 이미지의 데이터 추출 시스템 및 카탈로그 이미지의 데이터 추출 방법

(57) 요약

개시된 발명의 일 실시예에 따른 카탈로그 이미지의 데이터 추출 방법은, (a) 텍스트 추출부에 의해, 적어도 하나 이상의 부품 카탈로그 이미지로부터 파라미터 명칭들 및 각 파라미터 명칭에 대응되는 파라미터 값을 포함하는 부품 속성 정보를 추출하는 단계; (b) 동종 파라미터 판단부에 의해, 상기 추출된 파라미터 명칭들 및 표준 용어 데이터에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단하는 단계; 및 (c) 디지털 카탈로그 생성부에 의해, 상기 동종 파라미터로 판단된 파라미터들의 파라미터 값들끼리 서로 대응되도록 상기 적어도 하나 이상의 부품 카탈로그 이미지들에 대한 디지털 부품 카탈로그를 생성하는 단계;를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 20/00 (2021.08)

G06T 7/60 (2013.01)

G06V 30/10 (2022.01)

(72) 발명자

김효태

경기도 성남시 분당구 하오개로 377, 102동 203호

여창모

서울특별시 종로구 숭인동1가길 24, 202호

명세서

청구범위

청구항 1

- (a) 텍스트 추출부에 의해, 적어도 하나 이상의 부품 카탈로그 이미지로부터 파라미터 명칭들 및 각 파라미터 명칭에 대응되는 파라미터 값을 포함하는 부품 속성 정보를 추출하는 단계;
- (b) 동종 파라미터 판단부에 의해, 상기 추출된 파라미터 명칭들 및 표준 용어 데이터에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단하는 단계; 및
- (c) 디지털 카탈로그 생성부에 의해, 상기 동종 파라미터로 판단된 파라미터들의 파라미터 값들끼리 서로 대응되도록 상기 적어도 하나 이상의 부품 카탈로그 이미지들에 대한 디지털 부품 카탈로그를 생성하는 단계;를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 2

제1항에 있어서,

상기 (a) 단계는,

상기 텍스트 추출부에 의해, 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 각 파라미터의 심볼 및 각 파라미터의 단위를 포함하는 부품 속성 정보를 추출하는 단계;를 포함하고,

상기 (b) 단계는:

상기 동종 파라미터 판단부에 의해, 추출된 상기 심볼 및 상기 단위에 기초하여, 상기 추출된 파라미터들이 동종 파라미터인지 여부를 판단하는 단계;를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 3

제1항에 있어서,

상기 (a) 단계는:

(a1) 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 상기 부품 속성 정보를 포함하는 사양 테이블 영역을 인식하는 단계;

(a2) 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 추출하는 단계를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 4

제3항에 있어서,

상기 (a1) 단계는:

상기 부품 카탈로그 이미지로부터 수평선 및 수직선을 인식하는 단계; 및

상기 인식된 수평선 및 수직선에 기초하여 상기 부품 카탈로그 이미지로부터 상기 사양 테이블 영역을 인식하는 단계;를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 5

제3항에 있어서,

상기 (a) 단계는:

상기 사양 테이블 영역으로부터 테이블 수평선 및 테이블 수직선을 인식하는 단계;를 더 포함하고,

상기 (a2) 단계는:

상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 인식하는 단계;를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 6

제5항에 있어서,

인공지능 학습부에 의해, 상기 테이블 수평선 및 상기 테이블 수직선을 입력 변수로 설정하고 상기 부품 속성 정보를 출력 변수로 설정하여 인공지능 모델을 학습하는 단계;를 더 포함하고,

상기 (a2) 단계는,

상기 인식된 테이블 수평선 및 테이블 수직선을 기초로 상기 인공지능 모델을 이용하여 상기 부품 속성 정보를 인식하는 단계;를 포함하는, 카탈로그 이미지의 데이터 추출 방법.

청구항 7

제5항에 있어서,

상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 사양 테이블 헤더를 추출하는 단계;를 더 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 8

제1항에 있어서,

텍스트 추출부에 의해, 상기 부품 카탈로그 이미지로부터 부품 번호를 추출하는 단계;

상기 부품 번호 및 부품 번호 규칙 데이터에 기초하여 상기 적어도 하나 이상의 부품 카탈로그 이미지들이 동종 부품의 카탈로그 이미지인지 여부를 판단하는 단계; 및

상기 적어도 하나 이상의 부품 카탈로그 이미지들을 상기 동종 부품의 카탈로그 이미지끼리 분류하는 단계;를 더 포함하고,

상기 (a) 단계는,

상기 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지로부터 부품 속성 정보를 추출하는 단계;를 더 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 9

제8항에 있어서,

상기 부품 번호를 추출하는 단계는,

상기 텍스트 추출부에 의해, 상기 부품 카탈로그 이미지의 상단의 미리 설정된 영역에서 텍스트를 추출하여 상기 부품 번호를 추출하는 단계;를 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 10

제1항에 있어서,

상기 추출된 부품 속성 정보의 텍스트를 상기 표준 용어 데이터와 비교하여 상기 부품 속성 정보의 텍스트 오류 여부를 판단하는 단계; 및

상기 부품 속성 정보의 텍스트에 오류가 있는 것으로 판단되면, 상기 표준 용어 데이터에 기초하여 상기 부품 속성 정보를 수정하는 단계;를 더 포함하는 카탈로그 이미지의 데이터 추출 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 카탈로그 이미지의 데이터 추출 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

청구항 12

적어도 하나 이상의 부품 카탈로그 이미지로부터 파라미터 명칭들 및 각 파라미터 명칭에 대응되는 파라미터 값을 포함하는 부품 속성 정보를 추출하도록 구성되는 텍스트 추출부;

상기 추출된 파라미터 명칭들 및 표준 용어 데이터에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단하도록 구성되는 동종 파라미터 판단부; 및

상기 동종 파라미터로 판단된 파라미터들의 파라미터 값들끼리 서로 대응되도록 상기 적어도 하나 이상의 부품 카탈로그 이미지들에 대한 디지털 부품 카탈로그를 생성하도록 구성되는 디지털 카탈로그 생성부;를 포함하는 카탈로그 이미지의 데이터 추출 시스템.

청구항 13

제12항에 있어서,

상기 텍스트 추출부는,

상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 각 파라미터의 심볼 및 각 파라미터의 단위를 포함하는 부품 속성 정보를 추출하도록 구성되고,

상기 동종 파라미터 판단부는,

추출된 상기 심볼 및 상기 단위에 기초하여, 상기 추출된 파라미터들이 동종 파라미터인지 여부를 판단하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 14

제12항에 있어서,

상기 텍스트 추출부는:

상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 상기 부품 속성 정보를 포함하는 사양 테이블 영역을 인식하고; 그리고

상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 추출하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 15

제14항에 있어서,

상기 텍스트 추출부는:

상기 부품 카탈로그 이미지로부터 수평선 및 수직선을 인식하고; 그리고

상기 인식된 수평선 및 수직선에 기초하여 상기 부품 카탈로그 이미지로부터 상기 사양 테이블 영역을 인식하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 16

제14항에 있어서,

상기 텍스트 추출부는:

상기 사양 테이블 영역으로부터 테이블 수평선 및 테이블 수직선을 인식하고; 그리고

상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 인식하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 17

제16항에 있어서,

상기 테이블 수평선 및 상기 테이블 수직선을 입력 변수로 설정하고 상기 부품 속성 정보를 출력 변수로 설정하

여 인공지능 모델을 학습하도록 구성되는 인공지능 학습부;를 더 포함하고,

상기 텍스트 추출부는,

상기 인식된 테이블 수평선 및 테이블 수직선을 기초로 상기 인공지능 모델을 이용하여 상기 부품 속성 정보를 인식하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 18

제12항에 있어서,

프로세서;를 더 포함하고,

상기 프로세서는:

부품 번호 및 부품 번호 규칙 데이터에 기초하여 상기 적어도 하나 이상의 부품 카탈로그 이미지들이 동종 부품의 카탈로그 이미지인지 여부를 판단하고; 그리고

상기 적어도 하나 이상의 부품 카탈로그 이미지들을 상기 동종 부품의 카탈로그 이미지끼리 분류하도록 구성되고,

상기 텍스트 추출부는:

상기 부품 카탈로그 이미지로부터 상기 부품 번호를 추출하도록 구성되고; 그리고

상기 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지로부터 부품 속성 정보를 추출하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 19

제18항에 있어서,

상기 텍스트 추출부는,

상기 부품 카탈로그 이미지의 상단의 미리 설정된 영역에서 텍스트를 추출하여 상기 부품 번호를 추출하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

청구항 20

제12항에 있어서,

프로세서;를 더 포함하고,

상기 프로세서는:

상기 추출된 부품 속성 정보의 텍스트를 상기 표준 용어 데이터와 비교하여 상기 부품 속성 정보의 텍스트 오류 여부를 판단하고; 그리고

상기 부품 속성 정보의 텍스트에 오류가 있는 것으로 판단되면, 상기 표준 용어 데이터에 기초하여 상기 부품 속성 정보를 수정하도록 구성되는, 카탈로그 이미지의 데이터 추출 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 엔지니어링 시트가 포함된 이미지 형식의 부품 카탈로그로부터 부품의 데이터를 추출하여 디지털 형식의 부품 카탈로그를 생성할 수 있는 추출 시스템 및 데이터 추출 방법에 관한 것이다.

배경 기술

[0002] 제품의 설계 과정에서 구매할 부품의 경우 부품 제작사가 제공한 부품 카탈로그로부터 고객의 요구에 가장 잘 맞는 부품을 탐색해야 한다.

[0003] 따라서, 적절한 부품을 자동으로 탐색하는 기술이 필요하지만, 현재 대부분의 부품 카탈로그는 스캐닝된 이미지

형식으로 보관되고 있고, 이러한 이미지 형식의 부품 카탈로그를 통해서도 부품간 사양 비교 및 부품의 검색에 많은 제약이 있다는 문제가 있다.

[0004] 따라서 이미지 형식으로 되어 부품의 정보가 혼재되어 있는 복수개의 부품 카탈로그 이미지를 통합하여 하나의 디지털 형태의 카탈로그를 생성하는 기술이 필요하다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 이미지 형식의 부품 카탈로그를 딥러닝 기술을 통해 인식하여 카탈로그에 포함된 부품 사양 데이터를 자동으로 추출하고, 추출된 데이터를 기초로 디지털 형식의 부품 카탈로그를 생성할 수 있는 데이터 추출 시스템, 데이터 추출 방법 및 컴퓨터 프로그램을 제공하기 위한 것이다.

과제의 해결 수단

[0006] 개시된 발명의 일 측면에 따른 카탈로그 이미지의 데이터 추출 방법은, (a) 텍스트 추출부에 의해, 적어도 하나 이상의 부품 카탈로그 이미지로부터 파라미터 명칭들 및 각 파라미터 명칭에 대응되는 파라미터 값을 포함하는 부품 속성 정보를 추출하는 단계; (b) 동종 파라미터 판단부에 의해, 상기 추출된 파라미터 명칭들 및 표준 용어 데이터에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단하는 단계; 및 (c) 디지털 카탈로그 생성부에 의해, 상기 동종 파라미터로 판단된 파라미터들의 파라미터 값들끼리 서로 대응되도록 상기 적어도 하나 이상의 부품 카탈로그 이미지들에 대한 디지털 부품 카탈로그를 생성하는 단계;를 포함 할 수 있다.

[0007] 또한, 상기 (a) 단계는, 상기 텍스트 추출부에 의해, 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 각 파라미터의 심볼 및 각 파라미터의 단위를 포함하는 부품 속성 정보를 추출하는 단계;를 포함하고, 상기 (b) 단계는: 상기 동종 파라미터 판단부에 의해, 추출된 상기 심볼 및 상기 단위에 기초하여, 상기 추출된 파라미터들이 동종 파라미터인지 여부를 판단하는 단계;를 포함할 수 있다.

[0008] 또한, 상기 (a) 단계는: (a1) 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 상기 부품 속성 정보를 포함하는 사양 테이블 영역을 인식하는 단계; (a2) 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 추출하는 단계를 포함할 수 있다.

[0009] 또한, 상기 (a1) 단계는: 상기 부품 카탈로그 이미지로부터 수평선 및 수직선을 인식하는 단계; 및 상기 인식된 수평선 및 수직선에 기초하여 상기 부품 카탈로그 이미지로부터 상기 사양 테이블 영역을 인식하는 단계;를 포함할 수 있다.

[0010] 또한, 상기 (a) 단계는: 상기 사양 테이블 영역으로부터 테이블 수평선 및 테이블 수직선을 인식하는 단계;를 더 포함하고, 상기 (a2) 단계는: 상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 인식하는 단계;를 포함할 수 있다.

[0011] 또한, 인공지능 학습부에 의해, 상기 테이블 수평선 및 상기 테이블 수직선을 입력 변수로 설정하고 상기 부품 속성 정보를 출력 변수로 설정하여 인공지능 모델을 학습하는 단계;를 더 포함하고, 상기 (a2) 단계는, 상기 인식된 테이블 수평선 및 테이블 수직선을 기초로 상기 인공지능 모델을 이용하여 상기 부품 속성 정보를 인식하는 단계;를 포함할 수 있다.

[0012] 또한, 상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 사양 테이블 헤더를 추출하는 단계;를 더 포함할 수 있다.

[0013] 또한, 텍스트 추출부에 의해, 상기 부품 카탈로그 이미지로부터 부품 번호를 추출하는 단계; 상기 부품 번호 및 부품 번호 규칙 데이터에 기초하여 상기 적어도 하나 이상의 부품 카탈로그 이미지들이 동종 부품의 카탈로그 이미지인지 여부를 판단하는 단계; 및 상기 적어도 하나 이상의 부품 카탈로그 이미지들을 상기 동종 부품의 카탈로그 이미지끼리 분류하는 단계;를 더 포함하고, 상기 (a) 단계는, 상기 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지로부터 부품 속성 정보를 추출하는 단계;를 더 포함할 수 있다.

[0014] 또한, 상기 부품 번호를 추출하는 단계는, 상기 텍스트 추출부에 의해, 상기 부품 카탈로그 이미지의 상단의 미리 설정된 영역에서 텍스트를 추출하여 상기 부품 번호를 추출하는 단계;를 포함할 수 있다.

[0015] 또한, 상기 추출된 부품 속성 정보의 텍스트를 상기 표준 용어 데이터와 비교하여 상기 부품 속성 정보의 텍스

트 오류 여부를 판단하는 단계; 및 상기 부품 속성 정보의 텍스트에 오류가 있는 것으로 판단되면, 상기 표준 용어 데이터에 기초하여 상기 부품 속성 정보를 수정하는 단계;를 더 포함할 수 있다.

[0016] 개시된 발명의 일 측면에 따른 컴퓨터 프로그램은, 상기 카탈로그 이미지의 데이터 추출 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 저장될 수 있다.

[0017] 개시된 발명의 일 측면에 따른 카탈로그 이미지의 데이터 추출 시스템은, 적어도 하나 이상의 부품 카탈로그 이미지로부터 파라미터 명칭들 및 각 파라미터 명칭에 대응되는 파라미터 값을 포함하는 부품 속성 정보를 추출하도록 구성되는 텍스트 추출부; 상기 추출된 파라미터 명칭들 및 표준 용어 데이터에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단하도록 구성되는 동종 파라미터 판단부; 및 상기 동종 파라미터로 판단된 파라미터들의 파라미터 값들끼리 서로 대응되도록 상기 적어도 하나 이상의 부품 카탈로그 이미지들에 대한 디지털 부품 카탈로그를 생성하도록 구성되는 디지털 카탈로그 생성부;를 포함할 수 있다.

[0018] 또한, 상기 텍스트 추출부는, 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 각 파라미터의 심볼 및 각 파라미터의 단위를 포함하는 부품 속성 정보를 추출하도록 구성되고, 상기 동종 파라미터 판단부는, 추출된 상기 심볼 및 상기 단위에 기초하여, 상기 추출된 파라미터들이 동종 파라미터인지 여부를 판단하도록 구성될 수 있다.

[0019] 또한, 상기 텍스트 추출부는: 상기 적어도 하나 이상의 부품 카탈로그 이미지로부터 상기 부품 속성 정보를 포함하는 사양 테이블 영역을 인식하고; 그리고 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 추출하도록 구성될 수 있다.

[0020] 또한, 상기 텍스트 추출부는: 상기 부품 카탈로그 이미지로부터 수평선 및 수직선을 인식하고; 그리고 상기 인식된 수평선 및 수직선에 기초하여 상기 부품 카탈로그 이미지로부터 상기 사양 테이블 영역을 인식하도록 구성될 수 있다.

[0021] 또한, 상기 텍스트 추출부는: 상기 사양 테이블 영역으로부터 테이블 수평선 및 테이블 수직선을 인식하고; 그리고 상기 인식된 테이블 수평선 및 테이블 수직선에 기초하여 상기 사양 테이블 영역으로부터 상기 부품 속성 정보를 인식하도록 구성될 수 있다.

[0022] 또한, 상기 테이블 수평선 및 상기 테이블 수직선을 입력 변수로 설정하고 상기 부품 속성 정보를 출력 변수로 설정하여 인공지능 모델을 학습하도록 구성되는 인공지능 학습부;를 더 포함하고, 상기 텍스트 추출부는, 상기 인식된 테이블 수평선 및 테이블 수직선을 기초로 상기 인공지능 모델을 이용하여 상기 부품 속성 정보를 인식하도록 구성될 수 있다.

[0023] 또한, 프로세서;를 더 포함하고, 상기 프로세서는: 부품 번호 및 부품 번호 규칙 데이터에 기초하여 상기 적어도 하나 이상의 부품 카탈로그 이미지들이 동종 부품의 카탈로그 이미지인지 여부를 판단하고; 그리고 상기 적어도 하나 이상의 부품 카탈로그 이미지들을 상기 동종 부품의 카탈로그 이미지끼리 분류하도록 구성되고, 상기 텍스트 추출부는: 상기 부품 카탈로그 이미지로부터 상기 부품 번호를 추출하도록 구성되고; 그리고 상기 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지로부터 부품 속성 정보를 추출하도록 구성될 수 있다.

[0024] 또한, 상기 텍스트 추출부는, 상기 부품 카탈로그 이미지의 상단의 미리 설정된 영역에서 텍스트를 추출하여 상기 부품 번호를 추출하도록 구성될 수 있다.

[0025] 또한, 프로세서;를 더 포함하고, 상기 프로세서는: 상기 추출된 부품 속성 정보의 텍스트를 상기 표준 용어 데이터와 비교하여 상기 부품 속성 정보의 텍스트 오류 여부를 판단하고; 그리고 상기 부품 속성 정보의 텍스트에 오류가 있는 것으로 판단되면, 상기 표준 용어 데이터에 기초하여 상기 부품 속성 정보를 수정하도록 구성될 수 있다.

발명의 효과

[0026] 개시된 발명의 일 측면에 따르면, 부품 카탈로그에 포함된 각종 부품의 사양 데이터를 자동으로 추출함으로써 제품 개발 과정에서 필요한 부품의 비교 및 검색에 소요되는 시간을 절감할 수 있는 디지털 형식의 부품 카탈로그를 생성할 수 있다.

[0027] 또한, 본 발명의 실시예에 의하면, 동종의 파라미터이지만 서로 다르게 표현된 파라미터들을 동종의 파라미터로 판단할 수 있어서, 부품 카탈로그들의 부품 제조사가 서로 다르더라도 하나의 통합된 디지털 형식의 부품 카탈

로그를 생성할 수 있다.

[0028] 또한, 부품 카탈로그에 개시된 부품 번호들에 기초하여 카탈로그들을 동종의 부품들끼리 분류하고, 동종의 부품들에 대한 부품 카탈로그로부터 부품의 사양 데이터를 추출함으로써 더욱 강건한 텍스트 인식이 가능할 수 있다.

도면의 간단한 설명

[0029] 도 1은 일 실시예에 따른 데이터 추출 시스템의 구성도이다.
 도 2는 일 실시예에 따라 디지털 부품 카탈로그가 생성되는 과정을 도시한 도면이다.
 도 3은 일 실시예에 따라 부품 속성 정보를 추출하는 것을 설명하기 위한 도면이다.
 도 4는 일 실시예에 따라 부품 속성 정보에 기초하여 디지털 부품 카탈로그를 생성하는 것을 설명하기 위한 도면이다.
 도 5는 일 실시예에 따른 표준 용어 데이터를 설명하기 위한 도면이다.
 도 6은 일 실시예에 따라 사양 테이블 영역을 인식하는 것을 설명하기 위한 도면이다.
 도 7은 일 실시예에 따라 테이블 수평선 및 테이블 수직선을 인식하는 것을 설명하기 위한 도면이다.
 도 8은 일 실시예에 따라 사양 테이블 헤더를 추출하는 것을 설명하기 위한 도면이다.
 도 9는 일 실시예에 따른 부품 번호를 도시한 도면이다.
 도 10은 일 실시예에 따라 부품 번호를 추출하는 것을 설명하기 위한 도면이다.
 도 11은 일 실시예에 따라 부품 속성 정보의 오류를 수정하는 것을 설명하기 위한 도면이다.
 도 12는 일 실시예에 따른 데이터 추출 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0030] 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 명세서가 실시예들의 모든 요소들을 설명하는 것은 아니며, 개시된 발명이 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한다. 명세서에서 사용되는 '~부'라는 용어는 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예들에 따라 복수의 '~부'가 하나의 구성요소로 구현되거나, 하나의 '~부'가 복수의 구성요소들을 포함하는 것도 가능하다.

[0031] 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.

[0032] 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 전술된 용어들에 의해 제한되는 것은 아니다.

[0033] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.

[0034] 각 단계들에 있어 식별부호는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.

[0035] 이하 첨부된 도면들을 참고하여 개시된 발명의 작용 원리 및 실시예들에 대해 설명한다.

[0036] 도 1은 일 실시예에 따른 데이터 추출 시스템의 구성도이다.

[0037] 도 1을 참조하면, 본 발명의 실시예에 따른 객체 인식 시스템(100)은, 텍스트 추출부(110), 동종 파라미터 판단부(120), 디지털 카탈로그 생성부(130), 프로세서(140), 메모리(150), 인공지능 학습부(160)를 포함할 수 있다.

[0038] 부품 카탈로그 이미지(200)는 이미지 데이터의 형태로 객체 인식 시스템(100)에 입력될 수 있다.

[0039] 부품 카탈로그 이미지(200)는 스캐닝된 부품 카탈로그일 수 있다. 부품 카탈로그 이미지(200)는 PDF(Portable Document Format) 형식 또는 JPG 형식의 이미지 데이터일 수 있다. 보다 텍스트 인식이 잘 이루어질 수 있도록 PDF 형식의 부품 카탈로그를 JPG 형식으로 변환하고, JPG 형식의 부품 카탈로그에 대하여 본 발명에 따른 데이

터 추출을 하는 것이 바람직할 수 있으나 이에 한정되는 것은 아니며, 어떠한 형식의 이미지 데이터라도 본 발명의 부품 카탈로그 이미지(200)로 입력될 수 있다.

- [0040] 부품 카탈로그 이미지(200)는 복수개일 수 있으며, 각각의 부품 카탈로그 이미지(200)는 전부 한 종류의 부품의 부품 카탈로그 이미지(200)일수도 있지만, 서로 다른 종류의 부품 카탈로그 이미지(200)일수도 있다.
- [0041] 또한 복수개의 부품 카탈로그 이미지(200)는 전부 한 부품 제조사의 부품 카탈로그 이미지(200)일수도 있지만, 서로 다른 부품 제조사의 부품 카탈로그 이미지(200)일수도 있다.
- [0042] 텍스트 추출부(110)는 적어도 하나 이상의 부품 카탈로그 이미지(200)로부터 부품 속성 정보(210)를 추출할 수 있다.
- [0043] 부품 속성 정보(210)는 파라미터 명칭(211)들 및 각 파라미터 명칭(211)에 대응되는 파라미터 값(212)을 포함할 수 있다.
- [0044] 파라미터는 부품의 제원, 특징, 성능을 나타내는 수치일 수 있다.
- [0045] 파라미터 명칭(211)은 해당 파라미터의 일반적인 명칭으로서 텍스트의 형태로 부품 카탈로그 이미지(200)에 표시되어 있을 수 있다.
- [0046] 한편, 어느 특정 파라미터의 파라미터 명칭(211)이 반드시 하나로 정해져 있는 것은 아닐 수 있다. 예를 들어, 동일한 파라미터의 파라미터 명칭(211)이라 하더라도 제작사 별로 다른 명칭을 이용할 수 있다. 따라서, 서로 다른 부품 카탈로그 이미지(200)에 도시된 부품이 동종의 부품이고, 동종의 파라미터들이 해당 부품의 성질을 기재하고 있더라도, 각각의 부품 카탈로그 이미지(200)에 도시된 동종의 파라미터의 파라미터 명칭(211)은 서로 다를 수 있다.
- [0047] 텍스트 추출부(110)는 부품 카탈로그 이미지(200)에 포함된 텍스트를 추출하는 방식으로 부품 속성 정보(210)를 추출할 수 있다. 텍스트 추출부(110)는 OCR(Optical character recognition) 기술을 기반으로 텍스트를 추출할 수 있으나, 부품 카탈로그 이미지(200)에 포함된 텍스트를 추출하여 부품 속성 정보(210)를 획득할 수 있다면 어떠한 방식을 이용해도 상관없다.
- [0048] 동종 파라미터 판단부(120)는 추출된 파라미터 명칭(211)들 및 표준 용어 데이터(151)에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단할 수 있다.
- [0049] 표준 용어 데이터(151)는 어느 특정 파라미터에 대한 파라미터 명칭(211), 파라미터의 심볼(213), 파라미터의 단위(214)가 정리되어 있는 데이터일 수 있다.
- [0050] 즉, 표준 용어 데이터(151)를 참고하면, 특정한 명칭, 특정한 심볼 및 특정한 단위가 어느 파라미터의 명칭, 심볼 및 단위인지 알 수 있다.
- [0051] 표준 용어 데이터(151)는 XML(Extensible Markup Language) 형식의 카탈로그 파일 형태일 수 있으나, 표준 용어 데이터(151)의 형식이 반드시 이에 한정되는 것은 아니다.
- [0052] 디지털 카탈로그 생성부(130)는 동종 파라미터로 판단된 파라미터들의 파라미터 값(212)들끼리 서로 대응되도록 적어도 하나 이상의 부품 카탈로그 이미지(200)들에 대한 디지털 부품 카탈로그(300)를 생성할 수 있다.
- [0053] 디지털 부품 카탈로그(300)는 XML(Extensible Markup Language) 형식의 카탈로그 파일 형태일 수 있으나, 디지털 부품 카탈로그(300)의 형식이 반드시 이에 한정되는 것은 아니다.
- [0054] 즉, 디지털 카탈로그 생성부(130)는 복수개의 부품 카탈로그 이미지(200)에 혼재되어 있는 부품들의 정보가 정리된 하나의 통합된 디지털 파일을 생성할 수 있다.
- [0055] 복수개의 이미지 형태의 부품 카탈로그들은 각 부품의 속성 정보들이 여기저기 혼재되어 있고, 동종 파라미터의 정보라 해도 부품의 제조사에 따라 다르게 표시되어 있었기 때문에, 사용자가 부품들끼리 사양을 비교하거나 부품을 검색하는데 어려움이 있었다.
- [0056] 본 발명의 실시예에 따라 생성된 디지털 부품 카탈로그(300)는 각 부품들의 속성 정보를 전부 모아서 하나의 데이터에 정리되어 있을 수 있다. 또한, 디지털 부품 카탈로그(300)는 이미지 형태의 부품 카탈로그에서 다르게 표현되어 있었던 동종 파라미터라 해도 동일한 종류의 파라미터로서 정리되어 있으므로, 사용자가 특정 파라미터의 비교를 통해 부품의 사양을 용이하게 비교할 수 있다.

- [0057] 결과적으로, 본 발명의 데이터 추출 시스템(100)은 사용자가 부품 카탈로그 이미지(200)의 부품들의 사양 비교 및 검색을 보다 용이하게 할 수 있는 디지털 부품 카탈로그(300)를 생성할 수 있다.
- [0058] 텍스트 추출부(110), 동종 파라미터 판단부(120), 디지털 카탈로그 생성부(130), 인공지능 학습부(160)는 데이터 추출 시스템(100)에 포함된 복수개의 프로세서(140) 중 어느 하나의 프로세서(140)를 포함할 수 있다. 또한, 지금까지 설명된 본 발명의 실시예 및 앞으로 설명할 실시예에 따른 객체 인식 방법은, 프로세서(140)에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다.
- [0059] 여기서 프로그램은, 프로그램 명령, 데이터 파일 및 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 프로그램은 기계어 코드나 고급 언어 코드를 이용하여 설계 및 제작된 것일 수 있다. 프로그램은 상술한 부호 수정을 위한 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 기 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 전술한 정보 표시 방법을 구현하기 위한 프로그램은, 프로세서(140)에 의해 판독 가능한 기록매체에 기록될 수 있다. 이때, 기록매체는 메모리(150)일 수 있다.
- [0060] 메모리(150)는 전술한 동작 및 후술하는 동작을 수행하는 프로그램을 저장할 수 있으며, 메모리(150)는 저장된 프로그램을 실행시킬 수 있다. 프로세서(140)와 메모리(150)가 복수인 경우에, 이들이 하나의 칩에 집적되는 것도 가능하고, 물리적으로 분리된 위치에 마련되는 것도 가능하다. 메모리(150)는 데이터를 일시적으로 기억하기 위한 S램(Static Random Access Memory, S-RAM), D램(Dynamic Random Access Memory) 등의 휘발성 메모리를 포함할 수 있다. 또한, 메모리(150)는 제어 프로그램 및 제어 데이터를 장기간 저장하기 위한 롬(Read Only Memory), 이피롬(Erasable Programmable Read Only Memory: EPROM), 이이피롬(Electrically Erasable Programmable Read Only Memory: EEPROM) 등의 비휘발성 메모리를 포함할 수 있다.
- [0061] 프로세서(140)는 각종 논리 회로와 연산 회로를 포함할 수 있으며, 메모리(150)로부터 제공된 프로그램에 따라 데이터를 처리하고, 처리 결과에 따라 제어 신호를 생성할 수 있다.
- [0062] 도 2는 일 실시예에 따라 디지털 부품 카탈로그가 생성되는 과정을 도시한 도면이다.
- [0063] 도 2를 참조하면, 스캔된 부품 카탈로그 이미지(200)는 데이터 추출 시스템(100)에 입력될 수 있다.
- [0064] 이후, 텍스트 추출부(110)는 OCR 기술에 의하여 문자열 객체를 인식하고, 그래픽스 기술에 의하여 선 객체를 인식할 수 있다.
- [0065] 동종 파라미터 판단부(120)는 표준 용어 데이터(151)에 기초하여 인식된 문자열들 사이의 연관 관계를 생성할 수 있다. 즉, 표시만 다른 텍스트로 되어 있었을 뿐이고 실제로는 동일한 파라미터끼리 연관 관계를 생성할 수 있다.
- [0066] 마지막으로, 디지털 카탈로그 생성부(130)는 생성된 문자열 연관 관계에 기초하여 부품의 속성 별로 부품 리스트가 정리된 디지털 부품 카탈로그(300)를 생성할 수 있다.
- [0067] 도 3은 일 실시예에 따라 부품 속성 정보를 추출하는 것을 설명하기 위한 도면이다.
- [0068] 도 3을 참조하면, 텍스트 추출부(110)는 적어도 하나 이상의 부품 카탈로그 이미지(200)로부터 각 파라미터의 심볼(213) 및 각 파라미터의 단위(214)를 포함하는 부품 속성 정보(210)를 추출할 수 있다.
- [0069] 한편, 어느 특정 파라미터의 심볼(213) 및 단위는 반드시 하나로 정해져 있는 것은 아닐 수 있다. 예를 들어, 동일한 파라미터의 심볼(213) 및 단위라 하더라도 제작사 별로 다른 심볼 및 단위를 이용할 수 있다. 따라서, 서로 다른 부품 카탈로그 이미지(200)에 도시된 부품이 동종의 부품이고, 동종의 파라미터들이 해당 부품의 성질을 기재하고 있더라도, 각각의 부품 카탈로그 이미지(200)에 도시된 동종의 파라미터의 심볼(213) 및 단위는 다를 수 있다.
- [0070] 동종 파라미터 판단부(120)는 추출된 심볼 및 단위에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단할 수 있다.
- [0071] 어느 특정한 파라미터가 서로 다른 부품 카탈로그 이미지(200)에 다른 파라미터 명칭(211)으로 표시되어 있다고 하더라도, 해당 파라미터에 대응되는 파라미터의 심볼(213) 및 파라미터의 단위(214)는 동일하게 표시되어 있을 수 있다.
- [0072] 따라서 동종 파라미터 판단부(120)는 각 파라미터들의 심볼 및 단위가 동일한지 여부를 판단하고, 심볼 및 단위

가 동일하다고 판단되면 비록 다른 명칭으로 표시되어 있었던 파라미터라 하더라도 동일한 파라미터라고 판단할 수 있다.

- [0073] 또한, 동종 파라미터 판단부(120)는 각 파라미터들의 심볼 및 단위가 동일하지 않더라도, 표준 용어 데이터(151)를 참조했을 때 해당 심볼들 및 해당 단위들이 동일한 파라미터에 대응되는 심볼 및 단위라면 해당 파라미터들을 동일한 파라미터라고 판단할 수 있다.
- [0074] 도 4는 일 실시예에 따라 부품 속성 정보에 기초하여 디지털 부품 카탈로그(300)를 생성하는 것을 설명하기 위한 도면이다.
- [0075] 도 4를 참조하면, 디지털 카탈로그 생성부(130)는 추출된 파라미터 명칭(211), 파라미터 값(212), 파라미터의 심볼(213) 및 파라미터의 단위(214)에 기초하여 디지털 부품 카탈로그(300)를 생성할 수 있다.
- [0076] 디지털 카탈로그는 부품 카탈로그 이미지(200)에서는 명칭, 단위 및 심볼이 다르게 표현되어 있었던 파라미터들이 서로 동일한 종류의 파라미터라면, 해당 파라미터에 대해선 명칭, 단위 및 심볼이 동일한 용어로 통일하여 디지털 부품 카탈로그(300)를 생성할 수 있다. 단위를 통일하는 과정에서 파라미터의 값 또한 단위에 맞도록 수치가 변경되도록 디지털 부품 카탈로그(300)가 생성될 수 있다.
- [0077] 한편, 텍스트 추출부(110)는 부품 카탈로그 이미지(200)로부터 부품 속성 정보(210)뿐 아니라 각 부품 카탈로그 이미지(200)의 소개 페이지에서 OCR 방식의 텍스트 인식을 통해 반복되는 텍스트를 기초로 해당 부품의 제조사, 제품명, 가격 등을 결정할 수 있다. 이렇게 결정된 제조사, 제품명, 가격 등의 정보는 사용자에게 의해 실제 정보에 부합되는지 검증될 수 있다.
- [0078] 디지털 카탈로그 생성부(130)는 부품 속성 정보(210)뿐 아니라 부품의 제조사, 제품명, 가격 등의 정보 또한 부품별로 정리하여 디지털 부품 카탈로그(300)를 생성할 수 있다.
- [0079] 도 5는 일 실시예에 따른 표준 용어 데이터를 설명하기 위한 도면이다.
- [0080] 도 5를 참조하면, 표준 용어 데이터(151)는 복수개의 참조 데이터를 참조하여 미리 생성될 수 있다.
- [0081] 참조 데이터는 부품 카탈로그 이미지(200)에 포함된 일부 부품들의 사양 데이터들에 기초하여 생성된 데이터일 수 있다. 이때, 참조 데이터가 반드시 한 제조사의 부품 카탈로그 이미지(200)들에 기초하여 생성되어야 하는 것은 아니다.
- [0082] 한편, 동일한 파라미터라 하더라도 제조사 별로 해당 파라미터의 명칭 및 심볼이 다를 수 있다. 예를 들어, GDGIT사의 부품 데이터 및 MOOG사의 부품 데이터에는 Peak Torque로 기재된 파라미터와 Kollmorgen사의 부품 데이터에는 Peak Torque Rating이라고 기재된 파라미터는 동일한 파라미터일 수 있다.
- [0083] 또한, GDGIT사의 부품 데이터에는 Friction Torque로 기재된 파라미터, MOOG사의 부품 데이터에는 Total Breakaway Torque로 기재된 파라미터와 Kollmorgen사의 부품 데이터에는 Static Friction (Max.)라고 기재된 파라미터는 동일한 파라미터일 수 있다.
- [0084] 표준 용어 데이터(151)는 이처럼 표현은 다르지만 실제로는 같은 파라미터를 나타내는 명칭들끼리 분류되어 메모리(150)에 저장될 수 있다.
- [0085] 동종 파라미터 판단부(120)는 어느 부품 카탈로그 이미지(200)로부터 추출된 파라미터의 명칭이 Friction Torque이고, 또다른 부품 카탈로그 이미지(200)로부터 추출된 파라미터의 명칭이 Total Breakaway Torque라 해도 두 파라미터를 동종 파라미터로 판단할 수 있다.
- [0086] 도 6은 일 실시예에 따라 사양 테이블 영역을 인식하는 것을 설명하기 위한 도면이며, 도 7은 일 실시예에 따라 테이블 수평선 및 테이블 수직선을 인식하는 것을 설명하기 위한 도면이다.
- [0087] 도 6을 참조하면, 텍스트 추출부(110)는 적어도 하나 이상의 부품 카탈로그 이미지(200)로부터 부품 속성 정보(210)를 포함하는 사양 테이블 영역(201)을 인식할 수 있다.
- [0088] 텍스트 추출부(110)는 부품 카탈로그 이미지(200)로부터 수평선 및 수직선을 인식할 수 있다. 이때, 텍스트 추출부(110)는 딥러닝 기반의 객체 검출 기술에 기초하여 부품 카탈로그 이미지(200)로부터 수평선과 수직선을 인식할 수 있다.
- [0089] 딥러닝 기반의 객체 검출 기술은 이미지로부터 추출되는 객체의 특징(feature)을 데이터를 기반으로 학습한다. 이때, 이미지로부터 특징(feature)을 추출하는 방식을 학습하기 위해 여러 단계의 컨볼루션 계층(convolution

layer)을 쌓은 CNN(Convolutional Neural Networks) 구조가 활용될 수 있으나 이에 한정되는 것은 아니다.

- [0090] 텍스트 추출부(110)는 인식된 수평선 및 수직선에 기초하여 부품 카탈로그 이미지(200)로부터 사양 테이블 영역(201)을 인식할 수 있다.
- [0091] 텍스트 추출부(110)는 사양 테이블 영역(201)으로부터 부품 속성 정보(210)를 추출할 수 있다.
- [0092] 도 7을 참조하면, 텍스트 추출부(110)는 사양 테이블 영역(201)으로부터 테이블 수평선(202) 및 테이블 수직선(203)을 인식할 수 있다.
- [0093] 테이블 수평선(202)은 테이블을 구성하는 선 중에서 부품 카탈로그 이미지(200)에 대해서 수평인 선이며, 테이블 수직선(203)은 테이블을 구성하는 선 중에서 부품 카탈로그 이미지(200)에 대해서 수직인 선일 수 있다.
- [0094] 텍스트 추출부(110)는 다른 테이블 수평선(202) 및 테이블 수직선(203)과 선의 길이가 다른 테이블 수평선(202) 및 테이블 수직선(203)을 인식할 수 있다.
- [0095] 텍스트 추출부(110)는 딥러닝 기반의 객체 검출 기술에 기초하여 테이블 수평선(202) 및 테이블 수직선(203)을 인식할 수 있다.
- [0096] 텍스트 추출부(110)는 인식된 테이블 수평선(202) 및 테이블 수직선(203)에 기초하여 사양 테이블 영역(201)으로부터 부품 속성 정보(210)를 인식할 수 있다. 이때, 다른 테이블 수평선(202) 및 테이블 수직선(203)과 선의 길이가 다른 테이블 수평선(202) 및 테이블 수직선(203)을 고려하여 텍스트를 인식한다면 보다 강건하게 테이블 영역에 포함되어 있는 부품 속성 정보(210)를 인식할 수 있다.
- [0097] 인공지능 학습부(160)는 학습용 데이터를 통해, 테이블 수평선(202) 및 테이블 수직선(203)의 특징을 입력 변수로 설정하고 부품 속성 정보(210)를 출력 변수로 설정하여 인공지능 모델을 학습하도록 구성될 수 있다.
- [0098] 테이블 수평선(202) 및 테이블 수직선(203)의 특징은 해당 선에 대한 다양한 특징을 나타내는 정보일 수 있다. 예를 들어, 테이블 수평선(202) 및 테이블 수직선(203)의 특징은 해당 선의 길이, 형태, 테이블 영역에서의 위치 등에 대한 정보일 수 있다.
- [0099] 이때, 인공지능 모델을 학습하는 것은 기계 학습(Machine Learning) 방식을 통한 학습일 수 있다.
- [0100] 기계 학습이란 다수의 파라미터로 구성된 모델을 이용하며, 주어진 데이터로 파라미터를 최적화하는 것을 의미할 수 있다. 인공지능 학습부(160)는 입력에 따라 인공지능 모델을 통해 최종적으로 나온 결과물(output)인 출력과, 입력에 대한 레이블(정답)을 이용하여 인공지능 모델을 학습할 수 있다.
- [0101] 인공지능 학습부(160)는 기계 학습 방식뿐만 아니라 딥 러닝 방식을 통해서도 학습을 수행할 수 있다.
- [0102] 인공지능 모델은 테이블 수평선(202) 및 테이블 수직선(203)의 특징 정보를 입력 변수로 하고, 부품 속성 정보(210)를 출력 변수로 설정한 모델일 수 있다. 인공지능 모델은 메모리(150)에 저장될 수 있다.
- [0103] 텍스트 추출부(110)는 부품 카탈로그 이미지(200)로부터 인식된 테이블 수평선(202) 및 테이블 수직선(203)을 기초로 인공지능 모델을 이용하여 부품 속성 정보(210)를 인식할 수 있다.
- [0104] 도 8은 일 실시예에 따라 사양 테이블 헤더를 추출하는 것을 설명하기 위한 도면이다.
- [0105] 도 8을 참조하면, 텍스트 추출부(110)는 인식된 테이블 수평선(202) 및 테이블 수직선(203)에 기초하여 사양 테이블 영역(201)으로부터 사양 테이블 헤더(204)를 추출할 수 있다. 이렇게 추출된 헤더는 사용자에게 의해 실제 부품 카탈로그 이미지(200)의 헤더에 부합되는지 검증될 수 있다.
- [0106] 도 9는 일 실시예에 따른 부품 번호를 도식한 도면이다.
- [0107] 도 9를 참조하면, 복수개의 부품 카탈로그 이미지(200) 중 일부 이미지는 개시하고 있는 부품에 대한 부품 번호(205)를 표시할 수 있다.
- [0108] 부품 번호(205)는 특정 부품에 하나씩 대응되는 번호일 수 있다. 또한, 부품 번호(205)의 체계는 제조사별로 다를 수 있다.
- [0109] 부품 번호 규칙 데이터는 제조사별로 부품에 부품 번호(205)를 부여하는 규칙을 모은 데이터일 수 있다. 부품 번호 규칙 데이터는 메모리(150)에 저장될 수 있다.
- [0110] 프로세서(140)는 부품 번호(205) 및 부품 번호 규칙 데이터에 기초하여 적어도 하나 이상의 부품 카탈로그 이미

지(200)들이 동종 부품의 카탈로그 이미지인지 여부를 판단할 수 있다.

- [0111] 예를 들어, GDGIT사의 부품 중 부품 번호(205)가 '1600C-059', '4500C-080C' 및 '1600C-059J'인 부품은 서로 부품 번호(205)가 다르지만 모두 'BRUSHLESS DC MOTOR'의 한 종류이다. 이때, 프로세서(140)는 부품 번호 규칙 데이터에 기초하여 해당 부품들의 부품 카탈로그 이미지(200)를 동종 부품의 카탈로그 이미지라고 판단할 수 있다.
- [0112] 또한, Kollmorgen사의 부품 중 부품 번호(205)가 'BMS-5905', 'BM-40702'인 부품 또한 도시된 GDGIT사의 부품들과 부품 번호(205) 및 부품 체계가 다르지만, 모두 'BRUSHLESS DC MOTOR'의 한 종류이다. 이때, 프로세서(140)는 부품 번호 규칙 데이터에 기초하여 해당 부품들의 부품 카탈로그 이미지(200)를 동종 부품의 카탈로그 이미지라고 판단할 수 있다.
- [0113] 도 10은 일 실시예에 따라 부품 번호를 추출하는 것을 설명하기 위한 도면이다.
- [0114] 도 10을 참조하면, 부품 번호(205)는 일반적으로 부품 카탈로그 이미지(200)의 상단에 위치하고 있음을 확인할 수 있다.
- [0115] 텍스트 추출부(110)는 부품 카탈로그 이미지(200)로부터 부품 번호(205)를 추출할 수 있다. 이때, 텍스트 추출부(110)는 부품 카탈로그 이미지(200)의 상단의 미리 설정된 영역(206)에서 텍스트를 추출하여 부품 번호(205)를 추출할 수 있다.
- [0116] 카탈로그 이미지의 상단의 미리 설정된 영역(206)은 부품 카탈로그 이미지(200) 중 상단 20% 면적을 차지하는 영역일 수 있으나, 이에 한정되는 것은 아니다.
- [0117] 프로세서(140)는 추출된 부품 번호(205) 및 부품 번호 규칙 데이터에 기초하여 해당 부품들의 부품 카탈로그 이미지(200)들이 동종 부품의 카탈로그 이미지인지 여부를 판단하고, 적어도 하나 이상의 부품 카탈로그 이미지(200)들을 동종 부품의 카탈로그 이미지끼리 분류할 수 있다.
- [0118] 텍스트 추출부(110)는 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지로부터 부품 속성 정보(210)를 추출할 수 있다.
- [0119] 동종의 부품이라면 카탈로그 이미지 및 사양 테이블에 기재된 텍스트의 내용과 텍스트가 나열된 순서, 특징 등이 유사할 수 있다. 따라서, 텍스트 추출부(110)는 동종 부품의 카탈로그 이미지로부터 부품 속성 정보(210)를 추출하고, 추후 이종의 부품과 부품 속성 정보(210)를 통합하는 방식으로 보다 강건하게 부품 속성 정보(210)를 추출할 수 있다.
- [0120] 도 11은 일 실시예에 따라 부품 속성 정보의 오류를 수정하는 것을 설명하기 위한 도면이다.
- [0121] 프로세서(140)는 추출된 부품 속성 정보(210)의 텍스트를 표준 용어 데이터(151)와 비교하여 부품 속성 정보(210)의 텍스트 오류 여부를 판단할 수 있다.
- [0122] 이때, 프로세서(140)는 잘못 인식된 텍스트를 표준 용어 데이터(151)에 해당 부품 속성 정보(210)의 오류의 예시로서 업데이트 하면서 표준 용어 데이터(151)를 업데이트할 수 있다.
- [0123] 프로세서(140)는 부품 속성 정보(210)의 텍스트에 오류가 있는 것으로 판단되면, 표준 용어 데이터(151)에 기초하여 부품 속성 정보(210)를 수정할 수 있다.
- [0124] 예를 들어, 프로세서(140)는 인식에 실패한 위 첨자, 아래 첨자, 그리스 문자, 수학 기호 등을 수정할 수 있다.
- [0125] 이상에서 설명된 구성요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성요소들의 상호 위치는 시스템의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.
- [0126] 도 12는 일 실시예에 따른 데이터 추출 방법의 순서도이다. 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 구성이 추가되거나 삭제될 수 있음은 물론이다.
- [0127] 도 12를 참조하면, 프로세서(140)는 부품 번호(205) 및 부품 번호 규칙 데이터에 기초하여 적어도 하나 이상의 부품 카탈로그 이미지(200)들이 동종 부품의 카탈로그 이미지인지 여부를 판단하고, 적어도 하나 이상의 부품 카탈로그 이미지(200)들을 동종 부품의 카탈로그 이미지끼리 분류할 수 있다(1001).
- [0128] 텍스트 추출부(110)는 동종 부품의 카탈로그 이미지로 분류된 복수개의 카탈로그 이미지부터 수평선 및 수직선

을 인식하고, 인식된 수평선 및 수직선에 기초하여 부품 카탈로그 이미지(200)로부터 사양 테이블 영역(201)을 인식할 수 있다(1002).

[0129] 텍스트 추출부(110)는 사양 테이블 영역(201)으로부터 테이블 수평선(202) 및 테이블 수직선(203)을 인식할 수 있다(1003).

[0130] 텍스트 추출부(110)는 인식된 테이블 수평선(202) 및 테이블 수직선(203)에 기초하여 사양 테이블 영역(201)으로부터 사양 테이블 헤더(204)를 추출할 수 있다(1004).

[0131] 텍스트 추출부(110)는 인식된 테이블 수평선(202) 및 테이블 수직선(203)에 기초하여 사양 테이블 영역(201)으로부터 부품 속성 정보(210)를 추출할 수 있다(1005). 이때, 텍스트 추출부(110)는, 인식된 테이블 수평선(202) 및 테이블 수직선(203)을 기초로 인공지능 모델을 이용하여 부품 속성 정보(210)를 인식할 수 있다.

[0132] 프로세서(140)는 추출된 부품 속성 정보(210)의 텍스트를 표준 용어 데이터(151)와 비교하여 부품 속성 정보(210)의 텍스트 오류 여부를 판단할 수 있다(1006).

[0133] 프로세서(140)는 부품 속성 정보(210)의 텍스트에 오류가 있는 것으로 판단되면, 표준 용어 데이터(151)에 기초하여 부품 속성 정보(210)를 수정할 수 있다(1007).

[0134] 동종 파라미터 판단부(120)는 추출된 부품 속성 정보(210) 및 표준 용어 데이터(151)에 기초하여, 추출된 파라미터들이 동종 파라미터인지 여부를 판단할 수 있다(1008).

[0135] 디지털 카탈로그 생성부(130)는 동종 파라미터로 판단된 파라미터들의 파라미터 값(212)들끼리 서로 대응되도록 적어도 하나 이상의 부품 카탈로그 이미지(200)들에 대한 디지털 부품 카탈로그(300)를 생성할 수 있다(1009).

[0136] 이상에서와 같이 첨부된 도면을 참조하여 개시된 실시예들을 설명하였다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고도, 개시된 실시예들과 다른 형태로 본 발명이 실시될 수 있음을 이해할 것이다. 개시된 실시예들은 예시적인 것이며, 한정적으로 해석되어서는 안 된다.

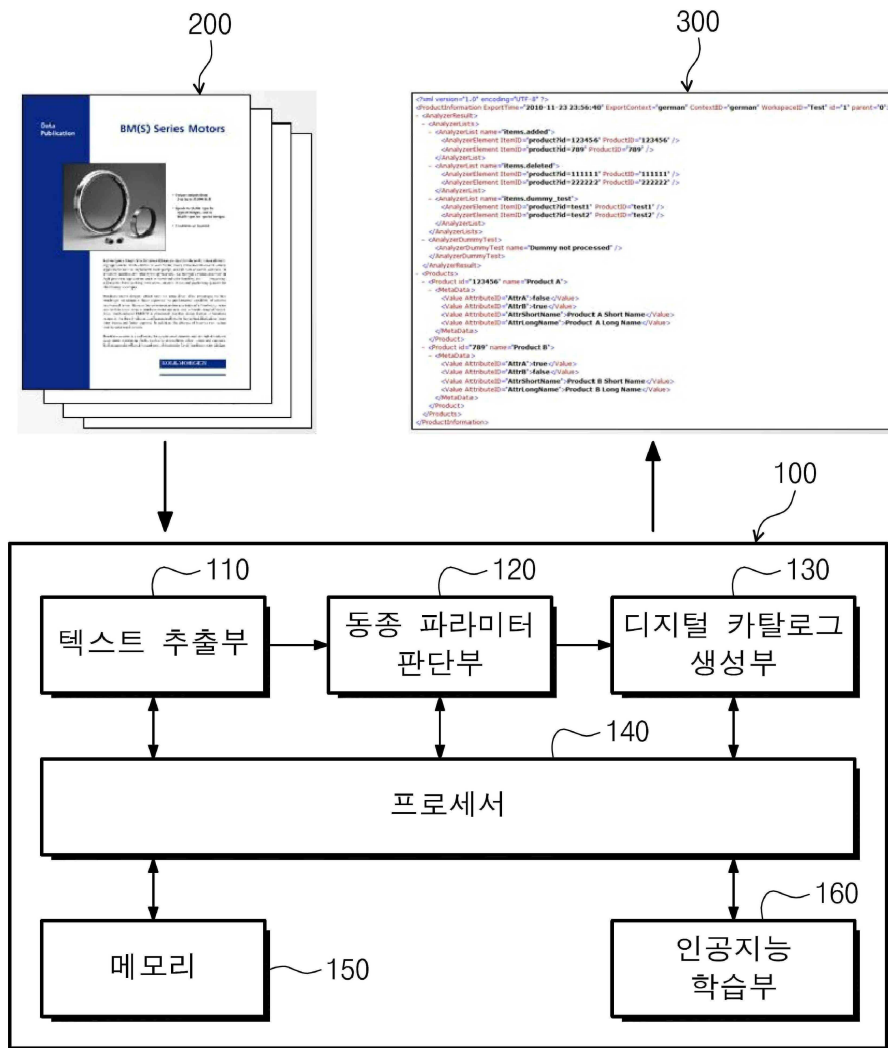
부호의 설명

- [0137]
- 100: 데이터 추출 시스템
 - 110: 텍스트 추출부
 - 120: 동종 파라미터 판단부
 - 130: 디지털 카탈로그 생성부
 - 140: 프로세서
 - 150: 메모리
 - 151: 표준 용어 데이터
 - 160: 인공지능 학습부
 - 200: 부품 카탈로그 이미지
 - 201: 사양 테이블 영역
 - 202: 테이블 수평선
 - 203: 테이블 수직선
 - 204: 사양 테이블 헤더
 - 205: 부품 번호
 - 206: 상단의 미리 설정된 영역
 - 210: 부품 속성 정보
 - 211: 파라미터 명칭

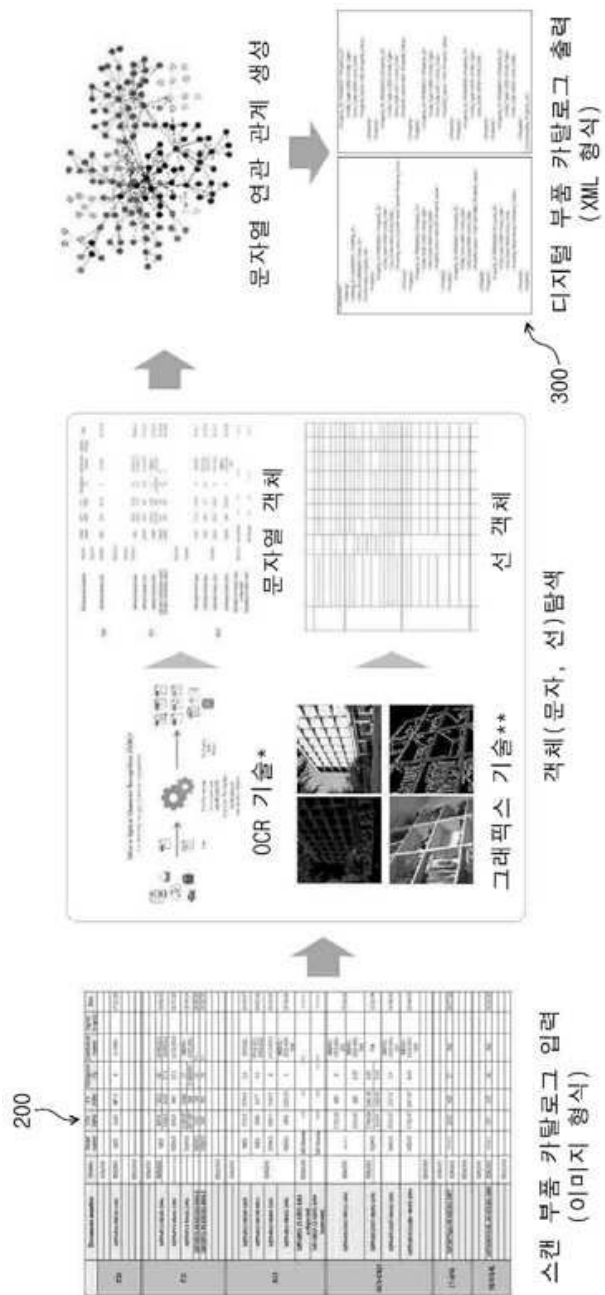
- 212: 파라미터 값
- 213: 파라미터의 심볼
- 214: 파라미터의 단위
- 300: 디지털 부품 카탈로그

도면

도면1



도면2

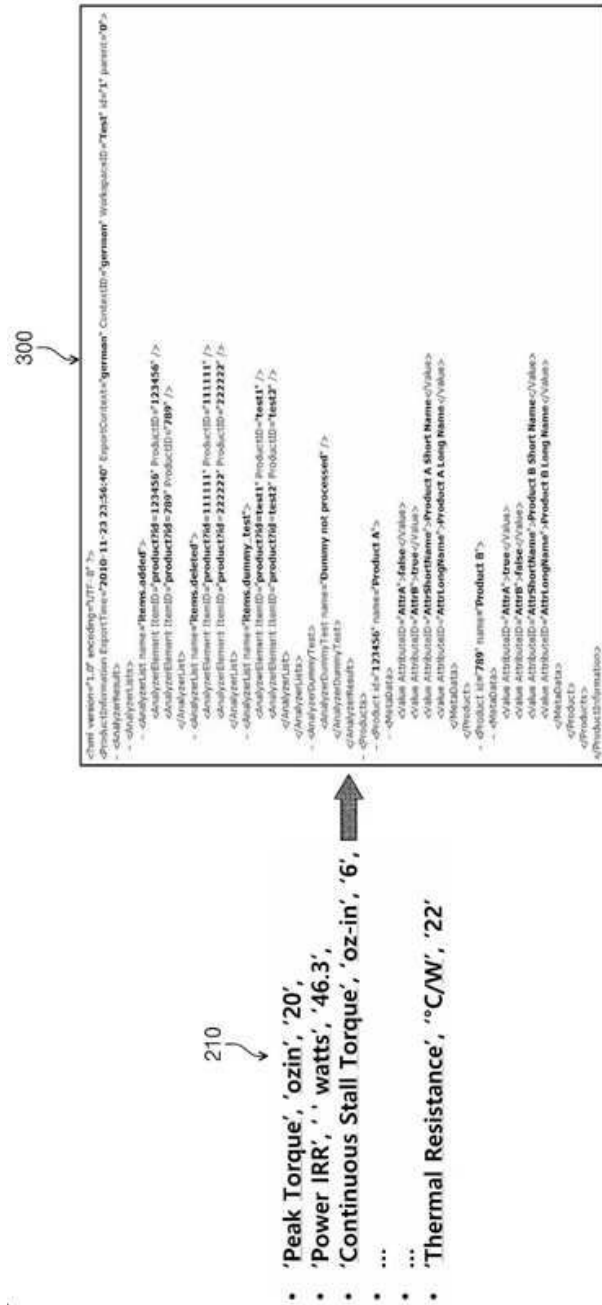


도면3

Size Constants	UNITS	SYMBOL	VALUE
Peak Torque, Stalled @ Vp:	oz-in	Tp	20
Power I ² R @ Tp:	watts	P	46.3
Continuous Stall Torque:	oz-in	Tcs	6
Motor Constant:	oz-in/ $\sqrt{W}$	Km	2.94
Electrical Time Constant	ms	Te	3.3
Mechanical Time Constant:	ms	Tm	3.3
Damping Factor: (zero impedance)	oz-in/ (rad/sec)	Fo	0.06
Max Breakaway Torque:	oz-in	Tf	0.4
Rotor Inertia:	oz-in-sec ²	Jm	0.0002
Number of Poles:		p	8
Weight:	OZ	WT	2.4
Rated Winding Temperature	°C	Temp	155
Thermal Resistance:	°C/W	tpr	22

'Size Constants', 'VALUE', 'SYMBOL', 'UNITS',
 'Peak Torque, Stalled', 'Vp:', 'Tp', '20', 'ozin',
 'Power IRR', 'Tp:', 'P', '46.3', 'watts',
 'Torque:', 'oz-in', 'Tcs', 'Continuous Stall', '6',
 '2.94', 'OZ-in/W', 'Km', 'Motor Constant:',
 'Time', 'Electrical', 'Constant', 'Te', '3.3', 'ms',
 'Mechanical Time Constant:', 'Tm', '3.3', 'ms',
 'oz-in/', 'Damping Factor:', 'Fo', '0.06', '(zero impedance)',
 'lrad', 'sec]',
 'Max Breakaway Torque:', 'Tf', '0.4', 'ozin',
 'OZ-in-sece', 'Rotor Inertia:', 'Jm', '0.0002',
 'Number of Poles:', '8', 'p',
 'WT', '2.4', 'Weight:', 'oz',
 'Temp', 'Rated Winding Temperature', '155',
 'Thermal Resistance:', 'OCIW', '22', 'tpr']

도면4



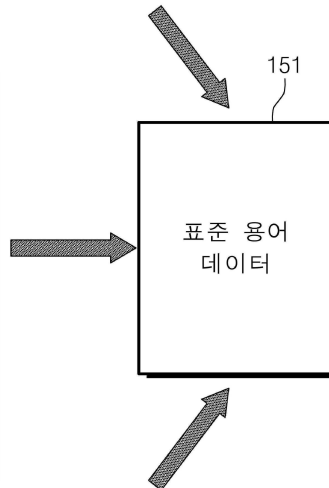
도면5

GDGIT

Size Constants	Value	Units
Peak Torque, T_p ①	27.6	oz-in
Motor Constant, K_M ②	1.10	oz-in/ $\sqrt{W}$
Number of Poles ⑥	20	
Weight ⑧	12	lb
Motor Inertia, J_M ⑦	1.18×10^{-2}	oz-in-s ²
Friction Torque, T_f ③	0.35	oz-in
Electrical Time Constant, τ_e ⑨	2.1	ms
Mechanical Time Constant, τ_m ⑩	7.19	ms
Temperature Rise, Housed TPR' ⑤	1.15	°C/W
Ripple Torque, T_R	4	max. avg. to peak (%)

MOOG

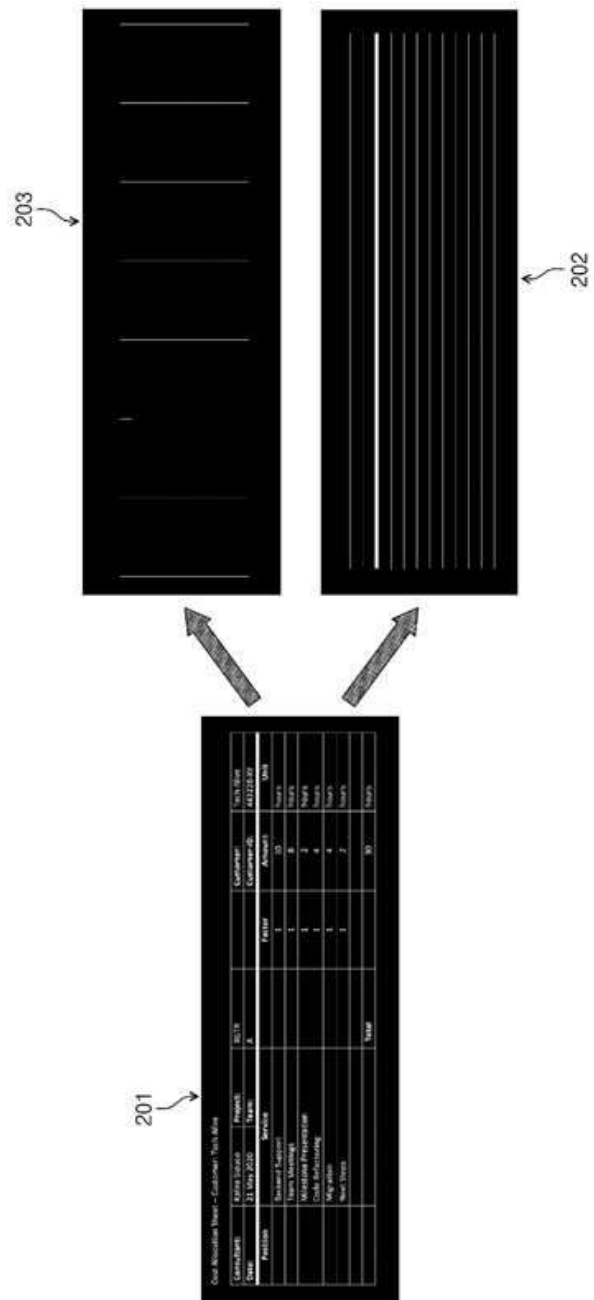
Performance Data		
Parameter	Units	Value
Peak Torque (Tp) ①	oz-in	6.0
Power at Tp (Pp)	watts	25
Motor Constant (Km) ②	oz-in/ $\sqrt{W}$	1.20
Continuous Torque (Tc)	oz-in	1.7
Power at Continuous Torque	watts	2.1
Input Volts at Continuous Torque	volts	5.6
Electrical Time Constant (Te) ⑨	milli-sec	0.30
Mechanical Time Constant (Tm) ⑩	milli-sec	-
Damping Factor (Fo)	oz-in/rad/sec	0.00101
Rotor Inertia (Jm) ⑦	oz-in-sec ²	0.00033
Angular Excursion (Page 46, Curve y)	degrees	±22.5
Total Breakaway Torque (Ti) ③	oz-in	-
Max. Allowable Wdg. Temp.	°C	155
Number of Poles ⑥		4
Weight ⑧	oz	1.5



Kollmorgen

Size Constants	Units	Value
Peak Torque Rating- T_p ①	N-m lb-ft	1.73 1.28
Motor Constant- K_m ②	N-m/ $\sqrt{Watt}$ lb-ft/ $\sqrt{Watt}$	.139 .102
Static Friction (Max.)- T_f ③	N-m lb-ft	2.7×10^{-2} 2.0×10^{-2}
Damping Coeff. INF Impedence- F_i ④	N-m per rad/s lb-ft per rad/s	2.3×10^{-5} 1.7×10^{-5}
Max. Winding Temperature	°C	200
Temperature Rise per Watt- TPR ⑤	°C/Watt	1.0
Number of Poles ⑥	-	4
Rotor Inertia- J_m ⑦	kg-m ² lb-ft-s ²	4.84×10^{-4} 3.57×10^{-4}
Motor Weight ⑧	kg(f) lbs	2.9 6.4

도면7

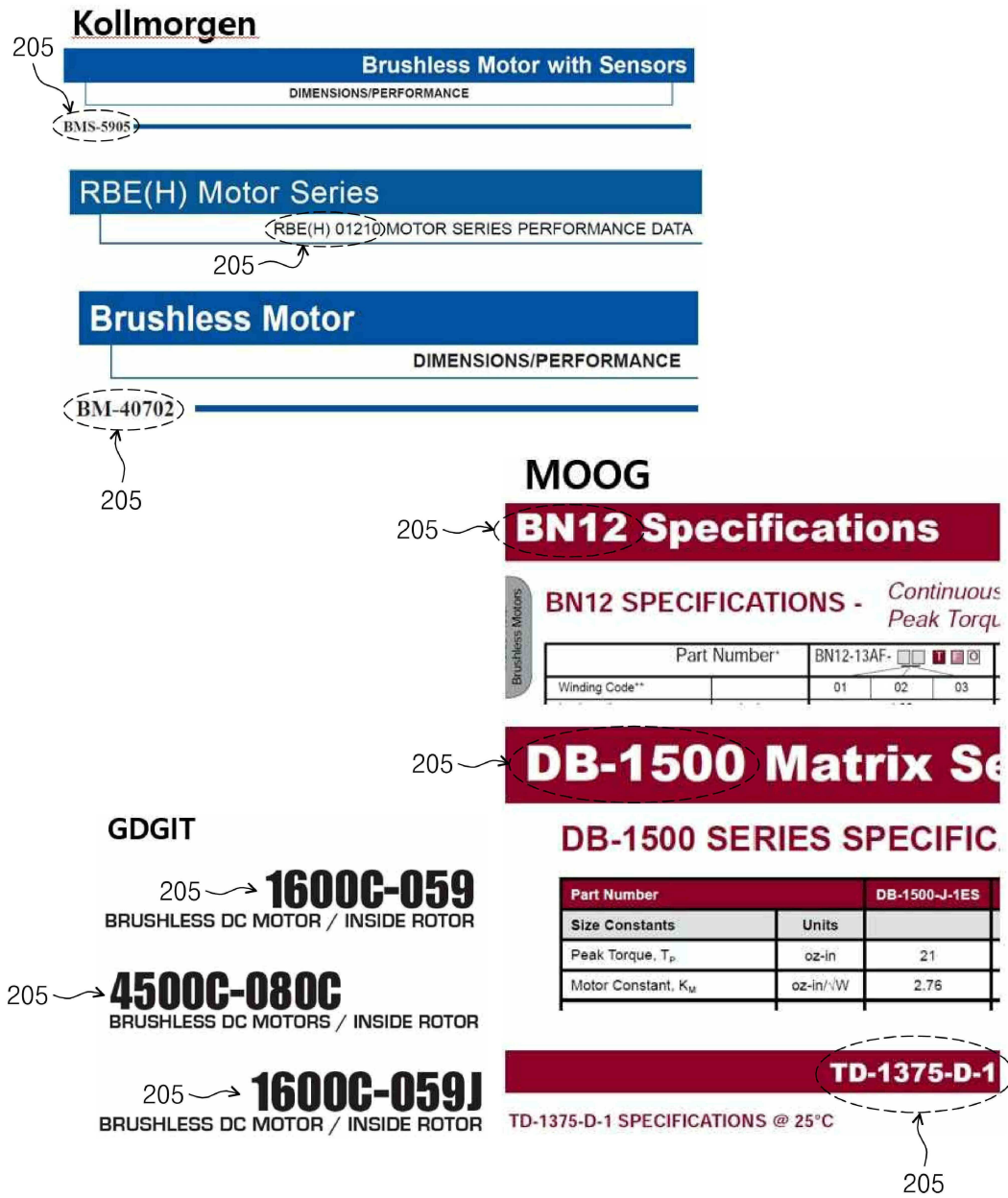


도면8

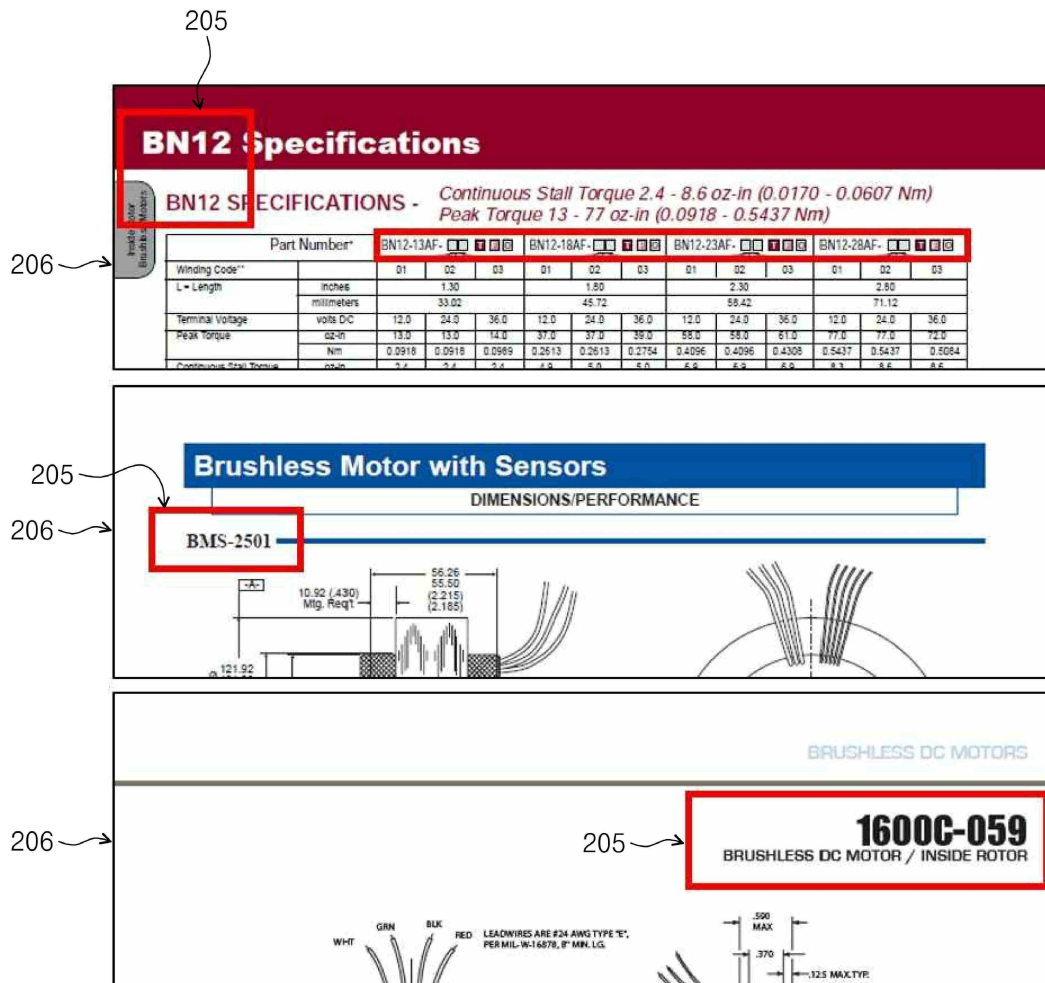
201

Header Row	Constant column	Unit column		Symbol column		Value column	
		UNITS		SYMBOL		VALUE	
	Size Constants						
	Peak Torque, Stalled @ Vp:	oz-in		Tp		20	
	Power I ² R @ Tp:	watts		P		46.3	
	Continuous Stall Torque:	oz-in		Tcs		6	
	Motor Constant:	oz-in/ $\sqrt{W}$		Km		2.94	
	Electrical Time Constant	ms		Te		3.3	
	Mechanical Time Constant:	ms		Tm		3.3	
	Damping Factor: (zero impedance)	oz-in/ (rad/sec)		Fo		0.06	
	Max Breakaway Torque:	oz-in		Tf		0.4	
	Rotor Inertia:	oz-in-sec ²		Jm		0.0002	
	Number of Poles:			p		8	
	Weight:	OZ		WT		2.4	
	Rated Winding Temperature	°C		Temp		155	
	Thermal Resistance:	°C/W		tpr		22	

도면9



도면10



카탈로그 이미지 파일의 상단 예시

도면11

201

Size Constants	UNITS	SYMBOL	VALUE
Peak Torque, Stated @ Vp	oz-in	Tp	20
Power IR @ Tp	watts	P	46.3
Continuous Stall Torque	oz-in	Tcs	6
Motor Constant	oz-in/A	Km	2.94
Electrical Time Constant	ms	Te	3.3
Mechanical Time Constant	ms	Tm	3.3
Damping Factor (zero impedance)	oz-in/[rad/sec]	Fo	0.06
Max Breakaway Torque	oz-in	Tf	0.4
Rotor Inertia	oz-in-sec ²	Jm	0.0002
Number of Poles	p	p	8
Weight	OZ	WT	2.4
Rated Winding Temperature	°C	Temp	155
Thermal Resistance	°C/W	tpc	22

210 'Power IRR' → 'Power IR'

210 'OZ-in/W' → oz-in/root(W)

210 'Damping Factor' & '(zero impedance)' → 'Damping Factor (zero impedance)'

도면12

