



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0083138  
(43) 공개일자 2021년07월06일

- |                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(51) 국제특허분류(Int. Cl.)<br/>G06F 40/232 (2020.01) G06F 40/242 (2020.01)</p> <p>(52) CPC특허분류<br/>G06F 40/232 (2020.01)<br/>G06F 40/242 (2020.01)</p> <p>(21) 출원번호 10-2020-0012743</p> <p>(22) 출원일자 2020년02월03일<br/>심사청구일자 2020년02월03일</p> <p>(30) 우선권주장<br/>1020190174628 2019년12월26일 대한민국(KR)</p> | <p>(71) 출원인<br/>울산과학기술원<br/>울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자<br/>이정혜<br/>울산광역시 울주군 언양읍 유니스트길 50</p> <p>김수현<br/>울산광역시 울주군 언양읍 유니스트길 50<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/>특허법인태백</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

전체 청구항 수 : 총 14 항

(54) 발명의 명칭 품목 기술내용을 반영한 품목명 오탈자 탐지 및 수정 시스템, 그리고 그 방법

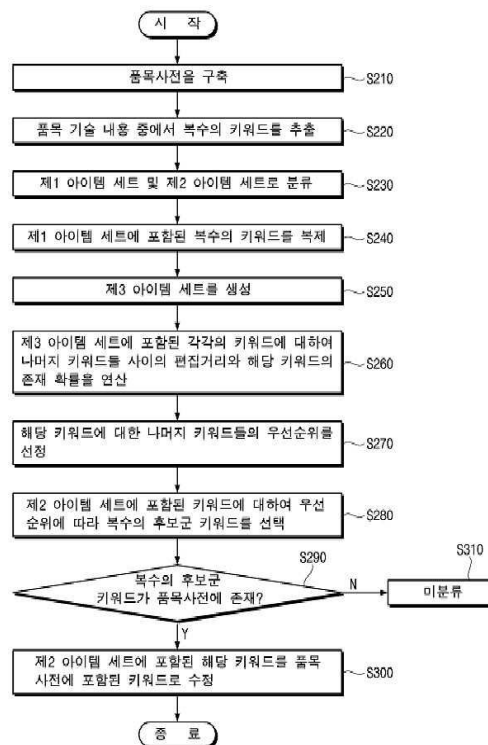
(57) 요약

본 발명은 품목 기술내용을 반영한 품목명 오탈자 탐지 및 수정 시스템, 그리고 그 방법에 관한 것이다

본 발명에 따르면, 품목명 자동 오탈자 탐지 및 수정 시스템을 이용한 오탈자 탐지 및 수정 방법에 있어서, 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축하는 단계, 사용자로부터 입력된 품목 기술 내용 중에서 복수

(뒷면에 계속)

대표도 - 도2



의 키워드를 추출하는 단계, 상기 추출된 복수의 키워드에 대하여 상기 품목사전에 포함되었는지 여부에 따라 제 1 아이템 세트 및 제2 아이템 세트로 분류하는 단계, 상기 제1 아이템 세트에 포함된 복수의 키워드를 복제하는 단계, 상기 제1 아이템 세트와 상기 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성하는 단계, 상기 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산하는 단계, 상기 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하는 단계, 상기 제2 아이템 세트에 포함된 키워드에 대하여 상기 우선 순위에 따라 복수의 후보군 키워드를 선택하는 단계, 그리고 상기 복수의 후보군 키워드와 상기 품목사전을 비교하여, 비교 결과에 따라 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 단계를 포함한다.

(72) 발명자

**손원호**

울산광역시 울주군 언양읍 유니스트길 50

**임동철**

울산광역시 울주군 언양읍 유니스트길 50

**길영재**

울산광역시 울주군 언양읍 유니스트길 50

**한석주**

울산광역시 울주군 언양읍 유니스트길 50

**김경빈**

울산광역시 울주군 언양읍 유니스트길 50

**김재호**

울산광역시 울주군 언양읍 유니스트길 50

이 발명을 지원한 국가연구개발사업

과제고유번호

2.190533.01

부처명

울산과학기술원 스마트 항만물류 지원센터

과제관리(전문)기관명

울산항만공사

연구사업명

스마트 항만 물류 지원센터 지원사업

연구과제명

스마트 항만 물류 지원센터(항만물류 데이터 사이언스 센터) 위탁

기 여 율

1/1

과제수행기관명

울산과학기술원

연구기간

2019.01.01 ~ 2019.12.31

## 명세서

### 청구범위

#### 청구항 1

품목명 자동 오탈자 탐지 및 수정 시스템을 이용한 오탈자 탐지 및 수정 방법에 있어서,  
 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축하는 단계,  
 사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출하는 단계,  
 상기 추출된 복수의 키워드에 대하여 상기 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류하는 단계,  
 상기 제1 아이템 세트에 포함된 복수의 키워드를 복제하는 단계,  
 상기 제1 아이템 세트와 상기 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성하는 단계,  
 상기 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산하는 단계,  
 상기 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하는 단계,  
 상기 제2 아이템 세트에 포함된 키워드에 대하여 상기 우선 순위에 따라 복수의 후보군 키워드를 선택하는 단계, 그리고  
 상기 복수의 후보군 키워드와 상기 품목사전을 비교하여, 비교 결과에 따라 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 단계를 포함하는 오탈자 탐지 및 수정 방법.

#### 청구항 2

제1항에 있어서,  
 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 단계는,  
 상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하면, 상기 제2 아이템 세트에 포함된 해당 키워드를 상기 품목사전에 포함된 키워드로 수정하고,  
 상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하지 않으면, 상기 제2 아이템 세트에 포함된 해당 키워드를 미분류 처리하는 오탈자 탐지 및 수정 방법.

#### 청구항 3

제1항에 있어서,  
 상기 제1 아이템 세트는,  
 상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드들의 집합인 오탈자 탐지 및 수정 방법.

#### 청구항 4

제3항에 있어서,  
 상기 제2 아이템 세트는,  
 상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하지 않는 키워드들의 집합인 오탈자 탐지 및 수정 방법.

#### 청구항 5

제1항에 있어서,

상기 복수의 키워드를 복제하는 단계는,

상기 제1 아이템 세트에 존재하는 키워드의 개수가 상기 제2 아이템 세트에 포함되는 키워드의 개수와 가장 가까운 개수가 되도록 상기 제1 아이템 세트에 포함된 키워드들을  $n$ 배 복제하는 오탈자 탐지 및 수정 방법.

#### 청구항 6

제1항에 있어서,

상기 품목사전은,

물류 화물, 유류 화물, 항공 화물 및 선박 화물 중에서 어느 하나를 포함하는 아이템에 대한 품목사전인 오탈자 탐지 및 수정 방법.

#### 청구항 7

제1항에 있어서,

상기 편집 거리는,

2개의 키워드들 사이에 삽입, 삭제, 대체를 통하여 동일한 키워드가 되기 위한 삽입, 삭제, 대체의 총 연산 횟수를 나타내는 오탈자 탐지 및 수정 방법.

#### 청구항 8

품목명 자동 오탈자 탐지 및 수정 시스템에 있어서,

카테고리 별로 키워드가 분류되어 있는 품목사전을 구축하는 구축부,

사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출하는 추출부,

상기 추출된 복수의 키워드에 대하여 상기 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류하는 분류부,

상기 제1 아이템 세트에 포함된 복수의 키워드를 복제하는 복제부,

상기 제1 아이템 세트와 상기 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성하는 생성부,

상기 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산하는 연산부,

상기 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하하고, 상기 제2 아이템 세트에 포함된 키워드에 대하여 상기 우선 순위에 따라 복수의 후보군 키워드를 선택하는 선택부, 그리고

상기 복수의 후보군 키워드와 상기 품목사전을 비교하여, 비교 결과에 따라 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 제어부를 포함하는 오탈자 탐지 및 수정 시스템.

#### 청구항 9

제8항에 있어서,

상기 제어부는,

상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하면, 상기 제2 아이템 세트에 포함된 해당 키워드를 상기 품목사전에 포함된 키워드로 수정하고,

상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하지 않으면, 상기 제2 아이템 세트에 포함된 해당 키워드를 미분류 처리하는 오탈자 탐지 및 수정 시스템.

#### 청구항 10

제8항에 있어서,

상기 제1 아이템 세트는,

상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드들의 집합인 오탈자 탐지 및 수정 시스템.

#### 청구항 11

제10항에 있어서,

상기 제2 아이템 세트는,

상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하지 않는 키워드들의 집합인 오탈자 탐지 및 수정 시스템.

#### 청구항 12

제8항에 있어서,

상기 복제부는,

상기 제1 아이템 세트에 존재하는 키워드의 개수가 상기 제2 아이템 세트에 포함되는 키워드의 개수와 가장 가까운 개수가 되도록 상기 제1 아이템 세트에 포함된 키워드들을  $n$ 배 복제하는 오탈자 탐지 및 수정 시스템.

#### 청구항 13

제8항에 있어서,

상기 품목사전은,

물류 화물, 유류 화물, 항공 화물 및 선박 화물 중에서 어느 하나를 포함하는 아이템에 대한 품목사전인 오탈자 탐지 및 수정 시스템.

#### 청구항 14

제8항에 있어서,

상기 편집 거리는,

2개의 키워드들 사이에 삽입, 삭제, 대체를 통하여 동일한 키워드가 되기 위한 삽입, 삭제, 대체의 총 연산 횟수를 나타내는 오탈자 탐지 및 수정 시스템.

### 발명의 설명

#### 기술 분야

[0001] 본 발명은 품목 기술내용을 반영한 품목명 오탈자 탐지 및 수정 시스템, 그리고 그 방법에 관한 것으로, 품목사전과 키워드를 이용하여 품목 기술 내용의 키워드를 수정하는 품목 기술내용을 반영한 품목명 오탈자 탐지 및 수정 시스템, 그리고 그 방법에 관한 것이다.

#### 배경 기술

[0002] PC 상에서 문자(영문, 한글, 숫자, 특수 문자 등 입력 가능한 모든 형태 포괄하는 의미임)를 입력할 때나 모바일 기기에서 문자를 입력할때의 전형적인 오타 수정 방식은 백스페이스(Backspace) 등의 삭제 기능키를 입력하여 오타가 난 단어 이후의 단어를 모두 지우고 다시 입력하거나, 마우스(PC의 경우) 또는 손가락 터치로(모바일 기기의 경우) 커서를 오타가 난 단어로 옮기고 단어를 수정하는 것이었다.

[0003] 모바일 기기에서 손가락 터치로 커서를 옮길 경우에는 터치의 정확성이 떨어져서 원하는 위치에 커서를 한 번에 옮기는 것이 힘들고 역시 번거로운 동작을 필요로 하였다.

[0004] 위와 같은 오타 수정 방식의 번거로움을 줄이고자 등장한 것이 자동 수정 기능이다. 이 기능은 이미 상용화되어 있는 것으로서, 사용자가 A란 단어를 입력할 경우, 이를 오타로 판단하면 자동으로 B라는 단어로 고쳐주는 것이었다.

- [0005] 하지만, 자동 수정 기능에서는 수정 입력되는 단어가 사용자가 수정 의도를 가지고 입력한 단어가 아니고 시스템이 사용자가 실제로 입력하였을 것으로 추정하는 단어이므로, 추정의 정확성이 떨어지는 경우에는 문제가 발생하였다.
- [0006] 특히 물류의 경우, 품명에는 오탈자가 포함되어 있는 경우, 품목에 대한 정확한 분류가 이루어지지 않는 경우가 빈번하였다.
- [0007] 일반적으로 물류의 분류시에, 입력된 품목 텍스트에 의해 카테고리가 바로 분류되는 것이 아닌, 입력된 품목 텍스트를 기반으로 사람이 선하증권(Bill of Lading, BL) 문서에 국제 표준 분류 코드 중 하나인 6자리 숫자의 harmonized system (HS) 코드를 입력하게 되고, 이와 같은 HS 코드를 기반으로 카테고리 분류를 하게 된다.
- [0008] 그러나, 이 HS 코드의 설명이 굉장히 모호하고, 불분명하여 입력할 당시 오기입하는 경우가 많아 정확한 카테고리별 물동량 집계에 한계가 존재한다.
- [0009] 입력된 코드에 의해 분류되는 카테고리가 다르게 결정되므로, 품목의 카테고리로 결정되는 문제점이 있어 품명으로 분류 하는 것에 대한 필요성이 대두된다.
- [0010] 다만 품명에는 오탈자가 포함될 수 있으므로 품목에 대한 정확한 분류를 위하여 품목명에 대한 오탈자를 탐지하고 수정하는 기술이 필요하게 되었다.
- [0011] 본 발명의 배경이 되는 기술은 대한민국 등록특허 제10-1946545호(2019.02.11. 공고)에 개시되어 있다.

## 발명의 내용

### 해결하려는 과제

- [0012] 본 발명이 이루고자 하는 기술적 과제는 품목사전과 키워드를 이용하여 품목 기술 내용의 키워드를 수정하는 품목 기술내용을 반영한 품목명 오탈자 탐지 및 수정 시스템, 그리고 그 방법을 제공하기 위한 것이다.

### 과제의 해결 수단

- [0013] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따르면, 품목명 자동 오탈자 탐지 및 수정 시스템을 이용한 오탈자 탐지 및 수정 방법에 있어서, 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축하는 단계, 사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출하는 단계, 상기 추출된 복수의 키워드에 대하여 상기 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류하는 단계, 상기 제1 아이템 세트에 포함된 복수의 키워드를 복제하는 단계, 상기 제1 아이템 세트와 상기 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성하는 단계, 상기 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산하는 단계, 상기 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하는 단계, 상기 제2 아이템 세트에 포함된 키워드에 대하여 상기 우선 순위에 따라 복수의 후보군 키워드를 선택하는 단계, 그리고 상기 복수의 후보군 키워드와 상기 품목사전을 비교하여, 비교 결과에 따라 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 단계를 포함한다.
- [0014] 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 단계는, 상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하면, 상기 제2 아이템 세트에 포함된 해당 키워드를 상기 품목사전에 포함된 키워드로 수정하고, 상기 복수의 후보군 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드가 존재하지 않으면, 상기 제2 아이템 세트에 포함된 해당 키워드를 미분류 처리할 수 있다.
- [0015] 상기 제1 아이템 세트는, 상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하는 키워드들의 집합일 수 있다.
- [0016] 상기 제2 아이템 세트는, 상기 추출된 복수의 키워드 중에서 상기 품목사전에 포함된 키워드와 일치하지 않는 키워드들의 집합일 수 있다.
- [0017] 상기 복수의 키워드를 복제하는 단계는, 상기 제1 아이템 세트에 존재하는 키워드의 개수가 상기 제2 아이템 세트에 포함되는 키워드의 개수와 가장 가까운 개수가 되도록 상기 제1 아이템 세트에 포함된 키워드들을  $n$ 배 복제할 수 있다.
- [0018] 상기 품목사전은, 물류 화물, 유류 화물, 항공 화물 및 선박 화물 중에서 어느 하나를 포함하는 아이템에 대한 품목사전일 수 있다.

[0019] 상기 편집 거리는, 2개의 키워드들 사이에 삽입, 삭제, 대체를 통하여 동일한 키워드가 되기 위한 삽입, 삭제, 대체의 총 연산 횟수를 나타낼 수 있다.

[0020] 본 발명의 다른 실시예에 따르면, 품목명 자동 오탈자 탐지 및 수정 시스템에 있어서, 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축하는 구축부, 사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출하는 추출부, 상기 추출된 복수의 키워드에 대하여 상기 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류하는 분류부, 상기 제1 아이템 세트에 포함된 복수의 키워드를 복제하는 복제부, 상기 제1 아이템 세트와 상기 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성하는 생성부, 상기 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산하는 연산부, 상기 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하하고, 상기 제2 아이템 세트에 포함된 키워드에 대하여 상기 우선 순위에 따라 복수의 후보군 키워드를 선택하는 선택부, 그리고 상기 복수의 후보군 키워드와 상기 품목사전을 비교하여, 비교 결과에 따라 상기 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류하는 제어부를 포함한다.

### 발명의 효과

[0021] 이와 같이 본 발명에 따르면, 기존의 방법보다 정확하게 오탈자를 체크하고 수정할 수 있으며, 새로운 분류체계를 구축할 수 있다.

[0022] 또한, 본 발명에 따르면 HS 코드 기반의 카테고리를 분류하는 것이 아니라, 품목 텍스트 자체를 이용하여 품목 키워드 사전을 제작하고 해당 사전을 기반으로 분류하므로, 품목 키워드에 오탈자가 존재하더라도 더욱 정확하게 품목을 분류할 수 있다.

### 도면의 간단한 설명

[0023] 도 1은 본 발명의 실시예에 따른 품목명 자동 오탈자 탐지 및 수정 시스템의 구성을 나타낸 구성도이다.

도 2는 본 발명의 실시예에 따른 오탈자 탐지 및 수정 방법을 설명하기 위한 순서도이다.

도 3은 본 발명의 실시예에 따른 품목사전을 설명하기 위한 도면이다.

도 4는 본 발명의 실시예에 따른 제1 아이템 세트, 제2 아이템 세트 및 제3 아이템 세트를 설명하기 위한 도면이다.

도 5는 본 발명의 실시예에 따른 편집거리 알고리즘을 설명하기 위한 도면이다.

### 발명을 실시하기 위한 구체적인 내용

[0024] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시 예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0025] 명세서 전체에서, 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.

[0026] 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.

[0027] 도 1은 본 발명의 실시예에 따른 품목명 자동 오탈자 탐지 및 수정 시스템의 구성을 나타낸 구성도이다.

[0028] 도 1에서 나타난 것처럼, 본 발명의 실시예에 따른 품목명 자동 오탈자 탐지 및 수정 시스템(100)은 구축부(110), 추출부(120), 분류부(130), 복제부(140), 생성부(150), 연산부(160), 선택부(170) 및 제어부(180)를 포함한다.

[0029] 먼저, 구축부(110)는 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축한다.

[0030] 여기서, 품목사전은 물류 화물, 유류 화물, 항공 화물 및 선박 화물 중에서 어느 하나에 해당되는 아이템에 대한 키워드를 포함한다.

- [0031] 다음으로, 추출부(120)는 사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출한다.
- [0032] 이때, 복수의 키워드는 품목사전에 포함된 키워드이거나 품목사전의 키워드와 유사한 키워드를 포함한다.
- [0033] 다음으로, 분류부(130)는 추출된 복수의 키워드에 대하여 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류한다.
- [0034] 여기서, 제1 아이템 세트(Clean Item Set)는 추출된 복수의 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드들의 집합이다.
- [0035] 또한, 제2 아이템 세트(Dirty Item Set)는 추출된 복수의 키워드 중에서 품목사전에 포함된 키워드와 일치하지 않는 키워드들의 집합이다.
- [0036] 다음으로, 복제부(140)는 제1 아이템 세트에 존재하는 키워드를 복제한다.
- [0037] 즉, 복제부(140)는 제1 아이템 세트에 존재하는 키워드의 개수가 제2 아이템 세트에 포함되는 키워드의 개수와 가장 가까운 개수가 되도록 제1 아이템 세트에 포함된 키워드들을  $n$ 배 복제한다.
- [0038] 다음으로, 생성부(150)는 복제된 제1 아이템 세트와 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성한다.
- [0039] 그리고, 연산부(160)는 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산한다.
- [0040] 여기서, 편집거리는 2개의 키워드들 사이에 삽입, 삭제, 대체를 통하여 동일한 키워드가 되기 위한 삽입, 삭제, 대체의 총 연산 횟수를 나타낸 것이다.
- [0041] 그리고, 존재 확률은 제3 아이템 세트에 존재하는 해당 키워드의 확률을 나타낸다.
- [0042] 다음으로, 선택부(170)는 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 정하고, 제2 아이템 세트에 포함된 키워드에 대하여 우선 순위에 따라 복수의 후보군 키워드를 선택한다.
- [0043] 다음으로, 제어부(180)는 복수의 후보군 키워드와 품목사전을 비교하여, 비교 결과에 따라 제2 아이템 세트에 포함된 키워드를 수정 또는 미분류한다.
- [0044] 이때, 제어부(180)는 복수의 후보군 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드가 존재하면, 제2 아이템 세트에 포함된 해당 키워드를 품목사전에 포함된 키워드로 수정한다.
- [0045] 또한, 제어부(180)는 복수의 후보군 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드가 존재하지 않으면, 제2 아이템 세트에 포함된 해당 키워드를 미분류 처리한다.
- [0046] 여기서, 미분류 처리는 해당 품목사전에 존재하지 않는 것으로 판단한다.
- [0047] 이하에서는 도 2 내지 도 4를 이용하여 본 발명의 실시예에 따른 품목명 자동 오탃자 탐지 및 수정 시스템(100)을 이용한 오탃자 탐지 및 수정 방법에 대하여 설명한다.
- [0048] 도 2는 본 발명의 실시예에 따른 오탃자 탐지 및 수정 방법을 설명하기 위한 순서도이다.
- [0049] 먼저, 도 2에서 나타난 것처럼, 구축부(110)는 카테고리 별로 키워드가 분류되어 있는 품목사전을 구축한다(S210).
- [0050] 도 3은 본 발명의 실시예에 따른 품목사전을 설명하기 위한 도면이다.
- [0051] 도 3에서 나타난 것처럼, 품목사전은 카테고리(Category), 서브카테고리(Subcategory) 및 키워드(Keyword)로 저장된다.
- [0052] 예를 들어, 도 3에서 나타난 것처럼, 품목사전의 카테고리가 "GAS"이면, 서브카테고리는 "LPG GAS" 및 "LNG GAS"로 나누어지고, 키워드는 "LPG", "PROPYLENE", "BUTANE", "LNG" 및 "GAS CONDENSATE"로 나누어진다.
- [0053] 즉, 하나의 카테고리에는 복수의 서브카테고리를 포함하고 있으며, 서브카테고리의 경우 복수의 키워드를 포함한다.
- [0054] 다음으로, 추출부(120)는 사용자로부터 입력된 품목 기술 내용 중에서 복수의 키워드를 추출한다(S220).
- [0055] 이때, 추출부(120)는 사용자가 기술한 품목 내용 중에서 품목사전에 해당되는 키워드와 해당되지 않는 키워드

각각을 추출한다.

- [0056] 그러면, 분류부(130)는 추출된 복수의 키워드에 대하여 품목사전에 포함되었는지 여부에 따라 제1 아이템 세트 및 제2 아이템 세트로 분류한다(S230).
- [0057] 도 4는 본 발명의 실시예에 따른 제1 아이템 세트, 제2 아이템 세트 및 제3 아이템 세트를 설명하기 위한 도면이다.
- [0058] 여기서, 제1 아이템 세트(Clean Item Set)는 추출된 복수의 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드들의 집합이다.
- [0059] 그리고, 제2 아이템 세트(Dirty Item Set)는 추출된 복수의 키워드 중에서 품목사전에 포함된 키워드와 일치하지 않는 키워드들의 집합이다.
- [0060] 예를 들어, 도 4에서 나타난 것처럼, 제1 아이템 세트에는 "Crude", "Gasoline"과 같이 품목사전에 기재된 키워드가 포함되며, 제2 아이템 세트에는 품목사전에 포함되지 않은 "Cuude", "Curde" 및 "Cruude"와 같은 키워드가 포함된다.
- [0061] 다음으로, 복제부(140)는 제1 아이템 세트에 포함된 복수의 키워드를 복제한다(S240).
- [0062] 여기서, 복제부(140)는 제1 아이템 세트에 존재하는 키워드의 개수가 제2 아이템 세트에 포함되는 키워드의 개수와 가장 가까운 개수가 되도록 제1 아이템 세트에 포함된 키워드들을 n배 복제할 수 있다.
- [0063] 또한, 복제부(140)는 제1 아이템 세트에 존재하는 키워드의 개수가 제2 아이템 세트에 포함되는 키워드의 개수와 동일한 개수가 되도록 제1 아이템 세트에 포함된 키워드들을 복제할 수 있다.
- [0064] 예를 들어, 제1 아이템 세트의 키워드 개수가 28개이고, 제2 아이템 세트의 키워드 개수가 150개라고 가정하면, 복제부(140)는 제1 아이템 세트에 포함된 키워드 28개를 5배로 복제하여 총 140개의 키워드를 생성할 수 있다.
- [0065] 그리고, 생성부(150)는 복제된 제1 아이템 세트와 제2 아이템 세트를 결합하여 제3 아이템 세트를 생성한다(S250).
- [0066] 즉, 도 4에서 나타난 것처럼, 제3 아이템 세트는 복제된 제1 아이템 세트와 제2 아이템 세트를 포함한다.
- [0067] 다음으로, 연산부(160)는 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리와 해당 키워드의 존재 확률을 연산한다(S260).
- [0068] 여기서, 연산부(160)는 편집거리 알고리즘을 이용하여 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리를 연산한다.
- [0069] 도 5는 본 발명의 실시예에 따른 편집거리 알고리즘을 설명하기 위한 도면이다.
- [0070] 도 5에서 나타난 것처럼, 편집거리 알고리즘은 각각의 키워드를 비교하여 동일한 키워드로 수정시키기 위한 알고리즘으로, 본 발명에서는 제3 아이템 세트에 포함된 각각의 키워드에 대하여 나머지 키워드들 사이의 편집거리를 연산한다.
- [0071] 이때, 편집거리는 삽입, 삭제 및 대체 중에서 어느 하나를 진행한 횟수에 해당한다.
- [0072] 예를 들어, 도 5에서 나타난 것처럼, "Crude" 와 "Cuude"의 경우, "r"이 "u"로 대체되었으므로 편집 거리는 1로 연산된다. 마찬가지로, "Crude" 와 "Cruude"의 경우, "u"가 1개 더 포함되었으므로 편집 거리는 1로 연산된다.
- [0073] 그리고, 존재 확률은 제3 아이템 세트에 존재하는 키워드의 확률을 나타낸다.
- [0074] 예를 들어, 제3 아이템 세트의 키워드 개수가 200개 이고 제3 아이템 세트에 존재하는 키워드 "Crude"의 개수가 5개라고 하면, 존재 확률은 2.5%로 연산된다.
- [0075] 그러면, 연산부(160)는 존재확률과 편집거리를 각각 정규화하고, 정규화된 존재확률과 편집거리를 합산한다.
- [0076] 그러면, 선택부(170)는 S260 단계에서 연산된 결과를 이용하여 해당 키워드에 대한 나머지 키워드들의 우선순위를 선정한다(S270).
- [0077] 여기서, 선택부(170)는 제3 아이템 세트에 포함된 각각의 키워드에 대하여 우선순위를 선정한다.
- [0078] 예를 들어, 제3 아이템 세트의 키워드 개수가 200개 이고, 해당 키워드가 "CUUDE"일 경우, 선택부(170)는

"CUUDE"를 제외한 나머지 키워드들에 대하여 존재확률과 편집거리를 연산한 결과에 따라 우선순위를 결정한다.

- [0079] 다음으로, 선택부(170)는 제2 아이템 세트에 포함된 키워드 각각에 대하여 우선 순위에 따라 복수의 후보군 키워드를 선택한다(S280).
- [0080] 여기서, 선택부(170)는 우선순위가 가장 높은 키워드를 순서대로 복수의 후보군 키워드로 선택한다.
- [0081] 또한, 후보군 키워드의 개수는 제2 아이템 세트의 개수에 따라 달라질 수 있으며, 관리자의 선택에 따라 변경될 수 있다.
- [0082] 예를 들어, 후보군 키워드의 개수가 3개이고, 해당 키워드가 "CUUDE"일 경우, 선택부(170)는 우선순위가 높은 3개의 키워드인 "CRUDE", "CRUDE" 및 "CRUDEE"를 후보군 키워드로 설정한다.
- [0083] 다음으로, 제어부(180)는 선택된 복수의 후보군 키워드가 품목사전에 존재하는지 여부를 비교한다(S290).
- [0084] 즉, 상기의 예에서 제어부(180)는 후보군 키워드에 해당하는 "CRUDE", "CRUDE" 및 "CRUDEE" 중에서 품목사전에 존재하는 키워드와 일치하는 키워드가 존재하는지 여부를 판단한다.
- [0085] 만일, 복수의 후보군 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드가 존재하면, 제어부(180)는 제2 아이템 세트에 포함된 해당 키워드를 품목사전에 포함된 키워드로 수정한다(S300).
- [0086] 상기의 예에서, 후보군 키워드에 해당하는 "CRUDE", "CRUDE" 및 "CRUDEE" 중에서 "CRUDE"는 품목사전에 등록된 키워드이므로, 제어부(180)는 해당 키워드 "CUUDE"를 "CRUDE"로 수정한다.
- [0087] 또한, "CUUDE"가 제2 아이템 세트에 복수개 존재하는 경우, 제어부(180)는 제2 아이템 세트에 존재하는 "CUUDE"를 전부 "CRUDE"로 변경한다.
- [0088] 그리고, 후보군 키워드 중에서 품목 사전에 포함되는 키워드와 일치하는 키워드가 복수개 존재하면, 제어부(180)는 해당 후보군 키워드 중에서 우선순위가 가장 높은 후보군 키워드로 해당 키워드를 수정할 수 있다.
- [0089] 이와 같이, 제2 아이템 세트에 포함된 "CUUDE"에 대한 키워드가 품목사전에 포함된 "CRUDE"로 수정이 완료되면, 다른 제2 아이템 세트에 포함된 키워드에 대하여 S270 내지 S300 단계를 반복함으로써, 제2 아이템 세트에 포함되는 키워드의 개수를 줄여나가도록 한다.
- [0090] 따라서, 제2 아이템 세트에 포함된 키워드의 개수는 줄어들며, 상대적으로 제1 아이템 세트의 확률이 높아짐으로써 더욱 정교하게 오타자를 수정할 수 있다.
- [0091] 반면에, 복수의 후보군 키워드 중에서 품목사전에 포함된 키워드와 일치하는 키워드가 존재하지 않으면, 제2 아이템 세트에 포함된 해당 키워드를 미분류로 분류한다(S310).
- [0092] 즉, 제어부(180)는 제2 아이템 세트에 포함된 해당 키워드가 해당 품목사전에 존재하지 않는 것으로 판단한다.
- [0093] 예를 들어, 해당 키워드가 "CUUDE"이고 후보군 키워드가 "CUUDE", "CURDE" 및 "CURRDE"라고 가정한다.
- [0094] 그리고, 후보군 키워드 3개를 품목사전과 비교하여, 품목사전에 존재하는 키워드가 후보군 키워드에 포함되는지 비교하여, 품목사전에 "CUUDE", "CURDE" 및 "CURRDE"가 존재하지 않으면, 제어부(180)는 해당 키워드 "CUUDE"를 미분류 키워드로 판단하여 처리한다.
- [0095] 또한, 제어부(180)는 제2 아이템 세트에 포함된 해당 키워드가 해당 품목사전에 존재하지 않는 것으로 판단하면, 해당 키워드를 수정 없이 사용한다.
- [0096] 이와 같이 본 발명에 따르면, 기존의 방법보다 정확하게 오타자를 체크하고 수정할 수 있으며, 새로운 분류체계를 구축할 수 있다.
- [0097] 또한, 본 발명에 따르면 HS 코드 기반의 카테고리를 분류하는 것이 아니라, 품목 텍스트 자체를 이용하여 품목 키워드 사전을 제작하고 해당 사전을 기반으로 분류하므로, 품목 키워드에 오타자가 존재하더라도 더욱 정확하게 품목을 분류할 수 있다.
- [0098] 본 발명은 도면에 도시된 실시 예를 참고로 설명 되었으나 이는 예시적인 것이 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.

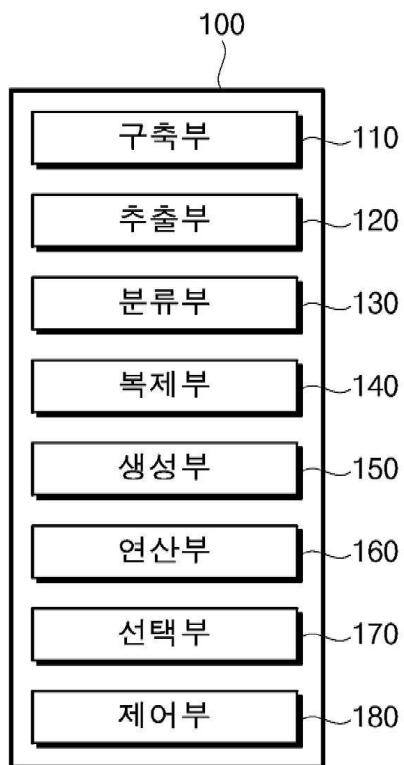
### 부호의 설명

[0099]

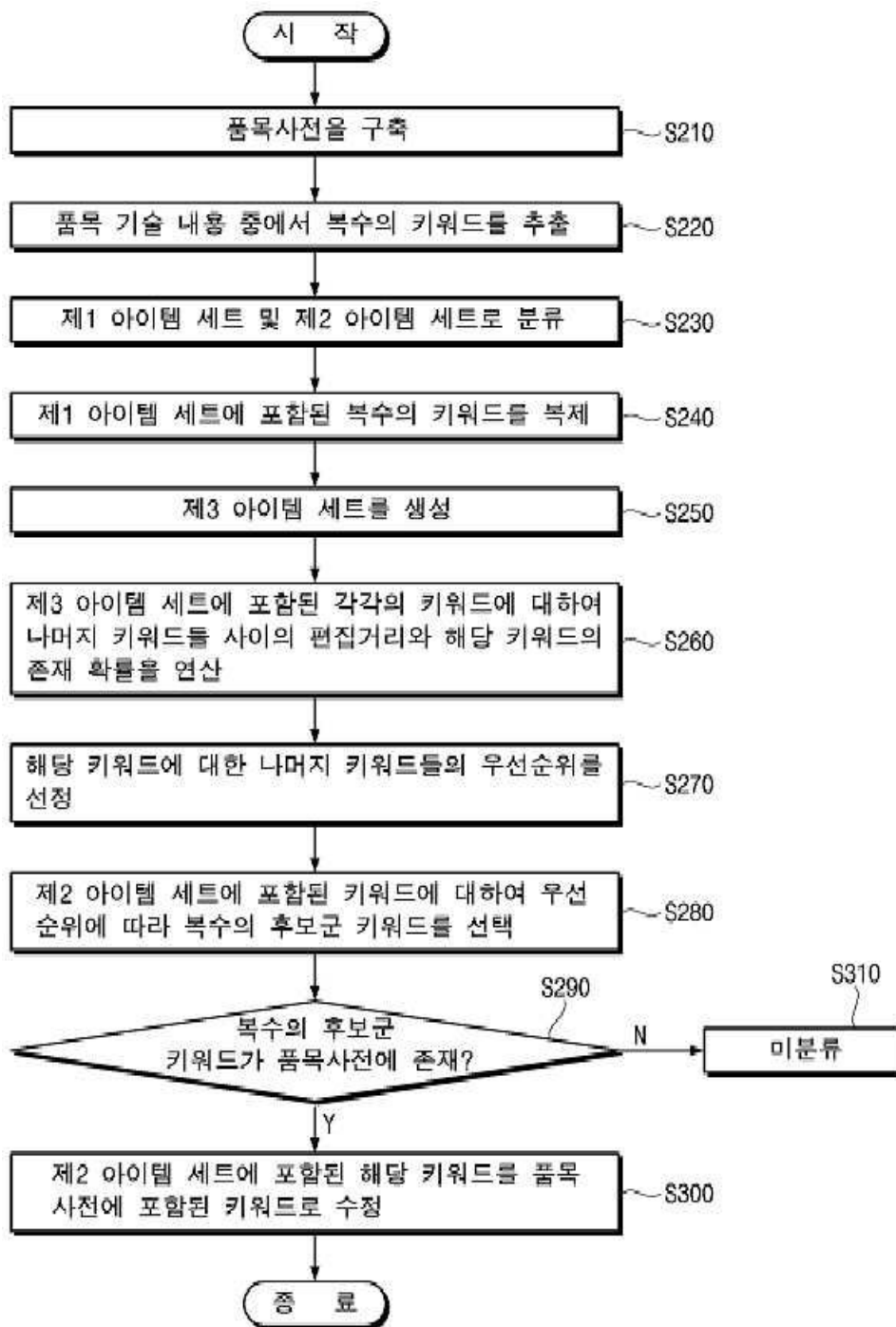
100: 품목명 자동 이탈자 탐지 및 수정 시스템,  
 110: 구축부, 120: 추출부,  
 130: 분류부, 140: 복제부,  
 150: 생성부, 160: 연산부,  
 170: 선택부, 180: 제어부

### 도면

도면1



도면2



도면3

| Category           | Subcategory        | Keywords                             |
|--------------------|--------------------|--------------------------------------|
| Crude Oil          | Crude Oil          | Crude Oil, Condensate, Arabian, etc. |
| Oil Products       | Gasoline           | Gasoline, PYGAS, etc.                |
|                    | Light Oil          | Gas Oil, Diesel, MGO, LCO, etc.      |
|                    | Asphalt            | Asphalt, Bitumen                     |
|                    | Naphtha            | Naphtha, Paraffin, Petroleum, etc.   |
|                    | Fuel Oil           | Fuel Oil, HSFO, LSFO, etc.           |
|                    | Heavy Oil          | Heavy Oil, Decant Oil, Bunker C      |
|                    | Jet Oil            | Jet                                  |
|                    | Kerosene           | Kerosene                             |
|                    | Base Oil           | Base Oil, Lubricating, etc.          |
|                    | Etc.               | Tudalen, Molten Sulphur, etc.        |
| Gas                | LPG Gas            | LPG, Propylene, Butane, etc.         |
|                    | LNG Gas            | LNG, Gas Condensate, etc.            |
| Chemical           | Chemical           | Paraxylene, Kokosol, Octene, etc.    |
| Plant & Animal Oil | Plant & Animal Oil | Parm Oil, Coconut Oil, etc.          |

도면4



도면5

