



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년08월19일
 (11) 등록번호 10-0977000
 (24) 등록일자 2010년08월13일

(51) Int. Cl.
G10L 15/22 (2006.01) *G06Q 50/00* (2006.01)
 (21) 출원번호 10-2008-0060941
 (22) 출원일자 2008년06월26일
 심사청구일자 2008년06월26일
 (65) 공개번호 10-2010-0001141
 (43) 공개일자 2010년01월06일
 (56) 선행기술조사문헌
 KR1020040054895 A
 JP10039894 A
 JP2006078829 A

(73) 특허권자
주식회사 에스피치
 서울 강남구 역삼동 770-9 성환빌딩 6층
 (72) 발명자
조주형
 서울특별시 서초구 양재동 7-47 204호
 (74) 대리인
김윤보

전체 청구항 수 : 총 12 항

심사관 : 정성윤

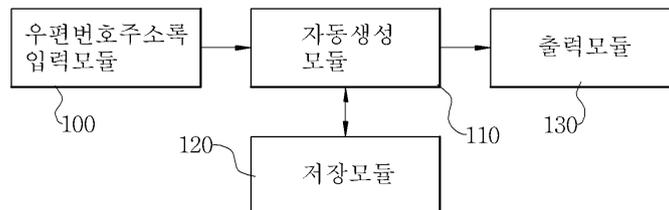
(54) 업데이트된 주소의 음성 인식 그래마 자동 생성 방법 및시스템

(57) 요약

본 발명은 그래마 자동 생성 방법 및 시스템에 관한 것이다. 보다 상세하게는, ARS 시스템과 같이 음성을 주요 매개로 정보를 전달하는 방법이나 시스템에 있어서, 사용자로부터 음성으로 녹음, 입력받은 정보, 특히 주소 정보를 분석하여 인식하기 위한 것으로서, 변경, 수정, 추가, 삭제된 주소 정보의 그래마 리스트를 자동으로 생성할 수 있도록 하고, 모듈화된 그래마 내용을 손쉽게 구성할 수 있도록 하며, 매 분기별로 업데이트 되는 주소 어휘를 자동으로 등록시킬 수 있도록 하는 그래마 자동 생성 방법 및 시스템에 관한 것이다.

본 발명은 새로운 주소 정보에 대한 그래마 리스트를 자동으로 생성할 수 있고, 모듈화된 그래마 내용을 손쉽게 구성할 수 있도록 하며, 매 분기별로 업데이트 되는 주소 어휘를 자동으로 등록시킬 수 있도록 하는 효과가 있다.

대표도 - 도1



특허청구의 범위

청구항 1

음성 주소인식 그래마 자동 생성 방법에 있어서,

- (1) 분기별로 업데이트 시킬 데이터를 추출하고, 통합하는 단계;
- (2) 표준 지역 주소와 유사 어휘로 이루어진 지역주소를 DB에 추가시키는 단계;
- (3) 발성 단계별로 주소 어휘의 단계(depth)를 구분하는 단계;
- (4) 단계별로 구분된 주소 어휘에 포함된 숫자 또는 특수문자 데이터를 전처리하는 단계;
- (5) 각 단계 내에 포함된 어휘만을 추출하고 중복 제거한 어휘군 리스트를 생성하는 단계;
- (6) 각 단계별 어휘군을 이용하여 하위군 최상위 그래마(top-grammar)로 구성되는 모듈별 어휘군을 구성하는 단계; 및
- (7) 단계별 어휘군, 모듈별 어휘군을 그래마 파일로 형성하는 단계;를 포함하여 이루어진 것을 특징으로 하는 업데이트된 주소의 음성 인식 그래마 자동 생성 방법.

청구항 2

제 1 항에 있어서,

상기 (3) 단계는:

하나의 지역 주소를 발성 단계(depth)별로 최대 5단계로 분리시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그래마 자동 생성 방법.

청구항 3

제 1 항에 있어서,

상기 (3) 단계에서, 주소어휘의 단계(depth)를 구분할 때 숫자를 포함하는 지역 단위 데이터에서 숫자를 분리시켜 인식가능한 형태로 변환시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그래마 자동 생성 방법.

청구항 4

제 1 항에 있어서,

상기 (4) 단계에서 특수문자를 처리할 때, 무규칙 데이터를 추출하여 삭제시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그래마 자동 생성 방법.

청구항 5

제 1 항에 있어서,

상기 (5) 단계에서, 단계(depth)별 어휘군을 구성할 때, 그래마 명, 그래마 내용 및 슬롯값을 자동으로 생성하는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그래마 자동 생성 방법.

청구항 6

제 1 항에 있어서,

상기 (6) 단계에서 생성된 그라마 파일은, 단계(depth)별, 모든 어휘 포함별, 지역별, 최상위 그라마(top grammar) 내용을 포함하여 구성되는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 방법.

청구항 7

그라마 자동 생성 시스템에 있어서,

주소 정보를 입력받기 위한 입력모듈;

상기 입력모듈을 통해 입력되는 주소 정보에 대해 분기별로 수정, 변경, 삭제될 데이터를 통합하고, 표준 지역 주소와 유사 어휘로 이루어진 지역주소를 추가시킨 후, 발성 단계별로 주소 데이터의 단계(depth)를 구분하고, 구분된 주소 데이터에 포함된 숫자 또는 특수문자를 처리하고, 각 단계(depth) 내에 포함된 어휘만을 추출하여 어휘군 리스트를 생성한 후 각 단계(depth)별 어휘군을 이용하여 하위군 최상위 그라마(top-grammar)를 구성하여 그라마 파일로 형성하는 자동생성 모듈;

상기 자동 생성 모듈의 제어에 의하여 상기 지역 주소의 유사어휘 및 그라마 파일을 DB 형태로 저장하고 있는 저장모듈; 및

상기 자동 생성 모듈의 제어에 의하여 상기 생성된 그라마 파일을 출력시키는 출력모듈;을 포함하여 이루어진 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 시스템.

청구항 8

제 7 항에 있어서,

상기 자동생성모듈은, 어휘군 리스트를 생성할 때 중복된 어휘는 제거시킨 상태에서 생성시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 시스템.

청구항 9

제 7 항에 있어서,

상기 자동생성 모듈은, 주소어휘의 단계(depth)를 구분할 때 숫자를 포함하는 지역 단위 데이터에서 숫자를 분리시켜 인식가능한 형태로 변환시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 시스템.

청구항 10

제 7 항에 있어서,

상기 자동생성 모듈은, 특수문자를 처리할 때, 무규칙 데이터를 추출하여 삭제시키는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 시스템.

청구항 11

제 7 항에 있어서,

상기 자동생성 모듈은, 단계(depth)별 어휘군을 구성할 때 그라마 명, 그라마 내용 및 슬롯값을 자동으로 생성하는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그라마 자동 생성 시스템.

청구항 12

제 7 항에 있어서,

상기 자동생성 모듈에 의해 생성된 그레마 파일은, 단계(depth)별, 모든 어휘 포함별, 지역별, 최상위 그레마(top grammar) 내용으로 구성되어 있는 것을 특징으로 하는 업데이트된 주소의 음성 인식 그레마 자동 생성 시스템.

명세서

발명의 상세한 설명

기술분야

- [0001] 본 발명은 그레마 자동 생성 방법 및 시스템에 관한 것이다.
- [0002] 보다 상세하게는, ARS 시스템과 같이 음성을 주요 매개로 정보를 전달하는 방법이나 시스템에 있어서, 사용자로부터 음성으로 녹음, 입력받은 정보, 특히 주소 정보를 분석하여 인식하기 위한 것으로서, 변경, 수정, 추가, 삭제된 주소 정보의 그레마 리스트를 자동으로 생성할 수 있도록 하고, 모듈화된 그레마 내용을 손쉽게 구성할 수 있도록 하며, 매 분기별로 업데이트 되는 주소 어휘를 자동으로 등록시킬 수 있도록 하는 그레마 자동 생성 방법 및 시스템에 관한 것이다.

배경기술

- [0003] 홈쇼핑, 통신회사, 대기업, 은행 등의 콜센터에서, 주문 정보나 기타 정보를 입력하기 위해, 자동 기록 시스템인 ARS를 많이 사용하고 있다. 이러한 자동 기록 시스템에서는 통화자의 주소를 입력받아야 할 일이 많은데, 이러한 단순한 작업을 위해 매번 콜센터 직원을 연결해서 직접 통화하여야 한다면, 이는 사업자는 물론 통화자의 입장에서도 매우 번거로운 일이며, 직원 확충에 따른 인건비나, 직원 연결을 위한 대기시간 증가 등의 문제점이 있었다.
- [0004] 예컨대, 업체에 사용자의 주소를 제공하거나, 기존 DB에 등록되어 있는 주소를 변경하고자 하는 경우, 사용자는 해당 DB를 이용할 수 있는 서비스 제공자와 전화가 연결된 상태에서 주소를 말해주고, 서비스 제공자가 상기 주소를 받아 적은 후 DB에 접속하여 DB에 기록되어 있는 주소 정보를 변경시키거나, 말해주는 주소를 바로 DB에 접속한 상태에서 DB에 기록되어 있는 주소 정보를 변경하는 방식을 사용한다.
- [0005] 이러한 방식은 실시간으로 받아쓰는 인력이 필요하고, 일일이 받아 적어야 하기 때문에 주소를 등록하는데 많이 시간이 소요되거나 입력하는 사람이 잘못 들을 경우 주소가 잘못 기록될 수 있다는 문제점이 있다.
- [0006] 이를 극복하기 위해, 사용자의 음성에 따른 주소를 우선 녹음하고, 서비스 제공자가 녹음된 파일을 일괄적으로 받아 적은 후에 이를 분석하여, DB에 수동 입력하는 방식을 사용하고 있으나, 이 역시 인력이 하나하나 단순 반복 업무를 해야 한다는 점에서, 위의 문제를 원천적으로 해결해주고 있지는 못하다.
- [0007] 따라서 녹음되는 주소를 자동적으로 인식하고 DB화 할 수 있는 음성인식 시스템이 요구되고 있으며, 사용자가 시스템에 불러 준 주소 정보를 인식하여 DB에서 자동으로 주소를 찾아내고 매칭시키는 방식이 도입될 필요성이 있다. 이 중 후자의 자동 매칭 방식을 사용하는 경우, 사람이 일일이 받아 적어야 하는 전사량을 줄여 비용을 절감시킬 수 있고, 받아 적는 인력을 줄일 수 있으며, 받아 적는 시간을 단축할 수 있는 장점이 있다.
- [0008] 이러한 장점이 있는 자동 매칭 방식을 사용하기 위해서는, 우리나라의 주소 체계에 대한 발성을 제대로 인식하여야 하며, 자주 변하는 주소 체계를 인식하여 DB가 계속적으로 업데이트될 수 있어야 한다는 전제 조건이 따르게 된다. 일정 기간마다 추가, 변경 또는 삭제된 주소 정보를 전달받아서, 이를 자동으로 DB에 업데이트시킬 수 있어야 하는데, 이를 단계별 모듈화 과정을 통해 시스템에서 인식하기 용이한 형태로 재편하여 자동 등록시킬 수 있어야, 상기한 음성 인식 시스템의 효과를 극대화할 수 있게 된다.

발명의 내용

해결하고자하는 과제

- [0009] 본 발명은 상기와 같은 종래 기술의 문제점을 음성인식 방법을 통해 해소시키기 위해 안출된 것으로, 음성 주소

를 자동으로 인식하는 주소 음성인식 방법, 시스템에 있어서, 새로이 추가, 수정, 변경, 삭제된 주소 체계나 정보에 대한 주소인식 그래마 리스트를 자동으로 생성할 수 있고, 모듈화된 그래마 내용을 손쉽게 구성할 수 있도록 하며, 매 분기별로 업데이트되는 주소 어휘를 자동으로 등록시킬 수 있도록 하는 음성 주소인식 그래마 자동 생성 방법 및 시스템을 제공하는 것을 목적으로 한다.

과제 해결수단

- [0010] 상기와 같은 목적을 달성하기 위한 본 발명은, 음성 주소인식 그래마 자동 생성 방법에 있어서, (1) 분기별로 업데이트 시킬 데이터를 추출하고, 통합하는 단계; (2) 표준 지역 주소와 유사 어휘로 이루어진 지역주소를 DB에 추가시키는 단계; (3) 발성 단계별로 주소 어휘의 단계(depth)를 구분하는 단계; (4) 단계별로 구분된 주소 어휘에 포함된 숫자 또는 특수문자 데이터를 전처리하는 단계; (5) 각 단계 내에 포함된 어휘만을 추출하고 중복 제거한 어휘군 리스트를 생성하는 단계; (6) 각 단계별 어휘군을 이용하여 하위군 최상위 그래마(top-grammar)로 구성되는 모듈별 어휘군을 구성하는 단계; 및 (7) 단계별 어휘군, 모듈별 어휘군을 그래마 파일로 형성하는 단계;를 포함하여 이루어진 것을 특징으로 하는 음성 주소인식 그래마 자동 생성 방법을 제공한다.
- [0011] 상기 (3) 단계는: 하나의 지역 주소를 발성 단계(depth)별로 5단계로 분리시키는 것을 특징으로 한다.
- [0012] 상기 (3) 단계에서, 주소어휘의 단계(depth)를 구분할 때 숫자를 포함하는 지역 단위 데이터에서 숫자를 분리시켜 인식가능한 형태로 변환시키는 것을 특징으로 한다.
- [0013] 상기 (4) 단계에서 특수문자를 처리할 때, 무규칙 데이터를 추출하여 삭제시키는 것을 특징으로 한다.
- [0014] 상기 (5) 단계에서, 단계(depth)별 어휘군을 구성할 때, 그래마 명, 그래마 내용 및 슬롯값을 자동으로 생성하는 것을 특징으로 한다.
- [0015] 상기 (6) 단계에서 생성된 그래마 파일은, 단계(depth)별, 모든 어휘 포함별, 지역별, 최상위 그래마(top grammar) 내용을 포함하여 구성되는 것을 특징으로 한다.
- [0016] 또한, 그래마 자동 생성 시스템에 있어서, 주소 정보를 입력받기 위한 입력모듈; 상기 입력모듈을 통해 입력되는 주소 정보에 대해 분기별로 수정, 변경, 삭제될 데이터를 통합하고, 표준 지역 주소와 유사 어휘로 이루어진 지역주소를 추가시킨 후, 발성 단계별로 주소 데이터의 단계(depth)를 구분하고, 구분된 주소 데이터에 포함된 숫자 또는 특수문자를 처리하고, 각 단계(depth) 내에 포함된 어휘만을 추출하여 어휘군 리스트를 생성한 후 각 단계(depth)별 어휘군을 이용하여 하위군 최상위 그래마(top-grammar)를 구성하여 그래마 파일로 형성하는 자동 생성 모듈; 상기 자동 생성 모듈의 제어에 의하여 상기 지역 주소의 유사어휘 및 그래마 파일을 DB 형태로 저장하고 있는 저장모듈; 및 상기 자동 생성 모듈의 제어에 의하여 상기 생성된 그래마 파일을 출력시키는 출력모듈;을 포함하여 이루어진 것을 특징으로 하는 음성 주소인식 그래마 자동 생성 시스템을 제공한다.
- [0017] 상기 자동생성모듈은, 어휘군 리스트를 생성할 때 중복된 어휘는 제거시킨 상태에서 생성시키는 것을 특징으로 한다.
- [0018] 상기 자동생성 모듈은, 주소어휘의 단계(depth)를 구분할 때 숫자, 동 데이터를 분리시키거나, 숫자를 포함하는 지역 단위 데이터에서 숫자를 분리시켜 인식가능한 형태로 변환시키는 것을 특징으로 한다.
- [0019] 상기 자동생성 모듈은, 특수문자를 처리할 때, 무규칙 데이터를 추출하여 삭제시키는 것을 특징으로 한다.
- [0020] 상기 자동생성 모듈은, 단계(depth)별 어휘군을 구성할 때 그래마 명, 그래마 내용 및 슬롯값을 자동으로 생성하는 것을 특징으로 한다.
- [0021] 상기 자동생성 모듈에 의해 생성된 그래마 파일은, 단계(depth)별, 모든 어휘 포함별, 지역별, 최상위 그래마(top grammar) 내용으로 구성되어 있는 것을 특징으로 한다.

효과

- [0022] 이상 설명한 바와 같이, 본 발명은 새로운 주소 정보를 다운로드하여 그래마 리스트를 자동으로 생성할 수 있고, 모듈화된 그래마 내용을 손쉽게 구성할 수 있도록 하며, 매 분기별로 업데이트 되는 주소 어휘를 자동으로 등록시킬 수 있도록 하는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- [0023] 이하, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 이해할 수 있을 정도로 상세히 설명하며, 이 발명의 바람직한 실시 예를 첨부된 도면을 참조로 설명하기로 한다.
- [0024] 통상 음성 입출력 기반의 음성인식 기술을 통해 음성서비스를 제공하기 위해 GRXML(GRAMMAR eXtensible Markup Language), GSL(GRAMMAR Specification Language), NGO(Nuance Grammar Object)와 같은 인식 기술을 사용할 수 있는데, 이러한 스펙의 그래마 태그는 음성인식 그래마를 설정한다. 음성인식 그래마는 사용자가 말할 수 있는 후보 어휘를 의미하며, 음성인식은 사용자가 발화한 음성과 가장 비슷한 후보 어휘를 그래마 중에서 찾는 것을 의미한다. 이와 같은 방법은 사용자가 말할 수 있는 어휘에 제한을 둘 수 있을 때 주로 사용되며, 사용자가 어떤 어휘를 발화할지 예상할 수 없는 상황에서는 사용이 어렵다.
- [0025] 주소 음성 인식의 경우, 어휘의 종류가 한정되어 있기 때문에, 이러한 그래마 태그의 사용이 가능하다. 현재 우리나라의 우편주소 체계 현황을 살펴보면, 15개의 시/도로 구분되어 있으며, 6 단계(Depth) 정도의 지역단위 발성 조합으로 구성된다. 예를 들면 "도-군-읍-리-가-개별주소"로 구성된다. 정확하고 효율적인 우리나라 주소발성 어휘의 음성인식을 위해서는 다음과 같은 그래마 구성요소가 필요하다.
- [0026] 첫째로는, 지역별 그래마 구분이 필요하다. 즉, 시, 도 단위로 그래마를 분리할 필요가 있다. 왜냐하면 현 우편주소 체계에서는 "경기 과천시 갈현동", "서울 은평구 갈현동", "인천 계양구 갈현동" 등과 같이 다른 지역 내 같은 이름의 지역단위가 존재하기 때문에, "경기도 과천시 갈현동"은 "경기도" 주소 그래마 내에서 인식될 필요가 있다.
- [0027] 둘째로는, 인식 그래마의 범위를 축소할 수 있어야 한다. 즉 5단계 발성 단계별로 그래마를 분리할 필요가 있다. 예컨대, "서울시 성북구 동선동4가 13번지..." 의 발성의 경우, "서울시", "성북구", "동선동4가"의 3단계 지역단위 어휘 조합만을 포함하는 그래마로 인식되도록 한다.
- [0028] 셋째로는, 원 슬롯(one slot) 구성으로 한다. 5단계 발성 그래마 구성시, 각 단계를 멀티 슬롯으로 구성함에 비해, 원 슬롯으로 구성시 트라이그램(trigram)률이 적용되어 인식율이 배가된다. 예컨대, 원슬롯 구성으로 인해 "용산구 갈월동"이 "용산구 갈현동"으로 오인식되는 것을 방지하게 되는데, 이는 멀티슬롯으로 구현시 "용산구 갈현동"과 같은 어휘 조합이 나타날 수 있으나, 용산구에는 갈현동이 없기 때문이다.
- [0029] 이러한 필수 요소를 기반으로, 본 발명의 구성을 설명하도록 한다. 본 발명은 새로이 개편되는 업데이트 정보, 즉, 변경되거나, 추가되거나, 삭제되는 주소록 정보를 입력받아, 이를 그래마로 구성하는 방법 및 시스템에 관한 것이다.
- [0030] 본 발명이 적용된 그래마 자동 생성 시스템은 도 1에 도시된 바와 같이, 입력모듈(100), 자료생성모듈(110), 저장모듈(120), 출력모듈(130)을 포함한다.
- [0031] 입력모듈(100)은 우편번호, 주소록 정보 등을 입력받기 위한 것이다
- [0032] 자료생성모듈(110)은 입력모듈(100)을 통해 입력되는 우편번호, 주소록 정보에 대해 분기별로 수정, 변경, 삭제될 데이터를 통합하고, 표준 지역 주소와 유사 어휘로 이루어진 지역주소를 추가시킨 후, 발성 단계별로 주소 데이터의 단계(depth)를 구분하고, 구분된 주소 데이터에 포함된 숫자 또는 특수문자를 처리하고, 각 단계(depth) 내에 포함된 어휘만을 추출하여 어휘군 리스트를 생성한 후 각 단계(depth)별 어휘군을 이용하여 하위군 최상위 그래마(top-grammar)를 구성하여 그래마 파일로 형성하는 기능을 한다.
- [0033] 저장모듈(120)은 자동생성모듈(110)과, 상기 자동생성모듈(110)의 제어에 의하여 상기 지역 주소의 유사어휘 및 그래마 파일을 DB 형태로 저장한다.
- [0034] 출력모듈(130)은 상기 자동생성모듈(110)의 제어에 의하여 상기 생성된 그래마 파일을 출력시키는 기능을 한다.
- [0035] 자동생성모듈(110)은 주소어휘의 단계(depth)를 구분할 때 숫자, 동 데이터를 분리시킬 수 있으며, 특수문자를

처리할 때, 무규칙 데이터를 추출하여 삭제시킬 수 있다. 또한, 자동생성모듈(110)은 단계(depth)별 어휘군을 구성할 때 그래마 명, 그래마 내용 및 인식 결과인 슬롯값을 자동으로 생성한다.

[0036] 상기와 같이 구성된 그래마 자동 생성 방법에 대해 설명하면 다음과 같다. 도 1 및 도 2를 참조하여 설명하면, 자동생성모듈(110)은 분기별로 업데이트 시킬 데이터, 즉 수정 또는 변경 또는 삭제시킬 주소 데이터를 추출하고, 통합한다(S100).

[0037] 그리고 자동생성모듈(110)은 표준 지역 주소에 대해, 지역적 또는 역사적인 이유로 발생되어 통계적으로, 실제로 자주 사용하는 지역 주소를 추가시킨다(S110). 예를 들어 표준 지역 주소는 "광주광역시"지만, "광주직할시"와 같은 유사어휘 데이터를 저장모듈(120)의 해당 DB에 추가시킨다.

[0038] 자동생성모듈(110)은 주소어휘단계(depth)를 구분하게 되는데, 발생 단계별 주소 데이터의 단계(depth)를 구분한다. 아울러 숫자와 동 데이터를 분리한다. "상수1동을 "상수", "동", "1"로 구분하는 것과 같이 숫자, 동 데이터를 분리하여 발생 단계별로 주소 어휘의 단계(depth)를 구분한다(S120).

[0039] 이 때 단계(depth)와 관련하여 도 3을 참조하여 설명하면, 우리나라의 주소 구조에서 지역단위는 마지막 단계를 제외하고, 최대 상위 5단계(depth)의 발생을 정형화하여 사용한다. 그 내용을 5단계 발생위치별로 분리한 뒤, 각 단계에서 발생 가능한 한 지역단위를 탐-다운방식 조합으로 그래마군을 단계(depth)별 그래마군으로 지정한다.

[0040] 첫 번째 단계(1 depth)로, 경기도, 서울시와 같이 가장 큰 분류를 지정하고, 두 번째 단계(2 depth)로, 첫 번째 단계 부속 주소, 예컨대 경기-수원, 경기-성남을 지정하고, 세 번째로 한 단계 더 작은 부속 주소를 설정하는 식이다.

[0041] 이렇게 결정된 그래마 구성은 그래마 범위를 축소할 수 있게 해준다. 이때 탐다운 방식에 따른 단계별 그래마 구성은, 중간 단어의 생략정도 포함하도록 한다. "경북 포항시 북구 청하면" 그래마는 "경북 북구 청하면" 그래마도 포함하도록 한다. 그리고 각 단계별 그래마는 해당 단계의 끝 어휘를 반드시 포함한다. 이렇게 구성된 단계(depth)별 어휘의 구성 예제는 다음과 같다. "경북 포항시 북구 청하면 서정리"와 같이 발생 단계별로 주소 어휘를 단계별로 구분한다.

[0042]

구분	내용	예제
1 Depth	레벨(01) 내용만을 포함하는 어휘군	경북
2 Depth	레벨(01+02) 내용을 포함	경북 포항시
3 Depth	레벨(01+02+03) 내용을 포함	경북 포항시 북구
4 Depth	레벨(01+02+03+04) 내용을 포함	경북 포항시 북구 청하면
5 Depth	레벨(01+02+03+04+05) 내용을 포함	경북 포항시 북구 청하면 서정리

[0043] 즉, ~~아파트, ~~빌딩과 같은, 세부 주소를 제외하고, 상위 5개 단계로 주소 발생을 구분하여 이것을 위주로 사용하게 된다. 자동생성모듈(110)은 주소 어휘의 단계(depth)를 상기와 같이 구분한 후, 구분된 주소 어휘에 포함된 숫자 또는 특수문자 데이터 등에 대해 전처리가 이루어지도록 하고 무규칙 데이터는 삭제시킨다(S130). 즉, "1동", "2동", "1.3동" 등과 같은 데이터는 인식 가능한 형태로 처리하고, "**토지 내", "*블록*구획" 등과 같은 무규칙 데이터는 삭제시킨다.

[0044] 그리고 자동생성모듈(110)은 각 단계(depth) 내에 포함된 어휘만을 추출하고 중복 제거한 어휘군 리스트를 생성하는 한편, 그래마명, 그래마 내용 및 인식 결과인 슬롯값(return value)을 자동으로 생성시킨다(S140).

[0045] 그리고 자동 생성 모듈(110)은 상기 S140 단계에서 생성된 각 단계(depth)별 어휘군을 이용하여 하위군 최상위 그래마(top-grammar)로 구성되는 모듈별 어휘군을 구성(S150)하고, 모듈별 어휘군을 구성한 후 그래마 파일로 형성(S160)한다.

[0046] 상기 S150 단계에서 모듈별 어휘군을 구성할 때에도 단계(depth)별 어휘군을 구성할 때와 마찬가지로 그래마명, 그래마 내용 및 슬롯값(return value)을 자동으로 생성시킨다.

[0047] 모듈별 어휘군 구성은, 단계별 그래마군의 해당 단계의 어휘로 시작하여, 그 뒷단계 어휘를 미들다운(middle-down) 방식으로 조합한 내용의 그래마 군이다. 도 4를 참조하면, 모듈3은 3단계(depth) 인 경북-포항-북구로 시

작하여, 그 하부의 흥해읍, 청하면, 남송리, 마산리 등을 조합한 것을 하나의 모듈로 한다. 모듈4는 4단계인 경부-포항-북구-흥해읍으로 시작하여, 그 하부의 남송리, 마산리 등을 조합한 것이다. 본 발명에서는 3단계부터 모듈을 구성하게 되며, 1, 2단계에서는 모듈을 형성하지 않는다.

[0048] 즉, 모듈별 그래마 구성은 이전 단계에서 인식된 내용을 앞단에 반드시 포함한다. 예컨대 3단계(depth)를 이용, "부산-금정구-부곡일동"을 인식 한후에, 모듈3그래마는, 부산-금정구-부곡일동 삼팔팔 다시 이십구번지와 같이 3단계 내용을 앞단에 포함하고 4단계에 적용될 어휘군을 형성한다. 모듈별로 그래마를 구성하므로 그래마 범위의 축소를 가져온다. 모듈별 그래마 구성과 그에 따른 선택을 모듈 순으로 나타내면 아래의 표와 같으며, L01, L02로 표시된 것은 각 주소별 단계를 의미한다.

[0049]

구분	내용	예제
모듈2	L01_L02+(ex) 포함	경북 포항시 북구 청하면 서정리... 남구 동해면 발산리...
모듈3	L01_L02_L03+(ex) 포함	경북 포항시 북구 청하면 서정리... 흥해읍 남송리...
모듈4	L01_L02_L03_L04+(ex) 포함	경북 포항시 북구 청하면 서정리... 월포리...
모듈5	L01_L02_L03_L04_L05+(ex) 포함	경북 포항시 북구 청하면 서정리 A빌딩 B아파트

[0050] 본 발명은 3단계부터 모듈을 구성하여 사용하므로, 상기 모듈2는 사용하지 아니하며, 모듈3, 4, 5만을 사용한다. 모듈 3을 보면, "북구 청하면~, 북구 흥해읍~" 등의 모듈 어휘군을 갖게 된다. 업데이트되어 제공된 정보는 상기와 같은 과정과 형태를 따라서, 각 단계 및 모듈 형태로 재편되고, 그래마 파일로 저장된다. 이 때 업데이트된 정보의 형태가 기존 주소에서 추가되는 것인지(추가형), 또는 삭제하는 것인지(삭제형)를 같이 판단하도록 한다.

[0051] 그리고 상기 S160 단계에서 생성된 그래마 파일(*.grammer)은 단계(depth)별, 모든 어휘 포함별, 지역별, 최상위 그래마(top grammar) 내용을 포함하고 있다.

[0052] 자동생성모듈(110)에 의해 생성된 그래마 군은 하기 표와 같다.

[0053]

그래마군 종류	발생 가능한 최대 그래마군 개수 $O(n)$	설명	파일군 내용
단계(depth)별 그래마군	5	모든 지역의 내용을 포함하는 각 단계(depth)별 그래마군	Depth01.h Depth02.h,
	$5*15$	각 15개 지역별로 구분된 각 단계(depth)별 그래마군	Depth01_강원.h...
	1	모든 단계(depth)별 내용을 포함하는 그래마군	A.grammar
	$1*15$	모든 단계(depth)별 내용을 포함하고, 지역별로 구분된 그래마군	A.grammar
모듈별 그래마군	$3*n$	모든 지역의 내용을 포함하는 각 모듈별 그래마군	A.grammar
	$3*n*15$	각 지역별로 구분된 각 모듈별 그래마군	Module03.h Module04.h
	※ n= 발생가능하고 중복 제거된 어휘의 조합 개수		

[0054] 한편, 업데이트 되는 내용에 따라 자동생성모듈(110)은 다음과 같이 동작한다. 지명이 변경되었거나, 동수가 변경되었거나, 우편번호, 동수가 변경된 경우에는 그래마의 어휘를 변경시킨다. 우편번호가 추가되는 경우에는 그래마 어휘를 추가하며, 우편번호가 삭제되는 경우에는 그래마 어휘를 삭제하되, 일정 기간 그래마 어휘를 유지하도록 조정하며, 유사 데이터와 중복되는지 확인하도록 한다.

[0055]

변경 내용	예제	향후 업데이트 방향	데이터 처리 형태
지명 변경	동해지방해양수산청 → 동해지방해양항만청	그래마 어휘 변경	추가형
동수 변경	마두2동 KT고양지사 → 마두1동 KT고양지사	그래마 어휘 변경	추가형
우편번호, 동수 변경	수원시 팔달구 인계동 중부지방 국세청 → 수원시 장안구 과장동 중부지방 국세청	그래마 어휘 변경	추가형
우편번호 추가	경기 고양시 일산동구 장항2동 SK엠티아파트	그래마 어휘 추가	추가형
우편번호 삭제	국민연금관리공단 고양 지사	그래마 어휘 삭제	삭제형

[0056]

위와 같이 새로 업데이트 된 주소는, 기존의 지명, 동수, 우편번호가 변경되거나 추가되거나 삭제되는 것인데, 이를 판단하여 데이터 처리 형태가 추가형인지, 삭제형인지 판단한다.

[0057]

도 5는 주소 업데이트 순서를 나타내는 순서도로서, 이를 참조하면, 업데이트 된 주소를 단계, 모듈별로 재편한 후, 그래마 파일을 형성한 후에, 추가형 데이터가 있는지, 삭제형 데이터가 있는지 확인한다.

[0058]

분기별로 주소 업데이트 정보에 대한 그래마 파일을 수신(S200)하고, 수신된 자료의 처리 형태별로, 추가형 데이터가 있는 경우(S210)에는 기존 데이터와 중복된 것이 있는지 확인하고(S220), 이를 체크하며 데이터를 추가시킨다(S230).

[0059]

또한 삭제형 데이터가 있는 경우(S240), 이를 체크(S250)하나 바로 삭제하지는 않도록 한다. 왜냐하면 사용자들은 주소가 변경된다고 해서 반드시 변경된 주소만을 말하지 않고 일정 기간 동안은 이전 주소를 동시에 말하기 때문이다. 이후에는 이전 체크 데이터의 체크 기간을 확인한 후(S260), 일정 기간이 경과한 데이터에 대해서만 일괄적으로 삭제하도록 한다(S270).

[0060]

도 6은 본 발명에 의해 자동 생성된 파일의 내용 구조를 나타내는 도면이다. 도 6을 참조하면, "#include 상위 단계 그래마 파일"은 파일간 참조시 참조파일이 명시되는 부분이고, Depth/Module 그래마명 이하의 부분은, 해당 그래마 군의 조합 내용 및 인식결과 내용, 슬롯값을 포함한다. 조합 내용은 상세 그래마명으로 구성되며, 상세 그래마명의 정의 및 어휘 내용은 해당 그래마 파일의 하단에 기술된다. 인식결과 내용은 인식 주소내용 및 어휘 검색 그래마명의 조합으로 구성된다. 인식결과 내용에 다음 하위단계 그래마명이 포함됨으로써, 인식 후 다른 검색절차 없이 바로 해당 하위단계 모듈로 검색이 가능하다.

[0061]

"상세그래마명" 이하의 부분은, 그래마군의 조합내용을 구성하는 상세그래마명 및 그 내용이 정의된다. 여기서 정의된 상세그래마명 및 그 내용은 해당단계의 그래마 군에서 뿐 아니라 하위단계의 또 다른 그래마군에서도 역시 이 내용을 참조하게 된다.

[0062]

이와 같이 그래마를 자동생성하는 방법 및 프로세스에 따라, 보다 손쉽게 그래마를 생성할 수 있게 되는데, 각 지역별로 그래마를 구분하는 방식에 따르며, 기존의 인식율에 비해 높은 수치를 나타내게 된다. 이는 그래마로 단계 지워진 지역 이하에 존재하는 주소의 어휘만을 갖도록 범위를 제한하여, 실제 존재하지 않는 어휘로의 오인식이 줄어들기 때문이다.

[0063]

이상, 본 발명의 바람직한 실시 예에 대하여 상세히 기술하였지만, 본 발명이 속하는 기술 분야에 있어서 통상

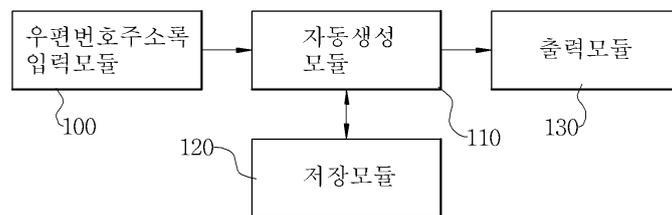
의 지식을 가진 사람이라면, 첨부된 청구 범위에 정의된 본 발명의 정신 및 범위를 벗어나지 않으면서 본 발명을 여러 가지로 변형 또는, 변경하여 실시할 수 있음을 알 수 있을 것이다.

도면의 간단한 설명

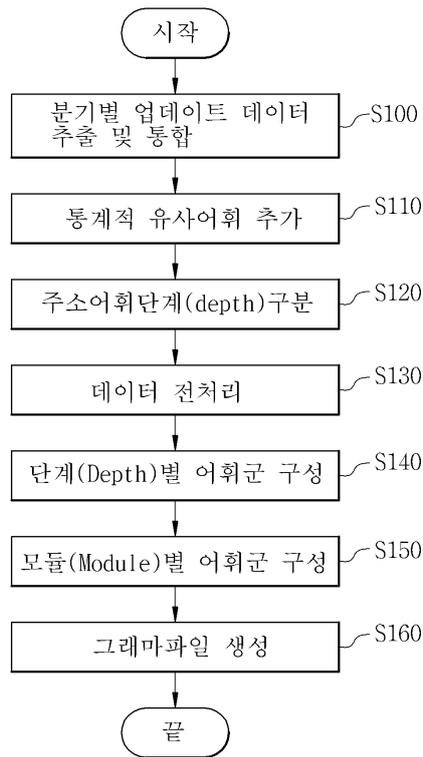
- [0064] 도 1은 본 발명에 따른 그레마 자동 생성 방법을 설명하기 위한 도면이다.
- [0065] 도 2는 본 발명에 따른 그레마 자동 생성 시스템을 설명하기 위한 도면이다.
- [0066] 도 3은 본 발명에 적용된 단계별 그레마의 구성을 설명하기 위한 도면이다.
- [0067] 도 4는 본 발명에 적용된 모듈별 그레마의 구성을 설명하기 위한 도면이다.
- [0068] 도 5는 본 발명에 적용된 업데이트 반영 순서를 나타내는 순서도이다.
- [0069] 도 6은 본 발명에 의해 자동 생성된 파일의 내용 구조를 나타내는 도면이다.
- [0070] <도면의 주요 부분에 대한 부호 설명>
- [0071] 100 : 우편번호 주소록 입력모듈
- [0072] 110 : 자동생성모듈
- [0073] 120 : 저장모듈
- [0074] 130 : 출력모듈

도면

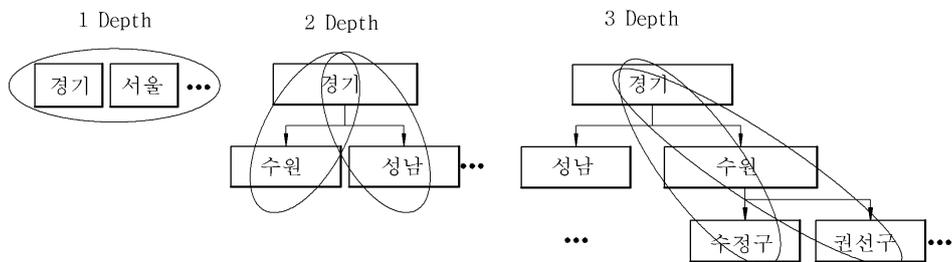
도면1



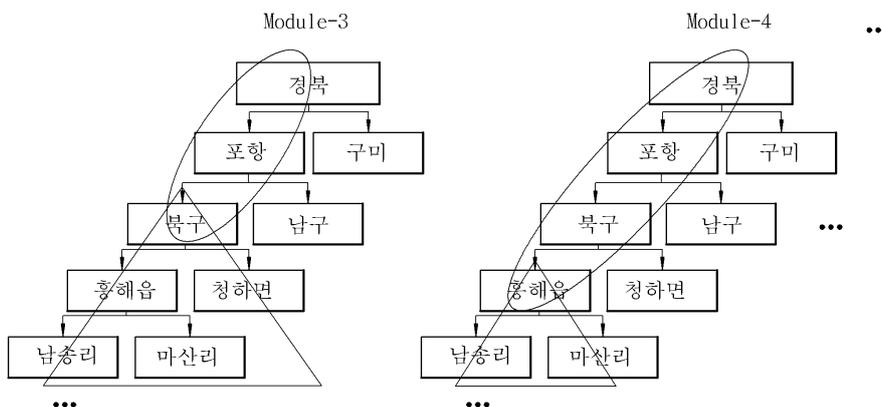
도면2



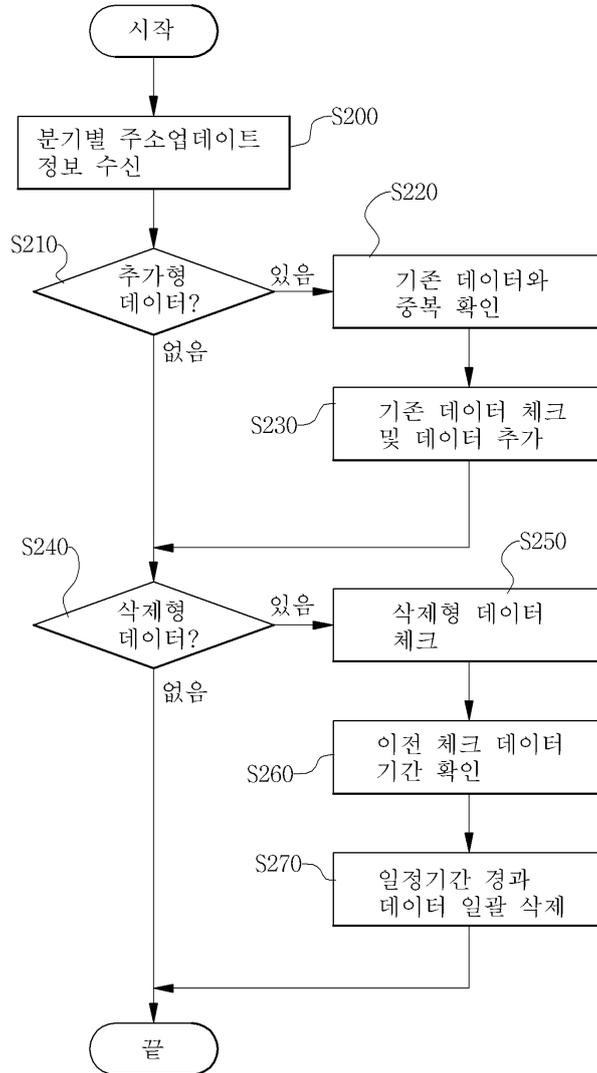
도면3



도면4



도면5



도면6

```

#include 상위단계 그램마 파일

Depth/Module 그램마명
[ (?전단계그램마명 상세그램마명)
{return
("해당 인식결과#하위module그램마명")}
...
]

상세그램마명
[
주소어휘
] ..
    
```