

(19) 대한민국특허청(KR) (12) 공개특허공보(A)

(51) Int. Cl.⁷
G06F 17/21

(11) 공개번호 특2001-0051459
(43) 공개일자 2001년06월25일

(21) 출원번호	10-2000-0065393
(22) 출원일자	2000년11월04일
(30) 우선권주장	99-322113 1999년11월12일 일본(JP)
(71) 출원인	가부시키가이샤 히타치세이사쿠쇼 가나이 쓰토무
(72) 발명자	일본 도쿄도 치요다쿠 간다스루가다이 4쵸메 6반치 이케다히사시 일본국도쿄도치요다쿠마루노우치1쵸메5반1고신마루비루가부시키가이샤히타 치세이사쿠쇼치테키쇼유크혼부나이 카토아쓰시 일본국오와리아사히시히가시인바초이치리야마3300-1 후지사와히로미치 일본국도쿄도치요다쿠마루노우치1쵸메5반1고신마루비루가부시키가이샤히타 치세이사쿠쇼치테키쇼유크혼부나이 코가마사시 일본국도쿄도치요다쿠마루노우치1쵸메5반1고신마루비루가부시키가이샤히타 치세이사쿠쇼치테키쇼유크혼부나이 후루카와나오히로 일본국도쿄도치요다쿠마루노우치1쵸메5반1고신마루비루가부시키가이샤히타 치세이사쿠쇼치테키쇼유크혼부나이
(74) 대리인	특허법인 원전 임석재

심사청구 : 없음

(54) 지명표기사전 작성방법 및 지명표기사전 작성장치

요약

지명표기에 대한 이표기의 생성이나 추가를 특별하게 행하지 않고, 이들을 가능한 한 자동적으로 추가함과 동시에, 자동적으로 추가할 수 없었던 표기나 불필요한 표기의 추가나 삭제를 2회째 이후는 자동적으로 행하는 지명표기사전 작성방법 및 장치. 지명표기를 문맥자유문법에 의해 표현하고, 지명표기의 문자열과 문맥자유문법에 있어서 구문카테고리를 대응하는 지식, 구문카테고리 및 구문카테고리의 배열을 별도의 구문카테고리에 대응하는 지식을 가지므로써 지명표기의 이표기를 추가할 수 있도록 한다. 또한 추가 또는 삭제되는 표기만을 지명표기사전을 기술하는 형식과 동일한 문맥자유문법에 의해 기술하고 파일로 출력하므로써, 그것 이후의 지명표기사전작성의 경우에 상기 추가, 삭제처리에 의한 변경을 자동적으로 반영시킨다.

대표도

도1

명세서

도면의 간단한 설명

- 도 1은 실시예의 지명사전 작성처리의 전체의 흐름을 나타내는 구성도.
- 도 2는 동일한 지명표기 리스트에 있어서 도도부현명(都道府縣名)을 나타내는 단어테이블에.
- 도 3은 표준적인 지명표기 리스트에 있어서 시구군정촌명(市區郡町村名)을 나타내는 단어테이블에.
- 도 4는 표준적인 지명표기 리스트에 있어서 정명(町名)·대자명(大字名)을 나타내는 단어테이블의 예.
- 도 5는 표준적인 지명표기 리스트에 있어서 자명(字名)을 나타내는 단어테이블의 예.
- 도 6은 표준적인 지명표기 리스트에 있어서 지명표기를 나타내는 코드테이블의 예.
- 도 7은 지명표기의 문맥자유문법에 의한 기술예.

- 도 8은 지명표기의 형식언어표현으로의 변환처리의 흐름도.
- 도 9는 지명표기 전체의 이표기 추가전의 정의파일의 예.
- 도 10은 지명단어 정의파일 작성처리에(예 1)
- 도 11은 시구군정촌(市區郡町村) 필드의 단어분할처리의 흐름도.
- 도 12는 지명단어 정의파일 작성처리에(예 2).
- 도 13은 지명단어 정의파일 작성처리에(예 3).
- 도 14는 지명표기 정의파일 작성처리에(예 1).
- 도 15는 지명표기 정의파일 작성처리의 흐름도.
- 도 16은 지명표기 정의파일 작성처리에(예 2).
- 도 17은 지식에 의한 이표기의 추가처리의 흐름도.
- 도 18A, 18'B는 지식에 의한 문자열의 구문카테고리로의 치환의 예.
- 도 19는 공통문자열의 구문카테고리로의 변환의 예.
- 도 20A, 20B, 20C는 지식에 의한 지명표기 정의로의 이표기의 추가의 예.
- 도 21은 지명단어 정의로의 이표기 추가를 위한 지식의 예.
- 도 22는 지명표기 정의로의 이표기 추가를 위한 지식의 예.
- 도 23A, 23B, 23C는 수작업에 의한 지명표기의 이표기의 추가를 위한 사용자 인터페이스(user interface) 예.
- 도 24A, 24B, 24C는 동일하고, 수작업에 의한 지명단어의 추가를 위한 사용자 인터페이스(user interface) 예.
- 도 25는 상황의존 이표기 추가·삭제처리에 있어서 이표기 추가의 흐름.
- 도 26은 상황의존 이표기 추가·삭제처리에 있어서 이표기 삭제의 흐름도.
- 도 27은 추가, 수정 이표기 정의의 예.
- 도 28은 사례에 의한 이표기·삭제처리의 흐름도.
- 도 29는 그래프형식의 지명표기사전의 예.
- 도 30은 영어권의 지명표기의 이표기의 예.
- 도 31은 도 30의 표기예를 문맥자유문법으로 기술한 예.
- 도 32는 지명표기사전 작성장치의 하드웨어(hardware)구성의 예.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 우편구분기 등에서 실행되는 지명표기 판독에 있어서, 주로 문자열 조합처리의 경우에 사용되는 지명표기사전의 작성방법과 작성장치에 관한 것이다. 특히 표준적인 표기로 이루어진 지명표기 리스트(list)로부터, 지명표기에 있어서 단어 배열의 변형이나 문자의 변형(이후, 이표기(異表記)라 칭한다)을 포함하는 지명표기사전의 작성방법 및 장치에 관한 것이다.

문자열을 판독하기 위해서는 이하의 3개의 단계로 이루어지는 처리가 행해지는 것이 일반적이다.

- (1) 문자절단 : 문자행 화상으로부터의 문자패턴의 절단
 - (2) 문자식별 : 각 문자패턴의 자중(문자코드)의 식별
 - (3) 문자열 조합 : 미리 기억한 판독대상의 문자열과 문자식별결과를 조합하여 문자열 판독결과를 확정
- (1) 문자 절단이나 (2) 문자식별에 관한 기술로서 예컨대, 문헌(Koga et al., Segmentation of Japanese Handwritten Characters Using Peripheral Feature Analysis International Conference for Pattern Recognition, pp. 1137-1141, 1998) 등이 알려져 있다.
- (3)의 문자열 조합에 관한 기술로서는 문자식별결과와 격자(lattice)로부터 유한의 자동기계장치(automaton)를 생성하고, 이것에 판독대상의 문자열을 입력함으로써 후보(候補)단어를 추출하는 방식(환천지 「손으로 쓴 한자주소 인식을 위한 에러수정 알고리즘」 정보처리학회논문지 등 35권 제 6호) 등이 있다. 또한, 숨겨진(hidden) 마코브모델을 사용함으로써 문자절단, 문자식별, 문자열 조합을 동시에 행하는 방식(A. Kaltenmeier, "Sophisticated Topology of Hidden Markov Models for Cursive Script Recognition, Proceedings of International Conference of Document Analysis and Recognition '93, pp.139-142, 1993)이나 탐색적으로 문자열을 인식하는 방법(고하지 「주소인식방법」

특원평 9-238032호)이 있다. 여기에서 사용되는, 미리 준비한 판독대상의 문자열의 집합을 사전이라 하고, 지명판독용으로 지명표기가 저장된 사전을 지명표기사전이라 한다.

지명표기사전의 계산기 메모리상에서의 형태는 예컨대, 환천 등의 방식에서는 나무 구조로 표시되어 있고, 또한 숨겨진 마코브모델에 따른 방식이나 탐색적으로 문자열을 인식하는 방법에서는 네트워크형식을 채택한다. 문자열 조합처리에서는 문자식별결과를 판독대상의 문자열과 조합하므로써 문자식별처리에서의 잘못을 정정하는 기능을 가지므로, 문자열판독 정도의 향상을 위해서는 어느 방법을 채용하는 경우도, 인식대상의 문자열, 즉 어휘를 미리 빠짐 없이 사전파일에 기억시켜 둘 필요가 있다. 결국, 판독대상의 전체 표기수에 대한 등록표기수의 비율인 사전완비율을 가능한 한 높이는 것이 필요하다.

지명표기에는 「上之町」에서 「之(노)」가 「ノ(노)」나 「の(노)」라고도 쓰여지고, 또는 「大字」라는 문자열이 지명표기에서는 생략되어 있는 것과 같은 여러가지 이표기가 존재한다. 우편구분기에서의 성이름(宛名)판독처리를 상정한 경우, 실제의 우편물상에 쓰여진 성이름도 상술한 바와 같은 표기의 변형이 존재하므로, 지명표기사전에 이표기를 등록하여 사전완비율을 높이는 것이 성이름 판독 정도의 향상에는 필수이다. 그러나, 문자열 조합처리를 실현하는 때에, 이들의 이표기를 완전하게 망라한 사전을 처음부터 준비한다는 것은 곤란하다. 따라서, 작성한 지명표기사전에 대해서 이표기를 추가하는 작업이 필요하게 된다.

지명표기사전에 대한 이표기의 추가라고 하는 문제에 대해서, 종래에는 「大田區」라는 문자열에 대해서 「太田區」라는 일부의 한자의 다른 문자열을 수작업하므로써 추가하는 것에 의해 사전완비율을 향상시키는 방법(공과, 「성이름 판독구분기」, 특개평 제 5-169031호)나, 「川越市」에 대한 「越市」나 「市」와 같은 부분문자열을 이표기로서 수작업으로 추가하는 방법(소도 「성이름 판독구분기」, 특개평 7-39819호)이 알려져 있다. 또한 데이터베이스에 등록되어 있는 문자열을 증가하는 방법으로서 미리 문자열의 이표기의 대응표를 준비하고, 이것에 따라서 이표기를 기계적으로 추가하는 방법(구전타 「표준명 부여시스템」, 특개평 제 5-165619호) 등도 알려져 있다.

일본의 지명표기는 크게 이하의 4가지로 분류할 수 있다.

(1) 사용하는 문자의 변형에 의한 이표기, 「문자형 이표기」라 칭함.

「야?하」와 「야노하」 「야노하」 등

(2) 단어의 생략에 의한 이표기, 「생략형 이표기」라 칭함.

도도부현명(都道府縣名)을 생략하는 이표기, 「대자(大字)」 「자(字)」를 생략하는 이표기 등

(3) 부가적으로 문자열에 의한 이표기, 「추가형 이표기」라 칭함.

자명(字名)(아자) 등의 본래 주소의 특징에는 불필요한 문자열이 부가된 이표기

「사이타마켄, 카와고에시, 오하자, 오가야」에 대한 「사이타마켄, 카와고에시, 오하자, 오가야, 아자, 히가시제키」 등

(4) 도로명·통칭에 의한 이표기, 「별명형 이표기」라 칭함.

경도(京都) 등에서 많이 나타나고, 또한 다른 단어에 의한 지명을 표기한다.

「쿄토시, 시모쿄쿠, 오만도쿄로마치」에 대하여 「쿄토시, 시모쿄쿠, 카라수마, 부코우지, 쿠다루」 등.

예컨대, 「사이타마켄, 카와고에시, 오가야」라 하는 지명을 예로 들면, (1), (2)의 문자형 이표기, 생략형 이표기만으로

「埼玉縣川越市小ヶ谷」(사이타마켄, 카와고에시, 오가야)

「埼玉縣川越市小ヶ谷」(사이타마켄, 카와고에시, 오가야)

「埼玉縣川越市小が谷」(사이타마켄, 카와고에시, 오가야)

「埼玉縣川越市大字小ヶ谷」(사이타마켄, 오하자, 오가야)

「埼玉縣川越市大字小ヶ谷」(사이타마켄, 오하자, 오가야)

「埼玉縣川越市大字小が谷」(사이타마켄, 오하자, 오가야)

「川越市小ヶ谷」(카와고에시, 오가야)

「川越市小ヶ谷」(카와고에시, 오가야)

「川越市小が谷」(카와고에시, 오가야)

「川越市大字小ヶ谷」(카와고에시 오하자, 오가야)

「川越市大字小ヶ谷」(카와고에시 오하자, 오가야)

「川越市大字小が谷」(카와고에시 오하자, 오가야)

의 12개의 표기가 존재한다. 또한 「사이타마켄, 카와고에시, 오가야, 히가시다」 「사이타마켄, 카와고에시, 오가야, 히가시제키」 「사이타마켄, 카와고에시, 오가야, 니시제키」 등, 소자명(小字名)이 병존하여 사용되고, (3)의 추가형 이표기를 고려하여, 상술한 12개의 이표기와 조합시키면, 84개의 이표기가 존재하는 것으로 된다. 또한, 경도시(京都市) 등에 있어서 현저하게 나타나는 정명(町名)이나 도로명에 의한 (4)의 별명형 이표기를 고려하면, 예컨대 동경시 이하의 경구(京區)의 지명표기에 있어서 이표기의

수는 수천에서 수만개에나 달한다.

우편구분기나 주소판독에 있어서는 최소한 하나 또는 복수의 시구군정촌(市區郡町村)의 주소, 지명판독 처리의 응용선에 따라서는 최대로 전국의 주소를 판독할 필요가 있게 되어, 지명표기의 총 수는 수만 이상으로 된다. 판독정도 향상에는 이들 지명표기의 이표기를 추가하고, 사전완비율을 높인 지명표기사전을 작성하는 것이 필요하다.

그러나, 수만에 이르는 표기를 특별히 지명표기사전에 추가해 가는 것은 곤란하다. 또한 단어의 수 만큼 이표기의 대응표를 준비하여, 자동적으로 이표기의 추가를 행한다 하더라도, 이표기의 대응표의 작성은 단어마다 특별히 행하지 않으면 안되므로 마찬가지로 곤란하다. 또한 단어의 배열에 관한 이표기(특정의 단어, 문자열의 생략 등)에 관해서는 그 조합의 수가 많게 되므로, 특별하게 추가하는 것은 물론, 이표기의 대응표를 갖는 것도 곤란하다. 더욱이, 특정의 문자에 관한 치환규칙을 준비하여 두고, 이표기를 자동적으로 추가하도록 하면, 예컨대 단어의 선두의 「野(야)」를 「ノ(노)」로 치환하는 등 잘못된 이표기를 추가하게 되고, 이와 같이 작성된 지명표기사전은 그 용량이 크게 될 뿐만 아니라, 판독정밀도에 악영향을 미친다.

이표기에 의한 지명표기사전의 용량의 증대를 억제하기 위하여, 문맥 자유문법의 생성규칙을 사용하여 지명표기의 이표기를 표현하는 방법이 알려져 있다(고하, 「지명표현방법, 지명문자열 인식방법 및 장치」, 특원평 제 11-187753호). 즉, 지명문자열의 일부 또는 전부를 구성하는 부분문자열마다에 문자 또는 구문카테고리의 배열을 정의하고, 문자 또는 정의완료의 구문카테고리의 배열로 이루어진 구문카테고리에 의해 지명문자열을 나타낸다. 예컨대 「ヶ」「ケ」「ガ」를 하나의 구문카테고리로서 정의하고, 상기의 문자가 사용되는 문자열을 이 구문카테고리에 의해 정의하면, 「ヶ」「ケ」「ガ」에 관한 이표기를 전체의 지명표기에 대해서 추가한 것으로 된다. 여기에서, 구문카테고리는 동등한 의미, 동등한 용법, 동등한 발음을 갖는 등 어느 정도의 공통성이 존재하는 문자열에 대해서 이들 문자열을 요소로 하는 집합적인 것이다. 또한, 이와 같은 집합에 붙여진 이름을 구문카테고리명이라 칭한다.

문맥자유문법을 사용하여 지명표기를 기술하면, 지명표기의 복수의 개소에 나타나는 단어나 부분문자열의 이표기가 동일한 구문카테고리로 표시되므로, 이표기의 추가작업의 회수가 감소된다. 그러나, 각각의 지명표기중 이표기가 존재하는 부분을 정의된 구문카테고리로 치환해 갈 필요가 있다. 이때, 지명표기의 어느 부분문자열을 대응하는 구문카테고리로 치환하는 작업은 수작업에 의해 행해지지 않으면 안되므로, 지명표기사전의 작성은 역시 곤란하다.

발명이 이루고자하는 기술적 과제

본 발명에 있어서 제 1의 목적은 지명표기에 대한 이표기의 생성이나 추가를 특별하게 행하지 않고, 이들을 가능한 한 자동적으로 추가하여, 사전완비율이 높은 지명표기사전의 작성방법을 제공하는 것이다.

지명표기사전으로 이표기를 자동적으로 추가할 수 없었던 경우, 사전완비율을 높이기 위해서는 수작업에 의한 표기의 추가를 행할 필요가 있다. 이 경우, 지명표기사전으로의 이표기의 추가분은 자동적으로 추가된 표기와 수작업에 의해 추가된 것의 2개의 집합으로 이루어진다. 지명표기사전의 작성의 경우는 표준적인 지명표기만으로 이루어진 지명표기 리스트가 존재하고, 이것을 기초로 하여 필요에 따라 이표기를 추가하여 지명표기사전을 작성하는 것이 일반적이다. 표준적인 지명표기 리스트에 변경이 가해질 때, 그 변경을 지명표기사전에 반영시키기 위해서는 다시 한번 이표기의 추가처리를 실행할 필요가 있다. 이 자동적으로 추가된 표기 이외에 수작업에 의해 추가된 표기에 관해서, 이전에 추가한 경우와 동일한 작업을 지명표기사전 작성자에 강요하는 것은 사전작성 비용이 크게 됨과 동시에, 이전에 추가한 표기의 추가누락으로 연결되어 사전완비율의 저하로 연결될 가능성이 있다.

본 발명에 있어서 제 2의 목적은 지명표기사전에 자동적으로 생성·추가될 수 있었던 지명표기의 이표기를 수작업에 의해 추가하는 수단을 제공함과 동시에, 지명사전작성시에 기초가 되었던 표준적인 지명표기 리스트가 변경된 경우에도 사전작성시에 추가되었던 여러가지의 이표기나 지명표기사전에 가해진 변경을 보존하면서, 새롭게 지명표기 리스트로부터 지명표기사전을 작성하는 방법을 제공하는 것이다.

본 발명에서는 표준적인 지명표기 리스트로서, 도도부현(都道府縣), 시구군정촌(市區郡町村), 정명(町名)·대자(大字), 자(字)이었던 주소의 계층마다의 단어로 구분된 텍스트 데이터(text data)를 상정한다. 지명표기가 구분되지 않은 경우에 있어서도, 「懸(현)」이나 「市(시)」이었던 키(key)로 되는 문자에 의해 지명표기를 분할함으로써 용이하게 변화가 가능하다. 또한 본 발명에 있어서 지명표기사전으로서 문맥자유문법으로 기술된 지명표기의 집합을 상정한다. 즉, 지명표기의 일부 또는 전부를 구성하는 부분문자열마다에 문자 또는 구문카테고리의 배열을 정의하고, 문자 또는 정의완료의 구문카테고리의 배열로 이루어진 구문카테고리에 의해 지명문자열을 나타낸다.

지명표기의 이표기를 가능한 한 자동적으로 지명표기사전에 추가한다고 하는 상기의 과제를 해결하기 위해서, 본 발명에서는 주소의 계층마다의 단어로 구분된 지명표기의 텍스트 데이터를 문맥자유문법에 의한 표현으로 변환하는 방법과, 문자열과 문맥자유문법에 있어서 구문카테고리를 대응시키는 지식, 문맥자유문법에 있어서 구문카테고리 및 구문카테고리의 배열을 별개의 구문카테고리에 대응시킨 지식을 갖고, 이들 지식을 사용하여 문맥자유문법으로 표기된 지명표기에 이표기를 추가하는 방법을 갖는 지명표기사전 작성방법을 제공한다.

여기에서 사용되는 문자열과 문맥자유문법에 있어서 구문카테고리를 대응시킨 지식, 문맥자유문법에 있어서 구문카테고리 및 구문카테고리의 배열을 별도의 구문카테고리에 대응시킨 지식으로서, 이하의 2종류의 지식을 제공한다.

A1 지명단어의 정의에 있어서 문자의 변형에 관한 이표기를 추가하기 위한 것, 문맥자유문법에 의한 구문카테고리의 정의문과 문자열로부터 구문카테고리로의 치환처리를 기동하기 위한 조건으로 이루어지는 지식. 이것을 「문자변형 보완형 지식」이라 칭한다.

「상지정(上之町)」에 있어서, 「之」「ノ」「の」 또는 「ヶ」「ヶ」「ガ」 등의 문자를 하나의 구문카테고리로서 정의하고, 지명표기 문자열중의 상기 문자를 구문카테고리로 치환한다.

A1 지명표기 정의문에 있어서 구문카테고리의 배열을 조건으로 하여, 구문카테고리에 대한 치환, 생략 등 새롭게 추가하는 정의문의 내용을 기재한 항목으로 이루어진 지식. 이것을 「단어변형 보완형지식」이라 칭한다.

도도부현명(都道府縣名), 시구군정촌명(市區郡町村名)과 계속되는 구문카테고리의 배열에 대해서는 도도부현명(都道府縣名)의 생략은 가능하였던 것 처럼 특정의 구문카테고리의 배열에 대해서 별개의 구문카테고리의 배열을 정의한다.

지명표기사전으로 자동적으로 추가할 수 없는 표기가 존재하는 경우, 해당하는 구문카테고리 및 구문카테고리의 배열을 지정하면, 그 카테고리가 표현하는 문자열을 표시하는 수단과, 문자열을 입력하는 수단과, 지명표기사전 작성자가 입력한 문자열을 구문카테고리가 표현하는 지명표기에 추가하는 수단을 가지므로써, 자동적으로 추가할 수 없었던 지명표기를 지명표기사전에 추가할 수 있도록 한다.

또한, 상기의 입력한 문자열을 구문카테고리가 표현하는 지명표기에 추가하는 방법에 있어서, 추가된 지명표기문자열만으로 이루어지는 구문카테고리를 메모리상에 작성하거나 또는 파일에 출력하는 방법과, 문맥자유문법으로 표현된 지명표기사전에 대해서 상기의 파일출력된 추가된 지명표기문자열을 추가하는 수단을 가지므로써 지명표기사전에 자동적으로 생성·추가할 수 없었던 이표기를 수작업에 의해 추가할 수 있음과 동시에, 주소의 계층마다의 단어로 구분된 지명표기 리스트가 변경된 경우에도 이전의 사전작성의 경우에 가해진 변경을 보존하면서, 새롭게 지명표기사전을 작성하는 것이 가능하게 된다.

지금까지 지명표기사전으로는 지명표기문자열을 열거한 텍스트, 문맥자유문법으로 기술한 텍스트형식이었지만, 우편구분기에 있어서 성이를 판독, 또는 다른 주소문자열 판독처리에 사용되는 지명표기사전은 이하의 3종류의 형태가 고려된다.

B1 파일에 저장된 「사전 소스 파일(source file)」

B2 메모리상에 저장된 「사전테이블」

B3 (1)과 (2)의 중간단계의 「사전 바이너리 파일(binary file)」

등록된 지명표기의 보수성, 시인성을 높인 경우에는 「사전 소스 파일」의 형태를 취한다. 또한 문자열 조합처리와의 친화성, 처리속도의 고속화를 도모하기 위해서, 계산기상에서는 「사전테이블」의 형태를 취한다. 또한 지명표기사전을 보존하는 경우의 디스크 스페이스(disc space)를 작게 하기 위해서 「사전 바이너리 파일」의 형태를 취한다. 이들 어느 형태를 취한 경우에서도, 그 정보량, 요컨대 등록된 지명표기는 동등하다. 또한 문맥자유문법으로 쓰여진 사전 소스 파일을 계산기에 의해 번역하는 사전 바이너리 파일, 또는 사전 테이블의 형태로 변환하는 것은 문헌(중전, 「컴파일러(compiler)」 ISBN4-7828-5057-3) 등에 기술된 방법을 사용하므로써 가능하다.

이후 본 명세서에서는 지명표기사전의 형태로서 B1의 사전 소스 파일을 상정하지만, 본 발명은 지명판독처리에 사용되는 지명표기사전이 사전 테이블이나 사전 바이너리 파일의 형태를 취하는 경우에서도 동일하게 적용가능하다.

발명의 구성 및 작용

본 발명의 의한 지명표기사전 작성방법의 실시형태의 일례로서, 일본의 지명표기에 대한 이표기를 포함한 지명표기사전의 작성에 관해서 서술한다. 여기에서, 일본의 지명표기로는 도도부현명(都道府縣名), 시구군정촌명(市區郡町村名), 정명(町名)(또는 대자(大字), 자명(字名))까지를 표현하는 문자열의 것으로 한다. 즉, 「도쿄도, 코쿠분지시, 히가시코이가쿠보, 1초메, 280반치」라 하는 주소표기에 있어서, 「도쿄도, 코쿠분지시, 히가시 코이가쿠보」의 부분을 지명표기라 칭한다.

최초에, 지명표기사전 작성처리의 흐름을 도 1을 사용하여 서술한다.

지명표기사전 작성처리의 입력으로서, 표준적인 지명표기를 열거한 파일인 지명표기 리스트(101)를 준비한다. 표준적인 지명표기는 단어로 구분되고, 단어의 배열로서 지명표기 리스트에 저장된다. 이 단어로 구분된 지명표기 리스트로부터, 형식언어표현 변환처리(102)에 의해, 미리 정의된 지명표기를 표현하는 문맥자유문법으로 정해진 형식으로 기술된 지명조합용 지명표기사전(이하, 지명표기사전이라 칭한다)(103)으로 변환된다. 이 단계에서는 표기사전중에는 표준적인 지명표기가 포함되어 있지 않다. 이후의 처리에 있어서는, 지명표기는 문맥자유문법에 의해 표현된 형식으로 처리된다. 문맥자유문법으로 표현된 표준적인 표기에 대해서 지식에 의한 이표기 추가처리(104)를 행한다. 여기에서는 「上之町(카미노마치)」라 하는 단어가 지명판독처리로의 입력에서 「上ノ町」「上の町」라 씌여지는 경우에 대해서, 한자에 끼워진 「之」는 「ノ」「の」를 이표기에도 적용된 규칙이나 「카와고에시, 오하자」의 「대자(大字)」는 생략하여도 좋다는 지명 표기법에 관한 규칙을 모은 지명표기추가용 지식(105)을 준비하여 둔다.

지명표기추가용 지식(105)을 사용하여 이표기의 추가된 지명표기에 대해서, 상황의존 이표기 추가처리(106)에 있어서, 수작업 등에 의한 특별한 형으로 이표기를 추가하고(107), 지명조합용 지명표기사전(103)을 출력한다.

여기에서 추가되는 이표기는 지명표기추가용 지식(105)을 사용하여 자동적으로 추가할 수 없는 「城内市本町」을 「城内市大手前」으로 표기하도록 한 이표기이다. 상황의존 이표기 추가처리(106)에 있어서 추가된 표기를 포함하는 지명조합용 지명표기사전(103)중, 추가·수정표기출력(108)에 있어서, 추가·수정된 부분만을 추출하여, 문맥자유문법의 형식으로 추가·수정 이표기정의(109)로서 출력, 보존된다.

표준적인 지명표기 리스트(101)에 변경이 가해지는 등 하여(110), 지명사전 작성처리를 다시 고치는 경우, 이 추가·수정 이표기정의(109)는 사례에 의한 이표기 추가처리(111)에 있어서, 과거에 가했던 이표기로서 지명표기사전(103)에 추가된다. 이것에 의해 지명표기사전 작성시에 수작업에 의한 특별한 이표기 추가를 하지 않아도 되도록 하여, 지명사전작성에 있어서 수작업의 부가를 최소한으로 억제할 수 있다. 여기에서 사용되는 지명표기추가용 지식(105)은 수작업에 의해 새로운 지식의 추가 등의 편집작업(112)을 행하는 것도 가능하다.

여기에서 작성된 지명표기사전(103)은 네트워크(113) 등을 통하여 우편구분기(114)나 그 외의 지명판독 장치(115)에 로우드(load)되고, 지명표기판독처리에 사용된다.

더욱이 작성된 지명표기사전(103)은 미디어변환장치(116)에 의해 기억매체의 변환이 행해져서, 지명표기 미디어(117)로서 배포하는 것도 가능하다.

이하, 도 1에 있어서 각 데이터, 지식 및 처리에 관해서 상세하게 서술한다.

도 1에서 사용하는 데이터중, 최초로 지명표기 리스트(101)에 관해서 서술한다. 일본의 지명표기는 (1) 도도부현(都道府縣), (2) 시구군정촌(市區郡町村), (3) 정명(町名)·대자(大字), 또한 필요에 따라서 (4) 자(字)(아자)의 계층의 단어로 나눌 수 있다. 예컨대 지명표기 「토쿄도, 코쿠분지시, 히가시코이카쿠보」는 (1) 도도부현명(都道府縣名)으로서 「東京都(도쿄도)」, (2) 시구군정촌(市區郡町村)으로서 「쿠쿠분지시」, (3) 정명(町名)·대자(大字)로서 「히가시코이카쿠보」의 단어로 분할될 수 있다. 또한 별개의 지명표기 「아이치켄, 나고야시, 미나미쿠, 모토호시자키마치, 아자, 미야우라」에 있어서는 (1) 도도부현명(都道府縣名) 「아이치켄」, (2) 시구군정촌명(市區郡町村名) 「나고야시, 미나미쿠」, (3) 정명(町名)·대자(大字) 「모토호시자키마치」, (4) 자명(字名)으로서 「아자, 미야우라」라 하는 단어로 분할할 수 있다. 본 실시예에서의 지명표기 리스트로서 각 지명표기는 주소의 계층에 따라 단어로 분할되고, 동일한 계층의 단어를 동일한 파일로 모음과 동시에, 각 단어에 코드를 할당하고, 지명표기는 단어에 할당된 코드를 지시함으로써 표현한다고 하는 형식을 상정한다. 이와 같은 동일계층의 단어가 모여진 파일을 지명단어 리스트 파일이라 칭한다. 또한 다이렉트 메일 등의 성이름에 잘 쓰이고 있고, 카타카나에 의한 성이름을 판독하기 위해서 카타카나에 의한 지명표기사전도 작성하는 것을 고려하고, 상기 주소계층마다의 단어에는 후리가나가 써여 있는 것으로 한다. 카타카나의 지명표기의 사전을 작성하지 않는 경우는 이상의 기술중, 카타카나에 관한 부분을 삭제하면 된다. 지명표기는 상기 지명단어 리스트 파일중의 지명단어로서의 포인터의 배열에 의해 구성된다. 코드 리스트 파일에 의해 표현된다.

구체적으로는 도도부현명(都道府縣名)에 관한 지명단어 리스트 파일은 도 2와 같이 이루어진다. 도도부현(都道府縣)코드(201)는 도도부현명(都道府縣名)(202)에 대해서 일의로 할당된 번호이다. 도도부현명(都道府縣名) 후리가나(203)는 도도부현명(都道府縣名)의 읽기가 카타카나로 저장되어 있다. 시구군정촌명(市區郡町村名)에 대한 지명단어 리스트 파일은 도 3과 같이 된다. 도도부현(都道府縣) 코드(301)는 그 시구군정촌(市區郡町村)이 존재하는 도도부현명(都道府縣名)의 코드이다. 이것은 도 2에 나타난 도도부현지명(都道府縣地名) 단어 리스트 파일에 기재된 도도부현(都道府縣) 코드와 동일하다. 시구군정촌(市區郡町村) 코드(302)는 시구군정촌명(市區郡町村名)(303)에 대해서 일의로 할당된 번호이다. 정부령 지정도시와 같이 구(區)가 존재하는 경우는 명고옥시남구(名古屋市南區)를 하나의 단어로서 저장한다. 시구군정촌명(市區郡町村名) 후리가나(304)는 시구군정촌명(市區郡町村名)의 읽기가 카타카나로 저장되어 있다. 정명(町名)·대자(大字)에 대한 지명단어 리스트 파일은 도 4와 같이 된다. 도도부현(都道府縣)코드(401), 시구군정촌(市區郡町村)코드(402)는 그의 정명(町名)·대자(大字)가 존재하는 도도부현(都道府縣), 시구군정촌(市區郡町村)의 코드이다. 정명(町名)·대자(大字)코드(403)는 정명(町名)·대자명(大字名)(404)에 대해서 할당된 번호이다. 정명(町名)후리가나(405)는 정명(町名)·대자명(大字名)에 대한 읽기를 카타카나로 저장하고 있다. 표준지명표기에 대해서 미리 이표기를 갖는 것도 가능하다. 지명단어 리스트 파일에 이표기수(406)로 저장된 수만큼 이표기란(407)에 이표기를 저장하여 둔다. 예컨대 「石元町(이시모토초)」(408)에 대한 이표기 「石本町(이시모토초)」(409)가 존재할 때, 이는 파일에 저장된다. 이표기는 후에 규칙에 의해, 또는 수작업에 의해 추가되므로 지명단어 리스트 파일에 있어서 이표기 필드는 생략가능하다. 자명(字名)에 관한 지명단어 리스트 파일은 도 5와 같다. 도도부현(都道府縣) 코드(501), 시구군정촌(市區郡町村)코드(502), 정명(町名)·대자(大字)코드(503)는 각각 이 자(字)가 존재하는 도도부현(都道府縣), 시구군정촌(市區郡町村), 정명(町名)·대자(大字)의 코드이다. 자(字)코드(504)는 자명(字名)(505)에 대해서 일의로 할당된 코드이다. 자(字)후리가나(506)는 자명(字名)에 대한 읽기를 카타카나로 저장하고 있다. 자명(字名)에 이표기가 등록되어 있으면, 이표기수(507), 이표기(508)필드에 드는 것도 가능하다. 각 지명표기는 도 6에 나타난 바와 같은 코드 리스트 파일에 의해 표현된다. (601)은 각 지명표기를 식별하기 위한 기호이고, 여기에서는 지명표기에 대해서 일의로 할당되어 있는 7행의 우편번호(601)를 사용한다. 우편번호에 이어서, 도도부현(都道府縣)코드(602), 시구군정촌(市區郡町村)코드(603), 정명(町名)·대자(大字)코드(604), 자(字)코드(605)를 지정함으로써 지명표기를 기술한다.

본 실시예에서는 지명표기사전 작성처리의 입력으로서, 상기에 서술하는 바와 같은 지명단어 리스트 파일과 코드 리스트 파일을 상정하지만, 「토쿄도, 코쿠분지시, 히가시코이카쿠보」와 같은 지명을 나타내는 문자열의 리스트형식이어도 좋다. 예컨대, 지명을 표기하는 문자열에 대해서 「縣」「市」「町」「大字」「字」라 하는 키(key)로 이루어진 문자에 의해 문자열을 단어로 분할하고, 이것에 코드를 할당함으로써 상기 지명단어 리스트 파일과 코드 리스트 파일의 형식으로 변환할 수 있다. 따라서, 지명사전작성처리의 입력으로서 지명단어 리스트 파일과 코드 리스트 파일의 형식을 전제로 하여도 일반성을 잃지 않는다.

도 1에 있어서 지명표기사전(103)을 기술하는 형식에 있어서, 지명을 표기하기 위한 문맥자유문법의 예를 나타낸다.

정의하는 문맥자유문법의 종단기호로서 이하의 기호를 사용한다. 이들은 문맥자유문법의 기술법으로서 일반적인 BNF기법(Backus-Naur-Form)(중전 「컴파일러」ISBN4-7828-5057-3)과 동일한 의미를 갖는다.

::= 치환. 좌변의 구문카테고리를 우변의 구문카테고리 또는 문자의 배열로 치환할 수 있는 것을 의미한다.

[] 옵션. []내의 기재가 있거나 없더라도 좋은 경우를 의미한다.

| 선택. 좌변, 우변의 어느 것을 의미한다.

() 결합, 전후의 변수보다 먼저 괄호내를 평가한다.

< > 구문카테고리를 나타내는 변수명.

; 문장의 마침.

begin 변수의 유효범위의 시작.

end 변수의 유효범위의 마침.

이것 외에, 아라비아수자 및 한자·히라가나·카타카나의 배열이 종단기호로서 사용된다. 비종단기호로서 「지명표기」 「지역블록군」 「지명블록군」 「지역블록」 「지명블록」 「지명선언」 「변수정의」 「변수식」 「변수항」 「지명표기정의」 「숫자열」 「문자열」 「영자」의 13종을 준비한다.

이 문맥자유문법을 나타내는 생성규칙은 이하와 같이 된다.

「지명표기」 → 「지명선언」 「지역블록군」

「지명표기」 → 「지명선언」 「정역블록군」

「지역블록군」 → 「지역블록군」 「지역블록」

「지역블록군」 → 「지역블록」

「지명블록군」 → 「지명블록군」 「지명블록」

「지명블록군」 → 「지명블록」

「지역블록」 → begin 「지명선언」 「지명블록군」 end

「지역블록」 → begin 「지명선언」 「지명블록군」 end

「지명블록」 → begin (「숫자열」) 「지명선언」 「지명표기정의」 end

「지명선언」 → 「지명선언」 「변수정의」

「지명선언」 → NULL

「변수정의」 → 「정역변수명」 ::= 「변수식」 :

「변수식」 → 「변형식」 | 「변수항」

「변수식」 → 「변형식」 「변수항」

「변수식」 → 「변수항」

「변수항」 → [「변수식」]

「변수항」 → (「변수식」)

「변수항」 → 「정역변수명」

「변수항」 → 「문자열」

「지명표기정의」 → < 「숫자열」 > ::= 「변수식」 :

「정역변수명」 → < 「영자」 「문자열」 >

여기에서 비종단 기호 「숫자열」 「문자열」 「영자」는 각각 숫자의 배열, 한자·히라가나·카타카나의 배열, 알파벳 문자의 종단기호를 나타낸다. 또한, 이들의 기호는 이하의 순서로 우선순위로 평가한다.

(1) <W「문자열」> 및 <「숫자」> 에서 기술되는 구문카테고리의 각 정의

(2) [] 및 () 의 2류, 2중 이상의 괄호 2종을 사용한 경우에는 내측의 괄호를 우선으로 평가.

(3) |

(4) ::=

이것 외에 매크로(macro)명령으로서 「#include <파일명>」을 준비한다. 이 문장에 대해서 「파일명」으로 지정된 문맥자유문법으로 기술된 지명표기의 파일을 매크로전개한다. 이것에 의해 문맥자유문법으로 기술된 지명표기를 복수로 분할할 수 있다. 예컨대, 주소의 계층에 대응하여 파일을 분할하면, 동일한 문맥자유문법의 정의문을 쓸 필요가 없고, 관리가 용이하게 된다. 또한, 「/*」와 「*/」로 둘러싼 부분은 코멘트로서 언급한다.

예로서, 「아이치켄, 나고야시, 미나미쿠, 모토호시자키초」의 이표기를 포함한 지명표기를 문맥자유문법으로 기술한 예를 도 7에 나타낸다. 본 실시예에서는 문맥자유문법으로 기술된 지명표기의 파일로서,

C1 지명단어를 구문카테고리로서 정의하는 파일, 지명단어 정의파일이라 칭한다.

C2 지명표기를 구문카테고리에 의해 정의하는 파일, 지명표기 정의파일이라 칭한다.

C3 상기 C1, C2에 의해 지명표기 전체를 정의하는 파일, 전체 표기정의 파일이라 칭한다.

C4 지역에 관계 없이 사용되는 문자열을 구문카테고리로서 정의한 파일, 공통표기 정의파일이라 칭한다.

상기 4종류로 분할하여, 지명표기사전을 작성한다. 상기 C4의 파일은 미리 지명사전 작성처리에 있어서 이표기를 추가하기 위한 지명표기추가용 지식(105)으로서 사전에 준비하여 두는 파일이다.

다음에 도 1에서의 처리중, 형식언어 표현변환처리(102)에 관해서 서술한다. 도 8은 처리의 흐름을 나타낸 플로우차트이다. 표준적인 지명표기 리스트인, 주소의 각 계층의 지명단어 리스트 파일과 코드 리스트 파일로부터 문맥자유문법에 의한 표현형식으로의 변환처리를 「나고야시, 미나미쿠」의 지명표기를 예로 사용하여 서술한다.

형식언어표현 변환처리(102)에 있어서 제 1 스텝(스텝(801))으로서 상기 C3의 지명표기 전체를 정의하는 파일을 작성한다. 명고옥시남구(名古屋市南區)에 관한 지명표기 전체를 정의하는 파일의 1군을 도 9에 나타낸다. 지명표기 전체를 정의하는 파일의 작성에 있어서, 우선 상기 C1, C2의 명고옥시남구(名古屋市南區)에서의 지명단어 정의파일과 지명표기 정의파일의 이름을 설정하고, 메모리상에 기억한다. 다음에, 지역에 의존하지 않는 구문카테고리를 참조하기 위해서 전체 표기정의 파일의 1행째에 상기 C4파일을 포함하는 명령(901)을 출력한다. 이어서, 명고옥시남구(名古屋市南區)의 지명표기를 정의하는 것을 명시적으로 나타내기 위해, 「begin」(902)을 파일의 2행째에 출력한다. 그리고, 지명단어 정의파일, 지명표기 정의파일을 참조하기 위해서 3행째와 4행째에 각각, 명고옥시남구(名古屋市南區)에 있는 지명단어 정의파일의 포함문장(903)과 명고옥시남구(名古屋市南區)의 지명표기 정의파일의 포함문장(904)을 출력한다. 최후에 명고옥시남구(名古屋市南區)의 지명표기가 끝난 것을 명시적으로 나타내기 위하여 「end」문(905)을 5행째에 출력한다. 복수의 지역의 지명표기를 정의하는 경우는 상술한 「begin」과 「end」의 사이의 문장의 생성을 반복하면 좋다. 또한, 도도부현(都道府縣), 시구정촌(市區町村)의 계층에서 각각 지명단어의 정의파일과 지명표기의 정의파일을 분할하므로써, 예컨대 명고옥시(名古屋市)내의 복수의 구에 걸친 지명표기에 관한 정의파일을 작성하는 경우, 도도부현(都道府縣)이나 시(市)에 관한 정의파일을 한번 더 포함할 뿐만 아니라, 지명표기사전의 용장성(冗長性)이 감소된다. 이를 위해서는 「begin」 「end」를 넣은 글자구조로 하면 좋다.

형식언어표현 변환처리(102)에 있어서 제 2의 스텝으로서 지명단어정의 파일작성처리(802)를 행한다. 주소의 각 계층의 지명단어 리스트 파일로부터 지명단어정의하는 파일을 생성한다. 우선, 지명단어 리스트 파일에 있어서, 도도부현명(都道府縣名), 시구군정촌명(市區郡町村名), 정명(町名)·대자명(大字名), 자명(字名)에 기재된 단어에 대해서 구문카테고리를 작성하고, 지명단어 정의파일로 출력한다. 구문카테고리의 이름은 그 단어 자체로 한다. 이때 구문카테고리명의 앞의 영자로서 도도부현명(都道府縣名)의 카테고리에는 대해서는 「P」, 시구군정촌(市區郡町村)에 대해서는 「C」, 정명(町名)·대자(大字)에 대해서는 「T」, 자명(字名)에 대해서는 「V」를 붙인다. 시구군정촌명(市區郡町村名) 필드에 있어서, 시(市)와 구(區)나 군(郡)과 정촌(町村)이 혼재되어 있는 경우에도, 그들을 분할하지 않고, 하나의 변수로서 정의하는 문장을 생성한다. 예컨대 「애지현(愛知縣)」에 대한 구문카테고리의 이름은 「〈P愛知縣〉」으로 된다. 더욱이 구문카테고리를 지명단어 리스트 파일에 있어서 도도부현(都道府縣)코드, 시구군정촌(市區郡町村) 코드, 정명(町名)·대자(大字)코드, 자(字)코드를 사용하여 나타내고, 각 행의 선두에 코멘트의 형으로 부가한다. 명고옥시남구(名古屋市南區)의 지명단어에 대해서 여기까지의 처리를 행하여 생성된 지명단어 정의파일의 예를 도 10에 나타낸다. 행두의 코멘트기호 「/*」 「*/」로 둘러싸인 부분은 「!」에 이어서 도도부현(都道府縣) 등을 나타내는 코드가 기재된다. 예컨대 행(1001)은 愛知縣을 나타내는 코드 「23」에 이어서 명고옥시남구(名古屋市南區)를 증명하는 코드 「112」가 「!」를 끼워 기재된다. 도도부현(都道府縣)코드로부터 주소의 계층에 따라 코드를 배열하므로써 지명단어 사이의 접속관계를 유지하는 것이 완성된다. 또한 지명표기 리스트의 지명단어 리스트 파일중에 도 3이나 도 4에서 나타낸 것과 같은 지명단어에 관한 이표기가 저장되어 있으면, 행 (1002)와 같이 정의문의 우변을 |으로 구분하여 추가한다.

여기에서, 행 (1001)이나 (1003)과 같이, 시구군정촌명(市區郡町村名)으로서 「~시, ~군」나 「~군, ~마치」로 되어 있는 경우, 「~시」나 「~군」의 부분을 공통화하므로써 지명표기사전의 용장성을 감소할 수 있다. 따라서, 지명단어 정의파일작성처리에 있어서 다음의 처리로서 이들 문자열을 지명단어로 분할한다. 시구군정촌명(市區郡町村名)에 관한 지명단어 리스트 파일에 기재된 문자열은 이하의 6종류이다.

- (1) ~시
- (2) ~시~군
- (3) ~군
- (4) ~군 ~정
- (5) ~정
- (6) ~군 ~촌
- (7) ~촌

상기의 6종류에 관해서 지명단어분할의 처리의 흐름을 도 11을 사용하여 서술한다. 스텝 (1101)에 있어서, 최초에 필드내의 최후의 문자를 보므로써, (1), (2) (3), (4)(5), (6)(7)의 4개의 그룹으로 구별할 수 있다. (1)에 관해서는 스텝 1102으로 나아가고, 그대로 문자열을 지명단어로 한다. (2)(3)에 관해서는 문자열중에 「시」가 포함되어 있는 것으로 구별할 수 있고, 후의 2개의 그룹에 관해서는 문자열중에

「군」이 포함되어 있으므로, 각각 구별할 수 있다. (2), (4), (6)의 3그룹에 관해서는 2개의 단어로 분할된다. (2)에 관해서는 「시」의 전후에서 단어로 나뉘어진다. (4)(6)에 관해서도 동일하게 「군」을 구분할 수 있지만, 필드중에 「군교군, 야하타마치」등 「군」이 복수 포함되는 경우가 있다. 이와 같은 경우에도 (4), (6)을 바르게 단어를 분할하기 위해서 이하와 같은 규칙을 설정한다.

필드중의 2문자째로부터 문자열을 검색하고, 최초로 발견한 「군」의 다음에서 분할한다, 분할한 후의 단어에 대해서 정부령 지정도시 또는 군을 나타내는 첫번째의 단어에 대해서 카테고리명은 <0을, 구, 정촌을 나타내는 두번째의 단어에 대해서는 카테고리명에는 <W를 붙여 구별한다. 「군」이 나타나지 않으면, 분할은 행해지지 않고, 문자열 전체를 <W에서 시작하는 카테고리로서 정의한다(스텝1103, 1104). 여기에서, 분할되지 않고, 행 (1002)의 「오와리아사히시」와 같은 단어에 대해서도 생성된 「<C」로 시작하는 정의문을 「<0」으로 시작하는 정의문으로서 재정의한다. 이 때, 「<0」으로 시작하는 변수 명으로부터는 말미의 「시」「정」「촌」을 제거한다. 또한, 지명단어 정의파일의 행두의 지명코드를 잡고, 「<P」로 시작하는 구문카테고리로 표시되는 도도부현명(都道府縣名)과 <C로 시작하는 구문카테고리로 표시되는 시정촌명을 조합시킨다. 이것에 의해서 「<C」로 시작하는 구문카테고리의 정의에 도도부현명(都道府縣名)으로부터 시구군정촌(市區郡町村)까지의 표기를 정의하는 것으로 된다. 여기에서의 처리로 생성된 지명단어 정의파일의 예를 도 12에 나타낸다.

지명단어 정의파일 작성처리(802)에 있어서 3번째의 처리로서 정의문의 우변이 구문카테고리명만으로 이루어지는 정의문을 지명표기 정의파일로 옮긴다. 「<C」로 시작하는 구문카테고리의 정의문은 그 우변에 관해서도 이미 정의된 구문카테고리명으로 되어 있다. 따라서, 이들의 「<C」에서 시작하는 정의문만을 추출하고, 별도로 다음에 작성하는 지명표기 정의파일에 출력한다. 이렇게 하여 지명단어의 구문카테고리에 의한 정의와, 지명표기의 구문카테고리에 의한 정의가 완전히 분할된다. 이와 같이 하므로써 한자에 의한 지명표기사전과 카타카나에 의한 지명표기사전을 작성하는 경우, 지명단어의 구문카테고리에 의한 정의파일을 각각 작성하는 것 만으로, 지명표기의 구문카테고리에 의한 파일은 한자와 카타카나의 사전에서 공통화할 수 있으므로, 지명표기사전의 용장성을 작게 하는 효과가 있다. 이상의 처리에 의해 생성되는 지명단어 정의파일을 도 13에, 추출된 「<C」에서 시작하는 구문카테고리의 정의문을 도 14에 나타낸다.

형식언어표시 변환처리(102)에 있어서 최후의 스텝으로서 지명표기 정의파일 작성처리(803)를 행한다. 처리의 흐름을 도 15와 처리에 도 16을 사용하여 서술한다.

우선, 스텝 (1501)에 있어서 코드 파일의 말미가 어떤지를 조사하고, 말미이면 처리를 마친다. 미처리의 지명표기가 남아 있으면 스텝 (1520)으로 진행하고, 코드파일의 도도부현(都道府縣), 시구군정촌(市區郡町村), 정명(町名)·대자(大字)코드로부터 지명표기를 특정하는 코드를 지명표기 정의파일의 코멘트행으로 하여 출력한다. 다음에 스텝 (1530)으로 진행하고, 「begin」을 출력한다. 이것은 구문카테고리의 유효범위를 명확하게 함과 동시에, 사전파일의 시인성을 높이기 위한 것이다. 더욱이 스텝 (1504)에 있어서, 코드파일의 우편번호란에 기재된 지명표기의 식별자를 괄호로 둘러싸서 출력한다. 다음에 스텝 (1505)로 진행하여, 우편번호를 이름앞에 갖는 구문카테고리를 정의한다. 구체적으로는 구문카테고리의 식별자를 「< >」으로 둘러싼 좌변으로 하여, 기호 「::=」를 끼우고, 우변에는 미리 기재한 코드에 대응하는 구문카테고리를 지명단어 정의파일로부터 검색하여, 그 카테고리의 이름을 차례로 배열하고, 행 (1601)과 같은 정의문을 작성한다. 자명(字名)이 존재하면, 해당하는 지명표기의 정명(町名)·대자(大字)의 카테고리명의 후에 「()」를 추가하고, 그 내부에 자명(字名)의 구문카테고리명을 추가한다. 자명(字名)이 복수 존재하는 경우는 「|」로 끼워 행 (1602)와 같이 추가한다.

이상, 한자에 의한 지명표기사전에 관한 것으로서, 형식언어표현으로의 변환에 대해서 설명하였지만, 카타카나에 의한 지명표기사전의 작성에 관해서도 동일한 방법으로 실현가능하다.

다음에 도 1에 있어서, 지식에 의한 이표기 추가처리(104)에 관해서 서술한다. 사전에 준비한 지명표기 추가용 지식(105)에 의해 자동적으로 추가되는 표기는 이전에 서술한 4개의 이표기중, 이하의 3종류이다.

(1) 문자형 이표기 : 지명단어에 있어서 문자사용의 변형 등의 이표기

「之」「ノ」「の」 또는 「ガ」「ケ」「が」 등

(2) 생략형 지명표기 : 단어·구문카테고리가 생략되는 이표기

현명의 생략은 가능하다고 하는 표기의 차이

(3) 추가형 이표기 : 단어·구문카테고리가 추가되는 이표기

자명(字名)(아자)을 추가할 수 있다고 하는 이표기

이들 이표기를 추가하기 위해서 도 17에 나타난 것과 같은 순서를 취한다. 이것을 도 13, 도 14의 지명표기를 예로 서술한다. 지식에 의한 이표기 추가처리의 제 1의 스텝 (1701)으로서 특정의 문자열을 이미 정의하고 지식으로서 가지고 있는 구문카테고리로 치환한다. 여기에서의 지식은 본 실시예에서 사용하는 문맥자유문법에 의한 구문카테고리의 정의문의 형식으로 기술되어 있고, 정의문의 우변의 문자열이 처리 대상의 지명단어 정의파일중에 존재하는 경우에, 그 문자열을 우변의 구문카테고리명으로 치환한다. 예컨대, 도 18A의 행 (1801)의 「大字(오하자)」라고 하는 문자열은 정명(町名)·대자명(大字名)을 나타내는 구문카테고리의 선두에 위치하므로, 이것을 「<V대자>」라 하는 정의완료의 구문카테고리로 치환하고, 도 18B의 행 (1804)와 같이 한다. 행 (1802), 행 (1803)의 「ケ」「の」라는 문자도 지명단어의 선두, 말미에서는 없고, 또한 그 양측이 한자로 되어 있으므로, 각각 「<Vケ>」「<Vノ>」라고 하는 정의완료의 구문카테고리로 치환하고, 행 (1805), 행 (1806)과 같이 한다. 지식의 형식에 관해서는 후술한다.

지식에 의한 이표기 추가처리의 제 2의 스텝 (1702)로서 지명단어 문자열의 복수개소에 나타난 문자열의

구문카테고리화를 행한다. 여기에서 문자형 이표기를 생성한다. 이것은 예컨대 愛知縣은 「오와리」로 시작하는 지명이 많고, 또는 경도시(京都市)의 지명에서 「니시쿠쵸」나 「이치쵸지」등의 넓은 범위를 나타내는 지명을 접두어로 갖는 지명이 많으므로, 복수의 정역(町域)에 걸쳐 공통으로 접두구처럼 사용되고 있는 문자열을 구문카테고리화한다. 이것에 의해 복수의 지명에 공통하여 나타내는 문자열에 대해서, 이표기의 추가처리가 일원화될 수 있음과 동시에 이표기의 문자열을 지명표기사전중에 복수개소에서 갖는 것이 없으므로, 사전의 용장성을 작게할 수 있다. 지명단어 정의파일에 대해서 정의문의 우변의 선두로부터 2문자 또는 3문자를 취출하고 메모리상에 유지하여 두고, 복수의 정의문에 중복하여 놓으면, 이 공통부분의 문자열을 구문카테고리로서 정의하고, 시작의 문자열을 구문카테고리명으로 치환한다. 도 19에 있어서, 행 (1901)과 행 (1902)의 우변의 선두 3문자 「니시쿠쵸」가 공통으로 존재하고 있으므로, 이것을 구문카테고리로 하여 행 (1903)을 작성하고, 행 (1901), 행 (1902)의 문자열 「니시쿠쵸」를 구문카테고리 「〈V니시쿠쵸〉」로 치환하여, 행 (1904), 행 (1905)를 생성한다.

지식에 의한 이표기 추가처리의 제 3의 스텝 (1703)으로서 구문카테고리에 의한 지명표기의 정의의 추가를 행한다. 여기에서 추가되는 이표기는 생략형 이표기와 추가형 이표기이고, 주로 지명표기에 있어서 단어, 즉 구문카테고리의 배열의 차이이다. 예컨대, 도도부현명(都道府縣名)이 생략가능하거나 정명(町名)·대자(大字)에 있어서 「大字(오하자)」가 생략가능하거나, 상술한 동경시에 있어서 「니시쿠쵸」등의 넓은 범위의 지명을 생략가능하다고 하는 예가 있다. 이들을 후술하는 형식의 규칙으로서 유지하여 두고, 지명표기 정의파일 및 지명단어 정의파일에 대해서 정의문의 추가를 행한다. 예컨대 도 20A의 지명표기의 정의에 대해서 지정촌명(市町村名)으로부터 시작하는 표기를 추가한다고 하는 규칙이 유효하게 되는 경우, 「미나미쿠」와 「오와리아사히시」의 정의문의 다음에 각각 「」가 생략된 정의문을 부가한다. 그 결과가 도 20B이고, 행 (2001)과 행 (2002)가 추가된 정의문이다. 다음에 표기가 정의되면, 좌변의 구문카테고리가 동일한 정의문에 대해서 또는 결합연산자 「|」를 사용하여 1개의 정의문으로 한다. 이 결과, 도 20C에 나타난 행 (2003)과 행 (2004)와 같이 된다. 개시위치의 지정 외에, 어느 문자열의 생략 등도 동일한 순서에 의해 TPDL중에 반영한다. 지명단어 정의파일로의 이표기의 추가도 동일한 순서에 의해 가능하다.

이상이 지식에 의한 이표기 추가처리이다.

다음에 도 1에 있어서, 이표기추가를 위해 사용되는 지명표기 추가용 지식(105)의 지식표현에 관해서 서술한다.

이표기추가에 사용되는 지식은 상술한 바와 같이, 이하의 2종류이다.

D1 지명단어의 정의에 있어서 문자형 이표기를 추가하기 위한 것, 문맥자유문법에 의한 구문카테고리의 정의문과 문자열로부터 구문카테고리로의 치환처리를 기동하기 위한 조건으로 되는 지식.

D2 생략형 이표기, 추가형 이표기를 생성하기 위한 것, 지명표기 정의문에 있어서 구문카테고리의 배열을 조건으로서 구문카테고리에 대한 치환, 생략 등 새롭게 추가하는 정의문의 내용을 기재한 항목으로 이루어지는 지식.

상기 D1에 관한 구체적인 예를 도 21에 나타낸다. 이 지식은 구문카테고리 변환규칙부(2101)와 구문카테고리 지식정의부(2102)로 이루어진다. 구문카테고리 변환규칙부는 4개의 항목으로 이루어지고, 좌로부터 구문카테고리의 타입을 나타내는 카테고리 타입 필드(2103), 변환의 대상으로 되는 문자열의 배열을 규정하는 문자열 조건 필드(2104), 구문카테고리 정의부에 있어서, 변환되는 구문카테고리의 정의문의 좌변을 「|」 기호로 구분한 경우에 몇번째 문자열을 변환의 조건으로 하는가를 나타내는 비교대상 위치필드(2105), 구문카테고리 정의부에 있어서 변환되는 구문카테고리의 정의문을 가리키는 정의 포인터 필드(2106)로 된다. 구문카테고리 지식정의부는 문맥자유문법으로 씌여진 지명표기 파일중 D2의 지역에 의존하지 않는 문자열에 대한 구문카테고리를 정의하는 공통지명표기 정의파일이다. 이와 같이 표현되는 지식의 의미를 규칙 (2107)을 예로 서술한다.

1. 지명단어정의 파일중의 구문카테고리의 정의문에 있어서, 그 좌변의 구문카테고리의 종류, 즉, 「〈C, O, W, T, V〉」의 다음의 영자가 카테고리 타입 필드에 씌여진 「C, O, W, T, V」의 어느 것이고,

2. 한자로 끝나는 형으로 지명단어정의 파일중의 구문카테고리의 정의문의 좌변이 이하의 조건, 즉 정의 포인터 필드가 가리키는 구문카테고리 지식정의부중의 정의문의 우변에서 표현되는 문자중, 비교대상 위치필드에 씌여진 1번째와 2번째의 「ヶ」「ヶ」의 어느 것이, 문자열 조건필드에 나타나는 바와 같이, 한자로 시작하고, 도중에 상기 2문자를 포함하는 것을 만족시키면,

지명단어 정의파일중의 구문카테고리의 정의문의 우변에 있는 상기 문자를 구문카테고리 「〈Vヶ〉」로 치환한다.

여기에서, 구문카테고리 변환규칙부(2101)의 각 필드의 기법을 서술한다.

카테고리 타입 필드에는 이하의 영자를 기입하므로써, 당해 지식을 적용하는 지명단어 정의파일중의 구문카테고리의 정의문을 한정한다.

P 도도부현명(都道府縣名)의 구문카테고리

C 지구군정촌명의 구문카테고리

O 군명, 시명의 구문카테고리

W 구명, 정촌명의 구문카테고리

T 정명(町名)·대자명(大字名)의 구문카테고리

V 자명(字名), 그 외의 구문카테고리

다음에 문자열 조건필드에 사용되는 기호를 기재한다.

- x 구문카테고리 지식정의부에 씌여진 치환의 대상으로 되는 문자(열)
- K 문자열의 조건을 기술하는 경우에 임의의 한자를 나타낸다.
- H 문자열의 조건을 기술하는 경우에 임의의 평가명을 나타낸다.
- T 문자열의 조건을 기술하는 경우에 임의의 카타카나를 나타낸다.
- () 상기 문자의 종류가 씌여진 범위를 명시적으로 하기 위한 기호
- + 그 전에 씌여진 문자의 일회 이상의 반복
- * 그 전에 씌여진 문자의 제로회 이상의 반복
- ^ 문자열의 선두를 나타낸다.
- \$ 문자열의 말미를 나타낸다.

구문카테고리 지식정의부 즉, 지역에 의존하지 않고 사용되는 문자열에 대한 구문카테고리를 정의하는 정의문을 받아들이는 파일에는 이하와 같은 구문카테고리를 준비하여 둔다.

<Vヶ> ::= ヶ|ヶ|が;

<Vノ> ::= ノ|之|の;

<V々> ::= ノ|々|の;

<V老> ::= 老|一|1;

<V武> ::= 武|二|2;

<V参> ::= 参|三|3;

<V通> ::= 通[リ];

<V大字> ::= 大字;

<V字> ::= 字;

<V入ル> ::= 入([[い]る] | [[イ]ル]);

<V上ル> ::= 上([[ガ]ル] | [[が]る]);

<V下ル> ::= 下([[ガ]ル] | [[が]る]);

이표기추가를 위한 지식중 02에 관해서 서술한다. 구체적인 예를 도 22에 나타낸다. 이 지식은 4개의 필드로 이루어지고, 좌로부터 이표기 추가대상으로 되는 지명표기의 정의문에 있어서 구문카테고리의 배열방향을 기재한 카테고리 조건필드(2201), 카테고리 조건부의 조건을 만족한 정의문에 대해서 추가되는 변경의 종류를 기재하는 변형처리 지정필드(2202), 변형처리 지정부에서의 처리의 대상으로 되는 구문카테고리수를 나타내는 카테고리수 지정필드(2203), 추가처리의 경우에 새롭게 추가되는 구문카테고리의 이름을 나타내는 추가카테고리 지정필드(2204)로 된다. 이와 같이 표현되는 지식의 의미를 규칙(2205)을 예로 서술한다.

지명표기 정의파일에 있어서 구문카테고리의 정의문중, 그 우변이 카테고리조건구에 씌여진 것과 같이, 도도부현명(都道府縣名)을 나타내는 구문카테고리에 이어진 시명을 나타내는 구문카테고리로 구성되어 있으면, 변형처리 지정필드, 카테고리수 지정필드에 씌여진 바와 같이, 도도부현명(都道府縣名)과 시명(市名)의 구문카테고리의 2개의 구문카테고리의 변형에 「<V市内>」이라 하는 구문카테고리로부터 시작하는 표기를 추가한다.

여기에서, 각 필드의 기법을 서술한다.

카테고리 조건필드에서는 이하의 문자, 기호에 의해 지명표기 정의파일에 있어서 구문카테고리의 배열방향을 기술한다.

< > 구문카테고리명을 나타낸다.

P 도도부현명(都道府縣名)의 구문카테고리

C 지구군정촌명의 구문카테고리

O 도명, 시명의 구문카테고리

W 구명, 정촌명의 구문카테고리

T 정명(町名)·대자명(大字名)의 구문카테고리
V 자명(字名), 그 외의 구문카테고리
() 상기 문자의 종류가 씌여진 범위를 명시적으로 하기 위한 기호
+ 그 전에 씌여진 문자의 일회 이상의 반복
* 그 전에 씌여진 문자의 제로회 이상의 반복
^ 정의문에 있어서 우변의 선두를 나타낸다.
\$ 정의문에 있어서 우변의 말미를 나타낸다.
. 임의의 1문자를 나타낸다.
! 다음에 계속되는 구문카테고리의 부정

변경처리 지정필드에는 이하의 어떠한 처리를 기술한다.

SKIP 카테고리 조건부에 기재된 구문카테고리의 생략을 허가하는 표기를 추가한다.

REPLACE 표기의 선두로부터, 카테고리수 지정필드에 기재된 수의 카테고리를 추가카테고리필드에 기재된 카테고리로 치환한 표기를 추가한다.

우편물에 관한 성이름을 판독하기 위해 사용하는 지명표기사전을 작성하기 위해서, 이하와 같은 지명표기에 관한 이표기를 추가하기 위한 지식으로서, 이하와 같은 지식을 준비한다.

도도부현명(都道府縣名)에 대한 구문카테고리는 생략하여도 좋다.

도도부현명(都道府縣名), 시군명에 대한 구문카테고리는 생략하여도 좋다.

도도부현명(都道府縣名), 시군명, 더욱이 구명, (군으로 이어진다) 정촌명은 생략하여도 좋다.

「시」에 대해서는 「 시내」라 하는 표기를 추가한다

「정」에 대해서는 「정내」라 하는 표기를 추가한다.

「촌」에 대해서는 「촌내」라 하는 표기를 추가한다.

「구」에 대해서는 「구내」라는 표기를 추가한다.

정명(町名)·대자명(大字名)의 선두의 「대자(大字)」는 생략하여도 좋다

자명(字名)의 선두의 「아자(字)」는 생략하여도 좋다.

이상이 이표기추가를 위해서 사용되는 지식의 표현형식이다.

이어서, 도 1에 있어서, 상황의존 이표기추가처리(106)와 추가·수정이표기정의(109)에 관해서 상술한다.

여기에서는 지명단어 정의파일 및 지명표기 정의파일로의 구문카테고리의 정의문의 추가를 수작업에 의해 행한다. 즉, 수작업에 의한 구문카테고리의 정의문을 작성한다. 그러나, 문맥자유문법형식의 구문카테고리를 직접 기술하는 것은 작업자의 부담이 크기 때문에 이 작업을 지원하는 I/F를 준비한다.

이하, 전술한 이표기의 2개의 타입마다에 표기추가의 방법을 서술한다.

우선 지명표기에 대한 이표기의 추가의 순서에 관해서 서술한다. 지명표기에 대해서 이표기의 추가작업은 도 23 A, B, C와 같은 사용자 인터페이스를 통하여 행한다. 최초로 작업자가 편집의 대상으로 되는 지역의 식별자로서 우편번호를 우편번호입력란(2301)에 입력하면(도 23A 참조), 이것을 함께 지명표기 정의파일중으로부터 대응하는 지명표기정의문을 검색한다. 예컨대 「〈4570012〉 ::= 〈C명고옥시남구(名古屋市南區)〉〈T입사(笠寺)〉〈V본전(本殿)〉」이라 하는 지명표기 정의문이 검색되면, 그 우변에 기재된 구문카테고리를 취출하고, 시구군정촌명(市區郡町村名)에 대응하는 구문카테고리를 좌변으로 하는 정의문을 또한 검색하고, 도도부현명(都道府縣名), 시군명에 대응하는 구문카테고리를 취출한다. 이들의 구문카테고리의 이름을 도도부현(都道府縣)표시란(2302), 시군표시란(2303), 구정촌표시란(2304), 정명(町名)·대자(大字)표시란(2305), 자명(字名)표시란(2306)에 각각 표시한다. 복수의 구문카테고리가 사용되고 있으면, 이들 전부를 병기하여 표시한다(도 23B 참조). 다음에 작업자는 추가할 수 있는 표기를 입력한다(도 23C 참조). 현명, 시명, 구명, 정명(町名)·대자(大字)를 생략한 표기를 추가하는 경우는 정명 대자표시란(2306), 구정촌명란 표시란(2307)을 공백으로 하고, 자명(字名)표시란(2308)에 「본전(本殿)」을 입력한다. 그리고 추가단추(2309)를 누르므로써 지정한 표기가 생성·추가된다.

이 표기의 추가처리의 흐름을 도 25에 나타낸다.

우선, 스텝(2501)에 있어서 도 23A, B, C와 같은 인터페이스를 사용하여 추가한 표기에 대해서 문맥자유문법의 정의문의 형으로 한다. 우편번호가 4570012로 표시되는 지명표기에 대해서 도 24에서 추가된 「본전」이라 하는 표기에 대해서 「〈4570012〉 ::= 〈V본전〉」이라고 하는 구문카테고리 정의문이 생성된다. 지명표기사전중의 구문카테고리의 정의문에 대해서, 그 좌변이 상기 추가정의문의 좌변과 동일한 문장을 검색한다(스텝 2502). 검색의 결과, 해당하는 정의문이 존재하면, 스텝 (2503)에 있어서 각 지명표기문자열이 OR결합되도록, 검색결과의 정의문의 좌변을 전개한다. 요컨대

A ::= (B | C) D

라는 정의문을

$A ::= BD \mid CD$

로 변형한다. 스텝 (2504)에서는 변형된 지명표기사전중의 정의문에 대해서 추가정의문을 그 좌변을 OR 결합하므로써 통합한다. 이것에 의해,

$\langle 4570012 \rangle ::= \langle C\text{명고육시남구} \rangle \langle T\text{입사} \rangle \langle V\text{본전} \rangle ;$

이라 하는 지명표기지식에

$\langle 4570012 \rangle ::= \langle V\text{본전} \rangle ;$

을 결합하고,

$\langle 4570012 \rangle ::= \langle C\text{명고육시남구} \rangle \langle T\text{입사} \rangle \langle V\text{본전} \rangle \mid \langle V\text{본전} \rangle ;$

이라 하는 지명표기정의문이 생성된다.

다음에 스텝 (2505)에 있어서, 정의식의 최적화, 즉 사용하는 구문카테고리명의 수의 삭감을 행한다. 즉,

$A ::= BD \mid CD;$

라 하는 정의식을

$A ::= (B \mid C) D;$

또는

$A ::= BC \mid B;$

라 하는 정의식을

$A ::= B [C];$

로 변형하는 처리이다. 이것에 의해,

$\langle 4570012 \rangle ::= \langle C\text{명고육시남구} \rangle \langle T\text{입사} \rangle [\langle V\text{본전} \rangle] ;$

가 얻어진다.

또한, 스텝 (2506)에 있어서는, 추가한 정의문

$\langle 4570012 \rangle ::= \langle V\text{본전} \rangle ;$

를 후술하는 도 1에 있어서 추가·수정이표기 정의파일(109)에 저장한다.

이표기를 삭제하는 경우는 도 26과 같은 처리의 흐름으로 이루어진다.

우선 스텝 (2601)에 있어서 도 24A, 24B, 24C와 같은 인터페이스를 사용하여 선택한 표기에 대해서, 문맥자유문법의 정의문의 형으로 한다. 우편번호가 4570012로 표시되는 지명의 「본전」이라 하는 표기를 삭제하도록 하면, 「 $\langle 4570012 \rangle ::= \langle V\text{본전} \rangle$ 」라 하는 구문카테고리 정의문이 생성된다. 지명표기사전중의 구문카테고리의 정의문에 대해서, 그 좌변이 상기 추가정의문의 좌변과 동일한 문장을 검색한다(스텝 2602). 검색의 결과, 해당하는 정의문이 존재하면, 스텝 (2603)에 있어서, 각 지명표기 문자열이 OR결합되도록, 검색결과와 정의문의 좌변을 전개한다. 요컨대

$A ::= (B \mid C) D$

라 하는 정의식을

$A ::= BD \mid CD$

로 변형한다. 스텝 (2604)에서는 추가정의문의 좌변과 동일요소를, 변형한 지명표기사전중의 정의문의 좌변의 OR결합된 요소로부터 삭제하는 것에 의해 처리된다. 이것에 의해,

$\langle 4570012 \rangle ::= \langle C\text{명고육시남구} \rangle \langle T\text{입사} \rangle \langle V\text{본전} \rangle \mid \langle V\text{본전} \rangle ;$

라 하는 지명표기지식으로부터

$\langle 4570012 \rangle ::= \langle V\text{본전} \rangle ;$

를 삭제하여,

$\langle 4570012 \rangle ::= \langle C\text{명고육시남구} \rangle \langle T\text{입사} \rangle \langle V\text{본전} \rangle ;$

라 하는 지명표기 정의문이 생성된다.

다음에 스텝 (2505)에 있어서, 정의식의 최적화 즉, 사용하는 구문카테고리명의 수의 삭감을 행한다. 즉,

$A ::= BD \mid CD;$

라 하는 정의식을

$A ::= (B \mid C) D;$

또는

$A ::= BC \mid B;$

라 하는 정의식을

$A ::= B [C];$

로 변형하는 처리이다.

또한 스텝 (2506)에 있어서는, 삭제한 정의문

$\langle 4570012 \rangle ::= \langle V\text{본전} \rangle;$

을 후술하는 도 1에 있어서 추가·수정이표기 정의파일(109)에 저장한다.

수작업에 의한 이표기의 추가의 2번째의 예로서, 어느 특정의 지명단어에 관한 이표기를 추가·수정하는 경우에 관해서 서술한다. 이 작업은 도 24A, 24B, 24C와 같은 사용자 인터페이스에 의해 행한다. 지명표기에 관한 이표기의 추가의 경우와 동일하게, 작업자가 편집의 대상으로 되는 지역의 식별자로서 우편번호를 우편번호입력란(2401)에 입력하면(도 24A 참조), 이것을 함께 지명표기 정의파일중으로부터 대응하는 지명표기 정의문을 검색한다. 이어서 도 24B와 같이 해당하는 계층의 구문카테고리명을 선택한 후, 단어정의 추가·수정 인터페이스를 기동한다. 단어정의 추가·수정 인터페이스가 기동되면, 추가·수정의 대상으로 되는 구문카테고리명을 좌변으로 하는 지명단어 정의문을 검색한다. 검색한 지명단어 정의문의 우변에 기술된 문자열을 도 24C의 지명단어표시·입력란(2403)에 표시한다. 다음에 작업자는 추가·수정하고 싶은 표기(문자열)를 지명단어표시·입력란(2404)에 추가하면, 추가·수정의 대상으로 되는 구문카테고리명을 좌변으로 하고, 추가된 문자열을 우변으로 하는 지명단어 정의문을 지명단어 정의파일에 추가한다. 이것에 의해 도 25에 표시된 처리가 행해지고, 「 $\langle 0\text{毛張旭市} \rangle ::= \text{愛知旭市};$ 」라 하는 정의문이 생성되고, 전술한 지식에 의한 이표기추가처리와 동일한 순서로 지명단어 정의파일에 추가된다. 여기에서 생성된 구문카테고리 정의문이 추가·수정이표기 출력처리(108)에 의해 출력되고, 도 1에 있어서, 추가·수정이표기정의에 저장된다. 단어정의의 삭제의 경우도 도 26에 나타난 방법에 의해서 행하여진다.

상기 상황의존 추가처리에 있어서 출력되는 추가·수정이표기 정의의 예를 도 27에 나타낸다.

여기에서는 추가하는 정의문, 삭제하는 정의문이 지명표기사전과 동일한 형식으로 저장된다. 단, 추가처리가 삭제처리화를 나타내므로, 행두에 코멘트문의 형식을 사용하고, 처리내용의 식별자를 쓴다. 처리내용의 식별자는

#ADD# 추가

#DEL# 삭제

의 2종류이다.

이어서 도 1에 있어서 사례에 의한 이표기추가처리(111)에 관해서 서술한다. 사례에 의한 이표기 추가처리에 있어서 행해지는 처리는 추가·수정이표기 정의(109)를 읽어 들여, 그 중에 기재된 처리내용의 식별자에 의해 추가와 삭제의 처리가 대체되는 것 이외에는 상황의존 이표기 추가처리에 있어서 행해지는 이표기의 추가, 삭제처리와 동일하다. 사례에 의한 이표기 추가처리에 의해 과거에 수작업으로 추가한 이표기를 자동적으로 지명표기사전에 추가할 수 있다.

사례에 의한 이표기 추가처리의 흐름을 도 28에 나타낸다.

우선 스텝 (2801)에 있어서 도 27에 나타난 바와 같은 추가·수정이표기 정의 (109)를 읽어 들이고, 그 행두의 코멘트 내부에 기재된 처리내용의 식별자와, 그것에 읽은 표기의 정의문을 읽어 들인다. 우편번호가 4570012로 표시되는 지명표기의 「본전」이라 하는 표기를 추가하도록 추가·수정 이표기 정의파일에 써여 있는 경우, 지명표기사전중의 구문카테고리의 정의문에 대해서 그 좌변이 「 $\langle 4570012 \rangle ::= \langle V\text{본전} \rangle$ 」라 하는 추가정의문의 좌변과 동일문장을 검색한다(스텝 2802). 검색의 결과, 해당하는 정의문이 존재하면, 스텝 2803에 있어서 각 지명표기 문자열이 OR결합되도록 검색결과의 정의문의 좌변을 전개한다. 스텝 2804에서는 변형한 지명표기사전중의 정의문에 대해서 추가정의문을 그 좌변을 OR결합함으로써 통합한다. 이것에 의해

$\langle 4570012 \rangle ::= \langle C\text{명고옥시남구} \rangle \langle T\text{입사} \rangle \langle V\text{본전} \rangle \mid \langle V\text{본전} \rangle;$

라 하는 지명표기정의문이 생성된다.

다음에 스텝 (2805)에 있어서, 정의식의 최적화, 즉 사용하는 구문카테고리명의 수의 삭감을 행한다. 이것에 의해

$\langle 4570012 \rangle ::= \langle C\text{명고옥시남구} \rangle \langle T\text{입사} \rangle [\langle V\text{본전} \rangle];$

가 얻어진다.

이상의 처리를 행하여 도 1에 있어서 지명조합용 지명표기사전(103)을 생성한다. 전에도 서술한 바와 같이, 이 지명표기사전의 형식은 사전 소스 파일의 형식을 갖고 있다. 문맥자유문법으로 표현된 지명표기 사전으로부터 그래프형식의 사전 바이너리 형식으로 변환하는 수법으로서, 예컨대 「자연언어처리입문」(근대과학사, ISBN4-7649-0143-9) pp. 19-31에 있는 것과 같은, 생성규칙에 따른 천이그래프를 생성하는 방법 등을 사용할 수 있다. 본 실시예에서 사용된 지명표기의 문맥자유문법 표현의 예로부터 제조되는 그래프형식의 지명표기사전의 일예를 도 29에 나타낸다. 그래프형식의 지명표기사전은 각 변이 부분문자열에 각 항점이 부분문자열의 경계에 대응하는 유향그래프이다. 각 변의 방향은

문자열중의 문자의 순서에 일치한다. NULL로 기재된 변은 그 개소에 어느 것도 문자열이 없어도 좋은 것을 나타낸다. 도면중 오른쪽 아래에 선을 넣은 원은 지명문자열의 개시위치를 나타낸다. 또한 중앙에 사선이 들어 있는 원은 문자열의 마지막의 위치를 나타낸다. 또한, 문자열의 마무리에 대응하는 노드(node)에는 〈숫자〉표기에서 지정한 지명표기에 대한 식별자로서의 숫자(본 실시예에서는 우편번호)가 저장된다.

영어권의 지명표기의 이표기에 관해서도 동일하게 취급하는 것이 가능하다.

예컨대 "South 9th street"에 있어서, 도 30에 나타난 바와 같이 "South"를 "S"로 약기하거나, 전체를 생략하거나 한다. 또한 "9th"에 관해서도 "9"를 영어표기하는 것이다. 또한 "Street"에 관해서는 "St"로 약기하거나, 전체를 생략하는 것이다. 그리고, "South"나 "9th"에 상당하는 부분의 어순이 변화하는 것도 있다.

상기의 지명표기를 본 발명에 있어서 나타나는 문맥자유문법에 따라 기술하면, 도 31과 같이 된다.

본 실시예에 있어서 하드웨어의 구성을 도 32에 나타낸다. 지명표기사전 생성장치(3001)는 표준적인 지명표기가 단어마다 분할되어 저장된 표준적 지명표기 리스트(3002)를 입력하고, 문맥자유문법으로 표현된 지명표기사전(3003)을 출력하는 장치이다. 그 내부는 외부로부터의 파일입력을 처리하는 입력용 인터페이스(3004), 각종 파일을 보존하여 두는 외부기억장치(3005), 연산처리장치(3006), 메모리(3007), 외부로의 결과출력을 처리하는 네트워크 인터페이스(3008), 출력용 인터페이스(3009)를 갖추고, 이들이 버스(3010)를 통하여 접속되어 있다.

표준적 지명표기 리스트는 입출력장치(3004)를 통하여 지명표기사전 생성장치에 입력되고, 외부기억장치(3005)에 저장되고, 표준적 지명표기 리스트 파일(3011)로 되어, 메모리(3007)상에 복사되고, 표준지명표기(3016)로 된다. 지명표기사전 생성처리가 실행되면, 외부기억장치중에 저장되어 있는 지명표기사전 생성처리를 위한 프로그램파일(3012), 이표기추가용 지식 파일(3013)이 메모리상에 로드되고, 프로그램(3014), 이표기추가용 지식(3015)으로 된다. 입력용 인터페이스 또는 외부기억장치로부터 메모리상에 로드된 표준적 지명표기(3016)에 대해서 연산처리장치는 본 발명의 방식에 따라 지명표기사전(3017)을 메모리상에 생성한다.

그 결과는 외부기억장치상에 지명표기사전 파일(3021)로서 씌여지지만, 디스플레이 등의 출력 인터페이스(3009)에 의해 출력된다. 프로그램(3014)을 통하여 추가·수정된 지명표기는 이표기추가용 지식파일(3022)로서 외부기억장치상에 보존되고, 이후의 지명표기사전 작성의 경우에 이용된다.

또한, 본 실시예로서 서술된 사전 소스 파일 형식으로부터 사전 바이너리 형식으로 변환하는 사전형식 변환장치(3018)에 의해 사전 바이너리 형식의 지명표기사전으로 변환하고, 이것을 네트워크(3019)를 통하여 우편구분기(3020)에 로드하고, 성이를 판독처리를 행하는 것도 가능하다.

본 발명은 설명의 편의상, 일본의 지명이나 주소를 예로 설명하였지만, 본 발명은 일본의 지명이나 주소 또는 일본어로 표기된 지명이나 주소에 한정되는 것은 아니다.

제외국의 지명이나 표기 또는 영어나 다른 언어로 표기된 지명이나 주소이더라도, 본 발명을 실시하는 것은 가능하다. 상술한 카테고리만이 아니고, 카테고리를 수시로 새롭게 추가·변경하거나, 그들 카테고리에 속하는 단어의 추가·변경을 행하는 것, 본 발명을 외국의 지명·주소 또는 외국어로 표기된 지명·주소로도 적용가능하게 된다.

발명의 효과

이상 서술한 바와 같이, 본 발명의 실시예에 의하면 1지역에 대해서 표준적인 지명표기만을 기재한 지명표기 리스트로부터 이표기를 가능한 한 자동적으로 추가하고, 사전완비율이 높은 지명표기사전을 용이하게 작성할 수 있다. 또한 자동적으로 추가할 수 없었던 이표기에 대해서도 이것을 수작업에 의해 추가하는 수단을 제공하는 것에 의해 지명표기사전을 작성하는 대상으로 되는 지역에 특유한 이표기를 추가할 수 있고, 사전완비율을 또한 높힐 수 있다. 또한, 여기에서의 수작업에 의한 추가작업의 결과를 보존하고, 이후의 사전작성처리의 경우에 이용함으로써, 수작업에 의해 추가작업을 행하는 회수를 억제하고, 사전작성작업을 간단화하는 것이 가능하게 된다.

(57) 청구의 범위

청구항 1

제 1의 지명을 유지하는 지명표기 리스트(101)를 포함하는 지명표기사전으로부터 상기 제 1의 지명을 입력하고, 출력하는 인터페이스 수단과,

상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가 가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전과,

상기 제 1의 지명과, 상기 이표기사전에 포함되는 상기 이표기의 정보를 입력하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 이표기지명 생성수단을 갖는 사전작성장치.

청구항 2

제 1의 지명을 유지하는 지명표기 리스트를 포함하는 지명표기사전과,

상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가

가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전과,

상기 제 1의 지명과, 상기 이표기사전에 포함되는 상기 이표기의 정보를 입력하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 이표기지명 생성수단을 갖는 사전작성장치.

청구항 3

제 2항에 있어서,

상기 이표기지명 작성수단은 또한 상기 제 1의 지명을 다른 언어형식으로 변환하는 형식언어표현변환수단을 포함하고, 상기 다른 언어형식으로 변환된 제 1의 지명을 참조하여 상기 제 2의 지명을 생성하는 이표기지명 작성수단인 것을 특징으로 하는 사전작성장치.

청구항 4

제 2항에 있어서,

상기 지명표기 리스트는 복수의 구문카테고리로 나누어 상기 제 1의 지명을 유지한 지명표기 리스트이고,

상기 이표기 지명 생성수단은 상기 복수의 구문카테고리의 어느 것을 참조하여, 상기 제 2의 지명을 생성하는 이표기 지명 생성수단인 것을 특징으로 하는 사전작성장치.

청구항 5

제 4항에 있어서,

상기 지명표기 리스트는 상기 복수의 구문카테고리로 나누어진 상기 제 1의 지명의 각각에 코드정보를 부여하여 유지한 지명표기 리스트인 것을 특징으로 하는 사전작성장치.

청구항 6

제 5항에 있어서,

상기 제 2의 지명은 상기 복수의 구문카테고리로 나누어진 상기 제 1의 지명의, 상기 복수의 구문카테고리의 배열방향을 변화시킨 지명인 것을 특징으로 하는 사전작성장치.

청구항 7

제 1의 지명을 유지하는 지명표기 리스트를 포함하는 지명표기사전과,

상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가 가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전과,

상기 제 1의 지명과, 상기 이표기사전에 포함되는 상기 이표기의 정보를 입력하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 이표기지명 생성수단과,

상기 제 2의 지명과 표기가 다른 제 3의 지명을 추가하는 지정을 접수하는 접수수단과,

상기 지정에 따라서 상기 제 3의 지명을 추가하여 출력하는 추가수단을 갖는 사전작성장치.

청구항 8

제 1의 지명을 유지하는 지명표기 리스트를 포함하는 지명표기사전과,

상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가 가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전과,

상기 제 1의 지명과, 상기 이표기사전에 포함되는 상기 이표기의 정보를 입력하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 이표기지명 생성수단과,

상기 제 2의 지명과 표기가 다른 제 3의 지명을 추가하는 지정을 접수하는 접수수단과,

상기 지정에 따라서 상기 제 3의 지명을 추가하여 출력하는 추가수단과,

상기 제 2의 지명과 상기 제 3의 지명을 입력하고, 상기 제 2의 지명과 상기 제 3의 지명의 쌍을 보존하는 보존수단을 갖는 사전작성장치로서,

상기 이표기지명 생성수단은 상기 보존수단을 참조하는 수단을 포함하고,

상기 보존수단에 상기 제 2의 지명과 상기 제 3의 지명의 쌍이 이미 보존되어 있는 경우에는 상기 쌍을 참조하여 상기 제 1의 지명으로부터 상기 제 3의 지명을 생성하여 출력하는 수단을 포함하는 이표기지명 생성수단인 것을 특징으로 하는 사전작성장치.

청구항 9

인터페이스수단은 제 1의 지명을 유지하는 지명표기 리스트를 포함하는 지명표기사전으로부터 상기 제 1

의 지명을 입력하여 출력하고,

이표기지명 생성수단은 상기 제 1의 지명과, 상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가 가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전으로부터 상기 이표기의 정보를 수작업하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 사전작성방법.

청구항 10

이표기지명 생성수단은, 제 1의 지명을 유지하는 지명표기 리스트를 포함하는 지명표기사전으로부터 상기 제 1의 지명과, 상기 제 1의 지명의 표기에 관계되는 표기가 다른 지명, 문자가 다른 지명, 생략이 가능한 지명, 추가가 가능한 지명의 어느 것을 포함하는 이표기의 정보를 유지하는 지명표기추가용 지식을 포함하는 이표기사전으로부터 상기 이표기의 정보를 입력하여, 상기 제 1의 지명과 표기가 다른 제 2의 지명을 생성하고, 상기 제 2의 지명을 출력하는 사전작성방법.

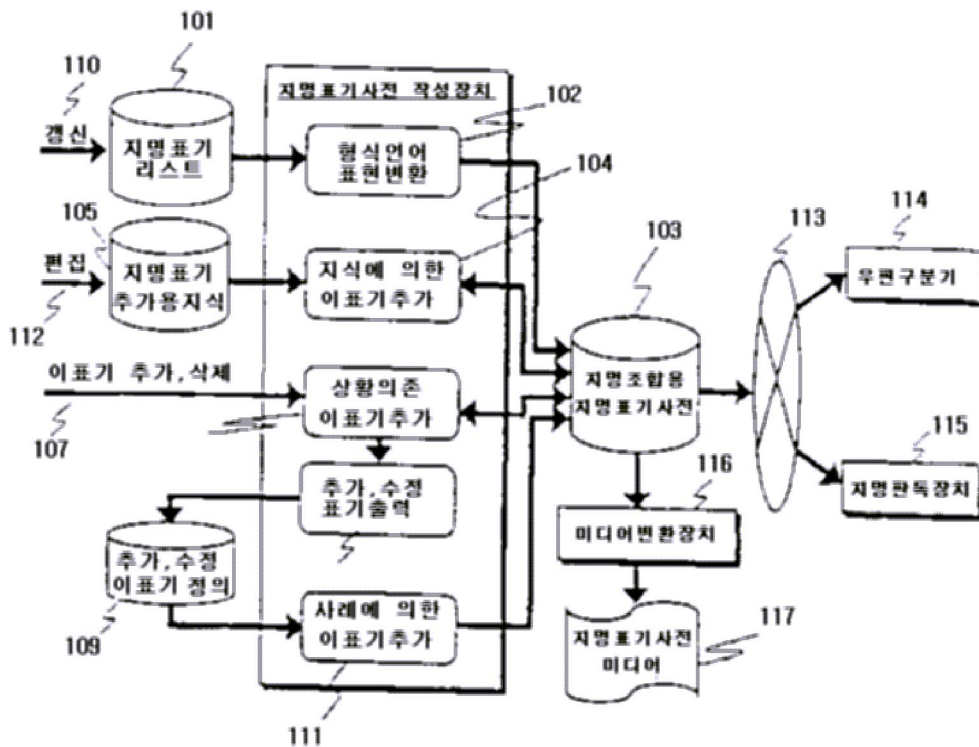
청구항 11

제 10항에 있어서,

상기 이표기지명 작성수단은 형식언어표현 변환수단으로 상기 제 1의 지명을 다른 언어형식으로 다시 변환하고, 상기 다른 언어형식으로 변환된 제 1의 지명을 참조하여 상기 제 2의 지명을 생성하는 것을 특징으로 하는 사전작성방법.

도면

도면1



도면2

201	202	203
지명표기 코드	지명표기명	지명표기후리가나
22	경기도	경기도
23	충청도	충청도
24	전라도	전라도

도면3

301 都道府県コード	302 市区郡町村コード	303 市区郡町村名	304 市区郡町村名ふり仮名
23	111	名古屋市港区	ナゴヤミナト
23	112	名古屋市南区	ナゴヤミナミ
23	113	名古屋市守山区	ナゴヤモリヤマ

도면4

401 都道府県 コード	402 市区郡町村 コード	403 町名・大字 コード	404 町名・大字	405 町名ふり仮名	406 異表記数	407 異表記
23	112	37	石元町	イシモト	1	石本町
23	112	40	笠原町	カサハラ	0	
23	112	44	塩田町	シホタ	0	

도면5

501 都道府県 コード	502 市区郡町村 コード	503 町名・大字 コード	504 字コード	505 字名	506 字 ふり仮名	507 異表記数	508 異表記
23	112	13	906	字宮瀬	ミヤノセ	0	0
23	112	13	907	字宮西	ミヤニシ	0	0
23	112	13	908	字寺坂	テラサカ	0	0

도면6

郵便番号	601 都道府県コード	602 市区郡町村コード	603 町名・大字コード	604 字コード	605
4570075	23	112	37		
4570051	23	112	40	15	
4570078	23	112	44		

도면7

```

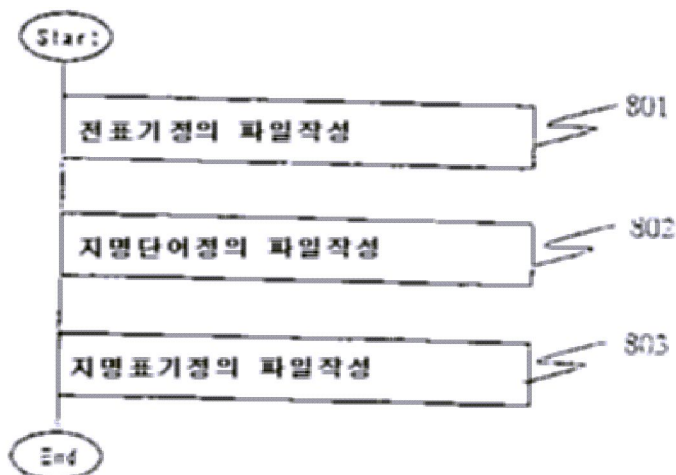
<字> ::= 字; /* 全国共通表現 */
begin /* 愛知県共通 */
  <愛知県> ::= 愛知県;
  <名古屋市> ::= 名古屋市;
  <南区> ::= 南区;
  <名古屋市南区> ::= [(愛知県)(名古屋市)(南区)];

  begin /* 名古屋市南区共通 */
    <本郷町> ::= 本郷町;
    <通間> ::= 通間;
    <宮崎> ::= 宮崎;
    <瀬部> ::= 瀬部;

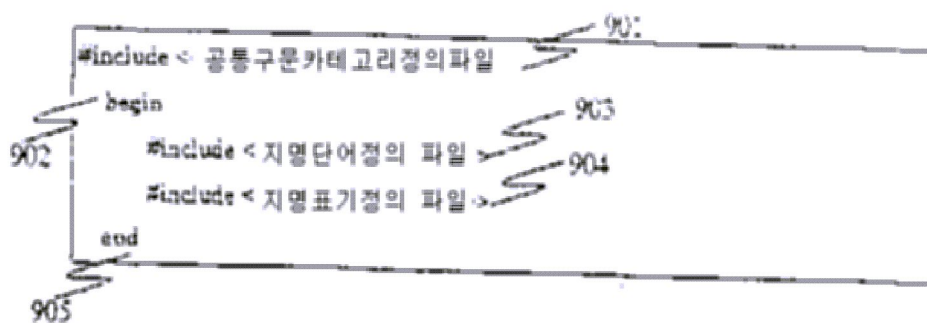
    begin(4570055)
      (4570055) ::= <名古屋市南区>(本郷町)
                    [(字)(通間)(宮崎)(瀬部)];
    end
  end
end
end

```

도면8



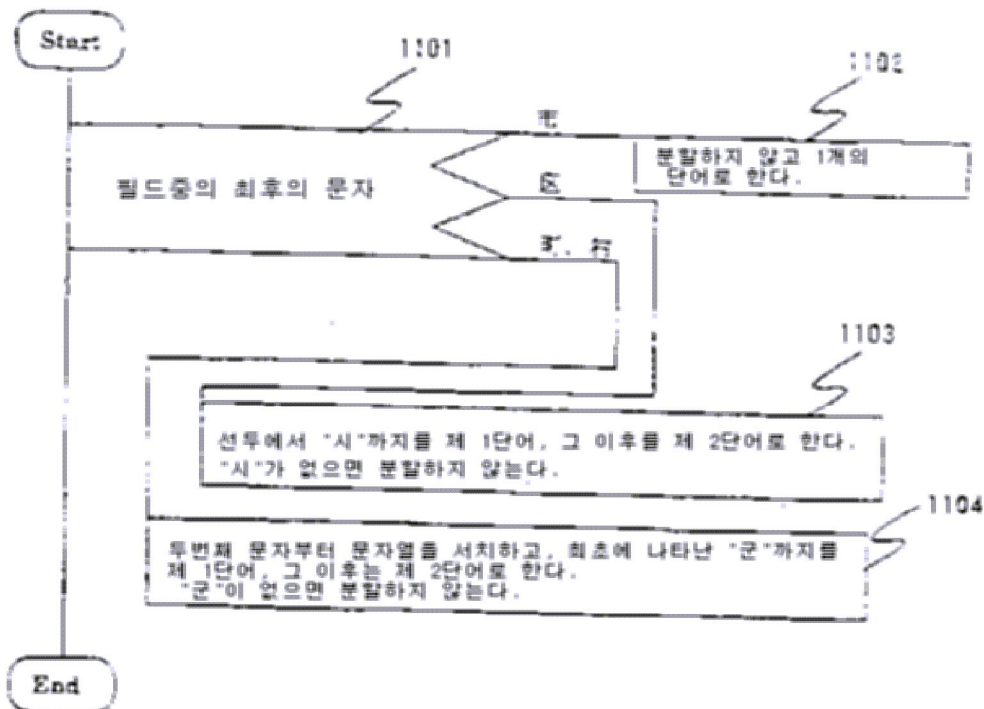
도면9



도면10

/*123*/	<P 愛知県> ::= 愛知県;	
/*123:112*/	<C 名古屋市南区> ::= 名古屋市南区;	1001
/*123:226*/	<C 尾張旭市> ::= 尾張旭市 旭市;	1002
/*123:304*/	<C 愛知県長久手町> ::= 愛知県長久手町;	1003
/*123:112:40*/	<T 大幸寺> ::= 大幸寺;	1004
/*123:112:44*/	<T 瑞穂町> ::= 瑞穂町;	
/*123:112:45*/	<T 大蔵通> ::= 大蔵通;	
/*123:112:40:908*/	<V 松葉> ::= 松葉;	
/*123:112:40:907*/	<V 上新町> ::= 上新町;	

도면11



도면12

/*123*/	<P 愛知県> ::= 愛知県;
	<O 名古屋市> ::= 名古屋市;
	<W 南区> ::= 南区;
/*123:112*/	<C 名古屋市南区> ::= <P 愛知県><O 名古屋市><W 南区>;
	<O 尾張旭市> ::= 尾張旭市 旭市;
/*123:226*/	<C 尾張旭市> ::= <P 愛知県><O 尾張旭市>;
	<O 愛知県> ::= 愛知県;
	<W 長久手町> ::= 長久手町;
/*123:304*/	<C 愛知県長久手町> ::= <P 愛知県><O 愛知県><W 長久手町>;

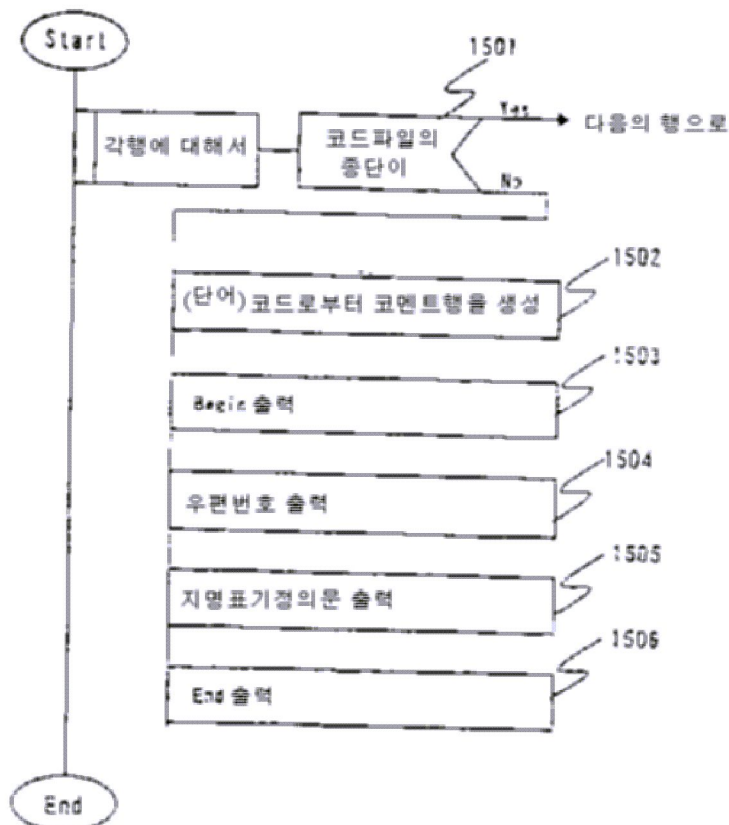
도면13

/*23*/	<P 愛知県> ::= 愛知県;
	<O 名古屋市> ::= 名古屋市;
	<W 南区> ::= 南区;
	<O 尾張旭市> ::= 尾張旭市 ; 旭市;
	<O 愛知県> ::= 愛知県;
	<W 長久手町> ::= 長久手町;
/*23/112/40*/	<T 大字> ::= 大字 ;
/*23/112/44*/	<T 大字町> ::= 大字町;
/*23/112/45*/	<T 大字通> ::= 大字通;
/*23/112/40/906*/	<V 松葉> ::= 松葉;
/*23/112/40/907*/	<V 上新町> ::= 上新町;

도면14

/*A23/112*/	<C 名古屋市南区> ::= <P 愛知県><O 名古屋市><W 南区>;
/*A23/226*/	<C 尾張旭市> ::= <P 愛知県><O 尾張旭市>;
/*A23/304*/	<C 愛知県長久手町> ::= <P 愛知県><O 愛知県><W 長久手町>;

도면15

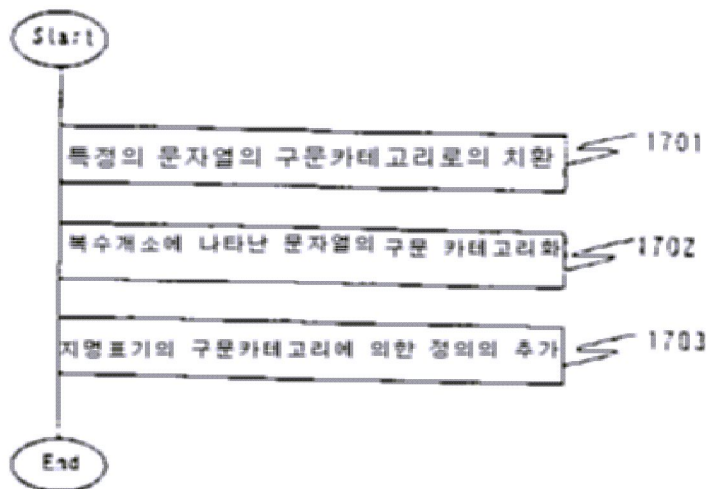


도면 16

```

/*23:112:75*/ begin (4570034)
    <4570034> ::= <C 名古屋市南区><T 春日野町>;
end
/*23:112:40*/ begin (4570051)
    <4570051> ::= <C 名古屋市南区><T 笠寺町>
    (<V 下新町>|<V 市場>);
end
    
```

도면 17



도면 18

18A

```

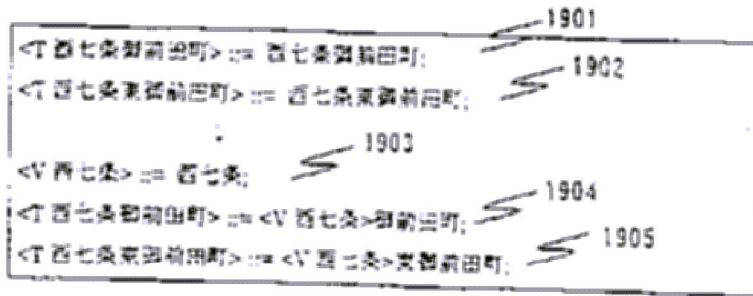
/*23:112:40*/ <T 笠寺> ::= 笠寺;
/*23:112:45*/ <T 大字大藏通> ::= 大字大藏通;
               <T 堀ヶ丘> ::= 堀ヶ丘;
               <T 上の山町> ::= 上の山町;
    
```

18B

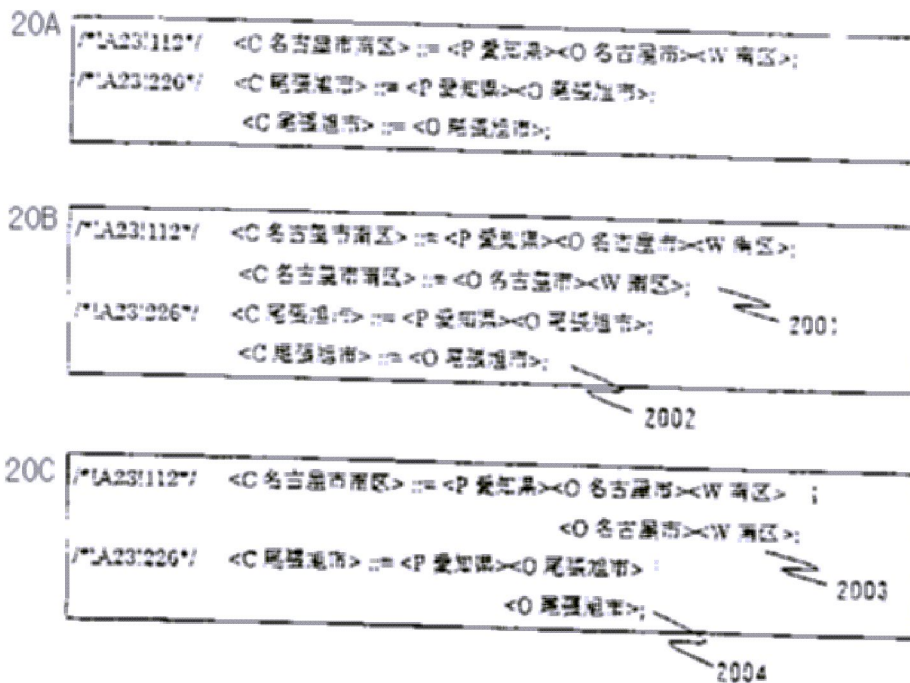
```

/*23:112:40*/ <T 笠寺> ::= 笠寺;
/*23:112:45*/ <T 大字大藏通> ::= <V 大字>大藏通;
               <T 堀ヶ丘> ::= 堀<V ケ>丘;
               <T 上の山町> ::= 上<V ノ>山町;
    
```

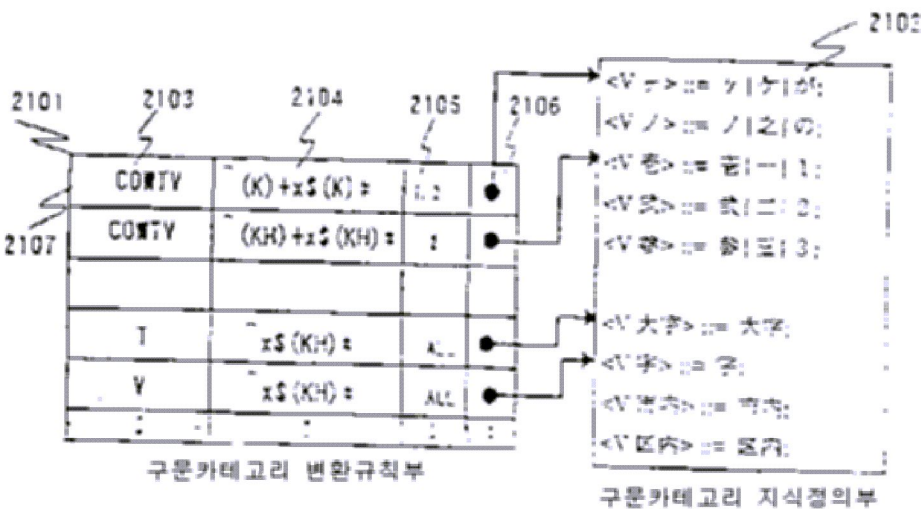

도면19



도면20



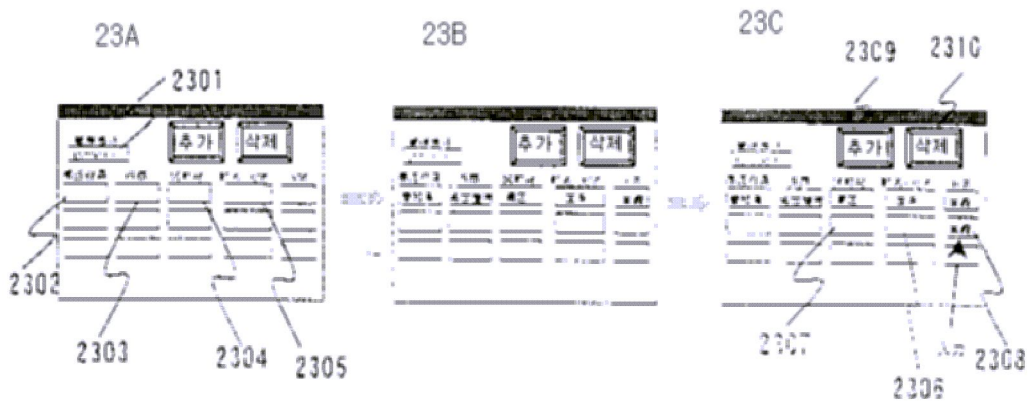
도면21



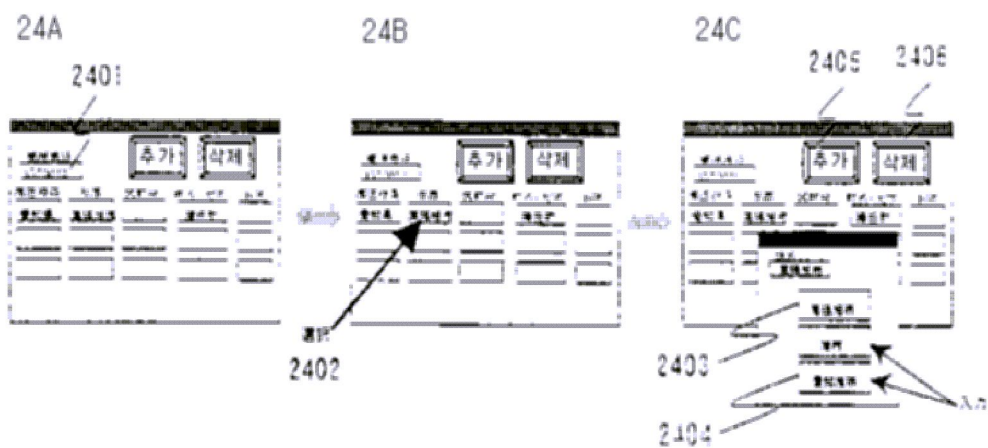
도면22

2201	2202	2203	2204
<P.+都>	SKIP	1	
<P.+道>	SKIP	1	
<P.+府>	SKIP	1	
<P.+県>	SKIP	1	
<P.+県><O.+><市.+>	SKIP	2	
<C.+区>	SKIP	1	
<P.+都><O.+市>	REPLACE	2	市内
<P.+道><O.+市>	REPLACE	2	市内
<P.+府><O.+市>	REPLACE	2	市内
<P.+県><O.+市>	REPLACE	2	市内
...

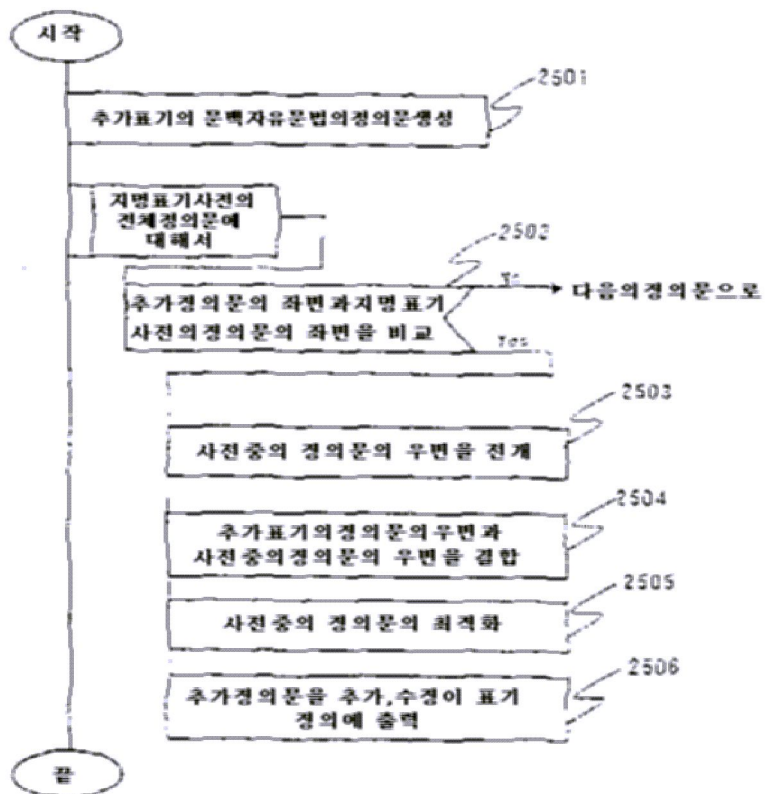
도면23



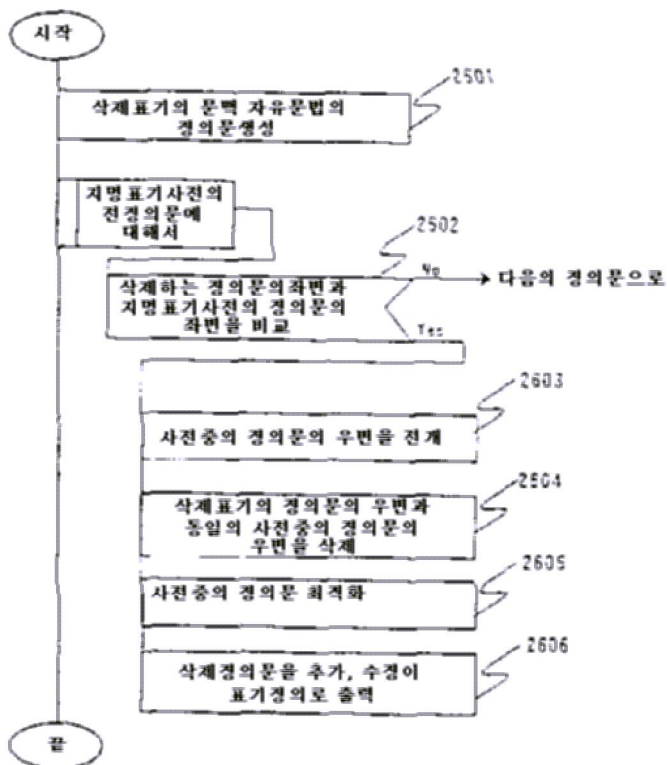
도면24



도면25



도면26



도면30

South 9th Street

S 9th St	S Ninth Street
South 9th St	South Ninth Street
9th St	Ninth Street
S Ninth St	S 9th
South Ninth St	South 9th
Ninth St	9th
S 9th Street	S Ninth
South 9th Street	South Ninth
9th Street	Ninth

도면31

<Wsouth> :- South | south | S ;
 <W9th> :- Ninth | ninth | 9th | 9 ;
 <Wstreet> :- Street | street | St | st ;
 ~ Wsouth9thstreet ~ :- ~ Wsouth ~ W9th ~ Wstreet ~ ;

도면32

