



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0082559
(43) 공개일자 2020년07월08일

(51) 국제특허분류(Int. Cl.)

G06F 40/20 (2020.01)

(52) CPC특허분류

G06F 40/268 (2020.01)

(21) 출원번호 10-2018-0173254

(22) 출원일자 2018년12월31일

심사청구일자 2018년12월31일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자

권오병

서울특별시 강남구 삼성로 150, 107동 1001호

최석재

서울특별시 구로구 구일로2길 60, 201동 501호

(74) 대리인

김홍석

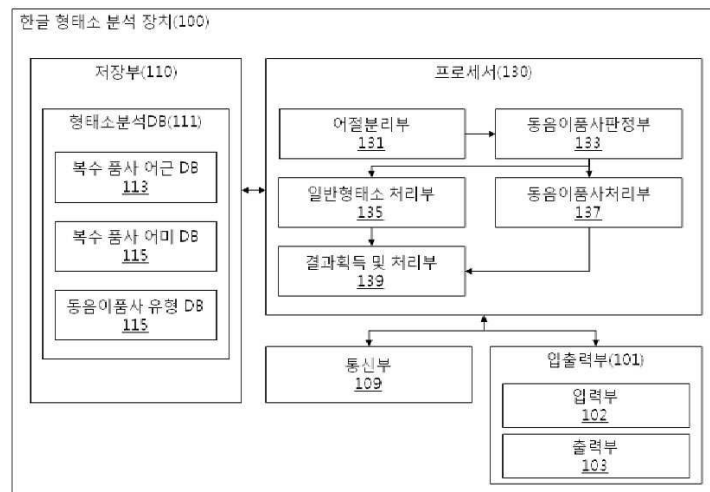
전체 청구항 수 : 총 21 항

(54) 발명의 명칭 한글 형태소 분석 장치 및 방법

(57) 요약

한글 형태소 분석 장치 및 방법에 관한 것으로, 한글 형태소 분석 장치는, 형태소 분석 데이터베이스 및 상기 형태소 분석 데이터베이스를 기반으로 분석 대상 어절이 동음이품사인지 여부를 판단하고, 상기 분석 대상 어절이 동음이품사인 경우, 복수 품사 판정 함수를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 프로세서를 포함할 수 있다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 2017-0-01122

부처명 미래창조과학부

연구관리전문기관 정보통신기술진흥센터

연구사업명 ICT유망기술개발지원(ICT융합핵심기술개발지원)

연구과제명 지능형 잇츠마이스토리 서비스 제공을 위한 언어 사용패턴 분석 기반 개인 프로파일링 및
개인맞춤형 챗봇 기술 개발

기 여 율 1/1

주관기관 경희대학교 산학협력단

연구기간 2017.05.01 ~ 2018.12.31

명세서

청구범위

청구항 1

형태소 분석 데이터베이스; 및

상기 형태소 분석 데이터베이스를 기반으로 분석 대상 어절이 동음어파사인지 여부를 판단하고, 상기 분석 대상 어절이 동음어파사인 경우, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 프로세서;를 포함하는 한글 형태소 분석 장치.

청구항 2

제1항에 있어서,

상기 프로세서는, 상기 분석 대상 어절이 동음어파사가 아니라고 판단되면, 상기 분석 대상 어절에 대한 일반적인 형태소 분석을 수행하는 한글 형태소 분석 장치.

청구항 3

제1항에 있어서,

상기 형태소 분석 데이터베이스는,

복수 품사로 활용될 수 있는 적어도 하나의 어근으로 구축된 복수 품사 어근 데이터베이스; 및

복수 품사로 활용될 수 있는 적어도 하나의 어미로 구축된 복수 품사 어미 데이터베이스;를 포함하는 한글 형태소 분석 장치.

청구항 4

제3항에 있어서,

상기 프로세서는,

상기 분석 대상 어절 내에서 추출된 어느 하나의 형태소가 상기 복수 품사 어근 데이터베이스에 존재하고, 다른 하나의 형태소가 상기 복수 품사 어미 데이터베이스에 존재하면, 단일 품사 판정 함수를 기반으로 상기 분석 대상 어절의 적어도 하나의 형태소의 품사를 판단하는 한글 형태소 분석 장치.

청구항 5

제4항에 있어서,

상기 단일 품사 판정 함수는, 일반 명사 판정 함수, 고유 명사 판정 함수, 의존 명사 판정 함수, 수사 판정 함수, 일반 동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및 어근 판정 함수 중 적어도 하나를 포함하는 한글 형태소 분석 장치.

청구항 6

제3항에 있어서,

상기 형태소 분석 데이터베이스는,

동음어파사에 대한 적어도 하나의 유형을 포함하는 동음어파사 유형 데이터베이스;를 더 포함하는 한글 형태소 분석 장치.

청구항 7

제6항에 있어서,

상기 프로세서는, 상기 분석 대상 어절 내에서 추출된 적어도 하나의 형태소가 복수 품사 어근 데이터베이스 및

복수 품사 어미 데이터베이스에 존재하지 않는 경우, 상기 동음이품사 유형 데이터베이스를 열람하는 한글 형태소 분석 장치.

청구항 8

제7항에 있어서,

상기 프로세서는, 상기 적어도 하나의 형태소가 상기 동음이품사 유형 데이터베이스의 열람 결과를 기반으로 상기 복수 품사 판정 함수를 기반으로 상기 적어도 하나의 형태소의 품사를 판단하는 한글 형태소 분석 장치.

청구항 9

제8항에 있어서,

상기 복수 품사 판정 함수는, 서로 상이한 단일 품사 판정 함수를 적어도 둘 이상 조합하여 획득된 것인 한글 형태소 분석 장치.

청구항 10

제9항에 있어서,

상기 서로 상이한 품사 판정 함수는, 동사 판정 함수, 명사 판정 함수 및 어근 판정 함수 중 적어도 둘을 포함하는 한글 형태소 분석 장치.

청구항 11

문장이 입력되는 단계;

문장으로부터 어절을 분리하는 단계;

분석 대상 어절이 동음이품사인지를 여부를 형태소 분석 데이터베이스를 기반으로 판단하는 단계; 및

상기 분석 대상 어절이 동음이품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계;를 포함하는 한글 형태소 분석 방법.

청구항 12

제11항에 있어서,

상기 분석 대상 어절이 동음이품사가 아니라고 판단되면, 상기 분석 대상 어절에 대한 일반적인 형태소 분석을 수행하는 단계;를 더 포함하는 한글 형태소 분석 방법.

청구항 13

제11항에 있어서,

상기 형태소 분석 데이터베이스는,

복수 품사로 활용될 수 있는 적어도 하나의 어근으로 구축된 복수 품사 어근 데이터베이스; 및

복수 품사로 활용될 수 있는 적어도 하나의 어미로 구축된 복수 품사 어미 데이터베이스;를 포함하는 한글 형태소 분석 방법.

청구항 14

제13항에 있어서,

상기 분석 대상 어절이 동음이품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는,

상기 분석 대상 어절 내에서 추출된 어느 하나의 형태소가 상기 복수 품사 어근 데이터베이스에 존재하고, 다른 하나의 형태소가 상기 복수 품사 어미 데이터베이스에 존재하면, 단일 품사 판정 함수를 기반으로 상기 분석 대상 어절의 적어도 하나의 형태소의 품사를 판단하는 단계;를 포함하는 한글 형태소 분석 방법.

청구항 15

제14항에 있어서,

상기 단일 품사 판정 함수는, 일반 명사 판정 함수, 고유 명사 판정 함수, 의존 명사 판정 함수, 수사 판정 함수, 일반 동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및 어근 판정 함수 중 적어도 하나를 포함하는 한글 형태소 분석 방법.

청구항 16

제13항에 있어서,

상기 형태소 분석 데이터베이스는,

동음이품사에 대한 적어도 하나의 유형을 포함하는 동음이품사 유형 데이터베이스;를 더 포함하는 한글 형태소 분석 방법.

청구항 17

제16항에 있어서,

상기 분석 대상 어절이 동음이품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는,

상기 분석 대상 어절 내에서 추출된 적어도 하나의 형태소가 복수 품사 어근 데이터베이스 및 복수 품사 어미 데이터베이스에 존재하지 않는 경우, 상기 동음이품사 유형 데이터베이스를 열람하는 단계;를 포함하는 한글 형태소 분석 방법.

청구항 18

제17항에 있어서,

상기 분석 대상 어절이 동음이품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는,

상기 적어도 하나의 형태소가 상기 동음이품사 유형 데이터베이스의 열람 결과를 기반으로 상기 복수 품사 판정 함수를 기반으로 상기 적어도 하나의 형태소의 품사를 판단하는 단계;를 더 포함하는 한글 형태소 분석 방법.

청구항 19

제18항에 있어서,

상기 복수 품사 판정 함수는, 서로 상이한 단일 품사 판정 함수를 적어도 둘 이상 조합하여 획득된 것인 한글 형태소 분석 방법.

청구항 20

제19항에 있어서,

상기 서로 상이한 품사 판정 함수는, 동사 판정 함수, 명사 판정 함수 및 어근 판정 함수 중 적어도 둘을 포함하는 한글 형태소 분석 방법.

청구항 21

분석 대상 어절을 포함하는 문장이 입력되는 단말 장치;

상기 문장을 수신하고, 상기 문장으로부터 상기 분석 대상 어절을 획득하고, 형태소 분석 데이터베이스를 기반으로 상기 분석 대상 어절이 동음이품사인지 여부를 판단하고, 상기 분석 대상 어절이 동음이품사인 경우, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 한글 형태소 분석 장치;를 포함하는 한글 형태소 분석 시스템.

발명의 설명

기술 분야

[0001] 한글 형태소 분석 장치 및 방법에 관한 것이다.

배경 기술

[0002] 텍스트 분석은, 다양한 분야에서 이용되고 있다. 텍스트 분석은 각종 텍스트(예를 들어, 문학 작품, 소셜 네트워크 서비스 등의 게시물, 기사, 답글, 후기 및/또는 상품 평 등)의 내용을 분석하고, 이를 기반으로 텍스트 자체로부터 가치 있는 정보를 추출하거나 통계적 처리를 수행하여 통계적 분석 결과를 획득하는 것을 의미한다. 예를 들어, 기업 또는 공공기관에서는 소비자나 국민의 인식이나 여론 등을 인터넷 상에서 작성된 문장들을 통해 확인하기 위해 텍스트 분석을 이용하고 있다. 이에 따라 텍스트 분석을 전문적으로 하는 사이트가 생겨나고 있고, 각 기업이나 공공기관의 사이트 내에도 텍스트 분석의 결과를 포함하려는 시도가 늘고 있다.

[0003] 이와 같은 텍스트 분석의 핵심은 형태소 분석의 성능에 달려 있다. 형태소 분석 결과에 따라서 텍스트의 의미가 상이하게 될 수 있기 때문이다. 다시 말해서, 형태소 분석이 적절하게 수행되지 않는 경우, 원래의 의도와는 상이한 다른 의도의 문장으로 해석될 수도 있기 때문에, 분석의 정확성 및 일관성이 저하될 수 있다.

[0004] 한국어는 형태소 분석이 매우 어려운 언어에 속한다. 이는 교착어의 특성에서 기인한다. 교착어는 실질 형태소에 형식 형태소가 부가된 언어로, 두 개 이상의 형태소가 하나의 음절에 융합되기 때문에 형태소 분리의 난이도가 높고, 또한 실질 형태소와 형식 형태소의 의미 영역이 상이하기 때문에 형태는 동일하나 그 의미는 상이한 동음이의어가 빈번히 발생하는 문제점이 존재한다. 특히, 동음이의어 중에서 그 외형적 형태는 동일하나 품사가 상이한 동음이품사는, 형태소 분석의 어려움을 더욱 증가시켜 형태소 분석 프로그램의 성능을 저하시키는 주요 원인이 되고 있다. 그러므로, 이와 같은 문장 내의 동음이품사의 효과적인 추출 및 처리는, 형태소 분석 및 전체 문장 분석의 정확성을 높이기 위해 필수적으로 요구될 수밖에 없다.

발명의 내용

해결하려는 과제

[0005] 동음이품사에 해당하는 단어나 형태소의 품사를 효율적으로 적절하게 결정함으로써, 텍스트를 보다 정확하게 분석하여 텍스트에 대한 분석 성능의 향상을 도모할 수 있는 한글 형태소 분석 장치 및 방법을 제공하는 것을 해결하고자 하는 과제로 한다.

과제의 해결 수단

[0006] 상술한 과제를 해결하기 위하여 한글 형태소 분석 장치, 시스템 및 방법이 제공된다.

[0007] 한글 형태소 분석 장치는, 형태소 분석 데이터베이스 및 상기 형태소 분석 데이터베이스를 기반으로 분석 대상 어절이 동음이품사인지를 여부를 판단하고, 상기 분석 대상 어절이 동음이품사인 경우, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 프로세서를 포함할 수 있다.

[0008] 상기 프로세서는, 상기 분석 대상 어절이 동음이품사가 아니라고 판단되면, 상기 분석 대상 어절에 대한 일반적인 형태소 분석을 수행할 수 있다.

[0009] 상기 형태소 분석 데이터베이스는, 복수 품사로 활용될 수 있는 적어도 하나의 어근으로 구축된 복수 품사 어근 데이터베이스 및 복수 품사로 활용될 수 있는 적어도 하나의 어미로 구축된 복수 품사 어미 데이터베이스를 포함할 수 있다.

[0010] 상기 프로세서는, 상기 분석 대상 어절 내에서 추출된 어느 하나의 형태소가 상기 복수 품사 어근 데이터베이스에 존재하고, 다른 하나의 형태소가 상기 복수 품사 어미 데이터베이스에 존재하면, 단일 품사 판정 함수를 기반으로 상기 분석 대상 어절의 적어도 하나의 형태소의 품사를 판단할 수 있다.

[0011] 상기 단일 품사 판정 함수는, 일반 명사 판정 함수, 고유 명사 판정 함수, 의존 명사 판정 함수, 수사 판정 함수, 일반 동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및 어근 판정 함수 중 적어도 하나를 포함할 수 있다.

[0012] 상기 형태소 분석 데이터베이스는, 동음이품사에 대한 적어도 하나의 유형을 포함하는 동음이품사 유형 데이터

베이스를 더 포함할 수도 있다.

- [0013] 상기 프로세서는, 상기 분석 대상 어절 내에서 추출된 적어도 하나의 형태소가 복수 품사 어근 데이터베이스 및 복수 품사 어미 데이터베이스에 존재하지 않는 경우, 상기 동음어품사 유형 데이터베이스를 열람할 수 있다.
- [0014] 상기 프로세서는, 상기 적어도 하나의 형태소가 상기 동음어품사 유형 데이터베이스의 열람 결과를 기반으로, 상기 복수 품사 판정 함수를 기반으로 상기 적어도 하나의 형태소의 품사를 판단할 수 있다.
- [0015] 상기 복수 품사 판정 함수는, 서로 상이한 단일 품사 판정 함수를 적어도 둘 이상 조합하여 획득된 것일 수 있다.
- [0016] 상기 서로 상이한 품사 판정 함수는, 동사 판정 함수, 명사 판정 함수 및 어근 판정 함수 중 적어도 둘을 포함할 수 있다.
- [0017] 한글 형태소 분석 방법은, 문장이 입력되는 단계, 문장으로부터 어절을 분리하는 단계, 분석 대상 어절이 동음어품사인지를 여부를 형태소 분석 데이터베이스를 기반으로 판단하는 단계 및 상기 분석 대상 어절이 동음어품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계를 포함할 수 있다.
- [0018] 한글 형태소 분석 방법은, 상기 분석 대상 어절이 동음어품사가 아니라고 판단되면, 상기 분석 대상 어절에 대한 일반적인 형태소 분석을 수행하는 단계를 더 포함할 수 있다.
- [0019] 상기 형태소 분석 데이터베이스는, 복수 품사로 활용될 수 있는 적어도 하나의 어근으로 구축된 복수 품사 어근 데이터베이스 및 복수 품사로 활용될 수 있는 적어도 하나의 어미로 구축된 복수 품사 어미 데이터베이스를 포함할 수 있다.
- [0020] 상기 분석 대상 어절이 동음어품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는, 상기 분석 대상 어절 내에서 추출된 어느 하나의 형태소가 상기 복수 품사 어근 데이터베이스에 존재하고, 다른 하나의 형태소가 상기 복수 품사 어미 데이터베이스에 존재하면, 단일 품사 판정 함수를 기반으로 상기 분석 대상 어절의 적어도 하나의 형태소의 품사를 판단하는 단계를 포함할 수 있다.
- [0021] 상기 단일 품사 판정 함수는, 일반 명사 판정 함수, 고유 명사 판정 함수, 의존 명사 판정 함수, 수사 판정 함수, 일반 동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및 어근 판정 함수 중 적어도 하나를 포함할 수 있다.
- [0022] 상기 형태소 분석 데이터베이스는, 동음어품사에 대한 적어도 하나의 유형을 포함하는 동음어품사 유형 데이터베이스를 더 포함할 수 있다.
- [0023] 상기 분석 대상 어절이 동음어품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는, 상기 분석 대상 어절 내에서 추출된 적어도 하나의 형태소가 복수 품사 어근 데이터베이스 및 복수 품사 어미 데이터베이스에 존재하지 않는 경우, 상기 동음어품사 유형 데이터베이스를 열람하는 단계를 포함할 수 있다.
- [0024] 상기 분석 대상 어절이 동음어품사이면, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 단계는, 상기 적어도 하나의 형태소가 상기 동음어품사 유형 데이터베이스의 열람 결과를 기반으로 상기 복수 품사 판정 함수를 기반으로 상기 적어도 하나의 형태소의 품사를 판단하는 단계를 더 포함할 수 있다.
- [0025] 상기 복수 품사 판정 함수는, 서로 상이한 단일 품사 판정 함수를 적어도 둘 이상 조합하여 획득된 것일 수 있다.
- [0026] 상기 서로 상이한 품사 판정 함수는, 동사 판정 함수, 명사 판정 함수 및 어근 판정 함수 중 적어도 둘을 포함할 수 있다.
- [0027] 한글 형태소 분석 시스템은, 분석 대상 어절을 포함하는 문장이 입력되는 단말 장치 및 상기 문장을 수신하고, 상기 문장으로부터 상기 분석 대상 어절을 획득하고, 형태소 분석 데이터베이스를 기반으로 상기 분석 대상 어절이 동음어품사인지를 여부를 판단하고, 상기 분석 대상 어절이 동음어품사인 경우, 단일 판정 함수 및 복수 품사 판정 함수 중 어느 하나를 기반으로 상기 분석 대상 어절에 대응하는 적어도 하나의 품사를 결정하는 한글

형태소 분석 장치를 포함할 수 있다.

발명의 효과

- [0028] 상술한 한글 형태소 분석 장치 및 방법에 의하면, 동음어품사에 해당하는 단어나 형태소의 품사를 간단하고 효율적이면서도 적절하게 결정할 수 있게 되고, 이에 따라 문자나 음성 등의 형태의 전체 텍스트를 보다 정확하게 분석할 수 있게 되어, 텍스트에 대한 분석 성능의 향상을 도모할 수 있게 된다.
- [0029] 상술한 한글 형태소 분석 장치 및 방법에 의하면, 형태소 분석이 매우 어려운 교착어(일례로 한국어)의 텍스트를 분석함에 있어서, 그의 형태소 분석을 보다 정확하게 용이하게 수행할 수 있게 되는 효과도 얻을 수 있다.
- [0030] 상술한 한글 형태소 분석 장치 및 방법은, 기사, 게시물, 댓글, 후기 및/또는 상품 평 등의 분석 시, 각 분야의 한글 말뭉치의 도구로 이용될 수 있거나, 및/또는 공공 기관이나 기업 등이 여론 조사나 평가 등을 수행하는 경우에 있어서 텍스트 분석의 핵심 알고리즘으로 이용될 수도 있는 등 다양한 텍스트 분석 시스템이나 프로그램에 적용 가능한 장점도 존재한다.

도면의 간단한 설명

- [0031] 도 1은 한글 형태소 분석 시스템의 일 실시예에 대한 도면이다.
- 도 2는 한글 형태소 분석 장치의 일 실시예에 대한 블록도이다.
- 도 3은 동음어품사 처리부의 일 실시예에 대한 블록도이다.
- 도 4는 단일 품사 판정 함수의 일례에 대한 프로그램 코드를 도시한 도면이다.
- 도 5는 한글 형태소 분석을 설명하기 위한 도표이다.
- 도 6은 복수 품사 판정 함수의 어느 하나의 일례에 대한 프로그램 코드를 도시한 도면이다.
- 도 7은 단일 품사 판정 함수의 연결의 일례를 설명하기 위한 도면이다.
- 도 8은 단일 품사 판정 함수의 연결의 일례에 대응하는 프로그램 코드를 도시한 도면이다.
- 도 9는 복수 품사 판정 함수의 일례를 설명하기 위한 도면이다.
- 도 10은 복수 품사 판정 함수의 일례에 대응하는 프로그램 코드를 도시한 도면이다.
- 도 11은 동사-일반 명사 함수의 일례에 대한 프로그램 코드를 도시한 도면이다.
- 도 12는 한글 형태소 분석 방법의 일 실시예에 대한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0032] 이하 명세서 전체에서 동일 참조 부호는 특별한 사정이 없는 한 동일 구성요소를 지칭한다. 이하에서 사용되는 '부'가 부가된 용어는, 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예에 따라 '부'가 하나의 부품으로 구현되거나, 하나의 '부'가 복수의 부품들로 구현되는 것도 가능하다.
- [0033] 명세서 전체에서 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 어떤 부분과 다른 부분에 따라서 물리적 연결을 의미할 수도 있고, 또는 전기적으로 연결된 것을 의미할 수도 있다. 또한, 어떤 부분이 다른 부분을 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 부분 이외의 또 다른 부분을 제외하는 것이 아니며, 설계자의 선택에 따라서 또 다른 부분을 더 포함할 수 있음을 의미한다.
- [0034] 제 1 이나 제 2 등의 용어는 하나의 부분을 다른 부분으로부터 구별하기 위해 사용되는 것으로, 특별한 기재가 없는 이상 이들이 순차적인 표현을 의미하는 것은 아니다. 또한, 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함할 수 있다.
- [0035] 이하 도 1 내지 도 12를 참조하여, 한글 형태소 분석 장치 및 이를 포함하는 한글 형태소 분석 시스템의 일 실시예에 대해서 설명하도록 한다.
- [0036] 도 1은 한글 형태소 분석 시스템의 일 실시예에 대한 도면이다.
- [0037] 도 1에 도시된 바를 참조하면, 한글 형태소 분석 시스템(1)은, 일 실시예에 있어서, 한글 형태소 분석 장치(100)를 포함할 수 있다. 한글 형태소 분석 장치(100)는, 사용자 또는 다른 컴퓨팅 장치(예를 들어, 단말 장치

(20)나 다른 서버 장치 등)로부터 입력되는 문장으로부터 한글 형태소 분석을 수행할 수 있도록 마련된다.

- [0038] 한글 형태소 분석 장치(100)는, 하나의 컴퓨팅 장치(예를 들어, 서버 장치)를 기반으로 구현될 수도 있고, 또는 둘 이상의 컴퓨팅 장치를 이용하여 구현될 수도 있다. 둘 이상의 컴퓨팅 장치를 이용하는 경우, 둘 이상의 컴퓨팅 장치는 직접적으로 또는 외부의 네트워크(9)를 통하여 상호 통신 가능하게 연결되어 있을 수 있다. 또한, 둘 이상의 컴퓨팅 장치를 이용하는 경우, 실시예에 따라서, 적어도 하나의 컴퓨팅 장치(들)는, 형태소 분석 데이터 베이스(도 2의 111)를 저장하고, 다른 적어도 하나의 컴퓨팅 장치(들)는 프로세서(도 2의 130)의 동작을 수행하도록 설계되는 것도 가능하다.
- [0039] 한글 형태소 분석 장치(100)는, 네트워크(9)에 접속하여 텍스트 또는 음성 형태의 적어도 하나의 문장(상황에 따라서 복수의 단어로 구성될 수도 있고, 또는 하나의 단어로만 구성될 수도 있다. 이하 동일하다)을 수신할 수도 있고, 또는 사용자로부터 직접 텍스트 또는 음성 형태의 적어도 하나의 문장을 입력 받을 수도 있다.
- [0040] 실시예에 따라서, 한글 형태소 분석 시스템(1)은, 적어도 하나의 단말 장치(10)를 더 포함할 수도 있다. 적어도 하나의 단말 장치(10)는, 예를 들어, 텍스트나 음성 형태의 문장을 입력 받거나 및/또는 한글 형태소 분석 결과를 수신하고 이를 외부로 출력할 수도 있다.
- [0041] 단말 장치(10)는, 유선 통신 네트워크, 무선 통신 네트워크 또는 이들의 조합을 기반으로 한글 형태소 분석 장치(100)와, 상호 통신을 수행할 수 있다. 여기서, 유선 통신 네트워크는, 케이블을 이용하여 구축 가능하며, 케이블은, 예를 들어, 페어 케이블, 동축 케이블, 광섬유 케이블 또는 이더넷 케이블 등을 이용하여 구현된 것일 수 있다. 무선 통신 네트워크는 근거리 통신 네트워크 및 원거리 통신 네트워크 중 적어도 하나를 이용하여 구현된 것일 수 있다. 근거리 통신 네트워크는, 예를 들어, 와이 파이(Wi-Fi), 와이파이 다이렉트(Wi-Fi Direct), 지그비(zigbee), 블루투스(Bluetooth), 저전력 블루투스(Bluetooth Low Energy), 캔(CAN) 통신 및/또는 엔에프 씨(NFC, Near Field Communication) 등을 이용하여 구현된 것일 수 있다. 원거리 통신 네트워크는, 유선 통신 네트워크예를 들어, 3GPP, 3GPP2 또는 와이맥스 계열 등의 이동 통신 표준을 기반으로 구현된 것일 수 있다.
- [0042] 단말 장치(10)는 셀룰러 폰, 스마트 폰, 태블릿 피씨, 데스크톱 컴퓨터, 랩톱 컴퓨터, 스마트 시계, 두부 장착형 디스플레이(HMD, Head Mounted Display) 장치, 디지털 텔레비전, 셋톱 박스, 내비게이션 장치, 휴대용 게임기, 개인용 디지털 보조기(PDA, Personal Digital Assistant), 전자 칠판, 전자 광고판 및/또는 음향 재생 장치 등을 포함할 수 있으나, 이에 한정되는 것은 아니다.
- [0043] 이하 한글 형태소 분석 장치(100)에 대해 보다 상세히 설명한다.
- [0044] 도 2는 한글 형태소 분석 장치의 일 실시예에 대한 블록도이다.
- [0045] 도 2에 도시된 바를 참조하면, 한글 형태소 분석 장치(100)는, 일 실시예에 의하면, 저장부(110) 및 프로세서(130)를 포함할 수 있으며, 실시예에 따라서 입출력부(101)나 통신부(109)를 더 포함할 수도 있다.
- [0046] 입출력부(101)는 외부로부터 문장 등이나 각종 데이터를 입력 받거나 및/또는 출력할 수 있도록 마련된다. 입출력부(101)는 입력부(102) 및 출력부(103) 중 적어도 하나를 포함 가능하다. 입력부(102)는, 사용자 등으로부터 문장 등을 입력받을 수 있도록 마련되며, 예를 들어, 키보드, 키패드, 터치스크린, 터치패드, 트랙볼, 트랙패드, 마우스 장치, 마이크로 폰, 스타일러스 펜, 데이터 입력 단자 및/또는 동작 감지 센서 등을 포함할 수 있다. 출력부(103)는 사용자에게 한글 형태소의 분석 결과를 시각적 및/또는 청각적으로 제공 가능하도록 마련된다. 출력부(103)는, 예를 들어, 디스플레이 패널, 조명 장치, 데이터 출력 단자 및/또는 스피커 장치 등을 포함할 수 있다.
- [0047] 통신부(109)는 통신 네트워크(9)에 접속하여 한글 형태소 분석 장치(100) 내의 다른 장치(100A, 100B 등) 및/또는 외부의 다른 장치(예를 들어, 단말 장치(10), 또는 외부의 서버 장치(미도시) 등)로 데이터를 전송하거나 및/또는 외부의 다른 장치로부터 데이터를 수신할 수 있도록 마련된다. 예를 들어, 통신부(109)는 단말 장치(10)로부터 문장을 소정의 형태(예를 들어, 텍스트 형태)로 수신할 수도 있다.
- [0048] 입출력부(101) 및/또는 통신부(109)는, 프로세서(130)와 전기적으로 연결되어, 프로세서(130)로 데이터를 전송하거나 또는 프로세서(130)로부터 데이터를 수신한다. 필요에 따라서, 입출력부(101) 및/또는 통신부(109)는 저장부(110)와도 전기적으로 연결되어 있을 수도 있다.
- [0049] 저장부(110)는, 한글 형태소 분석 장치(100)의 동작에 필요한 각종 데이터나 애플리케이션을 저장할 수 있다. 예를 들어, 저장부(110)는, 형태소 처리, 동음이의어 판정 및 처리 및/또는 결과 처리 등과 관련된 적어도 하나의 애플리케이션(프로그램이나 앱 등으로 지칭 가능하다)을 저장할 수 있다. 여기서, 저장부(110)에 저장된 애플리케이션

플리케이션은, 설계자에 의해 미리 작성되어 저장부(110)에 저장된 것일 수도 있고, 또는 유선 또는 무선 통신 네트워크(9)를 통해 접속 가능한 전자 소프트웨어 유통망을 통하여 획득 또는 갱신된 것일 수도 있다.

- [0050] 일 실시예에 의하면, 저장부(110)는 형태소 분석 데이터베이스(111)를 저장할 수도 있다. 형태소 분석 데이터베이스(111)는 프로세서(130)에 의해 호출되어 동음어품사의 판정이나, 동음어품사의 처리 등에 이용될 수 있다.
- [0051] 형태소 분석 데이터베이스(111)는, 예를 들어, 도 2에 도시된 바와 같이 복수 품사 어근 데이터베이스(113), 복수 품사 어미 데이터베이스(115) 및 동음어품사 유형 데이터베이스 중 적어도 하나를 포함할 수 있다.
- [0052] 복수 품사 어근 데이터베이스(113)는, 복수 품사를 가지는 어근(들)로 구축될 것일 수 있다. 복수 품사를 가지는 어근은, 문장 내에서 둘 이상의 품사로 해석될 수 있는 어근을 의미하며, 보다 상세하게는 [고려]와 같이 하나의 문장 내에서 명사(나라 이름) 및 동사(생각하다) 중 어느 하나로 이용될 수 있는 어근을 의미한다. 이러한 복수 품사를 가지는 어근으로는, 예를 들어, 가, 가까이, 강하, 견고, 경미, 그림, 꼬꼬, 내일, 너댓, 노래, 다행이, 대한, 도연, 매회, 모모, 반하, 사서, 사진도, 사실, 아연, 앞서, 잠깐, 있, 잉, 자, 자라, 지인, 차지, 참고, 타, 터, 피디, 한적, 할, 할거 화하 또는 회구 등이 존재한다. 이들 외에도 다양한 복수 품사를 가지는 어근이 존재한다. 이와 같은 복수 품사를 가지는 어근은 대략 1,690개 정도로 추산된다.
- [0053] 복수 품사 어근 데이터베이스(113)는, 이와 같은 복수 품사를 가지는 어근을 하나 이상 포함하여 구축될 수 있다. 복수 품사 어근 데이터베이스(113)는 추산되는 모든 복수 품사를 가지는 어근을 포함할 수도 있고, 또는 이들 중의 일부의 어근만을 포함할 수도 있다.
- [0054] 일 실시예에 의하면, 복수 품사 어근 데이터베이스(113)에 저장되는 어근은 체언, 부사, 감탄사, 관형사 또는 일부 어근과 같이 어형 변화를 하지 않는 용언을 포함하도록 마련된 것일 수도 있다.
- [0055] 복수 품사 어미 데이터베이스(115)는, 복수 품사를 가지는 어미(들)로 구축된 것일 수 있다. 복수 품사를 가지는 어미는, 문장 내에서 둘 이상의 품사로 판단될 수 있는 어미를 의미한다. 상세하게는 [가]와 같이 조사(-가), 명사(거리), 의문문 어미(-가?) 또는 동사(이동하다) 등과 같이 둘 이상의 품사로 해석될 수 있는 어미를 의미한다.
- [0056] 일 실시예에 의하면, 복수 품사 어미 데이터베이스(115)에 저장되는 어미는, 접사를 포함할 수 있으며, 실시예에 따라서 접사가 부가된 어형 변화가 발생하는 어근을 포함할 수도 있다.
- [0057] 복수 품사 어미 데이터베이스(115)에 저장되는 복수 품사를 가지는 어미는, 예를 들어, 가, 가지, 구간에, 구만, 나, 나는, 나니까, 낫, 낫다, 되셨습니다. 든, 만은, 만의, 시기에, 아사는, 아이가, 아이고, 우세요, 우실까, 울수록, 은, 을, 이가, 준, 준거야, 준다, 준다고, 준다구, 준다나봐, 준다냐, 준다는, 준다는데, 준다니, 준다니까, 준다니요, 준다던데, 준다데, 준다며, 준다면, 준다면서, 준다지만, 준다지요, 준다해도, 준단, 준단다, 준담, 준답니까, 준답니다, 준대, 준대니, 준대도, 준대요, 준대유, 준들, 준디야, 줄, 줄강, 줄걸, 줄게, 줄게요, 줄까, 줄까봐, 줄까요, 줄께, 줄께요, 줄는지, 줄달음친, 줄라, 줄라구, 줄라꼬, 줄라면, 줄란다, 줄랍니까, 줄래, 줄래도, 줄래요, 줄려고, 줄밖에, 줄뿐더러, 줄수록, 줄지, 줄니까, 줄니다 또는 줄디다 등을 포함할 수 있다. 이들 외에도 다양한 복수 품사를 가지는 어미가 존재한다. 이와 같은 복수 품사를 가지는 어미는 대략 265개로 추산된다.
- [0058] 상술한 복수 품사 어근 데이터베이스(113) 및 복수 품사 어미 데이터베이스(115)는 프로세서(113)의 동음어품사 판정부(133)에 의해 각각 호출될 수 있으며, 동음어품사의 판단을 위해 이용될 수 있다.
- [0059] 동음어품사 유형 데이터베이스(117)는, 동음어품사를 갖는 단어나 형태소 등의 유형으로 구축된 것일 수 있다. 예를 들어, 하나의 형태소는 문장 내에서 명사로 또는 동사로 이용될 수 있으며, 이는 명사-동사 유형으로 데이터베이스(117)에 저장된다. 이 경우, 동음어품사를 갖는 단어나 형태소는, 두 종류의 품사로 이용될 수 있는 경우도 있고, 또는 셋 이상의 품사로 이용될 수 있는 경우도 있다. 이와 같은 유형의 개수는 대략 78개로 추산된다.
- [0060] 구체적으로 예를 들어, 동음어품사 유형 데이터베이스(117)에 저장되는 유형은, 명사-감탄사, 명사-부사, 명사-관형사, 명사-의존명사, 명사-대명사, 명사-수사, 명사-형용사, 명사-동사, 명사-보조 용언, 명사-어근, 고유명사-감탄사, 고유명사-어근, 고유명사-관형사, 고유명사-부사, 고유명사-형용사, 고유명사-동사, 대명사-동사, 대명사-동사-부사, 대명사-어근, 수사-어근, 형용사-부사, 형용사-어근, 동사-부사, 동사-형용사, 동사-보조 용언, 동사-어근, 어근-부사, 감탄사-어근, 관형사-수사, 관형사-어근, 의존명사-부사, 의존명사-관형사, 의존명사-동사, 의존명사-어근, 명사-부사-어근, 명사-의존명사-부사, 명사-의존명사-부사, 명사-의존명사-관형사, 명

사-의존명사-대명사, 명사-의존명사-형용사, 명사-의존명사-동사, 명사-의존명사-어근, 명사-대명사-부사, 명사-대명사-관형사, 명사-대명사-수사, 명사-대명사-형용사, 명사-대명사-동사, 명사-대명사-어근, 명사-형용사-어근, 명사-동사-어근, 명사-동사-부사, 명사-동사-형용사, 명사-동사-보조 용언, 명사-어근-부사, 대명사-부사-어근, 대명사-동사-형용사, 대명사-동사-보조 용언, 대명사-동사-어근, 대명사-어근-부사, 동사-형용사-부사, 동사-형용사-보조 용언, 동사-보조 용언-부사, 동사-어근-부사, 명사-의존명사-부사-동사, 명사-의존명사-관형사-동사, 명사-의존명사-대명사-관형사, 명사-의존명사-대명사-동사, 명사-의존명사-형용사-보조 용언, 명사-의존명사-동사-형용사, 명사-의존명사-동사-보조 용언, 명사-의존명사-동사-어근, 명사-대명사-동사-어근, 명사-의존명사-대명사-관형사-어근, 명사-의존명사-대명사-동사-보조 용언, 명사-의존명사-대명사-동사-어근, 명사-의존명사-동사-형용사-보조 용언, 명사-의존명사-대명사-관형사-동사-보조 용언 및/또는 명사-의존명사-대명사-동사-형용사-어근 등을 포함할 수 있다. 실시예에 따라서, 동음이품사 유형 데이터베이스(117)는, 이들 유형 중 일부만을 포함할 수도 있고, 또는 이들 외에 다른 유형이 더 부가하여 포함할 수도 있다.

[0061] 이와 같은 동음이품사 유형 데이터베이스(117)는, 프로세서(130)의 동음이품사 처리부(137)에 의해 호출될 수 있으며, 동음이품사의 처리를 위해 이용될 수 있다.

[0062] 상술한 저장부(110)는, 실시예에 따라서, 주기억장치 및 보조기억장치 중 적어도 하나를 포함할 수 있다. 주기억장치는 롬(ROM) 및/또는 램(RAM)과 같은 반도체 저장 매체를 이용하여 구현될 것일 수 있다. 롬은, 예를 들어, 통상적인 롬, 이프롬(EPROM), 이이프롬(EEPROM) 및/또는 마스크롬(MASK-ROM) 등을 포함할 수 있다. 램은 예를 들어, 디램(DRAM) 및/또는 에스램(SRAM) 등을 포함할 수 있다. 보조기억장치는, 플래시 메모리 장치, SD(Secure Digital) 카드, 솔리드 스테이트 드라이브(SSD, Solid State Drive), 하드 디스크 드라이브(HDD, Hard Disc Drive), 자기 드럼, 콤팩트 디스크(CD), 디브이디(DVD) 또는 레이저 디스크 등과 같은 광 기록 매체(optical media), 자기 테이프, 광자기 디스크 및/또는 플로피 디스크 등과 같이 데이터를 영구적 또는 반영구적으로 저장 가능한 적어도 하나의 저장 매체를 이용하여 구현될 수 있다.

[0063] 프로세서(130)는, 한글 형태소 분석 장치(100)의 동작에 필요한 각종 연산, 판단, 처리 및/또는 제어 동작 등을 수행할 수 있다. 프로세서(130)는, 이와 같은 동작의 수행을 위해 저장부(110)에 저장된 애플리케이션을 구동시킬 수도 있다.

[0064] 이와 같은 프로세서(130)는, 예를 들어, 중앙 처리 장치(CPU, Central Processing Unit), 마이크로 컨트롤러 유닛(MCU, Micro Controller Unit), 마이컴(Micom, Micro Processor), 애플리케이션 프로세서(AP, Application Processor), 전자 제어 유닛(ECU, Electronic Controlling Unit) 및/또는 각종 연산 처리 및 제어 신호의 생성이 가능한 다른 전자 장치 등을 포함할 수 있다. 이들 장치는 예를 들어 하나 또는 둘 이상의 반도체 칩 및 관련 부품을 이용하여 구현 가능하다.

[0065] 도 2에 도시된 바를 참조하면, 프로세서(130)는, 어절 분리부(131), 동음이품사 판정부(133), 일반 형태소 처리부(135), 동음이품사 처리부(137) 및 결과 획득 및 처리부(139)를 포함할 수 있다. 어절 분리부(131), 동음이품사 판정부(133), 일반 형태소 처리부(135), 동음이품사 처리부(137) 및 결과 획득 및 처리부(139) 중 일부는, 필요에 따라 생략될 수도 있다. 또한, 이들(131 내지 139)은, 실시예에 따라서, 물리적으로 구분되는 것일 수도 있고, 논리적으로 구분되는 것일 수도 있다.

[0066] 어절 분리부(131)는, 문장에서 어절을 분리할 수 있다. 어절은 문장 성분의 최소 단위로, 하나 또는 둘 이상의 단어나 형태소로 이루어진다. 어절은, 입력부(102)로부터 입력되거나 및/또는 통신부(109)로부터 수신된 것일 수 있다. 이 경우, 저장부(110)는 입력 및/또는 수신된 문장을 먼저 저장하고, 동음이품사 처리 명령 등의 입력에 응하여 어절 분리부(131)는 저장부(110)에 저장된 문장을 호출하고 호출한 문장에 대한 어절 분리를 수행할 수도 있다. 문장에서 분리된 하나 또는 둘 이상의 어절은 동음이품사 판정부(133)로 전달될 수 있다.

[0067] 동음이품사 판정부(133)는, 어절이 동음이품사에 해당하는지 여부를 판단하거나, 또는 어절로부터 형태소(실질 형태소 또는 형식 형태소)를 추출한 후 추출한 형태소의 동음이품사 여부를 판단할 수 있다.

[0068] 일 실시예에 의하면, 동음이품사 판정부(133)는, 형태소 분석 데이터베이스(111)를 기반으로 어절이 동음이품사에 해당하는지 여부를 판정할 수도 있다. 예를 들어, 동음이품사 판정부(133)는 형태소 분석 데이터베이스(111)의 데이터베이스(113, 115, 117) 중에서 복수 품사 어근 데이터베이스(113) 및/또는 복수 품사 어미 데이터베이스(115)를 이용하여 어절이 동음이품사에 해당하는지를 판단할 수 있다.

[0069] 구체적으로 예를 들어, 동음이품사 판정부(133)는, 먼저 복수 품사 어근 데이터베이스(113)를 참조하여, 수신한 어절이나, 어절의 어근에 대응하는 형태소에 대응하는 데이터(단어나 형태소 등)가 복수 품사 어근 데이터베이스

스(113)에 존재하는지 여부를 판단하고, 순차적으로 복수 품사 어미 데이터베이스(115)를 참조하여, 수신한 어절이나 어절 내의 어미에 대응하는 형태소에 대응하는 데이터가 복수 품사 어미 데이터베이스(115)에 존재하는지 여부를 판단하여 어절이 동음어품사에 해당하는지 여부를 판정할 수도 있다. 이 경우, 동음어품사 판정부(133)는, 어절, 어근 및 어미 중 적어도 하나가 복수 품사 어근 데이터베이스(113) 또는 복수 품사 어근 데이터베이스(113)에 존재하는 경우, 어절이 동음어품사에 해당한다고 판단할 수 있다.

[0070] 만약 어절 또는 이에 포함된 형태소가 동음어품사에 해당되지 않는다고 판단되면, 어절 또는 형태소는 일반 형태소 처리부(135)로 전달되고, 통상적인 형태소 분석 방법에 따라 어절 또는 형태소는 분석된다. 즉, 동음어품사가 아니라는 것을 전제로 종래 알려진 분석 방법 중 적어도 하나에 따라 어절 또는 형태소 분석이 수행된다.

[0071] 반대로 만약 어절 또는 이에 포함된 형태소가 동음어품사라고 판단되면, 어절 또는 이에 포함된 형태소는 동음어품사 처리부(137)로 전달되고, 동음어품사 처리부(137)는 어절 또는 형태소에 대한 분석을 수행할 수 있다.

[0072] 도 3은 동음어품사 처리부의 일 실시예에 대한 블록도이다.

[0073] 도 3에 도시된 바에 의하면, 동음어품사 처리부(137)는, 일 실시예에 있어서, 복수 품사 어근 데이터베이스(113)를 확인하고(137a), 복수 품사 어미 데이터베이스(115)를 확인한 후(137b), 확인 결과를 기반으로 단일 품사 판정 함수를 이용하여 형태소 분석을 수행하거나(137c) 또는 동음어품사 유형 데이터베이스(117)를 확인할 수 있다(137d). 동음어품사 유형 데이터베이스(117)를 확인하는 경우(137d)에는, 동음어품사 처리부(137)는 동음어품사 유형 데이터베이스(117)에 대한 확인 결과를 기반으로 복수 품사 판정 함수를 이용하여 형태소 분석을 수행할 수 있다(137e).

[0074] 보다 구체적으로 동음어품사 처리부(137)는 어절로부터 어근을 추출하고 추출된 어근이 복수 품사 어근 데이터베이스(113)에 존재하는지 여부를 판단할 수 있다(137a). 이 경우, 동음어품사 처리부(137)는 어절의 첫 부분(한글 기재 방법에 의하면 좌측 부분)에서부터 순차적으로 읽으면서 어근을 추출할 수도 있다. 만약 어근이 복수 품사 어근 데이터베이스(113)에 존재하지 않는다면, 동음어품사 처리부(137)는 동음어품사 유형 데이터베이스(117)를 확인하도록 한다.

[0075] 반대로 만약 어근이 복수 품사 어근 데이터베이스(113)에 존재하면, 동음어품사 처리부(137)는 복수 품사 어미 데이터베이스(115)의 확인을 수행할 수 있다(137b). 상세하게는 동음어품사 처리부(137)는 어절로부터 어미를 추출하고 추출된 어미가 복수 품사 어미 데이터베이스(115)에 존재하는지 여부를 판단할 수 있다(137b). 이 경우, 동음어품사 처리부(137)는 어절의 마지막 부분(한글 기재 방법에 의하면 우측 말단 부분)에서부터 어절의 처음 부분에 대한 방향(한글 기재 방법에 의하면, 좌측 방향)으로 읽어가면서 어미를 추출할 수도 있다. 만약 어미가 복수 품사 어미 데이터베이스(115)에 부재한다면, 동음어품사 처리부(137)는 동음어품사 유형 데이터베이스(117)를 확인하도록 한다.

[0076] 반대로 만약 어미가 복수 품사 어미 데이터베이스(115)에 존재하면, 동음어품사 처리부(137)는 단일 품사 판정 함수를 이용하여 형태소 분석을 수행한다(137c).

[0077] 도 3에는 복수 품사 어근 데이터베이스(113)의 확인(137a) 후에 복수 품사 어미 데이터베이스(115)가 확인되는 실시예에 대해서만 도시되어 있으나, 복수 품사 어근 데이터베이스(113)의 확인 과정(137a) 및 복수 품사 어미 데이터베이스(115)의 확인 과정(137b)는, 도 3에 도시된 것과는 반대의 순서로 진행될 수도 있다. 또한, 실시예에 따라서, 복수 품사 어근 데이터베이스(113)의 확인 과정(137a)과 복수 품사 어미 데이터베이스(115)의 확인 과정(137b)은 동시에 수행되는 것도 가능하다.

[0078] 단일 품사 판정 함수는, 특정한 품사(예를 들어, 명사, 대명사, 수사, 동사, 형용사, 관형사, 부사, 조사 또는 감탄사 등)를 판정할 수 있는 함수를 의미한다. 단일 품사 판정 함수는, 어절, 단어 또는 형태소가 입력된 경우, 어절, 단어 또는 형태소가 특정한 품사(예를 들어 명사 등)에 해당하는지 여부를 판단하고, 판단 결과를 출력할 수 있다. 이 경우, 단일 품사 판정 함수는, 참(True) 또는 거짓(False)을 출력하여 어절, 단어 또는 형태소가 특정한 품사에 해당하는지 여부를 판단할 수 있다. 단일 품사 판정 함수는, 판정 대상이 되는 품사에 따라서 다양하게 결정될 수 있다. 예를 들어, 단일 품사 판정 함수는, 일반명사 판정 함수, 고유명사 판정 함수, 의존명사 판정 함수, 수사 판정 함수, 일반동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및/또는 어근 판정 함수 등을 포함할 수 있다.

[0079] 도 4는 단일 품사 판정 함수의 일례에 대한 프로그램 코드를 도시한 것으로, 일반 명사 판정 함수의 일례를 도시한 것이다.

- [0080] 도 4에 도시된 바에 의하면, 단일 품사 판정 함수(p1)는, 현재 검사 대상 형태소의 한 어절 앞에 특정 품사가 존재하는지를 여부를 판단할 수 있다(p11). 이 경우, 현재의 어절 앞에 형용사가 존재하면, 현재의 형태소를 명사로 판정하도록 설계될 수 있다. 맑은 예쁜 고운 집과 같이 형용사 뒤에 형용사가 올 수 있으므로 반드시 이와 같은 판정이 올바른 것은 아니다. 그러나, 이와 같은 판단은, 특정 형태가 명사라고 의심될 때 사용하는 것이므로 이 규칙을 사용하는 상황에서는 대체적으로 오분석이 발생하진 않는다.
- [0081] 또한, 단일 품사 판정 함수(p1)는, 현재 어절이 제1 문자열(w11, words)에 포함된 어휘로 끝나는지를 판단할 수도 있다(p12). 대체적으로 명사는 조사를 취하므로 이를 기반으로 판단을 수행하는 것이 불합리한 것은 아니다. 다만, [는]은 보조사(예를 들어, 특허는 등)로도 이용되고 어미(예를 들어, 가는 등)로도 이용 가능하여 형태소 분석에 있어서 오류를 발생시킬 수도 있으므로, 제1 문자열(w11)은 오직 조사로만 이용되는 경우만으로 포함하도록 설계될 수도 있다.
- [0082] 또한, 단일 품사 판정 함수(p1)는, 현재 어절이 제2 문자열(w12, words2)에 포함된 어휘로 끝나는지 여부도 판단할 수 있다. 특정한 형태소(예를 들어, [만큼]이나 [처럼] 등)들은 오로지 그 앞에 명사만을 취할 수 있으므로, 이를 기반으로도 일반 명사를 판정할 수 있다.
- [0083] 한편, 도 3에 도시된 바와 같이, 만약 어근이 복수 품사 어근 데이터베이스(113)에 존재하지 않거나 또는 어미가 복수 품사 어미 데이터베이스(115)에 존재하지 않으면, 동음이품사 처리부(137)는 동음이품사 유형 데이터베이스(117)를 확인할 수 있다(137d). 이 경우, 동음이품사 처리부(137)는, 동음이품사 유형 데이터베이스(117)로부터 적어도 하나의 동음이품사 유형을 추출하고, 추출된 동음이품사 유형을 기반으로 복수 품사 판정 함수를 이용하여 형태소 분석을 수행할 수 있다(137e).
- [0084] 도 5는 한글 형태소 분석을 설명하기 위한 도표이다.
- [0085] 예를 들어, 도 5에 도시된 바를 참조하면, [강하]라는 형태소는 문장의 맥락에 따라서 적어도 네 개의 의미를 가질 수 있다. 구체적으로 [언덕에서 넓은 강하를 바라보았다]라는 문장(이하 제1 문장) 내에서는 [강하]는 강과 하천을 의미하는 명사이고, [낙하산을 타고 하늘로부터의 강하를 경험하였다]라는 문장(이하 제2 문장) 내에서는 [강하]는 내려감(하강)을 의미하는 명사이다. 또한, [그는 매우 강하다]라는 문장(이하 제3 문장)에서는 [강하]는 힘이 센을 의미하는 형용사로 쓰인 것이고, [절벽 아래로 강하하였다]라는 문장(이하 제4 문장)에서는 내려가다(하강하다)는 의미의 어근으로 쓰인 것이다. 이와 같이 하나의 문장 내의 맥락에 따라서 동일한 형태소는 서로 유사하거나 또는 상이한 의미를 갖는 것으로 판단될 수 있으며, 또한 서로 상이한 품사로 판단될 수도 있다.
- [0086] 이 경우 제1 문장 및 제2 문장과 같이 [강하]가 명사로 사용된 경우에는 보통 조사와 같이 나타나는 경우가 많은 편이고, 제3 문장과 같이 형용사로 이용되는 경우에는 통상 어미 [-다]와 함께 쓰이는 경우가 많은 편이다. 또한, 제4 문장에서의와 같이 어근으로 쓰이는 경우에는, 어근의 뒤에 [-하다]/[-되다]/[-거리다]가 부가되는 경우가 많다. 이러한 점을 고려하면 [강하]의 품사를 적절하게 판단하여, 부가적인 정보(즉, [강하]의 품사 등에 대한 정보)를 텍스트에 부가할 수 있게 된다. 이와 같이 소정의 형태소에 소정의 형태소에 대응하는 품사 등을 부가하는 것을 품사 태깅이라고 한다.
- [0087] 도 6은 복수 품사 판정 함수의 어느 하나의 일례에 대한 프로그램 코드를 도시한 도면이다.
- [0088] 이와 같이 하나의 단어 또는 형태소가 복수의 품사로 이용될 수 있는 경우, 하나의 단어 또는 형태소에 대한 품사 태깅을 구현하기 위해서, 복수 품사 판정 함수가 필요하다. 복수 품사 판정 함수는 복수의 품사로 이용될 수 있는 단어 또는 형태소의 품사를 어느 하나로 결정할 수 있는 함수를 의미한다.
- [0089] 복수 품사 판정 함수는, 도 6에 도시된 바와 같이, 프로그램 코드(p2)로 작성되어 구현될 수 있다. 예를 들어, [강하]라고 정의된 복수 품사 판정 함수는, 어절이나 형태소 내에서의 [강하]라는 단어의 품사를 결정하도록 구현된다. 보다 상세하게 예를 들어, [강하]라고 정의된 복수 품사 판정 함수는, 현재 분석 대상이 되는 어절(이하 분석 대상 어절)이 변수로 입력되도록 마련되며, 변수로 입력된 분석 대상 어절이 [이니], [이다] 또는 [인] 등을 포함하거나(p21) 또는 [이], [가] 또는 [을] 등을 포함하는 경우(p22)에는 어절이나 형태소 내에서의 [강하]의 품사를 명사로 판단하도록 하고, 분석 대상 어절이 [다], [며] 또는 [어서] 등을 포함하는 경우(p23)에는 [강하]의 품사를 형용사로 판단하도록 하며, 분석 대상 어절이 [하], [할] 또는 [하니] 등을 포함하는 경우(p24)에는 [강하]의 품사를 어근으로 판단하도록 설계된 것일 수 있다.
- [0090] 도 7은 단일 품사 판정 함수의 연결의 일례를 설명하기 위한 도면이고, 도 8은 단일 품사 판정 함수의 연결의

일례에 대응하는 프로그램 코드를 도시한 도면이다.

- [0091] 그러나 가능한 모든 단어(예를 들어, [강하] 이외에 복수 품사 어근 데이터베이스(113)에 포함된 단어 등)에 대해 프로그램 코드를 구축하는 것은, 그 데이터 양이 방대해지고 이에 따라 성능이 저하되며, 갱신 시 코드의 일관성 유지 역시 어려워진다. 그러므로, 도 7 및 도 8에 도시된 바와 같이, 단일 품사 함수를 구축한 후 구축된 각각의 단일 품사 함수를 연결하여 특정 품사의 판정을 수행하면, 데이터 양의 절감, 코드의 간결화 및 갱신 시 코드의 일관성 유지 등을 이룰 수 있게 된다.
- [0092] 구체적으로 도 7에 도시된 바와 같이, 복수의 품사 각각에 대한 단일 품사 판정 함수(p3a, p3b, p3c)은 상호 연결되어 이용될 수 있다. 여기서, 단일 품사 판정 함수(p3a, p3b, p3c)는 일반명사 판정 함수, 고유명사 판정 함수, 의존명사 판정 함수, 수사 판정 함수, 일반동사 판정 함수, 형용사 판정 함수, 보조 동사 판정 함수 및 어근 판정 함수 중 적어도 둘을 포함할 수 있다.
- [0093] 구체적으로 예를 들어, 동음이품사에 속하는 단어 A(w11)의 품사 판정을 위해 동사 판정 함수(p3a) 및 명사 판정 함수(p3b)가 연결되어 사용된다. 다시 말해서, 동사 판정 함수(p3a) 및 명사 판정 함수(p3b)는 연결되어 하나의 함수로 구현될 수 있으며, 단어 A(w11)는 이들(p3a, p3b)이 연결된 함수에 대입되어 동사 또는 명사로 판단되고, 이에 따른 판단 결과(R11)가 획득된다. 동일하게 명사 판정 함수(p3b) 및 어근 판정 함수(p3c)도 서로 연결되어 하나의 함수로 구현될 수 있으며, 이는 다른 동음이품사에 속하는 단어 B(w11)를 명사 또는 어근으로 판단하기 위해 이용될 수 있다(R12).
- [0094] 이와 같이 복수의 단일 품사 판정 함수(p3a, p3b, p3c) 중 적어도 둘을 상호 연결하여 사용하는 경우, 프로그램 코드(p3)는 도 8에 도시된 바와 같이 작성될 수 있다. 이 경우, 프로그램 코드(p3)는, 도 6에 도시된 프로그램 코드(p2)와 비교하였을 때, 명사 판단과 관련된 부분(P21, P22)이 모듈화되어 상대적으로 간략하게 작성되게 된다(P31).
- [0095] 도 9는 복수 품사 판정 함수의 일례를 설명하기 위한 도면이다.
- [0096] 도 7 및 도 9에 도시된 바와 같이, 각각의 단어에 대한 단일 품사 함수들을 연결하여 사용하면 코드의 양이 획기적으로 감소할 뿐만 아니라 코드의 일관성과 관리의 용이성을 기대할 수 있게 된다. 그러나, 함수화해야 할 동음이품사의 개수는 다량이기 때문에 여전히 작성할 코드는 짧지 않다. 예를 들어, 명사, 형용사 및 어근으로 사용되는 단어들은 모두 [강하]에서 사용되었던 코드를 동일하게 반복해서 사용해야 하는 문제점이 존재한다.
- [0097] 따라서 코드의 양 감소, 작업의 자동화 및 이에 따른 코드의 일관성 유지를 위해, 도 9에 도시된 바와 같이, 둘 또는 그 이상의 단일 품사 판정 함수(f1 내지 f3)를 상호 연결하여, 적어도 하나의 복수 품사 판정 함수(f12, f23, f123 등)를 구현할 수 있다.
- [0098] 구체적으로 품사 A(예를 들어, 동사)에 대한 판정 함수(f1, 이하 품사 A 판정 함수)와, 품사 B(예를 들어, 명사)에 대한 판정 함수(f2, 이하 품사 B 판정 함수)와, 품사 C(예를 들어, 어근)에 대한 판정 함수(f3, 이하 품사 C 판정 함수)가 주어진 경우, 이를 기반으로 복수의 품사를 판정할 수 있는 함수를 설계하여 구축할 수 있다. 예를 들어, 품사 A 판정 함수(f1)와 품사 B 판정 함수(f2)를 조합 및 연결하여 품사 A-B 판정 함수(f12, 예를 들어, 동사-명사 판정 함수)를 획득하거나, 품사 B 판정 함수(f2)와 품사 C 판정 함수(f3)를 조합 및 연결하여 품사 B-C 판정 함수(f23, 예를 들어, 명사-어근 판정 함수)를 획득할 수 있다. 뿐만 아니라, 품사 A 판정 함수(f1), 품사 B 판정 함수(f2) 및 품사 C 판정 함수(f3) 모두를 연결하여 조합함으로써 품사 A-B-C 판정 함수(f123, 예를 들어, 동사-명사-어근 판정 함수)를 획득할 수도 있다. 이들 판정 함수(f12, f23, f123 등) 중 적어도 하나는, 설계자의 선택이나 사용자의 조작 등에 따라서, 복수의 품사로 이용될 수 있는 적어도 하나의 단어(예를 들어, 단어 A(w1), 단어 B(w2) 및 단어 C(w3) 등)에 각각 적용될 수 있으며, 이에 따라 적어도 하나의 단어(w1, w2, w3)의 품사가 결정될 수 있게 된다.
- [0099] 도 10은 복수 품사 판정 함수의 일례에 대응하는 프로그램 코드를 도시한 도면이다.
- [0100] 이와 같은 복수의 단일 품사 판정 함수(f1 내지 f3)를 조합하여 적어도 하나의 복수 품사 판정 함수(f12, f23, f123)를 구현하면, 도 6 또는 도 7에 도시된 [강하]의 품사를 판단하기 위한 함수는 도 10에 도시된 바와 같이 더욱 간단하고 단순한 형태의 프로그램(p4)으로 작성될 수 있다. 이 경우, 한 줄의 간단한 함수 호출 명령문(p41, 예를 들어, 동사-명사-어근 판정 함수(f123)의 호출 명령문)만으로도 전후 맥락에 따라 품사가 분석될 수 있게 된다. 예를 들어, 강하가 명사, 형용사 및 어근 중 어느 하나로 분석될 수 있게 된다. 이에 따라 코드의 간결화와 일관성 유지가 더욱 효과적으로 구현될 수 있게 된다.

- [0101] 도 11은 동사-일반 명사 함수의 일례에 대한 프로그램 코드를 도시한 도면이다.
- [0102] 도 10에 도시된 프로그램 코드(p4)는, 예를 들어, 도 11에 도시된 바와 같은 동사-일반 명사 함수(P5)를 호출할 수도 있다. 이 경우, 호출되는 동사-일반 명사 함수(P5)는, 예를 들어, 여러 판정 함수(p51 내지 p55)를 조합하여 구현된 것일 수 있다. 보다 구체적으로 동사-일반 명사 함수(P5)는 의존 명사 판정 함수(p51), 일반 동사 판정 함수(p52), 형용사 판정 함수(p53), 일반 명사 판정 함수(p54) 및/또는 기타 판정 함수(p55)를 포함할 수 있으며, 이들(p51 내지 p55)을 순차적으로 조합하여 구현된 것일 수 있다.
- [0103] 구체적으로 분석 대상 어절 또는 이에 포함된 형태소의 품사가 명사인지를 판정하기 위해, 의존 명사 함수(p51)가 이용될 수 있다. 의존 명사 함수(p51)는, 현재 분석 대상이 되는 어절(또는 형태소 등) 또는 분석 대상 어절의 전단의 어절(또는 형태소 등)에 수사(숫자)가 존재하는지 여부를 확인할 수 있다. 수사는 한글, 한자, 알파벳 또는 아라비아어 등으로 표현된 것일 수 있다. 명사 또는 동사 중 어느 하나로 결정해야 하는 경우에 있어서, 분석 대상이 되는 어절 앞의 어절에 수사가 존재한다면, 분석 대상이 되는 어절 명사일 가능성이 매우 높다. 의존 명사 함수(p51)는 이와 같이 거의 확실시되는 경우를 우선적으로 처리한다.
- [0104] 또한, 분석 대상 어절 또는 이에 포함된 형태소의 품사가 동사인지를 판정하기 위해 동사 판정 함수(p52)와 형용사 판정 함수(p53)를 이용될 수 있다. 동사 여부를 판정은, 명사 판정 함수(p54)를 이용한 명사 여부를 판정에 대해 선행하여 처리되도록 코딩될 수도 있다. 왜냐하면, 동사 또는 명사로 판정해야 하는 상황에서는 동사가 그 특징이 상대적으로 더 분명하기 때문이다. 구체적으로 예를 들어, [종이를 떼다]란 문장의 경우, 명사 여부를 먼저 판단하게 되면, [떼]가 명사로 판단될 가능성이 높아진다. 명사는 종결 어미 [-다]와 잘 결합되기 때문이다(예를 들어, 그는 의사[다]). 그러나, 동사는 그 앞에 목적어나 부사어를 갖는 특징을 가지므로, 분석 대상 어절의 앞의 어절에 목적어나 부사어가 존재하는지 여부를 먼저 확인하면, [종이를 떼다]의 [떼]는 동사로 적절하게 판정될 수 있게 된다. 동사의 판정은, 단일 품사 판정 함수인 동사 판정 함수(p52)를 기반으로 수행될 수 있다. 형용사도 동일하게 그 앞에 부사어가 올 수 있으므로, 명사에 대해 선행하여 판정될 수 있다. 형용사의 판정 역시 형용사 판정 함수(p53)를 이용하여 수행될 수 있다.
- [0105] 순차적으로 명사 판정 함수(p54)를 이용하여 명사인지를 여부가 확인될 수 있다. 명사 판정 함수(p54)는, 도 6에 도시된 바와 같이, 특정 품사 또는 특정 형태가 함께 오는 형태를 가지고 있으므로, 이를 기반으로 판단할 수 있다.
- [0106] 또한, 동사-일반 명사 함수(P5)는, 분석 대상 어절 내의 단어나 형태소의 길이가 분석 대상 어절의 길이와 동일한지 여부를 판단하기 위한 함수(p55)를 포함할 수도 있다. 만약 어절 내에 단어나 형태소가 혼자 존재한다면(다시 말해서, 어절이 하나의 단어 또는 하나의 형태소만을 갖는다면), 이와 같은 단어나 형태소는 통상 명사에 해당하지, 동사에 해당하는 경우는 극히 드물거나 없다. 그러므로, 명사 및 동사 양자로 판단될 수 있는 경우에 있어서, 동사-일반 명사 함수(P5)는 단어나 형태소의 길이와 어절의 길이를 비교하고, 만약 그 길이가 동일하면 명사로 판단하고, 그렇지 않으면 동사로 판단하도록 한다. 이에 따라, 동음이품사에 해당하는 단어나 형태소가 동사 및 명사 중 어느 하나로 판단될 수 있게 된다.
- [0107] 결과 획득 및 처리부(139)는, 일반 형태소 처리부(135) 및 동음이품사 처리부(137) 중 적어도 하나로부터 처리 결과를 전달받아, 입력된 문장의 어절에 대한 최종적인 분석 결과를 획득한다. 예를 들어, 만약 일반 형태소 처리부(135) 및 동음이품사 처리부(137)가 품사의 판단 결과만을 획득하는 경우, 결과 획득 및 처리부(139)는 획득한 품사 판단 결과를 기반으로 각 어절에 대해 품사 태깅을 수행할 수도 있다. 또한, 결과 획득 및 처리부(139)는 분석 결과를 단말 장치(10) 등으로 전송하거나, 분석 결과를 저장부(110)로 전달하여 저장되도록 하거나, 및/또는 분석 결과가 출력부(103)를 통해 외부로 출력되도록 제어할 수도 있다. 이외에도 결과 획득 및 처리부(139)는 설계자의 임의적 선택에 따라 분석 결과를 기반으로 각종 연산 처리 및/또는 제어 동작을 수행할 수도 있다.
- [0108] 이하 도 12를 참조하여 한글 형태소 분석 방법의 일 실시예에 대해서 설명하도록 한다.
- [0109] 도 12는 한글 형태소 분석 방법의 일 실시예에 대한 흐름도이다.
- [0110] 도 12에 도시된 한글 형태소 분석 방법의 일 실시예에 따르면, 먼저 한글 형태소 분석 장치 또는 이와 통신 가능하게 연결된 단말 장치에 문장이 입력된다(151). 이와 같은 문장의 입력은 사용자에게 의해 직접 수행되는 것일 수도 있고, 및/또는 다른 컴퓨터 장치 등에 의해 수행되는 것일 수도 있다. 또한, 문장의 입력은 텍스트의 형태로 입력될 수도 있고, 음성 등의 형태로 입력될 수도 있다. 음성 등의 형태로 문장이 입력된 경우, 음성 형태의 문장은 텍스트로의 변환 과정을 더 거칠 수 있다.

- [0111] 한글 형태소 분석 장치는, 직접 입력 받거나 단말 장치로부터 전송된 문장으로부터 어절을 분리할 수 있다(152). 실시예에 따라서 단말 장치가 문장으로부터 어절을 분리할 수도 있다. 이 경우, 한글 형태소 분석 장치는 어절을 수신하게 된다.
- [0112] 한글 형태소 분석 장치는 어절, 어절 내의 단어 또는 어절 내의 형태소가 동음어포사에 해당하는지 여부를 판단할 수 있다(153). 동음어포사에 대한 판단은, 예를 들어, 형태소 분석 데이터베이스를 이용하여 수행될 수도 있으며, 보다 구체적으로는 복수 품사 어근 데이터베이스 및/또는 복수 품사 어미 데이터베이스를 이용하여 수행될 수도 있다.
- [0113] 만약 동음어포사에 해당하지 않는다고 판단되면(153의 아니오), 한글 형태소 분석 장치는, 어절, 어절 내의 단어 또는 어절 내의 형태소에 대해 통상 알려진 일반적인 형태소 분석을 수행하게 된다(154).
- [0114] 반대로 만약 동음어포사에 해당한다고 판단되면(153의 예), 한글 형태소 분석 장치는 동음어포사 분석을 진행할 수 있다(155). 이 경우, 한글 형태소 분석 장치는, 형태소 분석 데이터베이스를 이용하여 동음어포사에 대한 분석을 진행하는 것도 가능하다. 여기서, 형태소 분석 데이터베이스는, 예를 들어, 복수 품사 어근 데이터베이스, 복수 품사 어미 데이터베이스 및/또는 동음어포사 유형 데이터베이스를 포함할 수 있다.
- [0115] 일 실시예에 의하면, 동음어포사 분석을 수행하기 위하여, 먼저 어절의 어근 등에 대응되는 데이터가 복수 품사 어근 데이터베이스에 존재하는지 여부가 판단될 수 있다. 만약 어절의 어근 등에 대응되는 데이터가 복수 품사 어근 데이터베이스로부터 검색되면, 어절의 어미 등에 대응되는 데이터가 복수 품사 어미 데이터베이스에 존재하는지 여부가 판단될 수 있다. 만약 어미 등에 대응되는 데이터가 복수 품사 어미 데이터베이스로부터 획득 가능하면, 단일 품사 판정 함수를 이용하여 동음어포사에 해당하는 어절이나 형태소에 대한 형태소 분석이 수행되게 된다. 실시예에 따라서, 어미 등에 대응되는 데이터가 복수 품사 어미 데이터베이스로부터 검출되는지 여부에 대한 판단은, 어근 등에 대응되는 데이터가 복수 품사 어근 데이터베이스로부터 검출되는 여부의 판단에 선행하여 수행되는 것도 가능하다. 또한, 이 둘은 동시에 수행될 수도 있다.
- [0116] 또한, 만약 어절의 어근 등에 대응되는 데이터가 복수 품사 어근 데이터베이스에 존재하지 않거나, 및/또는 어미 등에 대응되는 데이터가 복수 품사 어미 데이터베이스에 존재하지 않는 경우에는, 복수 품사 판정 함수를 이용하여 동음어포사에 해당하는 어절이나 형태소에 대한 형태소 분석이 수행될 수 있다. 이 경우, 동음어포사 유형 데이터베이스가 먼저 열람될 수 있으며, 열람 결과에 따라서 동음어포사의 유형이 적어도 하나 확인되고, 확인된 동음어포사의 유형을 기반으로 복수 품사 판정 함수를 이용한 형태소 분석이 수행될 수도 있다.
- [0117] 순차적으로 일반적인 형태소 분석 수행(154) 또는 동음어포사 분석 수행(155)에 따른 분석 결과가 획득된다(156). 분석 결과는 일반적인 형태소 분석 수행(154) 또는 동음어포사 분석 수행(155)이 종료된 이후에 획득될 수도 있다. 또한, 분석 결과는, 일반적인 형태소 분석 수행 과정(154) 또는 동음어포사 분석 수행 과정(155)에서 실시간으로 획득될 수도 있다. 획득된 분석 결과는, 분석 결과의 이용 목적 등에 따라서 추가적으로 가공 또는 변형될 수도 있다. 또한, 단말 장치로 전송되거나 또는 한글 형태소 분석 장치에 연결된 스피커나 모니터 장치 등을 통하여 사용자 등에게 제공될 수도 있다.
- [0118] 상술한 실시예에 따른 한글 형태소 분석 방법은, 컴퓨터 장치에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다. 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 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 또한, 여기서, 컴퓨터 장치는, 프로그램의 기능을 실현 가능하게 하는 프로세서나 메모리 등을 포함하여 구현된 것일 수 있으며, 필요에 따라 통신 장치를 더 포함할 수도 있다.
- [0119] 상술한 한글 형태소 분석 방법을 구현하기 위한 프로그램은, 컴퓨터에 의해 관독 가능한 기록 매체에 기록될 수 있다. 컴퓨터에 의해 관독 가능한 기록 매체는, 예를 들어, 하드 디스크나 플로피 디스크와 같은 자기 디스크 저장 매체, 자기 테이프, 콤팩트 디스크나 디브이디와 같은 광 기록 매체, 플롭티컬 디스크와 같은 자기-광 기록 매체 및 롬, 램 또는 플래시 메모리 등과 같은 반도체 저장 장치 등 컴퓨터 등의 호출에 따라 실행되는 특정 프로그램을 저장 가능한 다양한 종류의 하드웨어 장치를 포함할 수 있다.
- [0120] 이상 한글 형태소 분석 장치, 시스템 및 방법의 여러 실시예에 대해 설명하였으나, 한글 형태소 분석 장치, 시스템 및 방법은 오직 상술한 실시예에 한정되는 것은 아니다. 해당 기술 분야에서 통상의 지식을 가진 자가 상술한 실시예를 기초로 수정 및 변형하여 구현 가능한 다양한 장치, 시스템 또는 방법 역시 상술한 한글 형태소

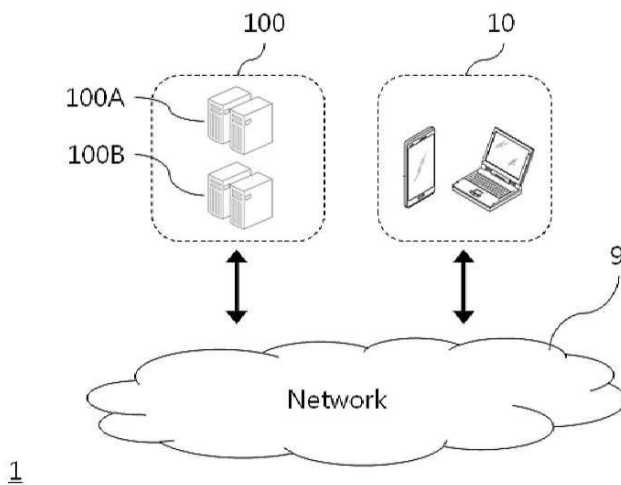
분석 장치, 시스템 또는 방법의 일 실시예가 될 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나 다른 구성 요소 또는 균등물에 의하여 대치되거나 또는 치환되더라도, 이는 상술한 한글 형태소 분석 장치, 시스템 및 방법의 일 실시예가 될 수 있다.

부호의 설명

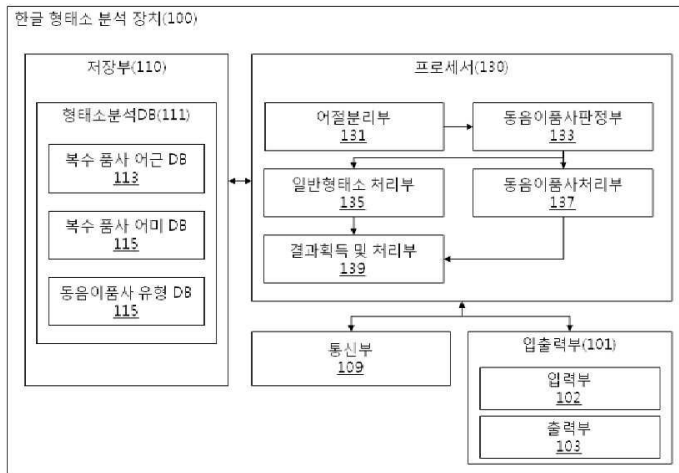
10: 단말 장치 100: 한글 형태소 분석 장치
101: 입출력부 109: 통신부
110: 저장부 111: 형태소 분석 데이터베이스
113: 복수 품사 어근 데이터베이스
115: 복수 품사 어미 데이터베이스
117: 동음어품사 유형 데이터베이스
130: 프로세서 131: 어절 분리부
133: 동음어품사판정부 135: 일반 형태소 처리부
137: 동음어품사 처리부 139: 결과 획득 및 처리부

도면

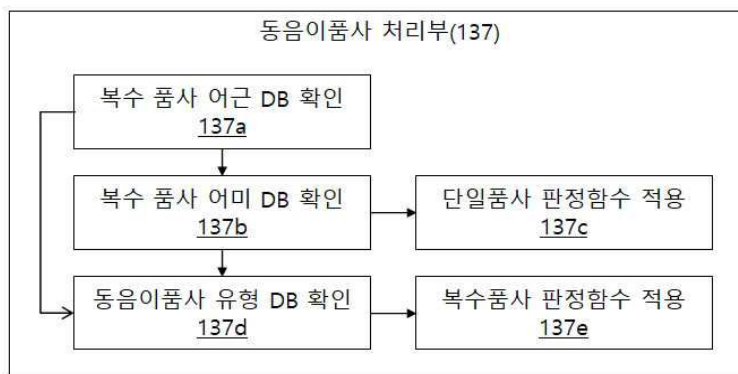
도면1



도면2



도면3



도면4

```

public boolean 일반명사_판정함수() {
    boolean 일반명사 = false;
    String[] words = {"이", "가", "을", "를", "와", "과", "의", "에", "에는", "에게", "에서", "에서", ← w11
    는", "만"};
    String[] words2 = {"만큼", "보다", "처럼", "임니", ← w12
    if(품사[현재어절하나앞].endsWith("/형용사")||품사[현재어절하나앞].endsWith("/관형사")||
    품사[현재어절하나앞].endsWith("/관형사형전성어미")||품사[현재어절하나앞].endsWith("/관형격조사")
    ||품사[현재어절하나앞].endsWith("/접속조사") } P11
        일반명사 = true;
    else if(this.endsWithWord(형태[현재어절], words)) } P12
        일반명사 = true;
    else if(this.containsWithWord(형태[현재어절], words2)) } P13
        일반명사 = true;
    return 일반명사;
}

```

P1

도면5

1	언덕에서 넓은 강하를 바라보았다	강과 하천	명사
2	낙하산을 타고 하늘로부터의 강하를 경험하였다	내려감	명사
3	그는 매우 강하다	힘이 셴	형용사
4	절벽 아래로 강하하였다	내려가다	어근

도면6

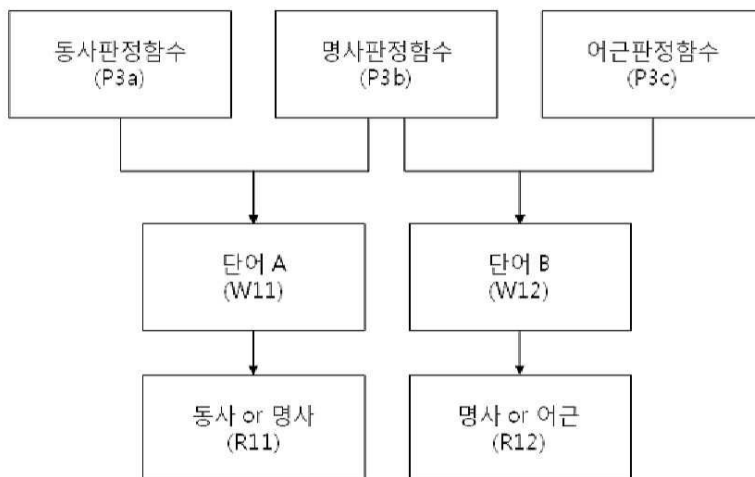
```

public String 강하(현재어절) {
    String result = "";
    if(현재어절.contains("이니", "이다", "이어", "일", "인", "일니", ...)) ← P21
        result = "명사"; ← P22
    else if(현재어절.contains("이", "가", "을", "를", "와", "과", "의", "에", ...)) ← P23
        result = "명사";
    else if(현재어절.contains("다", "며", "어서", "여서", "므로", "니까", ...)) ← P24
        result = "형용사";
    else if(현재어절.contains("하", "한", "할", "했", "하니", "되", "된", "둘", ...))
        result = "어근";
}

```

P2

도면7



도면8

```

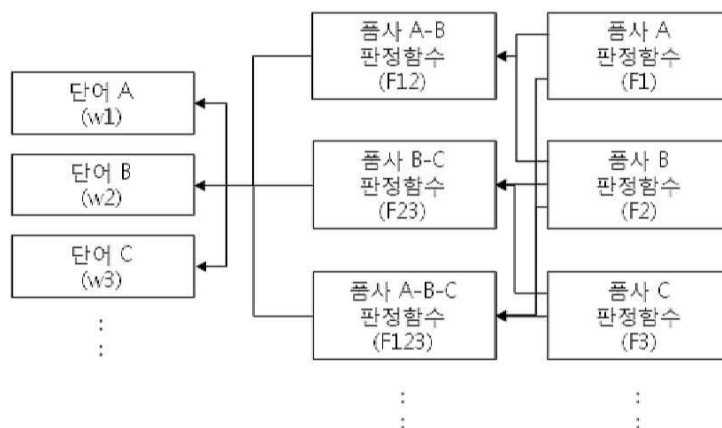
public String 강하(현재어절) {
    String result = "";
    if(명사함수(현재어절))
        result = "명사";
    else if(형용사함수(현재어절))
        result = "형용사";
    else if(어근함수(현재어절))
        result = "어근";
}

```

P31

P3

도면9



도면10

```

public String 강하(현재어절) {
    String result = 명사_형용사_어근함수0;
}

```

P41

P4

도면11

```

public String 동사-일반명사_판정함수(String 현재어절) {
P51   if(의존명사_판정함수(현재어절))
      결과 = "일반명사";
P52   else if(일반동사_판정함수(현재어절))
      결과 = "일반동사";
P53   else if(형용사_판정함수(현재어절))
      결과 = "일반동사";
P54   else if(일반명사_판정함수(현재어절))
      결과 = "일반명사";
P55   else if(inArr[paNum].equals(현재어절))
      결과 = "일반명사";
      else
        결과 = "일반동사";
      return 결과;
}

```

P5

도면12

