



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0072684
(43) 공개일자 2024년05월24일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06F 40/143</i> (2020.01) <i>G06F 16/34</i> (2019.01)
 <i>G06F 40/216</i> (2020.01) <i>G06F 40/258</i> (2020.01)
 <i>G06F 40/279</i> (2020.01) <i>G06F 40/56</i> (2020.01)
 <i>G06N 3/08</i> (2023.01)</p> <p>(52) CPC특허분류
 <i>G06F 40/143</i> (2022.01)
 <i>G06F 16/345</i> (2019.01)</p> <p>(21) 출원번호 10-2022-0154570
 (22) 출원일자 2022년11월17일
 심사청구일자 2022년11월17일</p> | <p>(71) 출원인
 충남대학교산학협력단
 대전광역시 유성구 대학로 99 (궁동, 충남대학교)</p> <p>(72) 발명자
 정상근
 대전광역시 서구 둔산북로 160, 한마루아파트 10
 2동 1502호
 서해인
 대전광역시 유성구 상대남로 26 도안신도시9블록
 트리폴시티아파트 919동 1003호
 정지수
 대전광역시 유성구 문화원로 131, 511호(봉명동,
 투유2)
 (74) 대리인
 이은철, 김재문</p> |
|---|---|

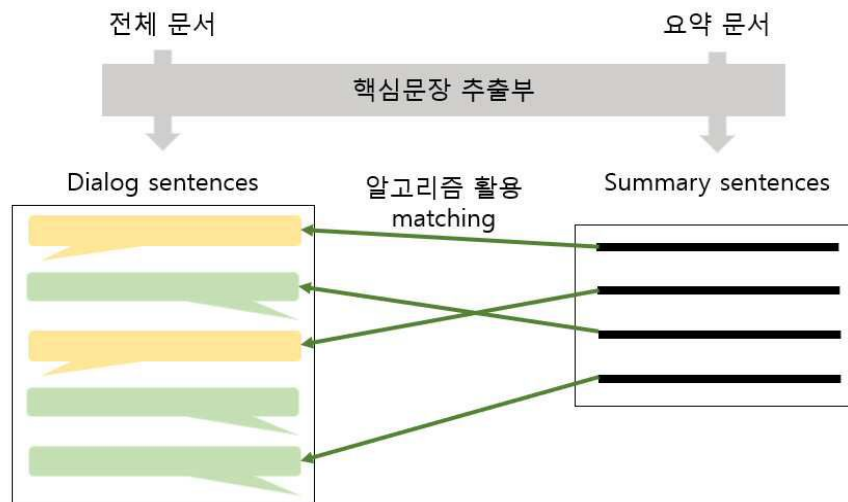
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **계약조건이 반영된 마크업 언어 생성방법, 이를 이용한 마크업 언어 생성장치, 및 마크업 언어를 이용한 문장 생성장치**

(57) 요약

본 발명은, 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신단계; 상기 전체 문서와 상기 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출단계; 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출단계; 및 추출된 핵심문장과 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성단계;를 포함하는 것을 일 특징으로 한다.

대 표 도 - 도1



(52) CPC특허분류

G06F 40/216 (2020.01)

G06F 40/258 (2020.01)

G06F 40/279 (2020.01)

G06F 40/56 (2020.01)

G06N 3/08 (2023.01)

명세서

청구범위

청구항 1

전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신단계;

상기 전체 문서와 상기 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출단계;

추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출단계; 및

추출된 핵심문장과 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성단계;를 포함하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 2

제 1 항에 있어서,

상기 핵심문장 추출단계는,

LCS(Longest Common Subsequence) 알고리즘을 이용하여 상기 전체 문서와 상기 요약 문서를 비교하는 것을 특징으로 하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 3

제 1 항에 있어서,

상기 핵심문장 추출단계는,

상기 전체 문서와 상기 요약 문서에 각각 설정된 키워드를 비교하여 상기 요약 문서에 설정된 키워드가 기설정된 빈도 이상 포함된 문장을 핵심문장으로 추출하는 것을 특징으로 하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 4

제 1 항에 있어서,

상기 핵심단어 추출단계는,

상기 핵심문장과 상기 요약 문서에서 각각 키워드를 추출하고, 추출된 키워드 중 상기 핵심문장과 상기 요약 문서에 공통되는 키워드를 핵심단어로 추출하는 것을 특징으로 하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 5

제 4 항에 있어서,

상기 핵심단어 추출단계는,

명사구를 키워드로 추출하는 것을 특징으로 하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 6

전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신부;

상기 문서 수신부에서 수신된 전체 문서와 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출부;

상기 핵심문장 추출부에서 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출부; 및

상기 핵심문장 추출부에서 추출된 핵심문장과 상기 핵심단어 추출부에서 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성부;를 포함하는 제약조건이 반영된 마크업 언어 생성방법.

청구항 7

전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신부, 상기 문서 수신부에서 수신된 전체 문서와 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출부, 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출부, 및 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 핵심단어 추출부에서 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성부를 구비한 마크업 언어 생성부; 및

상기 마크업 언어 생성부에서 생성된 마크업 언어와 상기 요약 문서를 인공신경망에 입력하여 요약문을 생성하는 인공신경망을 학습시키는 학습부;를 포함하는 마크업 언어를 이용한 문장 생성장치

청구항 8

제 7 항에 있어서,

상기 학습부에서 학습된 인공신경망에 문서를 입력하여 요약문을 생성하는 요약문 생성부;를 더 포함하는 것을 특징으로 하는 마크업 언어를 이용한 문장 생성장치.

발명의 설명

기술 분야

[0001] 본 발명은 제약조건이 반영된 마크업 언어 생성방법, 이를 이용한 마크업 언어 생성장치, 및 마크업 언어를 이용한 문장 생성장치에 관한 것으로서, 특히 요약기 훈련시 입력되는 데이터에 제약조건을 부여하여 요약기의 성능을 향상시키기 위한 제약조건이 반영된 마크업 언어 생성방법, 이를 이용한 마크업 언어 생성장치, 및 마크업 언어를 이용한 문장 생성장치에 관한 것이다.

배경 기술

[0003] 문서 생성이란 기계가 인간의 말이나 글의 의미가 무엇인지 기계가 알 수 있도록 구조화하는 과정을 거쳐, 특정 입력에 따른 새로운 문서를 생성하는 핵심기술이다.

[0004] 문서 자동 요약기란 문서의 내용을 일정한 크기로 줄여주는 것이다. 문서 자동 요약기는 전체 문서에서 중요하지 않은 부분이나 사소한 부분을 생략하면서 핵심적인 내용을 일관성 있게 추려내어 모아주는 것이 중요한 기능이며 성능의 지표라고 할 수 있다.

[0005] 최근에는 전체 문서와 전체 문서에 대한 요약 문서를 인공지능으로 학습시키고, 학습된 인공지능을 이용하여 요약 문서를 생성하는 문서 자동 요약기가 다수 연구되고 있다. 인공지능을 이용한 문서 자동 요약기는 필수정보를 모두 갖추고 동시에 잘 정형화된 요약문을 생성하는 것이 중요한 해결과제이다. 인공지능을 이용한 문서 자동 요약기는 위 해결과제를 달성하기 위해서 많은 데이터 세트가 요구된다.

[0006] 그러나, 요약을 원하는 문서는 종류에 따라 데이터를 확보하기 어려울 때도 있어, 인공지능을 학습시키는데 어려움이 존재한다. 또한, 인공지능에는 전체 문서와 요약 문서만이 입력되므로 사용자가 요약문에 포함되어야 하

는 중요한 정보를 포함하도록 인공지능을 제어할 수 없다는 문제점이 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국공개특허 제10-2005-0043280호

발명의 내용

해결하려는 과제

[0009] 본 발명은 요약기 훈련시 입력되는 데이터에 제약조건을 부여하여 요약기의 성능을 향상키기 위해 마크업 언어를 생성하고, 생성된 마크업 언어로 요약기를 학습 효율을 향상시키는 것을 목적으로 한다.

과제의 해결 수단

[0011] 상기 목적을 달성하기 위하여 본 발명은, 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신단계; 상기 전체 문서와 상기 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출단계; 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출단계; 및 추출된 핵심문장과 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성단계;를 포함하는 것을 일 특징으로 한다.

[0012] 바람직하게는, 상기 핵심문장 추출단계는, LCS(Longest Common Subsequence) 알고리즘을 이용하여 상기 전체 문서와 상기 요약 문서를 비교할 수 있다.

[0013] 바람직하게는, 상기 핵심문장 추출단계는, 상기 전체 문서와 상기 요약 문서에 각각 설정된 키워드를 비교하여 상기 요약 문서에 설정된 키워드가 기설정된 빈도 이상 포함된 문장을 핵심문장으로 추출할 수 있다.

[0014] 바람직하게는, 상기 핵심단어 추출단계는, 상기 핵심문장과 상기 요약 문서에서 각각 키워드를 추출하고, 추출된 키워드 중 상기 핵심문장과 상기 요약 문서에 공통되는 키워드를 핵심단어로 추출할 수 있다.

[0015] 바람직하게는, 상기 핵심단어 추출단계는, 명사구를 키워드로 추출할 수 있다.

[0016] 또한 본 발명은, 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신부; 상기 문서 수신부에서 수신된 전체 문서와 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출부; 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출부; 및 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 핵심단어 추출부에서 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성부;를 포함하는 것을 다른 특징으로 한다.

[0017] 또한 본 발명은, 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신부, 상기 문서 수신부에서 수신된 전체 문서와 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출부, 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출부, 및 상기 핵심문장 추출부에서 추출된 핵심문장과 상기 핵심단어 추출부에서 추출된 핵심단어를 상기 전체 문서에 마크업하는 마크업 언어 생성부를 구비한 마크업 언어 생성부; 및 상기 마크업 언어 생성부에서 생성된 마크업 언어와 상기 요약 문서를 인공지능망에 입력하여 요약문을 생성하는 인공지능망을 학습시키는 학습부;를 포함하는 것을 또 다른 특징으로 한다.

[0018] 바람직하게는, 상기 학습부에서 학습된 인공지능망에 문서를 입력하여 요약문을 생성하는 요약문 생성부;를 더 포함할 수 있다.

발명의 효과

- [0020] 본 발명은 요약문이 갖춰야 할 필수정보가 태깅된 마크업 언어를 생성할 수 있다는 이점이 있다.
- [0021] 또한, 본 발명은 제약조건이 구축된 마크업 문서로 인공지능망을 학습하여 요약문 생성 성능을 높일 수 있다는 이점이 있다.
- [0022] 또한, 본 발명은 인공지능망의 학습에 제약조건이 구축된 마크업 문서를 이용하여 상대적으로 적은 데이터만으로도 유효한 학습을 수행할 수 있다는 이점이 있다.

도면의 간단한 설명

- [0024] 도 1은 본 발명의 실시예에 따른 핵심문장 추출단계에서 수행되는 문장 비교 과정을 나타낸다.
- 도 2는 본 발명의 실시예에 따른 핵심문장 추출단계에서 수행되는 Dialogue와 그 요약문의 비교 과정을 나타낸다.
- 도 3은 본 발명의 실시예에 따른 핵심단어 추출단계에서 수행되는 단어 비교 과정을 나타낸다.
- 도 4는 본 발명의 실시예에 따른 핵심단어 추출단계에서 수행되는 핵심단어 추출 과정을 나타낸다.
- 도 5는 본 발명의 실시예에 따른 마크업 언어 생성단계에서 생성한 마크업 언어를 나타낸다.
- 도 6은 본 발명의 실시예에 따른 마크업 언어 생성장치의 구성도를 나타낸다.
- 도 7은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치의 구성도를 나타낸다.
- 도 8은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치의 흐름도를 나타낸다.
- 도 9는 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치가 그림 추출에 적용된 예를 나타낸다.
- 도 10은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치가 음악 추출에 적용된 예를 나타낸다.
- 도 11은 본 발명의 실시예에 따른 제약조건 반영 정확도 측정기의 개략도를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0025] 이하, 첨부된 도면들에 기재된 내용들을 참조하여 본 발명을 상세히 설명한다. 다만, 본 발명이 예시적 실시 예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일 참조부호는 실질적으로 동일한 기능을 수행하는 부재를 나타낸다.
- [0026] 본 발명의 목적 및 효과는 하기의 설명에 의해서 자연스럽게 이해되거나 보다 분명해 질 수 있으며, 하기의 기재만으로 본 발명의 목적 및 효과가 제한되는 것은 아니다. 또한, 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다.
- [0027] 본 발명의 실시예에 따른 마크업 언어 생성방법은 문서 수신단계, 핵심문장 추출단계, 핵심단어 추출단계, 및 마크업 언어 생성단계를 포함할 수 있다.
- [0028] 기존의 문장생성 기법은 대량의 데이터로 학습된 장치를 이용하여 유창한 문장을 잘 생성한다. 그러나, 기존의 문장생성 기법은 소량의 데이터로 학습된 장치를 이용한 경우 유창한 문장을 생성하기에는 한계점이 있다. 본 발명의 실시예에 따른 마크업 언어 생성방법은 소량의 데이터만으로도 효과적으로 인공지능을 학습시키기 위해 다양한 제약조건(constraints)이 반영된 마크업 언어를 생성할 수 있다.
- [0029] 마크업 언어 생성방법은 인공지능에 입력되는 전체 문서에서 중요한 정보를 마크업을 통해 강조한 마크업 언어(Constraints Markup Language, CML)를 생성할 수 있다. 마크업 언어 생성방법은 제약조건이 부여된 마크업 언어를 생성하고 이를 학습에 사용하여, 인공지능망 출력물의 질을 높일 수 있다. 여기에서, 인공지능망의 입력물이 전체 문서일 경우 출력물은 요약 문서가 되고, 입력물이 사진에 대한 설명인 경우 출력물은 설명과 같은 사진이 되며, 입력물이 음악에 대한 설명인 경우 출력물은 설명과 같은 음악이 될 수 있다. 마크업 언어 생성방법은 제약조건으로 필수적인 키워드, 문장의 길이, 서식 등을 부여할 수 있다.
- [0030] 마크업 언어 생성방법은 단어, 구문 단위로 입력 문서(ex, 전체 문서)와 출력 문서(ex, 요약 문서)를 각각 비교

하여 중요한 정보로 정의되는 어구를 매칭하고, 이러한 어구에 자동으로 마크업할 수 있다.

[0031] 마크업 언어 생성단계는 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신할 수 있다.

[0032] 도 1은 본 발명의 실시예에 따른 핵심문장 추출단계에서 수행되는 문장 비교 과정을 나타낸다. 도 1을 참조하면, 핵심문장 추출단계는 전체 문서와 요약 문서를 비교하여 전체 문서에서 핵심문장을 추출할 수 있다. 핵심문장 추출단계는 요약 문서가 가지는 중요한 정보를 포함하는 전체 문서의 핵심문장을 추출할 수 있다.

[0033] 핵심문장 추출단계는 알고리즘을 활용하여 요약 문서의 문장과 전체 문서의 문장을 매칭할 수 있다. 핵심문장 추출단계는 LCS(Longest Common Subsequence) 알고리즘을 이용하여 전체 문서와 요약 문서를 비교할 수 있다. 여기에서, LCS란 공통부분 문자열 중 가장 길이가 긴 문자열을 찾는 알고리즘을 의미한다.

[0034] 도 2는 본 발명의 실시예에 따른 핵심문장 추출단계에서 수행되는 Dialog와 그 요약문의 비교 과정을 나타낸다. 도 2를 참조하면, 핵심문장 추출단계는 전체 문서와 요약 문서에 각각 설정된 키워드를 비교하여 요약 문서에 설정된 키워드가 기설정된 빈도 이상 포함된 문장을 핵심문장으로 추출할 수 있다.

[0035] 도 2의 요약 문서는 ‘Apprentice loves girls with red hair(제1 구문)’와 ‘Red hair is common in 1-2% of the human population(제2 구문)’을 포함한다. 여기에서, 제1 구문의 키워드는 ‘Apprentice, girls’, 제2 구문의 키워드는 ‘hair, population’이라 한다면, 전체 문서의 ‘Apprentice: I love girls who have red hair, it looks great’ 구문에는 ‘Apprentice, girls’ 키워드가 모두 포함되어 있어 제1 구문과 매칭되는 구문으로 추출될 수 있다. 이때, LCS(Longest Common Subsequence) 알고리즘을 동시에 적용한다면 정확한 매칭 구문이 추출될 수 있다. 예를 들면, 전체 문서의 ‘Wizard: naturally in 1-2% of the human population’ 구문에는 ‘hair’ 키워드가 포함되어 있고, 요약 문서의 제2 구문과 공통되는 문자열이 가장 길어 제2 구문과 매칭되는 구문으로 추출될 수 있다.

[0036] 핵심문장 추출단계는 LCS(Longest Common Subsequence) 알고리즘과 키워드 비교 방식 중 어느 하나를 택일적으로 이용하거나, 두 가지 방식을 모두 이용할 수 있다. 핵심문장 추출단계는 기설정된 빈도 이상의 키워드를 포함하는 복수개의 문장을 추출한 후 LCS 알고리즘을 적용하여 핵심문장을 추출할 수 있다.

[0037] 핵심단어 추출단계는 전체 문서와 요약 문서를 비교하여 핵심단어를 추출할 수 있다. 핵심단어 추출단계는 전체 문서에서 추출된 핵심문장과 요약 문서를 비교하여 핵심단어를 추출할 수 있다.

[0038] 도 3은 본 발명의 실시예에 따른 핵심단어 추출단계에서 수행되는 단어 비교 과정을 나타낸다. 도 3을 참조하면, 핵심단어 추출단계는 전체 문서에 포함된 키워드와 요약 문서에 포함된 키워드를 모두 추출할 수 있다. 핵심단어 추출단계는 전체 문서에 포함된 키워드 중 요약 문서에 포함되어 있지 않은 키워드와 요약 문서에 포함되어 있는 키워드를 구분할 수 있다.

[0039] 도 4는 본 발명의 실시예에 따른 핵심단어 추출단계에서 수행되는 핵심단어 추출 과정을 나타낸다. 도 4를 참조하면, 핵심단어 추출단계는 핵심문장과 요약 문서에서 각각 키워드를 추출하고, 추출된 키워드 중 핵심문장과 요약 문서에 공통되는 키워드를 핵심단어로 추출할 수 있다. 핵심단어 추출단계는 전체 문서에서 ‘Olivia, liberals, election, great’의 키워드를 포함하고, 요약 문서에서 ‘Olivia, liberals, election’의 키워드를 추출한다. 핵심단어 추출단계는 전체 문서에 포함되어 있는 키워드 중 요약 문서에 포함된 키워드인 ‘Olivia, liberals, election’를 핵심단어로 추출한다. 핵심단어 추출단계는 전체 문서에 포함된 키워드 중 요약 문서에 포함되어 있지 않은 키워드는 ‘great’은 별도로 구분한다.

[0040] 핵심단어 추출단계에서 추출하는 키워드는 전체 문서에서 핵심이 되고 중요한 정보를 의미한다. 핵심단어 추출단계는 명사구를 키워드로 추출할 수 있다.

[0041] 도 5는 본 발명의 실시예에 따른 마크업 언어 생성단계에서 생성한 마크업 언어를 나타낸다. 도 5를 참조하면, 마크업 언어 생성단계는 추출된 핵심문장과 추출된 핵심단어를 전체 문서에 마크업할 수 있다. 마크업 언어 생성단계는 전체 문서에서 핵심문장은 <PP>(Part Positive)로 마크업하고, 핵심문장이 아닌 문장은 <PN>(Part Negative)로 마크업할 수 있다. 구체적으로 마크업 언어 생성단계는 핵심문장의 맨앞에 <PP>을 태그하고 맨뒤에 </PP>을 태그하는 방식으로 마크업할 수 있다. 마크업 언어 생성단계는 핵심문장이 아닌 문장의 맨앞에 <PN>을 태그하고 맨뒤에 </PN>을 태그하는 방식으로 마크업할 수 있다.

[0042] 마크업 언어 생성단계는 전체 문서에서 핵심단어는 <WP>(Word Positive)로 마크업하고, 핵심단어가 아닌 키워드(또는 명사구)에는 <WN>(Word Negative)로 마크업할 수 있다. 구체적으로 마크업 언어 생성단계는 핵심단어의 맨앞에 <WP>를 태그하고 맨뒤에 </WP>를 태그하는 방식으로 마크업할 수 있다. 마크업 언어 생성단계는 핵심단

어가 아닌 키워드(또는 명사구)의 맨앞에 <WN>을 태그하고 맨뒤에 </WN>dmf 태그하는 방식으로 마크업할 수 있다.

- [0043] 도 6은 본 발명의 실시예에 따른 마크업 언어 생성장치(10)의 구성도를 나타낸다. 도 1을 참조하면, 마크업 언어 생성장치(10)는 문서 수신부(100), 핵심문장 추출부(300), 핵심단어 추출부(500), 및 마크업 언어 생성부(700)를 포함할 수 있다.
- [0044] 마크업 언어 생성장치(10)는 인공지능망을 학습시키는 학습 장치와 별개의 엔진으로 개발될 수 있다. 따라서, 마크업 언어 생성 부분과 인공지능을 이용한 문서 생성 부분이 별개로 개발되어 필요시 결합 사용될 수 있다.
- [0045] 문서 수신부(100)는 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신할 수 있다. 문서 수신부(100)는 전술한 문서 수신단계의 동작을 수행할 수 있다.
- [0046] 핵심문장 추출부(300)는 문서 수신부(100)에서 수신된 전체 문서와 요약 문서를 비교하여 전체 문서에서 핵심문장을 추출할 수 있다. 핵심문장 추출부(300)는 LCS(Longest Common Subsequence) 알고리즘을 이용하여 전체 문서와 요약 문서를 비교할 수 있다. 핵심문장 추출부(300)는 전체 문서와 요약 문서에 각각 설정된 키워드를 비교하여 요약 문서에 설정된 키워드가 기설정된 빈도 이상 포함된 문장을 핵심문장으로 추출할 수 있다. 핵심문장 추출부(300)는 전술한 핵심문장 추출단계의 동작을 수행할 수 있다.
- [0047] 핵심단어 추출부(500)는 핵심문장 추출부에서 추출된 핵심문장과 상기 요약 문서를 비교하여 핵심단어를 추출할 수 있다. 핵심단어 추출부(500)는 핵심문장과 요약 문서에서 각각 키워드를 추출하고, 추출된 키워드 중 핵심문장과 요약 문서에 공통되는 키워드를 핵심단어로 추출할 수 있다. 핵심단어 추출부(500)는 명사구를 키워드로 추출할 수 있다. 핵심단어 추출부(500)는 전술한 핵심단어 추출단계의 동작을 수행할 수 있다.
- [0048] 마크업 언어 생성부(700)는 핵심문장 추출부(300)에서 추출된 핵심문장과 핵심단어 추출부(500)에서 추출된 핵심단어를 전체 문서에 마크업할 수 있다. 마크업 언어 생성부(700)는 전술한 마크업 언어 생성단계의 동작을 수행할 수 있다.
- [0049] 도 7은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치(1)의 구성도를 나타낸다. 도 7을 참조하면, 마크업 언어를 이용한 문장 생성장치(1)는 마크업 언어 생성부(10), 학습부(30), 및 요약문 생성부(50)를 포함할 수 있다.
- [0050] 도 8은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치의 흐름도를 나타낸다. 도 8을 참조하면, 마크업 언어를 이용한 문장 생성장치(1)는 제약조건이 적용된 마크업 언어로 인공지능망을 학습시킴으로써 적은 데이터를 이용하여 효과적으로 인공지능망을 학습시킬 수 있다. 또한, 마크업 언어를 이용한 문장 생성장치(1)는 사용자가 임의적으로 제약조건을 변경할 수 있어 학습 방향을 설정하거나 수정할 수 있다.
- [0051] 마크업 언어 생성부(10)는 전체 문서와 상기 전체 문서를 요약한 요약 문서를 수신하는 문서 수신부(100), 상기 문서 수신부에서 수신된 전체 문서와 요약 문서를 비교하여 상기 전체 문서에서 핵심문장을 추출하는 핵심문장 추출부(300), 핵심문장 추출부(300)에서 추출된 핵심문장과 요약 문서를 비교하여 핵심단어를 추출하는 핵심단어 추출부(500), 및 핵심문장 추출부(300)에서 추출된 핵심문장과 핵심단어 추출부(500)에서 추출된 핵심단어를 전체 문서에 마크업하는 마크업 언어 생성부(700)를 구비할 수 있다. 마크업 언어 생성부(10)는 전술한 마크업 언어 생성 장치(10)와 동일한 구성이며, 전술한 마크업 언어 생성방법을 이용하여 마크업 언어를 생성할 수 있다.
- [0052] 학습부(30)는 마크업 언어 생성부에서 생성된 마크업 언어와 상기 요약 문서를 인공지능망에 입력하여 요약문을 생성하는 인공지능망을 학습시킬 수 있다.
- [0053] 요약문 생성부(50)는 학습부(30)에서 학습된 인공지능망에 문서를 입력하여 요약문을 생성할 수 있다.
- [0054] 도 9는 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치(1)가 그림 추출에 적용된 예를 나타낸다. 도 9를 참조하면, 마크업 언어를 이용한 문장 생성장치(1)는 그림 추출에도 적용될 수 있다. 마크업 언어를 이용한 문장 생성장치(1)는 제약조건이 설정된 마크업 언어(CML)과 그림으로 인공지능망을 학습하여, 문장이 입력되었을 때 관련된 그림이 출력되게 할 수 있다.
- [0055] 도 10은 본 발명의 실시예에 따른 마크업 언어를 이용한 문장 생성장치가 음악 추출에 적용된 예를 나타낸다. 도 10을 참조하면, 마크업 언어를 이용한 문장 생성장치(1)는 제약조건이 설정된 마크업 언어(CML)과 음악으로 인공지능망을 학습하여, 문장이 입력되었을 때 관련된 음악이 출력되게 할 수 있다.

[0056] 도 11은 본 발명의 실시예에 따른 제약조건 반영 정확도 측정기의 개략도를 나타낸다. 도 11을 참조하면, 마크업 언어를 이용한 문장 생성장치(1)는 제약조건이 설정된 마크업 언어(CML)와 출력된 요약문의 일치도를 분석하여 인공지능망의 정확도를 측정할 수 있다. 여기에서 일치도는 (요약문에 포함된 핵심단어/전체핵심단어)의 공식으로 산출될 수 있다.

[0057] 이상에서 대표적인 실시예를 통하여 본 발명을 상세하게 설명하였으나, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리 범위는 설명한 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 특허청구범위와 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태에 의하여 정해져야 한다.

부호의 설명

[0059] 1 : 마크업 언어를 이용한 문장 생성장치

10 : 마크업 언어 생성장치

30 : 학습부

50 : 요약문 생성부

100 : 문서 수신부

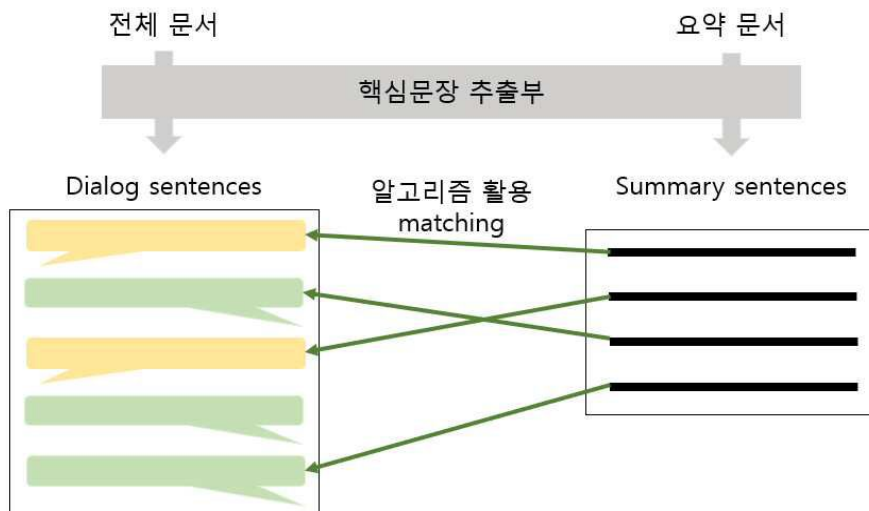
300 : 핵심문장 추출부

500 : 핵심단어 추출부

700 : 마크업 언어 생성부

도면

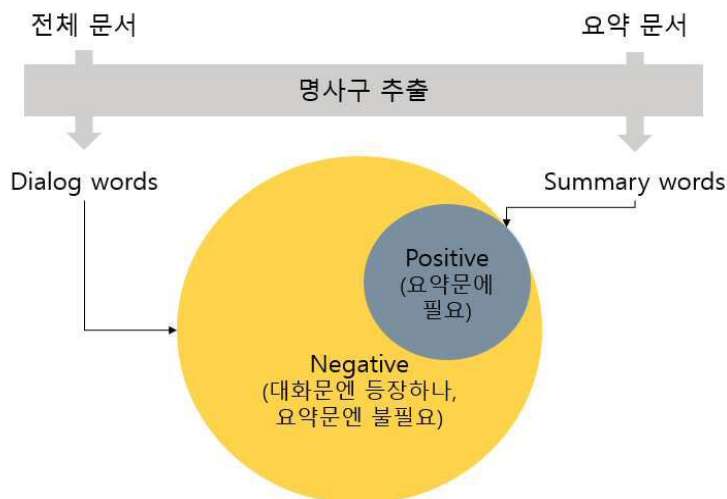
도면1



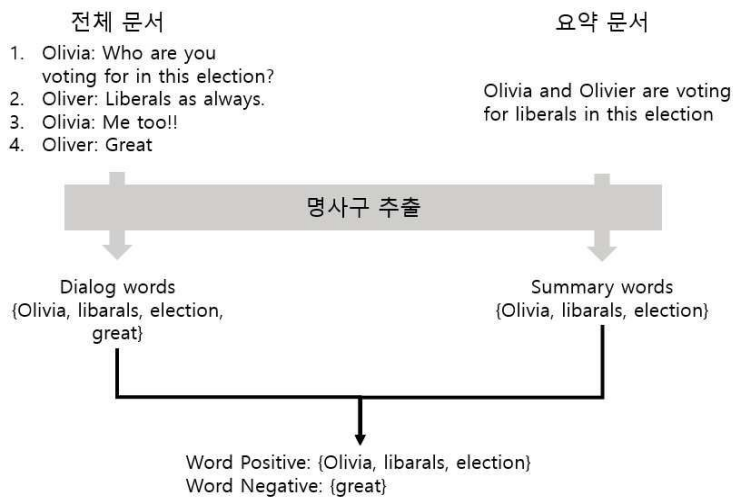
도면2



도면3



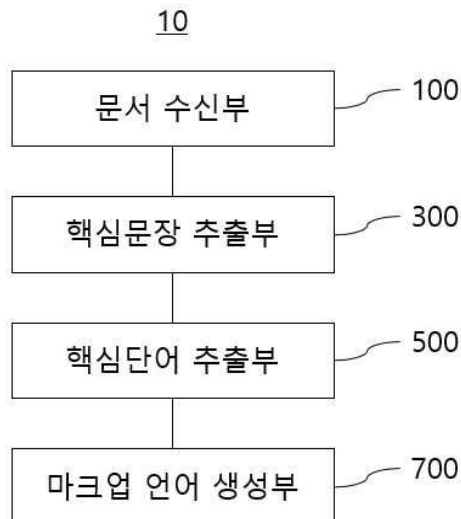
도면4



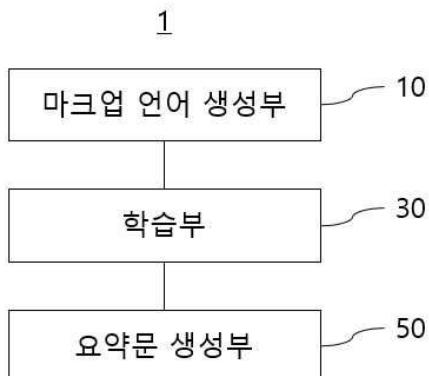
도면5

Dialogue Apprentice: I love girls who have red hair, it looks great. Wizard: naturally in 1-2% of the human population. Apprentice: Really? That uncommon? Wouldve thought closer to 10% Wizard: Its common with western European ancestry, a Apprentice: Oh yea like the irish and such? Makes sense, seems more like a british thing. Wizard: More so,aries in hues from a deep burgundy or bright copper (reddish-brown or auburn) through to burnt orange or red-orange and strawberry Apprentice: Yea the deeper reds are cool, most are more orange. Wizard: They have range pigments are largely in the ochre or cadmium families	
CML <DP> Hair Color</DP> <PP> <WP>apprentice</WP>: i love <WP>girls</WP> who have red <WP>hair</WP>. it looks great.</PP> <PP> wizard: naturally in 1-2<WP>%</WP> of the human <WP>population</WP>.</PP> apprentice: really? that uncommon? wouldve thought closer to 10% wizard: its common with <WN>western</WN> <WN>European</WN> <WN>ancestry</WN> a apprentice: oh yea like the irish and such? makes <WN>sense</WN> seems more like a british thing. <PP> wizard: more so,aries in <WP>hues</WP> from a deep <WP>burgundy</WP> or bright <WP>copper</WP> <WN>reddish</WN> <WN>brown</WN> or auburn) through to burnt <WP>orange</WP> or red-orange and <WN>strawberry</WN> apprentice: yea the deeper <WP>reds</WP> are cool, most are more orange.</PP> <PP> wizard: they have range <WP>pigments</WP> are largely in the <WP>ochre</WP> or <WP>cadmium</WP> <WP>families</WP>.</PP> <NP> apprentice: i want chocolate icecream.</NP>	
Summary Apprentice loves girls with red hair. Red hair is common in 1-2% of the human population. It comes in hues from deep burgundy or bright copper to burnt orange or red-orange and strawberry. The deeper reds are cool. The pigments are largely in the ochre or cadmium families.	

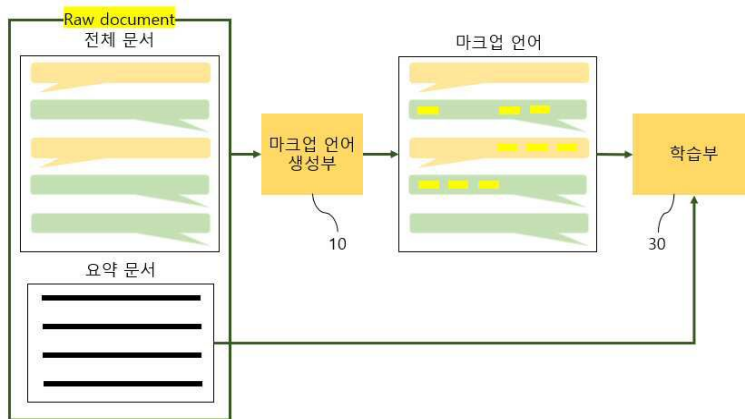
도면6



도면7

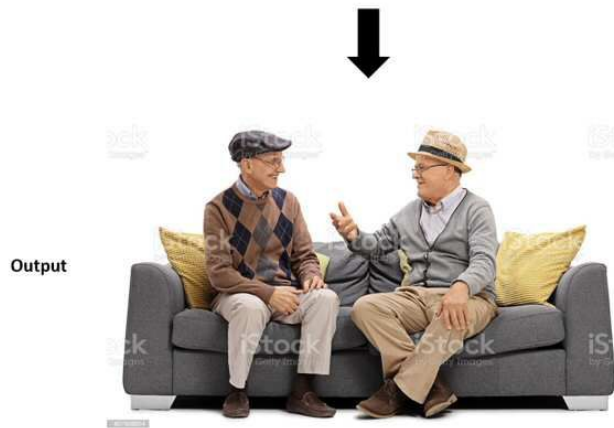


도면8



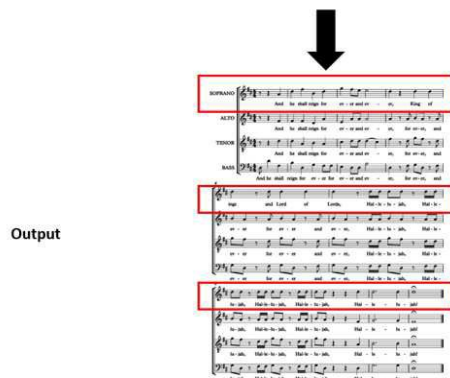
도면9

CML <WP>셔츠 <\WP>를 입은 <WP> 노인 두 명 <\WP>이 <WP> 대화 <\WP> 한다.



도면10

CML 경쾌한 <WP>소프라노 <\WP>의 목소리로 <WP>찬송가 <\WP>를 부른다.



도면11

