



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0098162
(43) 공개일자 2008년11월07일

(51) Int. Cl.

G06F 17/27 (2006.01)

(21) 출원번호 10-2007-0043497

(22) 출원일자 2007년05월04일

심사청구일자 2007년05월04일

(71) 출원인

건국대학교 산학협력단

서울 광진구 화양동 1 건국대학교내

(72) 발명자

정원석

서울 서초구 서초동 1548-11 202호

백현수

충북 청주시 흥덕구 사창동 동신아파트 다동 203호

(74) 대리인

이은철

전체 청구항 수 : 총 9 항

(54) 글짓기 검사모듈 및 검사방법

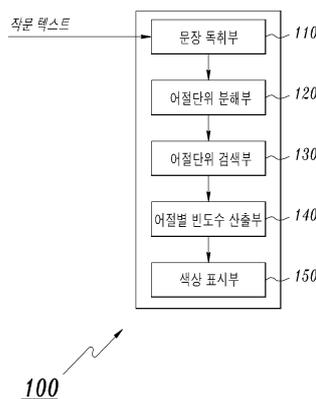
(57) 요약

본 발명은 글짓기 검사모듈 및 검사방법에 관한 것이다.

본 발명이 개시하는 글짓기 검사모듈은, 사용자가 입력한 작문 텍스트를 개별 문장으로 구분·독취하는 문장 독취부와, 분해된 개별 문장을 n개의 어절로 이루어진 어절단위로 분해하는 어절단위 분해부와, 검색서버를 매개로 분해된 어절단위들 각각에 대해 웹문서 기반의 검색결과를 취득하는 어절단위 검색부와, 개별 문장을 이루는 각 어절에 대한 빈도수를 산출하되 해당 어절이 포함된 어절단위 검색결과들의 평균으로 산출하는 어절별 빈도수 산출부, 그리고 개별 문장을 이루는 각 어절을 산출된 빈도수의 범위에 따라 색상으로 차등 표시하는 색상 표시부를 구성한다.

본 발명에 따르면, 사용자가 작성한 작문 텍스트에 대해 종래 단어 중심의 기계적 검사가 아닌 보다 객관적인 지표(웹문서)를 기반으로 검사를 수행할 수 있다. 또한, 문장을 이루는 각 어절의 적합 여부는 물론이고 어절을 포함하는 패턴의 적합 여부를 검사할 수 있다.

대표도 - 도2



특허청구의 범위

청구항 1

검색서버와 인터넷 통신 가능한 사용자단말기에 탑재되는 글짓기 검사모듈에 있어서,
 사용자가 입력한 작문 텍스트를 개별 문장으로 구분·독취하는 문장 독취부;
 분해된 개별 문장을 n개의 어절로 이루어진 어절단위(n-Gram)로 분해하는 어절단위 분해부;
 상기 검색서버를 매개로 분해된 어절단위들 각각에 대해 웹문서 기반의 검색결과를 취득하는 어절단위 검색부;
 상기 개별 문장을 이루는 각 어절에 대한 빈도수를 산출하되, 해당 어절이 포함된 어절단위 검색결과들의 평균
 으로 산출하는 어절별 빈도수 산출부; 및
 상기 개별 문장을 이루는 각 어절을 산출된 빈도수의 범위에 따라 색상으로 차등 표시하는 색상 표시부; 를 포
 함하는 것을 특징으로 하는 글짓기 검사모듈.

청구항 2

청구항 1에 있어서,
 상기 어절별 빈도수 산출부는,
 상기 어절단위 검색결과가 상한값을 초과할 경우, 그 검색결과를 상한값으로 대체하는 것을 특징으로 하는 것을
 특징으로 하는 글짓기 검사모듈.

청구항 3

청구항 1에 있어서
 상기 n은 3인 것을 특징으로 하는 글짓기 검사모듈.

청구항 4

검색서버와 인터넷 통신 가능한 사용자단말기를 기반으로 사용자가 입력한 작문 텍스트를 검사하는 방법에 있어
 서,
 상기 작문 텍스트를 개별 문장으로 구분·독취하는 제1 과정;
 분해된 개별 문장을 n개의 어절로 이루어진 어절단위로 분해하는 제2 과정;
 상기 검색서버를 매개로 분해된 어절단위들 각각에 대해 웹문서 기반의 검색결과를 취득하는 제3 과정;
 상기 개별 문장을 이루는 각 어절에 대한 빈도수를 산출하되, 해당 어절이 포함된 어절단위 검색결과들의 평균
 으로 산출하는 제4 과정; 및
 상기 개별 문장을 이루는 각 어절을 산출된 빈도수의 범위에 따라 색상으로 차등 표시하는 제5 과정; 을 포함하
 는 것을 특징으로 하는 글짓기 검사방법.

청구항 5

청구항 4에 있어서,
 상기 제4 과정은,
 상기 어절단위 검색결과가 상한값을 초과할 경우, 그 검색결과를 상한값으로 대체하는 단계를 포함하는 것을 특
 징으로 하는 글짓기 검사방법.

청구항 6

청구항 4에 있어서,
 상기 제2 과정은,

상기 개별 문장을 3개의 어절로 구성된 어절단위로 분해하는 단계를 포함하는 것을 특징으로 하는 글짓기 검사 방법.

청구항 7

컴퓨터를,

사용자가 입력한 작문 텍스트를 개별 문장으로 구분·독취하는 제1 기능;

분해된 개별 문장을 n개의 어절로 이루어진 어절단위로 분해하는 제2 기능;

검색서버를 매개로 분해된 어절단위들 각각에 대해 웹문서 기반의 검색결과를 취득하는 제3 기능;

상기 개별 문장을 이루는 각 어절에 대한 빈도수를 산출하되, 해당 어절이 포함된 어절단위 검색결과들의 평균으로 산출하는 제4 기능; 및

상기 개별 문장을 이루는 각 어절을 산출된 빈도수의 범위에 따라 색상으로 차등 표시하는 제5 기능; 으로 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 8

청구항 7에 있어서,

상기 제4 기능은,

상기 어절단위 검색결과가 상한값을 초과할 경우, 그 검색결과를 상한값으로 대체하는 기능을 포함하는 것을 특징으로 하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 9

청구항 7에 있어서,

상기 제2 기능은,

상기 개별 문장을 3개의 어절로 구성된 어절단위로 분해하는 기능을 포함하는 것을 특징으로 하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- <13> 본 발명은 글짓기 검사모듈 및 검사방법에 관한 것으로서, 특히 사용자가 입력한 문장을 어절단위(n개의 어절로 구성된 단위)로 구분하고, 이들을 검색서버들을 통해 검색한 결과의 누적 빈도수로부터 글짓기의 적합성을 검사하는 기술에 관한 것이다.
- <14> 영어를 제2 외국어로 구사하려는 사용자는 자신이 작성한 텍스트(영작 텍스트)가 적법한지 여부를 검증할 필요가 있다. 영작 텍스트의 적합성을 검증하기 위해 널리 이용되는 것은, 인터넷을 매개로 영어 능통자(혹은 능숙자)로부터 검수를 받는 방식이다. 그러나 이러한 방식은 시간소요는 물론이고 즉시성이 결여되는 문제점이 있다.
- <15> 한편, 영어 능통자의 검수가 아닌 교정 소프트웨어를 이용할 수도 있다. 교정 소프트웨어는 미리 축적된 소정의 단어 데이터베이스를 근간으로 기능한다. 이러한 교정 소프트웨어는 단어 중심의 교정에 한정될 뿐만 아니라 데이터베이스에 대한 지속적인 갱신이 요구된다. 나아가 기 축적된 단어 데이터베이스에 전적으로 의존하므로 단순한 기계적 결과 제공에 머물 뿐이다.

발명이 이루고자 하는 기술적 과제

- <16> 본 발명은 상기와 같은 문제점을 감안하여 안출된 것으로, 사용자가 작성한 텍스트의 문장을 기준으로, 문장을

이루는 각 어절(또는 단어)의 적합 여부를 다수의 웹문서를 참조하여 판별할 수 있도록 한다.

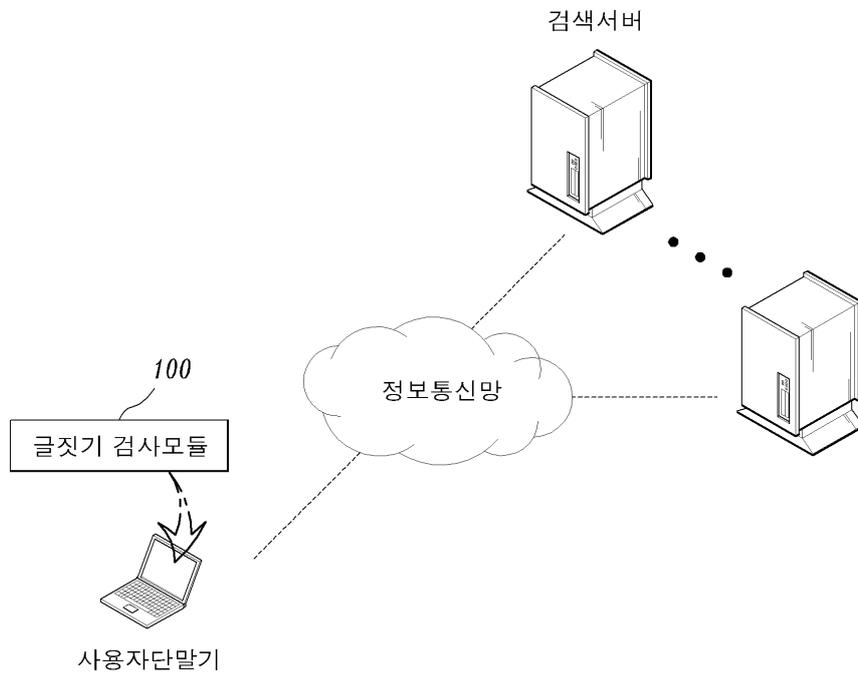
- <17> 구체적으로 본 발명은, 해당 문장을 어절단위(n-Gram)로 검색하여, 그 검색결과(문서의 개수)를 어절단위별로 축적하고, 문장을 이루는 각 어절에 대해 축적된 검색결과들을 바탕으로 빈도수를 산출하고, 산출된 빈도수에 의거하여 어절을 색상으로 차등 표시한다. 이를 통해 사용자에게 해당 어절의 적합성을 판단할 수 있도록 한다.
- <18> 본 발명의 구체적 특징 및 이점들은 첨부도면에 의거한 다음의 상세한 설명으로 더욱 명백해질 것이다. 이에 앞서 본 발명에 관련된 공지 기능 및 그 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는, 그 구체적인 설명을 생략하였음에 유의해야 할 것이다.

발명의 구성 및 작용

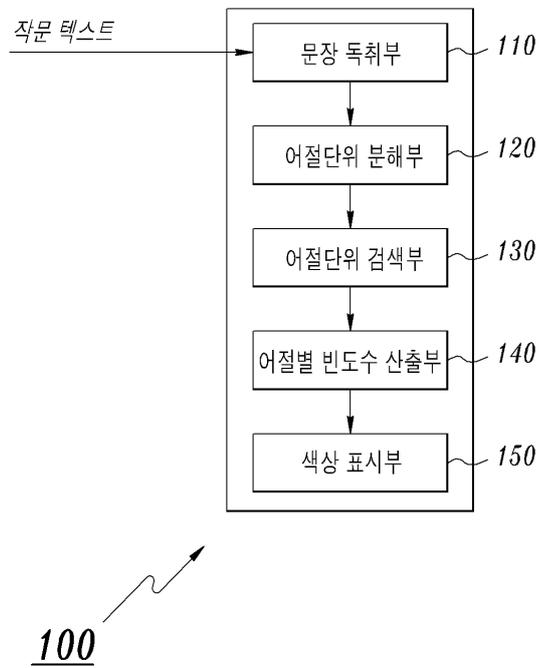
- <19> 첨부도면 도 1은 본 발명의 글짓기 검사모듈이 적용되는 시스템을 보인 일예시도이다. 도시된 바와 같이 본 발명의 글짓기 검사모듈(100)은 검색서버와 인터넷 통신 가능한 사용자단말기에 탑재된다. 사용자단말기는 개인컴퓨터(PC)를 비롯한 PDA, 휴대폰이 될 수 있다.
- <20> 사용자단말기에 탑재되는 글짓기 검사모듈(100)은, 도 2와 같이 기본 구성으로서, 문장 독취부(110), 어절단위 분해부(120), 어절단위 검색부(130), 어절별 빈도수 산출부(140) 및 색상 표시부(150)를 포함한다.
- <21> 문장 독취부(110)는 사용자의 작문 텍스트를 입력받아 개별 문장으로 분해하여 독취한다. 여기서, 문장(sentence)은 따옴표(", ') 및 마침표(.) 등의 특수기호로 구분될 수 있고, 띄어쓰기(space)에 의해 다수의 어절(문장 성분의 최소단위)로 이루어진다. 어절은 대개 단어로 취급될 수 있다.
- <22> 어절단위 분해부(120)는 상기 문장 독취부(110)로부터 분해된 문장에서 n개의 어절을 1개의 단위로 나눈다(n-Gram으로도 표현됨). 예컨대, 도 3과 같이 "It would be better to do now"라는 문장에 n=3인 어절단위를 적용할 경우, "It would be", "would be better", "be better to", "better to do" 및 "to do now"로 분해된다. 어절단위에서 n은 예시한 바와 같이 3으로 설정하는 것이 바람직하나, 2 또는 4로도 설정될 수 있다. 본 발명은 이하의 설명에서 n이 3인 경우를 기준으로 한다.
- <23> 어절단위 검색부(130)는 검색서버를 매개로 앞서 분해된 어절단위들 각각에 대한 검색결과를 얻고 이들을 저장한다. 여기서, 각 어절단위에 대한 검색결과는 검색된 웹문서의 개수를 의미하며, 검색서버는 어느 특정 검색서버에 한정되지 않는다.
- <24> 어절별 빈도수 산출부(140)는 상기 검색결과들을 근간으로 상기 문장을 구성하는 어절별로 빈도수를 산출한다. 본 발명의 특징에 따라, 어절에 대한 빈도수 산출은, 해당 어절을 포함하고 있는 검색결과들에 대한 평균값을 이용한다. 이를 부연하면, 앞서 예시한 문장에서 어절 "better"는, 3개의 어절단위("would be better", "be better to", "better to do")에 포함되어 있다(도 4 참조). 가령, 각 어절단위의 검색결과가 도면과 같이 100, 500, 150이라면 어절 "better"의 빈도수는 이들의 평균인 250이 된다. 이때, 고려되어야할 점은 어느 어절단위의 검색결과가 나머지 어절단위 검색결과들에 비해 너무 클 경우(예: 1억개), 평균의 의미는 없어진다. 따라서 본 발명의 어절별 빈도수 산출부(140)는 검색결과에 대한 상한값을 적용한다. 예를 들어, 상한값을 400으로 설정할 경우, 도 4에서 어절단위 "be better to"의 검색결과가 설정한 상한값을 상회하므로 500을 상한값 400으로 취하는 것이다.
- <25> 색상 표시부(150)는 문장을 구성하는 각 어절을 그 산출된 빈도수에 따라 색상으로 차등 표시하여, 사용자가 각 어절의 적합성을 용이하게 판단할 수 있도록 한다. 빈도수에 따른 색상 차등 표시를 위해서는, 빈도수의 범위와 각 범위에 따른 색상이 미리 지정되어야 한다. 예컨대, 임계치를 300으로 상정하고, 빈도수가 300이상인 경우 검은색으로, 299~100인 경우 갈색, 99~50인 경우 주황색, 49이하인 경우 빨간색으로, 각 범위와 색상이 설정될 수 있다. 물론, 본 발명이 이와 같이 예시한 범위 및 색상에 국한되는 것은 아니다.
- <26> 본 발명의 특징에 따르면, 어떤 어절의 빈도수가 많다는 것은, 그 어절의 전후 어절을 포함하는 패턴(예를 들어, 어절 "better"의 전후 어절은 "be"와 "to"가 되며, 그 패턴은 "be better to"가 됨)이, 그 만큼 많이 사용되고 있다는 것을 의미한다. 따라서 본 발명은, 종래 교정 소프트웨어가 제공하는 기계적 검사와는 달리 방대한 웹문서를 토대로 어절의 적합성은 물론이고 패턴에 대한 검사를 실시할 수 있다.
- <27> 이하, 도 5를 참조하여 본 발명의 글짓기 검사방법을 정리한다.
- <28> 사용자가 작성한 작문 텍스트에서, 특수기호 및 띄어쓰기를 기준으로 개별 문장으로 구분·독취한 후(S110), 독취한 문장을 n개의 어절로 이루어진 어절단위(n-Gram)로 분해한다(S120).

도면

도면1



도면2



도면3

It would be better to do now



It would be
 would be better
 be better to
 better to do
 to do now

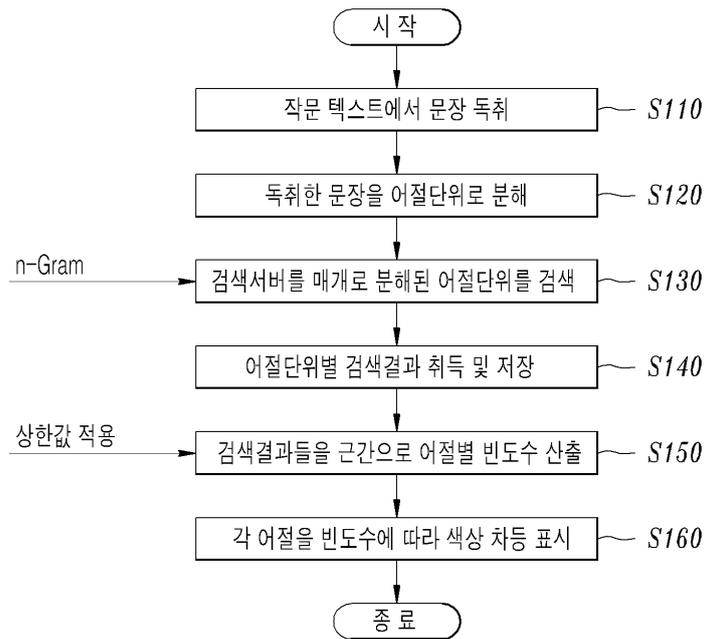
도면4

어절 'better'가 포함된 어절단위	검색결과
would be better	100
be better to	500
better to do	150



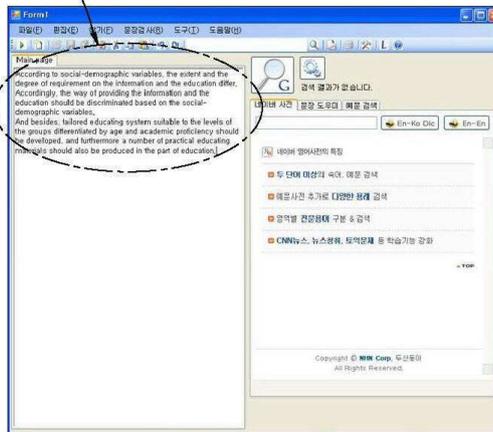
$$\frac{100+500+150}{3} = 250$$

도면5



도면6a

작문 텍스트(영문)



도면6b

검시수행 결과(어절별 색상 차등 표시)

