



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0083747
(43) 공개일자 2009년08월04일

(51) Int. Cl.

G06Q 50/00 (2006.01)

(21) 출원번호 10-2008-0009731

(22) 출원일자 2008년01월30일

심사청구일자 없음

(71) 출원인

삼성전자주식회사

경기도 수원시 영통구 매탄동 416

서강대학교산학협력단

서울 마포구 신수동 1-1 서강대학교

(72) 발명자

정지혜

서울 송파구 잠실본동 240-8번지 402호

이종호

경기 성남시 분당구 구미동 까치마을주공2단지아파트 203-1902호

(뒷면에 계속)

(74) 대리인

정홍식, 김종선, 김태현, 이현수

전체 청구항 수 : 총 15 항

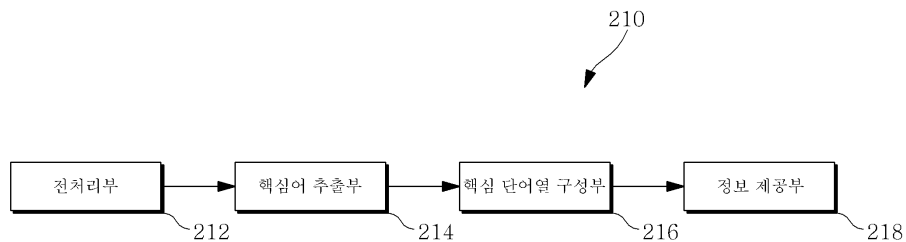
(54) 웹 문서 요약 제공을 위한 사용자 단말 장치 및 웹 문서제공 방법

(57) 요약

본 발명은 웹 페이지에 포함된 문서를 요약하여 제공하기 위한 사용자 단말 장치 및 웹 정보 문서 방법을 제안한다.

본 발명은 웹 문서에 포함된 형태소를 분석하는 전처리부; 분석 결과에 기초하여 적어도 하나의 핵심어를 추출하는 핵심어 추출부; 웹 문서에 포함된 각 문장별 형태소와 상기 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성하는 핵심 단어열 구성부; 및 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공하는 정보 제공부를 포함하는 것을 특징으로 하는 사용자 단말 장치를 제공하여, 사용자가 웹 문서의 내용을 신속하고 정확하게 파악하는 것을 가능하게 한다.

대표도 - 도3



(72) 발명자

이혜정

서울 송파구 풍납2동 쌍용아파트 101동 404호

김연배

경기 성남시 분당구 서현동 포스파크 103동 303호

서정연

서울 마포구 신수동 1 서강대학교 컴퓨터공학과

고영중

부산 사하구 하단2동 840 동아대학교 컴퓨터공학과

특허청구의 범위

청구항 1

웹 문서에 포함된 형태소를 분석하는 전처리부;

분석 결과에 기초하여 적어도 하나의 핵심어를 추출하는 핵심어 추출부;

상기 웹 문서에 포함된 각 문장별 형태소와 상기 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성하는 핵심 단어열 구성부; 및

상기 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공하는 정보 제공부를 포함하는 것을 특징으로 하는 사용자 단말 장치.

청구항 2

제1항에 있어서,

상기 전처리부는 각 형태소의 품사, 위치 및 출현 빈도수를 포함하는 정보를 생성하는 것을 특징으로 하는 사용자 단말 장치.

청구항 3

제1항에 있어서,

상기 핵심어 추출부는,

상기 웹 문서에 포함된 각 형태소의 위치 및 출현 빈도 중 적어도 하나에 기초하여 각 형태소의 가중치 정보를 생성하고, 상기 가중치 정보가 소정의 임계값 이상인 형태소를 핵심어로 추출하는 것을 특징으로 하는 사용자 단말 장치.

청구항 4

제1항에 있어서,

상기 핵심 단어열 구성부는,

상기 웹 문서에 포함된 문장 중에서 상기 적어도 하나의 핵심어와 관련된 적어도 하나의 중요 문장을 선택하고,

상기 중요 문장별로 상기 중요 문장에 포함된 형태소와 상기 적어도 하나의 핵심어를 각각 매칭시키고, 상기 중요 문장별로 상기 매칭된 순서대로 배열된 적어도 하나의 핵심 단어열을 구성하는 것을 특징으로 하는 사용자 단말 장치.

청구항 5

제4항에 있어서,

상기 핵심 단어열 구성부는,

상기 각 핵심어의 가중치 정보 및 상기 웹 문서에 포함된 문장의 위치 중 적어도 하나에 기초하여 상기 문장의 중요도 정보를 생성하고 중요도가 소정의 임계값 이상인 문장을 중요 문장으로 선택하는 것을 특징으로 하는 사용자 단말 장치.

청구항 6

제1항에 있어서,

상기 핵심 단어열 구성부는,

상기 적어도 하나의 핵심 단어열 중에서 상기 각 문장별 형태소와 가장 길이가 길게 매칭되는 핵심 단어열을 선택하고,

상기 정보 제공부는 상기 선택된 핵심 단어열을 우선적으로 이용하여 웹 문서 요약 정보를 제공하는 것을 특징

으로 하는 사용자 단말 장치.

청구항 7

제1항에 있어서,

상기 정보 제공부는,

상기 핵심 단어열이 매칭된 상기 문장들이 상기 웹 문서에 나타난 순서대로, 상기 핵심 단어열이 배열되도록 웹 문서 요약 정보를 제공하는 것을 특징으로 하는 사용자 단말 장치.

청구항 8

제1항에 있어서,

상기 정보 제공부는 사용자의 입력에 따라 상기 적어도 하나의 핵심 단어열을 이용하여 생성된 웹 문서 요약 정보, 상기 웹 문서의 문장 중 상기 핵심 단어열이 포함된 중요 문장 및 원래의 웹 문서를 순차적으로 제공하는 것을 특징으로 하는 사용자 단말 장치.

청구항 9

웹 문서에 포함된 형태소를 분석하는 단계;

분석 결과에 기초하여 적어도 하나의 핵심어를 추출하는 단계;

상기 웹 문서에 포함된 각 문장별 형태소와 상기 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성하는 단계; 및

상기 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공하는 단계를 포함하는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 10

제9항에 있어서,

상기 적어도 하나의 핵심어를 추출하는 단계는,

상기 웹 문서에 포함된 각각의 형태소의 위치 및 출현 빈도 중 적어도 하나에 기초하여 각 형태소의 가중치 정보를 생성하고, 상기 가중치 정보가 소정의 임계값 이상인 형태소를 핵심어로 추출하는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 11

제9항에 있어서,

상기 핵심 단어열을 구성하는 단계는,

상기 웹 문서에 포함된 문장 중에서 상기 적어도 하나의 핵심어를 포함하는 적어도 하나의 중요 문장을 선택하는 단계; 및

상기 중요 문장별로 상기 중요 문장에 포함된 형태소와 상기 적어도 하나의 핵심어를 각각 매칭시키고, 상기 중요 문장별로 상기 매칭된 순서대로 배열된 적어도 하나의 핵심 단어열을 구성하는 단계를 포함하는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 12

제11항에 있어서,

상기 적어도 하나의 중요 문장을 선택하는 단계는,

상기 각 핵심어의 가중치 정보 및 상기 웹 문서에 포함된 문장의 위치 중 적어도 하나에 기초하여 상기 문장의 중요도 정보를 생성하고 중요도가 소정의 임계값 이상인 문장을 중요 문장으로서 선택하는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 13

제9항에 있어서,

상기 핵심 단어열을 구성하는 단계는,

상기 적어도 하나의 핵심 단어열 중에서 상기 각 문장별 형태소와 가장 길이가 길게 매칭되는 핵심 단어열을 선택하는 단계를 포함하고,

상기 웹 문서 요약 정보를 제공하는 단계에서 상기 선택된 핵심 단어열을 우선적으로 이용하여 웹 문서 요약 정보가 구성되는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 14

제9항에 있어서,

상기 웹 문서 요약 정보를 제공하는 단계에서,

상기 핵심 단어열이 매칭된 상기 문장들이 상기 웹 문서에 나타난 순서대로, 상기 핵심 단어열이 배열되도록 제공하는 것을 특징으로 하는 웹 문서 제공 방법.

청구항 15

제9항에 있어서,

사용자의 입력에 따라 상기 적어도 하나의 핵심 단어열을 이용하여 생성된 웹 문서 요약 정보, 상기 웹 문서의 문장 중 상기 핵심 단어열이 포함된 중요 문장, 원래의 웹 문서를 순차적으로 제공하는 단계를 더 포함하는 것을 특징으로 하는 웹 문서 제공 방법.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 웹 문서 제공 장치 및 방법에 관한 것으로, 특히 웹 문서를 요약하여 제공하기 위한 사용자 단말 장치 및 웹 문서 제공 방법에 관한 것이다.

배경 기술

- <2> 유무선 인터넷의 이용이 증가함에 따라 개인용 컴퓨터에서만 아니라 IPTV나 PDA, 휴대폰 등과 같은 여러가지 형태의 단말 장치에서 웹 문서를 검색하여 이용하는 경우가 늘고 있다. 그러나, 디스플레이의 크기나 사용자 입력 수단이 통상의 컴퓨터와 다른, 익숙하지 않은 디바이스 환경에서 웹 문서의 가독성은 저하된다. 특히 소형 디스플레이를 가진 단말 장치에서 웹 문서의 내용을 빠르고 정확하게 파악하기 어려운 문제점이 있다.

발명의 내용

- <3> 따라서, 본 발명은 전술한 문제점을 해결하기 위해 안출된 것으로, 실시간으로 정확하게 웹 문서의 요약 정보를 제공하기 위한 사용자 단말 장치 및 방법을 제안한다.
- <4> 발명의 일 양상에 따른 웹 문서 제공 장치는 웹 문서에 포함된 형태소를 분석하는 전처리부; 분석 결과에 기초하여 적어도 하나의 핵심어를 추출하는 핵심어 추출부; 웹 문서에 포함된 각 문장별 형태소와 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성하는 핵심 단어열 구성부; 및 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공하는 정보 제공부를 포함한다.
- <5> 전처리부는 각 형태소의 품사, 위치 및 출현 빈도수를 포함하는 정보를 생성할 수 있다. 또한, 핵심어 추출부는 웹 문서에 포함된 각 형태소의 위치 및 출현 빈도 중 적어도 하나에 기초하여 각 형태소의 가중치 정보를 생성하고, 가중치 정보가 소정의 임계값 이상인 형태소를 핵심어로 추출할 수 있다.
- <6> 본 발명의 일 실시예에 따르면, 핵심 단어열 구성부는 웹 문서에 포함된 문장 중에서 적어도 하나의 핵심어와 관련된 적어도 하나의 중요 문장을 선택하고, 중요 문장별로 중요 문장에 포함된 형태소와 적어도 하나의 핵심

어를 각각 매칭시키고, 중요 문장별로 매칭된 핵심어의 배열 순서대로 배열된 적어도 하나의 핵심 단어열을 구성한다.

- <7> 또한, 핵심 단어열 구성부는 각 핵심어의 가중치 정보 및 웹 문서에 포함된 문장의 위치 중 적어도 하나에 기초하여 문장의 중요도 정보를 생성하고 중요도가 소정의 임계값 이상인 문장을 중요 문장으로 선택할 수 있다.
- <8> 핵심 단어열 구성부가 적어도 하나의 핵심 단어열 중에서 가장 길이가 길게 매칭되어 배열되는 핵심 단어열을 선택한 경우, 정보 제공부는 선택된 핵심 단어열을 우선적으로 이용하여 웹 문서 요약 정보를 제공할 수 있다.
- <9> 정보 제공부는, 핵심 단어열이 매칭된 문장들이 웹 문서에 나타난 순서대로 포함되도록 웹 문서 요약 정보를 제공할 수 있다.
- <10> 본 발명의 일 실시예에 따르면, 정보 제공부는 사용자의 입력에 따라 적어도 하나의 핵심 단어열을 이용하여 생성된 웹 문서 요약 정보, 웹 문서의 문장 중 핵심 단어열이 포함된 중요 문장 및 원래의 웹 문서를 순차적으로 제공할 수 있다.
- <11> 본 발명의 다른 양상에 따르면 웹 문서에 포함된 형태소를 분석하는 단계; 분석 결과에 기초하여 적어도 하나의 핵심어를 추출하는 단계; 웹 문서에 포함된 각 문장별 형태소와 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성하는 단계; 및 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공하는 단계를 포함하는 웹 문서 제공 방법이 제공된다.

발명의 실시를 위한 구체적인 내용

- <12> 이하, 첨부된 도면을 참조하여 본 발명의 일 실시예를 상세하게 설명한다. 본 발명을 설명함에 있어 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 또한, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- <13> 도 1은 웹 문서의 일례를 나타내는 도면이다. 도 1에 도시된 웹 페이지(100) 중에서 웹 문서(110)가 사용자 단말 장치에 제공될 수 있다. 도 1에서 웹 문서(120)는 웹 문서(110)를 xml 파일 형태로 나타낸 것이다.
- <14> 도 1에 도시된 바와 같은 웹 문서인 기사가 소형 디스플레이 화면을 가지는 사용자 단말 장치에 제공될 때에는 소형 디스플레이 화면에 표시되는 기사의 내용이 한정된다. 통상적으로 기사가 소형 디스플레이의 화면 크기에 제공될 때에는, 기사가 여러 개의 페이지로 분할되어 제공되며 화면에 '다음 페이지로' 아이콘이 제공된다. 이때, 사용자는 '다음 페이지로' 아이콘을 선택함으로써 기사의 나머지의 내용을 읽을 수 있게 된다. 따라서, 사용자는 기사의 내용을 한 번에 파악하기 어려우며, 내용을 파악하기 위해서는 여러 번의 사용자 입력 조작이 필요한 문제점이 있다.
- <15> 도 2는 본 발명의 일 실시예에 따른 웹 문서 요약 정보를 제공하는 사용자 단말 장치의 구성을 나타내는 블록도이다.
- <16> 본 발명의 일 실시예에 따른 웹 문서 요약 정보를 제공하는 사용자 단말 장치(200)는 제어부(210), 통신부(220), 사용자 입력부(230), 출력부(240) 및 저장부(250)를 포함한다. 사용자 단말 장치(200)는 웹 브라우저가 가능한 단말 장치로서 여러 가지 형태로 구현될 수 있다. 예를 들어, 사용자 단말 장치(200)는 대형 디스플레이를 가지는 IPTV, 소형 디스플레이를 가지는 UMPC(Ultra-Mobile Person Computers), 또는 휴대폰일 수 있다.
- <17> 제어부(210)는 사용자 단말 장치(200)의 구성요소 전반의 동작을 제어한다. 통신부(220)는 다수의 웹 서버(도시되지 않음)로부터 유무선 통신망을 통하여 웹 페이지를 수신한다. 사용자 입력부(230)는 키패드, 마우스, 터치 패드 또는 터치 스크린과 같은 다양한 사용자 조작 명령을 입력받는 입력 장치이다. 출력부(240)는 영상 및 오디오 출력 장치를 포함하여 구성될 수 있다. 저장부(250)는 사용자 단말 장치(200)의 동작에 필요한 애플리케이션이나 데이터를 저장하고 다양한 멀티미디어 데이터가 저장될 수 있다.
- <18> 제어부(210)는 수신된 웹 페이지를 분석하여 실시간으로 웹 페이지의 요약 정보를 생성하여 출력부(240)로 제공한다. 이를 위하여, 제어부(210)는 웹 페이지를 분석하여 웹 문서에 포함된 핵심어를 추출하고, 추출된 핵심어와 웹 문서의 문장 간에 매칭을 수행하여, 각 문장에 나타난 핵심 단어열을 구성하여 구성된 핵심 단어열을 요약 정보로서 제공한다. 제어부(210)의 상세 구성은 도 3을 참조하여 설명한다.

- <19> 도 3은 도 2의 웹 문서 요약 정보를 제공하는 사용자 단말 장치에 포함된 제어부의 구성을 상세하게 나타내는 블록도이다.
- <20> 제어부(210)는 전처리부(212), 핵심어 추출부(214), 핵심 단어열 구성부(216) 및 정보 제공부(218)를 포함한다.
- <21> 전처리부(212)는 웹 문서를 분석하여 텍스트를 추출한다. 예를 들어, 웹 문서는 콘텐츠, 스펙트럼, 레이아웃, 프리젠테이션 스타일을 독립적인 속성으로 포함하는 컴포넌트 기반 문서 트리(component-based document tree)로 분석되어 텍스트가 추출된다.
- <22> 또한, 전처리부(212)는 웹 문서에 포함된 형태소들을 분석한다. 전처리부(212)는 각 형태소의 품사, 위치 및 문서에서의 출현 빈도수를 포함하는 정보를 생성한다. 전처리부(212)는 각 형태소를 품사, 위치 및 출현 빈도수 등을 이용하여 인덱싱할 수 있다.
- <23> 핵심어 추출부(214)는 전처리부(212)의 분석 결과에 기초하여 적어도 하나의 핵심어를 추출한다. 핵심어 추출부(214)는 웹 문서에 포함된 각 형태소의 위치 및 출현 빈도 중 적어도 하나에 기초하여 각 형태소의 가중치 정보를 생성할 수 있다. 그런 다음, 핵심어 추출부(214)는 가중치 정보가 소정의 임계값 이상인 형태소를 핵심어로 추출할 수 있다.
- <24> 본 발명의 일 실시예에 따르면 핵심어 추출부(214)에서 타이틀 태그(ex. <TITLE>)에서 정의되거나 강조를 위한 글씨체를 위한 태그(ex.)에서 정의된 형태소의 경우에는 다른 위치에 있는 형태소보다 큰 가중치가 부여되도록 할 수 있다. 또한, 형태소의 품사 중 명사에 대해서만 가중치가 부여되도록 할 수 있으며, 출현 빈도수가 클수록 큰 가중치가 부여되도록 할 수 있다.
- <25> 본 발명의 다른 실시예에 따르면, 1차적으로 타이틀에 위치해 있는 명사를 핵심어로 추출하고, 타이틀이 없는 경우에는 문서 내 첫 번째로 위치한 문장에 있는 명사를 핵심어로 추출할 수 있다. 그런 다음, 1차적으로 추출된 핵심어를 기반으로 2차 확장 핵심어를 추출할 수 있다.
- <26> 예를 들어, 문서내의 문장을 1차 핵심어를 포함하는 문장(연관 문장이라 함)과 1차 핵심어를 포함하지 않는 문장(비연관 문장이라 함)으로 나누고, 연관 문장에 속한 형태소들에 대해 더 큰 가중치가 부여되도록 한 결과에 따라 2차 핵심어가 추출될 수 있다.
- <27> 도 1에 도시되어 있는 웹 문서를 예를 들면, 핵심어는 '탈레반', '협상', '인질', '정부', '아프간', '살해', '무장', '세력', '석방', '주장', '추가', '한국인', '희생자'로 추출될 수 있다.
- <28> 핵심 단어열 구성부(216)는 웹 문서에 포함된 각 문장별 형태소와 추출된 적어도 하나의 핵심어를 매칭하여, 매칭된 핵심어가 각 문장에서의 매칭 순서대로 배열된 핵심 단어열을 구성한다.
- <29> 이를 위해 본 발명의 일 실시예에 따르면 핵심 단어열 구성부(216)는 웹 문서에 포함된 문장 중에서 적어도 하나의 핵심어와 관련된 적어도 하나의 중요 문장을 선택한다. 핵심 단어열 구성부(216)는 각 핵심어의 가중치 정보 및 웹 문서에 포함된 문장의 위치 중 적어도 하나에 기초하여 문장의 중요도 정보를 생성하고 중요도가 소정의 임계값 이상인 문장을 중요 문장으로 선택할 수 있다. 예를 들어, 타이틀 문구 또는 문서의 첫 번째 내지 두 번째로 위치하는 문장이 중요 문장으로 선택될 수 있다.
- <30> 도 1에 도시된 웹 문서를 예를 들면, "탈레반 무장 세력이 한국인 인질을 추가로 살해하겠다고 위협, 긴장이 고조되는 가운데 인질 석방협상에 별다른 진척이 없는 것으로 전해졌다."가 중요 문장으로 선택될 수 있다.
- <31> 핵심 단어열 구성부(216)에서 중요 문장을 선택할 때 이용되는 핵심어는 전술한 2차 확장 핵심어일 수 있다. 또한, 핵심 단어열 구성부(216)는 핵심 단어열이 중복되어 구성되는 경우 MMR(Maximal Marginal Relevance) 기법을 이용하여 중복된 핵심 단어열을 제거할 수 있다.
- <32> 그런 다음, 핵심 단어열 구성부(216)는 중요 문장별로 중요 문장에 포함된 형태소와 적어도 하나의 핵심어를 각각 매칭시키고, 중요 문장별로 매칭된 핵심어의 배열 순서대로 배열된 적어도 하나의 핵심 단어열을 구성한다. 도 1에 도시된 웹 문서를 예를 들면, "탈레반 무장 세력 한국인 인질 추가 살해 인질 석방협상"이 핵심 단어열로 선택될 수 있다.
- <33> 정보 제공부(218)는 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공한다. 웹 문서 요약 정보는 도 2의 출력부(240)를 통해 표시되어 사용자에게 제공된다.
- <34> 본 발명의 일 실시예에 따르면 핵심 단어열 구성부(214)는 적어도 하나의 핵심 단어열 중에서 가장 길이가 길게

매칭되어 배열되는 핵심 단어열을 선택한다. 예를 들어, 핵심 단어열 구성부(214)는 단어열 간의 연결도를 측정하여 최장 길이 매핑 방식을 통해 핵심어를 선택한 다음 핵심 단어열을 구성할 수 있다. 그러면, 정보 제공부(218)는 선택된 핵심 단어열을 우선적으로 이용하여 웹 문서 요약 정보를 제공한다.

- <35> 정보 제공부(218)는 핵심 단어열이 매칭된 문장들이 웹 문서에 나타난 순서대로 핵심 단어열이 배열되도록 웹 문서 요약 정보를 제공할 수 있다. 이와 같이 핵심 단어열들이 웹 문서에서 나타난 순서대로 제공됨으로써 사용자는 웹 문서 전체의 문맥을 한 눈에 파악할 수 있는 장점이 있다.
- <36> 또한, 핵심 단어열 구성부(214)에서 적어도 하나의 핵심 단어열이 구성된 경우, 핵심 단어열의 중요도를 결정하여 중요도가 높게 결정된 핵심 단어열부터 웹 문서 요약 정보로 포함되어 제공될 수 있다. 핵심 단어열의 중요도는 예를 들어, 핵심 단어열을 구성하는 각 핵심어의 출현 횟수에 따른 가중치와, 핵심 단어열 자체의 출현 횟수에 따른 가중치의 합으로 계산될 수 있다.
- <37> 또한, 정보 제공부(218)는 사용자의 입력에 따라 적어도 하나의 핵심 단어열을 이용하여 생성된 웹 문서 요약 정보, 웹 문서의 문장 중 핵심 단어열이 포함된 중요 문장 및 원래의 웹 문서를 순차적으로 제공할 수 있다. 이와 같이 다양한 방식으로 웹 문서를 제공함으로써 사용자는 웹 문서를 우선 개략적으로 파악한 다음 원하는 문서의 경우에만 문서에서의 중요 문장이나 원래의 웹 문서를 제공받을 수 있게 된다.
- <38> 그러나, 웹 문서에 대한 요약 정보를 제공하기 위한 방법은 전술한 실시예에 한정되는 것이 아니고, 사용자 단말 장치의 디바이스 환경에 따라 적응적으로 구현될 수 있다.
- <39> 도 4a 내지 도 4c는 본 발명의 일 실시예에 따른 웹 문서 제공 방법에 따라 웹 문서가 제공되는 화면을 나타내는 도면이다.
- <40> 도 4a는 사용자가 도 1에 도시된 웹 문서를 검색했을 때, 종래의 웹 문서 제공 방식과 달리 핵심 단어열로 구성된 웹 문서 요약 정보가 제공되고 있는 화면을 도시한다. 따라서, 사용자에게 웹 문서의 핵심어들이 한 문장에서 연속되어 출현한 의미 단위로 묶여져 제공되므로, 사용자는 웹 문서의 내용을 한 번에 파악할 수 있다.
- <41> 도 4a에 도시된 화면에서 다른 형태의 웹 문서를 제공받기 위한 사용자 입력이 수신되면, 도 4b에 도시된 바와 같이 핵심 단어열을 포함하는 중요 문장이 제공될 수 있다. 그러면, 사용자는 웹 문서에서 중요한 의미를 가지는 중요 문장을 접할 수 있어 웹 문서의 내용을 좀 더 자세히 파악할 수 있게 된다.
- <42> 도 4b에 도시된 화면에서 다른 형태의 웹 문서를 제공받기 위한 사용자 입력이 수신되면, 도 4c에 도시된 바와 같이 원래의 웹 문서가 타이틀 부분부터 제공될 수 있다.
- <43> 따라서, 사용자의 입력에 따라 적어도 하나의 핵심 단어열을 이용하여 생성된 웹 문서 요약 정보, 웹 문서의 문장 중 핵심 단어열이 포함된 중요 문장, 원래의 웹 문서가 순차적으로 제공될 수 있다.
- <44> 도 5는 본 발명의 일 실시예에 따른 웹 문서 제공 방법을 나타내는 순서도이다.
- <45> 전처리부(212)는 웹 문서에 포함된 형태소를 분석한다(S 510).
- <46> 핵심어 추출부(214)는 전처리부(212)의 분석 결과에 기초하여 적어도 하나의 핵심어를 추출한다(S 520). 웹 문서에 포함된 각각의 형태소의 위치 및 출현 빈도 중 적어도 하나에 기초하여 각 형태소의 가중치 정보를 생성하고, 가중치 정보가 소정의 임계값 이상인 형태소를 핵심어로 추출할 수 있다.
- <47> 핵심 단어열 구성부(216)는 웹 문서에 포함된 각 문장별 형태소와 상기 핵심어를 매칭하여 각 문장에서 매칭된 순서대로 배열된 핵심 단어열을 구성한다(S 530).
- <48> 정보 제공부(218)는 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공한다(S 540). 핵심 단어열이 2개 이상 표시될 때에는 핵심어가 매칭된 문장들이 웹 문서에 나타난 순서대로 배열되어 제공될 수 있다.
- <49> 본 발명의 일 실시예에 따르면, 핵심 단어열을 선택할 때(S 530), 적어도 하나의 핵심 단어열 중에서 가장 길이가 길게 매칭되어 배열되는 핵심 단어열을 선택하고, 웹 문서 요약 정보를 제공하는 단계(S 540)에서 선택된 핵심 단어열을 우선적으로 이용하여 웹 문서 요약 정보가 구성될 수 있다.
- <50> 도 6은 본 발명의 다른 실시예에 따른 웹 문서 제공 방법을 나타내는 순서도이다.
- <51> 전처리부(212)는 웹 문서에 포함된 형태소를 분석한다(S 610).
- <52> 핵심어 추출부(214)는 전처리부(212)의 분석 결과에 기초하여 적어도 하나의 핵심어를 추출한다(S 620).

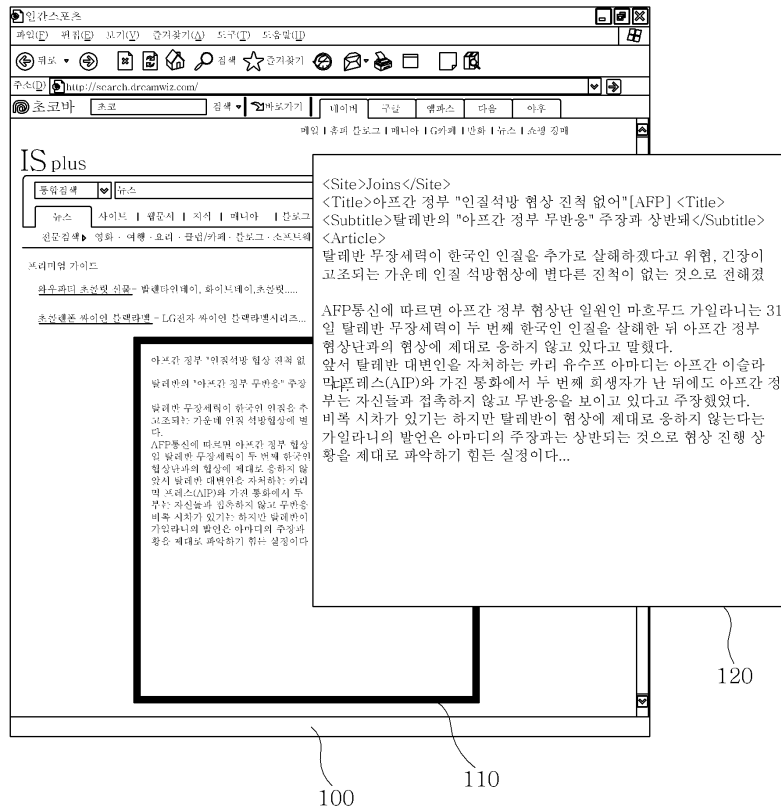
- <53> 핵심 단어열 구성부(216)는 웹 문서에 포함된 문장 중에서 적어도 하나의 핵심어를 포함하는 적어도 하나의 중요 문장을 선택한다(S 630). 이때, 각 핵심어의 가중치 정보 및 웹 문서에 포함된 문장의 위치 중 적어도 하나에 기초하여 각 문장에 대하여 문장의 중요도 정보를 생성하고 중요도가 소정의 임계값 이상인 문장을 중요 문장으로 선택할 수 있다.
- <54> 핵심 단어열 구성부(216)는 중요 문장별로 중요 문장에 포함된 형태소와 적어도 하나의 핵심어를 각각 매칭시킨다(S 640). 그런 다음, 핵심 단어 단어열 구성부(216)는 핵심어들이 중요 문장별로 매칭된 핵심어의 배열 순서대로 배열되도록 하여 적어도 하나의 핵심 단어열을 구성한다(S 650).
- <55> 정보 제공부(218)는 핵심 단어열을 이용하여 웹 문서 요약 정보를 제공한다(S 660).
- <56> 다른 형태로 웹 문서를 제공받기를 요청하는 사용자 입력 신호가 수신되면(S 670), 단계 S 630에서 선택된 중요 문장을 제공하거나 원래의 웹 문서를 제공할 수 있다(S 680).
- <57> 전술한 바와 같이 본 발명에 따르면, 사용자가 익숙하지 않은 디바이스 환경, 특히 소형 디스플레이 및 한정된 사용자 인터페이스 수단을 가지는 사용자 단말 장치를 이용하여 웹 문서를 검색할 때, 핵심어들의 의미 단위로 묶여진 핵심 단어열을 요약 정보로 제공함으로써 사용자가 웹 문서의 내용을 신속하고 정확하게 파악하는 것을 가능하게 한다.
- <58> 또한, 본 발명에 따르면, 웹 문서를 분석할 때 형태소 분석만을 이용하여 문서를 요약할 수 있으므로, 2가지 이상의 분석 방식(예를 들어, 형태소 분석 및 구문 분석)을 함께 이용하는 것보다 데이터 처리 성능이 향상되어 웹 문서의 실시간 요약을 가능하게 한다.
- <59> 이상의 설명은 본 발명의 일 실시예에 불과할 뿐, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 본질적 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현할 수 있을 것이다. 따라서, 본 발명의 범위는 전술한 실시예에 한정되지 않고 특히 청구범위에 기재된 내용과 동등한 범위 내에 있는 다양한 실시 형태가 포함되도록 해석되어야 할 것이다.

도면의 간단한 설명

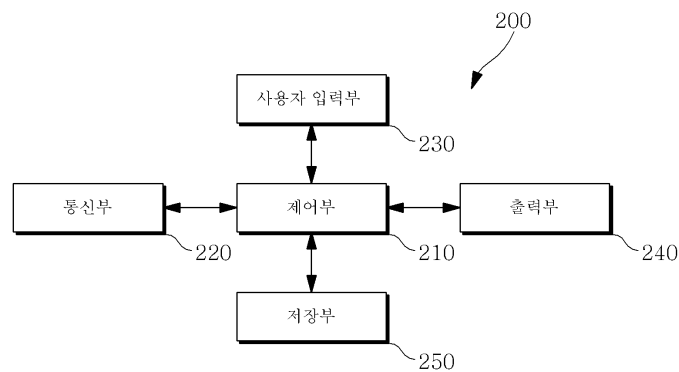
- <60> 도 1은 웹 문서의 일례를 나타내는 도면이고,
- <61> 도 2는 본 발명의 일 실시예에 따른 웹 문서 요약 정보를 제공하는 사용자 단말 장치의 구성을 나타내는 블록도이고,
- <62> 도 3은 도 2의 웹 문서 요약 정보를 제공하는 사용자 단말 장치에 포함된 제어부의 구성을 상세하게 나타내는 블록도이고,
- <63> 도 4a 내지 도 4c는 본 발명의 일 실시예에 따른 웹 문서 제공 방법에 따라 웹 문서가 제공되는 화면을 나타내는 도면이고,
- <64> 도 5는 본 발명의 일 실시예에 따른 웹 문서 제공 방법을 나타내는 순서도이고,
- <65> 도 6은 본 발명의 다른 실시예에 따른 웹 문서 제공 방법을 나타내는 순서도이다.

도면

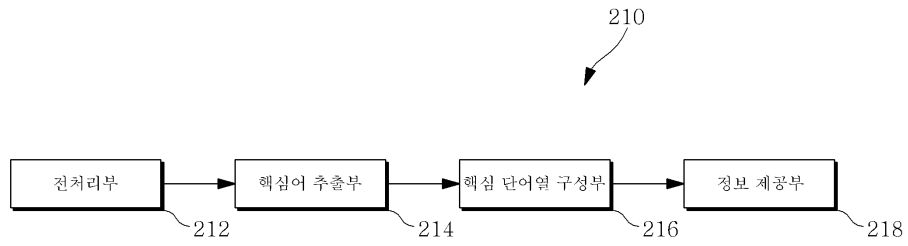
도면1



도면2



도면3



도면4a

탈레반 무장 세력
한국인 인질 추가 살해
인질 석방 협상

도면4b

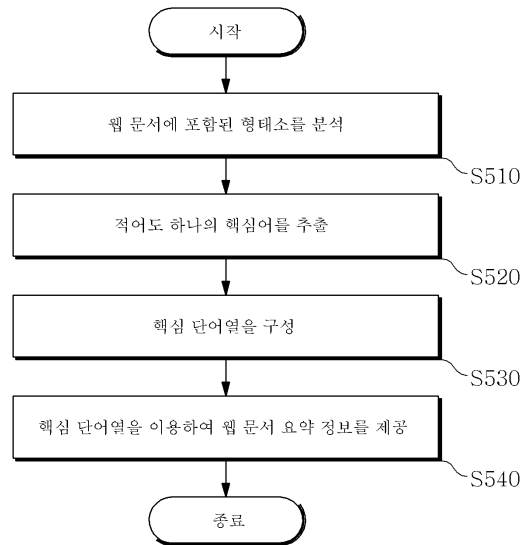
탈레반 무장 세력이 한국인 인질을 추가로 살해하겠다고 위협, 긴장이 고조되는 가운데 인질 석방 협상에 별다른 진척이 없는 것으로 전해졌다.

도면4c

아프간 정부 '인질협상' 진척 없어-탈레반의 '아프간 정부 무반응'과 상반돼

탈레반 무장 세력이 한국인...

도면5



도면6

