



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년06월01일
(11) 등록번호 10-1038087
(24) 등록일자 2011년05월24일

(51) Int. Cl.

H04B 1/40 (2006.01)

(21) 출원번호 10-2009-0116877

(22) 출원일자 2009년11월30일

심사청구일자 2009년11월30일

(56) 선행기술조사문헌

KR1020050022399 A

KR1020090002063 A

KR1020070065682 A

KR1020050073704 A

(73) 특허권자

전남대학교산학협력단

광주 북구 용봉동 300

한국과학기술원

대전 유성구 구성동 373-1

(72) 발명자

맹성현

대전광역시 유성구 용산동 우림아파트 1205-301

장혜주

서울특별시 광진구 자양동 553-403

(뒷면에 계속)

(74) 대리인

최태창

전체 청구항 수 : 총 12 항

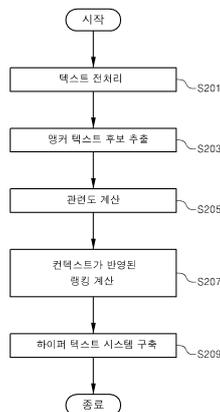
심사관 : 김기호

(54) 이동통신 단말기에서의 하이퍼텍스트 제공 방법

(57) 요약

본 발명은 이동통신 단말기에서 정보를 관리하고 검색하기 위한 방법과 그 방법을 구축하기 위한 방법론에 관한 것으로서, 본 발명은 텍스트가 제공되면, 제공된 텍스트를 전처리하는 단계, 상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계, 상기 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 단계, 상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계 및 상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계를 포함한다. 본 발명에 의하면, 하이퍼텍스트를 이용하여 정보 관리와 검색 기능을 제공하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법을 제공함으로써, 이동통신 단말기 내의 정보를 하이퍼텍스트로 구축하여 사용자의 편의에 기여하는 효과가 있다.

대표도 - 도2



(72) 발명자

신욱현

경상북도 안동시 옥동 3주공아파트 314-402

김성찬

전라북도 전주시 완산구 삼천동 신일 강변아파트
101동 107호

이철우

광주광역시 북구 용봉동 전남대학교 공과대학 7호
관 604호

이 발명을 지원한 국가연구개발사업

과제고유번호 C1090-0903-0008

부처명 지식경제부/정보통신산업진흥원

연구관리전문기관

연구사업명 대학 IT 연구센터 육성.지원사업

연구과제명 차세대 휴대폰 지능형 사용자 인터페이스 플랫폼 기술개발

기여율

주관기관 전남대학교산학협력단

연구기간 2009.03.01~2009.12.31

특허청구의 범위

청구항 1

텍스트가 제공되면, 제공된 텍스트를 전처리하는 단계;
상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계;
상기 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 단계;
상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계; 및
상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계를 포함하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 2

제1항에 있어서,
상기 제공된 텍스트를 전처리하는 단계는,
상기 제공된 텍스트에 대하여 품사 태깅(tagging), 스톱워드(stopword) 제거, 개체명 인식을 포함하는 처리를 하여 파싱(parsing)하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 3

제1항에 있어서,
상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계는,
상기 전처리된 결과를 하나 이상의 카테고리로 분류하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 4

제3항에 있어서,
상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계는,
상기 전처리된 결과를 사람 또는 그룹 이름, 장소 이름, 활동 또는 이벤트 이름, 시간 표현, 주제의 5가지 카테고리로 분류하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 5

제4항에 있어서,
개체명 인식기를 이용하여 어휘 사전 기반의 규칙 기반 분류기에 의해 상기 5가지 카테고리로 분류하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 6

제1항에 있어서,
상기 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 단계는,
상기 앵커 텍스트 후보에 속한 어휘를 이동통신 단말기에 저장된 기존 정보와 비교하여 관련도를 계산하는 것을

특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 7

제1항에 있어서,

상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계는,

시간의 컨텍스트 정보가 반영된 랭킹을 계산하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 8

제1항에 있어서,

상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계는,

장소의 컨텍스트 정보가 반영된 랭킹을 계산하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 9

제1항에 있어서,

상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계는,

상기 제공된 텍스트 중에서 관련정보가 있는 단어에 하이퍼링크 기능을 제공하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 10

제9항에 있어서,

상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계는,

상기 제공된 텍스트 중에서 2 이상의 관련정보가 존재하는 단어가 있으면, 해당 단어는 상기 관련정보 리스트를 디스플레이하는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 11

제10항에 있어서,

상기 관련정보 리스트는 상기 해당 단어의 관련도에 따른 우선순위에 의해 순번이 정해지는 것을 특징으로 하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법.

청구항 12

제9항에 있어서,

상기 관련 정보는 단문 메시지 서비스(SMS, Short message service)에 포함된 정보, 메일 서비스에 포함된 정보, 스케줄 서비스에 포함된 정보 및 사진 앨범 서비스에 포함된 정보를 포함하는 것임을 특징으로 하는 이동

통신 단말기에서의 하이퍼텍스트 제공 방법.

명세서

발명의 상세한 설명

기술분야

[0001] 본 발명은 이동통신 단말기에서 정보를 관리하고 검색하기 위한 방법과 그 방법을 구축하기 위한 방법론에 관한 것으로, 보다 상세하게는 이동통신 단말기에 저장되거나 새로 수신한 정보를 분석하여 관련 정보에 관한 하이퍼텍스트 망을 구축하는 시스템에 관한 것이다.

배경기술

[0002] 최근 모바일 장치들이 날로 진화하면서 점점 많은 능력을 갖추게 되고 있다. 모바일 장치는 모바일 폰, 핸드폰, 스마트 폰, PDA를 포함하는 이동통신 단말기를 의미한다. 예를 들어, 최근 출시되고 있는 스마트폰은 거의 PC의 경우처럼 소프트웨어 개발자들에게 플랫폼을 제공할 수 있는 완전한 운영체제를 탑재하고 있다. 운영체제를 탑재하지 않은 이동통신 단말기의 경우에도 통화와 관련된 기본 기능과 카메라 등과 더불어 이메일, 인터넷, 음악, GPS 등의 기능을 갖추고 있다. 이러한 추세에 따라, 현재 이동통신 단말기는 기존의 이동통신 단말기처럼 전화번호 또는 문자 메시지만 저장하고 있는 게 아니라, 사진, 음악, 동영상, 일정, 메모, 이메일 등의 다양한 정보를 저장하게 되었다.

[0003] 이렇게 이동통신 단말기의 저장 장소와 기능이 다양해짐에 따라 이동통신 단말기에서 정보를 검색하는 것의 중요성이 매우 부각되고 있다. 언제 어디서나 사용하는 이동통신 단말기는 가볍게 정보를 저장하는 수단으로 많이 사용된다. 이동통신 단말기에 저장된 정보는 사용자가 나중에 이용할지도 모르기 때문에 일정 기간 동안 삭제하지 않고 저장하고 있는 것이 일반적이다.

[0004] 이동통신 단말기에서의 정보 검색은 이동통신 단말기라는 제한된 크기에서 기인하는 제약요인과 동시에 휴대성이 용이하다는 기회요인을 모두 갖고 있는 기기이다. 즉, 이동통신 단말기의 작은 스크린은 많은 양의 정보를 한꺼번에 보여줄 수 없는 제한요인으로 작용한다. 또한, 이동통신 단말기의 불편한 입력 인터페이스는 PC와 비교하여 정보검색에 있어서의 제한 요인으로 작용한다. 그래서, 이동통신 단말기 내에서는 고급 검색 기능이 있다 해도 잘 사용되지 않고, 사용자가 입력하는 검색어가 일반PC에서 키보드를 이용할 때 입력하는 검색어에 비해 짧은 경향이 있다는 연구가 있다.

[0005] 이러한 상황에서 이동통신 단말기에서 사용자가 검색을 편리하게 수행하게 하기 위한 여러 가지 시도들이 있어 왔다. 즉, 기존의 웹 검색과 비슷한 검색창에서 질의 완성 기능을 제공한다든지, 미리 사용자들이 많이 하는 질의를 보여준다든지, 혹은 검색 인터페이스를 편리하게 만드는 연구들이 있었다. 하지만, 그 중 어떤 것도 이동통신 단말기 내에 있는 정보(사용자 정보, 사용자가 받은 정보, 사용자가 저장한 정보 등)를 분석하여 관련 정보를 제공하는 연구는 없었다. 또한, 요즘 나오는 스마트 폰의 경우 스마트 폰 내의 정보를 서비스(SMS, 이메일, 앨범 등)에 상관없이 키워드 기반으로 검색할 수 있지만, 스마트 폰 내의 정보가 방대해질수록 단순 키워드 기반 검색의 결과는 매우 많아질 것이고, 제한된 디스플레이부에 검색 결과를 모두 표현하는 것에는 한계가 있다.

발명의 내용

해결하고자하는 과제

[0006] 본 발명은 상기와 같은 종래의 이동통신 단말기 내에서의 정보 관리 및 검색의 문제점을 해결하기 위해 안출된 것으로서, 이동통신 단말기 내의 텍스트 정보를 분석하여 사용자의 검색이 예상되는 정보에 링크를 걸어 관련 정보를 미리 예측하여 보여주고 선택하게 함으로써, 사용자의 편의에 기여하고 디스플레이되는 정보량을 줄일 수 있는 하이퍼텍스트 제공 방법을 제공하는데 그 목적이 있다.

과제 해결수단

- [0007] 이와 같은 목적을 달성하기 위한 본 발명은 텍스트가 제공되면, 제공된 텍스트를 전처리하는 단계, 상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계, 상기 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 단계, 상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계 및 상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계를 포함한다.
- [0008] 상기 제공된 텍스트를 전처리하는 단계는, 상기 제공된 텍스트에 대하여 품사 태깅(tagging), 스톱워드(stopword) 제거, 개체명 인식을 포함하는 처리를 하여 파싱(parsing)하는 것일 수 있다.
- [0009] 상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계는, 상기 전처리된 결과를 하나 이상의 카테고리로 분류하는 것일 수 있다.
- [0010] 상기 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 단계는, 상기 전처리된 결과를 사람 또는 그룹 이름, 장소 이름, 활동 또는 이벤트 이름, 시간 표현, 주제의 5가지 카테고리로 분류하는 것일 수 있다. 이때, 개체명 인식기를 이용하여 어휘 사전 기반의 규칙 기반 분류기에 의해 상기 5가지 카테고리로 분류할 수 있다.
- [0011] 상기 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 단계는, 상기 앵커 텍스트 후보에 속한 어휘를 이동통신 단말기에 저장된 기존 정보와 비교하여 관련도를 계산하는 것일 수 있다.
- [0012] 상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계는, 시간의 컨텍스트 정보가 반영된 랭킹을 계산하는 것일 수 있다.
- [0013] 상기 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 단계는, 장소의 컨텍스트 정보가 반영된 랭킹을 계산하는 것일 수 있다.
- [0014] 상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계는, 상기 제공된 텍스트 중에서 관련정보가 있는 단어에 하이퍼링크 기능을 제공하는 것일 수 있다.
- [0015] 상기 랭킹을 기반으로 상기 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축하는 단계는, 상기 제공된 텍스트 중에서 2 이상의 관련정보가 존재하는 단어가 있으면, 해당 단어는 상기 관련정보 리스트를 디스플레이하는 것일 수 있다.
- [0016] 상기 관련정보 리스트는 상기 해당 단어의 관련도에 따른 우선순위에 의해 순번이 정해지는 것일 수 있다.
- [0017] 상기 관련 정보는 단문 메시지 서비스(SMS, Short message service)에 포함된 정보, 메일 서비스에 포함된 정보, 스케줄 서비스에 포함된 정보 및 사진 앨범 서비스에 포함된 정보를 포함하는 것일 수 있다.

효과

- [0018] 본 발명에 의하면, 하이퍼텍스트를 이용하여 정보 관리와 검색 기능을 제공하는 이동통신 단말기에서의 하이퍼텍스트 제공 방법을 제공함으로써, 이동통신 단말기 내의 정보를 하이퍼텍스트로 구축하여 사용자의 편의에 기여하는 효과가 있다.
- [0019] 예를 들어, 본 발명의 이동통신 단말기에서 디스플레이부에 표시되는 텍스트 중 관련 정보가 있는 텍스트에서는 관련 정보가 하나가 있을 경우 바로 하이퍼링크 기능을 수행하고, 관련 정보가 여러 개가 있을 경우에는 관련 정보 리스트가 나오게 되며 그 리스트 내 아이템을 선택하면 해당 하이퍼링크 기능을 수행한다. 이는 사용자가 별도로 검색어를 입력하기 전에 미리 관련 정보를 제공함으로써 사용자가 이동통신 단말기 내의 정보를 활용함에 있어 편의성을 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

- [0020] 이하, 첨부된 도면을 참조해서 본 발명의 실시예를 상세히 설명하면 다음과 같다. 우선 각 도면의 구성 요소들에 참조 부호를 부가함에 있어서, 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 그리고, 본 발명을 설명함에 있어서, 관련된 공지 기능

은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

- [0021] 도 1은 본 발명의 일 실시예에 따른 이동통신 단말기의 내부구조를 보여주는 블록도이다. 이동통신 단말기는 입력부(110), 표시부(120), 메모리(130), 제어부(140), 무선회로부(150), 아날로그 음성처리부(160)를 포함한다.
- [0022] 입력부(110)는 키패드, 조그셔틀, 포인트스틱, 터치스크린 등 다양한 소자가 사용될 수 있으며, 사용자가 원하는 기능을 선택하거나 정보를 입력할 때 사용된다.
- [0023] 표시부(120)는 사용자에게 여러 가지 정보를 보여주기 위한 것으로, LCD(Liquid Crystal Display)가 주로 사용된다.
- [0024] 메모리(130)는 이동통신 단말기가 동작하는데 필요한 프로그램과 데이터를 저장하고 있으며, EPROM(Erasable Programmable Read Only Memory), SRAM(Static Random Access Memory), 플래시 메모리 등 다양한 소자가 사용될 수 있다.
- [0025] 제어부(140)는 이동통신 단말기의 전체 동작을 제어하는 역할을 한다. 제어부(140)는 마이크로프로세서나 DSP(Digital Signal Processor) 등으로 이루어질 수 있다. 제어부(140)에는 또한 무선회로부(150)로부터의 기저대역신호의 처리를 위한 기저대역신호처리회로가 포함되어 있다.
- [0026] 무선회로부(150)는 무선으로 기지국과 음성 및 제어 정보를 주고받을 수 있도록 하기 위한 회로이다. 외부의 이동통신 단말기가 발송하는 호신호를 안테나를 통하여 수신하고 이를 제어부(140)에 전송하거나, 메모리(130)에 저장되어 있는 데이터를 안테나를 통해 발신자의 이동통신 단말기에 전송하는 기능을 수행한다.
- [0027] 아날로그 음성처리부(160)는 디지털 음성신호를 아날로그 음성신호로 변환하고 증폭하여 스피커(SP)로 출력하거나, 마이크(MIC)로부터의 음성신호를 증폭하고 디지털신호로 변환하는 역할을 한다.
- [0028] 이제 상술한 구조를 갖는 이동통신 단말기에서 하이퍼텍스트 시스템을 구축하고 제공하는 방법에 대하여 도면을 참조하여 상세히 설명하고자 한다.
- [0029] 도 2는 본 발명의 일 실시예에 따른 이동통신 단말기에서의 하이퍼텍스트 제공 방법을 보여주는 흐름도이다.
- [0030] 텍스트가 제공되면, 제공된 텍스트를 전처리한다(S201). 본 발명의 일 실시예에서, S201 단계에서 텍스트를 전처리하는 것은 제공된 텍스트에 대하여 품사 태깅(tagging), 스톱워드(stopword) 제거, 개체명 인식을 포함하는 처리를 하여 파싱(parsing)하는 것일 수 있다.
- [0031] 다음, 전처리된 텍스트에서 앵커 텍스트 후보를 추출한다(S203). 본 발명의 일 실시예에서, S203 단계에서 전처리된 텍스트에서 앵커 텍스트 후보를 추출하는 것은 전처리된 결과를 하나 이상의 카테고리로 분류하는 것일 수 있다. 보다 상세하게는 전처리된 결과를 사람 또는 그룹 이름, 장소 이름, 활동 또는 이벤트 이름, 시간 표현, 주제의 5가지 카테고리로 분류할 수 있다. 이때, 개체명 인식기를 이용하여 어휘 사전 기반의 규칙 기반 분류기에 의해 5가지 카테고리로 분류할 수 있다.
- [0032] 다음, 앵커 텍스트 후보에 속한 어휘의 관련도를 계산한다(S205). 본 발명의 일 실시예에서 S205 단계에서 앵커 텍스트 후보에 속한 어휘의 관련도를 계산하는 것은 앵커 텍스트 후보에 속한 어휘를 이동통신 단말기에 저장된 기존 정보와 비교하여 관련도를 계산하는 것일 수 있다.
- [0033] 다음, 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산한다(S207). 본 발명의 일 실시예에서 S207 단계에서 관련도를 이용하여 컨텍스트 정보가 반영된 랭킹을 계산하는 것은 시간 또는 장소의 컨텍스트 정보가 반영된 랭킹을 계산하는 것일 수 있다. 이에 대한 상세한 설명은 후술하기로 한다.
- [0034] 다음, 계산된 랭킹을 기반으로 제공된 텍스트에 하이퍼링크 기능을 제공하는 하이퍼텍스트 시스템을 구축한다(S209). 본 발명의 일 실시예에서 S209 단계는, 제공된 텍스트 중에서 관련정보가 있는 단어에 하이퍼링크 기능을 제공하는 것일 수 있다. 본 발명에서 제공된 텍스트 중에서 2 이상의 관련정보가 존재하는 단어가 있으면, 해당 단어는 관련정보 리스트를 디스플레이할 수 있다. 이때, 관련정보 리스트는 해당 단어의 관련도에 따른 우선순위에 의해 순번이 정해지는 것이 바람직하다.
- [0035] 본 발명의 일 실시예에서 관련 정보는 단문 메시지 서비스(SMS, Short message service)에 포함된 정보, 메일

서비스에 포함된 정보, 스케줄 서비스에 포함된 정보 및 사진 앨범 서비스에 포함된 정보를 포함할 수 있다.

[0036] 도 3은 본 발명의 일 실시예에 따른 이동통신 단말기에서의 하이퍼텍스트 시스템을 설명하기 위한 도면이다.

[0037] 도 3을 참조하면, SMS메시지, 스케줄, 메모, 사진 등의 새로운 정보 아이템이 이동통신 단말기에 수신되거나 생성되면 미리 구축된 하이퍼텍스트 시스템(102)과 연결된다. 그래서 하나 이상의 앵커 텍스트 후보를 질의로 하여 사용자에게 각각의 질의와 관련되어 있는 정보 아이템들의 랭크드 리스트를 보여주고 고를 수 있는 옵션을 제공하게 된다.

[0038] 도 3의 실시예에서는, 이동통신 단말기에 SMS 메시지를 수신한 화면예(101)가 도시되어 있다. 이 새로운 SMS 메시지가 하이퍼텍스트 시스템(102)에 연결된다. 그리고, 새로운 SMS 메시지에서 "meeting"이란 단어에 "1:Email, 2:SMS1"의 링크 리스트가 디스플레이되고, "class"란 단어에 "1:Note2, 2:SMS2"의 링크 리스트가 디스플레이된다(103).

[0039] 이제 본 발명에서 S207 단계에서 시간 또는 장소의 컨텍스트 정보가 반영된 랭킹을 계산하는 과정을 상세히 설명하기로 한다.

[0040] 예를 들어, "프로젝트 회의" 라는 단어가 여러 개의 스케줄 상의 정보 아이템과 매칭이 될 때는 "프로젝트 회의" 라는 단어와 함께 있는 "내일" 이라는 단어가 여러 개의 정보 아이템 중 관련도를 결정할 수 있는 주요 요인이 된다. 하이퍼텍스트가 생성되어야 하는지 아닌지는 질의와 텍스트 내의 어휘들의 유사도에 의해 결정된다. 이는 관련도를 이용해 tf*idf를 확장한 방법을 사용할 수 있다.

[0041] 먼저, 기존의 tf(term frequency)는 다음과 같다.

[0042] [수학식 1]

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

[0043] 여기서, $n_{i,j}$ 은 t_i 기간동안 문서 d_j 에서의 발생횟수이다.

[0045] 또한, 하나의 문서(document)를 단어의 집합으로 간주하면,

[0046] [수학식 2]

$$D = \{w_1, \dots, w_n\}$$

[0048] 이때, 특정 어휘의 출현 빈도는 다음과 같이 나타낼 수 있다.

[0049] [수학식 3]

$$n_{i,j} = \sum_{w \in D_i} f(w, t_i)$$

$$f(w, t) = \begin{cases} 1 & w = t \\ 0 & o.w \end{cases}$$

[0050] 즉, 하나의 문서 내에서의 특정 어휘 t_i 가 몇번 나오는지 확인할 수 있다.

[0052] 여기서, 위에서 계산한 관련도를 함수 f 에 적용하여 관련도가 적용된 tf 가 다음 [수학식 4]와 같이 표현될 수 있다.

[0053] [수학식 4]

$$f_s(w, t) = sim(w, t)$$

[0054] 이처럼 관련도를 적용하여 함수를 개선함으로써, 특정 어휘 뿐만 아니라 관련된 어휘까지도 반영할 수 있다.

[0056] 예를 들어, 문서가 "...symposium on...meeting schedule...at conference room..."이라고 할 때, 계수해야 하

는 단어가 ‘meeting’ 일 때, 기존의 tf 방식에서는 ‘meeting’ 이 문서에서 한 번 등장하기 때문에 tf는 1이 된다. 하지만, 본 발명에서 관련도를 적용하여 개선한 tf 방식에서는 ‘meeting’ 과 관련 있는 단어인 ‘symposium’, ‘conference’ 까지 tf에 고려하게 된다. 예를 들어, symposium과 meeting의 관련도가 0.6이고, conference와 meeting의 관련도가 0.7이라면, 개선된 방식에서의 tf는 1.0 + 0.6 + 0.7=2.3이 된다. 엄밀히 말하면 이는 프리퀀시(frequency)라고 할 수는 없지만, 텀 프리퀀시(term frequency)에 대응되는 방법이라고 할 수 있다.

[0057] 같은 방식으로 idf(inversed document frequency)를 설명하면, 기존의 idf는 다음과 같다.

[0058] [수학식 5]

[0059]
$$idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

[0060] 여기서, $|\{d : t_i \in d\}|$ 는 특정어휘를 포함하고 있는 문서의 수를 의미한다. 이는 다음 [수학식 6]과 같이 표현할 수 있다.

[0061] [수학식 6]

[0062]
$$df_i = |\{d : t_i \in d\}| = \sum_{d \in D} g(d, t_i)$$

[0063] 여기서, 함수 g는 문서가 해당 어휘를 포함하고 있으면 1이 된다. 즉,

[0064] [수학식 7]

[0065]
$$g(d, t) = \begin{cases} 1 & t \in d \\ 0 & o.w. \end{cases}$$

[0066] 함수 g에서는 문서가 해당 어휘를 한 번 포함하든지 여러 번 포함하든지 상관없이 1이다. 이는 또한 다음 [수학식 8]과 같이 나타낼 수 있다.

[0067] [수학식 8]

[0068]
$$g(d, t) = \max(f(w \in d, t))$$

[0069] 위 [수학식 8]에서, g는 하나의 문서에서 모든 어휘가 해당 어휘랑 같지 않으면 0이 된다는 것을 표현하고 있다. 관련도를 이용하여 개선된 함수 f를 이용하게 되면, 문서가 특정 어휘를 포함하고 있는지 아닌지를 나타내는 기존의 문서 프리퀀시(document frequency)는 문서 안에 특정 어휘와 얼마나 관련도가 높은 어휘가 있는가로 변할 수 있다.

[0070] 이제 시간과 장소의 컨텍스트를 활용하는 방법에 대하여 설명하도록 한다.

[0071] 시간 정보를 활용하기 위해서는 위의 확장된 tf*idf에서 나온 결과를 리오더링(reordering)하는 방법을 생각할 수 있다. 현재 문서(source document)가 멘셔닝(mentioning)하는 시간 주기(time period)와, 목표 문서(target document)가 멘셔닝(mentioning)하는 시간 주기(time period)의 전후 관계를 이용하여 결과를 그룹핑(grouping)하는 방법을 사용할 수 있다.

[0072] 여기서 현재 문서(source document)와 목표 문서(target document)가 멘셔닝(mentioning)하는 시간 주기(time period)의 예를 들면,

[0073] "Did you receive the package I' ve sent last week?"

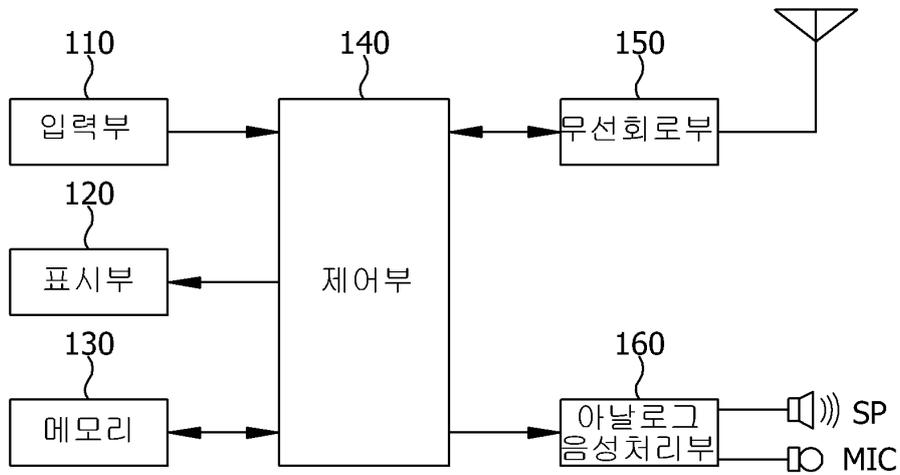
[0074] 2009-10-11 04:32"

[0075] 라는 문서가 있는 경우, 이 문서에서 언급하는 시간은 10월 11일 기준으로 지난 주간인 2009년 10월 4일부터 2009년 10월10일이 된다.

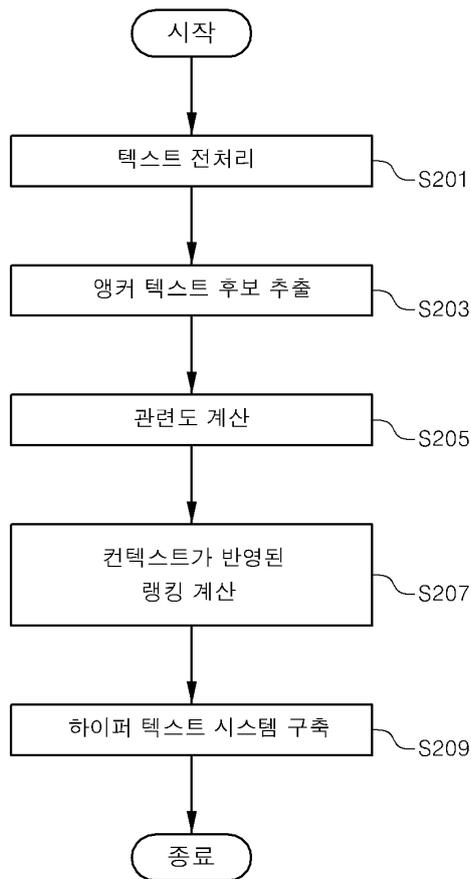
[0076] 다른 예로서,

도면

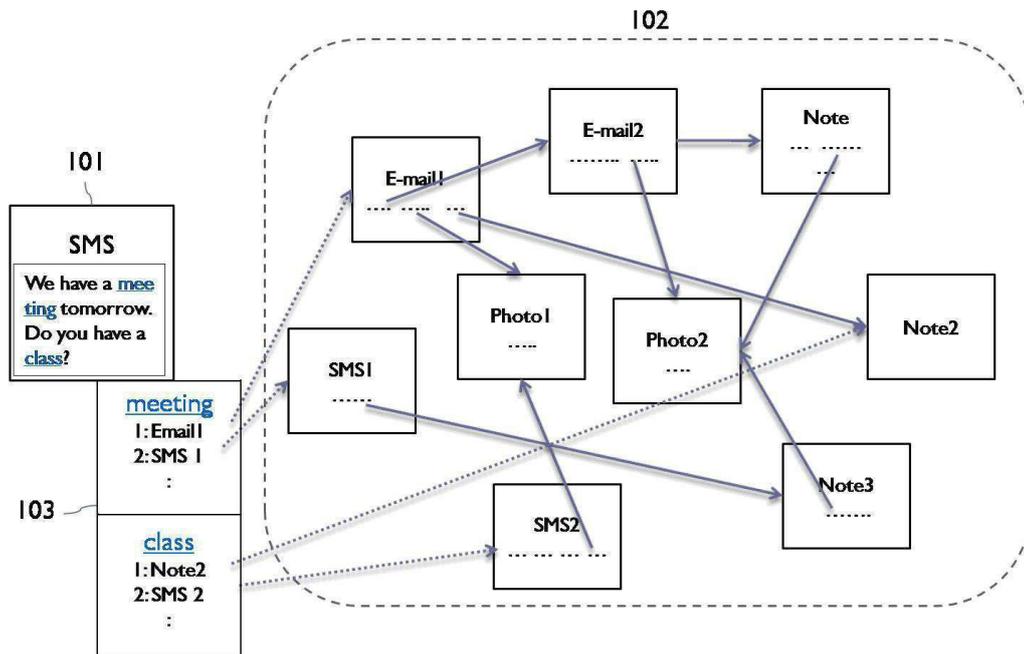
도면1



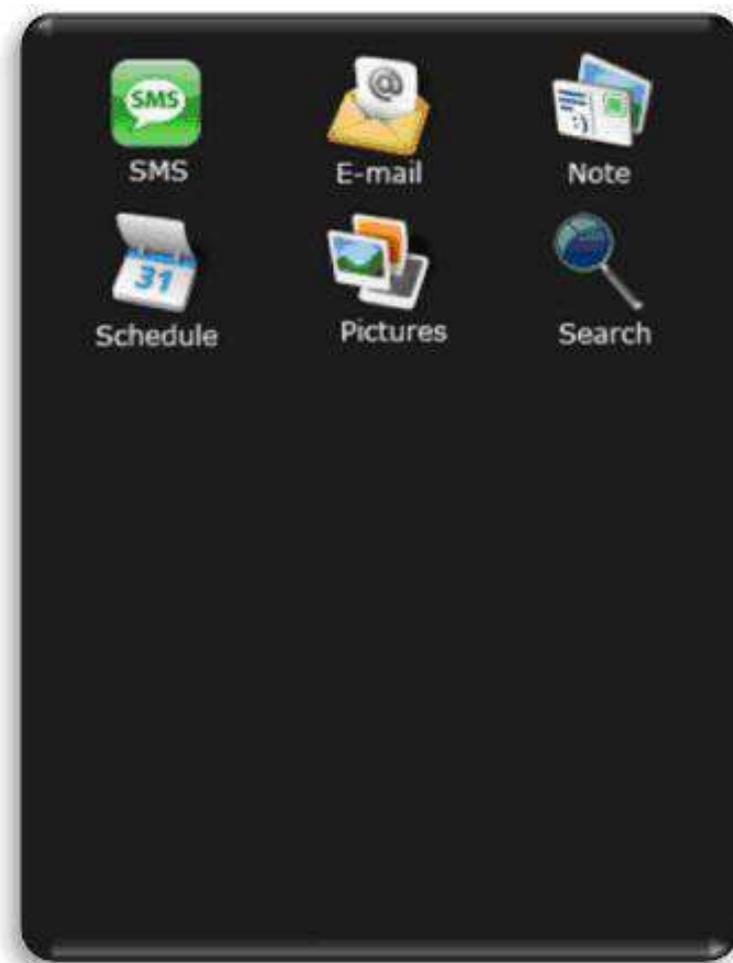
도면2



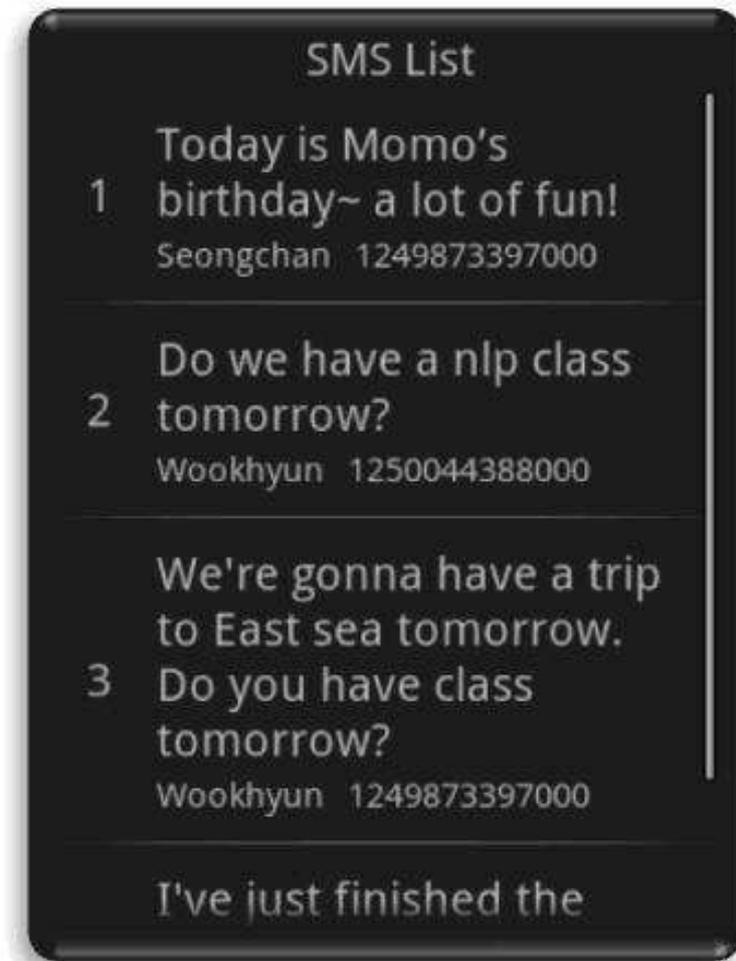
도면3



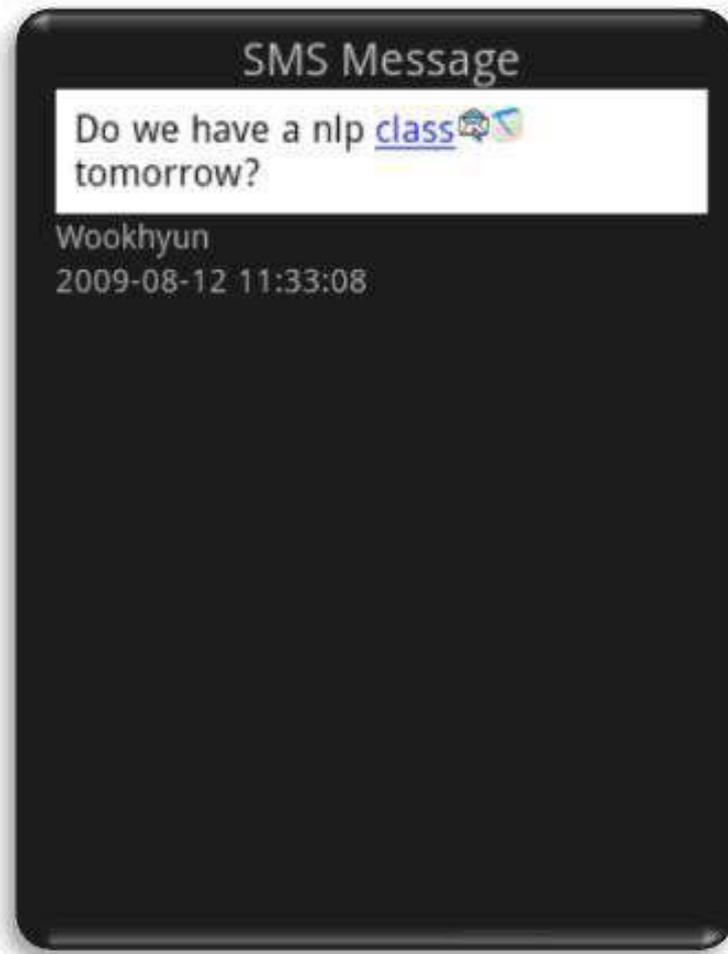
도면4a



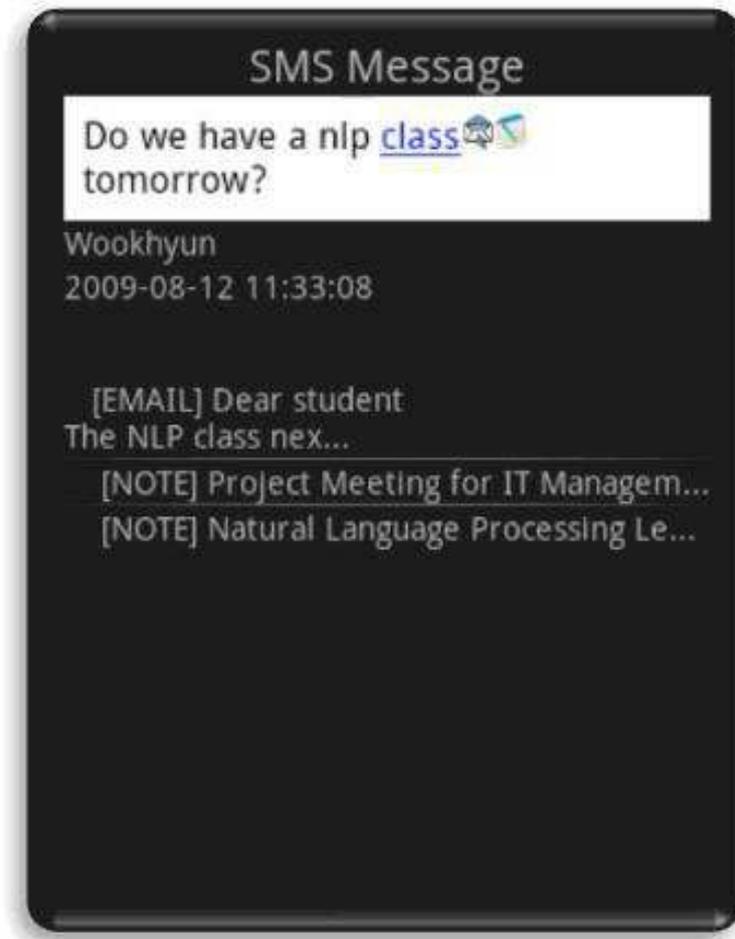
도면4b



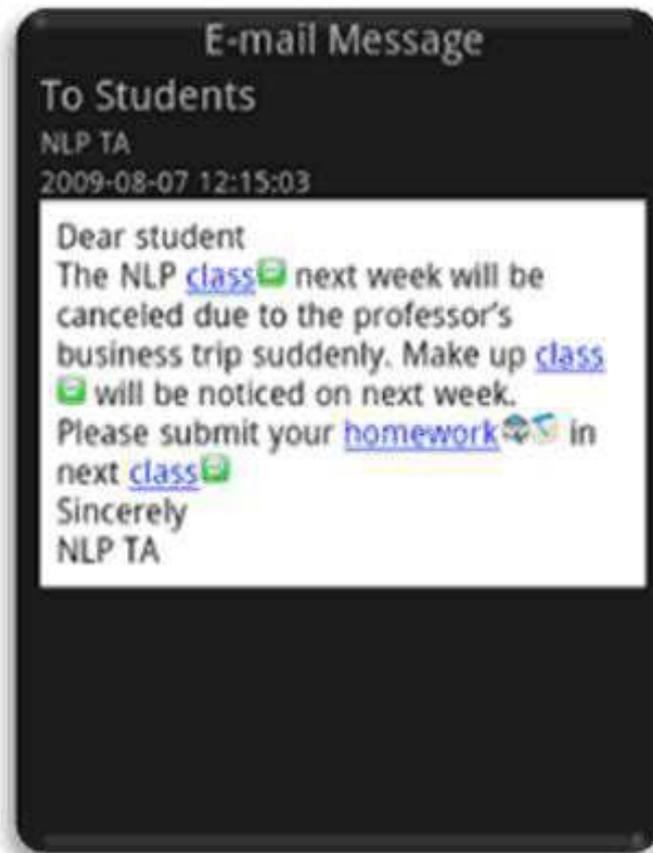
도면4c



도면4d



도면4e



도면4f

