(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2015년03월30일
(11) 등록번호 10-1506354
(24) 등록일자 2015년03월20일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
(21) 출원번호 10-2009-7015036
(22) 출원일자(국제) 2007년12월19일
심사청구일자 2012년12월18일
(85) 번역문제출일자 2009년07월17일
(65) 공개번호 10-2009-0100417
(43) 공개일자 2009년09월23일
(86) 국제출원번호 PCT/EP2007/011183
(87) 국제공개번호 WO 2008/074486
국제공개일자 2008년06월26일
(30) 우선권주장
MI2006A 002436 2006년12월19일 이탈리아(IT)
(56) 선행기술조사문헌
JP2003271670 A
JP2002236661 A
전체 청구항 수 : 총 17 항

(73) 특허권자
모올드텍 온트베르펜 비.브이.
네덜란드, 엔엘-지엑스 암스테르담, 스트라빈키란 3105
(72) 발명자
로티토, 쟝피에로, 쥬제페
이탈리아, 파비아 아이-27100, 비아 인디펜덴차, 26
테로니, 마루치아
이탈리아, 파비아 아이-27100, 비꼴로 산 세바스티아노, 8
(74) 대리인
김영철, 김 순 영

심사관 : 권현수

(54) 발명의 명칭 **웹 페이지의 분류 및 대응 콘텐트의 조직화 방법**

(57) 요 약

본 발명은 웹 페이지를 분류하고 그들의 콘텐트를 조직화하는 방법에 관한 것으로서, 복수의 자동 기록 절차의 실행을 포함하는 복수의 인터넷 주소를 기록하는 단계; 및 상기 복수의 인터넷 주소에 대응하는 적절성 값(pertinence value)을 설정(setting)하는 선택 단계를 포함한다. 상기 선택 단계는 복수의 인터넷 주소 내에서의 히트에 비례하는 값에 대응하는 확실한 적절성 값을 설정하고, 이미 결정된 임계값(threshold value)보다 큰 적절성 값을 가진 인터넷 주소를 선택한다. 상기 방법은, 임계값보다 큰 적절성 값을 가진 인터넷 주소로부터 시작되고, 하나 이상의 필수 기준에 부합하는 인터넷 주소로 인터넷 주소를 정리하는 정리 단계; 및 필수 기준에 부합하는 인터넷 주소의 서브셋(subset)을 확인하는 확인 단계를 더 포함하며, 상기 확인 단계는 인간의 행위를 포함한다.

대 표 도



특허청구의 범위

청구항 1

미리 결정된 수의 인터넷 주소의 복수의 자동 기록 절차의 실행을 포함하는 상기 복수의 인터넷 주소를 기록하는 기록 단계; 및

상기 복수의 인터넷 주소에 대응하는 적절성 값(pertinence value)을 선택 및 설정(setting)하기 위한 선택 단계를 포함하고,

상기 선택 단계는 상기 기록 단계에서의 기록에 비례하는 적절성 값에 대응하는 적절성 값을 설정하고, 이미 결정된 임계값(threshold value)보다 큰 적절성 값을 가진 인터넷 주소를 선택하는 것을 특징으로 하며,

상기 임계값보다 큰 적절성 값을 가진 상기 인터넷 주소 중에서, 하나 이상의 필수 기준(essentiality criteria)에 부합하는 인터넷 주소로 상기 인터넷 주소를 정리하는 정리 단계; 및

필수 기준에 부합하는 상기 인터넷 주소의 서브셋(subset)을 확인(validate)하기 위한 확인(validation) 단계로서, 확인 인터페이스(interface)를 실행하는 단계를 포함하는 확인 단계를 더 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 2

제1항에 있어서,

상기 기록 단계 전에, 복수의 주제 영역을 정의하고, 상기 주제 영역에 연결된 복수의 영역 키워드들을 정의하는 설정 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 3

제2항에 있어서,

상기 설정 단계는 각각의 상기 주제 영역에 대하여 복수의 레마(lemma)를 정의하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 4

제3항에 있어서,

상기 설정 단계는 하나 이상의 상기 주제 영역과 상기 레마 중 하나를 연결하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 5

제4항에 있어서,

상기 설정 단계는 상기 주제 영역의 정의, 상기 레마의 정의 및 상기 영역 키워드의 정의에 대하여 주(主) 언어를 정의하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 6

제5항에 있어서,

상기 설정 단계는 상기 이미 결정된 수의 인터넷 주소를 설정하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 7

제6항에 있어서,

상기 기록 단계는 상기 웹 페이지와 상기 레마의 연결 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 8

제7항에 있어서,

상기 연결 단계는 상기 웹 페이지의 상기 레마를 식별하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 9

제8항에 있어서,

상기 기록 단계는 상기 레마를 참조하여 제1 정보 층에 상기 인터넷 주소를 저장하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 10

제9항에 있어서,

상기 기록 단계는 상기 제1 정보 층에 상기 웹 페이지의 사본을 저장하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 11

제10항에 있어서,

상기 필수 기준은 허가된 언어의 알파벳을 포함하지 않는 복수의 웹 페이지는 제거하는 것을 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 12

제11항에 있어서,

상기 필수 기준은 동일한 도메인으로부터 나온 복수의 웹 페이지는 제거하는 것을 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 13

제12항에 있어서,

상기 정리 단계는 복수의 선택 질의(query)를 실행하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 14

제13항에 있어서,

상기 확인 단계는 상기 정리 단계에 의해 정리된 상기 인터넷 주소를 리스트(list)하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 15

제14항에 있어서,

상기 기록 단계는 하나 이상의 스파이더링(spidering) 절차를 실행하는 단계를 포함하는 것을 특징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 16

제15항에 있어서,

상기 기록 단계는 하나 이상의 메타-검색 엔진(meta-search engine) 기능을 실행하는 단계를 포함하는 것을 특

징으로 하는 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법.

청구항 17

상기 방법은 상기 복수의 인터넷 주소가 제1항 내지 제16항 중 어느 한 항의 방법에 의해 분류되고, 상기 인터넷 주소는 미리 결정된 임계값 이상의 적절성 값을 가지는 것을 특징으로 하는 검색 기준에 부합하는 복수의 인터넷 주소를 식별하기 위한 식별 프로그램을 실행하는 방법.

청구항 18

삭제

명 세 서

기 술 분 야

[0001] 본 발명은 웹 페이지를 분류하고 대응하는 형태의 콘텐트를 조직화하는 방법에 관한 것으로서, 복수의 자동 기록 처리 절차의 실행을 포함하는 복수의 인터넷 주소를 기록하는 단계 및 상기 복수의 인터넷 주소에 대응하는 적절성 값(pertinence value)을 설정(setting)하는 선택 단계를 포함한다.

배 경 기 술

[0002] 공지된 바와 같이, 인터넷 네트워크(internet network)에 연결되고 브라우저(browser)가 구비된 퍼스널 컴퓨터(personal computer)는 웹 페이지를 표시하는 것을 가능하게 한다.

[0003] 자세하게는, 퍼스털 컴퓨터란 용어는 인터넷 네트워크 연결 인터페이스가 구비된 전자 장치를 의미하는데, 원래 알려진 소프트웨어인 브라우저란 용어는 웹 페이지의 표시를 위해 이러한 전자 장치 상에 설치(install)된 공지된 소프트웨어를 뜻한다. 이러한 브라우저의 일례로 마이크로소프트(Microsoft®사의 인터넷 익스플로러 프로그램이 있다.

[0004] 좀더 자세하게는 웹 페이지들은 인터넷 네트워크와 차례로 연결되어 복수의 서버들에 저장된다. 웹 페이지의 인터넷 주소는 브라우저를 통해 지정(specify)된다. 예를 들면, 브라우저의 상부(upper portion)나 바(bar)에 지정된다. 그리고 이러한 인터넷 주소는 대응 웹 페이지의 표시, 일반적으로 브라우저의 주부(main portion)에의 표시를 위해서 제시된다.

[0005] 또한, 하나 이상의 웹 페이지의 식별 스크립트(identification script) 또는 프로그램은 사용자에 의해 설정된 검색 기준을 기초로 하여 기능한다고 공지되어 있다.

[0006] 이러한 식별 스크립트 또는 프로그램은 일반적으로 그래픽 인터페이스(graphic interface)를 통해 이용할 수 있다. 그래픽 인터페이스는 예를 들면 브라우저에 의해 그려질 수 있을 것이며, 검색 기준의 삽입부 및 식별 스크립트 또는 프로그램의 활성화 버튼을 포함한다.

[0007] 인터넷 네트워크에서 사용 가능한 막대한 양의 웹 페이지들이 주어지면, 식별 스크립트 또는 프로그램은 미리 결정된 검색 기준에 대한 많은 수의 웹 페이지들을 제공한다.

[0008] 사용자에게 이러한 웹 페이지들의 효과적인 참조를 안내하기 위해, 식별 스크립트 또는 프로그램은 각 웹 페이지에 대한 대응 중요도를 할당하는 분류 방법에 의해 지원된다. 달리 말하면, 식별 스크립트 또는 프로그램의 실행은 분류 방법에 의해 정해진(established) 중요도 값에 따라 정렬된 웹 페이지의 리스트(list)를 사용자에게 제공한다.

[0009] 보통 페이지랭크(PageRank)라는 용어로 나타내는 알려진 한 분류 방법은 랭크(rank)로 보통 나타내는 중요도 값과 웹 페이지를 결합(associate)시킨다. 자세하게는, 웹 페이지(A)의 랭크는 이 웹 페이지(A)를 참조로 하는 즉, 포인터를 포함하고 있는 부가적인 웹 페이지들(B1, ..., Bx)의 수에 의해 결정된다.

[0010] 좀더 자세하게는, 페이지랭크로 알려진 분류 방법에 따르면, 모든 웹 페이지는 다른 웹 페이지들에 대하여 연결(connection)로도 알려진 포인터를 통해 투표(vote)를 행사할 수 있다. 달리 말하면, 처음의 웹 페이지에서 유효한 연결을 통해 사용자는 하나 이상의 대상(destination) 웹 페이지로 연결 될 수 있고, 그 대상 웹 페이지의 랭크를 증가시킬 수 있다. 일반적으로, 웹 페이지의 인기가 높아질수록, 즉, 그 웹 페이지로의 연결의 수가 많을수록, 동일 페이지가 표현할 수 있는 투표의 값이 높아진다.

[0011] 그렇지만, 위에 나타낸 타입의 분류 방법은 아래의 기술로부터 명백히 드러날 다양한 한계를 가지고 있다.

[0012] 도 1을 참조하면, 인터페이스(10)는 웹 페이지의 식별 스크립트 또는 프로그램의 활성화를 위해 고려된다.

[0013] 예를 들면, 인터페이스(10)는 검색 기준의 삽입 바(insertion bar) 또는 삽입부(insertion portion)(1) 및 식별 스크립트 또는 프로그램을 활성화 시키는 활성화 버튼(2)을 포함한다.

[0014] 예로서, 도 1에서 검색 기준은 이탈리아 발명가 및 예술가 레오나르도 다빈치(Leonardo da Vinci)에 관련된 웹 페이지를 식별하기 위해 설정된다. 이러한 간단한 예에서, 검색 기준은 세 단어 "Leonardo", "da", "Vinci"의 연결에 의해 주어지고, 인터페이스(10)의 삽입 바(1)에 지정된다.

[0015] 이러한 경우, 인터페이스(10)의 주부(4)는 식별 스크립트 또는 프로그램의 검색 기준 "Leonardo da Vinci"에 기반한 실행 후의 결과를 표시한다. 검색 결과는 웹 페이지의 인터넷 주소의 리스트(3)를 포함하고, 단지 첫 번째 히트들(first hits)(3a 내지 3h)이 주부(4)에 표시된다.

[0016] 실제로, 식별된 웹 페이지의 수는 매우 많고, 약 수백만의 웹 페이지에 대한 지정된 예에서, 식별 스크립트 또는 프로그램은 표시될 다음 페이지들의 그래픽 인터페이스(10) 내부에서 웹 페이지들을 그룹화하기 위해 제공된다. 일반적으로, 1부터 n까지의 일련의 숫자(도시되지 않음)는 예컨대, 특정 페이지 및 인터넷 주소의 리스트(3)의 부분을 주부(4)에 사용자의 요청에 따라 표시하기 위해 인터페이스(10)의 하부에 이용될 수 있다.

[0017] 예를 들어, 식별 스크립트 또는 프로그램이 백 개의 인터넷 주소의 리스트(3)에 상응하는 백 개의 웹 페이지를 식별하는 것으로 가정하고, 모든 표시 페이지는 식별된 웹 페이지의 10개의 주소를 포함한다고 가정하면, 인터페이스(10)의 하부는 1에서 10의 숫자를 나타낼 것이다. 이러한 방법에서, 하부의 숫자 1의 선택은 1부터 10까지의 인터넷 주소 즉, 웹 페이지의 리스트(3)의 첫 페이지를 주부(4)에 표시하도록 한다. 자세하게는, 현재의 식별 스크립트 또는 프로그램에 있어서, 웹 페이지의 리스트(3)의 인터넷 주소는 대응 분류 방법에 의해 식별되어, 중요도의 내림차순에 따라 연속적으로 배치되고, 따라서 웹 페이지에 대응하는 리스트(3)의 페이지 1의 선택물은 이러한 분류 방법에 의해 가장 중요한 것으로 분류된다.

[0018] 숫자 2의 선택은 11부터 20까지의 인터넷 주소를 인터페이스(10)의 동일한 주부(4)에 표시하도록 한다. 이러한 인터넷 주소는 웹 페이지 리스트(3)의 두 번째 페이지에 대응한다. 하부의 숫자 10의 선택은, 항상 식별 스크립트 또는 프로그램에 의해 사용되는 분류 방법에 따라, 가장 부적절한 웹 페이지와 연관된 리스트(3)의 마지막 페이지를 참조하도록 한다.

[0019] 달리 말하면, 검색 결과의 표시의 순서는 분류 방법에 따라 얻어진다.

[0020] 좀더 자세하게는, 도 1을 참조하면, 웹 페이지(3a)는 이 도 1에서 예시로 사용된 구글(Google® 프로그램에서 사용되는 분류 방법에 의해 식별되는 모든 웹 페이지 중에서 가장 높은 중요도를 갖는 것이고, 발명자 및 예술가 레오나르도 다빈치의 정의를 제공하는 위키피디아(Wikipedia)라는 인터넷 서비스를 참조한다.

[0021] 웹 페이지(3b)는 발명자 및 예술가 레오나르도 다빈치에 의해 인류에 남겨진 유산을 기술하는 웹 페이지를 참조하고; 웹 페이지(3c)는 국립 과학 기술 박물관(National Museum of Science and Technology)을 참조하며; 웹 페이지(3e)는 로마 공항 "레오나르도 다빈치(Leonardo da Vinci)"를 참조한다.

[0022] 중요도의 순서로 첫 번째 다섯 개의 히트를 표시한 구글 프로그램의 상기 리스트된 결과로부터, 페이지랭크에 기반한 분류 방법은 아래의 한계를 가지고 있다는 것을 추론할 수 있다.

[0023] 식별 스크립트 또는 프로그램에 의해 제공되는 웹 페이지의 숫자는 중복된다(redundant). 이 경우, 첫 페이지 첫 번째 여덟 개의 결과에 있어서, 3개의 웹 페이지, 특히 웹 페이지들(3d, 3f 및 3h)은 중복된다. 즉, 그 페이지들은 이미 표시된 동일한 사이트 내의 인터넷 주소, 각각 인터넷 주소(3c, 3e 및 3g)과 반복된다. 또한, 사용자는 일반적으로 식별 방법에 의해 제공되는 수십 개의 결과 이상은 읽지 않고, 결과는 종종 다소 포괄적인 검색 기준에 대해 얻어진다.

[0024] 비록 식별 스크립트 또는 프로그램에 의해 제공되는 어떤 웹 페이지는 사용자에 의해 검색된 정보에 관해 거의 또는 전혀 적절하지 않고, 그 웹 페이지는 지나치게 높아진 랭크를 즐길 뿐이다. 웹 페이지(3e)를 고려해보면, "레오나르도 다빈치" 공항에 관심을 가지고 있는 사용자가 검색 기준으로 "공항"을 특정하지 않고 "레오나르도 다빈치"만 삽입했다고 믿기 힘들다. 그럼에도 불구하고, 뜻밖에, 로마 공항을 참조하는 웹 페이지(3e)가 구글 프로그램의 결과의 다섯 번째 히트로서 발견된다. 웹 페이지(3e)는 정보 노이즈(information noise)에 의해 영향을 받은 전형적인 결과의 예시를 구성하고, 따라서 주어진 검색의 목적에 절절하지 않을 뿐만 아니라 쓸모 없

으므로 매우 바람직하지 못하다. 이러한 결과를 입증하는 것이 가능하다. 즉, 검색 기준 "Leonardo da Vinci"를 설정한 결과로 웹 페이지의 리스트(3)의 높은 랭크에서의 로마 공항의 주소의 존재는 현재 이용 가능한 수많은 검색 스크립트 또는 프로그램을 사용함으로써 얻어진다.

[0025] 따라서, 페이지랭크에 기반한 분류 방법은, 웹 페이지의 자동화된 조직화를 제공하는 반면, 사용자가 사용자에게 가장 적절한 결과들로 한정된, 감소된 결과의 수에 집중할 수 없도록 한다.

[0026] 알려진 분류 방법의 다른 결점은, 실행 후에 식별 스크립트 또는 프로그램은 찾은 웹 페이지의 수로서 보고를 하는데, 이 수는 사용자에게 효과적으로 유용한 식별된 웹 페이지의 실제 숫자에 대응되지 않는다는 사실에서 발생한다.

[0027] 예를 들면, 도 1a를 참조하면, 검색 기준 "Leonardo da Vinci"에 대한 결과로서, 식별 스크립트 또는 프로그램에 의해 유용한 것으로 선언된(declared) 웹 페이지의 수는 3,880,000개이지만, 첫 번째 1000개의 웹 페이지만이 효과적으로 유용하다고 실례에 의해 입증하는 것은 쉬운 일이다. 더 자세하게는, 도 1b는 사용자에게 유용한 식별된 결과는 1000개를 넘지 않는다고 보고하는, 알려진 식별 프로그램에 의해 발생된 에러 메시지(error message)를 보여준다. 따라서 사용자들이 검색 기준의 결과와 관련된, 실질적으로 무한한 양의 웹 페이지를 갖고 있다고 믿어지기 때문에 유용하다고 선언된 웹 페이지의 숫자는 하찮은 것이고, 사용자는 그가 특별히 첫 번째, 높은 랭크의 웹 페이지에서 관심 콘텐트를 찾지 못하는 한, 다수의 웹 페이지를 열어보다가 단념하게 된다.

[0028] 또한, 상기 기술된 웹 페이지 분류 방법은 검색 기준에 대한 출처의 평판(notoriety), 인식 가능성(recognisability) 및 신빙성(credibility)으로 의도되는 출처의 신뢰도(reliability)와 실질적으로 관련된 문제를 더 가지고 있다. 달리 말하면, 발명가 및 예술가 레오나르도 다빈치와 효과적으로 관련된 웹 페이지는 부정확한 정보를 담고 있더라고 높은 랭크를 즐길 수 있다. 예를 들면, 웹 페이지는 그의 출생일에 대한 잘못된 정보를 담고 있거나, 검색 기준 "Leonardo da Vinci"를 설정한 사용자가 필연적으로 관심을 가질 만한 그의 가장 유명한 작품 또는 그의 삶 및 그의 작품과 관련된 다른 중요한 조짐과 같은 중요한 콘텐트가 부족할 수 있다.

[0029] 따라서, 높은 랭크의 웹 페이지는 권위 있는 출처가 아닌 개인적인 사용자에 의해 만들어진 웹 페이지와 관련된 웹 페이지(3b)의 경우와 같이 충분한 출처 신뢰도를 가지지 않더라도 적절한 것이 될 수 있다.

[0030] 반대로, 발명가 레오나르도 다빈치와 관련된 웹 페이지는 정확히 연관되지 않고 실질적으로 무용한 복수의 정보를 포함할 수 있고, 이러한 정보는 낮은 적절성을 가진 정보의 전형적인 예로서 사용자 관심 정보의 식별을 더디게 한다. 실질적으로, 현재 이용 가능한 웹 페이지의 분류 방법은 식별 스크립트 또는 프로그램이 적절하고, 기대하는 그리고 필터링된 웹 페이지의 리스트를 결과로 갖는 검색을 수행하도록 하지 못한다. 오히려, 그들은 사용자에 의해 설정된 검색 기준에 매우 제한된 관련성을 가지는 복수의 정보의 채집자(collector)처럼 행동한다.

[0031] 이같이, 알려진 분류 방법은 식별 스크립트 또는 프로그램에 의해 제공된 모든 웹 페이지들 중에서 오로지 관심 있는 것을 실질적으로 식별하기 위한 고된 필터링 및 선택 작업을 수행하도록 사용자에게 강요하고, 이러한 작업은 상당한 중요도의 웹 페이지를 놓칠 위험을 가지고 있다.

[0032] 본 발명의 기초가 되는 기술적인 문제는 식별 스크립트 또는 프로그램이 정보 노이즈가 없고, 사용자에 의해 설정된 검색 기준에 적절하며, 웹 페이지의 반복으로 중복되지 않고, 그들의 콘텐트가 신뢰할 만한 결과를 제공하도록 하는 웹 페이지의 분류 방법을 고안하는 동시에 검색 결과를 필터링하기 위해 사용자에 의해 채용되는 시간을 줄이며, 종래 기술에 따른 웹 페이지 분류 방법에 여전히 영향을 미치는 한계를 해결하는 것이다.

발명의 상세한 설명

[0033] 본 발명의 기초가 되는 해결안은, 주제 영역(thematic area) 및 레마(lemma)를 위한 웹 설정으로부터의 시작해서, 상이한 검색 엔진들 및 지시된 스파이더링(spidering) 행위 모두를 출처로 사용하여, 적절하다고 여겨지는 웹 페이지의 세트(set)를 결정하는 대응 콘텐트의 분류 및 조직화 방법을 제공하는 것이다.

[0034] 스파이더링 행위는 웹 설정 단계에서 신뢰할 만한 것으로 고려되는 사이트들에 대해 수행된다.

[0035] 이러한 웹 페이지의 세트에 근거하여, 선택이 수행된다. 선택은 예컨대, 동일한 웹 페이지의 상이한 번역물, 또는 여러 번 반복된 동일한 인터넷 주소와 결합된 중복된 웹 페이지들을 제거하고, 사용자의 기여(contribution)를 통해 가식적인(spurious) 웹 페이지 즉, 정보 노이즈를 가지고 있고/있거나 그다지 신뢰할 수 없는 출처로

부터 얻어진 것들을 제거한다.

[0036] 이러한 해결안에 따르면, 본 발명에 따른 웹 페이지를 분류하고 대응 콘텐트를 조직화하는 방법은,

[0037] 미리 결정된 수의 인터넷 주소의 복수의 자동 기록 처리 절차의 실행을 포함하는 복수의 인터넷 주소를 기록하는 단계; 및

[0038] 상기 복수의 인터넷 주소에 대응하는 적절성 값(pertinence value)을 선택하고 설정(setting)하는 단계를 포함한다.

[0039] 선택 단계는 상기 기록 단계 내에서의 기록에 비례하는 적절성 값에 대응하는 적절성 값을 설정하고, 이미 결정된 임계값(threshold value)보다 큰 적절성 값을 가진 인터넷 주소를 선택하는 것을 특징으로 한다.

[0040] 상기 방법은,

[0041] 상기 임계값보다 큰 적절성 값을 가진 인터넷 주소 중에서, 하나 이상의 필수 기준(essentiality criteria)을 가진 인터넷 주소로 인터넷 주소를 정리하는 정리 단계; 및

[0042] 필수 기준에 부합하는 인터넷 주소의 서브셋(subset)을 확인하기 위한 확인 단계를 더 포함하며, 확인 단계는 인간의 행위를 포함한다.

[0043] 유리하게, 본 발명의 웹 페이지 분류 방법에 따르면, 식별 스크립트(identification script) 또는 프로그램은 사용자에 의해 설정된 검색 기준과 관련되고, 웹 페이지 반복의 면에서 중복되지(redundant) 않으며, 웹 페이지 콘텐트가 신뢰할 만한 결과를 제공할 수 있고, 검색 결과를 필터링하기 위해 사용자에 의해 채용되는 시간을 상당히 줄일 수 있다.

[0044] 본 발명에 따른 분류 방법의 그 이상의 특징 및 장점은 첨부된 도면을 참조하여 직설적으로 비한정적으로 만들어진 아래의 본 발명의 일 실시예의 기재로부터 더욱 명확해질 것이다.

실 시 예

[0051] 첨부된 도면을 참조하며, 본 발명은 복수의 단계를 포함하는 웹 페이지 분류 방법을 기술한다.

[0052] 자세하게는, 본 발명은 웹 페이지 분류를 위한 주제 영역(thematic area)의 정의를 제공하는 처음의 설정 단계를 포함한다.

[0053] 본 발명의 보호 범위를 감소시키지 않는 예시로서, 주제 영역은 예컨대, 기술, 예술, 문학, 스포츠, 시사와 같은 지식(knowledge)의 부분이고, 이러한 지식 부분의 기술적인 단어들의 세트(set)에 의해 정의된다.

[0054] 설정 단계는 각 주제 영역에 대한 복수의 레마(lemma)들의 정의를 포함한다.

[0055] 단지 명확성을 위해 그리고 본 발명의 한도를 제한하지 않으면서, 레마라는 용어는 언어학에서 형태학상 분석(morphological analysis)의 표제어(headword) 및 다른 형태의 요약 단위(abstract unit)를 나타내는데 쓰이며, 각 요약 단위는 개략적으로 동일한 단어의 다른 형태들의 단어들의 세트에 대응된다.

[0056] 예를 들면, "run", "runs", "ran" 및 "running"은 영어에서 인습적으로 또는 일반적으로 RUN으로 쓰이는 동일한 요약 단위의 형태들이다. 요약 단위 RUN은 요약 단위 RUN과 관련성을 가지는 복수의 요약 단위를 포함하는 레마 X에 포함된다.

[0057] 또한, 각 주제 영역은 복수의 영역 키워드(key word)와 연결되고, 어떤 주제 영역은 단일 레마와 결합된다. 하나 이상의 주제 영역에 속한 레마는 대응하는 주제 영역에 속하는 복수의 키워드와 연결된다.

[0058] 예를 들면, 설정 단계에서, 검색 기준 "레오나르도 다빈치(Leonardo da Vinci)"는 주제 영역 "예술(Art)"에 속하는 단어들인 영역 키워드 "회화(Painting)", "르네상스(Renaissance)", "조각(Sculpture)"과 연결된 레마로서 설정된다. 표제어 "레오나르도 다빈치(Leonardo da Vinci)"는 또한 주제 영역 "공학(Engineering)"에 속하는 "운하(Canals)", "수력학(水力學, Hydaulics)"의 영역 키워드와 연결된다.

[0059] 좀더 자세하게는, 설정 단계에서 또한 레마 및 영역 키워드와 관련된 주제 영역의 생성을 위해서 예컨대, 이탈리아어와 같은 주(主) 언어가 정의된다.

[0060] 또한, 설정 단계에서, 허가된 복수의 언어들은 설정된다. 예컨대, 동일한 알파벳에 의해 구별되고, 아래에 설명할 바와 같이 정리 단계에 이용된다.

[0061] 주 언어 및 허가된 언어는 단순하고 복잡한 검색들을 그들 자신의 언어로 수행하도록 하고, 또한 다른 언어로 된 웹 페이지를 얻을 수 있게 한다. 달리 말하면, 레마 "Leonardo da Vinci"를 설정하면, 본 발명의 분류 방법에 따른 결과는 레마 "Leonardo da Vinci"를 포함하는 이탈리아어로 된 웹 페이지뿐만 아니라 설정 단계에서 허가된 언어의 웹 페이지 또한 식별하도록 할 수 있다.

[0062] 이어서, 본 발명에 따른 분류는 설정 단계에서 정의된 주제 영역 및 대응 레마에 기반하여 웹 페이지의 이미 결정된 수의 인터넷 주소를 저장하기 위한 기록 단계를 포함한다. 이러한 웹 페이지의 미리 결정된 수의 인터넷 주소는 설정 단계에서 설정될 수 있다.

[0063] 자세하게는, 기록 단계는 하나 이상의 레마와 웹 페이지의 연결을 결정한다. 예컨대, 웹 페이지 내에서 레마의 식별에 의해 결정한다. 레마와 연결된 모든 웹 페이지에 있어서, 그 인터넷 주소는 저장된다.

[0064] 좀더 자세하게는, 기록 단계의 실행은 예컨대, 스파이더링(spidering) 절차와 같은 복수의 자동화된 절차에 의해 수행된다.

[0065] 이러한 스파이더링 절차는, 특정 레마와 연결된 웹 페이지 x로부터 시작하여, 웹 페이지 x와 직접적으로 연결된 웹 페이지 z1, ..., zn의 인터넷 주소를 기록한다. 스파이더링 절차는 또한 웹 페이지 x와 직접적으로 연결되지 않은 웹 페이지 y1, ..., y2의 인터넷 주소의 기록도 웹 페이지 z1, ..., zn에 의해 가능하다.

[0066] 유리하게, 스파이더링 절차는 웹 페이지의 단 한번의 직접적인 기록 즉, 웹 페이지 x에 직접적으로 도달 가능한 웹 페이지의 기록을 수행하거나, 상기 기술한 바와 같이 간접적인 기록을 수행하도록 설정될 수 있다.

[0067] 이어서, 기록 단계는 분류 방법을 위한 정보 층(information layer)을 나타내는 콘텐트를 가진 데이터베이스 내부에서 특정 레마를 참조하여 인터넷 주소의 저장을 제공한다. 자세하게는, 정보 층의 적절성(pertinence), 중복성(redundancy), 신뢰성(reliability) 및 정보 노이즈는 아래의 기술로부터 명백해질 본 발명의 분류 방법에 따라 최적화된다.

[0068] 본 발명의 방법은 웹 페이지뿐만 아니라 실질적으로 소셜 네트워크 서비스(Social Network Service)를 분류하고 조직화하는데도 제공될 수 있는 것을 알아둘 가치가 있다.

[0069] 더 자세하게는, 소셜 네트워크 서비스란 용어는 다른 사람들의 관심 및 활동을 조사하는데 관심이 있는 사람들이 온라인 커뮤니티를 통해 관심과 활동을 공유하는 인터넷 기반 서비스를 가리키는데 쓰인다.

[0070] 대부분의 소셜 네트워크 서비스는 주로 웹 기반이고, 예컨대, 채팅(chatting), 메시지(message), 이메일(email), 영상, 음성 채팅, 파일 공유(file sharing), 블로깅(blogging), 디스커션 그룹(discussion group)등과 같은 사용자들이 상호작용할 수 있는 많은 방법을 제공한다. 소셜 네트워크 서비스의 주요 형태는 어떤 카테고리의 디렉토리(directory) 및 친구들과 연결할 수 있는 수단을 포함하는 것이다.

[0071] 본 발명의 방법에 따른 기록 단계는 설정 단계에서 정의된 주제 영역 및 대응 레마에 기반하여 웹 페이지뿐만 아니라 미리 결정된 수의 소셜 네트워크 서비스를 저장하는 것을 제공하고, 각 소셜 네트워크가 하나 이상의 레마와 연결하도록 결정하는 것을 제공한다. 이 경우에도, 레마와 연결된 모든 소셜 네트워크 서비스에 있어서, 대응 인터넷 주소는 저장된다. 이 방법에서, 레마 및 주제 영역의 정의는, 미리 결정된 검색 기준에 대해, 노이즈가 없고, 적절하며, 중복되지 않고, 신뢰할 만한 웹 페이지 및 소셜 네트워크 서비스의 모음을 추출하는 분류 및 조직화의 구현을 위해 제공된다.

[0072] 예를 들어, 검색 기준으로 단어 "child"를 고려하면, 종래 기술의 식별 스크립트 또는 프로그램은 단어 "child"를 포함하는 엄청난 수의 문서의 식별에 한정될 것이다. 반면에, 본 발명의 분류 및 기록 단계는 단어 "child"가 속한 요약 단위를 포함하는 레마와 연결된 복수의 웹 페이지 및 소셜 네트워크 서비스를 제공한다.

[0073] 상기 예시를 참조하면, 단어 "child" 뿐만 아니라 "children", "boy" 및 비슷한 단어를 포함하는 웹 페이지 및 소셜 네트워크 서비스가 본 발명의 방법에 따라 분류되고 조직화된다.

[0074] 명확성을 위하여, 기록 단계로부터 비롯된 정보 층 또는 정보 저장소(information repository)는 제1 정보 층으로서 하부에 나타난다.

[0075] 자세하게는, 기록 단계는 예컨대, 메타-검색 엔진(meta-search engine)과 같은 기능을 포함하는 다른 자동화된 절차의 실행을 포함한다. 알려진 바와 같이, 이러한 기능은 그들 자신의 분류 방법에 따라 웹 페이지를 분류하는 상이한 검색 엔진의 참고에 의해 레마에 속하는 복수의 웹 페이지의 식별을 제공한다.

[0076] 달리 말하면, 메타-검색 엔진 기능은, 그들 자신의 분류에 따라 검색 엔진에 의해 이미 저장된 특정 레마를 포함하는 복수의 웹 페이지를 결정하기 위해 복수의 검색 엔진을 질의(query)한다. 알려진 검색 엔진의 예로 구글(Google), 야후(Yahoo), 알타비스타(Altavista) 등이 있다.

[0077] 메타-검색 엔진 기능에 의해 식별된 웹 페이지의 인터넷 주소는 제1 정보 층 또는 저장소에 삽입된다. 또한, 기록 단계는 웹 페이지의 사본(copy)을 제1 정보 층에 저장한다. 이러한 사본은 참고 및 그들의 다음 버전과의 비교 모두를 위해 사용되며, 제1 정보 층에서 웹 페이지 자신의 갱신(update)을 가능하게 하기 위해 다음 스파이더링 행위에 의해 식별된다.

[0078] 본 발명에 따르면, 분류 방법은 기록 단계에서 자동화된 절차에 의해 저장된 인터넷 주소의 비교를 수행하는 선택 단계를 더 포함한다.

[0079] 자세하게는, 특정 레마에 대하여 스파이더링 절차에 의해 기록된 인터넷 주소는 메타-검색 엔진 기능에 의해 기록된 인터넷 주소와 비교된다. 만약 인터넷 주소가 스파이더링 절차에 의해 또는 설정 단계에서, 및 메타-검색 엔진 기능에 의해 모두 식별되었다면, 이러한 선택 단계에 의해 확인된다. 사실은, 상이한 검색 엔진에서의 그 주소의 존재 및 설정 단계에서의 그 주소의 식별은 특정 레마에 대한 그 내용의 적절성의 높은 가능성을 나타내는 것이다.

[0080] 기록 단계는 유리하게 설정 단계에서 또는 메타-검색 엔진 기능 상에서 반드시 스파이더링 절차에 기반하지는 않는 상이한 형태의 자동화 절차를 포함한다. 모든 자동화된 절차에 의해 수행되는 기록들에 기반하여, 선택 단계는 식별된 인터넷 주소의 비교를 수행한다. 즉, 인터넷 주소에 자동화된 절차에 의한 그 주소의 기록의 빈도수(frequency)와 함께 증가하는, 확실한 적절성 및 신뢰성 값을 할당한다.

[0081] 본 발명의 선택 단계에 따르면, 제1 정보 층은 미리 설정된 임계값 이상의 확실한 적절성 및 신뢰성 값을 가진 인터넷 주소를 포함하는 제2 정보 층 또는 저장소를 형성하기 하기 위해 정제(refine)된다. 제2 정보 층은 예컨대, 상기 언급한 데이터베이스 또는 개별 데이터베이스에 담긴다.

[0082] 제2 정보 데이터베이스는, 비록 그것이 확실히 높은 적절성 값을 가진 인터넷 주소를 포함하더라도, 잠재적으로 중복적인 콘텐트이다. 실은, 선택 동안에 할당된 확실히 높은 적절성 값에 의해, 상이한 인터넷 주소가 동일한 웹 페이지를 참조할 수 있다. 예를 들면, 상이한 언어로 번역된 동일한 웹 페이지의 다양한 버전들, 또는 동일한 웹 페이지의 갱신되지 않은 채로(non-updated) 공표된 버전, 또는 상이한 도메인에 속하는 웹 페이지에 저장된 동일한 정보 컨텐트와 같은 웹 페이지를 참조할 수 있다.

[0083] 제2 정보 저장소의 내용을 향상시키기 위해서, 본 발명에 따른 분류 방법은 선택 단계 동안 확인된 인터넷 주소의 정리 단계를 포함한다.

[0084] 이러한 정리 단계는 제2 정보 층의 인터넷 주소에 의해 직결(direct)되는 모든 웹 페이지가 필수 기준(essentiality criterion)을 고려하도록 하는 것을 제공한다.

[0085] 이러한 필수 기준은 예컨대, 설정 단계에서 설정된, 허가된 언어에 대응하지 않는 모든 웹 페이지 및 소셜 네트워크 서비스의 제거와 같은 것을 제공한다. 또한, 필수 기준은 그 메인 페이지 대신에 보고된, 동일한 도메인 또는 주제 영역으로부터의 동일한 레마에 대해 나온 웹 페이지의 제거를 포함한다.

[0086] 정리 단계는 필수 기준의 만족을 확인함에 의해, 더 정제된 제3 정보 층 또는 저장소에 대응 인터넷 주소의 존재를 자리잡게 한다.

[0087] 유리하게, 제3 정보 층 또는 저장소 상에서 수행되는 식별 스크립트 또는 프로그램은 실질적으로 중복성이 제거되고, 높은 적절성의 가능성을 가진 웹 페이지 및 소셜 네트워크 서비스의 인터넷 주소를 식별할 수 있다.

[0088] 예시적인 그리고 비한정적인 목적으로, 본 발명의 분류 방법에 의해 지원되는, 검색 기준 "Leonardo da Vinci"에 기반한 식별 스크립트 또는 프로그램은 높은 적절성 가능성을 가지고 실질적으로 중복되지 않는, 이미 이 단계에 있는 수백 개의 인터넷 주소를 결정할 수 있다. 반면에, 종래 기술에 따른 분류 방법에 의해 지원되는 식별 스크립트 또는 프로그램은 마지막 선택에서 수백만의 인터넷 주소를 식별한다.

[0089] 본 발명에 따르면, 제3 정보 층 상의 부가적인 확인(validation) 단계는, 매우 신뢰할 만한 웹 페이지를 참조하는 단일 인터넷 주소를 결정할 목적으로, 실질적으로 정보 노이즈를 줄이는 것을 제공한다.

[0090] 이러한 확인 단계는 제3 정보 층의 웹 페이지 또는 소셜 네트워크 서비스를 리스트(list)하는 확인 인터페이스(interface)의 실행을 포함한다. 좀더 자세하게는 확인 인터페이스는 앞의 기록, 선택 및 정리 단계에서 이미

할당된 확실한 적절성 값에 따라 이러한 인터넷 주소의 정렬을 제안한다.

[0091] 확인 인터페이스를 통해, 특별 오퍼레이터(especial operator)는 인터넷 주소와 연결된 웹 페이지 및 소셜 네트워크 서비스를 분석하고, 특정 레마에 대한 그들의 적절성을 확인(validate)한다. 더 자세하게는, 오퍼레이터는, 그래픽 인터페이스를 통해 그리고 자동 비교 절차 및 결과의 신호 알림의 원조를 통해, 비적절한 그리고 바람직하지 않은, 낮은 또는 0의 신뢰성을 갖는 인터넷 주소를 제거한다. 예컨대, "Leonardo da Vinci" 공항을 참조하는 인터넷 주소가 제4 정보 층에 삽입되는 것을 방지한다.

[0092] 본 발명에 따르면, 예컨대, 확인에 대한 체크 박스(check box)와 같은 특정 기능이 상기 기술한 확인 단계를 실행하는 마지막 사용자에게 제공될 수 있다. 마지막 사용자는 웹 페이지 및 소셜 네트워크 서비스를 분석하고, 고려된 것들을 보다 적절하게 확인한다. 이는 특별 오퍼레이터에 의해 동작되는 확인 양상의 한층 진보이다. 왜냐하면, 마지막 사용자는 웹 페이지 및 소셜 네트워크 서비스에 그들의 확인을 제공할 수 있고 정보 층 또는 저장소의 개발 및 정제에 기여할 수 있는 각 분야의 전문가이기 때문이다.

[0093] 바람직한 제1 실시예에 따르면, 만일, 마지막 사용자에 의해 확인된 웹 페이지 및 소셜 네트워크 서비스의 세트가 특별 오퍼레이터에 의해 확인된 웹 페이지 및 소셜 네트워크 서비스의 세트와 상이하다면, 각각의 결과를 보고하는 인터페이스의 개별 리스트로 리스트된다.

[0094] 바람직한 제2 실시예에 따르면, 제3 정보 층의 웹 페이지 및 소셜 네트워크 서비스가 마지막 사용자 및 특별 오퍼레이터 모두에 의해 확인된다.

[0095] 유리하게, 제3 정보 층의 인터넷 주소의 수동 선택으로부터 얻어지는 제4 정보 층의 정의에 있어 인간의 기여는 상당하다.

[0096] 따라서, 본 발명의 분류 방법에 따르면, 웹 페이지의 분류는 적어도 다음의 단계를 포함한다.

[0097] A) 미리 결정된 수의 인터넷 주소의 복수의 자동 기록 절차의 실행을 포함하는 상기 복수의 인터넷 주소를 기록하는 단계;

[0098] B) 상기 복수의 인터넷 주소에 대응하는 적절성 값(pertinence value)을 설정(setting)하기 위한 선택 단계로서, 상기 기록 단계에서의 기록에 비례하는 값에 대응하는 확실한 적절성 값을 설정하고, 이미 결정된 임계값(threshold value)보다 큰 적절성 값을 가진 인터넷 주소를 선택하는 선택 단계;

[0099] C) 상기 임계값보다 큰 적절성 값을 가진 인터넷 주소 중에서, 하나 이상의 필수 기준(essentiality criteria)에 부합하지 않는 인터넷 주소를 제거하는 정리 단계; 및

[0100] D) 필수 기준에 부합하는 인터넷 주소의 서브셋(subset)을 확인하기 위한 확인 단계로서, 사용자에 의해 실행되는 수동 행위를 포함하는 확인 단계.

[0101] 예를 들면, 검색 기준 "Leonardo da Vinci"에서, 본 발명에 따른 식별 스크립트 또는 프로그램은 도 2에 도시한 바와 같이, 수십 개의 인터넷 주소를 결정할 수 있다.

[0102] 자세하게는, 도 2는 예컨대, 검색 기준의 삽입 바 또는 삽입부(11) 및 식별 스크립트 또는 프로그램의 활성화를 위한 활성화 버튼(21)을 포함하는 웹 페이지와 같이, 인터페이스(101)로부터 그려질 수 있는 웹 페이지의 식별 스크립트 또는 프로그램을 도시한다.

[0103] 도 2에서, 검색 기준은 이탈리아 발명가 및 예술가 레오나르도 다빈치를 포함하는 웹 페이지의 식별을 위해 설정된다.

[0104] 인터페이스(101)의 주부(41)는 식별 스크립트 또는 프로그램의 검색 기준 "Leonardo da Vinci"에 기반한 실행 후의 결과를 표시한다. 검색 결과는 웹 페이지의 인터넷 주소의 리스트(31)를 포함하고, 단지 첫 번째 히트들(first hits)(3a1 내지 3c1)이 표시된다.

[0105] 유리하게, 식별 스크립트 또는 프로그램은 마지막 사용자에게 매우 적절하고, 신뢰할 만하며 노이즈 또는 중복이 없는 웹 페이지 및 소셜 네트워크 서비스를 식별한 수십 개의 인터넷 주소를 제공할 수 있다. 인터넷 주소(3a1)는 BBC의 웹 페이지를, 인터넷 주소(3b1)는 메트로폴리탄 미술관(Metropolitan Museum)을, 인터넷 주소(3c1)는 잘 알려진 이탈리아 텔레비전 방송국 RAI에 의해 발명가 레오나르도 다빈치에 헌정된 보고서를 참조한다.

[0106] 따라서, 식별 스크립트 또는 프로그램은 사용자에 의해 설정된 검색 기준에 대해 낮은 적절성을 갖는 다수의 웹

페이지를 제거하고, 제4 정보 층에 속한 웹 페이지만을 제안한다.

[0107] 전형적인 형태의 검색에 있어서, 인터페이스(101)는, 예컨대, 제2 정보 층과 같이 제4 정보 층보다 덜 정제된 정보 층 또는 저장소 상에서 식별 스크립트 또는 프로그램을 실행할 수 있도록 한다.

[0108] 예를 들면, 도 2의 버튼(61)을 선택함에 의해, 식별 스크립트 또는 프로그램은 제2 정보 층에 포함된 인터넷 주소의 리스트를 제공하고 따라서 또한 인간 확인 단계가 수행되지 않은 웹 페이지를 참고하도록 한다.

[0109] 더 자세하게는, 선택 단계로부터 비롯된 제3 정보 층 또는 저장소, 또는 정리 단계로부터 비롯된 제4 정보 층 또는 저장소가 실질적으로 비어있고, 특정 검색 기준에 대해 결과를 제공할 수 없는 경우에 분류 방법의 보조 실행의 활성화가 또한 제공된다. 이 경우, 보조 실행은, 웹 검색 엔진의 미리 결정된 세트에 의 접근 및 이러한 엔진들로부터 검색 기준에 대한 가장 적절한 결과의 추출을 제공하는 알고리즘에 기반한 것이다.

[0110] 유리하게, 인터페이스(101)은 또한, 하이퍼텍스트(hypertext) 형식으로 백과사전(encyclopaedic) 형태의 설정된 검색 기준과 관련된 정보를 제공하는 측부(42)를 포함한다. 예컨대, 검색 기준 "Leonardo da Vinci"에 대하여, 측부(42)는 전기(Biography), 생활 연대기(Life Chronology), 시간 연대기(Time Chronology), 레오나르도 코덱스(Leonardo Codices), 및 레오나르도 저서 목록(Leonardo Bibliography)을 제공한다.

[0111] 유리하게, 본 발명의 웹 페이지 분류 방법에 따르면, 식별 스크립트 또는 프로그램은 사용자가 설정한 검색 기준에 적절하고, 노이즈가 없으며, 웹 페이지의 숫자 및 반복 면에서 중복이 없고, 웹 페이지의 콘텐트가 신뢰할 만하며, 검색 결과를 필터링하는데 사용자에 의해 채용된 시간을 상당히 줄여주는 결과를 제공할 수 있다.

[0112] 상기 기술로부터 살펴볼 수 있듯이, 본 발명에 따라 웹 페이지 및 소셜 네트워크 서비스를 분류하는 방법은, 검색 기준에 관련된, 정확하고, 적절하며, 중복적이지 않으며, 노이즈가 없는 정보를 제공할 뿐만 아니라 사용에 매우 용이한 인터페이스를 제공한다.

[0113] 도 3을 참조하면, 웹 페이지 및 소셜 네트워크 서비스의 분류 및 그에 대응하는 콘텐트의 조직화의 실행을 시작점으로, 본 발명의 방법에 의해 제공되는 사용자 인터페이스(50)가 개략적으로 도시된다. 사용자 인터페이스(50)는 예컨대, 웹 어플리케이션의 인터넷 주소와 연결되고, 실질적으로 본 방법을 구현한다.

[0114] 사용자 인터페이스(50)는 검색 기준을 삽입하는 검색 박스(51), 본 방법의 실행으로부터 비롯된 웹 페이지들을 리스트하는 검색 결과 박스(52), 및 검색 기준과 관련된 부가적인 정보를 포함하는 복수의 부가적인 박스(53)를 포함한다.

[0115] 부가적인 박스(53)는 하나 이상의 소셜 네트워크 서비스가 리스트된 박스(54), 예컨대, 서드 파티(third party)에 의해 제공되는, 검색 기준에 관련된 논쟁에 대한 박스(55), 사용자의 코멘트(comment)를 포함하는 박스(56)를 포함한다.

[0116] 더 자세하게는, 서드 파티에 의해 제공되는 논쟁은, 실질적으로 본 발명의 방법에 따라 서드 파티에 의해 제공되고, 만들어지며, 재조직화된 하나 이상의 데이터베이스로부터 얻어진 복수의 정보이고, 박스(55)에 링크(link)의 리스트로 반환된다. 달리 말하며, 박스(55)의 콘텐트는 사용자에 의해 설정된 미리 결정된 검색 기준에 기반한 검색의 독점(proprietary) 결과이다.

[0117] 유리하게, 본 발명의 방법은 하나 이상의 출처로부터의 데이터를 단일 통합 사용자 인터페이스(50)로 합병시키는 웹 2.0 매시업(Web 2.0 Mashup) 기술에 기반한 것이다.

[0118] 사실은, 본 발명의 방법에 따른 기록, 선택, 정리, 및 확인 단계를 통해 검출된 결과는 웹 페이지뿐만 아니라 소셜 네트워크 서비스 및 서드 파티 콘텐트를 포함하는 복수의 출처로부터 얻어진다.

[0119] 매시업 기술은 웹 페이지, 소셜 네트워크 서비스 및 콘텐트 상에서 구현된다. 이 때문에, 이러한 웹 페이지, 소셜 네트워크 서비스 및 콘텐트로부터 얻어지는 단일 출처에 의해 원래 제공되는 것이 아닌 새롭고 구별되는 도구를 생성한다.

[0120] 매시업 기술로 구현된 알려진 서비스에 관하여 커다란 진보가 있음을 알 가치가 있다. 사실은, 알려진 서비스들은 상이한 출처로부터 얻어진 복수의 정보를 단일 인터페이스로의 그룹화를 제공하지만, 평범한 검색 기준에 대해 이러한 복수의 정보를 결합하지 못한다.

[0121] 유리하게, 본 발명은, 먼저 검색 박스의 평범한 검색 기준에 연결된 복수의 콘텐트를 검출하고, 이어서 사용자 인터페이스(50)의 상기 언급한 박스들(52, 53, 54, 55)에 이러한 콘텐트를 그룹화한다.

[0122] 도 4를 참조하면, 검색 기준 "Leonardo da Vinci"에 대한 본 방법의 결과로서 사용자 인터페이스(50)가 주요 세부 사항으로 도시된다.

[0123] 사용자 인터페이스(50)의 각 박스(51 내지 54)는, 예컨대, 인터페이스의 제1 부터 제2 위치까지 드래그 및 드롭 또는 다른 박스 즉, 웹 페이지의 박스의 링크들의 주요 숫자를 표시하기 위해 제1 박스의 크기를 늘이거나/줄이거나 하는 것과 같이 사용자에 의해 수정될 수 있도록 위젯(widget) 기술에 의해 구현된다.

[0124] 하나 이상의 박스는 또한 켜지거나 꺼질 수 있다. 달리 말하면, 사용자 인터페이스(50)로부터 삽입되거나 제거될 수 있다.

[0125] 유리하게, 사용자는 제공된 결과들 중에서 그가 관심있는 정보의 형태를 선택할 수 있다. 예컨대, 사용자 인터페이스(50)의 주부를 웹 페이지에 대해, 또는 특정 소셜 네트워크 서비스에 대해, 또는 서드 파티에 의해 제공되는 콘텐트에 대해, 또는 다른 사용자의 코멘트에 대해 지정할 수 있다.

도면의 간단한 설명

[0045] 도 1은 종래 기술에 따른 식별 스크립트 또는 프로그램을 위한 그래픽 인터페이스를 도시한 도면.

[0046] 도 1a는 도 1의 식별 스크립트 또는 프로그램에 의해 검출된 웹 페이지의 수를 도시한 도면.

[0047] 도 1b는 도 1의 식별 스크립트 또는 프로그램에 의해 보고된 에러 메시지(error message)를 도시한 도면.

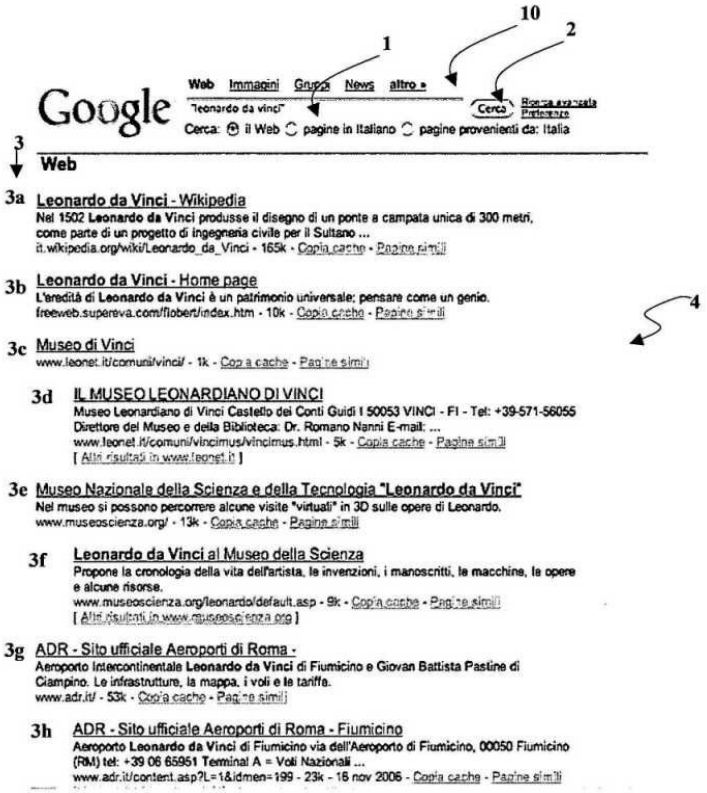
[0048] 도 2는 본 발명의 분류 방법에 따른 식별 스크립트 또는 프로그램을 위한 그래픽 인터페이스를 도시한 도면.

[0049] 도 3은 본 발명의 방법에 따른 콘텐트의 조직화를 개략적으로 도시한 도면.

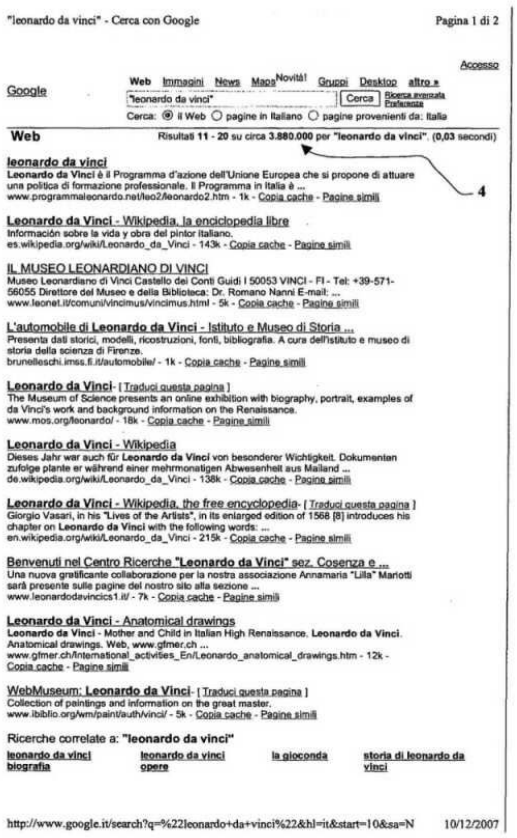
[0050] 도 4는 본 발명의 방법의 이미 결정된 검색 기준 상에서 실행 후, 콘텐트의 조직화를 주요 세부 사항으로 도시한 도면.

도면

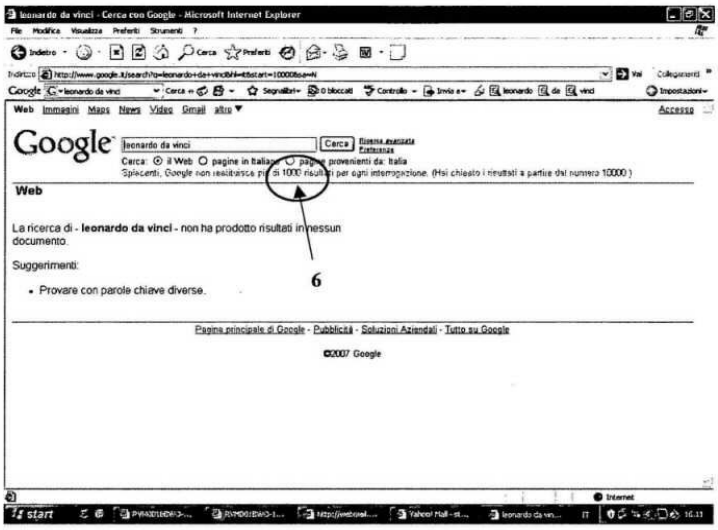
도면1


10
1
2
Google
Web Immagini Gruppi News altro »
"leonardo da vinci"
Cerca
Ricerca avanzata
Preferenze
Cerca: il Web pagine in Italiano pagine provenienti da: Italia
3
Web
3a Leonardo da Vinci - Wikipedia
Nel 1502 Leonardo da Vinci produsse il disegno di un ponte a campata unica di 300 metri, come parte di un progetto di ingegneria civile per il Sultano ...
it.wikipedia.org/wiki/Leonardo_da_Vinci - 165k - Copia cache - Pagine simili
3b Leonardo da Vinci - Home page
L'eredità di Leonardo da Vinci è un patrimonio universale; pensare come un genio.
freeweb.supereva.com/flobert/index.htm - 10k - Copia cache - Pagine simili
4
3c Museo di Vinci
www.leonet.it/comuni/vinci/ - 1k - Copia cache - Pagine simili
3d IL MUSEO LEONARDIANO DI VINCI
Museo Leonardiano di Vinci Castello dei Conti Guidi I 50053 VINCI - FI - Tel: +39-571-56055
Direttore del Museo e della Biblioteca: Dr. Romano Nanni E-mail: ...
www.leonet.it/comuni/vincimus/vincimus.html - 5k - Copia cache - Pagine simili
[Altri risultati in www.leonet.it]
3e Museo Nazionale della Scienza e della Tecnologia "Leonardo da Vinci"
Nel museo si possono percorrere alcune visite "virtuali" in 3D sulle opere di Leonardo.
www.museoscienza.org/ - 13k - Copia cache - Pagine simili
3f Leonardo da Vinci al Museo della Scienza
Propone la cronologia della vita dell'artista, le invenzioni, i manoscritti, le macchine, le opere e alcune risorse.
www.museoscienza.org/leonardo/default.asp - 9k - Copia cache - Pagine simili
[Altri risultati in www.museoscienza.org]
3g ADR - Sito ufficiale Aeroporti di Roma -
Aeroporto Intercontinentale Leonardo da Vinci di Fiumicino e Giovan Battista Pastine di Ciampino. Le infrastrutture, la mappa, i voli e le tariffe.
www.adr.it/ - 53k - Copia cache - Pagine simili
3h ADR - Sito ufficiale Aeroporti di Roma - Fiumicino
Aeroporto Leonardo da Vinci di Fiumicino via dell'Aeroporto di Fiumicino, 00050 Fiumicino (RM) tel: +39 06 65951 Terminal A = Voli Nazionali ...
www.adr.it/content.asp?L=1&idmen=199 - 23k - 18 nov 2006 - Copia cache - Pagine simili

도면1a


"leonardo da vinci" - Cerca con Google
Pagina 1 di 2
Accesso
Google
Web Immagini News Maps Novità! Gruppi Desktop altro »
"leonardo da vinci"
Cerca
Ricerca avanzata
Preferenze
Cerca: il Web pagine in italiano pagine provenienti da: Italia
Web
Risultati 11 - 20 su circa 3.880.000 per "leonardo da vinci". (0,03 secondi)
4
leonardo da vinci
Leonardo da Vinci è il Programma d'azione dell'Unione Europea che si propone di attuare una politica di formazione professionale. Il Programma in Italia è ...
www.programmaleonardo.net/leo2/leonardo2.htm - 1k - Copia cache - Pagine simili
Leonardo da Vinci - Wikipedia, la enciclopedia libre
Información sobre la vida y obra del pintor italiano.
es.wikipedia.org/wiki/Leonardo_da_Vinci - 143k - Copia cache - Pagine simili
IL MUSEO LEONARDIANO DI VINCI
Museo Leonardiano di Vinci Castello dei Conti Guidi I 50053 VINCI - FI - Tel: +39-571-56055 Direttore del Museo e della Biblioteca: Dr. Romano Nanni E-mail: ...
www.leonet.it/comuni/vinci/mus/vinci/mus.html - 5k - Copia cache - Pagine simili
L'automobile di Leonardo da Vinci - Istituto e Museo di Storia ...
Presenta dati storici, modelli, ricostruzioni, fonti, bibliografia. A cura dell'Istituto e museo di storia della scienza di Firenze.
brunelleschi.imss.fi.it/automobile/ - 1k - Copia cache - Pagine simili
Leonardo da Vinci - [Traduci questa pagina]
The Museum of Science presents an online exhibition with biography, portrait, examples of da Vinci's work and background information on the Renaissance.
www.mos.org/leonardo/ - 18k - Copia cache - Pagine simili
Leonardo da Vinci - Wikipedia
Dieses Jahr war auch für Leonardo da Vinci von besonderer Wichtigkeit. Dokumenten zufolge plante er während einer mehrmonatigen Abwesenheit aus Mailand ...
de.wikipedia.org/wiki/Leonardo_da_Vinci - 138k - Copia cache - Pagine simili
Leonardo da Vinci - Wikipedia, the free encyclopedia - [Traduci questa pagina]
Giorgio Vasari, in his "Lives of the Artists", in its enlarged edition of 1568 [8] introduces his chapter on Leonardo da Vinci with the following words: ...
en.wikipedia.org/wiki/Leonardo_da_Vinci - 215k - Copia cache - Pagine simili
Benvenuti nel Centro Ricerche "Leonardo da Vinci" sez. Cosenza e ...
Una nuova gratificante collaborazione per la nostra associazione Annamaria "Lilla" Mariotti sarà presente sulle pagine del nostro sito alla sezione ...
www.leonardodavincics1.it/ - 7k - Copia cache - Pagine simili
Leonardo da Vinci - Anatomical drawings
Leonardo da Vinci - Mother and Child in Italian High Renaissance. Leonardo da Vinci. Anatomical drawings. Web, www.gfmer.ch ...
www.gfmer.ch/International_activities_En/Leonardo_anatomical_drawings.htm - 12k - Copia cache - Pagine simili
WebMuseum: Leonardo da Vinci - [Traduci questa pagina]
Collection of paintings and information on the great master.
www.ibiblio.org/wm/paint/auth/vinci/ - 5k - Copia cache - Pagine simili
Ricerche correlate a: "leonardo da vinci"
leonardo da vinci biografia
leonardo da vinci opere
la gioconda
storia di leonardo da vinci
http://www.google.it/search?q=%22leonardo+da+vinci%22&hl=it&start=10&sa=N
10/12/2007

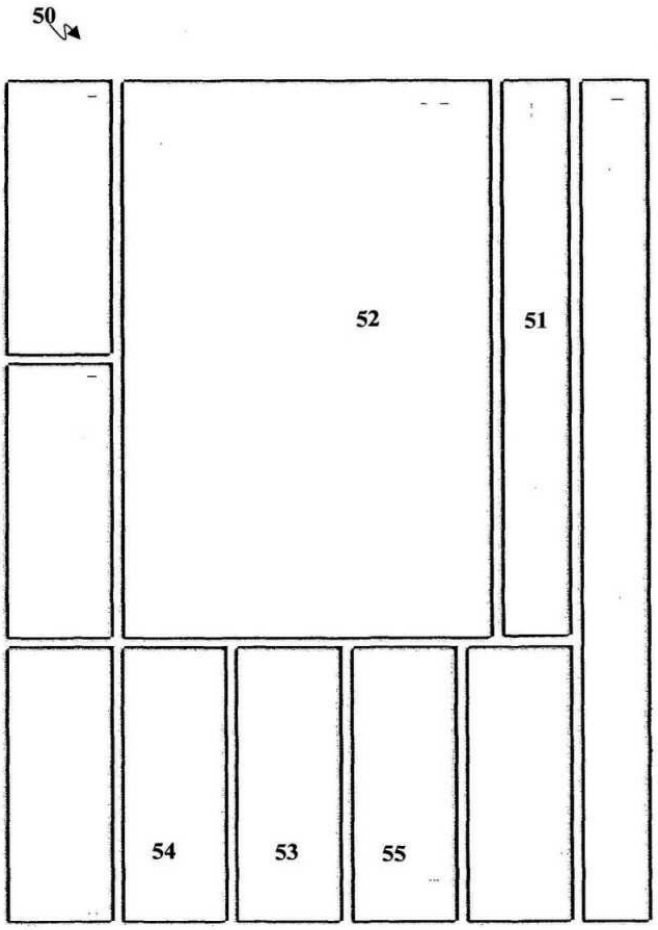
도면1b


leonardo da vinci - Cerca con Google - Microsoft Internet Explorer
Web Immagini Maps News Video Gmail altro
Accesso
Google
leonardo da vinci
Cerca
Cerca: il Web pagine in italiano pagine provenienti da: Italia
Web
La ricerca di - leonardo da vinci - non ha prodotto risultati in nessun documento.
Suggerimenti:
• Provare con parole chiave diverse.
6
Pagina principale di Google - Pubblicità - Soluzioni Aziendali - Tutto su Google
©2007 Google
start

도면2


101
11
21
31
3a1
BBC - Science & Nature - Leonardo - Homepage
3b1
The Metropolitan Museum of Art - Special Exhibitions
41
3c1
Italica - Storia dell'arte italiana - Arte Rinascimentale ...
42
Facility ti propone
Leonardo da Vinci
Libri da scaricare
Rassegna Stampa
Link a pagamento
Argomenti correlati
Biografia
Cronologia della vita
Cronologia del tempo
I Codici di Leonardo
Bibliografia di Leonardo
Altre voci correlate
Adorazione dei Magi
Annunciazione
Battesimo del Verrocchio
Belle Ferroniere
Bosch, Hyeronymus

도면3


50
52
51
54
53
55

도면4

