



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2016년07월15일  
 (11) 등록번호 10-1640051  
 (24) 등록일자 2016년07월11일

- |  |  |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/> <b>G06Q 50/10</b> (2012.01)</p> <p>(21) 출원번호 <b>10-2012-7014075</b></p> <p>(22) 출원일자(국제) <b>2012년10월25일</b><br/>                 심사청구일자 <b>2012년05월30일</b></p> <p>(85) 번역문제출일자 <b>2012년05월30일</b></p> <p>(65) 공개번호 <b>10-2012-0088792</b></p> <p>(43) 공개일자 <b>2012년08월08일</b></p> <p>(86) 국제출원번호 <b>PCT/JP2010/068820</b></p> <p>(87) 국제공개번호 <b>WO 2011/052526</b><br/>                 국제공개일자 <b>2011년05월05일</b></p> <p>(30) 우선권주장<br/>                 JP-P-2009-250594 2009년10월30일 일본(JP)<br/>                 JP-P-2009-250646 2009년10월30일 일본(JP)</p> <p>(56) 선행기술조사문헌<br/>                 KR1020070047579 A*<br/>                 *는 심사관에 의하여 인용된 문헌</p> | <p>(73) 특허권자<br/> <b>라쿠텐 인코포레이티드</b><br/>                 일본국 1580094 도쿄도 세타가야쿠 타마가와 1초<br/>                 메14만 1고</p> <p>(72) 발명자<br/> <b>이와부찌, 시가꾸</b><br/>                 일본 140-0002 도쿄도 시나가와구 히가시시나가와<br/>                 4-12-3 라쿠텐 인코포레이티드 내</p> <p>(74) 대리인<br/> <b>정석현, 박충범, 장수길, 이중희</b></p> |
|--|--|

전체 청구항 수 : 총 13 항

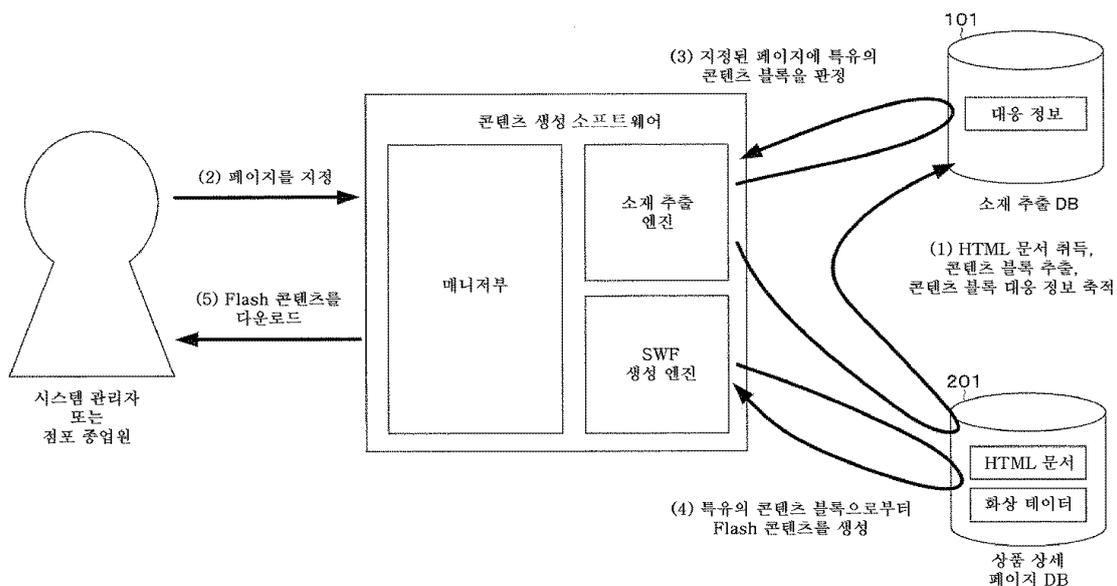
심사관 : 김상현

(54) 발명의 명칭 **특유 콘텐츠 판정 장치, 특유 콘텐츠 판정 방법, 기록 매체, 콘텐츠 생성 장치 및 관련 콘텐츠 삽입 장치**

**(57) 요약**

Web 페이지를 구성하고 있는 콘텐츠 중으로부터의 그 Web 페이지 특유의 콘텐츠를 용이하게 추출한다. 컴퓨터를, 지정된 Web 페이지를 구성하고 있는 콘텐츠를 추출하는 추출 수단, 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 계산 수단, 및 계산된 출현 빈도에 기초하여, 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 해당 Web 페이지 특유의 콘텐츠를 판정하는 판정 수단으로서 기능시킨다.

**대표도**



**명세서**

**청구범위**

**청구항 1**

삭제

**청구항 2**

삭제

**청구항 3**

삭제

**청구항 4**

삭제

**청구항 5**

삭제

**청구항 6**

삭제

**청구항 7**

삭제

**청구항 8**

삭제

**청구항 9**

삭제

**청구항 10**

삭제

**청구항 11**

삭제

**청구항 12**

삭제

**청구항 13**

삭제

**청구항 14**

소정의 사이트에 포함된 복수의 Web 페이지 중, 지정된 Web 페이지를 구성하고 있는 1개 이상의 콘텐츠를 포함하는 콘텐츠 그룹을 추출하는 추출 수단 - 각 콘텐츠는 블록화 태그에 의해 콘텐츠 그룹으로 블록화됨 - 과, 상기 복수의 Web 페이지 중, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠 그룹이 다른 Web 페이지에서

이용되는 출현 빈도를 계산하는 계산 수단과,

상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 다른 Web 페이지에서 이용되는 출현 빈도에 기초하여, 해당 Web 페이지 특유의 콘텐츠 그룹을 판정하는 판정 수단 - 상기 판정된 해당 Web 페이지 특유의 콘텐츠 그룹은, 동화상 콘텐츠 또는 정지 화상 콘텐츠를 생성하는데 이용됨 -

을 구비하는 것을 특징으로 하는 특유 콘텐츠 판정 장치.

**청구항 15**

제14항에 있어서,

상기 판정 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 출현 빈도가 가장 작은 콘텐츠 그룹을 해당 Web 페이지 특유의 콘텐츠 그룹이라고 판정하는 것을 특징으로 하는 특유 콘텐츠 판정 장치.

**청구항 16**

제14항에 있어서,

상기 판정 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 출현 빈도가 소정값 이하인 콘텐츠 그룹을 해당 Web 페이지 특유의 콘텐츠 그룹이라고 판정하는 것을 특징으로 하는 특유 콘텐츠 판정 장치.

**청구항 17**

삭제

**청구항 18**

제14항 내지 제16항 중 어느 한 항에 있어서,

상기 추출 수단은, 상기 소정의 사이트에 포함되는 미리 정해진 종류의 각 Web 페이지에 대해서 Web 페이지를 구성하고 있는 콘텐츠 그룹을 추출하고, 추출한 콘텐츠 그룹을 나타내는 콘텐츠 정보를 미리 기억 수단에 기억해 두고,

상기 계산 수단은, 상기 기억된 콘텐츠 정보에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠 그룹의 출현 빈도를 계산하는 것을 특징으로 하는 특유 콘텐츠 판정 장치.

**청구항 19**

삭제

**청구항 20**

제14항에 있어서,

상기 추출 수단은, 소정의 마크업 언어로 기술되고, Web 페이지를 구성하는 콘텐츠를 나타내는 문서 데이터에 기초하여, 콘텐츠 그룹을 추출하는 것을 특징으로 하는 특유 콘텐츠 판정 장치.

**청구항 21**

삭제

**청구항 22**

컴퓨터에 의해 행해지는 특유 콘텐츠 판정 방법으로서,

소정의 사이트에 포함된 복수의 Web 페이지 중, 지정된 Web 페이지를 구성하고 있는 1개 이상의 콘텐츠를 포함하는 콘텐츠 그룹을 추출하는 추출 행정 - 각 콘텐츠는 블록화 태그에 의해 콘텐츠 그룹으로 블록화됨 - 과,

상기 복수의 Web 페이지 중, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠 그룹이 다른 Web 페이지에서 이용되는 출현 빈도를 계산하는 계산 행정과,

상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 다른 Web 페이지에서 이용되는 출현 빈도에 기초하여, 해당 Web 페이지 특유의 콘텐츠 그룹을 판정하는 판정 행정 - 상기 판정된 해당 Web 페이지 특유의

콘텐츠 그룹은, 동화상 콘텐츠 또는 정지 화상 콘텐츠를 생성하는데 이용됨 -  
을 갖는 것을 특징으로 하는 특유 콘텐츠 판정 방법.

**청구항 23**

컴퓨터를,

소정의 사이트에 포함된 복수의 Web 페이지 중, 지정된 Web 페이지를 구성하고 있는 1개 이상의 콘텐츠를 포함하는 콘텐츠 그룹을 추출하는 추출 수단 - 각 콘텐츠는 블록화 태그에 의해 콘텐츠 그룹으로 블록화됨 -,

상기 복수의 Web 페이지 중, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠 그룹이 다른 Web 페이지에서 이용되는 출현 빈도를 계산하는 계산 수단, 및

상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 다른 Web 페이지에서 이용되는 출현 빈도에 기초하여, 해당 Web 페이지 특유의 콘텐츠 그룹을 판정하는 판정 수단 - 상기 판정된 해당 Web 페이지 특유의 콘텐츠 그룹은, 동화상 콘텐츠 또는 정지 화상 콘텐츠를 생성하는데 이용됨 -

으로서 기능시키는 특유 콘텐츠 판정 프로그램이 컴퓨터 관독 가능하게 기록되어 있는 것을 특징으로 하는 기록 매체.

**청구항 24**

제14항 내지 제16항 중 어느 한 항의 특유 콘텐츠 판정 장치와,

상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠 그룹이라고 판정된 콘텐츠 그룹에 기초하여, 새로운 콘텐츠를 생성하는 생성 수단 - 상기 새로운 콘텐츠는 동화상 콘텐츠 또는 정지 화상 콘텐츠임 -

을 구비하는 것을 특징으로 하는 콘텐츠 생성 장치.

**청구항 25**

제24항에 있어서,

상기 생성 수단은, 특유의 콘텐츠 그룹이라고 판정된 콘텐츠의 표시 사이즈를, 미리 설정된 표시 사이즈에 맞도록 조정하고, 표시 사이즈가 조정된 콘텐츠를 포함하는 새로운 콘텐츠를 생성하는 것을 특징으로 하는 콘텐츠 생성 장치.

**청구항 26**

제24항에 있어서,

상기 생성 수단은, 특유의 콘텐츠라고 판정된 콘텐츠 그룹에 이펙트가 실시되어 해당 콘텐츠가 재생되는 새로운 콘텐츠를 생성하는 것을 특징으로 하는 콘텐츠 생성 장치.

**청구항 27**

제14항 내지 제16항 중 어느 한 항의 특유 콘텐츠 판정 장치와,

상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠 그룹이라고 판정된 콘텐츠 그룹에 관련되는 관련 콘텐츠를, 상기 지정된 Web 페이지에 삽입하는 삽입 수단 - 상기 관련 콘텐츠는 동화상 콘텐츠 또는 정지 화상 콘텐츠임 -

을 구비하는 것을 특징으로 하는 관련 콘텐츠 삽입 장치.

**청구항 28**

제27항에 있어서,

상기 특유 콘텐츠 판정 장치는, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠로서, 블로그 기사의 텍스트 데이터가 포함되어 있는 경우에, 해당 텍스트 데이터를 해당 Web 페이지 특유의 콘텐츠라고 판정하고,

상기 삽입 수단은, 상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠라고 판정된 블로그 기사의 텍스트 데이터로부터 상기 지정된 Web 페이지의 특징어를 추출하고, 해당 특징어에 관련되는 관련 콘텐츠를, 해당 Web 페이지

에 삽입하는 것을 특징으로 하는 관련 콘텐츠 삽입 장치.

**청구항 29**

제25항에 있어서,

상기 생성 수단은, 특유의 콘텐츠 그룹이라고 판정된 콘텐츠에 이펙트가 실시되어 해당 콘텐츠가 재생되는 새로운 콘텐츠를 생성하는 것을 특징으로 하는 콘텐츠 생성 장치.

**발명의 설명**

**기술 분야**

[0001] 본 발명은, Web 페이지를 구성하는 콘텐츠를 추출하는 기술 분야에 관한 것이다.

**배경 기술**

[0002] 종래, Web 사이트 상에 공개되어 있는 Web 페이지를 구성하고 있는 소재인 콘텐츠를 취득하고, 취득한 콘텐츠에 기초하여 새로운 콘텐츠를 생성하는 기술이 알려져 있다. 예를 들면, 비특허 문헌 1에는, 유저에 의해 화상 데이터의 URL이 지정되면, 해당 URL에 대응하는 화상 데이터를 Web 상에서 취득하고, 취득한 화상 데이터에 기초하여 배너를 자동 작성하는 기술이 개시되어 있다.

**선행기술문헌**

**비특허문헌**

[0003] (비특허문헌 0001) 비특허 문헌 1 : “배너 자동 작성”, [online], [평성 21년(2009년) 10월 21일 검색], 인터넷<URL:ht tp://hyperbannermaker.com/>

**발명의 내용**

**해결하려는 과제**

[0004] Web 사이트를 구성하는 각 Web 페이지에는, 그 Web 사이트의 목적에 따른 내용이 게재된다. 그 때문에, Web 사이트를 구성하는 각 Web 페이지의 내용은, 기본적으로는 서로 관련성을 갖고 있는 것이지만, 각각 어떠한 특징을 갖고 있는 경우가 있다. 그리고, 그 Web 페이지의 내용을 결정하는 요인이, Web 페이지를 구성하고 있는 콘텐츠(예를 들면, 텍스트 데이터, 화상 데이터 등)의 내용이다. 따라서, Web 페이지를 구성하고 있는 콘텐츠 중에서, 그 Web 페이지를 특징짓는 콘텐츠, 즉, 그 Web 페이지 특유의 콘텐츠가 존재하는 경우가 있다.

[0005] 비특허 문헌 1에 기재된 기술은, Web 페이지 특유의 콘텐츠를 추출하는 것이지만, 자동적으로 추출하는 것이 아니라, 유저가 수작업으로 콘텐츠를 지정해야만 하여, 그 Web 페이지 특유의 콘텐츠를 용이하게 추출할 수는 없다. 그 때문에, 어느 콘텐츠가 Web 페이지 특유의 콘텐츠인지를 유저가 판단할 수 없는 경우, 또는 유저의 기호에 따라 원하는 콘텐츠가 치우치게 되는 경우 등, Web 페이지 특유의 콘텐츠를 정확하게 추출할 수 없다. 또한, 대상으로 하는 Web 페이지의 페이지수가 많으면, 유저의 작업이 심대하게 된다고 하는 문제가 있었다.

[0006] 또한, 예를 들면, HTML(Hyper Text Markup Language) 문서의 태그 기술에 기초하여, 화상만 또는 텍스트만과 같이, 특정 종류의 콘텐츠를 모두 추출하는 것은 가능하다. 그러나, 추출된 콘텐츠 중에는, Web 페이지 특유의 것이 아닌 흔히 있는 콘텐츠도 포함되므로, 추출 결과로서의 신뢰성이 낮고, 그 때문에 추출 결과로부터 유저가 특유의 콘텐츠를 찾아야만 하였다.

[0007] 본 발명은 이상의 점을 감안하여 이루어진 것으로, Web 페이지를 구성하고 있는 콘텐츠 중으로부터의 그 Web 페이지 특유의 콘텐츠를 용이하게 추출할 수 있는 특유 콘텐츠 판정 장치, 특유 콘텐츠 판정 방법, 특유 콘텐츠 판정 프로그램 등을 제공하는 것을 목적으로 한다.

**과제의 해결 수단**

[0008] 상기 과제를 해결하기 위해서, 청구항 1에 기재된 발명은, 컴퓨터를, 지정된 Web 페이지를 구성하고 있는 콘텐

츠를 추출하는 추출 수단, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 계산 수단, 및 상기 계산된 출현 빈도에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 해당 Web 페이지 특유의 콘텐츠를 판정하는 판정 수단으로서 기능시키는 것을 특징으로 한다.

- [0009] 본 발명에 따르면, 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도가 각각 계산된다. 이 출현 빈도가 작은 콘텐츠일수록, 지정된 Web 페이지 이외에는 그다지 출현하지 않는 콘텐츠이다. 그 때문에, 출현 빈도에 기초하여, 지정된 Web 페이지 특유의 콘텐츠를 특정할 수 있다. 따라서, Web 페이지 특유의 콘텐츠를 용이하게 추출할 수 있다.
- [0010] 청구항 2에 기재된 발명은, 청구항 1에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 판정 수단이, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 출현 빈도가 가장 작은 콘텐츠를 해당 Web 페이지 특유의 콘텐츠라고 판정하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0011] 본 발명에 따르면, 가장 출현 빈도가 작은 콘텐츠를 판단함으로써, 지정된 Web 페이지 특유의 콘텐츠가 특정된다.
- [0012] 청구항 3에 기재된 발명은, 청구항 1에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 판정 수단이, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 출현 빈도가 소정값 이하인 콘텐츠를 해당 Web 페이지 특유의 콘텐츠라고 판정하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0013] 본 발명에 따르면, 출현 빈도가 소정값 이하인지를 판정함으로써, 해당 조건을 충족시키는 모든 콘텐츠가, 지정된 Web 페이지 특유의 콘텐츠라고 특정된다.
- [0014] 청구항 4에 기재된 발명은, 청구항 1 내지 3 중 어느 한 항에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 계산 수단이, 소정의 사이트에 포함되는 복수의 Web 페이지 상에 있어서의 각 콘텐츠의 출현 빈도를 계산하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0015] 본 발명에 따르면, 소정의 사이트에 포함되는 어떤 복수의 Web 페이지 상에 있어서, 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도가 계산되므로, 사이트 내에서 공통해서 이용되는 콘텐츠는, Web 페이지 특유의 콘텐츠가 아니라고 판정하는 것이 가능하게 되고, Web 페이지 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0016] 청구항 5에 기재된 발명은, 청구항 1 내지 4 중 어느 한 항에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 추출 수단이, 상기 소정의 사이트에 포함되는 미리 정해진 종류의 각 Web 페이지에 대해서 Web 페이지를 구성하고 있는 콘텐츠를 추출하고, 추출한 콘텐츠를 나타내는 콘텐츠 정보를 미리 기억 수단에 기억해 두고, 상기 계산 수단이, 상기 기억된 콘텐츠 정보에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0017] 본 발명에 따르면, 소정의 사이트에 포함되는 미리 정해진 종류의 각 Web 페이지에 대해서 미리 추출해 둔 결과로서의 콘텐츠 정보에 기초하여, 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도가 계산되므로, 출현 빈도가 정확하게 계산되어, Web 페이지 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0018] 청구항 6에 기재된 발명은, 청구항 1 내지 5 중 어느 한 항에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 추출 수단이, 1개 이상의 콘텐츠로 구성된 콘텐츠 그룹의 단위로, Web 페이지를 구성하고 있는 콘텐츠를 추출하고, 상기 계산 수단이, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹의 출현 빈도를 계산하고, 상기 판정 수단이, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 해당 Web 페이지 특유의 콘텐츠 그룹을 판정하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0019] 본 발명에 따르면, 콘텐츠 그룹의 단위로 Web 페이지 특유의 콘텐츠가 판단되므로, 예를 들면, Web 페이지 상에 있어서 어떤 통합체로써 표시되어 있거나, 서로 관련성을 갖고 있는 콘텐츠를 콘텐츠 그룹으로 했을 때에, Web 페이지 특유의 콘텐츠로 되는 것을 추출할 수 있다.
- [0020] 청구항 7에 기재된 발명은, 청구항 6에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 추출 수단이, 소정의 마크업 언어로 기술되고, Web 페이지를 구성하는 콘텐츠를 나타내는 도큐먼트 데이터에 기초하여, 콘텐츠 그룹을 추출하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0021] 본 발명에 따르면, Web 페이지를 구성하는 콘텐츠를 나타내는 도큐먼트 데이터에 기초하여 콘텐츠 그룹이 추출되므로, 정확하게 콘텐츠 그룹을 추출할 수 있다.

- [0022] 청구항 8에 기재된 발명은, 청구항 7에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 추출 수단이, 상기 콘텐츠를 나타내는 도큐먼트 데이터에 있어서 미리 정해진 태그에 기초하여 콘텐츠 그룹을 정하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0023] 본 발명에 따르면, 미리 정해진 태그에 기초하여 콘텐츠 그룹이 추출되므로, Web 페이지 특유의 콘텐츠와, 특유하지 않은 콘텐츠가 각각 미리 정해진 태그로 그룹화되어 있는 경우에, Web 페이지 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0024] 청구항 9에 기재된 발명은, 청구항 1 내지 8 중 어느 한 항에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 특유의 콘텐츠라고 판정된 콘텐츠에 기초하여, 새로운 콘텐츠를 생성하는 생성 수단으로서 상기 컴퓨터를 더욱 기능시키는 것을 특징으로 한다.
- [0025] 본 발명에 따르면, Web 페이지 특유의 콘텐츠에 기초하여 새로운 콘텐츠가 생성되므로, 예를 들면, 해당 Web 페이지에 게재되어 있는 내용의 특징을 나타내는 콘텐츠를 생성할 수 있다.
- [0026] 청구항 10에 기재된 발명은, 청구항 9에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 생성 수단이, 특유의 콘텐츠라고 판정된 콘텐츠의 표시 사이즈를, 미리 설정된 표시 사이즈에 맞도록 조정하고, 표시 사이즈가 조정된 콘텐츠를 포함하는 새로운 콘텐츠를 생성하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0027] 청구항 11에 기재된 발명은, 청구항 9 또는 청구항 10에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 생성 수단이, 특유의 콘텐츠라고 판정된 콘텐츠에 이펙트가 실시되어 해당 콘텐츠가 재생되는 새로운 콘텐츠를 생성하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0028] 청구항 12에 기재된 발명은, 청구항 1 내지 8 중 어느 한 항에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 특유의 콘텐츠라고 판정된 콘텐츠에 관련되는 관련 콘텐츠를, 상기 지정된 Web 페이지에 삽입하는 삽입 수단으로서 상기 컴퓨터를 더욱 기능시키는 것을 특징으로 한다.
- [0029] 본 발명에 따르면, 특유의 콘텐츠라고 판정된 콘텐츠에 관련되는 콘텐츠가, 지정된 Web 페이지에 삽입되므로, Web 페이지의 특징과 관련되는 정보를 해당 Web 페이지에 추가할 수 있다.
- [0030] 청구항 13에 기재된 발명은, 청구항 12에 기재된 특유 콘텐츠 판정 프로그램에 있어서, 상기 판정 수단이, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠로서, 블로그 기사의 텍스트 데이터가 포함되어 있는 경우에, 해당 텍스트 데이터를 해당 Web 페이지 특유의 콘텐츠라고 판정하고, 상기 삽입 수단이, 상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠라고 판정된 블로그 기사의 텍스트 데이터로부터 상기 지정된 Web 페이지의 특징어를 추출하고, 해당 특징어에 관련되는 관련 콘텐츠를, 해당 Web 페이지에 삽입하도록, 상기 컴퓨터를 기능시키는 것을 특징으로 한다.
- [0031] 본 발명에 따르면, 각 블로그 기사의 텍스트 데이터에 그 기사 특유의 내용이 포함되어 있는 것이라면, 특유 콘텐츠 판정 장치에 의해 각 블로그 기사의 텍스트 데이터를 추출할 수 있다. 이에 의해, Web 페이지에 게재되어 있는 블로그의 내용에 관련되는 정보를 해당 Web 페이지에 추가할 수 있다.
- [0032] 청구항 14에 기재된 발명은, 지정된 Web 페이지를 구성하고 있는 콘텐츠를 추출하는 추출 수단과, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 계산 수단과, 상기 계산된 출현 빈도에 기초하여 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 해당 Web 페이지 특유의 콘텐츠를 판정하는 판정 수단을 구비하는 것을 특징으로 한다.
- [0033] 청구항 15에 기재된 발명은, 청구항 14에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 판정 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 출현 빈도가 가장 작은 콘텐츠를 해당 Web 페이지 특유의 콘텐츠라고 판정하는 것을 특징으로 한다.
- [0034] 청구항 16에 기재된 발명은, 청구항 14에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 판정 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 출현 빈도가 소정값 이하인 콘텐츠를 해당 Web 페이지 특유의 콘텐츠라고 판정하는 것을 특징으로 한다.
- [0035] 청구항 17에 기재된 발명은, 청구항 14 내지 16 중 어느 한 항에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 계산 수단은, 소정의 사이트에 포함되는 복수의 Web 페이지 상에 있어서의 각 콘텐츠의 출현 빈도를 계산하는 것을 특징으로 한다.
- [0036] 청구항 18에 기재된 발명은, 청구항 14 내지 17 중 어느 한 항에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기

추출 수단은, 상기 소정의 사이트에 포함되는 미리 정해진 종류의 각 Web 페이지에 대해서 Web 페이지를 구성하고 있는 콘텐츠를 추출하고, 추출한 콘텐츠를 나타내는 콘텐츠 정보를 미리 기억 수단에 기억해 두고, 상기 계산 수단은, 상기 기억된 콘텐츠 정보에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 것을 특징으로 한다.

[0037] 청구항 19에 기재된 발명은, 청구항 14 내지 18 중 어느 한 항에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 추출 수단은, 1개 이상의 콘텐츠로 구성된 콘텐츠 그룹의 단위로, Web 페이지를 구성하고 있는 콘텐츠를 추출하고, 상기 계산 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹의 출현 빈도를 계산하고, 상기 판정 수단은, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 그룹 중, 해당 Web 페이지 특유의 콘텐츠를 그룹을 판정하는 것을 특징으로 한다.

[0038] 청구항 20에 기재된 발명은, 청구항 19에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 추출 수단은, 소정의 마크업 언어로 기술되고, Web 페이지를 구성하는 콘텐츠를 나타내는 도큐먼트 데이터에 기초하여, 콘텐츠를 그룹을 추출하는 것을 특징으로 한다.

[0039] 청구항 21에 기재된 발명은, 청구항 20에 기재된 특유 콘텐츠 판정 장치에 있어서, 상기 추출 수단은, 상기 콘텐츠를 나타내는 도큐먼트 데이터에 있어서 미리 정해진 태그에 기초하여 콘텐츠 그룹을 정하는 것을 특징으로 한다.

[0040] 청구항 22에 기재된 발명은, 지정된 Web 페이지를 구성하고 있는 콘텐츠를 추출하는 추출 행정과, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 계산 행정과, 상기 계산된 출현 빈도에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 해당 Web 페이지 특유의 콘텐츠를 판정하는 판정 행정을 갖는 것을 특징으로 한다.

[0041] 청구항 23에 기재된 발명은, 컴퓨터를, 지정된 Web 페이지를 구성하고 있는 콘텐츠를 추출하는 추출 수단, 상기 지정된 Web 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하는 계산 수단, 및 상기 계산된 출현 빈도에 기초하여, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠 중, 해당 Web 페이지 특유의 콘텐츠를 판정하는 판정 수단으로서 기능시키는 특유 콘텐츠 판정 프로그램이 컴퓨터 판독 가능하게 기록되어 있는 것을 특징으로 한다.

[0042] 청구항 24에 기재된 발명은, 청구항 14 내지 21 중 어느 한 항에 기재된 특유 콘텐츠 판정 장치와, 상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠라고 판정된 콘텐츠에 기초하여, 새로운 콘텐츠를 생성하는 생성 수단을 구비하는 것을 특징으로 한다.

[0043] 청구항 25에 기재된 발명은, 청구항 24에 기재된 콘텐츠 생성 장치에 있어서, 상기 생성 수단은, 특유의 콘텐츠라고 판정된 콘텐츠의 표시 사이즈를, 미리 설정된 표시 사이즈에 맞도록 조정하고, 표시 사이즈가 조정된 콘텐츠를 포함하는 새로운 콘텐츠를 생성하는 것을 특징으로 한다.

[0044] 청구항 26에 기재된 발명은, 청구항 24 또는 청구항 25에 기재된 콘텐츠 생성 장치에 있어서, 상기 생성 수단은, 특유의 콘텐츠라고 판정된 콘텐츠에 이펙트가 실시되어 해당 콘텐츠가 재생되는 새로운 콘텐츠를 생성하는 것을 특징으로 한다.

[0045] 청구항 27에 기재된 발명은, 청구항 14 내지 21 중 어느 한 항에 기재된 특유 콘텐츠 판정 장치와, 상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠라고 판정된 콘텐츠에 관련되는 관련 콘텐츠를, 상기 지정된 Web 페이지에 삽입하는 삽입 수단을 구비하는 것을 특징으로 한다.

[0046] 청구항 28에 기재된 발명은, 청구항 27에 기재된 관련 콘텐츠 삽입 장치에 있어서, 상기 특유 콘텐츠 판정 장치는, 상기 지정된 Web 페이지를 구성하고 있는 콘텐츠로서, 블로그 기사의 텍스트 데이터가 포함되어 있는 경우에, 해당 텍스트 데이터를 해당 Web 페이지 특유의 콘텐츠라고 판정하고, 상기 삽입 수단은, 상기 특유 콘텐츠 판정 장치에 의해 특유의 콘텐츠라고 판정된 블로그 기사의 텍스트 데이터로부터 상기 지정된 Web 페이지의 특징어를 추출하고, 해당 특징어에 관련되는 관련 콘텐츠를, 해당 Web 페이지에 삽입하는 것을 특징으로 한다.

### **발명의 효과**

[0047] 본 발명에 따르면, 출현 빈도가 작은 콘텐츠일수록, 지정된 Web 페이지 이외에는 그다지 출현하지 않는 콘텐츠이다. 그 때문에, 출현 빈도에 기초하여, 지정된 Web 페이지 특유의 콘텐츠를 특정할 수 있다. 따라서, Web 페이지 특유의 콘텐츠를 용이하게 추출할 수 있다.

**도면의 간단한 설명**

[0048]

- 도 1은 일 실시 형태에 따른 쇼핑 시스템 S의 개요 구성의 일례를 도시하는 도면.
- 도 2는 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 개요 구성의 일례를 도시하는 블록도.
- 도 3은 Web 페이지가 지정되고나서 Flash 콘텐츠가 생성될 때까지의 처리의 개요를 도시하는 도면.
- 도 4는 Web 페이지의 구성예를 도시하는 도면.
- 도 5는 HTML 문서로부터 생성된 DOM 트리의 일례를 도시하는 도면.
- 도 6은 소개 추출 DB101에 등록되는 정보의 내용의 일례를 도시하는 도면이다.
- 도 7은 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 소재 추출 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 8은 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 1페이지 대응 추출 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 9는 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 트리 탐색 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 10은 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 콘텐츠 생성 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 11은 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 특유 콘텐츠 블록 판정 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 12는 일 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 Flash 콘텐츠 생성 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 13은 일 실시 형태에 따른 블로그 시스템 BS의 개요 구성의 일례를 도시하는 도면.
- 도 14는 일 실시 형태에 따른 블로그 서버(6)의 개요 구성의 일례를 도시하는 블록도.
- 도 15는 블로거가 지정되고나서 블로그 페이지에 광고 콘텐츠가 삽입될 때까지의 처리의 개요를 도시하는 도면.
- 도 16은 Web 페이지의 구성예를 도시하는 도면.
- 도 17은 HTML 문서로부터 생성된 DOM 트리의 일례를 도시하는 도면.
- 도 18은 기억부(65)에 기억된 콘텐츠 블록 대응 정보의 내용의 일례를 도시하는 도면.
- 도 19는 일 실시 형태에 따른 블로그 서버(6)의 시스템 제어부(70)의 광고 콘텐츠 삽입 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 20은 일 실시 형태에 따른 블로그 서버(6)의 시스템 제어부(70)의 1페이지 대응 추출 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 21은 일 실시 형태에 따른 블로그 서버(6)의 시스템 제어부(70)의 특유 콘텐츠 블록 판정 처리에 있어서의 처리예를 도시하는 플로우차트.
- 도 22는 일 실시 형태의 변형예에 따른 블로그 서버(6)의 시스템 제어부(70)의 블로그 갱신시 처리에 있어서의 처리예를 도시하는 플로우차트.

**발명을 실시하기 위한 구체적인 내용**

[0049]

[1. 제1 실시 형태]

[0050]

이하, 도면을 참조해서 본 발명의 실시 형태에 대해서 상세하게 설명한다. 또한, 이하에 설명하는 실시 형태는, 네트워크 상의 전자적인 교환에 의해 상품의 매매가 행하여지는 쇼핑 시스템에 있어서, 쇼핑 사이트의 Web 페이지로부터 추출된 Web 페이지 특유의 콘텐츠에 기초하여 새로운 콘텐츠를 생성하는 서버 장치에 대하여 본 발명을 적용한 경우의 실시 형태이다.

- [0051] [1-1. 쇼핑 시스템의 구성 및 기능 개요]
- [0052] 우선, 본 실시 형태에 따른 쇼핑 시스템 S의 구성 및 개요 기능에 대해서, 도 1을 이용하여 설명한다.
- [0053] 도 1은, 본 실시 형태에 따른 쇼핑 시스템 S의 개요 구성의 일례를 도시하는 도면이다.
- [0054] 도 1에 도시하는 바와 같이, 쇼핑 시스템 S는, 특유 콘텐츠 판정 장치 및 콘텐츠 생성 장치의 일례로서의 콘텐츠 생성 서버(1)와, 쇼핑 서버(2)와, 관리 단말기(3)와, 복수의 점포 단말기(4)와, 복수의 유저 단말기(5)를 포함해서 구성되어 있다. 그리고, 콘텐츠 생성 서버(1)와, 쇼핑 서버(2)와, 각 점포 단말기(4)와, 각 유저 단말기(5)는, 네트워크 NW를 거쳐서, 예를 들면, 통신 프로토콜에 TCP/IP 등을 이용해서 상호 데이터의 송수신이 가능하게 되어 있다. 또한, 네트워크 NW는, 예를 들면, 인터넷, 전용 통신 회선(예를 들면, CATV(Community Antenna Television) 회선), 이동체 통신망(기지국 등을 포함함), 및 게이트웨이 등에 의해 구축되어 있다. 또한, 콘텐츠 생성 서버(1)와 관리 단말기(3)는, LAN(Local Area Network) 등의 네트워크를 거쳐서 접속되어 있다. 또한, 콘텐츠 생성 서버(1)와 쇼핑 서버(2)가, 마찬가지로 LAN 등의 네트워크를 거쳐서 접속되어도 된다.
- [0055] 이러한 구성의 쇼핑 시스템 S에 있어서, 쇼핑 서버(2)는, 점포 단말기(4)나 유저 단말기(5)로부터의 리퀘스트에 따라서, 쇼핑 사이트를 구성하는 Web 페이지를 송신하는 Web 서버이다. 또한, 쇼핑 서버(2)는, 점포 단말기(4)로부터의 리퀘스트에 기초하여, 쇼핑 사이트에서 판매되는 상품을 등록하고, 그 상품의 상세한 설명 등이 게재되는 상품 상세 페이지를 생성한다. 그리고, 쇼핑 서버(2)는, 상품 상세 페이지 DB201을 구비하고, 생성한 상품 상세 페이지(상품 상세 페이지의 HTML 문서(도큐먼트 데이터의 일례), 상품 상세 페이지의 소재인 화상 데이터 등)를 상품 상세 페이지 DB201에 등록한다. 또한, 쇼핑 서버(2)는, 상품 상세 페이지를 열람한 유저의 유저 단말기(5)로부터의 리퀘스트에 기초하여, 상품의 구입의 처리를 행한다.
- [0056] 점포 단말기(4)는, 쇼핑 사이트에서 상품을 판매하는 점포의 종업원에게 사용되는 단말 장치이다. 점포 단말기(4)로서는, 예를 들면, 퍼스널 컴퓨터 등이 이용된다.
- [0057] 유저 단말기(5)는, 쇼핑 사이트에서 상품을 구입하는 유저에게 사용되는 단말 장치이다. 유저 단말기(5)로서는, 예를 들면, 퍼스널 컴퓨터, PDA, 휴대 전화기 등이 이용된다.
- [0058] 콘텐츠 생성 서버(1)는, 관리 단말기(3)나 점포 단말기(4)로부터의 리퀘스트에 기초하여, 지정된 상품 상세 페이지의 특징을 나타내는(나아가서는, 상품의 특징을 나타내는) Flash 콘텐츠(Adobe Systems사에 의해 규격화된 소프트웨어)를 생성한다. 생성되는 Flash 콘텐츠는, 예를 들면, 상품의 배너 화상, 상품을 소개하는 슬라이드 쇼 콘텐츠, 동화상 콘텐츠 등이다. 그리고, Flash 콘텐츠는, 예를 들면, 점포가 운영하는 Web 사이트에 게재되거나, 쇼핑 사이트를 구성하는 Web 페이지의 소재로서 이용되거나한다.
- [0059] 이러한 Flash 콘텐츠를 생성하기 위해서, 콘텐츠 생성 서버(1)는, 소재 추출 DB101을 구비하고, 상품 상세 페이지 DB201에 등록되어 있는 상품 상세 페이지를 구성하는 Web 소재로서의 콘텐츠(화상 데이터, HTML 문서에 기술되어 있는 텍스트 데이터 등)를 추출하여, 그 추출 결과를 소재 추출 DB101에 등록해 둔다. 그리고, 콘텐츠 생성 서버(1)는, 지정된 상품 상세 페이지로부터 추출된 콘텐츠 중으로부터 그 상품 상세 페이지 특유의 콘텐츠를 특정하고, 특정된 콘텐츠에 기초하여 Flash 콘텐츠를 생성한다.
- [0060] 관리 단말기(3)는, 쇼핑 시스템 S의 시스템 관리자에 의해 사용되는 단말 장치이다. 관리 단말기(3)로서는, 예를 들면, 퍼스널 컴퓨터 등이 이용된다.
- [0061] [1-2. 콘텐츠 생성 서버의 구성 및 기능]
- [0062] 다음으로, 콘텐츠 생성 서버(1)의 구성 및 기능에 대해서, 도 2를 이용하여 설명한다.
- [0063] 도 2는, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 개요 구성의 일례를 도시하는 블록도이다. 또한, 도 3은, Web 페이지가 지정되고나서 Flash 콘텐츠가 생성될 때까지의 처리의 개요를 도시하는 도면이다. 또한, 도 4는, Web 페이지의 구성예를 도시하는 도면이다. 또한, 도 5는, HTML 문서로부터 생성된 DOM 트리의 일례를 도시하는 도면이다. 또한, 도 6은, 소재 추출 DB101에 등록되는 정보의 내용의 일례를 도시하는 도면이다.
- [0064] 도 2에 도시하는 바와 같이, 콘텐츠 생성 서버(1)는, 조작부(11)와, 표시부(12)와, 통신부(13)와, 드라이브부(14)와, 기억 수단의 일례로서의 기억부(15)와, 입출력 인터페이스부(16)와, 시스템 제어부(20)를 구비하고 있다. 그리고, 시스템 제어부(20)와 입출력 인터페이스부(16)는, 시스템 버스(21)를 거쳐서 접속되어 있다.
- [0065] 조작부(11)는, 예를 들면, 키보드, 마우스 등에 의해 구성되어 있고, 시스템 관리자 등으로부터의 조작 지시를

접수하고, 그 지시 내용을 지시 신호로서 시스템 제어부(20)에 출력하게 되어 있다. 표시부(12)는, 예를 들면, CRT(Cathode Ray Tube) 디스플레이, 액정 디스플레이 등에 의해 구성되어 있고, 문자나 화상 등의 정보를 표시하게 되어 있다. 통신부(13)는, 네트워크 NW 등에 접속하여, 쇼핑 서버(2), 관리 단말기(3), 점포 단말기(4), 유저 단말기(5) 등과의 통신 상태를 제어하게 되어 있다. 드라이브부(14)는, 예를 들면, 플렉시블 디스크, CD(Compact Disc), DVD(Digital Versatile Disc) 등의 디스크 DK로부터 데이터 등을 읽어내는 한편, 해당 디스크 DK에 대하여 데이터 등을 기록하게 되어 있다. 기억부(15)는, 예를 들면, 하드디스크 드라이브 등에 의해 구성되어 있고, 각종 프로그램 및 데이터 등을 기억하게 되어 있다. 또한, 기억부(15)에는, 소재 추출 DB101이 구축되어 있다. 입출력 인터페이스부(16)는, 조작부(11)~기억부(15)와 시스템 제어부(20) 사이의 인터페이스 처리를 행하게 되어 있다. 시스템 제어부(20)는, CPU(Central Processing Unit)(17), ROM(Read Only Memory)(18), RAM(Random Access Memory)(19) 등에 의해 구성되어 있다.

[0066] 시스템 제어부(20)는, CPU(17)가, ROM(18)이나 기억부(15)에 기억된 각종 프로그램을 읽어내어 실행함으로써 콘텐츠 생성 서버(1)의 각 부를 제어한다. 또한, 시스템 제어부(20)는, 콘텐츠 생성 소프트웨어(특유 콘텐츠 편집 프로그램의 일례)를 실행함으로써, 추출 수단, 계산 수단, 편집 수단 및 생성 수단으로서 기능한다. 또한, 콘텐츠 생성 소프트웨어 등은, 예를 들면, 다른 서버 장치 등으로부터 네트워크 NW를 거쳐서 취득되도록 해도 되고, CD-ROM 등의 디스크 DK에 기록되어 드라이브부(14)를 거쳐서 읽어들이도록 해도 된다.

[0067] 콘텐츠 생성 소프트웨어는, 상품 상세 페이지 특유의 콘텐츠에 기초하여 Flash 콘텐츠를 생성하기 위한 프로그램이다. 도 3에 도시하는 바와 같이, 콘텐츠 생성 소프트웨어는, 매니저부, 소재 추출 엔진, SWF(ShockWave Flash Object) 생성 엔진 등에 의해 구성되어 있다. 매니저부는, 소재 추출 엔진 및 SWF 엔진의 실행을 제어함과 함께, 콘텐츠 생성 소프트웨어를 이용하는 유저(점포 종업원이나 시스템 관리자)에 대하여, Flash 콘텐츠를 생성하기 위한 GUI(Graphical User Interface)를 제공하기 위한 소프트웨어이다. 소재 추출 엔진은, 상품 상세 페이지의 HTML 문서로부터 Web 소재로서의 콘텐츠를 추출함과 함께, 상품 상세 페이지 특유의 콘텐츠를 편집하기 위한 소프트웨어이다. 콘텐츠의 추출은, 후술하는 콘텐츠 블록(콘텐츠 그룹의 일례)의 단위로 행하여진다. SWF 엔진은, 공급된 하나 또는 복수의 콘텐츠(Web 소재)에 기초하여 Flash 콘텐츠를 생성하기 위한 소프트웨어이다. 또한, 새로운 콘텐츠로서 Flash 콘텐츠 이외의 리치 인터넷 어플리케이션을 생성하는 경우, SWF 생성 엔진 대신에, 예를 들면, 마이크로소프트사의 Silverlight(상표)의 생성 엔진을 적용해도 된다. 또한, Ajax(Asynchronous JavaScript(등록상표)+XML) 등의 기술을 이용해서 동적 페이지를 실현하는 스크립트를 생성하는 소프트웨어를 적용해도 된다.

[0068] 이하에, Flash 콘텐츠의 생성의 개요에 대해서 설명한다. 도 3에 도시하는 바와 같이, 시스템 제어부(20)는, 쇼핑 서버(2)로부터 상품 상세 페이지 DB201에 등록되어 있는 HTML 문서를 취득해서 해석하고, Web 소재인 콘텐츠를 콘텐츠 블록 단위로 추출한다. 그리고, 그 추출 결과로서, 추출한 콘텐츠 블록마다 콘텐츠 블록 대응 정보(콘텐츠 정보의 일례)를 소재 추출 DB101에 등록한다(1). 이 처리는, Flash 콘텐츠의 생성 전에 미리 행해지고, 기본적으로는, 상품 상세 페이지 DB201에 등록되어 있는 전HTML 문서, 즉, 쇼핑 사이트를 구성하는 모든 상품 상세 페이지에 대해서 추출이 행하여진다.

[0069] 그 후, 시스템 관리자 또는 점포 종업원에 의해, Flash 콘텐츠의 생성 대상으로 되는 상품 상세 페이지의 HTML 문서의 URL이 지정 된다(2). 그러면, 시스템 제어부(20)는, 지정된 URL에 기초하여 쇼핑 서버(2)로부터 HTML 문서를 취득하고, 콘텐츠 블록을 추출한다. 그리고, 제어부(20)는, 소재 추출 DB101을 참조하여, 추출한 각 콘텐츠 블록의 전체 상품 상세 페이지에 있어서의 출현 빈도를 계산한다. 계산되는 출현 빈도로서는, 출현 횟수(도수(횟수))이어도 되고, 전체 상품 상세 페이지의 전체 콘텐츠 블록에 대한 출현 횟수의 비율(상대도수(횟수))이어도 된다. 그리고, 시스템 제어부(20)는, 출현 빈도에 기초하여, 지정된 URL에 대응하는 상품 상세 페이지 특유의 콘텐츠 블록을 편집한다. 구체적으로, 시스템 제어부(20)는, 출현 빈도가 가장 작은 콘텐츠 블록을, 지정된 URL에 대응하는 상품 상세 페이지 특유의 콘텐츠 블록이라고 편집한다(3).

[0070] 시스템 제어부(20)는, 특유의 콘텐츠 블록이라고 편집한 콘텐츠 블록에 포함되는 콘텐츠를 쇼핑 서버(2)를 거쳐서 상품 상세 페이지 DB201로부터 취득한다. 시스템 제어부(20)는, 취득한 콘텐츠에 기초하여 Flash 콘텐츠를 생성한다(4). 그리고, 시스템 관리자 또는 점포 종업원은, 생성된 Flash 콘텐츠를 다운로드한다(5). 또한, Flash 콘텐츠의 다운로드 전에, 시스템 관리자나 점포 종업원에 의해 적절히 Flash 콘텐츠를 수정할 수 있도록 해도 된다.

[0071] 다음으로, 콘텐츠 블록의 추출 방법에 대해서 설명한다. 또한, 본 실시 형태에 있어서는, 텍스트 데이터 및 화상 데이터가 Web 소재로서 추출되는 것으로 한다.

- [0072] 예를 들면, 상품 상세 페이지의 구성(레이아웃)이 도 4에 도시하는 바와 같은 것인 것으로 한다. Web 소제로서의 각 콘텐츠는, 상품 상세 페이지 상에 있어서, 각각 어떤 통합체(덩어리)마다 표시되어 있다. 그 각 통합체가 콘텐츠 블록에 상당한다. 각 콘텐츠는, HTML 문서에 기술되어 있는 DIV 태그 및 TABLE 태그(미리 정해진 태그의 일례)에 의해 각각 콘텐츠 블록으로 나뉜다. 즉, 각 콘텐츠는, DIV 태그 및 TABLE 태그에 의해 블록화(그룹화)된다. 또한, DIV 태그 및 TABLE 태그를, 「블록화 태그」라고 한다.
- [0073] 도 4에는, 콘텐츠 블록(301~306)이 표시되어 있다. 콘텐츠 블록(301)은, 예를 들면, 페이지의 헤더 부분의 콘텐츠 블록이며, 텍스트A 및 화상a에 의해 구성되어 있다. 또한, 콘텐츠 블록(302)은, 예를 들면, 각종 카테고리의 상품에 관한 Web 페이지로 이동하기 위한 네비게이션 부분의 콘텐츠 블록이며, 예를 들면, 다른 Web 페이지의 링크를 나타내는 텍스트B, 텍스트C 및 텍스트D에 의해 구성되어 있다. 또한, 콘텐츠 블록(303)은, 예를 들면, 상품에 관한 정보가 표시되는 콘텐츠 블록이며, 상품명 등의 표제를 나타내는 텍스트E, 콘텐츠 블록(304) 및 콘텐츠 블록(305)에 의해 구성되어 있다. 이와 같이, 콘텐츠 블록이 상자 형상, 즉, 계층 구조로 되어 있는 경우도 있다. 이 경우, 콘텐츠 블록(303)에 포함되는 콘텐츠는 텍스트E만으로 되고, 콘텐츠 블록(304) 및 콘텐츠 블록(305)은, 콘텐츠 블록(303)으로부터 독립되어 있는 것으로 된다. 콘텐츠 블록(304)은, 예를 들면, 상품의 상세를 나타내는 콘텐츠 블록이며, 상세한 설명을 나타내는 텍스트F, 상품 등의 화상으로서의 화상b 및 화상c에 의해 구성되어 있다. 콘텐츠 블록(305)은, 예를 들면, 상품 구입 시의 일반적인 주의 사항 등을 나타내는 콘텐츠 블록이며, 텍스트G 및 텍스트H에 의해 구성되어 있다. 콘텐츠 블록(306)은, 예를 들면, 카피라이트 표시를 나타내는 콘텐츠 블록이며, 텍스트I에 의해 구성되어 있다.
- [0074] 이들의 콘텐츠 블록 중, 콘텐츠 블록(301, 302, 305 및 306)은, 도 4에 도시하는 상품 상세 페이지 이외의 상품 상세 페이지 상에서도 비교적 빈번하게 나타난다. 한편, 콘텐츠 블록(303)(텍스트E) 및 콘텐츠 블록(304)은, 해당 상품 상세 페이지에만 이용되거나, 또는, 다른 상품 상세 페이지에서 이용된다고 해도, 그 빈도는 콘텐츠 블록(301, 302, 305 및 306)에 비해서 작다. 따라서, 예를 들면, 콘텐츠 블록(303) 또는 콘텐츠 블록(304)이, 해당 상품 상세 페이지 특유의 콘텐츠 블록이라고 판단되게 된다.
- [0075] 도 4에 도시하는 상품 상세 페이지의 HTML 문서를 DOM(Document Object Model)트리, 즉, 트리 구조로 도시한 것이 도 5이다. 또한, 도 5에 도시하는 DOM 트리에 있어서, 본 실시 형태의 설명에 필요 없는 태그의 노드의 도시는 생략하고 있다.
- [0076] DOM 트리에 있어서는, DIV 태그를 나타내는 DIV 노드 및 TABLE 태그를 나타내는 TABLE 노드가, 각 콘텐츠를 콘텐츠 블록으로 블록화하는 노드로 된다(이하, 「블록화 노드」라고 함). 시스템 제어부(20)는, 예를 들면, 깊이 우선 탐색에 의해 DOM 트리를 탐색하고, 콘텐츠 블록을 확정해 간다. 구체적으로, 시스템 제어부(20)는, 블록화 노드를 발견하면, 해당 노드를 정점으로 하는 부분 나무의 각 노드에 규정되어 있는 콘텐츠를 일괄해서 콘텐츠 블록으로 한다. 단, 콘텐츠 블록이 계층적으로 규정된 결과, 어떤 블록화 노드(이하, 「상위의 블록화 노드」라고 함)를 발견한 후, 해당 블록화 노드의 자손의 노드 중에서 더 블록화 노드(이하, 「하위의 블록화 노드」라고 함)를 발견하면, 콘텐츠 블록이 분할된다. 예를 들면, 루트 노드로부터의 거리가 가까운 노드일수록 계층적으로 보다 상위의 노드라고 하면, 상위의 블록화 노드를 정점으로 하는 부분 나무(이하, 「상위 부분 나무」라고 함)에 상당하는 콘텐츠 블록을, 하위의 블록화 노드를 정점으로 하는 부분 나무(이하, 「하위 부분 나무」라고 함)에 상당하는 콘텐츠 블록과, 상위 부분 나무 중 하위 부분 나무를 제외한 부분에 상당하는 콘텐츠 블록으로 나눈다(예를 들면, 콘텐츠 블록(304)과 콘텐츠 블록(303)). 이 경우에 있어서는, 전자의 콘텐츠 블록을, 후자의 콘텐츠 블록보다도 계층적으로 하위에 있는 것으로 한다. 예를 들면, 콘텐츠 블록(301, 302, 303 및 306)의 계층은 1이며, 콘텐츠 블록(304 및 305)의 계층은 2이다. 즉, 계층의 값이 작을수록 계층적으로 보다 상위로 된다.
- [0077] 이것을, HTML 문서의 태그의 기술로 설명하면, 기본적으로는 블록화 태그가 기술되어 있으면, 블록화 태그 사이에 끼워져 있던 범위 내에 규정이 기술되어 있는 콘텐츠가, 통합해서 해당 블록화 태그에 대응하는 콘텐츠 블록으로 된다. 단, 블록화 태그가 계층적으로 기술되어 있는 경우, 어떤 블록화 태그 사이에 끼워져 있던 범위 내에 규정이 기술되어 있는 콘텐츠 중, 해당 블록화 태그보다도 하위에 기술되어 있는 블록화 태그 사이에 끼워져 있던 범위에 규정이 기술되어 있는 콘텐츠를 제외한 것이, 통합해서 해당 블록화 태그에 대응하는 콘텐츠 블록으로 된다.
- [0078] 이와 같이 하여 콘텐츠 블록을 추출하면, 시스템 제어부(20)는, 추출 결과를 나타내는 콘텐츠 블록 대응 정보를 소재 추출 DB101에 등록한다. 도 6에 도시하는 바와 같이, 콘텐츠 블록 대응 정보(부호 401)는, 콘텐츠 블록마다 등록된다. 콘텐츠 블록 대응 정보에는, 추출원의 HTML 문서의 URL 설정 부분(부호 402)과 블록 구성 정보

(부호 403)에 의해 구성되어 있다. 블록 구성 정보에는, 추출된 각 콘텐츠가 설정된다. 여기서, 텍스트 데이터에 대해서는, DOM 트리에 있어서의 텍스트 노드의 내용이 설정된다. 한편, 화상 데이터에 대해서는, 화상 데이터 그 자체의 대신으로서, DOM 트리에 있어서 IMG 태그를 나타내는 IMG 노드의 src 속성으로서의 화상 데이터의 URL이 설정된다.

- [0079] [1-3. 쇼핑 시스템의 동작]
- [0080] 다음으로, 쇼핑 시스템 S의 동작에 대해서, 도 7 내지 도 12를 이용하여 설명한다.
- [0081] [1-3-1 전체 상품 상세 페이지로부터의 콘텐츠 블록 추출시]
- [0082] 도 7은, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 소재 추출 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0083] 소재 추출 처리는, 예를 들면, 정기적으로 실행되거나, 시스템 관리자의 조작에 기초하여 관리 단말기(3)로부터 소재 추출 처리의 실행의 리퀘스트가 송신되어 왔을 때에 개시된다. 소재 추출 처리에 있어서, 시스템 제어부(20)는, 상품 상세 페이지 DB201에 등록되어 있는 모든 HTML 문서를 해석한다. 그 때문에, 시스템 제어부(20)는, 예를 들면, 상품 상세 페이지 DB201에 등록되어 있는 HTML 문서의 일람의 정보를 미리 쇼핑 서버(2)로부터 취득해 두고, 이 일람의 정보에 기초하여 HTML 문서를 취득해도 되고, 쇼핑 사이트의 톱 페이지의 HTML 문서로부터 차례로 링크를 찾아가서, 상품 상세 페이지의 HTML 문서를 순차적으로 취득해도 된다.
- [0084] 도 7에 도시하는 바와 같이, 시스템 제어부(20)는, 소재 추출 DB101을 초기화한다(스텝 S1). 구체적으로, 시스템 제어부(20)는, 소재 추출 DB101에 콘텐츠 블록 대응 정보가 등록되어 있는 경우에는, 모든 콘텐츠 블록 대응 정보를 소재 추출 DB101로부터 삭제한다.
- [0085] 다음으로, 시스템 제어부(20)는, 전체 상품 상세 페이지 중 최초로 취득해야 할 상품 상세 페이지의 HTML 문서의 URL을 특정하고(스텝 S2), 특정한 URL을 설정한 리퀘스트를 쇼핑 서버(2)에 송신함으로써, 해당 쇼핑 서버(2)로부터 HTML 문서를 취득한다(스텝 S3). 다음으로, 시스템 제어부(20)는, 취득한 HTML 문서를 지정해서, 후술하는 1페이지 대응 추출 처리를 실행한다(스텝 S4). 이 1페이지 대응 추출 처리에서는, 취득한 HTML 문서로부터 콘텐츠 블록이 추출되고, 콘텐츠 블록 대응 정보가 등록된다.
- [0086] 다음으로, 시스템 제어부(20)는, 모든 상품 상세 페이지의 콘텐츠 블록을 추출하였는지의 여부를 판정한다(스텝 S5). 이때, 시스템 제어부(20)는, 콘텐츠 블록을 추출하지 않고 있는 상품 상세 페이지가 존재하는 경우에는(스텝 S5:NO), 다음의 상품 상세 페이지의 HTML 문서의 URL을 특정하고(스텝 S6), 스텝 S3으로 이행한다. 그리고, 시스템 제어부(20)는, 스텝 S3~S6의 처리를 반복해서 모든 상품 상세 페이지의 콘텐츠 블록을 추출하면(스텝 S5:YES), 소재 추출 처리를 종료시킨다.
- [0087] 또한, 시스템 제어부(20)는, 소재 추출 DB101을 초기화해서 콘텐츠 블록 대응 정보를 재등록하지 않아도 된다. 예를 들면, 시스템 제어부(20)는, 소재 추출 DB101을 초기화를 행하지 않고, 전회 소재 추출 처리를 실행한 후에 신규 생성된 상품 상세 페이지에 대해서 콘텐츠 블록 대응 정보를 생성해서 소재 추출 DB101에 추가 등록하고, 또한, 전회 소재 추출 처리를 실행한 후에 갱신된 상품 상세 페이지에 대해서 콘텐츠 블록 대응 정보를 생성해서 소재 추출 DB101에 갱신 등록해도 된다.
- [0088] 도 8은, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 1페이지 대응 추출 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0089] 도 8에 도시하는 바와 같이, 시스템 제어부(20)는, 우선, 취득한 HTML 문서의 DOM 트리를 RAM(19) 상에 생성한다(스텝 S21).
- [0090] 다음으로, 시스템 제어부(20)는, 블록수 NUM에 0을 설정함과 함께, 계층 LV에 0을 설정한다(스텝 S22). 블록수 NUM은, 현시점에서 발견 완료된 콘텐츠 블록의 개수이다. 또한, 계층 LV는, DOM 트리에 있어서 현재 탐색 중인 노드가 속하는 콘텐츠 블록의 계층이다. NUM 및 LV는 모두 글로벌 변수이며, 1페이지 대응 추출 처리 및 후술하는 트리 탐색 처리로부터 액세스가 가능하다.
- [0091] 다음으로, 시스템 제어부(20)는, DOM 트리의 루트 노드를 지정해서(스텝 S23), 트리 탐색 처리를 실행한다(스텝 S24). 트리 탐색 처리는 재귀 호출이 가능하며, 이 트리 탐색 처리에 의해, Web 페이지로부터 모든 콘텐츠 블록이 추출되고, 콘텐츠 블록 대응 정보가 생성된다.
- [0092] 다음으로, 시스템 제어부(20)는, 트리 탐색 처리에 의해 생성된 각 콘텐츠 블록 대응 정보를 소재 추출 DB101에

등록한다(스텝 S25). 시스템 제어부(20)는, 이 처리를 끝내면, 1페이지 대응 추출 처리를 종료시킨다.

- [0093] 도 9는, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 트리 탐색 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0094] 도 9에 도시하는 바와 같이, 시스템 제어부(20)는, 우선, 지정된 노드의 종류를 판정한다(스텝 S31). 이때, 시스템 제어부(20)는, 지정된 노드의 종류가, DIV 노드 또는 TABLE 노드(블록화 노드)인 경우에는, 즉, 콘텐츠 블록이 발견된 경우에는(스텝 S31:DIV 또는 TABLE), 스텝 S32로 이행한다.
- [0095] 스텝 S32에 있어서, 시스템 제어부(20)는, 블록수 NUM에 1을 가산함과 함께, 계층 LV에 1을 가산한다. 다음으로, 시스템 제어부(20)는, 블록 번호 BN[LV]에 NUM을 설정한다(스텝 S33). 블록 번호 BN[LV]은, 현재 탐색 중인 노드가 속하는 계층 LV에서 나타내지는 콘텐츠 블록의 블록 번호이다. 이 블록 번호는, 콘텐츠 블록의 발견 순으로 부여된다. 또한, BN[LV]은, 글로벌 변수이다.
- [0096] 다음으로, 시스템 제어부(20)는, 블록 번호 BN[LV]의 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보를 초기화한다(스텝 S34). 구체적으로, 시스템 제어부(20)는, 콘텐츠 블록 대응 정보를 저장하는 영역을 RAM(19) 상에 설정하고, 취득한 HTML 문서의 URL을, 해당 영역에 설정한다.
- [0097] 다음으로, 시스템 제어부(20)는, 지정된 노드의 자 노드 중, 아직 탐색되어 있지 않은 자 노드가 존재하는지의 여부를 판정한다(스텝 S35). 이때, 시스템 제어부(20)는, 아직 탐색되어 있지 않은 자 노드가 존재하는 경우에는(스텝 S35:YES), 스텝 S36으로 이행한다.
- [0098] 스텝 S36에 있어서, 시스템 제어부(20)는, 탐색되어 있지 않은 자 노드 중 1개의 자 노드를 지정해서, 트리 탐색 처리를 실행한다(스텝 S37). 시스템 제어부(20)는, 트리 탐색 처리를 끝내면, 스텝 S35로 이행한다.
- [0099] 그리고, 시스템 제어부(20)는, 스텝 S35~S37의 처리를 반복해서 모든 자 노드의 트리 탐색 처리를 끝내면(스텝 S35:NO), 스텝 S38로 이행한다. 또한, 시스템 제어부(20)는, 지정된 노드의 자 노드가 1개도 존재하지 않는 경우에도, 스텝 S38로 이행한다. 스텝 S38에 있어서, 시스템 제어부(20)는, 계층 LV로부터 1을 감산하여, 트리 탐색 처리를 종료시킨다.
- [0100] 스텝 S31에 있어서, 시스템 제어부(20)는, 지정된 노드의 종류가 텍스트 노드인 경우에는(스텝 S31:텍스트), 지정된 노드의 내용(텍스트 데이터)을, 블록 번호 BN[LV]의 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보 중의 블록 구성 정보에 추가 설정한다(스텝 S39). 시스템 제어부(20)는, 이 처리를 끝내면, 트리 탐색 처리를 종료시킨다.
- [0101] 스텝 S31에 있어서, 시스템 제어부(20)는, 지정된 노드의 종류가 IMG 노드인 경우에는(스텝 S31:IMG), 지정된 노드의 src 속성으로서 설정되어 있는 화상 데이터의 URL을 취득하고, 취득한 URL을, 블록 번호 BN[LV]의 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보 중의 블록 구성 정보에 추가 설정한다(스텝 S40). 시스템 제어부(20)는, 이 처리를 끝내면, 트리 탐색 처리를 종료시킨다.
- [0102] 스텝 S31에 있어서, 시스템 제어부(20)는, 지정된 노드의 종류가, DIV 노드, TABLE 노드, 텍스트 노드 및 IMG 노드의 어느 쪽도 아닌 경우에는(스텝 S31: 기타), 지정된 노드의 자 노드 중, 아직 탐색되어 있지 않은 자 노드가 존재하는지의 여부를 판정한다(스텝 S41). 이때, 시스템 제어부(20)는, 아직 탐색되어 있지 않은 자 노드가 존재하는 경우에는(스텝 S41:YES), 탐색되어 있지 않은 자 노드 중 하나의 자 노드를 지정해서(스텝 S42), 트리 탐색 처리를 실행한다(스텝 S43). 시스템 제어부(20)는, 트리 탐색 처리를 끝내면, 스텝 S41로 이행한다.
- [0103] 한편, 시스템 제어부(20)는, 지정된 노드의 모든 자 노드의 트리 탐색 처리를 끝낸 경우, 또는, 지정된 노드의 자 노드가 1개도 존재하지 않는 경우에는(스텝 S41:NO), 트리 탐색 처리를 종료시킨다.
- [0104] [1-3-2 Flash 콘텐츠 생성시]
- [0105] 도 10은, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 콘텐츠 생성 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0106] 콘텐츠 생성 처리는, 시스템 관리자의 조작에 기초하여 관리 단말기(3)로부터 콘텐츠 생성 처리의 실행의 리퀘스트가 송신되어 왔을 때, 또는, 점포 종업원의 조작에 기초하여 점포 단말기(4)로부터 콘텐츠 생성 처리의 실행의 리퀘스트가 송신되어 왔을 때에 개시된다.
- [0107] 그리고, 시스템 관리자 또는 점포 종업원이 Flash 콘텐츠의 생성 대상으로 하는 HTML 문서의 URL을 지정하면,

도 10에 도시하는 바와 같이, 시스템 제어부(20)는, 지정된 URL을 관리 단말기(3) 또는 점포 단말기(4)로부터 수신한다(스텝 S51). 다음으로, 시스템 제어부(20)는, 수신한 URL을 설정한 리퀘스트를 쇼핑 서버(2)에 송신함으로써, 해당 쇼핑 서버(2)로부터 HTML 문서를 취득한다(스텝 S52).

[0108] 다음으로, 시스템 제어부(20)는, 취득한 HTML 문서를 지정해서, 후술하는 특유 콘텐츠 블록 판정 처리를 실행한다(스텝 S53). 이 특유 콘텐츠 블록 판정 처리에서는, 취득한 HTML 문서로부터 콘텐츠 블록이 추출되고, 해당 HTML 문서 특유의 콘텐츠 블록이 판정된다.

[0109] 다음으로, 시스템 제어부(20)는, 특유라고 판정된 콘텐츠 블록을 구성하는 각 콘텐츠를 취득한다(스텝 S54). 이때, 시스템 제어부(20)는, 텍스트 데이터를 취득하는 경우에는, 특유라고 판정된 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보로부터 텍스트 데이터를 취득한다. 한편, 시스템 제어부(20)는, 화상 데이터를 취득하는 경우에는, 특유라고 판정된 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보로부터 화상 데이터의 URL을 취득하고, 취득된 URL을 설정한 리퀘스트를 쇼핑 서버(2)에 송신함으로써, 상품 상세 페이지 DB201에 등록되어 있는 화상 데이터를 쇼핑 서버(2)로부터 취득한다.

[0110] 다음으로, 시스템 제어부(20)는, 취득한 모든 콘텐츠를 지정해서, 후술하는 Flash 콘텐츠 생성 처리를 실행한다(스텝 S55).

[0111] 다음으로, 시스템 제어부(20)는, Flash 콘텐츠 생성 처리에 있어서 생성된 Flash 콘텐츠를, 생성 요구원의 관리 단말기(3) 또는 점포 단말기(4)에 송신한다(스텝 S56). 시스템 제어부(20)는, 이 처리를 끝내면, 콘텐츠 생성 처리를 종료시킨다.

[0112] 도 11은, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 특유 콘텐츠 블록 판정 처리에 있어서의 처리예를 도시하는 플로우차트이다.

[0113] 도 11에 도시하는 바와 같이, 시스템 제어부(20)는, 우선, 1페이지 대응 추출 처리와 마찬가지로, 취득한 HTML 문서의 DOM 트리 생성(스텝 S61), 블록수 NUM 및 계층 LV에 대하여 0의 설정을 행하고(스텝 S62), DOM 트리의 루트 노드를 지정해서(스텝 S63), 트리 탐색 처리를 실행한다(스텝 S64).

[0114] 다음으로, 시스템 제어부(20)는, 블록 번호i에 1을 설정한다(스텝 S65). 다음으로, 시스템 제어부(20)는, 블록 번호i의 콘텐츠 블록의 출현 빈도를 계산한다(스텝 S66).

[0115] 구체적으로, 시스템 제어부(20)는, 트리 탐색 처리에 있어서 생성된 콘텐츠 블록 대응 정보i(블록 번호i의 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보)의 블록 구성 정보와, 소재 추출 DB101에 등록되어 있는 각 콘텐츠 블록 대응 정보의 블록 구성 정보를 비교한다. 이때, 시스템 제어부(20)는, 블록 구성 정보의 내용이 일치하는 경우에는, 출현 횟수 1회로서 카운트한다. 이때, 시스템 제어부(20)는, 블록 구성 정보 중에 있어서의 콘텐츠의 규정 순서는 무시해도 상관없다. 또한, 시스템 제어부(20)는, 소재 추출 DB101에 등록되어 있는 콘텐츠 블록 대응 정보의 블록 구성 정보에 규정되어 있는 일부의 콘텐츠가 콘텐츠 블록 대응 정보i의 블록 구성 정보에 규정되어 있는 전부의 콘텐츠와 일치하는 경우도, 출현 횟수 1회로서 카운트해도 된다. 또한, 시스템 제어부(20)는, 콘텐츠 블록 대응 정보의 블록 구성 정보 중에 규정되어 있는 텍스트 데이터끼리를 비교하는 경우에는, 텍스트 데이터가 나타내는 문장 등 그 자체가 일치하는지의 여부를 판정하는 것은 아니고, 그 문장 등에 의해 표현되어 있는 실질적인 내용을 비교해도 된다. 예를 들면, 시스템 제어부(20)는, 각각의 텍스트 데이터의 형태소 해석 등을 행함으로써 텍스트 데이터로부터 단어를 추출하고, 추출한 단어끼리를 비교해도 된다. 그리고, 시스템 제어부(20)는, 모든 단어가 일치한 경우에 텍스트 데이터끼리가 일치하였다고 판단해도 되고, 소정의 비율 이상으로 단어가 일치한 경우에 텍스트 데이터끼리가 일치하였다고 판단해도 된다. 시스템 제어부(20)는, 이와 같이 해서 콘텐츠 블록 대응 정보i의 블록 구성 정보와, 소재 추출 DB101에 등록되어 있는 모든 콘텐츠 블록 대응 정보의 블록 구성 정보를 비교하여, 출현 빈도를 계산한다.

[0116] 시스템 제어부(20)는, 출현 빈도를 계산하면, 블록 번호i에 1을 가산해서(스텝 S67), 블록 번호i가 블록수 NUM의 값보다 큰지의 여부를 판정한다(스텝 S68). 이때, 시스템 제어부(20)는, 블록 번호i가 블록수 NUM의 값 이하인 경우에는(스텝 S68:NO), 스텝 S66으로 이행한다. 그리고, 시스템 제어부(20)는, 트리 탐색 처리에 있어서 추출된 모든 콘텐츠 블록의 출현 빈도를 계산하면(스텝 S68:YES), 스텝 S69로 이행한다.

[0117] 스텝 S69에 있어서, 시스템 제어부(20)는, 콘텐츠 블록1로부터 블록수 NUM이 나타내는 블록 번호까지의 모든 콘텐츠 블록의 출현 빈도끼리를 비교하여, 출현 빈도가 가장 적은 콘텐츠 블록을, 특유의 콘텐츠 블록이라고 판정한다(스텝 S69). 시스템 제어부(20)는, 이 처리를 끝내면, 특유 콘텐츠 블록 판정 처리를 종료시킨다.

- [0118] 도 12는, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(20)의 Flash 콘텐츠 생성 처리에 있어서의 처리예를 도시하는 플로우차트이다. 또한, 본 실시 형태에 있어서는, 상품 상세 페이지에 게재되어 있는 상품을 소개하는 콘텐츠로서, 특유의 콘텐츠 블록이라고 판정된 콘텐츠 블록에 포함되는 각 콘텐츠를 슬라이드쇼 표시하는 Flash 콘텐츠를 생성하는 것으로서 이하 설명한다.
- [0119] 도 12에 도시하는 바와 같이, 시스템 제어부(20)는, 우선, 지정된 각 콘텐츠의 표시 사이즈를 조정한다(스텝 S71). 예를 들면, 시스템 제어부(20)는, Flash 콘텐츠 재생시의 실제의 표시 사이즈에 맞도록, 화상 데이터의 종횡의 화소수를 조정하거나, 텍스트 데이터의 폰트 사이즈 등을 조정한다. 또한, 시스템 제어부(20)는, 콘텐츠의 표시 사이즈가 Flash 콘텐츠 재생시의 실제의 표시 사이즈에 비해서 지나치게 클 경우에는, 해당 콘텐츠를 복수로 분할한다. 또한, 시스템 제어부(20)는, 콘텐츠의 표시 사이즈가 Flash 콘텐츠 재생시의 실제의 표시 사이즈에 비해서 지나치게 작을 경우에는, 복수의 콘텐츠를 1개로 결합한다.
- [0120] 다음으로, 시스템 제어부(20)는, 각 콘텐츠의 표시 순서를 결정한다(스텝 S72). 각 콘텐츠의 표시 순서는, 기본적으로는 트리 탐색 처리에 있어서의 콘텐츠 블록 대응 정보에 대한 콘텐츠의 설정 순서와 동일한 순서이다. 즉, HTML 문서에 있어서 해당 문서의 선두의 근처에 규정되어 있는 콘텐츠일수록, 표시 순서가 빨라진다.
- [0121] 다음으로, 시스템 제어부(20)는, 각 콘텐츠의 천이 방법을 결정한다(스텝 S73). 즉, 시스템 제어부(20)는, 슬라이드쇼 표시에 있어서, 표시하는 콘텐츠를 전환할 때에 실시되는 이펙트(표시 효과)를 결정한다. 이펙트로서는, 예를 들면, 페이드 인/페이드 아웃, 슬라이드, 랜덤 블럭, 와이프, 이펙트 없음 등이 있다.
- [0122] 다음으로, 시스템 제어부(20)는, 스텝 S71에 있어서 조정이 행하여진 각 콘텐츠에 기초하여, 스텝 S72 및 S73에 있어서 결정된 조건에 기초하여 Flash 콘텐츠를 생성한다(스텝 S74). 시스템 제어부(20)는, 이 처리를 끝내면, Flash 콘텐츠 생성 처리를 종료시킨다.
- [0123] 이상 설명한 바와 같이, 본 실시 형태에 따르면, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 지정된 URL에 대응하는 상품 상세 페이지를 구성하고 있는 콘텐츠를 추출하고, 지정된 URL에 대응하는 상품 상세 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하고, 지정된 URL에 대응하는 상품 상세 페이지를 구성하는 콘텐츠 중, 출현 빈도가 가장 작은 콘텐츠를 해당 상품 상세 페이지 특유의 콘텐츠라고 판단한다.
- [0124] 따라서, 출현 빈도가 작은 콘텐츠일수록, 지정된 상품 상세 페이지 이외에는 그다지 출현하지 않는 콘텐츠이므로, 가장 출현 빈도가 작은 콘텐츠를 판단함으로써, 지정된 상품 상세 페이지 특유의 콘텐츠가 특정된다. 따라서, 상품 상세 페이지 특유의 콘텐츠를 용이하게 추출할 수 있다.
- [0125] 또한, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 상품 상세 페이지 특유의 콘텐츠라고 판단된 콘텐츠에 기초하여 Flash 콘텐츠를 생성한다.
- [0126] 따라서, 해당 상품 상세 페이지에 게재되어 있는 상품의 내용 등을 단적으로 나타내거나 또는 특징을 나타내는 콘텐츠를 생성할 수 있다.
- [0127] 또한, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 쇼핑 사이트에 포함되는 복수의 상품 상세 페이지 상에 있어서의 각 콘텐츠의 출현 빈도를 계산한다.
- [0128] 따라서, 쇼핑 사이트에 포함되는 복수의 Web 페이지 상에 있어서, 지정된 상품 상세 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도가 계산되므로, 쇼핑 사이트 내에서 공통해서 이용되는 콘텐츠는, 특유의 콘텐츠가 아니라고 판정하는 것이 가능하게 되어, 판단 정밀도를 높일 수 있다.
- [0129] 또한, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 쇼핑 사이트를 구성하는 모든 상품 상세 페이지에 대해서, 상품 상세 페이지를 구성하고 있는 콘텐츠를 추출하고, 추출한 콘텐츠를 나타내는 콘텐츠 블록 대응 정보를 미리 소재 추출 DB101에 등록해 두고, 소재 추출 DB101에 등록된 각 콘텐츠 블록 대응 정보에 기초하여, 지정된 URL에 대응하는 상품 상세 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산한다.
- [0130] 따라서, 모든 상품 상세 페이지에 기초하여 출현 빈도가 계산되므로, 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0131] 또한, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 1개 이상의 콘텐츠로 구성되는 콘텐츠 블록의 단위로, 상품 상세 페이지를 구성하고 있는 콘텐츠를 추출하고, 지정된 URL에 대응하는 상품 상세 페이지를 구성하고 있는 각 콘텐츠 블록 출현 빈도를 계산하고, 지정된 URL에 대응하는 상품 상세 페이지를 구성하는 콘텐츠 블록 중, 출현 빈도가 가장 작은 콘텐츠 블록을 해당 상품 상세 페이지 특유의 콘텐츠 블록이라고 판단한다.

- [0132] 따라서, 상품 상세 페이지 상에 있어서, 예를 들면, 헤더 부분, 네비게이션 부분, 상품의 상세를 나타내는 부분, 상품 구입 시의 일반적인 주의 사항 등을 나타내는 부분, 카피라이트 표시의 부분 등과 같이, 1개 이상의 콘텐츠가 통합체로써 콘텐츠 블록으로서 표시되어 있는 경우에, 상품 상세 페이지 특유의 콘텐츠 블록을 추출할 수 있다.
- [0133] 또한, 콘텐츠 생성 서버(1)의 시스템 제어부(20)가, 상품 상세 페이지를 구성하고 있는 콘텐츠를 해당 상품 상세 페이지의 HTML 문서에 기초하여 추출하고, HTML 문서에 있어서 DIV 태그 또는 TABLE 태그에 기초하여 콘텐츠 블록을 정한다.
- [0134] 따라서, DIV 태그에 의해, HTML 문서의 작성 시에 명시적으로 블록화된 1개 이상의 콘텐츠를 특정할 수 있고, 또한, TABLE 태그에 의해, 표 형식으로 블록화되어 표시되는 1개 이상의 콘텐츠를 특정할 수 있으므로, 예를 들면, 이들의 태그에 의해, 상품 상세 페이지 특유의 콘텐츠와, 특유하지 않은 콘텐츠가 블록화되어 있는 경우에, Web 페이지 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0135] 또한, 상기 제1 실시 형태에 있어서는, 지정된 상품 상세 페이지를 구성하는 각 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보를, 소재 추출 DB101에 등록되어 있는 모든 콘텐츠 블록 대응 정보와 비교함으로써 각 출현 빈도가 계산되어 있었다. 즉, 지정된 상품 상세 페이지를 구성하는 각 콘텐츠 블록의 출현 빈도를 계산하는 경우에, 쇼핑 사이트에 포함되는 모든 상품 상세 페이지를 대상으로 한 범위에 출현하는 빈도를 계산하게 되어 있었지만, 모든 상품 상세 페이지를 대상으로 하지 않아도 된다. 예를 들면, 대상으로 되는 점포의 지정을 가능하게 하고, 지정된 점포에 대응하는 모든 상품 상세 페이지를 대상으로 한 범위에 출현하는 빈도를 계산하도록 해도 된다. 또한, 예를 들면, 미리 정해진 페이지수 분의 상품 상세 페이지를 대상으로 해도 된다.
- [0136] 또한, 미리 각 상품 상세 페이지로부터 콘텐츠 블록의 추출해 두지 않아도 된다. 예를 들면, Flash 콘텐츠 생성시에, 출현 빈도를 계산하는 데에 필요한 각 상품 상세 페이지에 대해서 콘텐츠 블록의 추출을 행하면 된다.
- [0137] 또한, 상기 제1 실시 형태에 있어서는, 시스템 관리자 또는 점포 종업원으로부터 상품 상세 페이지의 HTML 문서의 URL을 지정받았을 때에, Flash 콘텐츠를 생성하도록 하고 있었지만, 예를 들면, 신규인 상품 상세 페이지가 작성되었을 때나, 상품 상세 페이지가 갱신되었을 때에, 신규로 작성되거나 또는 갱신된 상품 상세 페이지에 관한 Flash 콘텐츠를 생성해도 된다.
- [0138] 또한, 상기 제1 실시 형태에 있어서는, 출현 빈도가 가장 작은 콘텐츠 블록만을 Web 페이지 특유의 콘텐츠로 하고 있었지만, 예를 들면, 출현 빈도가 1번째로 작은 콘텐츠 블록으로부터 N번째(N은 2 이상의 자연수)로 작은 콘텐츠까지의 N개의 콘텐츠 블록을 Web 페이지 특유의 콘텐츠 블록으로 해도 된다. 이것은, 예를 들면, 원하는 처리에 필요한 콘텐츠 블록의 개수가 2개 이상으로 미리 정해져 있는 경우에 적용할 수 있다. 또한, 예를 들면, 원하는 처리에 필요한 콘텐츠(콘텐츠 블록이 아님)의 개수가 2개 이상으로 미리 정해져 있는 경우에, 출현 빈도가 1번째로 작은 콘텐츠 블록에 포함되는 콘텐츠만으로는 충분하지 않을 때에는, 출현 빈도가 2번째로 작은 콘텐츠 블록을 Web 페이지 특유의 콘텐츠라고 추가 인정하고, 출현 빈도가 1번째로 작은 콘텐츠 블록과 2번째로 작은 콘텐츠 블록에 포함되는 콘텐츠만으로도 충분하지 않을 때에는, 출현 빈도가 3번째로 작은 콘텐츠 블록을 Web 페이지 특유의 콘텐츠라고 추가 인정하는 등의 처리를 행해도 된다.
- [0139] 또한, 상기 제1 실시 형태에 있어서는, Web 페이지 특유의 콘텐츠를 이용해서 Flash 콘텐츠를 생성하고 있었지만, Flash 콘텐츠 이외의 콘텐츠(예를 들면, 동화상 데이터, 정지 화상 데이터, 전자 문서 등)를 생성해도 된다.
- [0140] 또한, Web 페이지 특유의 콘텐츠의 용도로서는 새로운 콘텐츠의 생성만에 한정되는 것은 아니다. 예를 들면, Web 페이지 특유의 화상 데이터를 판정하고, 특유의 화상 데이터라고 판정된 화상 데이터를, 해당 Web 페이지를 대표하는 화상 데이터로서, 검색 결과 등에 표시해도 된다.
- [0141] [2. 제2 실시 형태]
- [0142] 이하, 도면을 참조해서 본 발명의 실시 형태에 대해서 상세하게 설명한다. 또한, 이하에 설명하는 실시 형태는, 블로그 서비스를 제공하는 블로그 시스템에 있어서, 블로그 페이지를 송신하는 서버 장치에 대하여 본 발명을 적용한 경우의 실시 형태이다.
- [0143] [2-1. 블로그 시스템의 구성 및 기능 개요]
- [0144] 우선, 본 실시 형태에 따른 블로그 시스템 BS의 구성 및 개요 기능에 대해서, 도 13을 이용하여 설명한다.

- [0145] 도 13은, 본 실시 형태에 따른 블로그 시스템 BS의 개요 구성의 일례를 도시하는 도면이다. 도 13에 있어서, 도 1과 마찬가지로의 요소에 대해서는 마찬가지로의 부호를 붙이고 있다.
- [0146] 도 13에 도시하는 바와 같이, 블로그 시스템 BS는, 특유 콘텐츠 판정 장치 및 관련 콘텐츠 삽입 장치의 일례로서의 블로그 서버(6)와, 관리 단말기(3)와, 복수의 유저 단말기(5)를 포함해서 구성되어 있다. 그리고, 블로그 서버(6)와, 각 유저 단말기(5)는, 네트워크 NW를 거쳐서, 예를 들면, 통신 프로토콜에 TCP/IP 등을 이용해서 상호 데이터의 송수신이 가능하게 되어 있다. 또한, 블로그 서버(6)와 관리 단말기(3)는, LAN 등의 네트워크를 거쳐서 접속되어 있다.
- [0147] 이러한 구성의 블로그 시스템 BS에 있어서, 블로그 서버(6)는, 유저 단말기(5)로부터의 리퀘스트에 따라서, 블로그 서비스 사이트를 구성하는 Web 페이지를 송신하는 Web 서버이다. 유저 단말기(5)를 이용하는 유저가 블로그 서비스 사이트의 유저 등록을 하면, 그 유저는, 해당 블로그 서비스 사이트에 있어서 유저 자신의 블로그를 운영할 수 있게 되어 있다. 그리고, 유저 등록된 유저(블로거)는, 블로그 서비스 사이트에 액세스하여, 자기의 블로그를 갱신(블로그 기사(블로그1건 마다의 기록)를 추가)할 수 있게 되어 있다. 그 때문에, 블로그 서버(6)는, 블로그의 갱신에 따라서, 블로그의 Web 페이지로서, 하나 또는 복수의 블로그 기사가 게재되는 블로그 페이지를 생성 또는 갱신한다. 그리고, 블로그 서버(6)는, 블로그 페이지 DB(601)를 구비하고, 블로그 페이지를 해당 블로그 페이지 DB(601)에 등록한다.
- [0148] 또한, 블로그 서버(6)는, 시스템 관리자에 의해 지정된 블로거의 블로그 페이지에 대하여, 광고 콘텐츠(관련 콘텐츠의 일례)를 삽입한다. 광고 콘텐츠로서는, 예를 들면, 광고 문장의 텍스트 데이터, 배너의 화상 데이터, 동화상 데이터, Adobe Flash(상표)나 Silverlight(상표) 등에 의해 생성된 리치 인터넷 어플리케이션(RIA) 등이 있다. 각 블로그 페이지에 삽입되는 광고 콘텐츠는, 대상의 블로그 페이지에 게재되어 있는 블로그 기사에 관련되는 상품이나 서비스에 관한 광고를 나타내는 콘텐츠이다. 그 때문에, 블로그 서버(6)는, 복수의 광고 콘텐츠가 등록되어 있는 광고 DB(602)를 구비한다. 그리고, 블로그 서버(6)는, 블로그 페이지로부터 블로그 기사를 추출하고, 다시 블로그 기사로부터 특징어를 추출하여, 추출한 특징어에 관련되는 광고 콘텐츠를 선택한다.
- [0149] 유저 단말기(5)는, 블로거로서의 유저나, 블로그를 열람하는 유저에 의해 이용되는 단말 장치이다. 유저 단말기(5)로서는, 예를 들면, 퍼스널 컴퓨터, PDA, 휴대 전화기 등이 이용된다.
- [0150] 관리 단말기(3)는, 블로그 시스템 BS의 시스템 관리자에 의해 사용되는 단말 장치이다. 관리 단말기(3)로서는, 예를 들면, 퍼스널 컴퓨터 등이 이용된다.
- [0151] [2-2. 블로그 서버의 구성 및 기능]
- [0152] 다음으로, 블로그 서버(6)의 구성 및 기능에 대해서, 도 14를 이용하여 설명한다.
- [0153] 도 14는, 본 실시 형태에 따른 블로그 서버(6)의 개요 구성의 일례를 도시하는 블록도이다. 또한, 도 15는, 블로거가 지정되고나서 블로그 페이지에 광고 콘텐츠가 삽입될 때까지의 처리의 개요를 도시하는 도면이다. 또한, 도 16은, Web 페이지의 구성예를 도시하는 도면이다. 또한, 도 17은, HTML 문서로부터 생성된 DOM 트리의 일례를 도시하는 도면이다. 또한, 도 18은, 기억부(65)에 기억된 콘텐츠 블록 대응 정보의 내용의 일례를 도시하는 도면이다.
- [0154] 도 14에 도시하는 바와 같이, 블로그 서버(6)는, 조작부(61)와, 표시부(62)와, 통신부(63)와, 드라이브부(64)와, 기억 수단의 일례로서의 기억부(65)와, 입출력 인터페이스부(66)와, 시스템 제어부(70)를 구비하고 있다. 그리고, 시스템 제어부(70)와 입출력 인터페이스부(66)는, 시스템 버스(71)를 거쳐서 접속되어 있다.
- [0155] 조작부(61)는, 예를 들면, 키보드, 마우스 등에 의해 구성되어 있고, 시스템 관리자 등으로부터의 조작 지시를 접수하고, 그 지시 내용을 지시 신호로서 시스템 제어부(70)에 출력하게 되어 있다. 표시부(62)는, 예를 들면, CRT 디스플레이, 액정 디스플레이 등에 의해 구성되어 있고, 문자나 화상 등의 정보를 표시하게 되어 있다. 통신부(63)는, 네트워크 NW 등에 접속하여, 관리 단말기(3), 유저 단말기(5) 등과의 통신 상태를 제어하게 되어 있다. 드라이브부(64)는, 예를 들면, 플렉시블 디스크, CD, DVD 등의 디스크 DK로부터 데이터 등을 읽어내는 한편, 해당 디스크 DK에 대하여 데이터 등을 기록하게 되어 있다.
- [0156] 기억부(65)는, 예를 들면, 하드디스크 드라이브 등에 의해 구성되어 있고, 각종 프로그램 및 데이터 등을 기억하게 되어 있다. 또한, 기억부(65)에는, 블로그 페이지 DB(601) 및 광고 DB(602)가 구축되어 있다. 블로그 페이지 DB(601)에는, 블로그 서비스 사이트를 구성하는 각 블로그 페이지(블로그 페이지의 HTML 문서(도큐먼트 테

이터의 일례), 블로그 페이지의 소재인 화상 데이터 등)가, 예를 들면, 그 페이지의 URL 및 블로거의 식별 정보인 유저 ID에 대응지어서 등록되어 있다. 광고 DB(602)에는, 복수의 광고 콘텐츠가, 그 광고 콘텐츠에 의한 광고 대상의 상품이나 서비스에 관련되는 키워드에 대응지어 등록되어 있다. 여기서, 광고 콘텐츠에 텍스트 데이터 이외의 콘텐츠가 포함되어 있는 경우에는, 해당 콘텐츠의 URL도 대응지어 등록되어 있다. 또한, 블로그 페이지 상에 표시된 광고 콘텐츠를 유저가 선택했을 때에 광고 대상의 상품이나 서비스에 관한 Web 페이지로 이동하기 위해서, 해당 Web 페이지의 URL도 대응지어 등록되어 있다.

[0157] 입출력 인터페이스부(66)는, 조작부(61)~기억부(65)와 시스템 제어부(70) 사이의 인터페이스 처리를 행하게 되어 있다. 시스템 제어부(70)는, CPU(67), ROM(68), RAM(69) 등에 의해 구성되어 있다.

[0158] 시스템 제어부(70)는, CPU(67)가, ROM(68)이나 기억부(65)에 기억된 각종 프로그램을 읽어내어 실행함으로써 블로그 서버(6)의 각 부를 제어한다. 또한, 시스템 제어부(70)는, 광고 콘텐츠 삽입 소프트웨어(특유 콘텐츠 판정 프로그램의 일례)를 실행함으로써, 추출 수단, 계산 수단, 판정 수단 및 삽입 수단으로서 기능한다. 또한, 광고 콘텐츠 삽입 소프트웨어 등은, 예를 들면, 다른 서버 장치 등으로부터 네트워크 NW를 거쳐서 취득되도록 해도 되고, CD-ROM 등의 디스크 DK에 기록되어 드라이브부(64)를 거쳐서 읽어들이지도록 해도 된다.

[0159] 광고 콘텐츠 삽입 소프트웨어는, 블로그 페이지에 광고 콘텐츠를 삽입하기 위한 프로그램이다. 도 15에 도시하는 바와 같이, 광고 콘텐츠 삽입 소프트웨어는, 매니저부, 소재 추출 엔진, 문장 해석 엔진, 광고 선택부 등에 의해 구성되어 있다. 매니저부는, 소재 추출 엔진, 문장 해석 엔진 및 광고 선택부의 실행을 제어한다. 소재 추출 엔진은, 블로그 페이지의 HTML 문서로부터 Web 소재로서의 콘텐츠를 추출함과 함께, 블로그 페이지 특유의 콘텐츠를 판정하기 위한 소프트웨어이다. 콘텐츠의 추출은, 콘텐츠 블록(콘텐츠 그룹의 일례)의 단위로 행하여진다. 본 실시 형태에 있어서는, 예를 들면, 그 기사 특유의 내용을 포함하는 블로그 기사가, 블로그 페이지 특유의 콘텐츠 블록에 상당한다.

[0160] 문장 해석 엔진은, 블로그 페이지 특유의 콘텐츠로서 추출된 블로그 기사로부터 해당 블로그 페이지의 특징어를 추출하기 위한 소프트웨어이다. 광고 선택부는, 추출된 특징어를 키워드로 해서, 블로그 페이지에 관련되는 광고 콘텐츠를 선택하기 위한 소프트웨어이다.

[0161] 이하에, 광고 콘텐츠의 삽입의 개요에 대해서 설명한다. 도 15에 도시하는 바와 같이, 시스템 관리자에 의해 대상의 블로거의 유저 ID가 지정된다(1). 그러면, 시스템 제어부(70)는, 지정된 유저 ID에 대응하는 전체 블로그 페이지의 HTML 문서를 블로그 페이지 DB(601)로부터 취득해서 해석하고, Web 소재인 콘텐츠를 콘텐츠 블록 단위로 추출한다. 그리고, 그 추출 결과로서, 추출한 콘텐츠 블록마다 콘텐츠 블록 대응 정보(콘텐츠 정보의 일례)를 생성한다(2). 다음으로, 시스템 제어부(70)는, 추출한 각 콘텐츠 블록의, 지정된 유저 ID에 대응하는 전체 블로그 페이지에 있어서의 출현 빈도를 계산한다. 본 실시 형태에 있어서 계산되는 출현 빈도는, 예를 들면, 출현 횟수(도수(횟수))이다. 그리고, 시스템 제어부(70)는, 출현 빈도에 기초하여, 각 블로그 페이지에 있어서 특유의 콘텐츠 블록을 판정한다. 구체적으로, 시스템 제어부(70)는, 각 블로그 페이지에 있어서, 출현 빈도가 소정의 임계값 이하인 콘텐츠 블록을, 그 블로그 페이지 특유의 콘텐츠 블록이라고 판정한다(3).

[0162] 시스템 제어부(70)는, 특유의 콘텐츠 블록이라고 판정한 콘텐츠 블록, 즉, 블로그 기사에 대하여 형태소 해석 등의 분석을 행하여, 블록 페이지마다의 특징어를 추출한다(4). 특징어의 추출 방법으로서, 여러 가지의 방법이 있고, 또한 공지이므로, 상세한 설명은 생략한다. 일례로서는, 출현 빈도가 가장 높은 단어를 특징어로 한다.

[0163] 다음으로, 시스템 제어부(70)는, 광고 DB(602)를 참조하고, 추출한 특징어에 관련되는 광고 콘텐츠를 선택한다(5). 그리고, 시스템 제어부(70)는, 선택한 광고 콘텐츠를 블로그 페이지에 삽입해서 표시시키기 위한 규정(태그나 데이터 그 자체의 기술 등)을, 블로그 페이지의 HTML 문서에 삽입한다(6).

[0164] 다음으로, 콘텐츠 블록의 추출 방법에 대해서 설명한다. 또한, 본 실시 형태에 있어서는, 텍스트 데이터 및 화상 데이터가 Web 소재로서 추출되는 것으로 한다.

[0165] 예를 들면, 블로그 페이지의 구성(레이아웃)이 도 16에 도시하는 바와 같은 것인 것으로 한다. Web 소재로서의 각 콘텐츠는, 블로그 페이지 상에 있어서, 각각 어떤 통합체(덩어리)마다 표시되어 있다. 그 각 통합체가 콘텐츠 블록에 상당한다. 각 콘텐츠는, HTML 문서에 기술되어 있는 DIV 태그 및 TABLE 태그(미리 정해진 태그의 일례)에 의해 각각 콘텐츠 블록으로 나뉘인다. 즉, 각 콘텐츠는, DIV 태그 및 TABLE 태그에 의해 블록화(그룹화)된다.

[0166] 도 16에는, 콘텐츠 블록(701~706)이 표시되어 있다. 콘텐츠 블록(701)은, 예를 들면, 페이지의 헤더 부분의

콘텐츠 블록이며, 텍스트A 및 화상a에 의해 구성되어 있다. 또한, 콘텐츠 블록(702)는, 예를 들면, 다른 Web 페이지로 이동하기 위한 네비게이션 부분의 콘텐츠 블록이며, 예를 들면, 다른 Web 페이지에의 링크를 나타내는 텍스트B, 텍스트C 및 텍스트D에 의해 구성되어 있다. 또한, 콘텐츠 블록(703)은, 예를 들면, 블로그의 표시 영역에 대응하는 콘텐츠 블록이며, 블로그 등의 표제를 나타내는 텍스트E, 콘텐츠 블록(704) 및 콘텐츠 블록(705)에 의해 구성되어 있다. 이와 같이, 콘텐츠 블록이 상자 형상, 즉, 계층 구조로 되어 있는 경우도 있다. 이 경우, 콘텐츠 블록(703)에 포함되는 콘텐츠는 텍스트E만으로 되고, 콘텐츠 블록(704) 및 콘텐츠 블록(705)은, 콘텐츠 블록(703)으로부터 독립되어 있는 것으로 된다. 콘텐츠 블록(704 및 705)은, 각각 1건의 블로그 기사이다. 콘텐츠 블록(704)에는, 블로그 기사의 표제나 본문을 나타내는 텍스트F 및 G에 의해 구성되어 있다. 콘텐츠 블록(705)에는, 블로그 기사의 표제나 본문을 나타내는 텍스트H, I 및 J와, 블로그 기사에 관련되어 블로거에 의해 등록된 화상b 및 c에 의해 구성되어 있다. 콘텐츠 블록(706)은, 예를 들면, 카피라이트 표시를 나타내는 콘텐츠 블록이며, 텍스트I에 의해 구성되어 있다.

[0167] 이들의 콘텐츠 블록 중, 콘텐츠 블록(701, 702, 703 및 706)은, 도 16에 도시하는 블로그 페이지 이외의 블로그 페이지 상에서도 비교적 빈번하게 나타난다. 한편, 콘텐츠 블록(704) 및 콘텐츠 블록(705)은, 기본적으로 해당 블로그 페이지에만 이용된다. 따라서, 콘텐츠 블록(704) 또는 콘텐츠 블록(705)이, 해당 블로그 페이지 특유의 콘텐츠 블록이라고 판단되게 된다.

[0168] 본 실시 형태에 있어서는, 그 기사 특유의 내용을 포함하는 블로그 기사에 상당하는 콘텐츠 블록을 특유의 콘텐츠 블록이라고 판정되도록 할 필요가 있다. 이러한 특유한 내용을 포함하는 블로그 기사는, 1페이지 내에 복수 포함되어 있는 경우가 있다. 그 때문에, 출현 빈도가 소정의 임계값 이하인 콘텐츠 블록은, 모두 특유의 콘텐츠 블록으로 된다. 예를 들면, 임계값의 값을 1회로 설정한다. 그렇게 하면, 특유한 내용을 포함하는 블로그 기사는, 특유의 콘텐츠 블록이라고 판정되고, 다른 블로그 기사와 동일한 내용만을 포함하는 블로그 기사는, 특유의 콘텐츠 블록이라고는 판정되지 않는다. 또한, 헤더 부분, 네비게이션 부분, 카피라이트 표시 부분 등의 각 블로그 페이지에 공통되는 것 같은 콘텐츠 블록의 출현 빈도는 각각 2회 이상으로 되므로, 이들도 특유의 콘텐츠 블록이라고는 판정되지 않는다. 또한, 임계값은 기억부(65)에 미리 기억된다.

[0169] 도 16에 도시하는 블로그 페이지의 HTML 문서를 DOM 트리, 즉, 트리 구조로 나타낸 것이 도 17이다. 또한, 도 17에 나타내는 DOM 트리에 있어서, 본 실시 형태의 설명에 필요 없는 태그의 노드의 도시는 생략하고 있다.

[0170] 제1 실시 형태의 경우와 마찬가지로 콘텐츠 블록을 추출하면, 시스템 제어부(70)는, 추출 결과를 나타내는 콘텐츠 블록 대응 정보를 일시적으로 기억부(65)에 보존한다. 도 18에 도시하는 바와 같이, 콘텐츠 블록 대응 정보(부호 401)는, 콘텐츠 블록마다 보존된다. 또한, 본 실시 형태에 있어서는, 블로그 페이지 특유의 콘텐츠 블록이라고 판정된 콘텐츠 블록, 즉, 블로그 기사로부터 특징어를 추출하므로, 텍스트 데이터가 추출되면 되고, 화상 데이터에 대해서는 추출하지 않아도 된다.

[0171] [2-3. 블로그 시스템의 동작]

[0172] 다음으로, 블로그 시스템 BS의 동작에 대해서, 도 19 내지 도 21을 이용하여 설명한다.

[0173] 도 19는, 본 실시 형태에 따른 블로그 서버(6)의 시스템 제어부(70)의 광고 콘텐츠 삽입 처리에 있어서의 처리 예를 도시하는 플로우차트이다.

[0174] 광고 콘텐츠 삽입 처리는, 예를 들면, 시스템 관리자의 조작에 기초하여 관리 단말기(3)로부터 광고 콘텐츠 삽입 처리의 실행의 리퀘스트가 송신되어 왔을 때에 개시된다.

[0175] 그리고, 시스템 관리자가 광고 콘텐츠의 삽입 대상으로 하는 블로그를 운영하는 블로거의 유저 ID를 지정하면, 도 19에 도시하는 바와 같이, 시스템 제어부(70)는, 지정된 유저 ID를 관리 단말기(3)로부터 수신한다(스텝 S101).

[0176] 다음으로, 시스템 제어부(70)는, 블록수 NUM에 0을 설정한다(스텝 S102). 블록수 NUM은, 현시점에서 발견 완료된 콘텐츠 블록의 개수이다. 또한, NUM은 글로벌 변수이며, 후술하는 1페이지 대응 추출 처리 및 트리 탐색 처리로부터 액세스가 가능하다.

[0177] 다음으로, 시스템 제어부(70)는, 수신한 유저 ID에 대응하는 최초의 블로그 페이지의 HTML 문서를 블로그 페이지 DB(601)로부터 취득한다(스텝 S103). 다음으로, 시스템 제어부(70)는, 취득한 HTML 문서를 지정해서, 후술하는 1페이지 대응 추출 처리를 실행한다(스텝 S104). 이 1페이지 대응 추출 처리에서는, 취득한 HTML 문서로부터 콘텐츠 블록이 추출되고, 콘텐츠 블록 대응 정보가 보존된다.

- [0178] 다음으로, 시스템 제어부(70)는, 수신한 유저 ID에 대응하는 모든 블로그 페이지의 콘텐츠 블록을 추출하였는지의 여부를 판정한다(스텝 S105). 이때, 시스템 제어부(70)는, 콘텐츠 블록을 추출하고 있지 않은 블로그 페이지가 존재하는 경우에는(스텝 S105:NO), 다음의 블로그 페이지의 HTML 문서를 블로그 페이지 DB(601)로부터 취득해서(스텝 S106), 스텝 S104로 이행한다. 그리고, 시스템 제어부(70)는, 스텝 S104~S106의 처리를 반복해서 모든 블로그 페이지의 콘텐츠 블록을 추출하면(스텝 S105:YES), 스텝 S107로 이행한다.
- [0179] 스텝 S107에 있어서, 시스템 제어부(70)는, 수신한 유저 ID에 대응하는 최초의 블로그 페이지의 HTML 문서를 특정한다.
- [0180] 다음으로, 시스템 제어부(70)는, 취득한 HTML 문서를 지정해서, 후술하는 특유 콘텐츠 블록 판정 처리를 실행한다(스텝 S108). 이 특유 콘텐츠 블록 판정 처리에서는, 특정한 HTML 문서로부터 콘텐츠 블록이 추출되고, 블로그 페이지 특유의 콘텐츠 블록이 판정된다.
- [0181] 다음으로, 시스템 제어부(70)는, 특유라고 판정된 콘텐츠 블록을 구성하는 각 텍스트 데이터로부터 블로그 페이지의 특징어를 추출한다(스텝 S109). 다음으로, 시스템 제어부(70)는, 추출한 특징어에 기초하여, 블로그 페이지에 관련되는 광고 페이지를 해당 블로그 페이지에 삽입한다(스텝 S110). 구체적으로, 시스템 제어부(70)는, 추출한 특징어를 키워드로 하고, 광고 DB(602)를 참조해서 해당 키워드에 대응하는 광고 콘텐츠를 선택한다. 다음으로, 시스템 제어부(70)는, 특정한 HTML 문서 상의 소정의 위치에, 선택한 광고 콘텐츠의 규정을 삽입한다. 예를 들면, 시스템 제어부(70)는, 광고 콘텐츠에 텍스트 데이터가 포함되어 있는 경우에는, 해당 텍스트 데이터의 내용을 HTML 문서에 추가한다. 또한, 예를 들면, 시스템 제어부(70)는, 광고 콘텐츠에 화상 데이터가 포함되어 있는 경우에는, 해당 화상 데이터를 표시하기 위한 IMG 태그를 HTML 문서에 추가한다. 또한, 예를 들면, 시스템 제어부(70)는, 광고 대상의 상품이나 서비스에 관한 Web 페이지의 링크 정보를 HTML 문서에 추가한다.
- [0182] 시스템 제어부(70)는, 특정한 HTML 문서에 광고 콘텐츠의 규정을 삽입하면, 해당 HTML 문서에서, 블로그 페이지 DB(601)에 등록되어 있는 HTML 문서를 갱신한다(스텝 S111).
- [0183] 다음으로, 시스템 제어부(70)는, 수신한 유저 ID에 대응하는 모든 블로그 페이지에 광고 콘텐츠를 삽입하였는지의 여부를 판정한다(스텝 S112). 이때, 시스템 제어부(70)는, 광고 콘텐츠를 삽입하지 않고 있는 블로그 페이지가 존재하는 경우에는(스텝 S112:NO), 다음의 블로그 페이지의 HTML 문서를 특정해서(스텝 S113), 스텝 S108로 이행한다. 그리고, 시스템 제어부(70)는, 스텝 S108~S113의 처리를 반복해서 모든 블로그 페이지에 광고 콘텐츠를 삽입하면(스텝 S112:YES), 기억부(65)에 보존시켜 둔 모든 콘텐츠 블록 대응 정보를, 기억부(65)로부터 삭제한다(스텝 S114). 시스템 제어부(70)는, 이 처리를 끝내면, 광고 콘텐츠 삽입 처리를 종료시킨다.
- [0184] 도 20은, 본 실시 형태에 따른 블로그 서버(6)의 시스템 제어부(70)의 1페이지 대응 추출 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0185] 도 20에 도시하는 바와 같이, 시스템 제어부(70)는, 우선, 취득한 HTML 문서의 DOM 트리를 RAM(69) 상에 생성한다(스텝 S121).
- [0186] 다음으로, 시스템 제어부(70)는, 계층 LV에 0을 설정한다(스텝 S122). 계층 LV는, DOM 트리에 있어서 현재 탐색 중인 노드가 속하는 콘텐츠 블록의 계층이다. LV는 글로벌 변수이며, 1페이지 대응 추출 처리 및 트리 탐색 처리로부터 액세스가 가능하다.
- [0187] 다음으로, 시스템 제어부(70)는, DOM 트리의 루트 노드를 지정해서(스텝 S123), 트리 탐색 처리를 실행한다(스텝 S124). 트리 탐색 처리의 처리 내용은, 제1 실시 형태의 경우와 마찬가지로, 상세한 설명은 생략한다.
- [0188] 다음으로, 시스템 제어부(70)는, 트리 탐색 처리에 의해 생성된 각 콘텐츠 블록 대응 정보를 기억부(65)에 보존한다(스텝 S125). 시스템 제어부(70)는, 이 처리를 끝내면, 1페이지 대응 추출 처리를 종료시킨다.
- [0189] 도 21은, 본 실시 형태에 따른 콘텐츠 생성 서버(1)의 시스템 제어부(70)의 특유 콘텐츠 블록 판정 처리에 있어서의 처리예를 도시하는 플로우차트이다.
- [0190] 도 21에 도시하는 바와 같이, 시스템 제어부(70)는, 우선, 1페이지 대응 추출 처리와 마찬가지로, 지정된 HTML 문서의 DOM 트리 생성(스텝 S161), 블록수 NUM 및 계층 LV에 대하여 0의 설정을 행하고(스텝 S162), DOM 트리의 루트 노드를 지정해서(스텝 S163), 트리 탐색 처리를 실행한다(스텝 S164).
- [0191] 다음으로, 시스템 제어부(70)는, 블록 번호i에 1을 설정한다(스텝 S165). 다음으로, 시스템 제어부(70)는, 블

록 번호*i*의 콘텐츠 블록의 출현 빈도를 계산한다(스텝 S166).

- [0192] 구체적으로, 시스템 제어부(70)는, 스텝 S164의 트리 탐색 처리에 있어서 생성된 콘텐츠 블록 대응 정보*i*의 블록 구성 정보와, 기억부(65)에 보존되어 있는 각 콘텐츠 블록 대응 정보의 블록 구성 정보를 비교하여, 출현 빈도를 계산한다. 출현 빈도의 계산 방법은, 제1 실시 형태의 경우와 마찬가지로이다.
- [0193] 시스템 제어부(70)는, 출현 빈도를 계산하면, 계산한 출현 빈도가, 기억부(65)에 기억되어 있는 임계값 이하인지의 여부를 판정한다(스텝 S167). 이때, 시스템 제어부(70)는, 출현 빈도가 임계값 이하인 경우에는(스텝 S167:YES), 블록 번호*i*의 콘텐츠 블록을, 특유의 콘텐츠 블록의 1개라고 판정한다(스텝 S168). 즉, 시스템 제어부(70)는, 블록 번호*i*의 콘텐츠 블록을, 지정된 HTML 문서가 대응하는 블로그 페이지 특유의 콘텐츠 블록에 추가한다.
- [0194] 시스템 제어부(70)는, 출현 빈도가 임계값보다도 클 경우(스텝 S167:NO), 또는, 스텝 S168의 처리를 끝낸 경우에는, 블록 번호*i*에 1을 가산해서(스텝 S169), 블록 번호*i*가 블록수 NUM의 값보다 큰지의 여부를 판정한다(스텝 S170). 이때, 시스템 제어부(70)는, 블록 번호*i*가 블록수 NUM의 값 이하인 경우에는(스텝 S170:NO), 스텝 S166으로 이행한다. 그리고, 시스템 제어부(70)는, 트리 탐색 처리에 있어서 추출된 모든 콘텐츠 블록의 출현 빈도를 계산하면(스텝 S170:YES), 특유 콘텐츠 블록 판정 처리를 종료시킨다.
- [0195] 또한, 시스템 제어부(70)는, 스텝 S164의 트리 탐색 처리에 의해 콘텐츠 블록을 추출하고 있었지만, 광고 콘텐츠 삽입 처리로부터 실행된 1페이지 대응 추출 처리(도 19 스텝 S104)에 있어서, 수신한 블로그의 유저 ID에 대응하는 모든 블로그 페이지에 대해서 콘텐츠 블록이 추출되고, 그 결과로서 콘텐츠 블록 대응 정보가 기억부(65)에 기억되어 있으므로, 다시 콘텐츠 블록을 추출하지 않아도 된다. 그 경우에는, 지정된 HTML 문서의 URL에 기초하여, 해당 HTML 문서가 대응하는 블로그 페이지를 구성하는 각 콘텐츠 블록의 콘텐츠 블록 대응 정보를 기억부(65)로부터 취득할 수 있다.
- [0196] [2-4. 변형예1]
- [0197] 다음으로, 본 실시 형태의 변형예에 대해서, 도 22를 이용하여 설명한다.
- [0198] 지금까지의 설명에 있어서는, 시스템 관리자에 의해 블로거가 지정되었을 때에, 지정된 블로거의 블로그 페이지에 광고 콘텐츠를 삽입하고 있었지만, 블로거가 갱신된 타이밍에서 광고 콘텐츠를 삽입해도 된다.
- [0199] 도 22는, 본 실시 형태의 변형예에 따른 블로그 서버(6)의 시스템 제어부(70)의 블로그 갱신시 처리에 있어서의 처리예를 도시하는 플로우차트이다. 또한, 도 22에 있어서, 도 19와 마찬가지로의 처리에 대해서는 마찬가지로의 스텝 번호를 붙여 있다.
- [0200] 우선, 블로그의 갱신에 앞서서, 블로거는, 유저 단말기(5)를 조작해서 블로그 서비스 사이트에 액세스하고, 자신의 유저 ID와 패스워드를 입력하는 것에 의해 블로그 서비스 사이트에 로그인한다. 이 로그인에 의해, 블로그 서버(6)는 유저 단말기(5)에 대하여 세션 ID를 발행하고, 세션 ID와 유저 ID를 대응지어 관리한다. 유저 단말기(5)로부터 블로그 서버(6)에의 리퀘스트에는 세션 ID가 포함되어 있으므로, 블로그 서버(6)는, 어느 블로거로부터의 리퀘스트인지를 특정할 수 있다.
- [0201] 그리고, 블로거가 새로운 블로그 기사의 등록 조작을 행하면, 유저 단말기(5)는, 블로그 기사의 데이터(표제나 본문 등의 텍스트 데이터, 화상 데이터 등)를 블로그 서버(6)에 송신하고, 도 22에 도시하는 바와 같이, 블로그 서버(6)의 시스템 제어부(70)는, 블로그 기사의 데이터를 수신한다(스텝 S171). 다음으로, 시스템 제어부(70)는, 블로거의 유저 ID에 대응하는 블로그 페이지 중으로부터, 갱신해야 할 블로그 페이지의 HTML 문서를 블로그 페이지 DB(601)로부터 취득한다(스텝 S172). 다음으로, 시스템 제어부(70)는, 수신한 블로그 기사의 데이터에 기초하여, 취득한 HTML 문서를 갱신한다(스텝 S173). 예를 들면, 시스템 제어부(70)는, 취득한 HTML 문서에, 블로그 기사용의 TABLE 태그 또는 DIV 태그를 추가하고, 해당 태그 사이에 끼워진 형태로, 수신한 블로그 기사의 표제나 본문의 텍스트 데이터 등을 추가한다. 다음으로, 시스템 제어부(70)는, 블로그 기사의 데이터를 추가한 HTML 문서에서, 블로그 페이지 DB(601)에 등록되어 있는 HTML 문서를 갱신한다(스텝 S174).
- [0202] 다음으로, 시스템 제어부(70)는, 블로거의 유저 ID에 대응하는 모든 블로그 페이지로부터 콘텐츠 블록을 추출한다(스텝 S103~S106).
- [0203] 다음으로, 시스템 제어부(70)는, 스텝 S173에 있어서 갱신한 HTML 문서를 지정해서, 특유 콘텐츠 블록 판정 처리를 실행하고(스텝 S108), 특유라고 판정된 콘텐츠 블록을 구성하는 각 텍스트 데이터로부터 블로그 페이지의

특징어를 추출한다(스텝 S109).

- [0204] 다음으로, 시스템 제어부(70)는, 지정된 HTML 문서로부터, 기존의 광고 콘텐츠의 규정을 삭제하고(스텝 S775), 추출한 특징어를 키워드로 해서, 관련되는 광고 콘텐츠의 규정을 삽입한다(스텝 S110). 즉, 시스템 제어부(70)는, 블로그 페이지 상에 표시되는 광고 콘텐츠를 변경한다.
- [0205] 그리고, 시스템 제어부(70)는, 광고 콘텐츠의 규정이 삽입된 HTML 문서에서, 블로그 페이지 DB(601)에 등록되어 있는 HTML 문서를 갱신하고(스텝 S111), 기억부(65)로부터 모든 콘텐츠 블록 대응 정보를 삭제한다(스텝 S114).
- [0206] 또한, 블로그의 갱신에 따라 블로그 페이지를 신규로 생성해야만 하는 경우의 처리도, 기본적으로 상술한 처리와 마찬가지로도 된다. 단, 신규로 생성된 블로그 페이지에는, 광고 콘텐츠는 아직 삽입되어 있지 않으므로, 스텝 S175에 있어서의 광고 콘텐츠의 규정의 삭제는 행해지지 않는다.
- [0207] [2-5. 변형예2]
- [0208] 지금까지의 설명에 있어서는, 블로그 페이지 특유의 콘텐츠의 판정에 이용되는 임계값으로서 1회를 설정하고 있었지만, 2회 이상의 값을 임계값으로서 설정해도 된다.
- [0209] 예를 들면, 임계값을 1회로 한 경우에는, 출현 빈도가 1회인 콘텐츠 블록(블로그 기사)이 블로그 페이지 특유의 콘텐츠로서 추출되고, 추출된 블로그 기사의 텍스트 데이터로부터 특징어가 추출된다. 이때, 추출된 각 블로그 기사의 텍스트 데이터의 데이터량이 적으면, 거기에서 추출되는 단어의 수는 적어진다. 그리고, 충분한 수의 단어를 추출할 수 없으면, 어느 단어가 특징어인지를 전혀 판단할 수 없는 경우나, 적확하게 판단할 수 없는 경우가 있다. 그래서, 임계값의 값을 올리고, 블로그 페이지 특유의 콘텐츠라고 판정되는 조건을 느슨하게 함으로써, 특징어를 추출하는 대상으로 되는 블로그 기사를 늘려 간다. 이에 의해, 특징어를 추출하는 것이 가능하게 된다.
- [0210] 구체적으로는, 블로그 서버(6)의 시스템 제어부(70)가, 처음에는 임계값을 1회로 설정해서, 블로그 페이지 특유의 콘텐츠 블록을 판정함으로써, 출현 횟수가 1회인 블로그 기사를 추출해서 특징어를 추출한다. 이때, 시스템 제어부(70)는, 특징어를 추출할 수 없다고 판정한 경우에는, 임계값을 2회로 변경해서, 블로그 기사의 추출 및 특징어의 추출을 행한다. 시스템 제어부(70)는, 그래도 특징어를 추출할 수 없다고 판정한 경우에는, 임계값을 3회로 변경해서, 블로그 기사의 추출 및 특징어의 추출을 행한다. 시스템 제어부(70)는, 이러한 처리를, 특징어를 추출할 수 있을 때까지 계속한다. 즉, 특유의 콘텐츠 블록의 추출 결과에 기초하는 처리를 정상적으로 행할 수 없었던 경우에, 임계값을 올리는 것이다.
- [0211] 단, 임계값을 무제한으로 올려 가면, 블로그 기사가 아닌 것도 추출되게 되므로, 임계값이 어느 정도까지 올라 가면 처리를 중단하는 것으로 한다. 예를 들면, 임계값이, 지정된 블로거에 대응하는 블로그 페이지의 페이지수의 값으로까지 오르면, 각 블로그 페이지에서 공통해서 이용되는 콘텐츠 블록을 추출하게 되므로, 임계값이 블로그 페이지의 페이지수의 값으로 되면 처리를 중단해도 된다.
- [0212] 또한 예를 들면, 블로그 페이지의 소정 페이지수당 1회만 출현하는 콘텐츠 블록이 블로그 페이지 특유의 콘텐츠 블록이면, 시스템 관리자측에서 미리 정해도 된다. 이 경우, 지정된 블로거에 대응하는 블로그 페이지의 페이지수에 비례해서, 임계값으로서의 출현 횟수를 변경해도 된다.
- [0213] [2-6. 변형예3]
- [0214] 지금까지의 설명에 있어서는, 블로그 페이지 특유의 콘텐츠의 판정에 이용되는 출현 빈도로서, 출현 횟수(도수(횟수))를 이용하고 있었지만, 지정된 블로거에 대응하는 블로그 페이지의 전체 콘텐츠 블록에 대한 출현 횟수의 비율(상대도수(횟수))을 이용해도 된다.
- [0215] 예를 들면, 블로거가 등록한 블로그 기사에 대하여, 다른 유저로부터 코멘트를 등록할 수 있고, 블로그 기사와 함께 코멘트를 열람 가능하게 되어 있는 것으로 한다. 이 코멘트의 텍스트 데이터도 블로그 페이지를 구성하는 콘텐츠의 1개로 된다. 블로그 서버(6)의 시스템 제어부(70)는, 코멘트의 텍스트 데이터를 블로그 페이지에 추가하는 경우, 해당 블로그 페이지의 HTML 문서에, 블록화 태그의 기술을 추가한 후에 해당 텍스트 데이터를 추가함으로써, 코멘트의 텍스트 데이터를, 블로그 기사나 다른 코멘트의 텍스트 데이터와는 독립된 콘텐츠 블록으로 한다. 그리고, 시스템 제어부(70)는, 콘텐츠 블록으로서 코멘트의 텍스트 데이터를 추출하고, 추출한 코멘트의 텍스트 데이터가 특유한 내용을 갖고 있는 경우에는, 그 코멘트에 관련되는 광고 콘텐츠를 블로그 페이지에 삽입하도록 한다.

- [0216] 그런데, 어떤 블로그 기사에 대하여 복수의 코멘트가 등록된 경우에, 복수의 코멘트의 각 내용이, 예를 들면, 다수파의 의견과 소수파의 의견 등과 같이, 빈번하게 출현하는 내용과, 그다지 빈번하게는 출현하지 않는 내용으로 분리되는 경우가 있다. 이때, 다수파의 의견은, 일반적인 의견이며, 그다지 특징적인 내용이 아니라고 생각할 수 있다. 한편, 소수파의 의견은, 특이한 의견이며, 블로그 페이지 특유의 내용이라고 생각할 수 있다. 그러한 경우에, 소수파의 의견을 나타내는 코멘트를 블로그 페이지 특유의 콘텐츠로서 추출하고자 한다.
- [0217] 그러나, 다수파의 의견의 수와 소수파의 의견의 수는, 상대적인 것이며, 코멘트의 총수에 따라 변화한다. 이러한 경우에, 출현 빈도로서 도수(횟수)를 이용하고, 임계값을 예를 들면 1회로 하면, 빈번하게 나타나지 않는 내용(소수파의 의견)을 적절하게 추출할 수 없는 경우가 있다. 그래서, 출현 빈도로서 상대도수(횟수)를 이용하고, 임계값을 소정의 비율로 설정하는 것이다. 이때의 임계값은 임의로 설정할 수 있다. 예를 들면, 추출된 콘텐츠 블록의 내용이 N개의 패킷(N은 2 이상의 정수)으로 나뉘지는 경우, 소수파의 의견을 구별하기 위해서, 임계값에는  $1 \div N$  미만의 범위에서 임계값을 설정해도 된다. 이와 같이, 시스템 제어부(70)가, 그때마다의 상황에 따라서 임계값을 변경해도 된다.
- [0218] 또한, 블로그 등과 같은 기사에 대하여 코멘트 등을 등록할 수 있는 시스템으로서, 예를 들면, 어떤 유저가 등록한 중얼거림에 대하여, 이것을 팔로우하는 중얼거림을 다른 유저가 등록할 수 있는 Twitter(상표)이나, 전자 게시판 등이 있다.
- [0219] 이상 설명한 바와 같이, 본 실시 형태에 따르면, 블로그 서버(6)의 시스템 제어부(70)가, HTML 문서가 지정됨으로써 순차적으로 지정된 블로그 페이지를 구성하고 있는 콘텐츠를 추출하고, 지정된 블로그 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도를 계산하고, 지정된 블로그 페이지를 구성하는 콘텐츠 중, 출현 빈도가 소정의 임계값 이하인 콘텐츠를 해당 블로그 페이지 특유의 콘텐츠라고 판단한다.
- [0220] 따라서, 출현 빈도가 작은 콘텐츠일수록, 지정된 블로그 페이지 이외에는 그다지 출현하지 않는 콘텐츠이므로, 출현 빈도가 임계값 이하인지를 판정함으로써, 해당 조건을 충족시키는 모든 콘텐츠가, 지정된 블로그 페이지 특유의 콘텐츠라고 특정된다. 따라서, 블로그 페이지 특유의 콘텐츠를 용이하게 추출할 수 있다.
- [0221] 또한, 블로그 서버(6)의 시스템 제어부(70)가, 지정된 블로그 페이지 특유의 콘텐츠에 관련되는 광고 콘텐츠를 해당 블로그 페이지에 삽입한다.
- [0222] 따라서, 블로그 페이지의 특징과 관련되는 정보를 Web 페이지에 추가할 수 있다.
- [0223] 또한, 블로그 서버(6)의 시스템 제어부(70)가, 지정된 블로그 페이지를 구성하고 있는 콘텐츠로서, 블로그 기사의 텍스트 데이터가 포함되어 있는 경우에, 해당 텍스트 데이터를, 해당 블로그 페이지 특유의 콘텐츠라고 판정하고, 블로그 기사의 텍스트 데이터로부터 해당 블로그 페이지의 특징어를 추출하고, 해당 특징어를 키워드로 해서, 미리 관련지어져 있는 광고 콘텐츠를, 해당 블로그 페이지에 삽입한다.
- [0224] 따라서, 블로그 페이지에 게재되어 있는 블로그의 내용에 관련되는 광고를 해당 블로그 페이지에 추가할 수 있다.
- [0225] 또한, 블로그 서버(6)의 시스템 제어부(70)가, 블로그 서비스 사이트에 포함되는 복수의 블로그 페이지 상에 있어서의 각 콘텐츠의 출현 빈도를 계산한다.
- [0226] 따라서, 블로그 서비스 사이트에 포함되는 복수의 Web 페이지(예를 들면, 지정된 블로거의 유저 ID에 대응하는 복수의 블로그 페이지) 상에 있어서, 지정된 블로그 페이지를 구성하고 있는 각 콘텐츠의 출현 빈도가 계산되므로, 블로그 서비스 사이트 내에서 공통해서 이용되는 콘텐츠는, 특유의 콘텐츠가 아니라고 판정하는 것이 가능하게 되어, 판단 정밀도를 높일 수 있다.
- [0227] 또한, 블로그 서버(6)의 시스템 제어부(70)가, 1개 이상의 콘텐츠로 구성되는 콘텐츠 블록의 단위로, 블로그 페이지를 구성하고 있는 콘텐츠를 추출하고, 지정된 블로그 페이지를 구성하고 있는 각 콘텐츠 블록 출현 빈도를 계산하고, 지정된 블로그 페이지를 구성하는 콘텐츠 블록 중, 출현 빈도가 임계값 이하인 콘텐츠 블록을 해당 블로그 페이지 특유의 콘텐츠 블록이라고 판단한다.
- [0228] 따라서, 블로그 페이지 상에 있어서, 예를 들면, 헤더 부분, 네비게이션 부분, 블로그가 표시되는 부분, 카피라이트 표시의 부분 등과 같이, 1개 이상의 콘텐츠가 통합체로써 콘텐츠 블록으로서 표시되어 있는 경우에, 블로그 페이지 특유의 콘텐츠 블록을 추출할 수 있다.
- [0229] 또한, 블로그 서버(6)의 시스템 제어부(70)가, 블로그 페이지를 구성하고 있는 콘텐츠를 해당 블로그 페이지의

HTML 문서에 기초하여 추출하고, HTML 문서에 있어서 DIV 태그 또는 TABLE 태그에 기초하여, 콘텐츠 블록을 정한다.

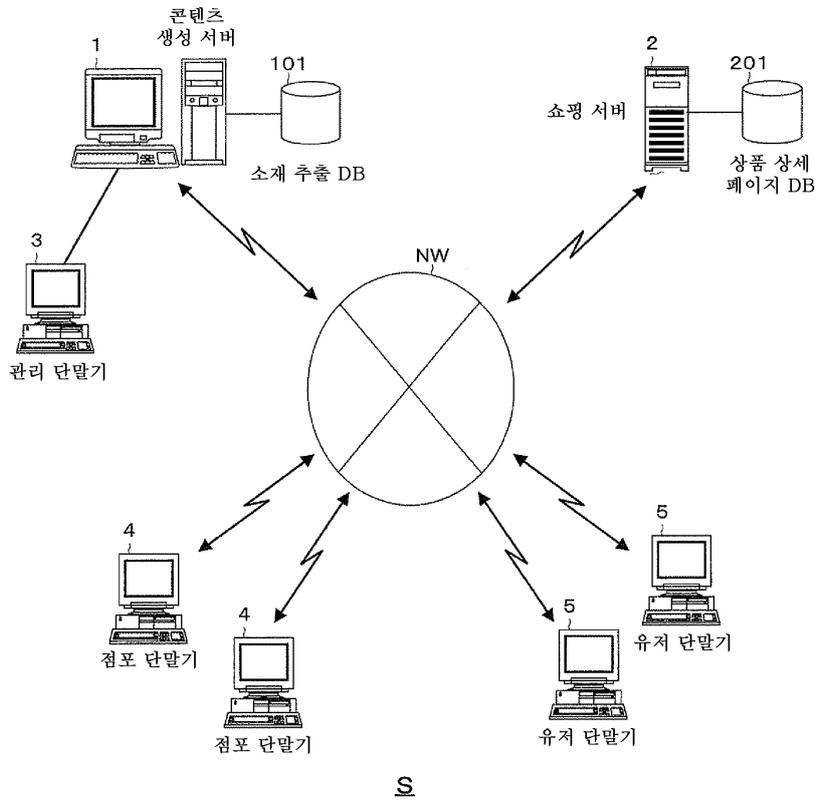
- [0230] 따라서, DIV 태그에 의해, HTML 문서의 작성 시에 명시적으로 블록화된 1개 이상의 콘텐츠를 특정할 수 있고, 또한, TABLE 태그에 의해, 표 형식으로 블록화되어 표시되는 1개 이상의 콘텐츠를 특정할 수 있으므로, 예를 들면, 이들의 태그에 의해, 블로그 페이지 특유의 콘텐츠와, 특유하지 않은 콘텐츠가 블록화되어 있는 경우에, Web 페이지 특유의 콘텐츠를 판단하는 정밀도를 높일 수 있다.
- [0231] 또한, 상기 제2 실시 형태에 있어서는, 지정된 블로그 페이지를 구성하는 각 콘텐츠 블록에 대응하는 콘텐츠 블록 대응 정보를, 지정된 블로거의 유저 ID에 대응하는 모든 블로그 페이지를 구성하는 각 콘텐츠 블록 대응 정보와 비교함으로써 각 출현 빈도가 계산되어 있었다. 즉, 지정된 블로그 페이지를 구성하는 각 콘텐츠 블록의 출현 빈도를 계산하는 경우에, 지정된 블로거에 대응하는 모든 블로그 페이지를 대상으로 한 범위에 출현하는 빈도를 계산하게 되어 있었지만, 대상으로 하는 범위는 이것에만 한정되는 것은 아니다. 예를 들면, 미리 정해진 페이지수 분의 블로그 페이지를 대상으로 해도 되고, 블로그 서비스 사이트를 구성하는 모든 블로그 페이지를 대상으로 해도 된다.
- [0232] 또한, 상기 제2 실시 형태로서, Web 페이지 특유의 콘텐츠에 관련되는 콘텐츠로서, 상품이나 서비스에 관한 광고를 나타내는 광고 콘텐츠를, 해당 Web 페이지에 삽입하고 있었지만, 관련되는 콘텐츠이면 광고 콘텐츠에 한정되는 것은 아니다. 예를 들면, 특유의 콘텐츠라고 판정된 블로그 기사 등의 콘텐츠에 관련되는 화상 데이터(정지 화상이나 동화상)를, 배경 화상이나 삽입 화상(삽화 등)으로서 삽입해도 된다. 구체적으로는, 예를 들면, 화상 데이터용의 데이터베이스를 구축하고, 해당 데이터베이스에, 화상 데이터와 키워드를 대응지어 등록해 둔다. 화상 데이터에 대응지어지는 키워드는, 그 화상 데이터에 의해 표시되는 화상을 나타내는 단어나 해당 화상에 관련되는 단어이다. 그리고, 특유의 콘텐츠라고 판정된 콘텐츠로부터 특징어를 추출하고, 추출한 특징어를 키워드로 해서, 관련되는 화상 데이터를 데이터베이스로부터 선택한다. 그리고, 대상의 HTML 문서의 BODY 태그에, 선택한 화상 데이터의 URL을 background 속성으로서 삽입하거나, 대상의 HTML 문서의 소정 위치에, 선택한 화상 데이터를 표시하는 IMG 태그를 삽입하거나 한다. 이에 의해, 특유의 콘텐츠라고 판정된 블로그 기사 등의 콘텐츠의 내용에 적합한 화상을 Web 페이지에 삽입할 수 있다.
- [0233] 또한, Web 페이지 특유의 콘텐츠의 용도로서는, 관련되는 콘텐츠를 Web 페이지에 삽입하는 것에만 한정되는 것은 아니다. 예를 들면, Web 페이지 특유의 콘텐츠에 기초하여, 새로운 콘텐츠를 생성해도 된다.
- [0234] 또한, 상기 각 실시 형태에 있어서는, Web 페이지를 구성하고 있는 콘텐츠로서, 텍스트 데이터 및 화상 데이터를 추출하고 있었지만, 추출 대상의 콘텐츠는 이들에 한정되는 것은 아니다. 예를 들면, Web 페이지 상에 표시되는 콘텐츠, 또는, Web 페이지가 표시되어 있을 때에 재생되는 콘텐츠(예를 들면, 동화상 데이터, 음성 데이터, 전자 문서 등)이면 된다. 또한, 소정의 종류의 콘텐츠만을 추출해도 된다.
- [0235] 또한, 상기 각 실시 형태에 있어서는, DIV 태그 사이에 끼워져 있는 콘텐츠, 및 TABLE 태그 사이에 끼워져 있는 콘텐츠를, 콘텐츠 블록으로서 그룹화해서 추출하고 있었지만, 콘텐츠를 그룹화하는 태그로서는, 이들에만 한정되는 것은 아니다.
- [0236] 또한, 상기 각 실시 형태에 있어서는, Web 페이지 특유의 콘텐츠를 콘텐츠 블록의 단위로 추출하고 있었지만, 각 콘텐츠를 그대로 1개씩 추출해도 된다.
- [0237] 또한, 상기 각 실시 형태에 있어서는, 서버 장치에 대하여 본 발명의 특유 콘텐츠 판정 장치를 적용하고 있었지만, 기억 수단이나 네트워크 상에서 HTML 문서를 취득할 수 있으면, 단말 장치 등에 대하여 특유 콘텐츠 판정 장치를 적용해도 된다.
- [0238] 또한, 상기 각 실시 형태에 있어서는, HTML 문서에 대하여 본 발명의 도큐먼트 데이터를 적용하고 있었지만, 마크업 언어로 기술되고, Web 페이지를 구성하는 콘텐츠를 나타내는 데이터(예를 들면, XHTML(Extensible HyperText Markup Language) 문서 등)에 대하여 도큐먼트 데이터를 적용해도 된다.
- [0239] 또한, 상기 실시형태에 있어서는, 쇼핑 사이트에 있어서의 상품 상세 페이지를 구성하는 콘텐츠나, 블로그 서비스 사이트에 있어서의 블로그 페이지를 구성하는 콘텐츠를 추출하고 있었지만, 대상으로 하는 사이트 및 페이지의 종류는 이들에만 한정되는 것은 아니다.

**부호의 설명**

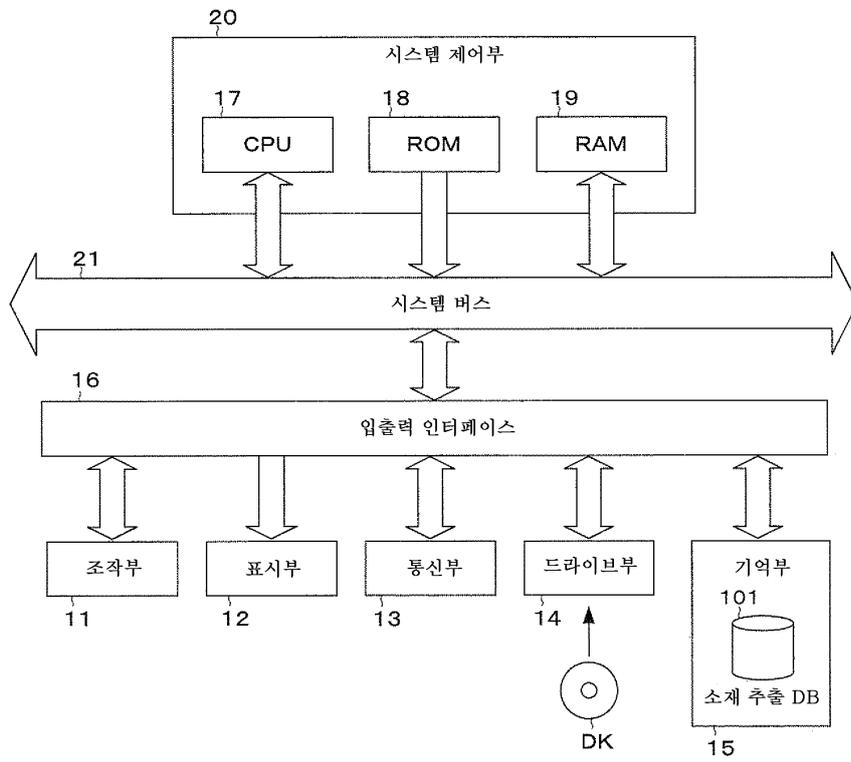
- [0240]
- 1 : 콘텐츠 생성 서버
  - 2 : 쇼핑 서버
  - 3 : 관리 단말기
  - 4 : 점포 단말기
  - 5 : 유저 단말기
  - 11 : 조작부
  - 12 : 표시부
  - 13 : 통신부
  - 14 : 드라이브부
  - 15 : 기억부
  - 16 : 입출력 인터페이스부
  - 17 : CPU
  - 18 : ROM
  - 19 : RAM
  - 20 : 시스템 제어부
  - 21 : 시스템 버스
  - 101 : 소재 추출 DB
  - 201 : 상품 상세 페이지 DB
  - NW : 네트워크
  - S : 쇼핑 시스템
  - 6 : 블로그 서버6
  - 61 : 조작부
  - 62 : 표시부
  - 63 : 통신부
  - 64 : 드라이브부
  - 65 : 기억부
  - 66 : 입출력 인터페이스부
  - 67 : CPU
  - 68 : ROM
  - 69 : RAM
  - 70 : 시스템 제어부
  - 71 : 시스템 버스
  - 601 : 블로그 페이지 DB
  - 602 : 광고 DB
  - BS : 블로그 시스템

도면

도면1

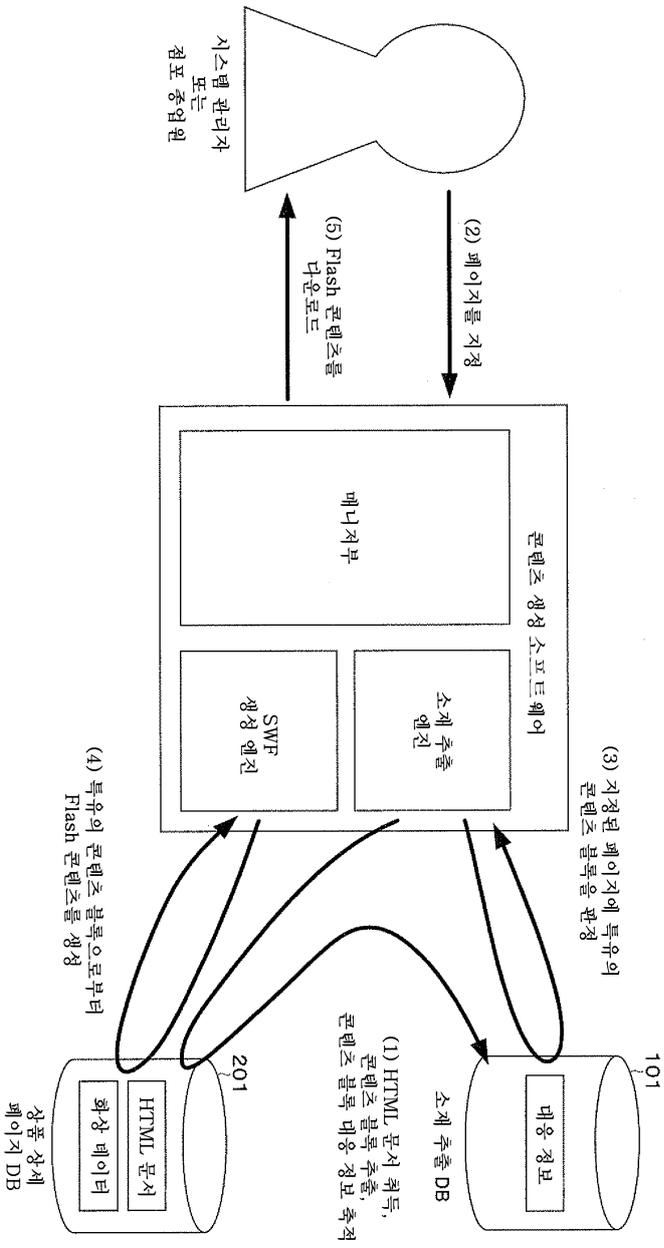


도면2

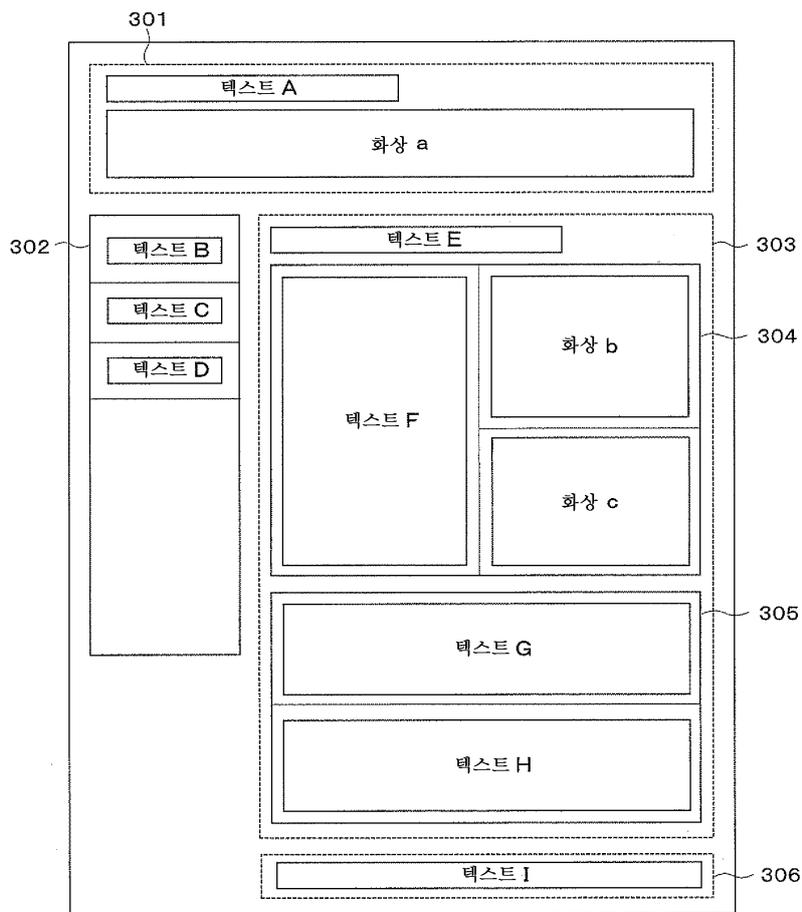


1

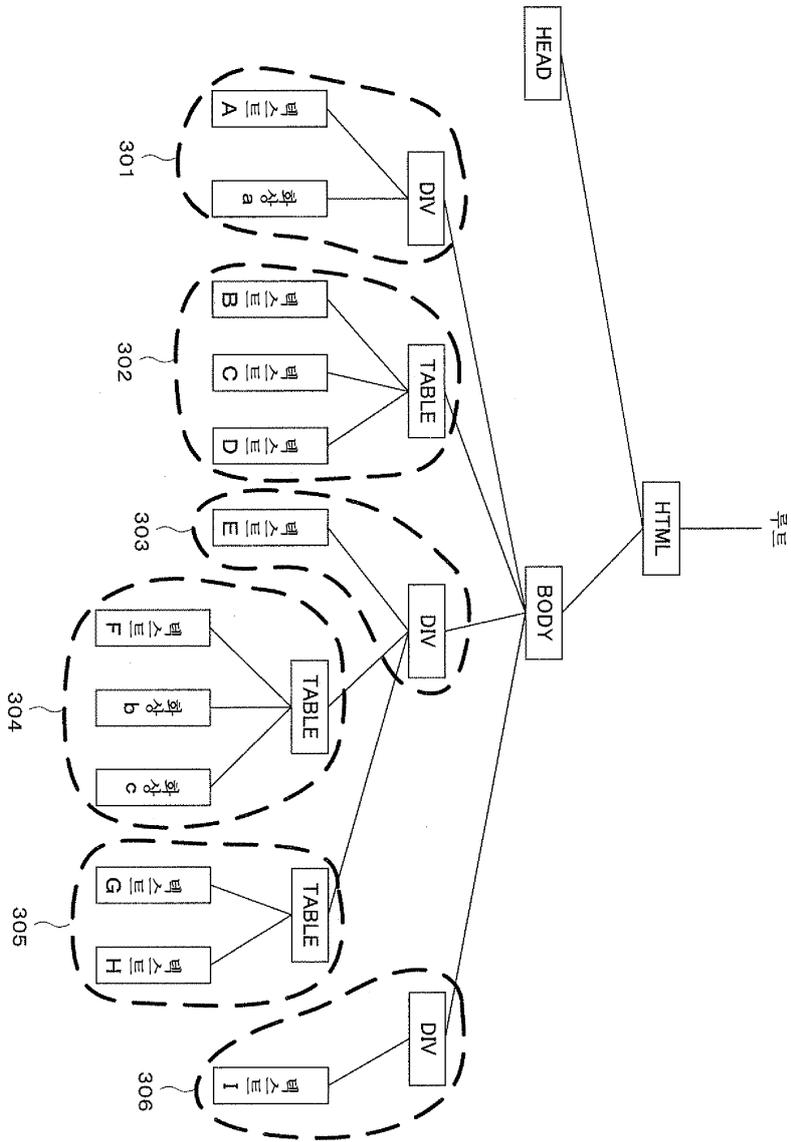
도면3



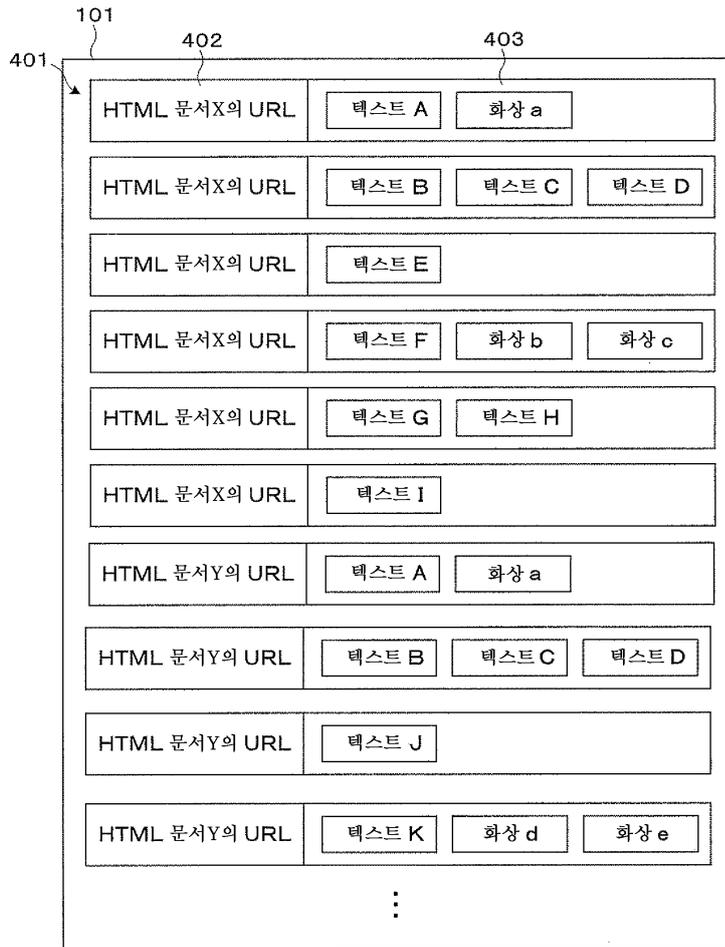
도면4



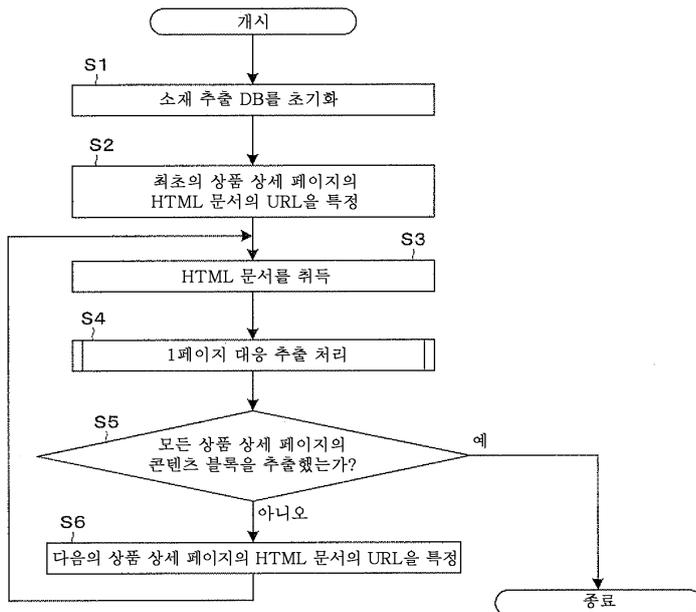
도면5



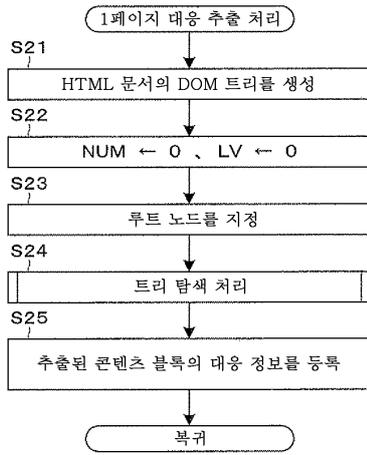
도면6



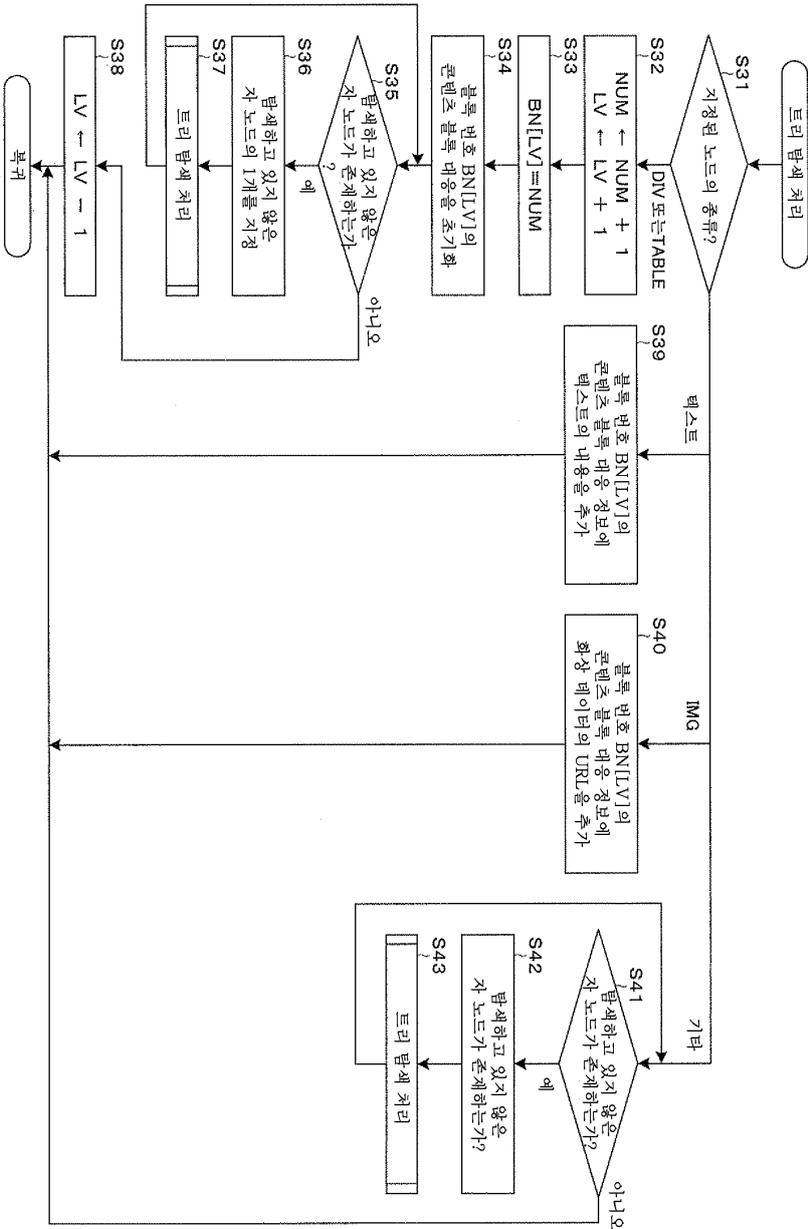
도면7



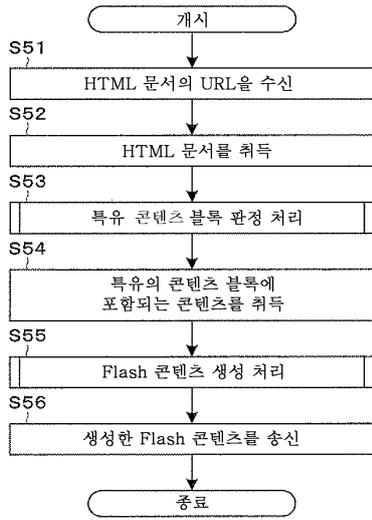
도면8



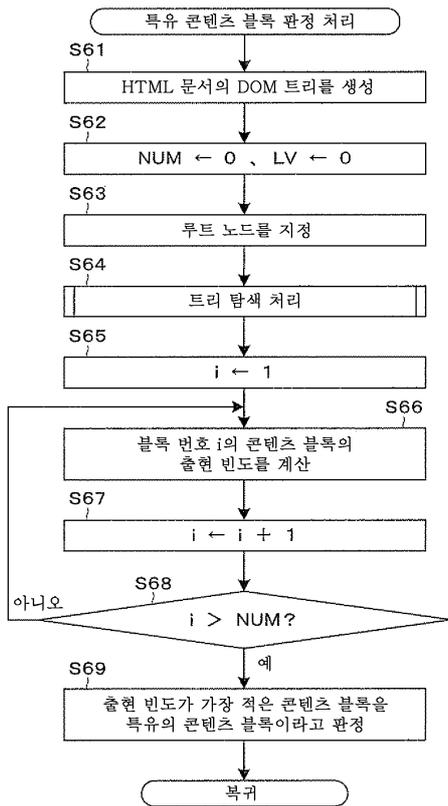
도면9



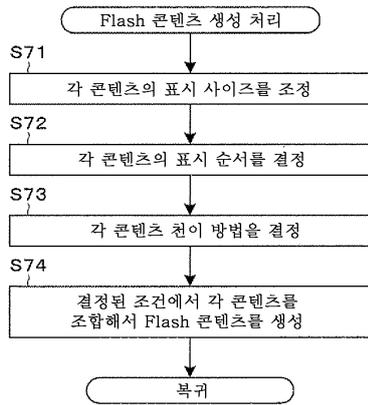
도면10



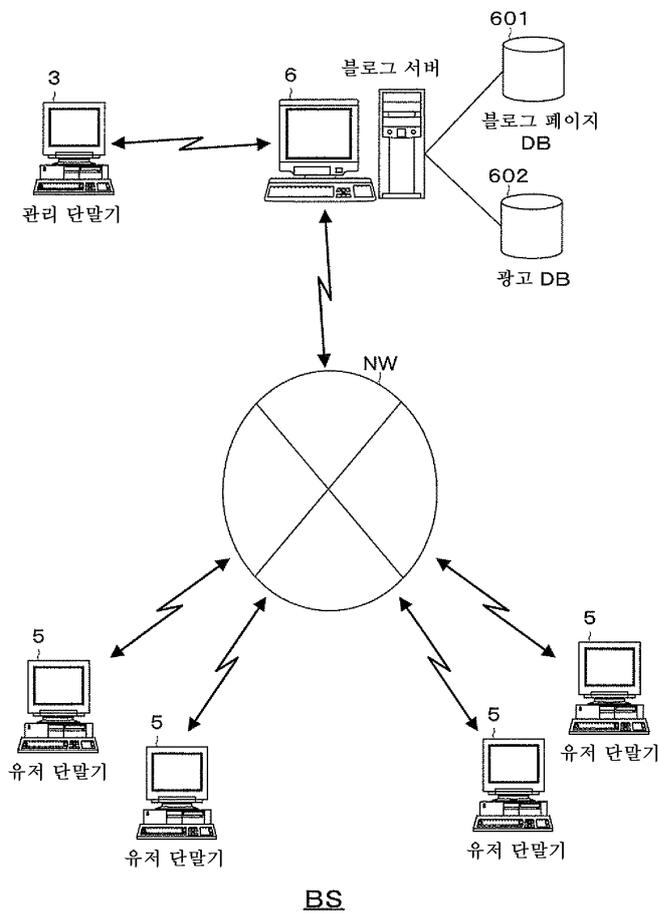
도면11



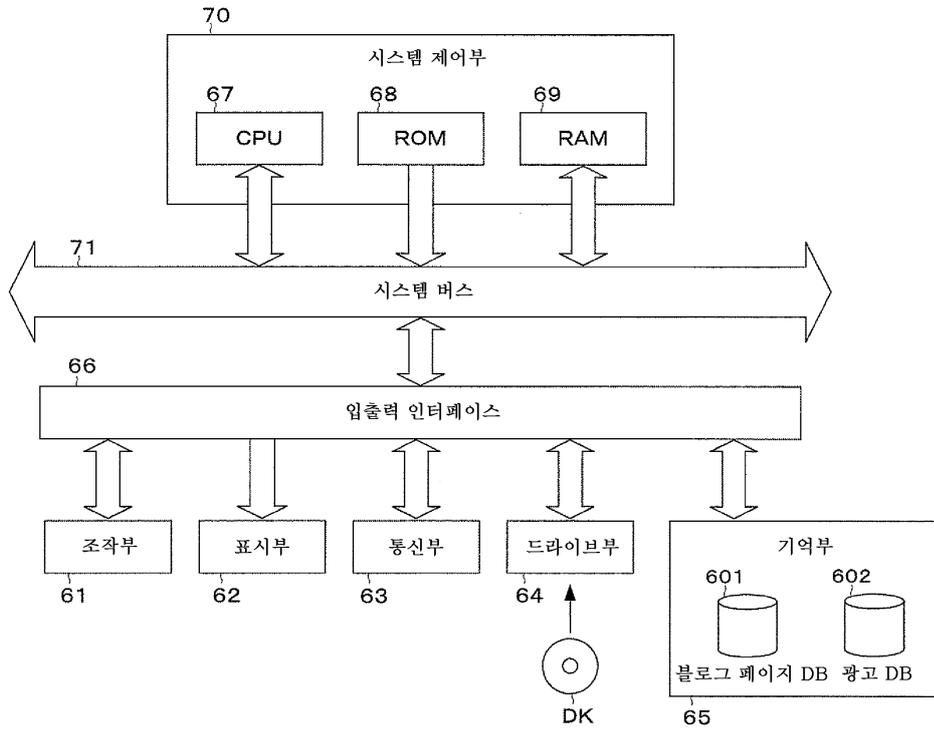
도면12



도면13

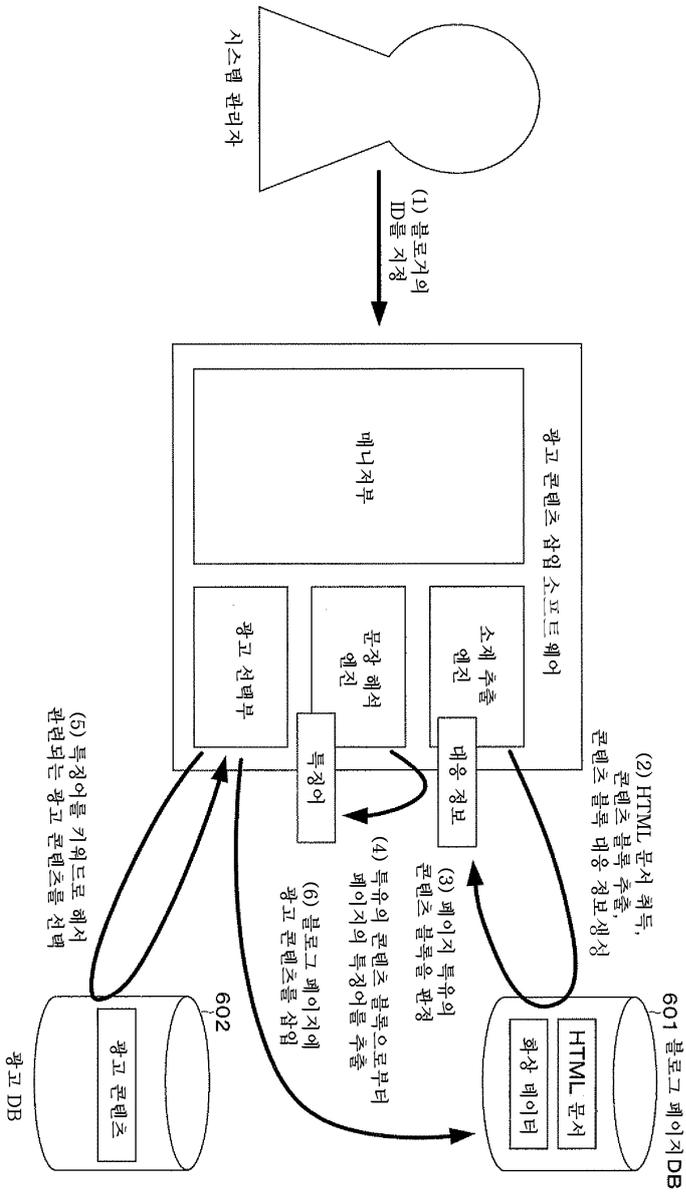


도면14

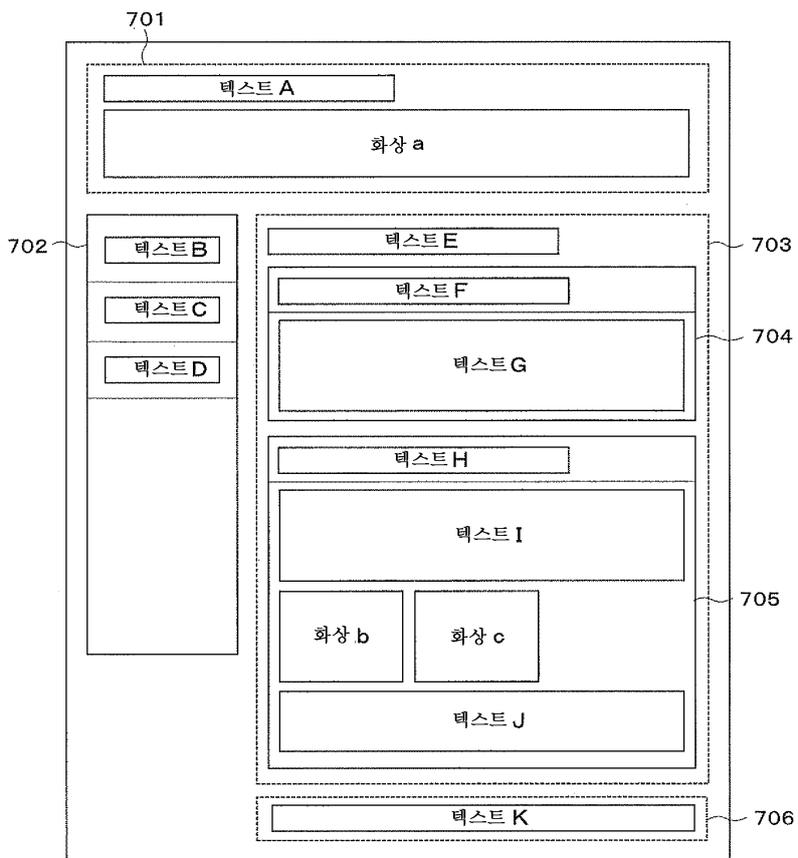


6

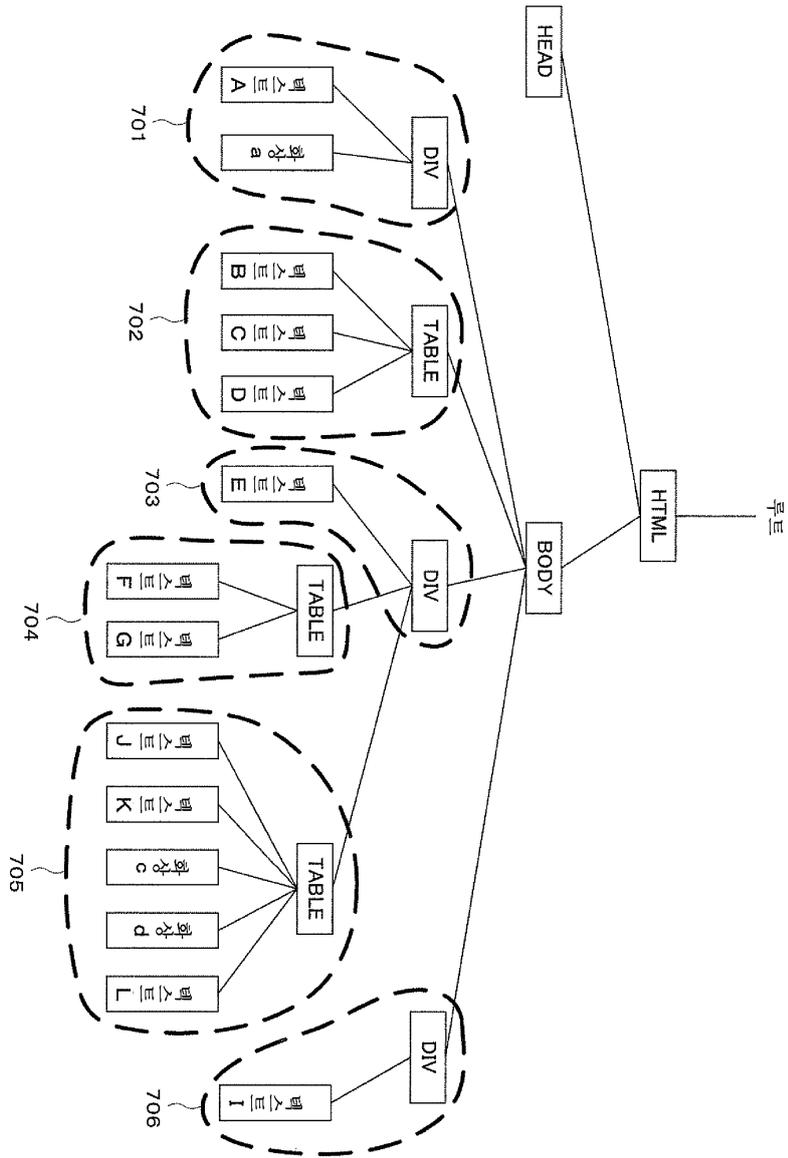
도면15



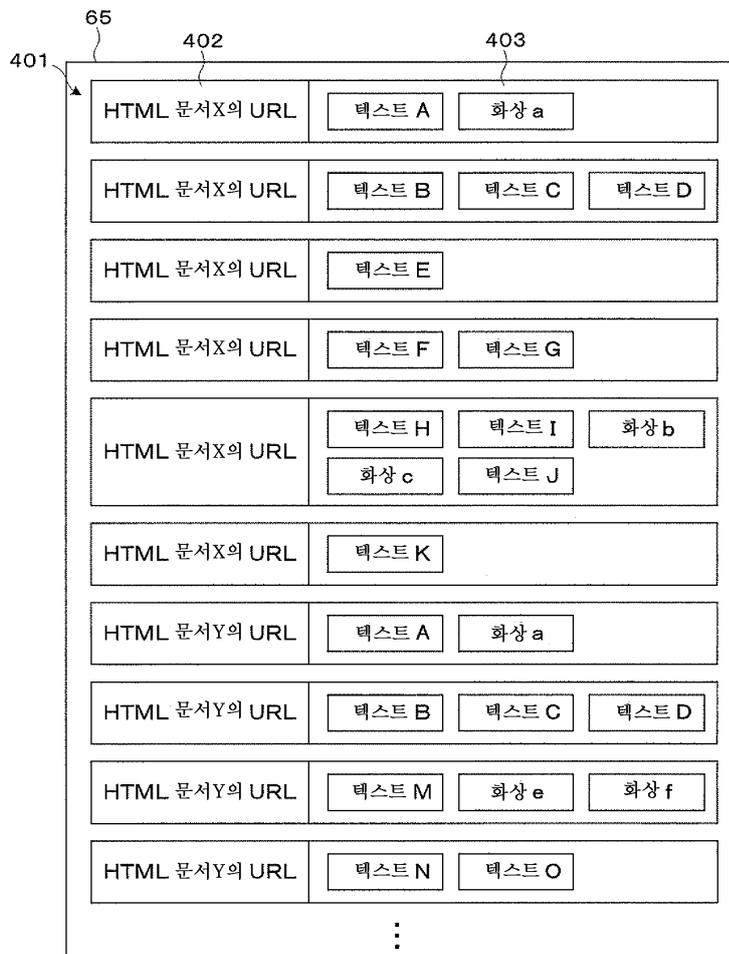
도면16



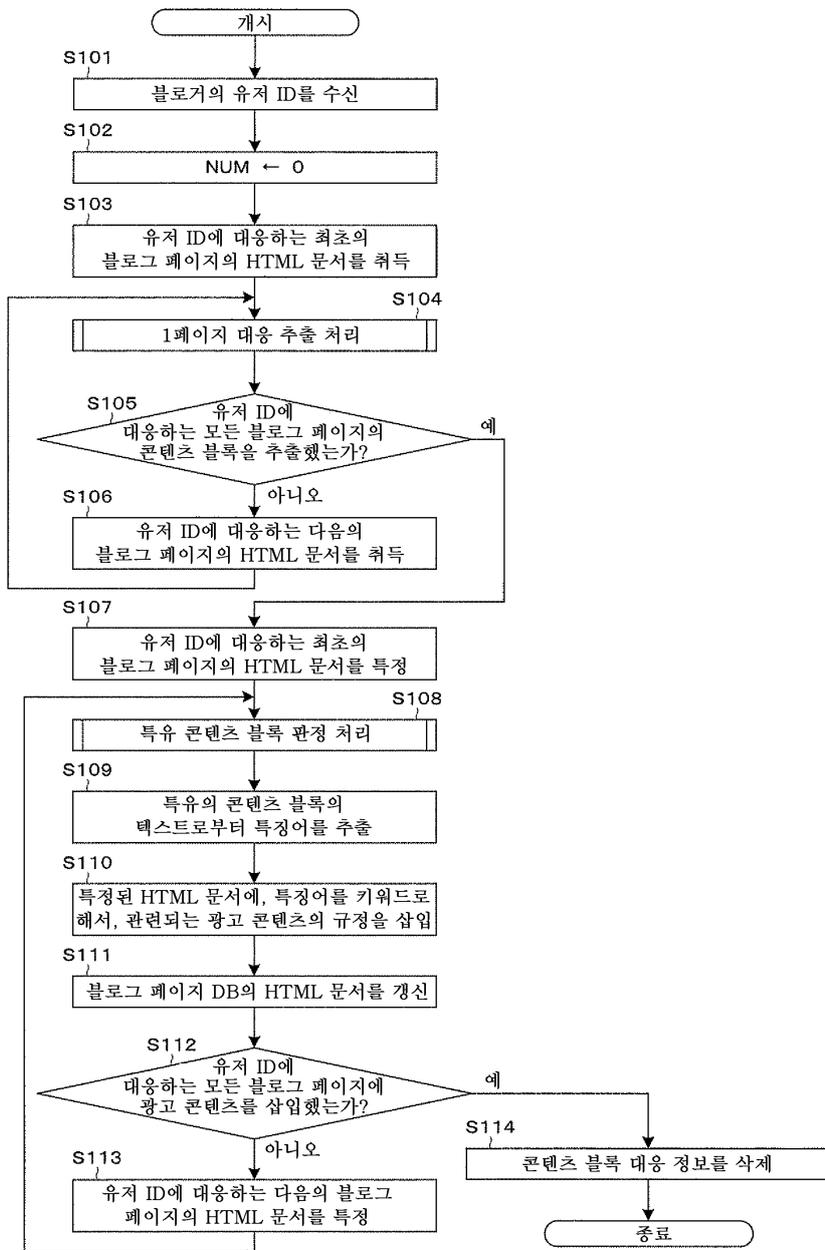
도면17



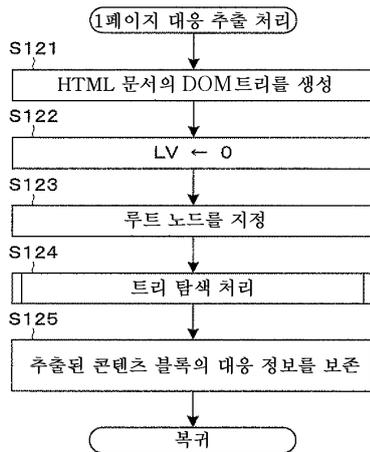
도면18



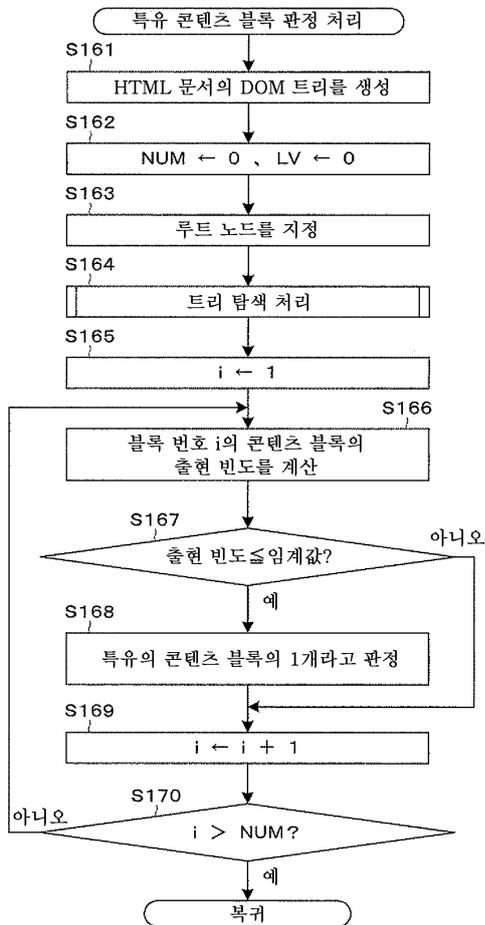
도면19



도면20



도면21



도면22

