

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5313295号
(P5313295)

(45) 発行日 平成25年10月9日(2013.10.9)

(24) 登録日 平成25年7月12日(2013.7.12)

(51) Int.Cl. F I
G06F 17/30 (2006.01) G06F 17/30 210D
 G06F 17/30 170A

請求項の数 7 (全 23 頁)

(21) 出願番号	特願2011-114168 (P2011-114168)	(73) 特許権者	505205812
(22) 出願日	平成23年5月20日 (2011.5.20)		エヌエイチエヌ コーポレーション
(62) 分割の表示	特願2008-260587 (P2008-260587) の分割		大韓民国 キュンギード・463-844 ・ソンナム-シ・ブンダン-グ・ジョンジ ャードン・25-1・ブンダン・ベンチャ ー・タウン
原出願日	平成20年10月7日 (2008.10.7)	(74) 代理人	110000408
(65) 公開番号	特開2011-154739 (P2011-154739A)		特許業務法人高橋・林アンドパートナーズ
(43) 公開日	平成23年8月11日 (2011.8.11)	(72) 発明者	郭 兌 榮
審査請求日	平成23年10月6日 (2011.10.6)		大韓民国ソウル廣津區廣壯洞 現代アパー トメント905-902
(31) 優先権主張番号	10-2007-0102231	(72) 発明者	李 恩 芝
(32) 優先日	平成19年10月10日 (2007.10.10)		大韓民国松坡區三田洞 55-5 402 号
(33) 優先権主張国	韓国 (KR)		

最終頁に続く

(54) 【発明の名称】 文書探索サービス提供方法及びシステム

(57) 【特許請求の範囲】

【請求項1】

分類された文書を文書探索サービス提供システムによって探索できる文書探索サービス
 を使用者端末機に提供する方法であって、

前記使用者端末機から使用者が関心を持つ主題に対する情報を受信し、

前記分類された文書のうち、前記関心を持つ主題に対する情報に関連された文書を抽出
 して、前記抽出された文書を前記使用者端末機に提供し、

前記文書の分類は、

文書を主題によって分類し、

前記文書の題目を抽出し、

前記抽出された題目に基づいてクラスターを形成し、

前記クラスターを前記主題に属する所定のディレクトリにマッピングすることからなる
 文書構造形成方法によって形成されることを特徴とする、文書探索サービス提供方法。

【請求項2】

前記文書は、少なくとも一つのフィールドを含み、

前記文書の題目の抽出は、前記文書を構成するフィールドの属性を考慮して前記題目を
 抽出することを特徴とする、請求項1に記載の文書探索サービス提供方法。

【請求項3】

前記クラスターの形成は、

前記抽出された題目を音節単位で区分し、前記題目のうち他の文書と共有される部分を

前記クラスターの中心概念候補として選定することを特徴とする、請求項 1 に記載の文書探索サービス提供方法。

【請求項 4】

請求項 1 乃至請求項 3 のうちいずれか 1 項に記載の文書探索サービス提供方法をコンピュータで実行するためのプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項 5】

ツリー構造で分類された文書を探索できる文書探索サービスを使用者端末機に提供するシステムであって、

前記文書の分類構造を形成するツリー構造形成部と；

前記分類構造で分類された文書のうち、前記使用者端末機から受信される使用者が関心を持つ主題に対する情報に関連された文書を抽出して、前記抽出された文書を前記使用者端末機に提供する文書提供部と；を含み、

前記ツリー構造形成部は、

前記文書を主題によって分類する文書分類部と；

前記文書の題目を抽出する題目抽出部と；

前記抽出された題目に基づいてクラスターを形成するクラスター形成部と；

前記クラスターを前記主題に属する所定のディレクトリにマッピングするディレクトリマッピング部と；を含む、文書探索サービス提供システム。

【請求項 6】

前記文書は、少なくとも一つのフィールドを含み、

前記題目抽出部は、前記文書を構成するフィールドの属性に基づいて前記題目を抽出することを特徴とする、請求項 5 に記載の文書探索サービス提供システム。

【請求項 7】

前記クラスター形成部は、

前記抽出された題目を音節単位で区分し、前記題目のうち他の文書と共有される部分を前記クラスターの中心概念候補として選定することを特徴とする、請求項 5 に記載の文書探索サービス提供システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文書探索サービス提供方法及びシステムに関するものである。

【背景技術】

【0002】

ウェブ上では、多様な関心事に対する数多くの文書が存在する。使用者たちは、自身の望む情報に対する質疑語情報（クエリー）を検索エンジンに伝送することで、情報を取得することができる。しかしながら、使用者自身が関心を持つ主題に対する質疑語をその都度毎回入力することは、非常に面倒なことである。

【0003】

一方、質疑語入力などの手順を経ずに使用者自身の望む情報にアクセスするために、特定の分野に専門性を有するパーティカルサイト（vertical site（専門分野サイト））及びブログなどに接続し、該当の分野の最新情報を取得することができる。

【0004】

このようなパーティカルサイト及びブログに存在する各情報の水準は日々向上しており、パーティカルサイト及びブログは、該当の分野で最も速くかつ深みのある有益な情報を取得可能なメディアとして発展しつつある。

【0005】

しかしながら、多様なパーティカルサイト及びブログに分散されている情報を閲覧するために各サイトを訪問することも、使用者にとって不便である。これに対する補完策として、パーティカルサイト及びブログは、RSS フィード（Really Simple Syndication Feed；RSS Feed）を提供しており、これを購読するた

10

20

30

40

50

めにRSSリーダー（RSS Reader）などのプログラムが使用される。

【0006】

ところが、各RSSフィードは、互いに独立的に情報を提供し、同一の内容または極めて類似した内容の文書も別個の情報として取り扱うので、使用者には情報の探索・閲覧過程で効率性を向上させるための追加的な努力が要求される。

【発明の概要】

【発明が解決しようとする課題】

【0007】

そこで、本発明の目的は、ツリー構造で分類された文書を探索できる文書探索サービスを提供する方法にあたり、使用者の反応情報を考慮することでツリー構造を好適に更新する文書探索サービス提供方法及びシステムを提供することにある。

10

【0008】

また、本発明の他の目的は、文書探索サービスを用いる使用者の文書閲覧性向を反映してツリー構造を更新することで、より直観的でありながらも正確に調整されたツリー構造を形成する文書探索サービス提供方法及びシステムを提供することにある。

【課題を解決するための手段】

【0009】

本発明の一側面によると、ツリー構造で分類された文書を探索できる文書探索サービスを提供する方法であって、文書に対するアクセス手段を提供する段階と、アクセス手段に対する使用者反応情報を取得する段階と、反応情報を考慮してツリー構造を更新する段階とを含む文書探索サービス提供方法が提供される。

20

【0010】

文書に対するアクセス手段を提供する段階は、ツリー構造を視覚化するための手段を提供することができる。使用者反応情報は、文書のツリー構造が視覚化される手段に対する使用者の反応情報を収集することで取得される。

【0011】

文書探索サービス提供方法でツリー構造を更新する段階は、ツリー構造で使用者の反応頻度が所定値以下である連結を断絶させることができる。

【0012】

また、ツリー構造を更新する段階は、文書の閲覧者が入力した質疑語情報をさらに考慮してツリー構造を更新することができる。

30

【0013】

文書探索サービスを提供するために使用されるツリー構造は、文書を主題によって分類する段階と、文書の題目を抽出する段階と、抽出された題目に基づいてクラスターを形成する段階と、クラスターを主題に属する所定のディレクトリにマッピングする段階とを含むツリー構造形成方法によって形成される。

【0014】

このようなツリー構造形成方法において、文書は、少なくとも一つのフィールドを含み、文書の題目を抽出する段階は、文書を構成するフィールドの属性を考慮して題目を抽出することができる。

40

【0015】

一方、クラスターを形成する段階で、抽出された題目を音節単位で区分し、題目のうち他の文書と共有される部分を前記クラスターの中心概念候補として選定することができる。中心概念候補は、抽出された題目に対するn-gram分析を用いて選定される。

【0016】

また、本発明の他の側面によると、ツリー構造で分類された文書を探索できる文書探索サービスを提供するシステムであって、文書に対するアクセス手段を提供する文書提供部と、アクセス手段に対する使用者反応情報を取得する反応情報獲得部と、反応情報を考慮してツリー構造を更新するツリー構造更新部とを含む文書探索サービス提供システムが提供される。

50

【0017】

文書探索サービス提供システムにおいて、文書提供部は、ツリー構造を視覚化するための手段を提供することができ、反応情報獲得部は、文書提供部によって提供された視覚化手段に対する使用者の反応情報を収集することができる。

【0018】

文書探索サービス提供システムにおいて、ツリー構造更新部は、ツリー構造で使用者の反応頻度が所定値以下である連結を断絶させることができる。

【0019】

また、ツリー構造更新部は、文書の閲覧者が質疑語情報をさらに考慮し、文書のツリー構造を更新することができる。

10

【0020】

文書探索サービス提供システムは、ツリー構造を形成するツリー構造形成部を含むことができる。ツリー構造形成部は、文書を主題によって分類する文書分類部と、文書の題目を抽出する題目抽出部と、抽出された題目に基づいてクラスターを形成するクラスター形成部と、クラスターを主題に属する所定のディレクトリにマッピングするディレクトリマッピング部とを含むことができる。

【0021】

一方、文書は、少なくとも一つのフィールドを含み、題目抽出部は、文書を構成するフィールドの属性を考慮して題目を抽出することができる。

【0022】

また、クラスター形成部は、抽出された題目を音節単位で区分し、題目のうち他の文書と共有される部分をクラスターの中心概念候補として選定することができる。中心概念候補は、抽出された題目に対する $n - g r a m$ 分析を用いて選定される。

20

【0023】

一方、本発明の文書探索サービス提供方法は、コンピュータによって行われ、コンピュータで実行するためのプログラムを記録するコンピュータ可読記録媒体に記録される。

【0024】

その他の側面、特徴及び利点は、添付された図面、特許請求の範囲及び発明の詳細な説明によって明確になるだろう。

【発明の効果】

30

【0025】

本発明の好適な一実施例によると、ツリー構造で分類された文書を探索できる文書探索サービスを提供する場合において、使用者の反応情報を考慮することで好適にツリー構造を更新することが可能な文書探索サービス提供方法及びシステムを提供することができる。

【0026】

また、本発明の好適な一実施例によると、文書探索サービスを用いる使用者の文書閲覧性向を反映してツリー構造を更新することで、より直観的でありながらも正確に調整されたツリー構造を形成する文書探索サービス提供方法及びシステムを提供することができる。

40

【図面の簡単な説明】

【0027】

【図1】本発明の一実施例に係る文書探索サービス提供方法のフローチャートである。

【図2】本発明の一実施例に係る文書探索サービス提供システムの構成図である。

【図3】本発明の一実施例に係る更新前の文書ツリー構造を例示した図である。

【図4】本発明の一実施例に係る更新後の文書ツリー構造を例示した図である。

【図5】本発明の一実施例に係る更新前の文書探索サービス提供画面を例示した図である。

。

【図6】本発明の一実施例に係る更新後の文書探索サービス提供画面を例示した図である。

。

50

【図7】本発明の一実施例に係る広告表示領域を含む文書探索サービス提供画面を例示した図である。

【発明を実施するための形態】

【0028】

以下、本発明に係る文書探索サービス提供方法及びシステムの実施例を添付された図面に基づいて詳細に説明する。本発明は、特定の実施形態に限定されるものでなく、本発明の技術的思想及び技術範囲に含まれる全ての変更、均等物及び代替物を含むものとして理解されるべきである。本発明の説明において、関連した公知技術に対する具体的な説明が本発明の要旨を不明確にすると判断される場合には、それに対する詳細な説明を省略している。また、添付された図面に基づく説明において、同一の構成要素または対応する構成要素には同一の図面番号を付与し、それに対する重複的な説明を省略する。

10

【0029】

図1は、本発明の一実施例に係る文書探索サービス提供方法の処理フローを示すフローチャートであり、図2は、本発明の一実施例に係る文書探索サービス提供システムの構成図である。

【0030】

図1及び図2を参照すると、本発明の一実施例に係る文書探索サービス提供システムは、探索サービス提供サーバ210、文書分類部211、題目抽出部212、クラスター形成部213、ディレクトリマッピング部214、文書提供部215、反応情報獲得部216、ツリー構造更新部217、原本文書データベース221、探索サービスデータベース222、反応情報データベース223、使用者端末機230及びツリー構造形成部240を含んでいる。

20

【0031】

ツリー構造形成段階(S110)は、文書探索サービス提供方法が適用された文書探索サービス提供システムによって提供される各文書間の連結関係を表現するツリー構造を形成する段階である。ツリー構造形成段階(S110)は、ツリー構造形成部240によって行われる。

【0032】

ツリー構造形成段階(S110)は、文書を主題によって分類する段階(S112)と、分類された各文書の題目を抽出する段階(S114)と、抽出された題目に基づいて文書のクラスターを形成する段階(S116)と、文書のクラスターをディレクトリにマッピングする段階(S118)とを含むことができ、以下、各段階の細部的な動作を説明する。

30

【0033】

文書を主題によって分類する段階(S112)は、文書分類部211が原本文書データベース221内の各文書を分類する段階である。

【0034】

文書分類部211は、原本文書データベース221から文書に対する情報を取得し、分類の構造に関する情報などを探索サービスデータベース222から取得する。取得した各情報に基づいて文書がマッチングされる分類を決定し、文書と分類との間のマッチング関係に対する情報を探索サービスデータベース222に保存する。

40

【0035】

本段階で、文書分類部211は、文書に含まれた情報を使用して文書が特定のキーワード及び特定の内容を含んでいるかを判断することで、文書を主題別に分類することができる。

【0036】

一例として、'ワイン'という分類とマッチングされるのに適した文書は、'ワイン'という分類名自体及び分類名の同義語がその文章に含まれているか、及び'ワイン'と深い関係があると判断されるキーワードである'ソムリエ'、'デキャンティング'などがその文書に含まれているかなどを考慮して決定される。

50

【 0 0 3 7 】

一方、文書と分類との間のマッチング可否を決定する場合において、分類関連キーワードの包含可否を数値化し、これを基準として活用することができる。一例として、特定の主題と関連した各キーワードが含まれる場合、所定の点数を付与し、これら点数の合計が一定の基準を越える場合、該当の分類とマッチングされると決定することができる。

【 0 0 3 8 】

なお、文書の分類段階において一つの文書が必ずしも一つの分類のみにマッチングされると判断されることはない。一例として、'ワイン'という分類と'日本漫画'という分類が存在する場合、ワインを主題とした日本漫画である'神の雫'に対する鑑賞評(レビュー)などを取り扱う文書は、'ワイン'に対する分類だけでなく、'日本漫画'という分類にも同時にマッチングされる。

10

【 0 0 3 9 】

原本文書データベース221には、探索サービス提供サーバー210によって分類・再構成される各原本文書に対する情報が保存される。これら原本文書は、最終的にユーザー端末機230に提供される。原本文書は、ウェブ上でウェブロボットなどによって収集される。

【 0 0 4 0 】

一方、本発明の原本文書データベース221に保存される各原本文書は、所定の属性を有する文書を含むことができる。一例として、パーティカルサイト及びブログで使用される構造化された文書が原本文書として使用される。

20

【 0 0 4 1 】

このような構造化された各文書は、文書内のコンテンツを少なくとも一つの領域または区画に分けて保存することができる。これら領域または区画は、フィールドと命名される。一例として、ブログのポストのような文書は、題目フィールド、本文フィールド、作成時刻フィールド及び該当のポストに対するキーワードフィールドなどを含むことができる。

【 0 0 4 2 】

このような文書に関して、その作成者は、それぞれのフィールド名に対応するコンテンツを入力することで文書を生成できるので、フィールド名及びそれに該当するコンテンツは、後述する題目抽出段階などで有用に使用される。

30

【 0 0 4 3 】

また、このようなパーティカルサイト及びブログでは、各文書の間関係も構造化されている可能性がある。このような文書の構造化された関係は、該当のサイトでのディレクトリ形態で表れる。

【 0 0 4 4 】

一例として、映画を主題とするパーティカルサイトは、当該サイトの文書を分類するディレクトリとして'映画鑑賞評'、'映画順位'及び'最新公開作'などのディレクトリを含むことができ、ブログもそれぞれのポストを分類するディレクトリに関する情報を有することができる。

【 0 0 4 5 】

このようなパーティカルサイト及びブログでのディレクトリ名は、そのサイトが取り扱う主題と関連したキーワードとして使用される。これらキーワードは、上述した文書の主題別分類段階で活用されるもので、その分類の正確度を向上させるために使用される。

40

【 0 0 4 6 】

本発明において、文書という用語は、電子的に記録された文書を通称する用語として理解される。文書は、HTMLなどのマークアップランゲージを使用して記述され、「* . h t m」などの拡張子を有するが、特定の記述形態及び拡張子を有するファイルに限定されるものとして解析されることはない。

【 0 0 4 7 】

探索サービスデータベース222には、文書分類部211によって決定された文書と分

50

類との間のマッチング関係に対する情報が保存される。すなわち、探索サービスデータベース222は、文書別に各分類に対するマッチング可否を保存し、各分類別関連キーワードの包含可否を数値化して保存することができる。

【0048】

一方、上述した原本文書データベース221及び探索サービスデータベース222での情報保存方法及び形態などは、本発明の目的範囲内で多様に変化される。

【0049】

各文書の題目を抽出する段階(S114)は、題目抽出部212が原本文書データベース221に保存された各文書の題目を抽出する段階である。文書の題目とは、文書の内容及び主題を含んでいる単語、句または文章を意味する。

10

【0050】

題目抽出部212は、原本文書データベース221に保存された文書の情報を用いて各文書の題目を抽出し、抽出された題目を探索サービスデータベース222に保存する。

【0051】

本段階で、題目抽出部212は、各文書に含まれた情報を用いて文書の題目を抽出することができる。文書の構造、文書に含まれた各単語の出現頻度及び文書が使用者端末機230でブラウジングされる場合の属性などは、題目決定の基準として使用される。

【0052】

すなわち、題目を抽出する過程で使用される文書情報は、文書に直接的に含まれたコンテンツテキストのみならず、文書が使用者端末機230で閲覧される形態に関する情報などを含む概念として理解される。

20

【0053】

一例として、ブログなどのウェブサイトは、構造化された各文書を含むことができる。これら文書は、少なくとも一つの名を有するフィールドに情報を保存することができる。すなわち、'題目'、'(title)'などの名を有するフィールドに含まれたテキストを題目として選定することができる。

【0054】

他の例として、文書が使用者端末機230のウェブブラウザなどを通してブラウジングされる場合、文書内の他の内容より相対的に大きく表示されたり、差別化される属性を有して表現されて強調されるテキストも題目の候補句として考慮される。

30

【0055】

抽出された題目に基づいて文書のクラスターを形成する段階(S116)は、クラスター形成部213が文書の題目情報に基づいてクラスタリングを行う段階である。

【0056】

クラスター形成部213は、探索サービスデータベース222から取得した文書の題目情報に基づいて各文書のクラスターを形成する。形成されたクラスターに関する情報は、探索サービスデータベース222に保存される。

【0057】

文書のクラスターは、中心概念を共有する各文書のグループを意味する。文書のクラスターは、各文書の題目における共通部分の存在可否を考慮して形成される。各クラスターの名称は、この中心概念を用いて命名される。

40

【0058】

文書の題目のうち、他の文書と共通される文字列がクラスターの中心概念の候補となり、共通の文字列を有する文書の個数が所定値以上である場合、一つの独立したクラスターが形成される。

【0059】

一例として、一つの文書の題目が'ソムリエ追跡：ワインエチケット - 一緒に楽しむワイン'で、他の文書の題目が'テーブルマナー5編 - ワインエチケット'である場合、二つの題目で共通される部分である'ワインエチケット'が中心概念として抽出される。

【0060】

50

文書の題目における重複部分を中心概念として抽出する過程で、n - g r a m分析方法を使用することができる。この場合、題目は、音節単位で分離され、所定個数の音節を有する文字列に再び組み合わされる。

【 0 0 6 1 】

このように再び組み合わされた文字列のうち、重複部分を中心概念の候補となり得る。上述した例示の場合、二つの音節を有する‘ワイン’と5個の音節を有する‘ワインエチケット’が中心概念の候補として抽出される。

【 0 0 6 2 】

なお、上記のように、各文書の題目で同一の重複部分が多数個ある場合、そのうち一つの重複部分を中心概念として決定する過程が要求されるが、このとき、重複部分の音節数、重複部分と文書の分類名との関係、及び該当の重複部分を有する文書の数などが決定基準として使用され、一つの中心概念を決定することができる。

10

【 0 0 6 3 】

上述した例において、重複部分は、‘ワイン’、‘エチケット’及び‘ワインエチケット’である。この場合、‘ワイン’は、各文書が含まれた主題である‘ワイン’と同一であるので、一つのクラスターの中心概念としては適切でない。

【 0 0 6 4 】

重複部分を有する各文書の個数の側面でも、‘ワイン’を共有する文書の個数は、一つのクラスターに形成するには過度に大きい値である。このようにクラスターの中心概念の候補を決定する場合、その中心概念の候補を共有する各文書の個数を所定の範囲に制限することが要求される。

20

【 0 0 6 5 】

また、中心概念候補の長さも考慮対象になり得る。過度に短い中心概念候補の場合、その品詞が助詞であったり、特定の分類に使用されるのに不適切な一般的な用語である可能性がある。

【 0 0 6 6 】

一方、長い中心概念候補の場合、その候補を共有するクラスター内の各文書の間に関連度が高いこと、及びノイズが含まれる確率が小さいことが期待されるので、中心概念候補のうち最も長い候補を優先的に考慮することができる。

【 0 0 6 7 】

上述した例の場合、残りの二つの中心概念候補が‘ワインエチケット’に含まれるので、まず、最も長い‘ワインエチケット’を候補として考慮することができ、‘ワインエチケット’を共有する文書数などの他の基準を満足すると判断される場合、単一のクラスターを構成する中心概念として選定される。

30

【 0 0 6 8 】

また、‘ワイン’という主題に含まれた文書で、‘エチケット’を共有する文書の個数と‘ワインエチケット’を共有する文書の数が極めて類似している場合、より具体的な‘ワインエチケット’を中心概念として選定することが効率的である。

【 0 0 6 9 】

これら中心概念のうち所定の個数以上の文書が関連したことを基準にしてクラスターを構成することができる。クラスターをなす中心概念及びその中心概念を共有するクラスターに属する文書に関する情報は、探索サービスデータベース222に保存される。

40

【 0 0 7 0 】

文書のクラスターをディレクトリにマッピングする段階(S118)は、ディレクトリマッピング部214が文書のクラスターを各クラスターの中心概念を基準にしてディレクトリにマッピングする段階である。

【 0 0 7 1 】

ディレクトリは、文書の分類、すなわち、主題の下位概念として一つまたは複数のクラスターを含むことができる小主題を意味する。一例として、文書が分類された主題が‘ワイン’である場合、そのディレクトリとして‘ワインの産地’、‘ワインの歴史’及び‘

50

ワインエチケット'などが含まれ、'ワインの産地'ディレクトリは、ワインの生産地として知られた'ボルドー'及び'ブルゴーニュ'などの地名が中心概念として形成されたクラスターを含むことができる。

【0072】

ディレクトリマッピング部214は、探索サービスデータベース222からディレクトリ構造に関する情報及びクラスターに対する情報を取得し、各ディレクトリにマッピングされるクラスターを決定する。決定されたディレクトリマッピング情報は、探索サービスデータベース222に保存される。

【0073】

クラスターがマッピングされるディレクトリは、該当のクラスターの中心概念がそのディレクトリと関連したキーワードを含んでいるかによって決定される。

10

【0074】

一例として、ディレクトリが'ワインエチケット'である場合、ディレクトリ名では、'エチケット'がディレクトリの包含可否を決定するためのキーワードになり得る。既に'ワイン'分類に該当すると判断された各文書に対して形成されたクラスターをマッピングする過程では、分類名である'ワイン'自体を除いたキーワードでディレクトリをマッピングすることが効率的である。

【0075】

一方、これらキーワードに対しては、辞典式羅列法を使用して該当のキーワードを拡張することができる。'エチケット'は、同義語、類義語及び表記言語を異にするキーワードに拡張される。

20

【0076】

この場合、'礼節'、'etiquette'、'マナー'及び'manner'などのキーワードがディレクトリマッピングのための追加的なキーワードとして考慮される。これを通じて、ディレクトリマッピングの効率性を向上させることができる。

【0077】

このようなディレクトリマッピングのためのキーワードも、ディレクトリ構造に関する情報の一部として探索サービスデータベース222に保存される。

【0078】

以上では、ツリー構造形成段階(S110)の各段階の細部的な動作を説明した。形成されたツリー構造は、分類される文書自体に含まれる情報を用いて形成される。この場合、ツリー構造は、好ましくない文書、クラスター及びそれらの連結を含むことができる。このようにノイズを含むツリー構造は、更新前のツリー構造として図3に基づいて説明される。

30

【0079】

文書を提供する段階(S120)は、文書提供部215が使用者端末機230にクラスタリングされた各文書をディレクトリ別に提供する段階である。

【0080】

本段階は、所定の主題、すなわち、分類に属するディレクトリ構造及びディレクトリに属するクラスターの包含関係を視覚化して提供することで、使用者が自身の関心分野の各文書を容易に探索できることに特徴がある。

40

【0081】

使用者は、使用者端末機230を通じて自身が関心を持つ主題に関する情報を探索サービス提供サーバー210に伝送する。これは、該当の主題に対する探索サービスを提供するウェブページに対するリンクのクリック動作などによって行われる。

【0082】

文書提供部215は、使用者の関心主題に関する情報が含まれた使用者端末機230からの要請を受信し、これに対する応答として、上述した各段階で分類されてクラスタリングされた各文書に対するアクセスリンクを含むウェブページを提供することができる。これを通じて、探索サービスが使用者に提供される。

50

【 0 0 8 3 】

一方、文書探索サービスを提供するために、文書提供部 2 1 5 は、クラスタリングされた各文書に対する情報を探索サービスデータベース 2 2 2 から取得する。

【 0 0 8 4 】

文書提供部 2 1 5 の応答が提供される形態及び様式は、CSS (cascading style sheets) などを用いて調節される。また、別途のコンテンツマネジメントシステム (content management system: CMS) を用いることも可能である。

【 0 0 8 5 】

使用者の反応情報を取得する段階 (S 1 3 0) は、反応情報獲得部 2 1 6 が使用者の反応情報を取得する段階である。

10

【 0 0 8 6 】

本実施例において、探索サービスは、文書提供部 2 1 5 によって視覚化されたツリー構造を使用者に提供することができる。このようなツリー構造の視覚化は、図 5 及び図 6 に示した分類構造表示領域 5 1 0、探索位置表示領域 5 2 0 及びクラスター表示領域 5 3 0 などで行われる。

【 0 0 8 7 】

このような領域で、使用者は、各領域において自身が探索しようとするディレクトリ、クラスター及び文書と関連した部分をクリックすることができる。そして、使用者がこのように自身の探索位置を決定して文書を閲覧する使用者のアクション、すなわち、文書を閲覧するための使用者のツリー構造で分類された文書に対する反応を、当該使用者の文書に対する反応に関する情報として反応情報取得部 2 1 6 が収集する。

20

【 0 0 8 8 】

取得された反応情報は、反応情報データベース 2 2 3 に保存され、使用者反応情報は、使用者に対する識別情報、反応時刻及び反応対象に対する情報を含むことができる。このように収集・保存された使用者反応情報は、ツリー構造を更新する基礎資料として活用される。

【 0 0 8 9 】

一方、文書提供部 2 1 5 は、ツリー構造の文書に含まれた情報を検索するための検索手段を提供することができる。このような検索手段は、文書提供部 2 1 5 が使用者端末機 2 3 0 に提供するウェブページなどに含まれた検索ウィンドウなどの形態で提供される。このような検索ウィンドウに入力された質疑語も、使用者の反応情報としてツリー構造の更新に使用される。

30

【 0 0 9 0 】

文書のツリー構造を更新する段階 (S 1 4 0) は、ツリー構造更新部 2 1 7 が文書のツリー構造を更新する段階である。ツリー構造更新部 2 1 7 は、反応情報データベース 2 3 2 に保存された使用者の反応情報に基づいて文書のツリー構造を更新することができる。ツリー構造の更新は、不必要なディレクトリ、クラスター及び文書をツリーから排除する形式などで行われる。

【 0 0 9 1 】

一例として、特定のクラスターに属する文書に対するリンクが使用者に持続的に提供されていたにもかかわらず、使用者による閲覧頻度が他の文書に比べて相対的に低い場合、該当の文書とクラスターとの連結を断絶 (解除) することで、ツリー構造を更新することができる。

40

【 0 0 9 2 】

これと同様に、特定のクラスターに対する使用者の反応頻度が相対的に低い場合、該当のクラスターをディレクトリから排除 (削除) することも可能である。

【 0 0 9 3 】

また、含まれた下位分類に対する使用者の閲覧頻度を考慮し、上位分類を変更することができる。一例として、特定のディレクトリに含まれた各クラスターに対する使用者の訪

50

問頻度が過度に低い場合、そのディレクトリをツリー構造から排除（削除）することができる。

【0094】

このようなツリー構造の変更は、ツリー構造の各構成要素の間の連結程度を数値化することで行われる。一例として、使用者が頻繁に用いる探索経路に対して所定の点数を付与することで、頻繁に用いられる探索経路に含まれた各連結に対して高い点数を付与することができ、その反対に、該当の連結に対して低い点数を付与することができる。この場合、所定値以下の点数を有する各連結を断絶（解除）対象の候補として選定する方法によってツリー構造の更新が可能である。

【0095】

また、特定の文書に対して使用者たちがアクセスした探索経路の比率も、ツリー構造を更新する情報として考慮される。文書が二つ以上の上位概念と関連している場合、使用者の流入経路（アクセス経路）が特定の経路に偏重されていると、その経路のみを残す形態でツリー構造を更新することができる。特定の経路で流入（アクセス）する比率が極めて低い場合には、その経路を排除することも可能である。

【0096】

このように、探索経路に対する使用者の照会頻度及び特定の文書と関連した流入経路などは、文書分類部215によって提供された情報に対する使用者の反応情報に基づいて算出される。

【0097】

一方、ツリー構造は、ツリー構造に属する文書を閲覧した閲覧者たちが入力した質疑語情報をさらに考慮して更新される。ウェブ上の文書を検索して閲覧する場合、その文書を含む検索結果を入力するために入力した質疑語は、その文書と関連しているものと判断される。一例として、検索エンジンで‘ボルドーワイン’という質疑語を入力した使用者が閲覧した文書は、‘ボルドーワイン’と関連した文書であると判断される。

【0098】

一方、ツリー構造は、新しいディレクトリ及び新しいクラスターを形成することによっても更新される。原本文書データベース221に新しい文書が追加される場合、これら新しい文書は、既存のツリー構造に新しく連結される。

【0099】

上述したように、文書のクラスターを形成する場合において、クラスターに含まれた文書の個数も考慮対象になり得る。新しい文書の追加によって特定のクラスターに含まれた文書の個数が過度に大きくなる場合、該当のクラスターを多数個のクラスターに分離することも可能である。この場合、クラスターをディレクトリに変更することも考慮される。

【0100】

このように更新されたツリー構造は、探索サービスデータベース222に保存される。また、更新されたツリー構造は、文書提供部215によって使用者に提供され、これを通じて、使用者は、より正確な分類結果を有するツリー構造を探索することができる。

【0101】

一方、本発明の一実施例に係る探索サービス提供サーバー210は、広告提供部218

【0102】

をさらに含むことができる。広告提供部218は、広告コンテンツを使用者端末機230に提供することができる。広告コンテンツは、広告データベース224に保存されるもので、広告提供部218によって呼び出されて使用者端末機230に伝送される。

【0103】

使用者端末機230に伝送される広告コンテンツを決定する要素として、使用者に関する情報及び使用者が探索する文書に関する情報などが考慮される。

【0104】

一例として、使用者が文書探索サービスを用いる過程でログイン手順を行った場合、使

10

20

30

40

50

用者の年齢、職業、性別及び居住地域などの使用者の個人情報広告コンテンツ決定要素として考慮される。

【0105】

一方、広告コンテンツ決定要素として、使用者端末機230を通じて閲覧される文書に関する各情報が考慮され、使用者が入力した質疑語情報も考慮される。

【0106】

また、使用者が本発明の一実施例に係る文書探索サービスを用いる過程で取得される各情報も、コンテンツ決定要素として考慮される。

【0107】

このように、使用者に関する情報及び使用者が探索する文書に関する情報などを用いて提供される広告コンテンツを決定することで、提供される広告の効果が極大化されるという長所がある。

10

【0108】

図3は、本発明の一実施例に係る更新前の文書ツリー構造を例示した図で、図4は、本発明の一実施例に係る更新後の文書ツリー構造を例示した図である。

【0109】

図3及び図4を参照すると、原本文書データベース221に属する各文書は、分類(主題)、ディレクトリ、クラスター及び文書の順に連結される階層構造によって構造化されている。

【0110】

20

図3及び図4を参照すると、各文書が構造化される上位概念である分類(主題)は'ワイン'である。特定の文書が'ワイン'分類に該当するかどうかは、文書分類部211によって判断される。

【0111】

分類(主題)は、その下位概念として少なくとも一つのディレクトリを含むことができる。'ワイン'分類は、'ワインの産地'、'ワインの歴史'及び'ワインエチケット'と命名されたディレクトリを含む。

【0112】

ディレクトリの名称は、使用者が文書を探索しようとするグループの名として機能できるので、原本文書データベース221に保存された各文書の出处であるパーティカルサイト及びブログなどで使用される文書グループの名称をディレクトリ名として使用することで、使用者の文書探索効率を高めることができる。

30

【0113】

ディレクトリは、その下位に少なくとも一つのクラスターを含むことができる。'ワインの産地'ディレクトリは、'ボルドー'、'ブルゴーニュ'、'カンパーニャ'及び'ボルドーTV'と命名されたクラスターを含む。

【0114】

クラスターも少なくとも一つの文書を含むことができ、図3に例示した文書2の場合、その題目に'ボルドー'及び'ブルゴーニュ'を含むので、'ボルドー'クラスターと'ブルゴーニュ'クラスターの全てに含まれる。

40

【0115】

一方、パーティカルサイト及びブログで使用されるディレクトリの名称及びこのようなサイトに含まれた構造化された各文書のフィールド情報が、文書を主題別に分類してクラスタリングするために使用可能であることは、図1及び図2の詳細な説明で述べた通りである。

【0116】

一方、'ワインに似たボルドーTV'という題目を有する文書3は、ワインに対する文書ではない韓国の電子製品業体で生産したTVに関する文書で、ワインに関する文書として分類されるのに不適切なノイズ文書である。

【0117】

50

しかしながら、文書3は、その題目に‘ワイン’及び‘ボルドー’という文字列を含み、文書のコンテンツとしてワインと関連したマーケティング活動、製品開発コンセプトを含むので、‘ワイン’分類で別途のクラスターに含まれる結果をもたらす。

【0118】

これら文書は、ワインに対する関心を持って文書を探索する訪問者の注意を引くことができないので、他の文書に比べて低い照会数が記録されるようになる。このような訪問者の反応情報は、ツリー構造の更新に反映される。

【0119】

一方、上述したように、文書の閲覧者が入力した質疑語情報もツリー構造の更新に反映される。文書3を閲覧する使用者が入力した質疑語の分布が‘ワイン’や‘ボルドー’でなく、製品の生産者である‘三星電子’や‘TV’に偏重されている場合、これをツリー構造に反映し、‘ワイン’分類から文書3を排除（削除）することができる。

10

【0120】

一方、上述したように文書に流入する経路の比率がツリー構造の更新に反映される。一例として、文書3が分類（主題）‘ワイン’のみならず、分類（主題）‘TV’（図示せず）に連結されている場合、文書3を閲覧する使用者のうち、後者と関連した探索経路で文書3にアクセスする使用者の比率が相対的に大きくなり得る。この場合、その偏重度が所定の値を超えると、文書3と関連した‘ワイン’分類の経路を遮断することで、ツリー構造を更新することができる。

【0121】

これによって、更新後のツリー構造を表す図4では、クラスター‘ボルドーTV’と文書3が除去された。

20

【0122】

一方、‘カンパーニャ産モッツアレッタチーズ - ボルドーワインにうってつけ’という題目を有する文書4の場合、含まれたコンテンツは、モッツアレッタチーズに対するもので、ワイン生産地としてのボルドーに対する内容を取り扱った文書ではない。

【0123】

したがって、ワインの生産地としてのボルドーに対する関心を持ってクラスター‘ボルドー’を探索する使用者の照会頻度が、文書1に比べて低くなり得る。この場合、ツリー構造更新過程で、文書4とクラスター‘ボルドー’との連結が断絶（解除）され、図4のような結果を表す。

30

【0124】

一方、この過程で照会頻度だけではなく、流入経路の比率が考慮されることは、上述した通りである。

【0125】

しかしながら、‘カンパーニャ’は、イタリアワインの生産地でもあるので、文書3の場合と異なり、クラスター自体が除去されないこともある。このようにクラスター及び文書自体を主題から排除すること以外にも、ツリー構造での連結を変更することでツリー構造が更新される。

【0126】

図5は、本発明の一実施例に係る更新前の文書探索サービス提供画面を例示した図で、図6は、本発明の一実施例に係る更新後の文書探索サービス提供画面を例示した図である。すなわち、図5は、図3に関する文書探索サービス提供画面で、図6は、図4に関する文書探索サービス提供画面である。

40

【0127】

図5及び図6を参照すると、文書探索サービス提供画面は、分類表示領域500、分類構造表示領域510、探索位置表示領域520、クラスター表示領域530及び文書リンク532を含んでいる。

【0128】

上述したように、本実施例の文書探索サービスを提供するために、文書提供部215は

50

、使用者端末機 2 3 0 に提供されるウェブページを生成して伝送することができる。

【 0 1 2 9 】

文書探索サービスで提供されるウェブページは、図 5 及び図 6 に例示したような画面構成を有することができる。このような画面構成には、探索対象文書の構造化を視覚化するための表示領域が含まれる。

【 0 1 3 0 】

分類表示領域 5 0 0 は、文書が構造化される上位概念である分類（主題）に関する情報が表示される領域である。本実施例では、分類名である‘ワイン’が相対的に差別化された属性として表示されている。

【 0 1 3 1 】

分類構造表示領域 5 1 0 は、図 3 及び図 4 に例示したツリー構造を使用者に提供する領域である。図 4 の場合、使用者が探索しているディレクトリである‘ワインの産地’とクラスター‘ボルドー’は、他の項目と差別化された属性として表示されている。

【 0 1 3 2 】

また、使用者が探索しているディレクトリとクラスターを表示する探索位置表示領域 5 2 0 が追加的に提供される。

【 0 1 3 3 】

クラスター表示領域 5 3 0 は、使用者が探索しているクラスターに属する各文書に対するアクセス手段を提供する領域である。本実施例において、使用者が探索しているクラスターの中心概念は‘ボルドー’で、クラスター表示領域 5 3 0 には、‘ボルドー’という中心概念と関連した文書に対するアクセス手段としての文書リンク 5 3 2 が提供される。

【 0 1 3 4 】

クラスター表示領域 5 3 0 には、使用者が探索しているクラスターに属する各文書に対する文書リンク 5 3 2 が提供される。文書リンク 5 3 2 は、参照する文書の題目情報をアンカーテキストとして表示することができる。

【 0 1 3 5 】

文書リンク 5 3 2 は、各文書に対するリンクである。このリンクを選択することで、使用者は、自身が探索しようとする情報を含む文書の内容にアクセスすることができる。この場合、選択された文書のコンテンツは、使用者端末機 2 3 0 で新しいブラウザウィンドウを生成することで提供され、図 5 及び図 6 のような文書探索サービス提供画面が表示されたブラウザウィンドウの一部または全部を更新することによっても提供される。

【 0 1 3 6 】

使用者は、図 5 及び図 6 で提供される各領域の項目をクリック方法などで選択することで、自身の探索対象を変更することができる。これに対する応答として、文書提供部 2 1 5 は、分類構造表示領域 5 1 0、探索位置表示領域 5 2 0 及びクラスター表示領域 5 3 0 に視覚化される文書に関する情報を提供することができる。

【 0 1 3 7 】

一方、このような使用者たちの文書探索行為に関する情報は、使用者反応情報として反応情報データベース 2 3 2 に保存される。これら使用者反応情報に基づいてツリー構造更新部 2 1 7 がツリー構造を更新可能であることは、上述した通りである。

【 0 1 3 8 】

上述した画面構成を通して使用者にディレクトリ構造及びクラスタリング構造を視覚的に伝達することで、使用者が訪問するパーティカルサイト及びブログなどを個別的に訪問せずにも、関心分野に対する情報を効率的に探索ことができ、上述した画面に対する使用者の反応情報などをツリー構造に反映することで、ツリー構造をより正確でかつ効率的に更新することができる。

【 0 1 3 9 】

図 7 は、本発明の一実施例に係る広告表示領域を含む文書探索サービス提供画面を例示した図である。図 7 を参照すると、本発明の一実施例に係る広告表示領域を含む文書探索サービス提供画面は、分類表示領域 5 0 0、分類構造表示領域 5 1 0、探索位置表示領域

10

20

30

40

50

520、クラスター表示領域530及び広告表示領域710を含んでいる。

【0140】

広告表示領域710は、広告提供部218が使用者端末機230に提供する広告コンテンツが表示される領域である。広告表示領域710には、テキスト広告コンテンツ711及びアニメーション広告コンテンツ712が表示されている。

【0141】

テキスト広告コンテンツ711及びアニメーション広告コンテンツ712は、広告主と関連した追加的な情報を含んでいるサイトに接続可能なリンクなどを追加的に含むことができる。

【0142】

広告表示領域710に表示される広告は、クリック回数に相応して広告費用が執行されるPPC (pay per click) モデル、及び露出回数のうち少なくとも一つに相応して広告費用が執行されるPPV (pay per view) モデルなどに基づいて運営される。

【0143】

一方、広告表示領域710に表示される広告コンテンツを決定する要素として、使用者に関する情報及び使用者が探索する文書に関する情報などが考慮されることは、上述した通りである。

【0144】

一例として、文書の主題(分類)、ディレクトリ、クラスター及び文書の題目に関する情報が広告コンテンツ決定要素として考慮される。図5を参照すると、使用者が探索している主題は「ワイン」で、ディレクトリは「ワインの産地」で、クラスターは「ボルドー」である。このような情報に基づいて、「ボルドーワイン共同購買申し込み」というタイトルを有するテキスト広告コンテンツ711を提供することで、広告の効果を最大化することができる。

【0145】

一方、文書探索サービス提供方法は、コンピュータプログラムで作成可能である。前記プログラムを構成する各コード及び各コードセグメントは、当該分野のコンピュータプログラマーによって容易に推論される。また、前記プログラムは、コンピュータ可読情報保存媒体に保存され、コンピュータによって読まれて実行されることで、文書探索サービス提供方法を具現する。前記情報保存媒体は、磁気記録媒体、光記録媒体及びキャリアウェーブ媒体を含む。

【0146】

本出願で使用した用語は、特定の実施例を説明するために使用されたもので、本発明を限定する意図を持つものではない。単数の表現は、文脈上に明白に異なった意味を持たない限り、複数の表現を含む。

【0147】

本出願で、「含む」または「有する」などの用語は、明細書に記載された特徴、数字、段階、動作、構成要素、部品またはこれらの組み合わせが存在することを指定するもので、一つまたはそれ以上の他の特徴や数字、段階、動作、構成要素、部品またはこれらの組み合わせの存在または付加可能性を予め排除しないものとして理解されるべきである。

「第1、第2」などの用語は、多様な構成要素を説明するために使用されるが、前記各構成要素は、前記各用語によって限定されてはならない。前記各用語は、一つの構成要素を他の構成要素から区別する目的のみに使用される。

【0148】

以上、本発明の実施例に対して説明したが、上述した実施例以外にも多くの実施例が本発明の特許請求の範囲内に存在する。本発明の属する技術分野で通常の知識を有する者は、本発明がその本質的な特性から逸脱しない範囲で変形された形態で具現可能であることを理解するであろう。したがって、開示された実施例は、限定的な観点でなく、説明的な観点で考慮されるべきである。本発明の範囲は、上述した説明でなく、特許請求の範囲に

10

20

30

40

50

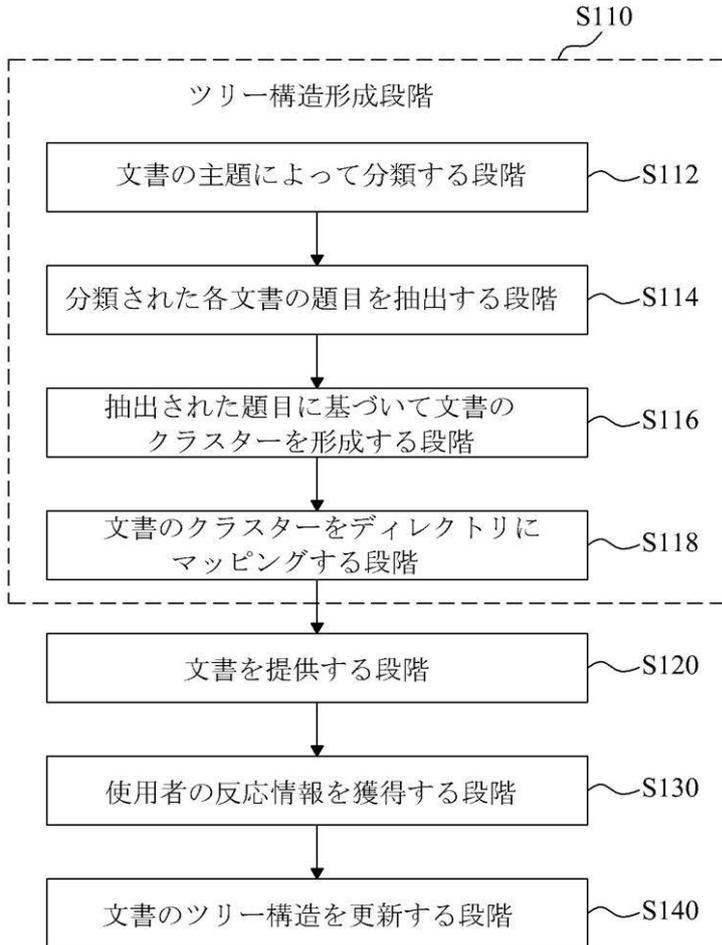
示されており、それと同等な範囲内にある全ての差異点は、本発明に含まれたものとして解析されるべきである。

【符号の説明】

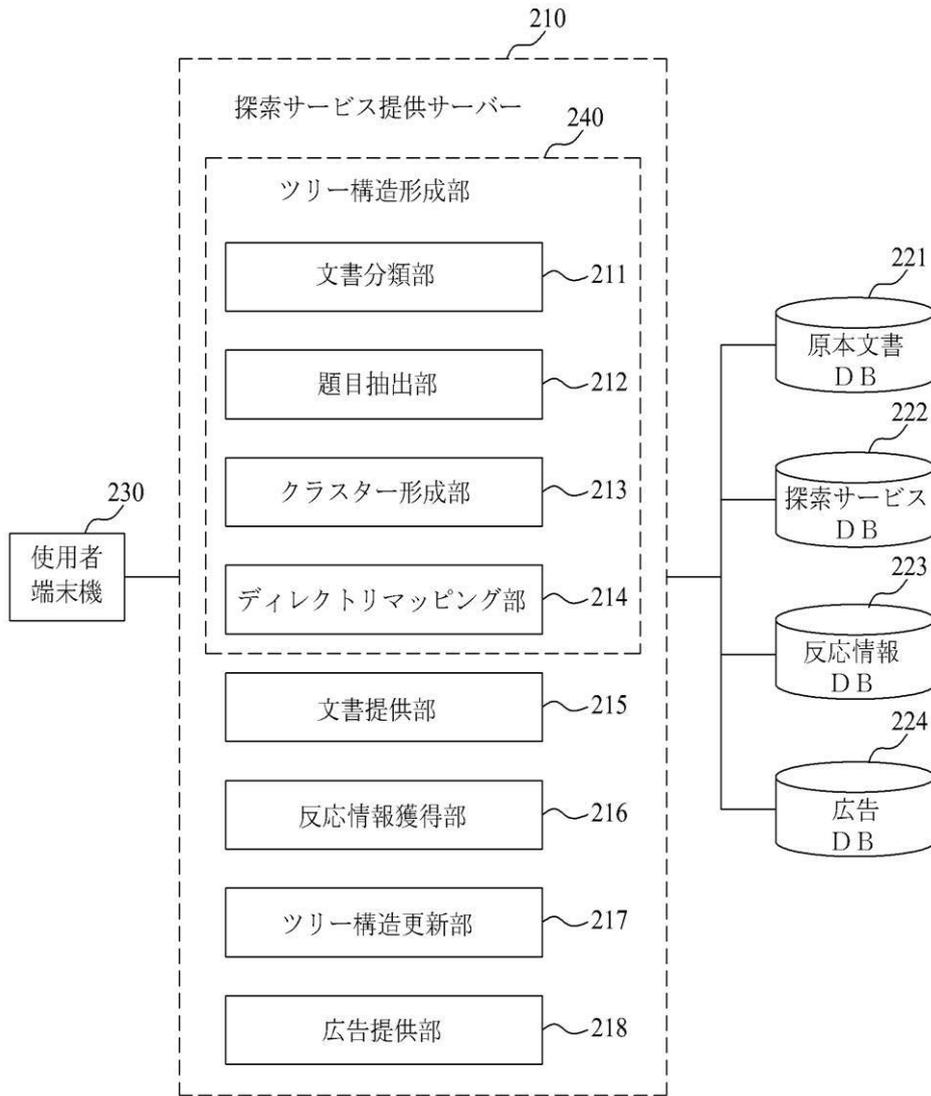
【 0 1 4 9 】

2 1 0	探索サービス提供サーバー	
2 1 1	文書分類部	
2 1 2	題目抽出部	
2 1 3	クラスター形成部	
2 1 4	ディレクトリマッピング部	
2 1 5	文書提供部	10
2 1 6	反応情報獲得部	
2 1 7	ツリー構造更新部	
2 1 8	広告提供部	
2 2 1	原本文書データベース	
2 2 2	探索サービスデータベース	
2 2 3	反応情報データベース	
2 2 4	広告データベース	
2 3 0	使用者端末機	
2 4 0	ツリー構造形成部	
5 0 0	分類表示領域	20
5 1 0	分類構造表示領域	
5 2 0	探索位置表示領域	
5 3 0	クラスター表示領域	
5 3 2	文書リンク	
7 1 0	広告表示領域	
7 1 1	テキスト広告コンテンツ	
7 1 2	アニメーション広告コンテンツ	

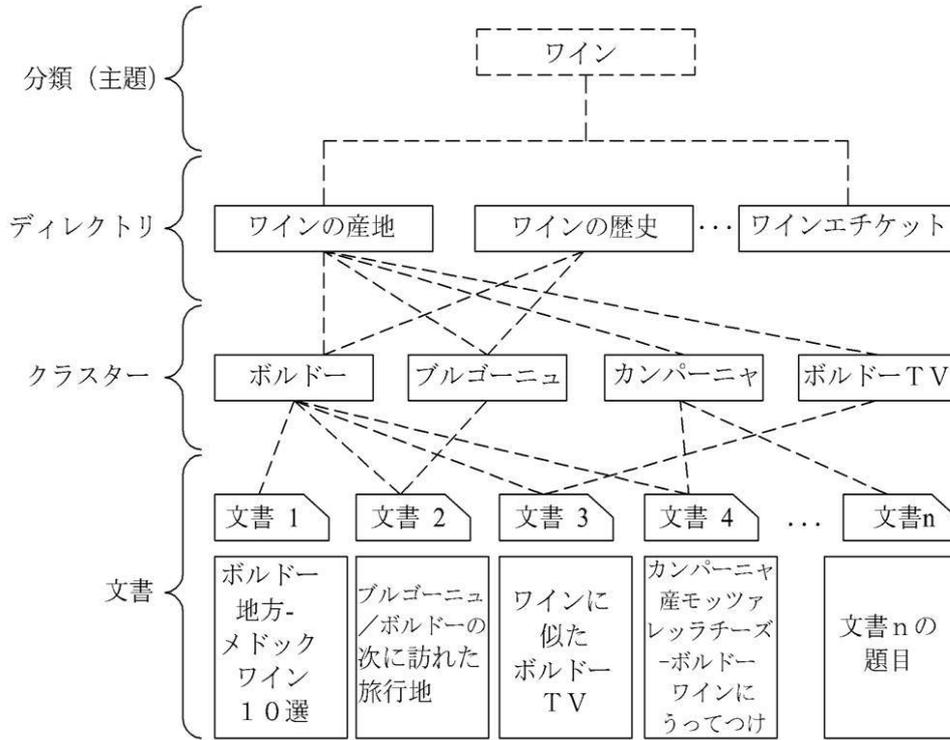
【図1】



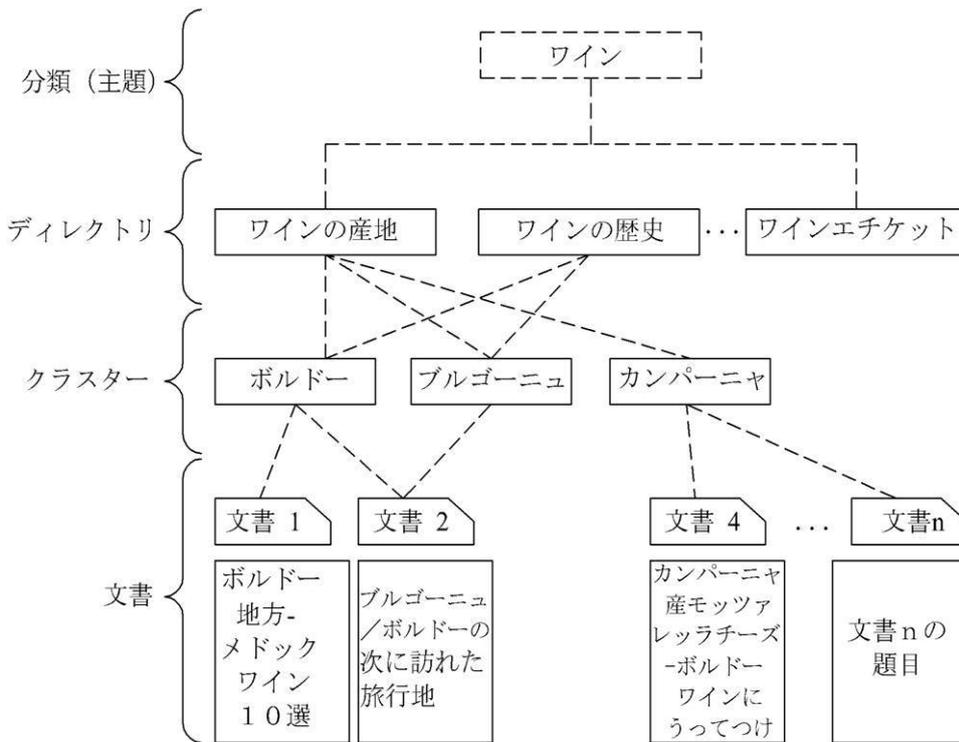
【図2】



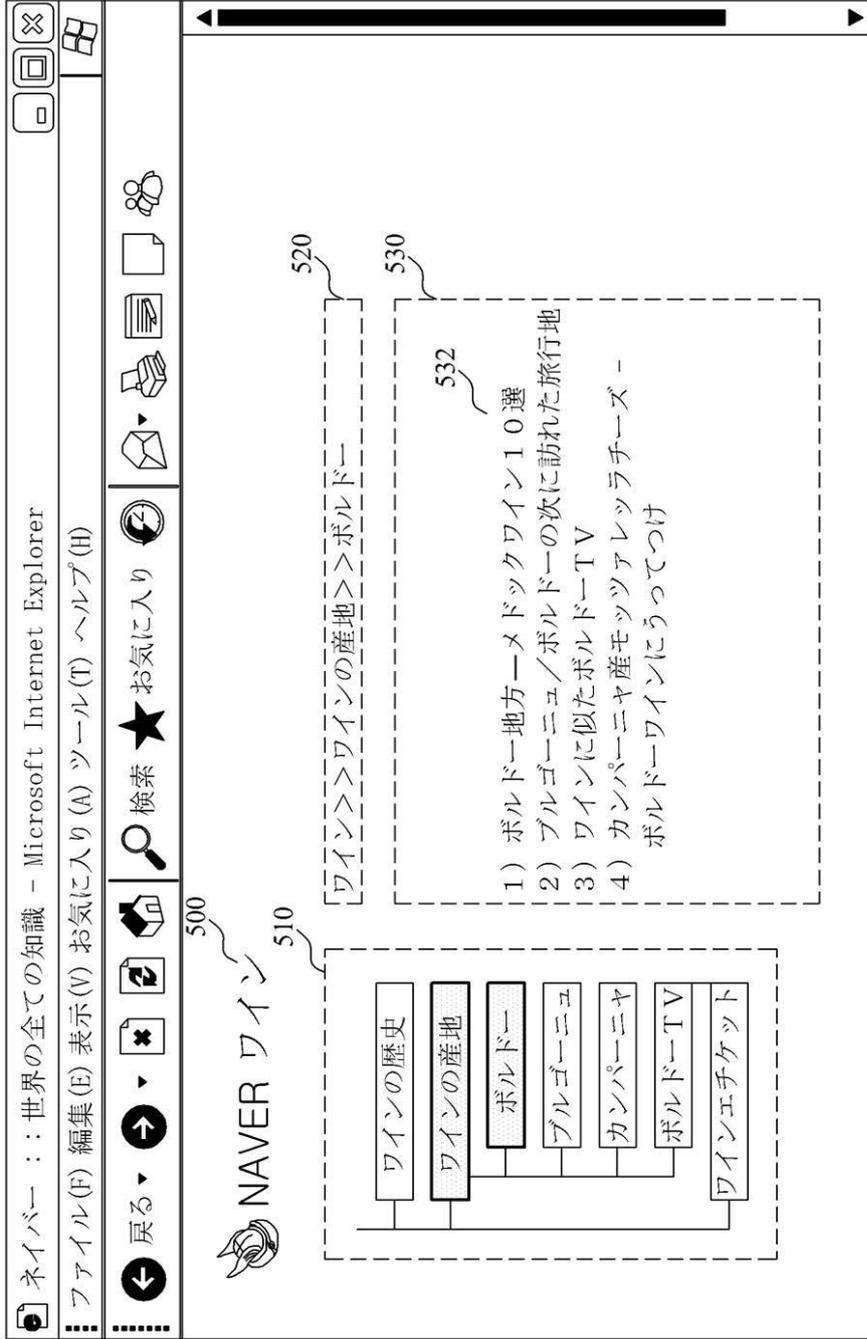
【図3】



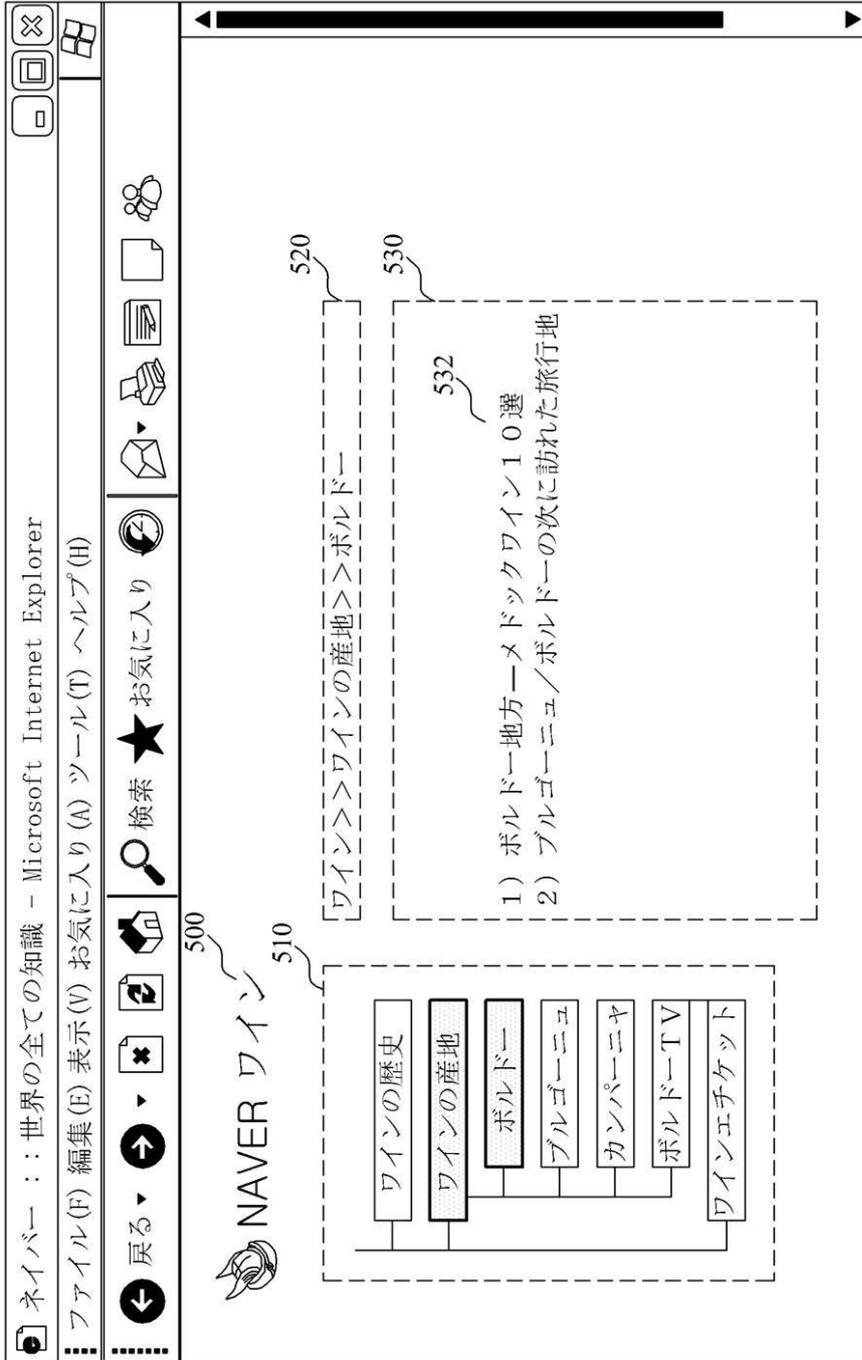
【図4】



【図 5】



【図 6】



フロントページの続き

(72)発明者 金 秉 學

大韓民國ソウル松坡區蠶室3洞 住公アパートメント レイク パレス102 - 1801

審査官 鈴木 和樹

(56)参考文献 特開2002 - 297642 (JP, A)

特開2002 - 288189 (JP, A)

特開平11 - 203319 (JP, A)

特開2007 - 213209 (JP, A)

特開2001 - 147940 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 12/00