

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7545061号
(P7545061)

(45)発行日 令和6年9月4日(2024.9.4)

(24)登録日 令和6年8月27日(2024.8.27)

(51)国際特許分類

F I

G O 6 F 16/34 (2019.01)

G O 6 F 16/332 (2019.01)

G O 6 F 16/338 (2019.01)

G O 6 F 16/34

G O 6 F 16/332

G O 6 F 16/338

請求項の数 9 (全19頁)

(21)出願番号	特願2022-29784(P2022-29784)	(73)特許権者	390002761
(22)出願日	令和4年2月28日(2022.2.28)		キャノンマーケティングジャパン株式会
(65)公開番号	特開2023-125592(P2023-125592		社
	A)		東京都港区港南2丁目16番6号
(43)公開日	令和5年9月7日(2023.9.7)	(73)特許権者	592135203
審査請求日	令和4年5月26日(2022.5.26)		キャノンITソリューションズ株式会
前置審査			社
			東京都港区港南2丁目16番6号
		(74)代理人	100189751
			弁理士 木村 友輔
		(74)代理人	100227857
			弁理士 中山 圭
		(72)発明者	進 義治
			東京都港区港南2丁目16番6号 キャ
			ノンITソリューションズ株式会社内
		審査官	齊藤 貴孝

最終頁に続く

(54)【発明の名称】 情報処理システム、情報処理方法、プログラム

(57)【特許請求の範囲】

【請求項1】

ユーザから指定された検索クエリを用いて文書の検索を行う検索手段と、
前記検索手段による検索の結果を表示するよう制御する表示制御手段と、
前記検索手段により検索されたそれぞれの文書毎に、当該文書の特徴づける単語である
特徴語を取得する特徴語取得手段と、
を備え、

前記表示制御手段は、前記ユーザから指定された検索クエリから取得された単語及び前
記特徴語取得手段により取得された特徴語を含む識別単語候補のうち、取得元を含む情報
に基づき算出される優先度に従い特定される単語を、当該文書の識別単語として当該文書
に対応付けて識別可能に表示するよう制御することを特徴とする情報処理システム。

10

【請求項2】

前記特徴語取得手段は、前記文書に含まれる単語のt f - i d f値に基づき、当該文書
における特徴語を取得することを特徴とする請求項1に記載の情報処理システム。

【請求項3】

前記表示制御手段は、前記優先度の高い順に所定数の識別単語を識別可能に表示するよ
う制御することを特徴とする請求項1または2に記載の情報処理システム。

【請求項4】

ユーザにより識別単語の指定を受け付ける受付手段を備え、
前記表示制御手段は、前記受付手段によりユーザから識別単語の指定を受け付けている

20

場合、当該識別単語を識別可能に表示するよう制御することを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載の情報処理システム。

【請求項 5】

前記表示制御手段は、前記識別単語の候補を表示するよう制御することを特徴とし、

前記受付手段は、前記表示された識別単語の候補の中から識別可能に表示する単語の選択を受け付けることで、識別単語の指定を受け付けることを特徴とする請求項 4 に記載の情報処理システム。

【請求項 6】

前記表示制御手段は、前記検索手段により検索された文書から、前記識別可能に表示する対象である識別単語の周辺の文字列をスニペットとして表示するよう制御することを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の情報処理システム。

【請求項 7】

前記表示制御手段は、前記スニペットに含まれる前記識別単語を識別可能に表示することを特徴とする請求項 6 に記載の情報処理システム。

【請求項 8】

情報処理システムにおける情報処理方法であって、

前記情報処理システムの検索手段が、ユーザから指定された検索クエリを用いて文書の検索を行う検索工程と、

前記情報処理システムの表示制御手段が、前記検索工程による検索の結果を表示するよう制御する表示制御工程と、

前記情報処理システムの特徴語取得手段が、前記検索工程において検索されたそれぞれの文書毎に、当該文書の特徴づける単語である特徴語を取得する特徴語取得工程と、
を備え、

前記表示制御工程は、前記ユーザから指定された検索クエリから取得された単語及び前記特徴語取得工程において取得された特徴語を含む識別単語候補のうち、取得元を含む情報に基づき算出される優先度に従い特定される単語を、当該文書の識別単語として当該文書に対応付けて識別可能に表示するよう制御することを特徴とする情報処理方法。

【請求項 9】

コンピュータを、請求項 1 乃至 7 のいずれか 1 項に記載の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理システム、情報処理方法、プログラムに関する。

【背景技術】

【0002】

企業内に電子化された文書が増えるにつれ、業務上必要な文書を効率的に検索するための文書検索システムの重要性が高まっている。ここで、文書検索システムは、ユーザーが入力した検索条件に関連する文書集合をユーザーに提示するシステムである。代表例としては、文字列で検索クエリを入力し、検索クエリ中に含まれる検索キーワードによって関連文書を取得する機能を持つものがある。

【0003】

検索システムにおいて、検索結果の文書において検索キーワードが現れる周辺の文字列を表示し、さらに検索キーワードをハイライトする（ハイライト機能）等により識別表示する技術が存在する（以下、識別表示の対象となる単語を識別単語という）。

【0004】

ハイライト機能により、ユーザーは検索結果として得られる複数の文書から自分にとって興味のある単語を効率的に見つけることができ、その結果、自分が求めている文書がどの文書であるか素早く把握することができる。

【0005】

非特許文献 1 には、文書検索システムにおけるハイライト機能について開示されている。

【先行技術文献】

【非特許文献】

【0006】

【文献】 <https://www.hitachi-systems.com/ind/srpartner/product/highlight/index.html>

【発明の開示】

【発明が解決しようとする課題】

【0007】

非特許文献 1 には、検索に用いたキーワードについてユーザが指定した色でハイライトして表示する機能について開示されている。

10

【0008】

しかし、検索の仕方によっては、検索に用いたキーワードがハイライトされるだけでは、検索された文書の特徴を効率的に把握できない等の課題がある。

【0009】

そこで、本発明は、検索結果を効率的に確認できる仕組みを提供することを目的とする。

【課題を解決するための手段】

【0010】

本発明の情報処理システムは、ユーザから指定された検索クエリを用いて文書検索を行う検索手段と、前記検索手段による検索結果を表示するよう制御する表示制御手段と、前記検索された文書から当該文書における特徴語を取得する特徴語取得手段と、を備え、前記表示制御手段は、前記検索手段により検索された文書における前記特徴語を識別単語として識別可能に表示するよう制御することを特徴とする。

20

【0011】

また、本発明の情報処理システムは、ユーザから指定された検索クエリを用いて文書検索を行う検索手段と、前記検索手段による検索結果を表示するよう制御する表示制御手段と、前記検索手段により検索された文書に関連する単語である関連語を取得する関連語取得手段と、を備え、前記表示制御手段は、前記検索手段により検索された文書に含まれる前記関連語を識別単語として識別可能に表示するよう制御することを特徴とする。

【発明の効果】

30

【0012】

本発明によれば、検索結果を効率的に確認することがかゝりとなる。

【図面の簡単な説明】

【0013】

【図 1】 本発明の実施形態における、文書検索システムのシステム構成の一例を示す図である。

【図 2】 本発明の実施形態における、文書検索システム、クライアント端末のハードウェア構成の一例を示すブロック図である。

【図 3】 本発明の実施形態における、文書 DB に保存された検索対象文書集合の一例を示す図である。

40

【図 4】 本発明の実施形態における、検索条件として用いられるデータの一例を示す図である。

【図 5】 本発明の実施形態における、検索結果として用いられるデータの一例を示す図

【図 6】 本発明の実施形態における、検索結果一覧に関する優先度ルール表の一例を示す図である。

【図 7】 本発明の実施形態における、検索結果詳細に関する優先度ルール表の一例を示す図である。

【図 8】 本発明の実施形態における、識別単語候補作成部が作成する識別単語候補ソース表の一例を示す図である。

【図 9】 本発明の実施形態における、識別単語候補作成部が作成する識別単語候補表の一

50

例を示す図である。

【図 1 0】本発明の実施形態における、検索処理部が実施する検索処理を示すフローチャートである。

【図 1 1】本発明の実施形態における、識別単語候補作成部が実施する識別単語候補作成処理を示すフローチャートである。

【図 1 2】本発明の実施形態における、検索結果画面の一例を示す図である。

【図 1 3】本発明の実施形態における、検索結果画面において識別単語候補を表示している状態の一例を示す図である。

【図 1 4】本発明の実施形態における、検索結果画面において識別単語を選択した状態の一例を示す図である。

【図 1 5】本発明の実施形態における、検索結果詳細画面において識別単語候補を表示している状態の一例を示す図である

【図 1 6】本発明の実施形態における、検索結果詳細画面において識別単語を選択した状態の一例を示す図である。

【発明を実施するための形態】

【0014】

以下、図面を参照して、本発明の実施形態を詳細に説明する。

【0015】

図 1 は、本発明の実施形態における文書検索システム 100 のシステム構成の一例を示す図である。

【0016】

文書検索システム 100 は、文書登録装置 110、文書 DB 120、文書検索装置 130、特徴語更新装置 140 から成る。

【0017】

文書登録装置 110 は、ユーザーによる検索の対象となる文書を登録するための装置であり、文書受信部 111、キーワード抽出部 112、文書登録処理部 113 から成る。

【0018】

文書受信部 111 は、登録対象の文書を受け付けるための機能部である。ユーザー（クライアント端末）は Web ブラウザなどを通じて任意の文書を文書受信部 111 に送信できる。あるいは、クローラーが機械的に文書を収集して送信するような構成をとってもよい。

【0019】

キーワード抽出部 112 は、文書受信部 111 が受け付けた文書から、当該文書における特徴語の候補となるキーワードとその出現頻度を抽出するための機能部である。特徴語の詳細については後述する。キーワード抽出部 112 におけるキーワード抽出処理は、公知の形態素解析技術を用いる。ここで、抽出する形態素は、文書検索システムの用途に応じて、固有名詞などの特定の品詞に限定してもよい。また、形態素解析を使用せずに、事前に定めたパターンに一致する文字列をキーワードとして抽出してもよい。

【0020】

文書登録処理部 113 は、文書受信部 111 で受け付けた文書と、キーワード抽出部 112 において抽出したキーワードとを紐づけて、文書 DB 120 へ格納する装置である。

文書 DB 120 は、文書を一意に識別するための文書 ID、文書名、本文、キーワード抽出部 112 が抽出した値を格納するキーワード：出現頻度、および、特徴語を格納する領域を備える。文書 DB 120 に格納されたデータの一例を図 3 に示す。特徴語の作成方法については後述する。なお、本アイデアを説明するための構成として、前述の 5 項目を例示しているが、文書の所在を示す URL、文書のサイズ、文書の作成者など、文書検索システムとして利用する項目を追加で備えてもよい。

【0021】

文書検索装置 130 は、検索処理部 131、検索条件保存部 132、検索操作保存部 133、検索結果保存部 134、識別単語候補作成部 135、優先度ルール表（検索結果一

10

20

30

40

50

覧) 136、優先度ルール表(検索結果詳細) 137からなる。

【0022】

検索処理部131は、ユーザーからの検索操作を受け付け、その検索操作を解釈して検索DBに問い合わせる検索条件を生成し、その検索条件に合致する文書を文書DB120から検索する機能部であり、検索条件に関連のある文書をスコア順に取得する機能を備える。検索処理部131が行う検索処理の詳細は後述する。さらに、ユーザーは検索処理部131へ検索操作を送ると同時に識別単語を送ることもできる。

【0023】

検索条件保存部132は、ユーザーが行った検索の検索条件を保存する機能部である。検索条件は図4のように「検索クエリ」と「類似文書検索クエリの特徴語」と「キーワードフィルター」からなる。

【0024】

検索操作保存部133は、ユーザーが行った検索の検索操作を保存する機能部である。検索操作保存部133に保存される値としては少なくとも「検索クエリによる検索」、「類似文書検索」、「キーワードフィルターの追加」の3種類の値がありうる。さらに、「キーワードフィルターの追加」の際には追加情報としてキーワードの文字列を保存できる。

【0025】

検索結果保存部134は、ユーザーが行った検索の検索結果を保存する機能部である。検索結果は図5のように、文書DB120から検索条件に合致する文書集合を抜き出した「文書一覧」と、検索結果に関連する単語の一覧である「関連語」と、文書一覧の文書IDごとに本文内から識別表示する部分を抜き出した「スニペット」からなる。

【0026】

検索処理部131は、検索処理を実行するたびに検索条件保存部132および検索操作保存部133および検索結果保存部134に保存している情報を更新する。

【0027】

識別単語候補作成部135は、図9のような識別単語候補表を作成する識別単語候補作成処理を行う。識別単語候補表は単語とその優先度を持つ表である。識別単語候補表作成処理については後述する。

【0028】

優先度ルール表(検索結果一覧) 136は、検索結果一覧に関する識別単語候補作成処理において用いる、図6のような優先度ルールを保持する。また、優先度ルール表(検索結果詳細) 137は、検索結果詳細に関する識別単語候補作成処理において用いる、図7のような優先度ルールを保持する。これらの表の利用方法については、識別単語候補表作成処理の説明時に合わせて説明する。なお、これらの優先度ルール表は文書検索システムを構築したときにシステム管理者が値を設定できる。

【0029】

特徴語更新装置140は、文書DBに格納された各文書について、特徴的なキーワードを特徴語として抽出して当該レコードを更新する装置である。特徴語の選出は、単語の特徴量を表す指標の1つであるtf-idfを用いることで実現できる。特徴語更新装置140は、文書DB120におけるキーワード：出現頻度の項目から、各単語の出現頻度を取得し、tf-idf値の高い順に、最大N件のキーワードを特徴語として抽出する。Nの値は文書検索システムを構築したときにシステム管理者が値を設定できる。例えば、図3の文書DB120において、文書1の特徴語は、「設計」、「画面」、「モバイル」の3件である。

【0030】

なお、図1において、文書検索システム100を構成する装置として、文書登録装置110、文書検索装置130、特徴語更新装置140の3つの装置と文書DB120とで構成される例を説明したが、本発明における文書検索システムはこの構成例に限定されるものではなく、各装置が備える機能を一つの装置が備えたシステムであっても良い。

【0031】

10

20

30

40

50

図2は、本発明の文書検索システム100や各装置として適用可能な情報処理装置のハードウェア構成の一例を示すブロック図である。

【0032】

図2に示すように、情報処理装置は、システムバス200を介してCPU（Central Processing Unit）201、ROM（Read Only Memory）202、RAM（Random Access Memory）203、記憶装置204、入力コントローラ205、音声コントローラ206、ビデオコントローラ207、メモリコントローラ208、および通信I/Fコントローラ209が接続される。

【0033】

CPU201は、システムバス200に接続される各デバイスやコントローラを統括的に制御する。

【0034】

ROM202あるいは外部メモリ213は、CPU201が実行する制御プログラムであるBIOS（Basic Input/Output System）やOS（Operating System）や、本情報処理方法を実現するためのコンピュータ読み取り実行可能なプログラムおよび必要な各種データ（データテーブルを含む）を保持している。

【0035】

RAM203は、CPU201の主メモリ、ワークエリア等として機能する。CPU201は、処理の実行に際して必要なプログラム等をROM202あるいは外部メモリ213からRAM203にロードし、ロードしたプログラムを実行することで各種動作を実現する。

【0036】

入力コントローラ205は、キーボード210や不図示のマウス等のポインティングデバイス等の入力装置からの入力を制御する。入力装置がタッチパネルの場合、ユーザがタッチパネルに表示されたアイコンやカーソルやボタンに合わせて押下（指等でタッチ）することにより、各種の指示を行うことができることとする。

【0037】

また、タッチパネルは、マルチタッチスクリーンなどの、複数の指でタッチされた位置を検出することが可能なタッチパネルであってもよい。

【0038】

ビデオコントローラ207は、ディスプレイ212などの外部出力装置への表示を制御する。ディスプレイは本体と一体になったノート型パソコンのディスプレイも含まれるものとする。なお、外部出力装置はディスプレイに限ったものはなく、例えばプロジェクタであってもよい。また、前述のタッチ操作を受け付け可能な装置については、入力装置も提供する。

【0039】

なおビデオコントローラ207は、表示制御を行うためのビデオメモリ（VRAM）を制御することが可能で、ビデオメモリ領域としてRAM203の一部を利用することもできるし、別途専用のビデオメモリを設けることも可能である。

【0040】

メモリコントローラ208は、外部メモリ213へのアクセスを制御する。外部メモリとしては、ブートプログラム、各種アプリケーション、フォントデータ、ユーザファイル、編集ファイル、および各種データ等を記憶する外部記憶装置（ハードディスク）、フレキシブルディスク（FD）、或いはPCMCIAカードスロットにアダプタを介して接続されるコンパクトフラッシュ（登録商標）メモリ等を利用可能である。

【0041】

通信I/Fコントローラ209は、ネットワークを介して外部機器と接続・通信するものであり、ネットワークでの通信制御処理を実行する。例えば、TCP/IPを用いた通信やISDNなどの電話回線、および携帯電話の4G回線、5G回線等を用いた通信が可能である。

10

20

30

40

50

【0042】

尚、CPU201は、例えばRAM203内の表示情報用領域へアウトラインフォントの展開（ラスターライズ）処理を実行することにより、ディスプレイ212上での表示を可能としている。また、CPU201は、ディスプレイ212上の不図示のマウ斯卡ーソル等でのユーザ指示を可能とする。

【0043】

次に、図10のフローチャートを用いて本発明の形態において検索処理部131がクライアント端末から検索リクエストを受けたときに実行する検索処理について説明する。

【0044】

まず、ステップS1001において、クライアントから受けた検索操作に含まれる検索条件と、検索条件保存部132に保存されている検索条件とを合わせることで、一つの検索条件を作る。

【0045】

検索条件には図4のように「検索クエリ」と「類似文書検索クエリの特徴語」と「キーワードフィルター」とを含む。図4に示す通り、検索クエリはキーワードである。検索クエリを含む検索条件によって、文書DB120に登録された文書の本文に対して全文検索が行われる。

【0046】

類似文書検索クエリの特徴語は、ユーザーが手持ちの文書を文書検索システム100へ送信し、文書検索システム100がその文書に類似する文書を検索DBから取得する類似文書検索において、検索条件として用いられるキーワードである。ユーザーから文書を受け付けると、文書検索システム100はその文書において特徴的な単語をtf-idfのような統計値から求めて特徴語として自動抽出する。類似文書検索クエリの特徴語を含む検索条件による検索処理により、文書DBに登録された文書のキーワードと、類似文書検索クエリにおける特徴語の一致率がしきい値以上の文書が取得される。

【0047】

キーワードフィルターはキーワードを含む絞り込み条件である。これを含む検索条件による検索処理により、文書DBから当該キーワードを含む文書のみの検索が行われる。

【0048】

「検索クエリ」と「類似文書検索クエリの特徴語」と「キーワードフィルター」のうち複数の種類を含む検索条件からは、それぞれの条件のAND条件をとることで最終的な検索結果の文書一覧が決まる。

【0049】

例えば、検索条件保存部132に何も保存されていない状態で、ユーザーから「操作種別：検索クエリによる検索、検索クエリ：『製品X 仕様』」という検索操作を受け付けたとする。このとき検索クエリが「製品X 仕様」である検索条件を生成する。そうすると、後の処理により、検索条件保存部132に「検索クエリ：『製品X 仕様』」という検索条件が保存される。

【0050】

その後、追加でユーザーから「操作種別：キーワードフィルターの追加、キーワード：『画面』」という検索操作を受け付けたとする。このとき、検索条件保存部132に保存された検索条件である「検索クエリ：『製品X 仕様』」と、ユーザーから受け付けた操作により生成されるキーワードフィルター「画面」を合わせて、「検索クエリ：『製品X 仕様』」かつ「キーワードフィルター：『画面』」であるような検索条件を生成する。

【0051】

ステップS1002において、検索条件に合致する文書を文書DB120から検索し、検索された文書を、検索条件に合致する程度を示す値であるスコアの高い順に並べる。効率的な検索処理を実現するためには、文書登録処理部113において、公知の技術である転置インデックスを作成して検索時に用いればよい。

【0052】

10

20

30

40

50

ステップS1003において、検索結果の関連語を求める。関連語としては、検索された文書に含まれるキーワードのうち一部を用いる。具体的な関連語の取得方法として、例えば検索された文書のうちおよそ半分の文書に共通に含まれるキーワードを選出する。この選出方法により、検索された文書において複数の話題を持つ文書が混在するとき、それらの話題を適切に分割する関連語が選ばれることを期待できる。

【0053】

ステップS1004において、生成された検索条件により検索条件保存部132を更新し、ユーザーから受け付けた検索操作により検索操作保存部133を更新する。

【0054】

ステップS1005において、ユーザーから識別単語を提示（指定）されたか否かにより処理を分岐する。識別単語を提示されている場合、それを識別単語として用いてステップS1010に進む。提示されていない場合、ステップS1006に進む。

10

【0055】

ステップS1006において、文書検索システムが識別単語簡易選択の動作モードであるか否かによって処理を分岐する。識別単語簡易選択の動作モードであるかどうかは、文書検索システムのシステム管理者により設定可能である。簡易選択の動作モードであればステップS1007に進み、簡易選択の動作モードでなければS1008に進む。

【0056】

識別単語簡易選択モードである場合（ステップS1006：YES）は、ステップS1007において、検索条件保存部に保存されている検索クエリから、空白区切りにより単語を抜き出し、それを識別単語として用いる。

20

【0057】

識別単語簡易選択モードではない場合（ステップS1006：NO）は、ステップS1008において、識別単語候補作成部に識別単語候補作成処理を行わせ、その結果を受け取る。識別単語候補作成処理については後述する。結果として図9のような、単語と優先度を持つ識別単語候補表を得られる。

【0058】

ステップS1009では、識別単語候補表から所定の条件を満たす単語（例えば優先度上位N件の単語や優先度が閾値以上の単語など）を識別単語として選出する。ここでNは文書検索システム100において定義された定数である。

30

【0059】

ユーザーは検索クエリに含まれる単語がハイライトされる動作に慣れている場合があるため、その場合にはシステム管理者が識別単語簡易選択の動作モードになるよう文書検索システムを設定することで、検索クエリの単語がハイライトされるようになり、ユーザーにとって違和感のない挙動を実現できる。

【0060】

一方で、よりユーザーにとって興味深い可能性が高い単語をハイライトさせたい場合、システム管理者は識別単語候補作成処理を行わせるよう文書検索システムを設定することもできる。

【0061】

40

ステップS1010において、検索結果として得られた各文書の本文から識別単語周辺の文字列をスニペットとして抽出する。周辺の文字列として、識別単語の前後N文字を用いることができる。ここでNは文書検索システム100において定義された定数である。

【0062】

さらに、クライアント端末において識別単語部分が識別表示されるように、抽出したスニペットの識別単語の部分の識別表示タグで囲む。識別表示の方法としては、識別単語を太字で表示する方法、他の文字列とは異なるフォントで表示する方法、マーカーで色付けして表示する方法、他の文字列とは異なる文字色で表示する方法など、識別単語が識別可能になる表示形態であればいずれでも良い。

【0063】

50

スニペットを効率よく抽出するために、公知の技術である転置インデックスを用いて本文内における識別単語の位置を取得することができる。

【0064】

識別単語が複数ある場合、各識別単語および周辺文脈の抽出結果を文字列結合したものをスニペットとして用いることができる。

【0065】

例えば図12に示す検索画面においては、ユーザーは明示的に識別単語を指定しておらず、検索クエリ1201に「製品X 仕様」が設定されており、類似文書検索クエリの特徴語1202に「製品Y」「仕様」「画面」という特徴語が設定されており、キーワードフィルター1203に「画面」「モバイル」というキーワードが設定されている。なお、画面上では「絞り込み条件」というラベルによりキーワードフィルターを表示している。ここで、仮にステップS1006において識別単語簡易選択の動作モードがYesであったとすると、ステップS1007により識別単語として「製品X」「仕様」という2件の単語が選出される。そのとき、検索処理部は検索結果の一番目の文書において、「製品X」および「仕様」の周辺文脈として本文内から「製品Xの管理画面の仕様は以下の通りとする」というスニペットを抽出する。さらに「製品X」「仕様」を識別表示タグで囲む。その結果、クライアント端末では検索結果の一番目の文書において、「製品Xの管理画面の仕様は以下の通りとする」「管理画面の仕様は以下の通りとする」という2つのテキストからなるスニペットが表示され、「製品X」「仕様」は識別表示される。

【0066】

また、図14の検索画面において、ユーザーは識別単語として「モバイル」を指定している。この場合、検索処理部は識別単語として「モバイル」を選出する。以下同様にして、クライアント端末では検索結果の一番目の文書において、「管理画面はモバイル向けには提供しない」「モバイル向けの検索画面の設計は以下の通り」という2つのテキストからなるスニペットが表示され、「モバイル」は識別表示される。

【0067】

ステップS1011において、文書DB120から得られた文書一覧と求めた関連語とスニペットにより検索結果保存部134を更新する。

【0068】

ステップS1012において、検索結果をクライアント端末へ返す。クライアント端末では図12のような検索結果画面が表示される。

【0069】

次に、図11のフローチャートを用いて、本発明の実施形態における識別単語候補作成部135が実行する識別単語候補作成処理について説明する。

【0070】

また、参考例として、検索条件保存部132に図4の検索条件が、検索結果保存部133に図5の検索結果が保存されているものとし、検索操作保存部133に保存されている検索操作が「キーワードフィルターの追加・キーワード『モバイル』」であるとする。また、優先度ルール表として優先度ルール表（検索結果一覧）136を用いて、その中身が図6であるとする。

【0071】

ステップS1101からS1105にかけて、識別単語候補ソース表の作成処理が行われる。識別単語候補ソース表は図8のように単語と、その単語の取得元および取得元詳細からなる表である。取得元詳細は空のことがありうる。同じ単語が2回以上出現する場合もある。

【0072】

ステップS1101において、検索条件保存部132に保存されている検索クエリから単語を識別単語候補ソース表に加える。

【0073】

例えば、図4の検索条件には検索クエリ「製品X 仕様」が含まれるが、この検索クエ

10

20

30

40

50

りの文字列を空白で区切り「製品 X」「仕様」という単語が得られる。「製品 X」は取得元を「検索クエリ」とし、検索クエリ内の左から 1 番目に得られた単語であるため取得元詳細を「前から 1 番目」として識別単語候補ソース表に加える。同様に「仕様」は取得元を「検索クエリ」とし、検索クエリの左から 2 番目に得られた単語であるため取得元詳細を「前から 2 番目」として識別単語候補ソース表に加える。

【0074】

ステップ S 1 1 0 2 において、検索条件保存部 1 3 2 に保存されている類似文書検索クエリの特徴語から単語を識別単語候補ソース表に加える。

【0075】

例えば、図 4 の検索条件には類似文書検索クエリの特徴語として「製品 Y」「仕様」「画面」が含まれるため、「製品 Y」「仕様」「画面」という単語が得られる。これらの単語は、取得元を「類似文書検索クエリの特徴語」として識別単語候補ソース表に加える。

【0076】

ステップ S 1 1 0 3 において、検索条件保存部 1 3 2 に保存されているキーワードフィルターから単語を識別単語候補ソース表に加える。

【0077】

例えば、図 4 の検索条件にはキーワードフィルターとして「画面」「モバイル」が含まれるため、「画面」「モバイル」という単語が得られる。これらの単語は、取得元を「キーワードフィルター」として識別単語候補ソース表に加える。

【0078】

ステップ S 1 1 0 4 において、検索結果保存部 1 3 4 に保存されている関連語から単語を識別単語候補ソース表に加える。

【0079】

例えば、図 5 の検索結果には関連語として「企画」「設計」「提案」が含まれるため、「企画」「設計」「提案」という単語が得られる。これらの単語は、取得元を「検索結果の関連語」として識別単語候補ソース表に加える。

【0080】

ステップ S 1 1 0 5 において、検索結果保存部 1 3 4 に保存されている文書一覧の特徴語（その文書において特徴的な単語であって、t f - i d f のような統計値から求められる単語）を識別単語候補ソース表に加える。

【0081】

例えば、図 5 の検索結果の一番目の文書には特徴語として「設計」「画面」「モバイル」が含まれるため、「設計」「画面」「モバイル」という単語が得られる。これらの単語は、取得元を「検索結果の特徴語」とし、検索スコアが最も高い文書から得た特徴語であるため取得元詳細を「文書の検索スコア 1 位」として識別単語候補ソース表に加える。検索結果の 2 番目、3 番目の文書についても同様に特徴語を加えることができる。

【0082】

ステップ S 1 1 0 6 において、識別単語候補ソース表のエントリ一覧について繰り返す処理を開始する。

【0083】

ステップ S 1 1 0 7 において、取得元および検索操作保存部 1 3 3 に保存されている検索操作に応じて、優先度ルールから加算する優先度を計算（算出）する。

【0084】

例えば、図 8 の識別単語候補ソース表におけるエントリ 8 0 1 の単語「モバイル」の優先度は、図 6 の優先度ルール表（検索結果一覧）により以下のように計算される。まず取得元がキーワードフィルターであるため、ルール 6 0 4 より優先度は + 3 0 0 される。

【0085】

また、検索操作保存部 1 3 3 に保存されている直近の検索操作の操作種別が「キーワードフィルターの追加」であり、かつ追加されたキーワードが「モバイル」であるため、ルール 6 0 5 より優先度は + 8 0 0 される。

10

20

30

40

50

【0086】

一般に、ユーザーが最後に行った検索操作はユーザーが直前に興味を持った内容を反映していると考えられる。そのため、ユーザーが最後に行った検索操作に関連する識別単語は、優先度を上げて優先的に表示することが有益と考えられる。ルール601、602、603も同様に、ユーザーが直前に興味を持った内容の優先度を高めるためのルールである。

【0087】

最終的にはエントリ801の単語「モバイル」の優先度は1100となる。

【0088】

もう一つの例として、図8の識別単語候補ソース表におけるエントリ802の単語「設計」の優先度は以下のように計算される。まず取得元が「検索結果の特徴語」由来であるため、ルール604より優先度は+50される。

さらに、取得元詳細において文書の検索スコア順位が1位であるため、ルール605より優先度は-1される。

【0089】

このルールは、検索順位が高いほど優先度を高くする（下げ幅を小さくする）というルールである。

【0090】

検索順位が高い文書はユーザーの指定した検索条件への一致度合いが高く、ユーザーが興味を持つ可能性が高い文書といえる。そのため、検索順位が高い文書に特徴的に現れる識別単語は、優先度を上げて優先的に表示することが有益と考えられる。

【0091】

ステップS1108では、識別単語候補表におけるその単語エントリの優先度を更新する。その際の優先度の値は、識別単語候補表に既にその単語が存在する場合、既存の優先度とステップS1107で求めた値を足したものである。一方、識別単語候補表にその単語が存在していない場合、ステップS1107で求めた値を優先度の初期値として単語を追加する。

【0092】

例えば、図8の識別単語候補ソース表において、単語「モバイル」は取得元が「キーワードフィルター」であるエントリ801と、取得元が「検索結果文書の特徴語」であるエントリ803の2つのエントリがある。

エントリ801において優先度が1100得られ、エントリ803において優先度が49得られたとすると、最終的に単語「モバイル」の優先度はそれらの和である1149となる。

【0093】

ステップS1109では、識別単語の一覧に未処理の単語が残っていれば処理をステップS1106に戻し、全て処理が終了していればステップS1110に進む。

【0094】

ステップS1110では識別単語候補表を優先度の降順に並べ替える。なお、同一の優先度である単語はどのような順番にしてもよい。順番を一意にしたい場合、単語の文字コード順にしてもよい。

【0095】

最終的な識別単語候補表は図9のようになる。仮に検索処理のステップS1006で識別単語簡易選択の動作モードがNoであり、文書検索システムが優先度上位1単語を識別単語として選択する設定になっている場合、識別単語は「モバイル」となる。これは単純に検索クエリから識別単語を選んだ場合とは別の結果になる。こうして、直近の検索操作に関係が強いキーワードフィルターの単語など、ユーザーが興味を持つ可能性が高い単語を自動で優先して識別表示することができる。

【0096】

識別単語候補作成処理は、検索処理の一部として呼ばれるだけでなく、クライアント端

10

20

30

40

50

末からのリクエストに応じて単体で実行されることもある。ユーザーが識別単語を手動入力する際、候補をユーザーに提示して識別単語を簡易に入力できるようにすることを目的とする。

【0097】

図13を例として説明する。クライアント端末に検索結果画面が表示されているとき、ユーザーはクライアント端末を操作し、識別単語を手動で入力するフォーム1301にフォーカスを当てる。このとき、クライアント端末は文書検索システムへ識別単語候補表をリクエストする。リクエストを受け取った文書検索システムは識別単語候補作成処理を実施し、識別単語候補表をクライアント端末へ返す。なお、この場合には作成された識別単語候補表を上位だけに絞り込むことをせず、全件クライアント端末へ返す。なお、現在の検索結果において識別単語になっている単語のエントリを返さないこともできる。

10

【0098】

識別単語候補表を受けったクライアント端末は、フォーム1301の下部に識別単語候補一覧1302を表示する。識別単語候補一覧1302は、識別単語候補表の単語を並べたものである。上に表示される単語ほど優先度が高く、上下が同じ位置では左に表示される単語ほど優先度が高い。

【0099】

ユーザーがクライアント端末を操作して識別単語候補一覧1302に含まれる単語をクリックすると、フォーム1301にその単語が入力される。それと同時に、クライアント端末は、クリックされた単語を識別単語として現在と同じ検索条件で検索するよう、文書検索システムへリクエストを送る。その結果、文書検索システムにおいて検索処理が行われ、クライアント端末に検索結果が返ってくる。クライアント端末は返ってきた検索結果をもとに、検索結果の表示を更新する。なお、検索条件が同じであるため、検索結果の文書および順位は同一になり、実際にはスニペットの表示だけ更新されることになる。

20

【0100】

例えば図13において単語1303である「モバイル」がクリックされたとき、検索画面は図14のように変化する。すなわち、検索結果の文書は同一であり、スニペットの表示が「モバイル」周辺の文字列に変化する。

【0101】

さらに、クライアント端末に表示される画面として、特定の文書の詳細を表す図15のような検索結果詳細画面も存在する。

30

【0102】

これは、検索リクエストの返り値である検索結果として得られた文書から、1件の文書のみを詳細に表示する画面である。この画面によりユーザーは興味を持った文書について、本文をより詳細に確認することができる。

検索結果詳細画面においても、ユーザーが識別単語を手動で設定し、識別単語の周辺文脈を確認する機能がある。このときも、文書検索システムは識別単語の候補をユーザーに提示してユーザーが識別単語を入力できるよう支援を行う。

【0103】

識別単語候補提示の流れは、検索結果一覧画面のものと類似している。

40

ユーザーがクライアント端末を操作し、識別単語を手動で入力するフォーム1501にフォーカスを当てたとき、クライアント端末は文書検索システムへ識別単語候補表をリクエストする。ただしこのとき、クライアント端末は優先度ルール表（検索結果詳細）を用いて優先度を決定するようにリクエストを行う。文書検索システムは優先度ルール表（検索結果詳細）を用いて求めた識別単語候補表をクライアント端末へ返す。識別単語候補表を受けったクライアント端末は、フォーム1501の下部に識別単語候補一覧1502を表示する。

【0104】

ユーザーがクライアント端末を操作して識別単語候補一覧1502に含まれる単語をクリックすると、フォーム1501にその単語が入力される。それと同時に、クライアント

50

端末は、クリックされた単語を識別単語として、現在詳細表示している文書の文書IDを検索条件として検索するよう、文書検索システムへリクエストを送る。その結果文書検索システムにおいて検索処理が行われ、当該文書1件のスニペットを含む検索結果がクライアント端末に返される。端末は返された検索結果をもとに、本文の表示をスニペットに置き換える。なお、検索システムは本文以外でも、タイトルのようなテキスト項目からスニペットを抽出できる構成にすることが可能であり、本文以外のフィールドの表示を抽出されたスニペットで置き換えることができる。

【0105】

例えば図15において識別単語候補一覧にある単語1503「モバイル」がクリックされると、図16のように本文が「モバイル」周辺の文字列を表示するよう変化する。タイトルは「モバイル」を含まないため、空欄を表示する。

10

【0106】

検索結果詳細に関する識別単語候補表の作成処理について説明する。基本的には検索結果一覧によるものと同様、図11のフローチャートに従って処理を行う。ただし、ステップS1105において、詳細表示対象の文書の特徴語のみを識別単語候補ソース表に加える対象として、それ以外の文書の特徴語は加えない。他の文書の関連語が詳細表示中の文書に関係ある可能性は低いためである。また、ステップS1107において優先度ルールを計算する際、優先度ルール表（検索結果詳細）137を用いて計算を行う。図7に優先度ルール表（検索結果詳細）137の例を記載している。基本的には優先度ルール表（検索結果一覧）136と同じであるが、検索結果の特徴語に関してはルール701のように、取得元が当該文書の特徴語である場合に適用されるルールとなる。また、検索スコアの順位は利用しない。

20

【0107】

以上説明した通り、本発明では、検索クエリとして指定された単語だけでなく、検索された文書から取得される、当該文書の特徴付ける単語や、検索された文書に関連する単語についても識別表示の対象とすることが可能となるため、文書検索を行ったユーザは、検索結果を効率的に確認することが可能となる。

【0108】

本発明は、例えば、システム、装置、方法、プログラムもしくは記録媒体等としての実施態様をとることが可能である。具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

30

【0109】

また、本発明におけるプログラムは、図10、図11に示すフローチャートの処理方法をコンピュータが実行可能なプログラムであり、本発明の記憶媒体は図10、図11の処理方法をコンピュータが実行可能なプログラムが記憶されている。なお、本発明におけるプログラムは図10、図11の各装置の処理方法ごとのプログラムであってもよい。

【0110】

以上のように、前述した実施形態の機能を実現するプログラムを記録した記録媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記録媒体に格納されたプログラムを読み出し、実行することによっても本発明の目的が達成されることは言うまでもない。

40

【0111】

この場合、記録媒体から読み出されたプログラム自体が本発明の新規な機能を実現することになり、そのプログラムを記録した記録媒体は本発明を構成することになる。

【0112】

プログラムを供給するための記録媒体としては、例えば、フレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、DVD-ROM、磁気テープ、不揮発性のメモリカード、ROM、EEPROM、シリコンディスク等を用いることが出来る。

【0113】

50

また、コンピュータが読み出したプログラムを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0114】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPU等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

10

【0115】

また、本発明は、複数の機器から構成されるシステムに適用しても、ひとつの機器から成る装置に適用しても良い。また、本発明は、システムあるいは装置にプログラムを供給することによって達成される場合にも適応できることは言うまでもない。この場合、本発明を達成するためのプログラムを格納した記録媒体を該システムあるいは装置に読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。

【0116】

さらに、本発明を達成するためのプログラムをネットワーク上のサーバ、データベース等から通信プログラムによりダウンロードして読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。なお、上述した各実施形態およびその変形例を組み合わせた構成も全て本発明に含まれるものである。

20

【符号の説明】

【0117】

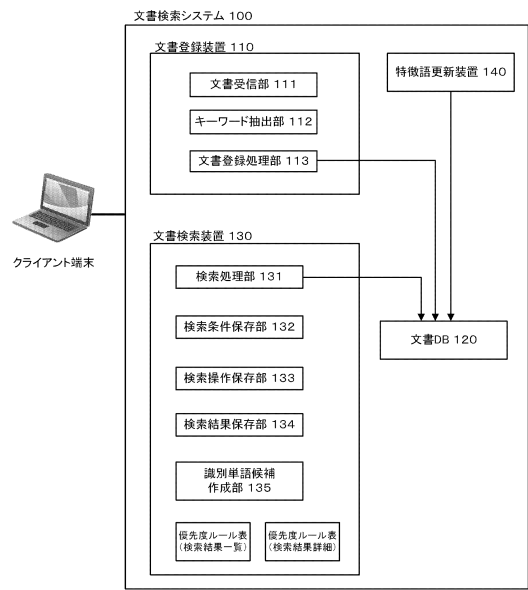
100 文書検索システム

30

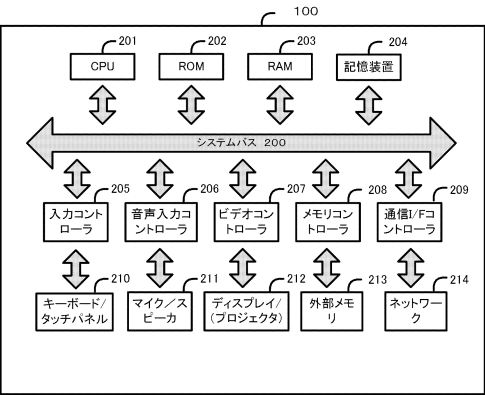
40

50

【図面】
【図 1】



【図 2】



【図 3】

キーワード出現頻度		特徴語
製品X、外部設計書		図面、画面、モバイル
文書1	製品X、仕様1、画面1、製品X、仕様2、モバイル3	図面、画面、モバイル
文書2	製品X、仕様1、画面1、製品X、仕様2、モバイル3	図面、画面、モバイル
文書3	製品X、仕様1、画面1、製品X、仕様2、モバイル3	図面、画面、モバイル

【図 4】

検索クエリ

製品X 仕様

類似文書検索クエリの特徴語

製品Y
仕様
画面

キーワードフィルター

画面
モバイル

【図 5】

文書一覧			
文書ID	文書名	本文	キーワード出現履歴
文書1	製品X/外部設計書	...	製品X-4, 管理-1, 画面-6, 仕様-2, モバイル-3
文書2	商品企画書 (製品X)	...	商品-1, 企画-5, 製品X-3, 企業-4, 文書-1, 検索-2, エンジン-1, 類似-1, 仕様-2
文書3	製品X-A仕様向け提案書	...	製品X-3, A社-3, 提案-5, 要件-8, システム-2, 検索-2

関連語	
企画	
設計	
提案	

スニペット

文書ID	スニペット部
文書1	本文書は製品Xの外部設計資料である。 管理画面の仕様は以下の通りとする
文書2	製品Xは企業内文書検索を行うための全文検索エンジンです... 類似文書検索の仕様について ...
文書3	貴社の競争システム構築において製品Xを利用し...

【図 7】

取得元が検索クエリである場合	膨大度 +600
	直近の検索操作の操作種別が「検索クエリ」の設定 ならば膨大度+800
取得元が類似文書検索クエリの特徴語である場合	検索クエリの前からN番目ならば膨大度 + N×100
	膨大度 + 200
取得元がキーワードフィルタである場合	直近の検索操作の操作種別が「類似文書検索」 ならば膨大度+800
	膨大度 + 300
取得元が検索結果の関連語である場合	直近の検索操作の操作種別が「キーワードフィルタ追加」 であり、追加したキーワードが優先度計算対象の単語と一致するならば膨大度+800
	膨大度 +200
取得元が当該文書の特徴語である場合	膨大度 +800

701

【図 6】

取得元が検索クエリである場合	膨大度 +600	601
	直近の検索操作の操作種別が「検索クエリ」による検索」ならば膨大度+800	602
取得元が類似文書検索クエリの特徴語である場合	検索クエリの前からN番目ならば膨大度 + N×100	603
	膨大度 +200	604
取得元がキーワードフィルタである場合	直近の検索操作の操作種別が「類似文書検索」ならば膨大度+800	605
	膨大度 + 300	606
取得元が検索結果の関連語である場合	直近の検索操作の操作種別が「キーワードフィルタ」であり、追加されたキーワードが膨大度計算対象の単語と一致するならば膨大度+800	607
	膨大度 +200	
取得元が検索結果の文書の特徴語である場合	膨大度 +50	
	検索スコアの順位がN位の文書である場合、膨大度 -N	

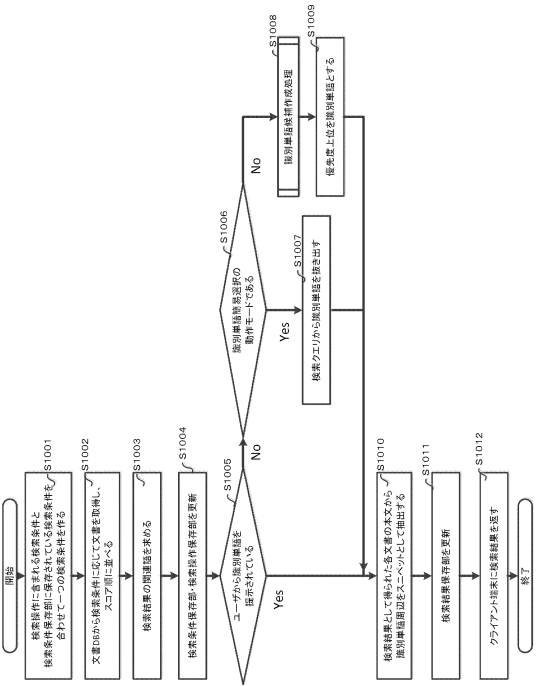
【図 8】

単語	取得元	取得元詳細	801
製品X	検索クエリ	前から1番目	
仕様	検索クエリ	前から2番目	
製品Y	類似文書検索クエリの特徴語		
仕様	類似文書検索クエリの特徴語		
画面	類似文書検索クエリの特徴語		
画面	キーワードフィルタ		
モバイル	キーワードフィルタ		
企画	検索結果の関連語		
設計	検索結果の関連語		
提案	検索結果の関連語		
設計	検索結果文書の特徴語	文書の検索スコア1位	802
画面	検索結果文書の特徴語	文書の検索スコア1位	
モバイル	検索結果文書の特徴語	文書の検索スコア1位	803
企画	検索結果文書の特徴語	文書の検索スコア2位	
企業	検索結果文書の特徴語	文書の検索スコア2位	
提案	検索結果文書の特徴語	文書の検索スコア3位	
基幹	検索結果文書の特徴語	文書の検索スコア3位	

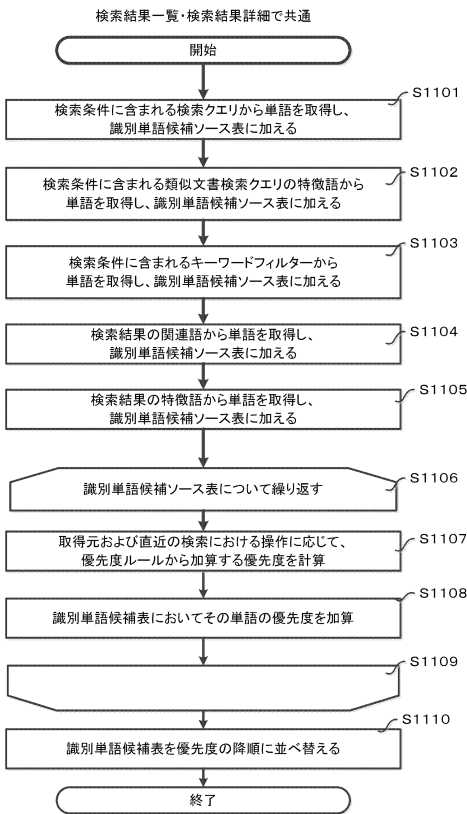
【図 9】

単語	優先度
モバイル	1149
仕様	1000
製品X	700
画面	549
設計	249
企画	248
提案	247
製品Y	200
企業	48
基幹	47

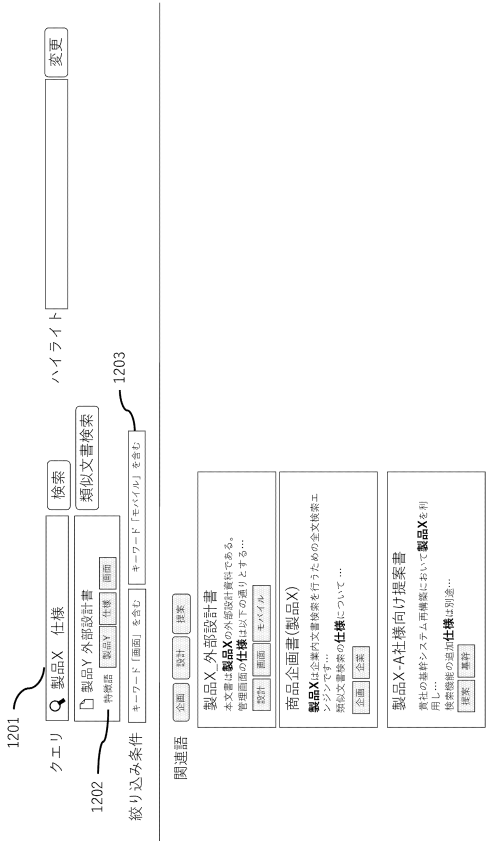
【図 10】



【図 11】



【図 12】



10

20

30

40

50

【図 13】

クエリ

製品X 仕様

仕様

製品X/外部設計書

仕様

キーワード「企画」を含む

キーワード「企画」を含む

キーワード「企画」を含む

変更に

モバイル

仕様

企画

基幹

企業

基幹

企画

設計

基幹

企業

製品X-外部設計書

本文書は製品Xの外部設計資料である。
管理組織の仕様は以下の通りとする...

設計

基幹

企業

商品企画書(製品X)

製品Xは企業内文書体例を行うための全文検索エンジンです...
類似文書体例の仕様について...

企画

企業

製品X-A仕様向け提案書

貴社の最新システム構築において製品Xを利用し...
検索機能の仕様は別添...

仕様

基幹

企業

基幹

企画

設計

基幹

企業

【図 14】

クエリ

Q製品X仕様

検索

変更

モバイル

ハイライト

類似文書探索

製品Y外部設計書

特許語

製品Y仕様

関連

キーワード「画面」を含む

キーワード「モバイル」を含む

関連語

企画

設計

検索

製品X外部設計書

管理画面は**モバイル**向けには提供しない…
モバイル向けの外部画面の設計は以下の通り…

設計

画面

モバイル

商品企画(製品X)

モバイル端末での操作には音入力を採用し、
入力の誤差を低減…

企画

企画

承認

製品X-A社様向け提案書

御社は**モバイル**環境に適した専用のカスタマイズ
を行いたい…

提案

承認

検討

【図 15】

タイトル	製品X_外部設計書
本文	本文書は製品Xの外部設計資料である。 管理画面の仕様は以下の通りとする。 1. ログインについて…

【図 16】

タイトル		
本文	管理画面は モバイル 向けには提供しない… モバイル 向けの検索画面の設計は以下の通り…	
		ハイライト モバイル 変更

フロントページの続き

- (56)参考文献 特開 2007-323238 (JP, A)
特開 2014-106665 (JP, A)
特開 2007-141059 (JP, A)
特開 2008-287517 (JP, A)
特開 2020-091607 (JP, A)
殷 成久、外 3 名、観光イベントについての「といえば検索」の提案、電子情報通信学会技術研究報告、日本、社団法人電子情報通信学会、2010年11月12日、第 110 巻、第 301 号、p. 43-47
- (58)調査した分野 (Int.Cl., DB 名)
G06F 16/00-16/958