

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2014-102625
(P2014-102625A)

(43) 公開日 平成26年6月5日(2014.6.5)

(51) Int.Cl. F I テーマコード (参考)
G06F 17/30 (2006.01) G06F 17/30 310A
 G06F 17/30 330C

審査請求 未請求 請求項の数 8 O L (全 18 頁)

(21) 出願番号 特願2012-253300 (P2012-253300)
 (22) 出願日 平成24年11月19日 (2012.11.19)

特許法第30条第2項適用申請有り 平成24年7月5日に株式会社イプロスが、http://issearch.ipros.jp/アドレスのウェブサイトに公開した『課題検索α版』により掲載して発表した。

(71) 出願人 512173704
 株式会社イプロス
 東京都千代田区丸の内3-8-1 住友不動産丸の内ビル5F
 (74) 代理人 100087848
 弁理士 小笠原 吉義
 (74) 代理人 100119161
 弁理士 重久 啓子
 (74) 代理人 100111822
 弁理士 渡部 章彦
 (72) 発明者 平田 英二
 東京都千代田区丸の内三丁目8番1号 株式会社イプロス内

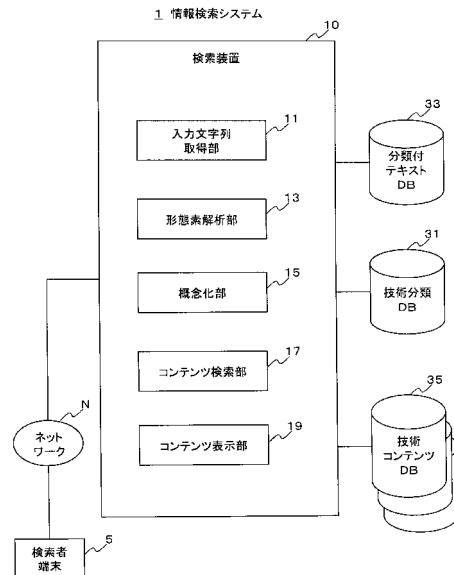
(54) 【発明の名称】 情報検索システム、プログラム、および方法

(57) 【要約】

【課題】 情報検索システムで、検索漏れや検索ノイズが少ない検索結果を出力することを目的とする。

【解決手段】 情報検索システム1は、技術のフィールド、該フィールドに属する技術範囲(テーマ)、各テーマの技術特性(ターム)を記憶する技術分類データベース31、技術コンテンツを蓄積する技術コンテンツデータベース35、検索者の入力文字列を取得する入力文字列取得部11、入力文字列で技術分類データベース31に記憶されたタームを照合し、一致するタームを含むテーマを抽出し、抽出したテーマから1つのテーマを特定し、特定したテーマが含むターム群を抽出する概念化部15、ターム群をキーワードとして技術コンテンツデータベース35をテキスト検索し、評価の良い技術コンテンツを取得するコンテンツ検索部17、取得した技術コンテンツを表示する表示用データを出力するコンテンツ表示部19を備える。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

情報を検索するシステムであって、

技術が適用される範囲を示すフィールドと、該フィールドに属する技術の範囲を示すテーマと、該テーマ毎に該技術の範囲に含まれる技術の特性を複数の観点に対応付けて示すタームとを定める技術分類を記憶する技術分類データベースと、

技術情報を示す技術コンテンツを蓄積する技術コンテンツデータベースと、

検索者によって入力された単語または単語の組合せである入力文字列を取得する入力文字列取得部と、

前記技術分類データベースに記憶された技術分類のタームを前記入力文字列で照合し、一致するタームを含む全てのテーマを抽出し、抽出した該テーマから少なくとも1つのテーマを特定し、該特定したテーマに含まれる複数のターム群を抽出する概念化部と、

前記抽出したターム群をキーワードとして前記技術コンテンツデータベースをテキスト検索し、予め定めた評価方法で該検索した技術コンテンツを評価して、評価の良い技術コンテンツを取得するコンテンツ検索部と、

前記取得した技術コンテンツを表示する表示用データを出力するコンテンツ表示部とを、備える

ことを特徴とする情報検索システム。

【請求項 2】

前記技術分類データベースに記憶された技術分類に基づくテーマおよびタームが付与された技術情報に関する分類付テキストを蓄積する分類付テキストデータベースを備えるとともに、

前記概念化部は、前記技術分類データベースの検索処理の代わりに、前記分類付テキストデータベースに記憶された分類付テキストに付与されたタームを前記形態素で照合し、一致するタームが付与されたテキストを抽出し、抽出したテキストに含まれる全てのテーマを抽出し、該抽出したテキスト数に基づいて該抽出したテーマから少なくとも1つのテーマを特定する

ことを特徴とする請求項 1 に記載の情報検索システム。

【請求項 3】

前記概念化部は、前記技術分類データベースから抽出した前記テーマが属するフィールドを抽出し、抽出した該フィールド各々の評価値を予め定めた評価方法により算出し、該評価値が近いフィールドが複数存在する場合に、該複数のフィールドから選択されたフィールドが属するテーマを特定する

ことを特徴とする請求項 1 または請求項 2 に記載の情報検索システム。

【請求項 4】

前記入力文字列取得部が、前記入力文字列として文を取得し、

前記概念化部が、前記入力文字列に対して形態素解析を行って形態素を抽出し、前記技術分類データベースに記憶された技術分類のタームを前記抽出した形態素で照合し、一致するタームを含む全てのテーマを抽出する

ことを特徴とする請求項 1 ないし請求項 3 のいずれか一項に記載の情報検索システム。

【請求項 5】

前記概念化部は、前記特定したテーマに含まれる全てのターム群を抽出する

ことを特徴とする請求項 1 ないし請求項 4 のいずれか一項に記載の情報検索システム。

【請求項 6】

前記概念化部は、前記技術分類のタームが階層構造を有する場合に、前記特定したテーマに含まれるタームから、前記入力文字列と一致したタームの上位に直接連なるタームまたは下位に属する全てのタームを抽出する

ことを特徴とする請求項 1 ないし請求項 4 のいずれか一項に記載の情報検索システム。

【請求項 7】

情報を検索するために、コンピュータに、

10

20

30

40

50

検索者によって入力された単語または単語の組合せである入力文字列を取得し、

技術分類データベースに記憶された、技術が適用される範囲を示すフィールドと、該フィールドに属する技術の範囲を示すテーマと、該テーマ毎に該技術の範囲に含まれる技術の特性を複数の観点に対応付けて示すタームとを定める技術分類に対して、該技術分類のタームを前記入力文字列で照合し、一致するタームを含む全てのテーマを抽出し、抽出した該テーマから少なくとも1つのテーマを特定し、

前記特定したテーマに含まれる複数のターム群を抽出し、

技術情報を示すコンテンツを蓄積する技術コンテンツデータベースに対して、前記抽出したターム群をキーワードとしてテキスト検索を行い、予め定めた評価方法で該検索した技術コンテンツを評価して、評価の良い技術コンテンツを取得し、

前記取得した技術コンテンツを表示する表示用データを出力する、処理を実行させることを特徴とする情報検索プログラム。

【請求項8】

情報を検索するために、コンピュータが、

検索者によって入力された単語または単語の組合せである入力文字列を取得する処理過程と、

技術分類データベースに記憶された、技術が適用される範囲を示すフィールドと、該フィールドに属する技術の範囲を示すテーマと、該テーマ毎に該技術の範囲に含まれる技術の特性を複数の観点に対応付けて示すタームとを定める技術分類に対して、該技術分類のタームを前記入力文字列で照合し、一致するタームを含む全てのテーマを抽出し、抽出した該テーマから少なくとも1つのテーマを特定する処理過程と、

前記特定したテーマに含まれる複数のターム群を抽出する処理過程と、

技術情報を示す技術コンテンツを蓄積する技術コンテンツデータベースに対して、前記抽出したターム群をキーワードとしてテキスト検索を行い、予め定めた評価方法で該検索した技術コンテンツを評価して、評価の良い技術コンテンツを取得する処理過程と、

前記取得した技術コンテンツを表示する表示用データを出力する処理過程とを、備えることを特徴とする情報検索方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報の検索処理に関する。より詳しくは、本発明は、与えられた文字列から検索したい技術の範囲を概念化して情報を検索する検索処理に関する。

【背景技術】

【0002】

エンジニアリング、製品企画等の業務では、技術情報の収集は欠かすことができない。技術情報として、インターネット、イントラネット等で標準的に用いられているドキュメントシステムであるwebに蓄積されている情報が広く利用されている。これらの情報を検索する手段として、テキスト検索処理が使用されている。

【0003】

従来、検索手法として、与えられた文字列をキーワードとする検索手法があり、例えば、製品名や仕様を示す文字列を入力して製品を検索する製品検索システムで実施されている(従来手法(1))。

【0004】

別の従来、検索手法として、大量の自由文等のテキストデータに対してデータマイニングを行い、相関関係や出現パターンなどを解析して情報を検索する手法があり、例えば、書籍検索システムで実施されている(従来手法(2))。

【0005】

さらに別の従来、検索手法として、検索対象となる技術文献に分類コードを付与し、分類コードを用いて検索する手法があり、例えば、分類が付与された特許公報を検索する特許検索装置が知られている(従来手法(3))。

10

20

30

40

50

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2002-351896公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

上記の従来手法(1)は、与えられたキーワードとの一致検索であるため、製品名や仕様、術語が入力されないと、所望する検索結果が得られないという問題がある。

【0008】

しかし、所望する技術情報を得るための適切なキーワードを入力できる検索者は多くはない。例えば、「ガラス板に穴を開ける手段」に関する情報を得たい場合に、既に知っている手段の「ドリル」に関連する検索結果を得ることが可能であっても、本当に知りたい他の手段、例えばレーザ、ウォータージェット、パンチング加工等に関連する技術情報を検索結果として得ることができない。

【0009】

一般的に、検索精度は、検索者が適正なキーワードを入力できるかに依存している。しかし、技術情報の検索の必要性は、検索したい技術情報について十分な知識がないから生ずるのであり、検索したい技術情報を十分に得ている時点では検索する必要性がなくなっていることが大半であるため、検索者が、適切なキーワードを思いつくことは困難であるとされている。

【0010】

上記の従来手法(2)は、従来手法(1)の問題を解決可能であるが、別の問題がある。すなわち、従来手法(2)は、与えられた語の共起性を利用した連想結果を提供するものであるため、本来の検索目的や意図に合致しない検索結果(検索ノイズ)が多いという問題がある。

【0011】

上記の従来手法(3)は、従来手法(1)、(2)の問題を解決可能であるが、別の問題がある。すなわち、検索対象の情報に付与されている分類コードが少ない場合には、検索対象の分類に柔軟性がなくなる反面、分類コードが多い場合には、検索ノイズの問題がある。さらに、検索者が入力するキーワードが適切でない場合には、分類コードによる検索範囲が不適切となり、検索漏れが発生する可能性がある。また、検索対象が、分類コードの付いた情報に限定され、収集できる情報が限定的であるという問題がある。また、他のデータベースに蓄積された論文、製品等の情報について横断的検索ができず検索者にとって不便であるという問題もある。

【0012】

本発明は、これらの従来手法を考慮してなされたものであり、その目的は、検索者の入力した単語や文が検索用キーワードとして適正ではない場合でも、検索ノイズや検索漏れが少なく、かつ、汎用的な検索対象に対する検索を行える情報検索技術を提供することである。

【課題を解決するための手段】

【0013】

本発明に係る情報検索システムは、技術が適用される範囲を示すフィールドと、該フィールドに属する技術の範囲を示すテーマと、該テーマ毎に該技術の範囲に含まれる技術の特性を複数の観点に対応付けて示すタームとを定める技術分類を記憶する技術分類データベースと、技術情報を示す技術コンテンツを蓄積する技術コンテンツデータベースとを備える。該情報検索システムは、さらに、検索者によって入力された単語または単語の組合せである入力文字列を取得する入力文字列取得部と、前記技術分類データベースに記憶された技術分類のタームを前記入力文字列で照合し、一致するタームを含む全てのテーマを抽出し、抽出した該テーマから少なくとも1つのテーマを特定し、該特定したテーマに含

10

20

30

40

50

まれる複数のターム群を抽出する概念化部と、前記抽出したターム群をキーワードとして前記技術コンテンツデータベースをテキスト検索し、予め定めた評価方法で該検索した技術コンテンツを評価して、評価の良い技術コンテンツを取得するコンテンツ検索部と、前記取得した技術コンテンツを表示する表示用データを出力するコンテンツ表示部とを備える。

【発明の効果】

【0014】

上記した情報検索システムによれば、検索者が入力した文字列から、検索者が本来検索したいと思われる技術概念を示す検索条件（検索キーワード）を生成し、技術情報を検索することができる。例えば、検索者から、適切な検索キーワードが与えられていない場合でも、入力された文字列をもとに検索者が検索したい技術の範囲を特定し、特定した技術範囲に関連する概念を用いて検索処理を行うことにより検索対象を拡張し、さらに検索結果を評価することにより、検索漏れおよび検索ノイズが少ない検索結果を提供することができる。

10

【図面の簡単な説明】

【0015】

【図1】一実施例における情報検索システムのブロック構成例を示す図である。

【図2】一実施例における概念化部のブロック構成例を示す図である。

【図3】一実施例における技術分類データベースのフィールド情報およびテーマ情報のデータ構成例を示す図である。

20

【図4】一実施例における技術分類データベースのターム情報のデータ構成例を示す図である。

【図5】一実施例における技術分類データベースのターム情報のデータ構成例を示す図である。

【図6】一実施例における分類付テキストデータベースのデータ構成例を示す図である。

【図7】一実施例における技術コンテンツデータベースのデータ構成例を示す図である。

【図8】一実施例における情報検索システムの概要処理の処理フロー例を示す図である。

【図9】一実施例において、情報検索システムが表示する画面例を示す図である。

【図10】一実施例において、情報検索システムが表示する検索結果を表す画面例を示す図である。

30

【図11】一実施例における第1のブロック構成例でのテーマ特定処理のより詳細な処理フロー例を示す図である。

【図12】一実施例における第2のブロック構成例でのテーマ特定処理のより詳細な処理フロー例を示す図である。

【図13】一実施例におけるフィールド特定処理のより詳細な処理フロー例を示す図である。

【図14】一実施例におけるフィールド選択域を備える画面例を示す図である。

【図15】一実施例における検索結果の文献情報と製品情報とを表示する画面例を示す図である。

【図16】一実施例における検索結果の企業情報と製品情報とを表示する画面例を示す図である。

40

【図17】一実施例における情報検索システムのハードウェア構成例を示す図である。

【発明を実施するための形態】

【0016】

以下、本発明の一態様として開示する情報検索システムについて説明する。

【0017】

図1は、一実施例における情報検索システムのブロック構成例を示す図である。

【0018】

情報検索システム1は、予め備えた技術分類体系に基づいて、検索者が入力した文字列から検索条件を生成し、生成した検索条件で技術情報を検索する情報処理システムである

50

。

【0019】

情報検索システム1が備える技術分類体系は、技術が適用される産業に基づく分野（フィールド）を示すフィールド情報、各フィールドに属する技術について、その範囲を示す区分（テーマ）を示すテーマ情報、および1つのテーマに分類される技術的特徴を示す術語（ターム）を複数の観点に分類して定義しているターム情報で構成されている。

【0020】

すなわち、技術分類体系では、複数のフィールドで構成され、各フィールドは、下位構成として1または複数のテーマを有する。さらに、各テーマは、1または複数の観点の分類に属する1または複数の階層化したタームを有する。技術分類体系を構成する情報の詳細は、後述する。

10

【0021】

情報検索システム1は、入力文字列取得部11、形態素解析部13、概念化部15、コンテンツ検索部17、コンテンツ表示部19を有する検索装置10と、技術分類データベース31、分類付テキストデータベース33、技術コンテンツデータベース35の記憶装置とを備える。また、情報検索システム1の検索装置10は、検索者が使用する情報処理端末である検索者端末5とネットワークNを介して接続している。

【0022】

検索装置10の入力文字列取得部11は、検索者端末5で検索者が入力した文字列（入力文字列）を取得する。入力文字列は、単語、単語の組合せ、または任意の文等である。

20

【0023】

形態素解析部13は、入力文字列取得部11が入力文字列として文の入力を受け付ける場合に、入力文字列に形態素解析を行い、形態素から自立語を取り出す。形態素解析部13は、既知の一般的な形態素解析処理を実施する。

【0024】

概念化部15は、入力文字列である単語、単語の組合せ、または入力文字列から取り出した自立語を用いて、検索者が検索したい技術の範囲を示すテーマを検索し、予め定めた評価方法により評価の良いテーマを特定して、特定したテーマに属する技術的特徴を示すターム群を取得する。

【0025】

コンテンツ検索部17は、取得したターム群を用いて、技術コンテンツデータベース35に対するOR条件の検索処理を行い、予め定めた評価方法により検索結果を評価し、評価の良い技術コンテンツを取得する。

30

【0026】

コンテンツ表示部19は、取得された技術コンテンツを検索者端末5に表示させる。

【0027】

検索装置10の概念化部15は、1つの構成例（第1の構成例）として、図2（A）に示すように、テーマ検索部151、ターム群取得部153を備える。さらに、第1の構成例をとる場合には、情報検索システム1の分類付テキストデータベース33は不要である。

40

【0028】

テーマ検索部151は、技術分類データベース31に記憶されている技術分類体系を構成するテーマの中から、取り出した自立語に合致するテーマを抽出し、予め定められた評価方法でテーマを評価し、評価の良いテーマを少なくとも1つ特定する。

【0029】

ターム群取得部153は、特定したテーマに含まれる全ターム（ターム群）を取得する。

【0030】

概念化部15は、第1の構成例において、さらに、フィールド特定部159を備えることができる。

50

【0031】

フィールド特定部159は、技術分類データベース31から、テーマ検索部151が抽出した各テーマが属するフィールドをそれぞれ抽出し、抽出した各フィールドを、そのフィールドに属するテーマの評価を用いて評価し、最良の評価のフィールドを特定する。さらに、フィールド特定部159は、評価が近いフィールドが複数ある場合に、検索者端末5に該当する複数のフィールドを表示させ、入力されたフィールド選択に示されたフィールドを特定する。

【0032】

検索装置10の概念化部15は、別の構成例(第2の構成例)として、図2(B)に示すように、分類付テキスト検索部157、ターム群取得部153を備える。

10

【0033】

分類付テキスト検索部157は、分類付テキストデータベース33に記憶されているテキストに付与されたタームに取り出した自立語に合致するタームを持つテキストおよびそのテキストに付与されたタームを抽出し、抽出したテキストを用いた評価方法で抽出した各テーマを評価し、評価の良いテーマを少なくとも1つ特定する。

【0034】

概念化部15は、第2の構成例において、さらに、フィールド特定部159を備えることができる。

【0035】

ターム群取得部153、フィールド特定部159は第1の構成例と同じであるので説明を省略する。

20

【0036】

技術分類データベース31は、属性、用途、目的を含む多観点で技術を分類した分類体系を保持するデータベースである。

【0037】

図3~図5は、技術分類データベース31のデータ構成例を示す図である。

【0038】

技術分類データベース31は、フィールド情報311、テーマ情報313、およびターム情報315を含む。

【0039】

フィールド情報311は、技術が適用される産業に基づいて分類した分野(フィールド)を示す情報である。図3(A)に示すように、フィールド情報311では、各分野の名称を示す「フィールド名」とフィールドの識別情報である「フィールドID」とが対応付けられて記憶されている。例えば、フィールドID“2B”、フィールド名“畜産、漁業”は、自然資源のうち畜産や漁業に適用される技術が分類される区分であることを表している。

30

【0040】

テーマ情報313は、技術の範囲を示す区分(テーマ)を示す情報である。図3(B)に示すように、テーマ情報313では、技術の範囲の名称を示す「テーマ名」とテーマの識別情報である「テーマID」とテーマが属するフィールドを示す「フィールドID」とが対応付けられて記憶されている。例えば、テーマID“104”、テーマ名“養殖”は、フィールドID“2B”の畜産、漁業に属する技術であって、養殖に関連する技術が分類される区分であることを表している。

40

【0041】

ターム情報315は、1つのテーマに分類される技術的特徴を示す術語(ターム)を示す情報である。ターム情報315のタームが示す技術的特徴は、複数の観点に基づいて分類されて定義されている。

【0042】

図4および図5に示すように、ターム情報315は、テーマ毎に構築され、観点上の分類の名称を示す「観点分類」と、観点分類の識別情報である「観点コード」とが対応付け

50

られ、さらに、観点分類毎に、技術的特徴を示す「ターム名」とタームの識別情報である「タームコード」との対応が定義されている。各観点分類は、1または複数のタームコードおよびターム名の組（以下、単にタームと呼ぶことがある）を含み、複数のタームが階層的に定義されていてもよい。

【0043】

観点分類は、多面的に、かつ、テーマに対応して設定される。各観点分類は、目的（課題を含む）や、テーマに応じて、対象（物等）、用途、構造、材料、製造方法、使用方法、管理方法等が任意に設定される。

【0044】

図4は、テーマ名が“養殖（104）”に対するターム情報315の例である。図4のターム情報315は、例えば、観点名（観点コード）が“対象（AA）”について、“魚類（010）”、“いか・たこ類（020）”、“甲殻類（030）”等の1次ターム群が設定され、ターム“魚類（010）”の下位タームとして、“サケ類（011）”が設定されていることを示している。

10

【0045】

図5は、テーマ名が“飼料（150）”に対するターム情報315の例である。図5のターム情報315は、例えば、観点名（観点コード）が“装置（BA）”について、“網ヒビ（010）”、“小型藻場（020）”、“陸上養殖装置（030）”等の1次ターム群が設定され、下位タームも設定されうることを示している。

【0046】

分類付テキストデータベース33は、分類コードとして、技術分類データベース31の分類体系に基づくテーマとタームとが付与されたテキストを蓄積するデータベースである。

20

【0047】

図6は、分類付テキストデータベース33のデータ構成例を示す図である。

【0048】

分類付テキストデータベース33に蓄積されるテキストは、文献ID、題名、テーマID、タームコード、およびテキスト（本文）を含む。「文献ID」はテキストを識別する情報、「題名」は文献の題名、「テキスト」は文献の本文の内容である。テキストには、1または複数の分類コード、すなわち、1または複数のテーマIDが設定され、各テーマIDに対して1または複数のタームコードが設定されてもよい。

30

【0049】

技術コンテンツデータベース35は、さまざまな技術情報を示す技術コンテンツを蓄積するデータベースである。情報検索システム1は、複数の技術コンテンツデータベース35を備えてもよい。技術コンテンツデータベース35には、情報検索システム1がインターネット、イントラネット等のネットワークを介して接続可能であるデータベースを含む。

【0050】

ここで、技術コンテンツに含まれる技術情報は、技術自体だけではなく、技術を提供する企業や研究機関等の技術に関連する情報を範囲とするものであり、技術コンテンツは、技術情報を示すテキストデータを含むデータであればよく、画像データ、音声データ等が含まれていてもよい。技術コンテンツは、具体的には、特許文献、論文、製品マニュアル、製品カタログ、仕様書、技法解説書、解析・適用事例集、実験報告書、技術ニュース、技術マガジン、ニュースリリース、企業、研究機関の事業報告書等である。

40

【0051】

図7は、技術コンテンツデータベース35に蓄積される技術コンテンツのデータ構成例を示す図である。

【0052】

図7(A)は、製品情報を提供する技術コンテンツ（製品コンテンツ）のデータ構成例を示す。図7(A)の製品コンテンツは、製品ID、製品名、企業名、説明テキスト、リ

50

ンク等のデータを含む。「製品ID」は製品を識別する情報、「製品名」は製品の名称、「企業名」は製品を製造、提供する企業の名称である。「説明テキスト」は、製品の特徴、用途等を説明する本文テキストである。「リンク」は、任意の情報であり、製品に関連する情報を提供する企業や関連団体へのサイトへのリンクを示す。

【0053】

図7(B)は、技術論文を提供する技術コンテンツ(論文コンテンツ)のデータ構成例を示す。図7(B)の論文コンテンツは、文献ID、題名、著者名、要約テキスト、本文テキスト等のデータを含む。「文献ID」は論文を識別する情報、「題名」は論文の題名、「著者名」は論文著作者の氏名である。「要約テキスト」は、論文の要約を記述するテキスト、「本文テキスト」は、論文本体のテキストである。

10

【0054】

次に、情報検索システム1の処理の流れを説明する。以降の説明において、検索者からの情報入力、検索者への情報出力に関し、一般的なインタフェース機能およびその動作については既知であるので説明を省略する。

【0055】

図8は、情報検索システム1の概要処理の処理フロー例を示す図である。

【0056】

ステップS1: 検索装置10の入力文字列取得部11が、検索者が検索者端末5で入力した文字列(入力文字列)を取得する。入力文字列は、単語、単語の組合せ、自由文等である。

20

【0057】

図9は、情報検索システム1が表示する画面例を示す図である。図9に示す画面100が検索者端末5に表示される。画面100は、検索者が自由に語句や文を入力できる文字入力域101、検索実行を指示する検索ボタン103、検索結果の技術コンテンツを表示する結果表示域105を備える。

【0058】

検索ボタン103がマウスクリック等で押下されると、入力文字列取得部11は、文字入力域101に入力された入力文字列を取得する。本例では、“養殖の改善”を取得する。本例は、“養殖”、“改善”を抽出する。

【0059】

ステップS2: 形態素解析部13が、入力文字列取得部11が取得した入力文字列に対し形態素解析処理を行って、自立語(語)を抽出する。

30

【0060】

ステップS3: 概念化部15が、抽出した語を用いて技術分類データベース31または分類付テキストデータベース33のいずれかを用いてテーマを検索し、検索したテーマから最良の評価のテーマを1つ特定する。

【0061】

本例では、テーマIDが“104(養殖)”、“150(飼料)”、...の順で評価値が良く、最良の評価のテーマとして、“104(養殖)”が特定されているとする。

【0062】

ステップS4: ターム群取得部153が、特定されたテーマに属する複数のタームを抽出する。または、技術分類データベース31のターム情報315が階層構造を持つ場合に、ターム群取得部153は、特定されたテーマに属するタームの中から入力文字列に一致したタームと上位下位関係(親子関係)を持つタームのみを抽出するようにしてもよい。すなわち、ターム群取得部153が、入力文字列に一致したタームの下位に属する全ターム、または、上位に連なるターム、または、これら両方のタームを抽出する。

40

【0063】

本例では、ターム群取得部153が、技術分類データベース31のテーマID“104(養殖)”、“150(飼料)”のターム情報315から、それぞれ、タームを全て抽出する。

50

【 0 0 6 4 】

ステップ S 5 : コンテンツ検索部 1 7 は、抽出したターム群を用いて、技術コンテンツ DB 3 5 の技術コンテンツについて OR 条件のテキスト検索を行う。コンテンツ検索部 1 7 は、既知の検索手法であればどの検索手法であっても実施できるが、例えば、重み付き OR 検索、評価 OR 検索と呼ばれている検索手法を実施する。

【 0 0 6 5 】

ステップ S 6 : コンテンツ表示部 1 9 は、特定された技術コンテンツから、ターム群と類似度が高い技術コンテンツを評価が良いものとし、良評価順の技術コンテンツを含む表示用データを生成し、検索者端末 5 に表示用データを表示させる。

【 0 0 6 6 】

図 1 0 は、検索結果を表示する画面例を示す図である。

【 0 0 6 7 】

コンテンツ表示部 1 9 は、検索者端末 5 に表示された図 1 0 に示す画面 1 0 0 の結果表示域 1 0 5 に、特定したテーマに属するターム群による検索結果の技術コンテンツを表示させる。本例では、図 7 (A) に示す製品情報、図 7 (B) に示す製品情報、文献情報を含む表示用データが生成、表示される。なお、画面 1 0 0 の結果表示域 1 0 5 に設けた次表示ボタン (次ボタン) 1 0 5 ' のマウスクリック操作 (押下) により、コンテンツ表示部 1 9 は、押下された表示領域に対応する技術情報の種類で未表示の技術コンテンツを、評価が良い順に表示する。

【 0 0 6 8 】

さらに、コンテンツ表示部 1 9 は、画面 1 0 0 に、抽出テーマから特定したテーマを除外したテーマを評価が良い順に表示し、検索者が選択できるテーマ選択域 1 1 3 を表示する。本例では、テーマ選択域 1 1 3 に、評価の順に、“飼料”、...等のテーマが表示される。

【 0 0 6 9 】

画面 1 0 0 のテーマ選択域 1 1 3 で、結果表示域 1 0 5 に検索結果を表示しているテーマと別のテーマが選択された場合に、概念化部 1 5 は、選択されたテーマを特定したテーマとし (ステップ S 3)、そのテーマに属する全ターム群を抽出し (ステップ S 4)、抽出したターム群でコンテンツ検索処理が実行される (ステップ S 5)。コンテンツ表示部 1 9 は、選択されたテーマから抽出されたターム群によるコンテンツ検索結果を示す画面 1 0 0 を表示する (ステップ S 6)。

【 0 0 7 0 】

以下に、ステップ S 3 のテーマ検索処理のより詳細な処理を説明する。

【 0 0 7 1 】

図 1 1 は、第 1 の構成例におけるテーマ検索処理のより詳細な処理フロー例を示す図である。

【 0 0 7 2 】

概念化部 1 5 のテーマ検索部 1 5 1 は、形態素解析部 1 3 が抽出した語を用いて、技術分類データベース 3 1 のタームに対する OR 条件の検索処理を行い、一致するタームを持つテーマを全て抽出する (ステップ S 3 1)。

【 0 0 7 3 】

本例では、“養殖”、“改善”のタームにより、図 3 (B) に示す技術分類データベースのテーマ情報 3 1 3 から、テーマ ID “ 1 0 4 (養殖) ”、“ 1 5 0 (飼料) ”等が検索されたとする。

【 0 0 7 4 】

テーマ検索部 1 5 1 は、抽出したタームを用いて、検索した各テーマを予め定めた評価方法で評価する (ステップ S 3 2)。例えば、テーマ検索部 1 5 1 は、タームの希少性、重複性、網羅性をもとに検索したテーマを評価する。希少性は、語に一致するタームを持つテーマ数を調べ、タームを持つテーマ数が少ないほど、そのタームを持つテーマの評価が良くなる尺度である。重複性は、1つのテーマ内で語に一致するタームが出現している

10

20

30

40

50

数を調べ、テーマ内での出現数が多いほど、そのタームを持つテーマの評価が良くなる尺度である。網羅性は、そのタームがテーマが有するターム全体に占める割合を調べ、語に一致するタームが全タームに占める割合が高いほど、そのタームを持つテーマの評価が良くなる尺度である。テーマ検索部 151 は、これらの尺度の 1 つまたは複数の組み合わせを用いて、抽出した各テーマの評価値を算出する。

【0075】

そして、テーマ検索部 151 は、最良の評価のテーマを特定する（ステップ S33）。

【0076】

図 12 は、第 2 の構成例におけるテーマ検索処理のより詳細な処理フロー例を示す図である。

10

【0077】

分類付テキスト検索部 157 は、形態素解析部 13 が抽出した自立語（語）を用いて、分類付テキストデータベース 33 に記憶されている分類付テキストに OR 条件でテキスト検索を行い、分類付テキストを抽出する（ステップ S35）。

【0078】

分類付テキスト検索部 157 は、取得する分類付テキストを絞り込み、予め定めておいた数の分類付テキストを取得するようにしてもよい。本例では、分類付テキスト検索部 157 が、語の重み付けとして、図 11 のステップ S32 の処理と同様に、語の希少性、重複性、網羅性を算出して分類付テキストを評価し、評価の良い所定数の分類付テキストを取得する。なお、分類付テキスト検索部 157 は、既知の他の絞り込み処理を実行するものであってもよい。

20

【0079】

次に、分類付テキスト検索部 157 は、取得した各分類付テキストに付与されたテーマを抽出し（ステップ S36）、抽出した各テーマを評価し、評価の良い順に所定数のテーマを抽出する（ステップ S37）。例えば、抽出したテーマ毎に分類付テキスト数を算出し、分類付テキスト数が多いほど、良評価とする。また、分類付テキスト数の代わりに語と一致するタームの類似度を用いた評価を行ってもよく、類似度と分類付テキスト数とを組み合わせた評価を行ってもよい。

【0080】

分類付テキスト検索部 157 は、最良の評価のテーマを特定する（ステップ S38）。

30

【0081】

また、情報検索システム 1 は、抽出したテーマが属するフィールドを検索者に選択させ、選択されたフィールドの検索結果を表示することができる。この場合に、情報検索システム 1 のフィールド特定部 159 は、図 8 に示すステップ S3 と S4 との間に、フィールド特定処理を実行する。さらに、以下に説明するフィールド特定処理で特定されたフィールドについて、ステップ S3 の処理を再実行する。

【0082】

情報検索システム 1 は、検索装置 10 の概念化部 15 がフィールド特定部 159 を備える場合に、フィールド特定部 159 は以下の処理を行う。

【0083】

図 13 は、フィールド特定処理のより詳細な処理フロー例を示す図である。

40

【0084】

フィールド特定部 159 は、第 1 の構成例におけるテーマ検索部 151 または第 2 の構成例における分類付テキスト検索部 157 によって実行されたステップ S3 の処理で抽出された各テーマが属するフィールドを取得する（ステップ S41）。

【0085】

フィールド特定部 159 は、フィールド毎に、抽出されたテーマをまとめ（ステップ S42）、各フィールドを評価する（ステップ S43）。本例では、フィールド特定部 159 は、フィールドの評価として、まとめたテーマの評価値の合算値をフィールドの評価値とする。

50

【 0 0 8 6 】

算出した評価値が近いフィールドが複数ある場合に（ステップ S 4 4 の Y）、フィールド特定部 1 5 9 は、検索者に複数のフィールド名を表示して、検索者のフィールド選択を取得し（ステップ S 4 5）、フィールド選択に示されたフィールドを特定する（ステップ S 4 6）。評価値が近いフィールドが複数ない場合には（ステップ S 4 4 の N）、最も良い評価値のフィールドを特定する（ステップ S 4 7）。

【 0 0 8 7 】

図 1 4 は、フィールドを選択できるフィールド選択域 1 2 1 を持つ画面例を示す図である。図 1 4 に示す画面 1 0 0 は、入力文字列から抽出された語が“ラーメン”である場合の検索結果を示す画面例である。

10

【 0 0 8 8 】

抽出された語“ラーメン”が橋等の構築物の構造に関連する術語であり、かつ、穀物加工製品の名称でもあり、異なる 2 つのフィールドに属するテーマのタームとして存在しているため、検索された 2 つのフィールドの評価値が近くなる。

【 0 0 8 9 】

フィールド特定部 1 5 9 は、画面 1 0 0 のフィールド選択域 1 2 1 に、“ラーメン”から特定された 2 つのフィールド“橋または陸橋”と“穀類誘導製品、スープ、その他食品”とを表示し、検索者のフィールド選択を受け付ける。

【 0 0 9 0 】

これにより、入力文字列から抽出される語が、複数のフィールドで使用される術語であって検索対象とする技術分野の判別が難しい場合にも、検索者に技術分野を選択させ、より適切な検索結果を表示させることが可能となる。

20

【 0 0 9 1 】

図 1 5 および図 1 6 は、情報検索システム 1 により検索者端末 5 に表示される他の画面例を示す図である。

【 0 0 9 2 】

情報検索システム 1 は、検索結果を技術情報の種類毎に技術コンテンツを表示することができる。この場合に、コンテンツ検索部 1 7 は、製品情報、文献情報、企業情報等をそれぞれ蓄積する複数の技術コンテンツデータベース 3 5 を横断的に検索し、検索した技術コンテンツを技術情報の種類に基づいて分類し、良い評価の技術コンテンツから表示する。

30

【 0 0 9 3 】

図 1 5 は、検索結果として文献情報と製品情報とを表示する場合の画面例を示す図である。図 1 5 に示す画面 1 0 0 は、結果表示域 1 0 5 に相当する検索結果の表示領域として、文献情報を示す文献検索結果表示域 1 1 5、製品情報を示す製品検索結果表示域 1 1 7 を有する。

【 0 0 9 4 】

画面 1 0 0 の文献検索結果表示域 1 1 5 に設けた次表示ボタン（次ボタン）1 1 5'（図示しない）、または、製品検索結果表示域 1 1 7 に設けた次表示ボタン（次ボタン）1 1 7' のマウスクリック操作（押下）により、コンテンツ表示部 1 9 は、押下された表示領域に対応する技術情報の種類で未表示の技術コンテンツを、評価が良い順に表示する。

40

【 0 0 9 5 】

図 1 6 は、検索結果として企業情報と製品情報とを表示する場合の画面例を示す図である。図 1 6 に示す画面 1 0 0 は、結果表示域 1 0 5 に相当する検索結果の表示領域として、製品検索結果表示域 1 1 7 と企業情報を示す企業検索結果表示域 1 1 9 とを有する。

【 0 0 9 6 】

さらに、情報検索システム 1 は、画面 1 0 0 に表示する技術情報に、関連する他の技術情報や副次的情報に対するリンク情報を設定することができる。

【 0 0 9 7 】

コンテンツ表示部 1 9 は、画面 1 0 0 に表示する検索結果の表示用データを生成する際

50

に、検索結果の技術コンテンツ毎に、関連するサイト等へのリンク情報を付加することができる。例えば、図16に示すように、コンテンツ表示部19は、画面100の製品検索結果表示域117に表示させる各製品情報に、その製品に関連する企業サイトへのリンク情報が設定、表示する。

【0098】

以上説明した情報検索システム1の検索装置10は、図1に示す処理部を備える専用ハードウェアとして実施することができる。

【0099】

また、検索装置10は、図17に示すような、CPU301、メモリ302、表示制御部304、表示装置303、記憶装置(ハードディスク)305、入力装置(キーボード)306、通信制御部307が内部のネットワーク等で接続されたコンピュータ300として実施することができる。

10

【0100】

さらに、検索装置10は、コンピュータ300で実行可能なプログラムとして実施することができる。この場合に、図1に示す情報検索システム1の処理部の機能を実現するプログラムを実装し、コンピュータ100上で実行することにより、実施する。

【0101】

すなわち、図1に示した入力文字列取得部11、形態素解析部13、概念化部15、コンテンツ検索部17、コンテンツ表示部19の機能をコンピュータ300に実行させる実行プログラムをコンピュータ300に読み込ませ、実行させることによって、検索装置10を実施することができる。なお、実行プログラムは、CD-ROM、CD-RW、DVD-R、DVD-RAM、DVD-RW等やフレキシブルディスク等の記録媒体だけでなく、通信回線の先に備えられた他の記憶装置やコンピュータのハードディスク等に記憶されるものであってもよい。

20

【0102】

なお、情報検索システム1を構成する要素は、任意の組合せで実現されてもよい。複数の構成要素が1つの部材として実現されてもよく、1つの構成要素が複数の部材から構成されてもよい。また、情報検索システム1は、上述した実施形態に限定されず、本発明の要旨を逸脱しない範囲において各種の改良および変更を行ってもよいことは当然である。例えば、上述の実施例において、情報検索システム1の検索対象を、文献情報、製品情報、企業情報等の技術情報を例に説明したが、情報検索システム1は、例えば論文、事例集、技術ニュース、実験報告書等の他の情報検索にも適用することができる。

30

【0103】

以上説明したように、本発明に係る情報検索システム1によれば、検索者によって入力された文字列“養殖の改善”から抽出された単語(養殖、改善)のみによる検索処理ではなく、検索したい技術に関連するテーマとして特定した範囲の技術の概念を形成するターム群を用いた検索処理を実行することができる。これにより、検索者が入力文字列によって提示できなかった概念を表す術語(例えば、図4に示すターム群)による検索処理が行われ、その検索結果の全体から、評価の良い検索結果(技術情報)が検索者端末5に表示される。

40

【0104】

すなわち、情報検索システム1によれば、従来の入力文字列による検索手法では検索結果中での評価が低くなり表示される可能性が低くなるような技術コンテンツ、例えば、図7(A)の製品名「使い捨て製品」の製品情報等を、検索結果の上位に表示させることができ、検索ノイズや検索漏れを配慮した検索処理を実現することができる。

【0105】

よって、検索者は、入力した単語の組合せや文が適切でなくても、検索したい技術概念に合致した情報コンテンツを取得できる可能性が非常に高くなる。

【0106】

また、情報検索システム1によれば、入力された文字列に関連する複数のテーマを表示

50

して、検索者に選択させることができる。

【0107】

よって、検索者は、検索したい技術と類似または関連する範囲を認識して、より適切な検索結果を取得できる可能性が高くなる。

【符号の説明】

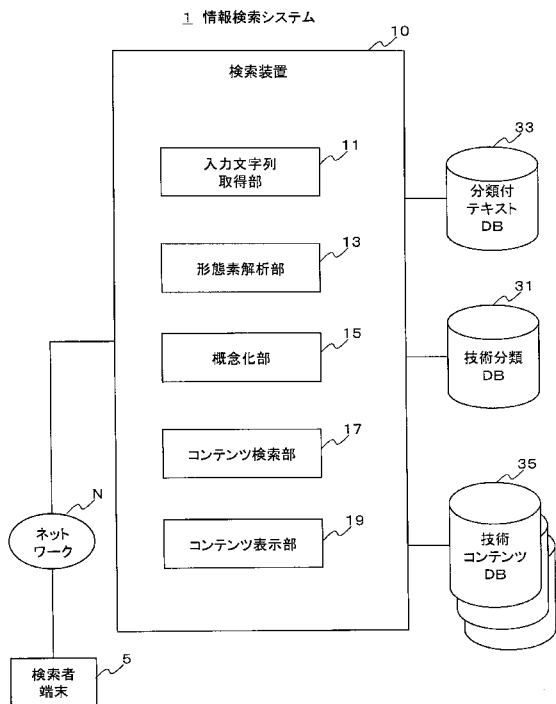
【0108】

- 1 情報検索システム
- 10 検索装置
- 11 入力文字列取得部
- 13 形態素解析部
- 15 概念化部
- 151 テーマ検索部
- 153 ターム群取得部
- 157 分類付テキスト検索部
- 159 フィールド特定部
- 17 コンテンツ検索部
- 19 コンテンツ表示部
- 31 技術分類データベース(DB)
- 33 分類付テキストデータベース(DB)
- 35 技術コンテンツデータベース
- 5 検索者端末

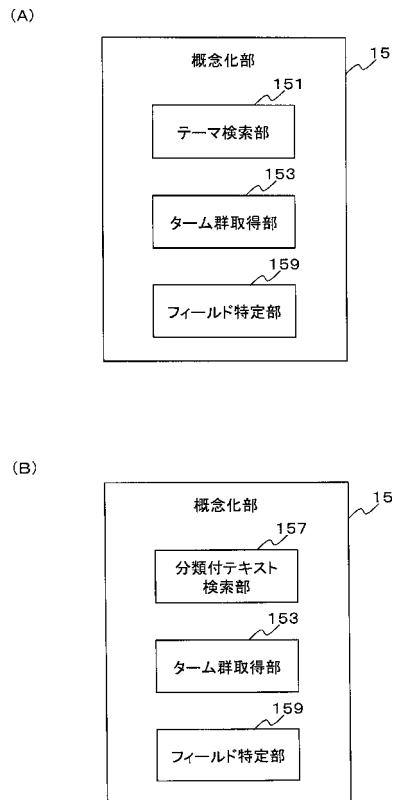
10

20

【図1】



【図2】



【図3】

(A)

フィールドID	フィールド名
2B	畜産、漁業
...	...
4C	医療
4B	食品、微生物
...	...

311

(B)

テーマID	テーマ名	フィールドID
104	養殖	2B
...
150	飼料	2B
...

313

【図4】

観点コード	観点分類	タームコード	ターム	タームコード	ターム	タームコード	ターム	タームコード	ターム
AA	対象	010	魚類	011	サケ類
		020	いかいたこ類	031	えび類
		030	甲殻類	041	えび類
		040	貝類
BA	目的	010	特性成長時期	011	産卵準備期
		020	体色対策
		030	病氣予防	031	薬剤使用
CA	装置の用途	010	養殖用
		010	水槽の形態
DA	装置の構造
...

315

【図5】

観点コード	観点分類	タームコード	ターム	タームコード	ターム	タームコード	ターム
AA	対象	010	のり
		020	昆布・わかめ類
		030	まりも
		040	クロレラ
BA	装置	010	網ヒビ
		020	小型薬場
		030	陸上養殖装置
CA	目的	010	増殖
		020	予防・駆除
...

315

【図6】

文献ID	题名
2012001001	養殖従用ローブ
テーマID	タームコード
2B026	BA011
テーマID	タームコード
2B104	AA040;DA030
テーマID	タームコード
3B153	BB040;CA021
...	...
テキスト(本文)	

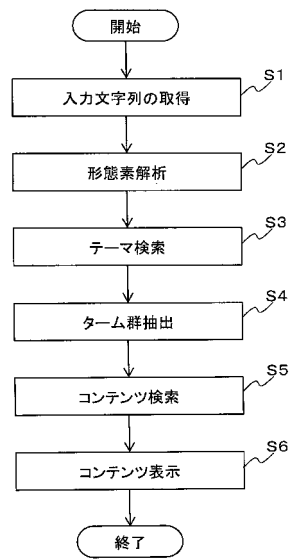
33

分類コード

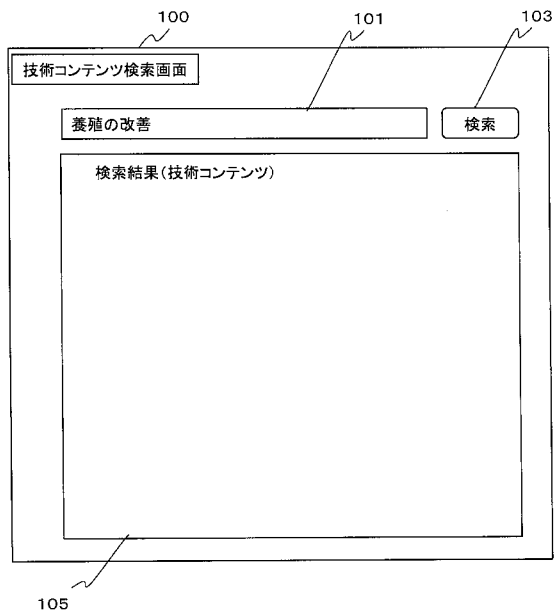
【 図 7 】



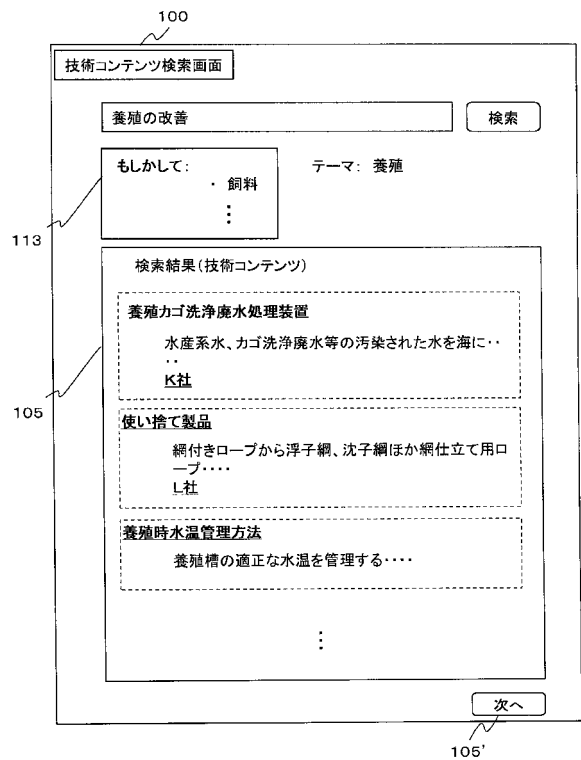
【 図 8 】



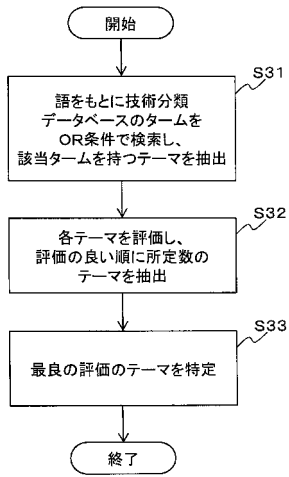
【 図 9 】



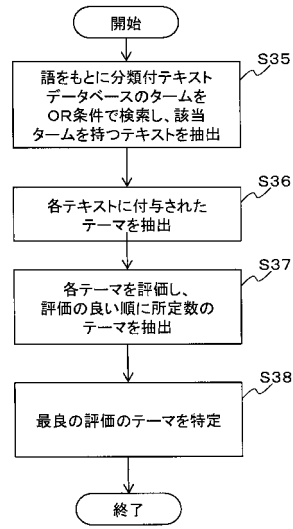
【 図 10 】



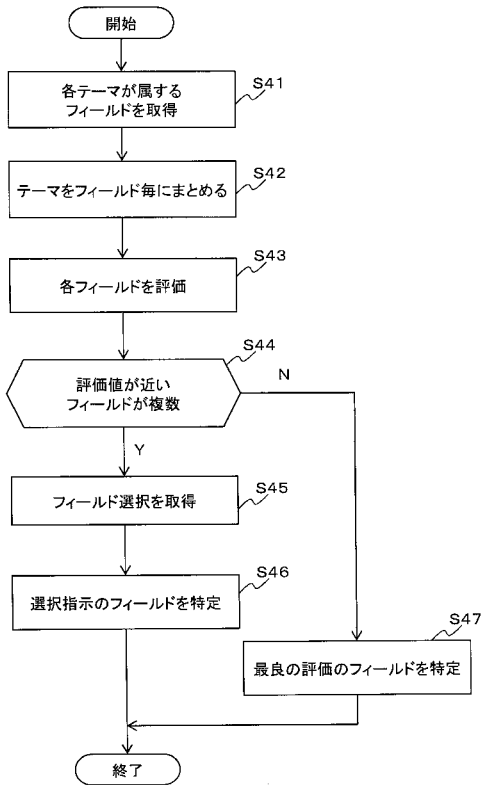
【 図 1 1 】



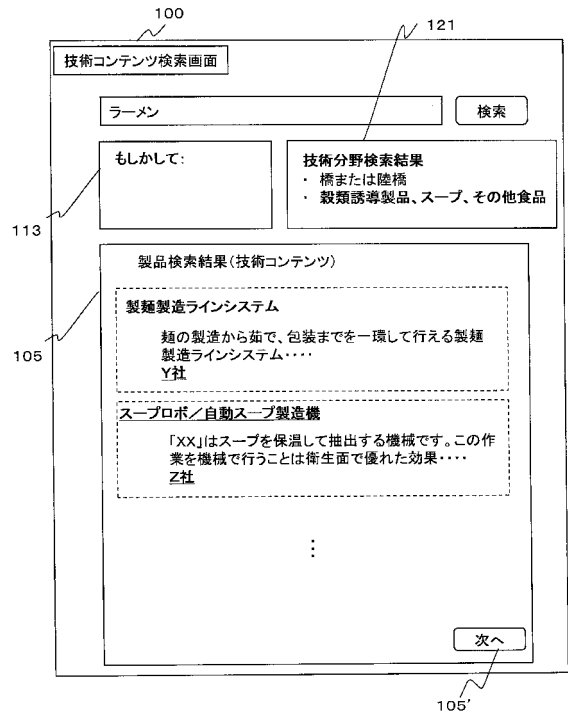
【 図 1 2 】



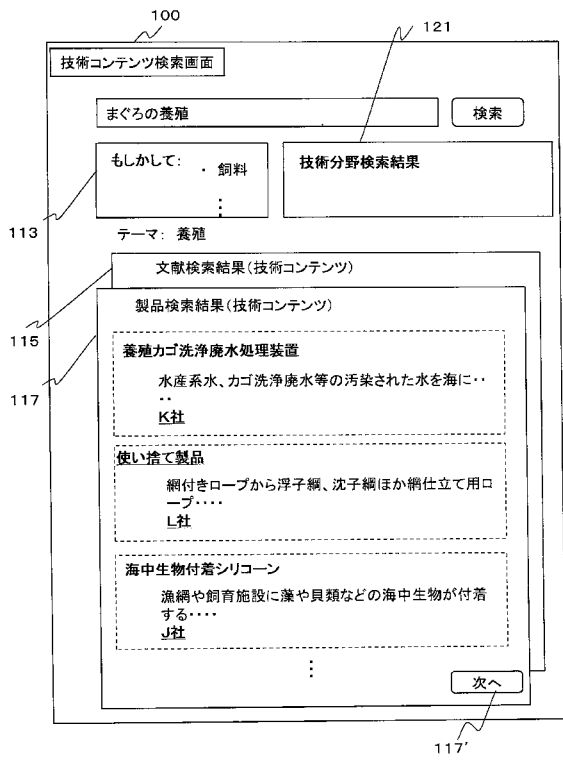
【 図 1 3 】



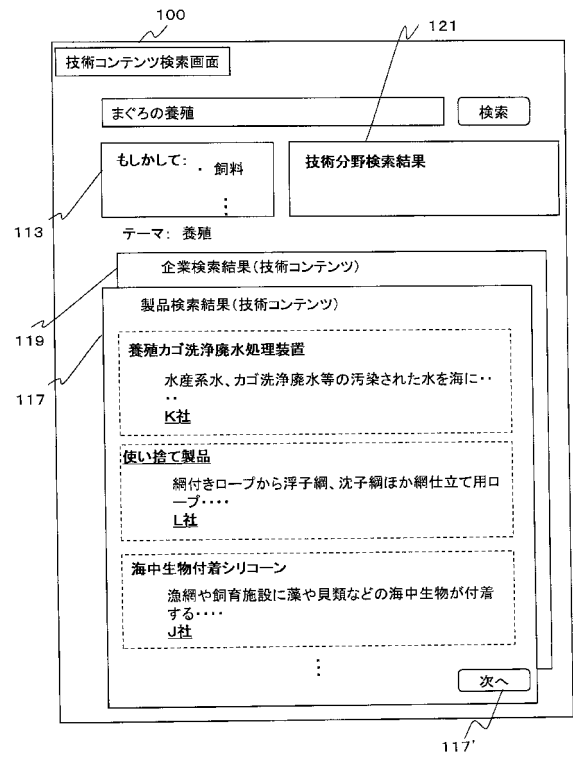
【 図 1 4 】



【 図 1 5 】



【 図 1 6 】



【 図 1 7 】

