

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4951407号
(P4951407)

(45) 発行日 平成24年6月13日(2012.6.13)

(24) 登録日 平成24年3月16日(2012.3.16)

(51) Int.Cl.

G06F 9/44 (2006.01)
G06F 13/00 (2006.01)

F 1

G06F 9/06 620C
G06F 9/06 620K
G06F 13/00 560A

請求項の数 4 (全 43 頁)

(21) 出願番号	特願2007-128824 (P2007-128824)
(22) 出願日	平成19年5月15日 (2007.5.15)
(65) 公開番号	特開2008-287311 (P2008-287311A)
(43) 公開日	平成20年11月27日 (2008.11.27)
審査請求日	平成21年11月26日 (2009.11.26)

(73) 特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(74) 代理人	110000279 特許業務法人 ウィルフォート国際特許事務所
(72) 発明者	小川 祐一 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内
(72) 発明者	八高 克志 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内

最終頁に続く

(54) 【発明の名称】コンテンツ部品検索方法及び装置

(57) 【特許請求の範囲】

【請求項 1】

We b コンテンツを作成するために該We b コンテンツの部品とするコンテンツ部品を検索するための装置において、

テキストを表すテキスト情報と入力コントロールの種類を表す入力コントロール種類情報との組合せを含んだ検索条件の入力を受け付ける検索条件入力部と、

前記テキスト情報に対応したテキストと前記入力コントロール種類情報に対応した入力コントロールとの組み合わせからなるコンテンツ部品を複数含み且つ前記テキスト情報と前記入力コントロール種類情報との組み合わせを複数含む情報パッケージを複数記憶する記憶部と、

前記入力された検索条件に含まれているテキスト情報及び入力コントロール種類情報を用いて前記記憶部から前記情報パッケージを検索して、該検索条件で検索された前記情報パッケージの該検索条件に対する類似度を算出する類似度算出部と、

前記検索条件で検索された複数の前記情報パッケージをそれぞれの前記算出された類似度に基づいて表示画面に表示する部品候補表示部とを備えるコンテンツ部品検索装置。

【請求項 2】

前記類似度は、前記検索条件に含まれている少なくとも1つの前記テキスト情報と前記情報パッケージに含まれている少なくとも1つのテキスト情報との第一の類似度と、前記検索条件に含まれている少なくとも1つの入力コントロール種類情報と前記情報パッケー

ジに含まれている少なくとも1つの入力コントロール種類情報との第二の類似度に基づいて算出された値である、

ことを特徴とする請求項1記載のコンテンツ部品検索装置。

【請求項3】

Webコンテンツを作成するために該Webコンテンツの部品とするコンテンツ部品を検索するための装置における方法において、

テキストを表すテキスト情報と入力コントロールの種類を表す入力コントロール種類情報との組合せを含んだ検索条件の入力を受け付け、

前記テキスト情報に対応したテキストと前記入力コントロール種類情報に対応した入力コントロールとの組み合わせからなるコンテンツ部品を複数含み且つ前記テキスト情報と前記入力コントロール種類情報との組み合わせを複数含む情報パッケージを複数記憶した記憶部から、前記入力された検索条件に含まれているテキスト情報及び入力コントロール種類情報を用いて前記情報パッケージを検索して、該検索条件で検索された前記情報パッケージの該検索条件に対する類似度を算出し、

前記検索条件で検索された複数の前記情報パッケージをそれぞれの前記算出された類似度に基づいて表示画面に表示する、

ことを特徴とするコンテンツ部品検索方法。

【請求項4】

Webコンテンツを作成するために該Webコンテンツの部品とするコンテンツ部品を検索するための装置としてコンピュータを機能させるためのプログラムにおいて、

テキストを表すテキスト情報と入力コントロールの種類を表す入力コントロール種類情報との組合せを含んだ検索条件の入力を受け付ける検索条件ステップと、

前記テキスト情報に対応したテキストと前記入力コントロール種類情報に対応した入力コントロールとの組み合わせからなるコンテンツ部品を複数含み且つ前記テキスト情報と前記入力コントロール種類情報との組み合わせを複数含む情報パッケージを複数記憶する記憶部から、前記入力された検索条件に含まれているテキスト情報及び入力コントロール種類情報を用いて前記情報パッケージを検索して、該検索条件で検索された前記情報パッケージの該検索条件に対する類似度を算出する類似度算出ステップと、

前記検索条件で検索された複数の前記情報パッケージをそれぞれの前記算出された類似度に基づいて表示画面に表示する部品候補表示ステップと

を前記コンピュータに実行させるためのコンテンツ部品検索プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンピュータを利用した情報検索技術に関する。

【背景技術】

【0002】

例えば、情報検索技術の一つとして、特許文献1（特開2003-15872）に開示の技術がある。特許文献1に開示の技術によれば、オブジェクト指向で開発されたソフトウェアの各オブジェクトを検索キーに基づいて効率的に検索するオブジェクト検索装置が提供される。

【特許文献1】特開2003-15872号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

ところで、テキストや入力コントロール（例えば、テキストボックス、ラジオボタンあるいはチェックボックスなど）を有する画面コンテンツの設計が必要となるケースがある。例えば、電子申請システムでは、そのような画面コンテンツとして、たくさんの入出力画面（例えば、電子申請システムが大規模であれば数千の入出力画面）の設計が必要となる。このケースにおいて、そのような画面コンテンツを人手で一から設計するとなると、開

10

20

30

40

50

発者の負担が大きい。

【0004】

そのような画面コンテンツの設計を支援する方法として、例えば、所望のテキストを検索キーとして入力し、その検索キーと一致するテキストを有した既存の画面コンテンツを検索し、その既存の画面コンテンツを流用する方法が考えられる。しかし、この方法よりも開発者の負担を軽減することが望ましい。

【0005】

従って、本発明の目的は、テキストや入力コントロールを有する画面コンテンツの開発者の負担を、テキストを検索キーとして既存の画面コンテンツを検索し流用する方法よりも軽減することにある。

10

【0006】

本発明の他の目的は、後の説明から明らかになるであろう。

【課題を解決するための手段】

【0007】

記憶部が、一の画面コンテンツにつき一以上のコンテンツ部品情報グループを記憶する。コンテンツ部品情報グループは、テキストと入力コントロールとを含んだコンテンツ部品を表し、一以上の情報パッケージを含む。情報パッケージには、表示されるテキストを表すテキスト情報要素と入力コントロールの種類を表す入力コントロール種類情報要素とが含まれる。検索装置が、テキスト情報要素と入力コントロール種類情報要素との組合せを含んだ検索条件の入力を受け、その検索条件と記憶部内の各コンテンツ部品情報グループとの類似度を、該検索条件に含まれているテキスト情報要素及び入力コントロール種類情報要素とコンテンツ部品情報グループに含まれているテキスト情報要素及び入力コントロール種類情報要素とに基づいて算出する。そして、検索装置が、その検索条件についてのコンテンツ部品候補に関する情報を上記算出された類似度に基づいて表示画面に表示する。

20

【発明を実施するための最良の形態】

【0008】

一つの実施形態では、検索装置に、検索条件入力部、部品検索部及び部品候補表示部を備えることで、コンテンツ部品検索装置が構築される。検索条件入力部は、テキストを表すテキスト情報要素と入力コントロールの種類を表す入力コントロール種類情報要素との組合せを含んだ検索条件の入力を受け付ける。部品検索部が、記憶部（例えば記憶資源）内の各コンテンツ部品情報グループと入力された検索条件との類似度を、該検索条件に含まれているテキスト情報要素及び入力コントロール種類情報要素とコンテンツ部品情報グループに含まれているテキスト情報要素及び入力コントロール種類情報要素とに基づいて算出する。部品候補表示部が、検索条件についてのコンテンツ部品候補に関する情報を算出された類似度に基づいて表示画面に表示する。

30

【0009】

ここで、コンテンツ部品とは、テキスト又は入力コントロールを最小単位とし、テキストと入力コントロールとの一つの組合せ、その組合せを複数個含んだグループ、或いは画面コンテンツそれ自体であっても良い。具体的には、例えば、階層構造を有する画面コンテンツについて、階層レベルが最高の場合には、コンテンツ部品は、画面コンテンツそれ自体であり、階層レベルが最低の場合には、コンテンツ部品は、一のテキスト、一の入力コントロール又はそれらの一の組合せであり、階層レベルが最高でも最低でも無い場合には、コンテンツ部品は、テキスト及び入力コントロールの組合せを複数個有したグループであって良い。

40

【0010】

また、「表示画面」は、一又は複数の表示装置がそれぞれ有するハードウェアとしての一又は複数の表示画面であっても良いし、一又は複数のハードウェアとしての表示画面に表示されるソフトウェアとしての表示画面であっても良い。

【0011】

50

また、「コンテンツ部品候補に関する情報」としては、コンテンツ部品の識別子、そのコンテンツ部品に対応した類似度、或いは、そのコンテンツ部品に対応したコンテンツ部品情報グループの或る情報パッケージに含まれているテキスト情報要素など、種々の情報を採用することができる。

【0012】

また、「コンテンツ部品候補に関する情報を算出された類似度に基づいて表示する」とは、例えば、類似度の降順に表示することであっても良いし、所定の閾値以上の類似度であるコンテンツ部品候補に関する情報を表示することであっても良い。

【0013】

一つの実施形態では、類似度は、検索条件に含まれている一以上のテキスト情報要素とコンテンツ部品情報グループに含まれている一以上のテキスト情報要素との第一の類似度と、検索条件に含まれている一以上の入力コントロール種類情報要素とコンテンツ部品情報グループに含まれている一以上の入力コントロール種類情報要素との第二の類似度に基づいて算出された値（例えば所定の計算式に第一の類似度及び第二の類似度を代入することで算出された値）とすることができます。

10

【0014】

一つの実施形態では、類似度は、検索条件に含まれているテキスト情報要素及び入力コントロール種類情報要素の一以上のセットと、コンテンツ部品情報グループに含まれているテキスト情報要素及び入力コントロール種類情報要素の一以上のセットとの類似度であるとすることができます。

20

【0015】

一つの実施形態では、コンテンツ部品検索装置が、更に、コンテンツ解析部、グルーピング部、パッケージグループ生成部、テキスト付与部及び部品登録部を備えることができる。コンテンツ解析部は、画面コンテンツ（例えば入力フォーム）を表すコンテンツファイルを解析することで、テキスト情報要素、テキスト配置位置、入力コントロール種類情報要素、及び入力コントロール配置位置を特定することができる。具体的には、例えば、コンテンツ解析部は、コンテンツファイルを解析することで、テキスト情報要素及びテキスト配置位置情報要素の組合せであるテキスト／位置組合せと、入力コントロール種類情報要素及び入力コントロール配置位置情報要素の組合せである入力コントロール／位置組合せとを特定し、特定されたテキスト／位置組合せをテキスト管理情報に格納し、特定された入力コントロール／位置組合せを入力コントロール管理情報に格納することができる。グルーピング部は、コンテンツファイルの解析の結果を基に（例えば入力コントロール管理情報を参照することにより）、それぞれの入力コントロール配置位置関係を特定し、特定された位置関係が所定の第一の条件を満たす場合に、前記条件を満たす複数の入力コントロールをグループとして抽出することができる。パッケージグループ生成部は、抽出したグループを構成する複数の入力コントロールに対応した複数の入力コントロール種類情報要素を含んだ複数の情報パッケージを有するパッケージグループを生成することができる。テキスト付与部は、例えば、コンテンツファイルの解析の結果を基に（例えばテキスト管理情報から）、情報パッケージにおける入力コントロール種類情報要素に対応するテキスト情報要素を取得し、取得したテキスト情報要素をその情報パッケージに含めることができる（例えば、取得したテキスト情報要素をラベルとして情報パッケージに含めることができる）。部品登録部は、テキスト情報要素がそれぞれ含まれられた複数の情報パッケージを有するパッケージグループであるコンテンツ情報グループを、解析されたコンテンツファイルに関連付けて記憶部に格納することができる。

30

【0016】

一つの実施形態では、所定の第一の条件は、入力コントロール同士の配置位置間の距離が所定の値以下であるとすることができる。

【0017】

一つの実施形態では、情報パッケージに含められるテキスト情報要素は、その情報パッケージ内の入力コントロール種類情報要素に対応した入力コントロールとの間の配置位置

40

50

関係が所定の第二の条件を満たすテキストを表す情報要素であるとすることができる。

【0018】

一つの実施形態では、所定の第二の条件は、テキストの配置位置と入力コントロールの配置位置との間の距離が所定の値以下であるとすることができる。

【0019】

一つの実施形態では、コンテンツ情報グループは、最低階層レベルの末端に対応する情報パッケージをリーフとし、最低階層レベル以外の階層レベルの先頭に対応する情報パッケージをプランチとする情報パッケージツリーであるとすることができる。コンテンツ部品情報グループは、コンテンツ情報グループそれ自体又はそのサブセットであるとすることができます。

10

【0020】

一つの実施形態では、コンテンツ部品候補に関する情報には、コンテンツ部品情報グループの最上位にある情報パッケージ内のテキスト情報要素が含まれてもよい。

【0021】

一つの実施形態では、コンテンツ部品候補は、算出された類似度が所定の閾値以上であるコンテンツ部品情報グループに対応したコンテンツ部品であるとすることができる。

【0022】

一つの実施形態では、部品候補表示部が、更に、コンテンツ部品候補の数を表示することができる。

【0023】

一つの実施形態では、部品候補表示部は、入力された検索条件が複数個存在し、それらの検索条件が階層関係（例えば親子関係）にある場合、上位（例えば親）の検索条件についてのコンテンツ部品候補に関する情報を優先的に表示することができる。

20

【0024】

一つの実施形態では、部品登録部が、同一又は類似のコンテンツ部品情報グループの集まりである部品グループを作成することができる。部品検索部が、入力された検索条件と各部品グループとの類似度を算出することができる。部品候補表示部が、コンテンツ部品候補に関する情報として部品グループに関する情報を表示することができる。

【0025】

ここで、検索条件と部品グループとの類似度としては、その部品グループに属する一のコンテンツ部品情報グループと検索条件との類似度であっても良いし、その部品グループに属する複数のコンテンツ部品情報グループと検索条件との複数の類似度を用いた所定の計算により算出された値（例えば複数の類似度の平均）であっても良い。

30

【0026】

また、コンテンツ部品情報グループと類似する部品グループとしては、類似度が所定の閾値を越えている部品グループのうちの一の（例えば最も類似度が高い）部品グループであるとすることができる。

【0027】

一つの実施形態では、部品登録部が、更に、一の部品グループについて複数のスタイルグループを作成することができる。スタイルグループは、同一又は類似のスタイルを有する複数のコンテンツ部品情報グループの集まりであるとすることができる。部品候補表示部が、部品グループに関する情報に加えスタイルグループに関する情報を表示することができる。

40

【0028】

ここで、テキストのスタイルは、例えば、テキストのフォントサイズ、フォントタイプ（例えば、明朝、ゴシックなど）、フォントスタイル（例えば、通常、太字、斜体など）、フォントの色、或いはテキストの長さなど以上のテキストスタイル属性により定義することができる。

【0029】

また、スタイルが類似しているとは、スタイルに関する複数の要素（例えば、テキスト

50

、テキストの並び順、テキストのフォントサイズ、フォントタイプなど)が所定の割合(例えば80%)以上に一致していることとすることができます。

【0030】

一つの実施形態では、部品登録部が、新規のコンテンツファイルを記憶部に登録する際に、該新規のコンテンツファイルについてのコンテンツ部品情報グループを、そのコンテンツ部品情報グループと同一又は類似の部品グループに分類することができる。

【0031】

一つの実施形態では、部品登録部が、登録性能よりも検索性能を重視することが入力された場合に、新規のコンテンツファイルを前記記憶部に登録する際に、該新規のコンテンツファイルについてのコンテンツ部品情報グループを、そのコンテンツ部品情報グループと同一又は類似の部品グループに分類することができる。 10

【0032】

一つの実施形態では、部品検索部が、どの部品グループにも分類されていないコンテンツ部品情報グループと前記検索条件との類似度を算出することができる。部品登録部が、どの部品グループにも分類されていないコンテンツ部品情報グループに対応したコンテンツ部品がコンテンツ部品候補とされた場合に、そのコンテンツ部品情報グループをそれに該当する部品グループに分類することができる。

【0033】

一つの実施形態では、部品登録部が、検索性能よりも登録性能を重視することが入力された場合に、どの部品グループにも分類されていないコンテンツ部品情報グループに対応したコンテンツ部品がコンテンツ部品候補とされた場合に、そのコンテンツ部品情報グループをそれに該当する部品グループに分類することができる。 20

【0034】

一つの実施形態では、複数のコンテンツ部品情報グループが二以上の部品グループに分類されており、部品検索部が、前記入力された検索条件と各部品グループとの類似度を算出し、部品候補表示部が、コンテンツ部品候補に関する情報として、部品グループに関する情報を表示することができる。

【0035】

一つの実施形態では、一の部品グループについて、複数のスタイルグループがあり、スタイルグループは、同一又は類似のスタイルを有する複数のコンテンツ部品情報グループの集まりであり、部品候補表示部が、部品グループに関する情報に加えスタイルグループに関する情報を表示することができる。 30

【0036】

一つの実施形態では、第一の部品グループに属する既存部品サブツリー内の情報パッケージが、第一の部品グループとは別の第二の部品グループにも属していれば、その第二の部品グループが、第一の部品グループに含まれる部品グループであり、部品候補表示部が、第一の部品グループに含まれる第二の部品グループに関する情報を表示することができる。

【0037】

一つの実施形態では、部品候補表示部が、スタイルグループに対応したコンテンツ部品情報グループの数を表示することができる。 40

【0038】

一つの実施形態では、検索条件は、階層構造を持つ画面コンテンツの定義情報又はそのサブセットであるとすることができます。

【0039】

上述した複数の実施形態のうちの二以上の実施形態を組み合わせることができる。また、上述した各部(例えば、検索条件入力部、部品検索部、部品候補表示部、コンテンツ解析部、グレーピング部、パッケージグループ生成部、テキスト付与部及び部品登録部)は、各手段と言い換えてよい。各部は、ハードウェア(例えば回路)、コンピュータプログラム、或いはそれらの組み合わせ(例えば、一部をコンピュータプログラムで実行しー 50

部をハードウェア回路で実行すること)によって実現することもできる。各コンピュータプログラムは、コンピュータマシンに備えられる記憶資源(例えばメモリ)から読み込むことができる。その記憶資源には、CD-ROMやDVD(Digital Versatile Disk)等の記録媒体を介してインストールすることもできるし、インターネットやLAN等の通信ネットワークを介してダウンロードすることもできる。

【0040】

以下、図面を参照して、本発明の幾つかの実施形態を詳細に説明する。

【0041】

<第一の実施形態>。

【0042】

画面コンテンツとして、Webコンテンツを例に採ることができる。Webコンテンツとしては、HTML(HyperText Markup Language)コンテンツ、独自形式のWebコンテンツ、あるいは文書作成アプリケーションで作成された電子化文書など、画像や動画を除くテキストを主体としたデジタルコンテンツとすることができる。第一の実施形態では、Webコンテンツとして、HTML形式の入力フォームを例に採ることができる。以下の説明では、入力フォームなどのようなWebコンテンツを、単に「コンテンツ」と呼ぶことにする。

【0043】

第一の実施形態において行われる処理を、部品登録ステップと部品検索ステップとに大別することができる。

【0044】

部品登録ステップでは、既存のコンテンツ中に含まれる要素(テキストや入力コントロール)の位置情報などのレイアウト情報を用いて要素間をグループ化するグルーピングステップと、各グループおよび要素に対してラベルを付与するラベル付与ステップと、それらグルーピングステップとラベル付与ステップの結果からコンテンツの構成を表すツリー(以下、コンテンツ構成ツリーと呼ぶ)を生成するコンテンツ構成ツリー生成ステップと、コンテンツ構成ツリーをコンテンツと関連付けてレポジトリに登録するコンテンツ登録ステップとが行われる。

【0045】

部品検索ステップでは、階層構造を持つ新規コンテンツの定義情報のサブツリー(以下、クエリツリーと呼ぶ)と、レポジトリに蓄積されている各コンテンツ構成ツリーのサブツリー(以下、既存部品サブツリーと呼ぶ)との類似度を算出する部品検索ステップと、クエリツリーに対して類似度の高い(例えば所定の閾値以上の)既存部品サブツリーを、クエリツリーに対する部品候補として出力する部品候補出力ステップとが行われる。

【0046】

以下、より詳細に説明する。

【0047】

図1は、本発明の第一の実施形態に係るコンテンツ部品検索システムの全体構成例を示す。

【0048】

コンテンツ部品検索システムは、一又は複数の計算機で構成することができる。コンテンツ部品検索システムには、例えば、CPU100、磁気ディスク装置101、主メモリ102、フロッピーディスクドライブ(以下、FDD103と呼ぶ)、及びこれらを結ぶバス104が備えられる。コンテンツ部品検索システムは、他の機器とネットワーク105を介して接続することもできる。

【0049】

磁気ディスク装置101は、二次記憶装置の一つであり、コンテンツのソースを含むコンテンツファイル160とコンテンツ構成ツリー161が格納される。FDD103を介してフロッピーディスク106に格納されている情報が、磁気ディスク装置101あるいは主メモリ102へ読み込まれる。

10

20

30

40

50

【0050】

主メモリ102には、システム制御処理部110、部品登録制御処理部111、部品検索制御処理部112、コンテンツ取得処理部121、コンテンツ内情報取得処理部122、コンテンツ構成ツリー生成処理部123、グルーピング処理部130、ラベル付与処理部131、部品登録処理部126、出入力定義情報取得処理部140、出入力定義情報解析処理部141、部品検索処理部142、部品候補出力処理部143が格納され、ワークエリア150が確保される。なお、本実施形態におけるシステム制御、部品登録制御、部品検索制御、コンテンツ取得、コンテンツ内情報取得、コンテンツ構成ツリー生成、グルーピング、ラベル付与、部品登録、出入力定義情報取得、出入力定義情報解析、部品検索および部品候補出力の各処理は、システム制御処理部110、部品登録制御処理部111、部品検索制御処理部112、コンテンツ取得処理部121、コンテンツ内情報取得処理部122、コンテンツ構成ツリー生成処理部123、グルーピング処理部130、ラベル付与処理部131、部品登録処理部126、出入力定義情報取得処理部140、出入力定義情報解析処理部141、部品検索処理部142および部品候補出力処理部143をCPU100で実行することにより実現するが、各処理を行なうプログラムとして集積回路化するなどしてハードウェアで実現することもできる。以下説明を簡略化するため、各種プログラムをCPU100が実行することで実現される各プログラムを各処理の主体として説明する。なお各処理部をハードウェアで実現した場合にはその各処理部が主体となって各処理を行なう。

【0051】

システム制御処理部110は、部品登録制御処理部111および検索制御処理部121の制御を行なう。

【0052】

部品登録制御処理部111は、コンテンツ取得処理部121、コンテンツ内情報取得処理部122、コンテンツ構成ツリー生成処理部123、グルーピング処理部130、ラベル付与処理部131および部品登録処理部126の制御を行なう。

【0053】

コンテンツ構成ツリー生成処理部123は、グルーピング処理部130およびラベル付与処理部131の制御を行なう。

【0054】

部品検索制御処理部112は、出入力定義情報取得処理部140、出入力定義情報解析処理部141、部品検索処理部142および部品候補出力処理部143の制御を行なう。

【0055】

本実施形態では、これらの処理部を主メモリ102に格納するものとしたが、磁気ディスク装置101、フロッピーディスク106、MO、CD-ROM、DVD等の記憶媒体(図1には示していない)に格納し、駆動装置を介して主メモリ102に読み込み、CPU100によって実行することも可能である。

【0056】

また、本実施形態では、コンテンツファイル160およびコンテンツ構成ツリー161は、磁気ディスク装置101に格納されるものとしたが、フロッピーディスク(登録商標)106、MO(Magneto-Optical disk)、CD-ROM、DVD等の記憶媒体(図1には示していない)に格納されるものとしてもよいし、あるいはネットワーク105を介して、他のシステムに接続された記憶装置(図1には示していない)に格納されるものとしてもよい。また、さらにはネットワーク105に直接接続された記憶媒体(図1には示していない)に格納されるものとしても構わない。

【0057】

以下、本実施形態に係るコンテンツ部品検索システムで行われる処理の概要を説明する。

【0058】

システム制御処理部110は、キーボード101(他種の入力装置でも良い)から入力

10

20

30

40

50

されたコマンドを解析する。この結果が、登録実行のコマンドであると解析された場合には、システム制御処理部110が、部品登録制御処理部111を起動し、部品登録制御処理部111が、コンテンツファイル160およびコンテンツ構成ツリー161を磁気ディスク装置101へ登録する。また、検索実行のコマンドであると解析された場合には、システム制御処理部110が、部品検索制御処理部112を起動し、部品検索制御処理部112が、指定されたクエリツリー171(図11参照)に基づいて部品検索の処理を行う。

【0059】

以下、部品検索制御処理部112が行う処理の流れを説明する。

【0060】

まず、部品検索制御処理部112は、入出力定義情報取得処理部140を起動し、入出力定義情報取得処理部140が、新規作成コンテンツの定義情報(以下、「入出力定義情報」と呼ぶ)170(図11参照)を取得し、取得された入出力定義情報170をワークエリア150に格納する。

【0061】

次に、部品検索制御処理部112は、入出力定義情報解析処理部141を起動し、入出力定義情報解析処理部141が、入出力定義情報170から、ユーザ(例えば開発者)によって指定されたクエリツリー171を取得し、取得されたクエリツリー171を、ワークエリア150に格納する。

【0062】

次に、部品検索制御処理部112は、部品検索処理部142を起動し、部品検索処理部142が、各コンテンツファイル160に含まれるすべての部品(以下、既存部品サブツリー174と呼ぶ)に対し、クエリツリー171との類似度を算出し、算出結果を評価結果183(図12参照)としてワークエリア150に格納する。

【0063】

最後に、部品検索制御処理部112は、部品候補出力処理部143を起動し、部品候補出力処理部143が、クエリツリー171に対して類似度の高い既存部品サブツリー174を、クエリツリー171に対する部品候補として出力する。

【0064】

入出力定義情報取得処理部140は、ユーザによって入力された入出力定義情報170をワークエリア150に格納する。入出力定義情報170は、例えば、図11に示すように、XML(Extensible Markup Language)形式で表すことができる。新規作成コンテンツの入力項目の構造情報をXMLの構造で表現し、入力項目のラベル(ラベルとは実質的に入力項目名を意味する)を、XMLのタグ名で表現し、入力コントロールの種類を、各エレメントのtype属性で表現することができる。なお、本実施形態では、入出力定義情報170のデータ形式を、XML形式としているが、入力項目の構造情報、入力項目のラベルおよび入力コントロールの種類を定義できるものならどのようなデータ形式であってもよい。また、入出力定義情報170は、ユーザから入力されるものとしたが、FDD103等のドライブを介してフロッピーディスク106、MO、CD-ROM、DVD等の記憶媒体(図1には示していない)に格納されているもの、あるいはネットワーク105を介して他のシステムに接続された記憶装置(図1には示していない)に格納されているもの、さらにはネットワーク105に直接接続された記憶媒体(図1には示していない)に格納されているものから入力されてもよい。

【0065】

入出力定義情報解析処理部141は、入出力定義情報取得処理部140によって取得された入出力定義情報170から、ユーザによって指定された構造情報(以下、「ユーザ指定構造情報」と呼ぶ)172(図11参照)に対応するクエリツリー171を取得し、取得されたクエリツリー171をワークエリア150に格納する。なお、クエリツリー171は、入出力定義情報170のサブセットである。

【0066】

10

20

30

40

50

部品検索処理部 142 が行う処理の流れを、図 10 及び図 12 を参照して説明する。

【0067】

まず、部品検索処理部 142 は、入出力定義情報解析処理部 141 で取得されたすべてのクエリツリー 171 に対して、ステップ 2010 ~ 2011 を繰り返す（ステップ 2000）。

【0068】

次に、部品検索処理部 142 は、クエリツリー 171 から、部品内情報 177A（ラベル情報 175A および入力コントロール情報 176A を含んだ情報）を取得し、取得された部品内情報 177A を、ワークエリア 150 に格納する（ステップ 2010）。

【0069】

次に、部品検索処理部 142 は、磁気ディスク装置 101 に登録されているすべてのコンテンツファイル 160 に対して、ステップ 2012 を繰り返す（ステップ 2011）。

【0070】

次に、部品検索処理部 142 は、該コンテンツファイル 160 に対応するコンテンツ構成ツリー 161 に含まれるすべての既存部品サブツリー 174 に対して、ステップ 2113 ~ 2015 を繰り返す（ステップ 2012）。

【0071】

次に、部品検索処理部 142 は、該既存部品サブツリー 174 の部品内情報 177B（ラベル情報 175B と入力コントロール情報 176B を含んだ情報）を取得し、取得された部品内情報 177B を、ワークエリア 150 に格納する（ステップ 2013）。

【0072】

次に、部品検索処理部 142 は、クエリツリー 171 の部品内情報 177A と既存部品サブツリー 174 の部品内情報 177B から、類似度算出のためのベクトル空間とベクトルを表現したテーブル形式のベクトル情報 182（ラベル情報 178、入力コントロール情報 179、クエリツリー・ベクトル情報 180 および既存部品サブツリー・ベクトル情報 181 を含んだ情報）を生成し、生成されたベクトル情報 182 を、ワークエリア 150 に格納する（ステップ 2014）。

【0073】

最後に、部品検索処理部 142 は、ベクトル情報 182 を用いて、該クエリツリー 171 に対する該既存部品サブツリー 174 の類似度を（式 1）により算出し（ステップ 2015）、その類似度算出結果を、該クエリツリー 171 に対する各既存部品サブツリー 174 の評価結果 183（コンテンツ ID 情報 1831、ノード番号情報 1832 および類似度情報 1833 を含んだ情報）を、ワークエリア 150 に格納する。類似度は、ベクトル情報 182 からベクトル空間法による余弦尺度により算出される。

【0074】

【数 1】

式1

クエリツリーと既存部品サブツリーの類似度

$$= \frac{\text{クエリツリー・ベクトルと既存部品サブツリー・ベクトルの内積}}{|\text{クエリツリー・ベクトル}| \cdot |\text{既存部品サブツリー・ベクトル}|}$$

40

【0075】

なお、ステップ 2015 では類似度の算出式におけるベクトル空間を、類似度を算出するクエリツリーと既存部品サブツリーのみが含まれる空間（次元）で計算するものとしたが、磁気ディスク装置 101 に含まれるすべてのコンテンツ 160 およびクエリツリーに関する共通空間（次元）で計算するものとしてもよい。

【0076】

まず、部品候補出力処理部 143 は、部品検索処理部 142 によって算出された評価結

50

果 183 から、各クエリツリー 171 に対して、類似度の降順で既存部品サブツリー 174 に対応する部品の候補に関する情報の一覧（以下、「部品候補一覧」と呼ぶ）を部品候補一覧表示画面 187（図 13 参照）に表示する。なお、部品候補一覧には、各部品候補について、部品 ID（例えば、“コンテンツ ID_コンテンツ構成ツリーのノード番号”で表現される部品の識別子）、類似度、および既存部品サブツリーのルートノードに格納されているラベルが含まれる。

【0077】

次に、部品候補出力処理部 143 は、部品候補一覧よりユーザによって選択された部品候補に対して、コンテンツファイル 160 から、選択された部品に対応するコードを取得し、取得したコードを、ワークエリア 150 に格納する。

10

【0078】

最後に、部品候補出力処理部 143 は、上記で取得されたコードを用いて、ユーザによって選択された部品を、プレビュー画面 184（図 13 参照）に表示する。

【0079】

以上が、部品候補出力処理部 143 の処理手順である。

【0080】

なお、部品候補一覧よりユーザによって部品候補が選択されることにより該部品候補がプレビュー画面 184 に表示されるものとしたが、部品候補出力処理部 143 は、ユーザ操作を介さずにクエリツリー 171 に対して類似度が最も高い既存部品サブツリーに対応する部品をプレビュー画面 184 に表示してもよい。

20

【0081】

また、ユーザによって指定されたクエリツリーが複数個存在する場合は、クエリツリーの情報（以下、「クエリツリー情報」と呼ぶ）として、複数のクエリツリー間の関係をクエリツリー情報画面 186（図 13 参照）に表示してもよい。

【0082】

また、部品候補一覧において、或る閾値以上の類似度を持たない既存部品サブツリーについては、部品候補として含まれなくてよい。この際、クエリツリー情報表示画面 186 に、各クエリツリーに対して検索された部品数（既存部品サブツリーの数）が表示されてもよい。

【0083】

30

また、ユーザによって指定された複数のクエリツリー間に階層関係（例えば親子関係）が存在する場合（入出力定義情報 170 に含まれるすべてのサブツリーをクエリツリー 171 とした場合も含む）、上位（例えば親）のクエリツリーの部品候補から優先的に部品候補一覧表示画面 187 に表示されてもよい。優先的に表示するとは、例えば、上位のクエリツリーの部品候補の類似度がそれよりも下位のクエリツリーの部品候補の類似度よりも低くても、上位のクエリツリーの部品候補が、下位のクエリツリーの部品候補よりも、降順で上位に位置することを言う。

【0084】

また、現在表示されている部品候補一覧に対するクエリツリー 171 の内容が判断しやすいように、部品候補出力処理部 143 が、入出力定義情報表示画面 185 に表示される入出力定義情報において、クエリツリー 171 の部分が強調表示（例えば網掛け表示）してもよい。

40

【0085】

以上が、部品検索制御処理部 112 が行う処理についての説明である。

【0086】

次に、部品登録制御処理部 111 が行う処理について説明する。

【0087】

まず、部品登録制御処理部 111 は、コンテンツ取得処理部 121 を起動し、コンテンツ取得処理部 121 が、FDD104 を介してフロッピーディスク 108 に格納されているコンテンツファイル 160 を読み込む。

50

【0088】

次に、部品登録制御処理部111は、コンテンツ内情報取得処理部122を起動し、コンテンツ内情報取得処理部122が、コンテンツファイル160のソース中に含まれるテキストや入力コントロール（例えばテキストボックスやラジオボタン等）と、これらのコンテンツ内要素の位置情報やサイズ（例えば、テキストの場合はフォントサイズ、入力コントロールの場合は領域サイズ）を取得する。

【0089】

次に、部品登録制御処理部111は、コンテンツ構成ツリー生成処理部123を起動し、コンテンツ構成ツリー生成処理部123が、コンテンツ内情報取得処理部122によって抽出されたコンテンツ内の要素情報を用いて、要素間のグルーピング、および要素やグループに対するラベルを付与することで、コンテンツ構成ツリー161を生成する。10

【0090】

最後に、部品登録制御処理部111は、部品登録処理部126を起動し、部品登録処理部126が、コンテンツファイル160とコンテンツファイル160に対応するコンテンツ構成ツリー161を磁気ディスク装置101へ登録する。

【0091】

コンテンツ取得処理部121は、FDD104を介してフロッピーディスク108に格納されているコンテンツファイル160を読み込み、ワークエリア150に格納する。なお、コンテンツファイル160はフロッピーディスク108に格納されているものとしたが、MO、CD-ROM、DVD等の記憶媒体（図1には示していない）に格納されるものとしてもよいし、ネットワーク105を介して、他のシステムに接続された記憶媒体（図1には示していない）に格納されるものとしてもよい。20

【0092】

コンテンツ内情報取得処理部122は、まず、コンテンツ取得処理部121によって取得されたコンテンツファイル160の中から、磁気ディスク装置101や主メモリ102に格納されたルールに基づいて、テキスト情報要素や入力コントロール種類情報要素（以下、これらを「コンテンツ内情報」と呼ぶことがある）を抽出する。抽出された全てのテキスト情報要素が記録されたテキストリスト情報162と、抽出された全ての入力コントロール種類情報要素が記録された入力コントロールリスト情報163とが、ワークエリア150に格納される。ここでいう「テキスト」とは、コンテンツファイルが例えばHTMLファイルの場合、画面上に表示されないタグなどではなく、実際に画面上で表示される文字列のことをいう。また、ここでいう「入力コントロール」とは、コンテンツファイルが例えばHTMLファイルの場合、INPUTタグ、SELECTタグおよびTEXTAREAタグで表現されたものなど、ユーザからの指定を受け付ける要素をいう。30

【0093】

なお、テキストリスト情報162や入力コントロールリスト情報163は、ワークエリア150に格納されるが、磁気ディスク装置101に格納されてもよい。コンテンツファイルの中のどのようなものをテキストとして抽出するかを示す情報や、コンテンツファイルの中のどのようなものを入力コントロールとして抽出するかを示す情報は、主メモリ102または磁気ディスク装置101に格納されている。40

【0094】

次に、コンテンツ内情報取得処理部122は、抽出されたコンテンツ内要素の配置位置情報をコンテンツファイルの中から抽出する。ここで、配置位置情報は、コンテンツ内要素がコンテンツ内での位置と占める範囲の定義を表す情報である。具体的には、例えば、テキストの配置位置情報には、基準位置情報とフォントサイズ情報とが含まれ、入力コントロールの配置位置情報には、基準位置情報と領域サイズ情報とが含まれる。テキストの基準位置情報やフォントサイズ情報は、テキストリスト情報162に記録され、入力コントロールの基準位置情報や領域サイズ情報は、入力コントロールリスト情報163に記録される。なお、テキストリスト情報162は、図5に例示するように、例えばテーブルであり、コンテンツファイルから検出されたテキスト毎に、テキスト番号700（例えば通

50

し番号)、テキスト情報要素710、基準位置情報720(例えばコンテンツの所定位置を原点とした場合の基準位置の座標)、およびフォントサイズ730が記録される。また、入力コントロールリスト情報163は、図6に例示するように、例えばテーブルであり、コンテンツファイルから検出された入力コントロール毎に、入力コントロール番号800(例えば通し番号)、入力コントロール種類情報要素810(例えばタグ付きの情報)、基準位置情報820、および領域サイズ830(例えば、基準位置を中心とした幅及び高さのサイズ)が記録される。

【0095】

コンテンツ構成ツリー生成処理部123は、まず、グルーピング処理部130を起動し、グルーピング処理部130が、コンテンツ内情報取得処理部122によって抽出された入力コントロールリスト情報163を用いて、コンテンツファイル160に含まれる入力コントロールのグルーピングを行ない、そのグルーピング結果(例えば図8参照)をワークエリア150に格納する。10

【0096】

次に、コンテンツ構成ツリー生成処理部123は、ラベル付与処理部131を起動し、ラベル付与処理部131が、グルーピング処理部130によって生成されたグルーピング結果に対して、テキストリスト情報162や入力コントロールリスト情報163に格納されている配置位置情報を用いて、各入力コントロールの名前やグルーピングによって生成されたグループの名前(以下、まとめてラベルと呼ぶ)を付与することで、最終的なコンテンツ構成ツリー161を生成し、その生成されたコンテンツ構成ツリー161をワークエリア150に格納する。20

【0097】

なお、コンテンツ構成ツリー161は、各ノードに、ノード属性情報として、ノード番号、ラベル、基準点、親ノード番号(直近の上位のノード番号)およびコントロール種類を持つ。コンテンツ構成ツリーの概念図と各ノードの属性情報の格納例を図9に示す。図9には、コンテンツ構成ツリーの上位部分が示されており、下位部分は図示が省略されている。コンテンツ構成ツリー161は、ワークエリア150に格納されるが、磁気ディスク装置101に格納されてもよい。

【0098】

部品登録処理部126は、まず、各コンテンツファイル160にコンテンツIDを付与する。コンテンツIDは、主メモリ102または磁気ディスク装置101に格納されているルールに基づいて生成され付与される。30

【0099】

次に、部品登録処理部126は、同コンテンツにおけるコンテンツファイル160とコンテンツ構成ツリー161を関連付けた上で、それぞれを磁気ディスク装置101へ登録する。

【0100】

図2は、グルーピング処理部130が行う処理の流れの一例を示す。

【0101】

グルーピング処理部130は、まず、コンテンツ内情報取得処理部122によって抽出された入力コントロールリスト情報163に含まれるN個の入力コントロールに対して、ステップ301～ステップ302を、 $n = 1 \sim (N - 1)$ まで繰り返し実行する(ステップ300)。40

【0102】

次に、グルーピング処理部130は、磁気ディスク装置101又は主メモリ102に格納された値Mに基づいて、M個の階層レベルに対して、ステップ302を、 $m = 2 \sim (M - 1)$ まで繰り返し実行する(ステップ301)。なお、階層レベルmとは、階層レベル1がコンテンツ全体を表し、mが大きいほど詳細な内容に関するグルーピングを行ない、最下位階層レベル(つまり $m = M - 1$)では、1グループ1入力コントロールとなる。

【0103】

次に、グルーピング処理部 130 は、階層レベル m におけるグルーピング条件を、階層レベル m に關し、n 番目と n + 1 番目の入力コントロール間が同じグループに屬するか否かの判定を行ない、判定結果をワークエリア 150 に格納する（ステップ 302）。なお、階層レベル m において、n 番目と n + 1 番目の入力コントロール間が同じグループと判定され、さらに n + 1 番目と n + 2 番目の入力コントロール間が同じグループと判定された場合、n 番目～n + 2 番目の入力コントロールは階層レベル m において同じグループとみなす。すなわち、同じグループと判定された入力コントロールペアが複数個存在し、入力コントロールペアにおける一方の入力コントロールが別の入力コントロールペアにおける他方の入力コントロールとなっている場合、複数の入力コントロールペアにおける入力コントロールが同一のグループとみなされる。なお、n 番目の入力コントロールとは、コンテンツ中で n 番目に入力される入力コントロール（例えば、n は図 6 の入力コントロール番号 800）を表す。10

【0104】

最後に、グルーピング処理部 130 は、ステップ 302 によって階層レベル別にグルーピングされた結果をツリー構造で管理するために、コンテンツ構成ツリー 161 における各ノードに属性情報として親ノード番号を格納する（ステップ 303）。なお、コンテンツ構成ツリー 161 は、各ノードの属性情報として、ノード番号、ラベル、コンテンツ内で該当する領域（入力コントロールが占める範囲）の基準点、親ノードのノード番号、および、入力コントロール種類（コントロール種類）を有するが、グルーピング処理部 130 の処理完了時点では、ラベルおよび基準点は格納されない。20

【0105】

ここで、ステップ 302 で行なわれる n 番目と n + 1 番目の入力コントロール間のグルーピング方法について詳しく説明する。

【0106】

n 番目と n + 1 番目の入力コントロール間のグルーピング方法については、n 番目と n + 1 番目の入力コントロール間の配置位置関係によって、以下の 8 つのケースで判定ルール（以下、グルーピング条件と呼ぶ）を決めて判定を行なう。なお、以下の説明で出てくる基準点は、入力コントロールの左上を表す。また、最右端、最左端、最上端および最下端は、基準点と領域サイズ 830 から定まる、入力コントロールの最も端の点を表す。30

【0107】

まず、n 番目の入力コントロールの基準点に対して n + 1 番目の入力コントロールの基準点が真右に位置する場合（以下、第 1 のケースと呼ぶ）、n 番目の入力コントロールの最右端と n + 1 番目の入力コントロールの最左端との水平距離が、第 1 のケースにおける階層レベル m でのグルーピング条件を満たすとき、n 番目と n + 1 番目の入力コントロールは、階層レベル m で同じグループに屬するとみなされる。

【0108】

n 番目の入力コントロールの基準点に対して n + 1 番目の入力コントロールの基準点が真下に位置する場合（以下、第 2 のケースと呼ぶ）、n 番目の入力コントロールの最下端と n + 1 番目の入力コントロールの最上端との垂直距離が、第 2 のケースにおける階層レベル m でのグルーピング条件を満たすとき、n 番目と n + 1 番目の入力コントロールは、階層レベル m で同じグループに属するとみなされ。40

【0109】

n 番目の入力コントロールの基準点に対して n + 1 番目の入力コントロールの基準点が右下に位置する場合（以下、第 3 のケースと呼ぶ）、n 番目の入力コントロールの最右端と n + 1 番目の入力コントロールの最左端との水平距離が第 3 のケースにおける階層レベル m でのグルーピング条件を満たし、かつ n 番目の入力コントロールの最下端と n + 1 番目の入力コントロールの最上端との垂直距離が第 3 のケースにおける階層レベル m でのグルーピング条件を満たすとき、n 番目と n + 1 番目の入力コントロールは、階層レベル m で同じグループに属するとみなされる。

【0110】

10

20

30

40

50

n 番目の入力コントロールの基準点に対して $n + 1$ 番目の入力コントロールの基準点が右上に位置する場合（以下、第 4 のケースと呼ぶ）、 n 番目の入力コントロールの最右端と $n + 1$ 番目の入力コントロールの最左端との水平距離が第 4 のケースにおける階層レベル m でのグルーピング条件を満たすとき、かつ n 番目の入力コントロールの最上端と $n + 1$ 番目の入力コントロールの最下端との垂直距離が第 4 のケースにおける階層レベル m でのグルーピング条件を満たすとき、 n 番目と $n + 1$ 番目の入力コントロールは、階層レベル m で同じグループに属するとみなされる。

【 0 1 1 1 】

n 番目の入力コントロールの基準点に対して $n + 1$ 番目の入力コントロールの基準点が左下に位置する場合（以下、第 5 のケースと呼ぶ）、 n 番目の入力コントロールの最下端と $n + 1$ 番目の入力コントロールの最上端との垂直距離が第 5 のケースにおける階層レベル m でのグルーピング条件を満たすとき、 n 番目と $n + 1$ 番目の入力コントロールは、階層レベル m で同じグループに属するとみなされる。 10

【 0 1 1 2 】

その他、 n 番目の入力コントロールの基準点に対して $n + 1$ 番目の入力コントロールの基準点が真左に位置する場合（以下、第 6 のケースと呼ぶ）、 n 番目の入力コントロールの基準点に対して $n + 1$ 番目の入力コントロールの基準点が真上に位置する場合（以下、第 7 のケースと呼ぶ）、 n 番目の入力コントロールの基準点に対して $n + 1$ 番目の入力コントロールの基準点が左上に位置する場合（以下、第 8 のケースと呼ぶ）、 n 番目と $n + 1$ 番目の入力コントロールは、階層レベル 1 を除くすべての階層レベルにおいて同じグループに属さないとみなされる。 20

【 0 1 1 3 】

図 7 は、グルーピング条件 930 を格納するグルーピング条件リスト情報 900 の例を示す図である。グルーピング条件情報 900 は、例えばテーブルであり、各ケースについて、ケース番号 910、 n 番目の入力コントロールに対する $n + 1$ 番目の入力コントロールの配置位置関係を表す情報（以下、「位置関係情報」と呼ぶ）920、およびグルーピング条件 930 を有する。たとえば、第 1 のケースの場合、ケース番号 910 には“1”、位置関係情報 920 には“真右”、グルーピング条件 930 には“ n 番目の最右端と $n + 1$ 番目の最左端との距離が 100 p × 未満”、といった情報を格納する。 30

【 0 1 1 4 】

グルーピング条件 920 は、階層レベル毎に設定される。このため、例えば、グルーピング条件リスト情報 900 が、一の階層レベルにつき一つである場合、各階層レベルに対応した各グルーピング条件リスト情報 900 が用意される。グルーピング条件リスト情報 900 は、例えば、主メモリ 102 または磁気ディスク装置 101 に格納されている。 40

【 0 1 1 5 】

なお、図 7 で示したグループピング条件 930 は、例えば、一般的な入力フォーム（コンテンツ）に基づいて、 n 番目と $n + 1$ 番目の入力コントロール間の相対距離に関する閾値が設定されたものとすることができる。しかし、それに限らず、例えば、コンテンツファイル 160 に対して実際に n 番目と $n + 1$ 番目の入力コントロール間の相対距離を抽出した上で、その平均値に基づいて動的にグルーピング条件 930 中の相対距離の閾値が設定されてもよい。具体的には、たとえば、第 2 のケースの場合、グルーピング条件として、“ n 番目の最下端と $n + 1$ 番目の最上端との垂直距離が、[垂直方向に関して、隣接する入力コントロール間の平均間隔] 以内” としてもよい。このような動的な閾値設定（条件設定）により、最適なグルーピング条件の設定が容易になることや、コンテンツ間の作成基準における開発者の個人差を吸収することができる。 40

【 0 1 1 6 】

また、ステップ 302 では、入力コントロールの位置情報を用いてグルーピングを行なったが、コンテンツ内情報取得処理部 122 によって生成されたテキストリスト情報 162 のフォントサイズ情報を用いて、ある閾値以上のフォントサイズを持つテキストの位置を階層レベル m における n 番目と $n + 1$ 番目の入力コントロール間をグループの境界をみ 50

なすといった方法をとってもよい。

【0117】

また、コンテンツ内情報取得処理部122によって生成されたテキストリスト情報162のH1などの文書構造情報を用いて、階層レベルmにおけるn番目とn+1番目の入力コントロール間をグループの境界をみなすといった方法をとってもよい。これらフォントサイズや文書構造情報と、位置情報を併用してグルーピング処理を行なうことで、グルーピングの精度が高くなる。

【0118】

図3は、ラベル付与処理部131が行う処理の流れの一例を示す。

【0119】

ラベル付与処理部131は、まず、コンテンツ構成ツリー161のM個の階層レベルに対し、ステップ410～ステップ411および415をm=M～1まで繰り返す(ステップ400)。

【0120】

次に、ラベル付与処理部131は、階層レベルが最下位階層レベルであるかどうか(m=Mであるかどうか)を判定する(ステップ410)。

【0121】

ステップ410の判定の結果、階層レベルが最下位階層レベル(m=M)である場合、ラベル付与処理部131は、コンテンツ全体に含まれるN個の入力コントロールに対し、ステップ412～ステップ414をn=1～Nまで繰り返し実行する(ステップ411)。

【0122】

次に、ラベル付与処理部131は、n番目の入力コントロールの基準点を領域の左上として設定し、ワークエリア150に格納する(ステップ412)。

【0123】

次に、ラベル付与処理部131は、n番目の入力コントロールの基準点に対し、テキストリスト情報162に格納されているコンテンツ中のテキストの中から上方向と左方向(距離が同じ場合は左より上優先)で最も近い距離にあるテキストを、n番目の入力コントロールのラベルとして抽出し、そのラベルをコンテンツ構成ツリー161における階層レベルMのn番目のノードの属性情報として、ワークエリア150に格納する(ステップ413)。

【0124】

次に、ラベル付与処理部131は、ラベルを含めたn番目の入力コントロールの基準点を、ラベル領域の左上とし、その基準点をコンテンツ構成ツリー161における階層レベルMのn番目のノードの属性情報として、ワークエリア150に格納する(ステップ414)。なお、ステップ413でラベルが抽出できなかった場合には、ラベル付与処理部131は、ステップ412で設定した入力コントロールの左上を基準点とする。

【0125】

次に、ラベル付与処理部131は、ステップ410の判定の結果、階層レベルが最下位階層レベルでない場合、階層レベルmに含まれるN個のグループに対し、ステップ415～ステップ418をn=1～Nまで繰り返し実行する(ステップ415)。

【0126】

次に、ラベル付与処理部131は、n番目のグループに含まれる要素(子ノード)の中で、最も左上(例えば、距離が同じ場合は左より上優先)の基準点をもつ要素の基準点をn番目のグループの基準点として、ワークエリア150に格納する(ステップ416)。

【0127】

次に、ラベル付与処理部131は、ステップ416で設定されたn番目のグループの基準点に対し、テキストリスト情報162に格納されているコンテンツ中のテキストの中から上方向と左方向(距離が同じ場合は左より上優先)で最も近い距離にあるテキストであって、未だラベルとして抽出されていないテキストを、n番目のグループのラベルとして

10

20

30

40

50

抽出し、そのラベルをコンテンツ構成ツリー 161 における該ノードの属性情報として、ワークエリア 150 に格納する（ステップ 417）。

【0128】

なお、未だラベルとして抽出されていないテキストが n 番目のグループではないグループに属してしまう場合や、未だラベルとして抽出されていないテキストが無い場合には、ラベル付与処理部 131 は、n 番目のグループの基準点のノードのラベルを、n 番目のグループのラベルとする。

【0129】

次に、ラベル付与処理部 131 は、ラベルを含めた n 番目のグループの基準位置を、ラベル領域の左上とし、その基準位置をコンテンツ構成ツリー 161 における該ノードの属性情報として、ワークエリア 150 に格納する（ステップ 418）。

10

【0130】

ラベル付与処理部 131 は、グルーピング結果を表すノードツリーのリーフからルートにかけてボトムアップ式に順次にラベルを付与していくことができる。

【0131】

なお、ステップ 413 では、n 番目の入力コントロールのラベルをテキストリスト情報 162 に格納されているコンテンツ中のテキストの中から、入力コントロールの基準点に対して上方向と左方向（距離が同じ場合は左より上優先）で最も近い距離にあるテキストとしたが、ラジオボタンやチェックボックスについては入力コントロールの基準点に対して右方向と下方向（距離が同じ場合は下より右優先）で最も近い距離にあるテキストとしてもよい。

20

【0132】

また、隣接する入力コントロール間が間隔なく配置されているグループについては、n 行 × m 列のテーブル構造とみなし、1 行目の各列に付与されたラベルを、n 行目以降の各列にも同じラベル付与する。例えば、1 行目の 1 列目の入力コントロールに付与されたラベルは、2 行目の 1 列目の入力コントロールに付与されるラベルとしても使われる。

【0133】

また、本実施形態におけるグループのラベル抽出方法は、グループ（ノード）の基準位置に対し、テキストリスト情報 162 に格納されているコンテンツ中のテキストの中から上方向と左方向で最も近い距離にあるテキストをラベルとして抽出するものとしたが、距離に関する閾値を抽出条件として設定してもよい。これにより、誤ったラベル付与が少なくなる。

30

【0134】

また、本実施形態におけるグループのラベル抽出方法は、グループ（ノード）の基準点に対し、テキストリスト情報 162 に格納されているコンテンツ中のテキストの中から上方向と左方向で最も近い距離にあるテキストをラベルとして抽出するものとしたが、グループの中に含まれる要素（子ノード）のラベルの中から代表値として抽出してもよいし、他のコンテンツによるトレーニングデータによって、トレーニングデータ中でラベル付与したいグループと類似するグループに付与されたラベルを用いてもよい。これにより、コンテンツ中にラベルに該当するテキストが存在しない場合でも、各コンテンツ内要素やグループにラベルを付与することができる。

40

【0135】

以上が、部品登録制御処理部 111 によって行われる処理の説明である。

【0136】

以下、部品検索制御処理部 112 及び部品登録制御処理部 111 によって行われる処理について、更に詳細に説明する。まず、部品検索制御処理部 112 の具体的な処理を説明する。

【0137】

図 11 は、入出力定義情報解析処理部 141 の具体的な処理を説明した図である。

【0138】

50

入出力定義情報解析処理部 141 は、入出力定義情報取得処理部 140 によって取得された入出力定義情報 170 から、ユーザによって指定されたユーザ指定構造情報 172 に対応するサブツリーを抽出し、クエリツリー 171 を取得する。本図の例では、ユーザ指定構造情報 172 として “Applicant Information” が指定されており、入出力定義情報 170 から “Applicant Information” の部分のサブツリーがクエリツリー 171 として取得されている。

【0139】

図 12 は、部品検索処理部 142 の具体的な処理を説明した図である。

【0140】

部品検索処理部 142 は、磁気ディスク装置 101 に登録されているすべてのコンテンツファイル 160 に含まれるすべての既存部品サブツリー 174 に対して、1つのクエリツリー 171 との類似度を算出する。なお、既存部品サブツリー 174 は、コンテンツ構成ツリー 161 のサブセットであり、最小単位として、コンテンツ構成ツリー 161 のリーフであることがある。

10

【0141】

図 12 の例では、“Applicant Information” の部分のクエリツリー 171 と、図 4 に示すコンテンツ 500 (コンテンツファイル 160 が示すコンテンツ) におけるコンテンツ構成ツリー 161 の既存部品サブツリー 174 (図 9 で示すコンテンツ構成ツリー 1100 のノード番号 2 のサブツリー) の類似度算出処理の詳細な具体例を示している。

【0142】

20

まず、部品検索処理部 142 は、クエリツリー 171 から部品内情報 177A (ラベル情報 175A と入力コントロール情報 174A を含んだ情報) を取得し、同様に、既存部品サブツリー 174 から、部品内情報 177B (ラベル情報 175B と入力コントロール情報 174B を含んだ情報) を取得する。本図の例では、クエリツリー 171 からは、7 個のラベル情報要素 (テキスト情報要素) および入力コントロール種類情報要素が取得され、既存部品サブツリー 174 からは、8 個のラベル情報要素および入力コントロール種類情報要素が取得される。

【0143】

次に、部品検索処理部 142 は、各部品内情報 177 から、類似度算出に必要なベクトル空間とベクトルを生成するために、ベクトル情報 182 として、ラベル情報 178、入力コントロール情報 179、クエリツリーベクトル 180 および既存部品サブツリーベクトル 181 を生成する。本図の例では、ベクトル空間を表す次元として、言い換えれば、ラベル情報 178 と入力コントロール情報 179 として、部品内情報 177A 及び 177B を用いて、(Name · text, Address · text, Telephone Number · text, Company Name · text, Year · select, Month · select, Day · select, First · text, Last · text, Email Address · text) が生成される (上記カッコ内では、ラベル情報要素・入力コントロール種類情報要素、となっている)。また、生成されたベクトル空間におけるクエリツリーベクトル 180 として (1, 1, 1, 1, 1, 1, 0, 0, 0)、既存部品サブツリーベクトル 181 として (0, 1, 1, 0, 1, 1, 1, 1, 1) が生成されている。各ベクトルの要素については、例えば、ラベル “Name” と入力コントロール種類 “text” の組合せが存在する場合は “1”、その組合せが存在しない場合は “0” となる。

30

【0144】

40

次に、部品検索処理部 142 は、生成されたベクトル情報 182 から、(式 1) によりクエリツリー 171 に対する既存部品サブツリー 174 の類似度を算出し、コンテンツ ID 情報 1831、ノード番号情報 1832 及び類似度情報 1833 を含んだ評価結果 183 を生成し、ワークエリア 150 に格納する。

【0145】

本図の例では、クエリツリー 171 に対する既存部品サブツリー 174 の類似度として “0.67” が算出され、該算出結果は、評価結果 183 のコンテンツ ID 情報 1831 として “0001”、ノード番号情報 1832 として “2”、類似度情報 1833 として

50

“0.67”が格納される。

【0146】

図13は、部品候補出力処理部143の具体的な処理を説明した図である。

【0147】

部品候補出力処理部143は、部品検索処理部143によって生成された類似度算出結果（評価結果183）から、部品候補一覧表示187を出力する。本図の例では、コンテンツID0001～0003のコンテンツに含まれる部品に対して、類似度の降順でクエリツリー171に対する部品候補の一覧が部品候補一覧表示画面187に表示される。その一覧によれば、第1位の部品候補は、部品IDが“0001_2”、類似度が“0.67”、ラベルが“Applicant Information”的部品である。

10

【0148】

また、本図の例では、部品のプレビューをするためのプレビューボタンが部品候補一覧表示画面187に各部品について表示される。ユーザが第1位の部品候補（部品ID：0001_2）のプレビューボタンを押下することで、部品候補出力処理部143は、第1位の部品候補に対応したコードを該部品候補を含んだコンテンツのコンテンツファイルから取得し、取得したコードに基づき、プレビュー画面184に、第1位の部品候補の部品イメージを表示する。

【0149】

また、本図の例では、入出力定義情報表示画面185が表示され、入出力定義情報表示画面185で、現在出力されている部品候補一覧に対応する入出力定義情報170中のクエリツリー171部分が、網掛けで示されている。

20

【0150】

また、本図の例では、複数のクエリツリー171として、“Applicant Information”、“Date of Birth”および“Service”が指定されており、それらのクエリツリー間の関係（例えば階層関係）が、クエリツリー情報表示画面186に表示される。クエリツリー情報表示画面186から、クエリツリー“Applicant Information”を親としクエリツリー“Date of Birth”を子とした親子関係があることがわかる。また、これら3つのクエリツリー171の中で、部品候補が存在しつつ最上位階層レベルであるクエリツリー“Applicant Information”に対する部品候補が、部品候補一覧表示187に優先的に表示される。

30

【0151】

また、本図の例では、類似度の閾値が“0.50”で設定されており、部品候補一覧表示画面187には、類似度が0.50以上の4件の部品が部品候補として表示されている。すなわち、部品候補出力処理部143は、類似度の閾値が“0.50”未満の部品を部品候補として表示しない。なお、類似度の閾値に代えて又は加えて、類似度が高い部品から上位所定件数の部品のみが表示されても良い。

【0152】

また、クエリツリー情報表示画面186に、各クエリツリー171に対する部品候補数が表示される。その画面186からは、クエリツリー“Applicant Information”に対しては4件の部品候補が存在することがわかる。また、クエリツリー“Service”に対しては、0件、すなわち部品候補がないことがわかる。

40

【0153】

また、図13によれば、入出力定義情報表示画面185、プレビュー画面184、クエリツリー情報表示画面186及び部品候補一覧表示画面187の組合せにより一つの表示画面が構成されているが、それらの画面は互いに離れていても重なっても良い。

【0154】

以上が、部品検索御処理部112の具体的な処理である。開発者は、部品候補一覧の中から所望の部品候補を選択し、その部品候補をそのまま或いは編集することで、新規のコンテンツを作成することができる。開発者は、その新規のコンテンツを表すコンテンツファイルを、磁気ディスク装置101に登録することができる。

50

【0155】

次に、部品登録制御処理部111の具体的な処理を説明する。

【0156】

コンテンツ取得処理部121とコンテンツ内情報取得処理部122の具体的な処理例を図4、図5および図6を用いて説明する。

【0157】

図4は、コンテンツ500の一例として、クライアント端末からアクセスされたHTMLコンテンツのカンファレンス申し込みフォームの例を示す。

【0158】

まず、コンテンツ内情報取得処理部122は、コンテンツ取得処理部121によって取得されたコンテンツ500のHTMLソース（コンテンツファイル160）から、テキスト情報要素およびそれらテキストの配置位置情報（例えば基準位置情報やフォントサイズ）を抽出し、テキスト番号700、テキスト情報710、基準位置情報720およびフォントサイズ730をテキストリスト情報162に記録する。テキストリスト情報162の具体例を図5に示す。図5の例では、テキストリスト情報162に、全部で13個のテキスト情報要素が記録されている（例えば、1番目のテキスト情報要素710として、“Conference Form”が記録され、基準位置情報720として“X：177px Y：35px”が記録され、フォントサイズ730として“+4”が記録される）。

【0159】

次に、コンテンツ内情報取得処理部122は、コンテンツ取得処理部121によって取得されたコンテンツ500のHTMLソースから、INPUTタグとSELECTタグ（入力コントロールの種類を表す情報要素）、およびそれら入力コントロールの配置位置情報（例えば基準位置情報及び領域サイズ）を抽出し、入力コントロール番号800、入力コントロール種類情報要素810、基準位置情報820および領域サイズ830を入力コントロールリスト情報163へ格納する。入力コントロールリスト情報163の具体例を図6に示す。図6の例では、入力コントロールリスト情報163に、全部で15個の入力コントロール種類情報要素が記録される（例えば、1番目の入力コントロール種類情報要素810として、“<INPUT type=“text”>”が記録され、基準位置情報820として“X：315px Y：161px”が記録され、領域サイズ830として“width：73px height：20px”が記録される）。

【0160】

図4に示すコンテンツ500に対し、5つの階層レベルを持つコンテンツ構成ツリーを生成する場合のコンテンツ構成ツリー生成処理部123の具体的な処理の流れを、図5～図9を用いて説明する。

【0161】

グルーピング処理部130は、図6に示す入力コントロールリスト情報163を読み込み、各階層レベル（2～M-1）に対して、入力順が隣り合う入力コントロール間が同じグループに属するか否かを、主メモリ102または磁気ディスク装置101に格納されたグルーピング条件リスト900に基づいて判定する。

【0162】

図7は、階層レベル2におけるグルーピング条件リスト900の例である。図5および図6の例では、階層レベル2のグルーピング処理に関して、まず、図6に示す入力コントロールリスト情報163より、1番目の入力コントロールに対する2番目の入力コントロールの位置関係は“真右”であることから、図7に示すグルーピング条件リスト900における第1のケースが用いられる。

【0163】

そして、入力コントロールリスト情報163に示されている基準位置情報と領域サイズにより、1番目の入力コントロールの最右端と2番目の入力コントロールの最左端との水平距離が52pxであることから、1番目の入力コントロールと2番目の入力コントロールの配置位置関係は、グルーピング条件900における第1のケースの条件である“n番目

10

20

30

40

50

の入力コントロールの最右端と $n + 1$ 番目の入力コントロールの最左端との水平距離が $100 p \times$ 未満”を満たす。この結果、1番目の入力コントロールと2番目の入力コントロールは階層レベル2において同じグループとみなされる。

【0164】

次に、入力コントロールリスト情報163により、2番目の入力コントロールに対する3番目の入力コントロールの位置関係は“左下”であることから、グルーピング条件900における第5のケースが用いられる。そして、入力コントロールリスト情報163に示されている基準位置情報と領域サイズより、2番目の入力コントロールの最下端と3番目の入力コントロールの最上端との垂直距離が $15 p \times$ であることから、2番目の入力コントロールと3番目の入力コントロールの配置位置関係は、グルーピング条件900における第5のケースの条件である“ n 番目の入力コントロールの最下端と $n + 1$ 番目の入力コントロールの最上端との垂直距離が $20 p \times$ 未満”を満たす。この結果、2番目の入力コントロールと3番目の入力コントロールは階層レベル2において同じグループとみなされる。
。

【0165】

次に、入力コントロールリスト情報163により、9番目の入力コントロールに対する10番目の入力コントロールの位置関係は“左下”であることから、グルーピング条件900における第5のケースが用いられる。そして、入力コントロールリスト情報163に示されている基準点と領域サイズの情報より、9番目の入力コントロールの最下端と10番目の入力コントロールの最上端との垂直距離が $108 p \times$ であることから、9番目の入力コントロールと10番目の入力コントロールの配置位置関係は、グルーピング条件900における第5のケースの条件である“ n 番目の入力コントロールの最下端と $n + 1$ 番目の入力コントロールの最上端との垂直距離が $20 p \times$ 未満”を満たさない。この結果、9番目の入力コントロールと10番目の入力コントロール間は、階層レベル2において異なるグループ、すなわちグループの境界とみなされる。

【0166】

この結果、1番目から9番目までの入力コントロールが、階層レベル2における1つのグループとみなされ、10番目の入力コントロールが次のグループの先頭とされる。

【0167】

このように、入力順が隣接するすべての入力コントロール間にに対して、上記のように各階層レベルでグルーピング判定を行ない、グループを生成していく。そして、グルーピング結果は、ノード番号、ラベル、基準点、親ノード番号およびコントロール種類を各ノードの属性情報として持つ図8に示すようなツリー構造で管理され、グルーピング処理部130の処理が終了した時点におけるコンテンツ構成ツリー1000が生成される。

【0168】

図8では、1番目から9番目までの入力コントロールをノード番号5～9およびノード番号13～17のノードとし、これを束ねるグループをプランチノードとしてノード番号2とする。また、ノード番号2～4の3つのプランチノードはコンテンツ全体を表す階層レベル1のルートノード（ノード番号1）に束ねられる。コンテンツ構成ツリーのデータ構造は、例えば、リーフからルート方向へ順次1つ上位階層のノードのデータ格納領域へのポインタを指定してツリー構造を形成する。また、ルートからリーフ方向へ順次1つ下位階層のノードのデータ格納領域へのポインタを指定して、ツリー構造を形成することもできる。また、ルートからリーフ方向、リーフからルート方向の両方のポインタをノードが有することもできる。

【0169】

なお、図8で示すようにコンテンツ構成ツリー1000の各ノードの属性情報であるラベルと基準点は、グルーピング処理部130の実行完了時点ではまだ格納されない。

【0170】

また、本実施形態では、入力コントロールのみを対象としてグルーピングを行なっているが、入力コントロールだけでなく、一般的にWebコンテンツに対してはテキストや画

10

20

30

40

50

像などをグルーピング処理の対象としてもよい。

【0171】

また、本実施形態では、コンテンツ構成ツリー生成処理部123が、グルーピング処理部130を起動し、グルーピング処理部130が、グループの生成とコンテンツ構成ツリー-1000の生成を行っているが、グルーピング処理部130はグループの生成だけ行い、コンテンツ構成ツリー生成処理部123が、コンテンツ構成ツリー-1000の生成を行つてもよい。

【0172】

ラベル付与処理部131は、階層レベルM（最下位階層レベル）から階層レベル1の順で、図8に示すコンテンツ構成ツリー-1000の各ノードにおけるラベルの抽出および基準点の設定を行なう。10

【0173】

ラベル付与処理部131は、階層レベルが階層レベルM（最下位階層レベル）のリーフノードの場合、各入力コントロールの基準位置を領域の左上とした上で、テキストリスト情報162および入力コントロールリスト情報163を用いて、n番目の入力コントロールの基準点に対し、上方向あるいは左方向（距離が同じ場合は左より上優先）で、配置位置的に最も近いテキストをn番目の入力コントロールのラベルとして抽出し、ラベルを含めたn番目の入力コントロールの基準点を再設定する。

【0174】

図5および図6の例では、入力コントロールリスト情報163より3番目の入力コントロールの基準点“X：285px Y：196px”に対し、テキスト情報162より6番目のテキストの基準点“X：100px Y：192px”が位置的に最も近いため、3番目の入力コントロールのラベルとして、6番目のテキスト“Address”が抽出され、抽出結果は、ラベル付与処理部131の処理実行後におけるコンテンツ構成ツリー-1100（図9参照）のノード番号6のノードに属性情報として格納されている。さらに、ラベルを含めた3番目の入力コントロール（ノード番号6）の基準位置を、6番目のテキストの基準位置“X：100px Y：192px”として再設定され、コンテンツ構成ツリー-1100のノード番号6のノードに属性情報として格納されている。20

【0175】

このように、コンテンツ500に含まれるすべての入力コントロールに対して、上述と同様の処理が行なわれることで、階層レベルM（最下位階層レベル）の各ノードのラベル名および基準位置が、コンテンツ構成ツリー-1100の各ノードの属性情報として格納される。30

【0176】

なお、本具体例では、n番目の入力コントロールのラベルをテキストリスト情報162に格納されているコンテンツ中のテキストの中から、入力コントロールの基準点に対して上方向と左方向（距離が同じ場合は左より上優先）で最も近い距離にあるテキストとしたが、ラジオボタンやチェックボックスについては入力コントロールの基準点に対して右方向と下方向（距離が同じ場合は下より右優先）で最も近い距離にあるテキストとされてもよい。40

【0177】

次に、ラベル付与処理部131は、階層レベルが階層レベルM（最下位階層レベル）以外のプランチノードの場合、テキストリスト情報162および入力コントロールリスト情報163を用いて、ノード番号nの基準点に対し、上方向あるいは左方向（距離が同じ場合は左より上優先）で位置的に最も近いテキストであって、未だラベルとして抽出されていないテキストをノード番号nのラベルとして抽出し、ラベルを含めたノード番号nの基準点を再設定する。

【0178】

図5および図6の例では、図8に示すコンテンツ構成ツリー-1000のノード番号2に関して、ノード番号2に属する要素（子ノード）の中で、最も左上（距離が同じ場合は左

より上優先)の基準位置をもつノード番号5の基準点“X：100px Y：158px”がノード番号2の基準位置に設定される。そのノード番号2の基準点に対し位置的に最も近いテキストの基準位置は、テキストリスト情報162の4番目のテキスト情報“First”であるが、これはすでにノード番号：5のラベルに使用されている。そこで、テキストリスト情報162の2番目のテキストの基準位置“X：90px Y：118px”が、未だラベルとして抽出されておらず位置的に最も近いテキスト情報要素であるため、ノード番号2のラベルとして、2番目のテキスト“Applicant Information”が抽出され、抽出結果がコンテンツ構成ツリー1100のノード番号2のノードに属性情報として格納される。さらに、ラベルを含めたノード番号2の基準位置をラベル(2番目のテキスト)の基準位置“X：90px Y：118px”として再設定され、図9で示すコンテンツ構成ツリー1100のノード番号2のノードに属性情報として格納される。10

【0179】

このように、階層レベルが階層レベルM(最下位階層レベル)以外のすべてのノードに対して、上述と同様の処理が行われることで、階層レベルが階層レベルM(最下位階層レベル)以外の各ノードのラベルおよび基準位置が、コンテンツ構成ツリー1100の各ノードの属性情報として格納される(なお、前述したように、図9のコンテンツ構成ツリーは一部が図示省略されている)。

【0180】

なお、本具体例におけるグルーピング処理部130では、グルーピング方法として、各入力コントロールの基準位置情報を用いて行ったが、図5に示すテキストリスト情報162のフォントサイズ情報を用いて、或る閾値以上のフォントサイズを持つテキストの位置を階層レベルmにおける入力コントロール間のグループ境界とみなすという方法をとってもよい。たとえば、フォントサイズが“+3”以上あるテキストの位置の直前を階層レベル2におけるグループの境界とみなすというグルーピング条件が設定されていた場合に、図5および図6の例では、テキストリスト情報162の2番目のテキスト“Applicant Information”と13番目のテキスト“Elective Conference”と26番目のテキスト“Payment”が“+3”的フォントサイズを持つので、入力コントロールリスト情報163の1番目と2番目と、9番目と10番目と、12番目と13番目の入力コントロール間が階層レベル2におけるグループの境界とみなされ、1番目～9番目、10番目～12番目および13番目～15番目の入力コントロールが階層レベル2における1つのグループとみなされる。20

【0181】

以上が、部品登録制御処理部111の具体的な処理の流れである。

【0182】

上述した第一の実施形態によれば、コンテンツを複数の部品に分け、複数の部品をそれぞれ表す複数の既存部品サブツリーもリポジトリ(例えば磁気ディスク装置101)に格納され、各既存部品サブツリーも、検索の対象となる。これにより、コンテンツファイルそれ自体のみを検索対象とすることに比べて、ユーザが作成しようとするコンテンツにより近いコンテンツとしての部品の検索でき、以って、ユーザの負担を軽減することができる。30

【0183】

また、上述した第一の実施形態によれば、クエリツリー及び既存部品サブツリーに、テキスト情報要素に加えて入力コントロール種類情報要素が含まれており、テキスト情報要素だけでなく入力コントロール種類情報要素をも基に、クエリツリーと既存部品サブツリーの類似度が算出される。これにより、テキストのみが検索キーとして入力される検索方法に比べて、ユーザが作成しようとするコンテンツにより近いコンテンツとしての部品を検索することができ、以って、ユーザの負担を軽減することができる。

【0184】

また、上述した第一の実施形態によれば、コンテンツファイルをリポジトリに登録するだけで、コンテンツ部品を表す既存部品サブツリーをリポジトリに蓄積できるとともに、50

それら蓄積された既存部品サブツリーに対して、人手によってあらかじめメタ情報を付与しなくとも、開発者が所望するコンテンツ部品を検索することができる。

【0185】

また、上述した第一の実施形態によれば、同一のコンピュータシステムや異なるコンピュータシステム（コンピュータシステムは例えば電子申請システム）でも、コンテンツ（例えば入力フォーム）のデザインの統一が図れる。

【0186】

<第二の実施形態>。

【0187】

次に、本発明の第二の実施形態について説明する。なお、以下、第一の実施形態の相違点を主に説明し、第一の実施形態との共通点については説明を省略或いは簡略する。10

【0188】

第二の実施形態では、レポジトリに登録される全ての既存部品サブツリーが、部品検索前に（具体的には部品登録の際に）、機能別およびスタイル別にグルーピングされ、部品候補一覧では、機能別およびスタイル別に部品候補に関する情報が提示される。この結果、複数の部品候補に関する情報を、機能単位およびスタイル単位で表示することができる。

【0189】

第一の実施形態と異なる点は、図14に示すとおり、部品登録処理部126に、機能別グルーピング処理部127とスタイル別グルーピング128が追加されている点と、部品検索処理部142と部品候補出力処理部143の処理がそれぞれ一部変更されている点である（変更後の部品登録処理部126、部品検索処理部142および部品候補出力処理部143を、それぞれ部品登録処理部126c、部品グループ検索処理部142cおよび部品候補出力処理部143cと呼ぶ）。20

【0190】

以下、第二の実施形態で行われる処理を、第一の実施形態で行われる処理と異なる部分を主に説明する。

【0191】

図15は、機能別グルーピング処理部127が行う処理の流れの一例を示す。

【0192】

まず、機能別グルーピング処理部127は、磁気ディスク装置101又はワークエリア150に登録されている本処理を実行していないすべてのコンテンツファイル160に対して、ステップ2511およびステップ2517～2519を繰り返す（ステップ2510）。

【0193】

次に、機能別グルーピング処理部127は、該コンテンツファイル160に対応するコンテンツ構成ツリー161に含まれるすべてのサブツリー（既存部品サブツリー174）に対して、ステップ2512～2513を繰り返す（ステップ2511）。

【0194】

次に、機能別グルーピング処理部127は、該既存部品サブツリーの部品内情報177B（ラベル情報175Bと入力コントロール情報176Bを含んだ情報）を取得し、取得した部品内情報177Bをワークエリア150に格納する（ステップ2512）。

【0195】

次に、機能別グルーピング処理部127は、すべての機能グループ188に対して、ステップ2514～2516を繰り返す（ステップ2513）。なお、機能グループとは、同一又は類似の機能を持つ部品に対応した既存部品サブツリーの集まりである。

【0196】

次に、機能別グルーピング処理部127は、該機能グループ188の部品内情報177C（ラベル情報175Cと入力コントロール情報176Cを含んだ情報）を取得し、取得された部品内情報177C（図19参照）を、ワークエリア150に格納する（ステップ4050）

2514)。

【0197】

次に、機能別グルーピング処理部127は、該既存部品サブツリー174の部品内情報177Bと該機能グループ188の部品内情報177Cから、類似度算出のためのベクトル空間とベクトルを表現したテーブル形式のベクトル情報182を生成し、生成したベクトル情報182をワークエリア150に格納する(ステップ2515)。

【0198】

次に、機能別グルーピング処理部127は、ベクトル情報182を用いて、該既存部品サブツリーに対する該機能グループの類似度を式2により算出する。類似度は、ベクトル情報182からベクトル空間法による余弦尺度(式2)により算出し、図18に例示する機能グループ別類似度算出結果189(機能グループID情報1891と類似度情報1892を含んだ情報)を、ワークエリア150に格納する(ステップ2516)。

【0199】

【数2】

式2

既存部品サブツリーと機能グループの類似度

$$= \frac{\text{既存部品サブツリーベクトル} \cdot \text{機能グループベクトル}}{|\text{既存部品サブツリーベクトル}| \cdot |\text{機能グループベクトル}|}$$

10

20

【0200】

次に、機能別グルーピング処理部127は、該既存サブツリー174に対して、閾値以上の類似度を持つ機能グループ188が存在するか否かを判定する(ステップ2517)。

【0201】

ステップ2517の判定の結果、閾値以上の類似度を持つ機能グループ188が存在する場合は、機能別グルーピング処理部127は、該既存部品サブツリー174を、最も高い類似度を持つ機能グループ188のメンバとし、該既存部品サブツリー174に対応する該コンテンツ構成ツリー161のノード195(図22参照)に機能グループIDを格納する(ステップ2518)。

30

【0202】

ステップ2517の判定の結果、閾値以上の類似度を持つ機能グループ188が存在しない場合は、機能別グルーピング処理部127は、該既存部品サブツリーをメンバとする新たな機能グループを生成し、機能グループIDを付与した上で(ステップ2519)、磁気ディスク装置101又はワークエリア150に、その既存部品サブツリーを登録する。なお、機能グループIDは、主メモリ102または磁気ディスク装置101に格納されているルールに基づいて機能グループIDを生成して付与することができる。また、一の機能グループについて、その機能グループに最後にメンバとされた既存部品サブツリーを、代表の既存部品サブツリーとすることができます。代表の既存部品サブツリーがどれであるかは、例えば、既存部品サブツリーのルートノードに代表であることを表すコード(以下、代表コード)を含めることで、特定することができる(代表となる既存部品サブツリーが変わった場合には、既存の代表の既存部品サブツリーのルートノードから代表コードを削除することができる)。

40

【0203】

図16は、スタイル別グルーピング処理部128が行う処理の流れの一例を示す。

【0204】

まず、スタイル別グルーピング処理部128は、磁気ディスク装置101又はワークエリア150に登録されている本処理を実行していないすべての機能グループ188に対して、ステップ2611を繰り返す(ステップ2610)。

50

【0205】

次に、スタイル別グルーピング処理部128は、該機能グループに含まれるすべての既存部品サブツリーに対して、ステップ2612～2613を繰り返す(ステップ2611)。

【0206】

次に、スタイル別グルーピング処理部128は、該既存部品サブツリーに対応するコンテンツファイル160のソースから、該既存部品サブツリーに対応した部品のスタイル情報を取得し(ステップ2612)、取得したスタイル情報を、ワークエリア150に格納する。なお、類似性を比較する部品間のテキスト数が二以上である場合、図20に例示するスタイル情報193(二以上のテキストとそれら二以上のテキストのコンテンツにおける並び順(図20では、一番左が先頭))が取得される。また、類似性を比較する部品間のテキスト数が1である場合、図21に例示するスタイル情報194(ラベル情報1941、type情報1942、size情報1943、font-size情報1944、color情報1945およびmaxlength情報1946を含んだ情報)が取得される。なお、スタイル情報193及び194は、HTMLソース(コンテンツファイル)の属性情報等から構築することができるが、ソースから抽出できる他の情報要素から構築されても良い。また、スタイル情報194に含まれる情報要素は、一のテキストに関する属性を表す情報要素であれば何でもよい。また、類似性を比較する部品間のテキスト数が二以上であっても、各テキストについてスタイル情報194が作成され、スタイル情報194同士の比較が行われても良い。

10

【0207】

次に、スタイル別グルーピング処理部128は、すべてのスタイルグループ190に対して、ステップ2614～2617を繰り返す(ステップ2613)。なお、スタイルグループとは、画面上に表示されるスタイルが同一又は類似している既存部品サブツリーの集まりである。

【0208】

次に、スタイル別グルーピング処理部128は、該スタイルグループからスタイル情報193又は194を取得し、取得したスタイル情報193又は194を、ワークエリア150に格納する(ステップ2614)。

20

【0209】

次に、スタイル別グルーピング処理部128は、既存部品サブツリーのスタイル情報193又は194とスタイルグループ190のスタイル情報193又は194とを比較し、スタイル情報が完全一致するか否かを判定する(ステップ2615)。

30

【0210】

ステップ2615の判定の結果、それぞれのスタイル情報が完全一致した場合は、スタイル別グルーピング処理部128は、該部品を該スタイルグループ190のメンバとし、該既存部品サブツリー174に対応する該コンテンツ構成ツリー161のノード195(図22参照)に、スタイルグループIDを格納する(ステップ2516)。

【0211】

ステップ2615の判定の結果、それぞれのスタイル情報が完全一致しない場合は、スタイル別グルーピング処理部128は、該既存部品サブツリーをメンバとする新たなスタイルグループを生成し、スタイルグループIDを付与した上で(ステップ2617)、磁気ディスク装置101又はワークエリア150に登録する。なお、スタイルグループIDは、主メモリ102または磁気ディスク装置101に格納されているルールに基づいて機能グループIDを生成して付与することができる。また、一のスタイルグループについて、最後にメンバとされた既存部品サブツリーを代表の既存部品サブツリーとすることができる。

40

【0212】

なお、ステップ2615の類似性判定処理では、既存部品サブツリーおよび該スタイルグループ190のそれぞれのスタイル情報が互いに完全一致しないと類似しているとみなさないとしたが、条件を詳細に設けて、小さな差異については吸収して類似しているとみ

50

なしてもよい。具体的には、例えば、スタイル情報 193 同士の比較において、テキストが完全一致であれば並び順が違っていても、比較対象の既存部品サブツリーが比較対象のスタイルグループのメンバとされても良いし、或いは、並び順が完全一致であれば所定数のテキストが違っていても、比較対象の既存部品サブツリーが比較対象のスタイルグループのメンバとされても良い。また、例えば、スタイル情報 194 同士の比較において、所定割合以上の情報要素が互いに一致していれば、比較対象の既存部品サブツリーが比較対象のスタイルグループのメンバとされても良い。

【0213】

なお、機能別グルーピング処理部 127 およびスタイル別グルーピング処理部 128 の処理結果は、図 22 に例示する既存部品グルーピング情報 196、すなわち、機能グループ ID 情報 1961、スタイルグループ ID 情報 1962、部品 ID 情報 1963（スタイルグループに含まれる既存部品サブツリーに対応した部品の ID）および出現数情報 1964（スタイルグループに含まれる既存部品サブツリーの数）が含まれている情報であってもよい。既存部品グルーピング情報 196 は、磁気ディスク装置 101 又はワークエリア 150 に格納されてもよい。10

【0214】

図 17 は、部品グループ検索処理部 142c が行う処理の流れの一例を示す。

【0215】

まず、部品グループ検索処理部 142c は、入出力定義情報解析処理部 141 で取得されたすべてのクエリツリー 171 に対して、ステップ 2710 ~ 2711 を繰り返す（ステップ 2700）。20

【0216】

次に、部品グループ検索処理部 142c は、該クエリツリー 171 の部品内情報 177A を取得し、取得された部品内情報 177A をワークエリア 150 に格納する（ステップ 2710）。

【0217】

次に、部品グループ検索処理部 142c は、磁気ディスク装置 101 又はワークエリア 150 に登録されているすべての機能グループ 188 に対して、ステップ 2712 ~ 2713 を繰り返す。30

【0218】

次に、部品グループ検索処理部 142c は、該機能グループ 188 の部品内情報 177C を取得し、取得された部品内情報 177C をワークエリア 150 に格納する（ステップ 2712）。

【0219】

最後に、部品グループ検索処理部 142c は、該クエリツリー 171 に対する該機能グループ 188 の類似度を算出し、図 23 に例示する機能グループ別類似度算出結果 189（機能グループ ID 情報 1891 と類似度情報 1892 を含んだ情報）をワークエリア 150 に格納する（ステップ 2713）。このステップ 2713 では、機能別グルーピング処理部 127 の処理流れにおけるステップ 2516 と実質的に同じ処理を行う（例えば、（式 2）で類似度を算出する）ことができる。40

【0220】

部品候補出力処理部 143c は、部品グループ検索処理部 142c で生成された機能グループ別類似度算出処理結果 189 と、機能別グルーピング処理部 127 とスタイル別グルーピング処理部 128 によって生成された既存部品グルーピング情報 196 とを用いて、クエリツリー 171 に対する機能グループ 188 に関する情報および機能グループ 188 内におけるスタイルグループに関する情報の一覧（以下、部品グループ候補一覧）を、部品グループ候補一覧表示画面 197 に表示することができる。この画面 197 は、例えば、第一の実施形態における部品候補一覧表示画面 187 に代わりに用意された画面である。

【0221】

10

20

30

40

50

部品グループ候補一覧には、機能グループのID、クエリツリー171に対する機能グループ188の類似度、該機能グループに含まれるサブ機能グループのID、該機能グループ188に含まれるスタイルグループのIDおよび該スタイルグループに含まれる部品数が含まれる。また、各スタイルグループの代表部品をプレビューするためのプレビューボタンが表示される。なお、該機能グループに含まれるサブ機能グループのIDは、該機能グループに含まれる既存部品サブツリー174内のノード195に記録されている機能グループIDである。すなわち、第一の機能グループに属する既存部品サブツリー内のノードに、第一の機能グループとは別の第二の機能グループのIDが記録されていれば、その第二の機能グループが、第一の機能グループのサブ機能グループとなる。

【0222】

10

以下、部品登録処理部126cの具体的な処理を、図18～図22を用いて説明する。

【0223】

図18は、機能別グルーピング処理部127の具体的な処理の説明図である。

【0224】

機能別グルーピング処理部127は、磁気ディスク装置101に登録されているすべてのコンテンツファイル160に含まれるすべての既存部品サブツリー174について、登録されている機能グループ188との類似度を算出することで、各既存部品サブツリー174に対応する部品がどの機能グループに属すかを決定する。

【0225】

20

図18の例では、グルーピング対象の既存部品サブツリー174が、3つの機能グループ188 (ID "001" ~ "003") の中のどの機能グループに属するかを決定するために、該既存部品サブツリー174に対する各機能グループ188の類似度が算出される。具体的には、機能グループID "001" については類似度 "0.50" が算出され、機能グループID "002" については類似度 "0.67" が算出され、機能グループID "003" については類似度 "0.42" が算出され、それぞれ機能グループID情報1891と類似度情報1892が機能グループ別類似度算出結果189に格納される。この結果、例えば類似度の閾値が "0.50" の場合、該既存部品サブツリー174は、類似度が最も高いID "002" の機能グループのメンバに追加される。また、類似度の閾値が "0.70" の場合、該既存部品サブツリー174の部品は、どの機能グループにも属さないものとして、該既存部品サブツリー174の部品をメンバとする新たな機能グループ (ID "004") が生成される。

30

【0226】

図19は、機能別グルーピング処理部127における類似度算出の具体的な処理の説明図である。

【0227】

まず、機能別グルーピング処理部127は、既存部品サブツリー174から部品内情報177Bを取得し、機能グループ188から部品内情報177Cを取得する。図19の例では、部品内情報177Bは、図12の部品内情報177Aと同じ内容であり、部品内情報177Cは、図12の部品内情報177Bと同じ内容である。このため、これらの部品内情報177B及び177Cを基に作成されたベクトル情報182を参照して上記(式2)を利用して算出された類似度は、0.67となる。機能別グルーピング処理部127は、機能グループ別類似度算出結果189に、機能グループID情報1891 "002" と、類似度情報1892 "0.67" とを記録する。

40

【0228】

なお、グルーピング対象の既存部品サブツリー174と機能グループ188との類似度算出では、機能グループ188における代表の既存部品サブツリーとグルーピング対象の既存部品サブツリー174とが比較されても良い。或いは、グルーピング対象の既存部品サブツリー174は、機能グループ188における二以上の(例えば全ての)既存部品サブツリー174と比較され、算出された二以上の類似度の平均が、グルーピング対象の既存部品サブツリー174と機能グループ188との類似度とされても良い。

50

【0229】

以上のように、機能グループとグルーピング対象の既存部品サブツリー 174 と

【0230】

図 20 は、類似性を比較する部品間のテキスト数が二以上である場合のスタイル別グルーピング処理部 128 の具体的な処理の説明図である。

【0231】

まず、スタイル別グルーピング処理部 128 は、コンテンツファイル 160 (ソース) から、グルーピング対象の既存部品サブツリーのスタイル情報 193 を取得する。スタイル情報 193 は、テキスト (例えば入力項目名) と順番を表現した情報であり、本図の例では、“氏名”、“住所”及び“電話番号”といった 3 つの入力項目名が、“氏名”、“住所”、“電話番号”の順でコンテンツ上にレイアウトされていることを示す。
10

【0232】

次に、スタイル別グルーピング処理部 128 は、スタイルグループ 190 についても、スタイルグループ 190 の代表既存部品ツリーからスタイル情報 193 を取得する。本図の例では、スタイルグループ ID “001”について、スタイル情報 193 として、“性”
名 住所 電話番号”が取得され、スタイルグループ ID “002”について、スタイル情報 193 として、“氏名 電話番号 住所”が取得され、スタイルグループ ID “003”について、スタイル情報 193 として、“氏名 住所 電話番号”が取得される。
20

【0233】

次に、スタイル別グルーピング処理部 128 は、グルーピング対象の既存部品サブツリー 174 のスタイル情報 193 と各スタイルグループ 190 の各スタイル情報 193 とを比較し、完全一致するかどうかを判定する。本図の例では、3 つのスタイルグループ 190 のうち、スタイルグループ ID “001”については入力項目名が異なり、スタイルグループ ID “002”については入力項目名の順番が異なり、スタイルグループ ID “003”に対応したスタイル情報 193 のみが、グルーピング対象の既存部品サブツリー 174 のスタイル情報 193 と完全一致することになる。このため、判定結果を表す情報であるスタイル別類似性判定結果 192 には、グルーピング対象の既存部品サブツリー 174 はスタイルグループ ID “003”のスタイルグループ 190 のメンバとなることが示される。
30

【0234】

図 21 は、類似性を比較する部品間のテキスト数が 1 である場合のスタイル別グルーピング処理部 128 の具体的な処理の説明図である。

【0235】

まず、スタイル別グルーピング処理部 128 は、コンテンツファイル 160 (ソース) から、グルーピング対象の既存部品ツリー 174 のスタイル情報 194 を取得する。スタイル情報 194 には、一つの入力項目名 (テキスト) のスタイルに関する属性を示す情報要素が含まれ、本図の例では、ラベル情報 1941 “氏名”、type 情報 1942 “text”、size 情報 1943 “10”、font-style 情報 1944 “normal”、font-size 情報 1945 “medium”、color 情報 1946 “black”、および max-length 情報 1947 “10” が含まれる。
40

【0236】

スタイル別グルーピング処理部 128 は、各スタイルグループ 190 についても、スタイルグループ 190 の代表既存部品サブツリーからスタイル情報 194 を取得する。本図の例では、スタイルグループ ID “001”に対応したスタイル情報 194 には、ラベル情報 1941 “名前”、type 情報 1942 “text”、size 情報 1943 “10”、font-style 情報 1944 “normal”、font-size 情報 1945 “medium”、color 情報 1946 “black” および max-length 情報 1947 “10” が含まれる。以下、スタイルグループ ID “002” と “003” についても、同様にスタイル情報 194 が取得される。
50

【0237】

次に、スタイル別グルーピング処理部128は、グルーピング対象の既存部品サブツリー174のスタイル情報194とスタイルグループ190のスタイル情報194とを比較し、それらが互いに完全一致するかどうかを判定する。本図の例では、スタイルグループID“001”はラベル情報1941が異なり、スタイルグループID“002”はfont-size情報1944が異なり、スタイルグループID“003”的スタイル情報194のみが、グルーピング対象の部品のスタイル情報194とすべての情報が一致している。スタイル別グルーピング処理部128は、そのことを検出し、スタイル別類似性判定結果192に、グルーピング対象の既存部品ツリー174はID“003”的スタイルグループのメンバとなることを記録する。

10

【0238】

なお、本図には示されていないが、仮にグルーピング対象の既存部品サブツリーのスタイル情報がどのスタイルグループのスタイル情報193或いは194とも一致しなかった場合は、グルーピング対象の既存部品サブツリーをメンバとする新たなスタイルグループを生成し、スタイルグループIDを付与した上で磁気ディスク装置101又はワークエリア150に登録することができる。また、この場合、グルーピング対象の既存部品サブツリーは新たに生成されたスタイルグループの代表既存部品サブツリーとなる。

【0239】

また、機能別グルーピング処理部127とスタイル別グルーピング処理部128のグルーピング結果は、グルーピング対象の既存部品サブツリーに対応するコンテンツ構成ツリー161のノード195の中に機能グループIDとスタイルグループIDとして格納される。図22の例では、コンテンツ構成ツリー161のノード195に、機能グループID“002”とスタイルグループID“003”がそれぞれ格納されている。

20

【0240】

また、前述したように、機能別グルーピング処理部127とスタイル別グルーピング処理部128のグルーピング結果は、部品候補出力処理部143cにおける情報取得の高性能化のために、既存部品グルーピング情報196（機能グループID情報1961、スタイルグループID情報1962、部品ID情報1963及び出現数情報1964を含んだ情報）が磁気ディスク装置101又はワークエリア150に格納されてもよい。本図の例では、既存部品グルーピング情報196の2レコード目の情報として、機能グループID情報1961“001”、スタイルグループID情報1962“002”、部品ID情報1963“0031_3、0065_2”および出現数情報1964“2”がそれぞれ格納されている。また、部品ID情報1963で先頭に格納された（下線で示された）部品IDについては、そのスタイルグループにおける代表既存部品サブツリーであることを示している。代表既存部品サブツリーがどれであるかを示す方法としては、代表既存部品サブツリーに対応した部品IDを先頭に格納することに代えて、下線で示すなど、他の方法を採用することができる。

30

【0241】

以上が、部品登録処理部126cの具体的な処理である。

【0242】

部品検索処理部142は、ユーザによって指定されたクエリツリー171に対して、磁気ディスク装置101又はワークエリア150に登録されているすべての機能グループ188との類似度を算出する。クエリツリー171に対する機能グループ188の類似度算出については、図19で示したグルーピング対象の既存部品サブツリーから部品内情報177Bを取得する処理が、クエリツリー171から部品内情報177Aを取得する処理に変わ以外は、図19で示した処理例と同じである。

40

【0243】

図23は、部品候補出力処理部143cの具体的な処理の説明図である。

【0244】

部品候補出力処理部143cは、部品グループ検索処理部142cで生成された機能グ

50

ループ別類似度算出結果 189 と、機能別グルーピング処理部 127 とスタイル別グルーピング処理部 128 によって生成された既存部品グルーピング情報 196 から、クエリツリー 171 に対する機能グループ 188 に関する情報および機能グループ 188 内におけるスタイルグループ 190 に関する情報の一覧を部品グループ候補一覧表示画面 197 に表示する。図 23 の例では、第 1 位の機能グループとして、機能グループ ID “002”、クエリツリー 171 に対する機能グループ 188 の類似度 “0.67”、該機能グループに含まれるサブ機能グループの ID “004 005” が表示されている。サブ機能グループ ID は、機能グループに含まれる既存部品サブツリー 174 におけるノード 195 内の、その機能グループの ID 以外の機能グループ ID である。また、第 1 位の機能グループ (ID : 002) の中に含まれるスタイルグループの一覧として、スタイルグループ ID “001”、“002” および “003” が表示されており、それぞれのスタイルグループに含まれる部品数 (出現数) として “30”、“2” および “1” がそれぞれ表示されている。
10

【0245】

また、本図の例では、各スタイルグループの代表既存部品サブツリーが表す部品をプレビューするためのプレビューボタンが設けられている。部品候補出力処理部 143c は、或るプレビューボタンが押下された場合、そのプレビューボタンに対応するスタイルグループ内の代表既存部品サブツリーが表すコードをソースから取得し、取得したコードを基に、代表既存部品サブツリーに対応した部品のイメージを表示することができる。
20

【0246】

以上が、本発明の第二の実施形態についての説明である。

【0247】

上述した第二の実施形態によれば、検索前に機能及びスタイルごとに既存部品サブツリーをグルーピングしておくことで、クエリツリーに対する部品候補を、機能およびスタイルの観点で重複することなく、異なる機能およびスタイルの部品ごとに一覧表示することができる。この結果、ユーザは効率よく所望の部品を部品候補一覧表示から探し出すことができる。
30

【0248】

<第三の実施形態>。

【0249】

次に、本発明の第三の実施形態について説明する。
30

【0250】

第三の実施形態では、第一の実施形態における部品検索方法で既存部品サブツリーの集合を或る程度絞った上で、その集合における各既存部品サブツリーについて、第二の実施形態で説明した機能別およびスタイル別のグルーピング処理を行なう。
40

【0251】

本実施形態において、第一の実施形態と異なる点は、図 24 に示すとおり、部品候補出力処理部 143 に機能別クルーピング処理部 127、スタイル別グルーピング 128、部品グループ検索処理部 142c および部品候補出力処理部 143c を追加する点である (変更後の部品候補出力処理部 143 を部品候補出力制御処理部 144 と呼ぶ)。
40

【0252】

以下、第一の実施形態及び第二の実施形態と異なる処理を主に説明する。

【0253】

部品候補出力制御処理部 144 は、まず、部品検索処理部 142 によって算出された各既存部品サブツリーの類似度に対して、磁気ディスク装置 101 又はワークエリア 150 に格納されている類似度の閾値と比較し、閾値以上の類似度を持つ既存部品サブツリーを抽出する。以下、抽出された既存部品サブツリーの集合を「一次検索結果部品集合」と呼ぶ。
50

【0254】

次に、部品候補出力制御処理部 144 は、機能別グルーピング処理部 127 を実行し、

機能別グルーピング処理部 127 が、一次検索結果部品集合における各既存部品サブツリーについて、機能別にグルーピングを行なう。機能別グルーピング処理部 127 の処理手順は、処理対象が一次検索結果部品集合における既存部品サブツリーとなる他は、第二の実施形態と同様である。

【0255】

次に、部品候補出力制御処理部 144 は、スタイル別グルーピング処理部 128 を実行し、スタイル別グルーピング処理部 128 が、登録された全ての機能グループ 188 に対して、スタイル別にグルーピングを行なう。スタイル別グルーピング処理部 128 の処理手順は、処理対象が一次検索結果部品集合に対して機能別にグルーピングされた各機能グループ 188 となる他は、第二の実施形態と同様である。

10

【0256】

次に、部品グループ検索処理部 142c を実行し、ステップ 3410 によって登録されたすべて機能グループ 188 に対して、クエリサブツリー 171 との類似度を算出する。部品グループ検索処理部 142c の処理手順は、処理対象が一次検索結果部品集合に対して機能別にグルーピングされた各機能グループ 188 となる他は第二の実施形態と同様である。

【0257】

次に、部品候補出力制御処理部 144 は、部品候補出力処理部 143c を実行し、部品候補出力処理部 143c が、生成された機能グループ別類似度算出結果 189 と、機能別グルーピング処理部 127 とスタイル別グルーピング処理部 128 によって生成された既存部品グルーピング情報 196 から、クエリツリー 171 に対する部品グループ候補一覧を部品グループ候補一覧表示画面 197 に表示する。部品候補出力処理部 143c の処理手順は、部品候補出力の対象が一次検索結果部品集合となる他は、第二の実施形態と同様である。

20

【0258】

第三の実施形態で行われる処理は、機能別およびスタイル別のグルーピングの処理対象が一次検索結果部品集合となる他は、第二の実施形態と略同様である。

【0259】

以上、第三の実施形態によれば、機能別およびスタイル別のグルーピングの処理対象を、第一の実施形態による部品検索方法によって、ある程度絞った部品集合として、新規に作成しようとしているコンテンツの分野の中で使われている部品グループのみを出力することができる。この結果、ユーザは効率よく所望の部品を部品グループ候補一覧から探し出すことができる。

30

【0260】

<第四の実施形態>。

【0261】

第四の実施形態では、テキストの類似度と入力コントロールの類似度とを別々に算出し、それらの類似度に基づいて、クエリツリーと既存部品サブツリー（或いは機能グループ）との類似度を算出することができる。具体的には、例えば、下記（式3）
クエリツリーと既存部品サブツリー（或いは機能グループ）との類似度 = (テキストの類似度) + (入力コントロールの類似度) ... (式3)

40

で算出することができる。及び は、それぞれ、余弦尺度を用いて求めることができる。

【0262】

例えば、クエリツリーに、5つのテキスト（入力項目名）として、氏名、住所、電話番号及びメールアドレスがあり、8つの入力コントロールとして、5つのテキストボックス、チェックリスト、ラジオボタン、プルダウンがあるとする。一方、比較対象の既存部品サブツリーに、7つのテキストとして、氏名、性、名、メールアドレス、職業、住所及び電話番号があり、6つの入力コントロールとして、5つのテキストボックス、プルダウンがあるとする。このため、クエリツリーと比較対象の既存部品サブツリーとの間で、テキ

50

ストについては、氏名、住所、電話番号及びメールアドレスの4つが一致し、入力コントロールについては、5つのテキストボックスとプルダウンの6つが一致する。

【0263】

また、テキスト及び入力コントロールの8つのセットうち、クエリツリーについては、一致するセットとして3つあり、比較対象の既存部品サブツリーについては一致するセットとして7つあり、クエリツリーと比較対象の既存部品サブツリーとの間で互いに一致するセットが3つあるとする。

【0264】

この場合、第一乃至第三の実施形態によれば、類似度が0.57として算出される。なぜなら、余弦尺度((式1)及び(式2))において、分子が、 $1^2 + 1^2 + 1^2 = 3$ となり、分母が、 $\{\sqrt{1^2 + 1^2 + 1^2 + 1^2}\} \times \{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}\}$ となるからである。

10

【0265】

しかし、第四の実施形態では、上記(式3)を用いて類似度が求められる。具体的には、例えば、は0.76であり、は0.96となり、そのため、類似度は、1.72となる。が0.76となる理由は、余弦尺度において、分子が、 $1^2 + 1^2 + 1^2 + 1^2 = 4$ となり、分母が、 $\{\sqrt{1^2 + 1^2 + 1^2 + 1^2}\} \times \{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}\}$ となるからである。が0.96となる理由は、余弦尺度において、分子が、 $5^2 + 1^2 = 26$ となり、分母が、 $\{\sqrt{5^2 + 1^2 + 1^2 + 1^2}\} \times \{\sqrt{5^2 + 1^2}\}$ となるからである。

20

【0266】

なお、及びは、余弦尺度により算出されることとしたが、他の方法により算出されても良い。

【0267】

<第五の実施形態>。

【0268】

第五の実施形態は、第二と第三の実施形態の組合せであり、且つ、登録性能を重視するか検索性能を重視するかの入力を受け付ける機能がコンテンツ部品検索システムに設けられる。コンテンツ部品検索システムでは、登録性能を重視することが開発者から入力された場合には、第三の実施形態で説明したように、一次検索結果部品集合について機能別及びスタイル別のグルーピング処理が行われ、検索性能を重視することが開発者から入力された場合は、第二の実施形態で説明したように、新規に登録されるコンテンツの各部品を表す各既存部品サブツリーについて、機能別及びスタイル別のグルーピング処理が行われる。

30

【0269】

以上、本発明の幾つかの実施形態を説明したが、これらの実施形態は本発明の説明のための例示にすぎず、本発明の範囲をそれらの実施形態にのみ限定する趣旨ではない。本発明は、その要旨を逸脱することなく、その他の様々な態様でも実施することができる。例えば、既存部品サブツリーや評価結果などの格納先は、主メモリ又は磁気ディスク装置に限らず、他種の記憶資源であってもよい。

40

【図面の簡単な説明】

【0270】

【図1】図1は、本発明の第一の実施形態におけるコンテンツ部品検索システムの全体構成例を示す。

【図2】図2は、グルーピング処理部が行う処理の流れの一例を示す。

【図3】図3は、ラベル付与処理部が行う処理の流れの一例を示す。

【図4】図4は、コンテンツの表示例を示す。

【図5】図5は、テキストリスト情報の一例を示す。

【図6】図6は、入力コントロール情報の一例を示す。

【図7】図7は、グルーピング条件リストの一例を示す。

50

【図8】図8は、ラベル付与処理部によるラベル付与前のコンテンツ構成ツリーの一例を示す。

【図9】図9は、ラベル付与処理部によるラベル付与後のコンテンツ構成ツリーの一例を示す。

【図10】図10は、部品検索処理部が行う処理の流れの一例を示す。

【図11】図11は、入出力定義情報からのクエリツリーの取得の説明図である。

【図12】図12は、部品検索処理部が行う処理の一例の説明図である。

【図13】図13は、部品候補出力処理部が行う処理の一例の説明図である。

【図14】図14は、本発明の第二の実施形態における部品登録処理部および部品検索制御処理部の構成例を示す。 10

【図15】図15は、機能別グルーピング処理部が行う処理の流れの一例を示す。

【図16】図16は、スタイル別グルーピング処理部が行う処理の流れの一例を示す。

【図17】図17は、本発明の第二の実施形態における部品グループ検索処理部が行う処理の流れの一例を示す。 11

【図18】図18は、機能別グルーピング処理部が行う処理概要の説明図である。

【図19】図19は、機能別グルーピング処理部の類似度算出処理の説明図である。

【図20】図20は、類似性を比較する部品間のテキスト数が二以上である場合のスタイル別グルーピング処理部の具体的な処理の説明図である。 12

【図21】図21は、類似性を比較する部品間のテキスト数が1である場合のスタイル別グルーピング処理部の具体的な処理の説明図である。 13

【図22】図22は、コンテンツ構成ツリーにおけるノードと、既存部品グルーピング情報との一例を示す。 14

【図23】図23は、本発明の第二の実施形態における部品候補出力処理部が行う処理の説明図である。 15

【図24】図24は、本発明の第三の実施形態における部品検索制御処理部の構成例を示す。 16

【符号の説明】

【0271】

100...中央演算処理装置(CPU) 101...磁気ディスク装置 102...主メモリ 17

103...フロッピーディスクドライブ(FDD) 104...バス 105...ネットワーク 18

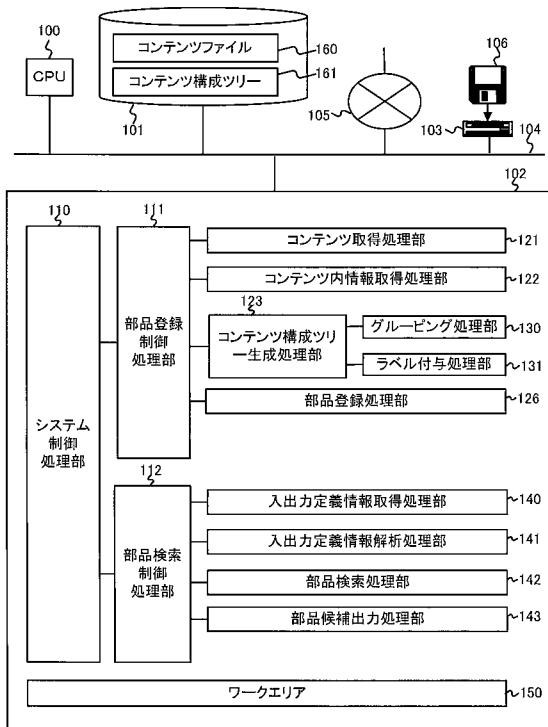
106...フロッピーディスク 110...システム制御処理部 111...部品登録制御処理部 112...部品検索制御処理部 112c...部品検索制御処理部 121...コンテンツ取得処理部 19

122...コンテンツ内要素情報取得処理部 123...コンテンツ構成ツリー生成処理部 126...部品登録処理部 126c...部品登録処理部 127...機能別グルーピング処理部 128...スタイル別グルーピング処理部 130...グルーピング処理部 131...ラベル付与処理部 140...入出力定義情報取得処理部 141...入出力定義情報解析処理部 142...部品検索処理部 142c...部品グループ検索処理部 143...部品候補出力処理部 143c...部品候補出力処理部 144...部品候補出力制御処理部 150...ワークエリア 160...コンテンツファイル 161...コンテンツ構成ツリー 20

40

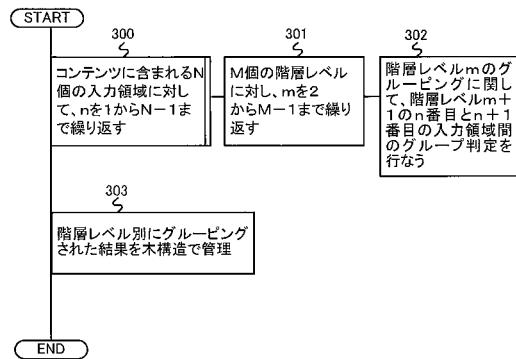
【図1】

図1



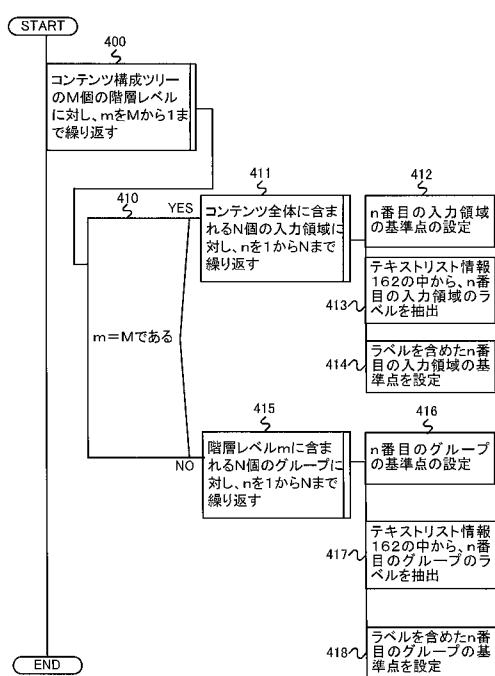
【図2】

図2



【図3】

図3



【図5】

図5

NO.	テキスト	基準位置情報	フォントサイズ
1	○○ Conference Form	X:177px Y:35px	+4
2	Applicant Information	X:90px Y:118px	+3
3	Name	X:100px Y:158px	+2
4	First	X:280px Y:158px	+1
5	Last	X:405px Y:158px	+1
6	Address	X:100px Y:192px	+2
7	Telephone Nuber	X:100px Y:225px	+2
8	Email Address	X:100px Y:258px	+2
9	Date of Birth	X:100px Y:291px	+2
10	Year	X:291px Y:300px	+1
11	Month	X:397px Y:300px	+1
12	Day	X:494px Y:300px	+1
13	Elective Conference	X:90px Y:358px	+3
14	Conference Name	X:155px Y:394px	+1
15	Time	X:293px Y:394px	+1
16	Room Number	X:438px Y:394px	+1
17	○○○○	X:155px Y:421px	+1
18	10:00 -12:00	X:293px Y:421px	+1
19	900	X:438px Y:421px	+1
20	△△△△	X:155px Y:457px	+1
21	13:00 - 14:00	X:293px Y:457px	+1
22	901	X:438px Y:457px	+1
23	× × × ×	X:155px Y:490px	+1
24	15:00 - 17:00	X:293px Y:490px	+1
25	902	X:438px Y:490px	+1
26	Payment	X:89px Y:550px	+3
27	Cash	X:139px Y:592px	+1
28	Credit	X:230px Y:592px	+1
29	Company	X:231px Y:614px	+1

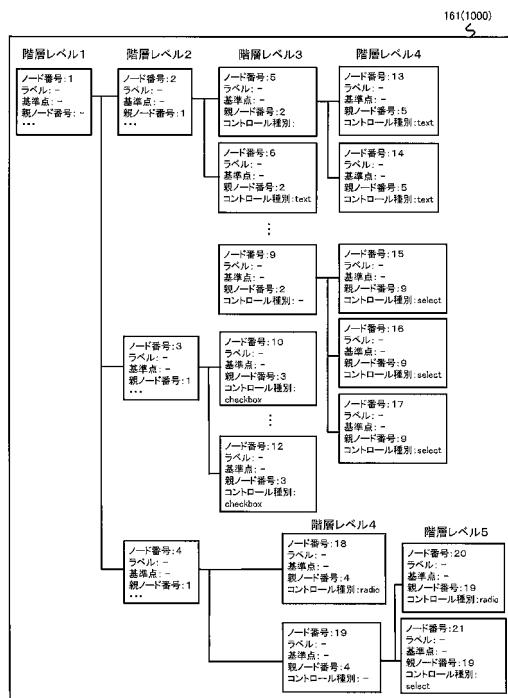
【図6】

図6

NO.	入力コントロール種類	基準位置情報	領域サイズ
1	<INPUT type="text">	X:315px Y:161px	width:73px height:20px
2	<INPUT type="text">	X:440px Y:161px	width:73px height:20px
3	<INPUT type="text">	X:285px Y:196px	width:68px height:20px
4	<INPUT type="text">	X:285px Y:229px	width:160px height:20px
5	<INPUT type="text">	X:483px Y:265px	width:160px height:20px
7	<SELECT>	X:325px Y:296px	width:49px height:16px
8	<SELECT>	X:443px Y:296px	width:25px height:16px
9	<SELECT>	X:523px Y:296px	width:25px height:16px
10	<INPUT type="checkbox">	X:120px Y:420px	width:15px height:15px
11	<INPUT type="checkbox">	X:120px Y:455px	width:15px height:15px
12	<INPUT type="checkbox">	X:120px Y:488px	width:15px height:15px
13	<INPUT type="radio">	X:118px Y:592px	width:15px height:15px
14	<INPUT type="radio">	X:202px Y:592px	width:15px height:15px
15	<SELECT>	X:291px Y:612px	width:60px height:16px

【図8】

図8



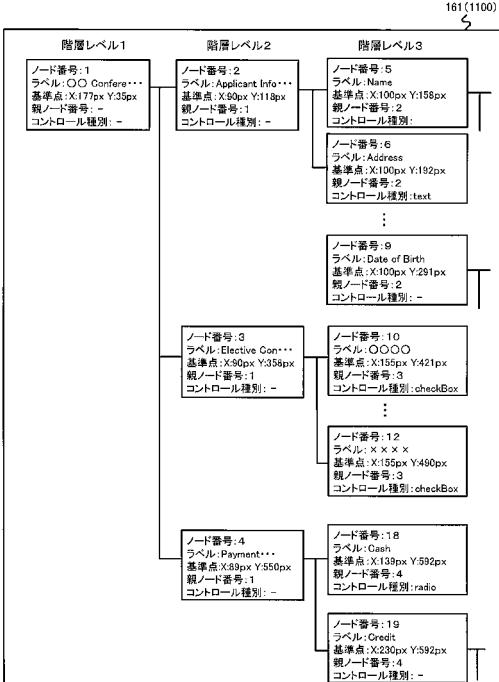
【図7】

図7

ケース	位置関係	グルーピング条件
1	真右	n番目の最右端とn+1番目の最左端との水平距離が100px未満
2	真下	n番目の最下端とn+1番目の最上端との垂直距離が20px未満
3	右下	n番目の最右端とn+1番目の最左端との水平距離が100px未満、かつn番目の最下端とn+1番目の最上端との垂直距離が20px未満
4	右上	n番目の最右端とn+1番目の最左端との水平距離が100px未満
5	左下	n番目の最下端とn+1番目の最上端との垂直距離が20px未満
6	真左	n番目とn+1番目の入力領域間はグルーピングされない。 → n番目とn+1番目の入力領域間は、階層レベルmでは必ずグループ境界となる
7	真上	
8	左上	

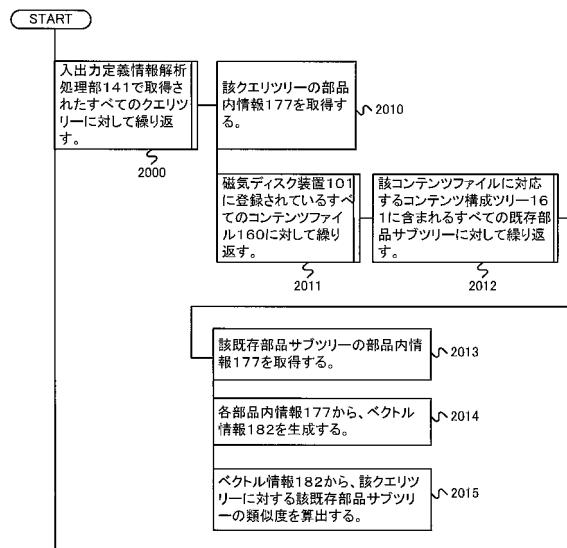
【図9】

図9



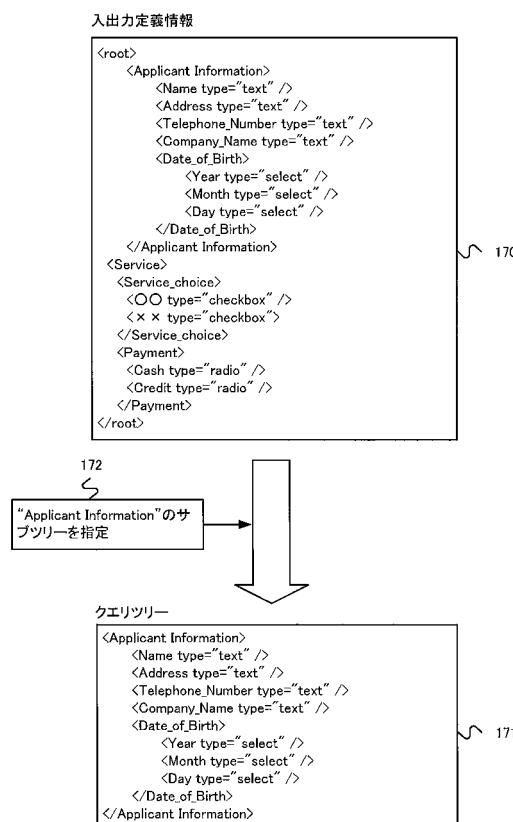
【図10】

図10



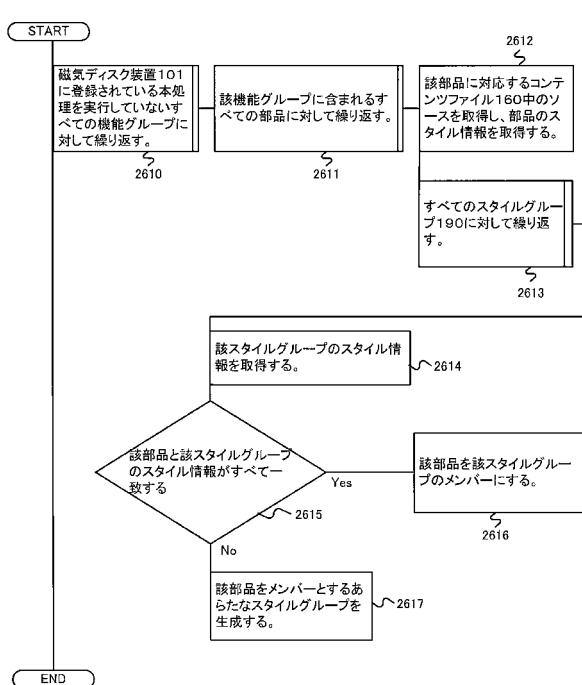
【図11】

図11



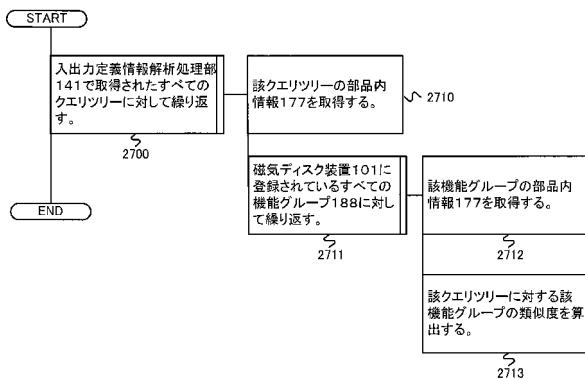
【図16】

図16



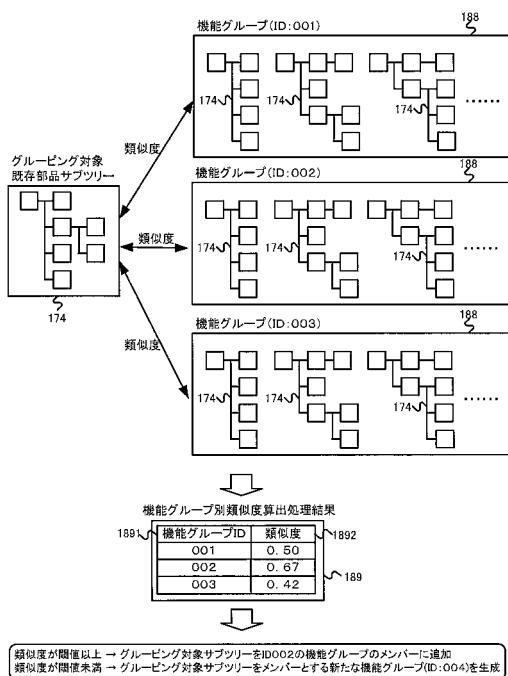
【図17】

図17



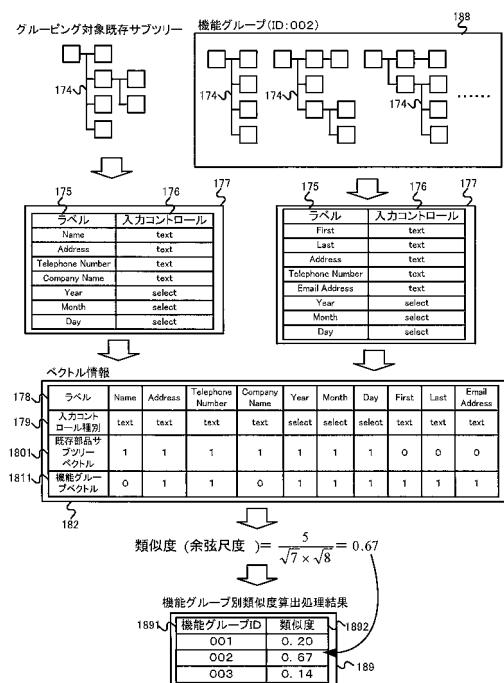
【図18】

図18

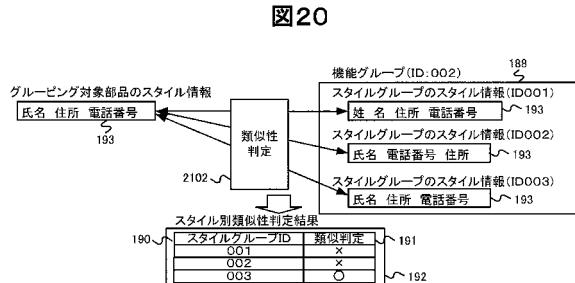


【図19】

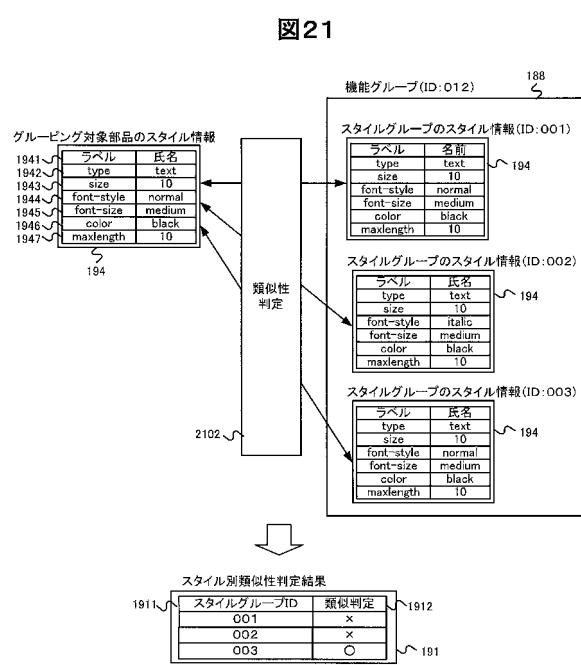
図19



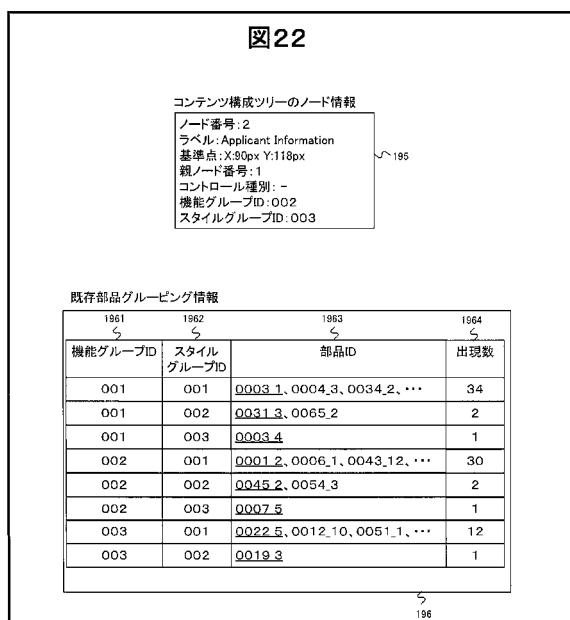
【図20】



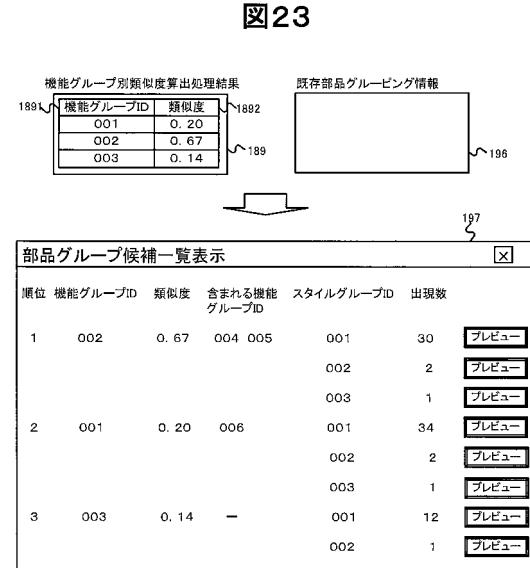
【図21】



【図22】

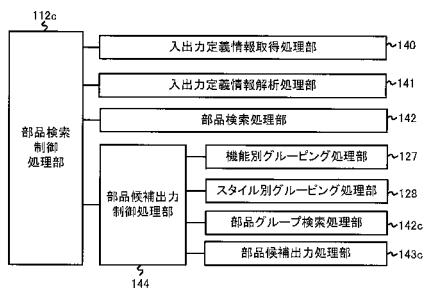


【図23】



【図24】

図24



【図4】

図4

500

OO Conference Form

Applicant Information

Name	First <input type="text"/>	Last <input type="text"/>	
Address	<input type="text"/>		
Telephone Number	<input type="text"/>		
Email Address	<input type="text"/>		
Date of Birth	Year <input type="text" value="1978"/>	Month <input type="text" value="3"/>	Day <input type="text" value="5"/>

Elective Conference

Conference Name	Time	Room Number
<input type="checkbox"/> ○○○○	10:00 - 12:00	900
<input type="checkbox"/> △△△△	13:00 - 14:00	901
<input type="checkbox"/> ××××	15:00 - 17:00	902

Payment

<input type="radio"/> Cash	<input checked="" type="radio"/> Credit
Company <input type="text" value="□□□"/>	

【 四 1 3 】

13

類似度算出処理結果(評価結果)		183
1831		コンテンツID 0001
1832	ノード番号	類似度
	1	0. 50
	2	0. 67
	⋮	⋮
	21	0. 00
1833		コンテンツID 0002
	ノード番号	類似度
	1	0. 60
	2	0. 43
	⋮	⋮
	8	0. 00
1834		コンテンツID 0003
	ノード番号	類似度
	1	0. 11
	2	0. 23
	⋮	⋮
	12	0. 58



185

184

<p>入出力定義情報</p> <pre> <root> <Applicant_Information> <Name type="text" /> <Address type="text" /> <Telephone_Number type="text" /> <Company_Name type="text" /> <Date_of_Birth> <Year type="select" /> <Month type="select" /> <Day type="select" /> </Date_of_Birth> </Applicant_Information> <Service> <Service_choice> <○○ type="checkbox" /> <×× type="checkbox" /> </Service_choice> <Payment> <Cash type="radio" /> <Credit type="radio" /> </Payment> </root> </pre> <p>クエリツリー情報</p> <ul style="list-style-type: none"> - Applicant_Information(4) <ul style="list-style-type: none"> - Date_of_Birth(17) - Service(0) 	<p>プレビュー画面</p> <h3>Applicant Information</h3> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">Name</td> <td style="width: 15%;">First</td> <td style="width: 15%;">Last</td> <td style="width: 45%;"></td> </tr> <tr> <td>Address</td> <td colspan="3"></td> </tr> <tr> <td>Telephone Number</td> <td colspan="3"></td> </tr> <tr> <td>Email Address</td> <td colspan="3"></td> </tr> <tr> <td>Date of Birth</td> <td>Year</td> <td>1978</td> <td>Month</td> <td>3</td> <td>Day</td> <td>5</td> </tr> </table> <h3>部品候補一覧表示</h3> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>順位</th> <th>部品ID</th> <th>類似度</th> <th>ラベル</th> <th></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0001_2</td> <td>0. 67</td> <td>Applicant Information</td> <td><button>プレビュー</button></td> </tr> <tr> <td>2</td> <td>0002_1</td> <td>0. 60</td> <td>Applicant Information</td> <td><button>プレビュー</button></td> </tr> <tr> <td>3</td> <td>0003_1</td> <td>0. 58</td> <td>Personal Information</td> <td><button>プレビュー</button></td> </tr> <tr> <td>4</td> <td>0001_1</td> <td>0. 50</td> <td>○○ Conference Form</td> <td><button>プレビュー</button></td> </tr> </tbody> </table>	Name	First	Last		Address				Telephone Number				Email Address				Date of Birth	Year	1978	Month	3	Day	5	順位	部品ID	類似度	ラベル		1	0001_2	0. 67	Applicant Information	<button>プレビュー</button>	2	0002_1	0. 60	Applicant Information	<button>プレビュー</button>	3	0003_1	0. 58	Personal Information	<button>プレビュー</button>	4	0001_1	0. 50	○○ Conference Form	<button>プレビュー</button>
Name	First	Last																																															
Address																																																	
Telephone Number																																																	
Email Address																																																	
Date of Birth	Year	1978	Month	3	Day	5																																											
順位	部品ID	類似度	ラベル																																														
1	0001_2	0. 67	Applicant Information	<button>プレビュー</button>																																													
2	0002_1	0. 60	Applicant Information	<button>プレビュー</button>																																													
3	0003_1	0. 58	Personal Information	<button>プレビュー</button>																																													
4	0001_1	0. 50	○○ Conference Form	<button>プレビュー</button>																																													

186

187

フロントページの続き

(72)発明者 茶谷 謙一

神奈川県横浜市戸塚区戸塚町 5030 番地 株式会社日立製作所 ソフトウェア事業部内

審査官 多胡 滋

(56)参考文献 特開2004-102839(JP,A)

特開2001-109742(JP,A)

特開平09-231062(JP,A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 9 / 06

G 06 F 9 / 44

G 06 F 13 / 00