

(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2010-108173

(P2010-108173A)

(43) 公開日 平成22年5月13日(2010.5.13)

(51) Int.Cl. F 1 テーマコード (参考)
G 0 6 F 1 7 / 3 0 (2 0 0 6 . 0 1) G 0 6 F 1 7 / 3 0 2 1 0 D 5 B 0 7 5

審査請求 未請求 請求項の数 5 O L (全 17 頁)

(21) 出願番号	特願2008-278644 (P2008-278644)	(71) 出願人	000003078
(22) 出願日	平成20年10月29日 (2008.10.29)		株式会社東芝
			東京都港区芝浦一丁目1番1号
		(71) 出願人	301063496
			東芝ソリューション株式会社
			東京都港区芝浦一丁目1番1号
		(74) 代理人	100058479
			弁理士 鈴江 武彦
		(74) 代理人	100108855
			弁理士 蔵田 昌俊
		(74) 代理人	100091351
			弁理士 河野 哲
		(74) 代理人	100088683
			弁理士 中村 誠

最終頁に続く

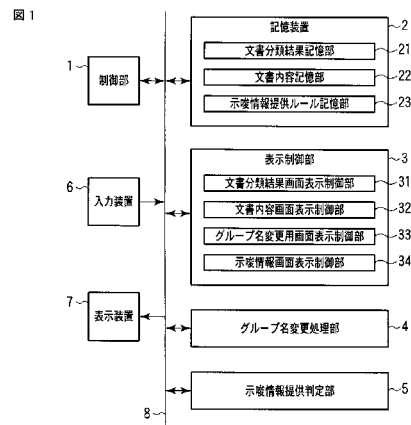
(54) 【発明の名称】 文書処理装置

(57) 【要約】

【課題】文書群のグループに対する識別情報の人為的な付与の妥当性を判定する。

【解決手段】表示制御部3の文書分類結果画面表示制御部31が文書分類結果記憶部21に記憶される文書分類情報をもとに文書分類結果画面を表示装置7に表示した状態で、変更対象のグループ名の変更先のグループ名が入力されると、示唆情報提供判定部5は、記憶装置2の示唆情報提供ルール記憶部23に記憶された示唆情報提供ルールに基づき、変更先のグループ名が不適切であることを示すための示唆情報を提供するかどうか、提供する場合はどのような示唆情報を提供するかを判定する。示唆情報提供判定部5は、示唆情報を提供すると判定した場合は、示唆情報提供ルールおよび所定のテンプレートを用いて生成した示唆情報を表示装置7に表示する。示唆情報提供判定部5が示唆情報を提供しないと判定した場合は、グループ名変更処理部4はグループ名の変更を行なう。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

複数の文書のそれぞれをグループ識別情報が付与される複数のグループのいずれかに分類した文書分類情報を管理する管理手段と、

前記管理される文書分類情報を出力する分類情報出力手段と、

当該出力された文書分類情報で示されるグループ識別情報のうち変更対象のグループ識別情報に対する変更先のグループ識別情報の入力を受け付ける入力手段と、

前記管理されるグループごとに、当該グループに属する文書数のうち前記入力された変更先のグループ識別情報の出現する文書数の割合を計算する計算手段と、

前記計算手段による計算結果をもとに、前記変更対象のグループ識別情報の前記変更先のグループ識別情報への変更が適切であるか否かを判定する判定手段と、

前記変更が適切であると前記判定手段により判定した場合に、前記管理手段により管理される変更対象のグループ識別情報を前記入力された変更先のグループ識別情報に変更する変更手段と、

前記変更が適切でないと前記判定手段により判定した場合に当該判定結果を出力する判定結果出力手段と

を備えたことを特徴とする文書処理装置。

【請求項 2】

前記判定手段は、

前記計算手段により計算した割合が所定の値を超えるグループの数が前記管理手段により管理されるグループ全体の数の所定割合以上である場合に前記変更が適切でないと判定する

ことを特徴とする請求項 1 に記載の文書処理装置。

【請求項 3】

前記判定手段は、

前記計算手段により計算した割合が所定の値を超えるグループの数が前記管理手段により管理されるグループ全体の数の所定割合未満である場合で、かつ、前記計算した割合が前記所定の値を超えるグループが前記変更対象のグループ識別情報に対応するグループ以外で所定数以下存在する場合に、前記変更が適切でなく、かつ、前記変更対象のグループ識別情報に対応するグループと前記変更対象のグループ識別情報に対応するグループ以外における前記計算した割合が前記所定の値を超えるグループとのマージが適切であると判定し、前記計算した割合が所定の値を超えるグループの数がグループ全体の数の所定割合未満である場合で、かつ、前記計算した割合が前記所定の値を超えるグループが前記変更対象のグループ識別情報に対応するグループ以外で所定数以下存在しない場合に、前記変更が適切であると判定する

ことを特徴とする請求項 1 に記載の文書処理装置。

【請求項 4】

前記判定手段は、

前記入力手段により入力した変更先のグループ識別情報が所定の不適切な識別情報である場合に、前記変更が不適切であると判定する

ことを特徴とする請求項 1 に記載の文書処理装置。

【請求項 5】

前記出力手段は、

前記変更が適切でないと前記判定手段により判定した場合に、前記入力した変更先のグループ識別情報は不適切であって、それでもなお前記変更対象のグループ識別情報を前記変更先のグループ識別情報に変更するか否かを問うメッセージを出力する

ことを特徴とする請求項 1 に記載の文書処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

10

20

30

40

50

本発明は、複数のグループに分けられた文書群に対するユーザによるグループ識別情報の付与を支援する文書処理装置に関する。

【背景技術】

【0002】

近年、計算機の高性能化や記憶媒体の大容量化、計算機ネットワークの普及などに伴い、電子化された大量の文書群を計算機システムに記憶管理して利用することが可能となった。ここでいう文書とは、例えば、帳票、企画書、設計書といった業務文書や、マニュアル、特許、技術文献、法令、規程、議事録、ニュース記事、電子メール、ウェブページ、書籍などを指す。

【0003】

ユーザが計算機システム上で各種作業を行う際、もし取り扱う文書群が少数である場合は、文書群を事前に整理しておく必要はない。しかし、もし文書が数百～数千と大量になる場合は、作業を行う前に文書を複数のグループに整理しておく方が、作業を効率よく実施しやすくなる。

【0004】

例えば、電子メール閲覧・送受信システムにおいて、過去に自分が受け取ったメールを閲覧するという作業を行う際、無整理のままではある場合に比べて、事前に過去のメールを「Aさんからのメール」、「プロジェクトBに関係のあるメール」などのグループに分けておいた方が、自分が見たいメールをより手早く見ることが出来るようになる。このように、文書分類は計算機システムを用いて文書処理を行う際に欠かせないものとなっている。

【0005】

一方、文書分類を行う際、文書群の傾向がわからないと、分け方そのものが思い付きにくいというケースがある。例えば、特許調査において、特許検索システムによって検索した数千件の自社・他社の特許を、記載されている技術内容をもとに複数のグループに分けて分析するという事があるが、このような場合、どのようなグループに分ければ最適な分析が出来るかというのは、実際に全ての特許を見てからでないとわからない。これは非常に手間のかかる作業であり、特許調査に多大な所要時間がかかる一因となっている。

【0006】

このように文書全体の傾向が分からない場合の分類作業を支援するため、例えば非特許文献1に開示されるように、大量の文書群を当該文書群に含まれる単語の統計的傾向などにより複数のグループに自動的に分類する自動分類技術が存在する。一般に「文書クラスタリング」と呼ばれるこの技術を用いることによって、文書群の傾向をある程度把握することができ、文書群の分け方のヒントとすることができる。これにより文書分類の際の所要時間を低減させることができる。

【0007】

前記の文書クラスタリング技術の大きな特徴は、事前に「どのようなグループに分けるのか」という情報が不必要であるという点にある。逆に、「どのようなグループに分けるのか」をユーザが指定できない（指定しない）ため、そのままでは各グループにどのような性質の文書が入っているのか分からないという特徴もある。

【0008】

そこで、各グループに含まれている文書群の特性を利用して、各グループに名称を付与することが必要となってくる。例えば特許文献1に開示されるように、各グループに対し、そのグループに含まれる単語の頻度などを利用して各グループに自動的に名称をつける技術が存在する。名称の付与方法は、単語や形態素、複数の単語や形態素の集合などさまざまである。

【特許文献1】特開2005-182611号公報

【非特許文献1】岸田和明,文書クラスタリングの手法:文献レビュー,Library and Information Science. No.49, 2003, p.33-75

【発明の開示】

10

20

30

40

50

【発明が解決しようとする課題】**【0009】**

ここで、各グループに付与された名称はあくまで機械的に付与されたものであり、これを人間がより識別しやすい名称に人手で変更したいというケースがある。例えば、遠隔通信関連特許文書1000件が文書クラスタリング技術によって6つのグループに自動的に分類された際、自動的に付与されたグループ名が、第1のグループ名「管理」、第2のグループ名「操作」、第3のグループ名「液晶」、第4のグループ名「モバイル」、第5のグループ名「表示」および第6のグループ名「突起」であったとする。

【0010】

この際、ユーザが第2のグループ名である「操作」という名称が分かりにくいと感じ、このグループ名に対応するグループに属する特許文書群の内容を見たところ、「画面操作」という名称の方が適切であると判断し、現在の「操作」というグループ名を、「画面操作」というグループ名に人手で変更するというケースがこれにあたる。

10

【0011】

しかし、ユーザは文書グループに名称を付与する際、分類結果全体から相対的に判断するのではなく、各グループに予め付けられた名称や、名称変更対象のグループに含まれる文書群の特徴にのみ依拠した名称を付けようとする傾向がある。このため、例えば、他のグループにも数多く含まれるキーワードを付けてしまう等、変更後の名称が分類結果全体で見た場合に適切な名称となっていない場合がある。

【0012】

前述した第1乃至第6のグループを例にとると、第4のグループの「モバイル」という名称がそのままでは分かりにくいいため、当該第4のグループ属する特許群を見たところ、「携帯電話」という名称の方が適切である事が分かり、現在の「モバイル」という名称を、「携帯電話」という名称に人手で変更したとする。

20

【0013】

しかし、分類対象となっている遠隔通信関連特許が、実はほとんどが携帯電話関連の特許であったとすると、「携帯電話」という名称を付けることは、他のグループに対して識別性があるとはいえず、適切な名称ではない。このような場合、従来 of 各種手法では、グループにユーザが付与する名称の妥当性を判定することは出来なかった。

【0014】

そこで、本発明の目的は、文書群を分類したグループに対する識別情報の人為的な付与の妥当性を判定することが可能になる文書処理装置を提供することにある。

30

【課題を解決するための手段】**【0015】**

すなわち、本発明に係わる文書処理装置は、複数の文書のそれぞれをグループ識別情報が付与される複数のグループのいずれかに分類した文書分類情報を管理する管理手段と、前記管理される文書分類情報を出力する分類情報出力手段と、当該出力された文書分類情報におけるグループ識別情報のうち変更対象のグループ識別情報に対する変更先のグループ識別情報の入力を受け付ける入力手段と、前記管理されるグループごとに、当該グループに属する文書数に対する前記入力された変更先のグループ識別情報の出現文書数の割合を計算する計算手段と、前記計算手段による計算結果をもとに、前記変更対象のグループ識別情報の前記変更先のグループ識別情報への変更が適切であるか否かを判定する判定手段と、前記変更が適切であると前記判定手段により判定した場合に、前記管理手段により管理される変更対象のグループ識別情報を前記入力された変更先のグループ識別情報に変更する変更手段と、前記変更が適切でない前記判定手段により判定した場合に当該判定結果を出力する判定結果出力手段とを備えたことを特徴とする。

40

【発明の効果】**【0016】**

本発明によれば、文書群を分類したグループに対する識別情報の人為的な付与の妥当性を判定することができる。

50

【発明を実施するための最良の形態】**【0017】**

以下図面により本発明の実施形態について説明する。

図1は、本発明の実施形態における文書処理装置の構成例を示すブロック図である。

図1に示すように、本発明の実施形態における文書処理装置は、装置全体の処理動作を司る制御部1、記憶装置2、表示制御部3、グループ名変更処理部4、示唆情報提供判定部5、入力装置6、表示装置7を備え、それぞれがバス8により相互に接続される。

入力装置6は、例えばキーボードやマウスであり、表示装置7は例えば液晶ディスプレイ装置である。

【0018】

記憶装置2は、例えば不揮発性メモリやHDD(Hard Disk Drive)などの記憶媒体であり、表示制御部3、グループ名変更処理部4、示唆情報提供判定部5による処理動作のためのプログラムを記憶する他、文書分類結果記憶部21、文書内容記憶部22、示唆情報提供ルール記憶部23を有する。

【0019】

文書分類結果記憶部21は、複数の文書の分類結果である文書分類結果情報を記憶する。図2は、本発明の実施形態における文書処理装置の記憶装置の文書分類結果記憶部に記憶される文書分類情報の構成例を表形式で示す図である。

図2に示すように、記憶装置2の文書分類結果記憶部21に記憶される文書分類結果情報では、文書群でなるグループに付与されたグループ識別情報であるグループ名、当該グループ名に対応するグループに分類された文書の文書名および当該グループに属する文書群の件数がグループごとに示される。

【0020】

ここでは、文書は特許文書であるとし、「特許1」～「特許20」までの20の特許文書のそれぞれが、「管理」、「操作」、「液晶」、「モバイル」、「押下」、「突起」の6つの名称が自動的に付与されたグループのいずれかに自動的に分類されているものとする。なお、自動分類の手法やグループへの名称付与の手法そのものは既知の手法が使われているものとし、ここでは説明を略する。

【0021】

文書内容記憶部22は、文書分類結果情報で示された複数の文書のそれぞれの記載内容を示す文書内容情報を記憶する。図3は、本発明の実施形態における文書処理装置の記憶装置の文書内容記憶部に記憶される文書内容情報の構成例を表形式で示す図である。

文書内容記憶部22に記憶される文書内容情報では、文書名、および当該文書名に対応する文書の中身を示す文書内容が文書ごとに示される。図3に示した例では、文書名「特許1」には「携帯電話における管理画面を...」という文書内容が記載されており、同様に「特許2」～「特許20」までの各文書名に対応する文書内容が記載されている。

【0022】

なお、実際の特許文書には、「出願人」、「実施例」、「図面の簡単な説明」など、多様な情報が記載されているが、ここでは簡便のため特許文書の要約文の一部を文書内容として記憶しているものとする。

【0023】

記憶装置2の示唆情報提供ルール記憶部23は、入力装置6による入力内容にしたがったグループ名の変更が妥当ではないことを示すための示唆情報をユーザに提供するかどうか、および示唆情報を提供する場合どのような情報をユーザに提供するかを定めたルールである示唆情報提供ルールを記憶する。

【0024】

グループ名変更処理部4は、入力装置6による入力内容を元に、文書分類結果情報における特定のグループのグループ名を変更する。

示唆情報提供判定部5は、入力装置6による入力内容にしたがったグループ名の変更が妥当であるか、つまり適切であるかどうか、もしくはグループ名の変更が妥当ではなく示

10

20

30

40

50

唆情報をユーザに提供すべきかを示唆情報提供ルール記憶部 2 3 に記憶された示唆情報提供ルールをもとに判定する。

【 0 0 2 5 】

表示制御部 3 は、文書分類結果画面表示制御部 3 1、文書内容画面表示制御部 3 2、グループ名変更用画面表示制御部 3 3、示唆情報画面表示制御部 3 4 を有する。

文書分類結果画面表示制御部 3 1 は、文書分類結果記憶部 2 1 に記憶された文書分類結果情報を表示装置 7 に表示することでユーザに提示する。本実施形態では、各種の表示処理は Web ブラウザを用いてなされるものとする。文書内容画面表示制御部 3 2 は、記憶装置 2 の文書内容記憶部 2 2 に記憶された文書内容情報を表示装置 7 に表示することでユーザに提示する。

10

【 0 0 2 6 】

グループ名変更用画面表示制御部 3 3 はグループ名を変更するための画面を表示装置 7 に表示する。

示唆情報画面表示制御部 3 4 は、示唆情報を提供することが適切であると示唆情報提供判定部 5 によって判定された場合、記憶装置 2 の示唆情報提供ルール記憶部 2 3 に記憶された示唆情報提供ルールに従って示唆情報を表示装置 7 に表示させることでユーザに提供する。

【 0 0 2 7 】

次に、図 1 に示した構成の文書処理装置の動作について説明する。

図 4 は、本発明の実施形態における文書処理装置の処理動作の一例を示すフローチャートである。

20

まず、表示制御部 3 の文書分類結果画面表示制御部 3 1 は、記憶装置 2 の文書分類結果記憶部 2 1 に記憶される文書分類情報をもとに、文書分類結果画面を表示装置 7 に表示する（ステップ S 1）。

【 0 0 2 8 】

図 5 は、本発明の実施形態における文書処理装置の表示装置に表示される文書分類結果画面の第 1 の例を示す図である。

図 5 に示すように、文書分類結果画面では、各グループのグループ名、当該グループ名に対応するグループに分類される文書の件数、および後述するグループ名変更画面に遷移するための、「グループ名変更画面へ」とラベリングされたグループ名変更画面遷移ボタンが表示される。

30

【 0 0 2 9 】

また、入力装置 6 により、文書分類結果画面上のグループ名が選択されると（ステップ S 2 の YES）、表示制御部 3 の文書内容画面表示制御部 3 2 は、文書分類結果記憶部 2 1 に記憶される文書分類結果情報と文書内容記憶部 2 2 に記憶される文書内容情報とを元に、文書内容画面を提示する（ステップ S 3）。

【 0 0 3 0 】

図 6 は、本発明の実施形態における文書処理装置の表示装置に表示される文書内容画面の一例を示す図である。

図 6 に示した文書内容画面は、図 5 に示した文書分類結果画面上で、グループ名「モバイル」にマウスポインタ（矢印マーク）をあわせてマウスによるクリックがなされた場合の当該クリックされたグループ名「モバイル」とそのグループに含まれる文書である特許文書名、および各特許文書の文書内容がリスト状に表示したものである。

40

【 0 0 3 1 】

また、前述した文書分類結果画面において、変更対象のグループ名に対応して設けられる「グループ名変更画面遷移ボタン」が入力装置 6 により選択されると（ステップ S 2 ステップ S 4）、表示制御部 3 のグループ名変更用画面表示制御部 3 3 は、グループ名を変更するための画面であるグループ名変更用画面を表示装置 7 に表示する（ステップ S 5）。

図 7 は、本発明の実施形態における文書処理装置の表示装置に表示される文書分類結果

50

画面の第 2 の例を示す図である。

【 0 0 3 2 】

図 8 は、本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更画面の第 1 の例を示す図である。

図 8 に示したように、グループ名変更画面は、今のグループ名と変更先のグループ名の入力欄、および変更する名称を決定するための、「グループ名変更」とラベリングされたグループ名変更ボタンから構成される。

【 0 0 3 3 】

図 8 に示したグループ名変更画面は、図 7 に示した文書分類結果画面上のグループ名「操作」の行にある、「グループ名変更画面へ」とラベリングされたグループ名変更画面遷移ボタンが入力装置 6 により選択された場合に表示される画面であり、現在のグループ名である「操作」と、現在のグループ名に代わるグループ名である変更先のグループ名を入力するための空の入力欄とが表示されている。

10

【 0 0 3 4 】

図 9 は、本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更画面の第 2 の例を示す図である。

ここではユーザはグループ名称を「画面操作」に変更したいとする。具体的にはユーザは、図 9 に示すように、入力装置 6 により変更先のグループ名を入力するための入力欄に当該変更先のグループ名「画面操作」を入力してグループ名変更ボタンを選択する（ステップ S 6）。

20

【 0 0 3 5 】

すると、示唆情報提供判定部 5 は、記憶装置 2 の示唆情報提供ルール記憶部 2 3 に記憶された示唆情報提供ルールに基づき、示唆情報を提供するかどうか、提供する場合はどのような示唆情報を提供するかを判定する（ステップ S 7）。

【 0 0 3 6 】

ここで、ステップ S 7 の処理の詳細について説明する。図 1 0 は、本発明の実施形態における文書処理装置による示唆情報提供判定処理の一例を示すフローチャートである。図 1 1 は、本発明の実施形態における文書処理装置の記憶装置の示唆情報提供ルール記憶部に記憶される示唆情報提供ルールの一例を示す図である。図 1 2 は、本発明の実施形態における文書処理装置の記憶装置の示唆情報提供ルール記憶部に記憶される示唆情報テンプレートの一例を示す図である。

30

【 0 0 3 7 】

本実施形態では、示唆情報を生成するために、示唆情報提供ルールおよび当該ルールに対応する所定の雛形（テンプレート）を用いており、図 1 2 に示すように、2 種類の示唆情報テンプレートが示唆情報提供ルール記憶部 2 3 に記憶されている。

【 0 0 3 8 】

まず、示唆情報提供判定部 5 は、記憶装置 2 の文書内容記憶部 2 2 に記憶される文書内容情報で示される文書の一つを選択し、当該文書中にユーザが入力した変更先のグループ名が含まれているか否かを調べる（ステップ S 7 1）。

【 0 0 3 9 】

示唆情報提供判定部 5 は、ステップ S 7 1 の処理を文書内容情報で示される全文書について繰り返して終了したら（ステップ S 7 2 の Y E S）、各グループのうち 1 つを選択し、当該選択したグループに属する全文書のうち実際に変更先のグループ名が含まれる文書の数、つまり変更先のグループ名の出現文書頻度をカウントする（ステップ S 7 3）。そして、示唆情報提供判定部 5 は、選択済みのグループにおける「変更先のグループ名の出現文書頻度 / グループに属する全文書数」D を計算する（ステップ S 7 4）。

40

【 0 0 4 0 】

示唆情報提供判定部 5 は、ステップ S 7 3、S 7 4 の処理をグループについて繰り返して終了したら（ステップ S 7 5 の Y E S）、もしステップ S 7 4 の処理で計算した D が 0 . 5 を越えるグループが全グループ数の半数以上ある場合は（ステップ S 7 6 の Y E S）

50

、図12に示したテンプレートAを用いて示唆情報を生成し、この生成した示唆情報を示す示唆情報画面を表示装置7に表示させる(ステップS78, S8 S9)。

【0041】

また、示唆情報提供判定部5は、ステップS74の処理で計算したDが0.5を超えるグループが全グループの半分未満しかなく(ステップS76のNO)、かつ、ステップS74の処理で計算したDが0.5を超えるグループがグループ名を変更するグループ以外に2グループ以下存在する場合は(ステップS77のYES)、図12に示したテンプレートBを用いて示唆情報を生成し、この生成した示唆情報を示す示唆情報画面を表示装置7に表示させる(ステップS79, S8 S9)。

【0042】

また、示唆情報提供判定部5は、ステップS74の処理で計算したDが0.5を超えるグループが全グループの半分未満しかなく(ステップS76のNO)、かつステップS74の処理で計算したDが0.5を超えるグループがグループ名を変更するグループ以外に存在しない、もしくは3グループ以上存在する場合は(ステップS77のNO)、示唆情報を提供しないと判定し、グループ名変更処理部4は、変更対象のグループ名を前述のように入力された変更先グループ名に変更する(ステップS8 S9, S10)。

【0043】

以下、ステップS7以降の処理の具体例を説明する。

図13は、本発明の実施形態における文書処理装置が管理する各文書における変更先グループ名の出現分布の一例を表形式で示す図である。

図13に示した出現分布はステップS71の処理結果であり、変更先のグループ名が「画面操作」である場合の各文書における出現有無、変更先のグループ名が「携帯電話」である場合の各文書における出現有無、および変更先のグループ名が「ボタン」である場合の各文書における出現有無を纏めて示したものである。

【0044】

ここでは「特許1」～「特許20」までの各文書のうち文言「画面操作」は、「特許5」、「特許6」、「特許13」、「特許17」に含まれており、文言「携帯電話」は、「特許1」、「特許2」、「特許4」～「特許7」、「特許9」～「特許11」、「特許13」～「特許15」、「特許17」、「特許19」、「特許20」に含まれており、文言「ボタン」は、「特許3」、「特許12」、「特許15」、「特許16」、「特許18」～「特許20」に含まれている。

【0045】

第1の例として、図5に示した文書分類結果画面上のグループ名「操作」の変更先のグループ名としてグループ名「画面操作」が入力された場合の処理を説明する。

図14は、本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第1のグループ名の出現文書頻度の割合の計算結果の一例を示す図である。

この処理では、図2に示した文書分類結果の状況、および図13に示す、文書内容における特定の単語の出現分布により、「グループに属する全文書件数」、「画面操作」の出現文書の頻度、「画面操作」の出現文書頻度/グループに属する全文書数」Dはそれぞれ図14のように算出される。

【0046】

この場合、計算した値Dが0.5を超えているグループがグループ数全体、つまり6グループのうちグループ名「操作」に対応するグループのみ、つまり1つしかなく、グループ数全体の半分に満たないため、図11に示した示唆情報提供ルール(1)は満たさない。

【0047】

また、計算した値Dが0.5を超えているグループが、名称を変更しようとするグループであるグループ名「操作」に対応するグループ以外に存在しないため、図11に示した示唆情報提供ルール(2)も満たさない。このため、図11に示した示唆情報提供

10

20

30

40

50

ルールの項目(3)が適用され、示唆情報は提供されず、「操作」から「画面操作」へのグループ名変更を行うため、グループ名変更処理部4が呼び出される。

【0048】

グループ名変更処理部4は、前述したグループ名「操作」を前述したように入力されたグループ名「画面操作」に変更する。具体的には、グループ名変更処理部4は、まず、文書分類結果記憶部21に記憶される文書分類結果情報のグループ名「操作」を新しいグループ名「画面操作」に書き換える。

図15は、本発明の実施形態における文書処理装置の記憶装置の文書分類結果記憶部に記憶される文書分類情報のグループ名変更後の構成例を表形式で示す図である。

【0049】

すると、図2に示した文書分類結果情報に記憶されている「操作」は図15のように「画面操作」に書き換えられる。その後、表示制御部3の文書分類結果画面表示制御部31は、図15に示したグループ名変更後の文書分類結果情報を元に新たな文書分類結果画面を表示装置7に表示する。

図16は、本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更後の文書分類結果画面の一例を示す図である。

【0050】

次に、第2の例として、図5に示した文書分類結果画面上のグループ名「モバイル」の変更先のグループ名として、グループ名「携帯電話」が入力された場合の処理を説明する。

図17は、本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第2のグループ名の出現文書頻度の割合の計算結果の一例を示す図である。

ここでは、各グループについての「グループに属する全文書件数」、「携帯電話」の出現文書の頻度、「携帯電話」の出現文書頻度/グループに属する全文書数」Dは図17のようになる。

この場合、計算した値Dが0.5を超えているグループが、グループ数全体の6グループのうちグループ名「押下」に対応するグループ以外の全て、つまり5グループにもものほり、グループ数全体の半分以上を超えている。よって、図11に示した示唆情報提供ルールの項目(1)に合致することになり、「所定のテンプレートAを用いて示唆情報を提示する」ことになる。

【0051】

すると、示唆情報提供判定部5は、テンプレートAを用いて、ユーザが入力した変更先のグループ名は他のグループにも高頻度で出現している言葉であって、変更先のグループ名として不適切であり、それでもなお変更を行なうかを問う示唆情報を提示するという情報、および当該テンプレートA中の「変更しようとするグループ名」つまり変更対象のグループ名が「携帯電話」であることを示す情報を表示制御部3の示唆情報画面表示制御部34に送る。

【0052】

示唆情報画面表示制御部34は、示唆情報提供判定部5からの情報を元に示唆情報を生成し、この情報を示す示唆情報画面を表示装置7に表示する。

図18は、本発明の実施形態における文書処理装置の表示装置に表示される示唆情報画面の第1の例を示す図である。

【0053】

ここでは示唆情報画面は図18のようになる。図18に示した例では、ユーザが入力した変更先のグループ名「携帯電話」は他のグループにも高頻度で出現している言葉であり、変更先のグループ名として不適切であることを明示している。

【0054】

この画面において、「はい」と表記されたアイコンが選択されると、グループ名変更処理部4による変更対象のグループ名の変更先のグループ名への変更処理がなされ、「いい

10

20

30

40

50

え」と表記されたアイコンが選択されると、前述のように入力された変更先のグループ名がクリアされて、元の文書分類結果画面が表示装置 7 に表示される。

【 0 0 5 5 】

次に、第 3 の例として、図 5 に示した文書分類結果画面上のグループ名「押下」の変更先のグループ名として、グループ名「ボタン」が入力された場合の処理を説明する。

図 1 9 は、本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第 3 のグループ名の出現文書頻度の割合の計算結果の一例を示す図である。

ここでは、各グループについての「グループに属する全文書件数」、「「ボタン」の出現文書の頻度」、「「ボタン」の出現文書頻度/グループに属する全文書数」D は図 1 9 のようになる。

【 0 0 5 6 】

この場合、計算した値 D が 0 . 5 を超えているグループは、グループ数全体である 6 グループのうちグループ名「押下」に対応するグループおよびグループ名「突起」に対応するグループ、つまり 2 グループしかいないため、図 1 1 に示した示唆情報提供ルールの項目 (1) には当てはまらない。

【 0 0 5 7 】

しかし、計算した値 D が 0 . 5 を超えているグループとして、変更対象のグループ名「押下」に対応するグループ以外に、グループ名「突起」のグループが存在する。よって、変更対象のグループ名以外における、値 D が 0 . 5 を超えているグループの個数が 2 以下であるため、図 1 1 に示した示唆情報提供ルールの項目 (2) を満たすことになり「所定のテンプレート B を用いて示唆情報を提示する」ことになる。

【 0 0 5 8 】

すると、示唆情報提供判定部 5 は、テンプレート B を用いて、ユーザが入力した変更先のグループ名は、他のグループ名、つまり変更対象のグループ名に対応するグループ以外における値 D が 0 . 5 を超えているグループにも高頻度で出現している言葉であり、これらのグループをマージ (合併) することを勧める示唆情報を提示するという情報、および当該テンプレート B 中の「変更しようとするグループ名」が「押下」であることを示す情報、およびテンプレート B 中の「他の該当グループ名 (複数ある場合は「 , 」区切りですべて記載)」が「突起」であることを示す情報を表示制御部 3 の示唆情報画面表示制御部 3 4 に送る。

示唆情報画面表示制御部 3 4 は、示唆情報提供判定部 5 からの情報を元に示唆情報を生成し、この情報を示す示唆情報画面を表示装置 7 に表示する。

【 0 0 5 9 】

図 2 0 は、本発明の実施形態における文書処理装置の表示装置に表示される示唆情報画面の第 2 の例を示す図である。

ここでは示唆情報画面は図 2 0 のようになる。図 2 0 に示した例では、ユーザが入力した変更先のグループ名「ボタン」は、他のグループ名「突起」に対応するグループにも高頻度で出現している言葉であり、これらのグループをマージしてはどうかとユーザに提案している。

【 0 0 6 0 】

この画面において、「はい」と表記されたアイコンが選択されると、グループ名変更処理部 4 による複数のグループのマージ処理がなされ、「いいえ」と表記されたアイコンが選択されると、前述のように入力された変更先のグループ名がクリアされて、元の文書分類結果画面が表示装置 7 に表示される。

この画面において「いいえ」と表記されたアイコンが選択された場合は、グループ名変更処理部 4 による変更対象のグループ名の変更先のグループ名への変更処理を行うようにしてもよい。

【 0 0 6 1 】

以上のように、本発明の実施形態における文書処理装置では、ユーザは複数のグループ

10

20

30

40

50

のグループ名を変更する際、自分が入力した変更先のグループ名が適切な名称なのかどうかを事前にユーザに確認させることができるので、文書群を分類した各グループに対し、ユーザが適切な名称を付与することを支援することができ、結果として各グループに適切な名称が付与されることになり、後に、ユーザは分類結果を用いた計算機システム上での各種作業を実行しやすくなる。

【0062】

なお、本実施形態に記載の手段や装置、部位については、本発明を実現するものであればどのような実装形態であってもよい。例えば各種処理はCPU (Central Processing Unit) 上で実行されるとしてもよい。また、本実施形態では、画面はWebブラウザを用いて表示されることを前提として説明を行ったが、同様の情報を表現可能であればWebブラウザ以外の表示手段によって表示されるとしてもよい。

10

【0063】

また、文書分類結果や文書内容についても本実施の形態以外の構造、名称、内容であってもよい。また、格納されている文書分類結果は、自動分類結果でなくてもよい。文書種別も特許文書以外の文書であってもよい。また、グループ名についても、本実施形態では単語1語であらわしたが、複数の単語や文章であるとしてもよい。

【0064】

また、各表示画面については、ここまで示した以外の表示形態、表示内容であるとしてもよい。さらには、複数の画面を組み合わせ、あるいは一つの画面を分割して複数の画面で表示するとしてもよい。

20

【0065】

また、示唆情報提供ルールは本実施形態に記載のルール以外のものであるとしてもよい。例えば、前述したテンプレートA、Bの文言が異なってもよく、図11に示したルール(1)、(2)の閾値が0.5以外の値、例えば0.3や0.7などであってもよく、図11の示したルール(2)の閾値が2つ以下以外の値(例えば3つ以下など)であってもよい。さらには、「特定のグループ名に変えようとする場合は示唆情報を出す」などの別の示唆情報提供ルール、および当該特定の不適切なグループ名を記憶装置2に記憶しておき、入力済みの変更先のグループ名が当該特定のグループ名に該当する場合には、当該グループ名が変更先のグループ名として適切でないことをしめす示唆情報を表示装置7に表示する形態であってもよい。さらには、前記のようにグループ名を複数の単語や文章で付与できる場合には、そこに含まれる複数の単語それぞれについて、本実施形態に記載の処理を実施するとしてもよい。

30

【0066】

なお、この発明は前記実施形態そのままに限定されるものではなく実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、前記実施形態に開示されている複数の構成要素の適宜な組み合わせにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を省略してもよい。更に、異なる実施形態に亘る構成要素を適宜組み合わせてもよい。

【図面の簡単な説明】

【0067】

40

【図1】本発明の実施形態における文書処理装置の構成例を示すブロック図。

【図2】本発明の実施形態における文書処理装置の記憶装置の文書分類結果記憶部に記憶される文書分類情報の構成例を表形式で示す図。

【図3】本発明の実施形態における文書処理装置の記憶装置の文書内容記憶部に記憶される文書内容情報の構成例を表形式で示す図。

【図4】本発明の実施形態における文書処理装置の処理動作の一例を示すフローチャート。

【図5】本発明の実施形態における文書処理装置の表示装置に表示される文書分類結果画面の第1の例を示す図。

【図6】本発明の実施形態における文書処理装置の表示装置に表示される文書内容画面の

50

一例を示す図。

【図 7】本発明の実施形態における文書処理装置の表示装置に表示される文書分類結果画面の第 2 の例を示す図。

【図 8】本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更画面の第 1 の例を示す図。

【図 9】本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更画面の第 2 の例を示す図。

【図 10】本発明の実施形態における文書処理装置による示唆情報提供判定処理の一例を示すフローチャート。

【図 11】本発明の実施形態における文書処理装置の記憶装置の示唆情報提供ルール記憶部に記憶される示唆情報提供ルールの一例を示す図。

【図 12】本発明の実施形態における文書処理装置の記憶装置の示唆情報提供ルール記憶部に記憶される示唆情報テンプレートの一例を示す図。

【図 13】本発明の実施形態における文書処理装置が管理する各文書における変更先グループ名の出現分布の一例を表形式で示す図。

【図 14】本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第 1 のグループ名の出現文書頻度の割合の計算結果の一例を示す図。

【図 15】本発明の実施形態における文書処理装置の記憶装置の文書分類結果記憶部に記憶される文書分類情報のグループ名変更後の構成例を表形式で示す図。

【図 16】本発明の実施形態における文書処理装置の表示装置に表示されるグループ名変更後の文書分類結果画面の一例を示す図。

【図 17】本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第 2 のグループ名の出現文書頻度の割合の計算結果の一例を示す図。

【図 18】本発明の実施形態における文書処理装置の表示装置に表示される示唆情報画面の第 1 の例を示す図。

【図 19】本発明の実施形態における文書処理装置におけるグループごとの当該グループに属する文書数に対する変更先の第 3 のグループ名の出現文書頻度の割合の計算結果の一例を示す図。

【図 20】本発明の実施形態における文書処理装置の表示装置に表示される示唆情報画面の第 2 の例を示す図。

【符号の説明】

【0068】

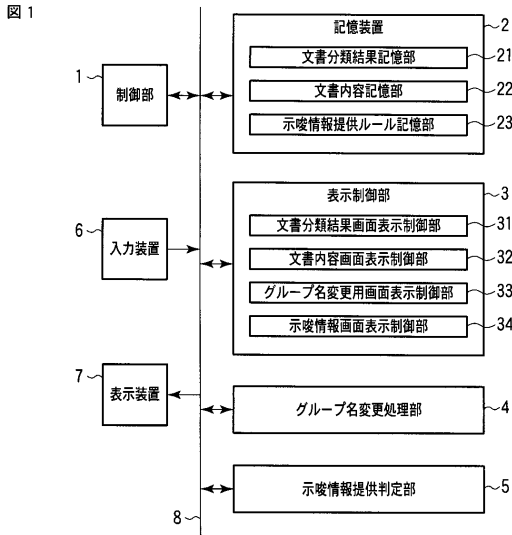
1 ... 制御部、2 ... 記憶装置、3 ... 表示制御部、4 ... グループ名変更処理部、5 ... 示唆情報提供判定部、6 ... 入力装置、7 ... 表示装置、8 ... バス、21 ... 文書分類結果記憶部、22 ... 文書内容記憶部、23 ... 示唆情報提供ルール記憶部、31 ... 文書分類結果画面表示制御部、32 ... 文書内容画面表示制御部、33 ... グループ名変更用画面表示制御部、34 ... 示唆情報画面表示制御部。

10

20

30

【 図 1 】



【 図 2 】

図 2

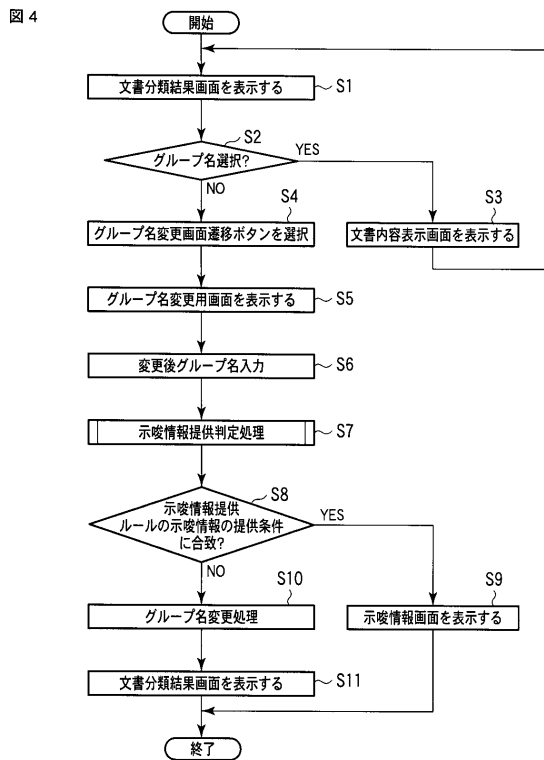
グループ名	分類された文書名	件数
管理	特許1、特許2、特許3、特許4	4件
操作	特許5、特許6	2件
液晶	特許7、特許8、特許9	3件
モバイル	特許10、特許11、特許12、特許13、特許14	5件
押下	特許15、特許16	2件
突起	特許17、特許18、特許19、特許20	4件

【 図 3 】

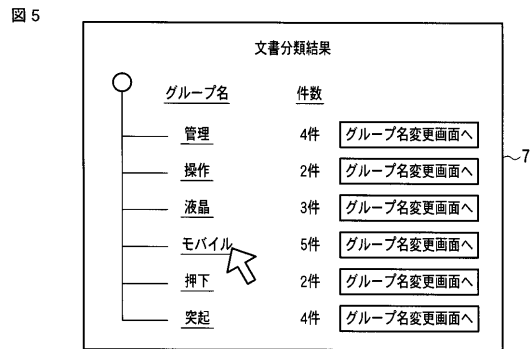
図 3

文書名	文書内容
特許1	携帯電話における管理画面を…
…	
特許10	モバイルシステム、特に携帯電話における…
特許11	携帯電話やノートPCといったモバイル機器に共通の…
特許12	電波が届きにくい地下で、モバイルフォンを正確に…
…	

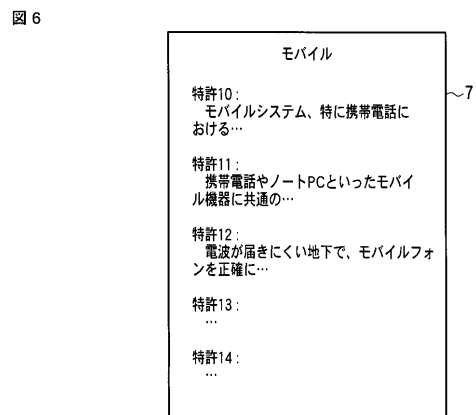
【 図 4 】



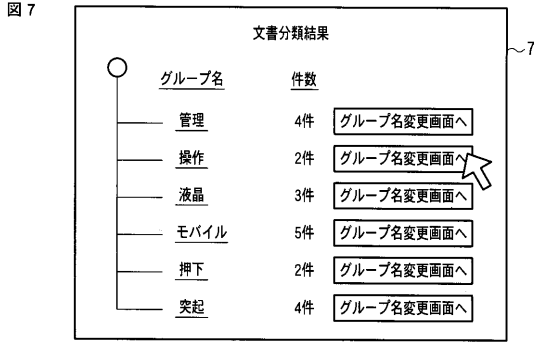
【 図 5 】



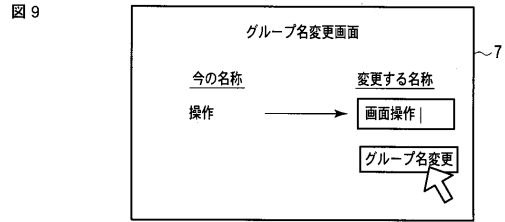
【 図 6 】



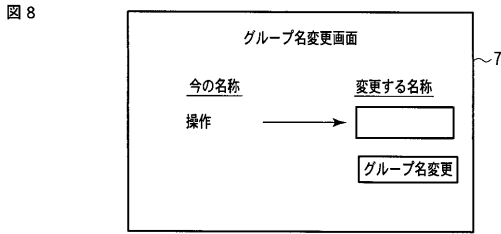
【 図 7 】



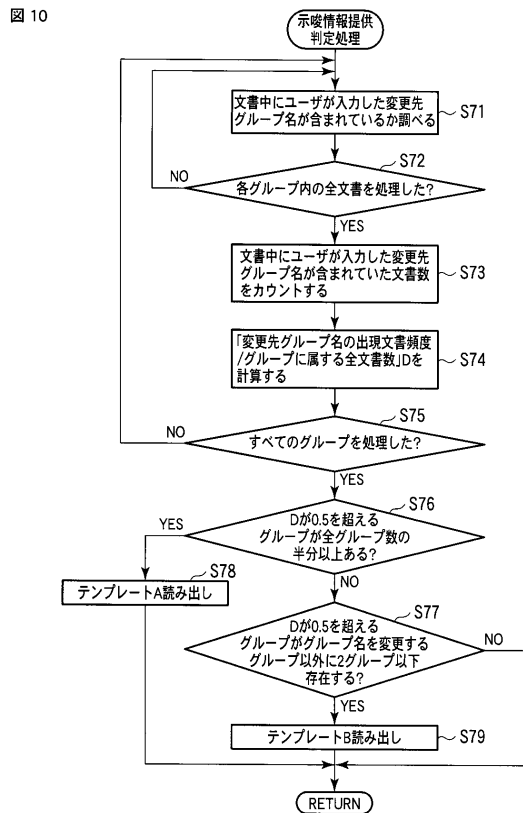
【 図 9 】



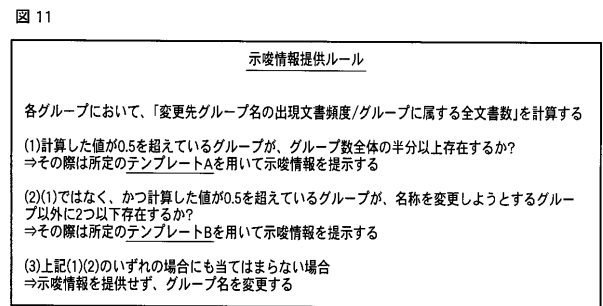
【 図 8 】



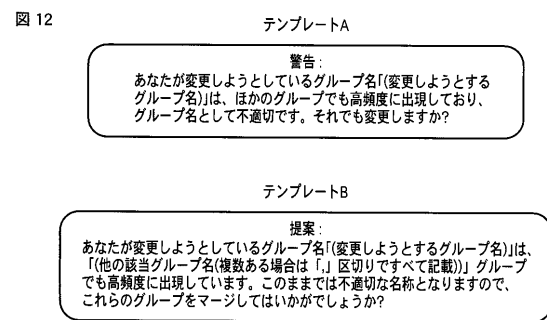
【 図 1 0 】



【 図 1 1 】



【 図 1 2 】



【 図 1 3 】

図 13

文書名	「画面操作」の有無	「携帯電話」の有無	「ボタン」の有無
特許1		○	
特許2		○	
特許3			○
特許4		○	
特許5	○	○	
特許6	○	○	
特許7		○	
特許8			
特許9		○	
特許10		○	
特許11		○	
特許12			○
特許13	○	○	
特許14		○	
特許15		○	○
特許16			○
特許17	○	○	
特許18			○
特許19		○	○
特許20		○	○

【 図 1 4 】

図 14

(「操作」⇒「画面操作」の場合)

グループ名	グループに属する全文書件数	「画面操作」の出現文書頻度	「画面操作」の出現文書頻度 / グループに属する全文書件数
管理	4件	0件	0/4=0
操作	2件	2件	2/2=1
液晶	3件	0件	0/3=0
モバイル	5件	1件	1/5=0.2
押下	2件	0件	0/2=0
突起	4件	1件	1/4=0.25

【 図 1 5 】

図 15

グループ名	分類された文書名	件数
管理	特許1、特許2、特許3、特許4	4件
画面操作	特許5、特許6	2件
液晶	特許7、特許8、特許9	3件
モバイル	特許10、特許11、特許12、特許13、特許14	5件
押下	特許15、特許16	2件
突起	特許17、特許18、特許19、特許20	4件

【 図 1 6 】

図 16

文書分類結果

グループ名	件数	
管理	4件	グループ名変更画面へ
画面操作	2件	グループ名変更画面へ
液晶	3件	グループ名変更画面へ
モバイル	5件	グループ名変更画面へ
押下	2件	グループ名変更画面へ
突起	4件	グループ名変更画面へ

【 図 1 8 】

図 18

示唆情報

警告:
あなた変更しようとしているグループ名「携帯電話」は、ほかのグループでも高頻度に出現しており、グループ名として不適切です。それでも変更しますか?

【 図 1 7 】

図 17

(「モバイル」⇒「携帯電話」の場合)

グループ名	グループに属する全文書件数	「携帯電話」の出現文書頻度	「携帯電話」の出現文書頻度 / グループに属する全文書件数
管理	4件	3件	3/4=0.75
操作	2件	2件	2/2=1
液晶	3件	2件	2/3=0.66
モバイル	5件	4件	4/5=0.8
押下	2件	1件	1/2=0.5
突起	4件	3件	3/4=0.75

【 図 1 9 】

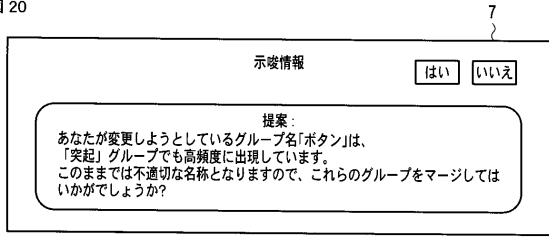
図 19

(「押下」⇒「ボタン」の場合)

グループ名	グループに属する全文書件数	「ボタン」の出現文書頻度	「ボタン」の出現文書頻度 / グループに属する全文書件数
管理	4件	1件	1/4=0.25
操作	2件	0件	0/2=0
液晶	3件	0件	0/3=0
モバイル	5件	1件	1/5=0.2
押下	2件	2件	2/2=1
突起	4件	3件	3/4=0.75

【図 20】

図 20



フロントページの続き

- (74)代理人 100109830
弁理士 福原 淑弘
- (74)代理人 100075672
弁理士 峰 隆司
- (74)代理人 100095441
弁理士 白根 俊郎
- (74)代理人 100084618
弁理士 村松 貞男
- (74)代理人 100103034
弁理士 野河 信久
- (74)代理人 100119976
弁理士 幸長 保次郎
- (74)代理人 100153051
弁理士 河野 直樹
- (74)代理人 100140176
弁理士 砂川 克
- (74)代理人 100100952
弁理士 風間 鉄也
- (74)代理人 100101812
弁理士 勝村 紘
- (74)代理人 100070437
弁理士 河井 将次
- (74)代理人 100124394
弁理士 佐藤 立志
- (74)代理人 100112807
弁理士 岡田 貴志
- (74)代理人 100111073
弁理士 堀内 美保子
- (74)代理人 100134290
弁理士 竹内 将訓
- (74)代理人 100127144
弁理士 市原 卓三
- (74)代理人 100141933
弁理士 山下 元
- (72)発明者 平 博司
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- (72)発明者 宮部 泰成
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- (72)発明者 加納 敏行
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- (72)発明者 松本 茂
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- (72)発明者 後藤 和之
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- (72)発明者 董 嘉挺
東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内
- Fターム(参考) 5B075 NR02