

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4683963号
(P4683963)

(45) 発行日 平成23年5月18日 (2011.5.18)

(24) 登録日 平成23年2月18日 (2011.2.18)

(51) Int.Cl.

F I

G O 6 F 17/21 (2006.01)

G O 6 F 17/21 5 7 O R

G O 6 F 17/24 (2006.01)

G O 6 F 17/24 5 5 4 G

請求項の数 9 (全 24 頁)

(21) 出願番号 特願2005-60022 (P2005-60022)
 (22) 出願日 平成17年3月4日 (2005.3.4)
 (65) 公開番号 特開2006-244199 (P2006-244199A)
 (43) 公開日 平成18年9月14日 (2006.9.14)
 審査請求日 平成19年12月21日 (2007.12.21)

(73) 特許権者 000002369
 セイコーエプソン株式会社
 東京都新宿区西新宿2丁目4番1号
 (74) 代理人 100095728
 弁理士 上柳 雅誉
 (74) 代理人 100107261
 弁理士 須澤 修
 (73) 特許権者 504180239
 国立大学法人信州大学
 長野県松本市旭三丁目1番1号
 (74) 代理人 100095728
 弁理士 上柳 雅誉
 (74) 代理人 100107076
 弁理士 藤網 英吉

最終頁に続く

(54) 【発明の名称】 電子ファイル書き込みシステム、プログラム、電子ファイルへの書き込み情報反映方法

(57) 【特許請求の範囲】

【請求項1】

1 以上のテキスト群を含む電子ファイルに書き込みを行うための情報端末と、当該情報端末に接続され、前記電子ファイルを記憶するサーバと、から構成される電子ファイル書き込みシステムであって、

前記情報端末は、

前記電子ファイルに書き込みを行う書き込み手段と、

前記書き込み手段によって書き込まれた画像データ、テキストデータ、および音声データの少なくともいずれか1のデータから成るメモ書き情報、または、前記テキスト群に含まれる文字のマーキングに関するマーキング情報を含む書き込み情報と、当該書き込み情報の前記電子ファイル上の位置を示す位置情報と、を含む書き込み詳細情報を前記サーバへ送信する送信手段と、を備え、

前記サーバは、

受信した前記書き込み詳細情報と、前記位置情報に基づいて生成した当該書き込み情報が属するテキスト群を特定するためのテキスト群特定情報と、前記マーキングが行われた文字を特定するための文字特定情報と、を記憶する記憶手段と、

記憶した前記書き込み情報を、前記テキスト群特定情報に基づいて、前記電子ファイルに反映させる書き込み情報反映手段と、を備え、

前記記憶手段は、前記記憶された複数のテキスト群の順序に関する順序情報及び前記文字特定情報をさらに記憶し、

10

20

前記書き込み情報反映手段は、

前記書き込み情報がメモ書き情報であって、前記電子ファイルの更新によって、前記書き込み情報が属するテキスト群が削除された場合、前記順序情報に基づいて、前記書き込み情報が属するテキスト群の直前または直後に位置するテキスト群に前記書き込み情報を反映させ、

前記書き込み情報がマーキング情報の場合、前記テキスト群特定情報に基づいて、前記マーキングが行われた文字が属するテキスト群を特定した後、記憶した前記文字特定情報に基づいて、前記マーキングが行われた文字をさらに特定し、前記書き込み情報を反映させることを特徴とする電子ファイル書き込みシステム。

【請求項 2】

前記サーバは、

前記書き込み情報記憶時の電子ファイルに含まれるテキスト群である旧テキスト群、および前記書き込み情報反映時の電子ファイルに含まれるテキスト群である新テキスト群、の全てについて、それらの特徴量であるテキスト群特徴量を算出するテキスト群特徴量算出手段、をさらに備え、

前記記憶手段は、前記テキスト群特定情報として、前記旧テキスト群のテキスト群特徴量および前記新テキスト群のテキスト群特徴量を記憶し、

前記書き込み情報反映手段は、

前記各旧テキスト群のテキスト群特徴量および前記各新テキスト群のテキスト群特徴量に基づいて、対応するテキスト群を特定し、前記書き込み情報を反映させることを特徴とする請求項 1 に記載の電子ファイル書き込みシステム。

【請求項 3】

前記サーバは、

前記電子ファイルを更新する更新手段と、

記憶した全テキスト群の前記テキスト群特徴量を用いて、前記各旧テキスト群と、前記各新テキスト群との対応付けを行うテキスト群マッチング手段と、

前記各旧テキスト群のテキスト群特徴量と、前記テキスト群マッチング手段によって当該各旧テキスト群に対応付けられた、前記各新テキスト群のテキスト群特徴量との差異が、所定の閾値未満であるかを判定するテキスト群判定手段と、をさらに備え、

前記電子ファイルが更新された場合、

前記テキスト群特徴量算出手段は、前記新テキスト群のテキスト群特徴量として、更新時の前記電子ファイルに含まれるテキスト群のテキスト群特徴量を算出し、

前記記憶手段は、前記新テキスト群のテキスト群特徴量として、書き込み情報反映直前の更新時における前記テキスト群特徴量を記憶し、

前記書き込み情報反映手段は、

前記テキスト群判定手段によって、前記両テキスト群特徴量の差異が前記所定の閾値未満であると判定された、旧テキスト群と新テキスト群の組み合わせを、対応するテキスト群として特定し、前記書き込み情報を反映させることを特徴とする請求項 2 に記載の電子ファイル書き込みシステム。

【請求項 4】

前記旧テキスト群に含まれる、句読点および / または改行位置で区切られる 1 以上の文字列である旧文字列、および当該旧テキスト群に対応する新テキスト群に含まれる前記 1 以上の文字列である新文字列、の全てについて、それらの特徴量である文字列特徴量を算出する文字列特徴量算出手段をさらに備え、

前記記憶手段は、前記文字特定情報として、前記各旧文字列の文字列特徴量および前記各新文字列の文字列特徴量を記憶し、

前記書き込み情報反映手段は、

前記各旧文字列の文字列特徴量および前記各新文字列の文字列特徴量に基づいて、対応する文字列を特定し、前記書き込み情報を反映させることを特徴とする請求項 2 に記載の電子ファイル書き込みシステム。

10

20

30

40

50

【請求項 5】

記憶した全文字列の前記文字列特徴量を用いて、前記各旧文字列と、前記各新文字列との対応付けを行う文字列マッチング手段と、

前記各旧文字列の文字列特徴量と、前記文字列マッチング手段によって当該各旧文字列に対応付けられた、前記各新文字列の文字列特徴量との差異が、所定の閾値未満であるかを判定する文字列判定手段と、をさらに備え、

前記書き込み情報反映手段は、

前記文字列判定手段によって、前記両文字列特徴量の差異が前記所定の閾値未満であると判定された、旧文字列と新文字列の組み合わせを、対応する文字列として特定し、前記書き込み情報を反映させることを特徴とする請求項 4 に記載の電子ファイル書き込みシステム。

10

【請求項 6】

前記書き込み情報反映手段は、

前記マーキングが行われた文字が属する前記文字列を特定した後、前記テキストデータに基づいて、前記マーキングが行われた文字をさらに特定し、前記書き込み情報を反映させることを特徴とする請求項 4 または 5 に記載の電子ファイル書き込みシステム。

【請求項 7】

前記記憶手段は、前記マーキングの種別を示す情報に関するマーキング種別情報をさらに記憶することを特徴とする請求項 1 ないし 6 のいずれか 1 項に記載の電子ファイル書き込みシステム。

20

【請求項 8】

コンピュータを、請求項 1 ないし 7 のいずれか 1 項に記載の電子ファイル書き込みシステムにおける各手段として機能させるためのプログラム。

【請求項 9】

1 以上のテキスト群を含む電子ファイルに書き込みを行うための情報端末と、当該情報端末に接続され、前記電子ファイルを記憶するサーバと、から構成されるネットワークシステムの電子ファイルへの書き込み情報反映方法であって、

前記情報端末が、

前記電子ファイルに書き込みを行う書き込みステップと、

前記書き込みステップによって書き込まれた画像データ、テキストデータ、および音声データの少なくともいずれか 1 のデータから成る メモ書き情報、または、前記テキスト群に含まれる文字のマーキングに関するマーキング情報を含む書き込み情報と、当該書き込み情報の前記電子ファイル上の位置を示す位置情報と、を含む書き込み詳細情報を前記サーバへ送信する送信ステップと、を実行し、

30

前記サーバが、

受信した前記書き込み詳細情報と、前記位置情報に基づいて生成した当該書き込み情報が属するテキスト群を特定するためのテキスト群特定情報と、前記マーキングが行われた文字を特定するための文字特定情報と、を記憶する記憶ステップと、

記憶した前記書き込み情報を、前記テキスト群特定情報に基づいて、前記電子ファイルに反映させる書き込み情報反映ステップと、を実行し、

40

前記記憶ステップにおいて、前記記憶された複数のテキスト群の順序に関する順序情報及び前記文字特定情報をさらに記憶し、

前記書き込み情報反映ステップは、

前記書き込み情報がメモ書き情報であって、前記電子ファイルの更新によって、前記書き込み情報が属するテキスト群が削除された場合、前記順序情報に基づいて、前記書き込み情報が属するテキスト群の直前または直後に位置するテキスト群に前記書き込み情報を反映させ、

前記書き込み情報がマーキング情報の場合、前記テキスト群特定情報に基づいて、前記マーキングが行われた文字が属するテキスト群を特定した後、記憶した前記文字特定情報に基づいて、前記マーキングが行われた文字をさらに特定し、前記書き込み情報を反映さ

50

せることを特徴とする電子ファイルへの書き込み情報反映方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、電子ファイルに書き込みを行うための、電子ファイル書き込みシステム、プログラム、電子ファイルへの書き込み情報反映方法に関するものである。

【背景技術】

【0002】

近年、インターネットやイントラネットを介して閲覧されるWebファイル等の電子ファイルは、単なる情報提供に限らず様々な利用方法が知られている。その一例として、電子教育システム(e-ラーニングシステム)が挙げられる(例えば特許文献1参照)。この電子教育システムとは、教材となる電子ファイルを、ユーザ(生徒)がインターネット等を利用して閲覧し、学習を進める自主自習型の教育システムである。

【0003】

この教育システムに依れば、図や動画等を用いたより効果的なマルチメディア教材の提供が可能であるだけでなく、大勢の生徒を一箇所に集合させて講義を行う必要がないので、コストを大幅に削減することが実現可能であるとともに、ユーザ(生徒)にとっても自身の都合に合わせて学習を進めることができるといった利点がある。さらに、多数の生徒に教材となる印刷物を配布する必要がなく、また、教材の内容変更に伴う更新も容易であることから、教材の作成コストを大幅に軽減させることができるので、従来の講義形式で学習を進める場合にも、この電子教育システムが導入されつつある。

【特許文献1】特開平11-282826号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかし、ユーザが学習を進めるうえで、教材である電子ファイルに対して、メモ書きやマーキングといった書き込み情報の書き込みを行う場合には、通常教材をプリントアウトし、プリントアウトした出力物に対して、メモ書きやマーキングを行うといった煩雑な作業を行っている。また、教材である電子ファイル自体に、メモ書きやマーキングを書き込むことができる場合でも、電子ファイルを表示する表示層、書き込みを行うための書き込み層との2層を構成し、これらを特定のサイズで固定的に表示しているため、電子ファイルの表示サイズは常に固定される。従って、ユーザは電子ファイルの表示サイズを所望のサイズに変更できない。さらに、教材の内容が変更され、新たに更新された場合には、書き込み情報の位置を特定することが困難であり、書き込み情報を反映することができない。

【0005】

本発明は、上記の問題に鑑み、電子ファイルのテキスト文書に含まれる任意のテキスト群に書き込まれた書き込み情報を、対応するテキスト群に反映させることができる電子ファイル書き込みシステム、プログラム、電子ファイルへの書き込み情報反映方法を提供することをその課題とする。

【課題を解決するための手段】

【0006】

本発明の電子ファイル書き込みシステムは、1以上のテキスト群を含む電子ファイルに書き込みを行うための情報端末と、当該情報端末に接続され、電子ファイルを記憶するサーバと、から構成される電子ファイル書き込みシステムであって、情報端末は、電子ファイルに書き込みを行う書き込み手段と、書き込み手段によって書き込まれた画像データ、テキストデータ、および音声データの少なくともいずれか1のデータから成るメモ書き情報、または、テキスト群に含まれる文字のマーキングに関するマーキング情報を含む書き込み情報と、当該書き込み情報の電子ファイル上の位置を示す位置情報と、を含む書き込み詳細情報をサーバへ送信する送信手段と、を備え、サーバは、受信した書き込み詳細情

報と、位置情報に基づいて生成した当該書き込み情報が属するテキスト群を特定するためのテキスト群特定情報と、マーキングが行われた文字を特定するための文字特定情報と、を記憶する記憶手段と、記憶した書き込み情報を、テキスト群特定情報に基づいて、電子ファイルに反映させる書き込み情報反映手段と、を備え、
前記記憶手段は、記憶された複数のテキスト群の順序に関する順序情報及び文字特定情報をさらに記憶し、書き込み情報反映手段は、書き込み情報がメモ書き情報であって、電子ファイルの更新によって、書き込み情報が属するテキスト群が削除された場合、順序情報に基づいて、書き込み情報が属するテキスト群の直前または直後に位置するテキスト群に前記書き込み情報を反映させ、書き込み情報がマーキング情報の場合、テキスト群特定情報に基づいて、マーキングが行われた文字が属するテキスト群を特定した後、記憶した文字特定情報に基づいて、マーキングが行われた文字をさらに特定し、書き込み情報を反映させることを特徴とする。

10

【0007】

また、本発明の電子ファイルへの書き込み情報反映方法は、1以上のテキスト群を含む電子ファイルに書き込みを行うための情報端末と、当該情報端末に接続され、電子ファイルを記憶するサーバと、から構成されるネットワークシステムの電子ファイルへの書き込み情報反映方法であって、情報端末が、電子ファイルに書き込みを行う書き込みステップと、書き込みステップによって書き込まれた画像データ、テキストデータ、および音声データの少なくともいずれか1のデータから成るメモ書き情報、または、テキスト群に含まれる文字のマーキングに関するマーキング情報を含む書き込み情報と、当該書き込み情報の電子ファイル上の位置を示す位置情報と、を含む書き込み詳細情報をサーバへ送信する送信ステップと、を実行し、サーバが、受信した書き込み詳細情報と、位置情報に基づいて生成した当該書き込み情報が属するテキスト群を特定するためのテキスト群特定情報と、マーキングが行われた文字を特定するための文字特定情報と、を記憶する記憶ステップと、記憶した書き込み情報を、テキスト群特定情報に基づいて、電子ファイルに反映させる書き込み情報反映ステップと、を実行し、記憶ステップにおいて、記憶された複数のテキスト群の順序に関する順序情報及び文字特定情報をさらに記憶し、書き込み情報反映ステップは、書き込み情報がメモ書き情報であって、電子ファイルの更新によって、書き込み情報が属するテキスト群が削除された場合、順序情報に基づいて、書き込み情報が属するテキスト群の直前または直後に位置するテキスト群に書き込み情報を反映させ、書き込み情報がマーキング情報の場合、テキスト群特定情報に基づいて、マーキングが行われた文字が属するテキスト群を特定した後、記憶した文字特定情報に基づいて、マーキングが行われた文字をさらに特定し、書き込み情報を反映させることを特徴とする。

20

30

【0008】

これらの構成によれば、テキスト群を特定するための情報に基づいて、例えばWebファイルといった電子ファイル上に書き込まれた書き込み情報が属するテキスト群を特定し、書き込み情報を反映させることができるので、表示画面の表示サイズが変更されても、対応するテキスト群に正しく書き込み情報を反映させることができる。

40

【0009】

この場合、サーバは、書き込み情報記憶時の電子ファイルに含まれるテキスト群である旧テキスト群、および書き込み情報反映時の電子ファイルに含まれるテキスト群である新テキスト群、の全てについて、それらの特徴量であるテキスト群特徴量を算出するテキスト群特徴量算出手段、をさらに備え、記憶手段は、テキスト群特定情報として、旧テキスト群のテキスト群特徴量および新テキスト群のテキスト群特徴量を記憶し、書き込み情報反映手段は、各旧テキスト群のテキスト群特徴量および各新テキスト群のテキスト群特徴量に基づいて、対応するテキスト群を特定し、書き込み情報を反映させることが好ましい。

【0010】

50

この構成によれば、電子ファイルに含まれるテキスト群全てについて、テキスト群のテキスト群特定情報、すなわちテキスト群の特徴量を算出し、算出したその特徴量に基づいて、電子ファイルに書き込まれた書き込み情報を反映するので、例えば、テキスト群の順序が入れ替わったような場合でも、書き込み情報が属するテキスト群を正しく特定し、書き込み情報を反映させることが可能である。

【 0 0 1 1 】

また、この場合、サーバは、電子ファイルを更新する更新手段と、記憶した全テキスト群のテキスト群特徴量を用いて、各旧テキスト群と、各新テキスト群との対応付けを行うテキスト群マッチング手段と、各旧テキスト群のテキスト群特徴量と、テキスト群マッチング手段によって当該各旧テキスト群に対応付けられた、各新テキスト群のテキスト群特徴量との差異が、所定の閾値未満であるかを判定するテキスト群判定手段と、をさらに備え、電子ファイルが更新された場合、テキスト群特徴量算出手段は、新テキスト群のテキスト群特徴量として、更新時の前記電子ファイルに含まれるテキスト群のテキスト群特徴量を算出し、記憶手段は、新テキスト群のテキスト群特徴量として、書き込み情報反映直前の更新時におけるテキスト群特徴量を記憶し、書き込み情報反映手段は、テキスト群判定手段によって、両テキスト群特徴量の差異が所定の閾値未満であると判定された、旧テキスト群と新テキスト群の組み合わせを、対応するテキスト群として特定し、書き込み情報を反映させることが好ましい。

【 0 0 1 2 】

この構成によれば、マッチングの結果、対応付けられた2つのテキスト群の特徴量の差が所定の閾値未満である組み合わせを、対応するテキスト群として特定する、すなわち一定以上の類似度を有するテキスト群の組み合わせを、対応するテキスト群として特定するので、たとえ電子ファイルが更新され、その内容に若干の変更が加えられた場合にも、更新前のテキスト群に対して、対応するテキスト群を正しく特定し、書き込み情報を反映することが可能となる。

【 0 0 1 3 】

さらに、記憶手段が、書き込み情報保存時のテキスト群特徴量、および書き込み情報反映直前の更新時における、テキスト群特徴量、を記憶しているので、サーバは、情報端末が複数の場合でも、各情報端末からの電子ファイル送信要求の度に、テキスト群特徴量を算出しなくてよい。特に情報端末の数が数十台、数百台に達する場合等には、かなりの負荷を軽減することができる。また、書き込み情報を反映する直前に算出されたテキスト群特徴量を新テキスト群特徴量として記憶するので、書き込み情報が反映されるまでの間に、電子ファイルが複数回に亘って更新された場合にも、その更新毎に算出された全てのテキスト群特徴量を記憶する必要がなくメモリ容量を節約できる。

【 0 0 1 5 】

この構成によれば、電子ファイルに書き込まれたメモ書き情報が属するテキスト群が、電子ファイルの更新によって削除された場合にも、メモ書き情報を反映させることができる。さらに、更新前にメモ書き情報が属していたテキスト群の、直前または直後に位置するテキスト群に対してメモ書き情報を反映するので、別のテキスト群にメモ書き情報を反映したことによる、ユーザが抱く虞のある違和感を可能な限り減少させ得る。また、メモ書き情報が画像や音声である場合には、ユーザが電子ファイルに描画した画像や貼り付けた音声をそのまま再現するので、ユーザによる書き込み処理の自由度をより高めることができる。

【 0 0 1 6 】

また、これらの場合、書き込み情報には、テキスト群に含まれる文字のマーキングに関するマーキング情報が含まれ、記憶手段は、マーキングが行われた文字を特定するための文字特定情報をさらに記憶し、書き込み情報反映手段は、テキスト群特定情報に基づいて、マーキングが行われた文字が属するテキスト群を特定した後、記憶した文字特定情報に基づいて、マーキングが行われた文字をさらに特定し、書き込み情報を反映させることが好ましい。

10

20

30

40

50

【 0 0 1 7 】

この構成によれば、文字列を特定するための情報に基づいて、電子ファイルに書き込まれた書き込み情報（マーキング情報）が属する文字列を特定し、書き込み情報を反映させることができるので、ユーザが表示画面の表示サイズに規制されることなく、マーキング情報を電子ファイルに書き込むことができる。

【 0 0 1 8 】

この場合、旧テキスト群に含まれる、句読点および／または改行位置で区切られる１以上の文字列である旧文字列、および当該旧テキスト群に対応する新テキスト群に含まれる前記１以上の文字列である新文字列、の全てについて、それらの特徴量である文字列特徴量を算出する文字列特徴量算出手段をさらに備え、記憶手段は、文字特定情報として、各旧文字列の文字列特徴量および各新文字列の文字列特徴量を記憶し、書き込み情報反映手段は、各旧文字列の文字列特徴量および各新文字列の文字列特徴量に基づいて、対応する文字列を特定し、書き込み情報を反映させることが好ましい。

10

【 0 0 1 9 】

この構成によれば、算出した文字列特徴量に基づいて、文字列の対応付けを行うことによって、マーキング情報が属している旧文字列に対応する新文字列を特定し、マーキング情報を反映させるので、たとえ、テキスト群を構成する文字列の順序が入れ替わったような場合でも、マーキング情報が属する文字列を特定し、反映させることが可能である。但しこの場合、マーキング情報は文字列単位で反映される。

【 0 0 2 0 】

20

また、この場合、記憶した全文字列の文字列特徴量を用いて、各旧文字列と各新文字列との対応付けを行う文字列マッチング手段と、各旧文字列の文字列特徴量と、文字列マッチング手段によって当該各旧文字列に対応付けられた、各新文字列の文字列特徴量との差異が、所定の閾値未満であるかを判定する文字列判定手段と、をさらに備え、書き込み情報反映手段は、文字列判定手段によって、両文字列特徴量の差異が所定の閾値未満であると判定された、旧文字列と新文字列の組み合わせを対応する文字列として特定し、書き込み情報を反映させることが好ましい。

【 0 0 2 1 】

この構成によれば、マッチングの結果、対応付けられた２つの文字列の特徴量の差が所定の閾値未満である組み合わせを、対応する文字列として特定する、すなわち一定以上の類似度を有する文字列の組み合わせを、対応する文字列として特定するので、たとえ電子ファイルが更新され、その内容に若干の変更が加えられた場合にも、更新前の文字列に対して、対応する文字列を正しく特定し、書き込み情報を反映することが可能となる。

30

【 0 0 2 2 】

これらの場合、文字特定情報には、マーキングが行われた文字を示すテキストデータがさらに含まれ、書き込み情報反映手段は、マーキングが行われた文字が属する文字列を特定した後、テキストデータに基づいて、マーキングが行われた文字をさらに特定し、書き込み情報を反映させることが好ましい。

【 0 0 2 3 】

この構成によれば、たとえマーキングが行われた文字が、句読点および／または改行位置で区切られる文字列に満たない文字（例えば単語）であるときにも、対応する文字列の中から、マーキングが行われた文字を特定して書き込み情報を反映させることができる。

40

【 0 0 2 4 】

また、これらの場合、記憶手段は、マーキングの種別を示す情報に関するマーキング種別情報をさらに記憶することが好ましい。

【 0 0 2 5 】

この構成によれば、直線、波線、二重線、下線、点線など多様なマーキングを電子ファイルに書き込み、反映させることが可能となる。

【 0 0 2 6 】

本発明のプログラムは、コンピュータを、上記のいずれか１に記載の電子ファイル書き

50

込みシステムにおける各手段として機能させることを特徴とする。

【 0 0 2 7 】

この構成によれば、電子ファイルのテキスト文書に含まれる任意のテキスト群に書き込まれた書き込み情報を、対応するテキスト群に反映させるためのプログラムを提供することができる。

【発明を実施するための最良の形態】

【 0 0 2 8 】

以下、添付の図面を参照して、本発明の電子ファイル書き込みシステム、プログラム、電子ファイルへの書き込み情報反映方法について、詳細に説明する。本発明の電子ファイル書き込みシステム等は、例えばインターネットやイントラネットを介して閲覧するWebファイル等の電子ファイルを教材とし、ユーザ（生徒）がこの電子ファイルを用いて学習を進める形態のe-ラーニングシステムに用いられるような、電子ファイルにユーザがメモ書きやマーキングといった書き込みを行うためのシステムである。

【 0 0 2 9 】

本実施形態に係る電子ファイル書き込みシステム1は、図1に示すように、電子ファイルEF（図3参照）および書き込み情報81（図2参照）を記憶するサーバ50と、このサーバ50にネットワーク150を介して接続され、電子ファイルEFに書き込みを行うための情報端末となるクライアントPC（以下単にPCと記載）100（図では3台表示）と、から構成されている。

【 0 0 3 0 】

PC100は、WWWブラウザ106（図2参照）を備え、利用者がWWWブラウザ106を介して電子ファイルEFを閲覧しようとする、サーバ50へ電子ファイルEFの送信を要求する。そして、その要求に対してサーバ50から送信された電子ファイルEFを受信し、ディスプレイ109（図2参照）に表示する。さらに、表示している電子ファイルEFに対して、マーキングやメモ書きといった書き込みを行うための書き込み手段120と、ユーザにより書き込み手段120を用いて書き込まれた書き込み情報81、および書き込み情報81の電子ファイル上の位置を示す位置情報82、を含む書き込み詳細情報80（いずれも図2参照）を取得し、サーバ50へ送信する送信手段130と、を有している。なお、書き込み手段120は、図2に示すキーボード110並びにマウス111によってその主要部が構成されている。

【 0 0 3 1 】

サーバ50は、電子ファイルEFを記憶する電子ファイルデータベースDB1と、PC100から受信した書き込み詳細情報80を記憶する書き込み情報データベースDB2と、を有している。また、サーバ50は、データ処理部60（書き込み情報反映手段）を有し、データ処理部60は、PC100からの電子ファイルEF送信要求に応じて、書き込み情報81を電子ファイルEFに反映させるための様々な処理を行い、ネットワークインタフェース（NT-I/F）54（図2参照）と共同して書き込み情報81を反映した電子ファイルEFをPC100に送信する。

【 0 0 3 2 】

続いて、図2を参照してサーバ50およびPC100のハードウェア構成について説明する。サーバ50は、ネットワーク150とのデータ送受信部であるネットワークインタフェース（NT-I/F）54、サーバ50で実行される各種処理を制御するCPU51、処理プログラム等を記憶するROM52、ワークエリアとなるRAM53、および電子ファイルデータベースDB1、書き込み情報データベースDB2等を記憶するハードディスク55を備え、これらは互いにシステムバス56を介して互いに接続されている。なお、ハードディスク55にはOS57がインストールされている。また、書き込み情報データベースDB2には、PC100から受信した書き込み詳細情報80（書き込み情報81、位置情報82）に加え、位置情報82に基づいて生成したテキスト群特定情報90および文字特定情報95（詳細については後述する）が記憶されている。

【 0 0 3 3 】

さらに、サーバ50は、電子ファイルEFを更新するための更新手段70（図1参照）を有しており、電子ファイルEFの内容が変更され更新された電子ファイルEFを、元のファイルに上書きして記憶しておくことが可能である。また、サーバ50のハードディスク55には、書き込み情報81を電子ファイルEFに反映させるための様々な処理を行うための、書き込み情報反映プログラムPがインストールされている。

【0034】

一方、PC100は、ネットワーク150とのデータ送受信部であるネットワークインタフェース（NT-I/F）104、PC100で実行される各種処理を制御するCPU101、処理プログラム等を記憶するROM102、ワークエリアとなるRAM103、WWWブラウザ106を含む各種アプリケーションプログラム107、およびOS108がインストールされているハードディスク105、電子ファイルEFを閲覧するためのディスプレイ109、閲覧している電子ファイルEFに対して、ユーザが書き込みや保存の操作を行うためのキーボード110、およびマウス111を備え、これらは互いにシステムバス112を介して接続されている。

【0035】

さらに、ユーザが電子ファイルEFに対して行う、メモ書きやマーキングといった書き込み操作について図3ないし図6を参照して説明する。図3に示すように、電子ファイルEFには、例えば空白行Wで区切られる1以上のテキスト群T（T1～T4）が含まれ、さらに各テキスト群Tには、句読点および/または改行位置で区切られる、1以上の文字列Sが含まれている（本実施形態では句点で区切られる文字列とする）。また、電子ファイルEF上部には、書き込み情報を保存するための保存ボタン200が配置されており、利用者は、メモ書きやマーキングといった書き込みを行った後、この保存ボタン200を選択することによって、書き込み情報を保存する。

【0036】

図4は、ユーザが電子ファイルEFにメモ書きを書き込む操作手順を示している。同図に示すように、ユーザはまずメモ書きを書き込む所望の位置に、マウス111やキーボード110を用いてカーソルKを合わせ、例えば右クリック操作等でメニューM1を表示させ、「新規メモ作成」を選択する（同図（a）参照）。「新規メモ作成」を選択したことによって、メモ書き用ボックスMBが表示され、メモ書き用ボックスMB内にメモ書きを書き込む（同図（b）参照）。最後に、書き込んだメモ書きを保存するため、保存ボタン200を選択する（同図（c）参照）。

【0037】

図5は、ユーザが電子ファイルEFに含まれる文字をマーキングするマーキング操作手順を示している。同図に示すように、まずマーキングを行う所望の位置に、マウス111やキーボード110を用いてカーソルKを合わせ、例えば右クリック操作等でメニューM1を表示させ、「新規マーキング作成」を選択する（同図（a）参照）。そして、マーキングを行う範囲指定を行う。例えばマウス111のドラッグ操作で行う場合、マーキングを行う範囲の開始位置でクリックし（同図（b）参照）そのままマウスをドラッグしてマーキングを行う範囲の終了位置でドラッグを終了する（同図（c）参照）。

【0038】

なお、この操作に限らずマーキングを行う範囲を指定できる操作なら何でもよい。最後に、書き込んだマーキングを保存するため、保存ボタン200を選択する（同図（d）参照）。このような手順で、ユーザがメモ書きやマーキングといった書き込みを電子ファイルEFに対して行い、次回ユーザが電子ファイルEFを閲覧すると、図6に示すように、電子ファイルEFに対してメモ書きやマーキングといった書き込み情報81が反映されて表示される。また、書き込み情報81を複数書き込んだ場合、複数の書き込み情報81を一括して保存するようにしてもよい。

【0039】

また、ユーザが書き込み情報を保存すると、PC100は、書き込み情報81および書き込み情報81の電子ファイルEF上の位置を示す位置情報82を取得し、書き込み詳細

10

20

30

40

50

情報 80 としてサーバ 50 へ送信する。なお、「書き込み情報」とは、書き込みがメモ書きの場合、「メモ書きされた画像データ、テキストデータ、および音声データの少なくともいずれか 1 のデータ」であり、書き込みがマーキングの場合、「マーキングされた文字のテキストデータ」である。また、「位置情報」とは、書き込みがメモ書きの場合、「メモ書きが属するテキスト群 T」に関する情報であり、書き込みがマーキングの場合、「マーキングされた文字が属するテキスト群 T、およびマーキングされた文字が属する文字列 S」に関する情報である。

【0040】

さらに、書き込み情報 81 には、書き込まれたメモ書きやマーキングのフォントに関するフォント情報やカラーに関するカラー情報が含まれている。なお、書き込みがメモ書きの場合、本実施形態では「メモ書きが属するテキスト群 T」内の位置情報については特に記憶していないが、もちろん「メモ書きが属するテキスト群 T」内の位置情報を記憶する構成でもよい。この場合、位置情報はマーキングと同様に、「メモ書きが属するテキスト群 T、およびメモ書きの作成位置として指定された位置が含まれる文字列 S」に関する情報である。

【0041】

また、サーバ 50 は、PC 100 から受信した書き込み詳細情報 80 を書き込み情報データベース DB2 に記憶する際、受信した書き込み詳細情報 80 に含まれる位置情報 82 に基づいて、書き込まれたメモ書きが属するテキスト群 T を特定するためのテキスト群特定情報 90 (図 2 参照)、およびマーキングされた文字を特定するための文字特定情報 95 (図 2 参照)、を生成し、書き込み情報データベース DB2 にさらに記憶する。

【0042】

ここで、テキスト群特定情報 90 および文字特定情報 95 の生成について、詳細に説明する。テキスト群特定情報 90 とは、電子ファイル EF に含まれる各テキスト群 T (図 3 参照) の特徴量を意味しており (以下テキスト群特徴量 Tv と記載)、以下、サーバ 50 が行うテキスト群 T のテキスト群特徴量 Tv の算出について述べる (テキスト群特徴量算出手段)。テキスト群特徴量 Tv は、テキスト群 T を形態素解析して得られた各単語の TFIDF 値を要素とする多次元ベクトルである。形態素解析について簡単に説明すると、例えば図 2 に示すハードディスク 55 に格納された日本語辞書 (図示省略) を用いてテキスト群 T を解析し、テキスト群 T を構成する各単語を求めるものである。

【0043】

図 3 に示すテキスト群 T1 の一部である「データリンクには、様々な特徴やアクセス方式がある。」という文章に対して形態素解析を行うと、「データ」「リンク」「に」「は」「様々」「な」「特徴」「や」「アクセス」「方式」「が」「ある」といった単語が得られる。また、TFIDF (Term Frequency Inverse Document Frequency) 値とは、ある単語が、所定のテキストデータ内で出現する回数を、そのテキストデータの量に応じて正規化したものである。

【0044】

本実施形態では、上記の TFIDF 値として、例えばテキスト群 T1 (図 3 参照) に含まれる「データ」という単語の TFIDF 値を求めるとすると、

$TF = \text{テキスト群 T1 に含まれる「データ」の単語数} / \text{テキスト群 T1 に含まれる全ての単語数} \times$

$IDF = \text{LOG (電子ファイル EF に含まれるテキスト群 T の総数 / 「データ」の単語を含むテキスト群 T の数)}$ の計算式で求めることができる。

【0045】

形態素解析によって得られたテキスト群 T1 に含まれる全ての単語について、上記の要領で TFIDF 値を求めることによって、各単語を軸とし、各単語の TFIDF 値を要素とした多次元ベクトルが得られ、テキスト群 T1 のテキスト群特徴量 Tv1 が算出されたことになる。同様にして、この処理を各テキスト群 T について行くと、各テキスト群 T のテキスト群特徴量 Tv を算出することができる。

【 0 0 4 6 】

図 7 は、3 つのテキスト群 $T_a \sim T_c$ のテキスト群特徴量 $T_{va} \sim T_{vc}$ を簡略して示したものである。同図 (a) に示すように、テキスト群 T_a のテキスト群特徴量 T_{va} は、 x_a 、 y_a 、 $z_a \cdots$ を要素とする多次元ベクトルであり、 x_a 、 y_a 、 z_a は、テキスト群 T_a における単語「 x 」、「 y 」、「 z 」のそれぞれの $TFIDF$ 値である。同様に、テキスト群 T_b のテキスト群特徴量 T_{vb} は、 x_b 、 y_b 、 $z_b \cdots$ を要素とする多次元ベクトルであり、 x_b 、 y_b 、 z_b は、テキスト群 T_b における単語「 x 」、「 y 」、「 z 」のそれぞれの $TFIDF$ 値である。

【 0 0 4 7 】

さらに、テキスト群 T_c のテキスト群特徴量 T_{vc} についても、 x_c 、 y_c 、 $z_c \cdots$ を要素とする多次元ベクトルであり、 x_c 、 y_c 、 z_c は、テキスト群 T_c における単語「 x 」、「 y 」、「 z 」のそれぞれの $TFIDF$ 値となる。なお、簡略化のため図示では 3 次元のベクトルとして表示しているが、実際のテキスト群特徴量 T_v は、テキスト群 T に含まれる単語数に応じた多次元空間上にマッピングされる。なお、テキスト群特定情報 90 として、テキスト群特徴量 T_v に代えて、例えば段落番号等を記憶する構成でもよい。

【 0 0 4 8 】

続いて、文字特定情報 95 について説明する。文字特定情報 95 とは、マーキングされた文字を特定するための情報であり、マーキングが行われた文字が属するテキスト群 T に含まれる、句読点や改行位置で区切られる所定の文字列 S の文字列特徴量 S_v を意味している。以下、サーバ 50 が行う文字列 S の文字列特徴量 S_v の算出について述べる (文字列特徴量算出手段)。文字列特徴量 S_v は、テキスト群特徴量 T_v と同様に、文字列 S を形態素解析して得られた各単語の $TFIDF$ 値を要素とする多次元ベクトルである。

【 0 0 4 9 】

従って、例えばテキスト群 T_1 における文字列 S_1 (図 3 参照) に含まれる「データ」という単語の $TFIDF$ 値を求めるとすると、

$$TF = \text{文字列 } S_1 \text{ に含まれる「データ」の単語数} / \text{文字列 } S_1 \text{ に含まれる全ての単語数} \\ \times$$

$$IDF = \text{LOG} (\text{テキスト群 } T_1 \text{ に含まれる文字列 } S \text{ の総数} / \text{「データ」の単語を含む文字列 } S \text{ の数}) \text{ で求めることができる。}$$

【 0 0 5 0 】

形態素解析によって得られた文字列 S に含まれる全ての単語について、上記の要領で $TFIDF$ 値を求めることによって、各単語を軸とし、各単語の $TFIDF$ 値を要素とした多次元ベクトルが得られ、文字列 S_1 の文字列特徴量 S_{v1} が算出されたことになる。同様にして、この処理を各文字列 S について行くと、各文字列 S の文字列特徴量 S_v を算出することができる。なお、 $TFIDF$ 値算出の対象となる単語は、日本語において、文書を意味する大きな部分を担うとされている、「名詞」、「サ変名詞」、「固有名詞」および「形容動詞」である。

【 0 0 5 1 】

テキスト群特徴量 T_v および文字列特徴量 S_v の算出は、書き込み情報の保存時と書き込み情報反映時と、の 2 回のタイミングで行われる。そして、書き込み情報保存時に算出されたテキスト群特徴量 T_v および文字列特徴量 S_v は、旧テキスト群特徴量 T_{ov} および旧文字列特徴量 S_{ov} として記憶され、書き込み情報反映時に算出されたテキスト群特徴量 T_v および文字列特徴量 S_v は、新テキスト群特徴量 T_{nv} および新文字列特徴量 S_{nv} として記憶される。また、書き込み情報が反映される以前に電子ファイル EF が更新された場合には、電子ファイル EF 更新のタイミングで、新テキスト群特徴量 T_{nv} および新文字列特徴量 S_{nv} が算出される。

【 0 0 5 2 】

さらに、その更新が複数回に及んだ場合には、更新の度にそれぞれの特徴量が算出され、前回算出された特徴量に上書きして書き込み情報データベース DB_2 へ記憶される。な

10

20

30

40

50

お、電子ファイルE Fが更新された場合についても、更新毎ではなく書き込み情報反映時に新テキスト群特徴量T n vおよび新文字列特徴量S n vを算出する構成でもよい。

【0053】

続いて、サーバ50が、書き込み情報81を書き込み情報データベースDB2に記憶し、次回電子ファイルE FをPC100に送信する際に、その書き込み情報を電子ファイルE Fに対して反映するまでの流れを、図8のフローチャートを参照して説明する。なお、以下の説明では、書き込み情報記憶時の電子ファイルE Fに含まれるテキスト群Tおよび文字列Sを、旧テキスト群T oおよび旧文字列S oと定義し、書き込み情報反映時の電子ファイルE Fに含まれるテキスト群Tおよび文字列Sを、新テキスト群T nおよび新文字列S nと定義して、説明する。

10

【0054】

まず、ユーザが書き込み情報の保存ボタン200をクリックすると(S01)、PC100は書き込み詳細情報80を取得しサーバ50へ送信する。そして、サーバ50は、上述した手順で、書き込み情報が属するテキスト群Tを始め、電子ファイルE Fに含まれるテキスト群T全てについて、テキスト群特徴量T vを算出する。また、書き込み情報がマーキングである場合、テキスト群特徴量T vに加えて、書き込まれたマーキングが属するテキスト群Tに含まれる、文字列S全ての文字列特徴量S vを算出する(S02)。そして、算出したテキスト群特徴量T vおよび文字列特徴量S vを、旧テキスト群特徴量T o vおよび旧文字列特徴量S o vとして、書き込み情報データベースDB2へ、書き込み詳細情報80とともに記憶する。

20

【0055】

その後、再びPC100においてユーザが電子ファイルE Fの閲覧をしようとする、まずPC100は、電子ファイル送信要求をサーバ50に対して行う(S03)。サーバ50は、PC100からの電子ファイル送信要求を受信すると、まず書き込み情報データベースDB2から、書き込み詳細情報80を取得する。そして、電子ファイルE Fを電子ファイルデータベースDB1から取得し、その電子ファイルE Fに含まれるテキスト群T全てについて、テキスト群特徴量T vを算出する(S04)。そして、算出したテキスト群特徴量T vを新テキスト群特徴量T n vとして、一旦書き込み情報データベースDB2へ記憶する。

【0056】

30

次に、サーバ50は、旧テキスト群特徴量T o vおよび新テキスト群特徴量T n vに基づいて、DPマッチングを含む対応付け処理を行い、各旧テキスト群T oに対して各新テキスト群T nを対応付ける処理を行う(S05)。なお、この処理の詳細については後述する。そして、対応付け処理の結果に基づいて、書き込み情報81が属する旧テキスト群T oに対応する新テキスト群T nを特定する(S06)。対応する新テキスト群T nが特定できた場合(S06: Yes)、書き込みがメモ書きであるときは、メモ書きを電子ファイルE Fの対応付けられたテキスト群Tの所定位置に反映する(S07)。また、書き込みがマーキングであるときは、さらに特定された新テキスト群T nに含まれる、文字列S全ての文字列特徴量S vを算出して(S08)、これを新文字列特徴量S n vとして一旦書き込み情報データベースDB2へ記憶する。そして旧文字列特徴量S o vおよび新文字列特徴量S n vに基づいてDPマッチングを含む対応付け処理を行い、対応するテキスト群Tにおいて、さらに各旧文字列S oに対して各新文字列S nを対応付ける処理を行う(S09)。

40

【0057】

そして、対応付け処理の結果に基づいて、マーキングが属する旧文字列S oに対応する新文字列S nを特定する(S10)。対応する新文字列S nを特定できた場合(S10: Yes)、書き込み情報81として記憶していた、「マーキングされた文字のテキストデータ」に基づいて、マーキングを電子ファイルE Fに対して反映する(S11)。また、対応する新文字列S nが存在しない場合(S10: No)、マーキングを削除する(S12)。

50

【0058】

一方、対応する新テキスト群 T_n が特定できない場合 (S06: No)、書き込み情報 81 を削除する (S13)。そして、記憶していた書き込み情報 81 を全て処理したか否かを判定し (S14)、未処理の書き込み情報 81 がある場合には (S14: No)、次の書き込み情報 81 について処理を続行する。このようにして、書き込み情報 81 の処理を全て終了すると (S14: Yes)、サーバ 50 は、書き込み情報 81 が反映された電子ファイル EF を PC100 へ送信する。

【0059】

なお、対応するテキスト群 T が特定できない場合でも (S06: No)、書き込み情報 81 がメモ書きである場合には、書き込み情報保存時に、電子ファイル上における各テキスト群の順序に関する順序情報をさらに記憶しておき、メモ書きが属していた旧テキスト群 T_o の、直前または直後に位置していたテキスト群 T_o に対応する新テキスト群 T_n に、メモ書きを反映させる構成でもよい。例えば、旧テキスト群 T_o2 にメモ書きが書き込まれ、電子ファイル EF の更新によって旧テキスト群 T_o2 が削除された場合、旧テキスト群 T_o2 の直前に位置していた旧テキスト群 T_o1 に対応する新テキスト群 T_n1 、あるいは旧テキスト群 T_o2 の直後に位置していた旧テキスト群 T_o3 に対応する新テキスト群 T_n3 にメモ書き情報を反映させる。

【0060】

また、書き込み情報 81 を記憶してから、書き込み情報 81 を反映するまでの間に、電子ファイル EF の更新を行っている場合には、電子ファイル更新時に、新テキスト群特徴量 T_nv および新文字列特徴量 S_nv を既に算出し記憶しているので、書き込み情報反映時に行われる新テキスト群特徴量 T_nv および新文字列特徴量 S_nv の算出処理 (S04、S08 に該当) は省略される。

【0061】

ここで、サーバ 50 が行う旧テキスト群 T_o に対する新テキスト群 T_n の対応付け処理について説明する。まず、書き込み情報 81 が保存されてから、書き込み情報 81 が反映されるまでの間に、電子ファイル EF が更新されていない場合について説明する。

【0062】

旧テキスト群 T_o に対する新テキスト群 T_n の対応付けには、最初に DP マッチング等のマッチング処理を実行する (テキスト群マッチング手段)。そして、マッチングによって対応付けられた、旧テキスト群 T_o と新テキスト群 T_n との組み合わせにおいて、旧テキスト群特徴量 T_ov と、新テキスト群特徴量 T_nv との差異が所定の閾値未満であるかどうかを判定する処理を行う (テキスト群判定手段)。電子ファイル EF が更新されていない場合、各旧テキスト群 T_o と各新テキスト群 T_n とにおいて、その内容は全く同一である。従って、当然対応付けられた新テキスト群特徴量 T_nv との差異は所定の閾値未満となり、マッチングによって対応付けられた旧テキスト群 T_o と新テキスト群 T_n との組み合わせを、対応するテキスト群 T として特定する。なお、「DP マッチング」とは、動的計画法によるパターンマッチングを意味し、二つのパターンの要素間の対応付けを行いながら類似度を計算する方法を指すものである。

【0063】

一方、書き込み情報 81 を記憶してから、書き込み情報 81 を反映するまでの間に、電子ファイル EF が更新されていた場合には、更新されなかった場合のように容易に対応するテキスト群 T が定まるとは限らない。そこで、電子ファイル EF が更新された場合について、図 9 ~ 図 12 に示す 4 種の例を挙げて説明する。

【0064】

図 9 に示す例は、書き込み情報記憶時の電子ファイル EF に、旧テキスト群 T_o1 、旧テキスト群 T_o2 、旧テキスト群 T_o3 、旧テキスト群 T_o4 、旧テキスト群 T_o5 の順序から成る 5 つのテキスト群 T が含まれ、旧テキスト群 T_o2 の内容を変更し電子ファイル EF を更新した例を示している。この場合、まず DP マッチング等のマッチング処理を行うと、旧テキスト群 $T_o1 \sim T_o5$ に対してそれぞれ新テキスト群 $T_n1 \sim T_n5$ が対

10

20

30

40

50

応付けられる（図9参照）。

【0065】

そして、マッチングによって対応付けられた、旧テキスト群 T_o と新テキスト群 T_n との組み合わせにおいて、旧テキスト群特徴量 T_{ov} と、新テキスト群特徴量 T_{nv} との差異が所定の閾値未満であるかどうかを判定する処理を行う。旧テキスト群 T_{o1} と新テキスト群 T_{n1} 、および旧テキスト群 $T_{o3} \sim T_{o5}$ と新テキスト群 $T_{n3} \sim T_{n5}$ の組み合わせについては、旧テキスト群特徴量 T_{ov} と、新テキスト群特徴量 T_{nv} との差異は所定の閾値未満であり、対応するテキスト群 T として特定される。

【0066】

しかし、旧テキスト群 T_{o2} の内容は変更され、新テキスト群 T_{n2} の内容とは異なることから、旧テキスト群 T_{o2} の旧テキスト群特徴量 T_{ov} と新テキスト群 T_{n2} の新テキスト群特徴量 T_{nv} とは異なるものとなる。例えば、図7(a)に示す電子ファイルEFにおいて、テキスト群 T_b の内容が変更され、テキスト群 T_b になった場合、テキスト群 T_b のテキスト群特徴量 T_{vb} は、図7(b)に示す通りとなる。

【0067】

よってこの場合、旧テキスト群 T_{o2} の旧テキスト群特徴量 T_{ov} と新テキスト群 T_{n2} の新テキスト群特徴量 T_{nv} との差異が所定の閾値未満である場合には、旧テキスト群 T_{o2} に対応するテキスト群を新テキスト群 T_{n2} として特定する。また、旧テキスト群 T_{o2} の旧テキスト群特徴量 T_{ov} と新テキスト群 T_{n2} の新テキスト群特徴量 T_{nv} との差異が所定の閾値以上である場合、旧テキスト群 T_{o2} に対応するテキスト群 T はないものと判定する。

【0068】

図10に示す例は、書き込み情報記憶時の電子ファイルEFに、旧テキスト群 T_{o1} 、旧テキスト群 T_{o2} 、旧テキスト群 T_{o3} 、旧テキスト群 T_{o4} 、旧テキスト群 T_{o5} の順序から成る5つのテキスト群 T が含まれ、旧テキスト群 T_{o2} を削除して電子ファイルEFを更新した例を示している。この場合、マッチングを行った結果の一例として、図10に示すとおり、新テキスト群 T_{n1} に対して、旧テキスト群 T_{o1} および旧テキスト群 T_{o2} の2つのテキスト群 T が対応付けられる。なお、旧テキスト群 $T_{o3} \sim T_{o5}$ に対しては、それぞれ新テキスト群 $T_{n3} \sim T_{n5}$ が対応付けられ、対応するテキスト群 T として特定される。

【0069】

このように、1のテキスト群 T に対して複数のテキスト群 T が対応付けられた場合、これらのテキスト群 T 全ての組み合わせにおいて、旧テキスト群特徴量 T_{ov} と新テキスト群特徴量 T_{nv} との差異を判定し、そして、その差異が所定の閾値未満となった組み合わせを、対応するテキスト群 T として特定する。図10に示す例の場合、旧テキスト群 T_{o1} と新テキスト群 T_{n1} 、および旧テキスト群 T_{o2} と新テキスト群 T_{n1} 、の2つの組み合わせにおいて、旧テキスト群特徴量 T_{ov} と新テキスト群特徴量 T_{nv} との差異を判定する。

【0070】

旧テキスト群 T_{o1} と新テキスト群 T_{n1} との内容は同一であり当然特徴量の差は所定の閾値未満となる。従って、旧テキスト群 T_{o1} と新テキスト群 T_{n1} の組み合わせを対応するテキスト群 T として特定する。なお、テキスト群特徴量 T_v の差異が所定の閾値未満となった組み合わせが複数存在する場合には、テキスト群特徴量 T_v の差異が最小となる組み合わせを対応するテキスト群 T として特定する。

【0071】

図11に示す例は、書き込み情報記憶時の電子ファイルEFに、旧テキスト群 T_{o1} 、旧テキスト群 T_{o2} 、旧テキスト群 T_{o3} 、旧テキスト群 T_{o4} 、旧テキスト群 T_{o5} の順序から成る5つのテキスト群 T が含まれ、旧テキスト群 T_{o2} を旧テキスト群 T_{o4} と旧テキスト群 T_{o5} との間に移動して電子ファイルEFを更新した例を示している。この場合、マッチングを行った結果の一例として、図11に示すとおり、新テキスト群 T_{n1}

10

20

30

40

50

に対して、旧テキスト群 T_{o1} および旧テキスト群 T_{o2} の2つのテキスト群 T が対応付けられ、さらに旧テキスト群 T_{o4} に対して、新テキスト群 T_{n4} および新テキスト群 T_{n2} の2つのテキスト群 T が対応付けられている。なお、旧テキスト群 T_{o3} 、 T_{o5} に対しては、それぞれ新テキスト群 T_{n3} 、 T_{n5} が対応付けられ、対応するテキスト群 T として特定される。

【0072】

この場合、新テキスト群 T_{n1} に対する旧テキスト群 T_{o1} および旧テキスト群 T_{o2} の組み合わせと、旧テキスト群 T_{o4} に対する新テキスト群 T_{n4} および新テキスト群 T_{n2} の組み合わせそれぞれについて、図10に示す例で説明したようにテキスト群特徴量 T_v の差を判定する構成でもよいが、ここでは異なる方法を説明する。

10

【0073】

図11に示す例では、対応付けが特定されていない旧テキスト群 T_{o1} 、旧テキスト群 T_{o2} 、旧テキスト群 T_{o4} 、新テキスト群 T_{n1} 、新テキスト群 T_{n4} 並びに新テキスト群 T_{n2} について、各旧テキスト群 T_o の旧テキスト群特徴量 T_{ov} と、各新テキスト群 T_n の新テキスト群特徴量 T_{nv} との差を、全ての組み合わせ分合計した合計値 D が最小となる組み合わせを算出する。例えば、旧テキスト群 T_{o1} と新テキスト群 T_{n1} 、旧テキスト群 T_{o2} と新テキスト群 T_{n4} 、旧テキスト群 T_{o4} と新テキスト群 T_{n2} 、の組み合わせにおける合計値 D は以下に示す計算式で算出される。

【0074】

【数1】

20

$$D = \sqrt{\sum (旧テキスト群特徴量T_{ov}(n) - 新テキスト群特徴量T_{nv}(m))^2}$$

$$= \sqrt{(旧テキスト群特徴量T_{ov1} - 新テキスト群特徴量T_{nv1})^2 + (旧テキスト群特徴量T_{ov2} - 新テキスト群特徴量T_{nv4})^2 + (旧テキスト群特徴量T_{ov4} - 新テキスト群特徴量T_{nv2})^2}$$

【0075】

すなわち、旧テキスト群 T_{o1} 、旧テキスト群 T_{o2} 、旧テキスト群 T_{o4} 、新テキスト群 T_{n1} 、新テキスト群 T_{n4} 並びに新テキスト群 T_{n2} において、旧テキスト群 T_o と新テキスト群 T_n との組み合わせは、6組存在する。この6つの各組み合わせにおいて、各旧テキスト群 T_o の旧テキスト群特徴量 T_{ov} と、各新テキスト群 T_n の新テキスト群特徴量 T_{nv} との差を、全ての組み合わせ分合計した合計値 D を算出し、その値が最小となる合計値 D の組み合わせを求める。このような処理を行った結果、図13に示すように、旧テキスト群 T_{o1} には新テキスト群 T_{n1} が、旧テキスト群 T_{o2} には新テキスト群 T_{n2} が、旧テキスト群 T_{o4} には新テキスト群 T_{n4} がそれぞれ対応付けられることによって、対応するテキスト群 T を全てのテキスト群 T について特定することができる。

30

【0076】

なお、上述した各旧テキスト群 T_o のテキスト群特徴量 T_{ov} と各新テキスト群 T_n のテキスト群特徴量 T_{nv} との差を、全ての組み合わせ分合計した合計値 D が最小となる組み合わせを算出する処理は、最大、テキスト群 T の数の2乗まで計算量が膨大となるので、遺伝的アルゴリズムを用いて算出することが好ましい。

40

【0077】

図12に示す例は、書き込み情報記憶時の電子ファイル EF に、旧テキスト群 T_{o1} 、旧テキスト群 T_{o2} 、旧テキスト群 T_{o3} 、旧テキスト群 T_{o4} 、旧テキスト群 T_{o5} 、旧テキスト群 T_{o6} の順序から成る6つのテキスト群 T が含まれ、旧テキスト群 T_{o3} と旧テキスト群 T_{o4} との間に、新テキスト群 T_{n7} および新テキスト群 T_{n8} を追加し、さらに旧テキスト群 T_{o5} を削除して電子ファイル EF を更新した例を示している。この場合、マッチングを行った結果の一例として、図12に示すとおり、旧テキスト群 T_{o4} に対して、新テキスト群 T_{n7} 、新テキスト群 T_{n8} 、新テキスト群 T_{n4} の3つのテキスト群 T がそれぞれ対応付けられ、新テキスト群 T_{n6} に対して、旧テキスト群 T_{o5} お

50

よび旧テキスト群 $T_o 6$ の2つのテキスト群 T が対応付けられている。なお、旧テキスト群 $T_o 1 \sim T_o 3$ に対しては、それぞれ新テキスト群 $T_n 1 \sim T_n 3$ が対応付けられ、対応するテキスト群 T として特定される。

【0078】

この場合、図11に示して説明した例と同様に、対応付けが特定されていない旧テキスト群 $T_o 4$ 、旧テキスト群 $T_o 5$ 、旧テキスト群 $T_o 6$ 、新テキスト群 $T_n 7$ 、新テキスト群 $T_n 8$ 、新テキスト群 $T_n 4$ 並びに新テキスト群 $T_n 6$ について、各旧テキスト群 T_o のテキスト群特徴量 $T_o v$ と、各新テキスト群 T_n のテキスト群特徴量 $T_n v$ との差を全ての組み合わせ分合計した合計値 D が、最小となる組み合わせを算出する。ところで、テキスト群 T の追加/削除によって、書き込み情報記憶時の電子ファイル $E F$ に含まれるテキスト群 T_o の数と、書き込み情報反映時の電子ファイル $E F$ に含まれるテキスト群 T_n の数とが異なる場合、旧テキスト群 T_o の数を n 個、新テキスト群 T_n の数を m 個とすると、その対応付けは、図14に示すように、 $n + 1 \times m + 1$ のマトリクスで表すことが可能である。

10

【0079】

すなわち、旧テキスト群 $T_o 4$ 、旧テキスト群 $T_o 5$ 、旧テキスト群 $T_o 6$ 、新テキスト群 $T_n 7$ 、新テキスト群 $T_n 8$ 、新テキスト群 $T_n 4$ 並びに新テキスト群 $T_n 6$ について、各旧テキスト群 T_o のテキスト群特徴量 $T_o v$ と、各新テキスト群 T_n のテキスト群特徴量 $T_n v$ との差を、全ての組み合わせ分合計した合計値 D が、最小となる組み合わせを算出すると、更新によって削除された旧テキスト群 $T_o 5$ は、同図に示すマトリクスにおいて $m + 1$ に該当する箇所にはじき出され、また更新によって新たに追加された新テキスト群 $T_n 7$ および新テキスト群 $T_n 8$ は、 $n + 1$ に該当する箇所にはじき出される。このようにして、対応するテキスト群 T をそれぞれ特定することができる。

20

【0080】

さらに、サーバ50が行う旧文字列 S_o に対する新文字列 S_n の対応付け処理について説明する。文字列 S の対応付け処理は図15に示すように、旧テキスト群 T_o に含まれる旧文字列 S_o と、旧テキスト群 T_o に対応付けされた新テキスト群 T_n に含まれる新文字列 S_n との対応付けを行うものである。文字列 S の対応付け処理には、テキスト群 T と同様に、各文字列 S の文字列特徴量 S_v を算出し、算出した文字列特徴量 S_v を用いて $D P$ マッチング等を行い対応付けを特定する。

30

【0081】

例えば、図15に示す旧テキスト群 $T_o 1$ における、旧文字列 $S_o 1 \sim S_o 5$ について、それぞれの旧文字列特徴量 $S_o v 1 \sim S_o v 5$ を算出する。また、新テキスト群 $T_n 1$ における新文字列 $S_n 1 \sim S_n 5$ について、それぞれの新文字列特徴量 $S_n v 1 \sim S_n v 5$ を算出する。そして、旧文字列 $S_o 1 \sim S_o 5$ の旧文字列特徴量 $S_o v 1 \sim S_o v 5$ と、新文字列 $S_n 1 \sim S_n 5$ の新文字列特徴量 $S_n v 1 \sim S_n v 5$ を用いて、 $D P$ マッチング等を行い対応付けを特定する。なお、文字列 S の対応付け処理については、テキスト群 T の対応付け処理において行った説明と同様であり詳細（文字列マッチング手段、文字列判定手段）については省略する。

【0082】

書き込み情報がマーキングである場合、対応する文字列 S を特定した後、さらに書き込み情報として記憶した「マーキングされた文字のテキストデータ」に基づいて、対応する文字列 S からマーキングされた文字を特定し、サーバ50は電子ファイル $E F$ に対してマーキングを反映する。

40

【0083】

例えば図16に示すように、 $P C 100$ から受信して取得した書き込み詳細情報80（書き込み情報81：「ネットワーク層」位置情報82：「テキスト群3、文字列1」）が書き込み情報データベース $D B 2$ に記憶され、旧テキスト群 T_o 全ての旧テキスト群特徴量 $T_o v$ 、および新テキスト群 T_n 全ての新テキスト群特徴量 $T_n v$ をテキスト群特定情報90として記憶し、旧テキスト群 $T_o 3$ に含まれる文字列 S_o 全ての文字列特徴量 S_o

50

v、および新テキスト群 T_n3 に含まれる文字列 S_n 全ての文字列特徴量 S_{nv} を文字特定情報 95 として書き込み情報データベース DB2 に記憶している場合、サーバ 50 は位置情報 82 およびテキスト群特定情報 90 に基づいてテキスト群 T_3 (旧テキスト群 T_o3) に対応するテキスト群 T を新テキスト群 T_n3 として特定する。

【0084】

さらに、位置情報 82 および文字特定情報 95 に基づいて、新テキスト群 T_n3 の中から、旧テキスト群 T_o3 に含まれる文字列 1「ネットワーク層～用いられる。」(旧文字列 S_o1) に対応する文字列「ネットワーク層～用いられる。」を新文字列 S_n1 として特定する。そして、新文字列 S_n1 「ネットワーク層～用いられる。」の中から、書き込み情報 81 として記憶した「ネットワーク層」をさらに特定し、マーキングを反映させる。

10

【0085】

なお、文字列 S 単位でマーキングを反映させる構成でもよい。この場合、ユーザが図 16 に示す旧文字列 S_o1 に含まれる「ネットワーク層」にマーキングを行うと、新文字列 S_n1 「ネットワーク層～用いられる。」全てにマーキングが反映される。この構成によれば、「マーキングされた文字列のテキストデータ」を記憶する必要がない。また、文字列 S の内容が若干変更したような場合にも、文字列 S 単位でマーキング情報を反映させることによる利点がある。

【0086】

例えば「日本の都市は、東京や大阪である。」という文字列 S にマーキングを行い、その文字列 S が「日本の都市は、東京、大阪、並びに名古屋である。」と変更された場合、対応する文字列 S を特定しさらに「マーキングされた文字列のテキストデータ」を特定する構成では、対応する文字列 S 「日本の都市は、東京、大阪、並びに名古屋である。」の中から「日本の都市は、東京や大阪である。」を特定するので、マーキングは「日本の都市は、東京」「大阪」および「である。」の文字列に対して反映される。しかし、文字列 S 単位でマーキング情報を反映させることによって、文字列 S 「日本の都市は、東京、大阪、並びに名古屋である。」全体に対してマーキングを反映させることができる。

20

【0087】

また、マーキングの書き込み時におけるユーザの選択に基づいて、図 17 に示すマーキング種別情報 97 を書き込み情報 81 としてさらに記憶し、記憶したマーキング種別情報 97 に基づいて、マーキング情報を反映させる構成でもよい。この構成によれば、直線、波線、二重線、下線、点線など多様なマーキングを電子ファイル EF に書き込み、反映させることが可能となる。

30

【0088】

なお、マーキングやメモ書きといった書き込み情報 81 として、画像を書き込む構成でもよい。この場合、位置情報 82 として、表示画面のサイズに関する情報と、テキスト群 T に対する書き込み情報 81 の相対的な位置に関する情報を PC100 が取得し、サーバ 50 に送信する構成となる。この構成によれば、ユーザが電子ファイル EF に描画した画像を、サーバ 50 はそのまま書き込み情報 81 として電子ファイル EF に対して反映することができるので、ユーザによる書き込み処理の自由度をより高めることができる。さらにこの場合、PC100 に接続されたペンを用いて、ユーザがディスプレイ 109 に画像を描画することも可能である。

40

【0089】

また、電子ファイルデータベース DB1、および書き込み情報データベース DB2 をそれぞれ別のサーバで構成してもよい。この構成によれば、書き込み情報データベース DB2 を記憶するサーバのメモリ負担が軽くなるので、書き込み情報を反映させるための処理能力を高めることができる。

【0090】

なお、電子ファイル EF にはテキスト群 T に加えて画像や動画が含まれていてもよい。また、位置情報 82 として、表示画面のサイズに関する情報と、表示画面に対する位置 (

50

座標)を記憶する構成でもよい。

【0091】

また、上述した実施例によらず、電子ファイル書き込みシステム1の装置構成や処理工程等について、本発明の要旨を逸脱しない範囲で、適宜変更も可能である。

【図面の簡単な説明】

【0092】

【図1】本実施形態に係る電子ファイル書き込みシステムの全体図である。

【図2】電子ファイル書き込みシステムのハードウェア構成を示す図である。

【図3】電子ファイルの一例を示す図である。

【図4】電子ファイルへのメモ書き操作の一例を示す説明図である。

10

【図5】電子ファイルへのマーキング操作の一例を示す説明図である。

【図6】書き込み情報が反映された電子ファイルの一例を示す図である。

【図7】テキスト群特徴量を説明するための説明図である。

【図8】書き込み情報の記憶から書き込み情報の反映までの流れを示すフローチャートである。

【図9】テキスト群のマッチング処理結果の一例を示す図である。

【図10】テキスト群のマッチング処理結果の他の例を示す図である。

【図11】テキスト群のマッチング処理結果の他の例を示す図である。

【図12】テキスト群のマッチング処理結果の他の例を示す図である。

【図13】テキスト群の対応付け処理の処理結果の一例を示す図である。

20

【図14】テキスト群の対応付け処理の処理結果の他の例を示す図である。

【図15】文字列の対応付け処理を説明するための説明図である。

【図16】文字列の対応付け処理を説明するための説明図である。

【図17】マーキング種別情報を表す図である。

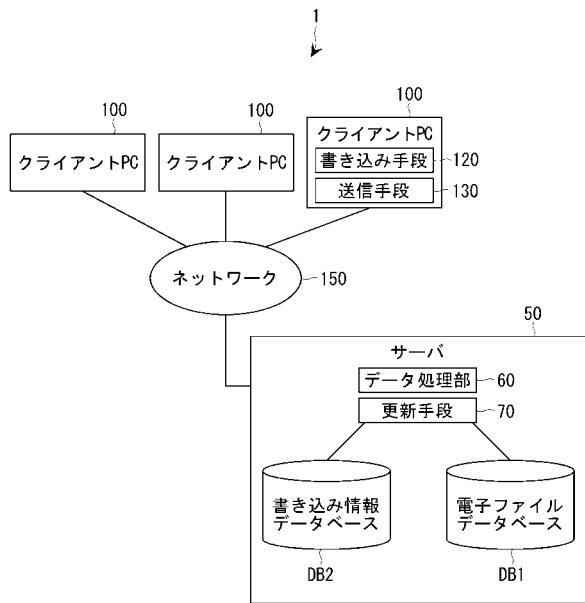
【符号の説明】

【0093】

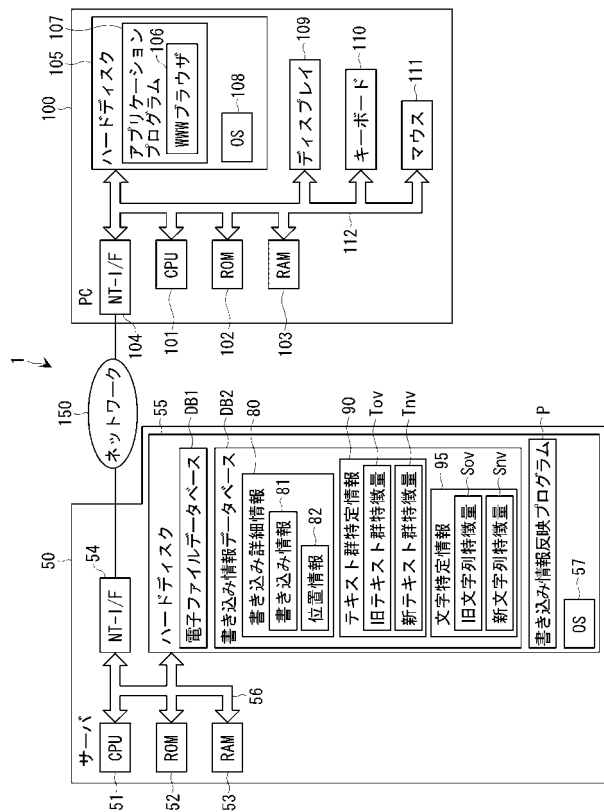
1	電子ファイル書き込みシステム	50	サーバ
60	データ処理部	70	更新手段
80	書き込み詳細情報	81	書き込み情報
82	位置情報	90	テキスト群特定情報
95	文字特定情報	100	クライアントPC
120	書き込み手段	130	送信手段
P	書き込み情報反映プログラム	DB1	電子ファイルデータベース
DB2	書き込み情報データベース	T	テキスト群
To	旧テキスト群	Tn	新テキスト群
Tv	テキスト群特徴量	Sv	文字列特徴量
S	文字列	So	旧文字列
Sn	新文字列		

30

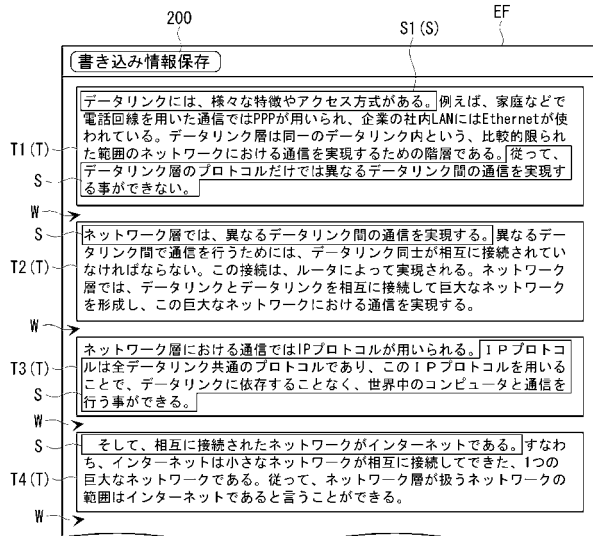
【図 1】



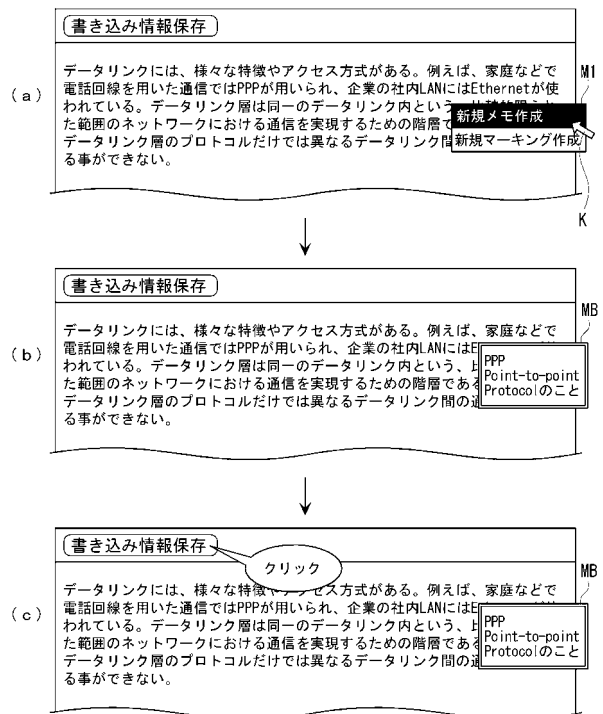
【図 2】



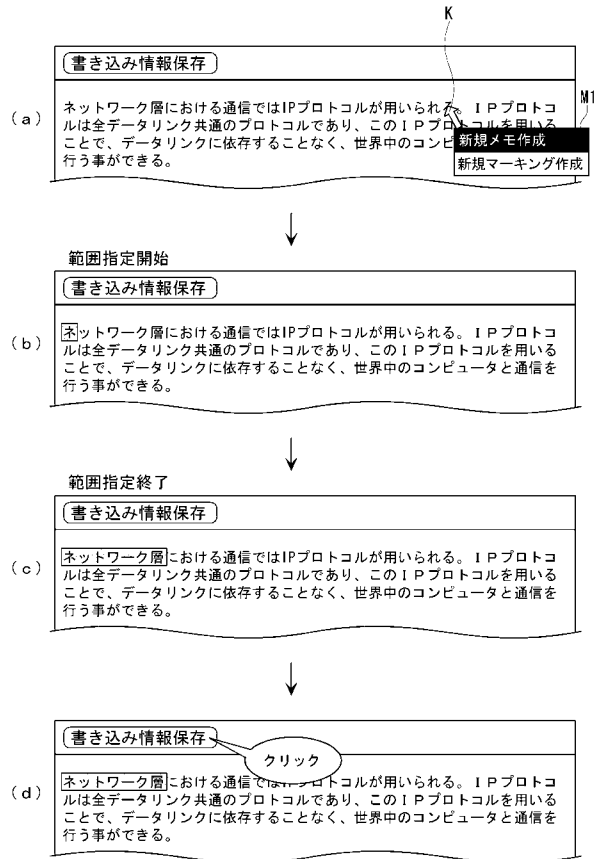
【図 3】



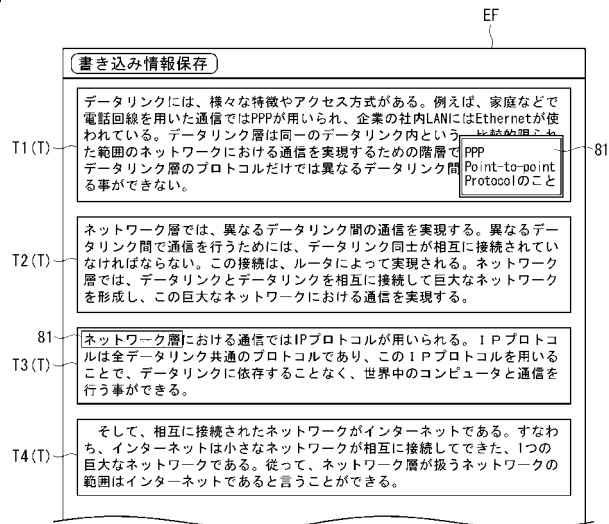
【図 4】



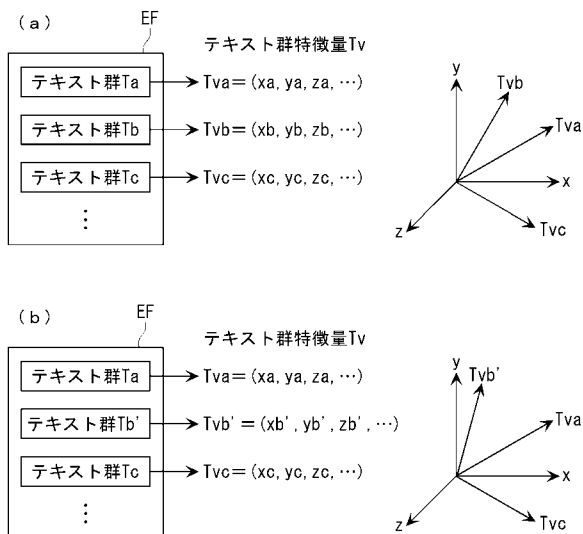
【図 5】



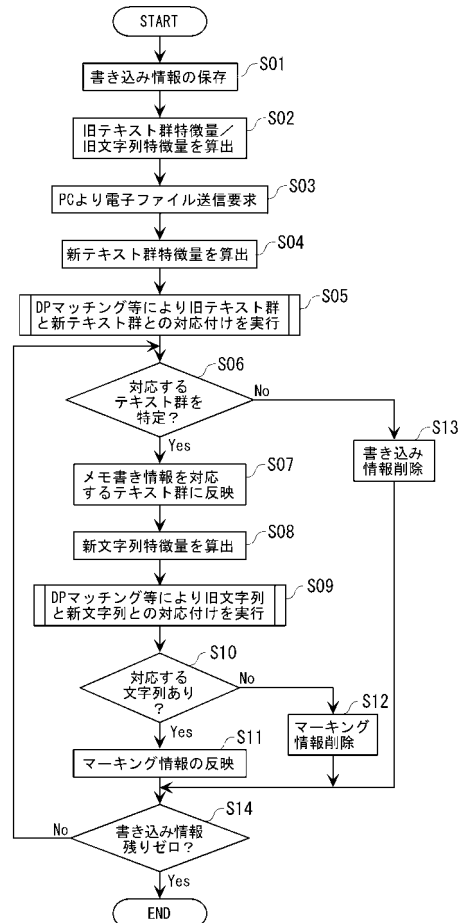
【図 6】



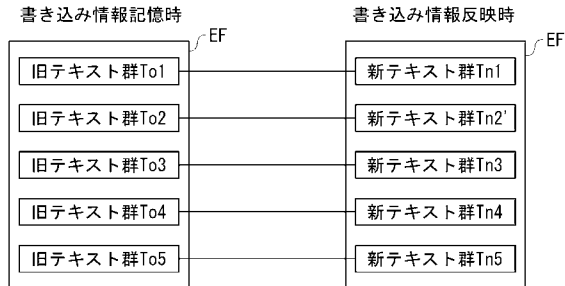
【図 7】



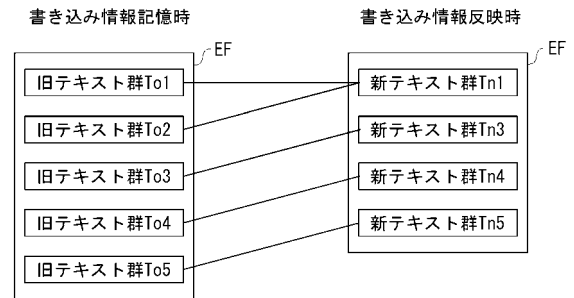
【図 8】



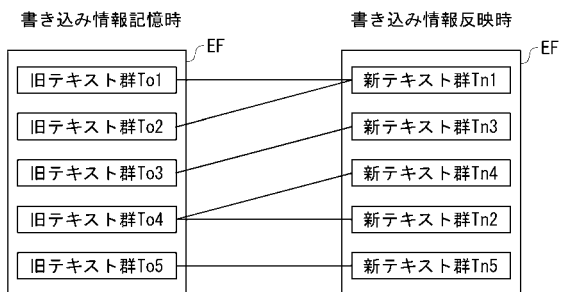
【図 9】



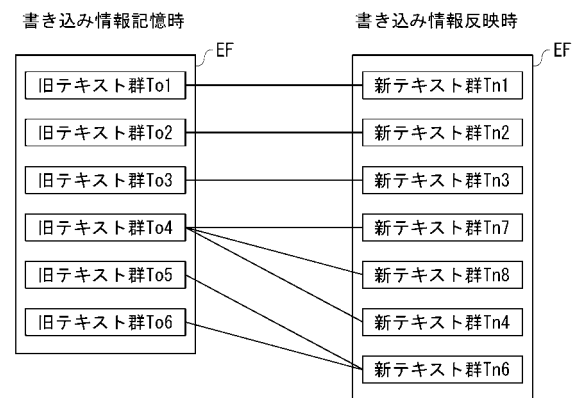
【図 10】



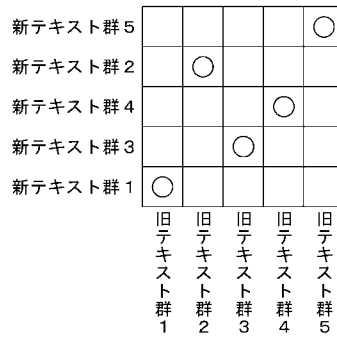
【図 11】



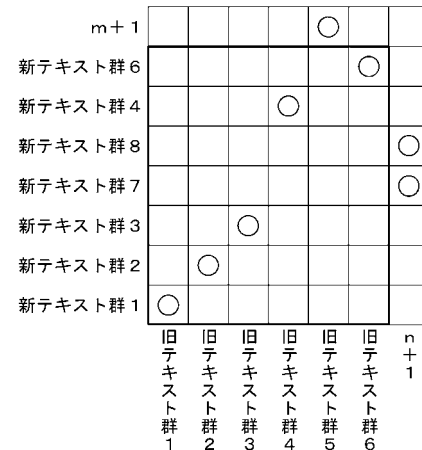
【図 12】



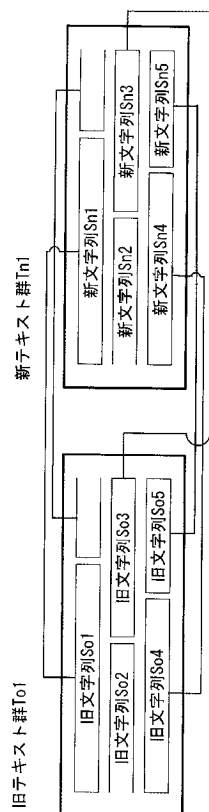
【図 13】



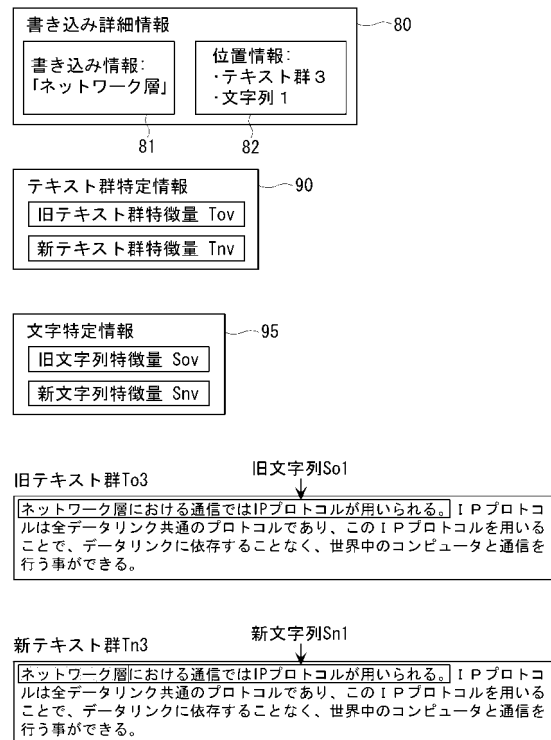
【図 14】



【図 15】



【図 16】



【図 17】

マーキング種別情報

97

マーキング種別コード	内 容
0	直線 ——
1	波線 ~~~~
2	二重線 ===
3	下線 ____
4	点線

フロントページの続き

(74)代理人 100107261

弁理士 須澤 修

(72)発明者 不破 泰

長野県長野市若里四丁目17番1号 信州大学工学部内

(72)発明者 青沼 正志

長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

(72)発明者 酒井 文隆

長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

審査官 岩田 淳

(56)参考文献 特開2003-058535(JP,A)

特開平07-295970(JP,A)

特開2000-339299(JP,A)

松永 聡彦, 外2名, 改版文書翻訳システムにおける文脈を考慮した文対応付け手法, 電子情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2003年 8月22日, 第103巻, 第280号, 第43-48ページ, NLC2003-15~24 言語理解とコミュニケーション

(58)調査した分野(Int.Cl., DB名)

G06F 17/21

G06F 17/24