

(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2011-76139
(P2011-76139A)

(43) 公開日 平成23年4月14日(2011.4.14)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 12/00 (2006.01)	G06F 12/00 545M	5B075
G06F 17/30 (2006.01)	G06F 17/30 220A	5B082
	G06F 17/30 350Z	
	G06F 12/00 533J	

審査請求 未請求 請求項の数 9 O L (全 28 頁)

(21) 出願番号 特願2009-223729 (P2009-223729)
(22) 出願日 平成21年9月29日 (2009. 9. 29)

(71) 出願人 303000372
 コニカミノルタビジネステクノロジーズ株式会社
 東京都千代田区丸の内一丁目6番1号
 (74) 代理人 100117651
 弁理士 高垣 泰志
 (72) 発明者 福留 憲治
 東京都千代田区丸の内一丁目6番1号 コニカミノルタビジネステクノロジーズ株式会社内
 Fターム(参考) 5B075 NR20 NS01 UU06
 5B082 HA03 HA05 HA08

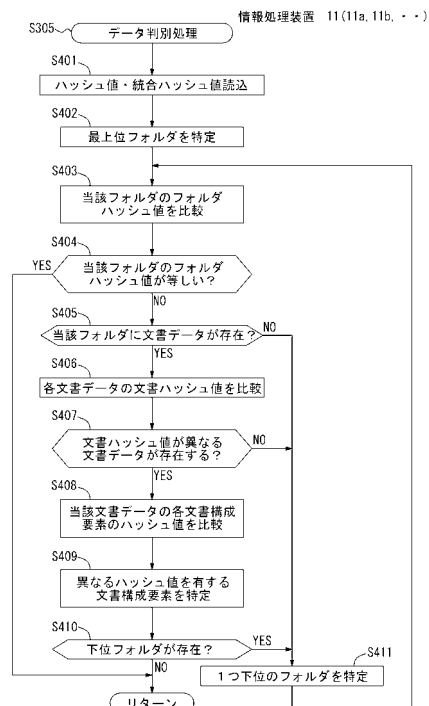
(54) 【発明の名称】 文書管理装置、情報処理装置、文書管理システムおよび文書管理プログラム

(57) 【要約】

【課題】 文書管理装置と情報処理装置とがネットワークを介してデータ通信可能に接続された文書管理システムにおいて、情報処理装置が保持する文書データを更新する際に、更新の必要性の有無を判断するために行われる文書データの同一性判別処理を効率よく行うことで、文書管理システムのパフォーマンスの低下を回避する。

【解決手段】 情報処理装置 11 とネットワークを介してデータ通信可能に接続され、複数の文書構成要素からなる文書データを共有して管理する文書管理装置 1 は、送信対象の文書構成要素の内容に対応したハッシュ値を生成し、更にハッシュ値を統合して統合ハッシュ値を生成し、情報処理装置 11 で保持される文書データが利用される際に、これらハッシュ値・統合ハッシュ値の比較によって文書データの同一性判別を行う構成である。

【選択図】 図 1 2



【特許請求の範囲】**【請求項 1】**

所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有し、ネットワークを介して接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するように管理する文書管理装置であって、

前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶された複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、

前記記憶部に記憶された複数の文書データと、前記所定の記憶領域に記憶された複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成された合成ダイジェストデータの比較により判定するデータ判別手段と、を備えることを特徴とする文書管理装置。

10

【請求項 2】

前記データ判別手段は、

前記記憶部に記憶された複数の文書データのうち、前記所定の記憶領域に記憶された文書データと内容が異なるものを、前記統合ダイジェストデータ生成手段により生成された文書ダイジェストデータの比較により特定することを特徴とする請求項 1 記載の文書管理装置。

20

【請求項 3】

前記記憶部に記憶される文書データは、複数の文書構成要素から構成されており、

各文書構成要素の内容に対応したダイジェストデータを生成するダイジェストデータ生成手段を更に備え、

前記統合ダイジェストデータ生成手段は、

前記記憶部に記憶された文書データを構成する複数の文書構成要素のそれぞれから生成した複数のダイジェストデータに基づいて、各文書データの内容に対応した文書ダイジェストデータを生成することを特徴とする請求項 1 または 2 記載の文書管理装置。

【請求項 4】

前記データ判別手段は、

前記記憶部に記憶された文書データを構成する複数の文書構成要素のうち、前記所定の記憶領域に記憶された文書データを構成する文書構成要素と内容が異なるものを、前記ダイジェストデータ生成手段により生成されたダイジェストデータの比較により特定することを特徴とする請求項 3 記載の文書管理装置。

30

【請求項 5】

所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有し、ネットワークを介して接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するように管理する文書管理装置であって、

前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶された複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、

40

前記記憶部に記憶された複数の文書データと、前記所定の記憶領域に記憶された複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成された合成ダイジェストデータの比較により判定させるために、前記情報処理装置に対して前記合成ダイジェストデータを送信するデータ送信手段と、を備えることを特徴とする文書管理装置。

【請求項 6】

請求項 5 に記載の文書管理装置とネットワークを介してデータ通信可能に接続される情

50

報処理装置であって、

所定の記憶領域に複数の文書データを記憶するデータ記憶手段と、

前記文書管理装置における所定の記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記文書管理装置において生成された合成ダイジェストデータの比較により判定するデータ判別手段と、

を備えることを特徴とする情報処理装置。

【請求項 7】

情報処理装置と文書管理装置とがネットワークを介して相互にデータ通信可能に接続され、前記情報処理装置と前記文書管理装置とのそれぞれで保持される複数の文書データの同一性を保持するように管理する文書管理システムであって、

10

前記文書管理装置は、

所定の記憶部に複数の文書データを記憶する文書データ記憶手段と、

前記記憶部に記憶される各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶される複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、

を備え、

前記情報処理装置は、

所定の記憶領域に複数に文書データを記憶するデータ記憶手段を備え、

20

前記文書管理装置および前記情報処理装置の少なくとも一方は、

前記記憶部に記憶されている複数の文書データと、前記記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成される合成ダイジェストデータの比較により判定するデータ判別手段を備えることを特徴とする文書管理システム。

【請求項 8】

所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有する文書管理装置によって実行され、ネットワークを介して前記文書管理装置に接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するための文書管理プログラムであって、前記文書管理装置に、

30

前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成するステップと、

前記記憶部に記憶された複数の文書データのそれぞれから生成された複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成するステップと、

前記記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記合成ダイジェストデータの比較により判定するステップと、

を実行させることを特徴とする文書管理プログラム。

40

【請求項 9】

所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有する文書管理装置によって実行され、ネットワークを介して前記文書管理装置に接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するための文書管理プログラムであって、前記文書管理装置に、

前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成するステップと、

前記記憶部に記憶された複数の文書データのそれぞれから生成された複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイ

50

ジェストデータを生成するステップと、

前記記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記合成ダイジェストデータの比較により判定させるために、前記情報処理装置に対して前記合成ダイジェストデータを送信するステップと、

を実行させることを特徴とする文書管理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文書管理装置、情報処理装置、文書管理システム、および文書管理プログラムに関し、特に文書データを共有する場合の文書データの管理を行うための技術に関する。

【背景技術】

【0002】

データ管理システムには従来から、データを管理するサーバと、当該サーバにネットワークを介してデータ通信可能に接続されたクライアントとによって構成される、クライアント・サーバ型のデータ管理システムが存在する。また、このようなデータ管理システムには、サーバで管理されるデータと同じデータをクライアントでも保持し、サーバで管理されるデータが更新された場合に、クライアントが保持するデータの更新を行うものが存在する。そしてこのような更新を行う場合、サーバとクライアントとの間でデータの送受信を行うことになるので、更新の回数が増大すれば、ネットワークのトラフィックに大きな負担がかかり、文書管理システムのパフォーマンスが低下する。そこで、このような更新を行う際に、予めデータの内容の同一性を判別して、更新対象となるデータの内容が同一である場合にまで更新を行わないようにし、無駄なデータの送受信を防止してデータ管理システムのパフォーマンスの低下を回避する技術が存在する（例えば、特許文献1）。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2003-122618号公報

【0004】

例えば、上記特許文献1では、自機が保持するマスターデータベースの列データからダイジェストデータを作成することができるサーバを備えるデータベースダウンロードシステムにおいて、最新のダイジェストデータと、クライアントによる直近の列データダウンロード時のダイジェストデータとを比較することにより、マスターデータベースの列データが直近のダウンロード時から更新されているか否かを判断する技術が開示されている。

【発明の概要】

【発明が解決しようとする課題】

【0005】

とこで、上記クライアント・サーバ型のデータ管理システムには、サーバが文書管理装置であり、クライアントが情報処理装置であって、管理対象が文書データであるような文書管理システムが存在する。このような文書管理システムにおいては、近年におけるパーソナルコンピュータの所有率の増加や、携帯性に優れるパーソナルコンピュータの普及により、クライアントである情報処理装置の台数が増加する傾向にある。従って、情報処理装置が保持するデータの更新を文書管理装置に対して要求することができる文書管理システムであれば、情報処理装置から文書管理装置に対して、非常に多くの更新要求が行われることになる。

【0006】

このような環境下では、上述したような、情報処理装置が保持する文書データの更新を行う際に更新対象となる文書データの内容が同一である場合にまで文書データの更新を行わないようにするため、データの内容の同一性を判別する技術が適用されることも多い。

このような技術が適用されれば、文書管理システムのパフォーマンスの低下はある程度回避される。しかし、それでも、少なくともその更新要求のたびに、文書管理装置と情報処理装置との間で、更新対象となるデータのやりとりを行い、それぞれが保持しているデータの中身を解析して同一性の判別を行う必要がある。ゆえに、上述のような文書管理システムにおいては、データ同一性の判別処理そのものが大きな負担となる。

【0007】

このような判別処理に関して、従来から、データが持つ更新日付を用いてデータの内容を判別する技術がある。すなわち、文書管理装置が保持するデータと情報処理装置が保持するデータの更新日付を比較し、データが持つ更新日付が同じであるなら、データの内容が同じであるものと判断して、文書管理装置から情報処理装置へデータの送信を行わないようにし、データが持つ更新日付が異なるならば、データの内容が異なるものであると判断して、データの送信を行う、というものである。この技術では、判別を行うために、文書管理装置と情報処理装置との間で、更新対象となるデータそのものではなく、更新日付データのみを送受信すればよいので、文書管理システムにかかる負担は大きく抑えられることになる。

10

【0008】

しかし、このような技術にも、次の2つの問題点がある。第1に、例えば、ひとつのフォルダに複数の文書データが記憶されている場合、それら複数の文書データの全てについて同一性の判別を行う際には複数の文書データのそれぞれについて更新日付の比較を行わなければならない。そのため、フォルダに格納される文書データの数が多くなればなるほど、更新日付の送受信回数や比較回数が増大し、文書管理システムにおける処理負担が大きくなる。また例えば、ひとつの文書データが複数の文書構成要素から構成される場合にも同様のことが言える。すなわち、ひとつの文書データの同一性を判断する際には複数の文書構成要素のそれぞれについて更新日付の比較を行わなければならないため、文書データを構成する文書構成要素の数が多くなればなるほど、更新日付の比較回数が増大し、処理負担が増大するという問題がある。第2に、データの更新日付が異なっても、データの内容が同じである場合があり、このような場合、文書管理装置と情報処理装置との間で無駄なデータの送受信が行われ、文書管理システムの処理負担が増大するという問題がある。

20

【0009】

上記第2の問題点に関して、例えば特許文献1のように、データの内容に対応したダイジェストデータを比較してデータ内容の同一性を判別する技術によれば、解決することが可能である。しかし、そのような技術であっても、上記第1の問題点を解決できない。すなわち、文書データが複数の文書構成要素から構成されており、かつそのような文書構成要素の数が多い場合には、ダイジェストデータの比較回数も多くなり、文書データが全体として同一であるか否かを判別する際には効率的な判別が行えず、文書管理システムに負担をかけることになる。

30

【0010】

本発明は、上記2つの問題点を解決するためになされたものであり、情報処理装置が保持する文書データの更新の際に行われる、送信対象の文書データの内容の同一性を判別する処理において、文書管理装置と情報処理装置との間で無駄なデータの送受信を行わないようにするとともに、データの比較回数を抑えることによって、文書管理システムのパフォーマンス低下を回避するようにした文書管理装置、情報処理装置、文書管理システム、文書管理プログラム、および情報処理プログラムを提供することを目的としている。

40

【課題を解決するための手段】

【0011】

上記目的を達成するため、請求項1にかかる発明は、所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有し、ネットワークを介して接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するよう

50

に管理する文書管理装置であって、前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶された複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、前記記憶部に記憶された複数の文書データと、前記所定の記憶領域に記憶された複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成された合成ダイジェストデータの比較により判定するデータ判別手段と、を備えることを特徴とする構成である。

【0012】

また請求項2にかかる発明は、請求項1記載の文書管理装置において、前記データ判別手段は、前記記憶部に記憶された複数の文書データのうち、前記所定の記憶領域に記憶された文書データと内容が異なるものを、前記統合ダイジェストデータ生成手段により生成された文書ダイジェストデータの比較により特定することを特徴とする構成である。

10

【0013】

また請求項3にかかる発明は、請求項1または2に記載の文書管理装置において、前記記憶部に記憶される文書データは、複数の文書構成要素から構成されており、各文書構成要素の内容に対応したダイジェストデータを生成するダイジェストデータ生成手段を更に備え、前記統合ダイジェストデータ生成手段は、前記記憶部に記憶された文書データを構成する複数の文書構成要素のそれぞれから生成した複数のダイジェストデータに基づいて、各文書データの内容に対応した文書ダイジェストデータを生成することを特徴とする構成である。

20

【0014】

また請求項4にかかる発明は、請求項3に記載の文書管理装置において、前記データ判別手段は、前記記憶部に記憶された文書データを構成する複数の文書構成要素のうち、前記所定の記憶領域に記憶された文書データを構成する文書構成要素と内容が異なるものを、前記ダイジェストデータ生成手段により生成されたダイジェストデータの比較により特定することを特徴とする構成である。

【0015】

また請求項5にかかる発明は、所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有し、ネットワークを介して接続された情報処理装置とデータ通信を行うことにより前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するように管理する文書管理装置であって、前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶された複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、前記記憶部に記憶された複数の文書データと、前記所定の記憶領域に記憶された複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成された合成ダイジェストデータの比較により判定させるために、前記情報処理装置に対して前記合成ダイジェストデータを送信するデータ送信手段と、を備えることを特徴とする構成である。

30

40

【0016】

また請求項6にかかる発明は、請求項5に記載の文書管理装置とネットワークを介してデータ通信可能に接続される情報処理装置であって、所定の記憶領域に複数の文書データを記憶するデータ記憶手段と、前記文書管理装置における所定の記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記文書管理装置において生成された合成ダイジェストデータの比較により判定するデータ判別手段と、を備えることを特徴とする構成である。

【0017】

また請求項7にかかる発明は、情報処理装置と文書管理装置とがネットワークを介して

50

相互にデータ通信可能に接続され、前記情報処理装置と前記文書管理装置とのそれぞれで保持される複数の文書データの同一性を保持するように管理する文書管理システムであって、前記文書管理装置は、所定の記憶部に複数の文書データを記憶する文書データ記憶手段と、前記記憶部に記憶される各文書データの内容に対応した文書ダイジェストデータを生成し、前記記憶部に記憶される複数の文書データのそれぞれから生成した複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成する統合ダイジェストデータ生成手段と、を備え、前記情報処理装置は、所定の記憶領域に複数の文書データを記憶するデータ記憶手段を備え、前記文書管理装置および前記情報処理装置の少なくとも一方は、前記記憶部に記憶されている複数の文書データと、前記記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記統合ダイジェストデータ生成手段により生成される合成ダイジェストデータの比較により判定するデータ判別手段を備えることを特徴とする構成である。

10

【0018】

また請求項8にかかる発明は、所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有する文書管理装置によって実行され、ネットワークを介して前記文書管理装置に接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するための文書管理プログラムであって、前記文書管理装置に、前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成するステップと、前記記憶部に記憶された複数の文書データのそれぞれから生成された複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成するステップと、前記記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記合成ダイジェストデータの比較により判定するステップと、を実行させることを特徴とする構成である。

20

【0019】

また請求項9にかかる発明は、所定の記憶部に複数の文書データを記憶する文書データ記憶手段を有する文書管理装置によって実行され、ネットワークを介して前記文書管理装置に接続された情報処理装置とデータ通信を行うことによって前記記憶部に記憶された複数の文書データと、前記情報処理装置における所定の記憶領域に記憶された複数の文書データとの同一性を保持するための文書管理プログラムであって、前記文書管理装置に、前記記憶部に記憶された各文書データの内容に対応した文書ダイジェストデータを生成するステップと、前記記憶部に記憶された複数の文書データのそれぞれから生成された複数の文書ダイジェストデータに基づいて、前記複数の文書ダイジェストデータの内容に対応した合成ダイジェストデータを生成するステップと、前記記憶部に記憶されている複数の文書データと、前記所定の記憶領域に記憶されている複数の文書データとで内容が異なるものが存在するか否かを、前記合成ダイジェストデータの比較により判定させるために、前記情報処理装置に対して前記合成ダイジェストデータを送信するステップと、を実行させることを特徴とする構成である。

30

40

【発明の効果】**【0020】**

本発明によれば、情報処理装置が保持するデータの更新の際に行われる、送信対象のデータの内容の同一性を判別する処理において、送信対象のデータの内容に対応したダイジェストデータを統合して生成する文書ダイジェストデータおよび合成ダイジェストデータを判定に用いることにより、文書管理装置と情報処理装置との間で無駄なデータの送受信を行わないようにするとともに、データの比較回数を抑えることができる。それ故、文書管理システムのパフォーマンス低下を回避することが可能となる。

【図面の簡単な説明】**【0021】**

50

- 【図 1】文書管理システムの一構成例を示す図である。
- 【図 2】文書データの一例を示す図である。
- 【図 3】文書管理装置のハードウェア構成の一例を示すブロック図である。
- 【図 4】情報処理装置のハードウェア構成の一例を示すブロック図である。
- 【図 5】文書管理装置の制御部がプログラムを実行することによって実現される詳細な機能構成を示すブロック図である。
- 【図 6】文書構成要素のハッシュ値および文書データの文書ハッシュ値生成の概念図である。
- 【図 7】フォルダのフォルダハッシュ値生成の概念図である。
- 【図 8】情報処理装置の制御部がシステムプログラムおよびアプリケーションプログラムを実行することによって実現される詳細な機能構成を示すブロック図である。 10
- 【図 9】文書管理装置が情報処理装置から文書構成要素を受信する場合の処理手順の一例を示すフローチャートである。
- 【図 10】文書管理装置が情報処理装置からハッシュ値・統合ハッシュ値要求を受信する場合の処理手順の一例を示すフローチャートである。
- 【図 11】情報処理装置においてユーザがフォルダをクリックする場合の処理手順の一例を示すフローチャートである。
- 【図 12】データ判定処理の詳細な処理手順の一例を示すフローチャートである。
- 【図 13】データ判別処理を、具体例により説明するための図である。
- 【図 14】データ判別処理を、具体例により説明するための図である。 20
- 【図 15】データ判別処理を、具体例により説明するための図である。

【発明を実施するための形態】

【0022】

以下、本発明に関する好ましい実施形態について図面を参照しつつ詳細に説明する。尚、以下に説明する実施形態において互いに共通する部材には同一符号を付しており、それらについての重複する説明は省略する。

【0023】

図 1 は、本実施形態における文書管理システムの一構成例を示す図である。この文書管理システムは、いわゆるクライアント・サーバ型の文書管理システムであり、サーバとして機能する文書管理装置 1 と、クライアントとして機能する複数の情報処理装置 11 (11a、11b、・・・) とを備えている。文書管理装置 1 および複数の情報処理装置 11 のそれぞれは、LAN や WAN などのネットワーク 10 を介して相互にデータ通信可能に接続されている。 30

【0024】

文書管理システムは、文書管理装置 1 と情報処理装置 11 とで同一の文書データを保持し、文書管理装置 1 で保持する文書データが更新されると、情報処理装置 11 がその更新された文書データを文書管理装置 1 からダウンロードすることにより、文書管理装置 1 と情報処理装置 11 とのそれぞれで保持される文書データの整合性を保持するように構成される。それ故、文書管理装置 1 は、ネットワーク 10 を介して、情報処理装置 11 において保持される文書データに対応した文書データを保持する。ここで、文書管理装置 1 に保持される文書データと、情報処理装置 11 に保持される文書データとの対応関係は、文書データの所定のプロパティ、例えばファイル名などによって判断される。つまり、文書管理装置 1 に保持される文書データのプロパティと、情報処理装置 11 に保持される文書データのプロパティとが互いに一致すれば、それらの文書データは互に対応するデータである。 40

【0025】

図 1 の例では、文書管理装置 1 は文書データ 100a を保持している。このとき、情報処理装置 11a は、文書データ 100a と同一プロパティを有する文書データ 100b を保持し、情報処理装置 11b は、文書データ 100a と所定の同一プロパティを有する文書データ 100c を保持する。この場合、文書管理装置 1 に保持される文書データ 100 50

aと、情報処理装置11a, 11aのそれぞれで保持される文書データ100b, 100cとは互いに対応するデータである。

【0026】

図2は、本実施形態における文書データ100の一例を示す図である。ここで、文書データ100は、文書管理装置1で保持される文書データ100aおよび情報処理装置11a, 11aのそれぞれで保持される文書データ100b, 100cを総称するものである。この文書データ100は、複数の文書構成要素190から構成される。文書構成要素190として、例えば文書本体データ101や、サムネイル102や、データベース103や、およびセキュリティ情報104などが挙げられる。文書本体データ101とは、例えばテキストファイルなどの文書ファイル本体のことをいう。サムネイル102は、文書本体データ101のサムネイルデータである。データベース103は、文書本体データ101のプロパティ情報、例えば文書本体データ101のオーナー情報やアクセス権等を管理するデータベースデータである。セキュリティ情報104は、電子署名等の、文書本体データ101のセキュリティに関するデータである。

10

【0027】

文書データ100は、文書構成要素190として、少なくとも文書本体データ101を含む。また文書データ100は、文書本体データ101以外の3つの文書構成要素を含むか否かは任意である。つまり、文書データ100が、文書本体データ101のみから構成されている場合もあり、また文書本体データ101、データベース103、およびセキュリティ情報104という3つの文書構成要素190から構成されている場合もある。尚、データベース103は必ずしも文書本体データ101のプロパティ情報を管理するものに限られず、他の情報を管理するものであってもよい。また、文書データ100は、文書構成要素190として、これ以外のデータを含んでいてもよい。

20

【0028】

図1に戻り、文書管理装置1は、文書データ100aを構成するデータとして、文書構成要素190aを保持している。この文書構成要素190aは、情報処理装置11aにおいて保持される文書データ100bを構成する文書構成要素190bに対応する。また文書構成要素190aは、情報処理装置11bにおいて保持される文書データ100cを構成する文書構成要素190cに対応する。文書管理装置1に保持される文書構成要素190aと、情報処理装置11a, 11bに保持される文書構成要素190b, 190cとの対応関係は、同種の文書構成要素であるか否かにより判断され、例えばデータの拡張子などで判断される。そして文書データ100aの文書構成要素190aと、文書データ100b, 100cの文書構成要素190b, 190cとが、同種の文書構成要素であれば、それらは互いに対応した文書構成要素となる。

30

【0029】

図1の例では、文書管理装置1が保持する文書データ100aには、文書構成要素190aが含まれている。このとき、情報処理装置11aが保持する文書データ100bには、文書構成要素190aと同種の文書構成要素190b含まれ、情報処理装置11bが保持する文書データ100cには、文書構成要素190aと同種の文書構成要素190cが含まれる。具体的には、文書管理装置1が保持する文書データ100aにサムネイル102が含まれていれば、情報処理装置11aが保持する文書データ100bにも、それに対応するサムネイル102が含まれる。同様に、文書データ100aにデータベース103が含まれれば、情報処理装置11bが保持する文書データ100cにもそれに対応するデータベース103が含まれる。

40

【0030】

本実施形態の文書管理システムは、文書管理装置1で保持される文書データ100aの内容と、各情報処理装置11で保持される文書データ100b, 100cの内容とが同一になるように、文書管理装置1と情報処理装置11との間でネットワーク10を介してデータの送受信を行うように構成されている。ここで、文書データ100aと文書データ100bの内容が同一であるとは、文書データ100aに含まれる文書構成要素190aの

50

内容と、文書データ100bに含まれる文書構成要素190bの内容が、全て同一であることをいう。例えば、文書データ100aが、文書本体データ101aと、サムネイル102aと、データベース103aと、セキュリティ情報104aとから構成されており、文書データ100bが、文書本体データ101bと、サムネイル102bと、データベース103bと、セキュリティ情報104bとから構成されている場合に、文書本体データ101aと101bとが同一内容であり、サムネイル102aと102bとが同一内容であり、データベース103aと103bとが同一内容であり、およびセキュリティ情報104aと104bとが同一内容であるなら、文書データ100aの内容と文書データ100bの内容は同一である。同様に、文書データ100aと文書データ100cの内容が同一であるとは、文書構成要素190aの内容と文書構成要素190cの内容が全て同一であることをいう。

10

【0031】

また文書管理装置1は、文書データ100aを所定のフォルダ200aに記憶して管理する。そして情報処理装置11aは、文書管理装置1におけるフォルダ200aに対応するフォルダ200bに文書データ100bを記憶して管理する。同様に、情報処理装置11aは、文書管理装置1におけるフォルダ200aに対応するフォルダ200cに文書データ100cを記憶して管理する。ここで、文書管理装置1におけるフォルダ200aと、情報処理装置11におけるフォルダ200b、200cとの対応関係は、それぞれのフォルダの所定のプロパティ、例えばフォルダ名などによって判断される。つまり、文書管理装置1のフォルダ200aのプロパティと、情報処理装置11のフォルダ200b、200cのプロパティとが互いに一致すれば、それらのフォルダは互に対応するフォルダである。図1の例では、文書管理装置1は、文書データ100aをフォルダ200aに格納している。このとき、情報処理装置11aは、文書データ100bを、フォルダ200aと同一フォルダ名のフォルダ200bに格納し、情報処理装置11bは、文書データ100cを、フォルダ200aと同一フォルダ名のフォルダ200cに格納する。

20

【0032】

したがって、本実施形態の文書管理システムでは、文書管理装置1における文書データ100aのデータ保持構造と、情報処理装置11における文書データ100b、100cのデータ保持構造とが同一となっており、文書管理装置1と情報処理装置11とのそれぞれで同一の文書データを共有して保持するようになっている。

30

【0033】

情報処理装置11(11a、11b...)は、例えば市販のパーソナルコンピュータ(PC)などで構成される。情報処理装置11aは、文書管理装置1が保持する文書データ100aを、ネットワーク10を介して更新することができる。すなわち、文書管理装置1が保持している文書データ100aの内容を、文書管理装置1に対して文書データ100bを送信することによって、文書管理装置1に書き換えさせることができる。また、情報処理装置11は、自機が保持する文書データ100bの内容を、ネットワーク10を介して文書管理装置1から受信した文書データ100aにより更新することができる。同様に、情報処理装置11bも、文書管理装置1が保持する文書データ100aを自機が保持する文書データ100cによって更新でき、また自機が保持する文書データ100cを、文書管理装置1が保持する100aによって更新できる。

40

【0034】

本実施形態の文書管理システムにおいて、情報処理装置11のユーザが文書データ100を利用する際、情報処理装置11が参照する文書データ100は当該情報処理装置11において保持している文書データ100である。すなわち、ユーザは、情報処理装置11で保持される文書データ100を閲覧等することになる。ここで例えば、情報処理装置11aのユーザが文書データ100bに対する編集作業を行って、文書データ100bを更新すると、それに伴い、情報処理装置11は、更新された文書データ100bを文書管理装置1に送信して文書データ100aの内容を、更新された文書データ100bの内容に更新する。すると、情報処理装置11bは、文書管理装置1が保持する文書データ100

50

aとは異なる内容の文書データ100cを保持することになる。この状態で、情報処理装置11bが、そのまま自機で保持する文書データ100cをユーザに表示すれば、ユーザは、文書管理装置1が保持する文書データ100aと異なる内容の文書データ100cを利用することになり、文書の統一が図れない。

【0035】

そのため、本実施形態におけるクライアント・サーバ型の文書管理システムでは、各情報処理装置11において保持される文書データ100b, 100cが利用される際、文書管理装置1において保持される文書データ100aが更新されて異なるデータとなっているか否かを判断し、その結果、異なるデータに更新されていれば、情報処理装置11が文書管理装置1から更新された文書データ100aをダウンロードし、その文書データ100aにより、各情報処理装置11において保持されている文書データ100b, 100cを更新するように構成される。その結果、各情報処理装置11では、この更新された文書データ100b, 100cに基づいてユーザに対する表示を行うことができるようになり、システム全体で利用される文書の統一が行える。

10

【0036】

以下、このような文書管理システムについて更に詳しく説明する。尚、文書管理システムに設けられる複数の情報処理装置11のそれぞれは情報処理装置11aと同様の機能および構成であるので、以下においては情報処理装置11として、情報処理装置11aを例に挙げて説明する。

20

【0037】

図3は、文書管理装置1のハードウェア構成の一例を示すブロック図である。図3に示すように文書管理装置1は、制御部20と、ネットワークインタフェース23と、記憶装置30とを備え、これらがデータバス24を介して接続されている。制御部20は、CPU21とメモリ22とを備えており、CPU21が、記憶装置30に記憶されているプログラム33を読み出して実行することにより、各部の動作を制御する。メモリ22は、CPU21がプログラム33を実行する際に一時的なデータなどを記憶するためのものである。

【0038】

ネットワークインタフェース23は、ネットワーク10を介して情報処理装置11とデータ通信を行うためのものである。

30

【0039】

記憶装置30は、例えばハードディスク装置などの不揮発性記憶装置によって構成される。この記憶装置30には、文書管理装置1に予めインストールされたプログラム33が記憶される。また記憶装置30には、文書データ100aを格納する文書データ記憶部31、および文書管理装置1において生成されるハッシュ値および統合ハッシュ値を格納するハッシュ値・統合ハッシュ値記憶部32が設けられる。尚、ハッシュ値および統合ハッシュ値とは、文書データの内容等に対応した一定長のデータのことであるが、これについては後述する。また文書データ記憶部31には、上述したように、情報処理装置11のフォルダ200bと対応するフォルダ200aが文書データ100aを記憶するための記憶領域(記憶部)として設けられる。

40

【0040】

図4は、情報処理装置11aのハードウェア構成の一例を示すブロック図である。図4に示すように情報処理装置11aは、制御部40と、ネットワークインタフェース43と、表示部44と、入力部45と、記憶装置50とを備え、これらがデータバス46を介して接続されている。

【0041】

制御部40は、CPU41とメモリ42とを備えており、CPU41が記憶装置50に記憶されているプログラム53および54を読み出して実行することにより、各部の制御や、各種の演算処理を行う。メモリ42はプログラム53および54の実行時に一時的に生成されるデータなどを記憶する。

50

【 0 0 4 2 】

ネットワークインタフェース 4 3 は、ネットワーク 1 0 を介して文書管理装置 1 とデータ通信を行うためのものである。

【 0 0 4 3 】

表示部 4 4 は、文書データに基づいて文書情報を表示したり、その他の情報を表示したりする表示手段であり、例えば液晶ディスプレイなどで構成される。また入力部 4 5 は、ユーザが操作することにより各種信号を入力する入力手段であり、例えばキーボードやマウスなどを備えて構成される。

【 0 0 4 4 】

記憶装置 5 0 は、例えばハードディスク装置などの不揮発性記憶装置によって構成される。この記憶装置 5 0 には、情報処理装置 1 1 に予めインストールされたプログラム 5 3 および 5 4 が記憶される。システムプログラム 5 3 は、オペレーティングシステムである。また、アプリケーションプログラム 5 4 は、自機が保持する文書データ 1 0 0 b を閲覧したり、また文書管理装置 1 が保持する文書データ 1 0 0 a を更新したり、等するためのプログラムである。また記憶装置 5 0 には、文書データ 1 0 0 b を格納する文書データ記憶部 5 1、および文書管理装置 1 から受信するハッシュ値および統合ハッシュ値を格納するハッシュ値・統合ハッシュ値記憶部 5 2 が設けられる。ここで、文書データ記憶部 5 1 には、上述したように、文書管理装置 1 のフォルダ 2 0 0 a と対応するフォルダ 2 0 0 b が文書データ 1 0 0 b を記憶するための記憶領域として設けられる。

【 0 0 4 5 】

上述したように、情報処理装置 1 1 a は、自機で保持する文書データ 1 0 0 b により、文書管理装置 1 で保持される文書データ 1 0 0 a を更新することができる。同様に、他の情報処理装置 1 1 b も、文書管理装置 1 で保持される文書データ 1 0 0 a を更新することがある。他の情報処理装置 1 1 b が文書管理装置 1 で保持される文書データ 1 0 0 a を更新した場合、情報処理装置 1 1 a で保持する文書データ 1 0 0 b は文書管理装置 1 の文書データ 1 0 0 a と異なった内容のデータとなる。そこで、情報処理装置 1 1 a は、自機が保持する文書データ 1 0 0 b を文書管理装置 1 が保持する文書データ 1 0 0 a で更新することによって、文書データ 1 0 0 b が文書データ 1 0 0 a と同一の内容となるようにしている。本実施形態では、情報処理装置 1 1 a が文書管理装置 1 からダウンロードして文書データ 1 0 0 b を更新する処理は、例えばユーザが情報処理装置 1 1 a に対して文書データ 1 0 0 b を閲覧等するなどの指示を行った際に行われるようになっている。

【 0 0 4 6 】

このとき、本実施形態では、文書管理装置 1 で保持される文書データ 1 0 0 a に含まれる文書構成要素 1 9 0 a のうち、情報処理装置 1 1 a で閲覧対象等として選択された文書データ 1 0 0 b に含まれる文書構成要素 1 9 0 b と内容が異なるものが存在するか否かを判別し、内容の異なるものが存在すればその文書構成要素 1 9 0 a を特定する。このような判別処理は、情報処理装置 1 1 a において行っても良いし、また文書管理装置 1 で行っても良い。そして文書構成要素 1 9 0 b と内容の異なる文書構成要素 1 9 0 a が特定されると、文書管理装置 1 が、情報処理装置 1 1 a に対して、その特定された文書構成要素 1 9 0 a のみを送信する。これにより、情報処理装置 1 1 a が保持する文書データ 1 0 0 b の更新を行うことができる。本実施形態では、このような判別処理を情報処理装置 1 1 a で行う場合を例示する。情報処理装置 1 1 a において判別処理を行うことで、例えば文書管理装置 1 とネットワーク 1 0 を介して接続されている情報処理装置 1 1 の台数が多い場合でも、文書管理装置 1 の処理パフォーマンスへの負担が大きく軽減することができる。

【 0 0 4 7 】

情報処理装置 1 1 a が判別処理を行う際、文書管理装置 1 で生成されるハッシュ値（ダイジェストデータ）を参照して判別を行う。以下、詳しく説明する。本実施形態においては、文書構成要素 1 9 0 a または文書構成要素 1 9 0 b の内容に対応したデータであるハッシュ値（ダイジェストデータ）と、文書データ 1 0 0 a または文書データ 1 0 0 b の内容に対応したデータである文書ハッシュ値（文書ダイジェストデータ）と、フォルダ 2 0

10

20

30

40

50

0 aまたはフォルダ200 bの内容に対応したデータであるフォルダハッシュ値（合成ダイジェストデータ）、の3種類のハッシュ値を用いて、文書構成要素190 bと異なる内容の文書構成要素190 aが存在するか否かを判別する。尚、フォルダハッシュ値が対応するフォルダの内容については、後述する。

【0048】

まず文書管理装置1において行われる処理について説明する。文書管理装置1は、複数の文書構成要素190 aのそれぞれから、各文書構成要素190 aの内容に対応したハッシュ値を生成する。次に文書データ100 aの内容に対応した文書ハッシュ値と、フォルダ200 aの内容に対応したフォルダハッシュ値とを生成する。文書管理装置1では、これら各種ハッシュ値を文書データ100 aのデータ保持構造に対応付けて記憶しておく。そして、情報処理装置11 aから、文書構成要素190 aの送信要求があった場合、文書管理装置1は、文書構成要素190 aと共に、それらハッシュ値を情報処理装置11 aに送信する。これにより情報処理装置11 aでは、文書データ100 bのデータ保持構造に対応したハッシュ値が保持されることになる。

10

【0049】

文書管理装置1で保持される文書データ100 a又はその文書構成要素190 aと、情報処理装置11 aで保持される文書データ100 b又はその文書構成要素190 bとが同一の内容であれば、文書管理装置1で保持される各種ハッシュ値と、情報処理装置11 aで保持される各種ハッシュ値とが一致した値となる。これに対し、文書管理装置1で保持される文書データ100 a又はその文書構成要素190 aと、情報処理装置11 aで保持される文書データ100 b又はその文書構成要素190 bとが互いに異なる内容となっていれば、文書管理装置1で保持される各種ハッシュ値と、情報処理装置11 aで保持される各種ハッシュ値とが異なる値となる。

20

【0050】

そのため、本実施形態では、情報処理装置11 aが上記判別処理を行う際、文書管理装置1においてその時点で保持されている各種ハッシュ値を取得し、情報処理装置11 aで保持しているハッシュ値との比較を行う。具体的には、文書ハッシュ値およびフォルダハッシュ値を比較することにより、文書データ100 aと文書データ100 bの内容が異なっているか否かを判別し、その後、文書構成要素ごとのハッシュ値を用いて文書構成要素190 bと内容の異なる文書構成要素190 aを特定する。これにより、例えば文書データ100 aと文書データ100 bの内容が同一である場合に、これらの文書データの文書ハッシュ値を比較すれば、文書構成要素190 aと文書構成要素190 bとを個別に比較することなしに、文書構成要素190 aと文書構成要素190 bとは全て同一データ内容の文書構成要素で構成されていることが判明する。また、フォルダ200 aのフォルダハッシュ値とフォルダ200 bのフォルダハッシュ値とを比較すれば、それぞれのフォルダ200 a, 200 bに複数の文書データが記憶されている場合であっても、それらを個別に比較することなしに全ての文書データが同一内容であることが判明する。この場合、各文書データに含まれる複数の文書構成要素190のそれぞれについて個別に比較することなく、文書構成要素190 aと文書構成要素190 bとが全て同一内容であることが明らかになる。

30

40

【0051】

このように本実施形態では、文書管理装置1で保持される複数の文書構成要素190 aのうち、情報処理装置11 aで保持される文書構成要素190 bと内容が異なるものが存在するかを判別する際、フォルダハッシュ値、文書ハッシュ値およびハッシュ値の順でハッシュ値比較を行っていく。このような判別手法を用いれば、異なる内容の文書構成要素を全て特定するために必要なデータの比較回数を抑えることができ、文書管理システムのパフォーマンスの低下を回避することができるようになる。

【0052】

ここで、本実施形態においては、文書ハッシュ値およびフォルダハッシュ値を総称して統合ハッシュ値という。従って、上述した判別手法では、まず統合ハッシュ値を用いて、

50

フォルダ200aの内容とフォルダ200bの内容が異なるか否かを判別し、異なる場合は次に文書データ100aの内容と文書データ100bの内容が異なるか否かを判別する。そしてさらに異なる場合には、その後、ハッシュ値を用いて文書構成要素190bと内容が異なる文書構成要素190aを特定することになる。

【0053】

このような本実施形態の文書管理システムにおいて、ユーザが情報処理装置11aを用いて文書データ100bを利用しようとした場合、まず、データ判別を行わせるために、情報処理装置11aに対して文書管理装置1が生成したハッシュ値および統合ハッシュ値を送信する。そして情報処理装置11aが、文書管理装置1から受信したハッシュ値および統合ハッシュ値を用いてデータ判別を行った後、文書構成要素190bと内容が異なる文書構成要素190aが存在すれば、その文書構成要素190aの送信を文書管理装置1に対して要求する。これにより、文書管理装置1は要求のあった文書構成要素190aのみを情報処理装置11aに送信する。そして情報処理装置11bは、自機で保持している文書構成要素190bを、文書管理装置1から受信した文書構成要素190aにより更新する。このような一連の処理を行うことで、文書データ100aと、文書データ100bとが同一のデータ内容として構成される。つまり、文書管理装置1と情報処理装置11aとでデータの同一性が保持される。

10

【0054】

そして本実施形態では、上述の判別処理をハッシュ値および統合ハッシュ値を用いて行うので、従来の方法に比べデータ処理の負担が軽減されることとなり、文書管理システム全体のパフォーマンスの低下を回避できる。尚、上述した例では、主として情報処理装置11aと文書管理装置1との関係について説明したが、他の情報処理装置11bについても同様である。

20

【0055】

次に上記のような動作を実現するための、文書管理装置1における具体的な内部機能について説明する。図5は、文書管理装置1の制御部20がプログラム33を実行することによって実現される詳細な機能構成を示すブロック図である。尚、図5では、文書管理装置1が保持する文書データ100aの更新後、新たに更新後の文書データ100aについてハッシュ値・統合ハッシュ値を生成する機能、および、ユーザが情報処理装置11aを操作して文書データ100bを閲覧しようとした場合に、文書データ100aと文書データ100bとを同一内容にするため、文書管理装置1が情報処理装置11aに文書構成要素190aを送信する機能に関するブロックのみを示しており、それ以外の機能については図示を省略している。

30

【0056】

図5に示すように、制御部20は、プログラム33を実行することにより、文書データ更新部60、ハッシュ値生成部61、統合ハッシュ値生成部62、文書ハッシュ値生成部63、フォルダハッシュ値生成部64、ハッシュ値・統合ハッシュ値送信部65、文書構成要素送信部66として機能する。

【0057】

文書データ更新部60は、文書管理装置1が保持する文書データ100aの更新を行う。具体的には、情報処理装置11aから文書構成要素190bが送信されてきた場合に、文書データ記憶部31に格納されている文書構成要素190aを、当該送信されてきた文書構成要素190bに書き換える。これにより、文書構成要素190aの内容は、送信されてきた文書構成要素190bの内容と同一になる。

40

【0058】

ハッシュ値生成部61は、文書データ更新部60により文書データ100aが更新された場合に機能し、文書データ更新部60により書き換えられた文書構成要素190aから、ハッシュ値を生成する。具体的には、文書データ更新部60により書き換えられた文書構成要素190aを、記憶装置30に格納されている所定のハッシュ関数に入力し、文書構成要素190aの内容に対応した一定長のデータであるハッシュ値の出力を得る。つま

50

り、ハッシュ値生成部 6 1 は、文書データ更新部 6 0 により書き換えられた文書構成要素 1 9 0 a の、書き換え後の内容に対応するハッシュ値を生成する。ここで、ハッシュ関数とは、入力したデータの内容に対応した、一定長のデータを出力する関数のことをいう。ハッシュ関数は、入力されるデータの内容が異なる場合は、異なるデータをハッシュ値として出力する。従って、ハッシュ関数の出力であるハッシュ値を比較すれば、文書構成要素 1 9 0 b と内容が異なる文書構成要素 1 9 0 a を特定できる。そしてハッシュ値生成部 6 1 は、上述のようにして生成したハッシュ値により、ハッシュ値・統合ハッシュ値記憶部 3 2 に設けられたハッシュ値記憶部 3 4 に記憶されているハッシュ値を更新する。具体的には、ハッシュ値記憶部 3 4 に記憶されている、文書データ更新部 6 0 による書き換え前の文書構成要素 1 9 0 a のハッシュ値を、上述のようにして新たに生成した書き換え後の文書構成要素 1 9 0 a のハッシュ値に書き換える。

10

【 0 0 5 9 】

図 6 (a) は、文書データ更新部 6 0 により更新された文書データ 1 1 0 a の文書ハッシュ値生成概念図である。ここで、文書データ 1 1 0 a は、文書データ 1 0 0 a の一例である。すなわち、文書管理装置 1 が文書データ 1 1 0 a を保持し、文書データ 1 1 0 a と同一のプロパティを有する文書データ 1 1 0 b を、情報処理装置 1 1 a が保持していると仮定する。尚、図 6 では、文書構成要素 1 9 0 a は文書本体データ 1 1 1 a、サムネイル 1 1 2 a、データベース 1 1 3 a、およびセキュリティ情報 1 1 4 a の 4 つであり、これらは全て文書データ更新部 6 0 により書き換えられているものとする。この場合、ハッシュ値生成部 6 1 は、文書本体データ 1 1 1 a をハッシュ関数 8 0 に入力し、その文書本体データ 1 1 1 a の内容に対応したハッシュ値 3 1 1 a を得る。またサムネイル 1 1 2 a をハッシュ関数 8 0 に入力し、そのサムネイル 1 1 2 a の内容に対応したハッシュ値 3 1 2 a を得る。またデータベース 1 1 3 a をハッシュ関数 8 0 に入力し、そのデータベース 1 1 3 a の内容に対応したハッシュ値 3 1 3 a を得る。またセキュリティ情報 1 1 4 a をハッシュ関数 8 0 に入力し、そのセキュリティ情報 1 1 4 a の内容に対応したハッシュ値 3 1 4 a を得る。以上のようにしてハッシュ値生成部 6 1 は、文書データ更新部 6 0 により更新された文書データ 1 1 0 a に含まれる各文書構成要素 1 9 0 a のハッシュ値を生成する。

20

【 0 0 6 0 】

統合ハッシュ値生成部 6 2 は、文書ハッシュ値生成部 6 3 とフォルダハッシュ値生成部 6 4 とを備えている。文書ハッシュ値生成部 6 3 は文書ハッシュ値を生成し、フォルダハッシュ値生成部 6 4 はフォルダハッシュ値を生成する。ハッシュ値生成部 6 1 により新たにハッシュ値が生成された場合に文書ハッシュ値生成部 6 3 が機能し、文書ハッシュ値生成部 6 3 による文書ハッシュ値の生成が終わった後、フォルダハッシュ値生成部 6 4 が機能する。

30

【 0 0 6 1 】

文書ハッシュ値生成部 6 3 は、文書データ更新部 6 0 により更新された文書データ 1 0 0 a につき、文書データ 1 0 0 a を構成する複数の文書構成要素 1 9 0 a のそれぞれに基づいて文書データ 1 0 0 a の内容に対応した文書ハッシュ値を生成する。具体的には、文書データ 1 0 0 a を構成する複数の文書構成要素 1 9 0 a のそれぞれの内容に対応した複数のハッシュ値を、ハッシュ関数に入力し、当該文書データ 1 0 0 a の内容に対応した一定長のデータである文書ハッシュ値の出力を得る。このとき、文書データ更新部 6 0 により書き換えられた文書構成要素 1 9 0 a については、ハッシュ値生成部 6 1 において生成されたハッシュ値を用い、文書データ更新部 6 0 により書き換えられていない文書構成要素 1 9 0 a については、ハッシュ値記憶部 3 4 から読み込んだハッシュ値を用いることになる。つまり、文書ハッシュ値生成部 6 3 は、文書データ更新部 6 0 により更新された文書データ 1 0 0 a の、更新後の内容に対応した文書ハッシュ値を生成する。このような文書ハッシュ値は、ハッシュ値と同様、ハッシュ関数に入力される複数のハッシュ値のうち少なくとも 1 つが異なる値であれば、異なる値のハッシュ値として出力される。そのため、文書ハッシュ値を比較すれば、文書データ 1 0 0 a を構成する複数の文書構成要素 1

40

50

90 aのそれぞれの内容と、文書データ100 bを構成する複数の文書構成要素190 bのそれぞれの内容とが異なるか否かを判別できる。

【0062】

図6(b)は、文書データ110 aの文書ハッシュ値生成概念図である。文書ハッシュ値生成部63は、図6(a)においてハッシュ値生成部61が文書本体データ111 aから生成したハッシュ値311 aと、ハッシュ値生成部61がサムネイル112 aから生成したハッシュ値312 aと、ハッシュ値生成部61がデータベース113 aから生成したハッシュ値313 aと、およびハッシュ値生成部61がセキュリティ情報114 aから生成したハッシュ値314 aとをハッシュ関数80に入力し、文書データ110 aの内容に対応した文書ハッシュ値410 aを得る。

10

【0063】

上述のようにして文書ハッシュ値生成部63が文書ハッシュ値を生成した後、フォルダハッシュ値生成部64は、文書データ更新部60により更新された文書データ100 aを格納しているフォルダ200 aについて、フォルダハッシュ値を生成する。具体的には、フォルダ200 aが下位フォルダを持つか否かにより、以下の2つの生成方法が実行される。

【0064】

(1) 下位フォルダを持たないフォルダ200 aについては、当該フォルダに格納されている全ての文書データ100 aの内容に対応した複数の文書ハッシュ値から、その複数の文書ハッシュ値に対応したフォルダハッシュ値を生成する。具体的には、当該フォルダに格納されている複数の文書データ100 aのそれぞれの内容に対応した複数の文書ハッシュ値を、記憶装置30に格納されているハッシュ関数に入力し、その複数の文書ハッシュ値に対応した一定長のデータであるフォルダハッシュ値の出力を得る。このとき、文書データ更新部60により更新された文書データ100 aについては、文書ハッシュ値生成部63において生成された文書ハッシュ値を用い、文書データ更新部60により更新されていない文書データ100 aについては、統合ハッシュ値記憶部35から読み込んだ文書ハッシュ値を用いることになる。

20

【0065】

図7(a)は、下位フォルダを持たないフォルダ210 aのフォルダハッシュ値生成概念図である。ここで、フォルダ210 aはフォルダ200 aの一例である。すなわち、文書管理装置1がフォルダ210 aを有し、フォルダ210 aと同一のプロパティを有するフォルダ210 bを、情報処理装置11aが有していると仮定する。尚、図7では全ての文書データ100 aが文書データ更新部60により更新されたものとする。フォルダ210 aには、文書データ100 aの一例である文書データ110 a、文書データ120 a、および文書データ130 aが格納されている。文書データ110 aの文書ハッシュ値410 a、文書データ120 aの文書ハッシュ値420 a、および文書データ130 aの文書ハッシュ値430 aを、上述した文書ハッシュ値生成方法により文書ハッシュ値生成部63が生成した後、フォルダハッシュ値生成部64は、文書ハッシュ値420と、文書ハッシュ値420と、文書ハッシュ値440とをハッシュ関数80に入力して、文書ハッシュ値420の内容、文書ハッシュ値430の内容、および文書ハッシュ値440の内容に対応した、フォルダ210 aのフォルダハッシュ値510 aを得る。

30

40

【0066】

(2) 下位フォルダを持つフォルダ200 aについては、当該フォルダ200 aに格納されている全ての文書データ100 aの内容に対応した複数の文書ハッシュ値と、1つ下位のフォルダのフォルダハッシュ値とから、その複数の文書ハッシュ値および1つ下位のフォルダのフォルダハッシュ値に対応したフォルダハッシュ値を生成する。具体的には、フォルダ200 aに格納されている各文書データ100 aの内容に対応した複数の文書ハッシュ値と、1つ下位のフォルダのハッシュ値とを、記憶装置30に格納されているハッシュ関数に入力し、その複数の文書ハッシュ値および1つ下位のフォルダのフォルダハッシュ値に対応した一定長のデータであるフォルダハッシュ値の出力を得る。このとき、文

50

書データ更新部 60 により更新された文書データ 100 a については、文書ハッシュ値生成部 63 において生成された文書ハッシュ値を用い、文書データ更新部 60 により更新されていない文書データ 100 a については、統合ハッシュ値記憶部 35 から読み込んだ文書ハッシュ値を用いることになる。また、1つ下位のフォルダが文書データ更新部 60 により更新された文書データを格納している場合はフォルダハッシュ値生成部 64 によって生成されたフォルダハッシュ値を用いる。さらに、1つ下位のフォルダが文書データ更新部 60 により更新されていない文書データのみを格納する場合は、統合ハッシュ値記憶部 35 から読み込んだフォルダハッシュ値を用いる。

【0067】

図 7 (b) は、下位フォルダが存在するフォルダ 220 a のフォルダハッシュ値生成の概念図である。ここで、フォルダ 220 a はフォルダ 200 a の一例である。フォルダ 220 a には、文書データ 100 a の一例である文書データ 140 a および文書データ 150 a が格納されており、その1つ下位のフォルダとしてフォルダ 210 a が存在する。文書データ 140 a の文書ハッシュ値 440 a および文書データ 150 a の文書ハッシュ値 450 a を、上述した文書ハッシュ値生成方法により文書ハッシュ値生成部 63 が生成した後、フォルダハッシュ値生成部 64 は、フォルダ 210 a のフォルダハッシュ値 510 a を、上述した下位フォルダが存在しない場合のフォルダハッシュ値生成方法により生成する。そしてフォルダハッシュ値生成部 64 は、文書ハッシュ値 440 a と、文書ハッシュ値 450 a と、フォルダハッシュ値 510 a とをハッシュ関数 80 に入力して、文書ハッシュ値 440 a の内容、文書ハッシュ値 450 a の内容、およびフォルダハッシュ値 510 a の内容に対応した、フォルダ 220 a のフォルダハッシュ値 520 a を得る。

【0068】

つまり、フォルダハッシュ値生成部 64 は、文書データ更新部 60 により文書データ 100 a が更新された場合、その更新後の文書データ 100 a を格納しているフォルダおよびその上位フォルダの全てについて、それらフォルダに格納されている全ての文書データ 100 a の文書ハッシュ値の内容と、1つ下位のフォルダのフォルダハッシュ値の内容とに対応したハッシュ値を生成する。ここで、下位のフォルダが存在しない場合のフォルダハッシュ値は、そのフォルダに格納されている全ての文書データ 100 a の文書ハッシュ値の内容に対応している。また、下位のフォルダが存在する場合のフォルダハッシュ値は、そのフォルダに格納されている全ての文書データ 100 a の内容と、そのフォルダよりも下位のフォルダ全てに格納されている全ての文書データ 100 a の内容に対応している。従って、フォルダハッシュ値を比較すれば、そのフォルダ、およびそのフォルダより下位のフォルダに格納されている文書データ 100 a のうちで、文書データ 100 b と内容が異なるものが存在するか否かを判定することができる。このようにフォルダハッシュ値は、そのフォルダに格納されている全ての文書データ 100 a およびその文書データ 100 a を構成する全ての文書構成要素 190 a の内容を反映したダイジェストデータとなっている。

【0069】

そして統合ハッシュ値生成部 62 は、上述のようにして生成した文書ハッシュ値およびフォルダハッシュ値を、ハッシュ値・統合ハッシュ値記憶部 32 に設けられた統合ハッシュ値記憶部 35 に格納する。このとき、統合ハッシュ値生成部 62 は、統合ハッシュ値記憶部 35 に既に格納されている文書ハッシュ値およびフォルダハッシュ値を、上述のようにして生成した文書ハッシュ値およびフォルダハッシュ値で書き換えることにより、文書ハッシュ値およびフォルダハッシュ値を更新する。

【0070】

ハッシュ値・統合ハッシュ値送信部 65 は、情報処理装置 11 a においてユーザが文書データ 100 b を閲覧等しようとした場合に情報処理装置 11 a から送信される、ハッシュ値・統合ハッシュ値要求を受信した場合に機能し、情報処理装置 11 b にハッシュ値および統合ハッシュ値を送信する。具体的には、フォルダ 200 a のフォルダハッシュ値、文書データ 100 a の文書ハッシュ値、および文書構成要素 190 a のハッシュ値を、ハ

10

20

30

40

50

ッシュ値・統合ハッシュ値記憶部 3 2 から読み込み、これを情報処理装置 1 1 a に送信する。

【0071】

文書構成要素送信部 6 6 は、データ判別処理を終えた情報処理装置 1 1 a から送信される文書構成要素送信要求を受信した場合に、当該送信要求された文書構成要素 1 9 0 a を送信する。具体的には、情報処理装置 1 1 a が送信した文書構成要素送信要求から、送信対象となる文書構成要素 1 9 0 a を特定し、特定した文書構成要素 1 9 0 a を文書データ記憶部 3 1 から読み込んで、情報処理装置 1 1 a に送信する。

【0072】

次に情報処理装置 1 1 a における具体的な内部機能について説明する。図 8 は、情報処理装置 1 1 a の制御部 4 0 がシステムプログラム 5 3 およびアプリケーションプログラム 5 4 を実行することによって実現される詳細な機能構成を示すブロック図である。尚、図 8 では、例えばユーザが文書データ 1 0 0 b が格納されているフォルダ 2 0 0 b を選択してクリック操作した場合に、文書データ 1 0 0 a と文書データ 1 0 0 b を同一内容にするため、文書管理装置 1 から送信されてくるハッシュ値・統合ハッシュ値により、文書構成要素 1 9 0 b と異なる内容の文書構成要素 1 9 0 a が存在するが否かを判定し、存在するならば、その文書構成要素 1 9 0 a の送信要求を行い、その文書構成要素 1 9 0 a によって情報処理装置 1 1 a が保持する文書構成要素 1 9 0 b を更新する機能に関するブロックのみを示しており、それ以外の機能については図示を省略している。

【0073】

図 8 に示すように、制御部 4 0 は、プログラム 5 3 および 5 4 を実行することにより、文書データ管理部 7 0、ハッシュ値・統合ハッシュ値要求部 7 1、データ判別部 7 2、文書構成要素要求部 7 3、データ更新部 7 4 として機能する。

【0074】

文書データ管理部 7 0 は、文書データ 1 0 0 b の内容を変更したり、閲覧等できるように、文書データを管理する。具体的には、文書データ管理部 7 0 は、アプリケーションプログラム 5 4 の機能によって、文書データ記憶部 5 1 に格納されている文書構成要素 1 9 0 b を読み込んだユーザが、その文書構成要素 1 9 0 b の内容を変更した場合、変更後の文書構成要素 1 9 0 b を文書管理装置 1 に送信し、また文書データ記憶部 5 1 に格納されている変更前の文書構成要素 1 9 0 b の内容を、当該変更後の文書構成要素 1 9 0 b の内容に書き換える。尚、文書管理装置 1 に送信された変更後の文書構成要素 1 9 0 b は、文書データ更新部 6 0 によって処理されることになる。また、ユーザが、例えば文書データ 1 0 0 b を閲覧等しようとして、入力部 4 5 を介してフォルダ 2 0 0 b をクリックすれば、文書データ管理部 7 0 に設けられたハッシュ値・統合ハッシュ値要求部 7 1 が機能し、文書管理装置 1 に対して、フォルダ 2 0 0 a のフォルダハッシュ値、文書データ 1 0 0 a の文書ハッシュ値、および文書構成要素 1 9 0 a のハッシュ値の送信を要求する。

【0075】

データ判別部 7 2 は、ハッシュ値・統合ハッシュ値要求部 7 1 が文書管理装置 1 に対してハッシュ値および統合ハッシュ値を要求した場合に機能し、文書構成要素 1 9 0 b と異なる文書構成要素 1 9 0 a が文書管理装置 1 で保持されている場合、これを特定する。具体的には、データ判別部 7 2 は、まず、文書管理装置 1 からハッシュ値および統合ハッシュ値を受信し、受信したハッシュ値および統合ハッシュ値を用いて、文書データ 1 0 0 b と内容が異なる文書データ 1 0 0 a が存在するが否かを判別する。異なるデータが存在している場合、次にその文書データ 1 0 0 a に含まれる複数の文書構成要素 1 9 0 a のうち、文書構成要素 1 9 0 b と内容が異なる文書構成要素 1 9 0 a を全て特定する。また、データ判別部 7 2 は、文書管理装置 1 から受信したハッシュ値および統合ハッシュ値を、メモリ 4 2 に一時的に格納する。

【0076】

文書構成要素要求部 7 3 は、文書管理装置 1 に対して、データ判別部 7 2 が行う上記データ判別方法によって特定された文書構成要素 1 9 0 a の送信要求を行う。

【 0 0 7 7 】

データ更新部 7 4 は、文書構成要素要求部 7 3 が送信要求を行った場合に機能し、文書構成要素 1 9 0 b と、ハッシュ値および統合ハッシュ値とを更新する。具体的には、まず、文書構成要素要求部 7 3 が文書管理装置 1 に対して送信要求した文書構成要素 1 9 0 a を受信し、文書データ記憶部 5 1 に格納されている文書構成要素 1 9 0 b の内容を、当該受信した文書構成要素 1 9 0 a の内容に書き換える。そして、ハッシュ値・統合ハッシュ値記憶部 5 2 に格納されている、上述の書き換えを行う前の文書構成要素 1 9 0 b のハッシュ値、その文書構成要素 1 9 0 b により構成される文書データ 1 0 0 b の文書ハッシュ値、その文書データ 1 0 0 b を格納するフォルダ 2 0 0 b のフォルダハッシュ値、およびそのフォルダ 2 0 0 a の上位フォルダのフォルダハッシュ値を、上述した、データ判別部 7 2 が一時的にメモリ 4 2 に格納したハッシュ値・統合ハッシュ値に書き換える。

10

【 0 0 7 8 】

このデータ更新部 7 4 による文書構成要素 1 9 0 b の更新により、文書管理装置 1 と情報処理装置 1 1 a とで、文書データ 1 0 0 a と文書データ 1 0 0 b は、全て同一内容となる。従って、ユーザは、情報処理装置 1 1 a により、文書管理装置 1 で保持されている文書データ 1 0 0 a と同一内容の文書データ 1 0 0 b を閲覧等することができるようになる。また、上述のように文書構成要素 1 9 0 b が更新されることにより、情報処理装置 1 1 a は、自機が保持する文書構成要素 1 9 0 b の内容と対応しないハッシュ値、自機が保持する文書データ 1 0 0 b の内容と対応しない文書ハッシュ値、および自機が保持するフォルダ 2 0 0 b の内容と対応しないフォルダハッシュ値、を保持することになるが、データ更新部 7 4 によりハッシュ値および統合ハッシュ値が更新されれば、情報処理装置 1 1 a は、自機が保持する文書構成要素 1 9 0 b の内容、文書データ 1 0 0 b の内容、およびフォルダ 2 0 0 b の内容と対応したハッシュ値および統合ハッシュ値を保持することになる。従って、再びユーザがフォルダ 2 0 0 b をクリックして、データ判別部 7 2 が、データ判別を行うことになっても、ハッシュ値・統合ハッシュ値記憶部 5 2 に格納されたハッシュ値・統合ハッシュ値を用いて行えばよいので、自機が保持する文書構成要素 1 9 0 b のハッシュ値、自機が保持する文書データ 1 0 0 b の文書ハッシュ値、および自機が保持するフォルダ 2 0 0 b のフォルダハッシュ値を生成する必要はなく、文書管理システムのパフォーマンスの低下を回避することができる。

20

【 0 0 7 9 】

また、本実施形態において、データ判別は、文書管理装置 1 が行わずに情報処理装置 1 1 が行うこととしている。これにより、文書管理装置 1 にかかる処理負担を軽減させている。特に、文書管理装置 1 とネットワーク 1 0 を介しデータ通信可能に接続されている情報処理装置 1 1 の数が多くなればなるほど、文書管理装置 1 に対する情報処理装置 1 1 からのデータ処理要求が多くなり文書管理装置 1 の処理負担が増えるので、情報処理装置 1 1 がデータ判別を行うことによる、文書管理システムのパフォーマンス低下回避の効果が大きくなる。

30

【 0 0 8 0 】

図 9 は、文書管理装置 1 が情報処理装置 1 1 a から文書構成要素 1 9 0 b を受信する場合の処理手順の一例を示すフローチャートである。文書管理装置 1 は、情報処理装置 1 1 a から、文書構成要素 1 9 0 b を受信したか否かを監視している（ステップ S 1 0 1）。文書管理装置 1 が文書構成要素 1 9 0 b を受信したならば（ステップ S 1 0 1 で Y E S）、文書データ更新部 6 0 が機能し、受信した文書構成要素 1 9 0 b により文書管理装置 1 が保持する文書データ 1 0 0 a の更新を行う（ステップ S 1 0 2）。その後、ハッシュ値生成部 6 1 が機能し、ステップ S 1 0 1 で受信した文書構成要素 1 9 0 b のハッシュ値を生成し（ステップ S 1 0 3）、生成したハッシュ値によりハッシュ値記憶部 3 4 に格納されているハッシュ値の更新を行う（ステップ S 1 0 4）。すると統合ハッシュ値生成部 6 2 が機能し、ステップ S 1 0 3 においてハッシュ値生成部 6 1 が生成したハッシュ値から、統合ハッシュ値を生成する（ステップ S 1 0 5）。そしてその後、統合ハッシュ値生成部 6 2 は、ステップ S 1 0 5 において生成した統合ハッシュ値により、ハッシュ値・統合

40

50

ハッシュ値記憶部 3 2 における統合ハッシュ値記憶部 3 5 に格納されている統合ハッシュ値を更新して（ステップ S 1 0 6 ）、処理を終了する。

【 0 0 8 1 】

図 1 0 は、文書管理装置 1 が情報処理装置 1 1 a からハッシュ値・統合ハッシュ値要求を受信する場合の処理手順の一例を示すフローチャートである。文書管理装置 1 は、情報処理装置 1 1 a から、ハッシュ値・統合ハッシュ値要求を受信したか否かを監視している（ステップ S 2 0 1 ）。文書管理装置 1 がハッシュ値・統合ハッシュ値要求を受信したならば（ステップ S 2 0 1 で Y E S ）、ハッシュ値・統合ハッシュ値送信部 6 5 が機能し、フォルダ 2 0 0 a、文書データ 1 0 0 a、および文書構成要素 1 9 0 a のハッシュ値・統合ハッシュ値を、情報処理装置 1 1 a に送信する（ステップ S 2 0 2 ）。その後、文書管理装置 1 は、所定時間以内に、情報処理装置 1 1 a から文書構成要素送信要求を受信したか否かを監視する（ステップ S 2 0 3 ）。文書構成要素送信要求を受信した場合は（ステップ S 2 0 3 で Y E S ）、文書構成要素送信部 6 6 が機能し、受信した文書構成要素送信要求から、送信対象の文書構成要素 1 9 0 a を特定する（ステップ S 2 0 4 ）。そしてステップ S 2 0 4 で特定した文書構成要素 1 9 0 a を情報処理装置 1 1 a に送信して（ステップ S 2 0 5 ）、処理を終了する。一方、ステップ S 2 0 3 で所定時間以内に文書構成要素送信要求を受信しなかった場合は（ステップ S 2 0 3 で N O ）、文書構成要素 1 9 0 a の送信をせずに処理を終了する。

10

【 0 0 8 2 】

図 1 1 は、情報処理装置 1 1 a においてユーザがフォルダ 2 0 0 b をクリックする場合の処理手順の一例を示すフローチャートである。文書データ管理部 7 0 は、フォルダ 2 0 0 b が入力部 4 5 を介してクリックされるか否かを監視している（ステップ S 3 0 1 ）。共有フォルダ 2 0 0 b がクリックされた場合（ステップ S 3 0 1 で Y E S ）、文書データ管理部 7 0 に備わるハッシュ値・統合ハッシュ値要求部 7 1 が機能し、文書管理装置 1 に対して、フォルダ 2 0 0 a、文書データ 1 0 0 a、および文書構成要素 1 9 0 a のハッシュ値・統合ハッシュ値の送信を要求する（ステップ S 3 0 2 ）。その後、データ判別部 7 2 が機能し、文書管理装置 1 からハッシュ値・統合ハッシュ値を受信したか否かを監視する（ステップ S 3 0 3 ）。ハッシュ値・統合ハッシュ値を受信したならば（ステップ S 3 0 3 で Y E S ）、データ判別部 7 2 が機能して、受信したハッシュ値・統合ハッシュ値をメモリ 4 2 に一時的に格納し（ステップ S 3 0 4 ）、その後データ判別処理を行う（ステップ S 3 0 5 ）。このデータ判別処理において、文書データ 1 0 0 b と異なる内容の文書データ 1 0 0 a が存在すると判定された場合（ステップ S 3 0 6 で Y E S ）、文書構成要素要求部 7 3 が機能し、データ判別処理において特定された文書構成要素 1 9 0 a の送信を、文書管理装置 1 に対して要求する（ステップ S 3 0 7 ）。すると、データ更新部 7 4 が機能し、文書管理装置 1 から文書構成要素 1 9 0 a を受信したか否かを監視する（ステップ S 3 0 8 ）。文書構成要素 1 9 0 a を受信した場合（ステップ S 3 0 8 で Y E S ）、文書データ記憶部 5 1 に格納されている文書構成要素 1 9 0 b の内容を、当該受信した文書構成要素 1 9 0 a の内容に書き換える（ステップ S 3 0 9 ）。また、ハッシュ値・統合ハッシュ値記憶部 5 2 に格納されているハッシュ値・統合ハッシュ値を、ステップ S 3 0 4 でデータ判別部 7 2 が一時的にメモリ 4 2 に格納したハッシュ値・統合ハッシュ値に書き換え（ステップ S 3 1 0 ）、処理を終了する。一方、ステップ S 3 0 5 のデータ判別処理において、文書データ 1 0 0 b と異なる内容の文書データ 1 0 0 a が存在しないと判定された場合（ステップ S 3 0 6 で N O ）、文書構成要素およびハッシュ値・統合ハッシュ値の更新は行わずに処理を終了する。

20

30

40

【 0 0 8 3 】

図 1 2 は、データ判定処理（図 1 1 のステップ S 3 0 5 ）の詳細な処理手順の一例を示すフローチャートである。まず、データ判別部 7 2 は、メモリ 4 2 に格納されたフォルダ 2 0 0 a、文書データ 1 0 0 a、および文書構成要素 1 9 0 a のハッシュ値・統合ハッシュ値を読み込み、更にハッシュ値・統合ハッシュ値記憶部 5 2 からフォルダ 2 0 0 b、文書データ 1 0 0 b、および文書構成要素 1 9 0 b のハッシュ値・統合ハッシュ値を読み込

50

む（ステップS401）。ここでメモリ42に格納されているハッシュ値・統合ハッシュ値は、図11のステップS304において格納したものであり、ハッシュ値・統合ハッシュ値記憶部52に格納されているハッシュ値・統合ハッシュ値は、図11のステップS310において更新されたハッシュ値・統合ハッシュ値を含んでいる。次にフォルダ200aのうちの最上位フォルダおよび200bのうちの最上位フォルダを特定し（ステップS402）、フォルダ200aのうちの最上位フォルダのフォルダハッシュ値とフォルダ200bのうちの最上位フォルダのフォルダハッシュ値とを比較する（ステップS403）。フォルダハッシュ値が等しい場合は（ステップS404でYES）、何もせずにデータ判別処理を終了する。フォルダハッシュ値が異なる場合は（ステップS404でNO）、次に当該フォルダ200bに格納されている文書データ100bと当該フォルダ200aに格納されている文書データ100aとで文書ハッシュ値が異なるものが存在するか否かを判定する（ステップS405）。存在する場合は（ステップS405でYES）、当該文書データ100aと当該文書データ100bとで文書ハッシュ値とを比較する（ステップS406）。文書ハッシュ値が異なる文書データ100aおよび文書データ100bが存在する場合は（ステップS407でYES）、当該文書データ100aに含まれる文書構成要素190aのハッシュ値と、当該文書データ100bに含まれる文書構成要素190bのハッシュ値とを比較する（ステップS408）。そして、文書構成要素190bと異なるハッシュ値を有する文書構成要素190aを特定し（ステップS409）、次に下位のフォルダが存在するか否かを判定する（ステップS410）。下位のフォルダが存在しない場合は（ステップS410でNO）、データ判別処理を終了する。一方、当該フォルダ200aに文書データ100aが存在しないか（ステップS405でNO）、当該フォルダに文書データは存在するが（ステップS405でYES）、文書ハッシュ値が異なる文書データが存在しないか（ステップS407でNO）、または当該フォルダに文書データが存在し、それらの文書データのうちの異なる文書ハッシュ値を有するものも存在するが、更に下位のフォルダが存在する場合は（ステップS410でYES）、当該フォルダの1つ下位のフォルダを特定し（ステップS411）、再びステップS403以降の上記プロセスを繰り返す。

【0084】

図13、図14、および図15は、いずれも図12におけるデータ判別処理を、具体例により説明するための図である。図例では、文書管理装置1における文書データ記憶部31のデータ保持構造として、最上位フォルダのフォルダ220aを備え、このフォルダ220aに、文書データ140a及び150aと、その下位フォルダとなるフォルダ210aとが格納されている。フォルダ210aには、文書データ110a、120a、130aが格納されている。また、情報処理装置11aにおいて文書データ記憶部51には、文書管理装置1のデータ保持構造と対応するように、最上位フォルダのフォルダ220bを備え、このフォルダ220bに、文書データ140b及び150bと、その下位フォルダとなるフォルダ210bとが格納されている。フォルダ210bには、文書データ110b、120b、130bが格納されている。そして図13、図14、および図15では文書管理装置1のデータ保持構造に対応したハッシュ値及び統合ハッシュ値と、情報処理装置11aのデータ保持構造に対応したハッシュ値及び統合ハッシュ値との比較を行う場合を例示する。

【0085】

図13は、文書データ110aの内容と文書データ110bの内容が異なる場合の図である。この例では、文書本体データ111aの内容と文書本体データ111bの内容とが異なり、またセキュリティ情報114aの内容とセキュリティ情報114bの内容とが異なっている。この場合、文書本体データ111aのハッシュ値と、文書本体データ111bのハッシュ値が異なる。またセキュリティ情報114aのハッシュ値と、セキュリティ情報114bのハッシュ値も異なる。そのため、文書データ110aの文書ハッシュ値と、文書データ110bの文書ハッシュ値が異なることとなり、さらにフォルダ210aのフォルダハッシュ値と、フォルダ210bのフォルダハッシュ値とが異なることと

なる。その結果、最上位フォルダであるフォルダ 2 2 0 a のフォルダハッシュ値と、フォルダ 2 2 0 b のフォルダハッシュ値とが異なることとなる。

【 0 0 8 6 】

この場合、データ判別処理において文書管理装置 1 および情報処理装置 1 1 a のデータ保持構造に基づき、最上位フォルダであるフォルダ 2 2 0 a のフォルダハッシュ値とフォルダ 2 2 0 b のフォルダハッシュ値との比較が行われる。これらフォルダハッシュ値が互いに異なる値であるので、次に、文書データ 1 4 0 a , 1 5 0 a の文書ハッシュ値と、文書データ 1 4 0 b , 1 5 0 b の文書ハッシュ値との比較が個別に行われる。これらはいずれも等しい値となるので、次に、その下位のフォルダ 2 1 0 a , 2 1 0 b が特定され、フォルダ 2 1 0 a のフォルダハッシュ値とフォルダ 2 1 0 b のフォルダハッシュ値との比較が行われる。これらフォルダハッシュ値が互いに異なる値である。そのため、次に、文書データ 1 1 0 a , 1 2 0 a , 1 3 0 a の文書ハッシュ値と、文書データ 1 1 0 b , 1 2 0 b , 1 3 0 b の文書ハッシュ値との比較が個別に行われる。その結果、文書ハッシュ値が互いに異なる文書データ 1 1 0 a , 1 1 0 b が特定される。そのため、文書データ 1 1 0 a を構成する文書構成要素 1 1 1 a , 1 1 2 a , 1 1 3 a , 1 1 4 a の各ハッシュ値と、文書データ 1 1 0 b を構成する文書構成要素 1 1 1 b , 1 1 2 b , 1 1 3 b , 1 1 4 b の各ハッシュ値とが比較され、文書管理装置 1 で保持されている文書本体データ 1 1 1 a とセキュリティ情報 1 1 4 a との 2 つの文書構成要素が、情報処理装置 1 1 a で保持している文書構成要素と異なった内容であることを特定することができる。

10

【 0 0 8 7 】

この例では、文書管理装置 1 と情報処理装置 1 1 a のそれぞれが、4 つの文書構成要素からなる 5 つの文書データを保持している。そのため、文書管理装置 1 と情報処理装置 1 1 a のそれぞれで保持される文書構成要素の数は 2 0 個である。この場合において、例えば従来のような比較処理を行えば、異なる内容の文書構成要素を全て特定するために 2 0 回の比較処理が必要になる。これに対し、本実施形態のように、ハッシュ値および統合ハッシュ値を用いた比較処理によれば、フォルダ 2 2 0 a 、文書データ 1 4 0 a 、文書データ 1 5 0 a 、フォルダ 2 1 0 a 、文書データ 1 3 0 a 、文書データ 1 2 0 a 、文書データ 1 1 0 a 、文書本体データ 1 1 1 a 、サムネイル 1 1 2 a 、データベース 1 1 3 a 、およびセキュリティ情報 1 1 4 a のハッシュ値又は統合ハッシュ値の比較処理を行えば良いので、合計 1 1 回の比較処理で、文書管理装置 1 と情報処理装置 1 1 a とで保持される異なる内容の文書構成要素を全て特定することが可能である。

20

30

【 0 0 8 8 】

次に、図 1 4 は、文書データ 1 4 0 a の文書構成要素であるサムネイル 1 0 7 の内容が異なっている場合の図である。この例では、サムネイル 1 0 7 a のハッシュ値と、サムネイル 1 0 7 b のハッシュ値とが異なっており、それに伴って文書データ 1 4 0 a の文書ハッシュ値と、文書データ 1 4 0 b の文書ハッシュ値とも異なっている。

【 0 0 8 9 】

この場合、データ判別処理において文書管理装置 1 および情報処理装置 1 1 a のデータ保持構造に基づき、最上位フォルダであるフォルダ 2 2 0 a のフォルダハッシュ値とフォルダ 2 2 0 b のフォルダハッシュ値との比較が行われる。これらフォルダハッシュ値が互いに異なる値であるので、次に、文書データ 1 4 0 a , 1 5 0 a の文書ハッシュ値と、文書データ 1 4 0 b , 1 5 0 b の文書ハッシュ値との比較が個別に行われる。そして文書データ 1 4 0 a の文書ハッシュ値と、文書データ 1 4 0 b の文書ハッシュ値とが異なることになる。そして文書データ 1 4 0 a を構成する文書構成要素 1 0 6 a , 1 0 7 a , 1 0 8 a , 1 0 9 a の各ハッシュ値と、文書データ 1 4 0 b を構成する文書構成要素 1 0 6 b , 1 0 7 b , 1 0 8 b , 1 0 9 b の各ハッシュ値とが比較され、文書管理装置 1 で保持されているサムネイル 1 0 7 a の 1 つの文書構成要素が、情報処理装置 1 1 a で保持している文書構成要素と異なった内容であることを特定することができる。

40

【 0 0 9 0 】

この例では、ハッシュ値および統合ハッシュ値を用いた比較処理により、上述のように

50

フォルダ 220 a、文書データ 140 a、文書データ 150 a、フォルダ 220 a、文書本体データ 106 a、サムネイル 107 a、データベース 108 a、およびセキュリティ情報 109 a のハッシュ値又は統合ハッシュ値の比較処理を行えば良いので、合計 8 回の比較処理で、文書管理装置 1 と情報処理装置 11 a とで保持される異なる内容の文書構成要素を全て特定することが可能である。

【0091】

次に、図 15 は、全ての文書データの内容が等しい場合の図である。この場合、文書管理装置 1 と情報処理装置 11 a とで保持される、フォルダ、文書データ、および文書構成要素のハッシュ値又は統合ハッシュ値は全て等しい。この例では、フォルダ 220 a のフォルダハッシュ値と、フォルダ 220 b のフォルダハッシュ値とが比較されると、それら 10

【0092】

このように本実施形態では、文書管理装置 1 と情報処理装置 11 とで異なる内容の文書データを保持しているか否かの判別を行う際、または、異なる内容の文書構成要素を特定する際に、ハッシュ値および統合ハッシュ値を用いることにより、従来よりも少ない比較回数で判別や特定を行うことができるようになる。そのため、文書管理システムのパフォーマンスの低下を回避することが可能である。そして、このような比較が行われた後、この比較結果に基づいて、文書管理装置 1 から情報処理装置 11 へ異なる内容の文書構成要素 20

【0093】

(変形例)

以上、本発明に関するいくつかの実施形態について説明したが、本発明は上述した内容に限られるものではなく、種々の変形例が適用可能である。以下でいくつかの変形例を挙げる。

【0094】

例えば上記実施形態においては、特に図 11 において示されるように、情報処理装置 11 がハッシュ値および統合ハッシュ値の送信要求を文書管理装置 1 に対して行うのは、ユーザがフォルダを選択してクリック操作した際に行われるように設定されているが、これに限らず、例えば定期的に行ってもよいし、また情報処理装置 11 の電源がオンとなった時に行うようにしてもよい。また、フォルダのクリック操作に限られず、例えば文書データの選択操作が行われた場合に行うこととしてもよい。またその他の入力操作であってもよい。

【0095】

また上記実施形態においては、ハッシュ値を生成するハッシュ関数と、文書ハッシュ値を生成するハッシュ関数と、フォルダハッシュ値を生成するハッシュ関数とが同一のハッシュ関数である場合を例示したが、これに限られるものではない。

【0096】

また上記実施形態においては、ハッシュ値および統合ハッシュ値の生成は、文書管理装置 1 が行うものとされているが、情報処理装置 11 が行ってもよいし、文書管理装置 1 と情報処理装置 11 との双方が行うようにしてもよい。

【0097】

また上記実施形態においては、文書管理装置 1 と情報処理装置 11 とで異なる内容の文書データを保持しているか否かのデータ判別処理を情報処理装置 11 で行う場合を例示したが、これに限られるものでもない。例えば、文書管理装置 1 で上述したデータ判別処理 (図 12) を行うようにしても良い。ただし、この場合、文書管理装置 1 が、情報処理装置 11 に対してハッシュ値および統合ハッシュ値の送信要求することにより、情報処理装 50

置 1 1 におけるデータ保持構造に対応したハッシュ値および統合ハッシュ値を情報処理装置 1 1 から取得することが必要になる。

【 0 0 9 8 】

また上記実施形態においては、データ判別処理を行う際、文書管理装置 1 から情報処理装置 1 1 に対してハッシュ値と統合ハッシュ値の双方をまとめて送信する場合を例示したが、これに限られるものでもない。すなわち、まず最上位フォルダに対応するフォルダハッシュ値を送信し、そのフォルダハッシュ値が異なる値と判定されれば、次に文書ハッシュ値や下位フォルダのフォルダハッシュ値などを送信するようにして、統合ハッシュ値とハッシュ値とを段階的に送信する構成としてもよい。この場合、ハッシュ値送受信時におけるネットワーク 4 のトラフィック量を軽減することができるという利点がある。

10

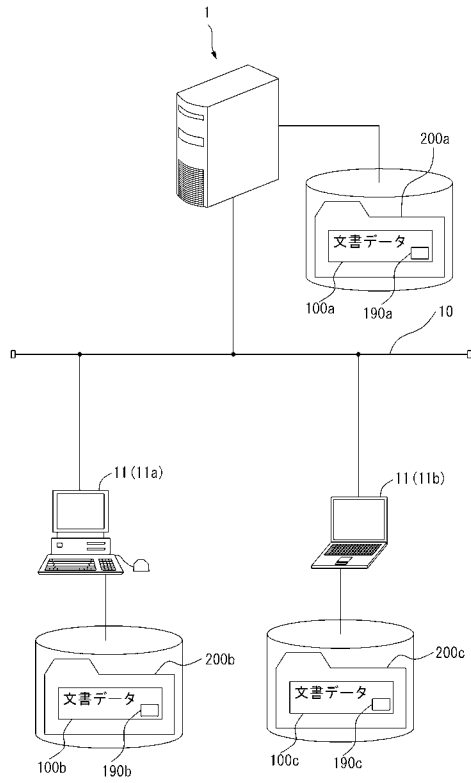
【符号の説明】

【 0 0 9 9 】

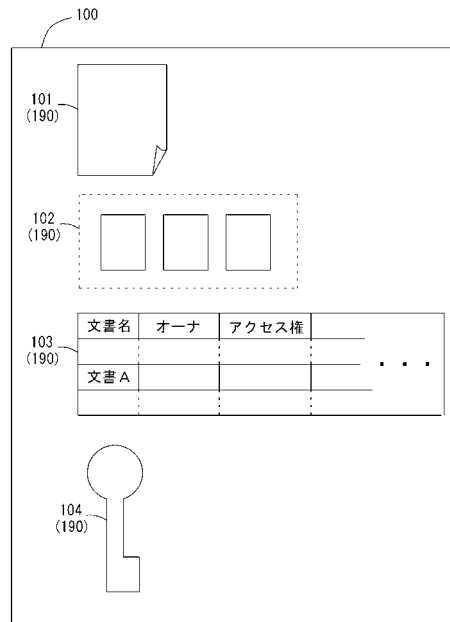
- 1 文書管理装置
- 1 1 情報処理装置
- 2 0、4 0 制御部
- 3 0 記憶装置（文書データ記憶手段）
- 5 0 記憶装置（データ記憶手段）
- 6 1 ハッシュ値生成部（ダイジェストデータ生成手段）
- 6 2 統合ハッシュ値生成部（統合ダイジェストデータ生成手段）
- 7 2 データ判別部（データ判別手段）
- 1 0 0 , 1 0 0 a , 1 0 0 b 文書データ
- 1 0 1 文書本体データ（文書構成要素）
- 1 0 2 サムネイル（文書構成要素）
- 1 0 3 データベース（文書構成要素）
- 1 0 4 セキュリティ情報（文書構成要素）
- 1 9 0 文書構成要素
- 2 0 0 , 2 0 0 a , 2 0 0 b フォルダ（記憶領域、記憶部）

20

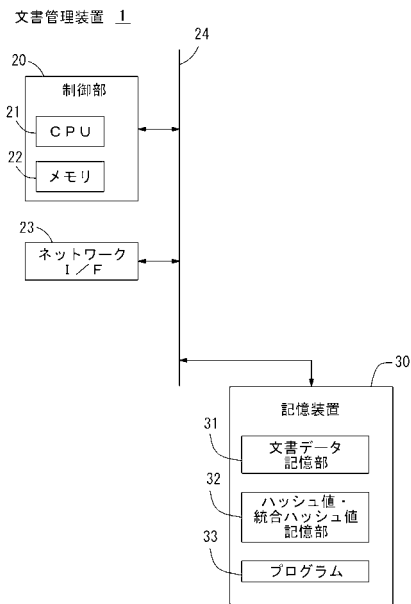
【 図 1 】



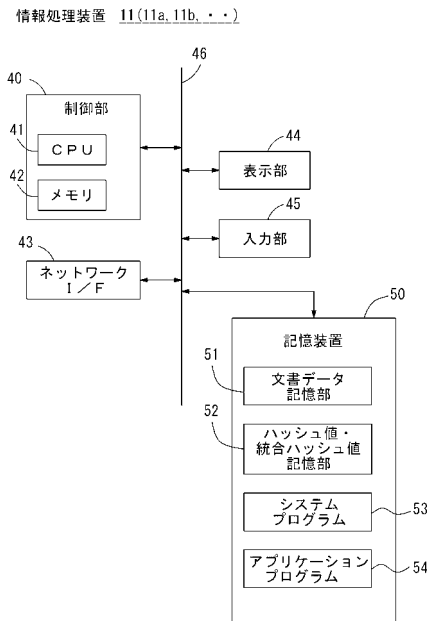
【 図 2 】



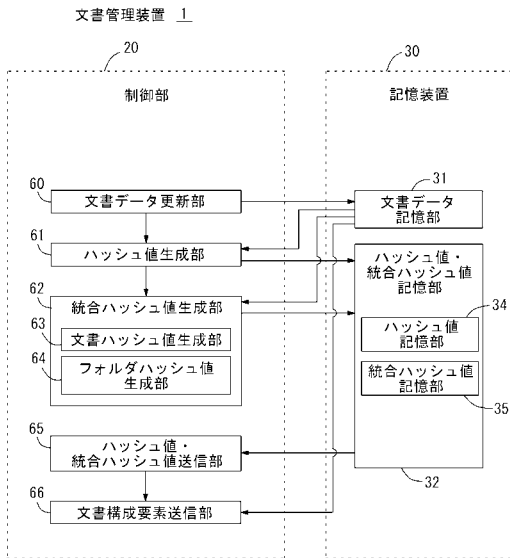
【 図 3 】



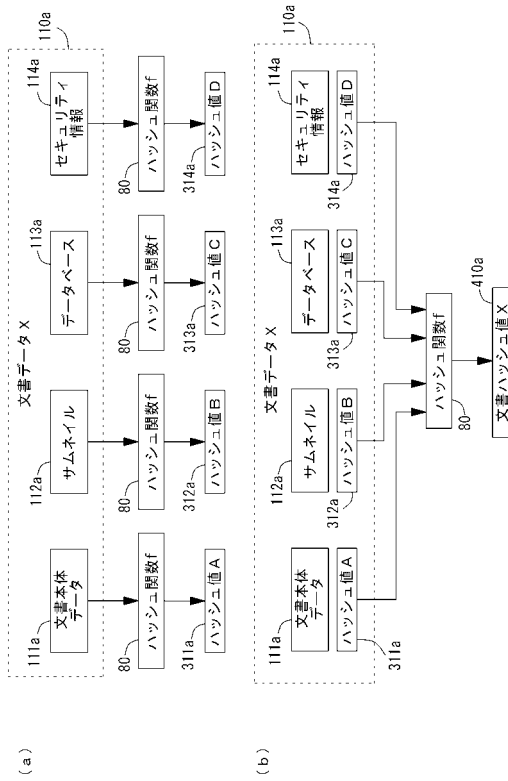
【 図 4 】



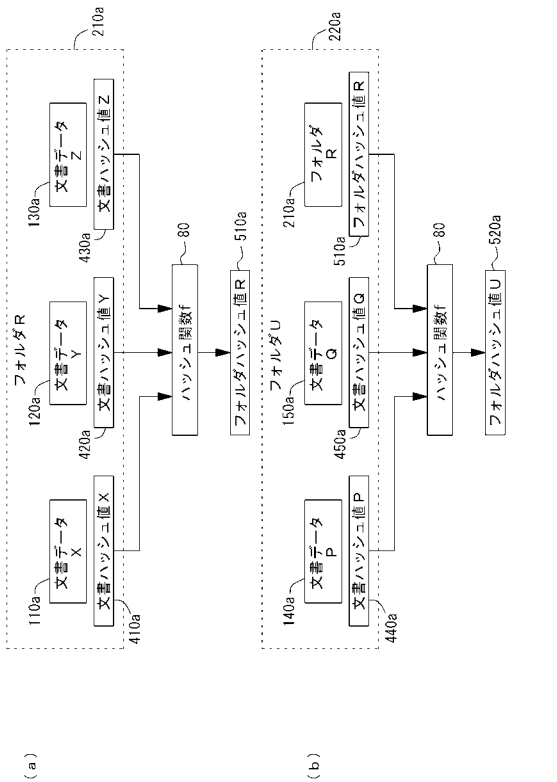
【図5】



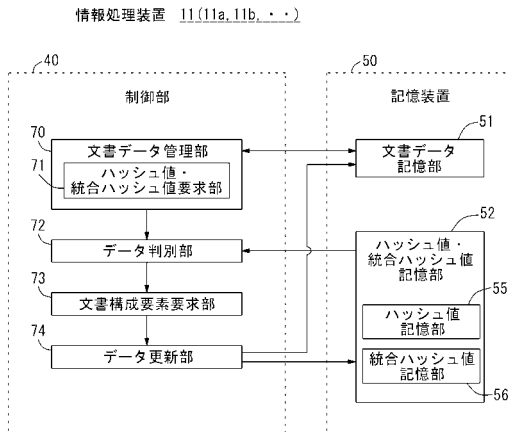
【図6】



【図7】



【図8】

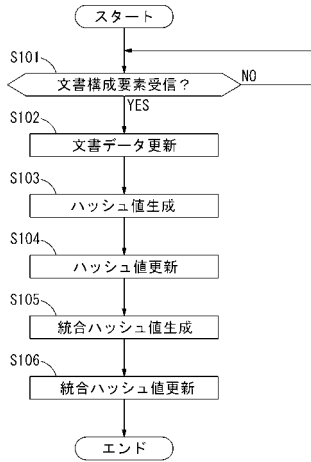


(a)

(b)

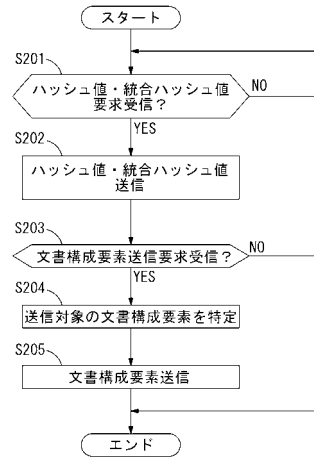
【図9】

文書管理装置 1



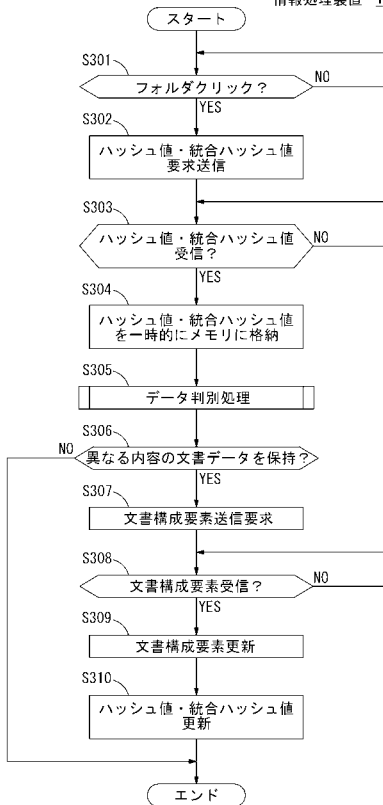
【図10】

文書管理装置 1



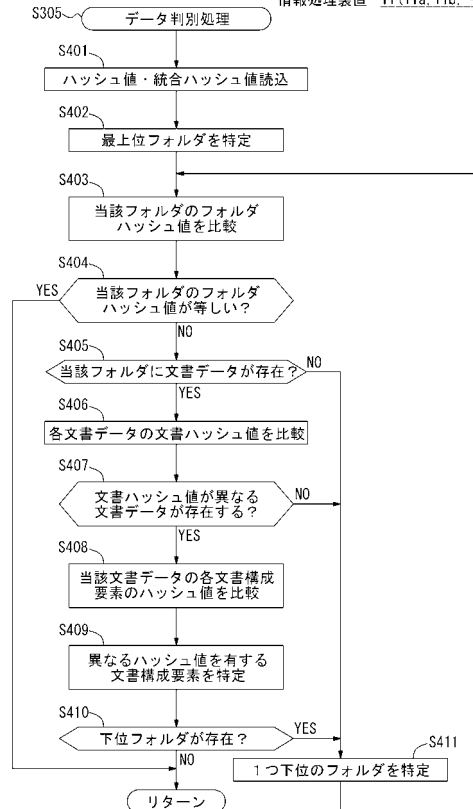
【図11】

情報処理装置 11(11a,11b,...)

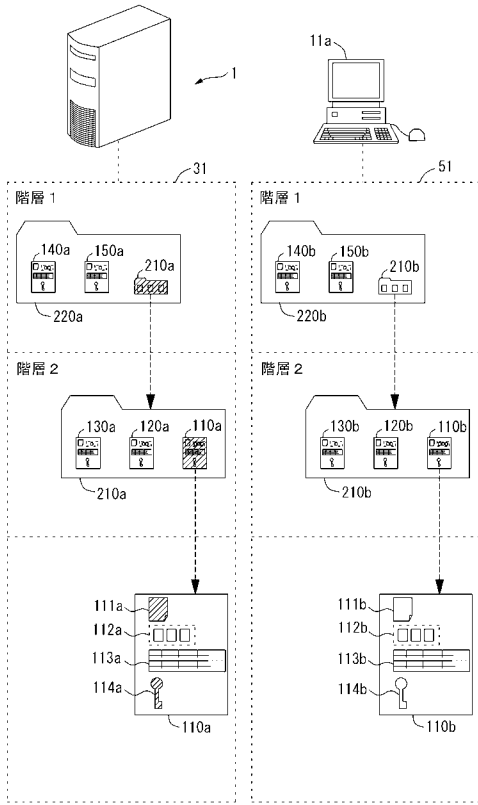


【図12】

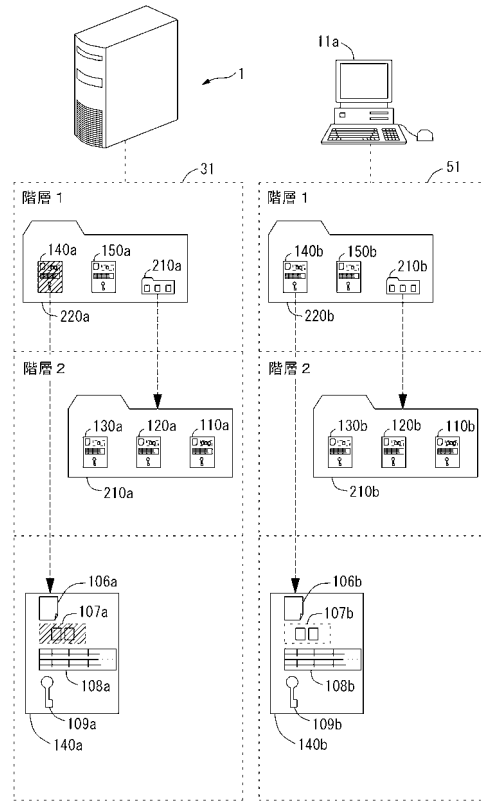
情報処理装置 11(11a,11b,...)



【図 1 3】



【図 1 4】



【図 1 5】

