

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2011-523133
(P2011-523133A)

(43) 公表日 平成23年8月4日(2011.8.4)

(51) Int.Cl.		F I			テーマコード (参考)
G06F 17/21	(2006.01)	G06F 17/21	536		5B082
G06F 12/00	(2006.01)	G06F 12/00	511C		5B109

審査請求 未請求 予備審査請求 未請求 (全 23 頁)

(21) 出願番号 特願2011-511963 (P2011-511963)
 (86) (22) 出願日 平成21年6月5日 (2009.6.5)
 (85) 翻訳文提出日 平成23年2月1日 (2011.2.1)
 (86) 国際出願番号 PCT/CN2009/072147
 (87) 国際公開番号 W02009/146657
 (87) 国際公開日 平成21年12月10日 (2009.12.10)
 (31) 優先権主張番号 200810114437.2
 (32) 優先日 平成20年6月5日 (2008.6.5)
 (33) 優先権主張国 中国 (CN)

(71) 出願人 507231932
 北大方正集▲団▼有限公司
 PEKING UNIVERSITY F
 OUNDER GROUP CO., L
 TD
 中華人民共和国北京市▲海▼淀区成府路2
 98号中▲関▼村方正大厦5▲層▼
 5 Floor, Zhongguanc
 un Founder Building
 , No. 298, Chengfu R
 oad, Haidian Distri
 ct, Beijing 100871,
 China

最終頁に続く

(54) 【発明の名称】 レイアウトファイルの構造処理方法及び装置

(57) 【要約】

レイアウトファイルの構造処理方法及び装置は、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得し、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行い、コンテンツブロック分割の結果に基づいてレイアウトファイルのドキュメントフロー情報を生成する。

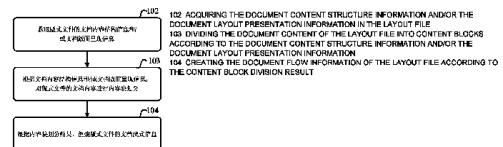


図 1 / Fig. 1

【特許請求の範囲】**【請求項 1】**

レイアウトファイルのドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報を取得するステップと、

前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報に基づいて前記レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うステップと、

前記コンテンツブロック分割の結果に基づいて前記レイアウトファイル用のドキュメントフロー情報を生成するステップと、

前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して構造処理を行うステップと、

を有することを特徴とするレイアウトファイルの構造処理方法。

10

【請求項 2】

(1) 前記レイアウトファイルのドキュメントコンテンツの 1 つ以上のソースに従って、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報を得るステップ、

(2) 前記レイアウトファイルを計算することにより、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報を得るステップ、

(3) 外部からの入力を受けることにより、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報を得るステップ、

20

上記(1)、(2)、(3)の少なくともいずれか 1 つのステップによって、前記レイアウトファイルの前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報を得ることを特徴とする請求項 1 に記載のレイアウトファイルの構造処理方法。

【請求項 3】

前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報に基づいて、前記レイアウトファイルにおけるドキュメントコンテンツに対してコンテンツブロック分割を行うステップは、

レイアウトファイルを形成する複数のコマンド文を、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報によってグループ毎のコマンド文の数が決定されるように、複数のグループに分割し、分割された各グループのコマンド文を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得するステップ、

30

または、

レイアウトファイルを形成する複数の対象を、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報によってグループ毎の対象の数が決定されるように、複数のグループに分割し、分割された各グループの対象を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得するステップ、

40

または、

レイアウトファイルを形成する複数のコンテンツを、前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報によってグループ毎のコンテンツの数が決定されるように、複数のグループに分割し、分割された各グループのコンテンツを一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得するステップにより行うことを特徴とする請求項 1 に記載のレイアウトファイルの構造処理方法。

【請求項 4】

前記ドキュメントコンテンツの構造情報及び / 又はドキュメントレイアウトの表示情報に基づいて、前記レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うステップは、

50

前記レイアウトファイルを形成する、コンテンツの参照シーケンスを取得し、

前記参照シーケンスを、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によって各サブ参照シーケンスのシーケンス数が決定されるように、複数のサブ参照シーケンスに分割するとともに、分割された各サブ参照シーケンスを一つのコンテンツブロックとし、

得られたコンテンツブロック分割の結果を記述することによってコンテンツブロック分割結果情報を取得することを特徴とする請求項1に記載のレイアウトファイルの構造処理方法。

【請求項5】

前記コンテンツブロック分割の結果に基づいて、前記レイアウトファイルのドキュメントフロー情報を生成するステップは、

前記コンテンツブロック分割結果情報と、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との関連を記述することによって、前記ドキュメントフロー情報を得ることを特徴とする請求項3または4に記載のレイアウトファイルの構造処理方法。

【請求項6】

前記コンテンツブロック分割結果情報と前記ドキュメントフロー情報とは、構造化マーキング言語を利用して記述するものであることを特徴とする請求項5に記載のレイアウトファイルの構造処理方法。

【請求項7】

前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して構造処理を行うステップは、

レイアウトファイルのコンテンツに対する検索操作、構造化記憶操作、修正操作、抽出操作、及び、レイアウトリアレンジ操作の内の少なくとも一つの操作を有し、

すなわち、

前記ドキュメントフロー情報に記述された、前記コンテンツブロック分割結果情報と前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との間の関連に基づいて、前記レイアウトファイルのコンテンツブロックと、前記ドキュメントコンテンツの構造及び/又はドキュメントレイアウトに対して行われる操作であることを特徴とする請求項5に記載のレイアウトファイルの構造処理方法。

【請求項8】

レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得するオリジナル情報取得モジュールと、

前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいて、前記レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うコンテンツブロック分割モジュールと、

前記コンテンツブロック分割の結果に基づいて、前記レイアウトファイル用のドキュメントフロー情報を生成するドキュメントフロー情報記述モジュールと、

前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して構造処理を行う構造処理モジュールと、

を備えることを特徴とするレイアウトファイルの構造処理装置。

【請求項9】

(1) 前記レイアウトファイルのドキュメントコンテンツの1つ以上のソースに従って、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を得るステップ、

(2) 前記レイアウトファイルを計算することにより、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を得るステップ、

(3) 外部からの入力を受けることにより、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を得るステップ、

前記オリジナル情報取得モジュールは、上記(1)、(2)、(3)の少なくともいずれ

10

20

30

40

50

れか1つのステップによって、前記レイアウトファイルの前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を得ることを特徴とする請求項8に記載のレイアウトファイルの構造処理装置。

【請求項10】

前記コンテンツブロック分割モジュールは、

レイアウトファイルを形成する複数のコマンド文を、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコマンド文の数が決定されるように、複数のグループに分割し、分割された各グループのコマンド文を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得し、

10

または、

レイアウトファイルを形成する複数の対象を、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎の対象の数が決定されるように、複数のグループに分割し、分割された各グループの対象を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得し、

または、

レイアウトファイルを形成する複数のコンテンツを、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコンテンツの数が決定されるように、複数のグループに分割し、分割された各グループのコンテンツを一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得することを特徴とする請求項8に記載のレイアウトファイルの構造処理装置。

20

【請求項11】

前記コンテンツブロック分割モジュールは、

前記レイアウトファイルを形成する、コンテンツの参照シーケンスを取得するコンテンツ参照シーケンス取得サブモジュールと、

前記コンテンツの参照シーケンスを、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によって各コンテンツのサブ参照シーケンスのシーケンス数が決定されるように、複数のコンテンツのサブ参照シーケンスに分割するとともに、分割された各コンテンツのサブ参照シーケンスを一つのコンテンツブロックとするコンテンツブロック分割サブモジュールと、

30

取得されたコンテンツブロック分割の結果を記述することによってコンテンツブロック分割結果情報を取得する記述サブモジュールと、
を備えることを特徴とする請求項8に記載のレイアウトファイルの構造処理装置。

【請求項12】

前記ドキュメントフロー情報記述モジュールは、

前記コンテンツブロック分割結果情報と、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との関連を記述することによって、前記ドキュメントフロー情報を得ることを特徴とする請求項10または11に記載のレイアウトファイルの構造処理装置。

40

【請求項13】

前記コンテンツブロック分割結果情報と前記ドキュメントフロー情報とは、構造化マーキング言語を利用して記述するものであることを特徴とする請求項12に記載のレイアウトファイルの構造処理装置。

【請求項14】

前記構造処理モジュールが前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して行う構造処理は、

レイアウトファイルのコンテンツに対する検索操作、構造化記憶操作、修正操作、抽出操作、及び、レイアウトリアレンジ操作により行い、

50

すなわち、

前記ドキュメントフロー情報に記述された、前記コンテンツブロック分割結果情報と前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との間の関連に基づいて、前記レイアウトファイルのコンテンツブロックと、前記ドキュメントコンテンツの構造及び/又はドキュメントレイアウトに対して行われる操作であることを特徴とする請求項1に記載のレイアウトファイルの構造処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンピュータ情報処理技術分野においてレイアウトファイルの構造処理方法及び装置に関するものである。 10

【背景技術】

【0002】

通常レイアウトファイルは、ユーザ定義の座標系において各ドキュメントの表示位置や表示サイズ等を明記するいわゆる完全記述方法が採用されているため、プリントされた結果としてのドキュメントがコンピュータでの閲覧時と一致する。それと共に、ドキュメントが忠実に再現できるように、異なるコンピュータにもかかわらず表示が一致する特性を持たせる。例えば、代表的なレイアウトファイルとして、PDFファイルが取上げられる。レイアウトファイルは、比較的安定するという特性があるため、電子公文書や、電子書籍や、電子ジャーナルや、電子新聞紙などの領域において電子ドキュメントの最終的な開示・伝播形式として広く用いられている。 20

【0003】

最近、コンピュータ技術の普及や情報技術の進歩に従い、レイアウトファイルの数は爆発的に増加してくると共に、例えば、PDA、スマートフォンなど、クライアント端末の種類も増加してくる。異なる種類のクライアント端末を利用してレイアウトファイルを支障なく閲覧できることが要求される。よって、クライアント端末では、レイアウトファイルの固定した表示という制限を克服し、表示装置の表示画面サイズに基づいてレイアウトファイルのコンテンツを再編する必要がある。

【0004】

本発明の発明者らは、レイアウトファイルにおいて各ドキュメントの表示位置やサイズが絶対値によって正確に指定されるため、編集などの構造処理が大変困難となるあることを見出した。例えば、ドキュメントコンテンツが修正される度に、レイアウトを再計算し、ドキュメント全体のレイアウト情報をリライトする必要がある。しかし、絶対値だけで記述されるドキュメントの表示位置やサイズに対するレイアウトの再計算やレイアウト情報のリライトが非常に困難である。また、レイアウトファイルのコンテンツに対する検索、構造化記憶、修正、抽出などの編集操作も非常に煩雑なものである。 30

【発明の概要】

【発明が解決しようとする課題】

【0005】

本発明は、前記した問題を解決するために、レイアウトファイルのコンテンツが修正された後にドキュメント構造やレイアウト表示などの情報の更新が容易になるとともに、レイアウトファイルのコンテンツ検索、構造化記憶、修正、抽出、リアレンジなどの操作が実現できるように、レイアウトファイルのドキュメントフロー情報を記述してレイアウトファイルに対して構造処理を行うレイアウトファイルの構造処理方法及び装置を提供する。 40

【課題を解決するための手段】

【0006】

本発明は、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得し、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいてレイアウトファイルのドキュメントコンテン 50

ツに対してコンテンツブロック分割を行い、コンテンツブロック分割の結果に基づいてレイアウトファイルのドキュメントフロー情報を生成するレイアウトファイルの構造処理方法を提供する。

【0007】

また、本発明は、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得するオリジナル情報取得モジュールと、前記ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいて、前記レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うコンテンツブロック分割モジュールと、前記コンテンツブロック分割の結果に基づいて、前記レイアウトファイルのドキュメントフロー情報を生成するドキュメントフロー情報記述モジュールと、を備えるレイアウトファイルの構造処理装置を提供する。

10

【発明の効果】

【0008】

上記実施態様は、少なくとも1つの以下の効果を有する。

【0009】

本発明の構成によれば、レイアウトファイルのドキュメントフロー情報を取得し、取得したドキュメントフロー情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行い、コンテンツブロック分割結果情報を記述し、コンテンツブロック分割結果情報に基づいて、コンテンツブロック分割された当該レイアウトファイルに基づくドキュメントフロー情報を記述することによって、レイアウトファイルに対する構造処理を容易にできる。例えば、レイアウトファイルはドキュメントコンテンツが修正された後、ドキュメント構造、レイアウト表示などの情報の更新が簡単にできる。また、レイアウトファイルのコンテンツに対する検索と、構造化記憶と、修正と、抽出と、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

20

【0010】

ここで説明される図面は、本発明に対して更に理解するためであり、本願の一部となる。本発明を示す実施例及びその説明は本発明を解釈するに用いられ、本発明にたいする不当な制限を構成しない。

【図面の簡単な説明】

【0011】

30

【図1】本発明の実施例におけるレイアウトファイルの構造処理方法を示すフローチャート図である。

【図2】本発明の実施例において、コンテンツブロック分割されたレイアウトファイルに基づくドキュメントフロー情報を説明する図である。

【図3】本発明の実施例におけるレイアウトファイル及びそのコンテンツ記述の図である。

【図4】本発明の実施例において、図3に示すレイアウトファイルのコンテンツブロック分割方法を示す図である。

【図5】本発明の実施例において、図3に示すレイアウトファイルのコンテンツ分割の結果情報を記述することを示す図である。

40

【図6】本発明の実施例において、図3に示すコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報におけるドキュメントの構造情報を示す図である。

【図7】本発明の実施例において、図3に示すコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報におけるドキュメントレイアウトの自己適応表示情報を示す図である。

【図8】本発明の実施例において、図3に示すコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報におけるドキュメントレイアウトのコンテンツのリアレンジを示す図である。

【図9】本発明の実施例におけるレイアウトファイルの構造処理装置の構成を示す図である。

50

【図10】本発明の実施例において、コンテンツ参照シーケンス分割の方法でレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うことを示す図である。

【発明を実施するための形態】

【0012】

以下は、図面に参照して本発明の具体的な実施方法を詳細に説明する。

【0013】

本発明の実施例では、先ず、レイアウトファイルのオリジナル情報を取得し、取得したオリジナル情報に基づいてレイアウトファイルのドキュメントコンテンツを複数のコンテンツブロックに分割し、次に、取得したコンテンツブロック分割の結果に基づいて、コンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報を記述することによって、レイアウトファイルに対する構造処理を容易にできる。例えば、レイアウトファイルはドキュメントコンテンツが修正された後、ドキュメント構造、レイアウト表示などの情報の更新が簡単にできる。また、レイアウトファイルのコンテンツに対する検索と、構造化記憶と、修正と、抽出と、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

10

【0014】

以下は、図面に基づいて本発明の実施例を詳細に説明する。

【0015】

図1は、レイアウトファイルの構造処理方法を示すフローチャート図である。具体的には、下記のステップを備える。

20

【0016】

ステップ102：レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得する。ここで、レイアウトファイルとは、レイアウトファイルの全体でもよいが、レイアウトファイルの一頁または何頁などでもよい。また、レイアウトファイルのオリジナル情報とは、レイアウトファイルにおけるドキュメントコンテンツの構造情報及び/又はレイアウトファイルにおけるドキュメントレイアウトの自己適応表示情報である。詳しくは、下記(1)～(3)を含むが、これに限らない。すなわち、

30

(1)ドキュメントコンテンツの構造情報：ドキュメントの章節情報と、各章節内のコンテンツブロックの順序と、コンテンツブロック内の各エンティティ(entity)の順序等。

(2)閲覧クルー情報(reading clue information)：具体的には、上記したドキュメントコンテンツの構造情報により提供される閲覧順序以外、必要によって提供される他の閲覧順序情報が含まれる。ユーザに提供される選択可能な閲覧順序情報である。閲覧クルー情報はレイアウトファイルのドキュメントコンテンツ全体の閲覧順序情報でもよいが、レイアウトファイルの一部分のドキュメントコンテンツの閲覧順序情報でもよい。

(3)レイアウト情報：具体的には、レイアウトファイルのレイアウトリアレンジを行う時にエンティティの最終的な表示効果を決める情報であり、エンティティ自身又はコンテンツブロック自身のレイアウトプロパティと、同じコンテンツブロックにおけるエンティティ同士又はコンテンツブロック同士のレイアウト関係とが含まれる。例えば、画像を指定する文字設定方式や、複数のコンテンツブロックを指定するコラム情報などである。上記したレイアウトリアレンジとは、レイアウトサイズ又はレイアウトコンテンツが変化する場合、一定の規則に従ってレイアウトにおける各エンティティを再編してレイアウト表示効果とする過程である。

40

【0017】

ここで、本発明の実施例において、下記(1)、(2)、(3)のいずれか又はこれらの組み合わせによって、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報が取得される。

【0018】

(1)ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情

50

報が含まれている電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、レイアウトファイルのさまざまなドキュメントコンテンツのソースを解析することによってレイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を直接に取得することができる。例えば、HTMLや、Microsoft Wordなど、レイアウトファイルに対応する、一部のドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を含んでいる電子ドキュメントの場合は、当該ドキュメントのドキュメント処理システムを利用してそのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を抽出することができる。特に、Microsoft Wordドキュメントに対して、Officeの自動化オブジェクトを利用してそのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。

10

【0019】

(2)ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報が含まれていない電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、さまざまな識別アルゴリズム又はインテリジェント理解アルゴリズムを利用してレイアウトファイルに対して計算することによってレイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。例えば、ドキュメント分析及びドキュメント理解に基づく処理システムを介してレイアウトファイルに対して計算することによって当該レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。

20

【0020】

(3)ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報が含まれていない電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、外部からユーザによって入力された、レイアウトファイルに関するドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を受け付けることによって、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得してもいい。例えば、ユーザはグラフィックインターフェイス付きコンピュータアプリケーションを介してレイアウトファイルのドキュメントコンテンツに対してマークすることによって当該レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報の入力を実現できる。

30

【0021】

ステップ103：ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行う。

【0022】

レイアウトファイルのドキュメントコンテンツを複数のコンテンツブロックに分割することは、レイアウトファイルを直接に組織することに基づいて実現できる。即ち、レイアウトファイルにおいて各グループのコマンド文、または、各グループの対象(object)、又は、各段落のコンテンツの記述を一つのコンテンツブロックユニットとして、当該レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行う。具体的には、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいて、文の番号、文の長さ、文のオフセット量、または、対象ID、対象のオフセット量、又は、コンテンツID、コンテンツのオフセット量、または、ある特殊な記号等を使用して、当該レイアウトファイルのドキュメントコンテンツを異なるコンテンツブロックに分割してもいい。ここで、分割されたコンテンツブロック同士は、重なり合うようにされてもいい。また、分割されたコンテンツブロック毎に唯一なIDが与えられてもいい。

40

【0023】

50

例えば、レイアウトファイルを形成する複数のコマンド文を、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコマンド文の数が決定されるように、複数のグループに分割し、分割された各グループのコマンド文を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得する。

【 0 0 2 4 】

また、レイアウトファイルを形成する複数の対象を、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎の対象の数が決定されるように、複数のグループに分割し、分割された各グループの対象を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツ

10

【 0 0 2 5 】

また、レイアウトファイルを形成する複数のコンテンツを、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコンテンツの数が決定されるように、複数のグループに分割し、分割された各グループのコンテンツを一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得する。

【 0 0 2 6 】

なお、レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行うことは、コンテンツの参照シーケンスを分割することによっても実現される。具体的には、レイアウトファイルを形成する、コンテンツの参照シーケンスを取得する。ここで、コンテンツの参照シーケンスとは、レイアウトファイルのドキュメントコンテンツにおけるテキストや、画像や、テーブルなどのさまざまなエンティティが、所定の順序に従ってシーケンスされる。この順序は、エンティティがレイアウトファイルのコンテンツのデータフローにおける順序でもよいが、ドキュメントツリーでのトラバーサル順序でもよい。取得した参照シーケンスを、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によって各サブ参照シーケンスのシーケンス数が決定されるように、順序付けの複数のサブ参照シーケンスに分割するとともに、分割された各サブ参照シーケンスを一つのコンテンツブロックとし、次に、得られたコンテンツブロック分割の結果を記述することによってコンテンツブロック分割結果情報を取得する。これらの

20

30

【 0 0 2 7 】

上記したコンテンツブロック分割の結果に基づいて当該レイアウトファイルのコンテンツブロック分割結果情報を記述する。構造化マークアップ言語（例えば、XML言語や、SGML言語等）を利用して記述してもよいが、これに限定されない。

40

【 0 0 2 8 】

ステップ104：コンテンツブロック分割の結果に基づいてレイアウトファイルのドキュメントフロー情報を生成する。

【 0 0 2 9 】

コンテンツブロックの分割された当該レイアウトファイルに基づくドキュメントフロー情報を記述する。具体的には、各コンテンツブロック自身のドキュメントフロー情報と、各コンテンツブロック間の関係、例えば、ドキュメントの構造情報や、閲覧クルー情報や、レイアウト情報等を記述する。ここで、XML言語や、SGML言語等でコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報を記述してもよいが、こ

50

れに限定されない。当該レイアウトファイルは、例えばPDFドキュメント等のレイアウトファイルである。

【0030】

具体的には、上記の記述によって取得されたコンテンツブロック分割結果情報と、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報とを関連付けて、関連付けられたコンテンツブロック分割結果情報とドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報とを対応して記憶してもよい。また、コンテンツブロック分割結果情報とドキュメントの構造情報及び/又はドキュメントレイアウトの表示情報は、レイアウトファイルとはそれぞれ記憶されてもよいが、レイアウトファイルに組み込まれてレイアウトファイルのデータブロックとしてもよい。

10

【0031】

構造化マークアップ言語を利用して、得られたコンテンツブロック分割結果情報とドキュメントフロー情報を記述することができる。

【0032】

ステップ105：前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して構造処理を行う。

【0033】

レイアウトファイルのドキュメントフロー情報を取得し、取得されたドキュメントフロー情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行い、コンテンツブロック分割結果情報を記述し、コンテンツブロック分割結果情報に基づいて、コンテンツブロック分割された当該レイアウトファイルの基づくドキュメントフロー情報を記述し、よって、レイアウトファイルに対する構造処理が容易に行われることができる。例えば、レイアウトファイルはドキュメントコンテンツが修正された後、ドキュメント構造、レイアウト表示などの情報の更新が簡単にできる。また、レイアウトファイルのコンテンツに対する検索と、構造化記憶と、修正と、抽出と、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

20

【0034】

図2は、本発明に係る方法に従って、コンテンツブロック分割されたレイアウトファイルに基づくドキュメントフロー情報を記述することを示す図である。レイアウトファイル205のドキュメントコンテンツを複数のコンテンツブロックに分割し、構造化マークアップ言語でコンテンツブロック分割結果情報204を記述し、コンテンツブロック分割結果情報204に基づいて、コンテンツ分割された当該レイアウトファイル205に基づくドキュメントフロー情報を記述し、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報は、ドキュメント構造情報201と、閲覧クルー情報202と、レイアウト情報203とを含む。ここで、本実施例において、コンテンツブロック分割結果情報204と、ドキュメントフロー情報(コンテンツブロック分割されたレイアウトファイル205に基づくドキュメント構造情報201、閲覧クルー情報202、レイアウト情報203、コンテンツブロック分割結果情報204の関係を含む)は、レイアウトファイル205とはそれぞれ記憶される。本実施例では、ドキュメントフロー情報は、インデックス構造であり、ドキュメント構造情報201と、閲覧クルー情報202と、レイアウト情報203とコンテンツブロック分割結果情報204との関係を示す。

30

40

【0035】

以下は、更に具体的な実施例を挙げる。

【0036】

図3に示すように、レイアウトファイル301は、そのドキュメントコンテンツが302及び303として記述されている。テキスト対象と画像エンティティ対象とを備える。302はレイアウトファイルのテキスト対象と画像エンティティ対象のコンテンツの定義である。各コンテンツの定義はそれぞれレイアウトファイルにおいて一つの対象ID(ID)を有する。303に示すように、レイアウトファイルにおいて対象ID(ID)に基づいて定義されたエンティティ対象またはテキスト対象を使用することによってレイアウト

50

トファイルの表示に際して 302 に定義されたエンティティ対象とテキスト対象とが表示される。

【0037】

図4と図5は、インテリジェント理解アルゴリズムを用いて図3におけるレイアウトファイル301に対して計算して当該レイアウトファイル301に対応するドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得した後、レイアウトファイル301のドキュメントコンテンツに対してコンテンツブロック分割を行ってコンテンツブロック分割結果情報を記述する実施例を示す図である。ここで、図4は、レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割処理を行う方法を示す。レイアウトファイルを形成する対象によって違うコンテンツブロックに分割されるように、レイアウトファイル301では、IDが1と3であるエンティティ対象を一つのコンテンツブロックに分割し、分割された当該コンテンツブロックの番号を9とし、IDが2であるエンティティ対象を一つのコンテンツブロックに分割し、分割された当該コンテンツブロックの番号を8とする。図5は、XML言語を用いてコンテンツブロック分割結果情報を記述することを示す図である。

10

【0038】

図6と図7は、コンテンツブロック分割されたレイアウトファイルに基づくドキュメントフロー情報を示す図である。図6は、コンテンツブロック分割されたレイアウトファイルに基づくドキュメントフロー情報におけるドキュメントの構造情報を示す。当該ドキュメントの構造情報はドキュメントの章節ツリーと各章節内のコンテンツブロックの順序(具体的には、図面では、コンテンツブロック番号で示される)を指定する。具体的には、図6では、レイアウトファイルにおける一つの段落が示される。当該段落には番号が8と9であるコンテンツブロックが含まれる。図7は、コンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報におけるドキュメントレイアウトの自己適応表示情報を示す図であり、具体的には、番号が9であるコンテンツブロックにおいて対象IDが1であるテキスト対象と、対象IDが3であるエンティティ対象の順序が調整される方法、即ち、対象IDが3であるエンティティ対象を対象IDが1であるテキスト対象

20

「这是一」

の第一字の後に挿入することを指定する。

30

【0039】

図8は、本発明の実施例において、図3に示されたコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報におけるドキュメントレイアウトがコンテンツリアレンジされることを示す図である。図3、4、5、6、7の結果によって当該段落のコンテンツをレイアウトリアレンジして図8の結果になる。レイアウトリアレンジする際に、先ず、図6に基づいて段落の構造を取得し、取得した段落の構造によると、コンテンツ9が前に位置され、コンテンツ8が後に位置されるため、

「这是一<Image.JPG>篇文章」

のようなシーケンスが形成される。また、図7による順序情報に基づいてシーケンスを

40

「这<Image.JPG>是一篇文章」

として調整する。このように、フロー情報を利用して正確なコンテンツを取得できる。次に、レイアウトサイズ(3文字の幅)がレイアウトリアレンジによって下記の図に示す結果になる。すると、この実施例は、レイアウトファイル及びこの前の処理によってフロー情報を取得してコンテンツ抽出とレイアウトリアレンジを行うプロセスを完成した。この実施例によると、レイアウトファイルのコンテンツに対する検索と、構造化記憶と、修正と、抽出と、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

【0040】

閲覧クルー情報は、特殊なドキュメントコンテンツの構造情報であり、現存のドキュメ

50

ントコンテンツの構造情報から直接に取得することができ、また、ユーザによって自由に指定することができる。閲覧クルー情報の処理方法は、ドキュメントコンテンツの構造情報の処理方法と同じであるため、ここでは、閲覧クルー情報の例を省略する。

【0041】

ステップ105の構造処理としては、レイアウトファイルのコンテンツに対する検索操作、構造化記憶操作、修正操作、抽出操作、レイアウトリアレンジ操作の少なくとも一つを備えてもよい。すなわち、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との関連に基づいて、レイアウトファイルのコンテンツブロックと、ドキュメントコンテンツの構造及び/又はドキュメントレイアウトに対して行われる操作である。

10

【0042】

例えば、検索、構造化記憶、修正、抽出の操作は、下記のように実現できる。

【0043】

まず、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントコンテンツの構造情報との関連に基づいて相応なレイアウトドキュメントのフロー構造及び順序正確なコンテンツフローとを生成する。次に、フロー構造またはコンテンツフローにおいて、順次アクセスやマルチセクション検索等の方法でコンテンツを検索することによって検索、構造化記憶、修正、抽出などの目的を実現する。

【0044】

例えば、レイアウトリアレンジは下記のように実現できる。

20

【0045】

まず、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントコンテンツの構造情報との関連に基づいてコンテンツフローにおけるコンテンツのために相応なレイアウト情報を提供し、レイアウトアルゴリズムを用いてレイアウトリアレンジを実現できる。例えば、レイアウトドキュメントを編集する時、正確なドキュメントフロー情報が得られたため、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントレイアウトの表示情報との関連に基づいてレイアウトファイルのドキュメント構造と、元のコンテンツの順序及び編集される位置とを取得でき、編集されるデータをドキュメントの構造情報またはドキュメントコンテンツフローの正確な位置に挿入でき、よって、編集を簡単に迅速にできると共に、編集後のドキュメントフロー情報を再建できる。

30

【0046】

また、本発明の実施例によって相応してレイアウトファイルの構造処理装置が提供される。図8に示すように、オリジナル情報取得モジュール802と、コンテンツブロック分割モジュール803と、ドキュメントフロー情報記述モジュール804と構造処理モジュール805とを備える。

【0047】

オリジナル情報取得モジュール802は、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得する。ここで、レイアウトファイルとは、レイアウトファイルの全体でもよいが、レイアウトファイルの一頁または何頁などでもよい。また、レイアウトファイルのオリジナル情報とは、レイアウトファイルにおけるドキュメントコンテンツの構造情報及び/又はレイアウトファイルにおけるドキュメントレイアウトの自己適応表示情報である。詳しくは、下記(1)~(3)を含むが、これに限らない。すなわち、

40

(1)ドキュメントコンテンツの構造情報：ドキュメントの章節情報と、各章節内のコンテンツブロックの順序と、コンテンツブロック内の各エンティティの順序等。

(2)閲覧クルー情報：具体的には、上記したドキュメントコンテンツの構造情報により提供される閲覧順序以外、必要によって提供される他の閲覧順序情報が含まれる。ユーザに提供される選択可能な閲覧順序情報である。閲覧クルー情報はレイアウトファイルのドキュメントコンテンツ全体の閲覧順序情報でもよいが、レイアウトファイルの一部分のド

50

キュメントコンテンツの閲覧順序情報でもよい。

(3) レイアウト情報：具体的には、レイアウトファイルのレイアウトリアレンジを行う時にエンティティの最終的な表示効果を決める情報であり、エンティティ自身又はコンテンツブロック自身のレイアウトプロパティと、同じコンテンツブロックにおけるエンティティ同士又はコンテンツブロック同士のレイアウト関係とが含まれる。例えば、画像を指定する文字引き立た方式や、複数のコンテンツブロックを指定するコラム情報など。上記したレイアウトリアレンジとは、レイアウトサイズ又はレイアウトコンテンツが変化する場合、一定の規則に従ってレイアウトにおける各エンティティを再編してレイアウト表示効果とする過程である。

【0048】

コンテンツブロック分割モジュール803は、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行う。

【0049】

ドキュメントフロー情報記述モジュール804は、コンテンツブロック分割の結果に基づいてレイアウトファイルのドキュメントフロー情報を生成する。

【0050】

構造処理モジュール805は、前記ドキュメントフロー情報に基づいて前記レイアウトファイルに対して構造処理を行う。

【0051】

レイアウトファイルのドキュメントフロー情報を取得し、取得されたドキュメントフロー情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行い、コンテンツブロック分割結果情報を記述し、コンテンツブロック分割結果情報に基づいて、コンテンツブロック分割された当該レイアウトファイルの基づくドキュメントフロー情報を記述し、よって、レイアウトファイルに対する構造処理が容易に行われることができる。例えば、レイアウトファイルはドキュメントコンテンツが修正された後、ドキュメント構造、レイアウト表示などの情報の更新が簡単にできる。また、レイアウトファイルのコンテンツに対する検索と、構造化記憶と、修正と、抽出と、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

【0052】

以下は、図9に基づいて本発明に係るレイアウトファイルの構造処理装置の動作について詳しく説明する。

【0053】

オリジナル情報取得モジュール802は、下記(1)、(2)、(3)の少なくとも一つによって、レイアウトファイルのドキュメントフロー情報を取得する。

【0054】

(1) ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報が含まれている電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、レイアウトファイルのさまざまなドキュメントコンテンツのソースを解析することによってレイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を直接に取得することができる。例えば、HTMLや、Microsoft Wordなど、レイアウトファイルに対応する、一部のドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を含んでいる電子ドキュメントの場合は、当該ドキュメントのドキュメント処理システムを利用してそのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を抽出することができる。特に、Microsoft Wordドキュメントに対して、Officeの自動化オブジェクトを利用してそのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。

【0055】

(2) ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情

10

20

30

40

50

報が含まれていない電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、さまざまな識別アルゴリズム又はインテリジェント理解アルゴリズムを利用してレイアウトファイルに対して計算することによってレイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。例えば、ドキュメント分析及びドキュメント理解に基づく処理システムを介してレイアウトファイルに対して計算することによって当該レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得することができる。

【0056】

(3) ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報が含まれていない電子ドキュメントは、レイアウトファイルのドキュメントコンテンツのソースである場合、外部からユーザによって入力された、レイアウトファイルに関するドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を受け付けることによって、レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報を取得してもいい。例えば、ユーザはグラフィックインターフェイス付きコンピュータアプリケーションを介してレイアウトファイルのドキュメントコンテンツに対してマークすることによって当該レイアウトファイルのドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報の入力を実現できる。

10

【0057】

コンテンツブロック分割モジュール803は、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行う。すなわち、レイアウトファイルの各グループのコマンド文、又は、各グループの対象、又は、各段落のコンテンツの記述を一つのコンテンツブロックユニットとして当該レイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行う。具体的には、ドキュメントフロー情報によって、文の番号、文の長さ、文のオフセット量、または、対象ID、対象のオフセット量、又は、コンテンツID、コンテンツのオフセット量、または、ある特殊な記号等を使用して、当該レイアウトファイルのドキュメントコンテンツを異なるコンテンツブロックに分割してもいい。ここで、分割されたコンテンツブロック同士は、重なり合うようにされて

20

30

【0058】

例えば、レイアウトファイルを形成する複数のコマンド文を、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコマンド文の数が決定されるように、複数のグループに分割し、分割された各グループのコマンド文を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得する。

【0059】

また、レイアウトファイルを形成する複数の対象を、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎の対象の数が決定されるように、複数のグループに分割し、分割された各グループの対象を一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得する。

40

【0060】

また、レイアウトファイルを形成する複数のコンテンツを、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によってグループ毎のコンテンツの数が決定されるように、複数のグループに分割し、分割された各グループのコンテンツを一つのコンテンツブロックユニットとし、得られたコンテンツブロックの分割結果を記述してコンテンツブロック分割結果情報を取得する。

【0061】

50

図10に示すように、レイアウトファイルを複数のコンテンツブロックに分割する場合は、コンテンツ参照シーケンス取得サブモジュール901を用いて、レイアウトファイルを形成する、コンテンツの参照シーケンスを取得することができる。コンテンツブロック分割サブモジュール902を用いて、コンテンツの参照シーケンスを、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報によって各サブ参照シーケンスにおけるシーケンス数が決定されるように、複数のコンテンツのサブ参照シーケンスに分割し、分割された各コンテンツのサブ参照シーケンスを一つのコンテンツブロックとされる。これらのコンテンツのサブ参照シーケンスは、コンテンツが重なり合うようにされてもよい。分割されたコンテンツのサブ参照シーケンス毎に、唯一な番号が与えられてもよい。また、記述サブモジュール903を用い、取得されたコンテンツブロック分割の結果を記述してコンテンツブロック分割結果情報を取得する。これらのコンテンツのサブ参照シーケンスは、コンテンツが重なり合うようにされてもよい。分割されたコンテンツのサブ参照シーケンス毎に、唯一な番号が与えられてもよい。ここで、コンテンツの参照シーケンスは、コンテンツの参照シーケンスにおけるエンティティのオフセット位置によって分割することができる。また、コンテンツの参照シーケンスは、コンテンツの参照シーケンスにおける一つ又は幾つかの特殊エンティティ記号の位置によって分割することができる。また、コンテンツの参照シーケンスは、コンテンツの参照シーケンスにおける一つ又は幾つかのIDの位置によって分割することができる。

10

20

30

40

50

【0062】

上記したコンテンツブロック分割の結果に基づいて当該レイアウトファイルのコンテンツブロック分割結果情報を記述する。構造化マークアップ言語（例えば、XML言語や、SGML言語等）を利用して記述してもよいが、これに限定されない。

【0063】

ドキュメントフロー情報記述モジュール804は、前記コンテンツブロック分割結果情報に基づいてレイアウトファイルのドキュメントフロー情報を生成する。コンテンツブロック分割された当該レイアウトファイルに基づくドキュメントフロー情報を記述する。具体的には、各コンテンツブロック自身のドキュメントフロー情報と、各コンテンツブロック間の関係、例えば、ドキュメントの構造情報や、閲覧クルー情報や、レイアウト情報等を記述する。ここで、XML言語や、SGML言語等でコンテンツブロック分割されたレイアウトファイルのドキュメントフロー情報を記述してもよいが、これに限定されない。

【0064】

具体的には、上記コンテンツブロック分割結果情報と、ドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報とを関連付けて、関連付けられたコンテンツブロック分割結果情報とドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報とを対応して記憶してもよい。ここで、コンテンツブロック分割結果情報とドキュメントフロー情報は、レイアウトファイルとはそれぞれ記憶されてもよいが、レイアウトファイルに組み込まれてレイアウトファイルのデータブロックとしてもよい。

【0065】

構造化マークアップ言語を利用して、得られたコンテンツブロック分割結果情報とドキュメントフロー情報を記述することができる。

【0066】

実際の応用において、記憶された上記コンテンツブロック分割結果情報とドキュメントフロー情報とを転送やコピー等を介して他の記憶デバイスに移すことができる。そして、他のユーザ端末はコンテンツブロック分割されたレイアウトファイルに基づくドキュメントフロー情報を直接に便利に利用することができる。

【0067】

また、本発明の実施例におけるレイアウトファイルの構造処理措置とやり取りを行う外部システムは、フォーマット変換システムと、レイアウトリアレンジシステムなどを備える場合がある。このようなシステムでは、コンテンツブロック分割されたレイアウトファ

イルに基づくドキュメントフロー情報を利用してレイアウトファイルに対して処理を行い、例えば、情報抽出、レイアウトリアレンジ、他のフォーマットのファイルへの変換などを行う。

【0068】

なお、ドキュメントフロー情報に基づくレイアウトファイルの構造処理としては、レイアウトファイルのコンテンツに対する検索操作、構造化記憶操作、修正操作、抽出操作、レイアウトリアレンジ操作の少なくとも一つを備えてもよい。すなわち、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントコンテンツの構造情報及び/又はドキュメントレイアウトの表示情報との関連に基づいて、レイアウトファイルのコンテンツブロックと、ドキュメントコンテンツの構造及び/又はドキュメントレイアウトに対して行われる操作である。

10

【0069】

構造処理モジュール805に行われる検索、構造化記憶、修正は、例えば、下記のようにして実現できる。

【0070】

まず、ドキュメントフロー情報におけるコンテンツブロック分割結果情報とドキュメントコンテンツの構造情報との関連に基づいて相応なレイアウトドキュメントのフロー構造及び順序正確なコンテンツフローとを生成する。次に、フロー構造またはコンテンツフローにおいて、順次アクセスやマルチセクション検索等の方法でコンテンツを検索することによって検索、構造化記憶、修正、抽出などの目的を実現する。

20

【0071】

構造処理モジュール805に行われるレイアウトリアレンジは、例えば、下記のようにして実現できる。

【0072】

まず、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントコンテンツの構造情報との関連に基づいてコンテンツフローにおけるコンテンツのために相応なレイアウト情報を提供し、レイアウトアルゴリズムを用いてレイアウトリアレンジを実現できる。例えば、レイアウトドキュメントを編集する時、正確なドキュメントフロー情報が得られたため、ドキュメントフロー情報に記述された、コンテンツブロック分割結果情報とドキュメントレイアウトの表示情報との関連に基づいてレイアウトファイルのドキュメント構造と、元のコンテンツの順序及び編集される位置とを取得でき、編集されるデータをドキュメントの構造情報またはドキュメントコンテンツフローの正確な位置に挿入でき、よって、編集を簡単に迅速にできると共に、編集後のドキュメントフロー情報を再建できる。

30

【0073】

以上のように、本発明に係るレイアウトファイルの構造処理方法及び装置によれば、レイアウトファイルのドキュメントフロー情報を取得し、取得したドキュメントフロー情報に基づいてレイアウトファイルのドキュメントコンテンツに対してコンテンツブロック分割を行い、コンテンツブロック分割結果情報を記述し、コンテンツブロック分割結果情報に基づいて、コンテンツブロック分割された当該レイアウトファイルに基づくドキュメントフロー情報を記述する。それによって、レイアウトファイルは、ドキュメントコンテンツが修正された後に、レイアウトを再計算する必要も、ドキュメント全体のレイアウト情報をリライトする必要もなく、容易にドキュメント構造処理を行うことができる。例えば、レイアウトファイルのコンテンツに対する検索、構造化記憶、修正、抽出、レイアウトリアレンジなどの編集操作が柔軟化、簡素化される。

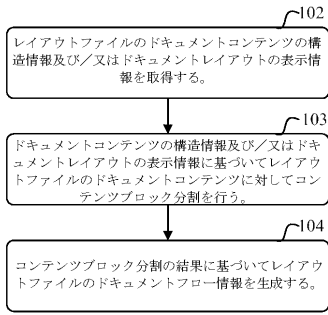
40

【0074】

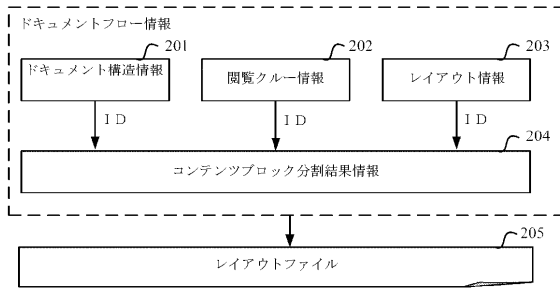
発明の詳細な説明の項においてなされた具体的な実施形態または実施例は、あくまでも、本発明の技術内容を明らかにするものであって、そのような具体例にのみ限定して狭義に解釈されるべきものではなく、本発明の精神と次に記載する特許請求事項の範囲内で、いろいろと変更して実施することができるものである。

50

【 図 1 】



【 図 2 】



【 図 3 】

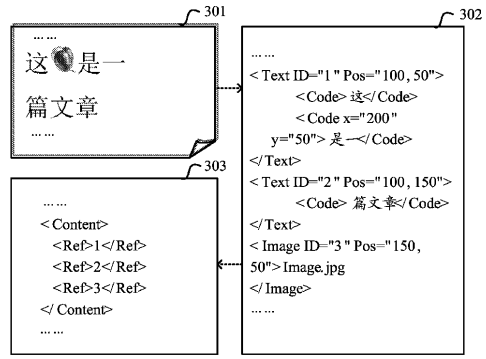


图 3

【 図 4 】

コンテンツブロック番号	含まれる対象 ID
8	2
9	1、3

【 図 5 】

```

    ...
    <block>
      <def BID="8">
        ...
      </def>
      <def BID="9">
        ...
      </def>
      ...
    </block>
    ...
  
```

图 5

【 図 6 】

```

    ...
    <Paragraph>
      <Content Ref="9" />
      <Content Ref="8" />
    </Paragraph>
    ...
  
```

图 6

【 図 7 】

ブロック番号	レイアウト情報	記述
8	順序設置	3を1の第一のCodeの後に挿入する。

【 図 8 】

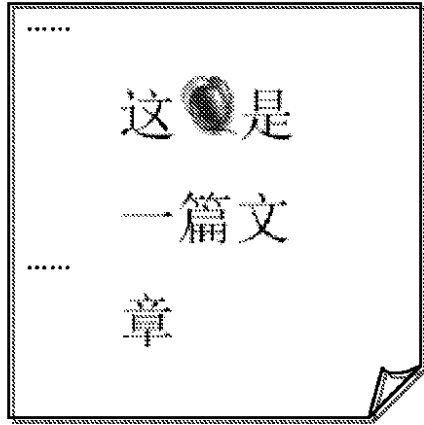
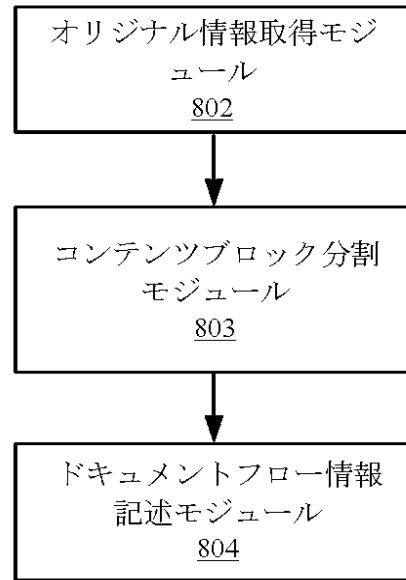
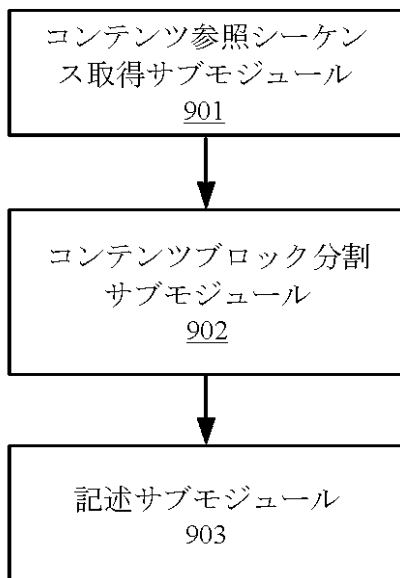


图 8

【 図 9 】



【 図 10 】



【 国际调查报告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/CN2009/072147
A. CLASSIFICATION OF SUBJECT MATTER		
H04W 4/00(2009.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC: H04W 4/-; H04L12/-; H04M3/-		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
WPI, EPODOC, PAJ, CNPAT, CNKI: format, document, structure, information, layout, file, content, divid???, creat+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN101308488A, (BEIDAFANGZHENG ELECTRONICS CO.,LTD.et al.), 19 Nov. 2008(19.11.2008), the whole document	1-14
A	CN1794218A, (WUXI YONGZHONG SCI & TECHNOLOGY CO.,LTD), 28 Jun. 2006(28.06.2006), the whole document	1-14
A	CN1848108A, (SHARP KK), 18 Oct. 2006(18.10.2006), the whole document	1-14
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&”document member of the same patent family</p>		
Date of the actual completion of the international search 20 Aug.2009 (20.08.2009)		Date of mailing of the international search report 17 Sep. 2009 (17.09.2009)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451		Authorized officer ZHAO, Qiang Telephone No. (86-10)62411672

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2009/072147

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101308488A	19.11.2008	NONE	
CN1794218A	28.06.2006	WO2007076717A1	12.07.2007
		CN100356372C	19.12.2007
		EP1975812A1	01.10.2008
		US2008320379A	25.12.2008
		JP2009522626T	11.06.2009
CN1848108A	18.10.2006	US2006221064A1	05.10.2006
		EP1710714A2	11.10.2006
		JP2006285896A	19.10.2006

国际检索报告		国际申请号 PCT/CN2009/072147
A. 主题的分类		
H04W 4/00(2009.01)i		
按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类		
B. 检索领域		
检索的最低限度文献(标明分类系统和分类号)		
IPC: H04W 4/-; H04L12/-; H04M3/-		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))		
WPI, EPODOC, PAJ, CNPAT, CNKI; 版式, 文件, 结构, 信息, 呈现, 文档, 内容, 划分, 创建, format, document, structure, information, layout, file, content, divid???, creat+		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN101308488A, (北大方正集团有限公司等), 19.11月 2008 (19.11.2008), 全文	1-14
A	CN1794218A, (无锡永中科技有限公司), 28.6月 2006 (28.06.2006), 全文	1-14
A	CN1848108A, (夏普株式会社), 18.10月 2006 (18.10.2006), 全文	1-14
<input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。		
* 引用文件的具体类型:		“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件
“A” 认为不特别相关的表示了现有技术一般状态的文件		“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性
“E” 在国际申请日的当天或之后公布的在先申请或专利		“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性
“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)		“&” 同族专利的文件
“O” 涉及口头公开、使用、展览或其他方式公开的文件		
“P” 公布日先于国际申请日但迟于所要求的优先权日的文件		
国际检索实际完成的日期 20.8月 2009 (20.08.2009)	国际检索报告邮寄日期 17.9月 2009 (17.09.2009)	
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号: (86-10)62019451	授权官员 赵强 电话号码: (86-10) 62411672	

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2009/072147

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101308488A	19.11.2008	无	
CN1794218A	28.06.2006	WO2007076717A1	12.07.2007
		CN100356372C	19.12.2007
		EP1975812A1	01.10.2008
		US2008320379A	25.12.2008
		JP2009522626T	11.06.2009
CN1848108A	18.10.2006	US2006221064A1	05.10.2006
		EP1710714A2	11.10.2006
		JP2006285896A	19.10.2006

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(71)出願人 510320737

北京方正阿 帕 比 技 術 有 限 公 司

BEIJING FOUNDER APABI TECHNOLOGY LIMITED

中華人民共和國北京市100080海淀区中 并 村北大街127-1号 北大科技 園 創新中心3屋

(71)出願人 507232478

北京大學

PEKING UNIVERSITY

中華人民共和國北京市 海 淀 區 頤 和 園 路 5 号

No.5, Yiheyuan Road, Haidian District, Beijing 100871, China

(74)代理人 100147485

弁 理 士 杉 村 憲 司

(74)代理人 100143568

弁 理 士 英 貢

(74)代理人 100134577

弁 理 士 石 川 雅 章

(72)發明者 仇 叡 恒

中華人民共和國北京市100080海淀区中 并 村北大街127-1号 北大科技 園 創新中心3屋

(72)發明者 王 毅

中華人民共和國北京市100080海淀区中 并 村北大街127-1号 北大科技 園 創新中心3屋

(72)發明者 沸 帆

中華人民共和國北京市100080海淀区中 并 村北大街127-1号 北大科技 園 創新中心3屋

Fターム(参考) 5B082 GA02

5B109 NC01