

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2002 年 12 月 27 日 (27.12.2002)

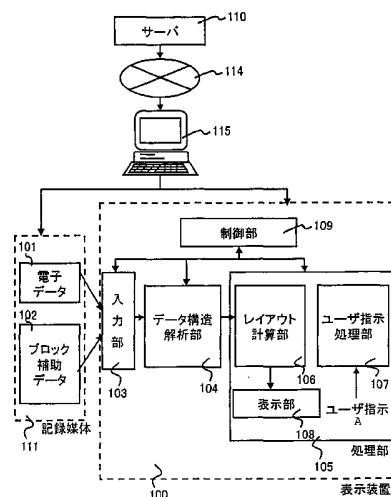
PCT

(10) 国際公開番号
WO 02/103554 A1

- (51) 国際特許分類: G06F 17/21 (SAWADA,Yuji) [JP/JP]; 〒536-0021 大阪府 大阪市城東区 諏訪 4-2-2 1-5 0 1 Osaka (JP).
- (21) 国際出願番号: PCT/JP02/05880
- (22) 国際出願日: 2002 年 6 月 12 日 (12.06.2002) (74) 代理人: 深見 久郎, 外(FUKAMI,Hisao et al.); 〒530-0054 大阪府 大阪市北区南森町 2 丁目 1 番 2 9 号 三井住友銀行南森町ビル 深見特許事務所 Osaka (JP).
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語 (81) 指定国 (国内): JP, KR, US.
- (30) 優先権データ: 特願2001-179415 2001 年 6 月 14 日 (14.06.2001) JP (84) 指定国 (広域): ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- (71) 出願人 (米国を除く全ての指定国について): シャープ株式会社 (SHARP KABUSHIKI KAISHA) [JP/JP]; 〒545-8522 大阪府 大阪市 阿倍野区長池町 2 2 番 2 2 号 Osaka (JP). 添付公開書類:
— 国際調査報告書
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 沢田 裕司
- 2 文字コード及び他の略語については、定期発行される各 PCT ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: DATA PROCESSING METHOD, DATA PROCESSING PROGRAM, AND DATA PROCESSING APPARATUS

(54) 発明の名称: データ処理方法、データ処理プログラム、およびデータ処理装置



110...SERVER
101...ELECTRONIC DATA
102...BLOCK AUXILIARY DATA
103...INPUT BLOCK
109...CONTROL BLOCK
104...DATA STRUCTURE ANALYSIS BLOCK
106...LAYOUT CALCULATION BLOCK
107...USER INSTRUCTION PROCESSING BLOCK
108...DISPLAY BLOCK
A...USER INSTRUCTION
105...PROCESSING BLOCK
100...DISPLAY APPARATUS
111...RECORDING MEDIUM

(57) Abstract: When a user instruction processing block (107) receives specification of electronic data (101) to be displayed, a data processing apparatus which is a display apparatus reads in block auxiliary data (102) prepared for the electronic data (101) from an input block (103). Among the block auxiliary data (102)

[続葉有]



which has been read in, only necessary blocks are read in according to the start/end position of each block and the position from the file head of the area to be displayed on the screen. Hierarchical structure of the electronic data (101) is analyzed and a layout within the display screen is calculated by a layout calculation block (106) and is displayed on a display block (108).

(57) 要約:

表示装置であるデータ処理装置は、ユーザ指示処理部（１０７）において表示したい電子データ（１０１）の指定を受付けると、その電子データ（１０１）に対して用意されたブロック補助データ（１０２）を入力部（１０３）より読込む。読込んだブロック補助データ（１０２）のうち、各ブロックの開始／終了位置と、画面に表示しようとしている領域のファイル先頭からの位置とに基づき、必要なブロックのみを読込む。そして、電子データ（１０１）の階層構造を解析してレイアウト計算部（１０６）で表示画面内でのレイアウトを計算し、表示部（１０８）に表示する。

明細書

データ処理方法、データ処理プログラム、およびデータ処理装置

技術分野

- 5 本発明は、データ処理方法、データ処理プログラム、データ処理装置、構造化データ、構造化データを記録したコンピュータ読取可能な記録媒体、および送信装置に関し、特に、階層構造を有する電子データを処理することのできるデータ処理方法、データ処理プログラム、データ処理装置、構造化データ、構造化データを記録したコンピュータ読取可能な記録媒体、および送信装置に関する。

10

背景技術

- 文書記述言語の定義に従って所定の処理を実行させるための構造化された電子データであって、階層構造を有する構造化文書を記録するためのデータ形式として、SGML (Standard Generalized Markup Language : 準一般化マーク付け言語)
- 15 や XML (Extensible Markup Language) などが知られている。特に XML は、インターネットで交換する電子文書や電子データなどで近年盛んに使用されている。

 XML 形式の電子データを扱う処理方法として、DOM (Document Object Model) 形式と SAX (Simple API for XML) 形式とが存在する。

- DOM 形式とは、対象とする電子データを全て読み込み、電子データ内の全要素の階層構造を把握してから、電子データの各要素にアクセスする処理方法である。
- 20 例えば、図 16 のような階層構造をもつ電子データがある場合、いったん電子データ全体を読み込み、全要素の階層構造を解析する。図 16 は、電子データの階層構造のみを示し、本文、内容は省略している。そして、図 16 の階層構造をもつ電子データから、図 17 のような木構造 (ツリー構造) を作成した後、各要素
- 25 (TITLE、AUTHOR など) にアクセスする。よって、DOM 形式の電子データの処理方法は、電子データの階層構造を把握した後に処理することになるため、任意の要素にアクセスしやすいという特徴がある。

 これに対し、SAX 形式は、電子データを先頭から順に読み込みながら、読み込んだ要素のみ構造解析し、処理していく形式の処理方法である。そのため、電子デー

タ全体の解析処理を待たずに、順次処理していくことができ、処理速度としても、メモリの容量としても、オーバーヘッドが少ないというメリットがある。

以上のように、DOM 形式および SAX 形式ともにそれぞれのメリットがあるが、反面、次のようなデメリットがそれぞれにある。

- 5 すなわち、DOM 形式の場合、電子データ内の一部のデータだけ进行处理したい場合であっても、木構造を生成するために、電子データ全体の構造解析をしなくてはならず、処理に無駄が生じる。また、電子データのサイズが大きくなると、木構造を作成するための処理時間、および木構造を記憶しておくためのメモリ使用量が大きくなるという欠点があった。
- 10 一方 SAX 形式の場合、電子データの先頭から順に処理することを前提にしたアクセス形式であるため、電子データの内容を先頭から順に処理せず、任意の要素を任意の順番に扱う場合には、無駄な読込処理、および構造解析処理が入るとい
- 15 う欠点が存在する。また、電子データの後半の要素进行处理した時であっても、電子データの先頭から読込み、構造解析をしないといけないため、処理時間として無駄が生じていた。

- 一方、DOM 形式や SAX 形式ではなく、取出したい要素のみを、階層構造を解析することなく取出す処理方法もある。しかしながら、取出したい領域が、どの要素の下に含まれるかによって、その要素の意味が変わったり、表示する位置やサイズ、属性などが変わってしまう場合があるため、この方法では取出した要素に
- 20 対して適切な処理ができない。例えば、図 18 のような、階層構造をもつ XHTML (Extensible HyperText Markup Language) 文書の例を考える。図 18 において、領域 71 の先頭部分は、`<html>`、`<body bgcolor="yellow">`、`<p>` に挟まれている。このうち、`<body bgcolor="yellow">` は、このタグで挟まれた領域が本文データであることを示すと共に、本文の背景を黄色にすることを意味しており、このタグがあるかないかにより、本文の背景が黄色になるか、白色（省略時のデフォルト値）になるかわかってしまう。したがって、領域 71 だけをファイルから読込んで表示処理を行なおうとしても、本文の背景が黄色であることを知ることができず、正しい表示処理ができない。
- 25 このように、一般に、階層構造をもつ電子データに対しては、処理したい領域

のみを読込んで処理しようとしても、処理する領域以外に書かれた制御情報（タグや属性等）を知ることができないため、正しく処理できないという問題があった。

そこで、本発明は、このような問題を解決することのできるデータ処理方法、
5 データ処理プログラム、データ処理装置、構造化データ、構造化データを記録したコンピュータ読取可能な記録媒体、および送信装置を提供することを目的としている。

発明の開示

10 本発明においては、上述の課題を解決するために、以下に示されるデータ処理方法、データ処理プログラム、データ処理装置、構造化データ、構造化データを記録したコンピュータ読取可能な記録媒体、および送信装置が提供される。

（１）階層構造を有する電子データと、電子データを複数のブロックに分割して処理するためのブロック補助データとからなる構造化データを処理するデータ
15 処理方法であって、電子データから所望の電子データが含まれるブロックを読込むステップと、読込まれたブロックのブロック補助データを用いて所望の電子データの階層構造を解析するステップと、解析された結果を用いて所定の処理をするステップとを備える、データ処理方法。

（２）（１）に記載されたデータ処理方法であって、電子データは表示用の文書
20 データであり、所定の処理は、読込んだブロックの表示レイアウトを計算し表示する処理であることを特徴とする、データ処理方法。

（３）（１）に記載されたデータ処理方法であって、電子データに対するブロック補助データが存在しない場合、電子データの処理をする前に、電子データに対するブロック補助データを作成するステップをさらに備えることを特徴とする、
25 データ処理方法。

（４）（１）に記載されたデータ処理方法であって、複数の分割された電子データのブロックは、ほぼ等しい大きさであることを特徴とする、データ処理方法。

（５）（１）に記載されたデータ処理方法であって、ブロック補助データは、複数の分割された電子データの各ブロックの位置情報と、ブロックの開始位置と終

了位置とにおける階層情報とを含むことを特徴とする、データ処理方法。

(6) (1)に記載されたデータ処理方法であって、電子データに、電子データ内の任意の位置にジャンプするリンク機能が含まれる場合、ブロック補助データは、リンク先の位置情報を含むことを特徴とする、データ処理方法。

5 (7)階層構造を有する電子データを複数のブロックに分解し、分割された各ブロックごとに所定の処理を実行させるための構造化データを生成するためのデータ処理方法であって、電子データを所定のブロックサイズに分割する際の、各ブロック境界における境界候補を抽出する抽出ステップと、抽出された境界候補に基づいて、ブロックの分割位置を決定する決定ステップと、決定された各ブロッ
10 クの先頭位置と終了位置とにおける階層構造の特徴を示す情報を取得する取得ステップと、決定ステップで決定された各ブロックの分割位置の位置情報と、取得された位置情報に対応する階層構造の特徴と示す情報とを含むブロック補助データを生成し、ブロック補助データを電子データに追加して構造化データを生成する生成ステップとを備える、データ処理方法。

15 (8) (7)に記載されたデータ処理方法であって、抽出ステップは、表示したときに行頭から表示される箇所から境界候補を抽出することを特徴とする、データ処理方法。

(9) (7)に記載されたデータ処理方法であって、抽出ステップは、電子データの構造を記録している制御コードの前後とその付近とから境界候補を抽出することを特徴とする、データ処理方法。
20

(10) (7)に記載されたデータ処理方法であって、ブロックサイズは、電子データを処理する装置の処理能力と、装置の画面に表示される文字数と、表示される文字数を決定する要因との少なくとも1つに応じて決められることを特徴とする、データ処理方法。

25 (11)サーバより階層構造を有する電子データを受取り、電子データに対して所定の処理をするデータ処理方法において、サーバに対して所定の処理をしたい電子データ名を送信する送信ステップと、サーバより電子データを複数のブロックに分割して処理するためのブロック補助データを受信する受信ステップと、電子データのうち、所定の処理をしたい電子データを含むブロックの送信をサーバ

に要求し、ブロックのデータを受信する受信ステップと、受信したブロックと、ブロックのブロック補助データとを用いて、ブロックの階層構造を解析し、その結果を用いて所定の処理をする再生ステップとを備える、データ処理方法。

5 (1 2) (1 1) に記載されたデータ処理方法であって、サーバよりブロック補助データを受信するステップにおいて、ブロック補助データが存在しない場合には、ブロック補助データを作成した後、サーバより受信することを特徴とする、データ処理方法。

10 (1 3) 階層構造を有する電子データと、電子データを複数のブロックに分割して処理するためのブロック補助データとを対として備える構造化データに対して、所定の文書記述言語の定義に従って所定の処理を実行させるためのデータ処理プログラムであって、ブロック補助データに基づいて、電子データのブロックデータを讀込む讀込みステップと、讀込まれたブロックデータと、ブロック補助データとに基づいて、ブロックデータに含まれる階層構造を解析する解析ステップと、解析結果と、文書記述言語の定義とに基づいて、ブロックデータに対して所定の
15 処理を実行させる実行ステップとをコンピュータに実行させる、データ処理プログラム。

(1 4) 階層構造を有する電子データを複数のブロックに分割し、分割された各ブロックごとに所定の処理を実行させるための構造化データを生成するためのデータ処理プログラムであって、電子データを所定のブロックサイズに分割する際
20 の、各ブロック境界における境界候補を抽出する抽出ステップと、境界候補に基づいて、ブロックの分割位置を決定する決定ステップと、決定された各ブロックの先頭位置と終了位置とにおける階層構造の特徴を示す情報を取得する取得ステップと、決定ステップで決定された各ブロックの分割情報の位置情報と、取得された位置情報に対応する階層構造の特徴を示す情報とを含むブロック補助データを生成し、ブロック補助データを電子データに追加して構造化データを生成する
25 生成ステップとをコンピュータに実行させる、データ処理プログラム。

(1 5) サーバより階層構造を有する電子データを受取り、電子データに対して所定の処理をするためのデータ処理プログラムであって、サーバに対して所定の処理をしたい電子データ名を送信する送信ステップと、サーバより電子データを

複数のブロックに分割して処理するためのブロック補助データを受信する受信ステップと、電子データのうち、所定の処理をしたい電子データを含むブロックの送信をサーバに要求し、ブロックのデータを受信する受信ステップと、受信したブロックと、ブロックのブロック補助データとを用いて、ブロックの階層構造を解析し、その結果を用いて所定の処理をする再生ステップとをコンピュータに実行させる、データ処理プログラム。

(16) 階層構造を有する電子データを処理するデータ処理装置であって、電子データには、電子データを複数のブロックに分割して処理するための補助情報であるブロック補助データが付随し、電子データのうち、処理をしたいデータが含まれるブロックと、ブロック補助データとを読み込む入力部と、ブロック補助データを用いて、読み込んだ前記ブロックの階層構造を解析するデータ構造解析部と、データ構造解析部の結果を用いて、所定の処理をする処理部を備える、データ処理装置。

(17) (16)に記載されたデータ処理装置であって、電子データは表示用の文書データであり、処理部は、データ構造解析部によって解析されたブロックの階層構造を用いて、読み込んだブロックの表示レイアウトを計算するレイアウト計算部と、レイアウト計算部によって求めた表示レイアウトに基づいて表示する表示部とをさらに備える、データ処理装置。

(18) サーバより階層構造を有する電子データを受取り、電子データに対して所定の処理をするデータ処理装置であって、電子データには、電子データを複数のブロックに分割して処理するための補助情報であるブロック補助データが付随し、電子データのうち、所定の処理をしたい電子データが含まれるブロックをサーバに伝え、ブロックのデータと、ブロック補助データとをサーバより受信する送受信部と、受信したブロックと、ブロック補助データとを用いて、受信したブロックの階層構造を解析するデータ構造解析部と、データ構造解析部の結果を用いて、所定の処理をする処理部とを備える、データ処理装置。

(19) (16)または(18)に記載されたデータ処理装置であって、携帯型の端末であることを特徴とするデータ処理装置。

(20) 文書記述言語の定義に従って記述された電子データと、電子データを複

数のブロックに分割して処理するためのブロック補助データとを対として備えたことを特徴とする、構造化データ。

5 (21) 文書記述言語の定義に従って所定の処理を実行させるための電子データと、電子データを複数のブロックに分割し、分割された各ブロックごとに所定の処理を実行させるためのブロック補助データとを対として備えたことを特徴とする、構造化データ。

(22) (20) または (21) に記載された構造化データであって、複数に分割された電子データのブロックは、ほぼ等しい大きさであることを特徴とする、構造化データ。

10 (23) (20) または (21) に記載された構造化データであって、ブロック補助データは、少なくとも各ブロックの位置情報と、ブロックの開始位置と終了位置とにおける階層情報とを有することを特徴とする、構造化データ。

(24) (20) または (21) に記載された構造化データであって、ブロック補助データは、ブロックの分割箇所を、電子データの構造を記録している制御コードの前後とその付近とから選ぶことを特徴とする、構造化データ。

15 (25) (20) または (21) に記載された構造化データであって、ブロック補助データは、ブロックの分割箇所を、表示したときに行頭から表示される箇所に限定することを特徴とする、構造化データ。

(26) (20) または (21) に記載された構造化データであって、電子データに、電子データ内の任意の位置にジャンプするリンク機能が含まれる場合、ブロック補助データは、リンク先の位置情報を含むことを特徴とする、構造化データ。

(27) 文書記述言語の定義に従って記述された電子データと、電子データを複数のブロックに分割して処理するためのブロック補助データとを対として備えたことを特徴とする構造化データを記録した、コンピュータ読取可能な記録媒体。

25 (28) (13) ～ (15) のいずれかに記載されたデータ処理プログラムを送信する送信部を備えることを特徴とする、送信装置。

(29) (20) または (21) に記載された階層化データを送信する送信部を備えることを特徴とする、送信装置。

図面の簡単な説明

第1図は、本発明の第1の実施の形態におけるデータ処理装置100のブロック図である。

5 第2図は、本発明の第1の実施の形態における表示装置を、具体例として携帯機器で実現した場合の概観図を示す図である。

第3図は、本実施の形態におけるブロック補助データの具体的な概略を示す図である。

10 第4図は、電子データを複数のブロックに分割した様子を具体的に示す図である。

第5図は、図4に示される電子データに対して設定されるブロック補助データの具体例を示す図である。

第6図は、本発明の第1の実施の形態におけるデータ処理装置100の処理を示すフローチャートである。

15 第7図は、1ブロック分のデータとブロック補助データとから作成したデータの具体例を示す図である。

第8図は、完全には階層構造を有しない電子データの具体例を示す図である。

第9図は、本発明の第2の実施の形態におけるデータ処理装置200のブロック図である。

20 第10図は、本発明の第2の実施の形態におけるデータ処理装置200の処理を示すフローチャートである。

第11図は、第2の実施の形態において、ブロック補助データを作成する際の処理を示すフローチャートである。

25 第12A図～第12C図は、行の途中でブロック分割するデータの具体例と表示の具体例とを示す図である。

第13図は、本発明の第3の実施の形態におけるデータ処理装置のブロック図である。

第14図は、第3の実施の形態において、ブロック補助データを作成する際の処理を示すフローチャートである。

第15図は、本発明の第3の実施の形態におけるデータ処理装置の処理を示すフローチャートである。

第16図は、階層構造をもつ電子データの具体例を示す図である。

5 第17図は、階層構造をもつ電子データから取出した木構造を説明するための図である。

第18図は、階層構造をもつ電子データの具体例を示す図である。

発明を実施するための最良の形態

本発明をより詳細に説明するために、添付の図面を用いてこれを説明する。

10 第1図は、第1の実施の形態におけるデータ処理装置100が表示装置である場合のデータ処理装置100のブロック図の一例を示す図である。

図1を参照して、サーバ110は、ユーザからの要求を受信し、データベースに記録された電子データを送信する。ネットワーク114は、サーバ110とユーザのパーソナルコンピュータ（以下、パソコンと称する）115とを接続する。
15 記録媒体111は、パソコン115から電子データを取出してデータ処理装置100に電子データを供給する。

なお、上述のパソコン115を、コンビニエンスストアや駅構内等に設置された電子データ受信装置（図示せず）に替えて、この電子データ受信装置から電子データを取出し、記録媒体111に記録する方式としてもよい。この場合、電子
20 データ受信装置から電子データを取出して記録媒体111に記録する際に、課金することが可能である。また、サーバ110から送信される電子データを、パソコン115を介することなく、データ処理装置100が受信して記録媒体111に記録する構成であってもよい。

さらに図1を参照して、電子データ101は、記録媒体111の中に記録された電子データであり、ブロック補助データ102は、記録媒体111に記録され、
25 電子データ101に付随するデータである。

上述の電子データ101は、文書記述言語の定義に従って所定の処理を実行させるための構造化された電子データであって、階層構造を有する構造化文書を記録するためのデータ形式である SGML や XML などを用いて記録されている。

ブロック補助データ 102 は、上述の構造化された電子データである電子データ 101 を複数のブロックに分割し、分割された各ブロックごとに所定の処理を実行させるデータであり、電子データ 101 と対になっている。

さらに図 1 を参照して、入力部 103 は、電子データ 101 およびブロック補助データ 102 を読み込む。データ構造解析部 104 は、電子データ 101 とブロック補助データ 102 とからデータの階層構造を解析する。処理部 105 は、データ構造解析部 104 で解析した階層構造に基づき、所定の処理を行なう。制御部 109 は、入力部 103、データ構造解析部 104、および処理部 105 の各処理部を制御する。

処理部 105 は、対象とする電子データの内容、および処理内容によって構成が異なるが、データ処理装置 100 が、例えばインターネットで交換する電子文書や電子データ、単行本、教科書、雑誌、小説、および論文のような文章などの表示を行なう表示装置である場合には、図 1 に示すように、データ構造解析部 104 の解析結果に基づいて、文章の表示レイアウトを計算するレイアウト計算部 106 と、求めた表示レイアウトに基づいて文章を表示する表示部 108 と、スクロールなどのユーザからの指示を処理するユーザ指示処理部 107 とから構成される。

また、電子データが音声である場合は、処理部 105 は、読上装置に変更され、また、表示部 108 には音声再生部が使用され、レイアウト計算部 106 は、読む部分と読まない部分とを判定したり、強調して読む箇所とそうでない箇所とを判定したり、また各読上げの間隔（ま）を取ったりする部分に変更される。この、電子データが音声である場合、電子データの階層構造に応じて、音声の属性を変えて読上げるようにしてもよい。さらに、電子データが音声である場合、データ処理装置 100 には、シナリオ解釈部、音声出力部、および、再生順を管理するために各要素の同期をとる同期部が必要になる。

図 2 に、上述のデータ処理装置 100 を携帯型機器で実現した場合の具体例を示す。

図 2 において、表示部 108 は、レイアウト計算部 106 で計算した表示レイアウトに基づき表示を行ない、ディスプレイなどから構成される。また、記録媒

体 1 1 1 は、図 1 に示したように、パソコン 1 1 5 や電子データ受信装置によつて、サーバ 1 1 0、およびネットワーク 1 1 4 を介して文書データベースから取出された、処理対象の電子データ 1 0 1 と、ブロック補助データ 1 0 2 とが記録された記録媒体である。この記録媒体 1 1 1 を本体のデータ処理装置 1 0 0 に挿入することで、表示装置であるデータ処理装置 2 0 0 内に設けられた入力部 1 0 3 を通してこの 2 つのデータが読込まれる。さらに十字キー 1 1 2 は、ユーザが文書のスクロール指示や表示させたい書籍や文書を選択する場合などに用いられる。また、ペン 1 1 3 は、電子データにリンク機能が含まれるとき、リンク先にジャンプするために使われる。また、表示装置であるデータ処理装置 1 0 0 から求められた確認事項に変更するためにも使用される。

なお、記録媒体 1 1 1 のように、記録媒体に電子データ 1 0 1 やブロック補助データ 1 0 2 を記録するものではなく、データ処理装置 1 0 0 の内部に記録する領域を設けて、そこに記憶するようにしてもよい。また、ネットワーク 1 1 4 上のサーバ 1 1 0、あるいはデータベースにこの 2 つのデータを記録し、ダウンロードしながら処理するようにしてもよい。

次に、ブロック補助データについて説明する。

ブロック補助データは、一般に、図 3 に示すように、電子データファイル名 1、ブロック情報 2、およびリンク先情報 3 の 3 つの領域から構成される。

電子データファイル名 1 は、このブロック補助データが、どの電子データに対応するものかを記録するために用意された領域である。ブロック補助データを、電子データ内、もしくは電子データに連結して記録する場合は、電子データファイル名 1 の領域は、省略してもよい。

また、処理する電子データによっては、リンク先情報 3 の領域は存在しないことがある。

以降、この 3 領域の書式について、文書記述言語の定義に従って記述され、階層構造を有する電子データの具体例として XHTML 文書を用いて説明する。図 1 8 に示される XHTML 文書を、図 4 に示されるように、4 つのブロック（ブロック 1 0、ブロック 1 1、ブロック 1 2、ブロック 1 3）に分割する場合を考える。このとき、この XHTML 文書に対するブロック補助データは、図 5 に示されるように

なる。

図 5 を参照して、ブロック補助データの領域 20 は電子データファイル名 1 の領域であり、図 4 の XHTML 文書のファイル名が記録される。

ブロック補助データの領域 21 ～ 37 は、ブロック情報 2 の領域である。

- 5 領域 21 はブロック数を記録する領域であり、4 つのブロックに分割する場合には 4 が記録される。

- 10 領域 22 ～ 25、26 ～ 29、30 ～ 33、34 ～ 37 は、それぞれブロック 10、11、12、13 に対するブロック情報の領域である。一般に、n 個のブロックに分割する場合、領域 22 ～ 25 のブロック情報の構造が n 回繰返し記録される。

- 15 領域 22、26、30、34 にはそれぞれのブロックの開始位置が、領域 23、27、31、35 にはそれぞれのブロックの終了位置が、ファイル先頭からのバイト数の形で記録される。したがって、ブロック 11 に属するデータを取り出した場合は、領域 26 と領域 27 とのブロック情報の値を調べ、ファイルの先頭から数えて 212 バイト目から 423 バイト目を読み込めばよい。

- 20 領域 24、28、32、36 には、各ブロックの開始位置において、終了されていない開始タグを記録する。領域 24 のブロック情報は、ブロック 10 の開始位置において終了されていない開始タグを記録する領域であるが、ブロック 10 はファイルの先頭から始まっているため、領域 24 に記録する制御コードは存在しない。ブロック 11 の開始位置では<html>が閉じられていないため、領域 28 には<html>を記録する。また、ブロック 12 の開始位置のように、複数のタグ (<html>、<body bgcolor="yellow">、<p>) が閉じられていない場合は、<html><body bgcolor="yellow"><p>のように、出現順に連結して領域 32 にブロック情報を記録する。なお、<body>タグのように属性 (bgcolor="yellow") をもつ場合は、属性を付けたまま記録する。ブロック 13 も同様にして、<html><body bgcolor="yellow">と領域 36 にブロック情報を記録する。

領域 25、29、33、37 には、各ブロックの終了位置において閉じられていないタグの終了タグを記録する。ブロック 10 の終了位置においては、<html>タグが終了されていないため、領域 25 にはブロック情報</html>を記録する。

また、ブロック 1 1 の終了位置においては、`<html>`、`<body bgcolor="yellow">`、`<p>`タグが閉じられていないため、出現順とは逆の順番で、上記の終了タグを
5 `</p></body></html>`と連結してブロック情報を領域 2 9 に記録する。同様に、
ブロック 1 2 に対してはブロック情報`</body></html>`を記録する。ブロック 1 3
の終了位置では全てのタグが閉じられているので、領域 3 7 には何も記録しない。

領域 3 8 ~ 4 1 は、リンク先情報 3 の領域であり、図 1 8 の XHTML 文書のうち、
リンク先として指定されたラベル位置を記録する。

ここで、XHTML 文書のリンクについて簡単に説明する。XHTML 文書では、`<a>`タグ
によって他のファイルやファイルの一部へリンクをはることができる。例えば、
10 図 4 のリンク先情報 3, 4 は、同一ファイルの一部へのリンクがはられた例である。
この例で、リンク先情報 3 の`<a>`タグで囲まれた文字列"BBB"がクリックされると、
`href` 属性で指定されたラベル"SUMMARY"が設定されている場所、つまり、
`<a>`の `name` 属性に"SUMMARY"が設定されているリンク先情報 4 に表示位置をジャンプさせる。

15 図 5 において、領域 3 8 ~ 4 1 には、リンク先のラベルの位置情報、つまり
`<a>`タグの `name` 属性により設定されているラベルの位置情報を記録する。図 4 の
XHTML 文書の場合、`name` 属性をもつ`<a>`タグは 1 つしか存在しないため、領域 3
8 には 1 を記録し、領域 3 9 ~ 4 1 の組合わせを 1 回だけ記録する。

一般的に、`name` 属性をもつ`<a>`タグが n 個ある場合は、領域 3 9 ~ 4 1 の構造
20 のリンク先情報が n 回繰返される。領域 3 9 には`<a>`タグの `name` 属性に記録された
ラベル名を、領域 4 0, 4 1 には`<a>`タグに挟まれた文字列の開始位置および
終了位置をファイル先頭からのバイト数で記録する。

なお、ブロックの大きさは、装置の処理能力から決定される。すなわち、ブロック
の大きさが大きいほど 1 ブロックあたりの処理量が増えるため、処理時間が
25 増大し、使用するメモリ容量やリソース容量も増大する。そのため、ブロックの
大きさは、装置の処理能力に応じて決定されることが望ましい。装置の処理能力
を決定する要因として、装置に搭載されている CPU (Central Processing
Unit) の処理能力、メモリ容量、およびリソース容量等が挙げられる。

一方、ブロックの大きさは、画面に表示される文字数やそれを決定する要因に

よっても決定される。例えば、電子書籍等の文章を表示する装置においては、1
画面分の表示をした後、ユーザによるページ移動の指示があるまで待機するよう
に設計されていることが多い。この場合、画面に表示される文字数に対してブロ
ックの大きさを極端に大きく設定した場合には、本実施の形態における処理方法
5 ではブロックを1つの単位として読込むため、画面に表示しない不要なデータも
読込まれることになり、無駄が生じてしまうことになる。したがって、画面に表
示される文字数に基づき、ブロックの大きさが決定されることが望ましい。なお、
画面に表示される文字数は、表示装置の備える画面の大きさおよび解像度、表示
する文字のフォント数、行間や字間の大きさ、および余白の大きさ等によって変
10 わるため、これらの要因によってブロックの大きさを変えるようにしてもよい。

このブロック補助データは、対象とする電子データの種類によって構成および
書式が若干異なる。一般にXML形式で書かれた電子データの場合、図3のリンク
先情報3を除き、電子データファイル名1、ブロック情報2の2領域は存在し、
書式も同じになる。リンク先情報3は、XHTML文書の機能であるリンク機能のた
15 めの情報であるため、XHTML文書などの電子書籍以外の電子データに対しては存
在しないこともある。しかし、他のブロック内のデータを参照しながら処理を進
めるような電子データや、ユーザの動作に応じて他ブロック内のデータを表示さ
せたり、もしくは再生させたりする機能をもつような電子データの場合は、アク
セスしやすくするために、そのデータがどこにあるかをこのリンク先情報3に記
20 録しておくといよい。

また、適用可能な電子データはXML形式に限らない。一般に階層構造をもつ構
造化文書に対して適用可能であり、具体的には、ブロック情報2の開始/終了位
置制御コードに、ブロックの開始/終了位置での階層構造がわかるように、それ
ぞれの書式で記録しておけばよい。

25 次に、この電子データとブロック補助データとによるデータ処理装置100の
動作について、表示装置を例に挙げて説明していく。図6に表示装置でのフロー
チャートを示す。

図6を参照して、まず、ユーザ指示処理部107を通じ、ユーザから表示した
い電子データが指定され（ステップS101）、その電子データに対して用意さ

れたブロック補助データを入力部103より読み込む（ステップS102）。

次に、ステップS102で読み込んだブロック補助データのうち、各ブロックの開始／終了位置と、画面に表示しようとしている領域のファイル先頭からの位置とを元に、どのブロックを読み込むかを判断し（ステップS103）、必要なブロックのみを入力部103より読み込む（ステップS104）。

そして、ステップS102で読み込んだブロック補助データの内容を参照し、読み込んだブロックの開始／終了位置における制御コードを調べる。そして、開始位置制御コード、ブロックデータ、終了位置制御コードの順に連結し、階層構造を解析し、木構造を作成する（ステップS105）。

例えば、図4のブロック12を読み込んだ場合には、図5のブロック補助データの領域32と領域33とにあるブロック情報を前後に連結した、図7のようなデータを作成し、階層構造を解析する。図7において、領域51のデータが領域32のブロック情報に記録された制御コード、領域52のデータがステップS104で読み込んだブロック12のデータ、および領域53が領域33に記録された制御コードである。なお、図7のように、必要ならば、先頭に領域50のデータのような、XML宣言と文書型宣言とを付けてもよい。このステップS105の処理はデータ構造解析部104で行なわれる。

次にレイアウト計算部160により、ステップS105で作成した1ブロック分の木構造を元に、表示画面内でのレイアウトを計算する（ステップS106）。この結果、表示部108がもつ表示画面全体のレイアウトが決まれば（ステップS107）、ステップS108に移り、表示部108に表示する。

一方、ステップS107において、まだ表示画面のうちレイアウトが決まっていない領域があればステップS104に戻り、次のブロックを読み込み、木構造を作成して（ステップS105）、1つ前のブロックに対して求めたレイアウト画面の続きから、現在のブロックのレイアウト処理を行なう（ステップS106）。続くステップS107、S108の処理は、先に説明した処理と同様の処理を行なう。

ステップS108で表示部108に表示した後は、ステップS109に移り、ユーザからの指示を待つ。ユーザからの指示があるまでステップS109で待機

する。

ここで、ユーザから表示処理の終了が指示された場合、表示処理を終了する。
ステップS109において、ユーザから次ページ、あるいは前ページのスクロール指示があった場合、ステップS111に移り、ユーザから受けたスクロール指示の結果、次に表示する内容が現在のブロックと同じかどうかを判断する。もし
5 同じならば、ステップS106に移り、先に作成した木構造を元にレイアウト処理をし、以下先に説明した処理と同様にして、ステップS107以降の処理を続ける。

また、ステップS111の判断の結果、現在と異なるブロックと判断された場合
10 合は、ステップS104に移り、表示に必要なブロックを読み込み、以下先に説明した処理と同様にして、ステップS105以降の処理を続ける。

一方、ステップS109において、リンクジャンプなどにより、他の領域へ飛ぶように指示された場合、ステップS110に移り、移動先が他ファイルへの移動か自ファイル内への移動かを調べる。例えば、XHTML 文書では、href 属性をも
15 つ<a>タグで挟まれた文字列がクリックされると、href 属性の値から、他ファイルへのリンクか、自ファイル内へのリンクかが判断される。その結果、自ファイル内へのリンクであった場合、ステップS111に移り、リンク先が現在のブロックと同じかどうか判断する。この時、ブロック補助データに記録されたリンク先情報3を参照して、リンク先がどのブロックに含まれるかを調べる。例えば、
20 図5の例の場合、領域38～41のブロック情報を参照し、<a>の href 属性で指定されたリンク先のラベルが、ファイル内のどの位置にあるかを調べる。その後、領域21～37のブロック情報を参照して、その位置がどのブロックに含まれるかを調べることでリンク先のブロックを調べる。

求められたリンク先のブロックが現在のブロックと同じであれば、ステップS
25 106に移り、以降、先に説明した処理と同様の処理を行なう。

一方、リンク先のブロックが現在のブロックと別のブロックであれば、ステップS104に移り、そのブロックを読み込み、以降、先に説明した処理と同様の処理を行なう。

また、ステップS110の判断の結果、リンク先が他のファイルであった場合

は、ステップ S 1 0 2 に移り、リンク先のファイルに対して用意されたブロック補助データを読み込み、以降、先に説明した処理と同様の処理を行なう。

上記のようにブロック補助データを用いて処理することで、電子データの一部のみのデータを読み込んで処理することができるため、高速かつ低メモリでの処理
5 ができる。

なお、本実施の形態では、所々で XHTML 文書を例に挙げながら説明したが、以上の説明から明らかなように XHTML 文書の表示装置に対象を限定するものではない。XHTML 文書のようなリンク機能をもち、階層構造を有する電子データの表示装置に対しては、図 3 のブロック補助データ、および図 6 のフローチャートが適用
10 可能である。また、リンク機能をもたない電子データの場合も、図 3 のリンク先情報 3 と図 6 のステップ S 1 1 0 とを省略することで適用可能である。例えば、日本電子出版協会が標準化を進めている JepaX (JEPA 電子出版交換フォーマット)、日本電子ブックコミッティーが提案しているネットワーク電子ブックフォーマット (Network Electronic Book Format) など、XML などにより記録された
15 階層構造を有する電子データであれば、図 6 と同様にして処理ができる。

また、本発明は、ブロック補助データを用意することにより、階層構造をもつ電子データの一部のみを読み出して処理できることに特徴がある。そのため、図 1 のブロック図において、処理部 1 0 5 を処理装置固有の処理に替えることで、表示装置に限らず、他の処理装置に対しても本発明を適用できる。この時、図 6
20 のフローチャートにおいては、ステップ S 1 0 6 ~ S 1 0 8 の処理が、処理装置固有の処理に替わる。例えば、文章の読上げ装置の場合、電子データのうち読上げる個所と読上げない個所とを判断するステップや、個所に応じて読上げる音質や強度を設定するステップや、音声再生するステップなどの処理に置換えられる。

一方、本実施の形態においては、XML などのように階層構造を有する電子データ
25 を対象に説明をしてきたが、HTML 文書のように、完全には階層構造を有しない電子データに対しても適用可能である。

例えば、HTML 文書には<basefont>という基準フォントのサイズを指定するタグがある。図 8 の領域 7 2 に示されるように指定すると、それ以降の文章の階層構造に関係なく、次の<basefont>タグの指定が来るまで、基準フォントのサイズ

が 3 (size="3"で指定) に設定される。例えば、領域 7 2 に示される<basefont> は、<p>タグや<u>タグなどに挟まれているが、<p>や<u>の終了タグが現れても設定が持続されるため、階層構造が崩れてしまっている。このように、階層構造を無視して効果が持続されるようなタグが存在する場合には、このタグを含むブロックの終了位置制御コードにそのタグの終了タグも記録し、次ブロック以降にも、開始位置制御コードにそのタグを、終了位置制御コードにそのタグの終了タグを追加しておくことにより、別のブロックのみを処理する場合でも、そのタグの効果がそのブロックにかかっていることを知ることができる。そのため、図 6 と同様な処理が可能となる。

10 (第 2 の実施の形態)

次に、本発明の第 2 の実施の形態について説明する

図 9 は、本発明の第 2 の実施の形態に係るデータ処理装置 2 0 0 のブロック図の一例である。

図 9 を参照して、電子データ 2 0 1 は、本データ処理装置 2 0 0 で処理される電子データである。ブロック補助データ 2 0 2 は、電子データ 2 0 1 に付随するブロック補助データである。

入力部 2 0 3 は、電子データ 2 0 1 とブロック補助データ 2 0 2 とを読み込む。データ構造解析部 2 0 4 は、電子データ 2 0 1 とブロック補助データ 2 0 2 とからデータの階層構造を解析する。ブロック補助データ作成部 2 0 5 は、ブロック補助データ 2 0 2 が存在しない場合に電子データ 2 0 1 からブロック補助データを作成する。処理部 2 0 6 は、データ構造解析部 2 0 4 で解析した階層構造に基づき所定の処理を行なう。制御部 2 1 0 は、入力部 2 0 3、データ構造解析部 2 0 4、ブロック補助データ作成部 2 0 5、および処理部 2 0 6 の各処理部を制御する。電子データ 2 0 1 とブロック補助データ 2 0 2 とは、第 1 の実施の形態と同様に記録媒体 1 1 1 に記録され、データ処理装置 2 0 0 に読み込まれる。

上述の処理部 2 0 6 は、対象とする電子データの内容、および処理内容によって構成が異なるが、このデータ処理装置 2 0 0 が文章などの表示を行なう表示装置である場合には、図 9 に示すように、データ構造解析部 2 0 4 の解析結果に基づいて、文章の表示レイアウトを計算するレイアウト計算部 2 0 7 と、計算した

表示レイアウトに基づいて表示する表示部 209 と、スクロールなどのユーザからの指示を処理するユーザ指示処理部 208 とから構成される。

図 10 に処理装置 200 での処理のフローチャートを示す。

図 10 を参照して、まず、ユーザより、キーボード、マウス、およびペン等を使用して、処理する電子データが指定されると（ステップ S 201）、ユーザ指示処理部 208 を通じステップ S 202 により、その電子データに対するブロック補助データが存在するかどうか判断される。もし、存在していれば、上述の第 1 の実施の形態と同様の処理となり、図 6 のステップ S 102 以降の処理が行なわれる。

もし、ステップ S 202 において対応するブロック補助データが存在しないと判断されれば、ステップ S 203 によりブロック補助データが作成され、引続き図 6 のステップ S 102 以降の処理が行なわれる。ステップ S 102 以降の処理は、第 1 の実施の形態で説明した処理と同じであるため、ここでの説明を繰返さない。

次に、ステップ S 203 の処理について、詳しく説明する。

ブロック補助データ作成部 205 では、入力部 202 より入力した電子データを複数のブロックに分割し、それぞれのブロックの開始／終了位置での制御コードを調べ、図 3 に示したブロック補助データを作成する。

この、ステップ S 203 での処理のフローチャートを図 11 に示す。

図 11 を参照して、まず、ブロックサイズの目標値 T を設定する（ステップ S 301）。先述の如く、適切なブロックサイズは、処置装置の処理能力や、画面に表示される文字数およびそれを決定する要因に基づいて決定されることが望ましい。そのため、これらのパラメータに基づきブロックサイズの目標値 T を設定する。なお、目標値 T の設定には、処理装置が予め備えるデフォルト値、もしくはユーザにより指定された値を用いてもよい。

図 5 に示されるブロック補助データは、ブロックサイズの目標値を 200 バイトに設定して作成されたものである。ここで、設定するブロックサイズを目標値としたのは、一般に、ブロック分割できる個所に制限があるためである。詳しくはステップ S 303 の説明で述べる。

さて、ステップS 3 0 1によりブロックサイズの目標値Tが設定されると、ステップS 3 0 2に移り、ファイル先頭からTバイト目を挟んだXバイトの領域を、ブロック境界の探索範囲に設定する。Xの値は、例えば、ブロックサイズの目標値Tの半分に設定する。そして、ステップS 3 0 3において、その探索範囲内における境界候補を抽出する。

境界候補は、対象とする電子データの種類によって変わるが、例えば、電子データがXML文書である場合は、区切る個所はタグやコメント、および文字列の途中ではなく、タグの直前か直後かで区切る。例えば、"

さらに、データ処理装置2 0 0がXHTML文書のような電子書籍を表示する表示装置である場合は、上記の制限に加え、改行タグの直後や、段落の先頭など、行頭から表示が始まる個所を境界候補とすることが望ましい。例えば、図1 2 Aに具体例が示されるXHTML文書を、行の途中で2つのブロック6 0、6 1に区切った場合を考える。この時、XHTML文書の先頭から表示した場合の表示例を図1 2 Bに示す。

ブロック6 0が行の途中で終わる場合、ブロック6 1は行の途中（3行目の7文字目以降）からレイアウトされることになる。一方、ユーザ等により、ブロック6 1の先頭から表示するよう指示された場合、ブロック6 1のみを読込んでレイアウト計算がされるため、図1 2 Cのようにブロック6 1は行の先端から表示される。そのため、ユーザによって、ブロック6 1から6 0へ、文章とは逆方向にスクロール指示がされると、図1 2 Bと図1 2 Cのようにブロック6 1の表示開始位置が異なるため、ブロック切替わり時に表示が乱れてしまう。

これに対し、ブロックの分割位置を行頭から表示が始まる個所に限定しておく

20

と、直前のブロックのレイアウト結果によることなく、常に行頭からレイアウトすることになるため、このような問題が生じない。よって、XHTML 文書のような電子書籍の場合は、ブロックの境界候補を、タグの直前もしくは直後であり、しかも表示が常に行頭となる個所から抽出する。

- 5 次に、ステップ S 3 0 3 により抽出された境界候補の数により、ステップ S 3 0 4 により条件が分岐される。

もし境界候補が見つからなければ、ステップ S 3 0 2 に移り、探索範囲をファイルの下方にずらして、先に述べたのと同様にしてステップ S 3 0 2 以降の処理をする。

- 10 一方、境界候補が存在していれば、ステップ S 3 0 2 で設定した探索範囲の中心に一番近い候補を選び、境界に設定する（ステップ S 3 0 5）。そして、ステップ S 3 0 6 により、ブロックの開始／終了位置の、ファイル先頭からの位置および階層関係を調べ、ブロック情報 2 に記録する 1 ブロック分の情報を取得する。

- 15 そしてこのブロックの終端がファイルの終端かどうかを調べ（ステップ S 3 0 8）、終端でなければステップ S 3 0 2 に移り、次のブロックの調査を続ける。

- 20 一方、ファイルの終端と一致すれば、ステップ S 3 0 8 により、S 3 0 1 から S 3 0 7 で調べたブロック情報と電子データのファイル名とをブロック補助データに記録する。この時、もし必要ならばリンク先情報も調査し、ブロック補助データに記録するようにする。ステップ S 3 0 8 が終わると、ブロック補助データ作成処理が終了する。

なお、本実施の形態は、ブロック補助データ作成部 2 0 5 がデータ処理装置 2 0 0 内にある場合について説明したが、ブロック補助データ作成部 2 0 5 を第 1 の実施の形態で説明したサーバ 1 1 0 に備え、そこでブロック補助データを作成することも可能である。

- 25 この場合の処理の流れは本実施の形態で説明した通りである。また、ブロック補助データ作成部 2 0 5 の機能を実現するハードウェアまたはソフトウェアを汎用のパーソナルコンピュータ（図示せず）に組込むことにより、一般的な文書記述言語で記述されたコンテンツを本願特有のデータ構造を備えるコンテンツに変換させることができる。このようにして生成されたコンテンツを図 1 のサーバ 1

10にアップロードして、サーバ110からユーザのパソコン115にダウンロードさせることができる。このような構成により、データ処理装置100に表示させるコンテンツを作成し販売するシステムを構築することができる。このようなシステムは、データ処理装置100が電子ブックビューアの場合には、汎用の文書記述言語で記述された電子ブックコンテンツをデータ処理装置100専用のデータ構造に変換して提供する場合に有効である。

また、本実施の形態では所々で表示装置を例に挙げて説明したが、以上の説明から明らかなように、本発明は、ブロック補助データを作成して処理することによって、階層構造をもつ電子データの一部分のみを読出して処理できることに特徴がある。そのため、図9のブロック図において、処理部206を処理装置固有の処理部に替えることで、表示装置に限らず、一般のデータ処理装置に対しても本発明を適用できる。

(第3の実施の形態)

次に、本発明の第3の実施の形態について説明する。

図13は、本発明の第3の実施の形態に係るデータ処理装置のブロック図の一例である。説明のため、データ表示装置を例に挙げて説明する。

図13に示すように、本実施の形態におけるデータ表示装置は、文書DB（データベース）301、サーバ処理部302、およびクライアント処理部304の大きく3つに分かれ、サーバ処理装置302とクライアント処理部304とはネットワーク回線303により接続されている。

文書DB301には処理する電子データおよびそれに付随するブロック補助データが格納されている。

サーバ処理部302は、文書DB301から電子データとブロック補助データとを読み込む入力部305と、電子データに対するブロック補助データが存在しないときにブロック補助データを作成し文書DB301に記録するブロック補助データ作成部306と、クライアント処理部304からの要求を受付けると共に、指定されたデータをクライアント処理部304に送信する送受信部307とから構成される。

クライアント処理部304は、処理する電子データやスクロール指示などのユ

ユーザからの指示を処理するユーザ指示処理部 309 と、ユーザ指示処理部 309 で解析したユーザからの指示内容などをサーバ処理部 302 に送信すると共に、サーバ処理部 302 から送られてきたデータを受信する送受信部 308 と、サーバ処理部 302 から送られてきた電子データの階層構造を解析するデータ構造解析部 310 と、データ構造解析部 310 で解析された階層構造を用いて、電子データの表示レイアウトを計算するレイアウト計算部 311 と、レイアウト計算部 311 によって計算したレイアウトに基づき表示する表示部 312 とから構成される。なお、レイアウト計算部 311 によって表示レイアウトを計算した結果、表示部 312 の途中までしか表示レイアウトが定まらなかった場合、送受信部 308 を通じて、必要なデータを送信するようにサーバ処理部 302 に要求を出す場合もある。

次に、図 14、図 15 を使って、本実施の形態のデータ表示装置における処理の流れを説明する。

図 14 を参照して、まず、ユーザから処理したい電子データが指定されると、サーバ処理部 302 に処理したい電子データのファイル名が送られる（ステップ S301）。サーバ処理部 302 では、その電子データに対するブロック補助データが文書 DB 301 内に存在するかどうかを調べ（ステップ S302）、存在していなければステップ S303 に移り、ブロック補助データを作成し、ステップ S304 に移る。ステップ S303 での処理は、図 11 を用いて説明した処理と同様であるため、ここでの説明は繰返さない。一方、ステップ S302 でブロック補助データが存在していた場合は、何もせずステップ S304 に移る。

そしてステップ S304 において、ブロック補助データがクライアント処理部 304 に送られ、クライアント処理部 304 で受信される（ステップ S305）。

なお、このうちステップ S301 において、表示部 312 の画面サイズやメモリ量などからブロック分割する際のサイズの目標値を、電子データのファイル名と共に伝え、サーバ処理部 302 はそのブロックサイズを元に、ブロック補助データを作成するようにしてもよい。このように構成することで、クライアント処理部 304 の処理能力に応じたブロック分割が可能となり、最終的にユーザの使い勝手の向上につながる。

次に、図 15 を参照して、クライアント処理部 304 は、ブロック補助データを受信すると、受取ったブロック補助データを解析し、各ブロックの開始／終了位置と、画面に表示しようとしている領域のファイル先頭からの位置を元に、どのブロックを読み込みかを判断し、読みみたいブロックをサーバ処理部 302 に伝える（ステップ S306）。

サーバ処理部 302 はその要求を受付けると、指定されたブロックを文書 DB 301 から読み込み、クライアント処理部 304 に返す（ステップ S307）。クライアント処理部 304 は、受取ったブロックデータとブロック補助データの内容とから、先に述べたステップ S105 と同様の処理を行ない、以降ステップ S106 から、先に述べたステップ S105 と同様の処理を行なう（ステップ S309～ステップ S314）。

このようにすることで、処理する電子データがサーバ上にある場合でも、階層構造を考慮しながら電子データの一部のみを読み込んで処理ができるため、電子データ全体を読み込んで処理する場合に比べ、高速に処理ができ、使用するメモリも少なく済む。また電子データの一部のみを処理する場合には、ネットワーク上でやり取りするデータ量を減らすことも可能である。

なお、本実施の形態においては、ブロック補助データ作成部 306 がサーバ処理部 302 内にある場合について説明したが、ブロック補助データ作成部 306 を文書 DB 301、あるいはクライアント処理部 304 に備え、ブロック補助データを文書 DB 301、あるいはクライアント処理部 304 で作成することも可能である。

また、上記の説明ではデータ表示装置を例に挙げて説明したが、以上の説明から明らかなように、本発明はブロック補助データを用いて処理することによって、階層構造をもつ電子データの一部のみを受信して処理できることに特徴がある。そのため図 13 のブロック図において、レイアウト計算部 311 および表示部 312 を処理装置固有の処理部に替えることで、表示装置に限らず、一般のデータ処理装置に対しても本発明を適用できる。

また、上述の第 1～3 の実施の形態における処理の一部または全部を、コンピュータによる処理に適した命令の順番付けられた列からなるもの（プログラム）

として提供することも可能である。さらに、これらのプログラムは、そのプログラムのインストール、実行、プログラムの流通のために、そのプログラムを記録したコンピュータ読取可能な記録媒体として提供することも可能である。

5 また、上述のプログラムまたは第1～3の実施の形態におけるデータ構造を備えるコンテンツデータを、ネットワークを介してサーバ装置からクライアント装置に送信して提供することも可能である。その場合は、例えば、図1のサーバ110にプログラムまたはコンテンツデータを送信する送信部を備えることにより実現可能である。

10 本実施の形態におけるデータベース処理装置が、以上のように構成されていることで、階層構造をもつ電子データであっても、電子データに対して用意したブロック補助データを用いて、階層構造を考慮しながら、電子データの一部分のみを読み込んで処理ができる。そのため、電子データ全体を読み込んで処理する場合に比べ、高速に処理ができ、使用するメモリも少なくて済む。また、ブロック補助データにリンク先の位置情報を用意することで、XHTML 文書のリンク機能のよう
15 に任意のブロックに移動する場合であっても、高速に移動ができる。

また、第2の実施の形態におけるデータ処理装置では、ブロック補助データをもたない電子データであっても、いったんブロック補助データを作成した後に処理するようにしたので、ブロック補助データをもたない電子データに対しても、同様に高速かつ低メモリ使用量で処理ができる。

20 さらに本実施の形態におけるデータ処理装置は、ブロックの大きさをほぼ均等な大きさに分割するので、ユーザからのスクロール指示によって前後のブロックに表示を移動する場合であっても、ほぼ同じ処理時間で表示される。そのため、ユーザにあまり違和感を感じさせないというメリットがある。また、サイズが極端に大きいブロックが存在すると、処理装置によっては、作業用メモリが不足して正常に動作しない可能性があるが、大きさがほぼ均等となっているため、その
25 ような問題も生じにくくなる。

また、第3の実施形態におけるデータ処理装置では、電子データおよびブロック補助データがネットワーク接続されたサーバ上にある場合でも、階層構造を考慮しながら、電子データの一部分のみを読み込んで処理ができる。そのため、電子

データ全体をサーバからダウンロードして処理する場合に比べ、高速に処理ができ、使用するメモリも少なく済む。

- 5 なお、今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

産業上の利用可能性

- 10 以上のように、この発明によれば、階層構造をもつ電子データを、高速にかつ低メモリで処理することができるので、本発明は、データ処理方法、データ処理プログラムおよびデータ処理装置に有利に適用することができる。

請求の範囲

1. 階層構造を有する電子データと、前記電子データを複数のブロックに分割して処理するためのブロック補助データとからなる構造化データを処理するデータ処理方法であって、
- 5 前記電子データから所望の電子データが含まれるブロックを読み込むステップと、
前記読み込まれたブロックのブロック補助データを用いて前記所望の電子データの階層構造を解析するステップと、
前記解析された結果を用いて所定の処理をするステップとを備える、データ処理方法。
- 10 2. 前記電子データは表示用の電子データであり、前記所定の処理は、前記読み込んだブロックの表示レイアウトを計算し表示する処理であることを特徴とする、請求の範囲の第1項に記載のデータ処理方法。
3. 前記電子データに対するブロック補助データが存在しない場合、前記電子データの処理をする前に、前記電子データに対するブロック補助データを作成するステップをさらに備えることを特徴とする、請求の範囲の第1項に記載のデータ処理方法。
- 15 4. 前記複数の分割された電子データのブロックは、ほぼ等しい大きさであることを特徴とする、請求の範囲の第1項に記載のデータ処理方法。
- 20 5. 前記ブロック補助データは、前記複数の分割された電子データの各ブロックの位置情報と、前記ブロックの開始位置と終了位置とにおける階層情報とを含むことを特徴とする、請求の範囲の第1項に記載のデータ処理方法。
6. 前記電子データに、前記電子データ内の任意の位置にジャンプするリンク機能が含まれる場合、前記ブロック補助データは、前記リンク先の位置情報を含むことを特徴とする、請求の範囲の第1項に記載のデータ処理方法。
- 25 7. 階層構造を有する電子データを複数のブロックに分解し、前記分割された各ブロックごとに所定の処理を実行させるための構造化データを生成するためのデータ処理方法であって、
前記電子データを所定のブロックサイズに分割する際の、各ブロック境界にお

ける境界候補を抽出する抽出ステップと、

前記抽出された境界候補に基づいて、前記ブロックの分割位置を決定する決定ステップと、

- 5 前記決定された各ブロックの先頭位置と終了位置とにおける前記階層構造の特徴を示す情報を取得する取得ステップと、

前記決定ステップで決定された各ブロックの分割位置の位置情報と、前記取得された前記位置情報に対応する前記階層構造の特徴と示す情報とを含むブロック補助データを生成し、前記ブロック補助データを前記電子データに追加して前記構造化データを生成する生成ステップとを備える、データ処理方法。

- 10 8. 前記抽出ステップは、表示したときに行頭から表示される箇所から前記境界候補を抽出することを特徴とする、請求の範囲の第7項に記載のデータ処理方法。

9. 前記抽出ステップは、前記電子データの構造を記録している制御コードの前後とその付近とから前記境界候補を抽出することを特徴とする、請求の範囲の第7項に記載のデータ処理方法。

- 15 10. 前記ブロックサイズは、前記電子データを処理する装置の処理能力と、前記装置の画面に表示される文字数と、前記表示される文字数を決定する要因との少なくとも1つに応じて決められることを特徴とする、請求の範囲の第7項に記載のデータ処理方法。

- 20 11. サーバより階層構造を有する電子データを受取り、前記電子データに対して所定の処理をするデータ処理方法において、

前記サーバに対して前記所定の処理をしたい電子データ名を送信する送信ステップと、

前記サーバより前記電子データを複数のブロックに分割して処理するためのブロック補助データを受信する受信ステップと、

- 25 前記電子データのうち、前記所定の処理をしたい前記電子データを含むブロックの送信を前記サーバに要求し、前記ブロックのデータを受信する受信ステップと、

前記受信したブロックと、前記ブロックの前記ブロック補助データとを用いて、前記ブロックの階層構造を解析し、その結果を用いて前記所定の処理をする再生

ステップとを備える、データ処理方法。

- 1 2. 前記サーバより前記ブロック補助データを受信するステップにおいて、前記ブロック補助データが存在しない場合には、前記ブロック補助データを作成した後、前記サーバより受信することを特徴とする、請求の範囲の第 1 1 項に記載のデータ処理方法。

1 3. 階層構造を有する電子データと、前記電子データを複数のブロックに分割して処理するためのブロック補助データとを対として備える構造化データに対して、所定の文書記述言語の定義に従って所定の処理を実行させるためのデータ処理プログラムであって、

- 10 前記ブロック補助データに基づいて、前記電子データのブロックデータを読み込む読み込みステップと、

前記読み込まれたブロックデータと、前記ブロック補助データとに基づいて、前記ブロックデータに含まれる階層構造を解析する解析ステップと、

- 15 前記解析結果と、前記文書記述言語の定義とに基づいて、前記ブロックデータに対して前記所定の処理を実行させる実行ステップとをコンピュータに実行させる、データ処理プログラム。

1 4. 階層構造を有する電子データを複数のブロックに分割し、前記分割された各ブロックごとに所定の処理を実行させるための構造化データを生成するためのデータ処理プログラムであって、

- 20 前記電子データを所定のブロックサイズに分割する際の、各ブロック境界における境界候補を抽出する抽出ステップと、

前記抽出された境界候補に基づいて、前記ブロックの分割位置を決定する決定ステップと、

- 25 前記決定された各ブロックの先頭位置と終了位置とにおける前記階層構造の特徴を示す情報を取得する取得ステップと、

前記決定ステップで決定された各ブロックの分割情報の位置情報と、前記取得された前記位置情報に対応する前記階層構造の特徴を示す情報とを含むブロック補助データを生成し、前記ブロック補助データを前記電子データに追加して前記構造化データを生成する生成ステップとをコンピュータに実行させる、データ処

理プログラム。

15. サーバより階層構造を有する電子データを受取り、前記電子データに対して所定の処理をするためのデータ処理プログラムであって、

5 前記サーバに対して前記所定の処理をしたい電子データ名を送信する送信ステップと、

前記サーバより前記電子データを複数のブロックに分割して処理するためのブロック補助データを受信する受信ステップと、

10 前記電子データのうち、前記所定の処理をしたい前記電子データを含むブロックの送信を前記サーバに要求し、前記ブロックのデータを受信する受信ステップと、

前記受信したブロックと、前記ブロックの前記ブロック補助データとを用いて、前記ブロックの階層構造を解析し、その結果を用いて前記所定の処理をする再生ステップとをコンピュータに実行させる、データ処理プログラム。

16. 階層構造を有する電子データを処理するデータ処理装置であって、

15 前記電子データには、前記電子データを複数のブロックに分割して処理するための補助情報であるブロック補助データが付随し、

前記電子データのうち、処理をしたいデータが含まれるブロックと、前記ブロック補助データとを読み込む入力部と、

20 前記ブロック補助データを用いて、前記読込んだ前記ブロックの階層構造を解析するデータ構造解析部と、

前記データ構造解析部の結果を用いて、所定の処理をする処理部を備える、データ処理装置。

17. 前記電子データは表示用の文書データであり、前記処理部は、

25 前記データ構造解析部によって解析された前記ブロックの階層構造を用いて、前記読込んだブロックの表示レイアウトを計算するレイアウト計算部と、

前記レイアウト計算部によって求めた前記表示レイアウトに基づいて表示する表示部とをさらに備える、請求の範囲の第16項に記載のデータ処理装置。

18. サーバより階層構造を有する電子データを受取り、前記電子データに対して所定の処理をするデータ処理装置であって、

前記電子データには、前記電子データを複数のブロックに分割して処理するための補助情報であるブロック補助データが付随し、

- 5 前記電子データのうち、前記所定の処理をしたい電子データが含まれるブロックを前記サーバに伝え、前記ブロックのデータと、前記ブロック補助データとを前記サーバより受信する送受信部と、

前記受信した前記ブロックと、前記ブロック補助データとを用いて、前記受信した前記ブロックの階層構造を解析するデータ構造解析部と、

前記データ構造解析部の結果を用いて、前記所定の処理をする処理部とを備える、データ処理装置。

- 10 19. 携帯型の端末であることを特徴とする、請求の範囲の第16項または第18項に記載のデータ処理装置。

20. 文書記述言語の定義に従って記述された電子データと、前記電子データを複数のブロックに分割して処理するためのブロック補助データとを対として備えたことを特徴とする、構造化データ。

- 15 21. 文書記述言語の定義に従って所定の処理を実行させるための電子データと、前記電子データを複数のブロックに分割し、前記分割された各ブロックごとに前記所定の処理を実行させるためのブロック補助データとを対として備えたことを特徴とする、構造化データ。

- 20 22. 前記複数の分割された電子データのブロックは、ほぼ等しい大きさであることを特徴とする、請求の範囲の第20項または第21項に記載の構造化データ。

23. 前記ブロック補助データは、少なくとも前記各ブロックの位置情報と、前記ブロックの開始位置と終了位置とにおける階層情報とを有することを特徴とする、請求の範囲の第20項または第21項に記載の構造化データ。

- 25 24. 前記ブロック補助データは、前記ブロックの分割箇所を、前記電子データの構造を記録している制御コードの前後とその付近とから選ぶことを特徴とする、請求の範囲の第20項または第21項に記載の構造化データ。

25 25. 前記ブロック補助データは、前記ブロックの分割箇所を、表示したときに行頭から表示される箇所限定することを特徴とする、請求の範囲の第20項または第21項に記載の構造化データ。

26. 前記電子データに、前記電子データ内の任意の位置にジャンプするリンク機能が含まれる場合、前記ブロック補助データは、前記リンク先の位置情報を含むことを特徴とする、請求の範囲の第20項または第21項に記載の構造化データ。
- 5 27. 文書記述言語の定義に従って記述された電子データと、前記電子データを複数のブロックに分割して処理するためのブロック補助データとを対として備えたことを特徴とする構造化データを記録した、コンピュータ読取可能な記録媒体。
28. 請求の範囲の第13項～第15項のいずれかに記載のデータ処理プログラムを送信する送信部を備えることを特徴とする、送信装置。
- 10 29. 請求の範囲の第20項または第21項に記載の階層化データを送信する送信部を備えることを特徴とする、送信装置。

FIG. 1

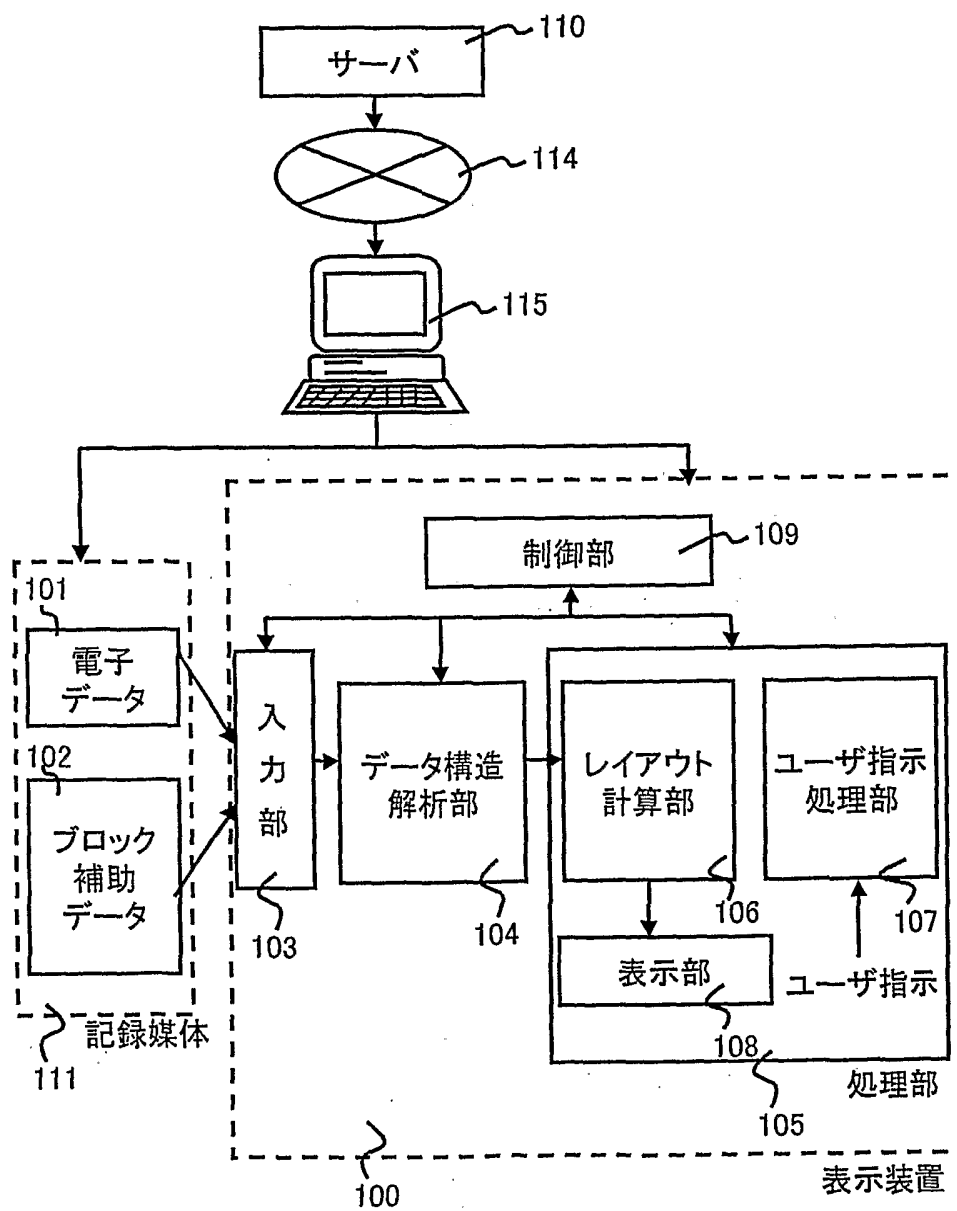


FIG. 2

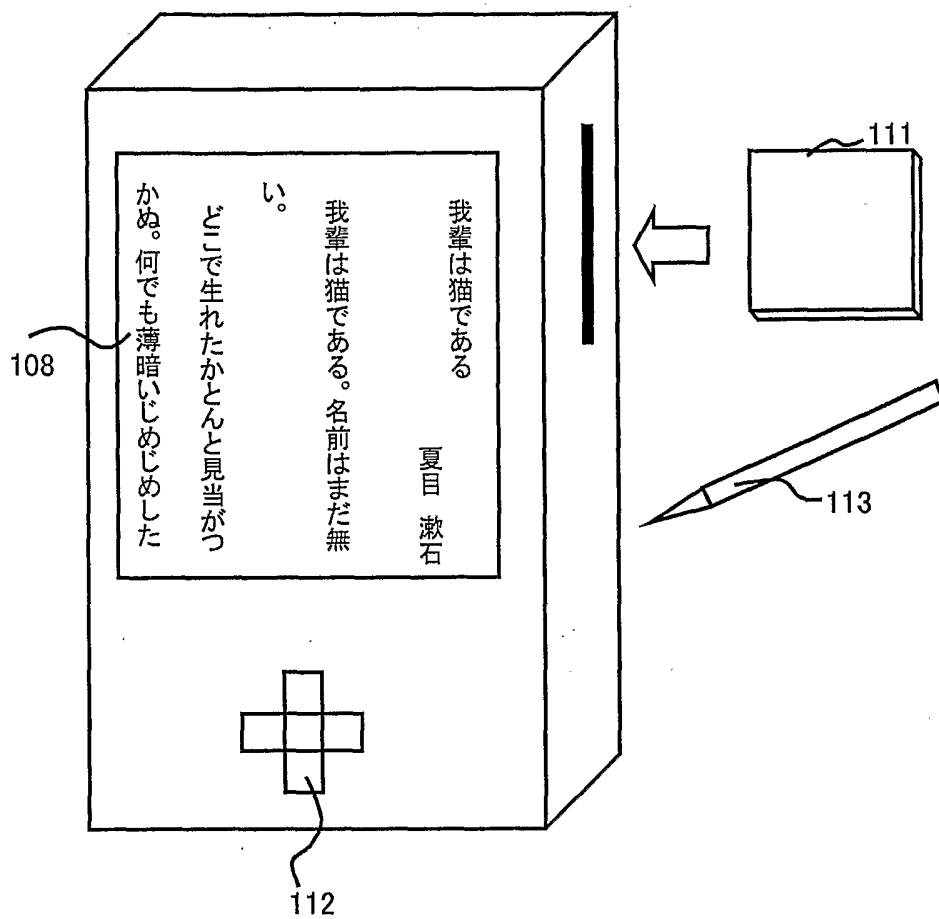


FIG. 3

電子データファイル名		1
ブロック数	ブロック情報	2
ブロック1開始位置		
ブロック1終了位置		
ブロック1開始位置制御コード		
ブロック1終了位置制御コード		
ブロック2開始位置		
ブロック2終了位置		
ブロック2開始位置制御コード		
ブロック2終了位置制御コード		
…以降、ブロック数分繰返される		
リンク先登録数	リンク先情報	3
リンク先1ラベル		
リンク先1開始位置		
リンク先1終了位置		
リンク先2ラベル		
リンク先2開始位置		
リンク先2終了位置		
…以降、リンク先登録数分繰返される		

FIG. 4

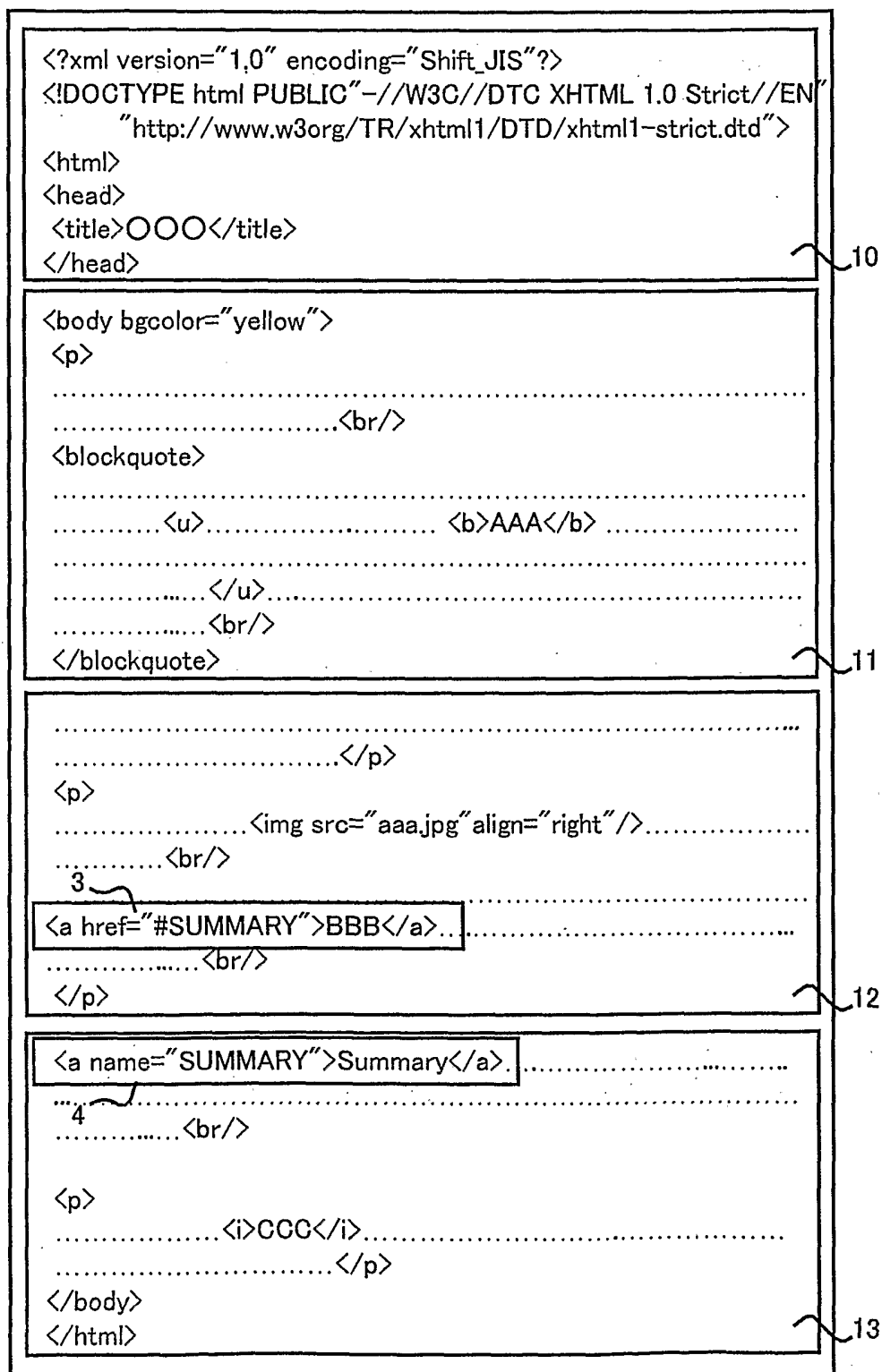


FIG. 5

電子データファイル名	document.html	20
ブロック数	4	21
ブロック1開始位置	0	22
ブロック1終了位置	211	23
ブロック1開始位置制御コード	なし	24
ブロック1終了位置制御コード	</html>	25
ブロック2開始位置	212	26
ブロック2終了位置	423	27
ブロック2開始位置制御コード	<html>	28
ブロック2終了位置制御コード	</p></body></html>	29
ブロック3開始位置	424	30
ブロック3終了位置	652	31
ブロック3開始位置制御コード	<html><body bgcolor="yellow"> <p>	32
ブロック3終了位置制御コード	</body></html>	33
ブロック4開始位置	653	34
ブロック4終了位置	723	35
ブロック4開始位置制御コード	<html><body bgcolor="yellow">	36
ブロック4終了位置制御コード	なし	37
リンク先登録数	1	38
リンク先1ラベル	SUMMARY	39
リンク先1開始位置	671	40
リンク先1終了位置	678	41

FIG. 6

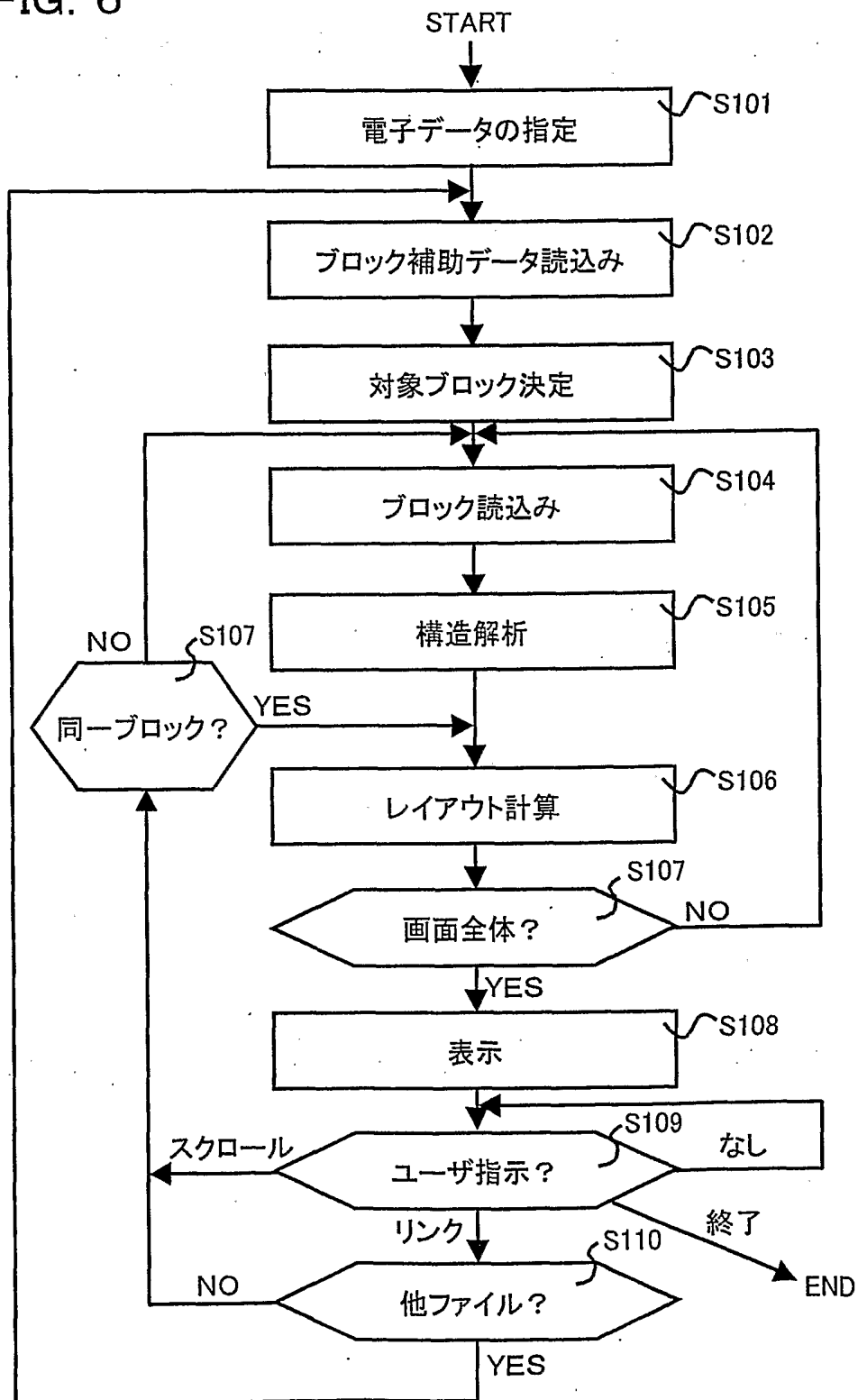


FIG. 7

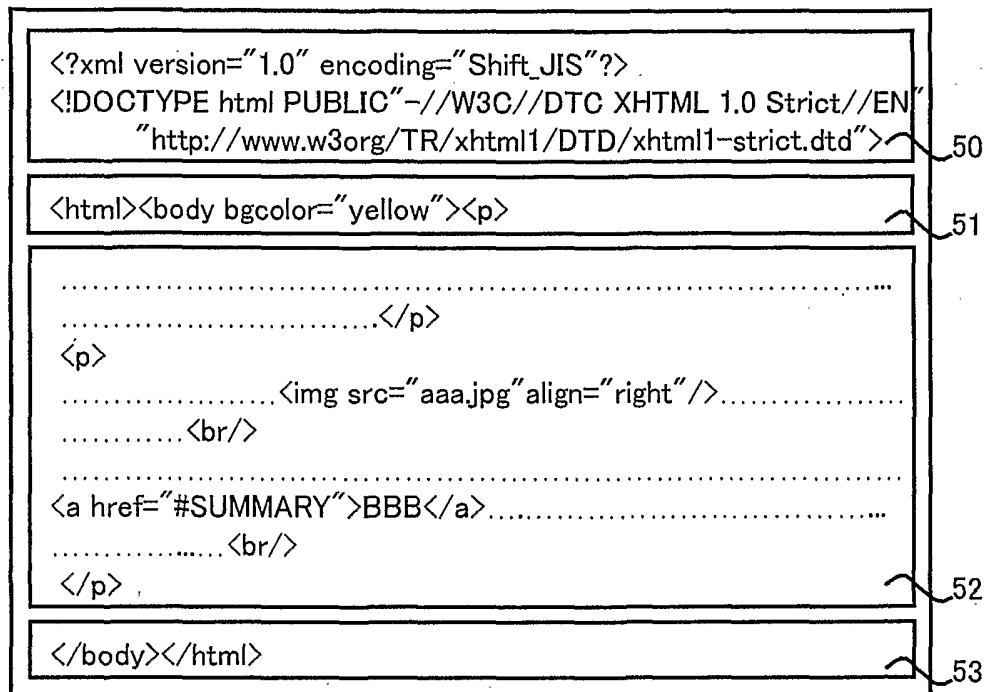


FIG. 8

```

<html>
<head>
  <title>〇〇〇</title>
</head>

<body>
  <p>
    .....<br/>
    .....
    .....<u>.....
    <basefont size="3">..... 72
    .....<b>.....</b>.....
    .....</u>.....
    .....<br/>
    .....
    .....</p>
  <p>
    ..........
    .....<br/>
    .....
  </body>
</html>

```

FIG. 9

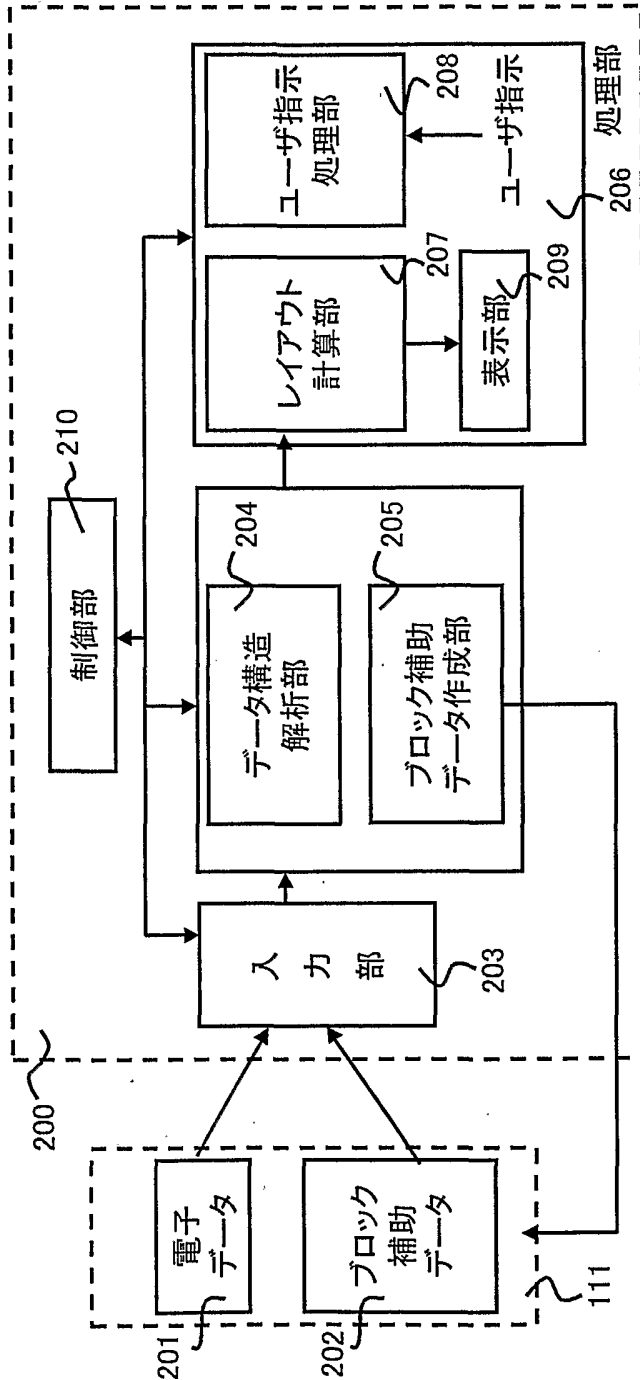


FIG. 10

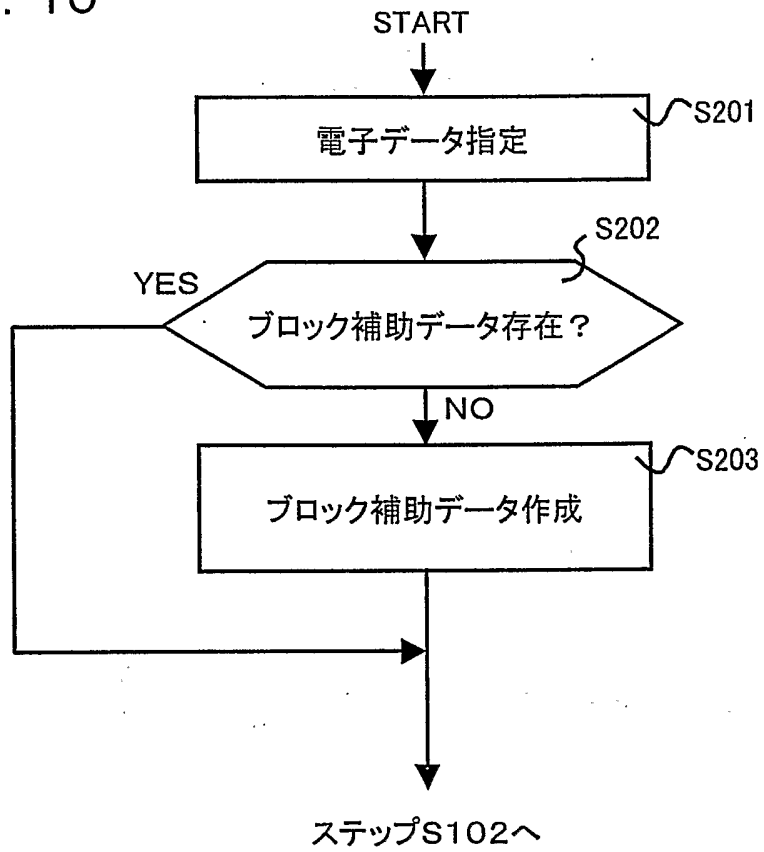


FIG. 11

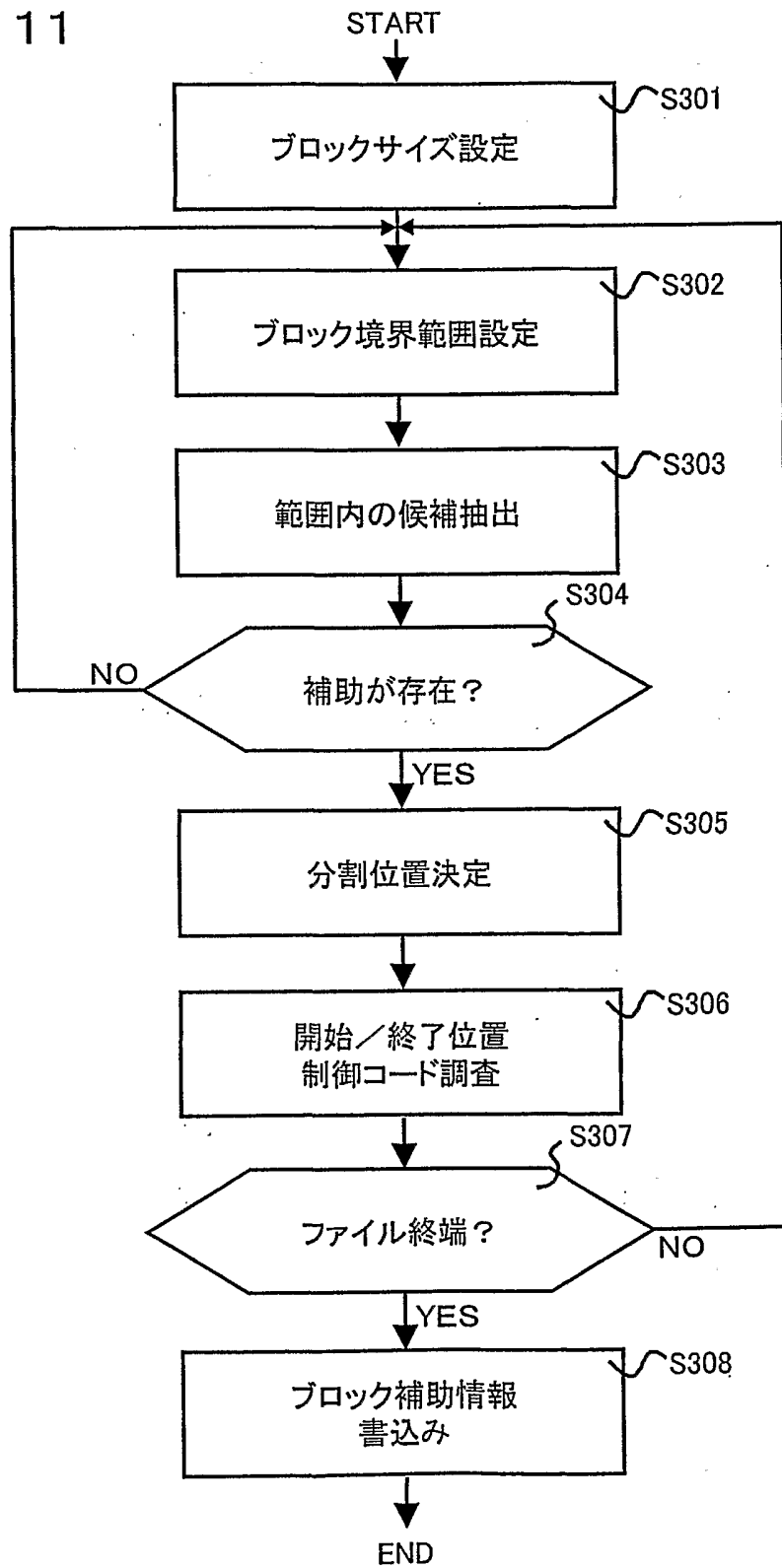


FIG. 12A

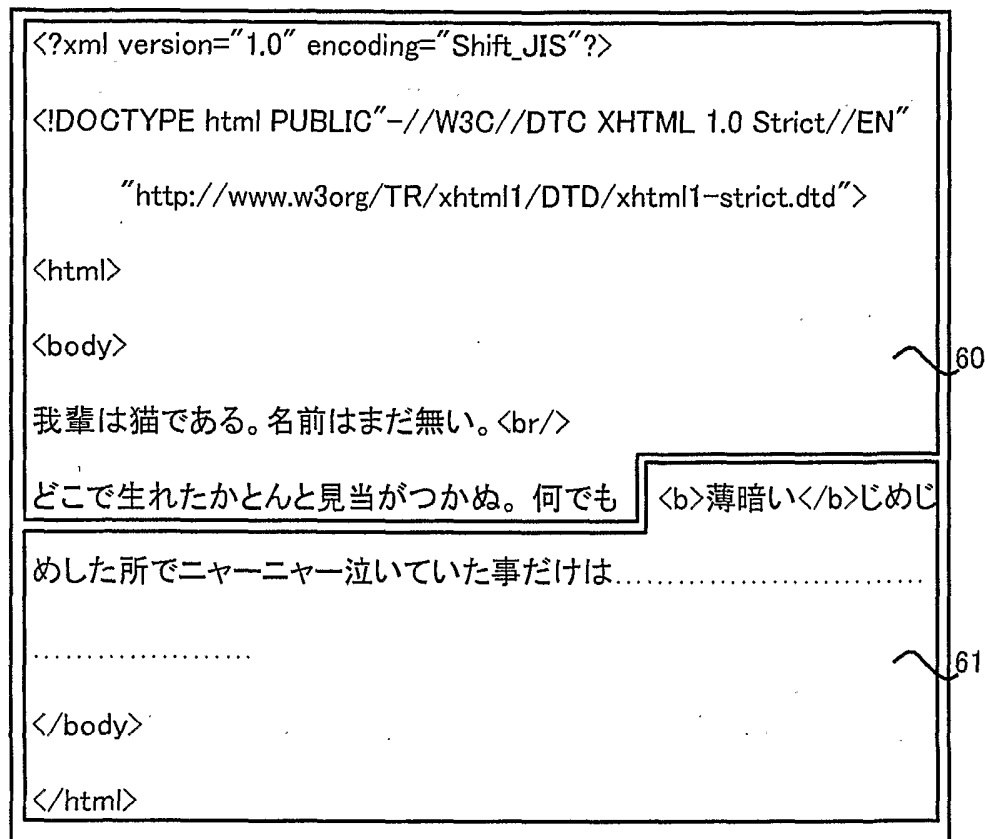


FIG. 12B

我輩は猫である。名前はまだ無い。
 どこで生れたかとうんと見当がつかぬ。
 何でも薄暗いじめじめした所でニャーニャー泣いていた事だ

FIG. 12C

薄暗いじめじめした所でニャーニャー泣いていた事だけは.....

FIG. 13

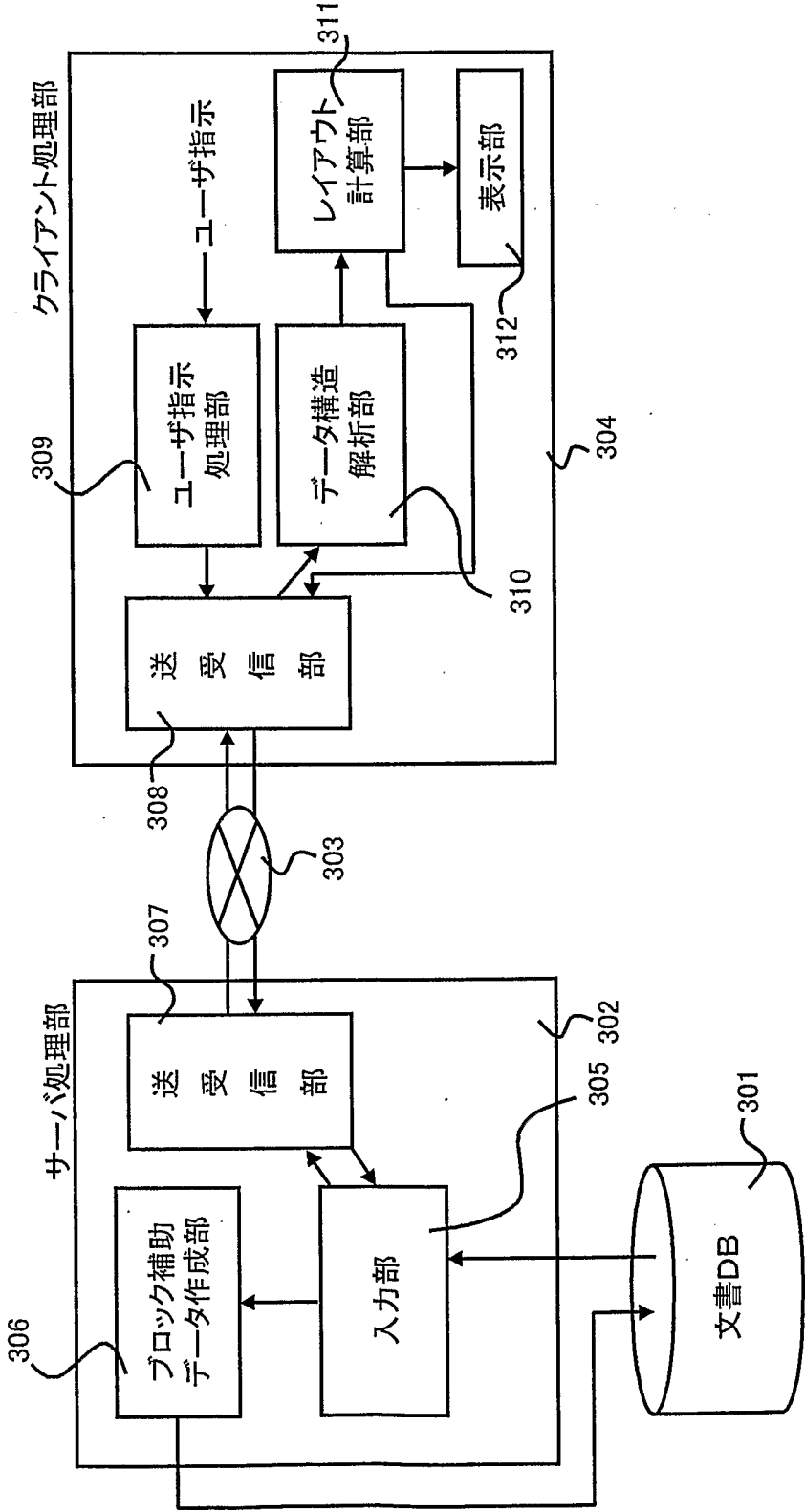


FIG. 14

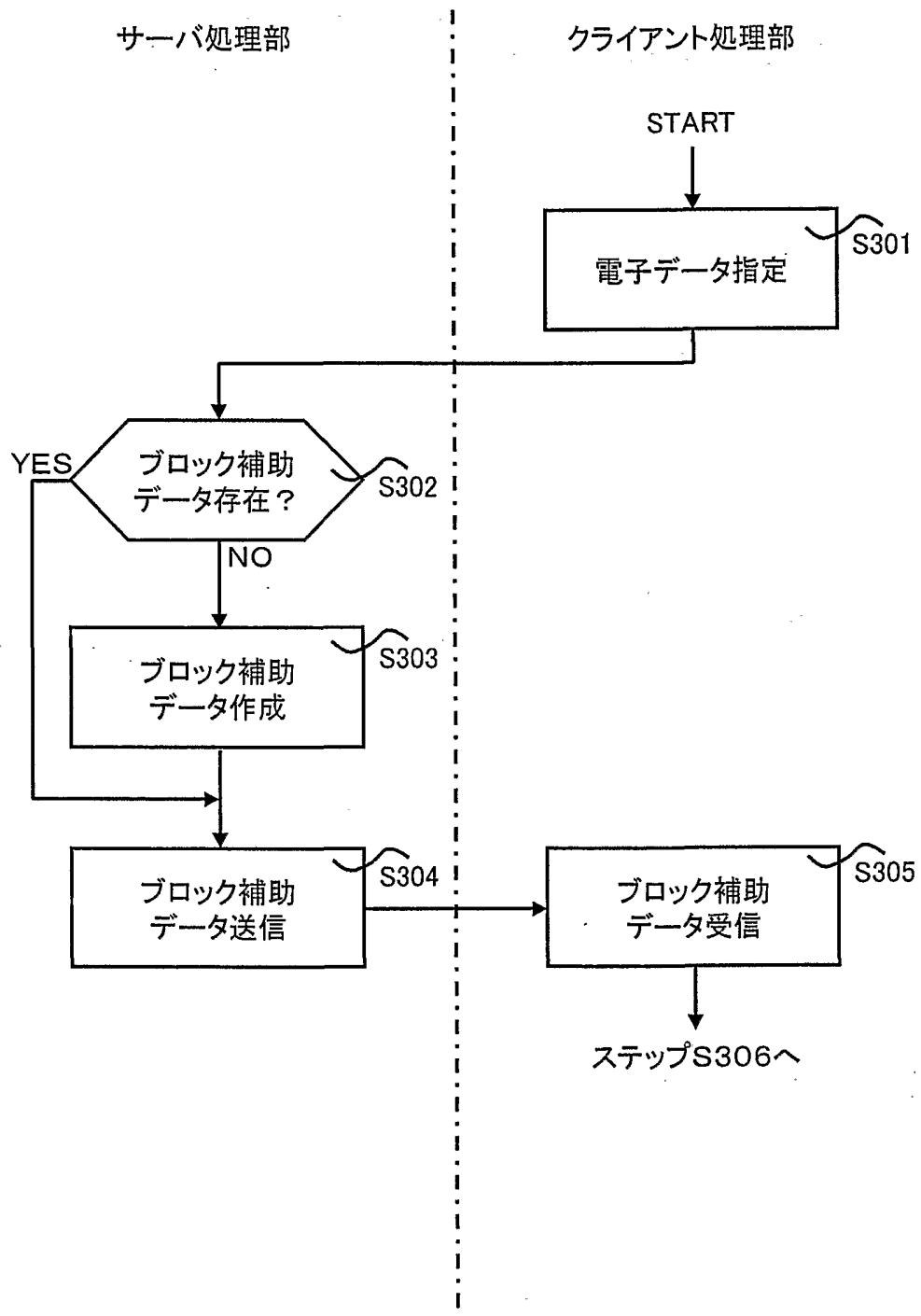


FIG. 15

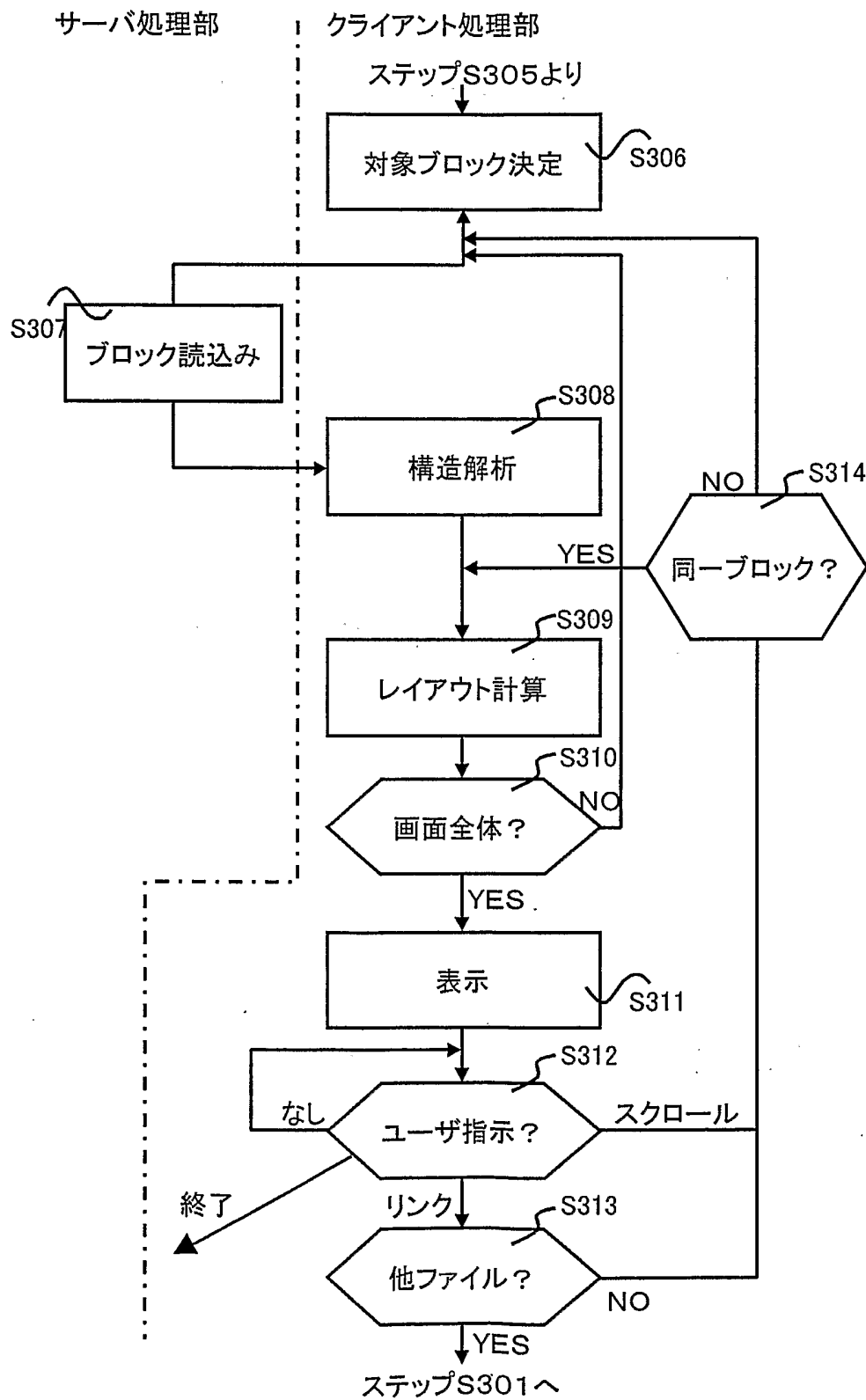


FIG. 16

```
<BOOK>
<BOOKINFO>
  <TITLE>about XML</TITLE>
  <AUTHOR>Yamada</AUTHOR>
</BOOKINFO>
<BODY>
  <CHAPTER>
    Nowadays,<B>XML</B>is ...
  </CHAPTER>
  <CHAPTER> ..... </CHAPTER>
</BODY>
</BOOK>
```

FIG. 17

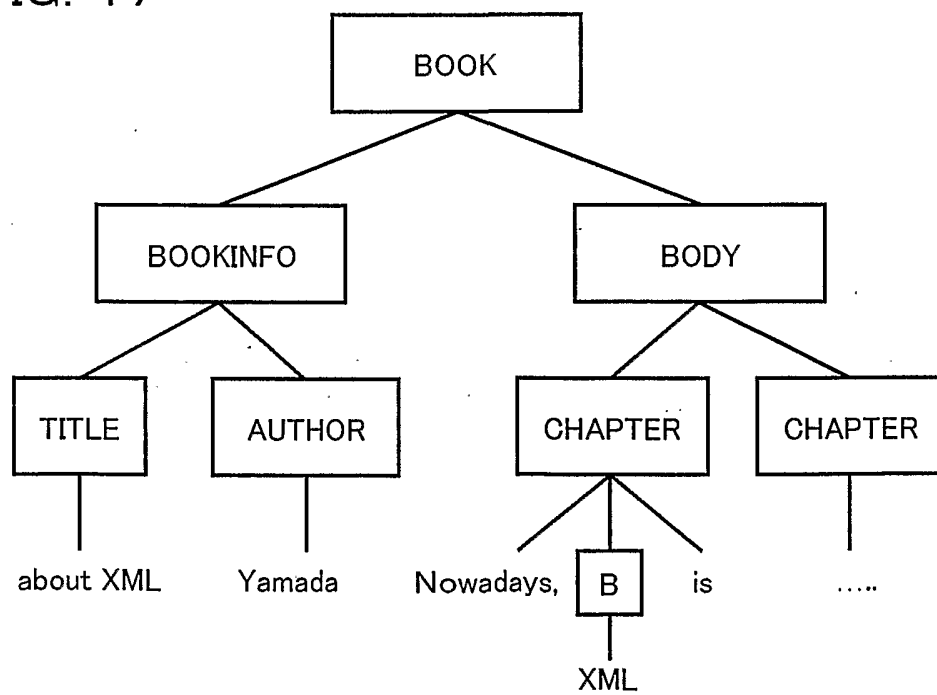


FIG. 18

```

<?xml version="1.0" encoding="Shift_JIS"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<title>〇〇〇</title>
</head>
<body bgcolor="yellow">
<p>
.....
.....<br/>
<blockquote>
.....
.....<u>.....<b>AAA</b>.....
.....</u>.....
.....<br/>
</blockquote>
.....
.....</p>
<p>
..........
.....<br/>
.....<a href="#SUMMARY">BBB</a>.....
.....<br/>
</p>
.....
.....<a name="SUMMARY">Summary</a>.....
.....<br/>
.....
<p>
.....<i>CCC</i>.....
.....</p>
</body>
</html>

```

71

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP02/05880

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G06F17/21

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)
Int.Cl⁷ G06F17/21-17/24

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2002
Kokai Jitsuyo Shinan Koho 1971-2002 Jitsuyo Shinan Toroku Koho 1996-2002

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 10-143498 A (Nippon Telegraph And Telephone Corp.), 29 May, 1998 (29.05.98), Full text; all drawings (Family: none)	1-29
A	JP 10-269160 A (Matsushita Electric Industrial Co., Ltd.), 09 October, 1998 (09.10.98), Full text; all drawings (Family: none)	1-29
P,A	JP 2001-195391 A (Nihon Denki Joho Service Kabushiki Kaisha), 19 July, 2001 (19.07.01), Fig. 4 (Family: none)	1-29

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "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) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"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 "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 "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 "&" document member of the same patent family
---	--

Date of the actual completion of the international search
26 August, 2002 (26.08.02)

Date of mailing of the international search report
10 September, 2002 (10.09.02)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))		
Int. Cl ⁷ G06F17/21		
B. 調査を行った分野		
調査を行った最小限資料 (国際特許分類 (IPC))		
Int. Cl ⁷ G06F17/21-17/24		
最小限資料以外の資料で調査を行った分野に含まれるもの		
日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2002年 日本国登録実用新案公報 1994-2002年 日本国実用新案登録公報 1996-2002年		
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 10-143498 A(日本電信電話株式会社), 1998. 05. 29, 全文、全図 (ファミリーなし)	1-29
A	JP 10-269160 A(松下電器産業株式会社), 1998. 10. 09, 全文、全図 (ファミリーなし)	1-29
PA	JP 2001-195391 A(日本電気情報サービス株式会社), 2001. 07. 19, 図4 (ファミリーなし)	1-29
☐ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献		
国際調査を完了した日	26. 08. 02	国際調査報告の発送日 10.09.02
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 成瀬 博之	5M 9192  電話番号 03-3581-1101 内線 3597