(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/65355 A1

(51) International Patent Classification[7]: G06F 7/00, 15/00

(21) International Application Number: PCT/IL01/00190

(22) International Filing Date: 28 February 2001 (28.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/516,119 1 March 2000 (01.03.2000) US

(71) Applicant *(for all designated States except US)*: **CELL-TREX LTD.** [IL/IL]; P.O. Box 193, 60910 Moshav Bnei Tzion (IL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HALAHMI, Erez** [IL/IL]; 9 Frug Street, Petach Tikva (IL).

(74) Agent: **BODNER, Marc**; Plinner, Bodner, Brass, 13 Noach Mozes, 67442 Tel Aviv (IL).

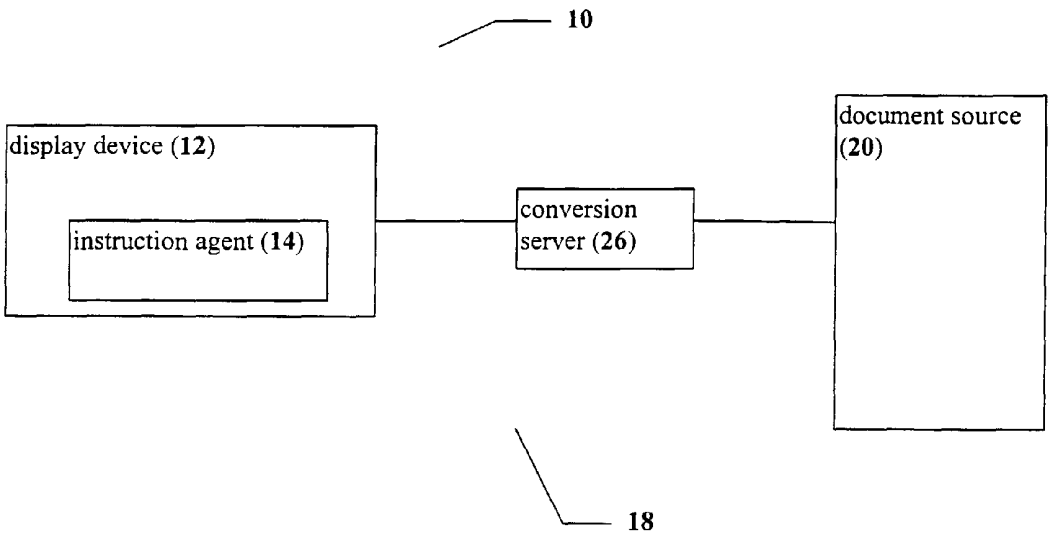
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DE (utility model), DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR RAPID DOCUMENT CONVERSION



(57) Abstract: A method and a system for converting a document in a streamed manner, resulting in rapid transmission and display of each part of a document as that part is converted. The system and method are preferred for operation in environments with limited bandwidth and/or display capacity, such as wireless handheld devices. Since such devices cannot easily receive large amounts of data, and also typically have relatively small display screens, the present invention allows the user to quickly receive and display each part of a document after being converted, rather than waiting for the entire document to be converted and then transmitted before any part is displayed. In addition, the invention is particularly useful for modulator file formats, such as word processing document file formats, in which each module of a file can only be fully interpreted with regard to at least one other module. The present invention utilizes conversion server (26) and instruction agent (14) for conversion of document source (20), with the resulting document displayed on display device (12).

WO 01/65355 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEM AND METHOD FOR RAPID DOCUMENT CONVERSION

FIELD AND BACKGROUND OF THE INVENTION

The present invention relates to a system and method for the rapid, automatic conversion of documents, and in particular, for a system and method which converts such documents in a streamed manner, for example for transmission and display by a WAP (wireless application protocol) enabled device.

Cellular telephones are becoming increasingly popular for portable telephone use, particularly for users who are interested in rapid, mobile communication. As the amount of computational power and memory space which are available in such small, portable electronic devices becomes increased, a demand has arisen for different types of communication services through such devices. In particular, users have demanded that cellular telephones receive many different types of multimedia data, including e-mail (electronic mail) messages and Web pages.

In response to such demands, and to extend the power and efficacy of operation of portable, wireless electronic communication devices, the WAP (wireless application protocol) *de facto* standard has been developed. WAP is now the standard for the presentation and delivery of wireless data, including

multimedia and other information, and telephony services, on mobile telephones and other types of wireless communication devices. WAP is designed to efficiently provide both multimedia and telephony services to such wireless communication devices, given the limitations of wireless networks and of the electronic devices themselves.

Wireless communication devices have requirements and drawbacks which are different than cable-linked electronic devices. For example, wireless networks are frequently significantly less stable than cable networks. Since users with such portable communication devices often operate these devices at different locations, the wireless network connection may not always be available, and may even suddenly become unavailable during a single communication session. In addition, the wireless communication devices themselves are more limited in terms of available resources than desktop computers. For example, such wireless communication devices typically have a less powerful CPU (central processing unit), less memory, a lower amount of available power since these devices are often battery-operated, and smaller display screens. Thus, wireless communication devices require adaptations of existing software and data transmission protocols in order to effectively deliver multimedia content from the Internet.

WAP provides the required adaptations and modifications to such software and data transmission protocols in order to meet the requirements of wireless communication devices. For example, HTML (Hyper-text Mark-up Language) has been adapted to form WML (Wireless Mark-up Language),

which provides a document mark-up language suitable for WAP-enabled devices and their corresponding limitations. WAP-enabled devices are able to receive and display documents written in WML, thereby enabling such devices to display Web pages which are written in WML, for example.

Unfortunately bandwidth considerations still limit the amount of data which can be rapidly received by WAP-enabled devices, such as cellular telephones for example. Therefore, the user may be forced to wait for a significant period of time before an entire document is downloaded for display by the WAP-enabled device. Furthermore, the user may not even wish to view the entire document, but only a portion of such a document. If that portion is located near the end of the document, then the user must wait for data which is not of interest to be downloaded, before the portion of interest can be received by the WAP-enabled device. Also, WAP-enabled devices are not able to display file formats such as Microsoft Word™ documents.

This problem is particularly acute for documents which are not originally designed for display by a WAP-enabled device, such as files which are composed of OLE (Object Linking and Embedding) file components (Microsoft Ltd., USA). Such components, or components of other types of files, are not necessarily sequentially assembled within the file, such that each component must be examined in order to determine its relationship to other such file components, before the component can be converted to a different file format.

For example, files produced by the word processing software program,

Word™ (Microsoft Ltd., USA), are actually assembled from OLE file components. Such files can be converted to text with formatting only after the relevant formatting block arrives for the text block, as the order of the formatting blocks parallels the order of the text blocks to which they refer. Therefore, the relative order of formatting and text blocks, and in particular the relationship between these blocks, must be maintained in order for the conversion to be successful. Thus, a simple solution to this problem is simply to wait until the entire file is received, and then to convert the entire file at once, thereby easily maintaining the relationship between the components.

A more useful solution would involve a "streamed" conversion, in which parts of the file are converted without waiting for the entire file to be received and/or without regard for the sequential order of the components within the file. Such a streamed conversion would enable the user to begin to receive and display the converted document in portions, without waiting for the entire document to be converted. Preferably, the user could also select a portion to be converted and viewed without regard to the location of that portion within the document, such that the user could optionally choose to view the last portion of the document before viewing other portions, for example. Such a solution would be particularly useful for low bandwidth devices such as wireless devices, since each part of the document could be downloaded to the device as soon as that part has been converted. For example, the document could be converted to WML (Wireless Markup Language) in a streamed manner, and then downloaded to, and displayed by, the WAP-enabled wireless device as

soon as each part is ready. Such a solution would clearly be more efficient and would also clearly enable the user to view the document more quickly. Unfortunately, such a solution is not currently available.

There is thus a need for, and it would be useful to have, a system and a method for converting a document in a streamed manner, for example to a WAP-enabled device such as a cellular telephone, such that the device is able to receive and display at least a part of the converted document before the entire document is converted.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, aspects and advantages will be better understood from the following detailed description of a preferred embodiment of the invention with reference to the drawings, wherein:

FIG. 1 is a schematic block diagram of a system according to the present invention;

FIGS. 2A-2C are schematic block diagrams illustrating the conversion system (Figures 2A and 2B) according to the present invention and an exemplary modular document format (Figure 2C); and

FIG. 3 is a flowchart of an exemplary method according to the present invention for converting a document in a streamed manner.

SUMMARY OF THE INVENTION

The present invention is of a method and a system for converting a

document in a streamed manner, for more rapid transmission and display of each part of the document as that part is converted. As described in greater detail below, the present invention is particularly useful for the conversion of documents which are based in discrete blocks with a particular relationship between the blocks, termed "modules" herein for a "modular document". Documents which are in a block format are more difficult to convert in a streamed manner, simply because the relationship between the blocks must be maintained during the conversion process. This necessitates maintaining a buffer in order to review previously examined blocks for enabling the relationship between the blocks to be preserved.

According to the present invention, there is provided a method for converting at least part of a modular document into a converted file format for display to a user, the method comprising the steps of: (a) analyzing at least a part of the modular document to form an analyzed document; (b) separating the analyzed document into a plurality of modules; (c) determining a relationship between at least a pair of modules; and (d) converting at least the pair of modules according to the relationship to form the converted file format.

According to another embodiment of the present invention, there is provided a system for converting a modular document to a converted file format for display to a user, the modular document featuring a plurality of modules having a relationship between at least a pair of modules, the system comprising: (a) a document source for serving the modular document; and (b) a conversion server for receiving the modular document and for converting at least part of

the modular document into the converted file format according to the relationship between at least the pair of modules.

According to still another embodiment of the present invention, there is provided a method for converting at least part of a document into a converted file format for display to a user, the document containing data in a non-sequential order, the method comprising the steps of: (a) analyzing at least a part of the document to form an analyzed document; (b) determining an order for the data in at least a part of the document; and (c) converting at least the part of the document according to the order for the data to form the converted file format.

Hereinafter, the term "network" refers to a connection between any two electronic devices which permits the transmission of data.

Hereinafter, the term "wireless device" refers to any type of electronic device which permits data transmission through a wireless channel, for example through transmission of radio waves. Hereinafter, the term "cellular phone" is a wireless device designed for the transmission of voice data and/or other data, optionally through a connection to the PSTN (public switched telephone network) system.

Hereinafter, the term "computational device" includes, but is not limited to, personal computers (PC) having an operating system such as DOS, Windows™, OS/2™ or Linux; Macintosh™ computers; computers having JAVA™-OS as the operating system; graphical workstations such as the computers of Sun Microsystems™ and Silicon Graphics™, and other computers

having some version of the UNIX operating system such as AIX™ or SOLARIS™ of Sun Microsystems™; or any other known and available operating system, or any device, including but not limited to: laptops, hand-held computers, cellular telephones, wearable computers of any sort, and WAP-enabled devices, as well as any device which can be connected to a network as previously defined and which have an operating system. Hereinafter, the term "Windows™" includes but is not limited to Windows95™, Windows 3.x™ in which "x" is an integer such as "1", Windows NT™, Windows98™, Windows CE™, Windows2000™, and any upgraded versions of these operating systems by Microsoft Corp. (USA).

Hereinafter, the term "Web browser" refers to any software program which can display text, graphics, or both, from Web pages on World Wide Web sites. Hereinafter, the term "Web page" refers to any document written in a mark-up language including, but not limited to, HTML (hypertext mark-up language) or VRML (virtual reality modeling language), dynamic HTML, XML (extensible mark-up language), WML (wireless mark-up language), or related computer languages thereof, as well as to any collection of such documents reachable through one specific Internet address or at one specific World Wide Web site, or any document obtainable through a particular URL (Uniform Resource Locator).

Hereinafter, the term "Web site" refers to at least one Web page, and preferably a plurality of Web pages, virtually connected to form a coherent group. Hereinafter, the term "Web server" refers to software, or a combination of

hardware and software, such as a software program operated by a computational device, which is capable of transmitting at least one Web page upon request by a Web browser.

Hereinafter, the phrase "display a Web page" includes all actions necessary to render at least a portion of the information on the Web page available to the computer user. As such, the phrase includes, but is not limited to, the visual display of graphical information, the audible production of audio information, the animated visual display of animation and the visual display of video stream data.

Hereinafter, unless otherwise noted, a WML card is assumed to be similar or identical to a Web page as previously described for the purposes of describing the present invention.

The method of the present invention could be described as a series of steps performed by a data processor, and as such could optionally be implemented as software, hardware or firmware, or a combination thereof. For the present invention, a software application could be written in substantially any suitable programming language, which could easily be selected by one of ordinary skill in the art. The programming language chosen should be compatible with the computer hardware and operating system according to which the software application is executed. Examples of suitable programming languages include, but are not limited to, C, C++, WMLscript and Java.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is of a method and a system for converting a document in a streamed manner, for more rapid transmission and display of each part of the document as that part is converted. The present invention is preferred for operation in environments with limited bandwidth and/or display capacity, such as for wireless handheld devices, for example. As previously described, such devices cannot easily receive large amounts of data, and also typically have relatively small display screens. Thus, the present invention enables the user to quickly receive and display each part of the document after being converted, rather than waiting for the entire document to be converted and then transmitted before any part is displayed.

As described in greater detail below, the present invention is particularly useful for the conversion of documents which are based in discrete blocks with a particular relationship between the blocks, termed "modules" herein for a "modular document". Documents which are already in a streamed format, such as streaming audio or video data for example, may also be converted according to the present invention, but the particular advantage of the present invention is the ability to handle documents which are not in such a streamed format. Documents which are in a block format are more difficult to convert in a streamed manner, simply because the relationship between the blocks must be maintained during the conversion process. This necessitates maintaining a buffer in order to review previously examined blocks for enabling the relationship between the blocks to be preserved.

For example, word processing documents, which may be written in either a standard or proprietary format such as that of the Word™ software program (Microsoft Ltd., USA), may be composed of separate blocks of text and formatting instructions. If the relationship between each block of text and the corresponding block of formatting instructions is not maintained, then the visual properties of the text may be either lost or corrupted. Thus, the relationship between components of a document is important for modular file formats, such as for word processing documents, in which each module can only be fully interpreted with regard to a relationship with at least one other module.

Another example of a modular document format is the MPEG (Motion Picture Expert Group) video data format, in which each frame may optionally be considered as a module, and in which intra-frames and inter-frames may each optionally be considered to be different types of modules.

For these reasons, the present invention is also particularly useful for documents which contain data in a non-sequential order, such that the conversion process depends upon determining the actual order of the data.

Although a portion of the description below is explained with regard to WAP and a WAP-enabled device, such as a cellular telephone for example, it is understood that this is for the purposes of description only and is without any intention of being limiting. For a reference to WAP, as well as a more detailed explanation, see for example "Programming Applications with the Wireless Application Protocol" (S. Mann, Wiley Computer Publishing, John Wiley and

Sons Inc., 1999), incorporated by reference as if fully set forth herein. Furthermore, both the display device and wireless network which are described below can be viewed as examples of a low bandwidth device and network for the purposes of the present invention.

The principles and operation of a system and a method according to the present invention may be better understood with reference to the drawings and the accompanying description, it being understood that these drawings are given for illustrative purposes only and are not meant to be limiting.

Referring now to the drawings, Figure 1 is a schematic block diagram of a system according to the present invention for converting a modular document in a streamed manner.

A system **10** has a display device **12** for interacting with a user, which operates an instruction agent **14**, such as a Web browser for example. Optionally and preferably, display device **12** could be a wireless communication device **12**, which more preferably operates according to WAP. Web browsers which operate according to WAP are also referred to as "microbrowsers". Requests are sent from display device **12** through a network **18**, such as a wireless network for example. As a non-limiting example, display device **12** is optionally a cellular telephone, while network **18** is optionally a cellular telephone communication channel.

The request for a document is sent from display device **12** to a document source **20**, which serves modular documents such as word processing documents, for example. However, the documents provided by document

source **20** need to be converted to a file format which is displayable by instruction agent **14**. One example of such a file format is a WML (wireless markup language) document, or WML card, for wireless communication devices which support WAP.

In order for the modular document of document source **20** to be converted to WML cards, or to another suitable file format, system **10** also features a conversion server **26** according to the present invention.

Conversion server **26** receives at least a part of a document from document source **20**, which is preferably a modular document, and then begins to convert the modular document in a streamed manner. By "streaming", it is meant that conversion server **26** is able to begin to convert the document into the converted format as soon as a sufficient part of the document is received. This process is explained in greater detail with regard to the schematic block diagrams in Figures 2A-2C and the flowchart in Figure 3 below.

Briefly, conversion server **26** analyzes the document, and then decomposes the document into its component modules according to the type of modular file format of the document. These modules are then converted in a streamed manner which is determined by the required relationship between every two or more modules, such that conversion server **26** may optionally not begin the process of converting a first module until the corresponding second module has been read, for example. More preferably, conversion server **26** includes a plurality of specific converters (not shown), each of which handles a particular type of module for the process of conversion. The minimum required

collection of a plurality of modules which are required before a particular module can be converted is termed herein a "set of modules".

Optionally and preferably, as each set of modules is converted by conversion server **26** to a converted file format, the converted data is sent to display device **12**. Instruction agent **14** then causes display device **12** to display the message. For example, if the converted file format is a WML deck containing a WML card, then preferably instruction agent **14** is a microbrowser.

Figures 2A-2C and 3 are illustrations for the process of converting a document in a streamed manner. Figure 2A is a schematic block diagram of a modular document in the system of the present invention, while Figure 2B is an exemplary illustration of the modular document as a directed graph. Figure 2C shows the basic structure of a Microsoft Word™ file, as an example of a modular file. Figure 3 is a flowchart for a method for converting the modular document into a converted file format. The process of Figure 3 could optionally be performed "off-line", before a specific user request for the document is received, or "on the fly", after such a request has been received.

Figure 2A is a schematic block diagram of a modular document **28**, which contains a plurality of modules **30**. Each module **30** is analyzed and converted by a modular machine **32**, which includes a converter **34** and a data buffer **36**. Data buffer **36** holds any data which is required for the operation of a subsequent modular machine **32**, and is preferably identical for each modular machine **32**.

Each modular machine **32** may optionally request specific information

from one or more modular machines **32**, such as information in a specified location in modular document **28** or information which is located in another, subsequent or previous, module **30**. In addition, each modular machine **32** may then respond to one or more modular machines **32**. Modular machine **32** from which the information is requested may optionally disregard such a request, or alternatively may decide to satisfy this request immediately. Preferably, modular machine **32** balances the satisfaction of the request against the requirement for optimized performance, for example with regard to answering requests sequentially, as opposed to a more efficient but non-sequentially performed group of responses. More preferably, modular machine **32** queues the incoming requests, for example by storing the requests in data buffer **36**. Modular machine **32** may then optionally answer requests sequentially or non-sequentially.

Modular machine **32** may optionally and preferably be required to wait until the requested data is available before performing the next action in the process of conversion, although again, the requirement for waiting is more preferably balanced against optimization of the conversion process. For example, depending upon the structure of modular document **28**, if modular machine **32** requires data from two other modular machines **32**, but only receives data from one such modular machine **32**, the requesting modular machine **32** may optionally be allowed to perform any action(s) which are possible with the current data, before waiting for the response to the other request.

Modular machine **32** may optionally and more preferably determine the type of module **30** for which information is supplied. The output of each modular machine **32** is optionally a generic file format, which is then more preferably rendered into a specific file format according to the profile of user preferences and/or device capabilities. This generic output format is preferably XML. An example of a specific file format is a WML deck containing a WML card.

The flow of information and modular machines **32** may be shown, statically or dynamically, as a directed graph, as in Figure 2B. In this example, document **28** is converted with a plurality of different types of modular machines **32**. For the purposes of illustration only and without any intention of being limiting, these different types of modular machine **32** include Microsoft Word™ document modular machines **38**, Microsoft Excel™ modular machines **40** and a graphic image modular machine **42**. Within these different types of modular machines **32**, the relationship between modules, according to which the data is analyzed and converted, is also different. For example, Microsoft Word™ modules are further divided into text modules and formatting modules. By contrast, Microsoft Excel™ modules do not have such different types, but these Microsoft Excel™ modules may optionally be arranged within the file in a non-sequential order. Both Microsoft Excel™ modules and the graphic image module are placed within document **28** according to particular locations, such that these modules also have a relationship to Microsoft Word™ modules.

As an example, the structure of Microsoft Word™ modular machines **32**

may be described as follows, with regard to the main OLE stream in a Microsoft Word™ file. The main stream contains the majority of the information of a Word document. Additional streams contain summary information for a document, and embedded OLE objects within the documents. Examples of such embedded objects include Microsoft Excel™ modules and the graphic image module as described with regard to Figure 2B. It should be noted that this description relates to a **non-complex** Word™ document, which is a document saved using the full save function, as opposed to the quick save function.

As shown in Figure 2C, a first type of module in the Microsoft Word™ file is a File Information Block, which is the first part of the file. This block contains pointers to most of the structures of the file, such as the blocks which are described in greater detail below.

Next, there are one or more modules containing the actual text of the document. Text can be stored in the Unicode character set. This section contains only basic formatting information (which is specified using special characters), such as spaces and tabs; paragraph structure, as determined by the end-of-paragraph character; page breaks; basic table information, such as cell end mark, and table row end mark; and special objects in the text (such as a date, a picture, line number and so forth). These special objects in the text must also be indicated in the Format Blocks, which are described below in greater detail.

The Format Blocks contain formatting information, which describes the

properties of sections of text. Formatting information is basically stored in blocks of 512 bytes in the file. Each such block contains information about several continuous sequences of characters in the text, particularly with regard to any difference(s) from the parent Style to which these sequences belong. These blocks are divided into two types. This first type is a paragraph property block, which usually contains information such as justification, frame information, line spacing, paragraph structure and so forth.

The second type is a character property block, which usually contains information relevant to specific character blocks, such as text type (bold, italic, underlined, and so forth), size, font type and other such information.

Other optional information may include Style Sheet descriptions, Document properties and so forth, each of which is present in separate modules in the file, and are not specified in the Format Blocks.

Figure 3 is a flowchart of an exemplary method according to the present invention for converting a Word™ document into a different file type, preferably XML as previously described, based upon the above description for the structure of such a document.

In step 1, at least a part of the document is received. In step 2, the modules of the document are analyzed, in order to separate these modules into the different types, as described in greater detail above. This step is preferably performed by first retrieving the File Information Block, and then analyzing this block in order to locate the remaining modules of the document, as this block contains pointers to the remaining blocks in the file.

In step 3, preferably all of the text blocks are analyzed in order to retrieve the text of the document. As described in greater detail above, the text blocks also contain simple format information, which is specified using special characters, such as spaces and tabs; paragraph structure, as determined by the end-of-paragraph character; page breaks; and basic table information, such as cell end mark, and table row end mark. This information is sufficient to enable the text to be correctly divided into paragraphs, and to show basic information regarding tables embedded within the text by using certain assumptions, for example that the first cell of the table contains a single paragraph.

The analysis of the document may optionally end at this step, for a text only conversion, in which almost all of the formatting information for the document is disregarded. In this embodiment, sections of the text are output for conversion, after basic formatting as previously described, such that the final conversion step is the conversion of the text to the generic file format such as XML for example. For conversion to XML, the minimal text formatting information which is available is easily converted directly to XML elements.

According to a second embodiment of the method, the analysis of the file continues after the text has been extracted, in order to obtain text with advanced formatting but without using Style information. In this embodiment, it is assumed that the Styles in the document are not changed from their default values. Therefore, each formatting information block is examined.

For the second embodiment, in step 4, the text section is stored rather than being converted. In step 5, each formatting information block is

examined. Again, each such block can be located from the File Information Block as previously described. As each formatting block is located for a particular text block, the changes specified in the formatting block are then applied to the relevant sections of text, based on the known default Style information, in step 6. In step 7, each formatted text section is output, such that steps 5-7 are optionally repeated at least once, and more preferably are repeated until the document has been fully analyzed. Again, the output sections are sent to the final conversion step, which again is the conversion of the text to the generic file format such as XML for example, and is similar to the previously described final conversion step, except that additional elements need to be added to incorporate the additional format information.

According to yet another embodiment of this method, the analysis of the file preferably continues, in order to produce converted text with full formatting, by using Style information. This embodiment may optionally be preferred if the modular machines support non-sequential data transference, which is supplying data from a specific location in the file, rather than converting only according to linear order. The Style Sheet information is then preferably requested in advance, based on its location which is stated in the File Information Block. Alternatively, such an embodiment may be supported for a full conversion, without regard to streaming considerations, for example for "off line" conversions.

According to this embodiment, the Style Sheet information is read before the text itself. Now, in step 8, changes are applied to the text as previously

described from the Style Sheet information, as for the other formatting information. Again, this embodiment ends with the final conversion step, which again is the conversion of the text to the generic file format such as XML for example, as previously described, except that further additional elements need to be added to incorporate the additional format information.

It should be noted that although the above description centers around visual data, the present invention is also applicable to audio data with at least one audio attribute. For example, an MP3 (MPEG layer 3) file includes stereo data, which is actually two mono channels or modules of data. The two mono channels can optionally be combined to a single mono channel, according to the relationship between these two channels, in order to form the converted file format data.

It will be appreciated that the above descriptions are intended only to serve as examples, and that many other embodiments are possible within the spirit and the scope of the present invention.

WHAT IS CLAIMED IS:

1. A method for converting at least part of a modular document into a converted file format for display to a user, the method comprising the steps of:

(a) analyzing at least a part of the modular document to locate a plurality of modules in at least a part of the modular document;

(b) separating said analyzed document into said plurality of modules;

(c) determining a relationship between at least a pair of modules; and

(d) converting at least said pair of modules according to said relationship to form the converted file format.

2. The method of claim 1, wherein at least said pair of modules are not arranged in a linear sequence in the modular document, such that at least a portion of at least one module is stored in step (d) for converting at least a second module.

3. The method of claim 2, wherein said relationship is such that said portion provides information for converting at least said second module.

4. The method of claim 3, wherein said information is at least one visual attribute of display data obtained from at least said second module.

5. The method of claim 4, wherein said display data is text, and said at least one visual attribute is a format of said text.

6. The method of claim 3, wherein said information is at least one audio attribute of display data obtained from at least said second module.

7. The method of claim 1, wherein at least said pair of modules are not arranged in a linear sequence in the modular document, such that step (d) further comprises the steps of:

(i) storing at least a first module before converting at least said first module;

(ii) analyzing at least a second module; and

(iii) converting at least said first module according to information obtained from at least said second module.

8. The method of claim 7, wherein an order for performing steps (i)-(iii) is determined according to an optimal order for efficient conversion.

9. The method of claim 1, wherein the modular document is a word processing document, such that at least one module of the plurality of modules contains text and at least one module of the plurality of modules contains information for determining a format of said text.

10. The method of claim 1, further comprising the steps of:

(e) providing a display device; and

(f) displaying the converted file format on said display device.

11. The method of claim 10, wherein said display device is a display device, and said network is a wireless network.

12. The method of claim 11, wherein said display device is a cellular telephone, and said wireless network is a cellular telephone network.

13. The method of claim 10, wherein the converted file format is a mark-up language format.

14. The method of claim 13, wherein said mark-up language is WML (Wireless Mark-up Language), such that said display device is a WAP (wireless application protocol) enabled device.

15. The method of claim 10, wherein said display device is a low bandwidth communication device.

16. The method of claim 1, wherein steps (a) - (d) are performed in advance, before the user requests the document.

17. The method of claim 1, wherein step (d) comprises the step of converting the modular document to a first generic file format, the method further comprising the step of:

(e) converting the modular document from said first generic file format to a specific file format.

18. The method of claim 17, wherein said specific file format is determined according to at least one preference of the user.

19. The method of claim 17, further comprising the steps of:

(f) providing a display device;

(g) determining said specific file format according to at least one characteristic of said display device; and

(h) displaying the modular document in said specific file format on said display device.

20. The method of claim 1, wherein the converted file format data is output as soon as the conversion is performed, such that step (d) includes the step of transmitting the converted file format data in a streamed manner.

21. A system for converting a modular document to a converted file format for display to a user, the modular document featuring a plurality of

modules having a relationship between at least a pair of modules, the system comprising:

(a) a document source for serving the modular document; and

(b) a conversion server for receiving the modular document and for converting at least part of the modular document into the converted file format according to the relationship between at least the pair of modules.

22. The system of claim 21, further comprising:

(c) a display device for displaying a converted part of the modular document to the user; and

(d) a network for connecting said display device to said conversion server.

23. A method for converting at least part of a document into a converted file format for display to a user, the document containing data in a non-sequential order, the method comprising the steps of:

(a) analyzing at least a part of the document to form an analyzed document;

(b) determining an order for the data in at least a part of the document; and

(c) converting at least said part of the document according to said order for the data to form the converted file format.

Figure 1

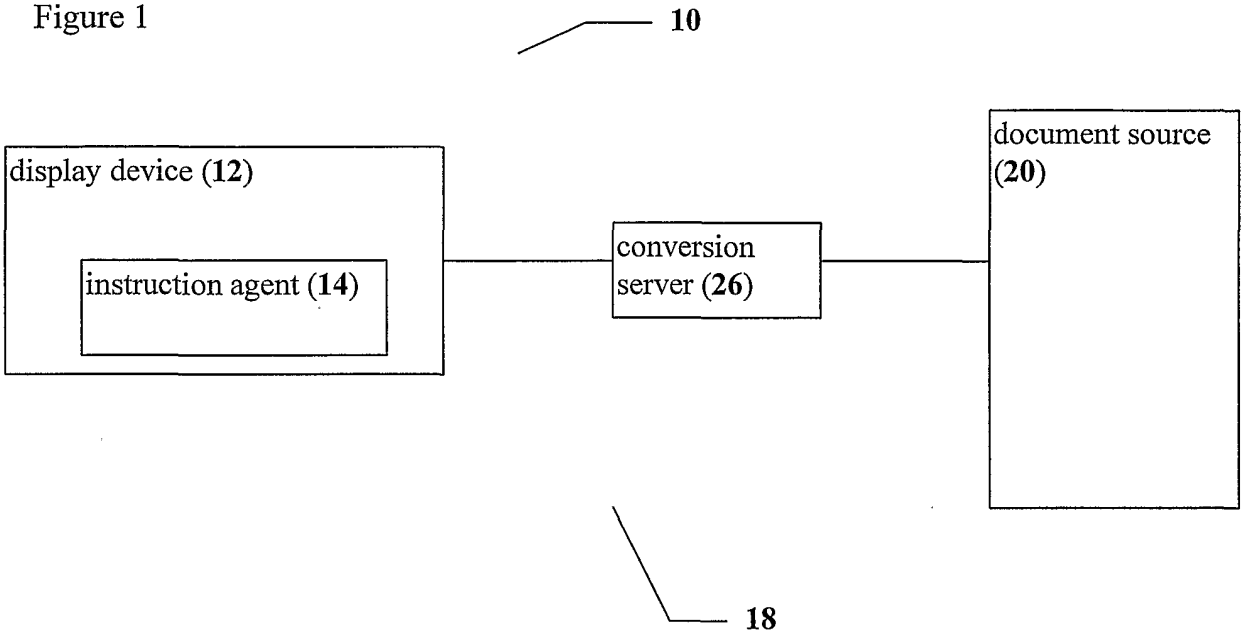

10
display device (12)
instruction agent (14)
conversion server (26)
document source (20)
18

Figure 2A

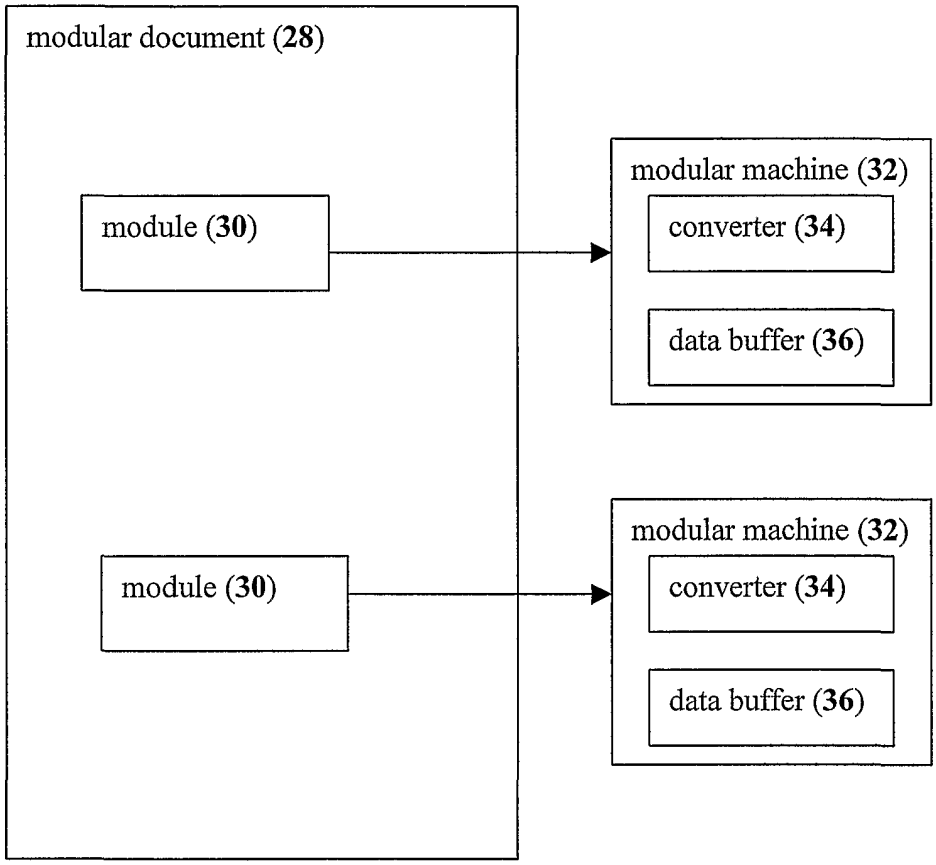

modular document (28)
module (30)
module (30)
modular machine (32)
converter (34)
data buffer (36)
modular machine (32)
converter (34)
data buffer (36)

Figure 2B

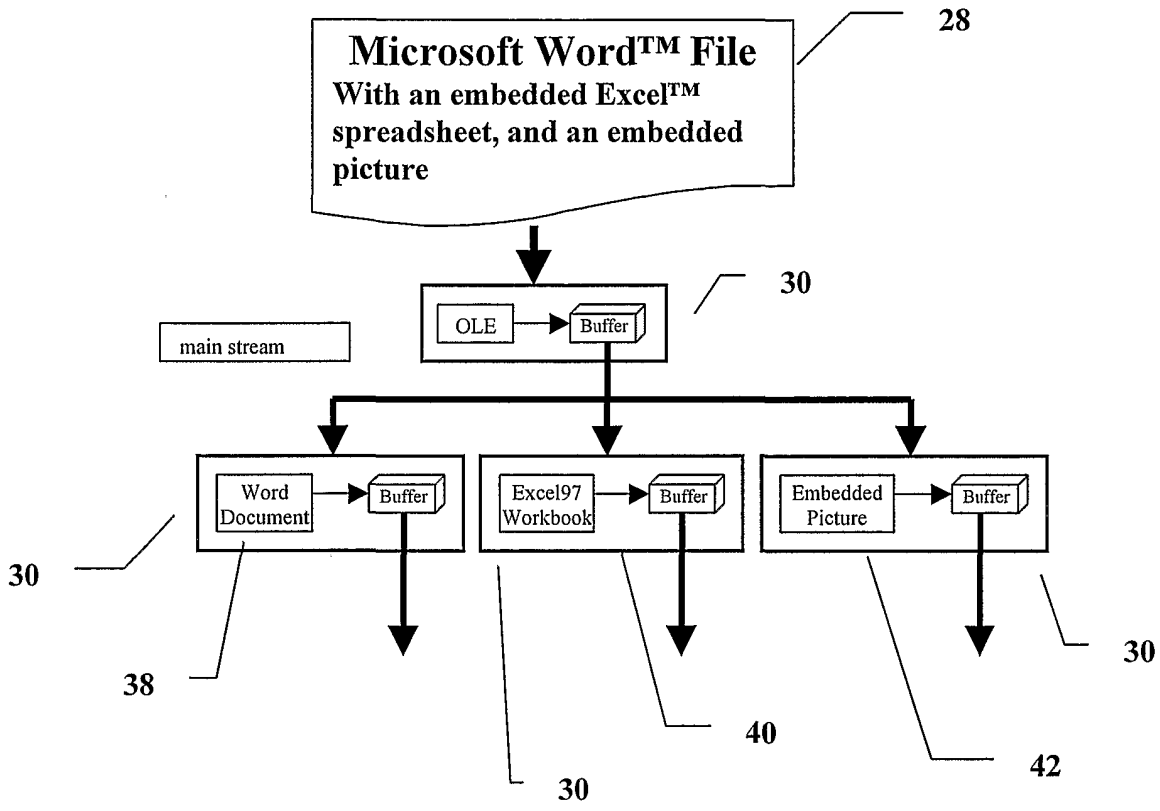

Microsoft Word™ File
With an embedded Excel™ spreadsheet, and an embedded picture
28
30
main stream
OLE
Buffer
Word Document
Buffer
Excel97 Workbook
Buffer
Embedded Picture
Buffer
30
38
40
42
30
30

Figure 2C

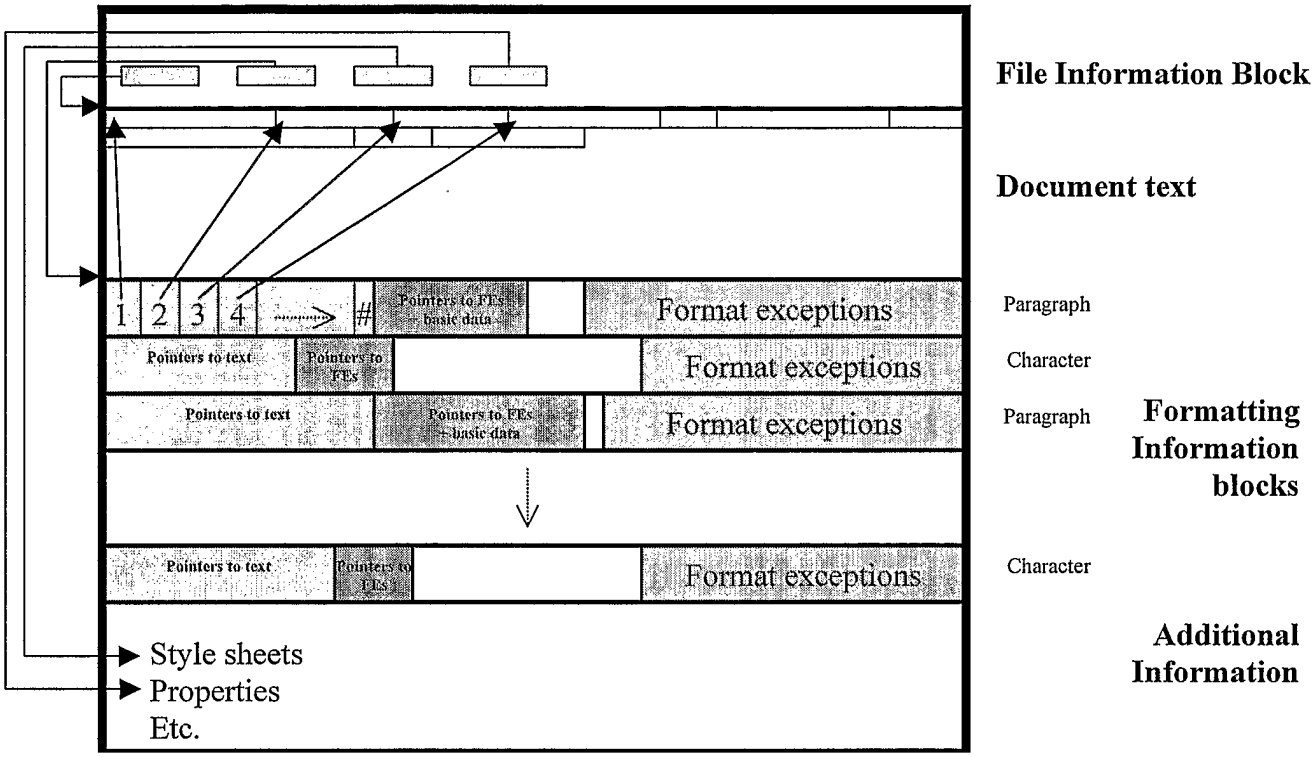
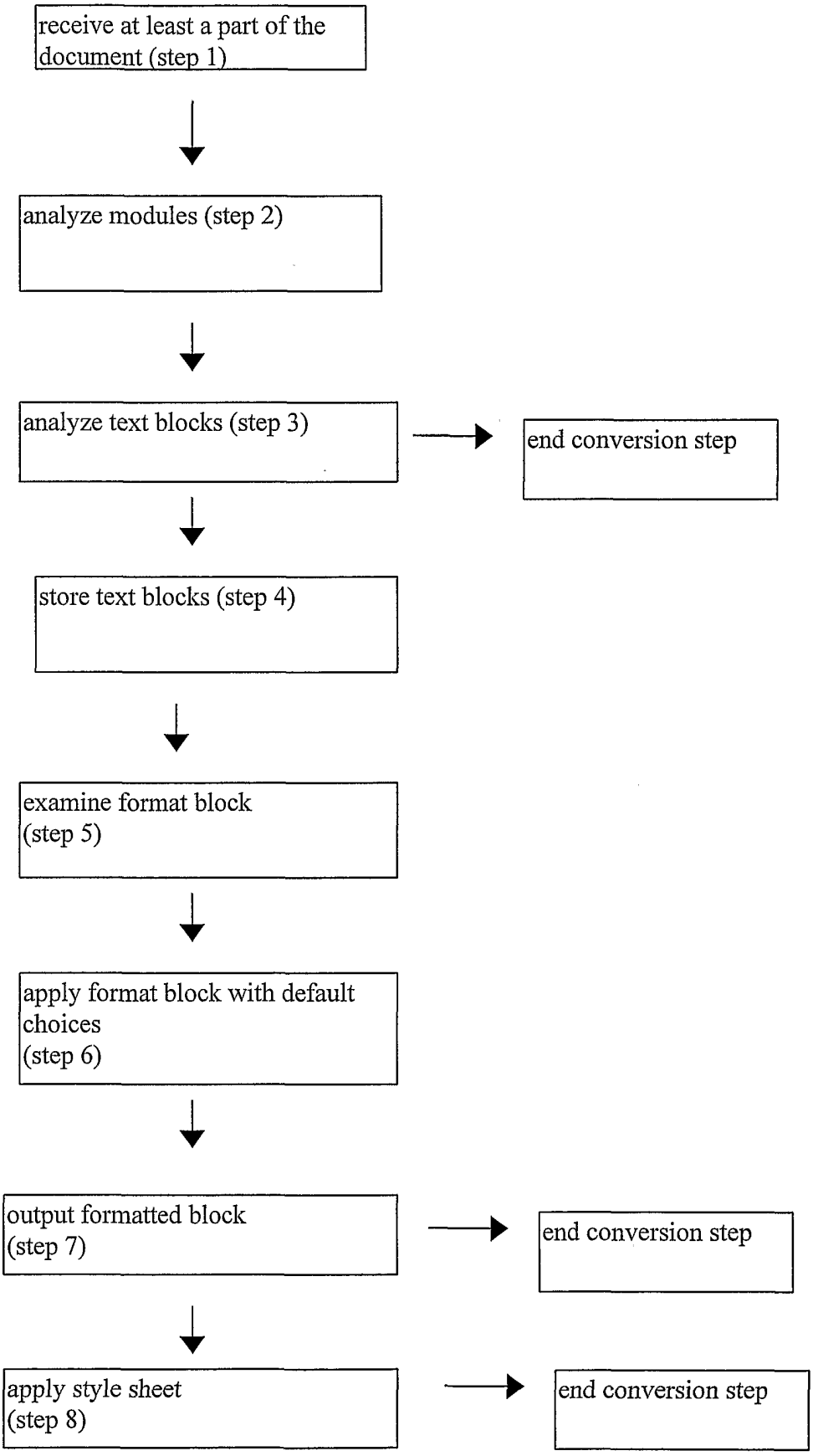

File Information Block
Document text
1
2
3
4
#
Pointers to FEs - basic data
Format exceptions
Paragraph
Pointers to text
Pointers to FEs
Format exceptions
Character
Pointers to text
Pointers to FEs - basic data
Format exceptions
Paragraph
Formatting Information blocks
Pointers to text
Pointers to FEs
Format exceptions
Character
Style sheets
Properties
Etc.
Additional Information

Figure 3


receive at least a part of the document (step 1)
analyze modules (step 2)
analyze text blocks (step 3)
end conversion step
store text blocks (step 4)
examine format block (step 5)
apply format block with default choices (step 6)
output formatted block (step 7)
end conversion step
apply style sheet (step 8)
end conversion step

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL01/00190

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G06F 7/00, 15/00
US CL :707/3, 517

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)

U.S. : 707/3, 517

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)

ACM
search terms: PDA, WML, HTML conversion, WAP translation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,895,476 A (ORR et al) 20 April 1999, whole document.	1-23
Y	US 5,907,837 A (FERREL et al) 25 May 1999, whole document.	1-23
A	FOX. A. Adapting to network and client variability via on-demand dynamic distillation ACM Architectural Support for Programming Languages and Operating Systems. October 1996. pages 160-170, whole document.	1-23

☐ 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 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	Date of mailing of the international search report
06 JUNE 2001	28 JUN 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Peggy Harrod HEATHER HERNDON Telephone No. (703) 305-3900