

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 August 2001 (23.08.2001)

PCT

(10) International Publication Number
WO 01/61517 A1

(51) International Patent Classification⁷: G06F 15/00,
17/00, 17/21, 17/24, H04N 1/00

(21) International Application Number: PCT/US01/04558

(22) International Filing Date: 14 February 2001 (14.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/182,674 15 February 2000 (15.02.2000) US
09/597,297 20 June 2000 (20.06.2000) US

(71) Applicant: FIRST USA BANK, N.A. [US/US]; Three
Christina Centre, 201 North Walnut Street, Wilmington,
DE 19801 (US).

(72) Inventors: BUISMAN, William, J.; 2252 Canoe Branch
Road, Lebanon, TN 37087 (US). COHEN, Gary; 780

Fawnhill Road, Broomall, PA 19008 (US). FOX, Steven,
R.; 200 Mount Vernon Avenue, Laurel Springs, NJ 08021
(US). KRUSE, Charles; 411 Taylor Street, Bear, DE
19701 (US). SULPZIO, Christel, K.; 21 John Beal Drive,
Boothwyn, PA 19061 (US).

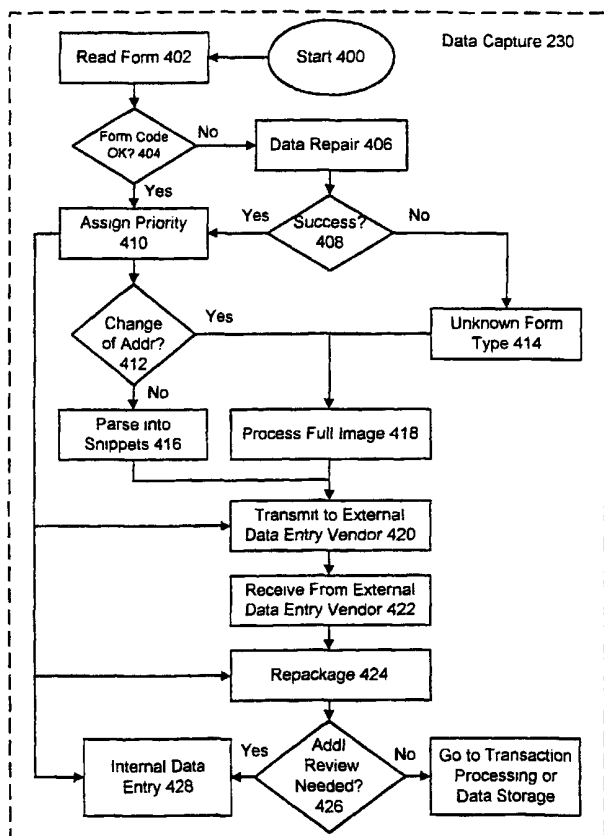
(74) Agents: SCOTT, Thomas, J. et al.; Hunton & Williams,
1900 K Street, N.W., Washington, DC 20006 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, 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, 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,

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR CONVERTING INFORMATION ON PAPER FORMS TO ELECTRONIC DATA



(57) Abstract: The invention is a system and method for converting paper forms to electronic data. More specifically, a system and technique are described for performing the conversion in a manner that is advantageous in terms of overall cost and cycle time, and is applicable to a variety of form types, such as credit card applications, license applications, and tax returns. Data capturing comprises reading a form (402) to determine if form is valid (404). A priority (410) is then assigned to said form. The form is either parsed into snippets (416), transmitted to an external data entry vendor (420), repackaged (424), and/or assigned to an internal data entry (428).



WO 01/61517 A1



IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

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.

Published:

— *with international search report*

**SYSTEM AND METHOD FOR CONVERTING INFORMATION ON PAPER
FORMS TO ELECTRONIC DATA**

FIELD OF THE INVENTION

5 The invention relates generally to a system and method for processing data, and more specifically for converting information on paper forms to electronic data which can be utilized by subsequent processes.

BACKGROUND OF THE INVENTION

10 Many commercial and government entities receive input data in paper format. Credit card applications, license applications, and tax returns are examples. Although input techniques such as automated telephone systems and Web-based utilities are increasingly provided as an input alternative, the volume of input made in paper format is still substantial. For instance, a bank or other financial institution may receive more than
15 10,000 credit card applications in paper format in a single day. Manual entry of this volume of data into data processing systems may be cost prohibitive.

 A common technique for processing such a large volume of input forms involves automated or semi-automated conversion of the completed paper forms to electronic format. In a typical approach, paper forms are first scanned by a digital image scanner to
20 yield an electronic bitmap image. The image is then converted to text via Optical Character Recognition (OCR) software (for reading machine printed characters) and Intelligent Character Recognition (ICR) software (for reading hand written characters). Not all data is correctly interpreted, however, due to functional limitations in scanning apparatus and recognition software. It is therefore common practice to employ data entry
25 operators for the purpose of correcting errors or omissions resulting from the automated conversion process. For an example of such a system and method, see U.S. Pat No. 5,054,096 issued to Beizer on Oct. 1, 1991.

 Known approaches have several drawbacks and limitations, however. One issue not addressed by existing technology is how to manage the cost and cycle time associated
30 with data entry staff. Another limitation is a failure to recognize that it may be

- 2 -

advantages to process some applications differently than others, according to a variety of factors.

In sum, existing systems and techniques for converting information on paper forms to electronic data have not adequately managed the conversion process. The resulting lack of efficiency, and other drawbacks, limit the utility of such systems for entities receiving a high volume of input data in paper format.

SUMMARY OF THE INVENTION

The invention overcoming these and other drawbacks in the art relates a system and method for converting credit card applications and other forms to electronic format. The system and method may operate on full forms or only portions (or snippets) of the forms, which are later reassembled. In one embodiment, conversion processing may be accomplished through the selective use of internal and external data entry operators. Work may also be prioritized according to its importance to the processing entity.

It is an object of the invention to reduce the cost and cycle time associated with the conversion process from paper format to electronic data format.

It is another object of the invention to enable an entity using the conversion process to tailor the work flow to the needs of their organization.

The following drawings and descriptions further describe the invention, including different embodiments of the major system components and processes. The construction of such a system, implementation of such a process, and advantages will be clear to a person skilled in the art of document conversion.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is schematic diagram of a system configured for the conversion of paper forms into electronic data, according to one embodiment of the invention.

Figure 2 is a high-level flow diagram for forms-based processing, according to one embodiment of the invention.

Figure 3 is a more detailed flow diagram of form preparation, according to one embodiment of the invention.

- 3 -

Figure 4 is a more detailed flow diagram of data capture, according to one embodiment of the invention.

Figure 5 is a schematic diagram illustrating how a form is parsed into snippets, according to one embodiment of the invention.

5

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is schematic diagram of the system, according to one embodiment of the invention. The diagram illustrates an overall architecture for conversion processing, including: paper forms 100, 102, and 114; a facsimile machine 112; image scanners 104, 106, and 116; servers 128, 132, 134, 136, and 138; database 130; clients 108, 110, 118, 120, 122, 124, and 126; and communication links 140, 141, 142, and 143.

Paper forms 100 and 102 may be received via an intake service, such as a mail delivery service. Paper form 114 is the output from facsimile machine 112. Paper forms 100, 102, and 114 may be imaged by digital image scanners 104, 106, and 116 or other image generator. Optical Character Recognition (OCR) and Intelligent Character Recognition (ICR) software, which may be resident on imaging devices 104, 106, and 116 or on input clients 108, 110, and 118, may convert the images into alpha-numeric text. The system may include internal data entry clients 120 and 122, and may also include external data entry clients 124 and 126.

Clients 108, 110, 118, 120, 122, 124, and 126 may be or include, for instance, a personal computer running the Microsoft Windows™ 95, 98, Millenium™, NT™, or 2000, Windows™CE™, PalmOS™, Unix, Linux, Solaris™, OS/2™, BeOS™, MacOS™ or other operating system or platform. Clients 108, 110, 118, 120, 122, 124, and 126 may include a microprocessor such as an Intel x86-based device, a Motorola 68K or PowerPC™ device, a MIPS, Hewlett-Packard Precision™, or Digital Equipment Corp. Alpha™ RISC processor, a microcontroller or other general or special purpose device operating under programmed control. Clients 108, 110, 118, 120, 122, 124, and 126 may furthermore include electronic memory such as RAM (random access memory) or EPROM (electronically programmable read only memory), storage such as a hard drive, CDROM or rewritable CDROM or other magnetic, optical or other media, and other associated components connected over an electronic bus, as will be appreciated by

30

persons skilled in the art. Clients 108, 110, 118, 120, 122, 124, and 126 may also be or include a network-enabled appliance such as a WebTV™ unit, radio-enabled Palm™ Pilot or similar unit, a set-top box, a networkable game-playing console such as Sony Playstation™ or Sega Dreamcast™, a browser-equipped cellular telephone, or other TCP/IP client or other device.

Server 128 may control access to database 130, and may also control work flow between servers 132, 134, 136, and 138. Servers 128, 132, 134, 136, and 138 may be or include, for instance, a workstation running the Microsoft Windows™ NT™, Windows™ 2000, Unix, Linux, Xenix, IBM AIX™, Hewlett-Packard UX™, Novell Netware™, Sun Microsystems Solaris™, OS/2™, BeOS™, Mach, Apache, OpenStep™ or other operating system or platform.

Clients 108, 110, 118, 120, 122, 124, and 126, and servers 128, 132, 134, 136, and 138 may utilize network enabled code to exchange data or instructions over communications links 140, 141, 142, and 143. The network enabled code may be, include or interface to, for example, Hyper text Markup Language (HTML), Dynamic HTML, Extensible Markup Language (XML), Extensible Stylesheet Language (XSL), Document Style Semantics and Specification Language (DSSSL), Cascading Style Sheets (CSS), Synchronized Multimedia Integration Language (SMIL), Java™, Jini™, C, C++, Perl, UNIX Shell, Visual Basic or Visual Basic Script, Virtual Reality Markup Language (VRML) or other compilers, assemblers, interpreters or other computer languages or platforms.

Communications links 140, 141, 142, and 143 may be, include or interface to any one or more of, for instance, the Internet, an intranet, a PAN (Personal Area Network), a LAN (Local Area Network), a WAN (Wide Area Network) or a MAN (Metropolitan Area Network), a storage area network (SAN), a frame relay connection, an Advanced Intelligent Network (AIN) connection, a synchronous optical network (SONET) connection, a digital T1, T3, E1 or E3 line, Digital Data Service (DDS) connection, DSL (Digital Subscriber Line) connection, an Ethernet connection, an ISDN (Integrated Services Digital Network) line, a dial-up port such as a V.90, V.34 or V.34bis analog modem connection, a cable modem, an ATM (Asynchronous Transfer Mode) connection, or an FDDI (Fiber Distributed Data Interface) or CDDI (Copper Distributed Data

- 5 -

Interface) connection. Communications links 140, 141, 142, and 143 may furthermore be, include or interface to any one or more of a WAP (Wireless Application Protocol) link, a GPRS (General Packet Radio Service) link, a GSM (Global System for Mobile Communication) link, a CDMA (Code Division Multiple Access) or TDMA (Time Division Multiple Access) link such as a cellular phone channel, a GPS (Global Positioning System) link, CDPD (cellular digital packet data), a RIM (Research in Motion, Limited) duplex paging type device, a Bluetooth radio link, or an IEEE 802.11-based radio frequency link. Communications links 140, 141, 142, and 143 may yet further be, include or interface to any one or more of an RS-232 serial connection, an IEEE-1394 (Firewire) connection, a Fibre Channel connection, an IrDA (infrared) port, a SCSI (Small Computer Systems Interface) connection, a USB (Universal Serial Bus) connection or other wired or wireless, digital or analog interface or connection.

The database 130 may be, include or interface to, for example, the OracleTM relational database sold commercially by Oracle Corp. Other databases, such as InformixTM, DB2 (Database 2) or other data storage or query formats, platforms or resources such as OLAP (On Line Analytical Processing), SQL (Standard Query Language), a storage area network (SAN), Microsoft AccessTM or others may also be used, incorporated or accessed in the invention.

Figure 2 is a high-level flow diagram for processing of financial applications or other forms, according to one embodiment of the invention. The illustrated process begins with form preparation 200, which is further described by Figure 3. Form distribution 210 may involve, for example, delivery of forms to existing or potential customers. Form distribution 210 may be accomplished by direct mail, by placement on countertop or rack, or by other suitable procedure. Form completion 220 may be satisfied with handwritten data entry onto the form, by using a pen or pencil, for example. Form completion 220 may also involve machine-aided data entry such as use of a typewriter or electromechanical printer. The data capture process 230 converts data added to the form into electronic format, and is further illustrated in Figure 4. The data capture process 230 may be internal to an organization, and may be supplemented with external data entry 240. The type and frequency of errors related to data capture process 230 may be continuously monitored or periodically audited by a quality control function

250. One use of the data capture process 230 may be to enable transaction processing 260, such as automated review of credit card applications. The output of data capture process 230 may also be exploited by a data storage process 270 to create a useful database.

5 Figure 3 is a more detailed flow diagram of form preparation 200, according to one embodiment of the invention. The diagram illustrates that a new form is created 310 only after determination of a need 300. The form creation process 310 may receive an assigned identification number 320 as an input. Identification number 320 may appear on the resulting form as printed text, or it may be encoded in machine-readable format
10 such as a barcode. The form creation process 310 may be further constrained by certain content and format requirements 330. A content requirement might be, for example, that all forms include the name and address of the form provider and request the name and address of the person completing the form. Format restrictions may include requirements such as font type, minimum character size, and line spacing. Before the form preparation
15 step 200 is considered complete, an entity may require that a sample form be tested for readability 340, for example with existing scanner hardware and character recognition software.

 Figure 4 is a more detailed flow diagram of data capture process 230, according to one embodiment of the invention. The process starts 400 with receipt of a completed
20 form. The first step is to scan or otherwise read the form 402, converting the form and all data contained thereon into electronic format. This may be accomplished via digital image scanners 104, bar code readers, Optical Character Recognition (OCR) software, Intelligent Character Recognition (ICR) software, or through similar reading techniques known to data conversion practitioners.

25 Processing may be contingent on the outcome 404 of the read 402. A form code identifies the particular type of form. If the form code was successfully interpreted, then a priority may be assigned to the form 410. Assignment of priority may be based, for example, on a preferred client list, according to the profit margin that a seller of goods or services expects to receive from a buyer, or according to other criteria. The priority
30 assigned in step 410 may be represented on a scale of 1 to 10, designated by high, medium, or low, or otherwise rank ordered. If the form code read was not successful in

- 7 -

step 402, then the form may be routed to data repair 406 for manual classification of the incoming form type.

Work flow may be further contingent on whether data repair 406 was successful 408. If data repair 406 was a success, for example where the form code was human-readable, then the form is promoted to step 410 for assignment of priority. If, on the other hand, data repair 406 was not successful in determining the form code, then the form may be designated as an unknown form type 414.

After forms have been assigned a priority 410, they may be reviewed for change of address 412. Change of address may be detected by a box that has been checked, for example, or by the presence of text outside of defined data input areas. Where a change of address has not been detected, the form may be routed by decision process 412 to parsing process 416. An unknown form 414, or a form with a change of address 412, may be processed as a full image 418.

Parsing 416 decomposes form data into snippets. Figure 5 illustrates a form 501 that has been parsed into a Name Snippet 502, an Address Snippet 503, and a Social Security Number Snippet 504. Subsequent processing of parsed data is advantageous because a majority of snippet types are common to different form types. The result is that new form types are easily introduced into the semi-automated data capture process 230. Another advantage of snippet processing is information security. For instance, a data entry operator verifying data originally contained on form 501 is not able to associate the social security number with a name: the data entry operator is merely processing a series of unassociated name and social security number snippets.

The daily volume of forms received 400 for data capture 230 may vary substantially. For this and other reasons, it may be advantageous to utilize external data entry vendors 240 capable of processing either parsed 416 or full image forms 418. Figure 4 therefore indicates that electronic form data may be transmitted 420 to external data entry vendors for the purpose of verifying the automated read 402. External data entry 240 may involve, for example, an on-screen comparison between a bit mapped image of a snippet or full image and the textual equivalent produced by character recognition software. After external data entry 240, the system may receive 422 snippets or full images from the external data entry vendor. In the case of snippet processing by

- 8 -

external data entry vendors 240, the data is repackaged 424 so that snippets are re-associated according to original format received in step 400.

When form data is received from external data entry vendors 422, and repackaged 424 if necessary, a decision 426 may be made as to the need for review by internal data entry operators 428. Such a review may be appropriate, for example, where data is missing or could not be discerned by operators employed by the external data entry operator. If no review is needed, the verified electronic forms may be sent to transaction processing 260 or data storage 270.

Finally, Figure 4 also illustrates that the order in which form data is operated on by the transmit 420, repackage 424, and internal data entry 428 processes may be dictated by the priority assigned in step 410. Processing in step 428 may also be prioritized or sorted according to the type of errors that could not be resolved by external data entry vendors in step 240. For instance, all errors resulting from missing data might be processed in step 428 only by internal data entry operators who are trained and equipped to reach applicants by telephone.

The specification and examples provided above should be considered exemplary only. It is contemplated that the appended claims will cover any other such embodiments or modifications as fall within the true scope of the invention.

What is claimed is:

1. A system for converting forms to electronic format, comprising:
 - an interface to at least one intake service for receiving forms;
 - at least one image generator, communicating with the intake service, to convert the
 - 5 forms into electronic format;
 - at least one processor for executing related processes and providing a contingent workflow;
 - at least one terminal, communicating with at least one processor, operable to edit form data;
 - 10 an interface to at least one external data entry vendor; and
 - an interface to at least one subsequent process that will utilize data on the electronic form.
2. The system of claim 1, wherein the intake service comprises a mail delivery service.
3. The system of claim 1, wherein the intake service comprises the output of a facsimile
- 15 machine.
4. The system of claim 1, wherein the image generator comprises optical character recognition software for reading machine printed text.
5. The system of claim 1, wherein the image generator comprises intelligent character recognition software for reading handwritten text.
- 20 6. The system of claim 1, wherein the processor alters workflow based at least on the ability to read the form type.
7. The system of claim 1, wherein the processor alters workflow based at least on the presence of a change of address.
8. The system of claim 1, wherein the processor alters workflow based at least on the
- 25 priority of the form to a using entity.
9. The system of claim 1, wherein the processor alters workflow based at least on errors received from external data entry operators.
10. The system of claim 1, wherein the subsequent process utilizing data on the electronic form comprises a transaction.

11. The system of claim 10, wherein the transaction comprises review of credit card applications.

12. The system of claim 1, wherein the subsequent process utilizing data on the electronic form comprises construction of a database.

· 5 13. A method for converting forms to electronic format, comprising:

(a) receiving forms;

(b) reading the forms into electronic format;

(c) processing the forms according to a contingent workflow; and

(d) making the form data available to a subsequent process.

· 10 14. The method of claim 13, wherein step (a) of receiving comprises receipt from a mail delivery service.

15. The method of claim 13, wherein step (a) of receiving comprises receipt of a form from a facsimile machine.

16. The method of claim 13, wherein step (b) of reading comprises image capture

· 15 17. The method of claim 13, wherein step (b) of reading comprises application of optical character recognition algorithms.

18. The method of claim 13, wherein step (b) of reading comprises application of intelligent character recognition algorithms.

· 20 19. The method of claim 13, wherein step (c) of processing comprises a workflow contingent on the ability to identify the form type.

20. The method of claim 13, wherein step (c) of processing comprises a workflow contingent on the presence of a change of address.

21. The method of claim 13, wherein step (c) of processing comprises a workflow contingent on the priority of the form.

· 25 22. The method of claim 13, wherein step (c) of processing comprises a workflow contingent on the type of error received from external data entry operators.

23. The method of claim 13, wherein step (d) of making comprises a process that writes data to another location.

· 30 24. The method of claim 13, wherein step (d) of making comprises a process that allows data to be read from another location.

25. The method of claim 13, wherein step (d) of making comprises data sharing with a transaction.

26. The method of claim 25, wherein the transaction is review of credit card applications.

• 5 27. The method of claim 13, wherein step (d) of making comprises data sharing with a database.

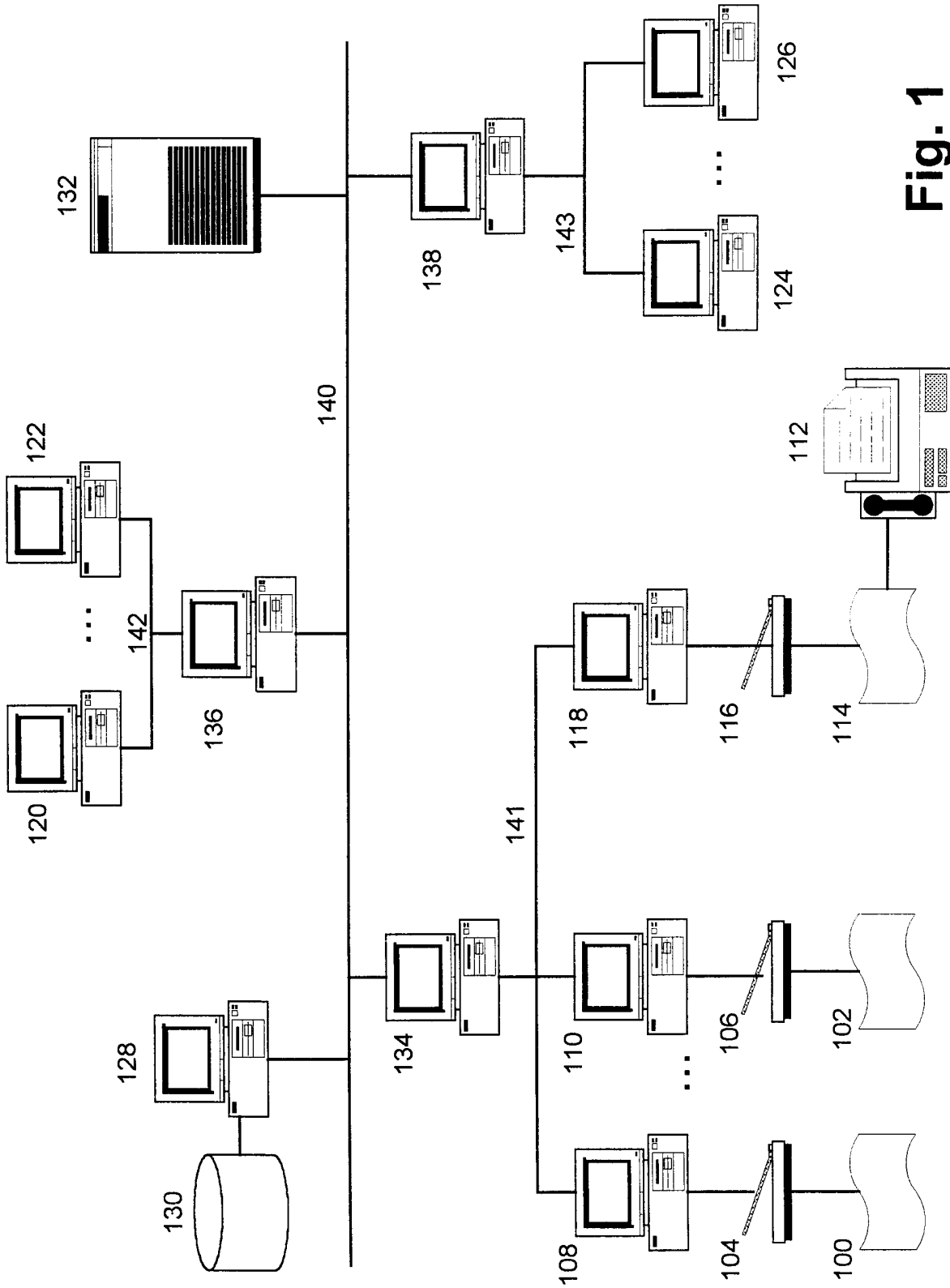


Fig. 1



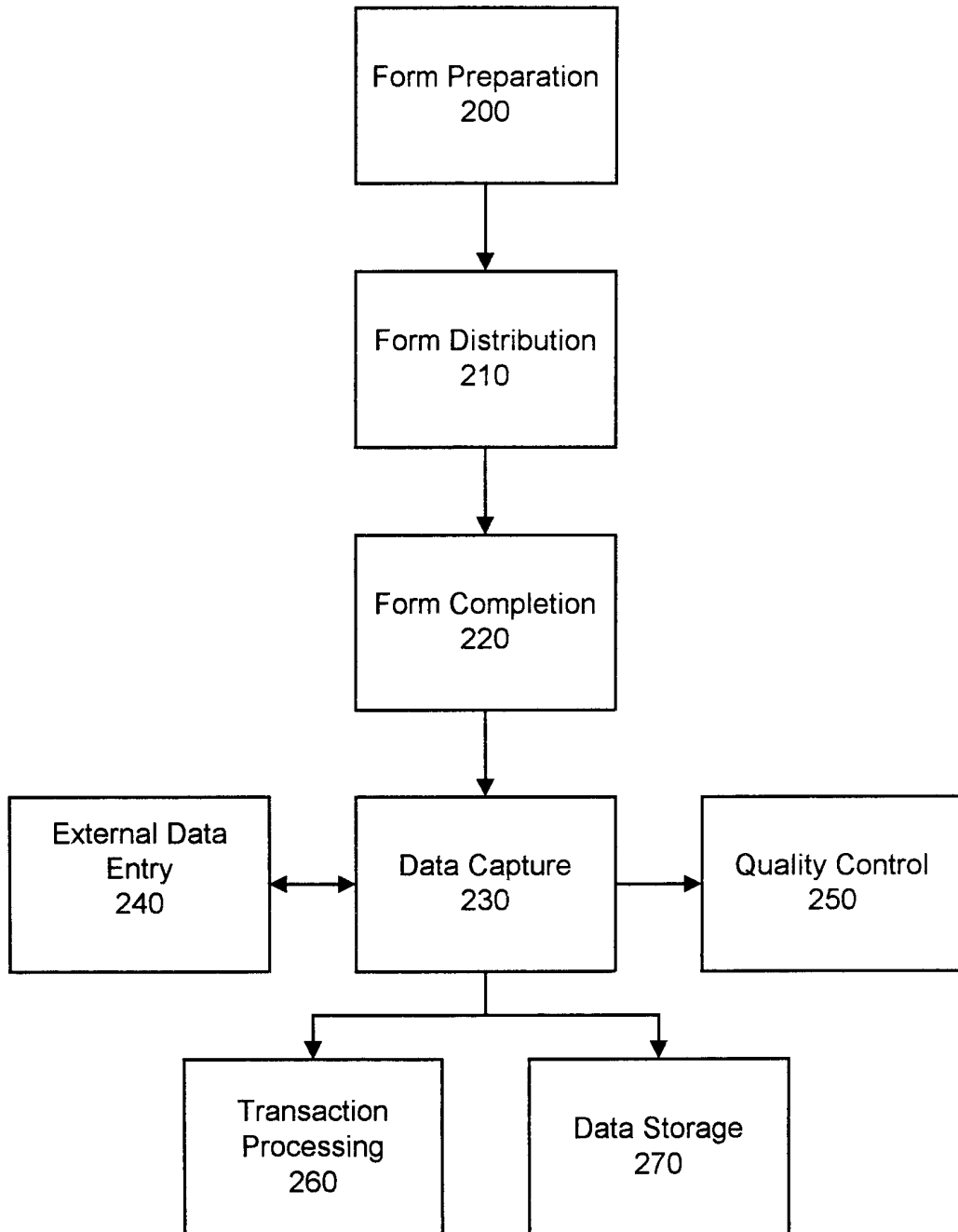


Fig. 2



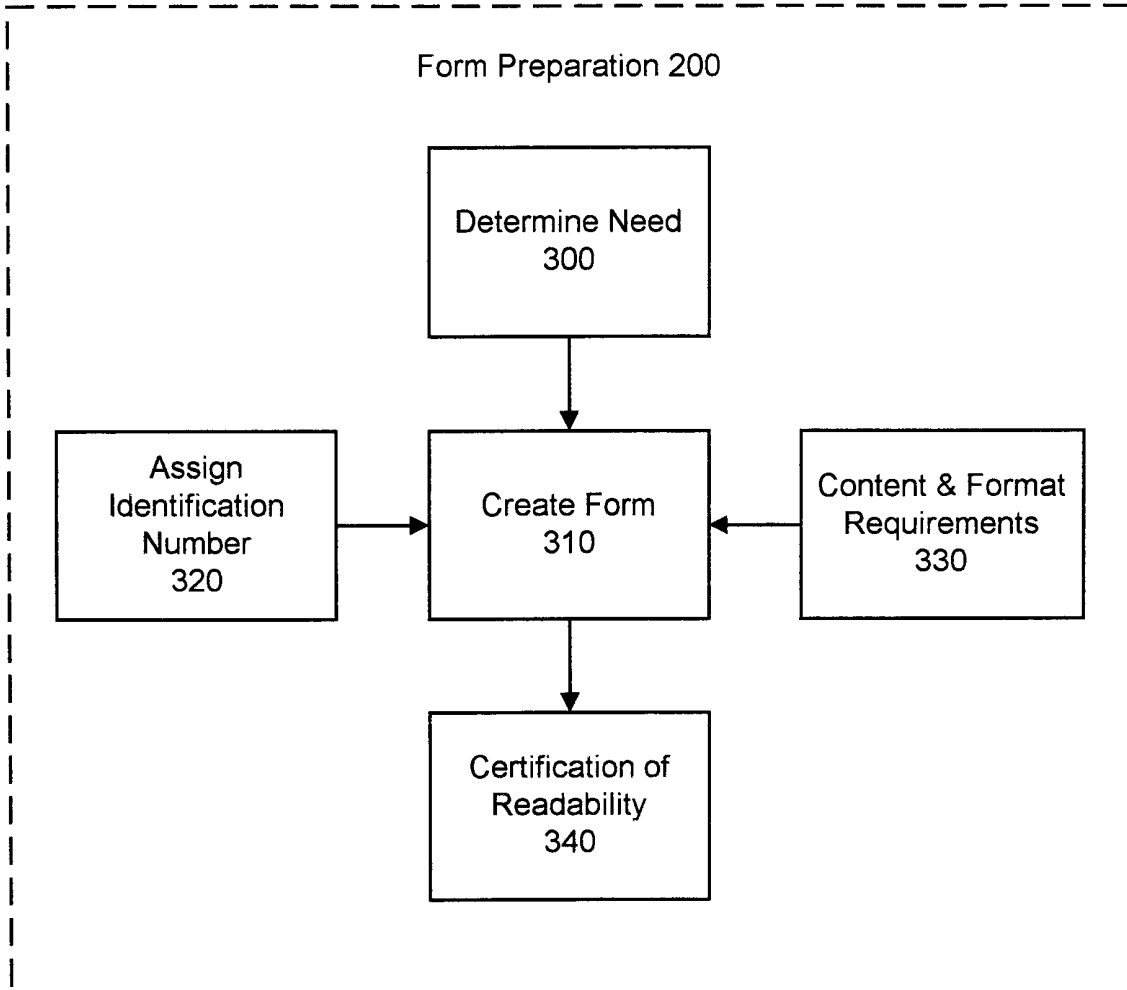


Fig. 3



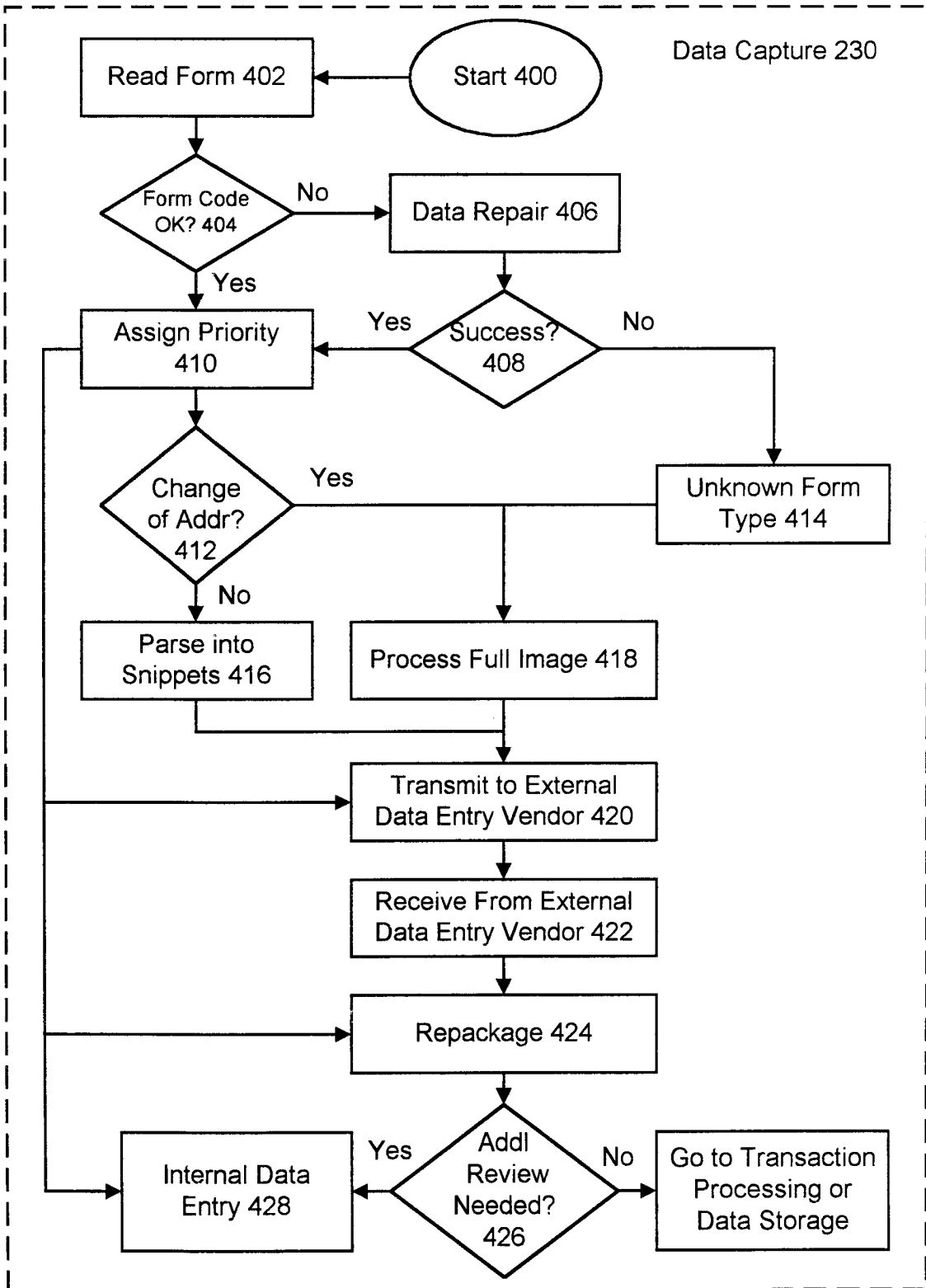


Fig. 4

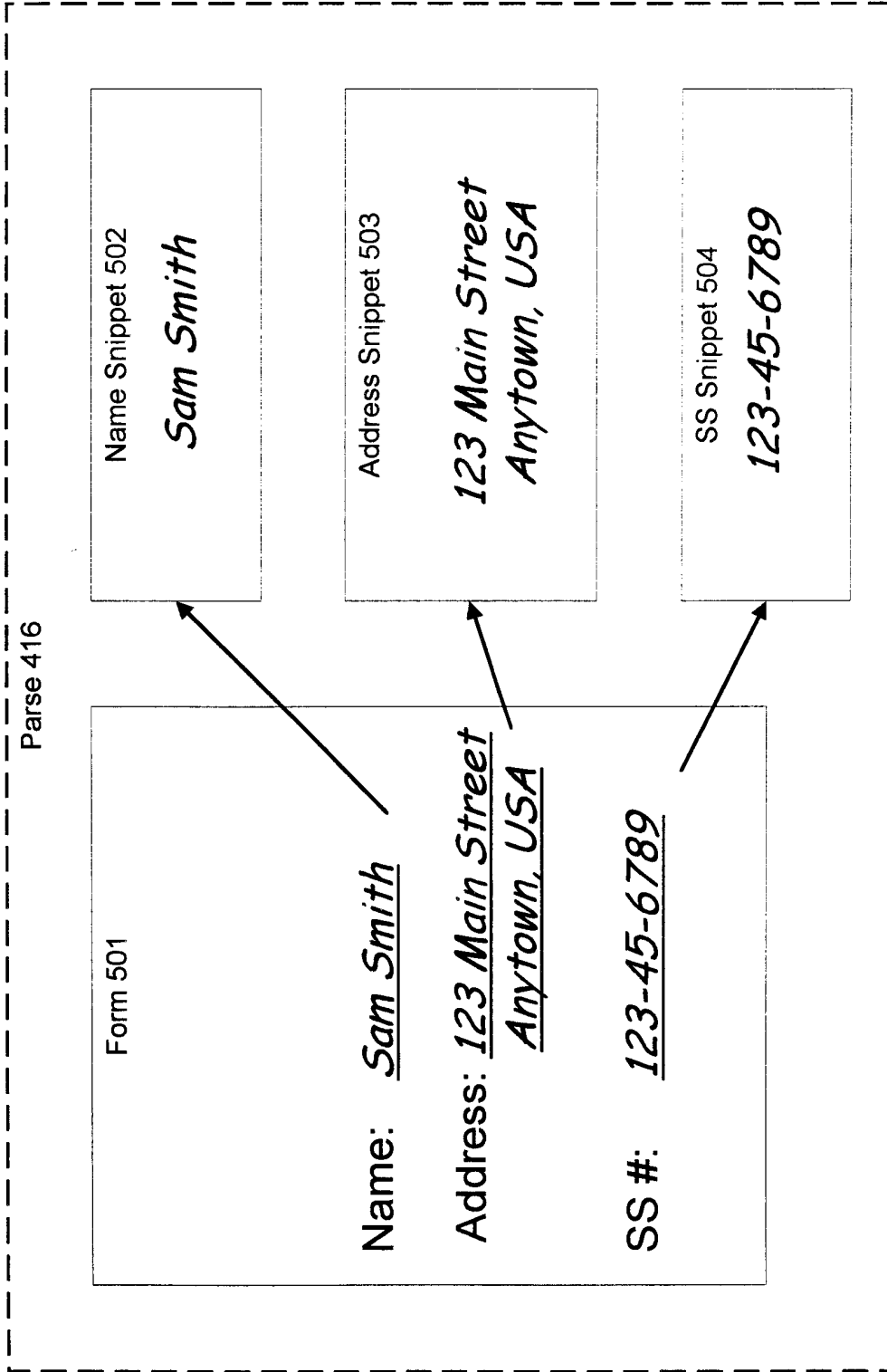


Fig. 5



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/04558

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) :G06F 15/00, 17/00, 17/21, 17/24; H04N 1/00 US CL :707/506; 358/403 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. : 707/506; 358/403 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: forms, data extraction		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y -- Y	US 5,721,940 A (LUTHER et al) 24 February 1998, whole document.	1-27 ----- 1-27
Y	US 5,555,101 A (LARSON et al) 10 September 1996, whole document.	1-27
A	US 5,999,948 A (NELSON et al) 07 December 1999, whole document.	1-27
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* *A* *E* *L* *O* *P*	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance earlier document published on or after the international filing date 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) document referring to an oral disclosure, use, exhibition or other means document published prior to the international filing date but later than the priority date claimed	*T* *X* *Y* *&*
		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 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 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 13 APRIL 2001		Date of mailing of the international search report 04 MAY 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 <i>Heather Herndon</i> HEATHER HERNDON Telephone No. (703) 305-3900