



- (51) **International Patent Classification:**
G06F 16/93 (2019.01) *G06V 30/416* (2022.01)
- (21) **International Application Number:**
PCT/US2024/036768
- (22) **International Filing Date:**
03 July 2024 (03.07.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/512,553 07 July 2023 (07.07.2023) US
- (71) **Applicant: NOTABLE SYSTEMS, INC.** [US/US]; 35 Eudora St., Denver, Colorado 80220 (US).
- (72) **Inventors: LIPPKE, David;** 35 Eudora St., Denver, Colorado 80220 (US). **DENTON, Sarah;** 35 Eudora St., Denver, Colorado 80220 (US).
- (74) **Agent: DIETRICH, Allison W.** et al.; P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,

(54) **Title:** AUTOMATED DATA EXTRACTION, VALIDATION, AND REPORTING

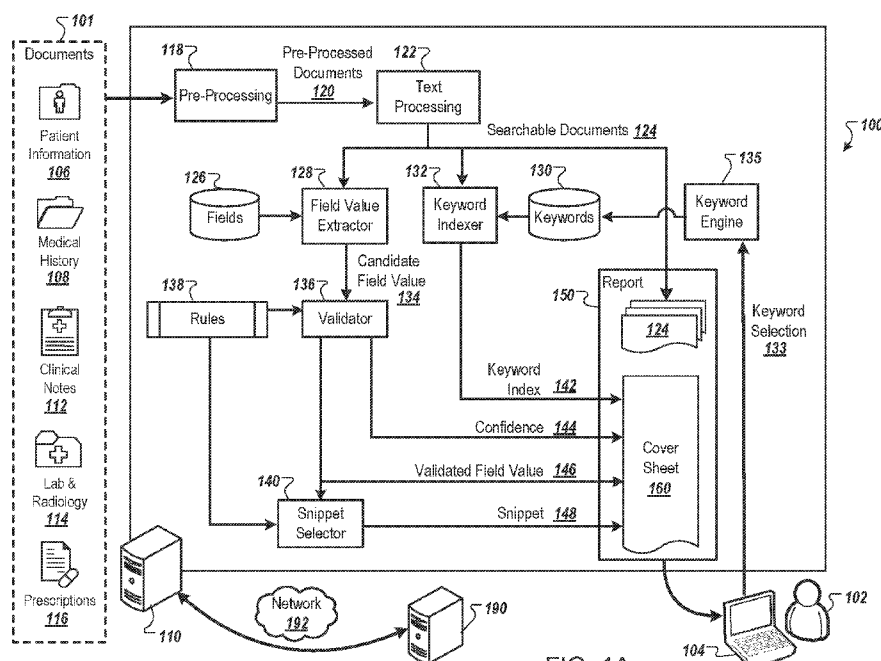


FIG. 1A

(57) **Abstract:** Methods, systems, and apparatus, including computer programs encoded on computer storage media, are disclosed. A method includes receiving documents that are associated with a first entity; converting the documents into searchable documents; extracting, from the searchable documents, values for a set of fields that relate to the first entity; identifying locations of the values within the searchable documents; generating a report including the set of fields and the respective values; and appending, to the report, a set of selectable user interface elements associated with the set of fields. User selection of a selectable user interface element associated with a field results in provision of a snippet of a document. The snippet shows a portion of the document including the location of the value for the field. The method includes providing, for display, the report and the appended set of selectable user interface elements associated with the set of fields.

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

AUTOMATED DATA EXTRACTION, VALIDATION, AND REPORTING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of the U.S. Provisional Patent Application No. 63/512,553 filed July 7, 2023, which is incorporated herein by reference in its entirety, including all Appendices.

TECHNICAL FIELD

[0002] Embodiments of the present disclosure related to a computer-implemented system and method for extraction, aggregation, validation, summarization, and visualization of data from documents.

BACKGROUND

[0003] Document collections can include many different types of documents. Data entries can be provided in a document in a typewritten or handwritten format. Different documents can have different fields from each other. There may be variability between documents with similar content. Document collections can include, for example, medical records, veterinary records, legal records, personnel record, and other types of records.

SUMMARY

[0004] Disclosed are systems and techniques for automated data extraction, validation, and visualization. In general, one innovative aspect of the subject matter described in this specification can be embodied in methods including the actions of: receiving, from a client device, a set of documents that are associated with a first entity; converting the set of documents into a searchable set of documents; extracting, from the searchable set of documents, values for a set of fields that relate to the first entity; identifying locations of the values within the searchable set of documents; generating a report including the set of fields and the respective values; appending, to the report, a set of selectable user interface elements associated with the set of fields. User selection of a selectable user interface element associated with a first field results in provision of a first snippet of a first document. The first snippet shows a portion of the first document including the location of the value for the first field. The actions include providing, for display on the client device, the report and the appended set of selectable user interface elements associated with the set of fields.

[0005] The foregoing and other implementations can each optionally include one or more of the following features, alone or in combination. In some implementations, extracting, from

the searchable set of documents, the values for the set of fields that relate to the first entity includes, for the first field: identifying, within the searchable set of documents, multiple instances of candidate values for the first field; selecting, from the multiple instances of candidate values, a first instance of a first candidate value for the first field; and assigning the first instance of the first candidate value as the value for the first field. The portion of the first document includes the location of the first instance of the first candidate value.

[0006] In some implementations, the actions include: determining a confidence for each of the multiple instances of candidate values for the first field; and selecting the first instance of the first candidate value for the first field based on the confidence for each of the multiple instances of candidate values.

[0007] In some implementations, the confidence of an instance of a candidate value is based at least in part on: a text recognition confidence for the instance; or a similarity of the instance of the candidate value to other instances of candidate values.

[0008] In some implementations, the actions include: determining whether the confidence for the selected instance satisfies a threshold confidence; and providing, for inclusion in the report, a visual indication of whether the confidence satisfies the threshold confidence.

[0009] In some implementations, extracting, from the searchable set of documents, the values for the set of fields that relate to the first entity includes, for the first field: identifying, within the searchable set of documents, multiple instances of candidate values for the first field; identifying, from the multiple instances of candidate values, a mode candidate value; and assigning the identified mode candidate value as the value for the first field.

[0010] In some implementations, generating the report including the set of fields and the respective values includes, for a particular field: determining, using a set of rules, a specified value format for the particular field; and converting the respective value for the particular field to the specified value format.

[0011] In some implementations, generating the report including the set of fields and the respective values includes, for a particular field: determining, using a set of rules, a specified unit of measurement for the particular field; and converting the respective value for the particular field to the specified unit of measurement.

[0012] In some implementations, the multiple instances of candidate values includes a second instance of the first candidate value for the first field, and user selection of the selectable user interface element results in provision of the first snippet and a second snippet.

The second snippet shows a portion of a document including the location of the second instance of the first candidate value for the first field.

[0013] In some implementations, the actions include generating the first snippet, including: determining, using a set of rules indicating snippet sizes for the set of fields, a specified snippet size of the first field; and selecting the portion of the first document based on the location of the value for the first field and the snippet size determined using the set of rules.

[0014] In some implementations, the set of fields, the respective extracted values, and the set of selectable user interface elements are provided in a cover sheet of the report, and the report further includes the set of documents.

[0015] In some implementations, generating the report including the set of fields and the respective values includes: identifying an error in the set of documents; and providing, for inclusion in the report, a visual indication of the error.

[0016] In some implementations, the actions include: in response to identifying the error, performing an error correction; and providing, for display on the client device, the report with the error correction. The error correction includes at least one of: omitting a value for at least one field of the set of fields; adding a value for at least one field of the set of fields; or modifying a value for at least one field of the set of fields.

[0017] In some implementations, the user selection of the selectable user interface element includes a first type of user selection; and a second type of user selection of the selectable user interface element results in opening a link to the first document.

[0018] In general, one innovative aspect of the subject matter described in this specification can be embodied in methods including the actions of: providing, to a client device, application data for a user interface of an application, the user interface including one or more user interface elements for changing configuration settings of a system for generating a report for a set of documents. The configuration settings include: a set of fields for inclusion in the report; selection criteria for values for the set of fields; and formatting rules for the values for the set of fields; receiving, from the client device, data indicating user interaction with the one or more user interface elements; and in response to receiving the data indicating the user interaction with the one or more user interface elements, changing the configuration settings from a first set of configuration settings to a second set of configuration settings specified by the user interaction.

[0019] The foregoing and other implementations can each optionally include one or more of the following features, alone or in combination. In some implementations, the configuration settings include at least one of: a set of keywords for indexing in the report; actions to be performed in response to detecting the error in the set of documents; a document format for the report.

[0020] In some implementations, the selection criteria for the values for the set of fields includes at least one of a threshold confidence or an error tolerance; and the formatting rules for the values for the set of fields includes at least one of a number format, a name format, or a measurement unit.

[0021] In some implementations, the actions include: receiving, from a client device, a set of documents that are associated with a first entity; converting the set of documents into a searchable set of documents; extracting, from the searchable set of documents, values for the set of fields included in the second set of configuration settings based on the selection criteria included in the second set of configuration settings; identifying locations of the values within the searchable set of documents; converting the values to respective formats specified by the formatting rules included in the second set of configuration settings; and generating a report including the set of fields and the respective values converted to the respective formats specified by the formatting rules in the second set of configuration settings.

[0022] In some implementations, the actions include: appending, to the report, a set of selectable user interface elements associated with the set of fields. User selection of a selectable user interface element associated with a first field results in provision of a first snippet of a first document. The first snippet shows a portion of the first document including the location of the value for the first field. The actions include providing, for display on the client device, the report and the appended set of selectable user interface elements associated with the set of fields.

[0023] Other implementations include corresponding computer systems, apparatus, computer program products, and computer programs recorded on one or more computer storage devices, each configured to perform the actions of the methods. A system of one or more computers can be configured to perform particular operations or actions by virtue of having software, firmware, hardware, or a combination of them installed on the system that in operation causes or cause the system to perform the actions. One or more computer programs can be configured to perform particular operations or actions by virtue of including

instructions that, when executed by data processing apparatus, cause the apparatus to perform the actions.

[0024] This specification uses the term “configured to” in connection with systems, apparatus, and computer program components. That a system of one or more computers is configured to perform particular operations or actions means that the system has installed on it software, firmware, hardware, or a combination of them that in operation cause the system to perform those operations or actions. That one or more computer programs is configured to perform particular operations or actions means that the one or more programs include instructions that, when executed by data processing apparatus, cause the apparatus to perform those operations or actions. That special-purpose logic circuitry is configured to perform particular operations or actions means that the circuitry has electronic logic that performs those operations or actions.

[0025] Particular embodiments of the subject matter described in this specification can be implemented to realize one or more of the following advantages. For example, the techniques described in this specification can achieve faster data entry (which in a medical application could enable faster order entry and fulfillment), fewer typing errors, less tedium of manual data entry, reduced erroneous communications between entities (e.g., clinics and hospitals) as they relate to the first entity (e.g., a patient).

[0026] The disclosed techniques can be used to automatically generate an indexed cover sheet of a large set of documents, which can be of any size and format and which includes data extracted from numerous documents of varying formats and content. A set of documents can include hundreds or thousands of pages of records. The pages can have varying formats, and can include data fields at different locations on a page. For example, some documents may include a field for a person’s name near the top of the page, and some documents may include a field for a person’s name near the bottom of the page. Information can be presented in typed or handwritten text, and in different fonts, colors, and sizes. The techniques described herein can be employed to parse through the documents and apply rule sets in order to elevate higher confidence values for data fields over lower confidence values. The system can present the values having the highest confidence, while also providing a visual representation of other high likelihood values. In this way, the, described solutions allow for comprehensive data extraction from differently structured records of varying formats.

[0027] The indexed cover sheet further includes user interface elements that enable previews, or snippets, of data within the documents, and hyperlinks to locations of data

values in the documents. The snippets allow the user to preview portions of a document from which information was obtained for inclusion in the cover sheet. The snippets overlap with a portion of the cover sheet. In some examples, the user interface shows confidences that are associated with each data value and/or each snippet. Thus, the user can view the extracted information, the source(s) of the extracted information, and the accuracy confidence of the extracted information on one screen and with a single operator action (e.g., a mouse hover, a mouse click). In some examples, such as when two or more sources of information have similar confidences, the system automatically presents multiple snippets, each showing a portion of one of the two or more sources of information. Therefore, the user interface is improved by automatically displaying snippets to the user based on the accuracy confidence of the information in the snippets.

[0028] A snippet can be presented when a user interaction is received such as a mouse hover over a data field. This enables a user to efficiently verify accuracy of extracted data values, and to quickly view original sources of information. The disclosed techniques identify potential errors and visually flag the potential errors. The user can investigate errors and discrepancies in a set of documents while viewing and interacting with a single page of a single document, without needing to scroll, print, or view multiple documents side-by-side. Additionally, a snippet vanishes when a different user interaction is initiated such as the mouse moving away from the data field. This can prevent overwhelming a user interface that is already populated with content. Therefore, the user interface is improved by automatically removing snippets in response to user interaction that moves away from a selected field, to enable the user to again view the entire unobstructed cover sheet.

[0029] The disclosed techniques can be implemented to improve accuracy of record keeping, such as in medical contexts. The improved accuracy record keeping can reduce the likelihood of errors in medical decision-making. The disclosed systems can improve handling of private and personal information by detecting cross-contamination between records such as health records for multiple different patients. Errors in records can be quickly identified and remedied, and duplicate records can be automatically flagged for deletion.

[0030] The details of one or more implementations of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. 1A shows an example system for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure.

[0032] FIG. 1B is a flow diagram of an example process for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure.

[0033] FIG. 2A shows an example system for changing configuration settings for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure.

[0034] FIG. 2B is a flow diagram of an example process for changing configuration settings for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure.

[0035] FIGS. 3A to 3F show an example interactive cover sheet of a report generated by the system of FIG. 1A in accordance with implementations of the present disclosure.

[0036] FIGS. 4A to 4C show another example interactive cover sheet of a report generated by the system of FIG. 1A in accordance with implementations of the present disclosure.

[0037] FIG. 5 shows an example interactive cover sheet for a report including a set of documents in which information for multiple entities were identified in accordance with implementations of the present disclosure.

[0038] FIGS. 6A to 6D show an example interactive document generated by the system of FIG. 1A in accordance with implementations of the present disclosure.

[0039] FIG. 7 is a diagram illustrating an example of a computing system.

[0040] Like reference numbers and designations in the various drawings indicate like elements.

DETAILED DESCRIPTION

[0041] This specification generally relates to data processing techniques, including extraction of data from a set of digital components (e.g., digital documents) and provision of an indexed report with hover functionality to enable quick visualization of and access to key data (e.g., text) in the set of digital components (e.g., the digital documents).

[0042] FIG. 1A shows an example system 100 for automated data extraction, validation, and reporting. The system 100 includes a server 110 and a client device 104. In general, the server 110 receives a set of documents 101 and a request for generation of a report 150. The server 110 can receive the set of documents 101 from the client device 104 or from one or

more other computing devices. The server 110 generates the report 150 and provides the report 150 to the client device 104. The report 150 includes a searchable set of documents 124, which include the set of documents 101 after text processing and conversion to a common format, size, and shape. The report 150 also includes a cover sheet 160.

[0043] The set of documents 101 are associated with an entity. In some examples, the entity is a living being such as a medical patient, a client (e.g., a legal client, a financial client), an employee, a student, or a veterinary patient. In some examples, the entity is a non-living entity such as an organization (e.g., a business, a school), a property (e.g., a building, a development), or a project.

[0044] For ease of description and without limiting the scope of this disclosure, the example system 100 generates the report 150 for an entity that is a human medical patient. The documents 101 in this example are medical documents relating to the patient. The documents 101 include patient information 106, medical history 108, clinical notes 112, lab and radiology reports 114, and prescriptions 116. One skilled in the art will appreciate that different types of documents may be provided for different use cases of the presently-described techniques.

[0045] The documents 101 can be in any appropriate format such as TIFF, JPG, PDF, Fax, Word, TXT, CSV, JSON, XLS, PNG, EML, MSG, OST, etc. The documents can include typed text, handwritten text, drawings, photographs, graphical images, or any combination thereof. Example documents are provided in Appendices C and D of the provisional application, as incorporated by reference herein in their entirety.

[0046] The server 110 performs pre-processing 118 on the documents 101 to produce pre-processed documents 120. Pre-processing a document can include any combination of the following document, image, or data processing operations: adjusting an orientation of the document, clarifying text and/or images in the document, converting the document from a first format to a second, different format, converting the document from a first size to a second, different size. In some examples, each document in the set of documents is aligned to the same orientation, converted to a same document format, converted to a same size, or any combination thereof.

[0047] In some examples, during pre-processing 118, the set of documents is combined into a single file (e.g., a pdf file), and the pages are numbered sequentially. The page numbers can then be used by the keyword indexer 132 and field value extractor 128 to identify locations of information extracted from the set of documents 101.

[0048] The server 110 performs text processing 122 on the pre-processed documents 120 to produce a searchable set of documents 124. In some examples, the text processing 122 includes processing the pre-processed documents 120 using optical character recognition (OCR) algorithm(s), natural language processing (NLP) algorithm(s), or both.

[0049] The OCR algorithms capture text elements from documents and convert the text elements into a machine-readable text format. The OCR algorithms can produce searchable text as output. In some examples, the OCR algorithms output information indicating the position of text in a document. The OCR algorithms, the NLP algorithms, or both, can include machine learning algorithms such as neural network models.

[0050] The NLP algorithms can perform operations such as part-of-speech tagging, sentence breaking, parsing, named entity recognition, terminology extraction, entity linking, relationship extraction, lexical semantics, or any combination thereof. The NLP algorithms can process textual information from the documents and output a characteristic of the textual information. For example, the NLP algorithms can output classifications, relationships, and/or meanings of words and phrases in the set of documents.

[0051] The keyword indexer 132 parses the searchable set of documents 124 to identify, within the set of documents, locations of keywords that are specified in a set of keywords 130. The keyword indexer 132 extracts the identified keywords from the searchable documents 124.

[0052] The set of keywords 130 includes a pre-determined set of words (or other sequence of text/characters and/or digits) that are designated by an operator as values to be extracted from the set of documents 101. In some implementations, the set of keywords 130 can be any pre-defined set of words or phrases (or other sequence of text/characters and/or digits) of interest that are to be searched for and identified within the set of documents. Such pre-defined set of words can be specified in a specification or another document or format that is recognizable and usable by the server to facilitate parsing of the searchable documents 124.

[0053] The system 100 can include a keyword engine 135 that is configured to generate the set of keywords 130 for use in parsing the searchable set of documents 124. The keyword engine 135 can provide a keyword configuration tool to the user 102 through the client device 104. The keyword configuration tool enables the user 102 to provide a keyword selection 133 that specifies words and/or other text strings that are to be added to the set of

keywords 130. The keyword selection 133 can include instructions to add, remove, and/or modify keywords in the set of keywords 130.

[0054] In some examples, the keyword engine 135 provides, to the client device 104, an initial set of keywords to be presented to the user 102. The initial set of keywords can be a default set of keywords for the particular context or use case. For example, the initial set of keywords can be a default set of keywords for a medical context. The user 102 can then provide a keyword selection 133 that modifies the initial set of keywords.

[0055] In some examples, the keyword engine 135 provides, to the client device 104 for presentation to the user 102, multiple options for sets of keywords. For example, the keyword engine 135 can provide an option for a medical-related keyword set, a veterinary keyword set, a financial keyword set, or an educational keyword set. The user 102 can then provide a keyword selection 133 that indicates the selected keyword set from the multiple options. In some examples, after selecting a particular keyword set, the user 102 can provide input through the client device 104 that modifies the selected keyword set by adding, removing, or modifying terms within the keyword set.

[0056] Based on the keyword selection 133, the keyword engine 135 generates and/or updates the set of keywords 130. In some examples, the keyword engine 135 generates and provides instructions to the keyword indexer 132 instructing the keyword indexer 132 to parsing the searchable documents 124 for the set of keywords 130.

[0057] In some examples, the server 110 stores different sets of keywords for different contexts. For example, a first set of keywords 130 for a medical context can include words such as “chronic,” “diagnosis,” “respiratory,” “weight,” “fracture,” and “infection.” Other example keywords for a medical context are shown in Appendices B, C, D, and E of the provisional application as incorporated herein by reference in its entirety. Different sets of keywords can be stored for use in different contexts (e.g., a legal context, a financial context, an educational context).

[0058] In some examples, the server 110 stores different sets of keywords for different clients. In some examples, the set of keywords 130 for a client includes words, phrases, and other text that are pre-selected by the client as described above with reference to FIG. 1. In some examples, the set of keywords 130 for a client are generated by tailoring a default set of keywords 130 according to the client’s specifications. For example, referring to FIG. 2, a user 202 can submit a configuration change request 220 to the server 110 to request a change in configuration, which can include a change in the set of keywords 130. A change in the set of

keywords 130 can include, for example, an addition of one or more keywords, a removal of one or more keywords, or a modification to one or more keywords.

[0059] In some examples, the keyword indexer 132 selects a set of keywords 130 from multiple different sets of keywords based on a context of the set of documents 101, based on a client that submitted the set of documents, or both. For example, the keyword indexer 132 can determine, by parsing the searchable documents 124, that the searchable documents 124 were submitted by a particular client, and can therefore select a set of keywords 130 specific to the particular client (e.g., by identifying the client using its client identifier and then obtaining the keywords 130 from storage that correspond to that client identifier). In another example, the keyword indexer 132 can determine, by parsing the searchable documents 124, that the searchable documents 124 are related to a medical context, and can therefore select a set of keywords 130 specific to the medical context (e.g., by identifying the medical context using a provision of a particular context identifier corresponding to a medical context and then obtaining the keywords 130 from storage that correspond to that client identifier). As will be appreciated, context identifiers and/or client identifiers can be provided in association with different sets of keywords, to enable identification of the appropriate set of keywords for a particular client, context, or identifier provided for a particular environment.

[0060] The keyword indexer 132 parses the searchable documents 124 to identify locations where the keywords 130 are located within the set of documents. The keyword indexer 132 outputs a keyword index 142. In some implementations, the keyword index 142 includes a list of keyword-page pairs. Each keyword-page pair includes a keyword and a page of the set of documents on which the keyword was located. The keyword index 142 is presented on the cover sheet 160 of the report, showing a list of the keywords and the respective locations within the set of documents. Example keyword indices are shown in FIGS. 3A and 4A.

[0061] The field value extractor 128 extracts, from the searchable documents 124, candidate values 134 for each of a set of fields 126. A candidate field value 134 is a value that is extracted by the field value extractor 128 and is a candidate for entry into a field 126 that is to be presented in the report 150. An example field is “Date of Birth,” a first example candidate field value for the field is “May 5, 1979,” a second example candidate field value for the field is “5/5/79,” and a third example candidate field value for the field is “05/06/1979.” Another example field is “First Name,” a first example candidate field value

for the field is “Daniel,” a second example candidate field value for the field is “Dan,” and a third example candidate field value for the field is “Dawn.”

[0062] The field value extractor 128 identifies locations of the candidate values within the searchable documents 124. The field value extractor 128 outputs the candidate field values 134 to the validator 136, which evaluates the candidate field values to select a particular candidate field value to include in the respective field on the cover sheet, as described in greater detail below.

[0063] The fields 126 are a pre-determined set of fields to be included on the cover sheet 160. In some examples, the server 110 stores different sets of fields 126 for different contexts. For example, a first set of fields 126 for a medical context can include “First Name,” “Last Name,” “Address,” “Date of Birth,” “Height,” and “Weight.” Other example fields for a medical context are shown in Appendices B and C of the provisional application as incorporated herein by reference in its entirety. Different sets of fields can be stored for use in different contexts (e.g., a legal context, a financial context, an educational context).

[0064] In some examples, the server 110 stores different sets of fields 126 for different clients. In some examples, the fields 126 for a client are pre-selected by the client. In some examples, the fields 126 for a client are generated by tailoring a default set of fields 126 according to a client’s specifications (or the specifications for a particular context or another specified environment). For example, referring to FIG. 2, a user 202 can submit a configuration change request 220 to the server 110 to request a change in configuration, which can include a change in the fields 126. A change in the fields 126 can include, for example, an addition of one or more fields, a removal of one or more fields, or a modification to one or more fields. In some examples, the user 202 can specify a format for a field. For example, the client can input, to the server 110, a preference for the “Date of Birth” field to be presented as “DOB” on the cover sheet 160.

[0065] In some examples, the field value extractor 128 identifies, within the searchable documents 124, multiple instances of candidate values for a particular field. The field value extractor 128 outputs candidate field values 134 to the validator 136. The candidate field values 134 can include alphanumeric values.

[0066] The validator 136 validates the candidate field values 134 using a set of rules 138. In general, the validator 136 identifies errors and discrepancies, and evaluates a confidence 144 for each of the field values. The confidence 144 of a field value indicates an estimated level of accuracy of the field value. The validator 136 selects a particular value for inclusion

in each field 126 of the report 150. In some examples, the validator 136 converts the selected value to a specified format according to the rules 138.

[0067] The specified format can include, for example, a numerical format and/or a unit of measurement. For example, the rules 138 may specify that a format for dates is DDMONYY (e.g., 08AUG95). In another example, the rules 138 may specify that any height or length is to be reported in units of inches out to one decimal place (e.g., 58.0 inches). In another example, the rules 138 may specify that a format for names is Last, First, MI (e.g., Smith-Hartley, Brenton, J).

[0068] The validator 136 can select, from the multiple instances of the candidate value, a particular instance. The validator 136 can assign the particular instance of the candidate value as the value for the field.

[0069] In an example, the set of fields 126 includes a Last Name field. Referring to FIG. 3B, the patient's last name, as presented in the Last Name field of cover sheet 300, is "De Gracia." The field value extractor 128 may extract multiple instances of the patient's last name from the set of documents 101. The validator can consider each instance of the last name extracted from the set of documents 101 as an instance of a candidate field value for the Last Name Field. In some cases, multiple instances of candidate field values have the same format (e.g., a first instance of "De Gracia" and a second instance of "De Gracia"). In some cases, multiple instances of candidate field values have different formats from each other (e.g., a first instance of "De Gracia," a second instance of "de Gracia," a third instance of "DeGracia"). The validator 136 can assign a particular instance of the candidate value as the field value for the Last Name field. The assigned field value is then presented in the Last Name field on the cover sheet 300. The validator 136 can assign a particular candidate field value by applying rules 138, as described below.

[0070] In some examples, the rules 138 specify that the validator 136 assigns a particular candidate field value to a field by determining the confidence 144 for each of the multiple instances of candidate values for a particular field. The validator 136 selects the particular instance of a candidate value based on the determined confidences (e.g., the particular instance having the highest confidence relative to the other determined confidences).

[0071] In some examples, the confidence is based at least in part on a text recognition confidence for the instance. The text recognition confidence can be determined by the OCR algorithm(s). For example, a text recognition confidence determined by the OCR algorithms for a typed document may be higher than a text recognition confidence for a handwritten

document. Similarly, a text recognition confidence for a higher color contrast document may be higher than a text recognition confidence for a lower color contrast document. The validator 136 can determine a higher confidence 144 for candidate values having higher text recognition confidence than for candidate values having lower text recognition confidences.

[0072] In some examples, the confidence is based at least in part on a similarity of the instance of a candidate value to other instances of candidate values. For example, three instances of candidate values for a patient's height may be extracted from the set of documents. The first instance is a value of 67.5 inches, the second instance is a value of 68 inches, and the third instance is a value of 64 inches. The validator 136 can determine a higher confidence 144 for the first instance and for the second instance, and a lower confidence for the third instance, due to the first instance and the second instance being more similar to each other than the third instance.

[0073] In some examples, the rules 138 specify that the validator 136 assigns a particular candidate field value to a field by identifying, from the multiple instances of candidate values for a field, a mode candidate value. The mode value is the value that appears most often in the set of candidate field values. The validator 136 can assign the identified mode candidate value as the value for the field. For example, referring to FIG. 3F, three instances of candidate values for the patient's weight are extracted from the set of documents. The first instance is a value of 93.3 kilograms (kg), the second instance is a value of 91.9 kg, and the third instance is a value of 93.3 kg. The validator 136 selects the value of 93.3 kg as the field value for the Weight field, due to 93.3 kg being the mode value. In some examples, instead of selecting the mode value of multiple candidate values, the rules 138 may specify that the validator 136 select a different value such as, e.g., a median value, a mean value, a maximum value, a minimum value, or a most recent value.

[0074] In some examples, the rules 138 specify that the validator 136 assigns a particular candidate field value to a field by verifying the candidate field values against record(s) associated with the field. For example, the validator can validate candidate field values against public sources of information to determine confidence. The public sources of information can be obtained, for example, from a third party server 190. The server 110 can communicate with the third party server 190 over a network 192. For an example of an Address field, the validator 136 can compare candidate field values to public records such as postal service records obtained from the third party server 190. The validator 136 can then

select a particular candidate field value that matches the public records more closely than other candidate field values.

[0075] In some examples, the rules 138 specify that the validator 136 assigns a particular candidate field value to a field by applying other selection criteria. The selection criteria can include, for example, a bias in favor of longer versions of candidate field values over shorter versions. For example, for a First Name field, the validator 136 may be configured to select a longer version of a name (e.g., “Jonathan”) over a shorter version of the name (e.g., “John.”)

[0076] The validator 136 can output the confidence 144 of field values for inclusion in the report 150. In some examples, the validator 136 determines, for each candidate field value 134, whether the confidence 144 satisfies a threshold confidence. The validator 136 can provide, for inclusion in the report 150, rendering data for a visual indication of whether the confidence 144 satisfies the threshold confidence. The visual indication can include, for example, a green checkmark to indicate that the confidence 144 satisfies (e.g., exceeds) the threshold confidence, and a yellow or red warning icon to indicate that the confidence 144 does not satisfy (e.g., does not exceed) the threshold confidence. The visual indication can be included in the report near the field associated with the field value.

[0077] The validator 136 outputs validated field values 146 and their respective confidences 144 for inclusion in the report 150. The validator 136 outputs the validated field values 146 to the snippet selector 140.

[0078] In some examples, the validator 136 outputs multiple instances of candidate values to the snippet selector 140. The snippet selector 140 can generate multiple snippets 148, each snippet 148 showing a location of one of the multiple instances of the candidate values. A snippet 148 is a text box that shows a snapshot of an underlying document from which a candidate value was extracted for a field of a cover sheet. When presented through a user interface, the snippet 148 is superimposed over the cover sheet such that the snippet 148 covers part of the cover sheet but does not cover the particular field that is associated with the snippet. Thus, the user 102 can view the snippet 148 and the associated field and validated field values 146 simultaneously.

[0079] In some cases, user selection of the selectable user interface element results in provision of multiple snippets 148. Each snippet shows a portion of a document including the location of the respective instance of the candidate value for the first field.

[0080] For example, referring to FIG. 4B, user selection of an Address field (e.g., a State field) on the cover sheet 400 results in presentation of four snippets. Each snippet shows a

respective instance of a patient address that was extracted from the set of documents. The snippets were generated by the snippet selector 140, which received the four instances of candidate values from the validator 136.

[0081] In some examples, the snippets 148 can be presented in order of confidence. For example, referring to FIG. 3C, the field value extractor 128 extracts different dates of birth on different pages of the set of documents, resulting in generation of multiple snippets. The snippet 318 shows the date of birth of 03/10/1950, which was found three times in the set of documents. The snippet 320 shows a date of birth of 03/11/1950, which was found one time in the set of documents. The validator 136 may determine a higher confidence for the date of birth of 03/10/1950, due to the date of birth being found in more instances in the set of documents.

[0082] In some cases, the validator 136 can evaluate the confidences for the candidate date of birth values using records obtained from the third party server 190. For example, the server 110 can communicate with the third party server 190 over the network 192 to request the date of birth for the patient. The server 110 can provide identifying information for the patient (e.g., name, address, phone number, etc.) to the third party server 190 and can receive in response, from the third party server 190, information indicating the patient's date of birth.

[0083] In the example of FIG. 3C, the validator 136 determines a higher confidence for the date of birth of 03/10/1950, shown in the snippet 318, and a lower confidence for the date of birth of 03/11/1950, shown in the snippet 320. Therefore, the snippet 318 is presented above the snippet 320 when eyeball icon 316 is selected or the mouse hovers over the Date of Birth field, due to the higher confidence for the value in the snippet 318.

[0084] In some examples, the number of snippets presented on the user interface depends on the confidences of the associated field values. For example, when the validated field value 146 presented on the cover sheet 160 has a confidence above a threshold confidence (e.g., 90%), a single snippet may be presented in response to user interaction with the field. When the validated field value 146 presented on the cover sheet 160 has a confidence below the threshold confidence, multiple snippets may be presented in response to user interaction with the field. In some examples, such as when the validated field value 146 presented on the cover sheet 160 has a confidence below the threshold confidence, a single snippet may be presented, and the user interface can present an option for the user to view additional snippets that may have a lower confidence than the validated field value 146.

[0085] In some examples, the snippets 148 can be presented in order of their location in the set of documents 101. For example, referring to FIG. 3F, a patient weight was extracted from three different locations within the set of documents 101, resulting in generation of multiple snippets. The snippet 352 may show a portion of page 4, while the snippet 354 shows a portion of page 7, and the snippet 356 shows a portion of page 10. The snippet 352 can therefore be presented above the snippet 354, which is presented above the snippet 356, when the eyeball icon 360 is selected or the mouse hovers over the Weight field, due to the sequential order of the pages from which the snippets were generated.

[0086] The snippet selector 140 selects a snippet 148 from the searchable documents 124. The snippet selector 140 provides the snippet 148 as output for inclusion in the report 150. The snippet selector 140 can generate a snippet by determining, using rules indicating snippet sizes for the set of fields, a specified snippet size of a particular field. For example, snippets for an Address field may be configured to be larger than snippets for a Date of Birth field. In some examples, the snippet size can be dynamically adjusted based on the content to be provided in the snippet 148. The snippet selector 140 can select the portion of the document based on the location of the value for the first field and the snippet size determined using the rules.

[0087] In some examples, the snippet selector 140 selects a snippet size based on a size of a field on a standard form. For example, a standard vaccine card may include fields for a patient name, a date of birth, a product number, a date of vaccination, and a clinic site. The server 110 can store data indicating the standard locations and sizes of the fields on the standard vaccine card. When the validator 136 outputs a validated field value 146 to the snippet selector 140 and the validated field value 146 is extracted from a particular field of a standard vaccine card, the snippet selector 140 can generate a snippet 148 that shows the particular field. The snippet 148 can conform to the shape, size, and/or location of the particular field on the standard vaccine card. In this way, the snippet 148 can be generated to show the entire field, without showing extra information.

[0088] In some examples, the location of a field value on a page can be specified by an x-y coordinate on the page. In some examples, the x and y values are in units of length (e.g., centimeters, inches). In some examples, the x and y values are in units of fractional width and fractional length (e.g., 20% across width from left to right, 80% across length from top to bottom).

[0089] The server 110 generates a report 150. The report includes the set of fields 126 and the respective validated field values 146 for the fields 126. The fields 126, the validated field values 146, and the keyword index 142 can be presented on the cover sheet 160 of the report 150. Data that is extracted from the searchable documents 124 can be highlighted or otherwise annotated in the underlying documents included in the report 150. For example, keywords and field values can be highlighted in the searchable documents 124 that are included in the report 150.

[0090] The cover sheet 160 can include links to the locations of the keywords and field values within the searchable documents 124. In some examples, annotations on a document can be added in real-time in response to a user selecting a hyperlink on the cover sheet 160 that causes presentation of the document. The annotations can be removed when the user navigates away from the document and/or returns to viewing the cover sheet 160.

[0091] The server 110 appends, to the cover sheet 160, a set of selectable user interface elements associated with the set of fields 126. User selection of a selectable user interface element associated with a particular field causes provision of a snippet 148 of a document that shows a portion of the document including the location of the value for the field. In some examples, the snippet is provided in java script.

[0092] In some examples, the cover sheet 160 is in pdf format, and the snippet 148 is provided as part of a layer, or content group, of the pdf file. A layer can include a snippet, multiple snippets, and/or other images. The server 110 can specify a positioning and size of each displayed part of a layer. As a mouse cursor moves around and hovers over or selects a selectable element, a corresponding layer is presented, while other layers remain hidden. For various positions of a mouse cursor on the cover sheet 160, a layer can be prescribed for presentation when the mouse cursor is at the corresponding position. Similarly, for various selectable elements on the cover sheet, a layer can be prescribed for presentation when the element is selected. For example, when a mouse cursor hovers over Patient Name field, a layer can be enabled that includes snippet(s) of document(s) in which the patient's name was found. Other layers (e.g., layers including snippets of a patient's date of birth) remain hidden until a user action is performed that causes the other layers to be presented. When the mouse cursor moves away from the Patient Name field, the layer is disabled so that the corresponding snippet(s) are no longer presented.

[0093] The report 150 can have a document format of, for example, PDF, HTML JSON, CSV, or any combination thereof. The set of fields 126, the validated field values 146, and

the set of selectable user interface elements are provided in the cover sheet 160 of the report 150. The report also includes the set of searchable documents 124.

[0094] The report 150 is provided to a client device 104. A user 102 can view and interact with the report 150 through the client device 104. The client device 104 can include personal computers, mobile communication devices, and other devices that can send and receive data over a network. The network (not shown), such as a local area network (“LAN”), wide area network (“WAN”), the Internet, or a combination thereof, connects the client device 104 and the server 110. The client device 104, the server 110, or both can use a single computer or multiple computers operating in conjunction with one another, including, for example, a set of remote computers deployed as a cloud computing service.

[0095] The server 110 can include several different functional components, including a field value extractor 128, a keyword indexer 132, a validator 136, and a snippet selector 140. The field value extractor 128, the keyword indexer 132, the validator 136, and the snippet selector 140, or a combination of these, can include one or more data processing apparatuses, can be implemented in code, or a combination of both. For instance, each of the field value extractor 128, a keyword indexer 132, a validator 136, and a snippet selector 140 can include one or more data processors and instructions that cause the one or more data processors to perform the operations discussed herein.

[0096] The various functional components of the server 110 can be installed on one or more computers as separate functional components or as different modules of a same functional component. For example, the components of the field value extractor 128, the keyword indexer 132, the validator 136, and the snippet selector 140 of the server 110 can be implemented as computer programs installed on one or more computers in one or more locations that are coupled to each through a network. In cloud-based systems for example, these components can be implemented by individual computing nodes of a distributed computing system.

[0097] FIG. 1B is a flow diagram of an example process 170 for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure. For example, the process 170 can be used by the server 110 from the system 100 or by another data processing apparatus.

[0098] At step 172, the process 170 includes receiving a set of documents. The set of documents can be received with a request for generation of an indexed report for the set of documents (as further described with reference to FIG. 1). The request can identify the set of

documents, with set of documents being associated with a first entity (e.g., a person, a corporation).

[0099] At step 174, the process 170 includes converting the set of documents to a searchable set of documents. Converting the set of documents to the searchable set of documents can include pre-processing the set of documents (as further described with reference to FIG. 1). The pre-processing can include straightening of the documents in the set of documents, clarification of the documents in the set of documents, or another pre-processing operation (e.g., blur correction) that refines quality of document to facilitate optimal optical character recognition. The server or another data processing apparatus can convert, using an optical character recognition algorithm, the set of documents into a searchable set of documents. This can include detecting and recognizing text present in the set of documents.

[0100] At step 176, the process 170 includes extracting, from the set of documents, values for a set of fields. The server can extract data for a first set of fields that relate to the first entity (as further described with reference to FIG. 1). Referring to the example interactive cover sheet 300 shown in FIG. 3A, the first set of fields can include fields identifying the first entity (e.g., name, date of birth, address, etc.).

[0101] At step 178, the process 170 includes identifying locations of the values (as further described with reference to FIG. 1). The location of a value can include a page number of a page on which the value was found. In some examples, the location of a value can include an x-y coordinate location of the value on the page. In some examples, the x-y coordinates are determined in units of distance (e.g., length, width). In some examples, the x-y coordinates are determined in units of fractional distance (e.g., fractional length, fractional width).

[0102] At step 180, the process 170 includes generating a report including the set of fields and the values. In some examples, the report includes (1) data for the first set of fields and (2) the set of keywords and the corresponding locations within the set of documents (as further described with reference to FIG. 1). An example of the report 150 is shown in FIGS. 3 to 6. It will be appreciated that the fields and the keywords are hyperlinked with positional references to the portion(s) of the set of documents where the underlying data or keywords are presented. As such, clicking on these hyperlinks results in the user (e.g., user 102) being navigated to the portion of the document that forms the basis for the field or keyword. This is

further illustrated with reference to Appendix C as well as Appendix D of the provisional application, as incorporated by reference herein in its entirety.

[0103] At step 182, the process 170 includes appending selectable user interface elements to the report. The server or another data processing apparatus can be configured to provide, on the report, a set of selectable user interface elements associated with the first set of fields (as further described with reference to FIG. 1).

[0104] At step 184, the process 170 includes providing a snippet of the location of the corresponding value in response to user selection of any selectable user interface element. In other words, selection of a user interface element results in provision of a snippet of a portion of the document where the data for the respective field in the set of fields is located (as further described with reference to FIG. 1). Example snippets are illustrated in FIGS. 3 to 6. Example cover sheets with snippet functionality enabled are further provided in Appendices B and C of the provisional application, as incorporated by reference herein in its entirety.

[0105] FIG. 2A shows an example system 200 for changing configuration settings for automated data extraction, validation, and reporting in accordance with implementations of the present disclosure. The system 200 can be used to change the configuration settings that are used for generating the report 150. Configuration settings can include, for example, the rules 138, the keywords 130, the fields 126, or any combination thereof.

[0106] The system 200 includes a computing device 204 and the server 110. The computing device 204 communicates with an application programming interface (API) 201 of the server 110 over a network 206. The server 110 stores the keywords 130, the fields 126, and the rules 138. The rules 138 can include, for example, value formats 212, document formats 216, confidence thresholds 214, error tolerances 218. Operational processes for changing configurations are described with reference to FIG. 2B.

[0107] FIG. 2B is a flow diagram of an example process 250 for changing configuration settings for automated data extraction, validation, and reporting. The process 250 can be used by the server 110 from the system 200 or by another data processing apparatus.

[0108] At step 252, the process 250 includes providing data for a user interface for changing configuration settings applied by a report generating system. For example, referring to FIG. 2A, the server 110 provides user interface data 208 to the computing device 204 over the network 206. The computing device 204 renders the user interface data 208 to present a user interface 222 on a display of the computing device 204. The user 202 can interact with

the user interface elements provided through the user interface 222 to request changes to configuration settings for the server 110.

[0109] The user interface 222 presents options to the user 202 for changing configuration settings of the server 110. The user interface 222 can include selectable icons for requesting a change. In some examples, the user interface 222 includes drop-down menus, toggle buttons, text fields, or any combination thereof.

[0110] In the example of FIG. 2, the user selects a setting of “Confidence Threshold” 224. Several options for the confidence threshold are provided in a drop-down menu. Other user interface elements can be used, such as textbox-based numerical input, toggle buttons, a sliding scale, and other types of elements. In the example of FIG. 2, the user 202 selects a new confidence threshold of 90%.

[0111] At step 254, the process 250 includes receiving data indicating user interaction with the user interface. For example, the user 202 submits the configuration change request 220 by interacting with user interface elements provided in the user interface 222, and the computing device 204 sends the configuration change request 220 to the server 110.

[0112] At step 256, the process 250 includes changing configuration settings from a first set of configuration settings to a second set of configuration settings specified by the user interaction. For example, referring to FIG. 2A, the server 110 receives the configuration change request 220 and implements the requested change(s). As a result, the configuration of the server 110 changes from Configuration C0 at time T0 to Configuration C1 at time T1.

[0113] Example changes to configuration settings can include any of the following: adding keywords 130, removing keywords 130, modifying keywords 130, adding fields 126, removing fields 126, modifying fields 126, changing value formats 212 for field values, changing document formats 216 for the report 150, changing confidence thresholds 214 for field values, and changing error tolerances 218 for field values, changing selection criteria for field values, changing actions to be performed when errors are detected, or any combination thereof. Changing value formats 212 for field values can include, for example, changing a number format, a name format, and/or a measurement unit for field values.

[0114] In an example scenario, a default set of fields 126 includes separate Phone fields for each of: Mobile Phone, Home Phone, and Work Phone. The user 202 requests to replace the three Phone fields with a single Phone field labeled “Phone Number.” Another user can request to remove the Work Phone field and replace it with an Emergency Contact Phone field.

[0115] In some examples, the server 110 has an initial configuration. After receiving the configuration change request 220, the server 110 implements the change, and as a result the server 110 has a new configuration that is applicable to a particular client, context, or other environment. For example, referring to FIG. 2A, the server 110 initially has a configuration C0 at time T0. The server 110 receives the configuration change request 220 from the computing device 104. The configuration change request 220 specifies a change in a confidence threshold from 80% to 90%. After implementing the change request, the server 110 has a configuration C0 at time T1. The configuration C0 includes the updated confidence threshold of 90%. In this way, the configuration settings can be customized according to user preference.

[0116] In some examples, the initial configuration for the server is a default configuration. The default configuration can be a generic configuration, or can be tailored to a particular context. For example, the default configuration can be tailored to a medical context, such that the keywords 130 and fields 126 relate to medical information, and the rules 138 are applicable to the types of information that are to be extracted from medical documents.

[0117] At step 258, the process 250 includes receiving a set of documents. For example, referring to FIG. 1A, the server 110 can receive the set of documents 101 and a request to generate a report 150.

[0118] At step 260, the process 250 includes generating a report from the set of documents using the second set of configuration settings. For example, referring to FIG. 1A, the server 110 generates the report 150 using the fields 126, keywords 130, and rules 138 specified by the second set of configuration settings.

[0119] In some examples, the server 110 can include multiple APIs that are exposed over the Network 206. Each API can be accessed individually to perform one or more data extraction and reporting functions. For example, a first API can be accessed in order to extract fields from a document. A second API can be accessed in order to identify locations of the fields within the document. A third API can be accessed in order to generate snippets at identified locations of documents. A fourth API can be accessed in order to append a document with the generated snippets.

[0120] In some examples, a computing device 204 accesses the first API to extract a first set of fields from a set of documents. The computing device 204 accesses the second API to identify locations of field values for the set of fields within the set of documents. The

computing device 204 accesses the third API to generate snippets showing the location of the extracted values. The computing device 204 accesses the fourth API to append the snippets to a document. The computing device 204 can then generate a customized report using the data obtained through the first, second, third, and fourth APIs. The computing device 204 can generate the customized report with or without uploading the set of documents to the server 110.

[0121] In some examples, the server 110 exposes multiple APIs, with each API being associated with a particular functionality of the system 200. The server 110 can receive, from a computing device 204 over the network 206, a request to access one or more of the multiple APIs. The server 110 can provide the computing device 204 with access to the one or more of the multiple APIs.

[0122] FIGS. 3A to 3F show an example interactive cover sheet 300 of a report generated by the system of FIG. 1A. FIG. 3A shows the example cover sheet 300 with hover functionality disabled. The cover sheet 300 can be, for example, a portable document format (PDF) document. The cover sheet 300 includes a demographics section 302, a payor(s) section 304, an order information section 306, a diagnosis code section 308, and a keyword index section 310.

[0123] The cover sheet 300 includes a selectable button 301 for enabling and disabling hover functionality. When the hover functionality is enabled, hovering the mouse over the cover sheet results in presentation of snippets of documents showing locations at which the associated values were found. As described above, the snippets can be included in layers of a pdf file that are configured to be presented in response to user action such as a mouse cursor being positioned over a field, or a mouse click on a selectable element. In some examples, a single layer includes a single snippet to be presented in response to a user action. In some examples, a single layer includes multiple snippets to be presented simultaneously in response to a user action.

[0124] For example, referring to FIG. 3B, in which hovering is enabled, the mouse cursor 312 hovers over the last name 311 of a medical patient. Hovering of the mouse cursor 312 over the last name 311 causes presentation of a snippet 314. The snippet 314 shows a portion of a document in which the patient's last name was found. The snippet 314 is presented at or near the location of the mouse cursor 312 while the mouse cursor 312 hovers over the last name 311. The snippet 314 covers part of the cover sheet 300. When the mouse cursor 312

moves away from the last name 311, the snippet 314 is no longer visible (e.g., the application closes the snippet and reverts to the visualization of the cover sheet).

[0125] In some examples, different types of user selections result in different actions being performed. In addition to a mouse hover, other types of user selection can include, e.g., a left-button mouse click, a right-button mouse click, a mouse click while holding a keyboard button, or any other type of user selection (e.g., a user or stylus touch as would be the case in the context of mobile or handheld devices).

[0126] As discussed above with reference to FIG. 3B, hovering the mouse cursor 312 over a field value, such as the last name 311, causes temporary presentation of the snippet 314. In some examples, a mouse click on an eyeball icon next to a field causes fixed presentation of one or more snippets. For example, referring to FIG. 3C, eyeball icon 316 is located next to a Date of Birth field. A mouse click on the eyeball icon 316 causes presentation of snippets 318, 320. The snippets 318, 320, show two different document portions in which a date of birth was found for the patient. The snippets 318, 320 remain visible until another user selection of the eyeball icon 316 occurs.

[0127] The cover sheet 300 includes a warning 330 near the top of the cover sheet 300. The warning 330 indicates that two different dates of birth (DOBs) were found in the set of documents. Specifically, “03/10/1950” was found in three instances, and “03/11/1950” was found in one instance. The warning 330 can be presented in a way that is visually distinguishable from other parts of the cover sheet 300. For example, the warning 330 can be in larger text, bolder text, italicized text, a different font, and/or a different font color from other parts of the cover sheet 300.

[0128] The validator 136 can detect and identify multiple different types of errors. An error can include, for example, a record contamination, a data discrepancy, or another type of error. A record contamination can be identified, for example, by detecting that information for multiple patients is included in a single set of documents. A data discrepancy can be identified, for example, by determining that confidences for one or more field values are below threshold confidences.

[0129] In some examples, a visual indication of the error is presented at the top of the cover sheet 160, as shown in FIG. 3B. In some examples, a visual indication of the error is presented at or near a field to which the error applies. For example, the error can be indicated by a warning icon, an exclamation point, a red or yellow icon, red or yellow text, highlighting, bold text, or any combination thereof.

[0130] In some examples, the server 110 can perform an error correction. An error correction can include, for example, adding a field value, removing a field value, modifying a field value, leaving a field empty, or any combination of these. The cover sheet can include a visual indication that a field value was corrected. For example, referring to FIG. 3D, the Zip code field value is displayed along with the text: “(corrected)”. The label of “corrected” indicates that an erroneous field value was replaced with a corrected field value. The corrected field value can be obtained, for example, from the third party server 190 during validation of the candidate field values 134.

[0131] In some examples snippets can be selected in order to demonstrate potential errors. For example, according to the cover sheet 300, the set of documents included four different instances of the patient’s Date of Birth. When the eyeball icon 316 is selected, two snippets 318, 320 are presented that show two out of the four instances. The snippet selector 140 selected two instances that show different dates of birth, such that the snippet 318 shows a Date of Birth of 03/10/1950 and the snippet 320 shows a Date of Birth of 3/11/1950. This enables the user to quickly view the sources of the potential error.

[0132] In some examples, a user selection of a selectable user interface element results in opening a link to the document in which the field value was found. For example, a selection of the eyeball icon 316, a field name, a field value, or a snippet can cause the document to be presented in the same window or a different window. In some examples, selection of a snippet causes the pdf document to automatically scroll to the portion of the document from which the snippet was generated. The eyeball icon 316 (or another type of icon presented near a field) can be configured to achieve different results based on interaction and context. For example, in response to a first user interaction at a first time, a snippet can be presented. In response to a second user interaction that occurs while the snippet is being shown, the report can scroll to the portion of the document that corresponds to the snippet.

[0133] Referring to FIG. 3D, the cover sheet 300 shows a name “De Gracia” in the Last Name field. In this example, the validator 136 determined that “De” was part of the patient’s last name and not the middle name. The cover sheet 300 includes a Phone field 324, which is blank. Selection of an eyeball icon 326 next to the Phone field 324 causes presentation of a snippet from a document that had a “Phone” prompt, but was left blank. This enables the user to verify that the “Phone” prompt on the document was left blank and which may then prompt a separate process to request the entity to provide that missing information.

[0134] In some examples, the server 110 communicates with a third party service (e.g., third party server 190) in order to verify and/or correct addresses and/or other information found in the set of documents. In the example of FIG. 3D, an address lookup was performed for “1500 Fuller Dr, Cedar Hill TX,” and the results indicated that the correct Zip code was 75104 instead of 75105. The cover sheet 300 therefore indicates that the value presented in the Zip field 332 is “corrected.” When the mouse 336 hovers over the value, a snippet 334 is presented. The snippet 334 shows the original address found in the set of documents, which indicates a Zip code of 75105.

[0135] Referring to FIG. 3E, hovering a mouse 346 over the field value for the Height field 341 results in presentation of snippets 344, 348. The snippet 344 shows the patient’s height in feet (ft, ’), inches (in, ”), and meters (m), with the following text: “5’ 9.02” (1.753 m)”. The snippet 348 shows the patient’s height in meters, feet, and inches, with the following text: “1.753 m (5’ 9.02”)”. The validator 136 converts the different height formats of the different candidate field values from the snippets to a standardized format according to the rules 138. In this example, the height values are rounded to the nearest inch and displayed in units of inches, with feet (ft) and inches (in) in parentheses (e.g., 69 in (5 ft, 9 in)).

[0136] Similarly, referring to FIG. 3F, the validator 136 rounds the patient’s weight to the nearest pound and presents the value in pounds (lb), with kilograms in parentheses (e.g., 206 lb (93.3 kg)). The validator 136 also selects the mode value of multiple candidate values for the patient’s weight for inclusion in the cover sheet 300. Specifically, snippet 352 shows a weight of 93.3 kg, snippet 354 shows a weight of 91.9 kg, and snippet 356 shows a weight of 93.3 kg. The validator 136 selects the mode value of 93.3 kg as the field value for the Weight field on the cover sheet 300. In some examples, instead of selecting the mode value of multiple candidate values, the validator 136 can be configured to select a different value such as a median value, a mean value, a maximum value, or a minimum value.

[0137] FIGS. 4A to 4C show another example interactive cover sheet 400 of a report generated by the system of FIG. 1A. The cover sheet 400 includes a warning 430 near the top of the cover sheet 400. The warning 430 indicates that the submission of the set of documents is related to another submission. For example, the submission may be very similar, or exactly the same, as another submission. This may be due to a same set of documents being submitted to the server more than once. By presenting the warning 430 on the cover sheet 400, the user is alerted to the potential duplication of records.

[0138] Duplicate submissions can be detected by the validator 136 according to the rules 138. The validator 136 can compare each set of documents to previous sets of documents received over a specified time period (e.g., the previous twenty-four hours, the previous forty-eight hours).

[0139] Referring to FIG. 4B, the validator 136 normalized the patient's date of birth for inclusion of the date of birth value in the cover sheet 400. Specifically, the validator changed the date of birth from a format of Mon DD, YYYY, as shown in the snippet 404, to MM/DD/YYYY, as presented in the Date of Birth field 410 of the cover sheet 400.

[0140] In FIG. 4B, a mouse hover over an Address field 402 of the cover sheet results in presentation of four snippets 404, 406, 408, 412. The addresses in the four snippets have different formats from each other. For example, the addresses in snippet 404 and snippet 412 have a typed street address and a handwritten city, state, and a five-digit zip code, which are all provided in one line on the document. The address in snippet 406 and snippet 408 are typed, are provided in multiple lines, and have a nine-digit zip code. The street address in snippet 408 shows "Hatfield Place," whereas the street address in the other snippets 404, 406, 412 show "Hatfield Pl."

[0141] The field value extractor 128 extracts the different parts of the address (e.g., street address, city, state, Zip code, and provides the different parts of the address to the validator 136. The validator 136 validates the information, selects a field value for each Address field, and converts the field values to the specified format for the respective field according to the rules 138.

[0142] The cover sheet 400 shows a corrected street address 414. The street address was corrected, from "Hatfield Place" as shown in snippet 408 to "Hatfield Pl," based on performing a postal service address lookup. Additionally, referring to FIG. 4C, the cover sheet 400 shows a corrected Zip code 416. The Zip code found in the set of documents is 10314, and the Zip code as corrected through the address lookup is 10302.

[0143] The cover sheet 400 includes a list of products in an Order Information section 436. A mouse hover over the product "Auto CPAP Machine" results in presentation of a snippet 438 showing a portion of the document where the product was located. The snippet 438 shows that the Auto CPAP Machine was handwritten in the set of documents. The handwritten was converted to searchable text through text processing 122 performed by the server 110.

[0144] FIG. 5 shows an example interactive cover sheet 500 for a report including a set of documents in which information for multiple entities were identified in accordance with implementations of the present disclosure.

[0145] When data for one or more entity identifiers are distinct or do not match across different documents included in the set of documents 101, the validator 136 can determine that there are errors in the set of documents 101 and can generate a blank report with error messages indicating the error. For example, the cover sheet 500 includes a warning 530 near the top of the cover sheet 500. The warning 530 indicates that multiple patients are identified in the set of documents. Detection of multiple patients can occur due to extracting multiple different values for the same fields. For example, the server 110 may determine that records for multiple patients are included in the set of documents based on extracting three or more distinct dates of birth, or three or more distinct patient names from the set of documents. In some examples, the server 110 determines that records for multiple patients are included in the set of documents based on extracting at least two distinct dates of birth, at least two distinct patient names, and at least two distinct patient addresses.

[0146] Due to the inclusion of documents related to multiple patients in the set of documents, the field values are omitted from the cover sheet 500, such that the fields 502 remain blank. Thus, because information for multiple patients was found in the set of documents, the cover sheet 500 is empty, or mostly empty. The purpose for leaving the cover sheet 500 empty is to avoid mixed values from the multiple different patients, which could lead to erroneous medical records and potentially privacy issues from comingling of information of different patients. For example, the cover sheet 500 remains empty in order to avoid having a cover sheet 500 with a name of one patient and a date of birth of another patient.

[0147] In some examples, field value such as Zip code is provided in the cover sheet 500 in order to aid a user in identifying the set of documents that resulted in the error. A mouse selection of an eyeball icon 504 near the Zip code field results in presentation of snippets 506, 508. The snippet 506 shows information for a patient named “Dave Rojas,” and the snippet 508 shows information for a patient named “Brooke Stone.”

[0148] FIGS. 6A to 6D show an example interactive cover sheet 600 generated by the system of FIG. 1A in accordance with implementations of the present disclosure. The cover sheet 600 can be, for example, a document in HTML format, which may be displayed or

provided for display within a browser or another native or web application that can render HTML formats.

[0149] FIG. 6A shows the example interactive cover sheet 600 without any user selection. The cover sheet 600 includes a set of fields and respective field values. For example, the cover sheet 600 includes a Name field 602 with a respective field value of “Brenton.”

[0150] Each field includes a field value, an eyeball icon, an arrow icon, and a confidence indicator for the field value. For example, the Street Address field 608 includes a field value of “170 Hatfield Pl,” an eyeball icon 604, an arrow icon 606, and a confidence indicator 610. The confidence indicator 610 indicates a confidence for the field value of “170 Hatfield Pl.”

[0151] A user selection (e.g., a mouse hover, a mouse click) of the confidence indicator 610 results in presentation of a confidence value for the field value. For example, referring to FIG. 6B, a mouse hover over the confidence indicator 612 results in presentation of a confidence value 614 of 99.5. The confidence indicator 612 is a checkmark, which indicates higher confidence (e.g., a confidence value that satisfies a threshold confidence value). In the example of FIG. 6B, the threshold confidence is 90.0, such that values greater than or equal to 90.0 are presented with a checkmark, and values less than 90.0 are presented with a warning sign.

[0152] A mouse hover over the confidence indicator 622 results in presentation of a confidence value 624 of 87.0. The confidence indicator 622 is a warning sign, which indicates lower confidence (e.g., a confidence value that does not satisfy the threshold confidence).

[0153] In some examples, confidence indicators can be presented in different colors according to the respective confidence values. For example, for confidence values between 80.0 and 90.0, the confidence indicator may be a yellow warning sign. For confidence values less than 80.0, the confidence indicator may be a red warning sign. For confidence values greater than 90.0, the confidence indicator may be a green checkmark.

[0154] A user selection of an eyeball icon results in presentation of a snippet of a document where the field value is located. For example, referring to FIG. 6C, selection of the eyeball icon 632 results in presentation of a snippet 634 showing a location of a document from which a Height field value was extracted. The confidence indicator 636 for the Height field is a warning sign, indicating a lower confidence. In this example, the lower confidence

value may be due to a poor resolution and/or clarity of the document from which the Height field value was extracted.

[0155] A user selection of an arrow icon results in opening a document from which the snippet was generated. This can be implemented in several ways. In some examples, a web page presenting the cover sheet 600 is modified to resize the cover sheet 600 to fit in a side portion (e.g., a left portion) of the web page. The underlying document is then presented in a frame on the opposite side portion (e.g., a right portion) of the web page. In some examples, a new window is launched that includes the underlying document scrolled to the appropriate location of the document from which the snippet was generated. For example, referring to FIG. 6D, selection of the arrow icon 642 results in opening a document 644 in a window 640 next to the cover sheet 600.

[0156] A mouse hover over the eyeball icon 652 results in presentation of a snippet 654. As shown in FIG. 6D, the snippet 654 includes a portion of the document 644 from which the field value of “Auto CPAP Machine” was extracted. Keywords and field values are highlighted in the document 644 presented in the window 640. For example, the words “Tubing, Filter” 656 are highlighted in the document 644 because they are included as field values 658 in the “Products” section of the cover sheet 600.

[0157] Other example cover sheets are provided in Appendices B and E of the provisional application, as incorporated by reference herein in their entirety.

[0158] Although the above description is provided as a series of steps, not all the steps of the process need to be performed (for example, the parsing for keyword identification or data extraction steps need not be performed), nor do these steps need to be performed in the order presented above. Moreover, other embodiments of this aspect include corresponding methods, apparatus, and computer programs, configured to perform the actions of the methods, encoded on computer storage devices.

[0159] It is well understood that the use of personally identifiable information should follow privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining the privacy of users. In particular, personally identifiable information data should be managed and handled so as to minimize risks of unintentional or unauthorized access or use, and the nature of authorized use should be clearly indicated to users.

[0160] FIG. 7 is a diagram illustrating an example of a computing system, e.g., used for automated data extraction, validation, and reporting. The computing system includes

computing device 700 and a mobile computing device 750 that can be used to implement the techniques described herein. For example, one or more components of the system 100 or the system 200 could be an example of the computing device 700 or the mobile computing device 750.

[0161] The computing device 700 is intended to represent various forms of digital computers, such as laptops, desktops, workstations, personal digital assistants, servers, blade servers, mainframes, and other appropriate computers. The mobile computing device 750 is intended to represent various forms of mobile devices, such as personal digital assistants, cellular telephones, smart-phones, mobile embedded radio systems, radio diagnostic computing devices, and other similar computing devices. The components shown here, their connections and relationships, and their functions, are meant to be examples only, and are not meant to be limiting.

[0162] The computing device 700 includes a processor 702, a memory 704, a storage device 706, a high-speed interface 708 connecting to the memory 704 and multiple high-speed expansion ports 710, and a low-speed interface 712 connecting to a low-speed expansion port 714 and the storage device 706. Each of the processor 702, the memory 704, the storage device 706, the high-speed interface 708, the high-speed expansion ports 710, and the low-speed interface 712, are interconnected using various busses, and may be mounted on a common motherboard or in other manners as appropriate. The processor 702 can process instructions for execution within the computing device 700, including instructions stored in the memory 704 or on the storage device 706 to display graphical information for a GUI on an external input/output device, such as a display 716 coupled to the high-speed interface 708. In other implementations, multiple processors and/or multiple buses may be used, as appropriate, along with multiple memories and types of memory. In addition, multiple computing devices may be connected, with each device providing portions of the operations (e.g., as a server bank, a group of blade servers, or a multi-processor system). In some implementations, the processor 702 is a single threaded processor. In some implementations, the processor 702 is a multi-threaded processor. In some implementations, the processor 702 is a quantum computer.

[0163] The memory 704 stores information within the computing device 700. In some implementations, the memory 704 is a volatile memory unit or units. In some implementations, the memory 704 is a non-volatile memory unit or units. The memory 704 may also be another form of computer-readable medium, such as a magnetic or optical disk.

[0164] The storage device 706 is capable of providing mass storage for the computing device 700. In some implementations, the storage device 706 may be or include a computer-readable medium, such as a floppy disk device, a hard disk device, an optical disk device, or a tape device, a flash memory or other similar solid-state memory device, or an array of devices, including devices in a storage area network or other configurations. Instructions can be stored in an information carrier. The instructions, when executed by one or more processing devices (for example, processor 702), perform one or more methods, such as those described above. The instructions can also be stored by one or more storage devices such as computer- or machine-readable mediums (for example, the memory 704, the storage device 706, or memory on the processor 702). The high-speed interface 708 manages bandwidth-intensive operations for the computing device 700, while the low-speed interface 712 manages lower bandwidth-intensive operations. Such allocation of functions is an example only. In some implementations, the high-speed interface 708 is coupled to the memory 704, the display 716 (e.g., through a graphics processor or accelerator), and to the high-speed expansion ports 710, which may accept various expansion cards (not shown). In the implementation, the low-speed interface 712 is coupled to the storage device 706 and the low-speed expansion port 714. The low-speed expansion port 714, which may include various communication ports (e.g., USB, Bluetooth, Ethernet, wireless Ethernet) may be coupled to one or more input/output devices, such as a keyboard, a pointing device, a scanner, or a networking device such as a switch or router, e.g., through a network adapter.

[0165] The computing device 700 may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a standard server 720, or multiple times in a group of such servers. In addition, it may be implemented in a personal computer such as a laptop computer 722. It may also be implemented as part of a rack server system 724. Alternatively, components from the computing device 700 may be combined with other components in a mobile device, such as a mobile computing device 750. Each of such devices may include one or more of the computing device 700 and the mobile computing device 750, and an entire system may be made up of multiple computing devices communicating with each other.

[0166] The mobile computing device 750 includes a processor 752, a memory 764, an input/output device such as a display 754, a communication interface 766, and a transceiver 768, among other components. The mobile computing device 750 may also be provided with a storage device, such as a micro-drive or other device, to provide additional storage. Each of

the processor 752, the memory 764, the display 754, the communication interface 766, and the transceiver 768, are interconnected using various buses, and several of the components may be mounted on a common motherboard or in other manners as appropriate.

[0167] The processor 752 can execute instructions within the mobile computing device 750, including instructions stored in the memory 764. The processor 752 may be implemented as a chipset of chips that include separate and multiple analog and digital processors. The processor 752 may provide, for example, for coordination of the other components of the mobile computing device 750, such as control of user interfaces, applications run by the mobile computing device 750, and wireless communication by the mobile computing device 750.

[0168] The processor 752 may communicate with a user through a control interface 758 and a display interface 756 coupled to the display 754. The display 754 may be, for example, a TFT (Thin-Film-Transistor Liquid Crystal Display) display or an OLED (Organic Light Emitting Diode) display, or other appropriate display technology. The display interface 756 may include appropriate circuitry for driving the display 754 to present graphical and other information to a user. The control interface 758 may receive commands from a user and convert them for submission to the processor 752. In addition, an external interface 762 may provide communication with the processor 752, so as to enable near area communication of the mobile computing device 750 with other devices. The external interface 762 may provide, for example, for wired communication in some implementations, or for wireless communication in other implementations, and multiple interfaces may also be used.

[0169] The memory 764 stores information within the mobile computing device 750. The memory 764 can be implemented as one or more of a computer-readable medium or media, a volatile memory unit or units, or a non-volatile memory unit or units. An expansion memory 774 may also be provided and connected to the mobile computing device 750 through an expansion interface 772, which may include, for example, a SIMM (Single In Line Memory Module) card interface. The expansion memory 774 may provide extra storage space for the mobile computing device 750, or may also store applications or other information for the mobile computing device 750. Specifically, the expansion memory 774 may include instructions to carry out or supplement the processes described above, and may include secure information also. Thus, for example, the expansion memory 774 may be provide as a security module for the mobile computing device 750, and may be programmed with instructions that permit secure use of the mobile computing device 750. In addition, secure

applications may be provided via the SIMM cards, along with additional information, such as placing identifying information on the SIMM card in a non-hackable manner.

[0170] The memory may include, for example, flash memory and/or NVRAM memory (nonvolatile random access memory), as discussed below. In some implementations, instructions are stored in an information carrier such that the instructions, when executed by one or more processing devices (for example, processor 752), perform one or more methods, such as those described above. The instructions can also be stored by one or more storage devices, such as one or more computer- or machine-readable mediums (for example, the memory 764, the expansion memory 774, or memory on the processor 752). In some implementations, the instructions can be received in a propagated signal, for example, over the transceiver 768 or the external interface 762.

[0171] The mobile computing device 750 may communicate wirelessly through the communication interface 766, which may include digital signal processing circuitry in some cases. The communication interface 766 may provide for communications under various modes or protocols, such as GSM voice calls (Global System for Mobile communications), SMS (Short Message Service), EMS (Enhanced Messaging Service), or MMS messaging (Multimedia Messaging Service), CDMA (code division multiple access), TDMA (time division multiple access), PDC (Personal Digital Cellular), WCDMA (Wideband Code Division Multiple Access), CDMA2000, or GPRS (General Packet Radio Service), LTE, 6G/6G cellular, among others. Such communication may occur, for example, through the transceiver 768 using a radio frequency. In addition, short-range communication may occur, such as using a Bluetooth, Wi-Fi, or other such transceiver (not shown). In addition, a GPS (Global Positioning System) receiver module 770 may provide additional navigation- and location-related wireless data to the mobile computing device 750, which may be used as appropriate by applications running on the mobile computing device 750.

[0172] The mobile computing device 750 may also communicate audibly using an audio codec 760, which may receive spoken information from a user and convert it to usable digital information. The audio codec 760 may likewise generate audible sound for a user, such as through a speaker, e.g., in a handset of the mobile computing device 750. Such sound may include sound from voice telephone calls, may include recorded sound (e.g., voice messages, music files, among others) and may also include sound generated by applications operating on the mobile computing device 750.

[0173] The mobile computing device 750 may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a cellular telephone 780. It may also be implemented as part of a smart-phone 782, personal digital assistant, or other similar mobile device.

[0174] In general, use of “or” can refer to “and/or.” When providing a list of two or more items, the conjunction “or” can indicate any one of the items, any combination of a subset of the items, or all items in combination.

[0175] In this specification the term “engine” is used broadly to refer to a software-based system, subsystem, or process that is programmed to perform one or more specific functions. Generally, an engine will be implemented as one or more software modules or components, installed on one or more computers in one or more locations. In some cases, one or more computers will be dedicated to a particular engine; in other cases, multiple engines can be installed and running on the same computer or computers.

[0176] The subject matter and the actions and operations described in this specification can be implemented in digital electronic circuitry, in tangibly-embodied computer software or firmware, in computer hardware, including the structures disclosed in this specification and their structural equivalents, or in combinations of one or more of them. The subject matter and the actions and operations described in this specification can be implemented as or in one or more computer programs, e.g., one or more modules of computer program instructions, encoded on a computer program carrier, for execution by, or to control the operation of, data processing apparatus. The carrier can be a tangible non-transitory computer storage medium. Alternatively or in addition, the carrier can be an artificially-generated propagated signal, e.g., a machine-generated electrical, optical, or electromagnetic signal, that is generated to encode information for transmission to suitable receiver apparatus for execution by a data processing apparatus. The computer storage medium can be or be part of a machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device, or a combination of one or more of them. A computer storage medium is not a propagated signal.

[0177] The term “data processing apparatus” encompasses all kinds of apparatus, devices, and machines for processing data, including by way of example a programmable processor, a computer, or multiple processors or computers. Data processing apparatus can include special-purpose logic circuitry, e.g., an FPGA (field programmable gate array), an ASIC (application-specific integrated circuit), or a GPU (graphics processing unit). The apparatus

can also include, in addition to hardware, code that creates an execution environment for computer programs, e.g., code that constitutes processor firmware, a protocol stack, a database management system, an operating system, or a combination of one or more of them.

[0178] A computer program can be written in any form of programming language, including compiled or interpreted languages, or declarative or procedural languages; and it can be deployed in any form, including as a stand-alone program, e.g., as an app, or as a module, component, engine, subroutine, or other unit suitable for executing in a computing environment, which environment may include one or more computers interconnected by a data communication network in one or more locations.

[0179] A computer program may, but need not, correspond to a file in a file system. A computer program can be stored in a portion of a file that holds other programs or data, e.g., one or more scripts stored in a markup language document, in a single file dedicated to the program in question, or in multiple coordinated files, e.g., files that store one or more modules, sub-programs, or portions of code.

[0180] The processes and logic flows described in this specification can be performed by one or more computers executing one or more computer programs to perform operations by operating on input data and generating output. The processes and logic flows can also be performed by special-purpose logic circuitry, e.g., an FPGA, an ASIC, or a GPU, or by a combination of special-purpose logic circuitry and one or more programmed computers.

[0181] Computers suitable for the execution of a computer program can be based on general or special-purpose microprocessors or both, or any other kind of central processing unit. Generally, a central processing unit will receive instructions and data from a read-only memory or a random access memory or both. The essential elements of a computer are a central processing unit for executing instructions and one or more memory devices for storing instructions and data. The central processing unit and the memory can be supplemented by, or incorporated in, special-purpose logic circuitry.

[0182] Generally, a computer will also include, or be operatively coupled to, one or more mass storage devices, and be configured to receive data from or transfer data to the mass storage devices. The mass storage devices can be, for example, magnetic, magneto-optical, or optical disks, or solid state drives. However, a computer need not have such devices. Moreover, a computer can be embedded in another device, e.g., a mobile telephone, a personal digital assistant (PDA), a mobile audio or video player, a game console, a Global

Positioning System (GPS) receiver, or a portable storage device, e.g., a universal serial bus (USB) flash drive, to name just a few.

[0183] To provide for interaction with a user, the subject matter described in this specification can be implemented on one or more computers having, or configured to communicate with, a display device, e.g., a LCD (liquid crystal display) monitor, or a virtual-reality (VR) or augmented-reality (AR) display, for displaying information to the user, and an input device by which the user can provide input to the computer, e.g., a keyboard and a pointing device, e.g., a mouse, a trackball or touchpad. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback and responses provided to the user can be any form of sensory feedback, e.g., visual, auditory, speech, or tactile feedback or responses; and input from the user can be received in any form, including acoustic, speech, tactile, or eye tracking input, including touch motion or gestures, or kinetic motion or gestures or orientation motion or gestures. In addition, a computer can interact with a user by sending documents to and receiving documents from a device that is used by the user; for example, by sending web pages to a web browser on a user's device in response to requests received from the web browser, or by interacting with an app running on a user device, e.g., a smartphone or electronic tablet. Also, a computer can interact with a user by sending text messages or other forms of message to a personal device, e.g., a smartphone that is running a messaging application, and receiving responsive messages from the user in return.

[0184] This specification uses the term “configured to” in connection with systems, apparatus, and computer program components. That a system of one or more computers is configured to perform particular operations or actions means that the system has installed on it software, firmware, hardware, or a combination of them that in operation cause the system to perform the operations or actions. That one or more computer programs is configured to perform particular operations or actions means that the one or more programs include instructions that, when executed by data processing apparatus, cause the apparatus to perform the operations or actions. That special-purpose logic circuitry is configured to perform particular operations or actions means that the circuitry has electronic logic that performs the operations or actions.

[0185] The subject matter described in this specification can be implemented in a computing system that includes a back-end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front-end component,

e.g., a client computer having a graphical user interface, a web browser, or an app through which a user can interact with an implementation of the subject matter described in this specification, or any combination of one or more such back-end, middleware, or front-end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (LAN) and a wide area network (WAN), e.g., the Internet.

[0186] The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other. In some implementations, a server transmits data, e.g., an HTML page, to a user device, e.g., for purposes of displaying data to and receiving user input from a user interacting with the device, which acts as a client. Data generated at the user device, e.g., a result of the user interaction, can be received at the server from the device.

[0187] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of what is being claimed, which is defined by the claims themselves, but rather as descriptions of features that may be specific to particular embodiments of particular inventions. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially be claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claim may be directed to a subcombination or variation of a subcombination.

[0188] Similarly, while operations are depicted in the drawings and recited in the claims in a particular order, this by itself should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described

program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0189] Particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. For example, the actions recited in the claims can be performed in a different order and still achieve desirable results. As one example, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In some cases, multitasking and parallel processing may be advantageous.

CLAIMS

1. A computer-implemented method, comprising:
 - receiving, from a client device, a set of documents that are associated with a first entity;
 - converting the set of documents into a searchable set of documents;
 - extracting, from the searchable set of documents, values for a set of fields that relate to the first entity;
 - identifying locations of the values within the searchable set of documents;
 - generating a report including the set of fields and the respective values;
 - appending, to the report, a set of selectable user interface elements associated with the set of fields, wherein user selection of a selectable user interface element associated with a first field results in provision of a first snippet of a first document, wherein the first snippet shows a portion of the first document including the location of the value for the first field;
 - and
 - providing, for display on the client device, the report and the appended set of selectable user interface elements associated with the set of fields.
2. The computer-implemented method of claim 1, wherein extracting, from the searchable set of documents, the values for the set of fields that relate to the first entity comprises, for the first field:
 - identifying, within the searchable set of documents, multiple instances of candidate values for the first field;
 - selecting, from the multiple instances of candidate values, a first instance of a first candidate value for the first field; and
 - assigning the first instance of the first candidate value as the value for the first field, wherein the portion of the first document includes the location of the first instance of the first candidate value.
3. The computer-implemented method of claim 2, comprising:
 - determining a confidence for each of the multiple instances of candidate values for the first field; and
 - selecting the first instance of the first candidate value for the first field based on the confidence for each of the multiple instances of candidate values.

4. The computer-implemented method of claim 3, wherein the confidence of an instance of a candidate value is based at least in part on:
 - a text recognition confidence for the instance; or
 - a similarity of the instance of the candidate value to other instances of candidate values.
5. The computer-implemented method of claim 3, comprising:
 - determining whether the confidence for the selected instance satisfies a threshold confidence; and
 - providing, for inclusion in the report, a visual indication of whether the confidence satisfies the threshold confidence.
6. The computer-implemented method of claim 1, wherein extracting, from the searchable set of documents, the values for the set of fields that relate to the first entity comprises, for the first field:
 - identifying, within the searchable set of documents, multiple instances of candidate values for the first field;
 - identifying, from the multiple instances of candidate values, a mode candidate value;
 - and
 - assigning the identified mode candidate value as the value for the first field.
7. The computer-implemented method of claim 1, wherein generating the report including the set of fields and the respective values comprises, for a particular field:
 - determining, using a set of rules, a specified value format for the particular field; and
 - converting the respective value for the particular field to the specified value format.
8. The computer-implemented method of claim 1, wherein generating the report including the set of fields and the respective values comprises, for a particular field:
 - determining, using a set of rules, a specified unit of measurement for the particular field; and
 - converting the respective value for the particular field to the specified unit of measurement.

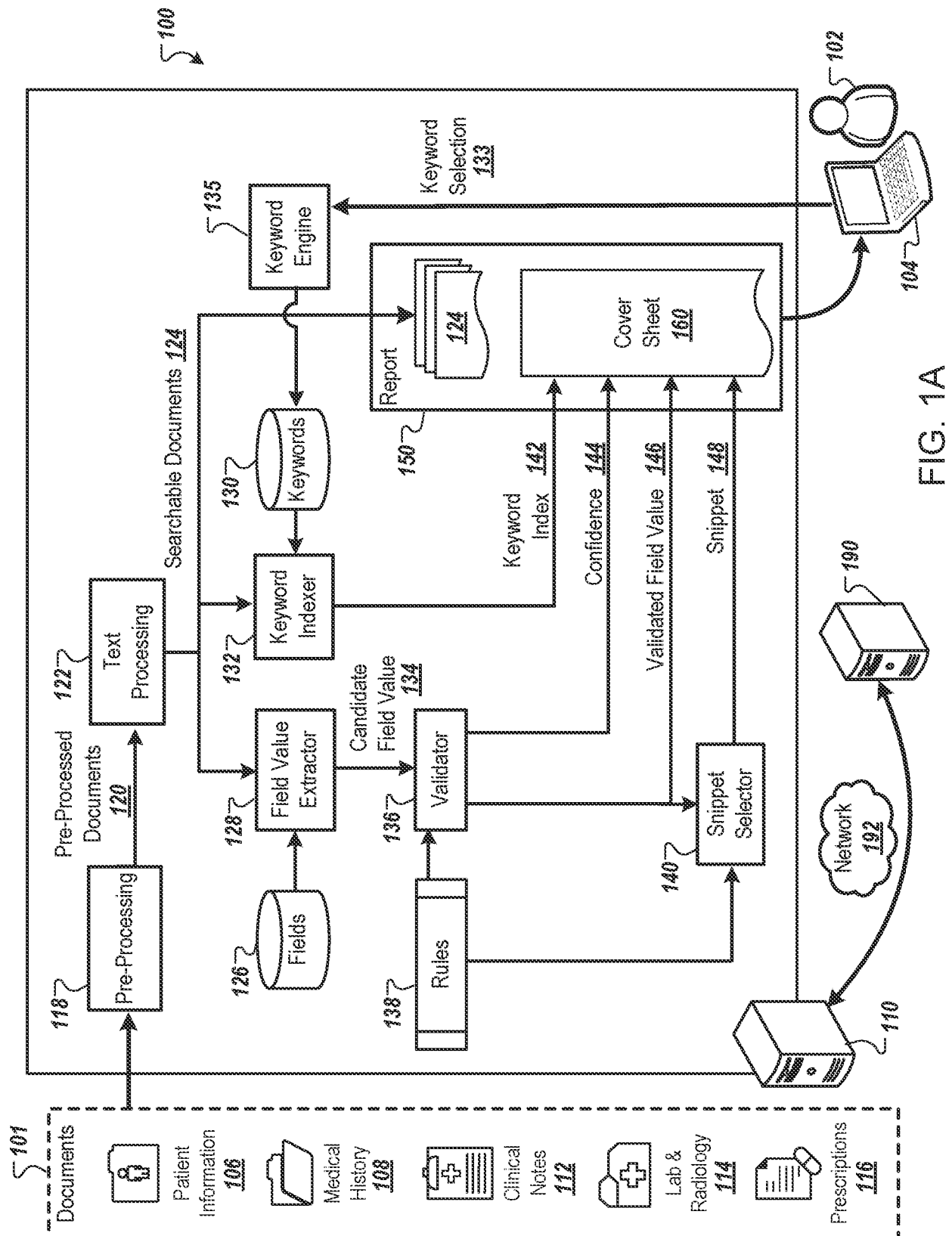
9. The computer-implemented method of claim 2, wherein:
the multiple instances of candidate values includes a second instance of the first candidate value for the first field, and
user selection of the selectable user interface element results in provision of the first snippet and a second snippet, wherein the second snippet shows a portion of a document including the location of the second instance of the first candidate value for the first field.
10. The computer-implemented method of claim 1, comprising generating the first snippet, including:
determining, using a set of rules indicating snippet sizes for the set of fields, a specified snippet size of the first field; and
selecting the portion of the first document based on the location of the value for the first field and the snippet size determined using the set of rules.
11. The computer-implemented method of claim 1, wherein:
the set of fields, the respective extracted values, and the set of selectable user interface elements are provided in a cover sheet of the report, and
the report further includes the set of documents.
12. The computer-implemented method of claim 1, wherein generating the report including the set of fields and the respective values comprises:
identifying an error in the set of documents; and
providing, for inclusion in the report, a visual indication of the error.
13. The computer-implemented method of claim 12, comprising:
in response to identifying the error, performing an error correction; and
providing, for display on the client device, the report with the error correction, wherein the error correction comprises at least one of:
omitting a value for at least one field of the set of fields;
adding a value for at least one field of the set of fields; or
modifying a value for at least one field of the set of fields.

14. The computer-implemented method of claim 1, wherein:
the user selection of the selectable user interface element comprises a first type of user selection; and
a second type of user selection of the selectable user interface element results in opening a link to the first document.
15. A computer-implemented method, comprising:
providing, to a client device, application data for a user interface of an application, the user interface comprising one or more user interface elements for changing configuration settings of a system for generating a report for a set of documents, wherein the configuration settings include:
a set of fields for inclusion in the report;
selection criteria for values for the set of fields; and
formatting rules for the values for the set of fields;
receiving, from the client device, data indicating user interaction with the one or more user interface elements; and
in response to receiving the data indicating the user interaction with the one or more user interface elements, changing the configuration settings from a first set of configuration settings to a second set of configuration settings specified by the user interaction.
16. The computer-implemented method of claim 15, wherein the configuration settings include at least one of:
a set of keywords for indexing in the report;
actions to be performed in response to detecting the error in the set of documents;
a document format for the report.
17. The computer-implemented method of claim 15, wherein:
the selection criteria for the values for the set of fields includes at least one of a threshold confidence or an error tolerance; and
the formatting rules for the values for the set of fields includes at least one of a number format, a name format, or a measurement unit.

18. The computer-implemented method of claim 15, comprising:
- receiving, from a client device, a set of documents that are associated with a first entity;
 - converting the set of documents into a searchable set of documents;
 - extracting, from the searchable set of documents, values for the set of fields included in the second set of configuration settings based on the selection criteria included in the second set of configuration settings;
 - identifying locations of the values within the searchable set of documents;
 - converting the values to respective formats specified by the formatting rules included in the second set of configuration settings; and
 - generating a report including the set of fields and the respective values converted to the respective formats specified by the formatting rules in the second set of configuration settings.
19. The computer-implemented method of claim 18, comprising:
- appending, to the report, a set of selectable user interface elements associated with the set of fields, wherein user selection of a selectable user interface element associated with a first field results in provision of a first snippet of a first document, wherein the first snippet shows a portion of the first document including the location of the value for the first field;
 - and
 - providing, for display on the client device, the report and the appended set of selectable user interface elements associated with the set of fields.
20. A system, comprising:
- at least one processor; and
 - a data store coupled to the at least one processor having instructions stored thereon which, when executed by the at least one processor, causes the at least one processor to perform operations comprising:
 - receiving, from a client device, a set of documents that are associated with a first entity;
 - converting the set of documents into a searchable set of documents;
 - extracting, from the searchable set of documents, values for a set of fields that relate to the first entity;

identifying locations of the values within the searchable set of documents;
generating a report including the set of fields and the respective values;
appending, to the report, a set of selectable user interface elements associated with the set of fields, wherein user selection of a selectable user interface element associated with a first field results in provision of a first snippet of a first document, wherein the first snippet shows a portion of the first document including the location of the value for the first field; and

providing, for display on the client device, the report and the appended set of selectable user interface elements associated with the set of fields.



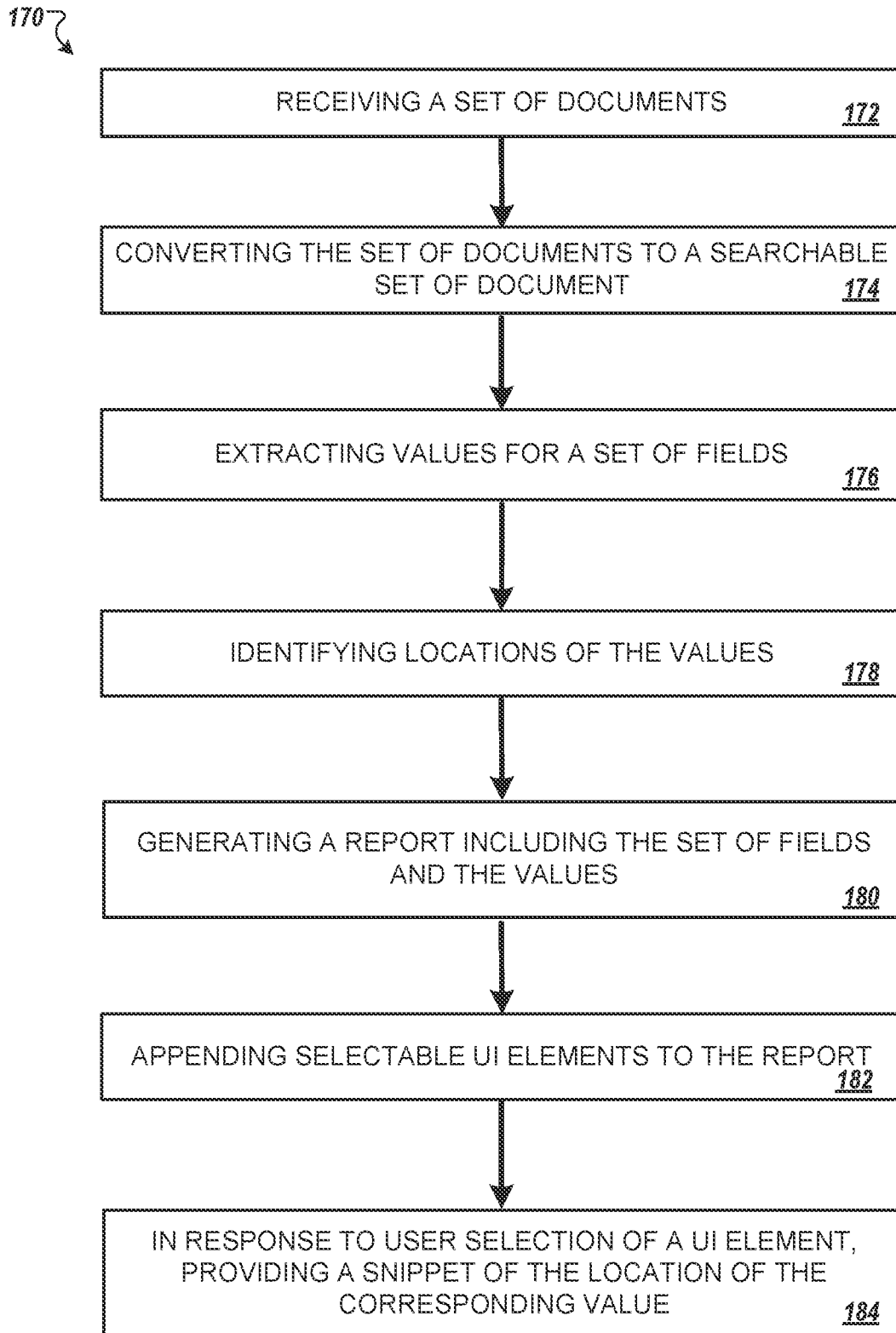


FIG. 1B

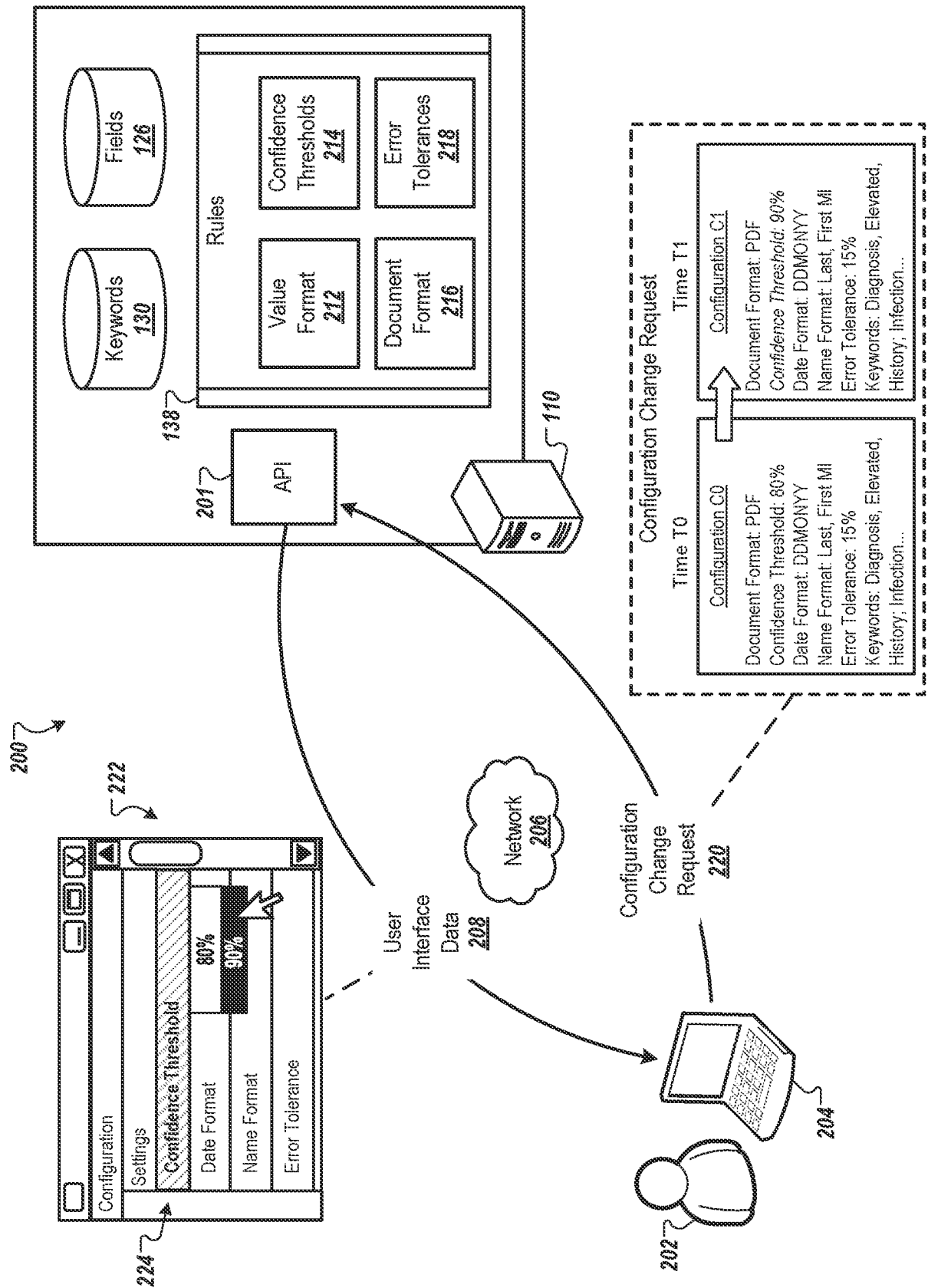


FIG. 2A

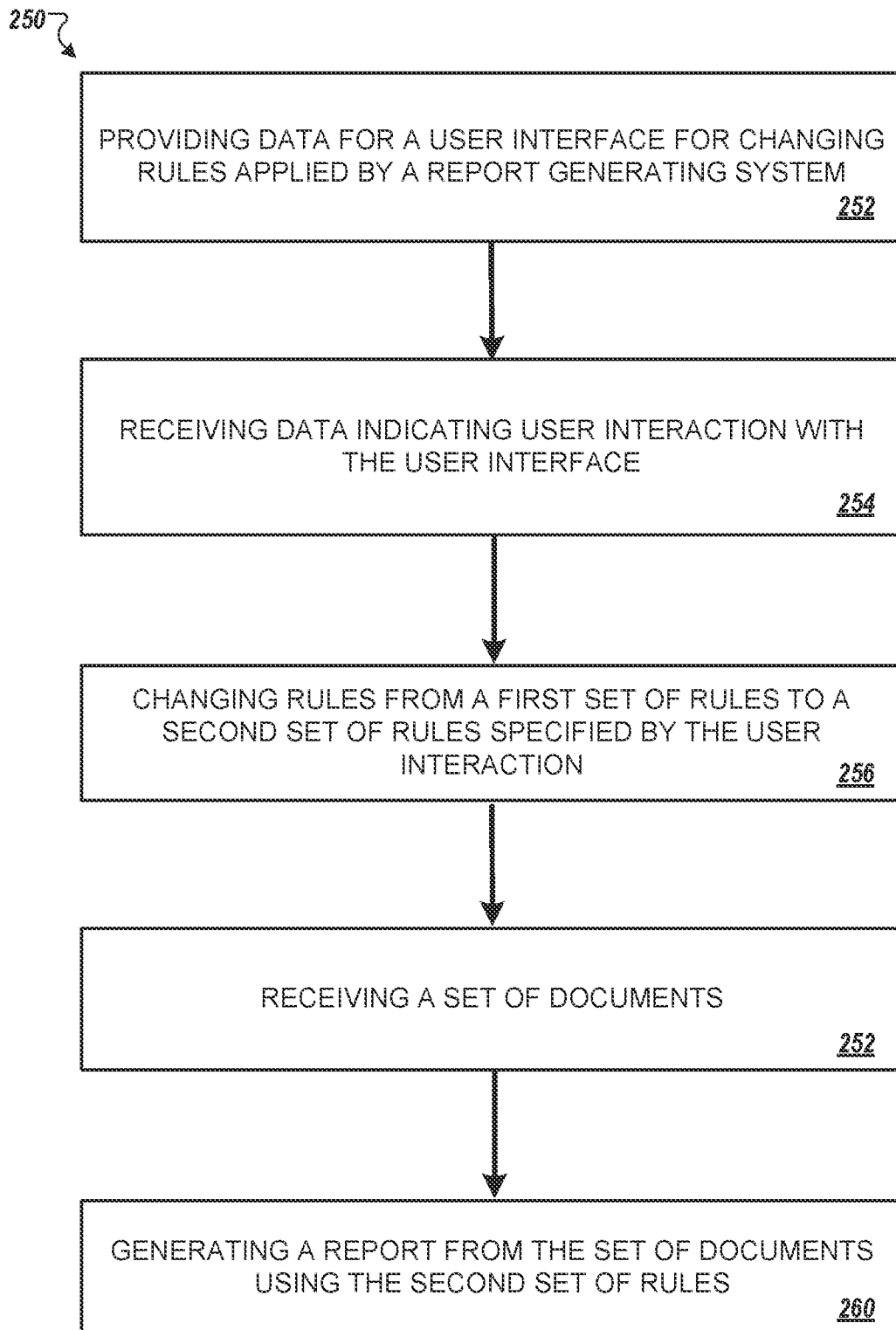


FIG. 2B

5/19

300

301

A1001001A23J06B35807D30489 Received 2023-10-23 at 13:01 PDT Enable Hover

34-page application/pdf

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

Demographics

302

◊ First Name: Robert	◊ Street Address: 1500 Fuller Dr
◊ Last Name: De Gracia	◊ Street Address 2:
◊ Date of Birth: 03/10/1950	◊ City: Cedar Hill
◊ Phone:	◊ State: TX
◊ Email: rob.degracia50@gmail.com	◊ Zip: 75104 (corrected)
◊ Height: 69 in (5 ft, 9 in)	◊ Weight: 260 lb (93.3 kg)

Payor(s)

304

◊ Primary Ins: MEDICARE	◊ Primary Ins ID: XUM0TNX4SS
◊ Secondary Ins: BCBS	◊ Secondary Ins ID: J9S8GQDPGJ

Order Info

306

◊ Product 1: Bedside Commode	◊ Product 2: wheelchair
◊ Physician: Osman, Houssam Galal Eldin	◊ Primary Diagnosis:
◊ Referral: Methodist Richardson Medical Center	

All Diagnosis Codes

308

⌘ P3 — K86.89

Keywords

310

Authorization --

(2) Auth p5 p12	(18) Electronically p4 p5 p10 p11 p12 p18 p23 p28 p29
(1) Reviewed by · p18	(3) Signed By · · · p4 p11 p23

DME --

(7) Bed p6 p15 p16 p17 p21 p33	(5) Commode · p11 p12 p21 p30 p33
(3) Elevated · · p10 p21 p33	(2) Manual · · · p30
(24) Mobility · · · p14 p15 p16 p17 p18 p20 p21 p22 p30 p32 p33 p34	(3) Walker · · · p17 p23 p35
(2) Walking · · · p22 p34	(4) Wheelchair p4 p5 p30

FIG. 3A

6/19

300 ↘

A1001001A23J06B35807D30489

Received 2023-10-23 at 13:01 PDT [Disable Hover](#)
34-page application/pdf

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

Demographics

- ◇ First Name: Robert 311
- ◇ Last Name: De Gracia 312 314
- ◇

Patient Name: Robert De Gracia
Today's Date: 10/4/2023
- ◇ Street Address: 1500 Fuller Dr
- ◇ Street Address 2:
- ◇ Email: rob.degracia50@gmail.com
- ◇ Zip: 75104 (corrected)
- ◇ Height: 69 in (5 ft, 9 in)
- ◇ Weight: 260 lb (93.3 kg)

Payor(s)

- ◇ Primary Ins: MEDICARE
- ◇ Primary Ins ID: XUM0TNX4SS
- ◇ Secondary Ins: BCBS
- ◇ Secondary Ins ID: J9S8GQDPGJ

Order Info

- ◇ Product 1: Bedside Commode
- ◇ Product 2: wheelchair
- ◇ Physician: Osman, Houssam Galal Eldin
- ◇ Primary Diagnosis:
- ◇ Referral: Methodist Richardson Medical Center

⑥ All Diagnosis Codes

P3 — K86.89

Keywords

FIG. 3B

7/19

300

A1001001A23J06B35807D30489

Received 2023-10-23 at 13:01 PDT
34-page application/pdf

Disable Hover

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

330

Demographics

◊ First Name: Robert

◊ Street Address: 1500 Fuller Dr

◊ Last Name: De Gracia

◊ Street Address 2:

◊ Date of Birth: 03/10/1950

◊ City: Cedar Hill

Gracia
DOB: 03/10/1950 Sex: M

◊ State: TX

◊ /C: —

◊ Zip: 75104 (corrected)

me
Gracia Date of Birth
3/11/1950

◊ Weight: 260 lb (93.3 kg)

Payor(s)

◊ Primary Ins: MEDICARE

◊ Primary Ins ID: XUM0TNX4SS

◊ Secondary Ins: BCBS

◊ Secondary Ins ID: J9S8GQDPGJ

Order Info

◊ Product 1: Bedside Commode

◊ Product 2: wheelchair

◊ Physician: Osman, Houssam Galal Eldin

Primary Diagnosis:

◊ Referral: Methodist Richardson Medical
Center

All Diagnosis Codes

P3 — K86.89

Keywords

FIG. 3C

8/19

300

A1001001A23J06B35807D30489 Received 2023-10-23 at 13:01 PDT Disable Hover
34-page application/pdf

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

Demographics

- 326

 - First Name: Robert
 - Last Name: De Gracia 322
 - Date of Birth: 03/10/1950
 - Phone: 324
 - Email: rob.degracia50@gmail.com
 - Height: 69 in (5 ft, 9 in)
- Street Address: 1500 Fuller Dr
 - Street Address 2:
 - City: Cedar Hill
 - State: TX
 - Zip: 75104 (corrected) 332
 - W 334 336

Payor(s)

- Primary Ins: MEDICARE
 - Secondary Ins: BCBS
- Address
1500 FULLER DR
CEDAR HILL TX 75105-4210

Order Info

- Product 1: Bedside Commode
 - Physician: Osman, Houssam Galal Eldin
 - Referral: Methodist Richardson Medical Center
- PCP and Center
Primary Care Provider
John D. Gabriel, MD

Primary Diagnosis:

All Diagnosis Codes

P3 — K86.89

Keywords

FIG. 3D

9/19

300

A1001001A23J06B35807D30489

Received 2023-10-23 at 13:01 PDT
34-page application/pdf

Disable Hover

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

Demographics

- ◊ First Name: Robert
- ◊ Last Name: De Gracia
- ◊ Date of Birth: 03/10/1950
- ◊ Phone:
- ◊ Email: rob.degracia50@gmail.com
- ◊ Height: 69 in (5 ft, 9 in)
- ◊ Street Address: 1500 Fuller Dr
- ◊ Street Address 2:
- ◊ City: Cedar Hill
- ◊ State: TX
- ◊ Zip: 75104 (corrected)
- ◊ Weight: 260 lb (93.3 kg)

341

Ht: 5' 9.02" (1.753 m)
Wt: 205 lb 11 oz (93.3 kg)

Payor(s)

◊ Primary Ins ID: XUMOTNX4SS

BP (I) 146/60 (BP Location: Left arm, Patient Position: Lying)
Resp 16 | Ht 1.753 m (5' 9.02") | Wt 93.3 kg (205 lb 11 oz) |

348

- ◊ Physician: Osman, Houssam Galal Eldin
- ◊ Referral: Methodist Richardson Medical Center
- Primary Diagnosis:

All Diagnosis Codes

P3 — K86.89

Keywords

FIG. 3E

10/19

300

A1001001A23J06B35807D30489

Received 2023-10-23 at 13:01 PDT [Disable Hover](#)
34-page application/pdf

2 DOBs: 3x 03/10/1950 | 1x 03/11/1950

Demographics

- ◇ First Name: Robert
- ◇ Last Name: De Gracia
- ◇ Date of Birth: 03/10/1950
- ◇ Phone:
- ◇ Email: rob.degracia50@gmail.com
- ◇ Height: 69 in (5 ft, 9 in)
- ◇ Street Address: 1500 Fuller Dr
- ◇ Street Address 2:
- ◇ City: Cedar Hill
- ◇ State: TX
- ◇ Zip: 75104 (corrected)
- ◇ Weight: 260 lb (93.3 kg)

- ◇ Primary Ins: MEDICARE
- ◇ Secondary Ins: BCBS

Order Info

- ◇ Product 1: Bedside Commode
- ◇ Physician: Osman, Houssam Galal Eldin
- ◇ Referral: Methodist Richardson Medical Center

: Right arm, Patient Position: Lying) |
| Wt 202 lb 9.6 oz (91.9 kg) | SPO2
A.1)

Ht: 5' 9.02" (1.753 m)
Wt: 205 lb 11 oz (93.3 kg)

All Diagnosis Codes

P3 — K86.89

Keywords

FIG. 3F

11/19

400

A1001001A23J09B10449B93643		Received 2023-10-23 at 12:49 PDT	Enable Hover
		15-page application/pdf	
Appears to be related to the following submission:			
A1001001A23J09B10449B93643 (exact)		2023-10-23 at 11:47 PDT	430
Demographics			
◊ First Name: Brenton	◊ Street Address: 170 Hatfield Pl (corrected)		
◊ Last Name: Smith-Hartley	◊ Street Address 2:		
◊ Date of Birth: 05/22/1981	◊ City: Staten Island		
◊ Phone: 544-658-5808	◊ State: NY		
Email:	◊ Zip: 10302 (corrected)		
◊ Height: 66 in	◊ Weight: 265 lb		
Payor(s)			
◊ Primary Ins: FIDELIS MEDICAID	◊ Primary Ins ID: 345634572-00		
Secondary Ins:	Secondary Ins ID:		
Order Info			
◊ Product 1: Machine level 5-20 cm H2O	◊ Product 2: Mask/Head Gear to fit		
◊ Product 3: Tubing, Filter	◊ Product 4: Humidifier		
◊ Product 5: Auto CPAP Machine	◊ Physician: Lin, Fred Y		
◊ Primary Diagnosis: G47.33	◊ Referral: Icahn School of Medicine at Mount Sinai		
All Diagnosis Codes			
P3 — G47.33			
P7 — G47.33			
Keywords			
Authorization --			
(1) Auth p15		(2) Electronically p7 p14	
NPWT --			
(1) Diabetes . p11		(2) Drainage . p12 p14	
(3) Operative p13 p14			
O2 --			
(1) Acute p12		(2) Respiratory p8 p11	
(2) SpO2 p8			
PAP --			
(3) AHI p7 p8		(1) APAP p2	
(2) Auto PAP . . p3 p7		(1) CM p3	
(1) CM pressure p3		(1) Central p8	
(1) Filter p3		(1) Humidifier . p3	
(1) Mask p3		(5) OSA p7 p13	
(5) PAP p2 p3 p7 p13		(1) Pressure . . . p13	
(3) Sleep Study . p7		(2) Supplies . . . p2 p3	
(1) Tubing p3			

FIG. 4A

12/19

400 ↗

A1001001A23J09B10449B93643

Received 2023-10-23 at 12:49 PDT Disable Hover
15-page application/pdf

Appears to be related to the following submission:

A1001001A23J09B10449B93643 (exact) 2023-10-23 at 11:47 PDT

Demographics

- ◊ First Name: Brenton
- ◊ Last Name: Smith-Hartley
- ◊ Date of Birth: 05/22/1981
- ◊ Phone: 544-658-5808
- ◊ Street Address: 170 Hatfield Pl (corrected)
- ◊ Street Address 2:
- ◊ City: Staten Island
- ◊ State: NY

DOB: May 22, 1981

Home #: (544) 658 5808

Mobile #: _____

Address: 170 Hatfield Pl, Staten Island, NY 10314

Insurance: Fidelis ID# 34563457200

To Whom It May Concern;

Secondary Ins:

- ◊ Product 1: Machine level 5-20 cm H₂O
- ◊ Product 3: Tubing, Filter
- ◊ Product 5: Auto CPAP Machine
- ◊ Primary Diagnosis: G47.33

DOB: May 22, 1981

Address: 170 Hatfield Pl
STATEN ISLAND NY 10314 - 1532

Phone : (544) 658-5808

PCP : _____

Demographics

May 22, 1981

42 year old male

170 Hatfield Place
STATEN ISLAND
NY 10314-1532
544-658-5808

Comm Pref: ☒ ☐

unknown at Other

© _____ All

P3 — Mobile #: _____

P7 — Address: 170 Hatfield Pl, Staten Island, NY 10314

Insurance: Fidelis ID# 34563457200

FIG. 4B

13/19

400

A1001001A23J09B10449B93643

Received 2023-10-23 at 12:49 PDT
15-page application/pdf[Disable Hover](#)

Appears to be related to the following submission:

A1001001A23J09B10449B93643 (exact) 2023-10-23 at 11:47 PDT

Demographics

- ◊ First Name: Brenton
- ◊ Last Name: Smith-Hartley
- ◊ Date of Birth: 05/22/1981
- ◊ Phone: 544-658-5808
- ◊ Email:
- ◊ Height: 66 in
- ◊ Street Address: 170 Hatfield Pl (corrected)
- ◊ Street Address 2:
- ◊ City: Staten Island
- ◊ State: NY
- ◊ Zip: 10302 (corrected)
- ◊ Weight: 265 lb

416

Payor(s)

- ◊ Primary Ins: FIDELIS MEDICAID
- ◊ Secondary Ins:
- ◊ Primary Ins ID: 345634572-00
- ◊ Secondary Ins ID:

Order Info

- ◊ Product 1: Machine level 5-20 cm H2O
- ◊ Product 2: Mask/Head Gear to fit
- ◊ Product 3: Tubing, Filter
- ◊ Product 4: Humidifier
- ◊ Product 5: Auto CPAP Machine
- ◊ Physician: Lin, Fred Y

436

438

Please supply all necessary equipment/supplies needed:

Machine level 5-20 cm H2O

Mask/Head Gear to fit

Tubing, Filter

Humidifier

Auto CPAP Machine

(1)

(1)

(3)

Operative pro p14

FIG. 4C

14/19

500

A1001001A23J03A93700D50629

Received 2023-10-23 at 12:06 PDT
2-page application/pdf

Enable Hover

Multiple patients identified, may need fax split

530

Demographics

502

First Name:

Street Address:

Last Name:

Street Address 2:

Date of Birth:

City:

Phone:

State:

Email:

504

Zip: 33125

506

Patient Name: Dave Rojas	Date of Birth: 04/08/1992	Phone#: 961-995-8728		
Address: 445 NW 35th Ave	City, State, Zip	Miami, FL 33125		
Insurance: Oscar				

Patient Name: Brooke Stone	Date of Birth: 09/18/1974	Phone#: 9303249644		
Address: 1780 SW 14th Ter	City, State, Zip	Miami, FL 33145		
Insurance: BCBS				

Order info

508

Product :

Physician:

Primary Diagnosis:

Referral:

All Diagnosis Codes

P2 — G47.33 G47.33 G47.39

P3 — G47.33 G47.33 G47.39

Keywords

FIG. 5

600 ↘

A1001001A23J09B10449B93643

Metadata: Received 2023-10-23 at 12:49 PDT

Metadata: 15-page application/pdf

Warning: Appears to be related to the following submission:

Warning: A1001001A23J09B10449B93643 (exact) 2023-10-23 at 11:47 PDT

Extraction Results

602

First Name: Brenton ✓

Last Name: Smith-Hartley ✓

Date of Birth: 05/22/1981 ✓

Phone: 544-658-5808 ⚠

Height: 66 in ⚠

Primary Ins: FIDELIS MEDICAID ✓

Physician: Lin, Fred Y ✓

Referral: Icahn School of Medicine at Mount Sinai ✓

604

606608

Street Address: 170 Hatfield Pl (corrected) ✓ ← 610

Street Address 2: ⚠

City: Staten Island ✓

State: NY ✓

Zip: 10302 (corrected) ✓

Weight: 265 lb ✓

Primary Ins ID: 345634572-00 ✓

Primary Diagnosis: G47.33 ✓

FIG. 6A

600 ↘

A1001001A23J09B10449B93643

Metadata: Received 2023-10-23 at 12:49 PDT

Metadata: 15-page application/pdf

Warning: Appears to be related to the following submission:

Warning: A1001001A23J09B10449B93643 (exact) 2023-10-23 at 11:47 PDT

Extraction Results

☒ ☐ First Name: Brenton ✓	☒ ☐ Street Address: 170 Hatfield Pl (corrected) ✓
☒ ☐ Last Name: Smith-Hartley ✓ 612	☒ ☐ Street Address 2: △ 99.5
☒ ☐ Date of Birth: 05/22/1981 ✓ 614	☒ ☐ City: Staten Island ✓
☒ ☐ Phone: 544-658-5808 622 87.0	☒ ☐ State: NY ✓
☒ ☐ Height: 66 in △ 624	☒ ☐ Zip: 10302 (corrected) ✓
☒ ☐ Primary Ins: FIDELIS MEDICAID ✓	☒ ☐ Weight: 265 lb ✓
☒ ☐ Physician: Lin, Fred Y ✓	☒ ☐ Primary Ins ID: 345634572-00 ✓
☒ ☐ Referral: Icahn School of Medicine at Mount Sinai ✓	☒ ☐ Primary Diagnosis: G47.33 ✓

FIG. 6B

600 ↘

A1001001A23J09B10449B93643

Metadata: Received 2023-10-23 at 12:49 PDT

Metadata: 15-page application/pdf

Warning: Appears to be related to the following submission:

Warning: A1001001A23J09B10449B93643 (exact) 2023-10-23 at 11:47 PDT

Extraction Results

⦿ ↗

First Name: Brenton ✓

⦿ ↗

Last Name: Smith-Hartley ✓

⦿ ↗

Date of Birth: 05/22/1981 ✓

⦿ ↗

Phone: 544-658-5808 ⚠

⦿ ↗

Height: 66 in ⚠

⦿ ↗

Referral: Icahn School of Medicine at Mount Sinai ✓

⦿ ↗

Street Address: 170 Hatfield Pl (corrected) ✓ ↖ 610

⦿ ↗

Street Address 2: ⚠

⦿ ↗

City: Staten Island ✓

⦿ ↗

State: NY ✓

⦿ ↗

Zip: 10302 (corrected) ✓

66 in.

265.0 lbs.

5634572-00

⦿ ↗

Primary Diagnosis: G47.33 ✓

632 ↗

⦿ ↗

Height: 66 in ⚠ ↖ 636 ↗ 634

Height:

Weight:

66 in.

265.0 lbs.

5634572-00

FIG. 6C

18/19

640

A1001001A23J09B10449B93643

600

Products

652 642

Machine level 5-20 cm H2O ✓

Auto CPAP Machine ✓

658

Please supply all necessary equipment/supplies needed:

Machine level 5-20 cm H2O

Mask/Head Gear to fit

Tubing, Filter

Humidifier

Auto CPAP Machine

654

Re: Brenton Smith-Hartley

DOB: May 22, 1981

Home #: (544) 658 5808

Mobile #: _____

Address: 170 Hatfield Pl, Staten Island, NY 10314

Insurance: Fidelity ID# 34563457200

To Whom It May Concern;

Please supply all necessary equipment/supplies needed:

Machine level 5-20 cm H2O

Mask/Head Gear to fit

Tubing, Filter

Humidifier

Auto CPAP Machine

Lifetime Supply

DX: G47.33

See attached office notes, prior test report, and Insurance card.

If there are any questions or concerns, please feel free to call my office at (718) 420-1279.

FIG. 6D

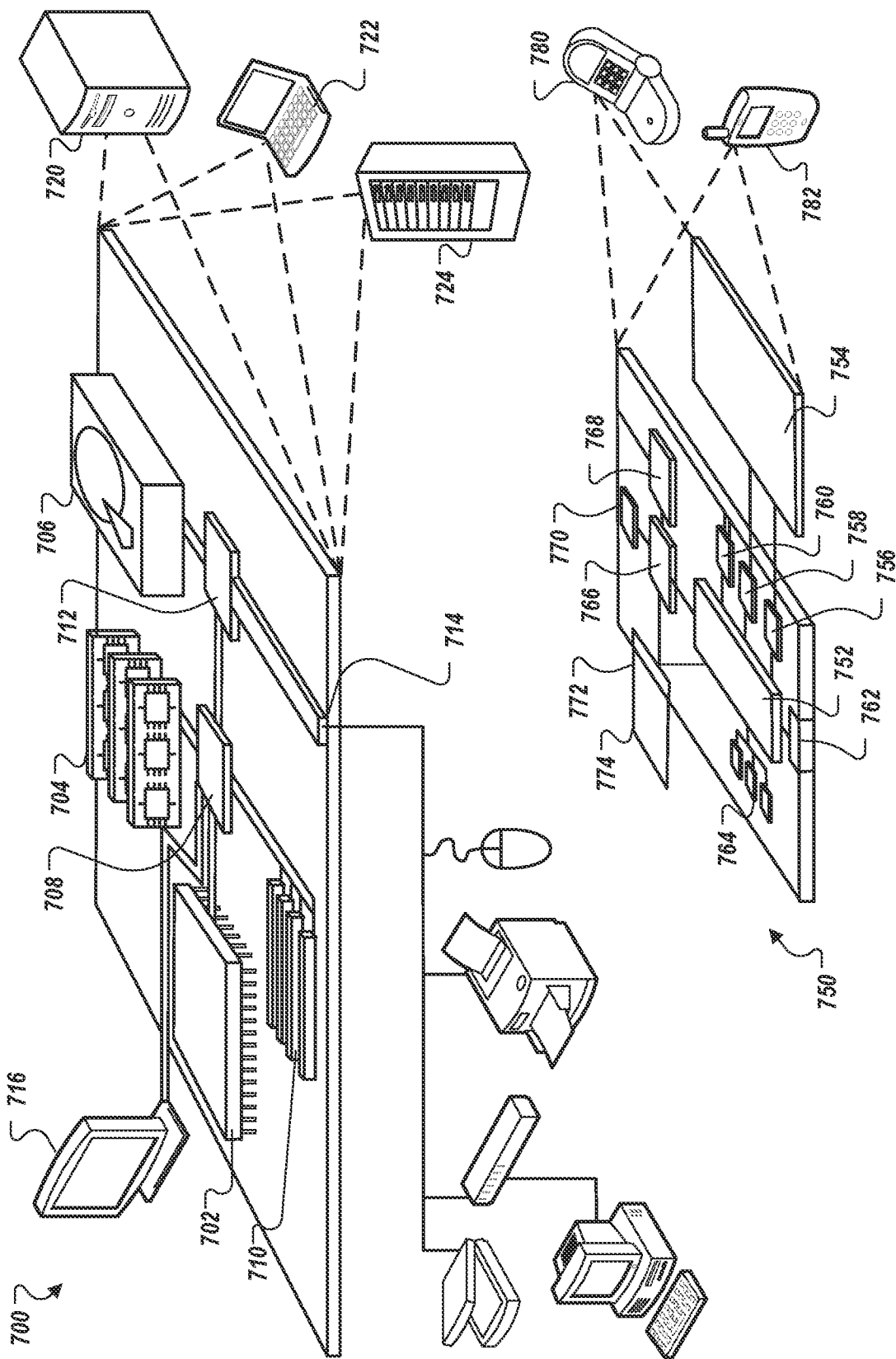


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2024/036768

A. CLASSIFICATION OF SUBJECT MATTER INV. G06F16/93 G06V30/416 ADD.		
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) G06F G06V 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) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/179313 A1 (HO MING FUNG [US]) 23 June 2016 (2016-06-23) paragraph [0020] - paragraph [0026]; figures 1-6 -----	1 - 20
☐ 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 application or patent 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 17 October 2024		Date of mailing of the international search report 30/10/2024
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Hauck, Rainer

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2024/036768

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2016179313 A1	23-06-2016	US 9317484 B1	19-04-2016
		US 2016179313 A1	23-06-2016
