

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 March 2012 (29.03.2012)

(10) International Publication Number  
**WO 2012/040677 A1**

- (51) **International Patent Classification:**  
*G06F 7/00* (2006.01)      *G06F 17/00* (2006.01)
- (21) **International Application Number:**  
PCT/US201 1/053 168
- (22) **International Filing Date:**  
24 September 201 1 (24.09.201 1)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
61/386,019    24 September 2010 (24.09.2010)      US
- (71) **Applicant (for all designated States except US):** INTERNATIONAL BUSINESS MACHINES CORPORATION [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (72) **Inventors; and**
- (75) **Inventors/ Applicants (for US only):** CHU-CARROLL, Jennifer [CA/US]; 19 Skyline Drive, Hawthorne, New York 10532 (US). FERRUCCI, David A. [US/US]; 19 Skyline Drive, Hawthorne, New York 10532 (US).
- (74) **Agents:** GROLZ, Edward W. et al; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Suite 300, Garden City, New York 11530 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**  
— with international search report (Art. 21(3))  
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) **Title:** EFFICIENT PASSAGE RETRIEVAL USING DOCUMENT METADATA

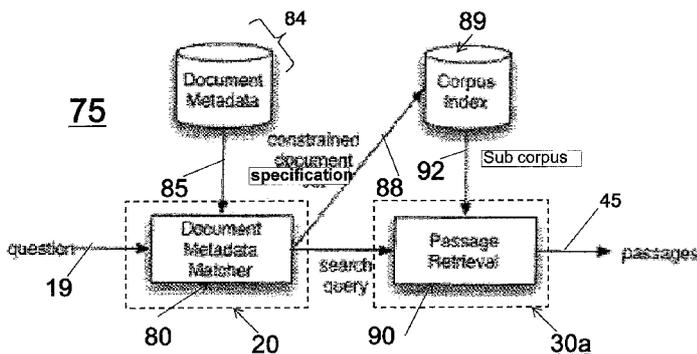


Fig. 2

(57) **Abstract:** A system, method and computer program product for efficiently retrieving relevant passages to questions based on a corpus of data. A processor device receives an input query and performs a query analysis to obtain searchable query terms. The processor performs: matching metadata associated with one or more documents against the query terms. The document metadata includes one or more of: a title of the documents, one or more user tags or clouds. Then the processor device performs: mapping matched document metadata to corresponding one or more documents; identifying corresponding matched documents to form a subcorpus of documents; and conducting a search in the data subcorpus using the searchable query terms to obtain one or more passages relevant input query from the identified documents.



WO 2012/040677 A1

## EFFICIENT PASSAGE RETRIEVAL USING DOCUMENT METADATA

## CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present invention relates to and claims the benefit of the filing date of commonly-owned, co-pending United States Provisional Patent Application No. 61/386,019, filed September 24, 2010, the entire contents and disclosure of which is incorporated by reference as if fully set forth herein.

## BACKGROUND

[0002] The invention relates generally to information retrieval systems, and more particularly, the invention relates to an automated query/answer system and method implementing a passage retrieval component to conduct a search that identifies passages relevant to a given question using document metadata from a collection including text-based resources.

## Description of the Related Art

[0003] An introduction to the current issues and approaches of question answering (QA) can be found in the web-based reference [http://en.wikipedia.org/wiki/Question\\_answering](http://en.wikipedia.org/wiki/Question_answering). Generally, QA is a type of information retrieval. Given a collection of documents (such as the World Wide Web or a local collection) the system should be able to retrieve answers to questions posed in natural language. QA is regarded as requiring more complex natural language processing (NLP) techniques than other types of information retrieval such as document retrieval, and it is sometimes regarded as the next step beyond search engines.

[0004] QA research attempts to deal with a wide range of question types including: fact, list, definition, How, Why, hypothetical, semantically-constrained, and cross-lingual questions.

Search collections vary from small local document collections, to internal organization documents, to compiled newswire reports, to the World Wide Web.

[0005] Closed-domain QA deals with questions under a specific domain, for example medicine or automotive maintenance, and can be seen as an easier task because NLP systems can exploit domain-specific knowledge frequently formalized in ontologies. Open-domain QA deals with questions about nearly everything, and can only rely on general ontologies and world knowledge. On the other hand, these systems usually have much more data available from which to extract the answer.

[0006] Alternatively, closed-domain QA might refer to a situation where only a limited type of questions are accepted, such as questions asking for descriptive rather than procedural information.

[0007] Access to information is currently dominated by two paradigms. First, a database query that answers questions about what is in a collection of structured records. Second, a search that delivers a collection of document links in response to a query against a collection of unstructured data, for example, text or html.

[0008] A major unsolved problem in such information query paradigms is the lack of a computer program capable of accurately answering factual questions based on information included in a collection of documents that can be either structured, unstructured, or both. Such factual questions can be either broad, such as "what are the risks of vitamin K deficiency?", or narrow, such as "when and where was Hillary Clinton's father born?"

[0009] It is a challenge to understand the query, to find appropriate documents that might contain the answer, and to extract the correct answer to be delivered to the user. There is a need to further advance the methodologies for answering open-domain questions.

## SUMMARY

[0010] In one aspect there is provided a computing infrastructure and methodology that conducts question and answering and performs automatic passage retrieval operations in a highly efficient manner.

[0011] In one aspect, there is provided a computer-implemented method for efficiently retrieving relevant passages to questions based on a corpus of data comprising: receiving an input query; performing a query context analysis upon the input query to obtain searchable query terms; matching metadata associated with one or more documents against the query terms; mapping matched document metadata to corresponding one or more documents; identifying corresponding matched documents to form a subcorpus of documents; and conducting a search in the data subcorpus using the searchable query terms to obtain one or more passages relevant to the input query from the identified documents, wherein one or more processor devices performs one or more the retrieving, performing, matching, mapping, identifying and conducting.

[0012] In this aspect, the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.

[0013] Further to this aspect, prior to matching of metadata associated with one or more documents against the query terms there is performed: extracting document metadata from one or more documents of a corpus of documents; providing the extracted document metadata as a dictionary in a storage device, each document metadata stored in the dictionary being associated with a corresponding document identification (ID), wherein the matching of metadata against the query terms comprises: performing, by the processor device, a dictionary matching.

[0014] In an alternate embodiment, there is provided a computer-implemented method for efficiently retrieving relevant passages to questions based on a corpus of data comprising: receiving, at a processor device, an input query; performing, at the processor device, a query

context analysis upon the input query to obtain searchable query terms; accessing a dictionary of document metadata obtained from one or more documents of the data corpus, each stored document metadata being associated with a corresponding document identification (ID); performing, by the processor device, a dictionary matching of the metadata associated with one or more documents against the query terms; mapping matched document metadata to corresponding one or more document IDs; identifying corresponding matched documents to form a subcorpus of documents; and conducting a search in the subcorpus using the searchable query terms to obtain passages relevant to the input query from the identified documents.

[0015] A computer program product is provided for performing operations. The computer program product includes a storage medium readable by a processing circuit and storing instructions run by the processing circuit for running a method(s). The method(s) are the same as listed above.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0016] The objects, features and advantages of the invention are understood within the context of the Detailed Description, as set forth below. The Detailed Description is understood within the context of the accompanying drawings, which form a material part of this disclosure, wherein:

[0017] FIG. 1 shows a prior art high level logical architecture 10 of a question/ answering method in which the present invention may be employed;

[0018] FIG. 2 is a schematic diagram depicting passage retrieval components 75 according to one embodiment;

[0019] FIG. 3 is a flow diagram illustrating a method 100 for performing passage retrieval operations in one embodiment; and,

[0020] FIG. 4 illustrates an exemplary hardware configuration to run method steps described in FIG. 3 in one embodiment.

## DETAILED DESCRIPTION

[0021] FIG. 1 shows a QA system diagram such as described in U.S. Patent Application 12/126,642 depicting a high-level logical architecture 10 and methodology in which the present system and method may be employed in one embodiment.

[0022] FIG. 1 illustrates the major components that comprise a canonical question answering system 10 and their workflow. The question analysis component 20 receives a natural language question 19 (e.g., *"Who is the 42<sup>nd</sup> president of the United States?"*) and analyzes the question to produce, minimally, the semantic type of the expected answer (in this example, "president"), and optionally other analysis results for downstream processing. The search component 30a formulates queries from the output 29 of question analysis and consults various resources such as the World Wide Web 41 or one or more knowledge resources, e.g., databases, knowledge bases 42, to retrieve "documents" including, e.g., whole documents or document portions 44, e.g., web-pages, database tuples, etc., having "passages" 44 that are relevant to answering the question. The candidate answer generation component 30b may then extract from the search results 48 potential (candidate) answers to the question, which are then scored and ranked by the answer selection component 50 to produce a final ranked list of answers with associated confidence scores.

[0023] In current questions and answer systems, one key component is the passage retrieval operations conducted when searching for candidate answers in heterogeneous collection of structured, semi-structured and unstructured information resources. Passage retrieval operations adapt a search engine at its core to identify passages relevant to a given question from the collection of sources, e.g., text-based sources. Passage retrieval is also relevant to any search application where selecting passages containing, for example, 1-3 sentences is more appropriate

than retrieving entire documents either for processing by downstream components, or for presentation to the end user.

[0024] Most existing systems performing a passage retrieval operation adopts one of two approaches. The first approach is to adopt a document search engine to retrieve a list of relevant documents using the search engine's internal document ranking criteria, and to apply a custom post-hoc passage scoring algorithm to identify the most relevant text segments from these documents. The second approach is to adopt a search engine with passage retrieval capability and to make use of the engine's internal ranking algorithm to return a set of relevant passages. In either approach, the retrieval process is performed over the entire collection, which typically contains millions of documents or more. This poses an efficiency issue for real-time question answering systems that must deliver answers to users in no more than a few seconds. A typical solution for this problem is to split the search index into multiple subindices on multiple machines so that retrieval against the subindices can be performed in parallel and their result merged. While this solution addresses the efficiency issue, it poses other problems related to merging search results from multiple indices.

[0025] It would be highly desirable to provide a system and method that improves the efficiency of passage retrieval based on dynamic subcorpus selection to constrain the number of relevant documents considered in the retrieval process.

[0026] In one embodiment, the present system and method for efficient passage retrieval against a corpus given a question is applicable and may be part of a Question Answering (QA) system. Alternatively, the system and method for efficient passage retrieval against a corpus given a question may be implemented in non-QA applications, i.e., applications implemented to return a passage, for example, a 1-sentence to 3-sentence passage most relevant to a question, as opposed to an answer per se.

[0027] Commonly-owned, co-pending U.S. Patent Application 12/126,642, titled "SYSTEM AND METHOD FOR PROVIDING QUESTION AND ANSWERS WITH DEFERRED TYPE

EVALUATION" and co-pending U.S. Patent Application 12/15241 1, titled "SYSTEM AND METHOD FOR PROVIDING ANSWERS TO QUESTIONS" are both incorporated by reference herein, and describe a QA (Question and Answer) system and method in which the present passage retrieval system may be incorporated.

[0028] In one embodiment, the present disclosure may extend and complement the effectiveness of a QA or non-QA system and method by improving the efficiency of passage retrieval operations based on dynamic subcorpus selection to constrain the number of relevant documents considered in the retrieval process.

[0029] In one embodiment, the subcorpus selection process is based on a matching algorithm that identifies relevant documents based on the question text and metadata associated with the documents in the collection, such as document titles, user tags ("clouds"), or automatically identified document labels. The passage retrieval process is then restricted to return passages only from this subcorpus, which typically contains several orders of magnitude fewer documents than the entire collection.

[0030] The approach to efficient passage retrieval significantly constrains the pool of documents from which passages may be retrieved based on metadata associated with documents, such as document titles and user tags ("clouds"). The efficiency of passage retrieval is improved by providing the ability to dynamically select a subcorpus from which search will take place based on terms in the user question and metadata associated with documents in the corpus. More specifically, the user's input question string is analyzed to extract all matches between question terms and document metadata. Those matched documents comprise a subcorpus from which the system will extract passages for this question.

[0031] In a non-limiting example, there is considered the following user question

*"which modern artist was Francoise Gilot, Dr. Jonas Salk's wife, once the companion of?"*

[0032] In the example, matching the instances of document titles to the terms in the question, yields five entities: "modern", "artist", "Francoise Gilot", "Jonas Salk", and "companion" are identified as document titles in the corpus. It is understood that a term may map to multiple documents with that title. For example, "companion" may map to an article that talks about a caregiver, or an architectural feature of ships, or a character in "Doctor Who". Using the document identifications (IDs) that corresponds to each document title, the documents with the identified document IDs are selected to form a subcorpus consisting of potentially highly relevant documents for answering the given question. The passage retrieval process is then constrained to finding the most relevant passages from this document subcorpus which may contain on the order of tens of documents, instead of from the entire collection which may contain millions of documents or more. In this example, several relevant passages, such as "Francoise Gilot (born 1921) is a French born painter and is known as a companion of Picasso between 1944 and 1953" from the document titled "Francoise Gilot", and "In 1968, they divorced, and in 1970 Salk married Francoise Gilot, the former mistress of Pablo Picasso" from the document titled "Jonas Salk".

[0033] FIG. 2 is a schematic diagram depicting passage retrieval components 75 that may be implemented in QA and non-QA systems according to one embodiment. In one embodiment, the system components 75 conducting passage retrieval operations make use of system modules from FIG. 1 such as: the question analysis processing component 20 that performs a query context analysis upon an received input query to break down said input query into query terms, and any searchable components thereof; and, the search component 30a that formulates queries from the output searchable components of question analysis unit and that consults various resources such as the World Wide Web 40 or one or more knowledge resources, e.g., databases, knowledge bases 42.

[0034] More particularly, as shown in FIG. 2, the question analysis processing component 20 includes a programmed matcher component 80 that functions to identify document metadata present in the question. It performs this by consulting a resource 84 containing document metadata information for all documents. Document metadata may include any information that

identifies the topic or domain of the document, such as the document title, manually or automatically derived category/domain classification, and crowdsourced or automatically derived tag clouds ("clouds") which indicate general topics of the document. It is against this data resource 84 where matching of terms in the input question to the document metadata information is performed. Data corpus 89 represents the entire data corpus that the QA or non-QA system is using and may include both open domain and closed domain topics.

[0035] For example, a document containing George W. Bush's 2007 State of the Union address may include the following metadata:

Title: 2007 State of the Union Address

Category: Presidential Addresses, George W. Bush Speeches,...

Tags: Security, Iraq, Terrorists, Health, America,...

[0036] A sample implementation of this matcher component 80 is to represent the metadata in dictionary form and to leverage a dictionary matcher to identify dictionary terms that appear in an input question. For example, any matching component can be used to identify closed or open domain dictionary terms in text (e.g., legal terms, medical terms, or generic named entities) may be used. Thus, given a piece of text (an input query), the matching algorithm determines from the question text those terms that match entries in the dictionary. In one embodiment, a dictionary matcher includes the open source ConceptMapper annotator available at <http://uima.apache.Org/sandbox.html#concept.mapper.annotator>, whose functionality is incorporated by reference as if fully set forth herein.

[0037] The matched dictionary entries (question terms) are used to identify a subset of documents for the passage retrieval process. That is, for the query terms that are mapped to the metadata (titles, tags, clouds) of a document in the resource 84, that document's index (or other document identifier) is flagged, tagged, or recorded for its inclusion in a subcorpus. In one embodiment, each dictionary entry in resource 84 encodes the document ID for each document that contains metadata matching that dictionary term. The metadata and associated document

information in the dictionary entry that match the terms in the input question is represented as 85 in FIG. 2.

[0038] The passage retrieval component can be any standard IR (Information Retrieval) search engine 90 that supports both of: Retrieval of relevant short passages, instead of full documents; and Runtime specification of a relevant subcorpus for retrieval. One example IR search engine that satisfies this requirement is the Indri engine from the Lemur Toolkit such as the search engine with passage retrieval capability, such as Indri, <http://www.lemurproject.org/indri/>, incorporated by reference as if fully set forth herein.

[0039] In further view of FIG. 2 the matched documents identified by the matcher component 80 form a constrained document set 88, indicated in the entire corpus 89 having the entire index and a subcorpus 92 is built including the constrained document set 88 on which passage retrieval operation via IR search engine 90 are performed to select the most relevant passages.

[0040] A passage retrieval method 100 employed by the passage retrieval components 75 for improving the efficiency of passage retrieval is described with respect to FIG. 3. As shown in FIG. 3, the method 100 includes at 101, receiving at a processor device, an input query and, using a parser device or function, breaking down the query into searchable query terms. In one embodiment, the obtained searchable query terms from said input query are terms that match document metadata. Then, at 105, there is performed accessing a semi-structured source of information containing document metadata (such as the title of the documents, a category, or user tags or clouds). In one embodiment, the semi-structured source of information is a dictionary or corpus that associates data (e.g., definitions) with a large set of vocabulary items including document metadata stored in memory storage device.

[0041] That is, in one embodiment, the semi-structured source of information may be formed via off-line processes that extract document metadata from one or more documents of a large corpus of documents. The extracted document metadata is stored as a dictionary in the memory storage device, with each document metadata stored in the dictionary having one or more associated

document identifications (IDs) that represent those documents matching the metadata in that dictionary entry.

[0042] Then, at 110, the programmed processor device performs invoking a matching component to match a document metadata against the query terms. As mentioned, a dictionary matcher may be invoked that includes the open source ConceptMapper annotator available at <http://uima.apache.org/sandbox.html#concept.mapper.annotator>.

[0043] Continuing to 115, there is next performed mapping of the matched document metadata to corresponding one or more document IDs. Then at 120, from the corresponding IDs, there is performed identifying the corresponding matched documents.

[0044] In one embodiment, for the matched document metadata found in the dictionary, the corresponding documents indicated by the mapped document IDs are identified, e.g., flagged, tagged or recorded in the corpus in which the actual documents are electronically stored with their ID. Thus, in one embodiment, the identified corresponding matched documents form the subcorpus 92 of documents including only the identified matched metadata documents of the larger corpus of documents. This step invokes corpus construction functionality to identify the subset of flagged, tagged or otherwise identified matched metadata documents obtained from the first corpus 84 (FIG. 2) during the matching step, which functionality for dynamically constructing subcorpora during runtime is provided for example in the above-incorporated Indri engine from the Lemur Toolkit.

[0045] In an alternate embodiment, there may be further performed at 125, extracting the identified corresponding matched documents are found in step 120 as the subcorpus 92.

[0046] Then, at 130, the method performs passage retrieval operations against those identified matched metadata documents obtained from the subcorpus 92 formed at step 120 or 125.

[0047] Finally, assuming a search engine has internal document ranking ability, then at 135, there is returned the resulting list of ranked passages at 125.

[0048] In one embodiment, the passage retrieval process 100, FIG. 3 when performed in parallel with traditional passage retrieval algorithms is more effective when the information sought in the question is present in documents whose relevant metadata field contains a term/phrase in the question. To increase recall, the dictionary can be constructed to include morphological variations for the given metadata information, such as including both the singular and plural forms of terms, as well as known synonyms. In one embodiment, redirect links between Wikipedia® titles (which, e.g., redirects requests for the document "artists" to the document titled "artist" and for example, "01' Blue Eyes" to "Frank Sinatra") are used to capture morphological variations and synonyms. Alternatively, morphological and synonym information can be mined from publicly available resources such as WordNet® (Trademark of The CORPORATION NEW JERSEY Princeton University) available at <http://wordnet.princeton.edu/>. For these questions, this approach significantly reduces execution time in those situations compared with performing passage retrieval against a large unconstrained corpus 89.

[0049] As mentioned, FIG. 1 shows a system diagram described in U.S. Patent Application 12/126,642 depicting a high-level logical architecture of a QA system 10 and methodology in which a system and method for deferred type evaluation using text with limited structure is employed in one embodiment.

[0050] Generally, as shown in FIG. 1, the high level logical architecture 10 includes the Query Analysis module 20 implementing functions for receiving and analyzing a user query or question. The term "user" may refer to a person or persons interacting with the system, or refers to a computer system 22 generating a query by mechanical means, and where the term "user query" refers to such a mechanically generated query and context 19'. A candidate answer generation module 30 is provided to implement a search for candidate answers by traversing structured, semi structured and unstructured sources contained in primary sources (e.g., the Web, a data corpus 41) and in an Answer Source or a Knowledge Base (KB), e.g., containing

collections of relations and lists extracted from primary sources. All the sources of information can be locally stored or distributed over a network, including the Internet.

[0051] The Candidate Answer generation module 30 of architecture 10 generates a plurality of output data structures containing candidate answers based upon the analysis of retrieved data. In FIG. 1, an Evidence Gathering module 50 further interfaces with the primary sources and knowledge base for concurrently analyzing the evidence based on passages having candidate answers, and scores each of candidate answers, in one embodiment, as parallel processing operations. In one embodiment, the architecture may be employed utilizing the Common Analysis System (CAS) candidate answer structures as is described in commonly-owned, issued U.S. Patent No. 7,139,752, the whole contents and disclosure of which is incorporated by reference as if fully set forth herein.

[0052] As depicted in FIG. 1, when the Search System 30a is employed in the context of a QA system, the Evidence Gathering and Scoring module 50 comprises a Candidate Answer Scoring module 40 for analyzing a retrieved passage and scoring each of candidate answers of a retrieved passage. The Answer Source Knowledge Base (KB) may comprise one or more databases of structured or semi-structured sources (pre-computed or otherwise) comprising collections of relations (e.g., Typed Lists). In an example implementation, the Answer Source knowledge base may comprise a database stored in a memory storage system, e.g., a hard drive.

[0053] An Answer Ranking module 60 may be invoked to provide functionality for ranking candidate answers and determining a response 99 returned to a user via a user's computer display interface (not shown) or a computer system 22, where the response may be an answer, or an elaboration of a prior answer or request for clarification in response to a question —when a high quality answer to the question is not found. A machine learning implementation is further provided where the "answer ranking" module 60 includes a trained model component (not shown) produced using a machine learning techniques from prior data.

[0054] The processing depicted in FIG. 1, may be local, on a server, or server cluster, within an enterprise, or alternately, may be distributed with or integral with or otherwise operate in conjunction with a public or privately available search engine in order to enhance the question answer functionality in the manner as described. Thus, the method may be provided as a computer program product comprising instructions executable by a processing device, or as a service deploying the computer program product. The architecture employs a search engine (e.g., a document retrieval system) as a part of Candidate Answer Generation module 30 which may be dedicated to searching the Internet, a publicly available database, a web-site (e.g., IMDB.com), a privately available collection of documents or, a privately available database. Databases can be stored in any storage system, non-volatile memory storage systems, e.g., a hard drive or flash memory, and can be distributed over the network or not.

[0055] In one embodiment, when employed in a QA system, the system and method of FIG. 1 makes use of the Common Analysis System (CAS), a subsystem of the Unstructured Information Management Architecture (UIMA) that handles data exchanges between the various UIMA components, such as analysis engines and unstructured information management applications. CAS supports data modeling via a type system independent of programming language, provides data access through a powerful indexing mechanism, and provides support for creating annotations on text data, such as described in (<http://www.research.ibm.com/journal/sj/433/gotz.html>) incorporated by reference as if set forth herein. It should be noted that the CAS allows for multiple definitions of the linkage between a document and its annotations, as is useful for the analysis of images, video, or other non-textual modalities (as taught in the herein incorporated reference U.S. Patent 7,139,752).

[0056] In one embodiment, UIMA may be provided as middleware for the effective management and interchange of unstructured information over a wide array of information sources. The architecture generally includes a search engine, data storage, analysis engines containing pipelined document annotators and various adapters. The UIMA system, method and computer program may be used to generate answers to input queries. The method includes inputting a document and operating at least one text analysis engine that comprises a plurality of coupled

annotators for tokenizing document data and for identifying and annotating a particular type of semantic content. Thus it can be used to analyze a question and to extract entities as possible answers to a question from a collection of documents.

[0057] In an alternative environment, modules of FIGs. 1, 2 can be represented as functional components in GATE (General Architecture for Text Engineering) (see: <http://gate.ac.uk/releases/gate-2.0alpha2-build484/doc/userguide.html>). GATE employs components which are reusable software chunks with well-defined interfaces that are conceptually separate from GATE itself. All component sets are user-extensible and together are called CREOLE - a Collection of REusable Objects for Language Engineering. The GATE framework is a backplane into which plug CREOLE components. The user gives the system a list of URLs to search when it starts up, and components at those locations are loaded by the system. In one embodiment, only their configuration data is loaded to begin with; the actual classes are loaded when the user requests the instantiation of a resource.). *GATE components are one of three types of specialized Java Beans:* 1) Resource: The top-level interface, which describes all components. What all components share in common is that they can be loaded at runtime, and that the set of components is extendable by clients. They have Features, which are represented externally to the system as "meta-data" in a format such as RDF, plain XML, or Java properties. Resources may all be Java beans in one embodiment. 2) ProcessingResource: Is a resource that is runnable, may be invoked remotely (via RMI), and lives in class files. In order to load a PR (Processing Resource) the system knows where to find the class or jar files (which will also include the metadata); 3) LanguageResource: Is a resource that consists of data, accessed via a Java abstraction layer. They live in relational databases; and, VisualResource: Is a visual Java bean, component of GUIs, including of the main GATE GUI. Like PRs these components live in .class or .jar files.

[0058] In describing the GATE processing model any resource whose primary characteristics are algorithmic, such as parsers, generators and so on, is modeled as a Processing Resource. A PR is a Resource that implements the Java Runnable interface. The GATE Visualisation Model implements resources whose task is to display and edit other resources are modeled as Visual

Resources. The Corpus Model in GATE is a Java Set whose members are documents. Both Corpora and Documents are types of Language Resources(LR) with all LRs having a Feature Map (a Java Map) associated with them that stored attribute/value information about the resource. FeatureMaps are also used to associate arbitrary information with ranges of documents (e.g. pieces of text) via an annotation model. Documents have a DocumentContent which is a text at present (future versions may add support for audiovisual content) and one or more AnnotationSets which are Java Sets.

[0059] As UIMA, GATE can be used as a basis for implementing natural language dialog systems and multimodal dialog systems having a question answering system as one of the main submodules. The references, incorporated herein by reference above (United States Patent Nos. 6,829,603 and 6,983,252, and 7,136,909) enable one skilled in the art to build such an implementation.

[0060] FIG. 4 illustrates an exemplary hardware configuration of a computing system 400 in which the present system and method may be employed. The hardware configuration preferably has at least one processor or central processing unit (CPU) 411. The CPUs 411 are interconnected via a system bus 412 to a random access memory (RAM) 414, read-only memory (ROM) 416, input/output (I/O) adapter 418 (for connecting peripheral devices such as disk units 421 and tape drives 440 to the bus 412), user interface adapter 422 (for connecting a keyboard 424, mouse 426, speaker 428, microphone 432, and/or other user interface device to the bus 412), a communication adapter 434 for connecting the system 400 to a data processing network, the Internet, an Intranet, a local area network (LAN), etc., and a display adapter 436 for connecting the bus 412 to a display device 438 and/or printer 439 (e.g., a digital printer of the like).

[0061] As will be appreciated by one skilled in the art, aspects of the present invention may be embodied as a system, method or computer program product. Accordingly, aspects of the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment

combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system." Furthermore, aspects of the present invention may take the form of a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

[0062] Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable signal medium or a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, or store a program for use by or in connection with a system, apparatus, or device running an instruction.

[0063] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including, but not limited to, electro-magnetic, optical, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that can communicate, propagate, or transport a program for use by or in connection with a system, apparatus, or device running an instruction.

[0064] Program code embodied on a computer readable medium may be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc., or any suitable combination of the foregoing.

[0065] Computer program code for carrying out operations for aspects of the present invention may be written in any combination of one or more programming languages, including an object oriented programming language such as Java, Smalltalk, C++ or the like and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The program code may run entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

[0066] Thus, in one embodiment, the system and method for efficient passage retrieval may be performed with data structures native to various programming languages such as Java and C++.

[0067] Aspects of the present invention are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which run via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer program instructions may also be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of

manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[0068] The computer program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other devices to cause a series of operational steps to be performed on the computer, other programmable apparatus or other devices to produce a computer implemented process such that the instructions which run on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0069] The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more operable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be run substantially concurrently, or the blocks may sometimes be run in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

[0070] The embodiments described above are illustrative examples and it should not be construed that the present invention is limited to these particular embodiments. Thus, various changes and modifications may be effected by one skilled in the art without departing from the spirit or scope of the invention as defined in the appended claims.

## CLAIMS

What is claimed is:

1. A computer-implemented method for efficiently retrieving relevant passages to questions based on a corpus of data comprising:
  - receiving an input query;
  - performing a query analysis upon said input query to obtain searchable query terms;
  - matching metadata associated with one or more documents against said query terms;
  - mapping matched document metadata to corresponding one or more documents;
  - identifying corresponding matched documents to form a subcorpus of documents; and
  - conducting a search in said data subcorpus using said searchable query terms to obtain one or more passages relevant to the input query from said identified documents,wherein one or more processor devices performs one or more said retrieving, performing, matching, mapping, identifying and conducting.
2. The computer-implemented method of Claim 1, wherein the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.
3. The computer-implemented method of Claim 2, wherein prior to matching of metadata associated with one or more documents against said query terms:
  - extracting document metadata from one or more documents of a corpus of documents;
  - providing said extracted document metadata as a dictionary in a storage device, each document metadata stored in said dictionary being associated with one or more corresponding document identifications.
4. The computer-implemented method of Claim 3, wherein said matching of metadata against said query terms comprises: performing, by said processor device, dictionary matching.

5. The computer-implemented method of Claim 2, wherein said data corpus comprising document metadata information includes variations of metadata including one or more of: singular and plural forms of metadata terms, and synonyms for metadata terms.

6. The computer-implemented method of Claim 2, wherein obtaining searchable query terms from said input query comprises parsing, by said processor device, said input query to obtain terms matching document metadata.

7. The computer-implemented method of Claim 2, wherein said identifying corresponding matched documents to form a subcorpus of documents includes tagging or flagging each matched metadata documents in said corpus of documents.

8. The computer-implemented method of Claim 2, further comprising: extracting said tagged or flagged identified corresponding matched documents to form said subcorpus of documents.

9. A computer program product for efficiently retrieving relevant passages to questions based on a corpus of data, the computer program device comprising a non-transitory storage medium readable by a processing circuit and storing instructions run by the processing circuit for performing a method, the method comprising:

receiving an input query;

performing a query context analysis upon said input query to obtain searchable query terms;

matching metadata associated with one or more documents against said query terms;

mapping matched document metadata to corresponding one or more documents;

identifying corresponding matched documents to form a subcorpus of documents; and

conducting a search in said data subcorpus using said searchable query terms to obtain one or more passages relevant to the input query from said identified.

10. The computer program product of Claim 9, wherein the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.

11. The computer program product of Claim 9, wherein prior to matching of metadata associated with one or more documents against said query terms:

extracting document metadata from one or more documents of a corpus of documents;

providing said extracted document metadata as a dictionary in a storage device, each document metadata stored in said dictionary being associated with one or more corresponding document identifications.

12. The computer program product of Claim 11, wherein said matching of metadata against said query terms comprises: performing, by said processor device, dictionary matching.

13. The computer program product of Claim 9, wherein said data corpus comprising document metadata information includes variations of metadata including one or more of: singular and plural forms of metadata terms, and synonyms for metadata terms.

14. The computer program product of Claim 10, wherein obtaining searchable query terms from said input query comprises parsing, by said processor device, said input query to obtain terms matching document metadata.

15. The computer program product of Claim 10, wherein said identifying corresponding matched documents to form a subcorpus of documents includes tagging or flagging each matched metadata documents in said corpus of documents.

16. The computer program product of Claim 10, further comprising: extracting said tagged or flagged identified corresponding matched documents to form said subcorpus of documents.

17. A computer-implemented method for efficiently retrieving relevant passages to questions based on a corpus of data comprising:

receiving an input query;

performing a query context analysis upon said input query to obtain searchable query terms;

accessing a dictionary of document metadata obtained from one or more documents of the data corpus, each stored document metadata being associated with one or more corresponding document identifications (IDs);

performing a dictionary matching of said metadata associated with one or more documents against said query terms;

mapping matched document metadata to corresponding one or more document IDs;

identifying corresponding matched documents to form a subcorpus of documents; and

conducting a search in said subcorpus using said searchable query terms to obtain one or more passages relevant to the input query from said identified documents, wherein one or more processor devices perform one or more said retrieving, performing query context analysis, accessing, performing dictionary matching, mapping, identifying and conducting.

18. The computer-implemented method of Claim 17, wherein the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.

19. The computer-implemented method of Claim 18, wherein obtaining searchable query terms from said input query comprises parsing, by said processor device, said input query to obtain terms matching document metadata.

20. The computer-implemented method of Claim 17, wherein said identifying corresponding matched documents to form a subcorpus of documents includes:

tagging or flagging each matched metadata documents in said data corpus; and,

extracting said tagged or flagged identified corresponding matched documents to form said subcorpus of documents.

21. A system for efficiently retrieving relevant passages to questions based on a corpus of data comprising:

a memory storage device;

a processor device in communication with the memory device that performs a method comprising:

receiving an input query;

performing a query context analysis upon said input query to obtain searchable query terms;

matching metadata associated with one or more documents against said query terms;

mapping matched document metadata to corresponding one or more documents;

identifying corresponding matched documents to form a subcorpus of documents; and

conducting a search in said data subcorpus using said searchable query terms to obtain one or more passages relevant to the input query from said identified documents.

22. The system of Claim 21, wherein the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.

23. The system of Claim 22, wherein prior to matching of metadata associated with one or more documents against said query terms:

extracting document metadata from one or more documents of a corpus of documents;

providing said extracted document metadata as a dictionary in a storage device, each document metadata stored in said dictionary being associated with a corresponding document identification, wherein said matching of metadata against said query terms comprises performing a dictionary matching.

24. A computer program product for efficiently retrieving relevant passages to questions based on a corpus of data, the computer program device comprising a storage medium readable by a processing circuit and storing instructions run by the processing circuit for performing a method, the method comprising:

receiving, at a processor device, an input query;  
performing, at said processor device, a query context analysis upon said input query to obtain searchable query terms;  
accessing a dictionary of document metadata obtained from one or more documents of the data corpus, each stored document metadata being associated with a corresponding document identification (ID);  
performing, by said processor device, a dictionary matching of said metadata associated with one or more documents against said query terms;  
mapping matched document metadata to corresponding one or more document IDs;  
identifying corresponding matched documents to form a subcorpus of documents; and  
conducting a search in said subcorpus using said searchable query terms to obtain one or more passages relevant to the input query from said identified documents.

25. The computer program product of Claim 24, wherein the document metadata includes one or more of: a title of the documents, one or more user tags, one or more automatically identified document labels.

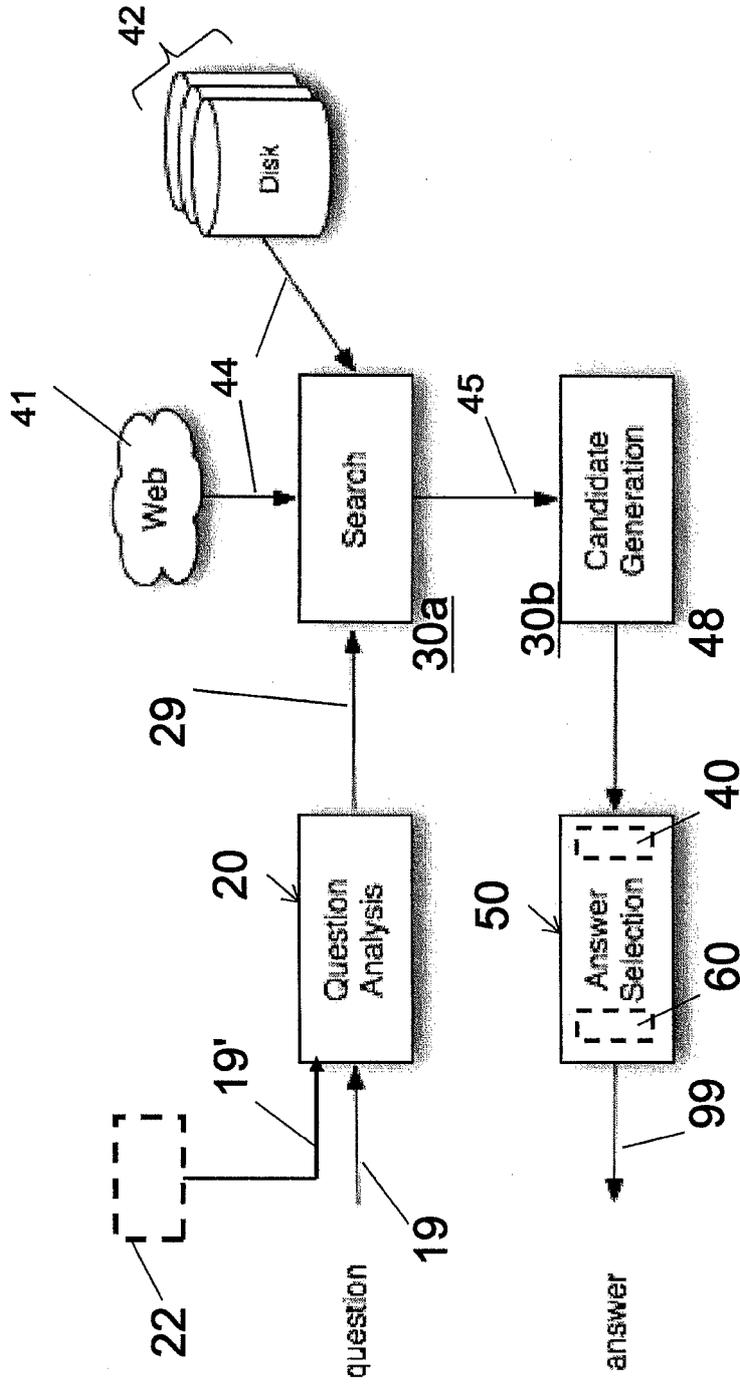


FIG. 1

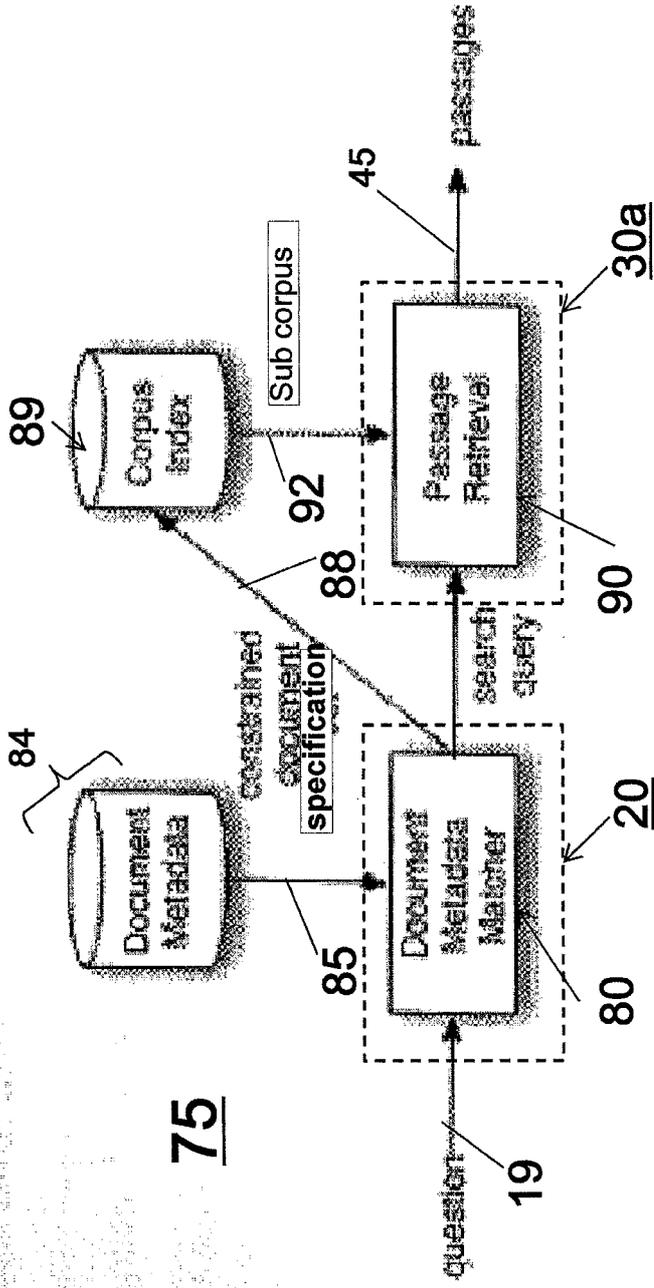


Fig. 2

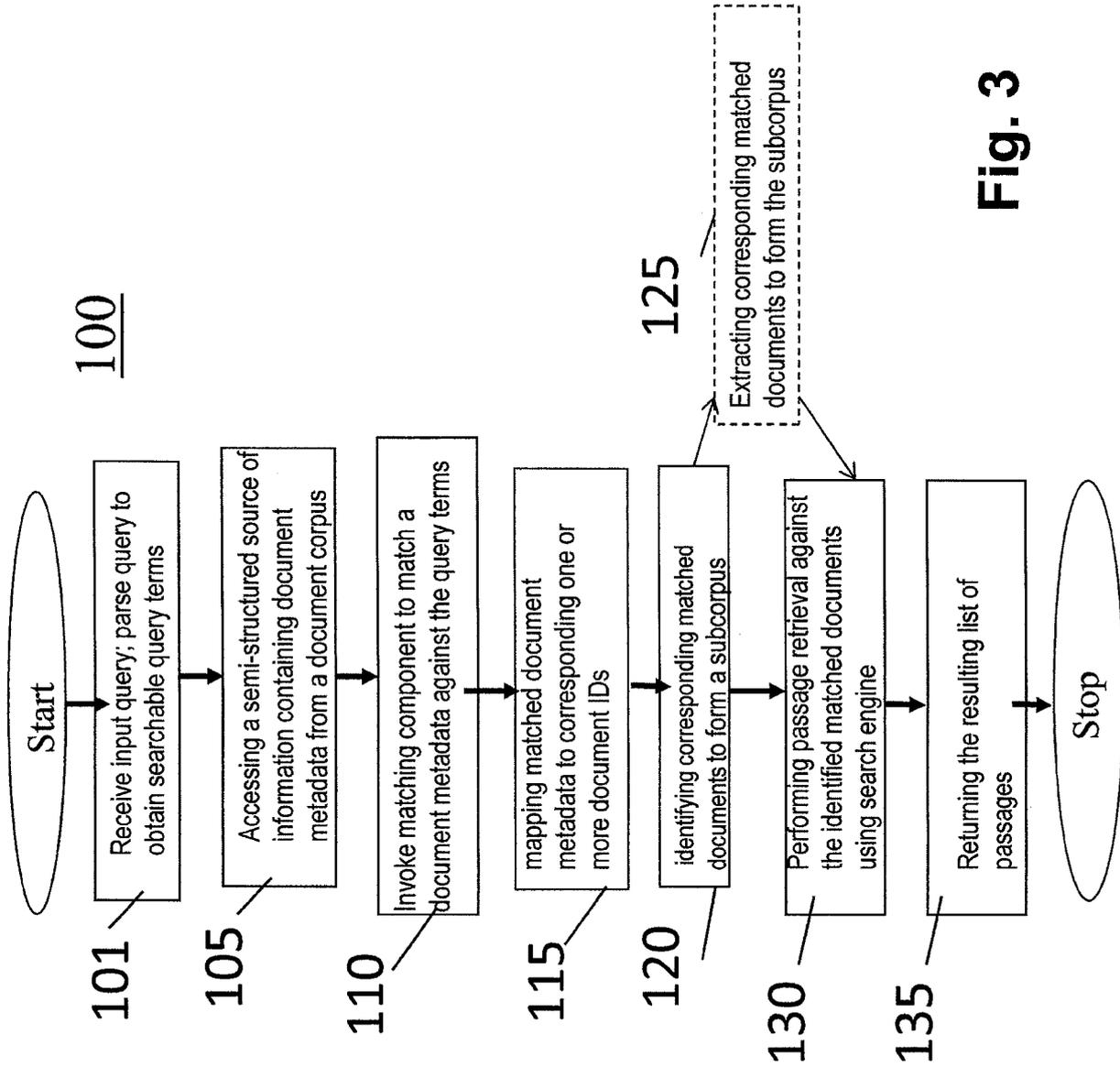
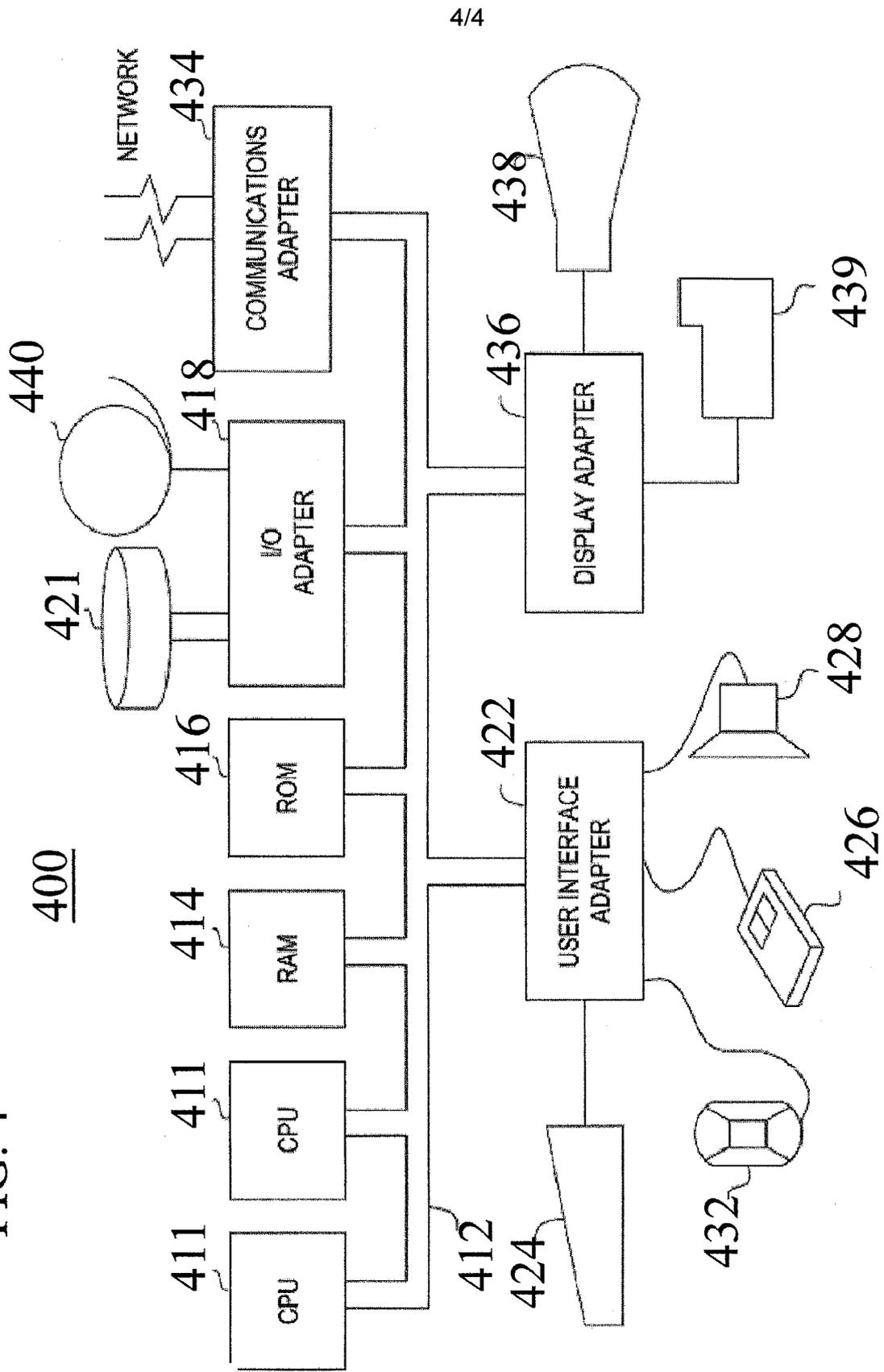


Fig. 3

FIG. 4



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/53168

| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>IPC(8) - G06F 7/00; G06F 17/00 (201 2.01 )<br>USPC - 707/769<br>According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|------------------------------------------------------------------------------|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|---|--------------------------------------------------------------------------|--------|--|
| <b>B. FIELDS SEARCHED</b><br>Minimum documentation searched (classification system followed by classification symbols)<br>IPC(8): G06F 7/00; G06F 17/00 (2012.01)<br>USPC: 707/769<br>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched<br>USPC: 707/999.003   707/999.102   707/E17.099   707/999.005   707/E17.075   707/E17.098   707/771 (keyword limited; terms below)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>PubWest; Google Scholar; Google Patents; FreePatentsOnline. Search terms used: query-answer question-response question-answer "Q/A" "Q&A" Q-and-A, search query, corpus sub-corpus collection aggregation accumulation assemblage body, term searchable-term query-term, context context-analysis query-analysis query-context-analysis, metadata...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>US 2009/0287678 A1 (BROWN et al.) 19 November 2009 (19.11.2009) entire document, especially Abstract; para [0026], [0027], [0029], [0031], [0032], [0044], [0053], [0065], [0067], [0078], [0085], [0085], [0090], [0095], [0121], [0125], [0185]</td> <td>1 - 25</td> </tr> <tr> <td>Y</td> <td>US 2002/0156846 A1 (RAWAT et al.) 24 October 2002 (24.10.2002) entire document, especially Abstract; para [0015], [0020], [0030], [0053], [0063], [0071]-[0073], [0075]</td> <td>1 - 25</td> </tr> <tr> <td>A</td> <td>US 2009/0292687 A1 (FAN et al.) 26 November 2009 (26.11.2009) entire document</td> <td>1 - 25</td> </tr> <tr> <td>A</td> <td>US 2002/0178163 A1 (MAYER) 28 November 2002 (28.11.2002) entire document</td> <td>1 - 25</td> </tr> </tbody> </table>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Category*                                                                                                                                                                                                                                         | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No.                                                                                                                                                                               | Y                                                                                                        | US 2009/0287678 A1 (BROWN et al.) 19 November 2009 (19.11.2009) entire document, especially Abstract; para [0026], [0027], [0029], [0031], [0032], [0044], [0053], [0065], [0067], [0078], [0085], [0085], [0090], [0095], [0121], [0125], [0185] | 1 - 25                                                                                    | Y                                                                                                                                                                                                                                                | US 2002/0156846 A1 (RAWAT et al.) 24 October 2002 (24.10.2002) entire document, especially Abstract; para [0015], [0020], [0030], [0053], [0063], [0071]-[0073], [0075] | 1 - 25                                        | A                                                                            | US 2009/0292687 A1 (FAN et al.) 26 November 2009 (26.11.2009) entire document | 1 - 25                                                                                                 | A | US 2002/0178163 A1 (MAYER) 28 November 2002 (28.11.2002) entire document | 1 - 25 |  |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                | Relevant to claim No.                                                              |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | US 2009/0287678 A1 (BROWN et al.) 19 November 2009 (19.11.2009) entire document, especially Abstract; para [0026], [0027], [0029], [0031], [0032], [0044], [0053], [0065], [0067], [0078], [0085], [0085], [0090], [0095], [0121], [0125], [0185] | 1 - 25                                                                             |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | US 2002/0156846 A1 (RAWAT et al.) 24 October 2002 (24.10.2002) entire document, especially Abstract; para [0015], [0020], [0030], [0053], [0063], [0071]-[0073], [0075]                                                                           | 1 - 25                                                                             |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | US 2009/0292687 A1 (FAN et al.) 26 November 2009 (26.11.2009) entire document                                                                                                                                                                     | 1 - 25                                                                             |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | US 2002/0178163 A1 (MAYER) 28 November 2002 (28.11.2002) entire document                                                                                                                                                                          | 1 - 25                                                                             |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| <table border="0"> <tr> <td>* Special categories of cited documents:</td> <td>"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</td> </tr> <tr> <td>"A" document defining the general state of the art which is not considered to be of particular relevance</td> <td>"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</td> </tr> <tr> <td>"E" earlier application or patent but published on or after the international filing date</td> <td>"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</td> </tr> <tr> <td>"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)</td> <td>"&amp;" document member of the same patent family</td> </tr> <tr> <td>"O" document referring to an oral disclosure, use, exhibition or other means</td> <td></td> </tr> <tr> <td>"P" document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </table> |                                                                                                                                                                                                                                                   | * Special categories of cited documents:                                           | "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 | "A" document defining the general state of the art which is not considered to be of particular relevance | "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                                                                      | "E" earlier application or patent but published on or after the international filing date | "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 | "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) | "&" document member of the same patent family | "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 |   |                                                                          |        |  |
| * Special categories of cited documents:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | "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                                               |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| "A" document defining the general state of the art which is not considered to be of particular relevance                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | "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                                                                      |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| "E" earlier application or patent but published on or after the international filing date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | "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  |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| "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)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | "&" document member of the same patent family                                                                                                                                                                                                     |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| "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                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                   |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| Date of the actual completion of the international search<br>03 February 2012 (03.02.2012)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Date of mailing of the international search report<br><b>22 FEB 2012</b>                                                                                                                                                                          |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |
| Name and mailing address of the ISA/US<br>Mail Stop PCT, Attn: ISA/US, Commissioner for Patents<br>P.O. Box 1450, Alexandria, Virginia 22313-1450<br>Facsimile No. 571-273-3201                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Authorized officer:<br>Lee W. Young<br>PCT Helpdesk: 571-272-4300<br>PCT OSP: 571-272-7774                                                                                                                                                        |                                                                                    |                                                                                                                                                                                                     |                                                                                                          |                                                                                                                                                                                                                                                   |                                                                                           |                                                                                                                                                                                                                                                  |                                                                                                                                                                         |                                               |                                                                              |                                                                               |                                                                                                        |   |                                                                          |        |  |