



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/CN2016/106462

(22) International Filing Date:
18 November 2016 (18.11.2016)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **MICROSOFT TECHNOLOGY LICENSING, LLC.** [US/US]; One Microsoft Way, Redmond, WA Washington 98052 (US).

(72) Inventor; and

(71) Applicant (*for SC only*): **LIU, Jing** [CN/CN]; c/o Microsoft Asia Pacific R&D, Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN).

(72) Inventors: **LIN, Chin-Yew**; c/o Microsoft Asia Pacific R&D, Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN). **CHEN, Xi**; c/o Microsoft Asia Pacific R&D, Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN). **WANG, Jinpeng**; c/o Microsoft Asia Pacific R&D,

Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN). **HUANG, Xiaojiang**; c/o Microsoft Asia Pacific R&D, Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN). **CAO, Yunbo**; c/o Microsoft Asia Pacific R&D, Headquarters 14F, Building 2, No. 5, Danling Street, Haidian District, Beijing 100080 (CN).

(74) Agent: **KING & WOOD MALLESONS**; 20th Floor, East Tower, World Financial Centre, No. 1 Dongsanhuan Zhonglu, Chaoyang District, Beijing 100020 (CN).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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,

(54) Title: SEARCH ENGINE BASED ON CITATION

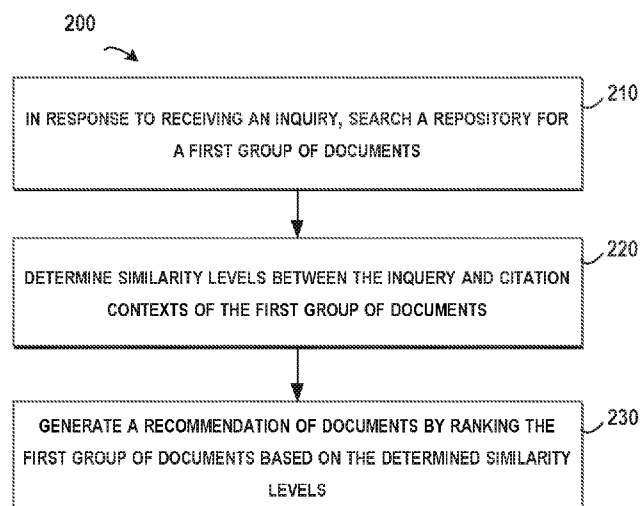


FIG. 2

(57) Abstract: An approach for generating a search result based on citation is proposed. Generally speaking, a repository is searched for a first group of documents in response to receiving an inquiry. Then similarity levels between the inquiry and citation contexts of the first group of documents are determined, where a citation context indicates a citation relationship between a respective document in the first group of documents and a target document. Next, a recommendation of documents is generated by ranking the first group of documents based on the determined similarity levels.

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

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

SEARCH ENGINE BASED ON CITATION

BACKGROUND

[0001] Data storage technology and networking technology have been becoming increasingly popular in recent years. Based on these technologies, centralized and distributed databases may provide a user with powerful storage capabilities and then the user may search these databases for a desired document via a search engine provided by a web page and/or a dedicated application.

[0002] Several approaches have been proposed to provide searching services and most of them conduct the search based on keywords matching and may return hundreds of or even more documents to the user. Then the user has to read each and every document so as to check whether the returned documents meet the user's requirement. Usually, the returned documents are ranked according to frequencies of occurrences of the keyword in the documents, where a true relevant document may be listed at the rear of the search result. As a result, the user has to spend a lot of time in going through the search result.

SUMMARY

[0003] In accordance with implementations of the subject matter described herein, a new approach for generating result based on a citation is proposed. Generally speaking, a repository is searched for a first group of documents in response to receiving an inquiry. Then similarity levels between the inquiry and citation contexts of the first group of documents are determined, where a citation context indicates a citation relationship between a respective document in the first group of documents and a target document. Next, a recommendation of documents is generated by ranking the first group of documents based on the determined similarity levels.

[0004] It is to be understood that the Summary is not intended to identify key or essential features of implementations of the subject matter described herein, nor is it intended to be used to limit the scope of the subject matter described herein. Other features of the subject matter described herein will become easily comprehensible through the description below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The details of one or more implementations of the subject matter described herein are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the disclosure will become apparent from the description, the drawings, and the claims, wherein:

5 [0006] Fig. 1 is a schematic diagram illustrating an architecture in which example implementations of the subject matter described herein may be implemented;

[0007] Fig. 2 is a flowchart illustrating a method for generating a recommendation of documents in response to an inquiry in accordance with an example implementation of the subject matter described herein;

10 [0008] Fig. 3 illustrates an example citation graph in accordance with an example implementation of the subject matter described herein;

[0009] Fig. 4A illustrates example edges in a citation graph in accordance with an example implementation of the subject matter described herein, and Fig. 4B illustrates another form of the edges in the citation graph in accordance with an example
15 implementation of the subject matter described herein;

[0010] Figs. 5A to 5C illustrate similarity graphs at respective phases for determining a search result in accordance with an example implementation of the subject matter described herein;

[0011] Fig. 6A illustrates an example similarity graph including various types of
20 documents in accordance with an example implementation of the subject matter described herein, and Fig. 6B illustrates an example similarity graph considering publication dates of the documents in accordance with an example implementation of the subject matter described herein;

[0012] Fig. 7A illustrates an example interface including a search result in response to
25 an inquiry in accordance with an example implementation of the subject matter described herein, and Fig. 7B illustrates an example interface including a list of cited laws in accordance with an example implementation of the subject matter described herein;

[0013] Fig. 8A illustrates an example interface including a recommendation list of
30 relevant legal cases in accordance with an example implementation of the subject matter described herein, and Fig. 8B illustrates an example interface including a recommendation

list of relevant laws in accordance with an example implementation of the subject matter described herein;

[0014] Fig. 9A illustrates an example interface including a recommendation list in response to a document being added into the cart in accordance with an example implementation of the subject matter described herein, and Fig. 9B illustrates an example interface including a recommendation list in response to a document being removed from the search result in accordance with an example implementation of the subject matter described herein; and

[0015] Fig. 10 is a block diagram of a device suitable for implementing one or more implementations of the subject matter described herein.

[0016] Throughout the figures, same or similar reference numbers will always indicate same or similar elements.

DETAILED DESCRIPTION

[0017] Principle of the subject matter described herein will now be described with reference to some example implementations. It is to be understood that these implementations are described only for the purpose of illustration and help those skilled in the art to understand and implement the subject matter described herein, without suggesting any limitations as to the scope of the disclosure. The disclosure described herein can be implemented in various manners other than the ones describe below.

[0018] As used herein, the term “include” and its variants are to be read as open terms that mean “include, but is not limited to”. The term “based on” is to be read as “based at least in part on”. The term “a” is to be read as “one or more” unless otherwise specified. The term “one implementation” and “an implementation” are to be read as “at least one implementation”. The term “another implementation” is to be read as “at least one other implementation”. Moreover, it is to be understood that in the context of the subject matter described herein, the terms “first”, “second” and the like are used to indicate individual elements or components, without suggesting any limitation as to the order of these elements. Further, a first element may or may not be the same as a second element. Other definitions, explicit and implicit, may be included below.

[0019] Conventionally, a search engine may be implemented in forms of a search

application, a website service and the like. In response to an inquiry from a user, the search engine may return a list of documents. The inquiry may include one or more keywords or describe a question. However, the relevance of search result may not be perfect, and the user has to spend a lot of time reading each document in the list to find the true relevant document(s). On one hand, the search result may include hundreds of or even more documents, and the user should analyze each and every document to avoid missing a true relevant document. On the other hand, even if the user reads all the documents in the list, some true relevant documents which are not included in the list may be missing.

[0020] Accordingly, it is desired to propose a more intelligent solution for ranking the returned documents based on the potential relevance between the documents and the inquiry. In order to at least partially solve the above and other potential problems, a new method and device for generating a search result for an inquiry are provided herein. According to implementations of the subject matter described herein, a repository is searched for a first group of documents in response to receiving an inquiry. Then similarity levels between the inquiry and citation contexts of the first group of documents are determined, where a citation context indicates a citation relationship between a respective document in the first group of documents and a target document. Next, a recommendation of documents may be generated by ranking the first group of documents based on the determined similarity levels.

[0021] In the proposed method and device, the citation context that indicates a citation relationship between a respective document and a target document is considered as a feature of the document found in the repository. By comparing the inquiry and the citation context, the similarity levels may be determined and then the found documents may be ranked according to the similarity levels.

[0022] Reference is first made to Fig. 1 which shows a block diagram illustrating an environment 100 in which example implementations of the subject matter described herein can be implemented. As shown, the environment 100 includes devices 110-1, 110-2, ..., 110-N (collectively referred to as "device 110") and a server 130. It is to be understood that although three devices 110 are shown, the environment 100 may include any suitable number of devices. Likewise, the environment 100 may include two or more servers 130.

[0023] A device 110 may be any suitable fixed, mobile, or wearable device. Examples of the devices 110 include, but are not limited to, cellular phones, smartphones, tablet computers, personal digital assistants (PDAs), digital watches, digital glasses, laptop computers, desktop computers, tablet computers, or the like. The devices 100 may have search applications installed and executed thereon, such as an application for accessing the server 130. In another example, the devices 100 may be installed with a browser such as Internet Explorer® and the like for accessing a search website supported by the server 130.

[0024] As shown in Fig. 1, the devices 110 and the server 130 are communicatively connected to one other via a network 120. Examples of the network 120 includes, but is not limited to, a wired or wireless network, such as a local area network (“LAN”), a metropolitan area network (“MAN”) a wide area network (“WAN”) or the Internet, a communication network, a near field communication connection or any combination thereof.

[0025] The server 130 is a device capable of processing the received inquiry and providing the search result. For example, the server 130 may host a search engine 140 which is capable of automatically providing the search result to users of the devices 110. That is, a user may have an interaction with the search engine 140 by submitting the inquiry via the search application / the website. The inquiry may include one or more keywords such as “marriage” and “marriage under California law.” Alternatively, or in addition, the inquiry may be a sentence for describing a question, such as “what is the definition of marriage?” In such interaction, the search engine 140 receives the inquiry from the device 110 and presents a search result on the device 110. The search engine 140 may access a repository (not illustrated in Fig. 1) for storing documents. The repository may reside on the server 130, or be deployed in the network 120. Alternatively, or in addition, the repository may be deployed in another place, as long as the search engine 140 may access the repository.

[0026] The search engine 140 includes a searching module 142, a determining module 144 and a generating module 146. These modules may be implemented in hardware, software or a combination thereof. For instance, in some implementations, these modules can be implemented as software modules in a computer program, which can be loaded in a memory for execution by a processing unit(s).

[0027] Some example implementations of the search engine 140 will be discussed in the following paragraphs. It is to be understood that the repository may include various types of documents such as legal documents, academic papers, patent documents, technical reports, news, and the like. In the following paragraphs, the legal documents may be taken as examples of the documents. The proposed solution may be adopted in searching for another type of document based on the specific environment of the requirements. Further, the documents may be organized in various data format such as plain text documents, multimedia documents, and so on. Meanwhile, although a search application is taken as an example for accessing the search engine 140, the search engine 140 may be accessed from a website or another tool.

[0028] In operation, upon receipt of an inquiry from the user, the device 110 (more particularly, the search application installed thereon) sends the inquiry to the server 130. The searching module 142 finds a first group of documents from the repository. The determining module 144 determines the similarity levels based on the inquiry and citation contexts of the first group of documents. In this example, a citation context indicates a citation relationship between a respective document in the first group of documents and a target document. The generating module 146 generates a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

[0029] For the sake of discussions, example implementations of the subject matter described herein will be described with reference to the environment 100. However, it is to be understood that such an environment is described merely for the purpose of illustration, without suggesting any limitations as to the scope of the subject matter described herein. For example, the ideas and principles are applicable to a stand-alone machine as well. That is, the search engine 140 may be installed and executed at least partially on a device(s) 110. In this example, the repository may also reside on the device(s) 110. Alternatively, or in addition, the repository may be accessed via the network 120.

[0030] For ease of discussions, some example implementations will be described in a scenario where a user wants to know “what is the definition of marriage?” Details of this example implementation are described with reference to Fig. 2 which shows a flowchart 200 illustrating a method for generating a recommendation of documents in response to an inquiry in accordance with an example implementation of the subject matter described

herein.

[0031] At 210, a repository is searched for a first group of documents in response to receiving an inquiry. In this step, the repository may be one or more storages for storing the documents. In the example of searching for legal documents, the repository may be a database includes laws, regulations, cases and the like. It is to be understood that the proposed implementation does not limit the algorithms of searching for the first group of documents that match the inquiry.

[0032] A variety of approaches may be used in the searching procedure. For example, a keyword(s) may be extracted from the inquiry, and then the keyword(s) may be matched with the full texts of the documents in the repository. Based on the specific requirements of the implementations, another algorithm that has been proposed or to be developed in the future may be adopted in the searching procedure. In the above example, the inquiry is "what is the definition of marriage" and the keyword "marriage" may be identified from the inquiry, for example, by Syntactic analysis. Then the documents in the repository may be searched with the keyword of "marriage," and a group of documents that contain the word "marriage" may be found.

[0033] The size of the group may be defined in advance. As an example, the top ten documents may be added into the group. Alternatively, portions or all the documents that contain "marriage" may be added into the group. It is also possible to define an appropriate data structure for storing the group of documents. In one implementation, an array illustrated as Table 1 may be utilized. In another implementation, Table 1 may include another column for storing other related information of the found documents. Although the found documents are listed in a queue, the documents in the queue may be arranged in a random order; alternatively, the documents may be ranked according to frequencies of "marriage" occurred in the respective document.

Table 1. Example Array

Title of Document
In re Marriage Cases 143 Cal. App. 4th 873 (referred to as "Doc 873")
In re Marriage of Scheppers 86 Cal. App. 4th 646 (referred to as "Doc 646")

In re Marriage Cases 43 Cal. 4th 757 (referred to as "Doc 757")
Lockyer v. City and County of San Francisco 33 Cal. 4th 1055 (referred to as "Doc 1055")
...

[0034] Table 2 illustrates a portion from "Doc 873" in Table 1, where the content of another document may be similar as that of the Doc 873.

Table 2. "Doc 873"

In re Marriage Cases 5, October, 2006 Reporter 143 Cal. App. 4th 873 49 Cal. Rptr. 3d 675 2006 Cal. App. LEXIS 1542 2006 Cal. Daily Op. Service 9451 2006 Daily Journal DAR 13485 Fullcase In re MARRIAGE CASES. [Six consolidated appeals.*] This is not to say that marriage can never be defined to include same-sex unions. As noted, civil marriage in California is based entirely on statutory law. (Lockyer, supra, 33 Cal.4th at p. 1074.) Thus, if the Legislature someday amends Family Code section 300 to omit gender references, the definition of marriage in this state will encompass same-sex unions. "The Court here does not hold marriage must remain a heterosexual institution." (Smelt v. County of Orange (C.D.Cal. 2005) 374 F. Supp. 2d 861, 878, fn. 22, vacated on another ground (9th Cir. 2006) 447 F.3d 673.) However, it is important to acknowledge the historical definition of marriage because this definition limits the precedential value of cases discussing the fundamental right to marriage. No authority binding upon us—from California appellate courts to the United States Supreme Court—has ever held or suggested that individuals have a fundamental constitutional right to enter the public institution of marriage with someone of the same sex. 16.
--

5

[0035] At 220, similarity levels between the inquiry and citation contexts of the first group of documents are determined, where the citation context indicates a citation

relationship between a respective document in the first group of documents and a target document. In this implementation, the respective document may be any document in the first group. Regarding the documents in the above Table 1, the document may be any of Doc 873, Doc 646, Doc 757, Doc 1055, ... , while the target document may be another one of Doc 873, Doc 646, Doc 757, Doc 1055, Alternatively, or in addition, the target document may be a further document not listed in Table 1.

[0036] In the context of the subject matter described herein, a citation in the document may be a reference to another document (referred to as the target document). The citation may be a description indicating a portion in the document cites another portion of the target document. In one example, the citation may be an “in-body citation,” where in Table 2, the content “Lockyer, supra, 33 Cal.4th at p. 1074” in the parentheses indicates an in-body citation; and the content “Smelt v. County of Orange (C.D.Cal. 2005) 374 F. Supp. 2d 861, 878, fn. 22, vacated on another ground (9th Cir. 2006) 447 F.3d 673” indicates another citation. Alternatively, the citation may be an abbreviated alphanumeric expression embedded in the body of the document that denotes an entry in the bibliographic references section of the work for the purpose of acknowledging the relevance of the works of others to the topic of discussion at the spot where the citation appears. Generally, the combination of both the in-body citation and the bibliographic entry constitutes what is commonly thought of as a citation (whereas bibliographic entries by themselves are not).

[0037] In the context of the subject matter described herein, the citation relationship may be explained broadly. As long as there is a citation between two documents, a citation relationship exists between these documents. In other words, the citation relationship may indicate: (1) the document in the first group is cited in the target document; or (2) the target document is cited in the document.

[0038] As the citation may relate to the further information on the concept of the document (such as important terms, viewpoints and the like) and the citation may reflect the general information of the document to a certain extent, the citation context may be taken as a brief of the document and compared with the inquiry to determine the similarity level. Table 3 illustrates an example result of the similarity level determined in this step. In the context of the subject matter described herein, the citation context may include various aspects of the citation, and various methods may be used in determining the

similarity level accordingly. Details for determining the similarity level will be described hereinafter.

Table 3. Similarity Levels of Documents

Title of Document	Similarity Level
In re Marriage Cases 143 Cal. App. 4th 873	15
In re Marriage of Scheppers 86 Cal. App. 4th 646	8
In re Marriage Cases 43 Cal. 4th 757	12
Lockyer v. City and County of San Francisco 33 Cal. 4th 1055	10
...	...

- 5 [0039] At 230, a recommendation of documents is generated by ranking the first group of documents based on the determined similarity levels. Based on the similarity levels determined at 220, the documents in the first group may be ranked and form the recommendation. Base on the similarity levels in Table 3, the found documents may be ranked and the recommendation may be illustrated as Table 4.

10

Table 4. Recommendation of Documents

Rank	Title of Document	Similarity Level
1	In re Marriage Cases 143 Cal. App. 4th 873	15
2	In re Marriage Cases 43 Cal. 4th 757	12
3	Lockyer v. City and County of San Francisco 33 Cal. 4th 1055	10
4	In re Marriage of Scheppers 86 Cal. App. 4th 646	8
...	...	...

- [0040] The general procedure in accordance with one method of the subject matter described herein has been described. As mentioned in the above paragraphs, the citation context may include rich information about the citation, and then the method of Fig. 2 may relate to several implementations. In accordance with implementations of the subject matter described herein, for a document from the first group of documents, a text fragment associated with a citation in the document may be determined based on the respective
- 15

citation context. Then the similarity level between the inquiry and the respective citation context may be determined based on a text similarity between the inquiry and the text fragment.

[0041] In this implementation, the citation context may provide the reason for citing and may be a part of the source document content. Specifically, the citation context may include a text fragment associated with a citation in the document. Referring back to Table 2, the content “Lockyer, supra, 33 Cal.4th at p. 1074” in the parentheses in the Doc 873 indicates a citation (where the Doc 873 cites a portion in the Doc 1055) and then the paragraph containing the citation may be taken as the citation context. In another example, the citation context may be a text string within a predefined arrange around the citation. The text fragment including twenty words before the citation and twenty words after the citation in the document may be identified as the citation context. Alternatively, it is possible to identify the text fragment including two sentences before the citation and two sentences after the citation as the citation context. Then, the inquiry and the text fragment may be compared and the text similarity therebetween may be used as the similarity level.

[0042] Any suitable data structure may be used to store the citation context. For instance, Table 5 below illustrates an example data structure for the citation context.

Table 5. Example Data Structure for Citation Context

Document	Target Document	Citation Context
Doc 873	Doc 1055	This is not to say that marriage can never be defined to include same-sex unions. As noted, civil marriage in California is based entirely on statutory law. (Lockyer, supra, 33 Cal.4th at p. 1074.)
.....		

[0043] It is to be understood that though the above Table 5 illustrates only one entry in the table, a plurality of entries may be stored therein, where each entry indicates a citation in the document. Further, one document may cite a plurality of target documents, and the document may cite the same target document several times at different portions in the document. At this point, the Table 5 may include multiple entries.

[0044] In another implementation, the citation context may be stored in a citation graph, where a node in the citation graph may indicate a document, an edge between a first node and a second node in the citation graph may indicate a citation relationship between a document indicated by the first node and a target document indicated by the second node.

5 Fig. 3 illustrates an example citation graph 300 in accordance with an example implementation of the subject matter described herein.

[0045] As shown in Fig. 3, nodes 310, 312, 314, and 318 may indicate the Doc 873, Doc 646, Doc 757, Doc 1055 in the above tables, and edges 320, 322, 324, 326, 328 and 330 may indicate that there is a citation between the nodes connected by one of the edges.

10 The edge 320 indicates a citation context such as the citation context associated with Doc 873 and Doc 1055. In other implementation, depending on the specific requirements and/or situations, other data structures may be used to store the citation context.

[0046] In implementations of the subject matter, multiple algorithms may be used in determining the similarity between the inquiry and the citation context. The citation
15 context may be indicated by a text fragment as shown in Table 5, then text analysis may be performed to the text strings:

- String 1 (the inquiry): “what is the definition of marriage;”
- String 2 (the citation context): “This is not to say that marriage can never be defined to include same-sex unions. As noted, civil marriage in California is
20 based entirely on statutory law. (Lockyer, supra, 33 Cal.4th at p. 1074.)”

[0047] In one simple implementation, the Strings 1 and 2 may be scanned and feature vectors 1 and 2 may be extracted from the two strings respectively. Based on a comparing of the two feature vectors 1 and 2, the similarity level may be obtained. In
25 another implementation, text and semantic analysis may be performed to the Strings 1 and 2, and then both of the text similarity and semantic similarity may be considered in determining the similarity level. In a further implementation, other algorithms such as deep learning and the like may be adopted to determining the similarity level.

[0048] Although Fig. 3 schematically illustrates only one edge from the node 310 to the
30 node 312, there may be multiple edges indicating multiple citation contexts from the node 310 to the node 312. Fig. 4A illustrates example edges in a citation graph in accordance

with an example implementation of the subject matter described herein. For simplicity, only two nodes 310 and 312 are illustrated in Fig. 4A, where there may be multiple citations (indicated by edges 410A and 320) between the documents indicated respectively by the two nodes. At this point, an individual similarity level may be determined by
5 comparing the inquiry and the citation context indicated by the edge 410A, and another individual similarity level may be determined by comparing the inquiry and the citation context indicated by the edge 320. Then, the two determined similarity levels may be added to generate a total similarity level for the ranking procedure.

[0049] In some implementations, for a document from the first group of documents, the
10 number of citations in the document may be determined based on the respective citation context; and then the similarity level between the inquiry and the respective citation context may be determined based on the number.

[0050] In the above implementation, the number of the citation between two documents may be considered, and thus the citation context may include a counter that counts the
15 number of citation. Fig. 4B illustrates another form of the edges in the citation graph 400A in accordance with an example implementation of the subject matter described herein. In Fig. 4B, the edge 410B from the node 310 to 320 shows that the document of the node 320 is cited twice in the document of the node 310. It is to be understood that the more citations exist between the two documents, the more relevance exist
20 therebetween. In one implementation, the number of the citation may be directly used as the similarity level. In another implementation, the similarity level determined based on text similarity comparison may be weighted by the number of the citation.

[0051] In the implementations described in the above paragraphs, for each of the documents that are found in the repository, a similarity level may be obtained, and then
25 the found documents may be ranked according to the similarity levels indicating matching degrees between the inquiry and the found documents, thereby forming a recommendation. In this implementation, the documents with higher matching degrees may be listed on the top of the recommendation, and the user may first read the documents listed at the top of the recommendation. Further, the user may judge whether the document is relevant to
30 his/her inquiry based on the personal judgement.

[0052] In some implementations, in response to receiving a user input that specifies a

document in the recommendation as a relevant document, the relevant document may be added into a result for the inquiry. In this implementation, if the user believes that the document well matches the inquiry and is a true relevant document, he/she may add the document into a final result. On the other hand, if the document is irrelevant according to the user, then the user may turn to the next document in the recommendation list.

[0053] Figs. 5A to 5B illustrate similarity graphs 500A and 500B at respective phases for determining a search result in accordance with an example implementation of the subject matter described herein. The structure of the similarity graph 500A in Fig. 5A is similar to that of the citation graph 300 in Fig. 3, where the difference lies in that the edges in Fig. 5A indicate the similarity levels determined by using the method described in the preceding paragraphs. Specifically, in this example, the nodes 310, 312, 314, and 318 indicate the documents (such as Doc 873, Doc 646, Doc 757, Doc 1055) and the edges 510, 512, 514, 516, 518, and 520 indicate the similarity levels determined from the corresponding citation contexts.

[0054] In one implementation, the similarity level may be further weighted based on edge distribution of the citation graph. The edge distribution reflects the number of edges originated from respective nodes. If a node has a number of edges originated from it, this may indicate that this node has more importance than another node with fewer edges. Thus the edge distribution may be used as a weighting factor for the similarity levels. For example, the similarity levels indicated by edges originated from the node may be summed up to generate an overall similarity level for the document indicated by the node. For the node 310, the overall similarity level may be determined as: $\text{SimilarityLevel}_{\text{node } 310} = 5 + 15 + 10 = 30$. For the node 312, the overall similarity level may be $\text{SimilarityLevel}_{\text{node } 312} = 15 + 8 = 23$. In another example, if the citation graph is represented by an undirected graph, then the edge distribution may reflect the number of edges connected to respective nodes.

[0055] In another implementation, if a document is cited in multiple documents, i.e. its corresponding node has multiple edges, a special score may be added to the similarity level for the document. Referring to Fig. 5A, the edges 516 and 520 indicate that the document of node 316 is cited in both the documents of nodes 310 and 312, thus the similarity level for node 316 may be increased with a special score (for example, 5 points).

[0056] In another implementation, if the number of citation is considered, the similarity level determined for each node may be weighted by the corresponding number. In a further implementation, the overall similarity level may be determined according to another method, for example, selecting a maximum from the similarity levels indicated by edges originated from the start node as the overall similarity level for the document indicated by the node. At this point, the overall similarity level for the node 310 may be 15, which is the maximum of 5, 15, and 10. With the above method, the similarity levels for all the four nodes may be determined and the four documents indicated by the four nodes may be listed in a recommendation list according to a descending order of the similarity levels. After reading the document of node 310, the user may specify the document as a relevant one, and a pin 530 may be dropped (as illustrated in Fig. 5B). At this point, the document of node 310 may be added into a result of the inquiry.

[0057] For a traditional search engine, although the engine may find hundreds of or even more documents based on the inquiry, some true relevant documents may not be returned in the search result. In this situation, even if the user reads all the returned documents, it is still possible that he/she cannot find the desired document. At this point, the user has to modify the inquiry and conduct another round of search. The implementations can take advantage of the true relevant document specified by the user in the prioritized recommendation list to start another round of search in the repository. In other words, the true relevant document may be used as a “seed” and initiate another searching for further documents in the repository, and then the found documents may be ranked according to corresponding similarity levels. Details for the further searching and ranking will be described hereinafter.

[0058] In some implementations, for a document in the result, the repository may be searched for a second group of documents based on a respective citation context of the document. Then similarity levels between the inquiry and the citation contexts of a third group of documents may be determined, where the third group includes documents both from the recommendation and the second group of documents. Next, the recommendation may be updated by ranking the third group of documents based on the determined similarity levels between the inquiry and the third group of documents.

[0059] The citation relationship between two documents may indicate a relevance degree therebetween. In this implementation, the documents that are closely relevant to

the document in the search result (which is specified by the user as the true relevant document) may be found based on the citation context. Then, the newly found documents (in the second group of documents) together with the previously found documents (in the first group of documents) may form a third group of documents. For each one in the third group of documents, a similarity level may be determined and then the documents in the third group may be ranked and the recommendation may be updated with the ranking.

[0060] Although the first and second groups of documents may have an intersection, the second group will usually include a candidate document that is not included in the first group returned from the inquiry. In this situation, the candidate document is not included in the first group because it has poor relevance with the inquiry. However, as it is closely associated with the true relevant document specified by the user, this candidate document may be considered for the recommendation. Comparing with the traditional search engine where this candidate document will never be added into the recommendation, the implementations of the subject matter use the documents in the search result as seeds to find more potential candidate documents.

[0061] Further, similar processing may be performed to each one in the third group of documents for determining the similarity levels and updating the recommendation by ranking, details of the steps will be omitted hereinafter. Fig. 5C illustrates similarity graphs during a further round of search. In this example, the document of the node 312 may be listed as the first one in the updated recommendation. In response to the user specifying the document of the node 312 as a relevant document, a pin 532 may be dropped and the document of the node 312 may be added into the search result.

[0062] In some implementations, in response to receiving a user input that specifies a document in the result as an irrelevant document, the irrelevant document may be removed from the result. In this implementation, if the user finds a document in the search result is not a desired one for some reason, then he/she may remove the unwanted document.

[0063] In the case that an irrelevant document being removed from the search result, the target document, which is found by using the citation context associated with the removed document, may be highlighted. Alternatively, it is also possible to highlight the

document that is in the second group but not in the first group. The highlighted document is found based on a precondition that the seed document is a true relevant document matching the inquiry, once the seed document is removed and labeled as an irrelevant one, the precondition collapses and then the target document may be highlighted for the user's reconsideration. Alternatively, the target document may be directed from the search result.

[0064] Additionally, in some implementations, the citation context may further comprise a citation weight that is associated with at least one of: a category of the target document and a position of the citation relationship in the document, and the determining similarity levels may further comprise: weighting the similarity level based on the citation weight.

[0065] In some implementations, the similarity level may be weighted by a citation weight indicating a potential importance factor associated with the citation. In one implementation, the category of a cited target document in a citing document may be considered for determining the citation weight. This implementation may be useful in legal documents search. For example, if the target document is for the constitution law of a country, then the citation weight may be set to a high value; and if the target document is for the common law, then the citation weight may be set to a relative low value. Accordingly, when different categories of target documents are cited by a same document, the corresponding citation weights may be different.

[0066] In one implementation, the citation weight may be associated with the position of the citation. Usually, the abstract section of a document is a brief of the document and thus may be more important than other sections of the document. Thereby, if a target document is cited in the abstract of the document, then the corresponding citation weight may be set to a high value; and if the target document is cited in another less important section of the document, then the citation weight may be set to a relative low value.

[0067] Continuing the above example, when citation weight is considered, the data structure for the citation context in Table 5 may be modified and an example is illustrated in Table 6. At this point, the similarity level determined according to the previous paragraphs may be further weighted by the citation weight.

Table 6. Example Data Structure for Citation Context

Document	Target Document	Citation Weight	Citation Context
Doc 873	Doc 1055	1.2	This is not to say that marriage can never be defined to include same-sex unions. As noted, civil marriage in California is based entirely on statutory law. (Lockyer, supra, 33 Cal.4th at p. 1074.)
.....			

[0068] Additionally, in some implementations, the recommendation may be filtered based on at least one of types and timestamps of documents. Usually, the repository may include multiple types of documents. In a legal database, there may be laws, implementing regulations, rules, cases and the like. The implementation may provide a classification to the documents in the recommendation. In one implementation, the user may search all the laws related to the inquiry in the legal database, and the user may search all the cases in the legal database. Alternatively, in some implementations, the user may search both the cases and the laws in the legal database. The type may relate to various aspects of the documents. For example, the legal documents may be classified into laws or cases; in another example, the legal documents may be classified according to the countries such as the legal documents in United State, the legal documents in China and so on; in a further example, the legal documents may be classified according to a further standard.

[0069] Fig. 6A illustrates an example similarity graph 600A including various types of documents in accordance with an example implementation of the subject matter described herein. In Fig. 6A, the nodes may be classified according to the types of the documents indicated by the nodes. Specifically, the shaded nodes 316 and 314 may indicate laws and the nodes 310 and 312 may indicate legal cases.

[0070] As the laws and cases change with the development of the society, then legal documents may also be classified according to the timestamp of the publication date of them. Fig. 6B illustrates an example similarity graph 600B considering publication dates of the documents in accordance with an example implementation of the subject matter described herein. The structure of the similarity graph may change according to the timestamp associated with the inquiry. If the inquiry specifies a timestamp of “Years:

2010 to 2016,” then the node related to a document failing to match the timestamp will be removed and the similarity graph may be illustrated as Fig. 6B.

[0071] In accordance with implementations of the subject matter described herein, the citation relationship may indicate at least one of: the target document being cited in the
5 respective document, or the respective document being cited in the target document.

[0072] Although the directed graphs are described as example citation/similarity graphs in the above implementations, this is merely for illustration without suggesting any limitations as to the scope of the subject matter described herein. In other implementations, the directions of the edges in the citation/similarity graphs may be
10 reversed to indicate that the document of the first node is cited in the target document of the second node. Further, the direction of edge is taken into account in the above implementations, in other implementations, the citation/similarity graphs may be represented by an undirected graph and the citation relationship between may be indicated by an undirected edge, as long as a document is cited in a target document or the
15 document cites the target document.

[0073] Some examples will now be described hereinafter by referring to Figs. 7A, 7B, 8A, 8B, 9A and 9B. Fig. 7A illustrates an example interface 700A including a search result in response to an inquiry in accordance with an example implementation of the subject matter described herein. As illustrated in Fig. 7A, the inquiry 710A including the
20 string “what is the definition of marriage” is inputted into the search engine. The left frame in the interface 700A is a recommendation of documents returned by using the method described above, where the document 720A is listed at the top of the recommendation due to a maximum similarity level, and the document 722A is listed as the second one in a descending order of the similarity levels.

[0074] Further, the search engine may display target documents that have citation relationship with the documents in the recommendation. When the types of the legal documents are considered, the target documents may be divided into two types: the laws (indicated by the reference number 732A) and the cases (indicated by the reference number 730A). Although Fig. 7A illustrates the recommendation in form of a list, other
25 data structures may be used in displaying the recommendation. For example, tables, icons, tabs, and other visual components may be used in the interface.

[0075] Fig. 7B illustrates an example interface 700B including a list of cited laws in accordance with an example implementation of the subject matter described herein. A list 710B of the target documents may be displayed in response to the button 732A in Fig 7A is pressed. Moreover, a button 720B may be displayed in the interface 700B to the user. In response to the button 720B being pressed, the document “In re Marriage Cases 143 Cal. App. 4th 873” may be added into the search result.

[0076] Figs. 8A and 8B illustrate the interfaces of the search engine after the Doc 873 is added into the search result. Fig. 8A illustrates an example interface 800A including a recommendation list of relevant legal cases in accordance with an example implementation of the subject matter described herein. A button 820A is provided in the interface 800A for receiving which type of documents may be displayed. In Fig. 8A, the document in the search result 810A is used as the seed document, and the list 830A in the right frame of the interface 800A shows the recommended documents obtained based on the search result 810A and the inquiry. As the user specifies that only cases may be displayed in the recommendation, the list 830A includes only candidate cases related to the inquiry.

[0077] Referring to Fig. 8B, this figure illustrates an example interface including a recommendation list of relevant laws in accordance with an example implementation of the subject matter described herein. As the type of the returned documents is set to “laws” by the button 820B, by taking the document in the search result 810B as the seed document, the list 830B shows the laws ranked according to the similarity levels.

[0078] Figs 9A and 9B illustrate the functions of adding/removing a document into/from the search result, respectively. Fig. 9A illustrates an example interface 900A including a recommendation list in response to a document being added into the search result in accordance with an example implementation of the subject matter described herein. By pressing the button 840 in Fig. 8A, another document “Perez v. Sharp, 32 Cal. 2d 711” may be added into the search result 910A. At this point, a further round of search may be performed by the search engine and the recommendation may be illustrated as the reference number 930A.

[0079] Fig. 9B illustrates an example interface 900B including a recommendation list in response to a document being removed from the cart in accordance with an example

implementation of the subject matter described herein. If the user finds an unwanted document is added into the search result 910B, the user may remove the unwanted document from the search result 910B. At this point, the repository may be searched based on the documents in the search result 910B and the document that are found by using the removed document may be highlighted or directly removed from the recommendation 930B.

[0080] It is to be understood that the above figures are only example interfaces that may implement the proposed subject matter. Those skilled in the art may design other interfaces to achieve the functions of the subject matter described herein.

[0081] Although the above paragraphs describe the implementations by considering the user interaction with the search engine, in another implementation, the user interaction is not necessary step. Instead, a rule may be defined that the top three documents in the recommendation may be added into the result of the inquiry, and the search engine will initiate a new round of search in the repository. Further, another rule may define that the search may stop if the size of the result reaches a predefined number.

[0082] In addition, though documents and inquiries written in English are described as examples in the above implementations, it is to be understood that the proposed search engine may be modified to conduct the search in a repository including documents in other languages such as Chinese, Japanese, French, Russian, Spanish, German, and other languages. Meanwhile, the subject matter does not limit the language of the inquiry, and the inquiries may be written in other languages. In one implementation, the documents and the inquiry may even be written in different languages.

[0083] Fig. 10 is a block diagram of a device suitable for implementing one or more implementations of the subject matter described herein. It is to be understood that the device 1000 is not intended to suggest any limitation as to scope of use or functionality of the subject matter described herein, as various implementations may be implemented in diverse general-purpose or special-purpose computing environments.

[0084] As shown, the device 1000 includes at least one processing unit (or processor) 1010 and a memory 1020. The processing unit 1010 executes computer-executable instructions and may be a real or a virtual processor. In a multi-processing system, multiple processing units execute computer-executable instructions to increase processing

power. The memory 1020 may be volatile memory (e.g., registers, cache, RAM), non-volatile memory (e.g., ROM, EEPROM, flash memory), or some combination thereof.

5 [0085] In the example shown in Fig. 10, the device 1000 further includes storage 1030, one or more input devices 1040, one or more output devices 1050, and one or more communication connections 1060. An interconnection mechanism (not shown) such as a bus, controller, or network interconnects the components of the device 1000. Typically, operating system software (not shown) provides an operating environment for other software executing in the device 1000, and coordinates activities of the components of the
10 device 1000.

[0086] The storage 1030 may be removable or non-removable, and may include computer-readable storage media such as flash drives, magnetic disks or any other medium which can be used to store information and which can be accessed within the device 1000. The input device(s) 1040 may be one or more of various different input
15 devices. For example, the input device(s) 1040 may include a user device such as a mouse, keyboard, trackball, etc. The input device(s) 1040 may implement one or more natural user interface techniques, such as speech recognition or touch and stylus recognition. As other examples, the input device(s) 1040 may include a scanning device; a network adapter; or another device that provides input to the device 1000. The output
20 device(s) 1050 may be a display, printer, speaker, network adapter, or another device that provides output from the device 1000. The input device(s) 1040 and output device(s) 1050 may be incorporated in a single system or device, such as a touch screen or a virtual reality system.

[0087] The communication connection(s) 1060 enables communication over a
25 communication medium to another computing entity. Additionally, functionality of the components of the device 1000 may be implemented in a single computing machine or in multiple computing machines that are able to communicate over communication connections. Thus, the device 1000 may operate in a networked environment using logical connections to one or more other servers, network PCs, or another common
30 network node. By way of example, and not limitation, communication media include wired or wireless networking techniques.

[0088] In accordance with implementations of the subject matter described herein, a search engine 140 may be executed on the device 1000 to provide ranked documents in response to an inquiry to a document repository.

5 [0089] Now only for the purpose of illustration, some example implemented will be listed below.

[0090] In some implementations, the subject matter described herein may be embodied as a device. The device comprises a processing unit and a memory. The memory is coupled to the processing unit and stores instructions for execution by the processing unit. The instructions, when executed by the processing unit, cause the device to perform acts
10 comprising: in response to receiving an inquiry, searching a repository for a first group of documents; determining similarity levels between the inquiry and citation contexts of the first group of documents, a citation context indicating a citation relationship between a respective document in the first group of documents and a target document; and
15 generating a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

[0091] In some implementations, the determining similarity levels comprises: for a document from the first group of documents, determining a text fragment associated with a citation in the document based on the respective citation context; and determining the
20 similarity level between the inquiry and the respective citation context based on a text similarity between the inquiry and the text fragment.

[0092] In some implementations, the determining similarity levels comprises: for a document from the first group of documents, determining the number of citations in the document based on the respective citation context; and determining the similarity level between the inquiry and the respective citation context based on the number.

25 [0093] In some implementations, the determining similarity levels comprises: obtaining the citation contexts from a citation graph, a node in the citation graph indicates a document, an edge between a first node and a second node in the citation graph indicates a citation relationship between a document indicated by the first node and a target document indicated by the second node; and determining the similarity based on edge distribution of
30 the citation graph.

[0094] In some implementations, the acts further comprise: in response to receiving a

user input that specifies a document in the recommendation as a relevant document, adding the relevant document into a result for the inquiry.

[0095] In some implementations, the acts further comprise: for a document in the result, searching the repository for a second group of documents based on a respective citation context of the document; determining similarity levels between the inquiry and the citation contexts of a third group of documents that includes the recommendation and the second group of documents; and updating the recommendation by ranking the third group of documents based on the determined similarity levels between the inquiry and the third group of documents.

[0096] In some implementations, the acts further comprise: in response to receiving a user input that specifies a document in the result as an irrelevant document, removing the irrelevant document from the result.

[0097] In some implementations, the acts further comprise: filtering the recommendation based on at least one of types and timestamps of documents.

[0098] In some implementations, the citation relationship indicates at least one of: the target document being cited in the respective document, or the respective document being cited in the target document.

[0099] In some implementations, the citation context further comprises a citation weight that is associated with at least one of: a category of the target document and a position of the citation relationship in the document, and the determining similarity levels may further comprise: weighting the similarity level based on the citation weight.

[00100] In some implementations, the subject matter described herein may be embodied as a computer-implemented method comprising: in response to receiving an inquiry, searching a repository for a first group of documents; determining similarity levels between the inquiry and citation contexts of the first group of documents, a citation context indicating a citation relationship between a respective document in the first group of documents and a target document; and generating a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

[00101] In some implementations, the determining similarity levels comprises: for a document from the first group of documents, determining a text fragment associated with

a citation in the document based on the respective citation context; and determining the similarity level between the inquiry and the respective citation context based on a text similarity between the inquiry and the text fragment.

5 [00102] In some implementations, the determining similarity levels comprises: for a document from the first group of documents, determining the number of citations in the document based on the respective citation context; and determining the similarity level between the inquiry and the respective citation context based on the number.

10 [00103] In some implementations, the determining similarity levels comprises: obtaining the citation contexts from a citation graph, a node in the citation graph indicates a document, an edge between a first node and a second node in the citation graph indicates a citation relationship between a document indicated by the first node and a target document indicated by the second node; and determining the similarity based on edge distribution of the citation graph.

15 [00104] In some implementations, the method further comprises: in response to receiving a user input that specifies a document in the recommendation as a relevant document, adding the relevant document into a result for the inquiry.

20 [00105] In some implementations, the method further comprises: for a document in the result, searching the repository for a second group of documents based on a respective citation context of the document; determining similarity levels between the inquiry and the citation contexts of a third group of documents that includes the recommendation and the second group of documents; and updating the recommendation by ranking the third group of documents based on the determined similarity levels between the inquiry and the third group of documents.

25 [00106] In some implementations, the method further comprises: in response to receiving a user input that specifies a document in the result as an irrelevant document, removing the irrelevant document from the result.

[00107] In some implementations, the method further comprises: filtering the recommendation based on at least one of types and timestamps of documents.

30 [00108] In some implementations, the citation relationship indicates at least one of: the target document being cited in the respective document, or the respective document being

cited in the target document.

[00109] In some implementations, the citation context may further comprise a citation weight that is associated with at least one of: a category of the target document and a position of the citation relationship in the document, and the determining similarity levels may further comprise: weighting the similarity level based on the citation weight.

[00110] In some implementations, the subject matter described herein may be embodied as a computer program product. The computer program product may be tangibly stored on a non-transient machine-readable medium and comprises machine-executable instructions. The instructions, when executed on an electronic device, cause the electronic device to: in response to receiving an inquiry, search a repository for a first group of documents; determine similarity levels between the inquiry and citation contexts of the first group of documents, a citation context indicating a citation relationship between a respective document in the first group of documents and a target document; and generate a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

[00111] In some implementations, the instructions further cause the electronic device to: for a document from the first group of documents, determine a text fragment associated with a citation in the document based on the respective citation context; and determine the similarity level between the inquiry and the respective citation context based on a text similarity between the inquiry and the text fragment.

[00112] In some implementations, the instructions further cause the electronic device to: for a document from the first group of documents, determine the number of citations in the document based on the respective citation context; and determine the similarity level between the inquiry and the respective citation context based on the number.

[00113] In some implementations, the instructions further cause the electronic device to: obtain the citation contexts from a citation graph, a node in the citation graph indicates a document, an edge between a first node and a second node in the citation graph indicates a citation relationship between a document indicated by the first node and a target document indicated by the second node; and determine the similarity based on edge distribution of the citation graph.

[00114] In some implementations, the instructions further cause the electronic device to:

in response to receiving a user input that specifies a document in the recommendation as a relevant document, add the relevant document into a result for the inquiry.

[00115] In some implementations, the instructions further cause the electronic device to: for a document in the result, search the repository for a second group of documents based on a respective citation context of the document; determine similarity levels between the inquiry and the citation contexts of a third group of documents that includes the recommendation and the second group of documents; and update the recommendation by ranking the third group of documents based on the determined similarity levels between the inquiry and the third group of documents.

[00116] In some implementations, the instructions further cause the electronic device to: in response to receiving a user input that specifies a document in the result as an irrelevant document, remove the irrelevant document from the result.

[00117] In some implementations, the instructions further cause the electronic device to: filter the recommendation based on at least one of types and timestamps of documents.

[00118] In some implementations, the citation relationship indicates at least one of: the target document being cited in the respective document, or the respective document being cited in the target document.

[00119] In some implementations, the citation context may further comprise a citation weight that is associated with at least one of: a category of the target document and a position of the citation relationship in the document, and the instructions further cause the electronic device to weight the similarity level based on the citation weight.

[00120] In general, the various example implementations may be implemented in hardware or special purpose circuits, software, logic or any combination thereof. Some aspects may be implemented in hardware, while other aspects may be implemented in firmware or software which may be executed by a controller, microprocessor or other computing device. While various aspects of the example implementations of the subject matter described herein are illustrated and described as block diagrams, flowcharts, or using some other pictorial representation, it will be to be understood that the blocks, apparatus, systems, techniques or methods described herein may be implemented in, as non-limiting examples, hardware, software, firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination

thereof.

[00121] In the context of the subject matter described herein, a machine readable medium may be any tangible medium that can contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device. The machine readable medium may be a machine readable signal medium or a machine readable storage medium. A machine readable medium may include 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 of the machine readable storage medium would include 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.

[00122] Computer program code for carrying out methods of the subject matter described herein may be written in any combination of one or more programming languages. These computer program codes may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus, such that the program codes, when executed by the processor of the computer or other programmable data processing apparatus, cause the functions or operations specified in the flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a computer, partly on the computer, as a stand-alone software package, partly on the computer and partly on a remote computer or entirely on the remote computer or server.

[00123] Further, while operations are depicted in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Likewise, while several specific implementation details are contained in the above discussions, these should not be construed as limitations on the scope of any disclosure or of what may be claimed, but rather as descriptions of features that may be specific to particular implementations of particular disclosures. Certain features that are

described in this specification in the context of separate implementations can also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation can also be implemented in multiple implementations separately or in any suitable sub-combination.

- 5 [00124] Various modifications, adaptations to the foregoing example implementations of this disclosure may become apparent to those skilled in the relevant arts in view of the foregoing description, when read in conjunction with the accompanying drawings. Any and all modifications will still fall within the scope of the non-limiting and example implementations of this disclosure. Furthermore, other implementations of the
- 10 disclosures set forth herein will come to mind to one skilled in the art to which these implementations of the disclosure pertain having the benefit of the teachings presented in the foregoing descriptions and the drawings.

- [00125] Therefore, it will be to be understood that the implementations of the disclosure are not to be limited to the specific implementations disclosed and that modifications and
- 15 other implementations are intended to be included within the scope of the appended claims. Although specific terms are used herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

WHAT IS CLAIMED IS:

1. A device comprising:

a processing unit; and

a memory coupled to the processing unit and storing instructions for execution by
5 the processing unit, the instructions, when executed by the processing unit, causing the device to perform acts comprising:

in response to receiving an inquiry, searching a repository for a first group
of documents;

determining similarity levels between the inquiry and citation contexts of
10 the first group of documents, a citation context indicating a citation relationship between a respective document in the first group of documents and a target document; and

generating a recommendation of documents by ranking the first group of
documents based on the determined similarity levels.

15 2. The device of claim 1, the acts further comprising:

in response to receiving a user input that specifies a document in the
recommendation as a relevant document, adding the relevant document into a result for
the inquiry.

20 3. The device of claim 2, the acts further comprising:

for a document in the result, searching the repository for a second group of
documents based on a respective citation context of the document;

determining similarity levels between the inquiry and the citation contexts of a
third group of documents that comprises the recommendation and the second group of
25 documents; and

updating the recommendation by ranking the third group of documents based on
the determined similarity levels between the inquiry and the third group of documents.

4. The device of claim 2, the acts further comprising:

30 in response to receiving a user input that specifies a document in the result as an irrelevant document, removing the irrelevant document from the result.

5. The device of claim 1, wherein the determining similarity levels comprises:
for a document from the first group of documents,

determining a text fragment associated with a citation in the document
based on the respective citation context; and

5 determining the similarity level between the inquiry and the respective
citation context based on a text similarity between the inquiry and the text fragment.

6. The device of claim 1, wherein the determining similarity levels comprises:
for a document from the first group of documents,

10 determining the number of citations in the document based on the
respective citation context; and

determining the similarity level between the inquiry and the respective
citation context based on the number.

15 7. The device of claim 1, wherein the determining similarity levels comprises:

obtaining the citation contexts from a citation graph, a node in the citation graph
indicating a document, an edge between a first node and a second node in the citation
graph indicating a citation relationship between a document indicated by the first node and
a target document indicated by the second node; and

20 weighting the similarity based on edge distribution of the citation graph.

8. The device of claim 1, wherein the citation context further comprises a
citation weight that is associated with at least one of: a category of the target document
and a position of the citation relationship in the document, and the determining similarity
25 levels further comprises:

weighting the similarity level based on the citation weight.

9. The device of claim 1, the acts further comprising:

30 filtering the recommendation based on at least one of types and timestamps of
documents.

10. The device of claim 1, wherein the citation relationship indicates at least one of: the target document being cited in the respective document, or the respective document being cited in the target document.

5 11. A computer-implemented method comprising:
in response to receiving an inquiry, searching a repository for a first group of documents;

determining similarity levels between the inquiry and citation contexts of the first group of documents, a citation context indicating a citation relationship between a
10 respective document in the first group of documents and a target document; and
generating a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

12. The method of claim 11, further comprising:
15 in response to receiving a user input that specifies a document in the recommendation as a relevant document, adding the relevant document into a result for the inquiry.

13. The method of claim 12, further comprising:
20 for a document in the result, searching the repository for a second group of documents based on a respective citation context of the document;
determining similarity levels between the inquiry and the citation contexts of a third group of documents that comprises the recommendation and the second group of documents; and
25 updating the recommendation by ranking the third group of documents based on the determined similarity levels between the inquiry and the third group of documents.

14. The method of claim 12, further comprising:
in response to receiving a user input that specifies a document in the result as an
30 irrelevant document, removing the irrelevant document from the result.

15. The method of claim 11, wherein the determining similarity levels comprises:

for a document from the first group of documents,

determining a text fragment associated with a citation in the document

5 based on the respective citation context; and

determining the similarity level between the inquiry and the respective citation context based on a text similarity between the inquiry and the text fragment.

16. The method of claim 11, wherein the determining similarity levels comprises:

for a document from the first group of documents,

determining the number of citations in the document based on the respective citation context; and

15 determining the similarity level between the inquiry and the respective citation context based on the number.

17. The method of claim 11, wherein the determining similarity levels comprises:

20 obtaining the citation contexts from a citation graph, a node in the citation graph indicating a document, an edge between a first node and a second node in the citation graph indicating a citation relationship between a document indicated by the first node and a target document indicated by the second node; and

weighting the similarity based on edge distribution of the citation graph.

25 18. The method of claim 11, wherein the citation context further comprises a citation weight that is associated with at least one of: a category of the target document and a position of the citation relationship in the document, and the determining similarity levels further comprises:

weighting the similarity level based on the citation weight.

30

19. The method of claim 11, further comprising:

filtering the recommendation based on at least one of types and timestamps of documents.

20. A computer program product being tangibly stored on a non-transient machine-readable medium and comprising machine-executable instructions, the instructions, when executed on an electronic device, causing the electronic device to:

search a repository for a first group of documents in response to receiving an inquiry;

determine similarity levels between the inquiry and citation contexts of the first group of documents, a citation context indicating a citation relationship between a respective document in the first group of documents and a target document; and

generate a recommendation of documents by ranking the first group of documents based on the determined similarity levels.

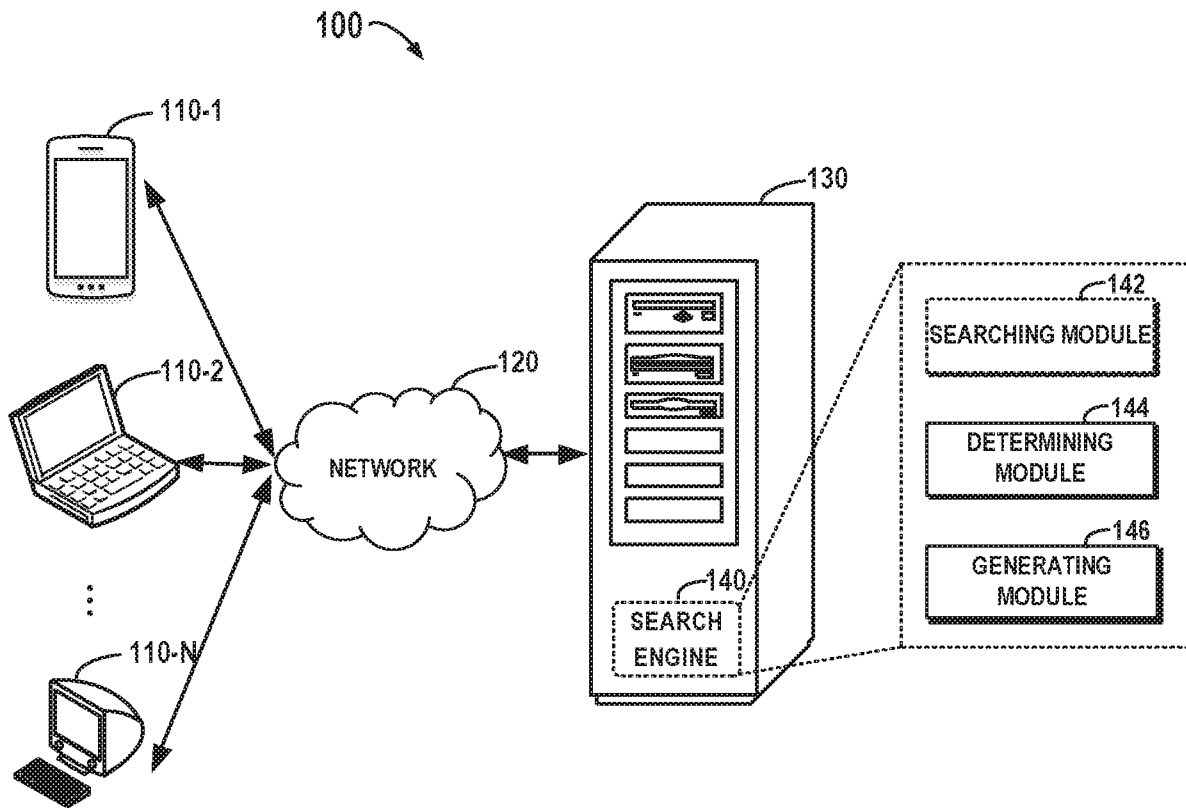


FIG. 1

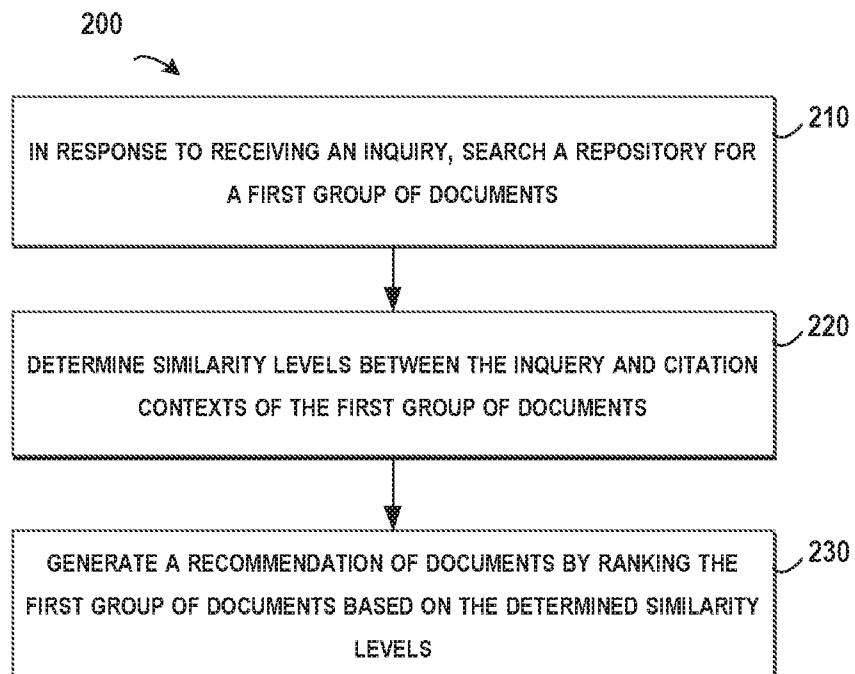


FIG. 2

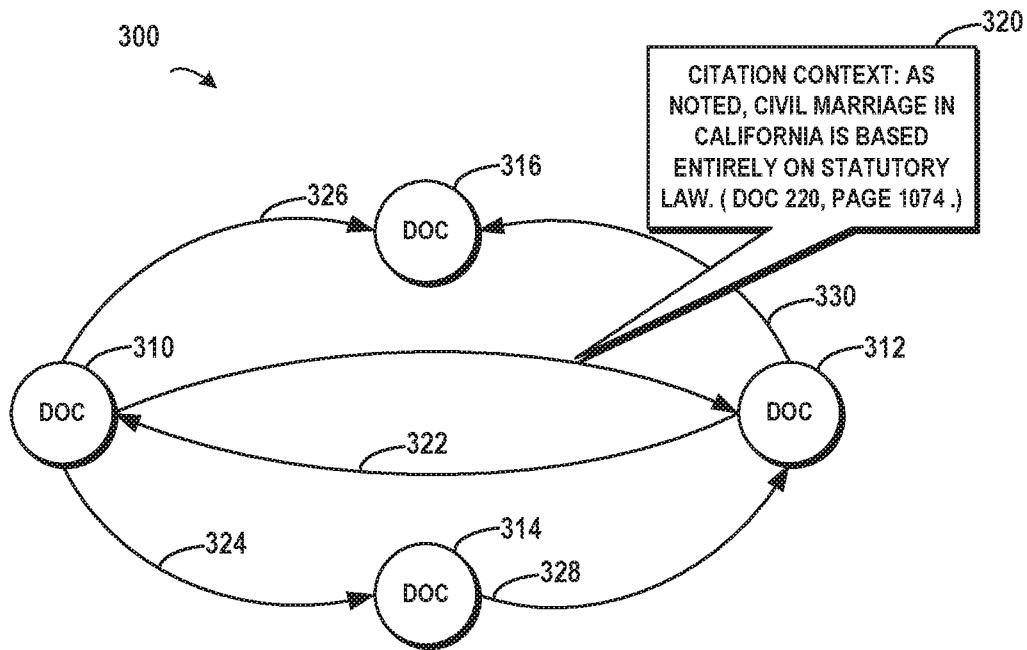


FIG. 3

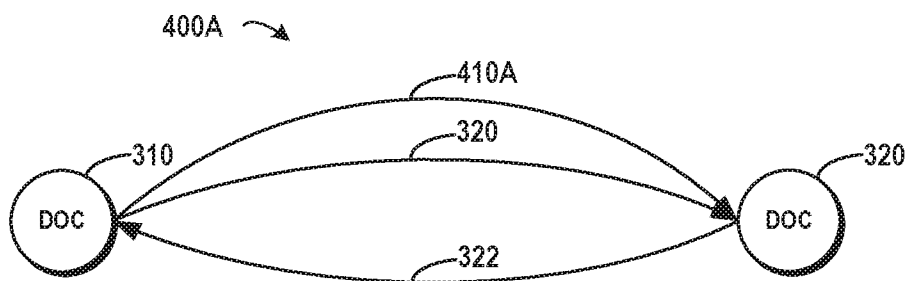


FIG. 4A

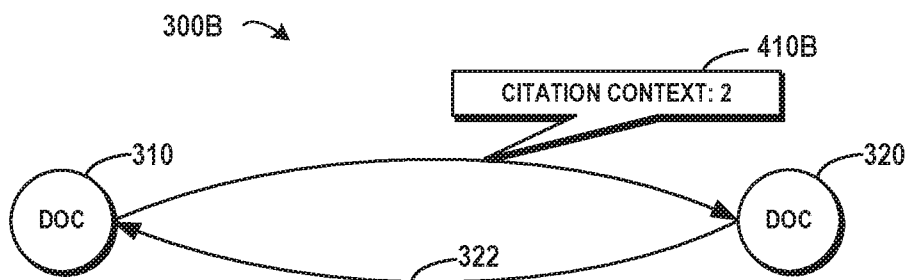


FIG. 4B

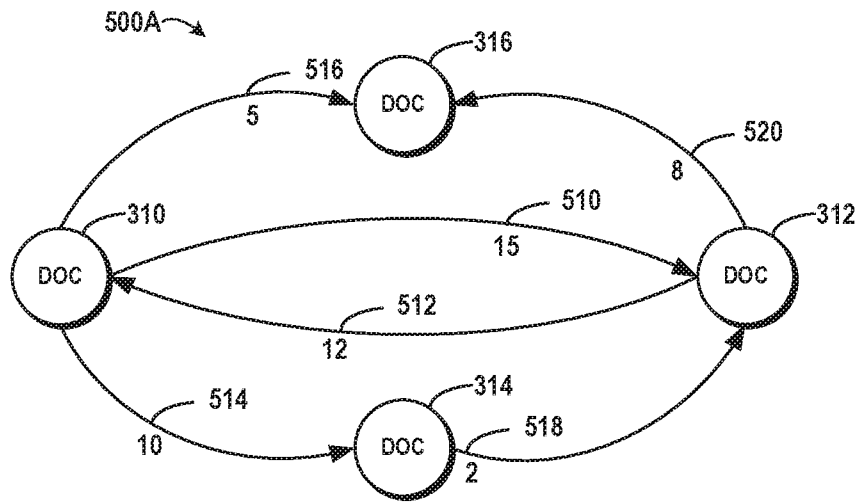


FIG. 5A

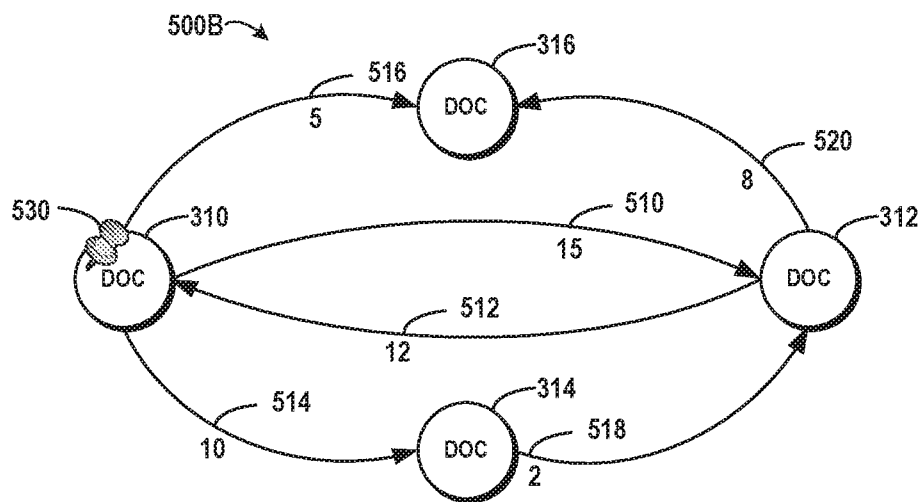


FIG. 5B

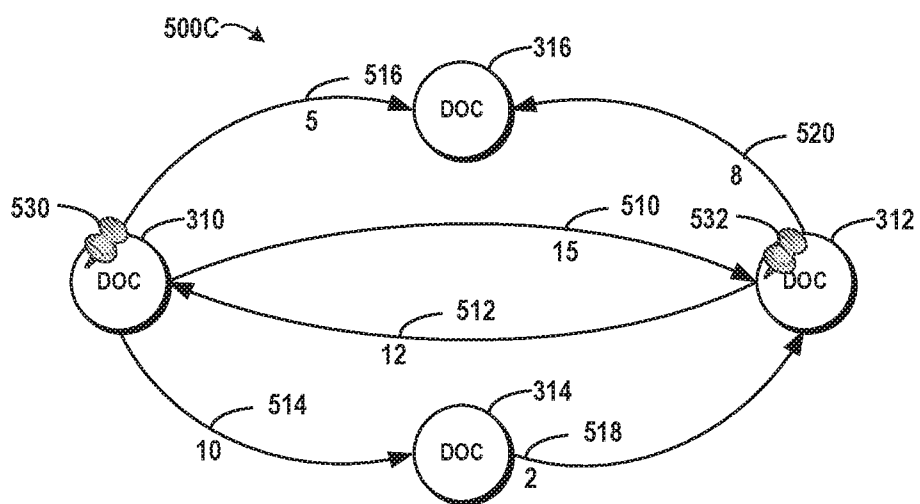


FIG. 5C

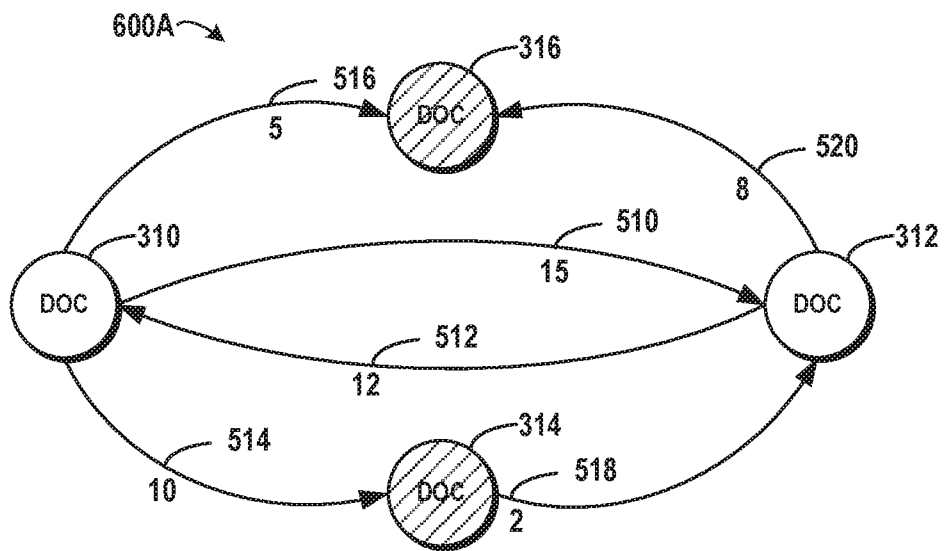


FIG. 6A

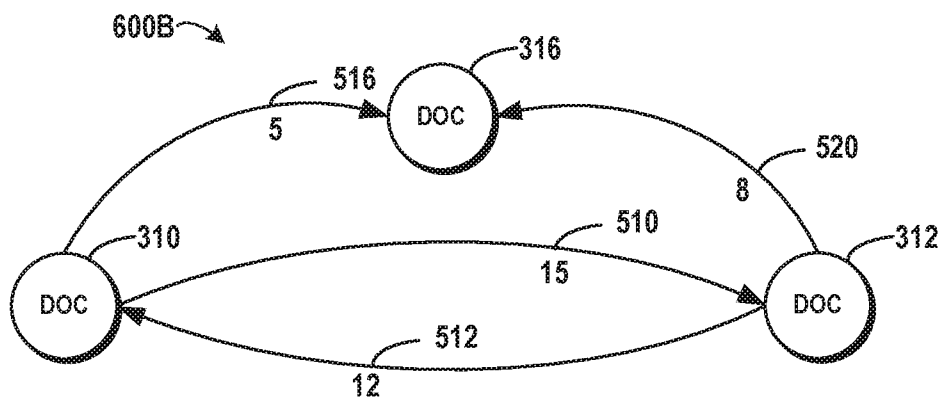


FIG. 6B



FIG. 7A



FIG. 7B

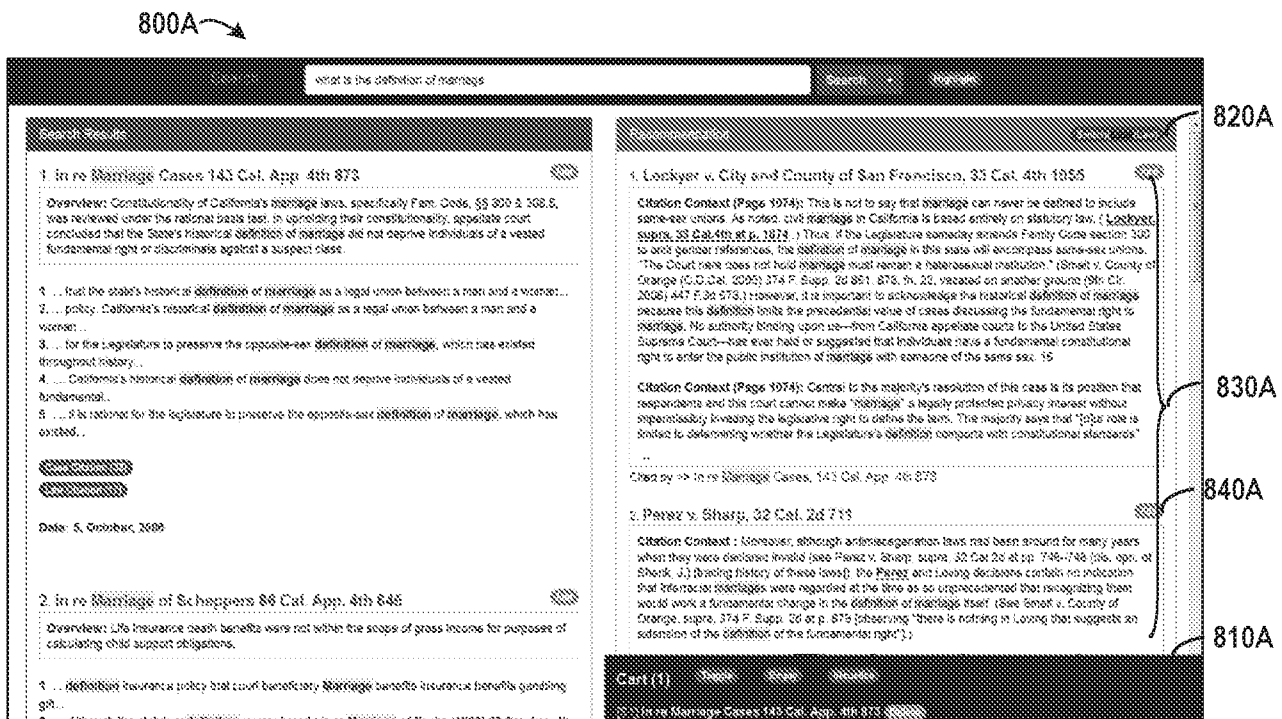


FIG. 8A

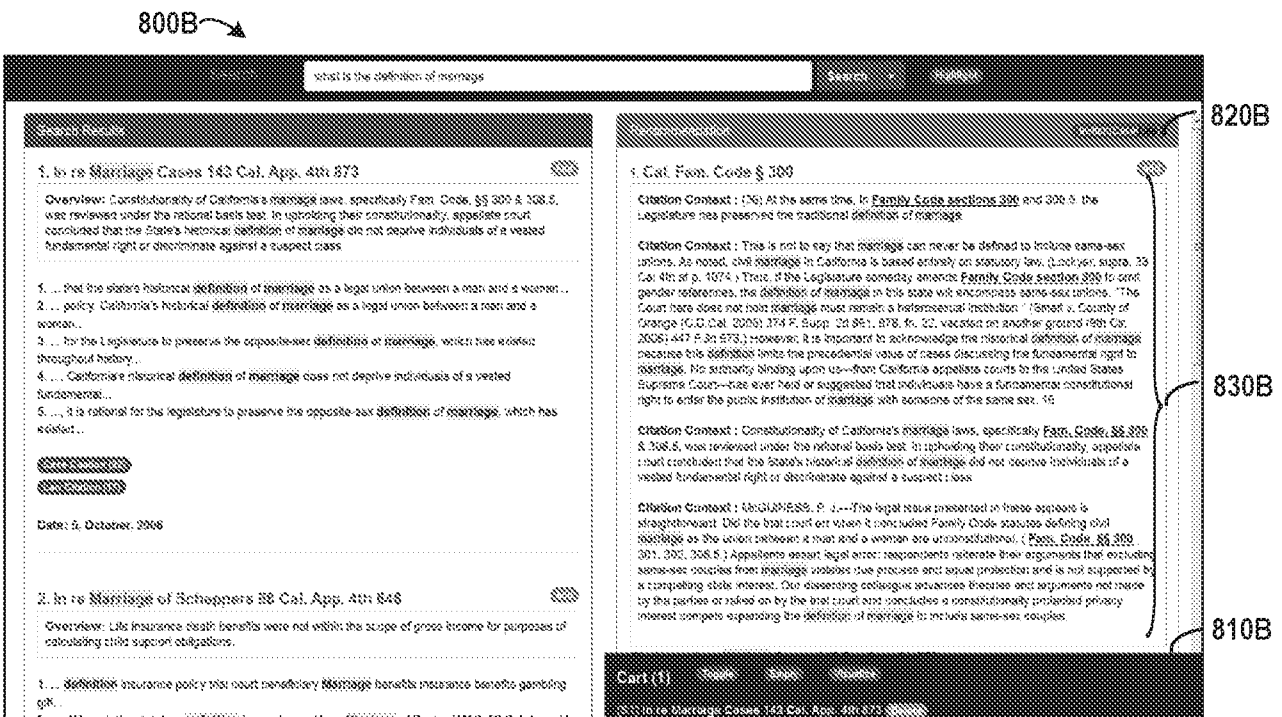


FIG. 8B

900A

what is the definition of marriage

Search

900A

1. In re Marriage Cases 143 Cal. App. 4th 873

Overview: Constitutionality of California's marriage laws, specifically Fam. Code §§ 300 & 308.5, was reviewed under the rational basis test. In upholding their constitutionality, appellate court concluded that the State's historical definition of marriage did not deprive individuals of a vested fundamental right or discriminate against a suspect class.

1. ... that the state's historical definition of marriage as a legal union between a man and a woman...

2. ... policy. California's historical definition of marriage as a legal union between a man and a woman...

3. ... for the Legislature to preserve the opposite-sex definition of marriage, which has existed throughout history...

4. ... California's historical definition of marriage does not deprive individuals of a vested fundamental...

5. ... it is rational for the legislature to preserve the opposite-sex definition of marriage, which has existed...

Date: 5. October, 2006

2. In re Marriage of Scheppers 86 Cal. App. 4th 848

Overview: Life insurance death benefits were not within the scope of gross income for purposes of calculating child support obligations.

1. ... definition insurance policy that court beneficiary marriage benefits insurance benefits gambling...

2. Although the statutory definition of gross income is in the Marriage of Burke (1997) 67 Cal. App. 4th...

3. Brown v. Merit, 8 Cal. 3d 836

Citation Context (Page 882): Some respondents claim the state's interest in preserving the

Cart (2)

1. Perez v. Sharp, 32 Cal. 2d 711

2. In re Marriage Cases 143 Cal. App. 4th 873

930A

910A

FIG. 9A

900B

what is the definition of marriage

Search

900B

1. In re Marriage Cases 143 Cal. App. 4th 873

Overview: Constitutionality of California's marriage laws, specifically Fam. Code §§ 300 & 308.5, was reviewed under the rational basis test. In upholding their constitutionality, appellate court concluded that the State's historical definition of marriage did not deprive individuals of a vested fundamental right or discriminate against a suspect class.

1. ... that the state's historical definition of marriage as a legal union between a man and a woman...

2. ... policy. California's historical definition of marriage as a legal union between a man and a woman...

3. ... for the Legislature to preserve the opposite-sex definition of marriage, which has existed throughout history...

4. ... California's historical definition of marriage does not deprive individuals of a vested fundamental...

5. ... it is rational for the legislature to preserve the opposite-sex definition of marriage, which has existed...

Date: 5. October, 2006

2. In re Marriage of Scheppers 86 Cal. App. 4th 848

Overview: Life insurance death benefits were not within the scope of gross income for purposes of calculating child support obligations.

1. ... definition insurance policy that court beneficiary marriage benefits insurance benefits gambling...

2. Although the statutory definition of gross income is in the Marriage of Burke (1997) 67 Cal. App. 4th...

3. Brown v. Merit, 8 Cal. 3d 836

Citation Context (Page 745): In Perez v. Sharp (1948) 32 Cal.2d 711 (198 P.2d 17), we struck down a law prohibiting interracial marriages. The majority placed great reliance on the Perez court's statement that "the right to marry is the right to join in marriage with the person of one's choice." (Id. at p. 715.) However, Perez and the many other cases establishing the fundamental right to marry were all based on the common understanding of marriage as the union of a man and a woman. (See maj. opn., ante, at pp. 813-815.) The majority recognizes this, as it must. (Id. at p. 820.) Because those cases involved the traditional definition of marriage, they do not support the majority's analysis. The question here is whether the meaning of the term as it was used in those cases must be changed.

Citation Context: In Perez v. Sharp (1948) 32 Cal.2d 711 (198 P.2d 17), we struck down a law prohibiting interracial marriages. The majority places great reliance on the Perez court's statement that "the right to marry is the right to join in marriage with the person of one's choice." (Id. at p. 715.) However, Perez and the many other cases establishing the fundamental right to marry were all based on the common understanding of marriage as the union of a man and a woman. (See maj. opn., ante, at pp. 813-815.) The majority recognizes this, as it must. (Id. at p. 820.) Because those cases involved the traditional definition of marriage, they do not support the majority's analysis. The question here is whether the meaning of the term as it was used in those cases must be changed.

Citation Context: Certainly Perez v. Sharp (1948) 32 Cal.2d 711 (198 P.2d 17) (Perez) does not support the majority's expansive view. There we struck down racial restrictions on the right of a man and a woman to marry. But nothing in Perez suggests an intent to alter the definition of marriage as a union of opposite-sex partners. In sum, there is no convincing basis in federal or California jurisprudence for the majority's claim that same-sex couples have a fundamental constitutional right to marry. 10

Citation Context: The circumstance that in the present instance the statutory limitation upon who may enter into the marriage relationship is contained in statutory provisions that may be viewed as defining the marriage relationship, rather than, for example, in a separate statutory provision stating that a marriage between persons of the same sex is void, does not render this aspect of the

Cart (1)

1. Perez v. Sharp, 32 Cal. 2d 711

930B

910B

FIG. 9B

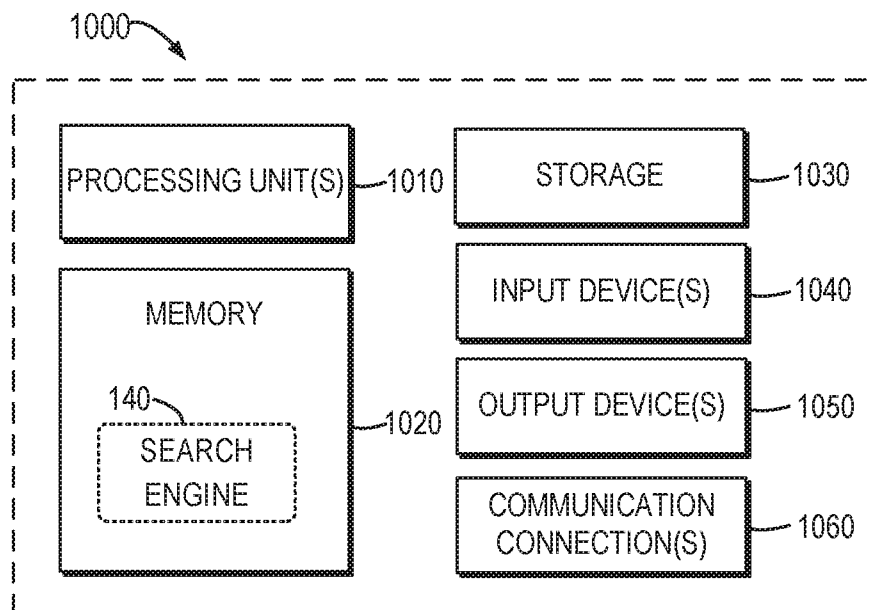


FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/106462

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

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

CNPAT, CNKI, EPODOC, WPI, IEEE: search, engine, query, inquiry, text, fragment, cit+, link, similarity, rank

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHANG, Ling et al. "CALA: A Web Analysis Algorithm Combined with Content Correlation Analysis Method" <i>Journal of Computer Science and Technology</i> , Vol. 18, No. 1, 31 January 2003 (2003-01-31), pp. 114-117	1-20
X	LIU, Jingjing et al. "Hyperlink Algorithm Based on Anchor Texts Similarity" <i>J. of Zhengzhou Univ. (Nat. Sci. Ed.)</i> , Vol. 39, No. 2, 30 June 2007 (2007-06-30), pp. 96-99	1-20
X	LIU, Jingjing et al. "Study on ranking Web pages based on pagerank and anchor text" <i>Computer Engineering and Applications</i> , Vol. 43, No. 10, 30 April 2007 (2007-04-30), pp. 170-173	1-20
A	CN 101526956 A (TSINGHUA UNIVERSITY) 09 September 2009 (2009-09-09) the whole document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

10 August 2017

Date of mailing of the international search report

18 August 2017

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

XU, Shuxian

Telephone No. (86-10)010-61648479

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/106462

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	101526956	A	09 September 2009	None	