



US 20060149800A1

(19) **United States**

(12) **Patent Application Publication**

Egnor et al.

(10) **Pub. No.: US 2006/0149800 A1**

(43) **Pub. Date: Jul. 6, 2006**

(54) **AUTHORITATIVE DOCUMENT
IDENTIFICATION**

(22) Filed: **Dec. 30, 2004**

(76) Inventors: **Daniel Egnor**, New York, NY (US);
Geeta Chaudhry, West Lebanon, NH
(US)

Publication Classification

(51) **Int. Cl.**
G06F 12/00 (2006.01)

G06F 17/30 (2006.01)

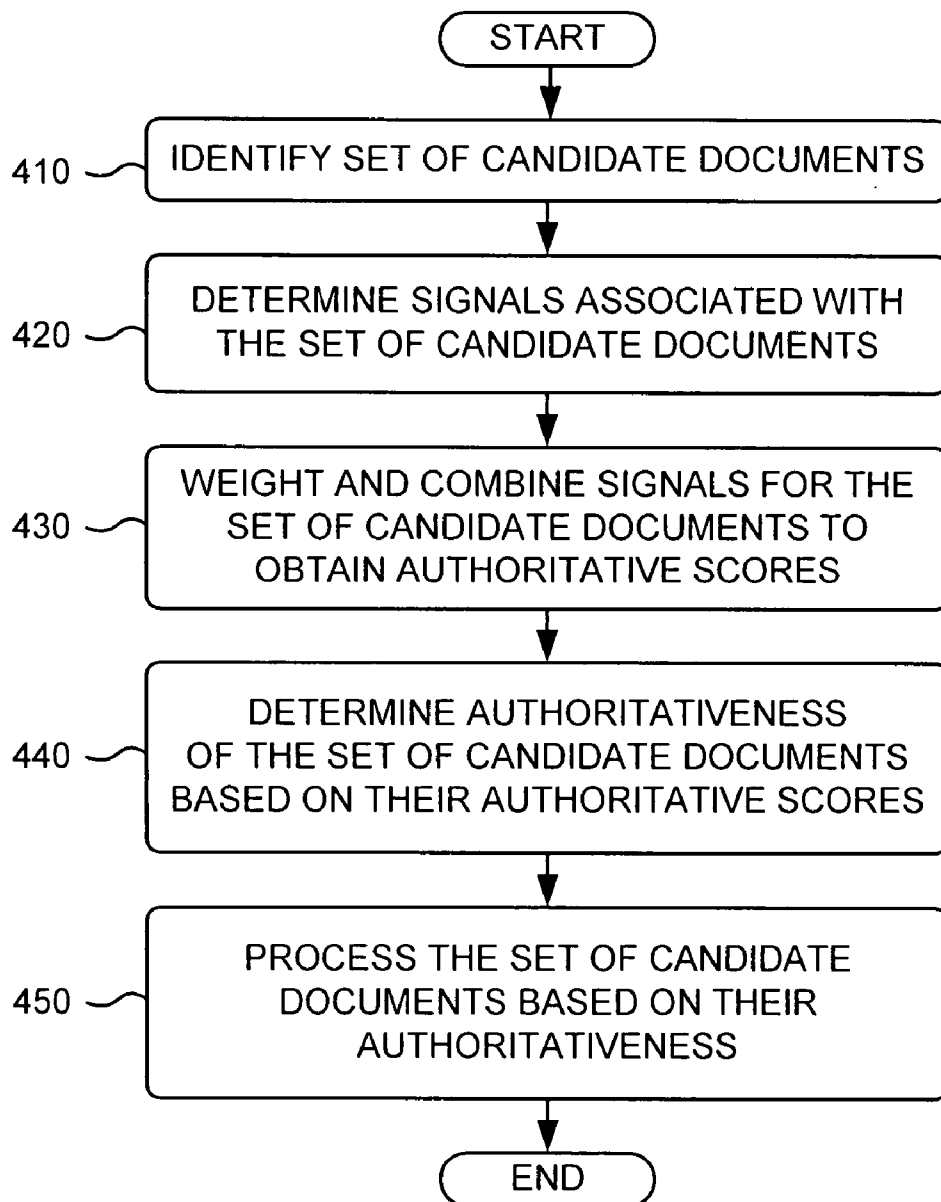
(52) **U.S. Cl.** **707/205**

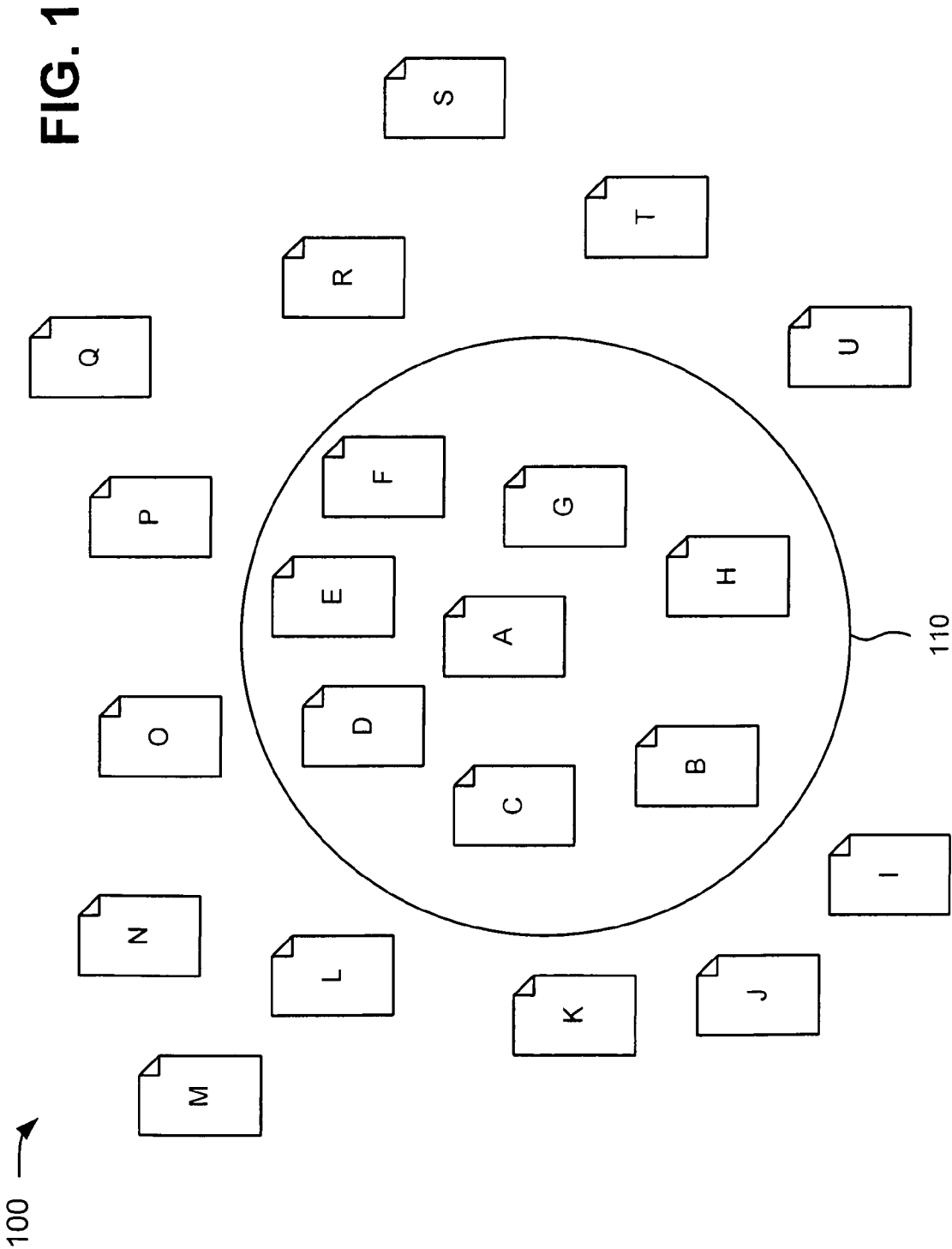
Correspondence Address:
HARRITY SNYDER, LLP
11350 Random Hills Road
SUITE 600
FAIRFAX, VA 22030 (US)

(57) **ABSTRACT**

A system determines documents that are associated with a location, identifies a group of signals associated with each of the documents, and determines authoritativeness of the documents for the location based on the signals.

(21) Appl. No.: **11/024,967**





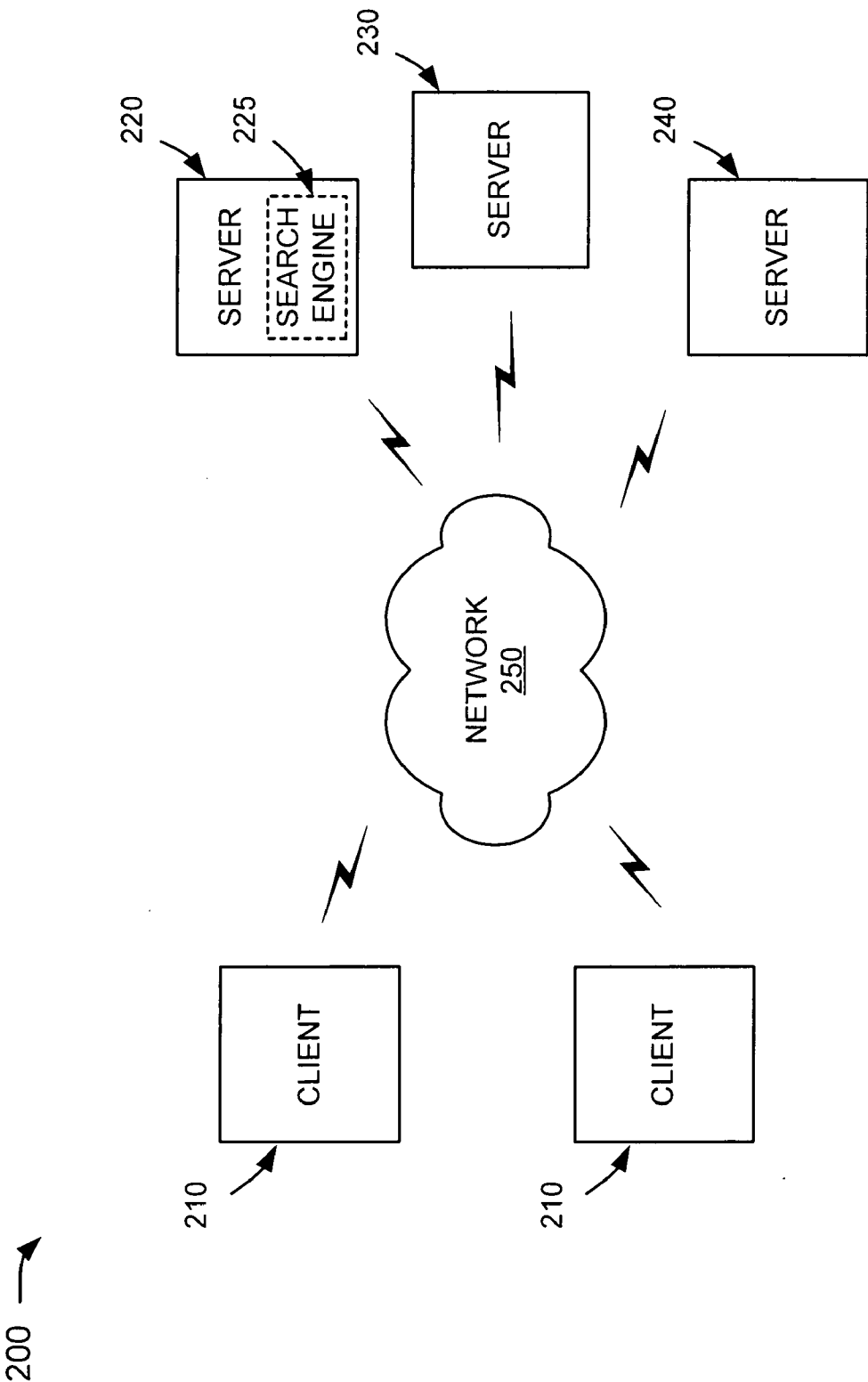


FIG. 2

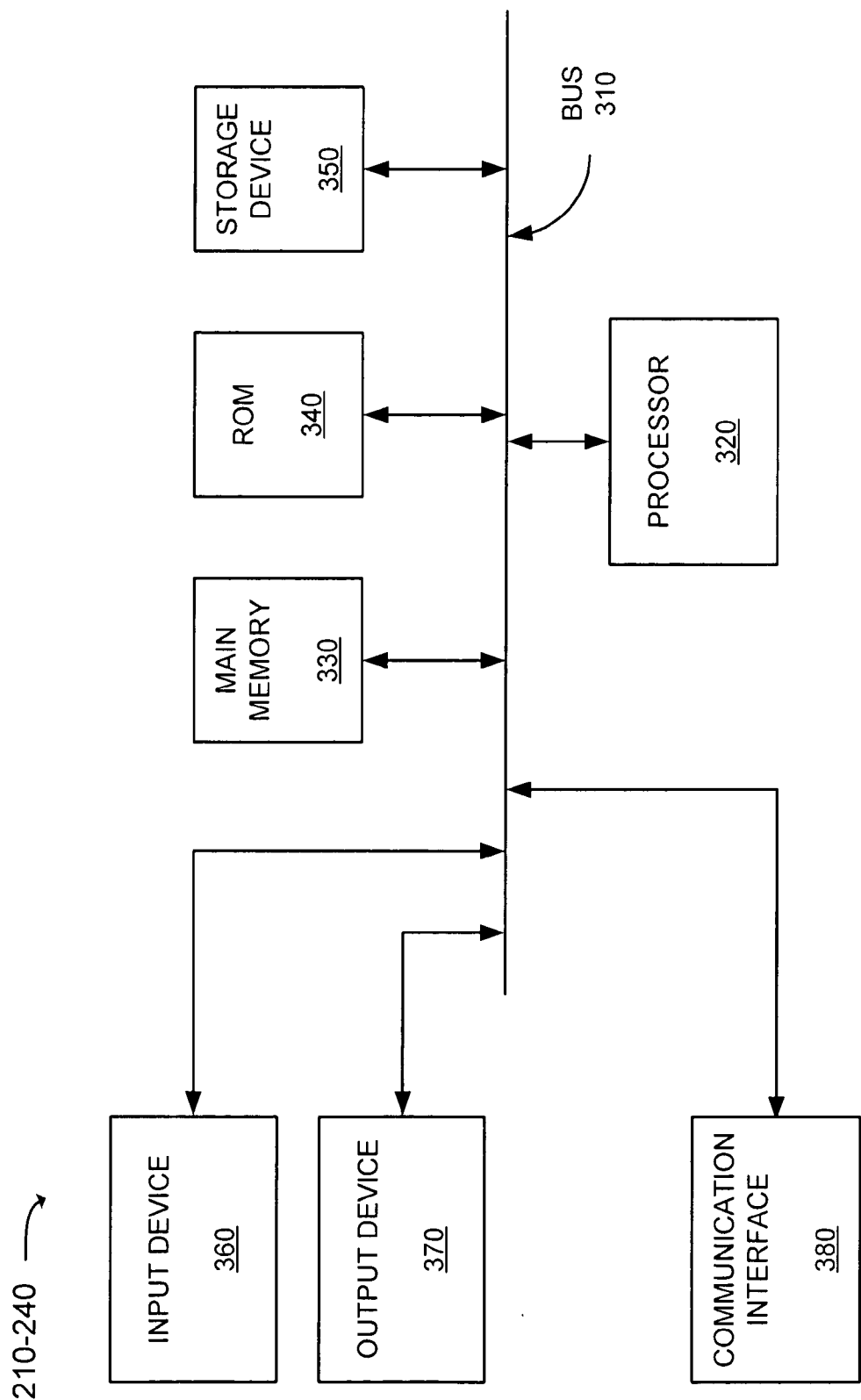
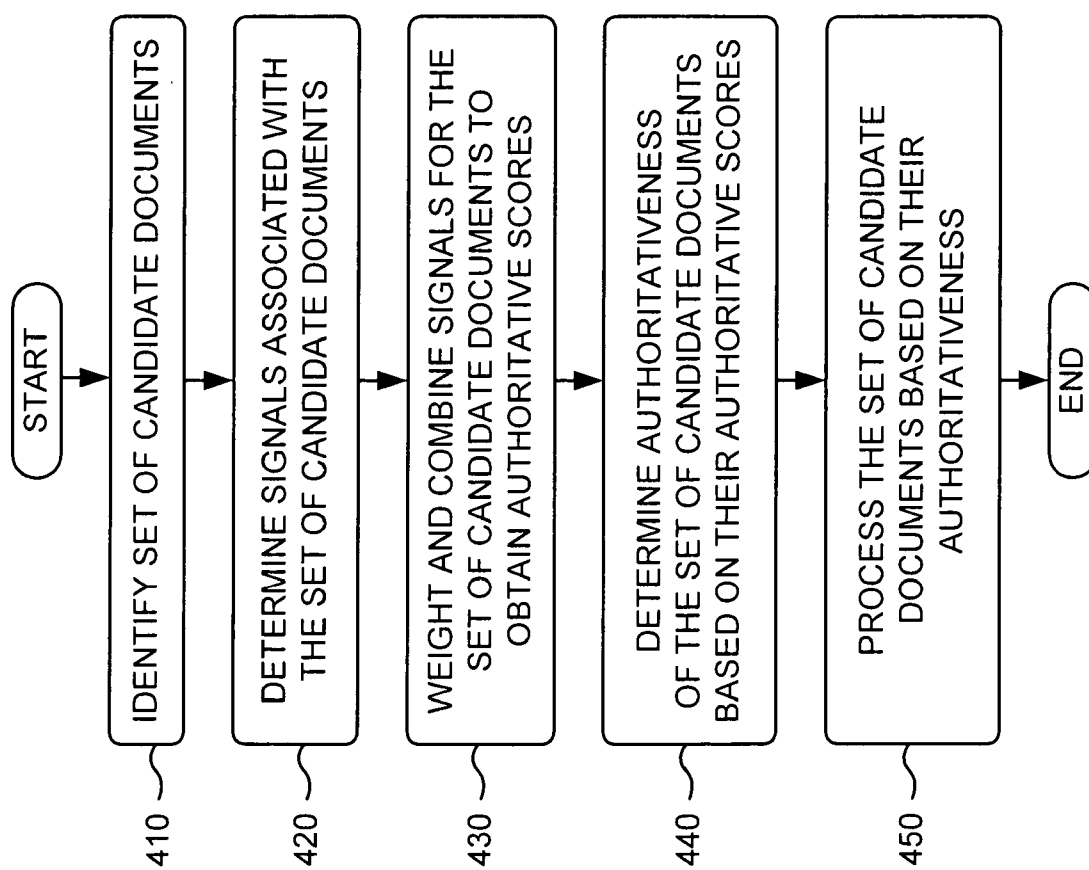


FIG. 3

FIG. 4



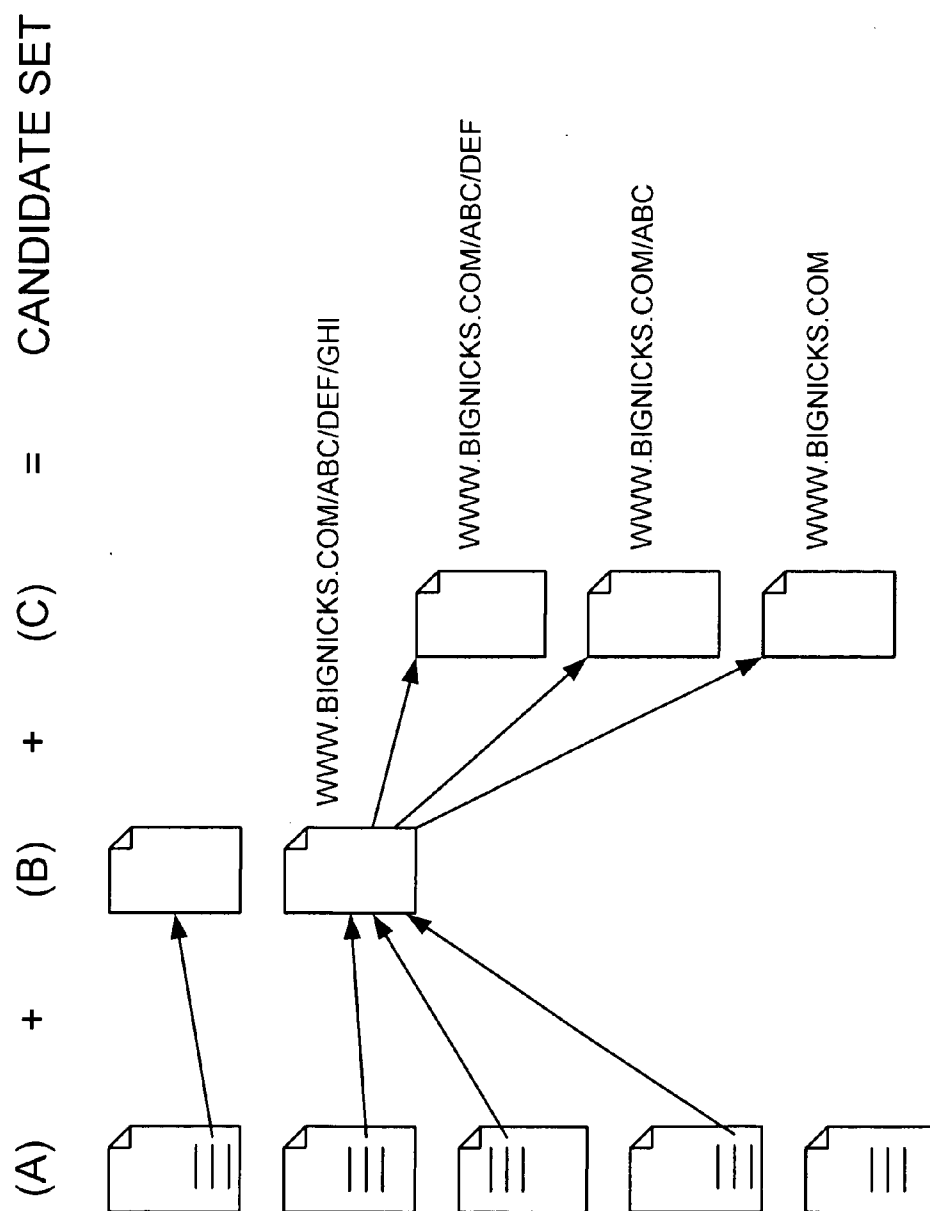
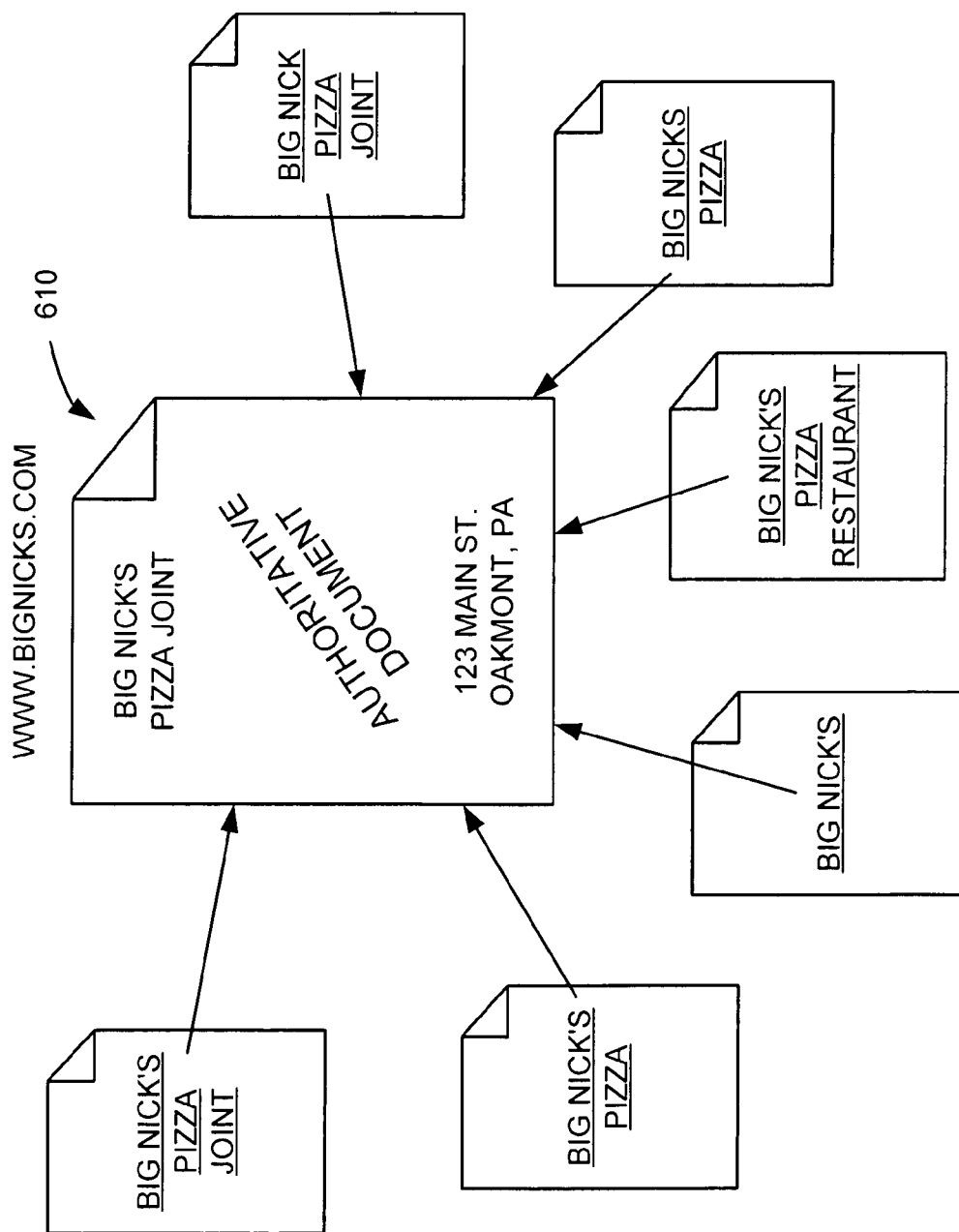


FIG. 5

FIG. 6



AUTHORITATIVE DOCUMENT IDENTIFICATION

BACKGROUND

[0001] 1. Field of the Invention

[0002] Implementations described herein relate generally to information retrieval and, more particularly, to the identification of authoritative documents for a location.

[0003] 2. Description of Related Art

[0004] Modern computer networks, and in particular, the Internet, have made large bodies of information widely and easily available. Internet search engines, for instance, index many millions of web documents that are linked to the Internet. A user connected to the Internet can enter a simple search query to quickly locate web documents relevant to the search query.

[0005] Frequently, users are interested in finding documents relating to a particular location. A document that is authoritative for the location, however, may not include the address of the location. Sometimes the address is located in a sub-document or in an image that cannot be analyzed (e.g., indexed). This makes it difficult to identify an authoritative document.

SUMMARY

[0006] According to one aspect, a computer-readable medium may store computer-executable instructions, including instructions for identifying documents that are associated with a location, instructions for determining a set of signals associated with the documents, and instructions for determining authoritative documents for the location based on the signals.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate an embodiment of the invention and, together with the description, explain the invention. In the drawings,

[0008] **FIG. 1** is an exemplary diagram illustrating a concept consistent with the principles of the invention;

[0009] **FIG. 2** is an exemplary diagram of a network in which systems and methods consistent with the principles of the invention may be implemented;

[0010] **FIG. 3** is an exemplary diagram of a client or server of **FIG. 2** according to an implementation consistent with the principles of the invention;

[0011] **FIG. 4** is a flowchart of exemplary processing for determining the authoritative documents for a location according to an implementation consistent with the principles of the invention;

[0012] **FIG. 5** is an exemplary diagram that illustrates how documents may be chosen as candidate documents according to an implementation consistent with the principles of the invention; and

[0013] **FIG. 6** is an exemplary diagram of signals that may be considered when identifying the authoritative documents for a location according to an implementation consistent with the principles of the invention.

DETAILED DESCRIPTION

[0014] The following detailed description of the invention refers to the accompanying drawings. The same reference numbers in different drawings may identify the same or similar elements. Also, the following detailed description does not limit the invention.

Overview

[0015] **FIG. 1** is an exemplary diagram illustrating a concept consistent with the principles of the invention. Consider a corpus **100** of local documents. The documents are local in the sense that they are associated with a particular geographic area—though not necessarily the same geographic area. A document that relates to a business listing, for example, can be considered a local document because it is associated with the particular address of the business.

[0016] Documents in corpus **100** may be analyzed to determine the locations with which they are associated. For example, assume that the documents in set **110** relate to the same location. Each of the documents in set **110** may refer in some way to the location. For example, a document in set **110** may mention a business at the location, the address of the business, and/or a telephone number associated with the business. One of the documents in the set may be more authoritative for the location than another one of the documents. For example, a document corresponding to the home page of a restaurant at the location may be considered more authoritative for the location than a document corresponding to a review of the restaurant. Systems and methods consistent with the principles of the invention may determine the authoritative documents associated with a location.

[0017] A “document,” as the term is used herein, is to be broadly interpreted to include any machine-readable and machine-storable work product. A document may include, for example, an e-mail, a web site, a business listing, a file, a combination of files, one or more files with embedded links to other files, a news group posting, a blog, a web advertisement, etc. In the context of the Internet, a common document is a web page. Web pages often include textual information and may include embedded information (such as meta information, images, hyperlinks, etc.) and/or embedded instructions (such as Javascript, etc.). A “link,” as the term is used herein, is to be broadly interpreted to include any reference to/from a document from/to another document or another part of the same document.

Exemplary Network Configuration

[0018] **FIG. 2** is an exemplary diagram of a network **200** in which systems and methods consistent with the principles of the invention may be implemented. Network **200** may include multiple clients **210** connected to multiple servers **220-240** via a network **250**. Two clients **210** and three servers **220-240** have been illustrated as connected to network **250** for simplicity. In practice, there may be more or fewer clients and servers. Also, in some instances, a client may perform the functions of a server and a server may perform the functions of a client.

[0019] Clients **210** may include client entities. An entity may be defined as a device, such as a wireless telephone, a personal computer, a personal digital assistant (PDA), a lap

top, or another type of computation or communication device, a thread or process running on one of these devices, and/or an object executable by one of these devices. Servers **220-240** may include server entities that gather, process, search, and/or maintain documents in a manner consistent with the principles of the invention.

[0020] In an implementation consistent with the principles of the invention, server **220** may include a search engine **225** usable by clients **210**. Server **220** may crawl a corpus of documents (e.g., web documents), index the documents, and store information associated with the documents in a repository of documents. Servers **230** and **240** may store or maintain documents that may be crawled or analyzed by server **120**.

[0021] While servers **220-240** are shown as separate entities, it may be possible for one or more of servers **220-240** to perform one or more of the functions of another one or more of servers **220-240**. For example, it may be possible that two or more of servers **220-240** are implemented as a single server. It may also be possible for a single one of servers **220-240** to be implemented as two or more separate (and possibly distributed) devices.

[0022] Network **250** may include a local area network (LAN), a wide area network (WAN), a telephone network, such as the Public Switched Telephone Network (PSTN), an intranet, the Internet, a memory device, or a combination of networks. Clients **210** and servers **220-240** may connect to network **250** via wired, wireless, and/or optical connections.

Exemplary Client/Server Architecture

[0023] FIG. 3 is an exemplary diagram of a client or server entity (hereinafter called "client/server entity"), which may correspond to one or more of clients **210** and/or servers **220-240**, according to an implementation consistent with the principles of the invention. The client/server entity may include a bus **310**, a processor **320**, a main memory **330**, a read only memory (ROM) **340**, a storage device **350**, an input device **360**, an output device **370**, and a communication interface **380**. Bus **310** may include a path that permits communication among the elements of the client/server entity.

[0024] Processor **320** may include a conventional processor, microprocessor, or processing logic that interprets and executes instructions. Main memory **330** may include a random access memory (RAM) or another type of dynamic storage device that may store information and instructions for execution by processor **320**. ROM **340** may include a conventional ROM device or another type of static storage device that may store static information and instructions for use by processor **320**. Storage device **350** may include a magnetic and/or optical recording medium and its corresponding drive.

[0025] Input device **360** may include a conventional mechanism that permits an operator to input information to the client/server entity, such as a keyboard, a mouse, a pen, voice recognition and/or biometric mechanisms, etc. Output device **370** may include a conventional mechanism that outputs information to the operator, including a display, a printer, a speaker, etc. Communication interface **380** may include any transceiver-like mechanism that enables the client/server entity to communicate with other devices and/

or systems. For example, communication interface **380** may include mechanisms for communicating with another device or system via a network, such as network **250**.

[0026] As will be described in detail below, the client/server entity, consistent with the principles of the invention, may perform certain document processing-related operations. The client/server entity may perform these operations in response to processor **320** executing software instructions contained in a computer-readable medium, such as memory **330**. A computer-readable medium may be defined as a physical or logical memory device and/or carrier wave.

[0027] The software instructions may be read into memory **330** from another computer-readable medium, such as data storage device **350**, or from another device via communication interface **380**. The software instructions contained in memory **330** may cause processor **320** to perform processes that will be described later. Alternatively, hardwired circuitry may be used in place of or in combination with software instructions to implement processes consistent with the principles of the invention. Thus, implementations consistent with the principles of the invention are not limited to any specific combination of hardware circuitry and software.

Exemplary Processing

[0028] FIG. 4 is a flowchart of exemplary processing for determining the authoritativeness of documents for a location according to an implementation consistent with the principles of the invention. In one implementation, the processing of FIG. 4 is performed by server **220** (FIG. 2). In another implementation, the processing of FIG. 4 is performed by another a device or a group of devices.

[0029] Processing may begin with identification of a set of candidate documents associated with a particular location (block **410**). A corpus of documents may be analyzed to identify snippets of text (where a snippet of text may be defined as a portion of a document or the entire document) that include information associated with the location, such as a full or partial address of the location, a full or partial telephone number associated with the location, and/or a full or partial name of a business associated with the location. The documents associated with these snippets may be included as a first group of candidate documents.

[0030] Often, a document that includes information associated with a location may link to an authoritative document for that location. Therefore, the documents linked to by the candidate documents in the first group may be included as a second group of candidate documents. A third group of candidate documents may be identified from addresses of candidate documents in the first and second groups, such as by stripping portions of the addresses of the candidate documents in the first and second groups. For example, assume that a candidate document includes the address <http://www.abcdef.com/ghijk/lmnop/qrst.htm>. Portions of the address may be stripped to identify additional candidate documents. For example, the following additional candidate documents may be included in the third group (if they exist): (1) <http://www.abcdef.com/ghijk/lmnop>; (2) <http://www.abcdef.com/ghijk>; and (3) <http://www.abcdef.com>. The set of candidate documents may be further expanded or expanded ways that would be apparent to one skilled in the art.

[0031] The first, second, and third groups of candidate documents may be combined to form the set of candidate

documents. **FIG. 5** is an exemplary diagram that illustrates how documents may be chosen as candidate documents according to an implementation consistent with the principles of the invention. As shown in **FIG. 5**, group (A) may include documents with snippets of text that include information associated with the location, such as a full or partial address of the location, a full or partial telephone number associated with the location, and/or a full or partial name of a business associated with the location; group (B) may include documents that are linked to by documents in group (A); and group (C) may include documents formed from addresses of documents in groups (A) and (B). The set of candidate documents may be formed from documents in groups (A), (B), and (C).

[0032] Returning to **FIG. 4**, signals associated with the set of candidate documents may be determined (block 420). The signals may correspond to meta data associated with the candidate documents. One type of signal may be associated with the number of outlinks in the candidate documents that point to another candidate document. An authoritative document may be a destination corresponding to outlinks from a large number of candidate documents.

[0033] Another type of signal may be associated with anchor text of outlinks that point to the candidate documents. The anchor text may be associated with any document in the document corpus and may be analyzed to determine whether the anchor text matches all or part of the name of the business associated with the location. Text matching may be tricky in this situation because business names can be phrased differently, including partial names and/or misspellings. As a result, a text similarity technique may be used to score words and/or bigrams based on the frequency of occurrence of the words and/or bigrams.

[0034] The frequency of occurrence of words and/or bigrams may be determined by analyzing documents on a per geographic area (e.g., zip code) basis. For example, all of the words and bigrams in a set of documents that are known to be associated with a particular geographic area may be counted. Assume that the bigram "New York" is very common to the New York city area and, therefore, has a high count value for the New York city area. Assume further that the bigram "Pandella Shop" is very uncommon to the New York City area and, therefore, has a low count value for the New York city area.

[0035] Histograms may be generated for the different geographic areas to identify the words and/or bigrams that are common, or uncommon, to those geographic areas. In the above example, the histogram associated with the New York city area may indicate that the bigram "New York" is common (i.e., occurs frequently in documents associated with the New York city area) and the bigram "Pandella Shop" is uncommon (i.e., occurs very infrequently in documents associated with the New York city area).

[0036] Any well known text similarity technique may be used to determine whether anchor text matches all or part of the name of the business associated with the location. More leeway may be given with regard to partial text matches and text matches with misspellings for uncommon words and/or bigrams (e.g., "Pandella Shop") than for common words and/or bigrams (e.g., "New York"). An authoritative document may be a destination corresponding to outlinks whose anchor text matches all or part of the name of the business.

[0037] Another type of signal may be associated with document titles of the candidate documents. The text of the candidate documents may be analyzed to determine whether the titles of the documents match all or part of the name of the business associated with the location. A text similarity technique similar to that described above may be used to determine when the title of a candidate document matches all or part of the business name. An authoritative document may include a title that matches all or part of the name of the business.

[0038] Another type of signal may be associated with domain names associated with the candidate documents. The text of the domain names may be analyzed to determine whether the text matches all or part of the name of the business associated with the location. Domain names are often pushed together and/or truncated versions of the business name (e.g., Bob's Billiard Shop might appear as BobsBilliard.com or BobsBilliardShop.com). Any well known sub-string matching technique may be used instead of, or in addition to, the text similarity technique described above to determine when the domain name associated with a candidate document matches all or part of the business name. An authoritative document may be associated with a domain name that matches all or part of the name of the business.

[0039] The signals for the different candidate documents in the set may be weighted and combined in some manner to obtain an authoritative score (block 430). For example, values (or scores) may be derived for the signals and the values (or scores) may be weighted in some manner. In one implementation, the values (or scores) associated with one or more of the signals, such as the signals associated with the anchor text and/or the domain name, may be weighted more than the values (or scores) associated with another one or more of the signals. The values (or scores) may be combined by, for example, adding them together to obtain an authoritative score for each of the candidate documents in the set.

[0040] In one implementation, the authoritative score for a candidate document may be increased if the candidate document is associated with a single location (as opposed to multiple locations). Some candidate documents may include snippets of text that mention different locations but refer (e.g., link) to the same document. For example, one candidate document may mention location A and link to document A, while another candidate document may mention location B and also link to document A. Also, some candidate documents may be associated with multiple locations. For example, a candidate document may mention locations A and B, such as in the case of a business with multiple locations. The authoritative score for a candidate document that is specific to one location may be increased.

[0041] The authoritativeness of the candidate documents may be determined based on their authoritative scores (block 440). A document with a higher authoritative score may be determined as more authoritative for the location than a document with a lower authoritative score.

[0042] The candidate documents may then be processed based on their authoritativeness (block 450). For example, the authoritative scores of the candidate documents may be used for later processing phases or to control ranking, placement, emphasis, and/or other user interface elements relating to the candidate documents. For example, when a

search query relating to a location is later received, a more authoritative document may be presented in a more prominent manner within the search results than a less authoritative document.

EXAMPLE

[0043] FIG. 6 is an exemplary diagram of signals that may be considered when identifying the authoritativeness of a document according to an implementation consistent with the principles of the invention. As shown in FIG. 6, document 610 may be determined to be authoritative (i.e., receive a high authoritative score) for the location associated with Big Nick's Pizza Joint located at 123 Main Street, Oakmont, Pa. 15302. As explained above, a combination of signals may be used to identify document 610 as authoritative for the location.

[0044] Authoritative document 610 is the destination corresponding to outlinks from a number of documents that mention all or part of the location or the business name. Authoritative document 610 is also the destination corresponding to outlinks whose anchor text matches all or part of the business name (e.g., Big Nick's Pizza Joint, Big Nick's Pizza, Big Nick's, Big Nick's Pizza Restaurant, Big Nick's Pizza, and Big Nick Pizza Joint). Authoritative document 610 also includes a title that matches all or part of the business name (e.g., Big Nick's Pizza Joint). Authoritative document 610 includes a domain name that matches all or part of the name of the business (e.g., www.bignicks.com). Authoritative document 610 is also associated with a single location (e.g., 123 Main St., Oakmont, Pa.).

[0045] When the various signals are weighted and combined, document 610 may receive a high authoritative score for the location associated with the business Big Nick's Pizza Joint at the address of 123 Main Street, Oakmont, Pa. 15302.

CONCLUSION

[0046] Systems and methods consistent with the principles of the invention may determine the authoritativeness of documents associated with a location. As used herein, "location" is intended to refer to an address and/or a business located at the address.

[0047] The foregoing description of preferred embodiments of the present invention provides illustration and description, but is not intended to be exhaustive or to limit the invention to the precise form disclosed. Modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention.

[0048] For example, while a series of acts has been described with regard to FIG. 4, the order of the acts may be modified in other implementations consistent with the principles of the invention. Further, non-dependent acts may be performed in parallel.

[0049] It will be apparent to one of ordinary skill in the art that aspects of the invention, as described above, may be implemented in many different forms of software, firmware, and hardware in the implementations illustrated in the figures. The actual software code or specialized control hardware used to implement aspects consistent with the principles of the invention is not limiting of the invention. Thus, the operation and behavior of the aspects were

described without reference to the specific software code—it being understood that one of ordinary skill in the art would be able to design software and control hardware to implement the aspects based on the description herein.

[0050] No element, act, or instruction used in the present application should be construed as critical or essential to the invention unless explicitly described as such. Also, as used herein, the article "a" is intended to include one or more items. Where only one item is intended, the term "one" or similar language is used. Further, the phrase "based on" is intended to mean "based, at least in part, on" unless explicitly stated otherwise.

What is claimed is:

1. A method comprising:

identifying a set of documents, as candidate documents, that are associated with a location;

determining signals associated with the candidate documents;

determining authoritativeness of the candidate documents for the location based on the signals; and

processing the candidate documents based on their authoritativeness for the location.

2. The method of claim 1, wherein identifying a set of documents includes:

analyzing documents in a document corpus to identify snippets of text that include information associated with the location, and

identifying documents that include the snippets of text as candidate documents.

3. The method of claim 2, wherein the information associated with the location includes at least one of a full or partial address of the location, a full or partial telephone number associated with the location, or a full or partial name of a business associated with the location.

4. The method of claim 2, wherein identifying a set of documents further includes:

determining documents that are linked to by the candidate documents, and

identifying the determined documents as candidate documents.

5. The method of claim 4, wherein identifying a set of documents further includes:

determining additional documents by stripping portions of addresses of the candidate documents, and

identifying the additional documents as candidate documents.

6. The method of claim 1, wherein determining signals associated with the candidate documents includes:

determining a number of outlinks from ones of the candidate documents that point to other ones of the candidate documents; and

wherein determining authoritativeness of the candidate documents includes:

generating an authoritative score for one of the candidate documents based on the number of outlinks from other ones of the candidate documents that point to the candidate document.

7. The method of claim 1, wherein determining signals associated with the candidate documents includes:

identifying anchor text associated with links to the candidate documents; and

wherein determining authoritativeness of the candidate documents includes:

generating an authoritative score for one of the candidate documents based on whether the candidate document is pointed to by one or more links whose anchor text matches all or part of a name of a business associated with the location.

8. The method of claim 1, wherein determining signals associated with the candidate documents includes:

identifying titles of ones of the candidate documents; and

wherein determining authoritativeness of the candidate documents includes:

generating an authoritative score for one of the candidate documents based on whether a title associated with the candidate document matches all or part of a name of a business associated with the location.

9. The method of claim 1, wherein determining signals associated with the candidate documents includes:

identifying domain names associated with ones of the candidate documents; and

wherein determining authoritativeness of the candidate documents includes:

generating an authoritative score for one of the candidate documents based on whether a domain name associated with the candidate document matches all or part of a name of a business associated with the location.

10. The method of claim 1, wherein determining signals associated with the candidate documents includes:

determining locations with which ones of the candidate documents are associated; and

wherein determining authoritativeness of the candidate documents further includes:

increasing the authoritativeness of one of the candidate documents based on whether the candidate document is associated with a single location.

11. The method of claim 1, wherein the signals are associated with at least one of:

a number of outlinks from ones of the candidate documents that point to another one of the candidate documents,

anchor text associated with links that point to ones of the candidate documents that matches all or part of a name of a business associated with the location,

titles of ones of the candidate documents that match all or part of the business name, and

domain names associated with ones of the candidate documents that match all or part of the business name.

12. The method of claim 1, wherein the signals are associated with a plurality of different types of data associated with the candidate documents; and

wherein the method further comprises:

weighting the different types of data;

combining the weighted data for ones of the candidate documents; and

assigning authoritative scores to the ones of the candidate documents based on the combined, weighted data.

13. The method of claim 12, wherein processing the candidate documents includes:

ranking one of the candidate documents based on its authoritative score.

14. A system comprising:

means for identifying a set of documents, as candidate documents, that are associated with a business;

means for determining a plurality of signals associated with each of the candidate documents; and

means for determining authoritativeness of the candidate documents for the business based on the signals.

15. A system, comprising:

a memory to store instructions; and

a processor to execute the instructions in the memory to:

determine documents that are associated with a location,

identify a plurality of signals associated with each of the documents,

assign authoritative scores to the documents based on the signals, and

process the documents based on the authoritative scores.

16. The system of claim 15, wherein when determining documents, the processor is configured to analyze documents in a document corpus to detect documents that include snippets of text with information associated with the location.

17. The system of claim 16, wherein the information associated with the location includes at least one of a full or partial address of the location, a full or partial telephone number associated with the location, or a full or partial name of a business associated with the location.

18. The system of claim 16, wherein when determining documents, the processor is further configured to identify documents that are linked to by the documents.

19. The system of claim 18, wherein when determining documents, the processor is further configured to identify additional documents by stripping portions of addresses of the documents.

20. The system of claim 15, wherein when identifying a plurality of signals, the processor is configured to determine a number of outlinks from ones of the documents that point to another one of the documents; and

when assigning authoritative scores to the documents, the processor is configured to generate an authoritative score for one of the documents based on a number of outlinks from other documents that point to the document.

21. The system of claim 15, wherein when identifying a plurality of signals, the processor is configured to identify anchor text associated with links to ones of the documents; and

when assigning authoritative scores to the documents, the processor is configured to generate an authoritative score for one of the documents based on one or more links to the document whose anchor text matches all or part of a name of a business associated with the location.

22. The system of claim 15, wherein when identifying a plurality of signals, the processor is configured to identify titles of ones of the documents; and

when assigning authoritative scores to the documents, the processor is configured to generate an authoritative score for one of the documents based on whether the document includes a title that matches all or part of a name of a business associated with the location.

23. The system of claim 15, wherein when identifying a plurality of signals, the processor is configured to identify domain names associated with ones of the documents; and

when assigning authoritative scores to the documents, the processor is configured to generate an authoritative score for one of the documents based on whether the document is associated with a domain name that matches all or part of a name of a business associated with the location.

24. The system of claim 15, wherein when identifying a plurality of signals, the processor is configured to determine locations with which ones of the documents are associated; and

when assigning authoritative scores to the documents, the processor is configured to increase the authoritative score assigned to one of the documents when the document is associated with a single location.

25. The system of claim 15, wherein the signals are associated with at least one of:

a number of outlinks from ones of the documents that point to another one of the documents,

anchor text associated with links to ones of the documents that matches all or part of a name of a business associated with the location,

titles of ones of the documents that match all or part of the business name, and

domain names associated with ones of the documents that match all or part of the business name.

26. The system of claim 15, wherein the signals are associated with a plurality of different types of data associated with the documents; and

wherein when assigning authoritative scores to the documents, the processor is configured to:

weight the different types of data,

combine the weighted data for ones of the documents, and

generate authoritative scores for the ones of the documents based on the combined, weighted data.

27. The system of claim 15, wherein when processing the documents, the processor is configured to rank one of the documents based on its authoritative score.

28. A computer-readable medium that stores computer-executable instructions, comprising:

instructions for identifying documents that are associated with a location;

instructions for determining a plurality of signals associated with the documents; and

instructions for determining authoritativeness of the documents for the location based on the signals.

29. A method comprising:

identifying a set of documents, as candidate documents, that are associated with a location;

determining, for each of the candidate documents, a first signal based on a number of outlinks from one or more of the candidate documents that point to the candidate document;

determining, for each of the candidate documents, a second signal based on whether there is anchor text, which matches all or part of a name of a business associated with the location, associated with a link that points to the candidate document;

determining, for each of the candidate documents, a third signal based on whether a title of the candidate document matches all or part of the business name;

determining, for each of the candidate documents, a fourth signal based on whether a domain name associated with the candidate document matches all or part of the business name;

determining, for each of the candidate documents, a fifth signal based on whether the candidate document is associated with a single location;

weighting the first, second, third, fourth, and fifth signals;

combining the weighted first, second, third, fourth, and fifth signals to identify a score for each of the candidate documents; and

processing the candidate documents based on the scores.

* * * * *