RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK

BACKGROUND

Ranking functions that rank documents according to their relevance to a given search query are known. Efforts continue in the art to develop ranking functions that provide better search results for a given search query compared to search results generated by search engines using known ranking functions.

SUMMARY

Described herein are, among other things, various technologies for determining a document relevance score for a given document on a network. The document relevance score is generated via a ranking function that comprises one or more query-independent components, wherein at least one query-independent component includes a biased click distance parameter that takes into account biased click distance values for multiple documents on the network. The ranking functions may be used by a search engine to rank multiple documents in order (typically, in descending order) based on the document relevance scores of the multiple documents.

This Summary is provided to generally introduce the reader to one or more select concepts describe below in the "Detailed Description" section in a simplified form. This Summary is not intended to identify key and/or required features of the claimed subject matter.

BRIEF DESCRIPTION OF THE FIGURES

FIG. **1** represents an exemplary logic flow diagram showing exemplary steps in a method of producing ranked search results in response to a search query inputted by a user;

FIG. **2** is a block diagram of some of the primary components of an exemplary operating environment for implementation of the methods and processes disclosed herein;

FIG. **3** depicts an exemplary web graph identifying documents in a network space, links between the documents, authoritative nodes having an assigned biased click distance value, and non-authoritative nodes having a calculated biased click distance value;

FIGS. **4A-4B** represent a logic flow diagram showing exemplary steps in a method of assigning and generating biased click distance values for nodes on a web graph;

FIGS. **5A-5B** represent a logic flow diagram showing exemplary steps in a method of generating biased click distance values for non-authoritative nodes on a web graph; and

FIG. **6** represents a logic flow diagram showing exemplary steps in a method of ranking search results generating using a ranking function containing a biased click distance value parameter.

DETAILED DESCRIPTION

To promote an understanding of the principles of the methods and processes disclosed herein, descriptions of specific embodiments follow and specific language is used to describe the specific embodiments. It will nevertheless be understood that no limitation of the scope of the disclosed methods and processes is intended by the use of specific language. Alterations, further modifications, and such further applications of the principles of the disclosed methods and processes discussed are contemplated as would normally occur to one ordinarily skilled in the art to which the disclosed methods and processes pertains.

Methods of determining a document relevance score for documents on a network are disclosed. Each document relevance score is calculated using a ranking function that contains one or more query-dependent components (e.g., a function component that depends on the specifics of a given search query or search query term), as well as one or more query-independent components (e.g., a function component that that does not depend on a given search query or search query term). The document relevance scores determined by the ranking function may be used to rank documents within a network space (e.g., a corporate intranet space) according to each document relevance score. An exemplary search process in which the disclosed methods may be used is shown as exemplary process **10** in FIG. **1**.

FIG. **1** depicts exemplary search process **10**, which starts with process step **80**, wherein a user inputs a search query. From step **80**, exemplary search process **10** proceeds to step **200**, wherein a search engine searches all documents within a network space for one or more terms of the search query. From step **200**, exemplary search process **10** proceeds to step **300**, wherein a ranking function of the search engine sorts the documents within the network

space based on the relevance score of each document, the document relevance score being based on one or more query-dependent components and one or more query-independent components. From step **300**, exemplary search process **10** proceeds to step **400**, wherein sorted search results are presented to the user, typically in decreasing order of relevance, identifying documents within the network space that are most relevant to the search query.

As discussed in more detail below, in some exemplary methods of determining a document relevance score, at least one query-independent component of a ranking function used to determine a document relevance score takes in to account a "biased click distance" of each document within a network space. The biased click distance for certain documents, referred to herein as "authoritative documents" within a network or "authoritative nodes" on a web graph, may be assigned an initial click distance value, in order to identify these documents as having different degrees of importance relative to each other, and possibly a higher degree of importance relative to the rest of the documents on the network. The remaining documents, referred to herein as "non-authoritative documents" within a network or "non-authoritative nodes" on a web graph, have a biased click distance value that is calculated based on their location to the closest authoritative document within a network space (or closest authoritative node on a web graph) resulting in click distance values biased towards the authoritative nodes.

In one exemplary embodiment, a biased click distance value may be assigned to *m* authoritative documents on a network comprising N total documents, wherein *m* is greater than or equal to 2 and less than N. In this exemplary embodiment, a system administrator manually selects or application code within a search system automatically identifies *m* authoritative documents within a given network space that have some degree of importance within the network space. For example, one of the *m* authoritative documents may be a homepage of a website or another page linked directly to the homepage of a website.

In another exemplary embodiment, at least two of the biased click distance values assigned to the *m* authoritative documents differ from one another. In this embodiment, different numerical values may be assigned to two or more *m* authoritative documents in order to further quantify the importance of one authoritative document to another authoritative document. For example, the importance of a given authoritative document may be indicated by a low biased click distance value. In this example, authoritative documents having a biased

click distance value equal to 0 will be considered more important than authoritative documents having a biased click distance value greater than 0.

The disclosed methods of determining a document relevance score may further utilize a ranking function that comprises at least one query-independent component that includes an edge value parameter that takes into account edge values assigned to each edge on the network, wherein each edge connects one document to another document within the hyperlinked structure of the network (or one node to another node on a web graph). Assigning edge values to one or more edges connecting documents to one another on a network provides a further method of affecting the document relevance score of documents on the network. For example, in the example described above wherein a lower biased click distance value indicates the importance of a given document, increasing an edge value between two documents, such as a first document and a second document linked to the first document, further reduces the importance of the second document (i.e., the linked document) relative to the first document. Conversely, by assigning a lower edge value to the edge between the first document and second document, the importance of the second document becomes greater relative to the first document.

In an exemplary embodiment, two or more edges linking documents within a network space may be assigned edge values that differ from one another. In this exemplary embodiment, different numerical values may be assigned to two or more edges in order to further quantify the importance of one document to another within a network space. In other exemplary embodiments, all of the edges linking documents within a network space are assigned the same edge value, wherein the assigned edge value is 1 or some other positive number. In yet another embodiment, the edge values are equal to one another and are equal to or greater than the highest biased click distance value initially assigned to one or more authoritative documents.

In yet a further exemplary embodiment, the disclosed methods of determining a document relevance score utilizes a ranking function that comprises at least one query-independent component, which includes both the above-described biased click distance parameter and the above-described edge value parameter.

The document relevance score may be used to rank documents within a network space. For example, a method of ranking documents on a network may comprise the steps of determining a document relevance score for each document on the network using the above-described method; and ranking

the documents in a desired order (typically, in descending order) based on the document relevance scores of each document.

The document relevance score may also be used to rank search results of a search query. For example, a method of ranking search results of a search query may comprise the steps of determining a document relevance score for each document in the search results of a search query using the above-described method; and ranking the documents in a desired order (typically, in descending order) based on the document relevance scores of each document.

Application programs using the methods disclosed herein may be loaded and executed on a variety of computer systems comprising a variety of hardware components. An exemplary computer system and exemplary operating environment for practicing the methods disclosed herein is described below.

Exemplary Operating Environment

FIG. **2** illustrates an example of a suitable computing system environment **100** on which the methods disclosed herein may be implemented. The computing system environment **100** is only one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of the methods disclosed herein. Neither should the computing environment **100** be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment **100**.

The methods disclosed herein are operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations that may be suitable for use with the methods disclosed herein include, but are not limited to, personal computers, server computers, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

The methods and processes disclosed herein may be described in the general context of computer-executable instructions, such as program modules, being executed by a computer. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The methods and processes disclosed herein may also be practiced in distributed computing

environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote computer storage media including memory storage devices.

With reference to FIG. **2**, an exemplary system for implementing the methods and processes disclosed herein includes a general purpose computing device in the form of a computer **110**. Components of computer **110** may include, but are not limited to, a processing unit **120**, a system memory **130**, and a system bus **121** that couples various system components including, but not limited to, system memory **130** to processing unit **120**. System bus **121** may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. By way of example, and not limitation, such architectures include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video Electronics Standards Association (VESA) local bus, and Peripheral Component Interconnect (PCI) bus also known as Mezzanine bus.

Computer **110** typically includes a variety of computer readable media. Computer readable media can be any available media that can be accessed by computer **110** and includes both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer readable media may comprise computer storage media and communication media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium, which can be used to store the desired information and which can be accessed by computer **110**. Communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes

wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. Combinations of the any of the above should also be included within the scope of computer readable media as used herein.

System memory **130** includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) **131** and random access memory (RAM) **132**. A basic input/output system **133** (BIOS) containing the basic routines that help to transfer information between elements within computer **110**, such as during start-up, is typically stored in ROM **131**. RAM **132** typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processing unit **120**. By way of example, and not limitation, FIG. **2** illustrates operating system **134**, application programs **135**, other program modules **136**, and program data **137**.

Computer **110** may also include other removable/non-removable, volatile/nonvolatile computer storage media. By way of example only, FIG. **2** illustrates a hard disk drive **140** that reads from or writes to non-removable, nonvolatile magnetic media, a magnetic disk drive **151** that reads from or writes to a removable, nonvolatile magnetic disk **152**, and an optical disk drive **155** that reads from or writes to a removable, nonvolatile optical disk **156** such as a CD ROM or other optical media. Other removable/non-removable, volatile/nonvolatile computer storage media that can be used in the exemplary operating environment include, but are not limited to, magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape, solid state RAM, solid state ROM, and the like. Hard disk drive **141** is typically connected to system bus **121** through a non-removable memory interface such as interface **140**, and magnetic disk drive **151** and optical disk drive **155** are typically connected to system bus **121** by a removable memory interface, such as interface **150**.

The drives and their associated computer storage media discussed above and illustrated in FIG. **2** provide storage of computer readable instructions, data structures, program modules and other data for computer **110**. In FIG. **2**, for example, hard disk drive **141** is illustrated as storing operating system **144**, application programs **145**, other program modules **146**, and program data **147**. Note that these components can either be the same as or different from operating system **134**, application programs **135**, other program modules **136**, and program data **137**. Operating system **144**, application programs **145**, other program modules **146**, and program data **147** are given different numbers here to illustrate that, at a minimum, they are different copies.

A user may enter commands and information into computer **110** through input devices such as a keyboard **162** and pointing device **161**, commonly referred to as a mouse, trackball or touch pad. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner, or the like. These and other input devices are often connected to processing unit **120** through a user input interface **160** that is coupled to system bus **121**, but may be connected by other interface and bus structures, such as a parallel port, game port or a universal serial bus (USB). A monitor **191** or other type of display device is also connected to system bus **121** via an interface, such as a video interface **190**. In addition to monitor **191**, computer **110** may also include other peripheral output devices such as speakers **197** and printer **196**, which may be connected through an output peripheral interface **195**.

Computer **110** may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer **180**. Remote computer **180** may be a personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to computer **110**, although only a memory storage device **181** has been illustrated in FIG. **2**. The logical connections depicted in FIG. **2** include a local area network (LAN) **171** and a wide area network (WAN) **173**, but may also include other networks. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.

When used in a LAN networking environment, computer **110** is connected to LAN **171** through a network interface or adapter **170**. When used in a WAN networking environment, computer **110** typically includes a modem **172** or other means for establishing communications over WAN **173**, such as the Internet. Modem **172**, which may be internal or external, may be connected to system bus **121** via user input interface **160**, or other appropriate mechanism. In a networked environment, program modules depicted relative to computer **110**, or portions thereof, may be stored in the remote memory storage device. By way of example, and not limitation, FIG. **2** illustrates remote application programs **185** as residing on memory device **181**. It will be appreciated that the network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

Methods and processes disclosed herein may be implemented using one or more application programs including, but not limited to, a search ranking application, which could be one of numerous application programs

designated as application programs **135**, application programs **145** and remote application programs **185** in exemplary system **100**.

As mentioned above, those skilled in the art will appreciate that the disclosed methods of generating a document relevance score for a given document may be implemented in other computer system configurations, including hand-held devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, networked personal computers, minicomputers, mainframe computers, and the like. The disclosed methods of generating a document relevance score for a given document may also be practiced in distributed computing environments, where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

Implementation of Exemplary Embodiments

As discussed above, methods of determining a document relevance score for a document on a network are provided. The disclosed methods may rank a document on a network utilizing (i) a ranking function that takes into account a biased click distance value of each document on the network, (ii) a ranking function that takes into account one or more edge values assigned to edges (or links) between documents on the network, or (iii) both (i) and (ii).

The disclosed methods of determining a document relevance score for a document on a network may comprise a number of steps. In one exemplary embodiment, the method of determining a document relevance score for a document on a network comprises the steps of storing document and link information for documents on a network; generating a representation of the network from the document and link information, wherein the representation of the network includes nodes that represent the documents and edges that represent the links; assigning a biased click distance value (CD_A) to at least two nodes on the network, wherein the nodes that are assigned a biased click distance value are authoritative nodes; computing a biased click distance for each of the non-authoritative nodes in the representation of the network, wherein a biased click distance for a given non-authoritative node is measured from the given non-authoritative node to an authoritative node closest to the given non-authoritative node, wherein the computing step results in a computed biased click distance value (CD_C) for each non-authoritative document; and using the biased click

distance value (i.e., (CD_A or CD_C) for each document to determine the document relevance score of a given document on the network.

The step of storing document and link information for documents on a network may be performed by indexing application code commonly found on computing systems. The indexing application code generates a representation of the network from the document and link information, wherein the representation of the network includes nodes that represent the documents and edges that represent the links. Such a representation of the network is commonly referred to as a "web graph." One exemplary method of generating a web graph comprises using data gathered by a process where link and anchor text information is gathered and attributed to specific target documents of the anchor. This process and the concept of anchor text is described more fully in U.S. Patent Application Serial Number 10/955,462 entitled "SYSTEM AND METHOD FOR INCORPORATING ANCHOR TEXT INTO RANKING SEARCH RESULTS" filed on August 30, 2004, the subject matter of which is hereby incorporated by reference in its entirety.

FIG. **3** depicts an exemplary web graph identifying documents in a network space and links between the documents. As shown in FIG. **3**, exemplary web graph **30** comprises nodes **31**, which represent each document within a given network space (e.g., a corporate intranet), and edges **32**, which represent links between documents within a given network space. It should be understood that exemplary web graph **30** is an overly simplified representation of a given network space. Typically, a given network space may comprise hundreds, thousands or millions of documents and hundreds, thousands or millions of links connecting documents to one another. Further, although exemplary web graph **30** depicts up to eight links connected to a given node (e.g., central node **33**), it should be understood that in an actual network setting, a given node may have hundreds of links connecting the node (e.g., document) to hundreds of other documents within the network (e.g., the home page of a network may be linked to every page within the network).

In addition, exemplary web graph **30** shows very few cycles (e.g., a first node linking to a second node, which may link to additional nodes, wherein the second node or one of the additional nodes links back to the first node). One such cycle is represented by nodes **41** and **42** in FIG. **3**. Other cycles could be represented if any of end nodes **40** linked back to any other node shown in FIG. **3**, such as central node **33**. Regardless of the simplicity or complexity of a given web graph, the disclosed methods of generating a document relevance

score for a given document may be used on any web graph, including those containing cycles.

Once a web graph has been generated, one or more techniques may be used to affect the relative importance of one or more documents within the network space, represented by the nodes of the web graph. As discussed above and below, these techniques include, but are not limited to, (i) designating two or more nodes as authoritative nodes; (ii) assigning each of the authoritative nodes a biased click distance value (CD_A), (iii) optionally, assigning two or more biased click distance values (CD_A) that differ from one another; (iv) assigning edge value to each edge of the web graph; (v) optionally, assigning a minimum edge value to each edge of the web graph, wherein the minimum edge value is greater than a maximum or highest assigned biased click distance values (CD_{Amax}); (vi) optionally, assigning two or more edge values that differ from one another; (vii) calculating a biased click distance value (CD_C) for each non-authoritative node; and (viii) optionally, downgrading any of the biased click distance values (CD_A or CD_C), when necessary, if test queries using the biased click distance values generate irrelevant search results. Some of the above-described exemplary techniques for affecting the biased click distance value of one or more documents within the network represented by exemplary web graph **30** are shown in FIG. **3**.

In exemplary web graph **30**, nodes **31** having a square shape are used to identify authoritative nodes within the network, while nodes **31** having a circular shape are used to identify non-authoritative nodes. It should be understood that any number of nodes within a given web graph may be designated as authoritative nodes depending on a number of factors including, but not limited to, the total number of documents within the network space, and the number of "important" documents that are within the network space. In exemplary web graph **30**, 9 of the 104 nodes are designated as authoritative nodes (i.e., represent 9 out of 104 documents as being as particular importance).

Further, although not shown on exemplary web graph **30**, each edge **32** between each pair of nodes **31** has an edge weight associated therewith. Typically, each edge **32** has a default edge weight of 1; however, as discussed above, an edge weight other than 1 can be assigned to each edge **32**. Further, in some embodiments, two or more different edge weights may be assigned to edges within the same web graph. In FIG. **3**, letters *p*, *q*, *r*, *s* and *t* shown on exemplary web graph **30** are used to indicate edge values of some of edges **32**. As discussed above, edge values *p*, *q*, *r*, *s* and *t* may have a value of 1, a value other than 1,

and/or values that differ from one another in order to further affect biased click distance values of nodes **31** within exemplary web graph **30**. Typically, edge values for *p*, *q*, *r*, *s* and *t*, as well as the other edges in exemplary web graph **30**, are the same number, and are typically equal to or greater than 1. In some embodiments, edge values for *p*, *q*, *r*, *s* and *t*, as well as the other edges in exemplary web graph **30**, are the same number, and are equal to or greater than the highest biased click distance value assigned to an authoritative node.

The one or more techniques used to modify a web graph in order to affect biased click distance values of documents on a network may be manually initiated and performed by a system administrator. The system administrator can view a given web graph, and edit the web graph as desired to increase and/or decrease the relative importance of one or more documents within a network space as described above. Application code, such as application code in a computing system capable of conducting a search query, may automatically produce a bias in a web graph using one or more of the above-described techniques (e.g., calculating a biased click distance value (CD_C) for each non-authoritative node).

FIGS. **4A-4B** represent a logic flow diagram showing exemplary steps in an exemplary method of assigning and generating biased click distance values for nodes on a web graph followed by an optional downgrading procedure by a system administrator. As shown in FIG. **4A**, exemplary method **401** starts at block **402** and proceeds to step **403**. In step **403**, a number of authoritative nodes (or URLs) are selected out of *N* total nodes (or URLs) within a network space. In exemplary method **401**, *m* authoritative nodes (or URLs) are selected, wherein *m* is greater than or equal to 2. Once the authoritative nodes (or URLs) are selected, exemplary method **401** proceeds to decision block **404**.

At decision block **404**, a determination is made by a system administrator whether to assign at least two different biased click distance values (CD_A) to two or more of the *m* authoritative nodes (or URLs). If a decision is made to assign at least two different biased click distance values (CD_A) to two or more of the *m* authoritative nodes (or URLs), exemplary method **401** proceeds to step **405**, wherein at least two different biased click distance values (CD_A) are assigned to two or more of the *m* authoritative nodes (or URLs). For example, referring to exemplary web graph **30** shown in FIG. **3**, authoritative nodes **33** and **34** may be assigned a biased click distance value of 0, authoritative nodes **35** and **36** may be assigned a biased click distance value of +3, and authoritative node **48**

may be assigned a biased click distance value of +2. From step **405**, exemplary method **401** proceeds to decision block **407**.

Returning to decision block **404**, if a decision is made not to assign at least two different biased click distance values (CD_A) to two or more of the m authoritative nodes (or URLs), exemplary method **401** proceeds to step **406**, wherein the same biased click distance value (CD_A) is assigned to each of the m authoritative nodes (or URLs). For example, referring to exemplary web graph **30** of FIG. **3** again, each of the authoritative nodes may be assigned a biased click distance value, such as 0, +2, or +5. From step **406**, exemplary method **401** proceeds to decision block **407**.

At decision block **407**, a determination is made by a system administrator or application code whether to assign an edge weight other than 1 to one or more edges of a web graph. If a decision is made to assign an edge weight other than 1 to one or more edges of a web graph, exemplary method **401** proceeds to decision block **408**. At decision block **408**, a determination is made by a system administrator whether to assign a minimum edge value to the edges of a web graph, wherein the minimum edge value is greater than the largest assigned biased click distance value (CD_{Amax}). If a decision is made to assign a minimum edge value to the edges of a web graph, wherein the minimum edge value is greater than the largest assigned biased click distance value (CD_{Amax}), exemplary method **401** proceeds to step **409**, wherein a minimum edge value greater than the largest assigned biased click distance value (CD_{Amax}) is assigned to each edge of a web graph. For example, referring to exemplary web graph **30** shown in FIG. **3**, if authoritative node **33** is assigned the largest biased click distance value (CD_{Amax}) and CD_{Amax} equals +3, a minimum edge value of greater than +3 is assigned to each edge **32** shown in FIG. **3**.

In some embodiments, applying a minimum edge value that is greater than the largest assigned biased click distance value (CD_{Amax}) to each edge of a web graph may have some advantages. In this embodiment, such a technique guarantees that the assigned biased click distance value (CD_A) of each authoritative node (or document or URL) is less than the calculated biased click distance value (CD_C) of every non-authoritative node (or document or URL) in a web graph. When importance of a document is based on a lower biased click distance value, such a technique enables all of the authoritative nodes (or documents or URLs) to be considered more important than the non-authoritative nodes (or documents or URLs) within a web graph.

From step **409**, exemplary method **401** proceeds to decision block **410** shown in FIG. **4B** and described below. Returning to decision block **408**, if a decision is made not to assign a minimum edge value to each edge, wherein the minimum edge value is greater than the largest assigned biased click distance value (CD_{Amax}), exemplary method **401** proceeds directly to decision block **410** shown in FIG. **4B** and described below. In this embodiment, it is possible for a non-authoritative node to have a biased click distance value less than an authoritative node (i.e., be considered more important than the authoritative node wherein importance of a document is based on a lower biased click distance value). For example, referring to exemplary web graph **30** of FIG. **3**, if authoritative node **34** is assigned a biased click distance value of +3, authoritative node **48** is assigned a biased click distance value of 0, and edge value *s* is +1, non-authoritative nodes **39** have a calculated biased click distance value of +1 (i.e., the sum of the assigned biased click distance value of the closest authoritative node **48**, 0, and edge value *s*, +1.

At decision block **410** shown in FIG. **4B**, a determination is made by a system administrator whether to assign at least two different edge values to two or more edges of a web graph. If a decision is made to assign at least two different edge values to two or more edges of a web graph, exemplary method **401** proceeds to step **411**, wherein at least two different edge values are assigned to two or more edges of a web graph. For example, referring to exemplary web graph **30** shown in FIG. **3**, any two of edge values *p*, *q*, *r*, *s* and *t* may be assigned at least two different numbers. From step **411**, exemplary method **401** proceeds to step **414** described below.

Returning to decision block **410**, if a decision is made not to assign at least two different edge values to two or more edges of a web graph, exemplary method **401** proceeds to step **412**, wherein the same edge value is assigned to each edge of a web graph, and the edge value is a value other than 1. For example, referring to exemplary web graph **30** shown in FIG. **3**, each of edge values *p*, *q*, *r*, *s* and *t* are assigned the same number and a number other than 1. From step **412**, exemplary method **401** proceeds to step **414** described below.

Returning to decision block **407** shown in FIG. **4A**, if a decision is made not to assign an edge weight to one or more edges of a web graph, exemplary method **401** proceeds to step **413**, wherein a default edge value (e.g., +1) is used for each edge of a web graph so that the edges of the web graph have a minimal effect on calculated biased click distance values. In this embodiment, factors such as the number and location of authoritative nodes have a greater

effect on calculated biased click distance values than the default edge values. From step **413**, exemplary method **401** proceeds to step **414** shown in FIG. **4B**.

In step **414**, biased click distance values (CD_c) for non-authoritative nodes (or documents or URLs) are calculated. As described in more detail below, the biased click distance value for a given target node (i.e., non-authoritative node) ($CD_{Ctarget}$) linked directly to an authoritative node may be calculated using the formula:

$$CD_{Ctarget} = \min(CD_{Aclosest} + EdgeWeight),$$

wherein $CD_{Aclosest}$ represents the assigned biased click distance value for the authoritative node closest to the target node; and *EdgeWeight* (also referred to herein as *EdgeValue*) represents the edge value or edge weight assigned to the edge linking the closest authoritative node to the target node. The $\min(x)$ function is used to indicate that a minimal calculated biased click distance value is used for a given node, for example, if the node is linked directly to two authoritative nodes. The biased click distance value for a given target node (i.e., non-authoritative node) ($CD_{Ctarget}$) other than those linked directly to an authoritative node may be calculated using the formula:

$$CD_{Ctarget} = \min(CD_{Cmin} + EdgeWeight),$$

wherein CD_{Cmin} represents the calculated biased click distance value of an adjacent node having the lowest calculated biased click distance value; and *EdgeWeight* represents the edge value or edge weight assigned to the edge linking the adjacent node having the lowest calculated biased click distance value and the target node. From step **414**, exemplary method **401** proceeds to step **415**.

In step **415**, the resulting biased click distance values, assigned (CD_A) and calculated (CD_c), are tested by a system administrator. Typically, the system administrator tests the system by executing one or more search queries using the resulting biased click distance values (assigned (CD_A) and calculated (CD_c)). If the system administrator notices obviously irrelevant content coming back, the system administrator can use the above-described biasing tools/techniques to downgrade one or more sites, for example, archive folders or web sites, generating the irrelevant content. The above-described test enables a system administrator to evaluate the biased click distance values for possible inconsistencies between (i) the actual importance of a given document within a

network space and (ii) the importance of the document as indicated by its biased click distance value. From step **415**, exemplary method **401** proceeds to decision block **416**.

At decision block **416**, a determination is made by a system administrator whether to downgrade any biased click distance values in order to more closely represent the importance of a given document within a network space. If a decision is made to downgrade one or more biased click distance values in order to more closely represent the importance of one or more documents within a network space, exemplary method **401** proceeds to step **417**, wherein the biased click distance values of one or more documents (or URLs) are adjusted either negatively or positively. From step **417**, exemplary method **401** proceeds to step **418**.

Returning to decision block **416**, if a decision is made not to downgrade one or more biased click distance values in order to more closely represent the importance of one or more documents within a network space, exemplary method **401** proceeds directly to step **418**. In step **418**, the biased click distance values assigned to authoritative nodes and calculated for non-authoritative nodes are utilized in a ranking function to determine an overall document relevance score for each document within a network space. From step **418**, exemplary method **401** proceeds to end block **419**.

As discussed above, biased click distance values (CD_c) for non-authoritative nodes (or URLs) on a web graph are calculated based on the shortest distance between a given non-authoritative node (or URLs), also referred to as a "target node," and the closest authoritative node (or URL). One exemplary process for calculating the biased click distance values (CD_c) for non-authoritative URLs within a network space is depicted in FIGS. **5A-5B**.

FIGS. **5A-5B** illustrate a logical flow diagram of an exemplary process **40** for calculating the biased click distance (CD_c) for non-authoritative nodes (or URLs) within a network space. Exemplary process **40** starts at block **4140** and proceeds to step **4141**, where a web graph comprising (i) authoritative nodes with their assigned biased click distance values (CD_A), (ii) non-authoritative nodes, (iii) links between nodes, and (iv) edge values for each link is loaded from a database into memory. (See, for example, exemplary web graph **30** in FIG. **3**). The web graph may have been previously generated using an indexing procedure as described above. From step **4141**, exemplary process **40** proceeds to step **4142**.

In step **4142**, biased click distance values (CD_C) for non-authoritative nodes are initialized to a maximum biased click distance value, such as infinity. Assigning a maximum biased click distance value, such as infinity, to the non-authoritative nodes identifies nodes for which a biased click distance value (CD_C) needs to be calculated. Once initialization of maximum biased click distance values is complete, exemplary process **40** proceeds to step **4143**.

In step **4143**, the m authoritative nodes are inserted into a queue. The m authoritative nodes inserted into the queue correspond to the m most authoritative nodes of the network space as pre-determined by a system administrator or some other system determinator. Once the m authoritative nodes are added to the queue, exemplary process **40** proceeds to decision block **4144**.

At decision block **4144**, a determination is made by the application code as to whether the queue is empty. An empty queue signifies that all nodes of the web graph have either (i) obtained an assigned biased click distance value (CD_A) or (ii) had their biased click distance value calculated (CD_C). If the queue is empty, exemplary process **40** proceeds to end block **4145** where exemplary process **40** ends. However, if the queue is not empty, exemplary process **40** continues to step **4146**.

In step **4146**, the node having the smallest biased click distance value (i.e., CD_A or CD_C) is removed from the queue. This node is referred to herein as "the current node." During the first iteration through exemplary process **40**, the authoritative node having the smallest assigned biased click distance value (i.e., CD_{Amin}) is the current node. During subsequent iterations through exemplary process **40**, the node having the smallest biased click distance value may be an authoritative node or a non-authoritative node. During the last iteration through exemplary process **40**, the node having the smallest assigned biased click distance value will typically be a non-authoritative node. Once the node having the smallest biased click distance value (i.e., CD_A or CD_C) is removed from the queue, exemplary process **40** proceeds to decision block **4147**.

At decision block **4147**, a determination is made by the application code as to whether the current node has any target nodes. As used herein, the term "target node" or "target nodes" refers to one or more nodes linked to the current node. If the current node does not have any target nodes, exemplary process **40** returns to decision block **4144** to again determine whether the queue is empty, and then proceeds as discussed above. However, if the current node has one or more target nodes, exemplary process **40** proceeds to step **4148**.

In step **4148**, a target node associated with the current node is retrieved from the web graph and evaluated. For example, referring to exemplary web graph **30** of FIG. **3**, if authoritative node **48** is the current node (i.e., the node having the smallest biased click distance value), any one of non-authoritative nodes **39** could be the target node (i.e., a node linked to authoritative node **48** and having an initial biased click distance value set to infinity). Once a current node and a target node are selected, exemplary process **40** proceeds to decision block **4149**.

At decision block **4149**, a determination is made by the application code whether the click distance associated with the target node biased click distance value is greater than the biased click distance value of the current node plus an edge weight value for the edge connecting the current node to the target node. If a determination is made that the target node biased click distance value is greater than the biased click distance value of the current node plus an edge weight value for the edge connecting the current node to the target node, exemplary process **40** proceeds to step **4150** (shown in FIG. **5B**), wherein the target node biased click distance value is updated to equal the biased click distance value of the current node plus the edge weight value of the edge connecting the current node to the target node.

During the first iteration through exemplary process **40**, all target nodes will have an initial target node biased click distance value set to infinity. Consequently, exemplary process **40** will proceed to step **4150**, wherein the biased click distance value of the target node is updated as described above. However, in subsequent iterations through exemplary process **40**, the selected target node may, for example, have an initial target node biased click distance value set to infinity (exemplary process **40** will proceed to step **4150**) or may have a biased click distance value previously configured by the system administrator (e.g., the target node is an authoritative node). From step **4150**, exemplary process **40** proceeds to step **4151**.

In step **4151**, the current node and the target node with an updated target node biased click distance value are both added to the queue. From step **4150**, exemplary process **40** returns to decision block **4146** (shown in FIG. **5A**) and continues as described above.

Returning to decision block **4149** (shown in FIG. **5A**), if a determination is made that the target node biased click distance value is not greater than the biased click distance value of the current node plus an edge weight value for the edge connecting the current node to the target node, (i) the

target node keeps its calculated target node biased click distance value, (ii) the target node remains out of the queue, and (iii) exemplary process **40** returns to decision block **4147** (shown in FIG. **5A**), where a determination is made whether the current node has any other target nodes. If a determination is made that the current node does not have another target node, exemplary process **40** returns to decision block **4144** and continues as described above. If a determination is made that the current node has another target node, exemplary process **40** proceeds to step **4148** and continues as described above.

When exemplary process **40** returns to step **4148**, another target node associated with the current node is selected and evaluated as described above. If the selected target node has not been selected before, the target node will have an initial biased click distance value set to infinity, and exemplary process **40** will proceed to step **4150** as described above.

The above-described exemplary method of providing a biased click distance value to all nodes on a web graph prevents a biased click distance value of a given target node from being changed if the biased click distance value is lower than the sum of a biased click distance value of a current node plus an edge value of the edge linking the target node to the current node.

Once all nodes of a given web graph have been determined and optionally downgraded (or optionally upgraded), if so desired, the biased click distance values for each document may be used as a parameter in a ranking function to provide a document relevance score for each document. Such a document relevance score may be used to rank search results of a search query. An exemplary method of ranking search results generating using a ranking function containing a biased click distance value parameter is shown in FIG. **6**.

FIG. **6** provides a logic flow diagram showing exemplary steps in exemplary method **20**, wherein exemplary method **20** comprises a method of ranking search results generating using a ranking function containing a biased click distance value parameter. As shown in FIG. **6**, exemplary method **20** starts at block **201** and proceeds to step **202**. In step **202**, a user requests a search by inputting a search query. Prior to step **202**, biased click distance values for each of the documents on the network have previously been calculated. From step **202**, exemplary method **20** proceeds to step **203**.

In step **203**, the biased click distance value for each document on a network is merged with any other document statistics (e.g., query-independent statistics) for each document stored in the index. Merging the biased click distance values with other document statistics allows for a faster query response

time since all the information related to ranking is clustered together. Accordingly, each document listed in the index has an associated biased click distance value after the merge. Once the merge is complete, exemplary method **20** proceeds to step **204**.

In step **204**, query-independent document statistics for a given document, including a biased click distance value, are provided as a component of a ranking function. Query-dependent data is also provided for the given document, typically as a separate component of the ranking function. The query-dependent data or content-related portion of the ranking function depends on the actual search terms and the content of the given document.

In one embodiment, the ranking function comprises a sum of at least one query-dependent (QD) component and at least one query-independent (QID) component, such as

$$Score = QD(doc, query) + QID(doc)$$

The QD component can be any document scoring function. In one embodiment, the QD component corresponds to a field weighted scoring function described in U.S. Patent Application Serial No. 10/804,326 entitled "FIELD WEIGHTING IN TEXT DOCUMENT SEARCHING," filed on March 18, 2004, the subject matter of which is hereby incorporated in its entirety by reference. As provided in U.S. Patent Application Serial No. 10/804,326, one equation that may be used as a representation of the field weighted scoring function is as follows:

$$QD(doc, query) = \sum \frac{wtf'(k_1 + 1)}{k_1 + wtf'} \times \log(\frac{N}{n})$$

wherein:

wtf' represents a weighted term frequency or sum of term frequencies of given terms in the search query multiplied by weights across all fields (e.g., the title, the body, etc. of the document) and normalized according to the length of each field and the corresponding average length,

N represents a number of documents on the network,

n represents a number of documents containing a query term, and

k_1 is a tunable constant.

The above terms and equation are further described in detail in U.S. Patent Application Serial No. 10/804,326, the subject matter of which is hereby incorporated in its entirety by reference.

The QID component can be any transformation of a biased click distance value and other document statistic (such as a URL depth) for a given document. In one embodiment, the QID component comprises a function as follows:

$$QID(doc) = w_{cd} \frac{k_{cd}}{k_{cd} + \frac{b_{cd}\frac{CD}{k_{ew}} + b_{ud}UD}{b_{cd} + b_{ud}}}$$

wherein:

w_{cd} represents a weight of a query-independent component such as a component containing a biased click distance parameter,

b_{cd} represents a weight of a biased click distance relative to the URL depth,

b_{ud} represents a weight of a URL depth,

CD represents a computed click distance or assigned biased click distance for a document,

k_{ew} represents a tuning constant that is determined by optimizing the precision of the ranking function, similar to other tuning parameters (i.e., k_{ew} may represent the edge weight value when all edges have the same edge weight value, or k_{ew} may represent the average or mean edge value when edge weight values differ from one another),

UD represents a URL depth, and

k_{cd} is the biased click distance saturation constant.

The weighted terms (w_{cd}, b_{cd}, and b_{ud}) assist in defining the importance of each of their related terms (i.e., the component containing a biased click distance parameter, the biased click distance value for a given document, and the URL depth of the given document respectively) and ultimately the outcome of the scoring functions.

The URL depth (UD) is an optional addition to the above-referenced query-independent component to smooth the effect that the biased click distance value may have on the scoring function. For example, in some cases, a document that is not very important (i.e., has a large URL depth) may have a short biased click distance value. The URL depth is represented by the number of slashes in a document's URL. For example, www.example.com\d1\d2\d3\d4.htm includes four slashes and would therefore have a URL depth of 4. This document however, may have a link directly from the main page www.example.com giving it a relatively low biased click distance

value. Including the URL depth term in the above-referenced function and weighting the URL depth term against the biased click distance value compensates for a relatively high biased click distance value to more accurately reflect the document's importance within the network. Depending on the network, a URL depth of 3 or more may be considered a deep link.

In one embodiment, the ranking function used to determine a document relevance score for a given document comprises a function as follows:

$$Score = \sum \frac{wtf'(k_1+1)}{k_1+wtf'} \times \log(\frac{N}{n}) + w_{cd} \frac{k_{cd}}{k_{cd} + \frac{b_{cd}\frac{CD}{k_{ew}} + b_{ud}UD}{b_{cd}+b_{ud}}}$$

wherein the terms are as described above.

In other embodiments, the URL depth may be removed from the ranking function or other components may be added to the ranking function to improve the accuracy of the query-dependent component, the query-independent component, or both. Furthermore, the above-described query-independent component containing a biased click distance parameter may be incorporated into other ranking functions (not shown) to improve ranking of search results.

Once document statistics for a given document are provided to a ranking function in step **204**, exemplary method **20** proceeds to step **205**. In step **205**, a document relevance score is determined for a given document, stored in memory, and associated with the given document. From step **205**, exemplary method **20** proceeds to decision block **206**.

At decision block **206**, a determination is made by application code whether a document relevance score has been calculated for each document within a network. If a determination is made that a document relevance score has not been calculated for each document within a network, exemplary method **20** returns to step **204** and continues as described above. If a determination is made that a document relevance score has been calculated for each document within a network, exemplary method **20** proceeds to step **207**.

In step **207**, the search results of the query comprising numerous documents are ranked according to their associated document relevance scores. The resulting document relevance scores take into account the biased click distance value of each of the documents within the network. Once the search results are ranked, exemplary method **20** proceeds to step **208** where ranked

results are displayed to a user. From step **208**, exemplary method **20** proceeds to step **209** where highest ranked results are selected and viewed by the user. From step **209**, exemplary method **20** proceeds to step **210** where exemplary method **20** ends.

In addition to the above-described methods of generating a document relevance score for documents within a network and using document relevance scores to rank search results of a search query, computer readable medium having stored thereon computer-executable instructions for performing the above-described methods are also disclosed herein.

Computing systems are also disclosed herein. An exemplary computing system contains at least one application module usable on the computing system, wherein the at least one application module comprises application code loaded thereon, wherein the application code performs a method of generating a document relevance score for documents within a network. The application code may be loaded onto the computing system using any of the above-described computer readable medium having thereon computer-executable instructions for generating a document relevance score for documents within a network and using document relevance scores to rank search results of a search query as described above.

While the specification has been described in detail with respect to specific embodiments thereof, it will be appreciated that those skilled in the art, upon attaining an understanding of the foregoing, may readily conceive of alterations to, variations of, and equivalents to these embodiments. Accordingly, the scope of the disclosed methods, computer readable medium, and computing systems should be assessed as that of the appended claims and any equivalents thereto.

WHAT IS CLAIMED IS:

1. A computer readable medium having stored thereon computer-executable instructions for ranking documents on a network, said computer-executable instructions utilizing a ranking function that comprises one or more query-independent components, wherein at least one query-independent component includes a biased click distance parameter that takes into account a biased click distance value for each document on the network.

2. The computer readable medium of Claim 1, wherein a biased click distance value is assigned to *m* authoritative documents on a network comprising N total documents with *m* being greater than or equal to 2 and less than N.

3. The computer readable medium of Claim 2, wherein at least two of the assigned biased click distance values differ from one another.

4. The computer readable medium of Claim 1, wherein the ranking function further comprises at least one query-independent component that includes an edge value parameter that takes into account edge values of each edge on the network, wherein one or more edge values are a number other than 1.

5. The computer readable medium of Claim 4, wherein the edge values are equal to one another and equal a number other than 1.

6. The computer readable medium of Claim 4, wherein the edge values are equal to one another and are equal to or greater than a highest biased click distance value initially assigned to one or more authoritative documents.

7. The computer readable medium of Claim 4, wherein the at least one query-independent component includes both the biased click distance parameter and the edge value parameter.

8. The computer readable medium of Claim 1, further comprising computer-executable instructions for assigning a score generated by the ranking function to each document on the network, said score being used to rank documents in ascending or descending order.

9. The computer readable medium of Claim 7, wherein the score for each document is generated using a formula:

$$\text{Score} = \sum \frac{wtf'(k_1+1)}{k_1+wtf'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{k_{cd} + \frac{b_{cd}\frac{CD}{k_{ew}} + b_{ud}UD}{b_{cd}+b_{ud}}}$$

wherein:

wtf' represents a weighted term frequency,

N represents a number of documents on the network,

n represents a number of documents containing a query term,

w_{cd} represents a weight of a query-independent component,

b_{cd} represents a weight of a click distance,

b_{ud} represents a weight of a URL depth,

CD represents a computed click distance or assigned biased click distance for a document,

k_{ew} represents a tuning constant related to edge weights,

UD represents a URL depth, and

k_{cd} and k_1 are constants.

10. The computer readable medium of Claim 1, further comprising computer-executable instructions for accepting a search inquiry inputted by a user, conducting a search of the documents on the network to generate search results comprising multiple documents, ranking the multiple documents of the search results using the ranking function to generate ranked search results, and displaying the ranked search results to the user.

11. A computing system containing at least one application module usable on the computing system, wherein the at least one application module comprises application code loaded thereon from the computer readable medium of Claim 1.

12. A method of determining a document relevance score for a document on a network, said method comprising the steps of:

storing document and link information for documents on a network;

generating a representation of the network from the document and link information, wherein the representation of the network includes nodes that represent the documents and edges that represent the links;

assigning a biased click distance value to at least two nodes on the network, wherein the nodes that are assigned a biased click distance value are authoritative nodes;

computing a click distance for each of the non-authoritative nodes in the representation of the network, wherein the click distance for a given non-authoritative node is measured from the given non-authoritative node to an authoritative node closest to the given non-authoritative node, said computing step resulting in a biased click distance value for each document; and

using the biased click distance value for each document to determine the document relevance score of a given document on the network.

13. The method of Claim 12, wherein at least two authoritative nodes have assigned biased click distance values that differ from one another.

14. The method of Claim 12, further comprising the step of:

assigning to each edge in the representation an edge value, wherein the edge values are equal to or greater than 1.

15. The method of Claim 14, wherein each edge value is greater than a highest biased click distance value assigned to any of the authoritative nodes.

16. The method of Claim 12, wherein the document relevance score for each document on the network is generated using a formula:

$$\text{Score} = \sum \frac{wtf'(k_1+1)}{k_1+wtf'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{k_{cd} + \frac{b_{cd}\frac{CD}{k_{ew}} + b_{ud}UD}{b_{cd}+b_{ud}}}$$

wherein:

wtf' represents a weighted term frequency,

N represents a number of documents on the network,

n represents a number of documents containing a query term,

w_{cd} represents a weight of a query-independent component,

b_{cd} represents a weight of a click distance,

b_{ud} represents a weight of a URL depth,

CD represents a computed click distance or assigned biased click distance for a document,

k_{ew} represents a tuning constant related to edge weights;

UD represents a URL depth, and

k_{cd} and k_1 are constants.

17. A method of ranking documents on a network, said method comprising the steps of:

determining a document relevance score for each document on the network using the method of Claim 12; and

ranking the documents in descending order based on the document relevance scores of each document.

18. A method of ranking search results of a search query, said method comprising the steps of:

determining a document relevance score for each document in the search results of a search query using the method of Claim 12; and

ranking the documents in descending order based on the document relevance scores of each document.

19. A computer readable medium having stored thereon computer-executable instructions for performing the method of claim 12.

20. A computing system containing at least one application module usable on the computing system, wherein the at least one application module comprises application code for performing a method of determining a document relevance score for a document on a network, said method comprising the steps of:

assigning a biased click distance value to at least two nodes on a representation of the network, wherein the nodes that are assigned a biased click distance value are authoritative nodes;

computing a click distance for each non-authoritative node in the representation of the network, wherein the click distance for a given non-authoritative node is measured from the given non-authoritative node to an authoritative node closest to the given non-authoritative node, said computing step resulting in a biased click distance value for each document; and

using the biased click distance value for each document to determine the document relevance score of a given document on the network.