

SYSTEM AND METHOD FOR CONFIRMING AUTHORSHIP OF DOCUMENTS

TECHNICAL FIELD

[0001] The disclosed embodiments relate generally to confirming authorship of documents.

BACKGROUND

[0002] Content on websites often include authorship information. For example, an article may include a byline listing entities that authored the article. However, the authorship information may not be accurate or may not be valid. For example, an article may include authorship information asserting that a particular entity authored the article when in fact the article was not authored by the particular entity. Inaccuracies and/or invalidity of authorship information cause inaccuracies when obtaining search results and/or confusion to users when the search results are presented. Thus, it would be desirable to confirm authorship of content in website so that the entities that authored the content are properly identified.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] The embodiments disclosed herein are illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings. Like reference numerals refer to corresponding parts throughout the drawings.

[0004] Figure 1 is a block diagram illustrating a network system, according to some embodiments.

[0005] Figure 2 is a block diagram illustrating authorship attributes that are usable to confirm authorship of documents, according to some embodiments.

[0006] Figure 3 is a block diagram illustrating authorship attributes that are usable to confirm authorship of documents, according to some embodiments.

[0007] Figure 4 is a block diagram illustrating an authorship identifier that is usable to confirm authorship of documents, according to some embodiments.

[0008] Figure 5 is a block diagram illustrating a profile for an entity, according to some embodiments.

[0009] Figure 6 is a block diagram illustrating a server system, according to some embodiments.

[0010] Figure 7 is a block diagram illustrating a content server system, according to some embodiments.

[0011] Figure 8 is a flowchart of a method for confirming authorship of documents, according to some embodiments.

[0012] Figure 9 is a flowchart of a method for traversing a link, according to some embodiments.

DESCRIPTION OF EXAMPLE EMBODIMENTS

[0013] The description that follows includes example systems, methods, techniques, instruction sequences, and computing machine program products that embody illustrative embodiments. In the following description, for purposes of explanation, numerous specific details are set forth in order to provide an understanding of the described embodiments. It will be evident, however, to those skilled in the art that some embodiments may be practiced without these specific details. In general, well-known instruction instances, protocols, structures and techniques have not been shown in detail.

[0014] It will also be understood that, although the terms “first,” “second,” etc. may be used herein to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. For example, a first contact could be termed a second contact, and, similarly, a second contact could be termed a first contact, which changing the meaning of the description, so long as all occurrences of the “first contact” are renamed consistently and all occurrences of the second contact are renamed consistently. The first contact and the second contact are both contacts, but they are not the same contact.

[0015] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the claims. As used in the description of the embodiments and the appended claims, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will also be understood that the term “and/or” as used herein refers to and encompasses any and all possible combinations of one or more of the associated listed items. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification,

specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0016] As used herein, the term “if” may be construed to mean “when” or “upon” or “in response to determining” or “in accordance with a determination” or “in response to detecting,” that a stated condition precedent is true, depending on the context. Similarly, the phrase “if it is determined (that a stated condition precedent is true)” or “if (a stated condition precedent is true)” or “when (a stated condition precedent is true)” may be construed to mean “upon determining” or “in response to determining” or “in accordance with a determination” or “upon detecting” or “in response to detecting” that the stated condition precedent is true, depending on the context.

[0017] The embodiments described herein provide techniques for confirming authorship of documents.

[0018] Figure 1 is a block diagram illustrating a network system 100, according to some embodiments. The network system 100 includes a server system 102 coupled to content server systems 104, 106, and 108 via network 120. The network 120 can generally include any type of wired or wireless communication channel capable of coupling together computing nodes. This includes, but is not limited to, a local area network, a wide area network, or a combination of networks. In some embodiments, the network 120 includes the Internet.

[0019] Server system 102 (sometimes herein called server 102 for ease of reference, noting that some embodiments of server system 102 include multiple servers) is configured to confirm authorship of documents hosted on content server systems (sometimes herein called content servers for ease of reference) 104, 106, and 108, as described in more detail below with reference to Figures 2-9. In some embodiments, server 102 includes a search engine.

[0020] In some embodiments, content servers 104, 106, and 108 include web servers that host documents 110, 112, and 114 for websites that are accessible via network 120. The websites include, but are not limited to, news websites, blogs, social networking websites, academic websites, and corporate websites. At least a subset of the documents 110, 112, and 114 include content where each piece of content is authored by respective entities. An entity is defined to be a person, a group of people, or an organization. The content includes, but is

not limited to, articles, blog entries, books, music, videos, literature, posts on online discussion forums, posts on social networking websites, and the like.

[0021] Note that although Figure 1 shows a single server (e.g., the server 102) and three content servers (e.g., the content servers 104, 106, and 108), any number of servers and content servers may be present in network system 100. For example, each of server 102, content server 104, content server 106, and content server 108 may include a plurality of distributed servers. The plurality of distributed servers may provide load balancing and/or may provide low-latency points of access to other computer systems. The distributed servers may be located within a single location (e.g., a data center, a building, etc.) or may be geographically distributed across multiple locations (e.g., data centers at various geographical locations, etc.).

[0022] Also note that although the embodiments described herein refer to server 102, content server 104, content server 106, and content server 108, the embodiments may be applied to multiple servers and content servers. Furthermore, the functionality of any of server 102, content server 104, content server 106, and content server 108 may be implemented within a single server (or a set of distributed servers). For example, server 102 and content server 104 may be located on the same server (or the same set of distributed servers).

[0023] In some embodiments, predefined authorship attributes included in links are used to confirm authorship of documents. Figure 2 is a block diagram 200 illustrating authorship attributes that are usable to confirm authorship of documents, according to some embodiments. As illustrated in Figure 2, a document 204 includes a link 206 and a document 214 includes a link 216. Documents 204 and 214 may be any of the documents from the documents 110, 112, and 114 hosted on the content servers 104, 106, and 108, respectively.

[0024] Document 204 is included in (i.e., stored at, or stored on, or accessible from) a website 202. A website includes a collection of documents that are authored, owned, or maintained by one or more entities. Furthermore, a website may be located in a domain (e.g., www.example.com), a subdomain (e.g., me.example.com), a subdirectory within a domain or subdomain (e.g., www.example.com/me/), and a set of dynamically-generated web pages in a domain or subdomain (e.g., www.example.com/profile?user_id=1001).

[0025] In some embodiments, document 214 is included in website 202 (i.e., the same website that includes document 204). In some embodiments, document 214 is included in a

website that is separate and distinct from website 202. In some embodiments, document 214 is a profile page for the entity that document 214 asserts (e.g., via a byline or the like) is an author or contributor to document 214.

[0026] In some embodiments, link 206 includes a target attribute 208 and an authorship attribute 210. The target attribute 208 includes information identifying a target document to which link 206 refers, or information specifying a location at which a target document can be accessed. As illustrated in Figure 2, the target document referred to by the target attribute 208 is document 214. In some implementations, the target attribute 208 includes a universal resource locator (URL). In some implementations, the target attribute 208 includes a universal resource identifier (URI). The authorship attribute 210 asserts authorship of document 204 by an entity associated with the target document to which link 206 refers. As illustrated in Figure 2, the authorship attribute 210 of link 206 asserts that the entity associated with document 214 is an author of document 204. In some implementations, the authorship attribute 210 is rel="me". In some implementations, the authorship attribute 210 is rel="author". In some implementations, the authorship attribute 210 is represented by a URL parameter in the target URL (target attribute 208), such as a rel="author" parameter in the URL (e.g., [http://www.example.com/article.php?article_id=1001&rel="author"](http://www.example.com/article.php?article_id=1001&rel='author')), instead of as a link parameter that is separate from the target attribute 208.

[0027] In some embodiments, link 216 includes a target attribute 218 and an authorship attribute 220. The target attribute 218 includes information identifying a target document to which link 216 refers, or information specifying a location at which a target document can be accessed. As illustrated in Figure 2, the target document referred to by the target attribute 218 is the website 202. For example, the target attribute 218 may refer to a root document of a domain (e.g., www.example.com), root document of a subdomain (e.g., me.example.com), a root of a subdirectory within a domain or subdomain (e.g., www.example.com/me/), and root of a set of dynamically-generated web pages in a domain or subdomain (e.g., www.example.com/profile?user_id=1001). In some implementations, the target attribute 218 includes a universal resource locator (URL). In some implementations, the target attribute 218 includes a universal resource identifier (URI). The authorship attribute 220 indicates that the entity associated with document 214 is an author of or contributor to content at the website 202 to which the target attribute 218 refers. As illustrated in Figure 2, the authorship attribute 220 of link 216 indicates that the entity

associated with document 214 is an author of or contributor to content at website 202. In some implementations, the authorship attribute 220 is rel="contributor-to". In some circumstances, the authorship attribute 220 is rel="me", which may be used to indicate that the entire website 202 contains content by the entity associated with document 214.

[0028] The following discussion provides is an example of how an entity may use the authorship attribute 210 and the authorship attribute 220 to confirm authorship of a document. Assume that document 214 is a profile page for the entity that authored document 204 and that the entity that authored document 204 desires to provide confirmation that the entity did in fact author document 204. In order to do so, the entity that authored document 204 (1) includes link 216 in document (profile page) 214 and (2) includes link 206 in document 204. Assuming that document 214 has the URL "http://www.example.com/profiles/johndoe", link 206 is View my profile, where rel="author" is the authorship attribute 210. Assuming that a root of the website that hosts document 204 has the URL "http://www.example2.com/articles/", link 216 is Read my articles at example2.com , where rel="contributor-to" is the authorship attribute 220. The loop formed by link 206 and link 216 confirms that the document is authored by the entity associated with document 214.

[0029] Note that although Figure 2 shows only one link for each of the documents, each of the documents may include multiple links in which at least one of the links includes the authorship attribute 210 and/or the authorship attribute 220.

[0030] Figure 2 illustrates a technique for confirming authorship of documents when the authorship attribute 210 and the authorship attribute 220 are included in links in documents that directly link to each other. However, the authorship of documents may be confirmed when the authorship attribute 210 and the authorship attribute 220 are included in links in documents that are not directly linked to each other. Figure 3 is a block diagram 300 illustrating authorship attributes that are usable to confirm authorship of documents, according to some embodiments. Figure 3 is similar to Figure 2 except that in Figure 3, the target attribute 208 of link 206 refers to a document 304, which in turn includes a link 306 having a target 308 that refers to document 214. Links 206, 306, and 216 form a chain of links to documents. When confirming authorship of document 204, server 102 traverses links in the chain of links until document 214 that includes the authorship attribute 220 (e.g., the

rel="contributor-to" attribute) is identified. In general, any number of links included in documents may exist between document 204 and document 214. Links 206 and 306 (and any other link in the chain of links between links 206 and 216) include the authorship attribute 210, which is the basis for establishing a chain of links. In other words, the chain of links is established by a series of links in which a predecessor document includes a predecessor link that has the authorship attribute 210 and the target of the predecessor link is a successor document that includes a successor link that has the authorship attribute 210. In some embodiments, document 304 (and any document between documents 204 and 214) is included in the website 202. In some embodiments, document 304 (and any document between documents 204 and 214) is included in a website that is separate and distinct from website 202.

[0031] The confirmation techniques discussed with reference to Figures 2 and 3 analyze attributes included in links to confirm authorship of documents. The confirmation techniques discussed below with reference to Figures 4 and 5 use authorship identifiers to confirm authorship of documents.

[0032] Figure 4 is a block diagram 400 illustrating an authorship identifier that is usable to confirm authorship of documents, according to some embodiments. A document 402 includes an authorship identifier 404 asserting authorship of the document 404 by an entity. In some embodiments, the authorship identifier 404 is an email address of the entity that authored the document 402. In some embodiments, the authorship identifier 404 is byline of the entity that authored the document 402. The fact that the document 402 includes the authorship identifier 404 is not sufficient by itself to confirm that the document 402 is authored by the entity. Thus, in some embodiments, the server 102 uses information from a profile for the entity that is associated with the authorship identifier 404 to confirm that the entity authored the document 402. The entity that is associated with the authorship identifier 404 is confirmed to be the author of the document 402 when the profile for the entity indicates that the entity has confirmed that the authorship identifier 404 is included in documents authored by the entity that are hosted on a website (e.g., website 202).

[0033] Figure 5 is a block diagram 500 illustrating profile 502 for an entity 504, according to some embodiments. The profile 502 for entity 504 includes an authorship identifier 506 that is associated with at least one website and a confirmation status 508 for the authorship identifier 506 that indicates whether entity 504 has confirmed that the authorship identifier 506 is included in documents authored by entity 504 that are hosted on the website.

When the authorship identifier 506 corresponds to an authorship identifier included in documents hosted on the at least one website, entity 504 is confirmed to be the author of the documents hosted on the at least one website. For example, when the authorship identifier 506 has been confirmed to be associated with a website that hosts the document 402, entity 504 is confirmed to be the author of document 402.

[0034] In some embodiments, when the authorship identifier 506 is an email address in a domain that hosts a website including documents authored by entity 504, server 102 sends a confirmation email that includes a confirmation link (and/or a confirmation code) to the email address. When entity 504 clicks on the confirmation link (and/or enters the confirmation code on a predetermined webpage), entity 504 confirms that the email address is included in documents authored by entity 504 that are hosted on the website of the domain. When documents hosted on the website of the domain include the email address, entity 504 is confirmed to be author of the documents

[0035] In some embodiments, after the email address has been confirmed, entity 504 provide a byline that is included in documents authored by entity 504 that are hosted on the website of the domain. In these embodiments, when documents hosted on the website of the domain include the byline, entity 504 is confirmed to be author of the documents.

[0036] In some embodiments, the profile 502 for entity 504 includes content 510 associated with entity 504. For example, profile 502 may include information relating to articles (e.g., links to the articles, titles of the articles, summaries of the articles) that entity 504 has authored. In some embodiments, profile 502 is a profile on a social networking website.

[0037] In some embodiments, the authorship of a document is confirmed using both the authorship identifier and the authorship attributes included in links.

[0038] Figure 6 is a block diagram illustrating server system 102, according to some embodiments. The server system 102 typically includes one or more processing units (CPU's, sometimes called processors) 602 for executing programs (e.g., programs stored in memory 610), one or more network or other communications interfaces 604, memory 610, and one or more communication buses 609 for interconnecting these components. The communication buses 609 may include circuitry (sometimes called a chipset) that interconnects and controls communications between system components. Server system 102 optionally includes (but typically does not include) a user interface 605 comprising a display

device 606 and input devices 608 (e.g., keyboard, mouse, touch screen, keypads, etc.). Memory 610 includes high-speed random access memory, such as DRAM, SRAM, DDR RAM or other random access solid state memory devices; and typically includes non-volatile memory, such as one or more magnetic disk storage devices, optical disk storage devices, flash memory devices, or other non-volatile solid state storage devices. Memory 610 optionally includes one or more storage devices remotely located from the CPU(s) 602. Memory 610, or alternately the non-volatile memory device(s) within memory 610, comprises a non-transitory computer readable storage medium. In some embodiments, memory 610 or the computer readable storage medium of memory 610 stores the following programs, modules and data structures, or a subset thereof:

- an operating system 612 that includes procedures for handling various basic system services and for performing hardware dependent tasks;
- a communication module 614 that is used for connecting the server 102 to other computers via the one or more communication interfaces 604 (wired or wireless) and one or more communication networks, such as the Internet, other wide area networks, local area networks, metropolitan area networks, and so on;
- an optional user interface module 616 that receives commands from the user via the input devices 608 and generates user interface objects in the display device 606;
- an authorship confirmation module 618 that confirms authorship of documents, as described herein;
- an optional search index 620 that facilitates identifying search requests, the search index 620 including information relating to documents, authorship of the documents, and a corresponding confirmation status of the authorship of the documents (e.g., confirmed authorship, non-confirmed authorship); and
- optional profiles 622 of entities that include authorship identifiers 624 (e.g., the authorship identifier 404, the authorship identifier 506), confirmation statuses 626 of the authorship identifiers 624 (e.g., the confirmation status 508), and content associated with the entity 628 (e.g., the content 510 associated with the entity 504), as described herein.

[0039] In some embodiments, search index 620 is located on a server that is separate and distinct from server system 102. In some embodiments, the profiles 622 are located on a server that is separate and distinct from server system 102.

[0040] Each of the above identified modules, applications or programs corresponds to a set of instructions, executable by the one or more processors (e.g., CPUs 602) of content server 700, for performing a function described above. The above identified modules, applications or programs (i.e., sets of instructions) need not be implemented as separate software programs, procedures or modules, and thus various subsets of these modules may be combined or otherwise re-arranged in various embodiments. In some embodiments, memory 610 stores additional modules and data structures not described above.

[0041] Although Figure 6 shows a “server,” Figure 6 is intended more as functional description of the various features which may be present in a set of servers than as a structural schematic of the embodiments described herein. In practice, and as recognized by those of ordinary skill in the art, items shown separately could be combined and some items could be separated. For example, some items shown separately in Figure 6 could be implemented on single servers and single items could be implemented by one or more servers. The actual number of servers used to implement server system 102 and how features are allocated among them will vary from one implementation to another, and may depend in part on the amount of data traffic that the system must handle during peak usage periods as well as during average usage periods.

[0042] Figure 7 is a block diagram illustrating a content server 700, according to some embodiments. The content server 700 can be any of the content servers 104, 106, and 108 in Figure 1. The content server 700 typically includes one or more processing units (CPU's, sometimes called processors) 702 for executing programs (e.g., programs stored in memory 710), one or more network or other communications interfaces 704, memory 710, and one or more communication buses 709 for interconnecting these components. The communication buses 709 may include circuitry (sometimes called a chipset) that interconnects and controls communications between system components. The content server 700 optionally includes (but typically does not include) a user interface 705 comprising a display device 706 and input devices 708 (e.g., keyboard, mouse, touch screen, keypads, etc.). Memory 710 includes high-speed random access memory, such as DRAM, SRAM, DDR RAM or other random access solid state memory devices; and typically includes non-volatile memory, such as one or more magnetic disk storage devices, optical disk storage devices, flash memory devices, or other non-volatile solid state storage devices. Memory 710 optionally includes one or more storage devices remotely located from the CPU(s) 702. Memory 710, or alternately the non-volatile memory device(s) within memory 710,

comprises a non-transitory computer readable storage medium. In some embodiments, memory 710 or the computer readable storage medium of memory 710 stores the following programs, modules and data structures, or a subset thereof:

- an operating system 712 that includes procedures for handling various basic system services and for performing hardware dependent tasks;
- a communication module 714 that is used for connecting the content server 700 to other computers via the one or more communication interfaces 704 (wired or wireless) and one or more communication networks, such as the Internet, other wide area networks, local area networks, metropolitan area networks, and so on;
- an optional user interface module 716 that receives commands from the user via the input devices 708 and generates user interface objects in the display device 706;
- a web server module 718 that responds to requests for documents 720 received from other computer systems;
- documents 720 that include content 722 authored by entities, optional authorship identifiers 724 (e.g., the authorship identifier 404, the authorship identifier 506), links 724 that each include a target attribute 728 (e.g., the target attribute 208, the target attribute 218, the target 308), an optional authorship attribute 730 (e.g., the authorship attribute 210, the authorship attribute 210), an optional confirmation status attribute 732 indicating whether the authorship of the document is confirmed or not, and an optional author profile attribute 734 that provides a link to a profile of the entity that authored the document, as described herein; and
- optional profiles 736 of entities that include authorship identifiers 738 (e.g., the authorship identifier 404, the authorship identifier 506), confirmation statuses 740 of the authorship identifiers 738 (e.g., the confirmation status 508), and content 742 associated with the entities (e.g., the content 510 associated with the entity 504), as described herein.

[0043] Content server 700 optionally further includes a security module 750 for ensuring that content that is stored at or by content server 700 and that includes an indicia of authorship (e.g., byline or email address), is received from or on behalf of the entity associated that the indicia of authorship. Security module 750 is configured to prevent, or substantially prevent, storage on content server 700 of content having false indicia of authorship.

[0044] Each of the above identified modules, applications or programs corresponds to a set of instructions, executable by the one or more processors (e.g., CPUs 702) of content server 700, for performing a function described above. The above identified modules, applications or programs (i.e., sets of instructions) need not be implemented as separate software programs, procedures or modules, and thus various subsets of these modules may be combined or otherwise re-arranged in various embodiments. In some embodiments, memory 710 stores additional modules and data structures not described above.

[0045] Although Figure 7 shows a “content server,” Figure 7 is intended more as a functional description of the various features which may be present in a set of servers than as a structural schematic of the embodiments described herein. In practice, and as recognized by those of ordinary skill in the art, items shown separately could be combined and some items could be separated. For example, some items shown separately in Figure 7 could be implemented on single servers and single items could be implemented by one or more servers. The actual number of servers used to implement a content server and how features are allocated among them will vary from one implementation to another, and may depend in part on the amount of data traffic that the system must handle during peak usage periods as well as during average usage periods.

Confirming Authorship of Documents

[0046] Figure 8 is a flowchart of a method 800 for confirming authorship of documents, according to some embodiments. The authorship confirmation module 618 accesses (802) a first document (e.g., document 204) hosted on a first website (e.g., website 202) of a first domain, where the first document is linked to a second document (e.g., document 214) through at least one link in a chain of links, and where a respective link in the chain of links including a first predefined authorship attribute (e.g., the authorship attribute 210) asserting authorship of a respective document including the respective link by a respective entity associated with a respective target document of the respective link.

[0047] In some embodiments, the second document includes a trusted document associated with the entity. In some embodiments, the second document includes a profile of the entity (e.g., the profile 502 for the entity 504).

[0048] In some implementations, the second document is hosted on a second domain that is distinct from the first domain. In some implementations, the second document is hosted on the first domain.

[0049] The authorship confirmation module 618 conditionally confirms (804) authorship of the first document by an entity associated with the second document when the second document includes a second link to the first website of the first domain, where the second link includes a second predefined authorship attribute (e.g., the authorship attribute 220) indicating that the entity associated with the second document is an author of or contributor to content at the first website of the first domain.

[0050] In some embodiments, the second link links to a root of the first website of the first domain. For example, as discussed above with reference to Figure 2, the second link may link to a root document of the first website of the first domain (e.g., www.example.com/index.html, www.example.com).

[0051] In some embodiments, the authorship confirmation module 618 determines that the first document has a non-confirmed authorship when no second document having the second link including the second predefined authorship attribute can be found. For example, the authorship confirmation module 618 determines that document 204 has a non-confirmed authorship if the loop formed between link 206 and link 216 does not exist for the documents 204 and 214.

[0052] When the authorship attribute 210 and the authorship attribute 220 are included in links in documents that directly link to each other, the authorship confirmation module 618 traverses the link including the authorship attribute 210 to the second document prior to conditionally confirming the authorship of the first document by the entity associated with the second document. Figure 9 is a flowchart of a method 900 for traversing a link, according to some embodiments. The authorship confirmation module 618 identifies (902) a first link (e.g., link 206) in the first document (e.g., document 204) that links to the second document (e.g., document 214), wherein the first link includes the first predefined authorship attribute (e.g., the authorship attribute 210). The authorship confirmation module 618 accesses (904) the second document (e.g., document 214) and determines (906) that the second document includes the second link including the second predefined authorship attribute (e.g., the authorship attribute 220).

[0053] In some embodiments, when more than one link in the first document includes the first predefined authorship attribute (e.g., the authorship attribute 210), the authorship confirmation module 618 traverses each link that includes the first predefined authorship attribute to search for a second document that includes the second predefined authorship attribute (e.g., the authorship attribute 220).

[0054] As discussed above, a chain of links may exist between the first and second documents. Thus, in some embodiments, prior to conditionally confirming the authorship of the first document by the entity associated with the second document, the authorship confirmation module 618 traverses links in the chain of links (e.g., links 206, 306, and 216), starting with the first document, until the second document that includes the second link including the second predefined authorship attribute is identified.

[0055] After confirming the authorship of the document, several operations may be performed. In some embodiments, responsive to confirming authorship of the first document by the entity, the authorship confirmation module 618 adds the first document (e.g., document 204) to the profile for the entity (e.g., profile 502 for entity 504).

[0056] In some embodiments, responsive to confirming authorship of the first document by the entity, the authorship confirmation module 618 adds the first document (e.g., document 204) to a social networking webpage for the entity.

[0057] In some embodiments, responsive to confirming authorship of the first document by the entity, the authorship confirmation module 618 adds information relating to the first document (e.g., document 204) to a webpage including information relating to documents that the entity authored. The webpage including information relating to documents that the entity authored may be an author webpage on the first website, an author webpage on a social networking website, a blog for the author, and the like.

[0058] In some embodiments, responsive to confirming authorship of the first document by the entity, the authorship confirmation module 618 associates the first document (e.g., document 204) with the entity (e.g., entity 504) in a search index (e.g., the search index 620).

[0059] In some embodiments, the methods illustrated in Figures 8-9 are governed by instructions that are stored in a computer readable storage medium and that are executed by at least one processor of at least one server. Each of the operations shown in Figure 8-9 correspond to instructions stored in a non-transitory computer memory or computer readable

storage medium. In various implementations, the non-transitory computer readable storage medium includes a magnetic or optical disk storage device, solid state storage devices such as Flash memory, or other non-volatile memory device or devices. The computer readable instructions stored on the non-transitory computer readable storage medium may be in source code, assembly language code, object code, or other instruction format that is interpreted and/or executable by one or more processors.

[0060] Plural instances may be provided for components, operations or structures described herein as a single instance. Finally, boundaries between various components, operations, and data stores are somewhat arbitrary, and particular operations are illustrated in the context of specific illustrative configurations. Other allocations of functionality are envisioned and may fall within the scope of the embodiment(s). In general, structures and functionality presented as separate components in the example configurations may be implemented as a combined structure or component. Similarly, structures and functionality presented as a single component may be implemented as separate components. These and other variations, modifications, additions, and improvements fall within the scope of the embodiment(s).

[0061] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the embodiments to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles and their practical applications, to thereby enable others skilled in the art to best utilize the embodiments and various embodiments with various modifications as are suited to the particular use contemplated.

What is claimed is:

1. A computer-implemented method for confirming authorship of documents, performed on a server having at least one processor and memory storing at least one program for execution by the at least one processor to perform the method, comprising:

accessing a first document hosted on a first website of a first domain, the first document being linked to a second document through at least one link in a chain of links, a respective link in the chain of links including a first predefined authorship attribute asserting authorship of a respective document including the respective link by a respective entity associated with a respective target document of the respective link; and conditionally confirming authorship of the first document by an entity associated with the second document when the second document includes a second link to the first website of the first domain, the second link including a second predefined authorship attribute indicating that the entity associated with the second document is an author of or contributor to content at the first website of the first domain.

2. The computer-implemented method of claim 1, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document includes:

identifying a first link in the first document that links to the second document, wherein the first link includes the first predefined authorship attribute;

accessing the second document; and

determining that the second document includes the second link including the second predefined authorship attribute.

3. The computer-implemented method of claim 2, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document includes starting with the first document, traversing links in the chain of links until the second document that includes the second link including the second predefined authorship attribute is identified.

4. The computer-implemented method of any of claims 1-3, including determining that the first document has a non-confirmed authorship when no second document having the second link including the second predefined authorship attribute can be found.

5. The computer-implemented method of any of claims 1-4, wherein the second document includes a trusted document associated with the entity.
6. The computer-implemented method of any of claims 1-5, wherein the second document includes a profile of the entity.
7. The computer-implemented method of any of claims 1-6, wherein the second link links to a root of the first website of the first domain.
8. The computer-implemented method of any of claims 1-7, wherein the second document is hosted on a second domain that is distinct from the first domain.
9. The computer-implemented method of any of claims 1-7, wherein the second document is hosted on the first domain.
10. The computer-implemented method of any of claims 1-9, further comprising responsive to confirming authorship of the first document by the entity, adding the first document to the profile for the entity.
11. The computer-implemented method of any of claims 1-10, further comprising responsive to confirming authorship of the first document by the entity, adding the first document to a social networking webpage for the entity.
12. The computer-implemented method of any of claims 1-11, further comprising responsive to confirming authorship of the first document by the entity, adding information relating to the first document to a webpage including information relating to documents that the entity authored.
13. The computer-implemented method of any of claims 1-12, further comprising responsive to confirming authorship of the first document by the entity, associating the first document with the entity in a search index.
14. A system to confirm authorship of documents, comprising:
 - at least one processor;
 - memory; and
 - at least one program stored in the memory and executable by the at least one processor, the at least one program comprising instructions to:

access a first document hosted on a first website of a first domain, the first document being linked to a second document through at least one link in a chain of links, a respective link in the chain of links including a first predefined authorship attribute asserting authorship of a respective document including the respective link by a respective entity associated with a respective target document of the respective link; and

conditionally confirm authorship of the first document by an entity associated with the second document when the second document includes a second link to the first website of the first domain, the second link including a second predefined authorship attribute indicating that the entity associated with the second document is an author of or contributor to content at the first website of the first domain.

15. The system of claim 14, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document, the at least one program includes instructions to:

identify a first link in the first document that links to the second document, wherein the first link includes the first predefined authorship attribute;

access the second document; and

determine that the second document includes the second link including the second predefined authorship attribute.

16. The system of claim 15, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document, the at least one program includes instructions to, starting with the first document, traverse links in the chain of links until the second document that includes the second link including the second predefined authorship attribute is identified.

17. The system of any of claims 14-16, including instructions to determine that the first document has a non-confirmed authorship when no second document having the second link including the second predefined authorship attribute can be found.

18. The system of any of claims 14-17, wherein the second document includes a trusted document associated with the entity.

19. The system of any of claims 14-18, wherein the second document includes a profile of the entity.
20. The system of any of claims 14-19, wherein the second link links to a root of the first website of the first domain.
21. The system of any of claims 14-20, wherein the second document is hosted on a second domain that is distinct from the first domain.
22. The system of any of claims 14-20, wherein the second document is hosted on the first domain.
23. The system of any of claims 14-22, further comprising instructions to add the first document to the profile for the entity responsive to confirming authorship of the first document by the entity
24. The system of any of claims 14-23, further comprising instructions to add the first document to a social networking webpage for the entity responsive to confirming authorship of the first document by the entity.
25. The system of any of claims 14-24, further comprising instructions to add information relating to the first document to a webpage including information relating to documents that the entity authored responsive to confirming authorship of the first document by the entity responsive to confirming authorship of the first document by the entity.
26. The system of any of claims 14-25, further comprising instructions to associate the first document with the entity in a search index responsive to confirming authorship of the first document by the entity.
27. A non-transitory computer readable storage medium storing at least one program configured for execution by at least one processor of a computer system, the at least one program comprising instructions to:
 - access a first document hosted on a first website of a first domain, the first document being linked to a second document through at least one link in a chain of links, a respective link in the chain of links including a first predefined authorship attribute asserting authorship of a respective document including the respective link by a respective entity associated with a respective target document of the respective link; and

conditionally confirm authorship of the first document by an entity associated with the second document when the second document includes a second link to the first website of the first domain, the second link including a second predefined authorship attribute indicating that the entity associated with the second document is an author of or contributor to content at the first website of the first domain.

28. The non-transitory computer readable storage medium of claim 27, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document, the at least one program includes instructions to:

identify a first link in the first document that links to the second document, wherein the first link includes the first predefined authorship attribute;

access the second document; and

determine that the second document includes the second link including the second predefined authorship attribute.

29. The non-transitory computer readable storage medium of claim 28, wherein prior to conditionally confirming the authorship of the first document by the entity associated with the second document, the at least one program includes instructions to, starting with the first document, traverse links in the chain of links until the second document that includes the second link including the second predefined authorship attribute is identified.

30. The non-transitory computer readable storage medium of any of claims 27-29, including instructions to determine that the first document has a non-confirmed authorship when no second document having the second link including the second predefined authorship attribute can be found.

31. The non-transitory computer readable storage medium of any of claims 27-30, wherein the second document includes a trusted document associated with the entity.

32. The non-transitory computer readable storage medium of any of claims 27-31, wherein the second document includes a profile of the entity.

33. The non-transitory computer readable storage medium of any of claims 27-32, wherein the second link links to a root of the first website of the first domain.

34. The non-transitory computer readable storage medium of any of claims 27-33, wherein the second document is hosted on a second domain that is distinct from the first domain.
35. The non-transitory computer readable storage medium of any of claims 27-33, wherein the second document is hosted on the first domain.
36. The non-transitory computer readable storage medium of any of claims 27-35, further comprising instructions to add the first document to the profile for the entity responsive to confirming authorship of the first document by the entity
37. The non-transitory computer readable storage medium of any of claims 27-36, further comprising instructions to add the first document to a social networking webpage for the entity responsive to confirming authorship of the first document by the entity.
38. The non-transitory computer readable storage medium of any of claims 27-37, further comprising instructions to add information relating to the first document to a webpage including information relating to documents that the entity authored responsive to confirming authorship of the first document by the entity responsive to confirming authorship of the first document by the entity.
39. The non-transitory computer readable storage medium of any of claims 27-38, further comprising instructions to associate the first document with the entity in a search index responsive to confirming authorship of the first document by the entity.
40. A non-transitory computer readable storage medium storing at least one program configured for execution by at least one processor of a system, the at least one program comprising instructions to be executed by the at least one processor so as to perform the method of any of claims 1-13.
41. A system, comprising:
at least one processor; and
memory storing at least one program for execution by the at least one processor;
the system including means for performing the method of any of claims 1-13.