(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
21 October 2010 (21.10.2010)

PCT

(10) **International Publication Number**
WO 2010/120350 A1

(51) **International Patent Classification**:
G06F 15/16 (2006.01)

(21) **International Application Number**:
PCT/US2010/001068

(22) **International Filing Date**:
9 April 2010 (09.04.2010)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
12/386,362 15 April 2009 (15.04.2009) US

(71) **Applicant** *(for all designated States except US)*: **ATTRIBUTOR CORPORATION** [US/US]; 1775 Woodside Road, Suite 100, Redwood City, CA 94061 (US).

(72) **Inventors**: **PITKOW, James, E.**; 1775 Woodside Road, Suite 100, Redwood City, CA 94061 (US). **DIKLIC, Dejan**; 1775 Woodside Road, Suite 100, Redwood City, CA 94061 (US). **MONGA, Rajat**; 1775 Woodside Road, Suite 100, Redwood City, CA 94061 (US).

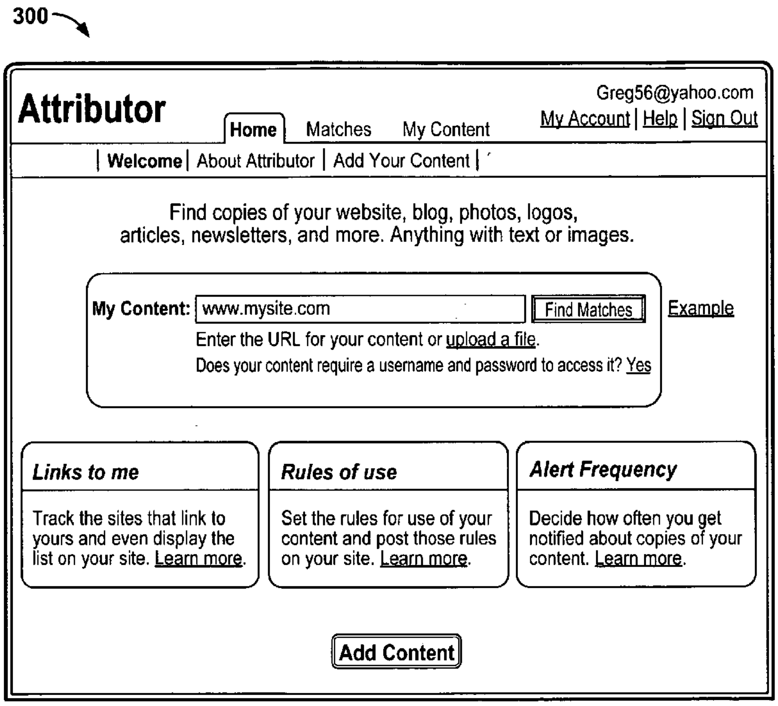
(74) **Agent**: **HUANG, Clover**; Van Pelt & Yi LLP, 10050 N. Foothill Blvd., Suite 200, Cupertino, CA 95014 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,

[Continued on next page]

(54) Title: MANAGING CONTROLLED CONTENT ON A WEB PAGE HAVING REVENUE-GENERATING CODE



Find copies of your website, blog, photos, logos, articles, newsletters, and more. Anything with text or images.

My Content: www.mysite.com

Enter the URL for your content or upload a file.

Does your content require a username and password to access it? Yes

Track the sites that link to yours and even display the list on your site. Learn more.

Set the rules for use of your content and post those rules on your site. Learn more.

Decide how often you get notified about copies of your content. Learn more.

FIG. 3

(57) **Abstract**: Managing controlled content on a web page is disclosed. A web page is analyzed. Controlled content and revenue-generating code are detected on the web page. A party to contact is determined based on the revenue-generating code or the controlled content.

WO 2010/120350 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

MANAGING CONTROLLED CONTENT ON A WEB PAGE HAVING REVENUE-GENERATING CODE

BACKGROUND OF THE INVENTION

[0001] Content, such as text, images, and video, may be stored and displayed on the Internet. Content owners may desire to permit the widest distribution and re-use of their content across web pages and applications, but also want to retain revenue, promotional and other benefits as their content is used and republished by others. Preserving these benefits may be impractical if it requires direct, on-going relationships between large numbers of disparate content users and individual content owners and their customized content use conditions. Similarly, parties that host content for users may bear responsibility when those users make improper use of content owned by third parties, but have no means to ensure that content use conditions are met without dealing individually with content rights holders. These are some examples of various issues that have arisen regarding ownership, use of content, and monetization on the Internet. Solutions to such problems would be desirable.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Various embodiments of the invention are disclosed in the following detailed description and the accompanying drawings.

[0003] Figure 1 is a block diagram illustrating an embodiment of a content monitoring system.

[0004] Figure 2A is a flow chart illustrating an embodiment of a process for monitoring content.

[0005] Figure 2B is a flow chart illustrating an embodiment of a process for monitoring for use of controlled content.

[0006] Figure 2C is a flow chart illustrating an embodiment of a process for evaluating context of a content object.

[0007] Figure 2D is a flow chart illustrating an embodiment of a process for monitoring for use of controlled content.

[0008] Figure 2E is a flow chart illustrating an embodiment of a process for engaging with a user of non-compliant content.

[0009] Figure 2F is a flow chart illustrating an embodiment of a process for displaying compliance information.

[0010] Figure 3 is an example of a graphical user interface (GUI) for providing controlled content.

[0011] Figure 4A is an example of a GUI for providing controlled content.

[0012] Figure 4B is an example of a GUI for providing usage rules.

[0013] Figure 5 is an example of a GUI for displaying search results.

[0014] Figure 6 is an example of a GUI for displaying use of a content object.

[0015] Figure 7 is a block diagram illustrating an embodiment of a system for handling controlled content on a web page having revenue-generating code (RGC).

[0016] Figure 8 is a flow chart illustrating an embodiment of a process for handling controlled content on a web page having revenue-generating code.

[0017] Figure 9 is a flow chart illustrating an embodiment of a process for notifying a party of royalty information.

[0018] Figure 10 is a flow chart illustrating an embodiment of a process for handling receipt of royalty information.

[0019] Figure 11 is a flow chart illustrating an embodiment of a process for notifying a party associated with revenue-generating code of information associated with controlled content.

[0020] Figure 12 is a flow chart illustrating an embodiment of a process for handling receipt of information associated with controlled content.

DETAILED DESCRIPTION

[0021] The invention can be implemented in numerous ways, including as a process; an apparatus; a system; a composition of matter; a computer program product embodied on a computer readable storage medium; and/or a processor, such as a processor configured to execute instructions stored on and/or provided by a memory coupled to the processor. In this specification, these implementations, or any other form that the invention may take, may be referred to as techniques. In general, the order of the steps of disclosed processes may be altered within the scope of the invention. Unless stated otherwise, a component such as a processor or a memory described as being configured to perform a task may be implemented as a general component that is temporarily configured to perform the task at a given time or a specific component that is manufactured to perform the task. As used herein, the term 'processor' refers to one or more devices, circuits, and/or processing cores configured to process data, such as computer program instructions.

[0022] A detailed description of one or more embodiments of the invention is provided below along with accompanying figures that illustrate the principles of the invention. The invention is described in connection with such embodiments, but the invention is not limited to any embodiment. The scope of the invention is limited only by the claims and the invention encompasses numerous alternatives, modifications and equivalents. Numerous specific details are set forth in the following description in order to provide a thorough understanding of the invention. These details are provided for the purpose of example and the invention may be practiced according to the claims without some or all of these specific details. For the purpose of clarity, technical material that is known in the technical fields related to the invention has not been described in detail so that the invention is not unnecessarily obscured.

[0023] Figure 1 is a block diagram illustrating an embodiment of a content monitoring system. In some embodiments, content monitoring system 100 is used by a content owner to monitor for non-compliant use of the content owner's content based on usage rules specified by the content owner. Examples of content owners include: a photographer (e.g., Ansel Adams), a film studio (e.g., Disney), or a columnist (e.g., Walter Mossberg), or a media outlet (e.g., The Wall Street Journal). The content owner is not necessarily the same as the content creator. Usage rules (including usage policies, terms of use, usage terms, etc.) are a set of rules regarding conditions under which content may be used, as specified by the content owner. Usage rules may vary depending on the content and/or the content owner and applicable law (such as "fair use"). Usage rules are more fully described below.

[0024] In some embodiments, content monitoring system 100 is used by a content host to monitor for non-compliant use of content based on a host policy specified by the content host. A content host refers to an entity that hosts, serves, stores, provides, and/or displays content. Examples of content hosts include OSPs, such as search engines (e.g., Google), photo or video sharing websites (e.g., YouTube, Yahoo), and blogging sites (e.g., TypePad). As used herein, an OSP is an entity that hosts and/or serves or provides content on behalf of itself or other entities. For example, an OSP includes an OSP as defined under DMCA. An OSP includes an electronic content management system (ECM). A host policy is a set of rules regarding conditions under which content may be hosted, as specified by a content host. A host policy may vary depending on the content host. As an example of a host policy, OSPs may have policies that apply to the posting of content by their users, in which they reserve the right to remove content or users in the event of non-compliance (determined at their discretion). In some embodiments, a configurable host policy governs the automatic handling of DMCA notices, as more fully described below.

[0025] A content user includes an entity that uses content that is not owned by the content user. A content user includes an entity that owns or posts content. Examples of content users include writers, photographers, bloggers, or any user who posts content on content hosts.

[0026] Controlled content refers to content associated with one or more compliance rules, where compliance rules include usage rules specified by a content owner and/or host policy rules specified by a content host. In the case where a content owner is monitoring for use of his content, controlled content is the content owner's content. In the case where a content host is monitoring for non-compliant content, controlled content is content that is non-compliant with the host policy. Monitored content refers to the set of content being searched (i.e., potential matches). In other words, content monitoring system 100 searches monitored content for use of controlled content. As used herein, a match, copy, or use of controlled content does not necessarily refer to an identical match, an identical copy, or use of identical content. A match, copy, or use of controlled content is identified based on criteria such as similarity scores and non-compliance scores, as more fully described below.

[0027] Compliant content refers to content that satisfies usage rules associated with the content. In the case where a content host such as an OSP is monitoring for non-compliant content, compliant content refers to content that not only satisfies the usage rules, but also satisfies the host policy of the content host (e.g., the OSP).

[0028] As used herein, a unit of content may be referred to as a content object. Content objects can include any object type. Examples of content objects include a text document, an image, video, audio, flash, animation, game, lyrics, code, or portions thereof (e.g., a phrase/sentence/paragraph, a subimage, or a video clip). Other examples include a single file (e.g., an image), all of the text on a web page (e.g., a news article), a chapter in a book, and a blog entry. The content object may be in various audio, image, or video formats, such as MP3, JPEG, MPEG, etc.

[0029] Content monitoring system 100 can be used to find copies of a set of content at a given point in time or regularly monitor for matches. Content monitoring system 100 may be used to monitor data associated with the Internet or any other appropriate environment in which there is a need to monitor content for compliance. Examples of appropriate environments include the Internet, an Intranet, a firewalled network, a private network, an Electronic Data Interchange (EDI) network, an ad hoc network, etc.

[0030] As shown, user 102 provides input to ingestor 104. Ingestor 104 provides input to subscriber database 105, content database 108, and crawler 112. Reporter 110 receives input from subscriber database 105 and content database 108. Crawler 112 provides input to digester 114. Digester 114 provides input to content database 108, controlled content store 116, and monitored content store 118. Matching engine 120 provides input to controlled content store 116 and monitored content store 118. Content database 108 interacts with matching engine 120.

[0031] Content ingestor 104 accepts controlled content from user 102. User 102 includes content owners or administrators of content monitoring system 100. The content may be specified in various ways. A user interface (UI) may be provided for user 102 to specify content. In some embodiments, the UI provides an interface for uploading content or specifying a link/set of links to the content, where the links may be local (e.g., on a local hard drive) or remote (e.g., on a remote server or on the Internet). An example of a remote link is a user's eBay account. User 102 may display, in his eBay store, images to be monitored. For example, user 102 is a photographer selling his photography. Using the UI, user 102 specifies a URL to the eBay store or particular auction. In some embodiments, instead of providing a URL to a particular auction, the content owner provides their username (such as an eBay seller ID), which allows the system to retrieve all of the user-posted content associated with that username, which could be associated with one or more auctions. In some embodiments, the content owner also provides a password if necessary or expedient to locate user-posted content. In some embodiments, a schedule for fetching content may be specified. For example, crawler 112 may be configured to fetch images from the user's eBay store every 24 hours. The raw content is passed to digester 114 for processing and storage.

[0032] In some embodiments, the ingesting of content is automatically triggered by content creation. For example, when a blogger posts a new entry, it is automatically ingested. When a writer updates a Word document, the content is automatically ingested.

[0033] In some embodiments, if the URL or username provided by the content owner contains some content of third parties, the user is presented with a means to exclude or include specific content objects (such as a single image) from monitoring and from the content owner's usage rules.

[0034] The controlled content may be from the Internet or from another source. A manual or automated API may be used to ingest content or perform any of the other processes described herein. A URL or any other appropriate identifier may be used to specify content. Credentials associated with accessing the content, such as a password, may be provided.

[0035] Besides controlled content, other data may be provided as input to content monitoring system 100, such as links (e.g., URLs or websites) identified by an administrator, content host, or content owner. These sites may have been identified because the user is aware of a specific instance of non-compliance at that location, they have historically posted non-compliant content or are of particular concern to the user. Other examples of additional data that may be input to content monitoring system 100 are more fully described below.

[0036] Crawler 112 fetches content from the network. The content to be fetched may include the Internet, a subset of the Internet, a complete domain, or a single piece of content from the web. Identifiers may be used to identify the content to be fetched. Some examples of identifiers include: a URL, a directory, a password protected website(s), all items for a seller on eBay, and all content of a given type or format (e.g., images only or JPEGs only). In some embodiments, crawler 112 is used with modules that provide different rules for crawling. In some embodiments, crawler 112 fetches content according to a specified schedule.

[0037] Controlled content store 116 includes controlled content. In some embodiments, controlled content store 116 includes the following information: a copy of the content, an index of fingerprints associated with the content, and metadata about the content (e.g., filename, URL, fetch date, etc.). In some embodiments, the copy of the content is stored in a separate cache. A fingerprint includes a signature of an object that can be used to detect a copy of an object as a whole or in part. A content object may have more than one fingerprint. A fingerprint may be associated with more than one content object. A fingerprint may be associated with a whole or part of a content object. A fingerprint may be multidimensional. For example, there may be multiple features associated with a fingerprint. A fingerprint may contain multiple fingerprints or subfingerprints.

[0038] Monitored content store 118 is a repository for crawled data. Monitored content store 118 may include any digital object collection or environment. In some embodiments, monitored content store 118 is a web store. In some embodiments, there are multiple content stores, e.g., one for each kind of data – text, images, audio, video, etc. In some embodiments, monitored content store 118 includes data from sites that copy the most often, and is updated most frequently. This data may be indicated as such (i.e., tagged or flagged as common copier) or stored separately. In some embodiments, a real-time store (not shown) is used to store various feeds coming in (e.g., from a content owner's blog each time the blog is updated, or from a content owner's eBay store every 24 hours). In some embodiments, a ping server or similar server is used to update feeds coming in. If the feeds contain links, the content is fetched by crawler 112. Over time, data moves from the real-time store to monitored content store 118 as it becomes older. Monitored content store 118 changes periodically, whereas the real-time store keeps changing as content comes in. In some embodiments, external stores (not shown), such as search engines, are accessed using application programming interfaces (APIs). Once data is fetched, they are stored in monitored content store 118. Some embodiments of this are more fully described below. In some embodiments, fingerprints of content are stored in monitored content store 118. In some embodiments, Gigablast is used to fetch and store content data.

[0039] Digester 114 receives content fetched by crawler 112, including controlled content or monitored content, analyzes, and processes it. Analysis of content is more fully described below. The content and associated metadata is stored in controlled content store 116 or monitored content store 118, as described above.

[0040] In some embodiments, matching engine 120 finds matches to controlled content by comparing controlled content from controlled content store 116 with monitored content from monitored content store 118 based on matching techniques including technical factors, compliance factors, and other factors, as more fully detailed below.

[0041] Reporter 110 reports match results to user 102 or an administrator of content monitoring system 100. Various user interfaces may be used. Examples of reporting and UIs for reporting results are more fully described below.

[0042] Subscriber database 106 contains information about customers. Content database 108 contains references to controlled content and to matched content corresponding to the controlled content. In some embodiments, a separate database is used for matched content.

[0043] In some embodiments, content monitoring system 100 is used as a content clearinghouse by content users wishing to use content. Before using a particular content object (i.e., unit of content), the content user checks with content monitoring system 100 to determine whether the conditions under which the content user wishes to the use the content complies with the usage policy set by the content owner.

[0044] Content monitoring system 100 may be implemented in various ways in various embodiments. For example, controlled content, web data, subscriber data, and/or content data may be organized and stored in one or more databases. Ingesting, crawling, digesting, matching, and/or reporting may be performed using one or more processing engines.

[0045] In some embodiments, any of the functions provided by content monitoring system 100, such as ingesting, crawling, digesting, matching, and reporting, may be provided as a web service. For example, content monitoring system 100 or an element of content monitoring system 100 is queried and provides information via XML.

[0046] Figure 2A is a flow chart illustrating an embodiment of a process for monitoring content. In some embodiments, this process is performed when a content owner is searching or monitoring for non-compliant use of the his controlled content. In some embodiments, this process is performed by content monitoring system 100.

[0047] In the example shown, the process begins at 202, and controlled content is specified. Controlled content may include text, images, video, or any other type of data. Controlled content may be specified in various ways, such as content located in a particular directory and/or all content contributed by a particular user (e.g., on eBay). A user (e.g., a content owner or an administrator) may specify controlled content using any appropriate interface. Examples of graphical user interfaces are described more fully below. The user may also request a one time search or regular monitoring for the controlled content. In the case of the latter, the user may specify options related to regular monitoring, such as frequency of monitoring, how often reports should be received, etc.

[0048] At 203, usage rules are specified. Usage rules include conditions under which a content owner permits the use of owned content. Usage rules may include terms under which a content owner permits the republication and/or modification of content. Usage rules may include different conditions depending on whether the use is for commercial or non-commercial uses, business or education uses, with or without attribution, in a limited amount, in a limited context, etc. The usage rules may be based on any appropriate compliance structure, such as "fair use," "copy left," "share alike," Creative Commons specified structures, user specific compliance rules, rules against associating the controlled content with objectionable content (e.g., obscenity, adult content, child pornography), rules requiring attribution, moral rights, rights of personality, or any legal or personal compliance structure. A usage rule may take into account editorial context. In other words, certain uses may be permitted that are not permitted in another context. For example, if the controlled content object is a book, portions from the book may be permitted to be used in a book review but not in another context (where other rules may apply).

[0049] A variety of user interfaces may be used to specify usage rules. For example, a list of terms, checkboxes (to apply a rule), and settings (specific to a rule) may be provided. The list may include, for example: whether attribution is required, amount of duplication allowed, whether commercial use is allowed, whether changes are allowed, whether permission is required, whether derivative content is allowed, geographical requirements, whether the owner requires advertisement revenue sharing (e.g., using Google AdSense) and associated terms and information, etc. The usage rules may be hierarchical. For example, a list of higher level rules or compliance structures may be displayed for selection, each of which may be expanded to display lower level rules that each of the high level rules comprises. Usage rules may have any number of levels. Checkboxes (or another appropriate object) may be located next to the higher level or lower level rules and may be selected (e.g., checked off) at any level of granularity. For example, selecting checkboxes next to a higher level rule automatically selectes all corresponding lower level rules. Alternatively, lower level rules may be individually selected. An example of a higher level rule is a particular type of license. Lower level rules under the license include the specific usage rules associated with the license.

[0050] Usage rules may be customized for each content owner (and for each content object). In some embodiments, a unique URL is provided to the content owner for his use (e.g., to include as a link associated with an icon placed in proximity to his content on his website, in his eBay store, etc.) When a content user wishes to use content on the content owner's website, the content user can then select the link, which leads to a page describing the content owner's usage rules (for that content object).

[0051] In some embodiments, rather than providing a unique URL to the content owner, the content owner could use a particular URL on his website or web page. For example, the particular URL could be “rules.attributor.com.” When a content user wishes to use content on the content owner's website, the content user can select the link, which leads to a page describing the content owner's usage rules (for the website or content on the website). In this case, the content monitoring system determines from which website the link was selected and can determine which usage rules to display. In some embodiments, the same URL is common to multiple content owner's websites. Further examples are discussed below.

[0052] Usage rules may be stored in the content monitoring system. For example, the usage rules for content owners may be stored in controlled content store 116 (e.g., as metadata associated with the content object) or in subscriber database 106.

[0053] At 204, controlled content is acquired. In some embodiments, 204 is performed by ingestor 104 in system 100. In various embodiments, controlled content is obtained from a source specified at 202. For example, controlled content is obtained from a particular directory or from one or more servers containing content contributed by a particular user. Controlled content acquisition may be automated or non-automated. For example, an automated process could poll for updates and acquire controlled content when an update is detected. In some embodiments, a ping server is used to detect updates. In some embodiments, controlled content is continuously acquired or ingested. For example, if the controlled content is specified as all content contributed by a particular user on eBay, then when the user contributes new content to eBay, that content is automatically acquired or acquired at configured times or time intervals. A variety of APIs may be used to acquire controlled content. In some embodiments, after controlled content is acquired, the user is given an opportunity to confirm that it is the correct controlled content or the controlled content the user intended. The acquisition of controlled content may involve any network, protocol (e.g., UDP, TCP/IP), firewall, etc.

[0054] At 206, controlled content is analyzed. In some embodiments, 206 is performed by digester 114 in system 100. In some embodiments, the acquired content is analyzed for unique identifying features. Any appropriate technique may be used to extract features from the content. For example, a fingerprint associated with the content may be determined. The technique may depend on the media type (e.g., spectral analysis for audio/video, histogram or wavelets for images/video, etc.) For example, in the case of text content, various techniques may be used, such as unique phrase extraction, word histograms, text fingerprinting, etc. An example is described in T. Hoad and J. Zobel, “Methods for identifying versioned and plagiarized documents,” in Journal of the American Society for Information Science and Technology, Volume 54, Issue 3, 2003. In the case of image content, various techniques may be used, including key point identification, color histograms, texture extraction, image signatures, or extraction of any other feature. An example is described in Y. Ke, R.Sukthankar, and L. Houston, “Efficient near-duplicate detection and sub-image retrieval,” in ACM Multimedia. ACM, Oct. 2004, pp. 1150–1157. In the case of video content, a video fingerprinting technique may be used. In another example, a signature is formed for each clip by selecting a small number of its frames that are most similar to a set of random seed images, as further described in S.-C. Cheung, A. Zakhor, "Efficient Video Similarity Measurement with Video Signature," Submitted to IEEE Trans. on CSVT, Jan., 2002. In the case of audio content, an audio fingerprinting technology may be used. For example, a spectral signature is obtained and used as input to a hash function. In various embodiments, other techniques may be used. Analyzing may include determining spectral data, wavelet, key point identification, or feature extraction associated with the controlled content. In some embodiments, results from the analysis are stored in controlled content store 116 in system 100.

[0055] At 208, monitored content is searched for use of controlled content. In some embodiments, monitored content is specified by a user, such as a content owner or administrator. The entire web may be searched, or a subset of the web (e.g., websites that have been identified as sites that copy the most often or data in a content store such as monitored content store 118). A database of sites that have been crawled and resulting data may be maintained that is updated at various times. Rather than searching the entire web, the database may be used instead. Searching may comprise a combination of searching the web and consulting a database of previously crawled websites. In some embodiments, monitored content store 118 in system 100 stores previously crawled websites. In some embodiments, 208 is performed by crawler 112 in system 100.

[0056] Searching may be performed in one or more stages, each stage refining the search further. For example, a first search may yield a first set of candidate content objects. A second search searches the first set of candidate content objects to yield a second set of content objects, and so forth. Eventually, the final set of content object(s) includes the content object(s) that match or most closely match the controlled content object. In some embodiments, less expensive and/or less complex techniques may be used to obtain candidate sets followed by one or more tighter, smaller granularity techniques to progressively enhance the resolution of the analysis. Which techniques may be used and in which order may be determined based on cost and/or complexity. In some embodiments, the second search comprises a manual search. For example, the second set of content objects may be a smaller set and may be searched by a human.

[0057] In some embodiments, a hash structure is used to obtain a candidate set of content objects. For example, a hash table is maintained such that similar content objects are hashed to the same or a nearby location in a hash table. This way, to search for content object A, a hash function associated with A is computed and looked up in a hash table, and a set of objects that are similar to A is obtained. A hash function associated with a content object may be computed in various ways. The hash function may be computed differently depending on the type of content object or one or more characteristics of the content object. For example, if the content object is a text document, a fingerprinting technique specific to text may be used to obtain a fingerprint of the document. The fingerprint may be input to the hash function to obtain a hash value that corresponds to a group of other content objects that have a similar fingerprint. Hash values that are nearby in the hash table correspond to content objects that have similar (though less similar than those in the same hash bin) fingerprints, to create a clustering effect. In this way, a candidate set of content objects may be obtained.

[0058] Other techniques such as cosine similarity, latent semantic indexing, keyword based methods, etc., may also be used.

[0059] In some embodiments, existing search engines or search facilities on websites, such as eBay, are used to obtain a candidate set of documents. This approach may be useful in an initial implementation of the system. For example, APIs provided by Google or other search engines may be used to perform this search. For example, to search for a document, a unique phrase within the document is selected. The unique phrase is input to a Google search using a Google API and the results are a candidate set of documents. Multimedia search engines (e.g., video, image) may be used to obtain a candidate set of documents. In the case of images, an image search engine may be used to obtain a candidate set of images. For example, Riya (www.Riya.com) includes an image search engine that may be used to obtain a candidate set.

[0060] In some embodiments, besides the Internet, databases may be searched using these techniques. Some examples of databases include Factiva, Corbis, and Hoover's. Although these databases do not allow indexing of their documents, they do have a search interface. This search interface may be used to perform searches for content using unique phrase extraction. For example, articles in the Factiva database containing a unique phrase from a controlled content object are more likely to be a match. A subsequent search may be performed by obtaining the full text of the articles and searching them using more refined techniques. Searching this way limits having to crawl the entire Internet. Also the more computationally intensive search techniques are limited to a smaller search space.

[0061] In some embodiments, once a candidate set of content objects is obtained, one or more refining searches are performed. For example, the candidate set of documents are crawled and advanced matching techniques can be applied to the candidate set of documents. A variety of content or document similarity techniques may be used. For example, the techniques described at 206 may be used on the candidate set of content objects.

[0062] In the case of text documents, a refining search may comprise computing a signature for each paragraph or other data set. A Levinstein distance could be used to determine the similarity between a document and the controlled content object. A byte by byte comparison could be used. Other techniques, such as anchoring or cosine similarity may be used, as described more fully in T. Hoad and J. Zobel, "Methods for identifying versioned and plagiarized documents," in Journal of the American Society for Information Science and Technology, Volume 54, Issue 3, 2003. Techniques such as PCA-sift or feature extraction of color, texture and signature generation may be used. For example, A.C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions, Technical Report, TR2004-515, Dartmouth College, Computer Science describes examples of such techniques.

[0063] In the case of images, images may be subsampled to be robust against cropping and subimage reuse using techniques such as key pointing (or key point extraction), which looks for unique signatures within a portion of an image, such as edges or extreme color gradations, and samples these portions to obtain a signature. Another way is to subsample distinctive portions of a color histogram of the image.

[0064] In some embodiments, different techniques are used depending on characteristics of the content object. For example, if a document has fewer than 20 paragraphs, a byte by byte comparison may be used. If a document has 20 or more paragraphs, a different technique may be used. Sampling and anchoring points may depend on the format of the document.

[0065] At 210, use of controlled content is detected. In some embodiments, 210-213 are performed by matching engine 110 in system 100. In some embodiments, detection is based on various criteria associated with technical factors that may result from searching at 208. An example of a technical factor is a similarity score. A similarity score is a measure of the similarity between two content objects and may be computed in a variety of ways. For example, the Levinstein distance is a similarity score. In some embodiments, if similarity scores meet one or more criteria, use of controlled content is detected. The criteria may be configurable by the user or administrator. One or more similarity scores may be computed for a controlled object and candidate object to represent various characteristics of the content. In some embodiments, one or more similarity scores may be weighted and combined into a single similarity score.

[0066] A similarity score may account for various degrees of copying. For example, the first and last paragraph of a document may be copied, a portion of a document may be copied, or the whole document may be copied. Different samples of music may be copied into a single audio file. Videos may be mixed from copied videos. One controlled document may have 15 samples, one or more of which may be copied. A similarity score may account for these factors. For example, a copying extent score may be used to indicate the percentage of a controlled content object that has been copied. A copying density score may be used to indicate the percentage of a match that is comprised of a controlled content object.

[0067] At 212, a context associated with the use of the controlled content is evaluated. The context refers to any attribute associated with the use of the content object. For example, the context includes compliance factors, technical factors, and reputation information. Context may be automatically and/or manually determined.

[0068] Compliance factors are based on usage rules specified by content owners. For example, compliance factors include information related to attribution and commercial context. Examples of compliance factors include whether the site is government, education, commercial, revenue producing, subscription based, advertising supported, or produces revenue in some other way (e.g., using a reputation bartering scheme associated with a compensation mechanism). This can be determined manually or automatically. For example, a human could review the website, or based on the top level domain (e.g., .edu, .com, .org), or the presence of advertising related HTML code, it can be determined whether the website is commercial.

[0069] In some embodiments, a non-compliance score is computed to represent the likelihood that a content object is non-compliant based on the compliance factors. In some embodiments, multiple compliance factors are used to determine a non-compliance score. For example, the non-compliance score takes multiple compliance factors, normalizes and weighs each one as appropriate, and takes the sum. In some embodiments, the weighting is based on usage rules and/or host policy rules. In addition an overall weight may be used to scale the non-compliance score. For example, content found on educational sites may be weighted differently. One or more non-compliance scores may be computed.

[0070] Besides technical factors and compliance factors, examples of other factors include reputation information. For example, a reputation database is maintained that includes reputation ratings of content users by other content owners. For example, Bob's blog may have a low reputation because it has posted numerous copyrighted content objects owned by others who have given Bob's blog a low reputation rating.

[0071] At 213, matching content (i.e., match content object(s)) is identified based on detection at 210 and/or evaluation at 212. As previously described, a match, copy, or use of controlled content does not necessarily refer to an identical match, an identical copy, or use of identical content.

[0072] In some embodiments, a match is a technical match and is selected based only on technical factors, such as similarity scores. In this case, technical matches are identified at 210, and at 212, the technical matches are evaluated based on context to determine whether they are compliant.

[0073] In other embodiments, a match is selected based on configurable criteria associated with technical factors (e.g., similarity scores), compliance factors (e.g., non-compliance scores), and/or other factors (e.g., reputation information). In some embodiments, it is determined that content objects with one or more similarity scores that exceed a similarity score threshold and one or more non-compliance scores that exceed a non-compliance score threshold are matches. In other words, a content object that is technically similar, but is compliant with applicable usage rules, would not be considered a match. In some embodiments, it is determined that any content object with one or more similarity scores that exceed a similarity score threshold is a match.

[0074] In some embodiments, a binary flagging is used. For example, it is determined that content objects with one or more similarity scores that exceed a similarity score threshold and/or one or more non-compliance scores that exceed a non-compliance score threshold are "interesting" and other content objects are "non-interesting." In some embodiments, "interesting" content objects are reported to the user at 214.

[0075] At 214, content is reported to the user (e.g., content owner). In some embodiments, which content to report is configurable and may depend on criteria based on technical factors (e.g., similarity scores), compliance factors (e.g., non-compliance scores), and/or other factors (e.g., reputation information). In some embodiments, matching content as identified at 213 is reported to the user. In some embodiments, a user views and manually confirms whether each matching content object is non-compliant. The results may be stored in a common database.

[0076] In some embodiments, 214 is performed by reporter 110 in system 100. Various interfaces could be used. Screenshots, links, buttons, tabs, etc. may be organized in any appropriate fashion. In some embodiments, a user interface is presented to the user that shows the matching content, one or more similarity scores, and one or more non-compliance scores. Example interfaces for reporting results are more fully described below.

[0077] In some embodiments, the interface provides a way for the user to confirm that content is the user's content or reject the content (i.e., indicate a false positive). This data may be fed back into the monitoring process. For example, this information may be stored in a database or with the content metadata. In some embodiments, the interface provides choices of actions for the user to select from (e.g., ask that the reusing party attributes it, offer license/licensing terms, remove under DMCA, etc.).

[0078] In some embodiments, 214 is not performed and the process continues at 216.

[0079] At 216, the user of the content is engaged. In some embodiments, user contact information is obtained from the IP address, the U.S. Copyright Office (e.g., a designated agent registered with the U.S. Copyright Office), or a known email address (e.g., of an OSP or a user of an OSP). A database or lookup table of contact information associated with various sites may be maintained and used to determine user contact information.

[0080] Depending on configuration settings, various types of communication may be sent to the content user. For example, a DMCA notice, information concerning usage rules, licensing information, etc. may be sent. For example, the content owner may have specified one or more usage rules associated with his content, such as "do not license any content," "replace content with an advertisement," "add watermark to content," "add Unicode overlay," "share advertisement revenue," or "ask permission prior to use." Based on the usage rules, an appropriate communication may be sent to the content user. In some embodiments, the content user is also configured to use the content monitoring system. The content user may have specified a set of compliance rules, such as "automatically debit my account up to $100 per year when licensed content is used," "offer to share advertising revenue when contacted by content owner," "remove content when contacted by content owner," etc. Based on the compliance rules, an appropriate response may be sent back to the content owner. In some embodiments, an engagement communication may be configured to be sent in a way that preserves the anonymity of the sender of the engagement communication (e.g., the content owner, or a content host, as more fully described below).

[0081] An example of an engagement communication includes an email that is automatically sent to a content user notifying the user that the content is owned and offering to license it for $9.99 per year, and including a link to the content owner's usage rules hosted by the content monitoring system. The content owner may configure his settings so that the email is not sent to content users whose sites are educational or non-profit or those settings may be default settings if the content owner's usage rules indicate free use by educational or non-profit sites. In response, the content user sends a response agreeing to the terms. The response may be created and/or sent automatically because the content user's compliance rules indicate the following rule: "automatically debit my account up to $100 per year when licensed content is used." The response may be sent manually, or the user may approve an automatically created response before it is sent.

[0082] In some embodiments, a series of communications may occur between the content user and content owner. On the content user and/or the content owner's side, the responses may be automatic. In this way, licensing terms can be negotiated and/or steps can be taken towards resolution.

[0083] In some embodiments, compensation is not necessarily monetary. For example, the content owner may just want to receive attribution, license revenue or advertising revenue sharing may be donated to charitable or other causes as directed by the content owner or may be treated as a credit towards a trade (e.g., if you use my content, I can use your content), or the content owner may require that the content and derivative works be presented in a manner that enables tracking of the number of uses or views of the content, or that derivative works must be available for use by others under specified usage rules.

[0084] In some embodiments, whenever new controlled content is provided, processes 202-206 are performed. In some embodiments, every prespecified search interval, processes 208-213 are performed. In some embodiments, every prespecified report interval, 214 is performed. For example, an email may be sent to the user indicating that new matches have been found, and a link to the web interface provided in the email message. In some embodiments, 214 is performed each time a user logs into the content monitoring system. In some embodiments, 208-213 are performed when a user logs into the content monitoring system, either automatically, or after a user selects an "update results" or "search" button upon logging in.

[0085] In some embodiments, the number of accesses to a controlled content object is tracked. For example, the content is associated with a web beacon or other element of code that enables the tracking of accesses of the content for purposes such as calculation of license fees or revenue sharing.

[0086] Figure 2B is a flow chart illustrating an embodiment of a process for monitoring for use of controlled content. In some embodiments, this process is performed when a content host, such as an OSP, is searching or monitoring for non-compliant use of content based on a host policy of the content host. Thus, the controlled content in this case is non-compliant content based on a host policy. In some embodiments, this process is performed by content monitoring system 100.

[0087] At 230, a host policy is specified. For example, an OSP may have a policy regarding what comprises non-compliant content. Non-compliant content may include material that violates third party copyrights or trademarks, is illegal (e.g., child pornography) or does not comply with an OSP's terms of service (e.g., adult content, pornography, obscenity). A host policy may include host rules that may be associated with any compliance structure, such as host specific compliance rules, rules against objectionable content (e.g., obscenity, adult content, child pornography), or any legal or personal compliance structure. A host policy may specify that content must comply with usage rules specified by the content owner, such as "copy left," "share alike," Creative Commons specified structures, etc.

[0088] A variety of user interfaces may be used to specify a host policy. For example, any of the user interfaces described at 203 for specifying usage rules may be used to specify a host policy. For example, a list of terms, checkboxes (to apply a rule), and settings (specific to a rule) may be provided. The list may include, for example: whether pornography is allowed, whether profanity is allowed, whether to comply with one or more usage rules, whether to comply with copyright or other legal structures, etc. The rules may be hierarchical. For example, a list of higher level rules or compliance structures may be displayed for selection, each of which may be expanded to display lower level rules that each of the high level rules comprises. Rules may have any number of levels. Checkboxes (or another appropriate object) may be located next to the higher level or lower level rules and may be selected (e.g., checked off) at any level of granularity.

[0089] At 232, content is monitored for use of controlled content. In this case, the monitored content comprises the content hosted by the content host (e.g., the content served by the OSP). In some embodiments, monitoring comprises checking each content object before it is hosted (or served) by the OSP. For example, an OSP such as youtube.com may check each video before it is made available for viewing on youtube.com. In some embodiments, monitoring comprises periodically checking content objects served by the OSP. For example, a new video is made available for viewing immediately after being posted, but the video may later be removed by a monitoring process that checks new content objects. If the video is determined to be non-compliant, it is removed and the video owner is optionally notified. The results of the check are stored in a database so that the video does not need to be checked again unless it is modified.

[0090] In some embodiments, if information obtained from the database is not enough to determine whether the content is compliant, an evaluation is performed, where the evaluation can include techniques described at 212. The evaluation may also include techniques used to detect objects or characteristics of objects in an image, such as faces, body parts, the age of a person being depicted, etc. Such techniques may be useful to detect pornography or child pornography, for example. The evaluation results may then be stored in the database.

[0091] Examples of monitoring are more fully described below with respect to Figure 2D.

[0092] In some embodiments, a common pool of objectionable content is maintained based on input from multiple content hosts. For example, the common pool may include content that has been identified by various content hosts as containing pornography, child pornography, profanity, or racial content. Depending on the compliance rules specified in their host policies, an OSP may have an interest in contributing to, sharing, and using the common pool to identify objectionable content and remove or reject it.

[0093] For example, an OSP such as eBay may desire to monitor content posted by its users. An eBay employee manually performs simple filtering for adult content. Each time the eBay employee flags an object as "adult content," that object is acquired by the content monitoring system and becomes part of a common pool of objectionable controlled content.

[0094] Content in the objectionable database may also be stored with a certainty rating. For example, the greater number of times the content object has been identified as violating a rule, the greater the certainty rating. In some embodiments, for each content object in the objectionable database, data is maintained regarding each usage/compliance rule that it violates. For example, content object 10034 may be non-compliant with rules 4, 7, and 112, but not other rules. This information may be stored in a table, metadata associated with content object 10034, or in any other appropriate way.

[0095] In some embodiments, if the content is being monitored for by a user at 202-213, data from that process may be re-used at 232. For example, similarity, compliance, and other factors may be determined based on data already obtained at 202-213. Additional compliance factors that take into account the host policy may also be determined and used.

[0096] At 234, content is reported. In some embodiments, which content to report is configurable and may depend on criteria based on technical factors (e.g., similarity scores), compliance factors (e.g., non-compliance scores), and/or other factors (e.g., reputation information) as described at 214. Content reported may include content determined to be non-compliant based on the host policy. Content reported may also include notices received from content owners who believe the content host is using their content in a non-compliant way.

[0097] For example, a web interface may be provided for viewing and managing reported content. In some embodiments, the web interface allows the host to track and manage past and/or pending engagement notices. The web interface includes information about matching content, reputation information, similarity scores, non-compliance scores, link(s) to usage rules associated with the content object, and any other appropriate information. Reputation information could be related to the content owner, e.g., how reputable the content owner is. For example, the content owner may not actually be the content owner, but a scam artist or spammer who has sent thousands of notices. On the other hand, a reputable content owner may have only sent 3 notices in the past year. In some embodiments, reputation is based on ratings by other content users, content hosts, and/or other users of the content monitoring system. For example, content users who have dealt with a particular content owner and felt that he was legitimate may have given him good reputation ratings. In some embodiments, APIs to the content monitoring system are provided to the OSP for managing notices and responding.

[0098] At 236, the report is responded to. In some embodiments, an automatic response is sent according to rules set by the OSP. For example, whenever the OSP receives a DMCA notice from a content owner with a reputation rating above a specified value, it automatically takes down the image. In another example, whenever a child pornography match is made with a similarity score above 90 and a non-compliance score above 80, an email is sent to the user and if no response is received within a set period of time, the content is removed. In some embodiments, an OSP administrator manually reviews each content match and selects a response for each content match.

[0099] Besides a common pool of objectionable content, various common/collaborative pools of data may be maintained. Other examples of common pools of data include reputation of content owners, reputation of content users, reputation of content hosts, content known to be in the public domain, sites known to copy the most often, etc. These common pools may be contributed to by content owners (e.g., end users), content hosts (e.g., an employee on an OSP's content review team), legal experts, experts in "fair use," other reputable entities, results from previous detection results (e.g., false positives), etc. APIs or other interfaces may be provided to assist with flagging content for inclusion in these pools. These common pools of data may then be accessed and used during the monitoring process (e.g., during 202-216 or 231-232).

[00100] For example, a negation database be maintained that includes content that is known to be in the public domain, content that has expired or lapsed in copyright, and/or content that is difficult to claim ownership of, e.g., because it is common, such as disclaimers and copyright notices. Any content in the negation database is designated as compliant.

[00101] Figure 2C is a flow chart illustrating an embodiment of a process for evaluating context of a content object. In some embodiments, this process is used to perform 212 when the context includes compliance information (e.g., compliance factors). Examples of compliance factors include the presence or absence of advertising on a page containing the content object, whether the page contains paid content, etc. In some embodiments, this process is performed by content monitoring system 100. In some embodiments, this process is performed when a content owner is monitoring for use of his content.

[00102] At 240, a detected content object associated with use of controlled content is obtained. In some embodiments, the detected content object is detected based on technical factors, as described at 210.

[00103] At 242, usage rules associated with the controlled content are obtained. In some embodiments, the usage rules specified by the content owner at 203 are obtained.

[00104] At 246, a usage rule is evaluated against the detected content object. The usage rule may be specified at a high level (e.g., do not permit use on for profit sites, permit use on nonprofit sites) or at lower level (e.g., do not permit use on pages containing advertising, offer to license on pages containing paid content, permit use on sites ending with .edu). For example, it is determined whether the page associated with the content object contains advertising, requires a subscription, contains affiliate links, or contains paid content.

[00105] At 248, it is determined whether the usage rule is satisfied. If not, one or more scores are adjusted. For example, a non-compliance score may be increased or decreased as appropriate. At 252, it is determined whether there are additional rules to check. If there are additional rules to check, the process returns to 246. If there are no additional rules to check, one or more scores are provided.

[00106] Figure 2D is a flow chart illustrating an embodiment of a process for monitoring for use of controlled content. In some embodiments, this process is performed by content monitoring system 100. In some embodiments, this process is performed when a content host, such as an OSP, is checking for non-compliant use of controlled content. For example, this process may be used to perform 232.

[00107] At 260, a content object is received. For example, a user is posting a new content object to an OSP site, and the OSP is receiving the content object for the first time. At 262, a fingerprint of the content object is generated. A fingerprint may be generated, feature(s) may be extracted, or other analysis performed, as described at 206. At 264, the fingerprint (or another analysis result) is checked against a database of known non-compliant (or known compliant) content objects. In some embodiments, the database includes a common pool of content that has previously been identified either manually or automatically

as non-compliant or compliant. The content can be looked up by fingerprint or any other appropriate index. At 266, it is determined whether the content object is non-compliant according to the database. If it is non-compliant according to the database, the content object is removed at 272. If it is not non-compliant according to the database, then the content object is evaluated at 268. (In some embodiments, if the content is compliant according to the database, then the content object is approved for posting.) Evaluating may include any of the processes described at 212-213 and/or at 240-256. In some embodiments, evaluating includes notifying the content host (e.g., the OSP) and receiving an evaluation of the content object from the content host. For example, the content host may perform a manual or automatic evaluation. The results of or data from the evaluation is stored in the database with the fingerprint. At 270, it is determined whether the content object is non-compliant according to the evaluation. For example, the determination can be made based on technical factors, compliance factors, or other factors, as previously described. If the content object is non-compliant, the content object is removed at 272. If not, the process ends. In some embodiments, if the content object is not non-compliant, then the content object is approved for posting.

[00108] Figure 2E is a flow chart illustrating an embodiment of a process for engaging with a user of non-compliant content. In this example, rather than automatically removing a non-compliant content object, the user may be contacted first. In some embodiments, this process is performed by content monitoring system 100. In some embodiments, this process is used to perform 236 when non-compliant content is found. For example, this process may be performed in place of 272. At 280, a content object is determined to be non-compliant. For example, the determination can be made based on technical factors, compliance factors, or other factors, as previously described. At 282, it is determined whether user contact is requested, which may be a configurable setting. In this example, the user refers to the entity that posted the content on the OSP. If user contact is not requested, then the content object is removed. If user contact is requested, then the user is contacted at 284. For example, the user is notified that the user's content has been identified as non-compliant content and to either take down the content, explain why the content is compliant, or cause the content to be compliant (e.g., based on usage rules for the content). At 286, it is determined whether the content object is in compliance. For example, the user is given a set amount of time to respond, and after that time, an evaluation of whether the content object is in compliance is performed. If it is still not in compliance, the content object is removed. In some

embodiments, if it is still not in compliance, the user is notified again, or another appropriate action is taken. If the content object is in compliance, the process ends. In some embodiments, if the content object is now in compliance a database is updated to include this information.

[00109] Figure 2F is a flow chart illustrating an embodiment of a process for displaying compliance information (e.g., rules) to a content user wishing to use content on a content owner's website (as described at 203). In this example, a content owner has created a web page of his content (e.g., "www.example.com") and included on the web page a link that is associated with a server that stores compliance information associated with his content. In some embodiments, the link is a common URL, where the common URL is not unique to the content owner or his web page (e.g., "rules.attributor.com"). At 290, the web page is viewed, e.g., by a potential content user. At 292, the "rules.attributor.com" link is selected. For example, the content user is interested in using the content, and would like to know if there are any usage rules associated with it.

[00110] A receiving system (e.g., a server that stores or has access to the compliance information) receives the request for "rules.attributor.com" at 296 and determines the appropriate compliance information at 298. In some embodiments, the compliance information is determined by looking up the web page from which the link was selected (e.g., the content owner's web page) in a table (or other appropriate structure) of compliance information. For example, next to "www.example.com" in the table are usage rules associated with content on "www.example.com." In some embodiments, the table includes information about content objects on the web page and associated usage rules. In some embodiments, the server retrieves the content on web page "www.example.com" and looks up associated compliance information based on the retrieved content information. For example, each content object may have a content object ID or fingerprint that may be used to identify it and look up usage rules associated with it. In some embodiments, both the URL "www.example.com" and information associated with the content object (such as a content object ID) are used to obtain the compliance information.

[00111] At 299, a web page with the compliance information is returned. At 294, the web page with the compliance information is viewed. For example, the potential content user views the compliance information and can decide whether to use the content.

[00112] Figure 3 is an example of a graphical user interface (GUI) for providing controlled content. In some embodiments, a user uses GUI 300 to specify content to be monitored at 202. As shown, a user can enter a URL or a link to controlled content or upload a file. Any number of content objects can be specified. A username and password to access content can be provided. In some embodiments, a user uses GUI 300 to specify input to ingestor 104 in Figure 1.

[00113] GUI 300 and the other GUIs described herein may vary depending on the embodiment. Which functionality to include and how to present the functionality may vary. For example, which objects (e.g., text, links, input boxes, buttons, etc.) to include and where to place the objects may vary depending on the implementation.

[00114] Figure 4A is an example of a GUI for providing controlled content. In some embodiments, GUI 400 opens in response to selecting a link in GUI 300, such as the "Add Content" button. In some embodiments, a user uses GUI 400 to specify content to be monitored at 202. In some embodiments, a user uses GUI 400 to specify input to ingestor 104 in Figure 1.

[00115] As shown, one or more files may be provided in the "My content" input box. A user can indicate whether the content is a single web page or file or a URL or feed. In the case of the URL or feed, the content includes all existing content plus any new content added in the future. In the "Nickname" input box, the user can specify a nickname for the controlled content. In this way, a user can manage or maintain multiple sets of controlled content using different nicknames.

[00116] In some embodiments, a "Sites to watch" input box is provided, in which the user may enter URLs where the user expects the content to appear. For example, the user may currently be aware that a particular site is using the user's content. In some embodiments, the content monitoring system searches the web, but searches the specified sites first or more frequently.

[00117] In some embodiments, a "Related Keywords" input box is shown, in which, the user may enter keywords associated with the specified controlled content. For example, if the user expects the content to be found primarily in children's websites, the keywords "kids" and "children" might be included. In some embodiments, the content monitoring system

automatically determines keywords (such as unique phrases) to search in addition to the related keywords specified by the user.

[00118] In some embodiment, a "Search Scope" input box, is shown, in which the user may specify whether the entire Internet should be searched or only domains specified by the user. In some embodiments, the user may specify to only search sites that copy the most often.

[00119] In some embodiments, a "Text" input box is provided, in which text may be entered. The text may be text in the content itself or text associated with the content, such as keywords, tags, depictions of the text (e.g., a photo of a street sign with text), etc. In addition, other search criteria may be specified, including a minimum similarity score, a minimum non-compliance score, a minimum percent of controlled content copied, a minimum percent of text copied, a minimum number of images copied, a minimum percent of match , whether the content is attributed (e.g., to the content owner), whether there is advertising on the page and what type, the minimum number of unique visitors per month, and what types of matches to find (e.g., images only, text only, video only, or combinations, etc.)

[00120] Figure 4B is an example of a GUI for providing usage rules. In some embodiments, GUI 402 is included as part of GUI 400. In some embodiments, GUI 402 opens in response to selecting a link in GUI 400, such as a "Specify Rules of Use" link (not shown in GUI 400). In some embodiments, a user uses GUI 402 to specify usage rules associated with the content specified in GUI 400. In some embodiments, a user uses GUI 402 to specify usage rules at 203.

[00121] As shown, a list of usage rules may be selected by selecting bullets and checkboxes. The rules listed in this example include: attribution required/not required; commercial use OK, OK if user shares a specified percentage of the revenue, or no commercial use; limit text copies to a specified percentage of the source (controlled) content; no changes may be made to controlled content; contact content owner first for permission; share alike; a specified Creative Commons license; all rights reserved; or public domain.

[00122] Graphical icons are displayed next to each usage rule. For example, "$%" indicates that commercial use is okay if the user shares a specified percentage of the revenue. "By" with a slash through it indicates that attribution is not required. "%" indicates that text copied must be limited to a specified percentage of the controlled content.

[00123] A similar GUI may be used to specify host rules for a host policy.

[00124] Figure 5 is an example of a GUI for displaying search results. In some embodiments, GUI 500 is used to report search results at 214, e.g., to a content owner. In some embodiments, reporter 110 in Figure 1 reports results using GUI 500.

[00125] In the example shown, a content owner owns a collection of photography related content, including images of cameras and text descriptions of cameras. The search results are shown in a grid based layout. In each grid cell, a controlled content object and a match content object are shown, where it has been determined that the match content object is similar to the controlled content object based on a similarity score and a non-compliance score. As shown, in grid cell 502, the controlled image (camera1) and the match image (camera2) have a similarity score of 98 and a non-compliance score of 88. In some embodiments, data displayed includes one or more of the following: similarity score, non-compliance score, URL of the match content object, percent of the controlled object copied, percent of the controlled text copied, the number of controlled images copied, the date found, whether there is advertising on the page, etc. In the case of text, a portion of the copied text is displayed along with the controlled text in grid cell 504.

[00126] In some embodiments, rather than or in addition to reporting a score, a binary flagging (e.g., "interesting" or not) is reported. For example, a score that aggregates similarity, non-compliance, and/or other factors into a combined/summary score may be displayed.

[00127] In some embodiments, if there is more than one matched content object, then the additional matched content objects are displayed using a 3D graphical effect indicating there are multiple pairs. Using forward and back arrows, the user can cycle through the various pairs. In some embodiments, the pairs are displayed in descending similarity score order.

[00128] Various other functionality may be provided in GUI 500. For example, the search results may be filtered and sorted in various ways using the “Showing” and “Sort by” pull down menus. Additional controlled content may be added in the “Controlled Content” input box, an email address may be entered for automatic notification (e.g., when more matches are found) in the “Email Address” input box, etc. Rather than use a grid based layout, other layouts may be used in other embodiments.

[00129] In the case of a content host monitoring for use of non-compliant content based on the host policy, an interface similar to interface 500 may be used to display resulting matches. For example, cell 502 may display a match with copyrighted content. Cell 504 may display a match with content associated with child pornography. For example, in place of text1 may be a known image that has been positively identified (either manually or automatically) as child pornography, and in place of text2 may be a new image that is being posted by a user to the content host. In this case, the known image in place of text1 may have been in a database of known non-compliant content, and the match determined as described at 264. In some cases, the new image is determined to be a match with child pornography based on an evaluation (e.g., 268) rather than a match with a content object in a database of known pornography. In this case, in place of text1, there may be no image displayed, or data related to the evaluation may be displayed instead.

[00130] Figure 6 is an example of a GUI for displaying use of a content object. In some embodiments, GUI 600 is displayed in response to selecting a "Match" link or the image or text corresponding to a match object in GUI 500.

[00131] In the example shown, the portions of the web page that include use of the controlled content are marked, i.e., boxed (e.g., a graphical box around the image or text that is being used). In this example, text1, text3, and photo2 are controlled content objects that are being used on this web page. In various embodiments, various indicators (e.g., visual cues) may be used to indicate the copied portions. Examples of indicators include: highlighting text, changing font appearance (e.g., using bold, underline, different fonts or font sizes, etc.), using different colors, displaying icons or other graphics in the vicinity of the copied portions, using time dependent indicators, such as causing the copied portions to flash, etc.

[00132] Various options or functionality may be provided for displaying information related to the use of the controlled content. For example, an archive date (May 31, 2006) may be displayed. Applicable usage rule(s) specified by the content owner may be displayed. In this case, the usage rules are displayed using the icons described with respect to Figure 4B. When selecting an icon, details regarding the associated usage rule may be displayed.

[00133] In some embodiments, the web page shown is the current version of the web page. In some embodiments, the web page shown is an archived version. For example, the archived version may be stored in monitored content store 118. Whether the web page is the

current version or an archived version may be indicated in the GUI. In addition, the user may be able to toggle between the two versions.

[00134] In some embodiments, a management GUI may be provided for managing content that provides links to one or more of the GUIs described above. In some embodiments, a user uses the management GUI to manage content, including add new controlled content, modify search parameters, report search results, etc. For example, various tabs may be provided, such as a "My Content" tab used to add/modify controlled content and search parameters and a "Matches" tab used to display search results. In some embodiments, selecting the "Matches" tab opens GUI 500.

[00135] A user can group content into categories, such as categories associated with the user's blog, camera reviews, the user's eBay account, and all files. In various embodiments, content may be grouped in folders, by tags, or in any other appropriate way. A list of controlled content (e.g., URLs, paths) associated with the category may be displayed, including the number of content objects associated with the controlled content, when the content objects were last archived (e.g., placed in controlled content store 116), rules associated with each content object, and the number of matches found for the content object(s).

[00136] ***Controlled Content on a Web Page Having Revenue-Generating Code***

[00137] Figure 7 is a block diagram illustrating an embodiment of a system for handling controlled content on a web page having revenue-generating code (RGC).

[00138] In the example shown, system 700 is shown to include web servers (or other content hosts) 702-710, detecting entity 712, and RGC parties 714-716. Web servers 702-710 may include any number of web servers, each hosting a web page that includes controlled content. For example, web server 702 includes web page 704, which includes controlled content 706 and revenue-generating code 708. Because web page 704 includes revenue-generating code 708, there is revenue generated each time controlled content 706 is displayed. For example, revenue-generating code 708 may be associated with an advertising network that displays advertisements based on CPMs (cost per thousand page impressions). Therefore, the owner of controlled content 706 may wish to share in some of this revenue.

[00139] In various embodiments, revenue-generating code may be associated with any party (referred to herein as RGC party), including, but not limited to: advertising network, an advertiser, e-commerce, an affiliate model, a survey, or any party that renumerates the publisher of the web page for inclusion of the revenue-generating code on the publisher's web page. For example, web page 704 could contain information about mountain climbing and controlled content 706 may be an excerpt from a book about climbing. Revenue-generating code 708 could be Google AdWords code, which provides advertisements to items related to climbing, such as offers from adventure travel companies. In another example, Revenue-generating code 708 could be Amazon.com affiliate code, which provides a link to buy the book. If a user buys a book through the link, Amazon.com (the RGC party) pays the publisher a royalty.

[00140] In some embodiments, detecting entity 712 is part of content monitoring system 100. Detecting entity 712 crawls web pages, such as web pages 704 and 718, searching for web pages that include controlled content and revenue-generating code. In some embodiments, when controlled content and revenue-generating code is detected in a web page, detecting entity 712 notifies a party associated with the revenue-generating code (e.g., RGC party 714). In some embodiments, detecting entity 712 compiles this information and then notifies the appropriate RGC party at preconfigured times. In various embodiments, detecting entity 712 includes a database of controlled content objects, including the content owner and royalty information (e.g., what percentage of revenue should be paid to the content owner for use of a controlled content object). Using this information, detecting entity 712 can compute royalty information and send that to appropriate RGC parties and/or the publisher of the web pages on which the controlled content and RGC was detected. The RGC party may then either distribute a share of the revenue to each of the content owner and detecting entity 712, or distribute a share of the revenue to detecting entity 712, which in turn provides a share to the content owner. Further examples are more fully provided below.

[00141] Figure 8 is a flow chart illustrating an embodiment of a process for handling controlled content on a web page having revenue-generating code. In some embodiments, this process is performed automatically by an entity such as detecting entity 712. At 802, a web page is analyzed. At 804, controlled content is detected. A variety of detection techniques may be used, including, for example, generating a fingerprint of the controlled

content object and checking the fingerprint against a database of controlled content objects, as discussed above.

[00142] At 806, revenue-generating code is detected. In some embodiments, pattern matching with known revenue-generating code patterns may be used. For example, Adblocker software may be used to identify code related to advertising. Some examples of revenue-generating code include:

[00143] Google AdSense revenue-generating code example:

```
[00144]     <script type="text/javascript"><!--
[00145]             google_ad_client = "pub-7732994450185580";
[00146]             /* 160x600, created 10/4/08 */
[00147]             google_ad_slot = "9666776673";
[00148]             google_ad_width = 160;
[00149]             google_ad_height = 600;
[00150]             //-->
[00151]             </script>
[00152]             <script type="text/javascript"
[00153]
src="http://pagead2.googlesyndication.com/pagead/show_ads.js">
[00154]             </script>
```

[00155] In this example, the pattern searched for would be "http://pagead2.googlesyndication.com".

[00156] Amazon affiliate revenue-generating code example:

```
[00157]     <SCRIPT charset="utf-8" type="text/javascript"
src="http://ws.amazon.com/widgets/q?ServiceVersion=20070822&MarketPlace=US&ID=V2
```

```
0070822/US/widgetsamazon-20/8005/6a93f3c7-8a6d-4210-be91-798d5a2fee1a">
</SCRIPT>
```

[00158] In this example, the pattern searched for would be "http://ws.amazon.com/widgets/q?ServiceVersion=20070822&MarketPlace=US&ID=V20070822/US/widgetsamazon-20/8005/6a93f3c7-8a6d-4210-be91-798d5a2fee1a".

[00159] In various embodiments, other patterns may be searched for. In addition, there may be other refinements to ensure that the search is accurate. For example, another refinement might be to check to make sure that the code is within JavaScript so that a web page that is simply displaying this text on the web page does not get matched.

[00160] At 808, a party to contact is automatically determined based on the revenue-generating code. In some embodiments, a table of revenue-generating code patterns and RGC parties is maintained and consulted to determine the RGC party. In the above examples, the RGC parties identified would be Google and Amazon. In some embodiments, the party to contact includes the publisher of the web page.

[00161] At 810, the process may continue in various ways, depending on the embodiment, as more fully described below.

[00162] Figure 9 is a flow chart illustrating an embodiment of a process for notifying a party of royalty information. In some embodiments, this process is performed automatically by an entity such as detecting entity 712. In this example, 810 continues at 902. At 904, it is determined whether the controlled content is monetizable. As used herein, controlled content is monetizable when monetary/financial or other compensation is or could be paid to the owner of the controlled content for use of the controlled content. For example, some cases where content might not be considered monetizable would be if the use of the controlled content in the web page complies with Fair Use, the web page is an educational or government site, or only a limited portion of the controlled content is used (e.g., whether less a prescribed number of words is used), etc. As such, 904 could include making a determination of whether the use of the controlled content in the web page complies with Fair Use, the web page is an educational or government site, or only a limited portion of the controlled content is used.

[00163] If it is determined that the controlled content is not monetizable, the process ends at 906. If it is determined that the controlled content is monetizable, then at 908, royalty information is determined based on a royalty scheme. In some embodiments, determining royalty information involves looking up the controlled content object in a table and determining the amount of royalty (e.g., a percentage or flat amount) to be paid to the content owner. In various embodiments, the royalty may be a function of the content owner and/or the publisher of the web page in which the controlled content is used. For example, a different royalty may be specified depending on who is publishing the content (e.g., a larger percentage may be specified if www.cnn.com (which has a large audience) uses the content as opposed to a personal website with a small audience). Different content objects owned by the same content owner may be associated with different royalty amounts. Different content owners may specify different royalty amounts. In various embodiments, the royalty scheme may depend on the amount of controlled content being used on the web page (e.g., what percentage of the controlled content is being used) and/or the amount that the controlled content contributes to the web page (e.g., what percentage of the content on the web page is the controlled content).

[00164] At 910, the party associated with the revenue-generating code is automatically notified of the royalty information. For example, an email or other message may be sent to the RGC party at preconfigured times. In various embodiments, instead of or in addition to sending royalty information, other information is sent. For example, a list of one or more web pages (e.g., URLs), a list of controlled content on each web page, and the content owners of each controlled content object could be sent. In some embodiments, the RGC party determines the royalty information based on the information sent to the RGC party at 910. In addition, information regarding the basis for and/or supporting evidence for the royalty determination may be sent, such as who the content owner is, a copy of the source controlled content, and/or a link to the URL that shows the web page with the controlled content highlighted.

[00165] Various examples of information that may be sent at 910 include: the URL of the web page; the percent royalty due; the license type (e.g., A, B, or C, where each associated with a different percent royalty due); identification of the content owner; copy of the source controlled content; the amount of controlled content being used on the web page (e.g., what percentage of the controlled content is being used); and the amount that the

controlled content contributes to the web page (e.g., what percentage of the content on the web page is the controlled content).

[00166] Figure 10 is a flow chart illustrating an embodiment of a process for handling receipt of royalty information. In some embodiments, this process is performed by a party such as RGC party 714, in response to receiving a notification of royalty information at 910. At 1002, royalty information is received. At 1004, it is determined whether the party disputes the royalty information. For example, the amount of royalty specified may be different from an agreed upon amount, or if there was not an agreed upon amount, the amount may not be acceptable to the party. If there is no dispute, then at 1006, the royalty is paid. In some embodiments, an amount is paid to the detecting entity and the detecting entity pays a portion to the content owner. In some embodiments, the RGC party pays the royalty directly to the content owner and some amount to the detecting entity as a service fee. The service fee may be a portion of the royalty paid to the content owner or a separate amount. The RGC party may then send a notification to the detecting entity that the royalty has been paid or will be paid.

[00167] If there is a dispute, then at 1008, the detecting entity (or other appropriate entity) is notified of the dispute so that it can be resolved, either automatically or offline. In some embodiments, when the detecting entity receives notice of the dispute, if the dispute meets the appropriate criteria (e.g., refusal to pay royalty to content owner), a DMCA takedown notice is issued automatically and sent to the RGC party.

[00168] Figure 11 is a flow chart illustrating an embodiment of a process for notifying a party associated with revenue-generating code of information associated with controlled content. This is an alternative to the process of Figure 9 (communicating royalty information) or can be used in combination with the process of Figure 9. In this example, rather than communicating royalty information to the RGC party, other information that may be useful to the RGC party (e.g., for advertisement optimization and/or targeting) is communicated. For example, the RGC party may choose to charge a higher CPM for advertising that appears on a page that include controlled content owned by a particular content owner. For example, high profile, highly credible, or highly reputable content owners, such as Forbes or The Wall Street Journal, may fall in a premium category whereby advertising that appears on web pages including their content demands a higher CPM from advertisers.

[00169] In some embodiments, this process is performed automatically by an entity such as detecting entity 712. In this example, 810 continues at 1102. At 1104, information associated with the source of the controlled content is identified. For example, the content owner is identified. At 1106, the RGC party is notified of the identified source information.

[00170] In some embodiments, the process of Figure 9 can be combined with the process of Figure 11. For example, the notification steps of 910 and 1106 can be combined into a single message sent to the RGC party.

[00171] Figure 12 is a flow chart illustrating an embodiment of a process for handling receipt of information associated with controlled content. In some embodiments, this process is performed by a party such as RGC party 714, in response to receiving a notification of information associated with controlled content at 1106. At 1202, identified source information is received. For example, the identity of the content owner is received. At 1204, the identified source information is used to affect a revenue generating scheme.

[00172] In the examples of Figures 11-12, information associated with the source of the controlled content was discussed. In various embodiments, other information associated with the controlled content may also be sent. Examples of other types of information include the number of pages that the controlled content is syndicated (or used) on. In other words, the more prevalent the controlled content is, then the higher (or lower) value might be placed on the controlled content and the RGC party may choose to charge a higher (or lower) CPM when revenue-generating code appears on a web page with such content. Another example of information that could be provided includes metadata about the controlled content. For example, metadata could be provided by the content owner or users. An example of metadata might be "auto" or "Ferrari" if the controlled content is an article about Ferraris. The RGC party could use metadata information to better target advertising.

[00173] In some embodiments, any one or more of the processes of Figures 8-12 may be performed by the RGC party (e.g., RGC party 714). For example, an ad network may perform the detection, determine royalty information, and/or make payment decisions.

[00174] In some embodiments, any one or more of the processes of Figures 8-12 may be performed by the content owner. For example, the detecting entity may detect the content owner's content on a web page having revenue-generating code, and notify the content owner. The content owner may then send a notice to the RGC party associated with revenue-

generating code or to a party associated with the web page having the RGC. For example, the Associated Press may use the output of the detecting entity to automatically send payment request notices to ad networks or publishers that are reusing the Associated Press' content.

[00175] In some embodiments, the detecting entity may automatically send payment request or royalty notices to a party associated with the web page having the revenue-generating code (e.g., parties associated with web server 702 and/or web page 704, such as a publisher of the web page, a content host, or a content owner). For example, detecting entity 712 may automatically send a payment/royalty notice to the publisher of the web page.

[00176] Although the foregoing embodiments have been described in some detail for purposes of clarity of understanding, the invention is not limited to the details provided. There are many alternative ways of implementing the invention. The disclosed embodiments are illustrative and not restrictive.

[00177] WHAT IS CLAIMED IS:

CLAIMS

1. A system for managing controlled content on a web page, comprising:

a processor configured to:

analyze a web page;

detect controlled content and revenue-generating code on the web page; and

determine a party to contact based on the revenue-generating code or the controlled content; and

a memory coupled to the processor and configured to provide the processor with instructions.

2. A system as recited in claim 1, wherein the processor is further configured to determine whether use of the controlled content on the web page is monetizable.

3. A system as recited in claim 1, wherein the processor is further configured to determine whether use of the controlled content on the web page complies with Fair Use.

4. A system as recited in claim 2, wherein the processor is further configured to, in the event that it is determined that use of the controlled content on the web page is monetizable, determine a royalty based on a royalty scheme.

5. A system as recited in claim 4, wherein the royalty scheme is based at least in part on the amount of the controlled content being used on the web page.

6. A system as recited in claim 4, wherein the royalty scheme is based at least in part on the amount that the controlled content contributes to the web page.

7. A system as recited in claim 4, wherein the processor is further configured to send a URL associated with the web page to the party.

8. A system as recited in claim 7, wherein the processor is further configured to send royalty information, content owner information, a license type, or a copy of the controlled content to the party.

9. A system as recited in claim 4, wherein the royalty is a function of the content owner.

10. A system as recited in claim 4, wherein the royalty is a function of the content host.

11. A system as recited in claim 4, wherein the party pays at least a portion of the royalty to a party that detected the controlled content and revenue-generating code.

12. A system as recited in claim 4, wherein the party pays at least a portion of the royalty to the content owner.

13. A system as recited in claim 4, wherein the party pays the royalty to an entity that detected the controlled content and revenue-generating code and wherein the entity pays at least a portion of the royalty to the content owner.

14. A system as recited in claim 4, wherein the processor is further configured to receive an indication that the party disputes the royalty.

15. A system as recited in claim 1, wherein the processor is further configured to identify information associated with the source of the content and inform the party of the identified source information.

16. A system as recited in claim 15, wherein the identified source information is used by the party for advertisement targeting.

17. A system as recited in claim 15, wherein the identified source information is used by the party to charge advertisers based on the identified source information.

18. A system as recited in claim 1, wherein the party is a party associated with the revenue-generating code.

19. A system as recited in claim 1, wherein the party is an owner of the controlled content.

20. A system as recited in claim 1, wherein the party is a host or a publisher of the web page.

21. A method of managing controlled content on a web page, comprising:

analyzing a web page;

detecting controlled content and revenue-generating code on the web page; and

determining a party to contact based on the revenue-generating code or the controlled content.

22. A computer program product for managing controlled content on a web page, , the computer program product being embodied in a computer readable medium and comprising computer instructions for:

analyzing a web page;

detecting controlled content and revenue-generating code on the web page; and determining a party to contact based on the revenue-generating code or the controlled content.

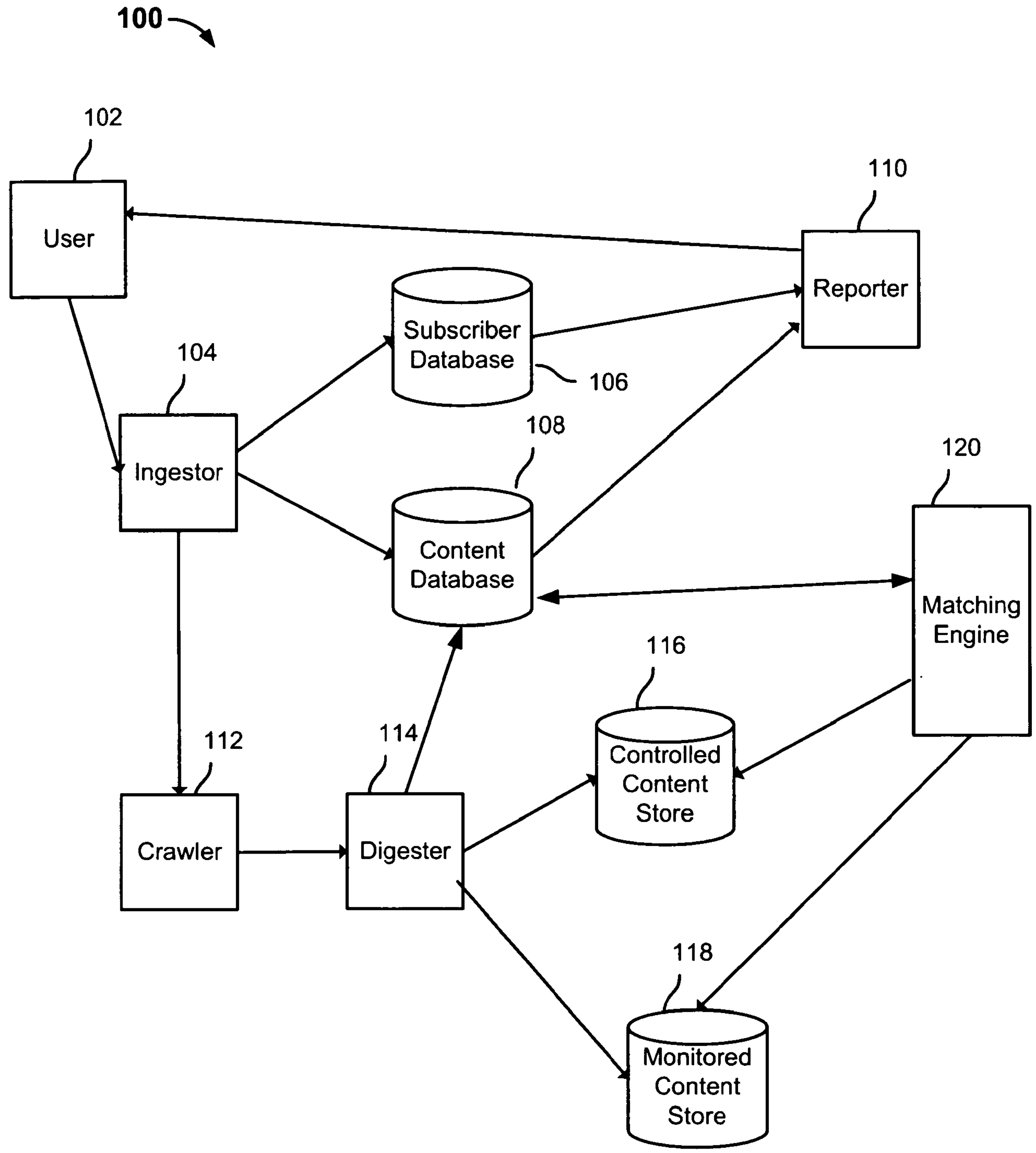

100
102
User
110
Reporter
Subscriber Database
106
104
Ingestor
108
Content Database
120
Matching Engine
116
Controlled Content Store
112
Crawler
114
Digester
118
Monitored Content Store

FIG. 1

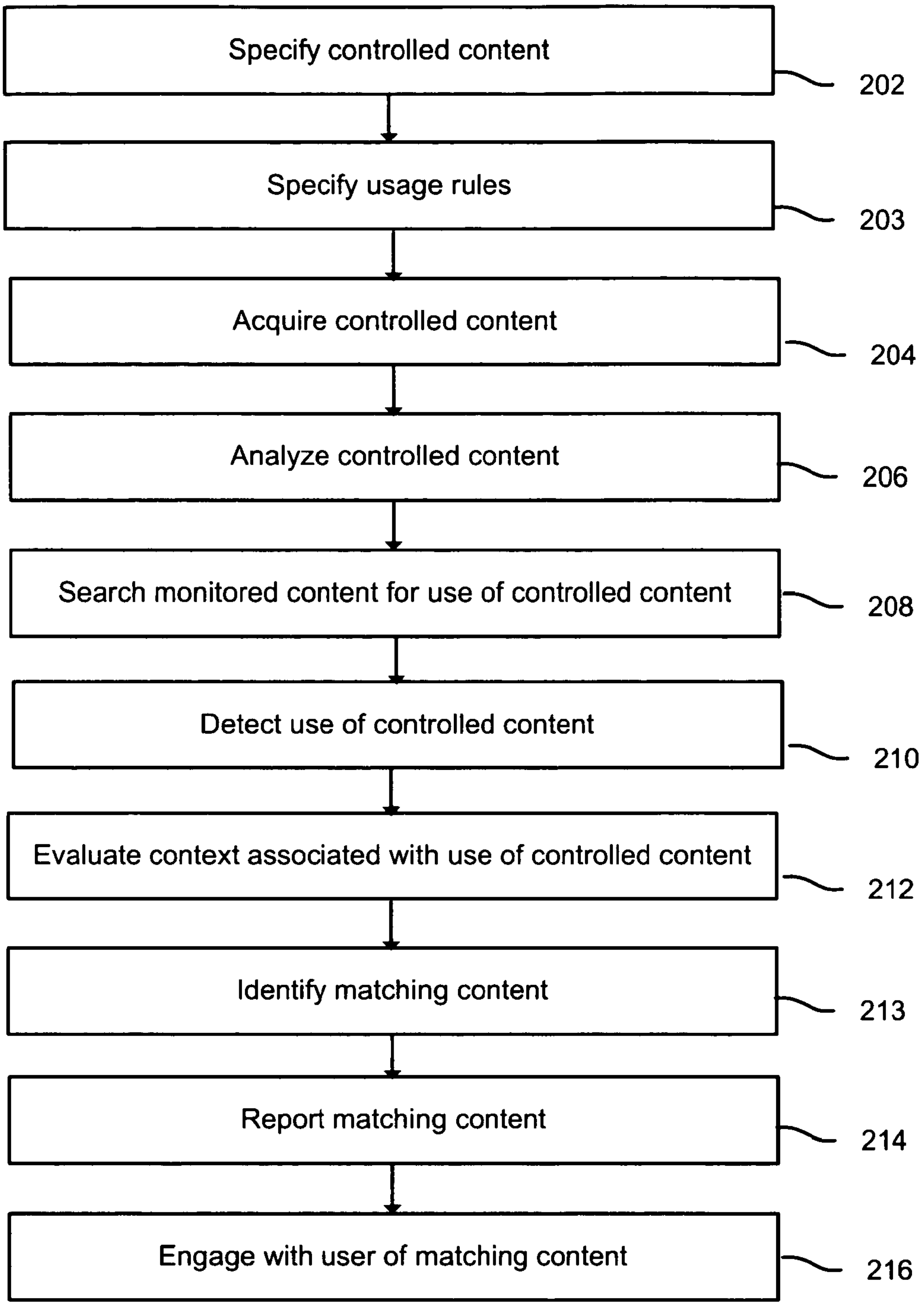

Specify controlled content
202
Specify usage rules
203
Acquire controlled content
204
Analyze controlled content
206
Search monitored content for use of controlled content
208
Detect use of controlled content
210
Evaluate context associated with use of controlled content
212
Identify matching content
213
Report matching content
214
Engage with user of matching content
216

FIG. 2A

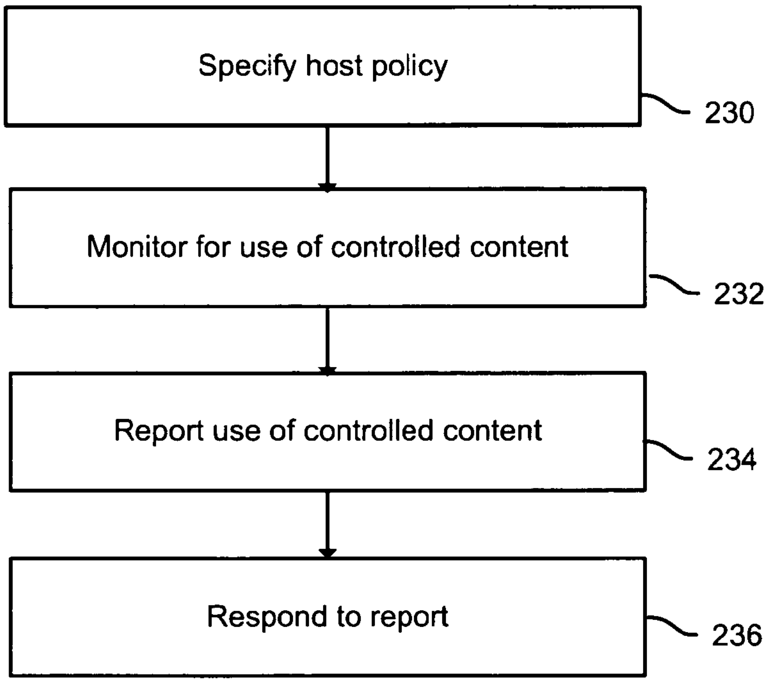

Specify host policy
230
Monitor for use of controlled content
232
Report use of controlled content
234
Respond to report
236

FIG. 2B

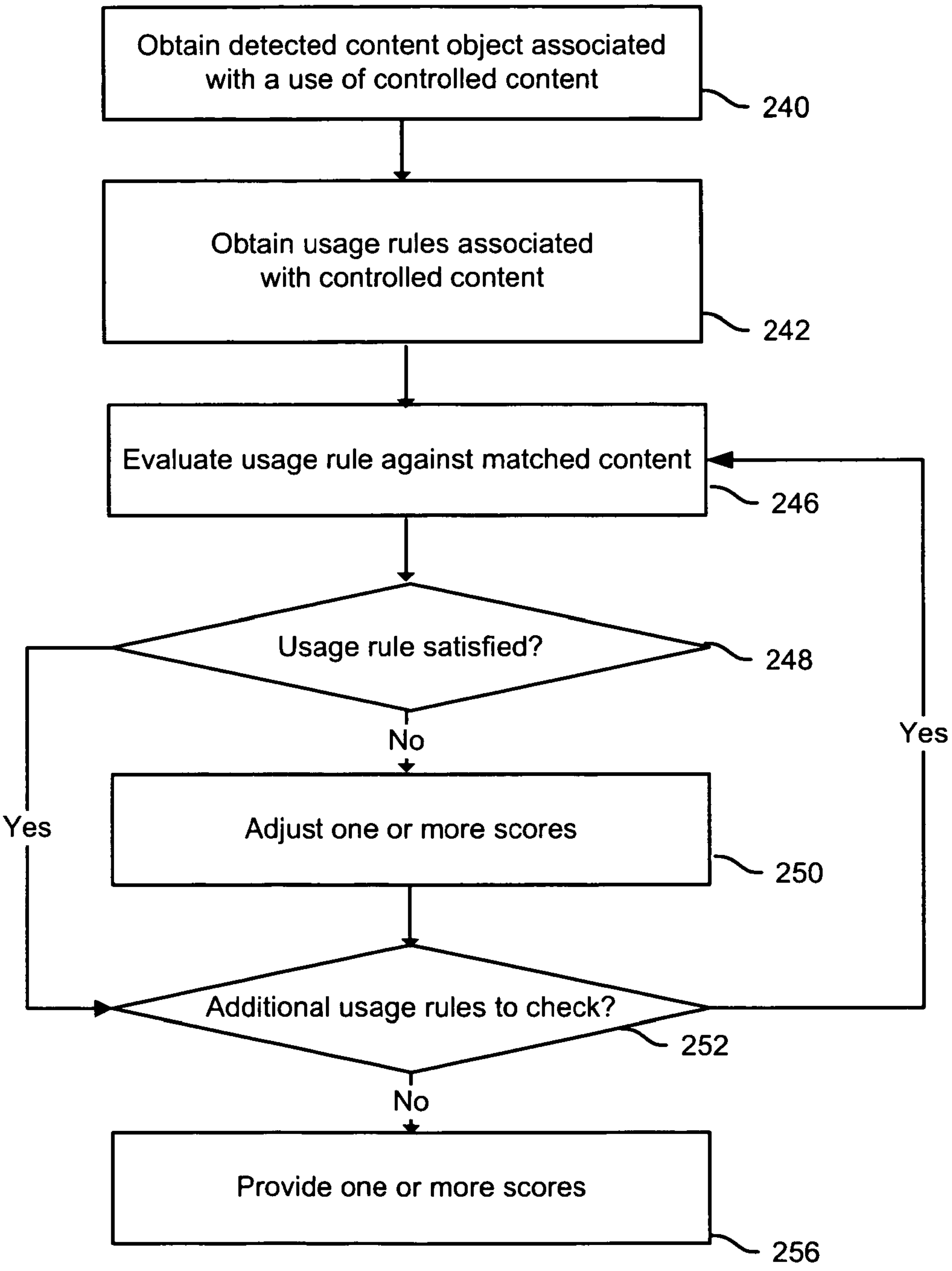

Obtain detected content object associated with a use of controlled content
240
Obtain usage rules associated with controlled content
242
Evaluate usage rule against matched content
246
Usage rule satisfied?
248
No
Yes
Yes
Adjust one or more scores
250
Additional usage rules to check?
252
No
Provide one or more scores
256

FIG. 2C

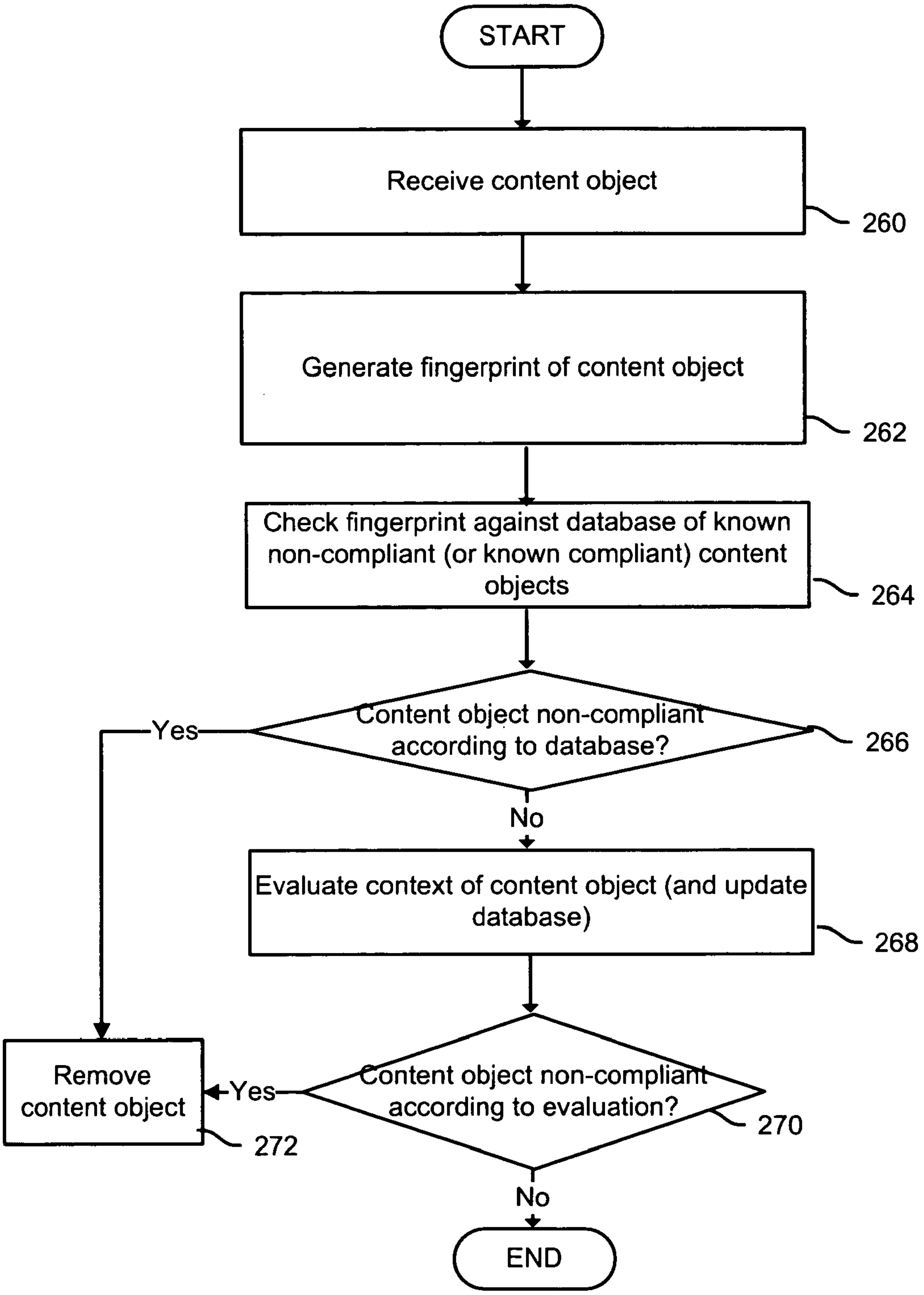

START
Receive content object
260
Generate fingerprint of content object
262
Check fingerprint against database of known non-compliant (or known compliant) content objects
264
Content object non-compliant according to database?
266
Yes
No
Evaluate context of content object (and update database)
268
Content object non-compliant according to evaluation?
270
Yes
No
Remove content object
272
END

FIG. 2D

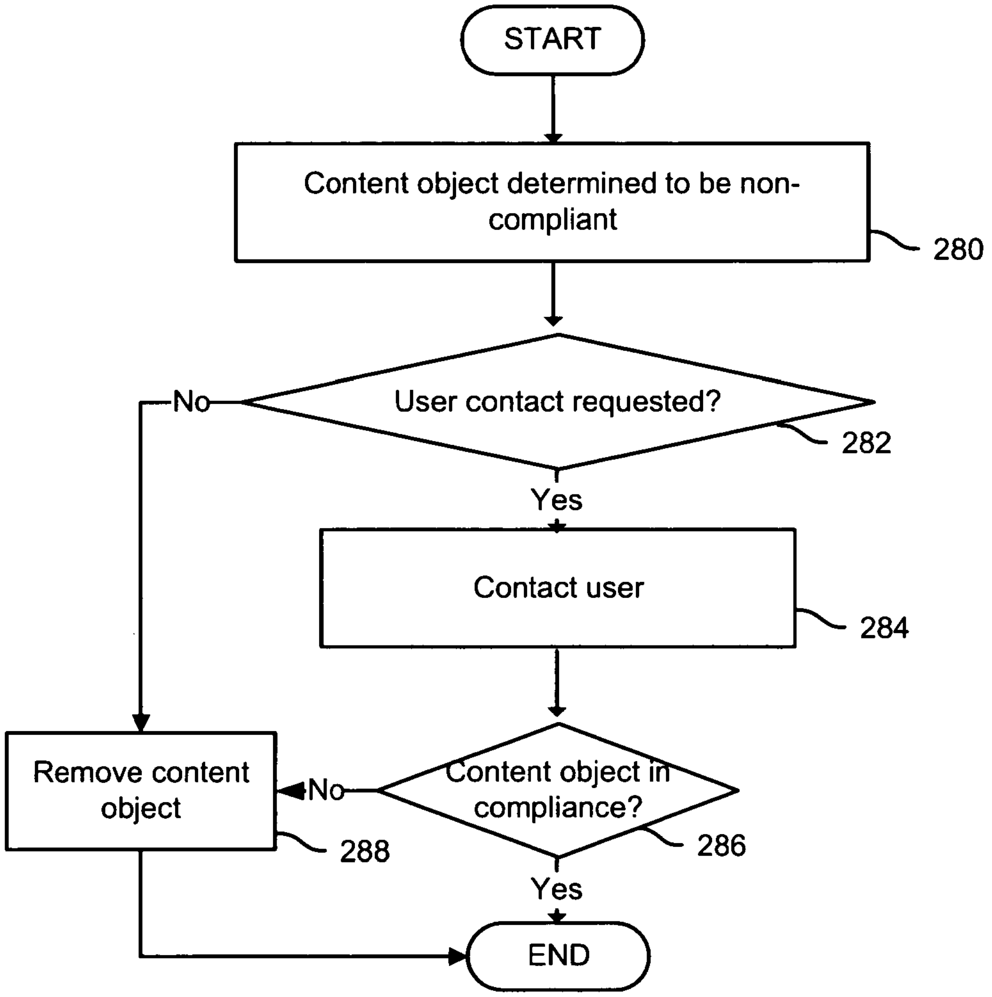

START
Content object determined to be non-compliant
280
User contact requested?
No
282
Yes
Contact user
284
Remove content object
No
Content object in compliance?
288
286
Yes
END

FIG. 2E

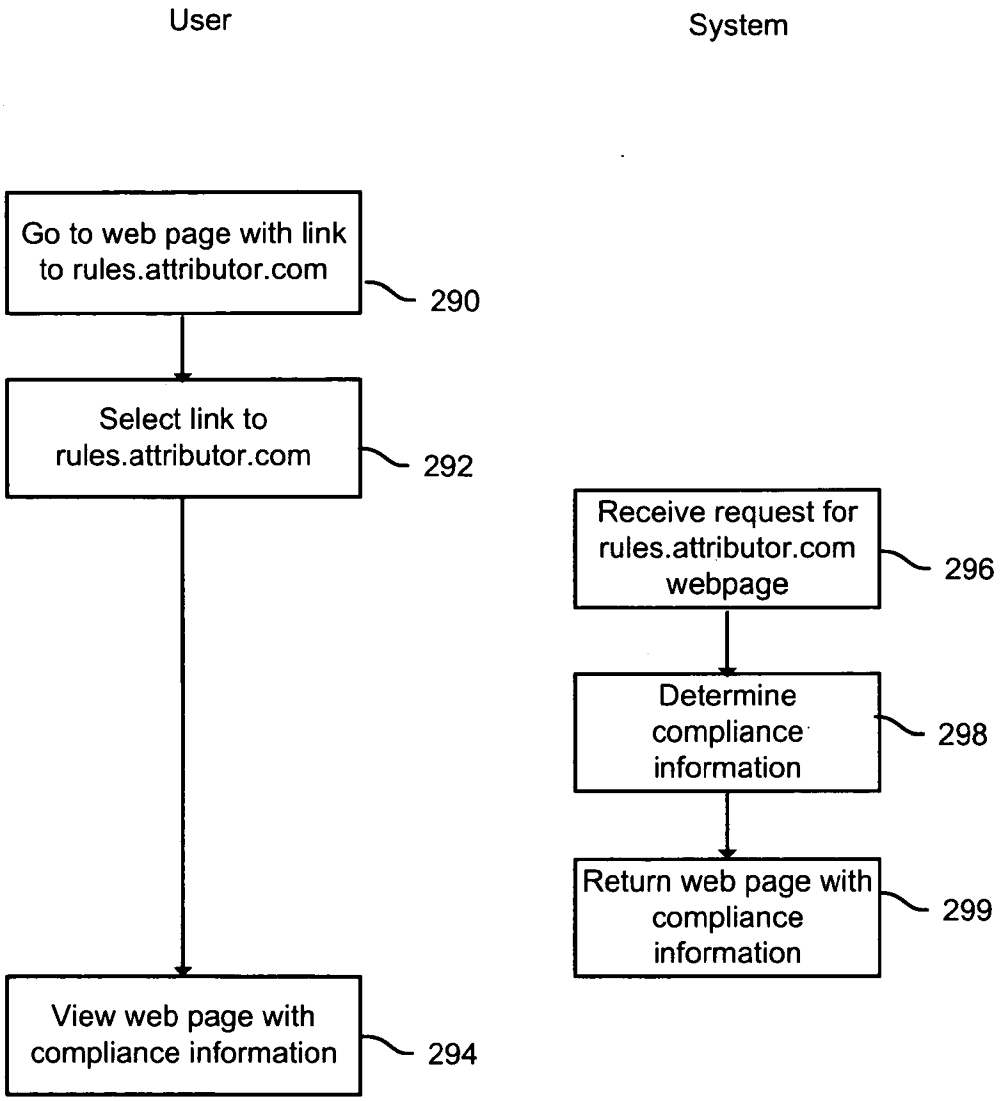
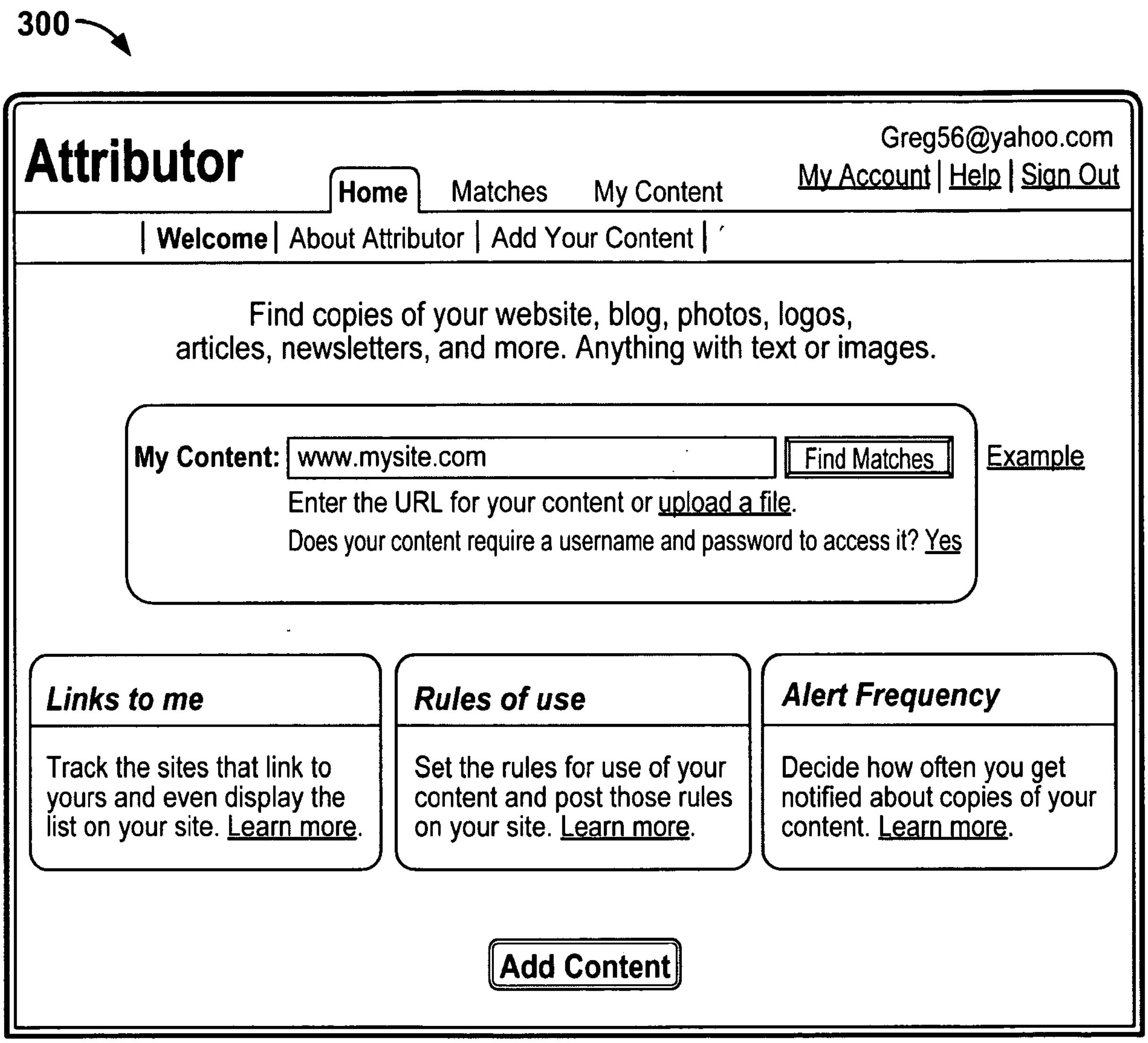

User
System
Go to web page with link to rules.attributor.com
290
Select link to rules.attributor.com
292
View web page with compliance information
294
Receive request for rules.attributor.com webpage
296
Determine compliance information
298
Return web page with compliance information
299

FIG. 2F



Find copies of your website, blog, photos, logos, articles, newsletters, and more. Anything with text or images.

My Content: www.mysite.com

Enter the URL for your content or upload a file.

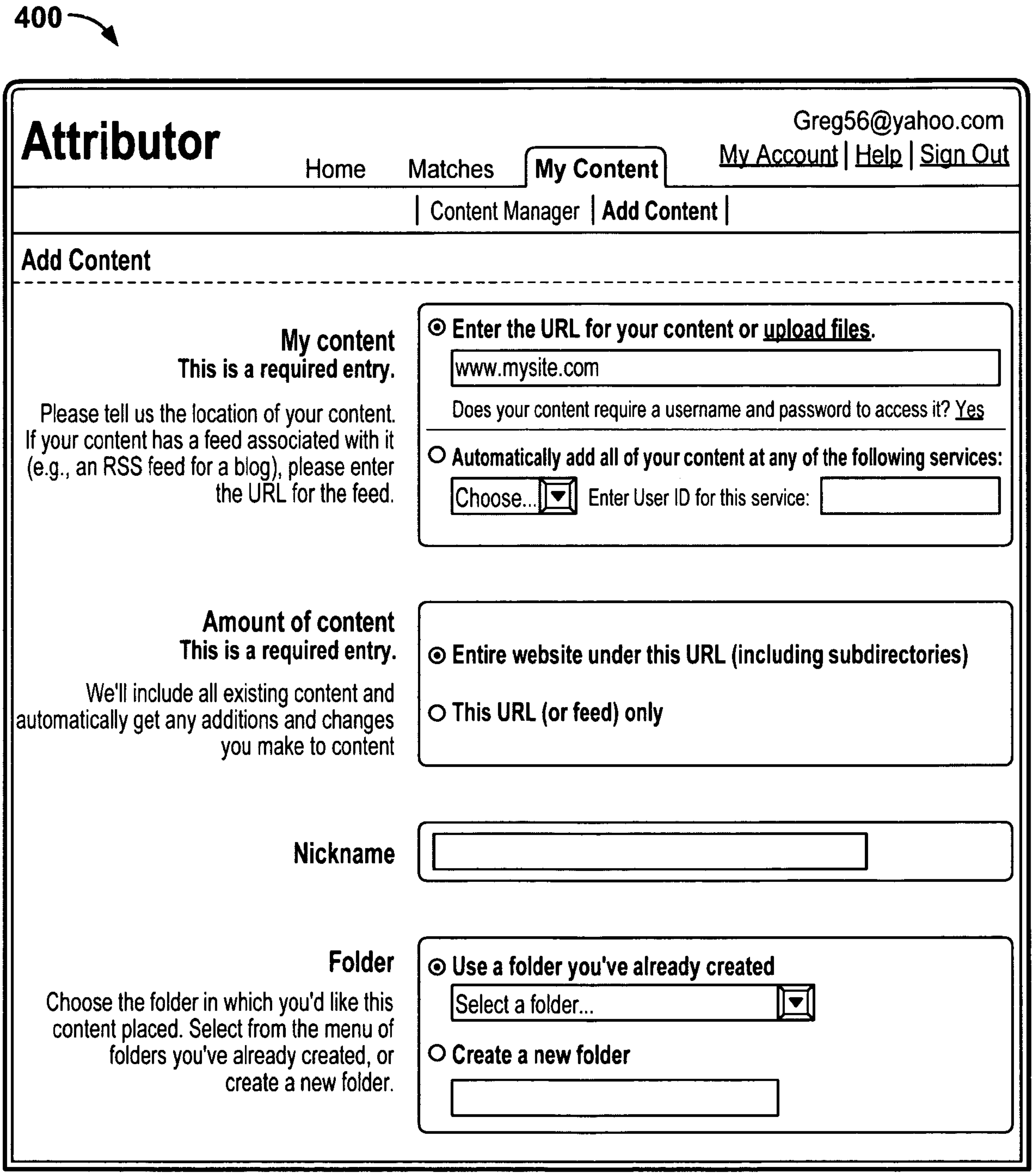
Does your content require a username and password to access it? Yes

Track the sites that link to yours and even display the list on your site. Learn more.

Set the rules for use of your content and post those rules on your site. Learn more.

Decide how often you get notified about copies of your content. Learn more.

FIG. 3



Attributor

Home | Matches | **My Content**

Greg56@yahoo.com

My Account | Help | Sign Out

| Content Manager | **Add Content** |

Add Content

My content
This is a required entry.

Please tell us the location of your content. If your content has a feed associated with it (e.g., an RSS feed for a blog), please enter the URL for the feed.

● **Enter the URL for your content or upload files.**

www.mysite.com

Does your content require a username and password to access it? Yes

○ **Automatically add all of your content at any of the following services:**

Choose... ▼ Enter User ID for this service:

Amount of content
This is a required entry.

We'll include all existing content and automatically get any additions and changes you make to content

● **Entire website under this URL (including subdirectories)**

○ **This URL (or feed) only**

Nickname

Folder

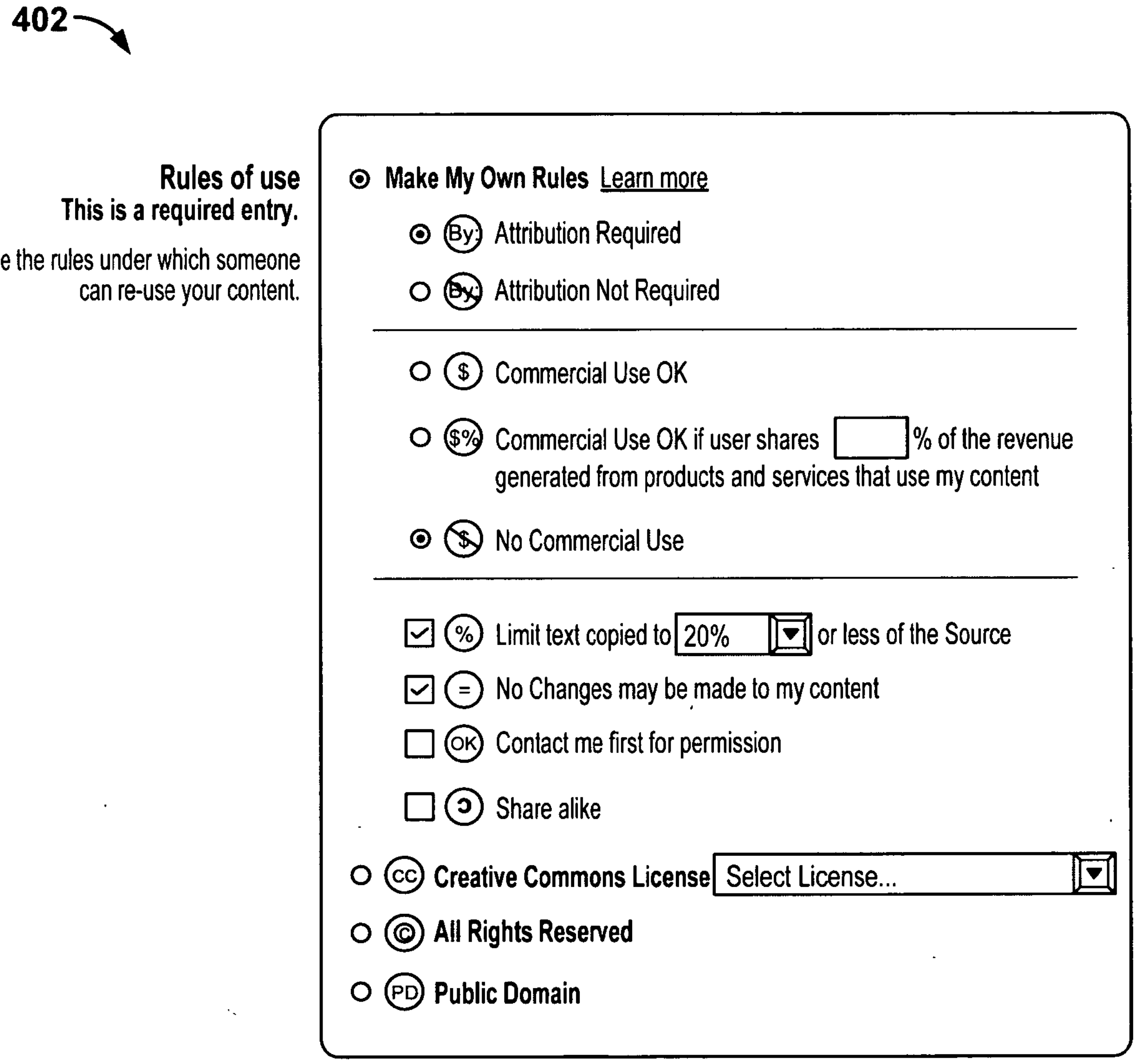
Choose the folder in which you'd like this content placed. Select from the menu of folders you've already created, or create a new folder.

● **Use a folder you've already created**

Select a folder... ▼

○ **Create a new folder**

FIG. 4A



Rules of use
This is a required entry.

Choose the rules under which someone can re-use your content.

● **Make My Own Rules** Learn more

● By Attribution Required

○ By Attribution Not Required

○ $ Commercial Use OK

○ $% Commercial Use OK if user shares [ ] % of the revenue generated from products and services that use my content

● $ No Commercial Use

☑ % Limit text copied to 20% ▼ or less of the Source

☑ = No Changes may be made to my content

□ OK Contact me first for permission

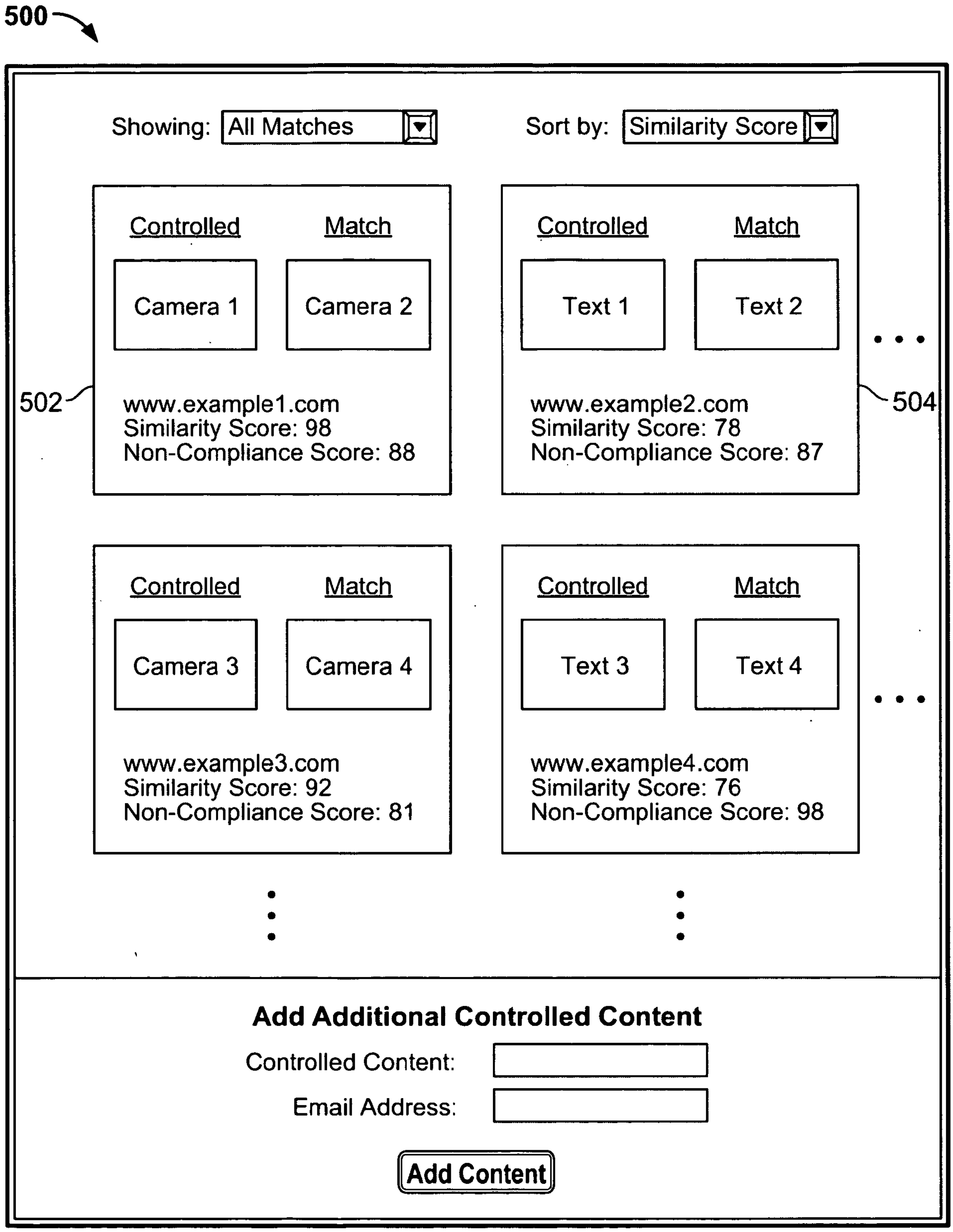
□ Share alike

○ cc **Creative Commons License** Select License... ▼

○ © **All Rights Reserved**

○ PD **Public Domain**

FIG. 4B



Showing: All Matches

Sort by: Similarity Score

Controlled: Camera 1 — Match: Camera 2
www.example1.com
Similarity Score: 98
Non-Compliance Score: 88

Controlled: Text 1 — Match: Text 2
www.example2.com
Similarity Score: 78
Non-Compliance Score: 87

Controlled: Camera 3 — Match: Camera 4
www.example3.com
Similarity Score: 92
Non-Compliance Score: 81

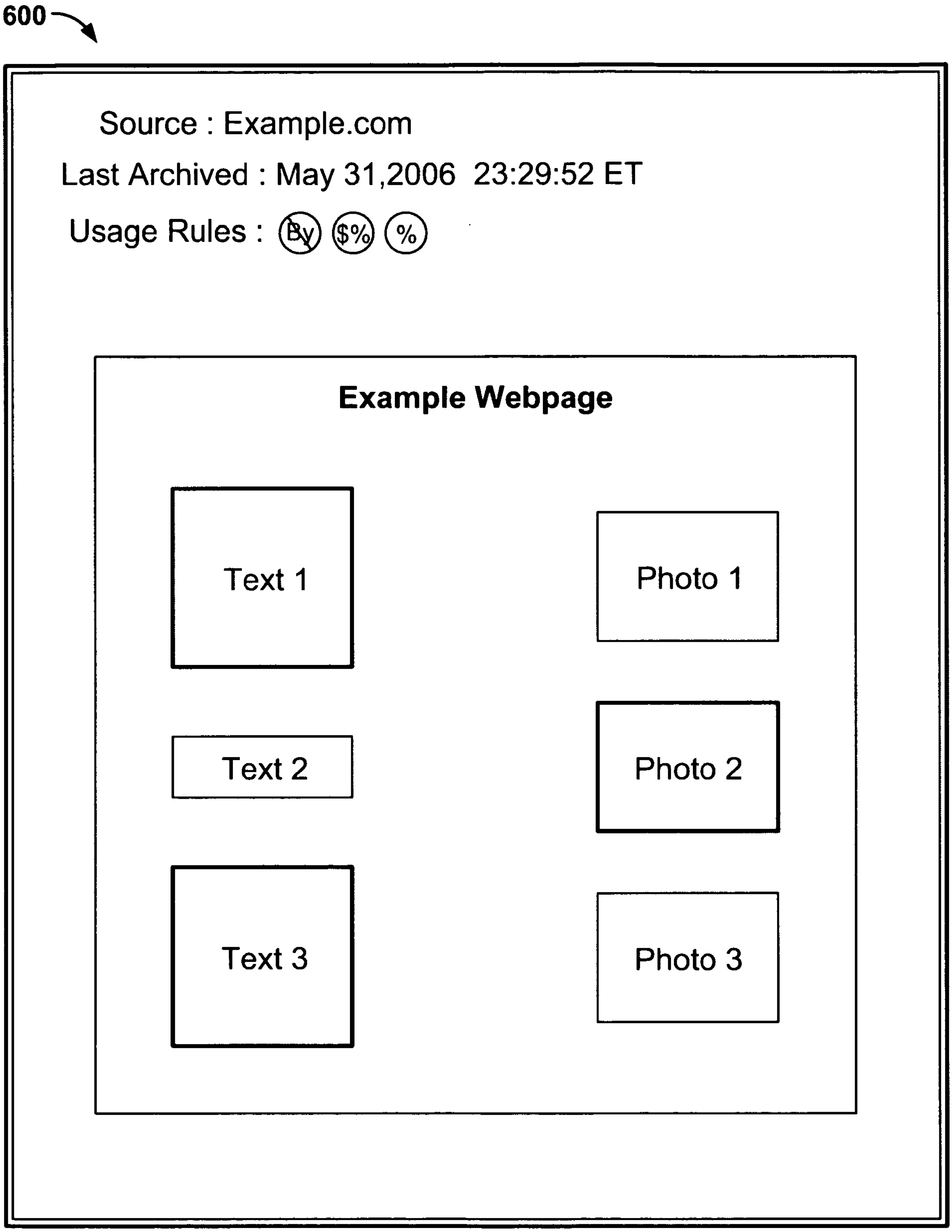
Controlled: Text 3 — Match: Text 4
www.example4.com
Similarity Score: 76
Non-Compliance Score: 98

Add Additional Controlled Content

Controlled Content:

Email Address:

FIG. 5



Source : Example.com

Last Archived : May 31,2006 23:29:52 ET

Usage Rules : By $% %

FIG. 6

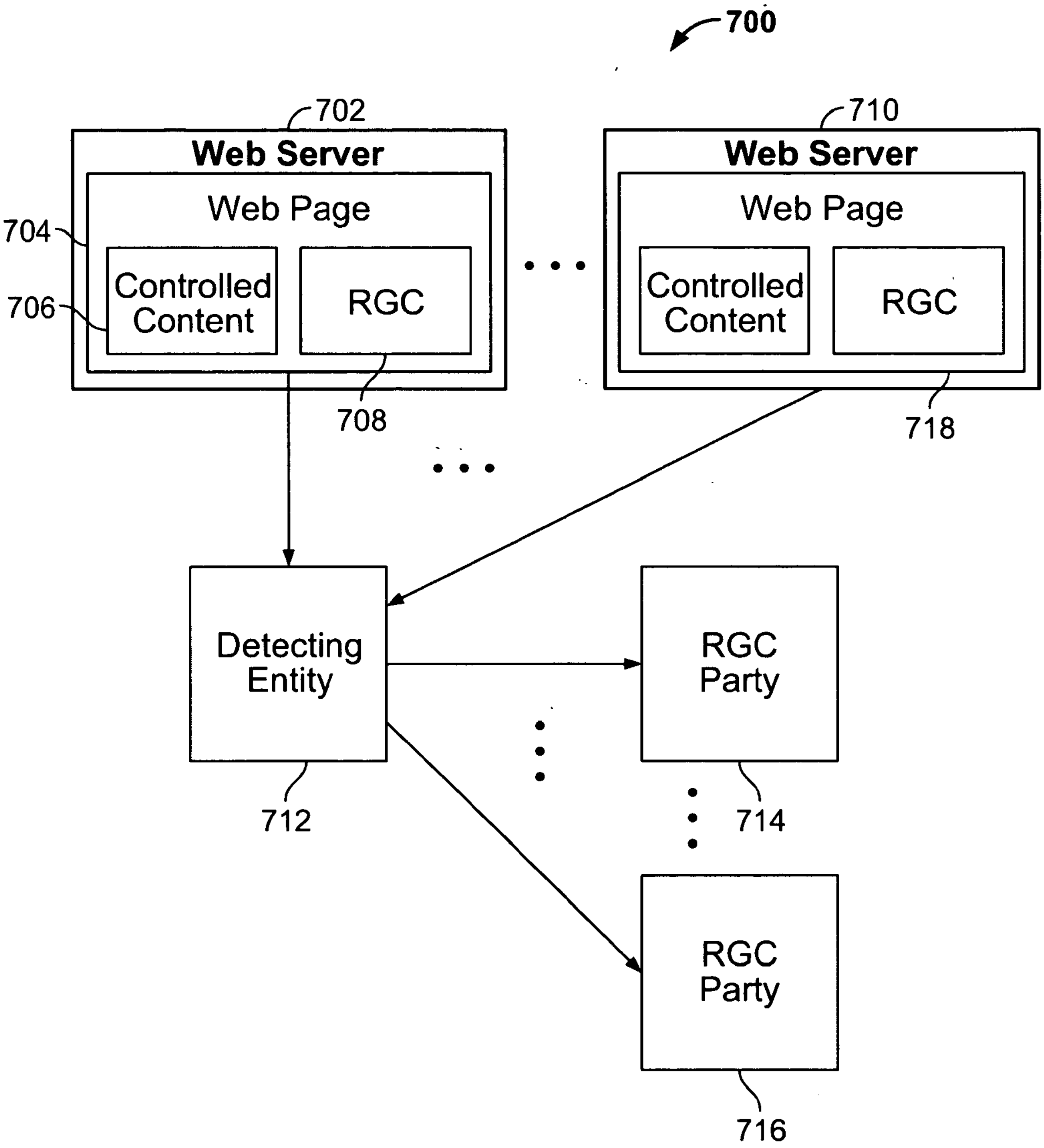

700
702
Web Server
Web Page
704
706
Controlled Content
RGC
708
710
Web Server
Web Page
Controlled Content
RGC
718
Detecting Entity
712
RGC Party
714
RGC Party
716

FIG. 7

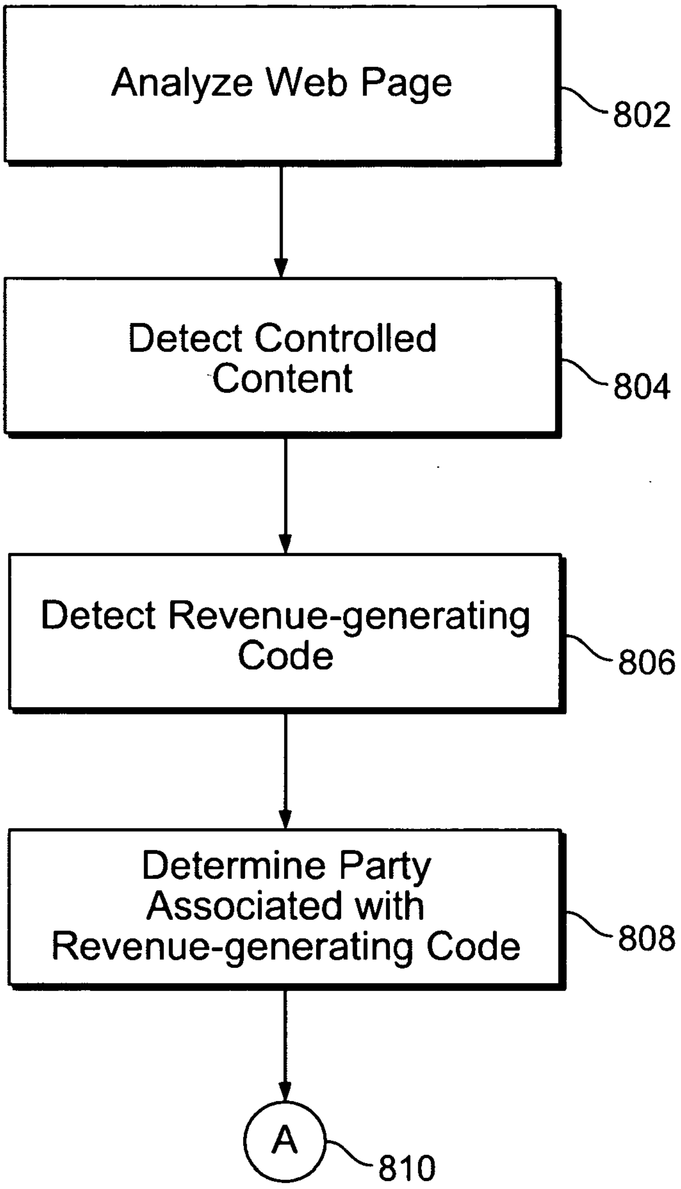

Analyze Web Page
802
Detect Controlled Content
804
Detect Revenue-generating Code
806
Determine Party Associated with Revenue-generating Code
808
A
810

FIG. 8

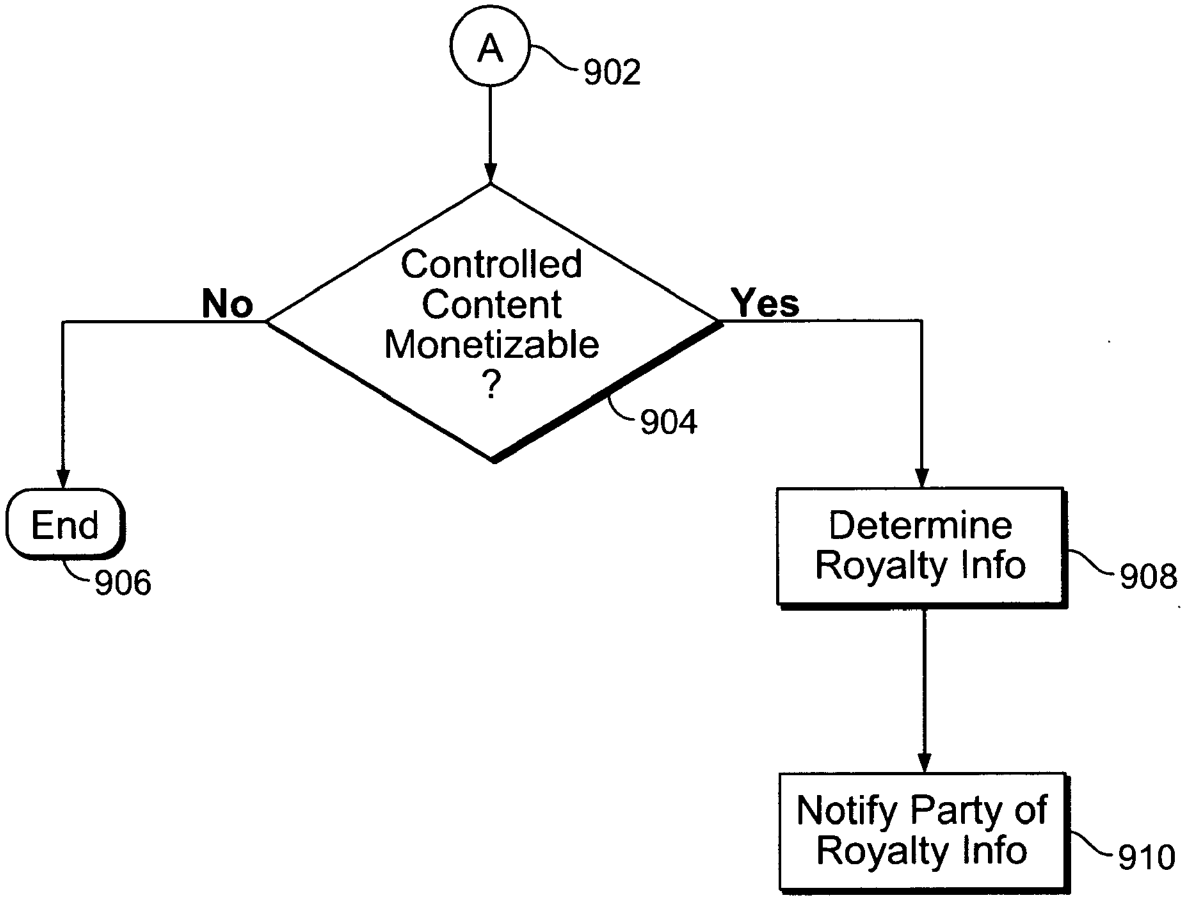

A
902
Controlled Content Monetizable ?
904
No
Yes
End
906
Determine Royalty Info
908
Notify Party of Royalty Info
910

FIG. 9

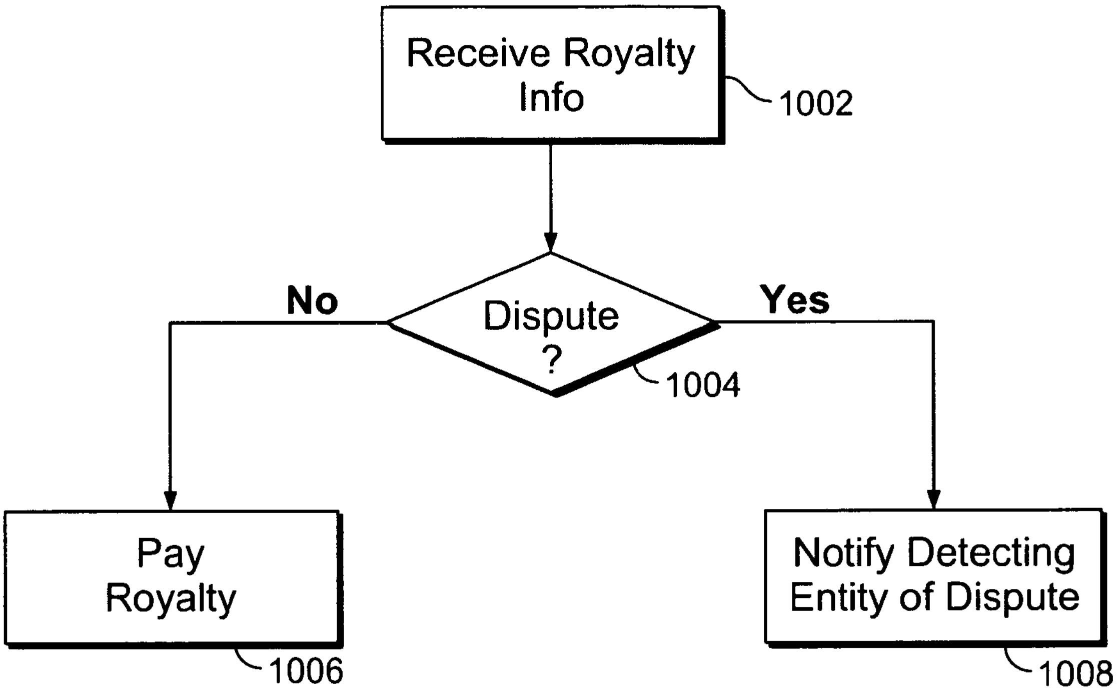

Receive Royalty Info
1002
Dispute ?
1004
No
Yes
Pay Royalty
1006
Notify Detecting Entity of Dispute
1008

FIG. 10

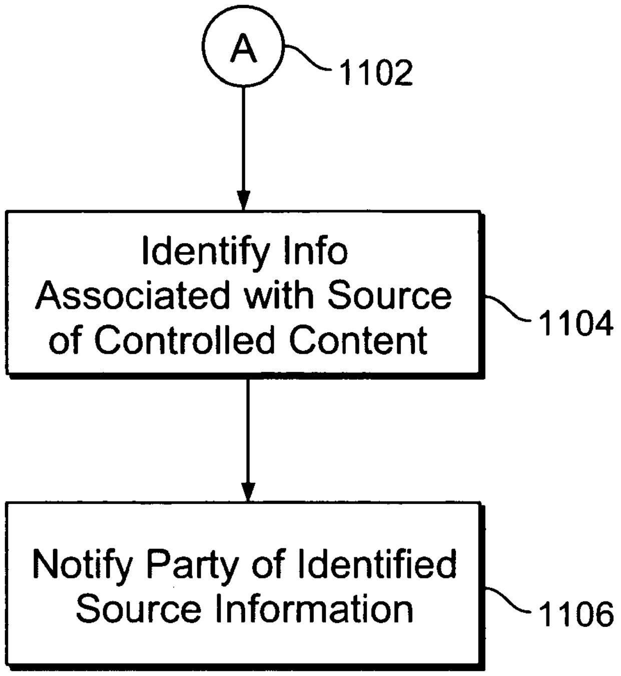

A
1102
Identify Info Associated with Source of Controlled Content
1104
Notify Party of Identified Source Information
1106

FIG. 11

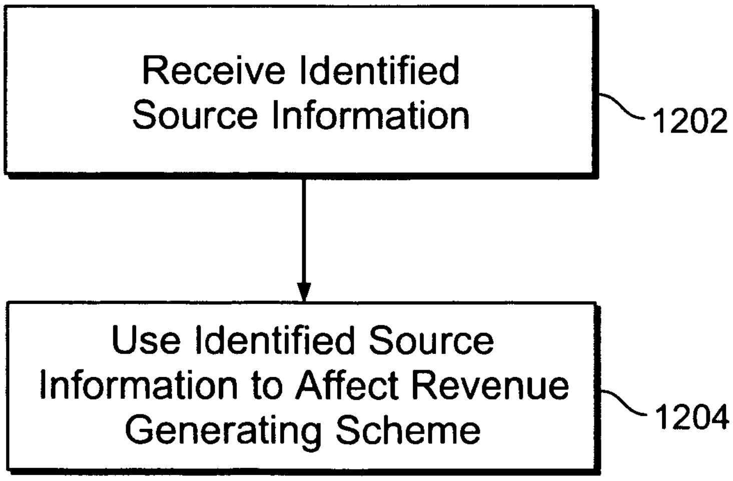

Receive Identified Source Information
1202
Use Identified Source Information to Affect Revenue Generating Scheme
1204

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/01068

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 15/16 (2010.01)
USPC - 709/217

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC - 709/217

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 709/201, 217, 218, 219; 705/1.1, 7, 10, 30, 32, 39

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Dialog; Google Patents; Google Scholar. Search Terms: revenue, advertisement, ads, royalt, scheme, plan, schedule, processor, computer, owner, licens, web, internet, disput, conflict, URL, link, hyperlink, publish, broadcast, post, print, display, code, source, instructions, fair, use, manag, control, distribut, split, divide, share, etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	US 2005/0119976 A1 (TAYLOR et al.) 02 June 2005 (02.06.2005), Para [0023], [0033], [0035], [0040], [0043], and [0046].	1, 2, 4-6, 9-13, 15-22 — 3, 7, 8, 14
Y	US 2006/0074913 A1 (O'SULLIVAN et al.) 06 April 2006 (06.04.2006), Para [0051].	3
Y	US 2005/0021398 A1 (MCCLESKEY et al.) 27 January 2005 (27.01.2005), Para [0145], [0177], [0180] and [0184].	7, 8, 14
A	US 2008/0071561 A1 (HOLCOMBE) 20 March 2008 (20.03.2008).	1-22

□ Further documents are listed in the continuation of Box C. □

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search 26 May 2010 (26.05.2010)	Date of mailing of the international search report 16 JUN 2010
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774