



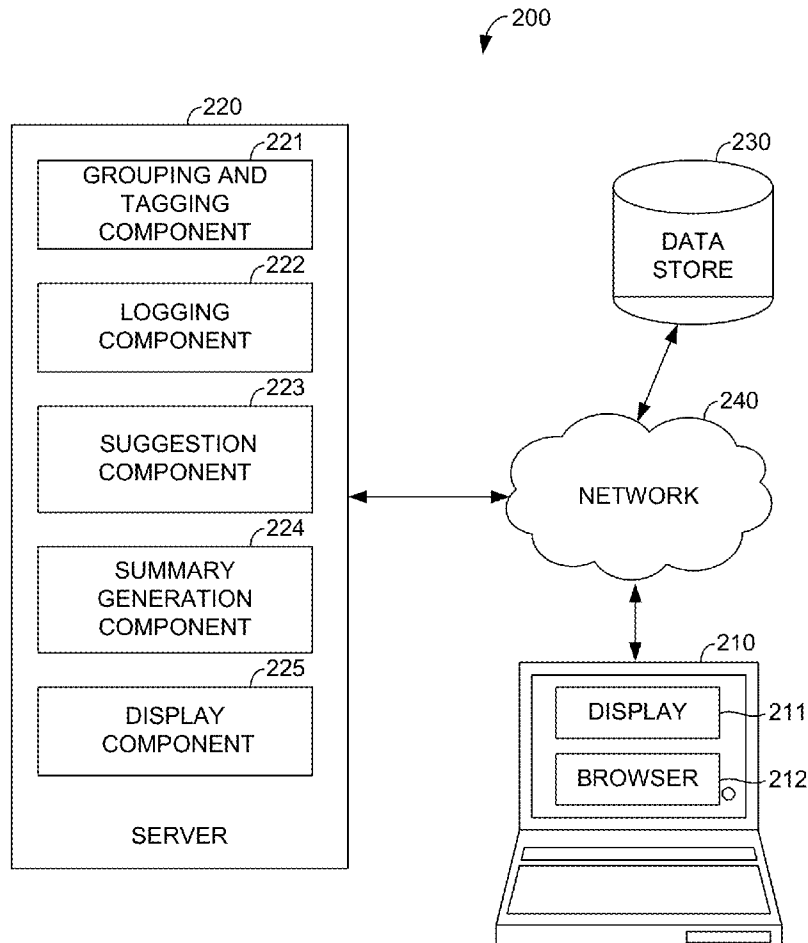
US 20150227531A1

(19) **United States**(12) **Patent Application Publication**
KULESZA et al.(10) **Pub. No.: US 2015/0227531 A1**(43) **Pub. Date: Aug. 13, 2015**(54) **STRUCTURED LABELING TO FACILITATE
CONCEPT EVOLUTION IN MACHINE
LEARNING**(71) Applicant: **MICROSOFT CORPORATION,**
REDMOND, WA (US)(72) Inventors: **TODD DAVID KULESZA,**
CORVALLIS, OR (US); **DENIS
XAVIER CHARLES,** REDMOND, WA
(US); **RICH CARUANA,** Woodinville,
WA (US); **SALEEMA AMIN
AMERSHI,** SEATTLE, WA (US);
DANYEL AHARON FISHER,
SEATTLE, WA (US)(73) Assignee: **MICROSOFT CORPORATION,**
REDMOND, WA (US)(21) Appl. No.: **14/176,990**(22) Filed: **Feb. 10, 2014****Publication Classification**(51) **Int. Cl.****G06F 17/30** (2006.01)**G06F 3/0481** (2006.01)**G06F 3/0486** (2006.01)**G06F 3/0482** (2006.01)(52) **U.S. Cl.****CPC G06F 17/30058** (2013.01); **G06F 17/30601**(2013.01); **G06F 3/0482** (2013.01); **G06F****3/04817** (2013.01); **G06F 3/0486** (2013.01)

(57)

ABSTRACT

A system, method, and media are provided for generating a structured labeling graphical user interface. The user interface receives user input that associates multimedia content with categories. The user input may include user-supplied tags that further define the category for the multimedia content. The user-supplied tags are rendered proximate to the categories. In turn, a database logs user events to store, among other things, the categories, the user-supplied tags, time associated with completing the user-supplied tags, and time for associating multimedia content with the categories or tags.



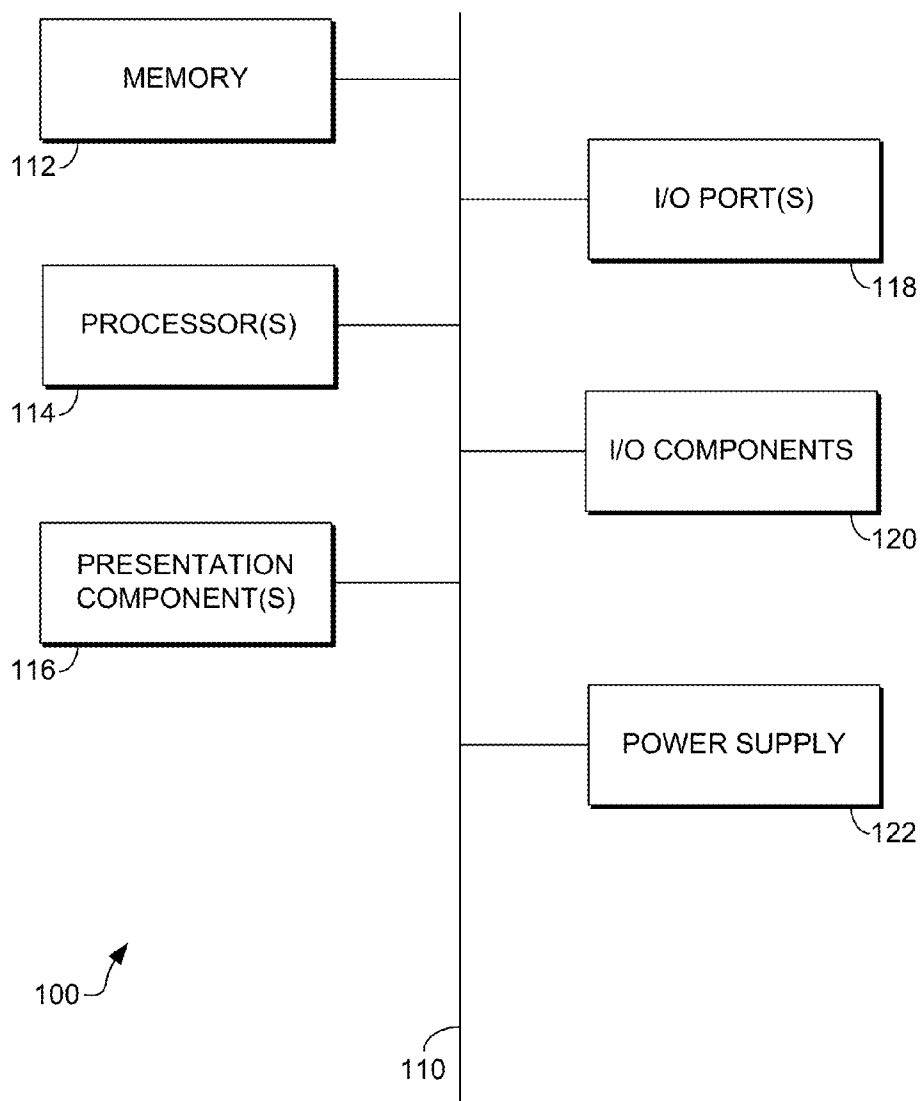


FIG. 1

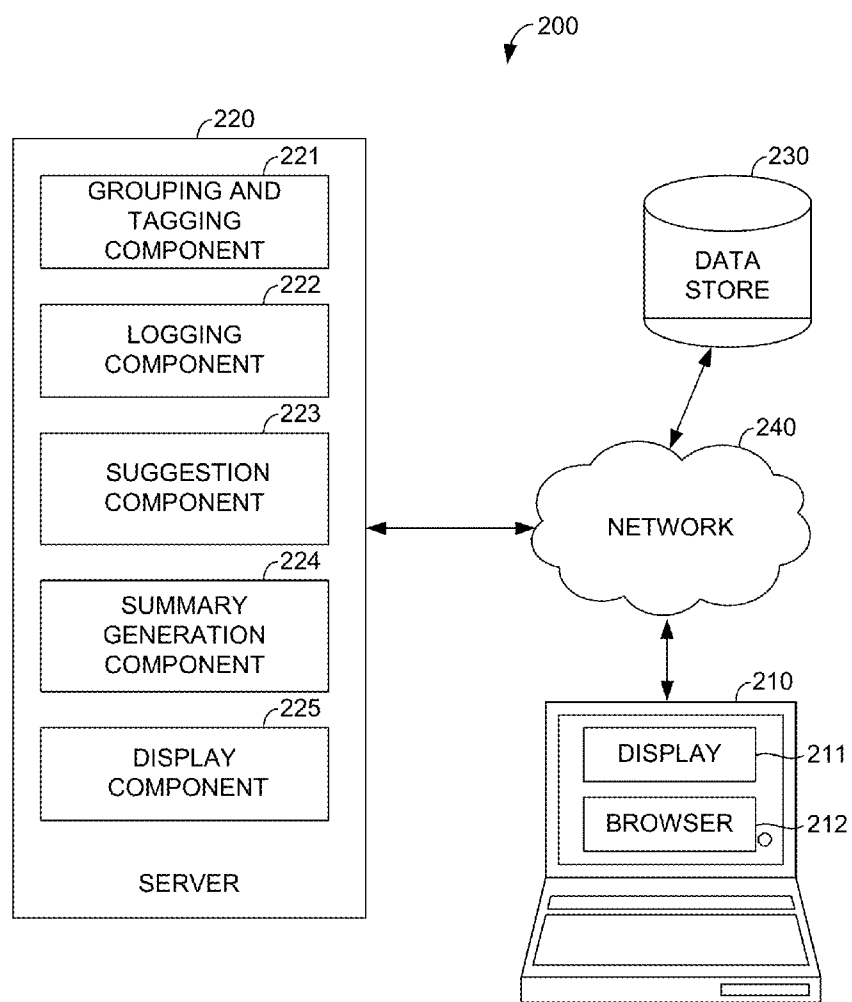


FIG. 2

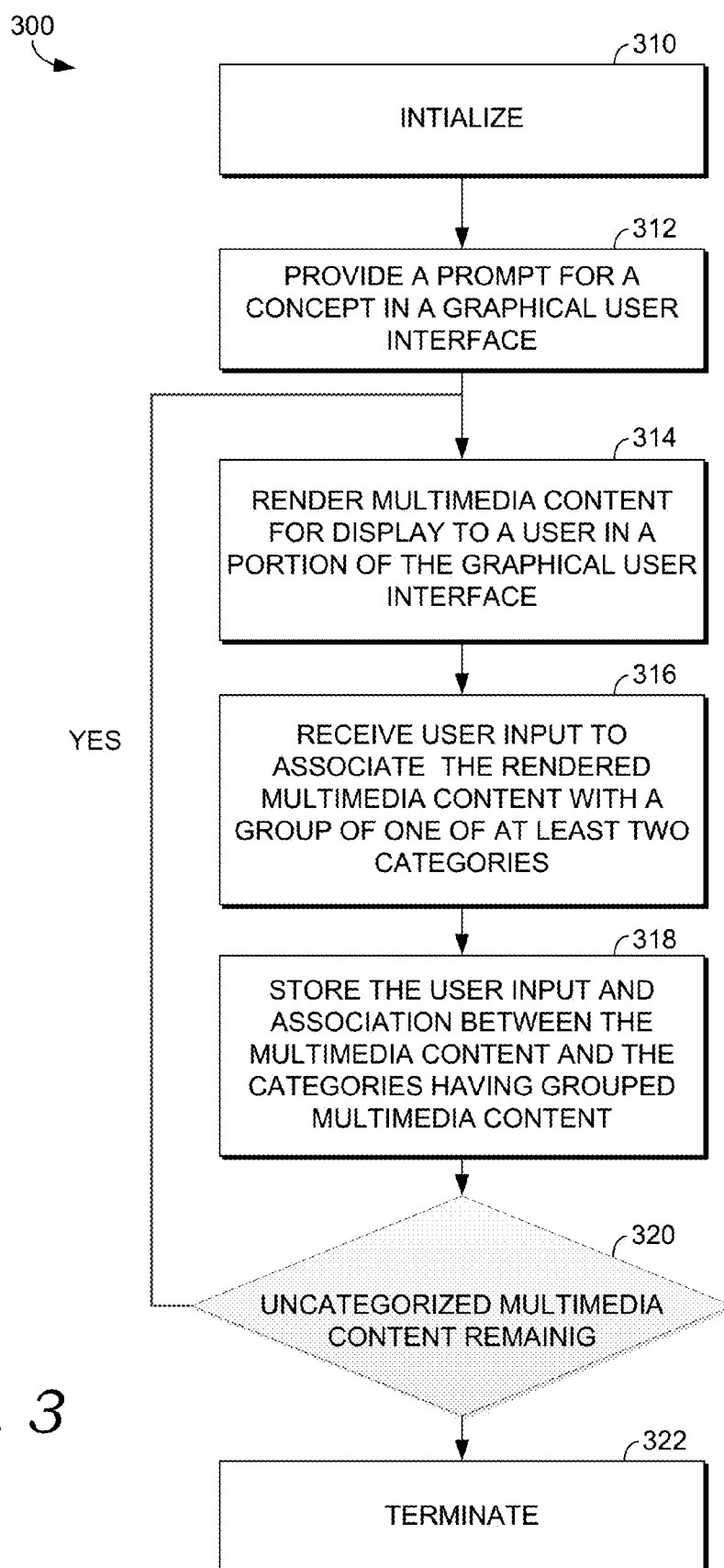


FIG. 3

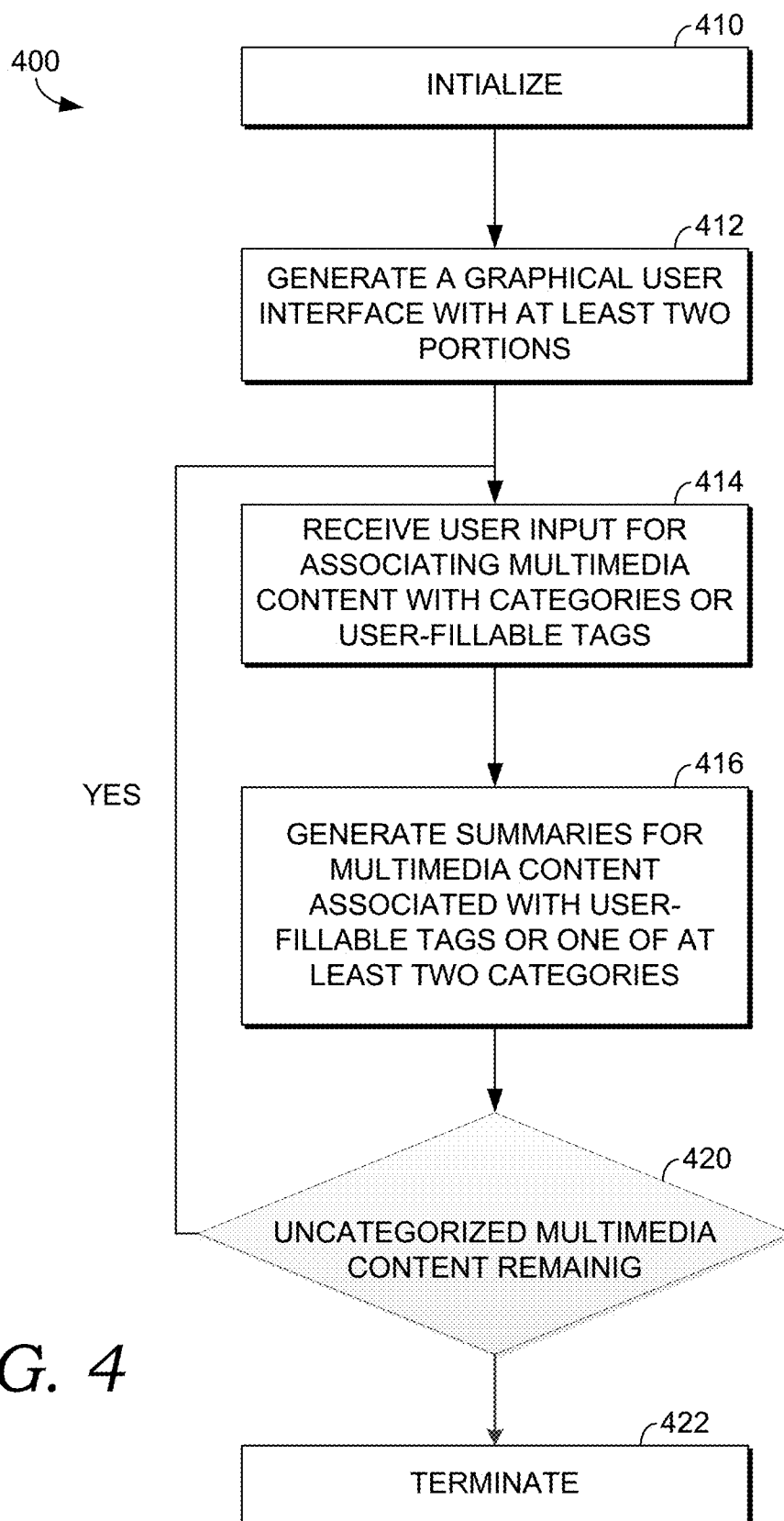


FIG. 4

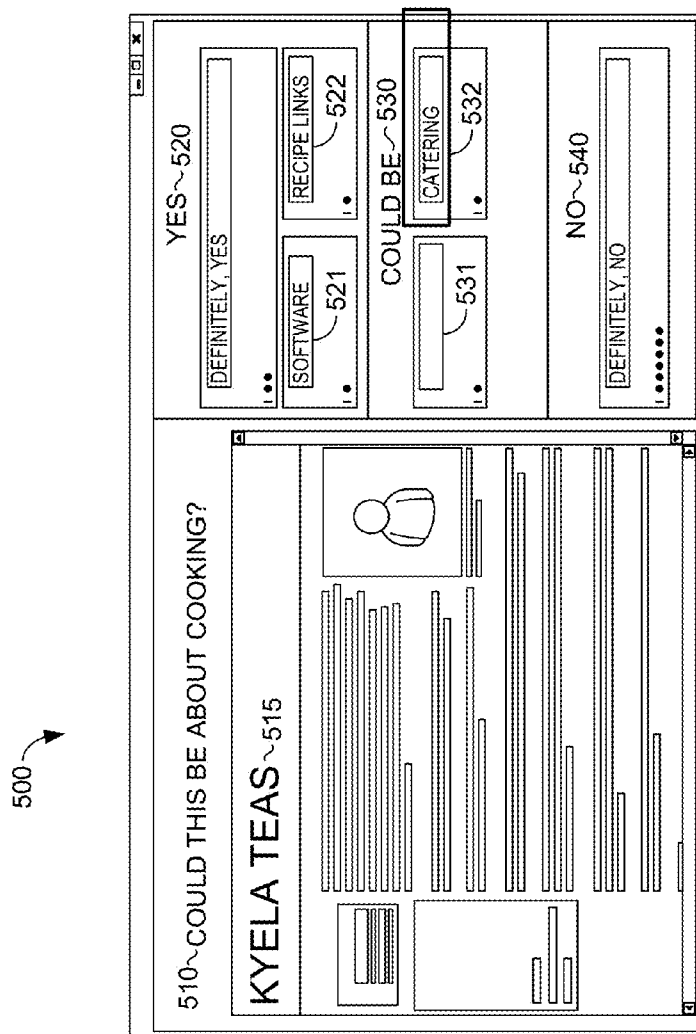


FIG. 5

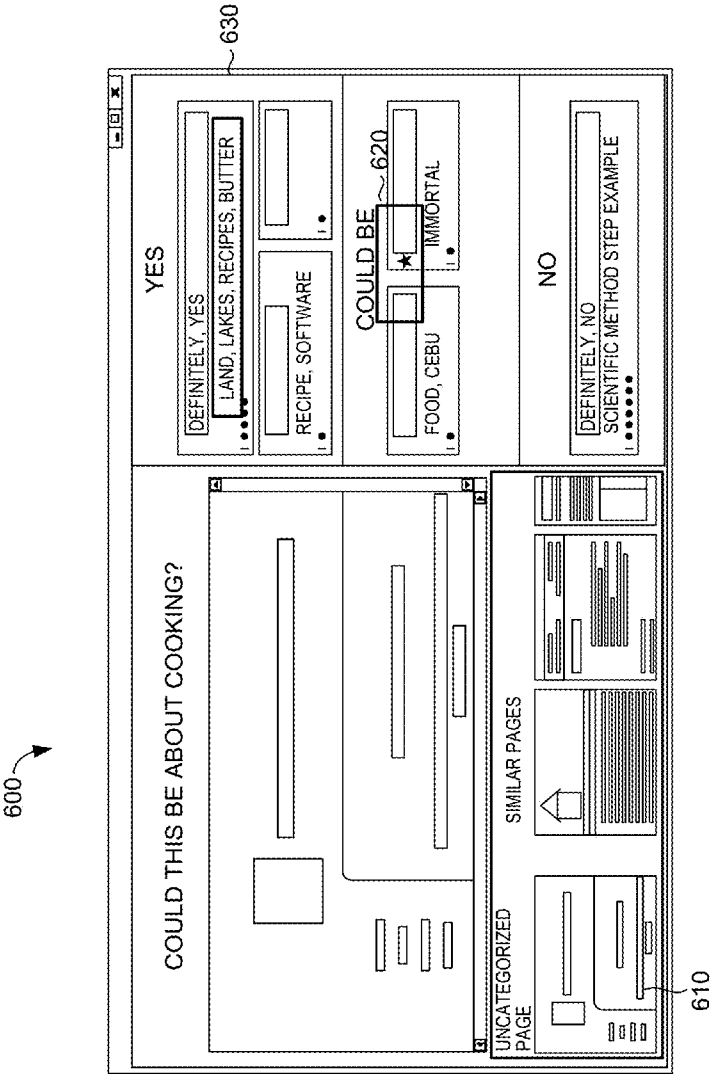


FIG. 6

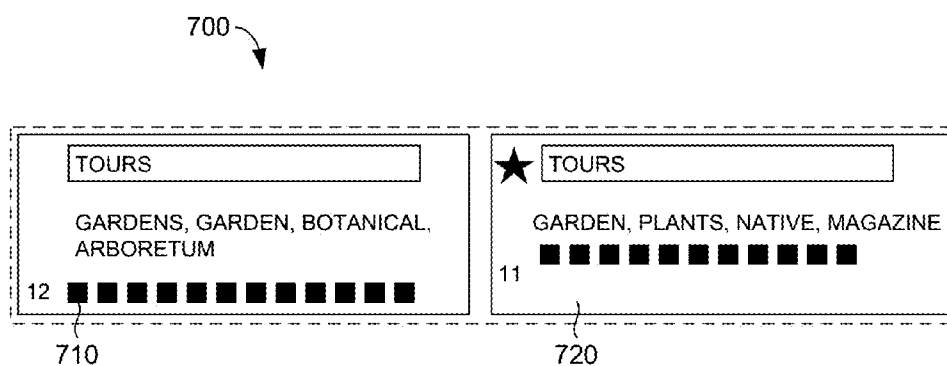


FIG. 7

STRUCTURED LABELING TO FACILITATE CONCEPT EVOLUTION IN MACHINE LEARNING

BACKGROUND

[0001] Conventional labeling technology allows users to label content. The labels are used by a computer to differentiate among different types of content. The labeling technology may assign content to different groups. However, users may label content differently at different times, and different users may label data differently. This uncertainty in label application may impact labeling consistency and label quality.

[0002] For instance, conventional spam filters employ labeling technology to identify spam messages. The performance of the conventional spam filters depend directly on label quality and label consistency. Spam filters may be trained from a large corpus of content (e.g. emails or web pages) labeled as spam or not spam. Poorly trained spam filters may admit unwanted spam or, worse yet, incorrectly classify important content as spam.

[0003] Improvements in label quality or label consistency may yield superior performance in spam filtering, product recommendation, prioritization, etc. Label quality is affected by factors such as the labeler's expertise or familiarity with the concept or data, the labeler's judgment ability and attentiveness during labeling, and the ambiguity and changing distribution of the content. The label quality may be particularly important in situations where data quantity is limited (e.g., when labels are expensive to obtain or when individuals are labeling data for their own purposes).

[0004] To improve label quality and consistency, label noise and concept drift should be managed. The label noise may be identified when several different labels are applied to the same content. Concept drift may be identified by when quickly changing content requires several different labels. The label noise and concept drift are managed by technologies that provide set-based label judgments and temporally applied labels (e.g., by discarding or weighting information according to a moving window that changes as the underlying content changes.)

SUMMARY

[0005] Embodiments of the invention relate to systems, methods, and computer-readable media for, among other things, generating a graphical user interface that structures labeling of multimedia content. The structured labeling of multimedia content via the graphical user interface may permit concept evolution as labels, categories, or user-supplied tags are applied to the multimedia content. The structured labeling allows the user to categorize multimedia content with an existing schema (e.g., 'YES', 'NO', 'COULD BE', etc.) The structured labeling graphical user interface may, in at least one embodiment, allow a labeler to postpone labeling decisions.

[0006] In one embodiment, a structured labeling graphical user interface having several portions is generated by a computer. A first portion of the structured labeling graphical user interface is configured to display multimedia content. A second portion of the structured labeling graphical user interface is configured to assign the multimedia content to one of at least two categories. In turn, a user provides input to the graphical user interface. The user input may associate multi-

media content displayed in the first portion with one of the at least two categories in the second portion of the graphical user interface. In an alternative embodiment, the user input may associate multimedia content displayed in the first portion with at least one group of the at least two categories. In certain embodiments, the user input may be, among other things, menu selections or drag-and-drop commands. For instance, hovering over the multimedia content with a selector, clicking on the multimedia content, dragging the multimedia content to one of the at least two categories, and dropping the multimedia content.

[0007] In some embodiments, the structured labeling graphical user interface is updated with summaries. The summaries are generated for the multimedia content associated with one or more groups. The groups may have, in one embodiment, user-supplied tags. In other embodiments, the summaries are generated for groups without user-supplied tags. Additionally, the structured labeling graphical user interface may provide recommendation for removing or adding multimedia content to the groups.

[0008] The user-supplied tags may correspond to the grouped multimedia content in at least one category. The user-supplied tags may receive user input that further define a concept corresponding to the multimedia content associated with one of the at least two categories or the user-supplied tags.

[0009] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The embodiments of the invention are described in detail below with reference to the attached drawing figures, wherein:

[0011] FIG. 1 is a block diagram illustrating an exemplary computing environment suitable for implementing embodiments of the invention;

[0012] FIG. 2 is a network diagram illustrating a network environment suitable for implementing embodiments of the invention;

[0013] FIG. 3 is a logic diagram illustrating a computer-implemented method of multimedia content labeling, in accordance with embodiments of the invention;

[0014] FIG. 4 is another logic diagram illustrating a computer-implemented method of generating a structured labeling graphical user interface, in accordance with embodiments of the invention;

[0015] FIG. 5 is a screen shot illustrating an exemplary graphical user for structured labeling, in accordance with embodiments of the invention;

[0016] FIG. 6 is a screen shot illustrating another exemplary graphical user for assisted structured labeling, in accordance with embodiments of the invention; and

[0017] FIG. 7 is a screen shot illustrating an exemplary user-supplied tag with visual representations for associated multimedia content, in accordance with embodiments of the invention.

DETAILED DESCRIPTION

[0018] The subject matter of this patent is described with specificity herein to meet statutory requirements. However, the description itself is not intended to necessarily limit the scope of the claims. Rather, the claimed subject matter might be embodied in other ways to include different steps or combinations of steps similar to the ones described in this document, in conjunction with other present or future technologies. Although the terms “step,” “block,” or “component,” etc., might be used herein to connote different components of methods or systems employed, the terms should not be interpreted as implying any particular order among or between various steps herein disclosed unless and except when the order of individual steps is explicitly described.

[0019] As utilized herein, “concept evolution” refers to defining and refining a target concept by a labeler such that different labels are applied to similar items due to changes in the labeler’s notion of the underlying concept over time. That is, similar items may be labeled differently not because the user perceives the items to be different from one another, but because their understanding of which items match their concept has evolved since labeling the earlier items. Noting this evolution may reduce inconsistency in labeling similar items when concept evolution is properly considered during label assignment.

[0020] As utilized herein, “category” is selected from a high level schema that includes at least two labels. In one embodiment, the labels may be “YES,” “NO,” and “COULD BE,” which are predefined.

[0021] As utilized herein, “groups” are collection of multimedia content that are associated with a category. Each category may have one or more groups in some embodiments. The groups are malleable. A group may be moved between categories. The groups may be merged, deleted, or edited.

[0022] As utilized herein, “user-supplied tags” are descriptions selected or defined by a labeler. These tags, in one embodiment, are temporary and modifiable. In at least one embodiment, the tags are applied to groups of multimedia content within at least one category. The tags, in additional embodiments, are displayed within at least two predefined categories selected from ‘YES,’ ‘NO,’ or ‘COULD BE.’ The tags could create a hierarchy within the predefined categories. In other words, tags may be nested (e.g. tag “B” may be nested in tag “1”).

[0023] Embodiments of the invention are directed to, among other things, multimedia content labeling. A server is configured to provide a prompt for a concept in a graphical user interface. The graphical user interface includes at least two categories for the prompt and one or more user-supplied tags for groups within each of the two categories. The server renders the multimedia content for display to a user in a portion of the graphical user interface. In turn, user input in the at least two categories or corresponding user-supplied tags is received. The user input, in an embodiment of the invention, may associate one of the at least two categories with the rendered multimedia content. The rendered multimedia content may be added to one or more groups corresponding to the at least two categories. Additionally, the user input may describe the grouped multimedia content with user-supplied tags. The server may store the user input, categories, and association between the multimedia content and the categories, groups, or user-supplied tags in a database.

[0024] The structured labeling graphical user interface assists labelers as they define and refine concepts for multi-

media content. This structured labeling graphical user interface allows people to organize their concept definition by grouping and tagging data (e.g., ‘YES,’ ‘NO,’ and ‘COULD BE’). The labeler’s interaction with the structured labeling graphical user interface is logged to gain insights into grouping multimedia content. Based on the logged data, the server may identify group-specific features, generate one or more recommendations for grouping multimedia content, and assign weights to group structures. The structured labels may be utilized to build machine learners or evaluate algorithms.

[0025] Having briefly described an overview of embodiments of the invention, an exemplary operating environment in which embodiments of the invention may be implemented is described below to provide a general context for various aspects of these embodiments.

[0026] FIG. 1 is a block diagram illustrating an exemplary computing environment suitable for implementing embodiments of the invention. Referring to the figures in general and initially to FIG. 1 in particular and computing device 100. The computing device 100 is but one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of embodiments of the invention. Neither should the computing device 100 be interpreted as having any dependency or requirement relating to any one component, or combination of components, illustrated.

[0027] The embodiments of the invention may be described in the general context of computer code or machine-useable instructions, including computer-executable instructions. These instructions may include program components being executed by a computer or other machine (e.g., a personal data assistant or other handheld device). Generally, program components, including routines, programs, applications objects, components, data structures, and the like, refer to code that performs particular tasks or implements particular abstract data types. Embodiments of the invention may be practiced in a variety of system configurations, including handheld devices, tablet computers, gaming devices, consumer electronics, general-purpose computers, specialty computing devices, etc. Embodiments of the invention may also be practiced in distributed computing environments, where tasks are performed by remote-processing devices that are linked through a communications network.

[0028] As one skilled in the art will appreciate, the computing device 100 may include hardware, firmware, software, or a combination of hardware and software. The hardware includes processors and memories configured to execute instructions stored in the memories. The logic associated with the instructions may be implemented, in whole or in part, directly in hardware logic. For example, and without limitation, illustrative types of hardware logic include field programmable gate array (FPGA), application-specific integrated circuit (ASIC), system-on-a-chip (SOC), or complex programmable logic devices (CPLDs). The hardware logic allows a device to generate a structured labeling graphical user interface for a user. The device is configured to receive user input at the graphical user interface and to log the user interaction with the graphical user interface. The device may recommend groupings or summaries based on the user interaction that associated multimedia content rendered in the graphical user interface with one or more categories. The device may, in an embodiment, display statistics associated with groupings and categories in the graphical user interface.

[0029] With continued reference to FIG. 1, computing device 100 includes a bus 110 that directly or indirectly couples the following devices: memory 112, one or more processors 114, one or more presentation components 116, input/output (I/O) ports 118, I/O components 120, and an illustrative power supply 122. Bus 110 represents what may be one or more busses (such as an address bus, data bus, or combination thereof). Although the various blocks of FIG. 1 are shown with lines for the sake of clarity, in reality, delineating various components is not so clear and, metaphorically, the lines would more accurately be grey and fuzzy. For example, one may consider a presentation component, such as a display device, to be an I/O component. Also, processors have memory. The inventors hereof recognize that such is the nature of the art and reiterate that the diagram of FIG. 1 is merely illustrative of an exemplary computing device that can be used in connection with one or more embodiments of the invention. Distinction is not made between such categories as “workstation,” “server,” “laptop,” “handheld device,” etc., as all are contemplated within the scope of FIG. 1 and refers to “computer” or “computing device.”

[0030] Computing device 100 typically includes a variety of computer-readable media. Computer-readable media can be any available media that is accessible by computing device 100 and includes both volatile and nonvolatile media, removable and non-removable media. Computer-readable media may comprise computer storage media and communication media.

[0031] Computer storage media includes volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, Random Access Memory (RAM), Read Only Memory (ROM), Electronically Erasable Programmable Read Only Memory (EEPROM), flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other holographic memory, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to encode desired data and that can be accessed by the computing device 100. In an embodiment, the computer storage media can be selected from tangible computer storage media like flash memory. These memory technologies can store data momentarily, temporarily, or permanently. Computer storage does not include and excludes communication media.

[0032] On the other hand, communication media typically embodies computer-readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term “modulated data signal” means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media, such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media.

[0033] Memory 112 includes computer-storage media in the form of volatile and/or nonvolatile memory. The memory may be removable, nonremovable, or a combination thereof. Exemplary hardware devices include solid-state memory, hard drives, optical-disc drives, etc. Computing device 100 includes one or more processors that read data from various

entities such as memory 112 or I/O components 120. Presentation component(s) 116 present data indications to a user or other device. Exemplary presentation components 116 include a display device, speaker, printing component, vibrating component, etc. I/O ports 118 allow computing device 100 to be logically coupled to other devices including I/O components 120, some of which may be built in. Illustrative I/O components include a microphone, joystick, game pad, satellite dish, scanner, printer, wireless device, a controller (such as a stylus, keyboard, and mouse) or a natural user interface (NUI), etc.

[0034] The NUI processes gestures (e.g., hand, face, body, etc.), voice, or other physiological inputs generated by a user. These inputs may be interpreted as terms for user-supplied tags, requests for associating multimedia content with categories, or requests for modifying groups having the multimedia content. The input of the NUI may be transmitted to the appropriate network elements for further processing. The NUI implements any combination of speech recognition, touch and stylus recognition, facial recognition, biometric recognition, gesture recognition both on screen and adjacent to the screen, air gestures, head and eye tracking, and touch recognition associated with displays on the computing device 100. The computing device 100 may be equipped with depth cameras, such as stereoscopic camera systems, infrared camera systems, RGB camera systems, and combinations of these, for gesture detection and recognition. Additionally, the computing device 100 may be equipped with accelerometers or gyroscopes that enable detection of motion. The output of the accelerometers or gyroscopes is provided to the display of the computing device 100 to render immersive augmented reality or virtual reality.

[0035] As previously mentioned, embodiments of the invention are generally directed to systems, methods, and computer-readable storage media for, among other things, generating a structured labeling graphical user interface configured to render multimedia content and associate the multimedia content with categories. At least a portion of the graphical user interface provides user-supplied tags that may be associated with the rendered multimedia content. The user-supplied tags may describe groups of multimedia content within the categories of the structured labeling graphical user interface. In some embodiments, the groups may be nested within each other. Accordingly, Group A may be nested in Group 1. This nesting may create a hierarchy for organizing the labeled multimedia content.

[0036] Various aspects of the technology described herein are generally employed in computer systems, computer-implemented methods, and computer-readable storage media for, among other things, associating the multimedia content with user-supplied tags and categories in a structured labeling graphical user interface. In one embodiment, a server executes suggestion components that detect associations between the user-supplied tags and the grouped multimedia content and provide grouping suggestions for uncategorized multimedia content or suggestions for modifying groups of categorized multimedia content. The server may display the suggestions to the user, who may decide to accept the suggestions or to decline the suggestions. The server may log user interaction with the structured labeling graphical user interface to gather metrics (e.g., dwell time, click frequency, revision per group, revision per user-supplied tag, and number of multimedia content per group, etc.) for storage in a data store.

[0037] In one embodiment, a computer system is configured to label multimedia content. The computer system includes, among other components, a grouping and tagging component and a summary generation component. The grouping and tagging component receives the user input, associates multimedia content with one or more categories or user-supplied tags based on the user input, and updates the user-supplied tags from terms provided in the user input. The summary generation component provides summaries for the multimedia content associated with the groups, one or more categories, or user-supplied tags. The summaries may be used by the user to modify the user-supplied tags or the groups having the multimedia content. The user may associate uncategorized multimedia content with the groups, categories, or the user-supplied tags based on the generated summaries.

[0038] FIG. 2 is a network diagram illustrating a network environment suitable for implementing embodiments of the invention. The computing system 200 may include client device 210, server 220, data store 230, and network 240. The network 240 may communicatively connect the client device 210, server 220, and data store 230. It should be understood that any number of client computing devices 210, servers 220 and data stores 230 may be employed in the computing system 200 within the scope of embodiments of the invention. Each may comprise a single device/interface or multiple devices/interfaces cooperating in a distributed environment. For instance, the server 220 may comprise multiple devices and/or modules arranged in a distributed environment that collectively provide the functionality of the server 220 described herein. Additionally, other components/modules not shown also may be included within the computing system 200.

[0039] In some embodiments, one or more of the illustrated components/modules may be implemented as stand-alone applications. In other embodiments, one or more of the illustrated components/modules may be implemented via the client device 210, as an Internet-based service, or as a module inside the server 220. It will be understood by those of ordinary skill in the art that the components/modules illustrated in FIG. 2 are exemplary in nature and in number and should not be construed as limiting. Any number of components/modules may be employed to achieve the desired functionality within the scope of embodiments hereof. Further, components/modules may be located on any number of search engines or user computing devices. By way of example only, the server 220 might be provided as a single server (as shown), a cluster of servers, or a computing device remote from one or more of the remaining components.

[0040] The client device 210 may be used to input one or more user-supplied tags and to associate multimedia content with categories or grouped multimedia content with user-supplied tags via a structured labeling graphical user interface. The client device 210 may communicate the user input received at the structured labeling graphical user interface to the server 220. In an embodiment, the client device 210 may include any type of computing device, such as the computing device 100 described with reference to FIG. 1, for example.

[0041] Generally, the client device 210 includes a display 211 and a browser 212. The display 211 is configured to present various content including, without limitation, a structured labeling graphical user interface having, among other things, a first portion for rendering multimedia content and a second portion for categories and user-supplied tags that may be associated with the multimedia content as described

herein. The structured labeling graphical user interface is designed to increase label consistency by assisting users by explicitly surfacing and recalling labeling decisions. Further, the structured labeling graphical user interface enables users to create, delete, split, and merge groups. This allows users to frequently refine their concept definition as they observe additional portions of multimedia content or other types of multimedia content. The graphical user interface, in certain embodiments, may provide assisted structured labeling. The assisted structured labeling is presented to the user as visual aids with a configurable level of automation to further assist users as they label multimedia content and while their understanding of the concept or multimedia content evolves.

[0042] The structured labeling graphical use interface, in one embodiment, may display one multimedia content at a time. In other embodiments, the graphical user interface may display several multimedia content at the same time. The graphical user interface may also prompt users to categorize the multimedia content into one of three high-level categories: 'YES', 'NO', or 'COULD BE' for a current concept prompt (e.g. "CARS"). For instance the prompt may be "WHETHER THE CURRENT MULTIMEDIA CONTENT IS AN EXAMPLE OF CONCEPT X." In addition, participants could create via the structured labeling graphical user interface user-supplied tags for groups within the categories. For a 'COULD BE' category the user may create a user-supplied tag as a reminder of the group of multimedia content that is associated with the current user-supplied tag. The user-supplied tag may include a description. In some embodiments, the user-supplied tag lacks a description and may be user a placeholder to display one or more suggestions generated by the summary generation component 224. The structured labeling graphical user interface organizes the multimedia content with mutually exclusive categories such as 'YES', 'NO', and 'COULD BE') and provides grouping and tagging of the multimedia content within the categories. The users may drag multimedia content from a rendering portion of the structured labeling graphical user interface to a labeling area. The user may drop the multimedia content over an existing user-supplied tag or over a category to create or update groupings of the multimedia content. The user may drop the multimedia content over a visual indicator to create a new user-supplied tag. In embodiments, the display 220 is further configured to enable touch or gesture inputs from a user or provide a NUI.

[0043] The browser 212 is configured to render multimedia content, for instance, web pages, video files, audio files, etc., in association with the display 211 of the client computing device 210. The browser 212 is further configured to receive user input for refining or defining the user-supplied tags, updating groupings of the multimedia content (generally inputted via a graphical user interface or NUI) and to receive multimedia content for presentation on the display 211, for instance, from the data store 230. The browser 212 may be any suitable type of web browser such as INTERNET EXPLORER®, FIREFOX®, CHROME®, SAFARI®, or other type of software configured to enable structured labeling of multimedia content as described herein. It should be noted that the functionality described herein as being performed by the browser 212 may be performed by any other application capable of rendering multimedia content. Any and all such variations, and any combination thereof, are contemplated to be within the scope of embodiments of the invention.

[0044] The server 220 is configured to receive user input from the client devices 210, provide group suggestions or summaries to assist the user, and log user interaction with a structured labeling graphical user interface transmitted to the client device 210. The server may implement any combination of the following components to process the user input: a grouping and tagging component 221, a logging component 222, a suggestion component 223, a summary generation component 224, and a display component 225. In one embodiment, the components of the server 220 may be executed locally by the client device 210 to process the multimedia content stored in the data store 230.

[0045] The grouping and tagging component 221, in one embodiment, receives user input from the client device 210. The user input may associate multimedia content rendered in the structured labeling interface with one or more categories. The one or more categories may be predefined in at least one embodiment. In other embodiments, the one or more categories may be configured based on user input in the graphical user interface. The user input, in some embodiments, includes user-supplied tags having terms that describe multimedia content grouped within a category. The grouping and tagging component 221 may group the multimedia content within categories based on the user input. The user-supplied tags, in at least one embodiment, are rendered with visual representation of the grouped multimedia content within the categories in the structured labeling graphical user interface. The terms received from the user are displayed in the user-supplied tags and a visual indicator representing multimedia content associated with the user-supplied tag may be updated to reflect an association between the grouped or rendered multimedia content and the user-supplied tag. In other embodiments, the user-supplied tags may be selected in the graphical user interface. For instance, a user may right-click on the multimedia content to select terms that describe the concept corresponding to the multimedia content. In additional embodiments, the grouping and tagging component 221 may receive user input that drags the user-supplied tags or categories over or near the multimedia content and drops the user-supplied tags or categories onto the multimedia content to add the multimedia content to the group having the user-supplied tag or category.

[0046] In certain embodiments, the logging component 222 receives the user inputs and logs the user inputs and other metadata associated with the structured labeling graphical user interface in the data store 230. The logging component 222 may log the number of multimedia content associated with the categories. In one embodiment, the logging component 222 may log the number of multimedia content associated with each group, including groups with user-supplied tags in the data store 230. The data store 230 may store the associations between the multimedia content and the categories and user-supplied tags. The data store 230 may receive log data from the logging component 222. The log data may include, among other things, the number of revisions to the groups, the length of time that transpired before a user associated multimedia content with a category or user-supplied tag, whether the user-supplied tag is completed by the user, whether the user-supplied tag is complete based on a summary provided by the summary generation component, the number of revisions to the user-supplied tags, and the length of time it takes to classify all multimedia content stored in the data store 230. The structured labeling graphical user interface allows the user to preform edits by moving user-supplied

tags between categories (e.g., preserving any accompanying tags and associated multimedia content), merging user-supplied tags, or moving individual multimedia content between user-supplied tags.

[0047] The suggestion component 223 receives the log data and generates suggestions for grouping multimedia content. The suggestion component 223, in one embodiment, provides suggestions for grouping multimedia content. For instance, the suggestion component may provide suggestions for modifying the groups of multimedia content that are associated with the categories or the user-supplied tags. Additionally, the suggestion component 223 may provide suggestions for associating the uncategorized multimedia content with the categories or the groups having the user-supplied tags.

[0048] In certain embodiments, the suggestions from the suggestion component 223 are computed based on similarity between an uncategorized multimedia content and a group of multimedia content associated with the categories or the user-supplied tags. The group of multimedia content that is most similar to the uncategorized multimedia content is suggested as a potential association for the uncategorized multimedia content. The suggestion component 223 computes item-to-group similarity as the similarity between the uncategorized multimedia content item and the most similar multimedia content in the group of the multimedia content of each user-supplied tag or category. For instance, the suggestion component 223 may compute similarity between the uncategorized multimedia content and all multimedia content associated with each of each group of multimedia content with a user-supplied tag or category. In turn, the suggestion component 223 selects the 'shortest-link' as the similarity value. The suggestion component 223 computes item-to-item similarity via the cosine similarity metric. In other embodiments, a term frequency-inverse document frequency (TF-IDF) may be used to calculate similarity. The recommendations for grouping may be displayed on the structured labeling graphical user interface as a 'wobble' animation on the group to draw the user's attention or a static indicator visible within the recommended group. In some embodiments, the suggestion component 223 may generate a similarity window. The suggestion component 223 may select the most similar unlabeled multimedia content to the multimedia content currently being labeled. The similar multimedia content, for display in the similarity window, may be identified using the same item-to-item similarity measure used to make group recommendations.

[0049] The summary generation component 224, in some embodiments, receives groups of multimedia content associated with each of the user-supplied tags. In an alternate embodiment, the summary generation component 224 may receive multimedia content corresponding to each category or one or more groups within each category. The summary generation component 224 generates summaries that may be included near the user-supplied tags. The groups may have, in one embodiment, user-supplied tags. In other embodiments, the summaries are generated for groups without user-supplied tags.

[0050] The summary generation component 224 may identify terms from the multimedia content associated with the user-supplied tags. The terms selected are best able to summarize the set of multimedia content associated with the user-supplied tag. The summary generation component 224 may analyze the document structure (e.g., heading, titles, metadata, and file type) to summarize the multimedia content.

In one embodiment, two bag-of-words algorithms are executed to create textual summaries. Initially, the summary generation component **224** considers the content of multimedia content associated with each user-supplied tag as a bag-of-words (i.e., the set of words from all the multimedia content associated with the user-supplied tag). In turn, the summary generation component **224** selects the most frequently occurring words from the bag, with frequency computed via the common term frequency inverse document frequency (TF-IDF). In an alternate embodiment, the summary generation component **224** accesses query logs of a search engine (not shown). The search terms associated with the multimedia content are extracted from the search log. The search terms include phrases that searchers used to locate that multimedia content via the search engine. Because these phrases are typically short and targeted, the summary generation component **224** may return the selected query terms as summaries. In one embodiment, the search query phrases with the highest TF-IDFs are displayed as summaries. These summaries may display the most prominent search terms used to find the web pages within each group and are updated in real-time as multimedia content changes.

[0051] In optional embodiments, the display component **225** is executed by the server **220** to render the user experiences for providing (i) entry of user input and (ii) viewing of multimedia content and associations between categories or user-supplied tags and the multimedia content. The display component **225** provides both user input entry and grouping of multimedia content user experiences. The viewing experiences on the structured labeling graphical user interface include providing visual representations of the multimedia content associated with the categories and user-supplied tags.

[0052] The display component **225** provides term highlighting as visual cues for suggested groupings of multimedia content in the graphical user interface. The visual cues may include icons such as “*” to assist users that are associating the multimedia content with the categories or user-supplied tags. The “*” may be rendered proximate to the suggested category or user-supplied tag for the multimedia content currently being considered by the user.

[0053] The data store **230** is accessed by the server **220** to select multimedia content for labeling. The data store **230**, in turn, stores search logs and log data from the user interaction with the structured labeling graphical user interface. The data store **230** may be a relational database that includes an index to content, including image, audio, video, text, webpages, etc. The data store **230** may also include a log that tracks statistics (e.g., decision time, click rate, term frequency, revision frequency, etc.) for each of the associated multimedia content, categories, and user-supplied tags. These statistics are sent to the suggestion component **223** (for creating group suggestions) or to the summary generation component **224** (for ranking of the summaries). The stored structure (e.g., categories and user-supplied tags) for each multimedia content may be shareable with other labelers that are considering labels for similar multimedia content or the same multimedia content. Thus, other labelers may have insight into how others have labeled the multimedia content.

[0054] The network **240** communicatively connects the search engine **220** and client device **210**. The network **240** may include, without limitation, one or more local area networks (LANs) and/or wide area networks (WANs). Such networking environments are commonplace in offices, enter-

prise-wide computer networks, intranets and the Internet. Accordingly, the network **240** is not further described herein.

[0055] The exemplary computing system **200** in which embodiments of the invention are employed is described above. Generally, the computing system **200** illustrates an environment in which group suggestions are provided to assist labelers as they associate multimedia content with user-supplied tags and categories. As will be described in further detail below, embodiments of the invention provide methods and graphical user interface elements for structured labeling of multimedia content. It should be understood that this and other arrangements described herein are set forth only as examples. Other arrangements and elements (e.g., machines, interfaces, functions, orders, and groupings of functions, etc.) can be used in addition to or instead of those shown, and some elements may be omitted altogether. Further, many of the elements described herein are functional components that may be implemented as discrete or distributed components or in conjunction with other components, and in any suitable combination and location. Various functions described herein as being performed by one or more entities may be carried out by hardware, firmware, and/or software. For instance, various functions may be carried out by a processor executing instructions stored in memory.

[0056] Accordingly, the structured labeling graphical user interface associates labels with multimedia content. Moreover, the structured labeling graphical user interface may provide assistance to labelers. The assisted structured labeling graphical user interface generates automated visual cues and recommendations that are explicitly surfaced to assist in recall of labeling decisions.

[0057] FIG. 3 is a logic diagram illustrating a computer-implemented method **300** of multimedia content labeling, in accordance with embodiments of the invention. The method initializes in step **310** on a server. The multimedia content may include a collection of webpages, audio files, or video files.

[0058] In step **312**, the server, optionally, provides a prompt for a concept in a graphical user interface. The graphical user interface, in one embodiment, includes at least two categories for the prompt. One or more user-supplied tags for groups within each of the two categories may be received via the graphical user interface. The graphical user interface may provide a viewable region in the graphical user interface with multimedia content similar to multimedia content associated with the at least two categories. In one embodiment, the at least two categories are selected from ‘YES’, ‘NO’, and ‘COULD BE’.

[0059] In turn, the server, in step **314**, renders multimedia content for display to a user in a portion of the graphical user interface. The graphical user interface receives, in step **316**, user input in the at least two categories to associate the multimedia content in the graphical user interface with one of the two categories. Additionally, the user input may correspond to user-supplied tags that describe the concept of multimedia content grouped within the at least two categories. The user-supplied tags are assigned to one or more groups of multimedia content. In some embodiments, the groups may be sub-groups of a group. All groups are associated with at least one of the at least two categories. In one embodiment, the user input is dragging and dropping a visual representation of the multimedia content onto or near one of the at least two categories or a group having the user-supplied tags. In certain embodiments, the user input may include removing, combining, or

adding user-supplied tags to the graphical user interface. In other embodiments, the user input may include removing, combining, or adding groups to the graphical user interface.

[0060] The server associates one of the at least two categories or the groups having the user-supplied tags with the rendered multimedia content. The server may update the graphical user interface with visual indicators. The visual indicators may represent a number of multimedia content associated with the groups of the at least two categories. These visual indicators may be displayed proximate to the at least two categories. In other embodiments, the visual indicators may be thumbnails of the multimedia content or icons associated with a file format of the multimedia content. In another embodiment, the visual indicators are numerical values. In yet another embodiment, the visual indicator reveals previews of additional content for the multimedia content in response to a hover command over the visual indicator.

[0061] The server, in step 318, stores the user input, categories, and association between the multimedia content and the categories or user-supplied tags in a database. The database may store a user interaction log. In one embodiment, the database stores a number of multimedia content associated with the at least two categories, and a number of multimedia content associated with each group having the user-supplied tags. The database may also store an amount of time that occurs between each input event received from the user. This database may also store a number of revisions made to user-supplied tags or a number of move, combine, add, or delete inputs received for each of the at least two categories.

[0062] In step 320, the server determines whether uncategorized multimedia content still exist. If uncategorized multimedia content exist, the server returns to step 314. If uncategorized multimedia content does not exist, the server continues to step 322. The method terminates in step 322.

[0063] Accordingly, the server provides a structure for hierarchically organizing and grouping multimedia content based on categories or user-supplied tags. The structured graphical user interface may provide improvements in label consistency by assisting labeler's to explicitly surface and recall their labeling decisions. The structured graphical user interface provides the user-supplied tags that receive metadata to aid in organization and recall. In turn, the structured graphical user interface may provide automated summaries of group multimedia content, suggestions for altering groups, and interactive revisions to group categories and user-supplied tags (e.g., moving groups within a category, adding and deleting user-supplied tags, merging and splitting user-supplied tags, adding or removing multimedia content within user-supplied tags, or revising user-supplied tags).

[0064] In additional embodiments of the invention, the structured graphical user interface may provide assistance during labeling. The structured graphical user interface renders various visual aids and some automation to assist labelers. For instance, visual aids may include, among other things, providing category recommendations (to improve label consistency or to reduce effort in tagging), providing tagging recommendations (to reduce effort in tagging), summarizing items within a user-supplied tags (to aid in recall of group contents), and rendering similar items or the number of similar items to the item being organized (to aid in structuring decisions).

[0065] FIG. 4 is another logic diagram illustrating a computer-implemented method 400 of generating a structured labeling graphical user interface label acquisition, in accordance

with embodiments of the invention. The method initializes in step 410 when loaded into memory by the server. In step 412, the server generates a graphical user interface with at least two portions. The first portion displays multimedia content. The second portion assigns the multimedia content to at least two categories. The server, in certain embodiments, may generate a third portion of the graphical user interface having uncategorized multimedia content that are similar to each other. The server may determine similarity based on term frequency within the multimedia content.

[0066] The server, in step 414, receives user input for associating multimedia content displayed in the first portion with at least one of two categories in the second portion of the graphical user interface. In some embodiments, the multimedia content is associated with at least one group of the at least two categories. In one embodiment, the user input is a gesture. In another embodiment, the user input is a voice command.

[0067] In turn, the server may, in step 416, generate assistance for the labeler. The assistance may include summaries. The summaries may correspond to multimedia content associated with the at least two categories. The summaries generated by the server may be based on multimedia content associated with at least one group within one of the at least two categories. Alternatively, the server may generate summaries for multimedia content associated with user-supplied tags that correspond to groups within each of the at least two categories. These generated summaries may be rendered in a visual representation of the user-supplied tags, in some embodiments of the invention.

[0068] The user-supplied tags may receive input that further describe a concept corresponding to the multimedia content grouped within one of the at least two categories or having the user-supplied tags. The graphical user interface is updated to reflect, among other things, the association between the selected one of the at least two categories and the multimedia content.

[0069] In step 420, the server determines whether uncategorized multimedia content exist. If uncategorized multimedia content exist, the server returns to step 414. If uncategorized multimedia content does not exist, the server continues to step 422. The method terminates in step 422.

[0070] Thus, the graphical user interface is updated to provide assistance via visual aids. These visual aids may include, among other things, providing category recommendations, providing tagging recommendations, summarizing multimedia content within a group of multimedia content having the user-supplied tags, and rendering similar multimedia content or the number of similar multimedia content to the multimedia content being organized.

[0071] In one embodiment, the structured labeling graphical user interface may include several portions. For instance, an exemplary structured labeling graphical user interface may be configured with a prompt portion, multimedia content portion, and category portion. The structured labeling graphical user interface may provide assistance to the user when the associated multimedia content rendered in the multimedia content portion is assigned to categories or groups of multimedia content having user-supplied tags rendered in the category portion.

[0072] FIG. 5 is a screen shot illustrating an exemplary graphical user 500 for structured labeling in accordance with embodiments of the invention. The structured labeling

graphical user interface **500** may include a prompt portion **510**, multimedia content portion **515**, and category portion **520**, **530**, or **540**.

[0073] The prompt portion **510** may provide a prompt for a concept under consideration by the labeler. For instance, the prompt may be about “COOKING.” In turn, the labeler may review multimedia content and assign the multimedia content to at least one category. The prompt portion **510** may be updated with a new concept after all the multimedia content is classified.

[0074] The multimedia content portion **515** may render the multimedia content. In one embodiment, the multimedia content is rendered individually (e.g., “one at a time”). The labeler may decide to associate the multimedia content with at least one category **520**, **530**, or **540**. The category **520**, **530**, or **540** may be predefined in an embodiment. In other embodiments, the category is alterable and may be selected based on user input.

[0075] The category portion **520**, **530**, or **540**, in an embodiment, provides three categories: “YES” **520**, “COULD BE” **530**, and “NO” **540**. The categories may each have user-supplied tags. For instance, category **520** may have two user-supplied tags (e.g., “SOFTWARE” **521** and RECIPE LINKS **522**). Category **530** may have two user-supplied tags of which one is blank (e.g., NULL **531** and CAETERING **532**). Category **540** may have zero user-supplied tags. The multimedia content may be associated with the categories **520**, **530**, or **540** or the groups having the user-supplied tags **521**, **522**, **531**, **532**.

[0076] In another embodiment, the structured labeling graphical user interface may include several visual indicators that provide assistance to the labeler. For instance, an exemplary structured labeling graphical user interface may be configured with a similar window visual indicator, summary visual indicator, and suggestion visual indicator. The visual indicators of the structured labeling graphical user may provide additional assistance to the user when associating multimedia content rendered in the multimedia content portion with categories or with groups having the user-supplied tags of the category portion.

[0077] FIG. 6 is a screen shot illustrating another exemplary graphical user **600** for assisted structured labeling in accordance with embodiments of the invention. The structured labeling graphical user interface **600** may provide visual indicators that include a similarity window visual indicator **610**, a suggestion visual indicator **620**, and a summaries visual indicator **630**.

[0078] The similarity window visual indicator **610** may display multimedia content that is similar to the multimedia content that is currently being categorized. In some embodiments, the similarity window visual indicator **610** may include thumbnails of the multimedia content that is currently being categorized and other multimedia content that is similar to it. In one embodiment, the other multimedia content that is identified as similar consist of uncategorized multimedia content.

[0079] The suggestion visual indicator **620** may provide a recommendation of which category to associate the currently being categorized multimedia content. The suggestion visual indicator **620** may include, among other things, a star. In other embodiments, the suggestion visual indicator **620** may be an animation that shows the multimedia content transitioning from the multimedia content portion of the graphical user

interface **600** to the appropriate location in the category portion of the graphical user interface **600**.

[0080] The summaries visual indicator **630** may include at least one or more terms that are generated to summarize multimedia content associated with the groups of multimedia content having user-supplied tags or the categories. The summaries visual indicator **630** may be highlighted in the graphical user interface. The highlighting may include font formatting (e.g., bold, italics, or color) or font animation (e.g., blinking, growing, shrinking, etc.)

[0081] Accordingly, these visual indicators rendered in the structured labeling graphical user interface are provided to improve user performance while labeling multimedia content. The structured labeling graphical user interface updates the visual indicators as the user associates multimedia content with the categories of groups having user-supplied tags. The server logs the interaction of the user with structured labeling graphical use interface to determine whether user performance is improved with the use of the visual aids.

[0082] In at least one embodiment, the structured labeling graphical user interface may render statistics associated with the multimedia content. For instance, the statistics may include the number of multimedia content associated each of the categories or the groups with the user-supplied tags. The statistics may be visualized as a number or an icon representing the multimedia content.

[0083] FIG. 7 is a screen shot illustrating an exemplary user-supplied tag **710** and **720** with visual representations for associated multimedia content in accordance with embodiments of the invention. The structured labeling graphical user interface may have a portion **700** for displaying statistics. The statistics may be rendered in the user-supplied tags **710** or **720** or categories. The statistics may include a numerical summary of the number of multimedia content associated with the group having the user-supplied tags **710** or **720**. For instance, user-supplied tag **710** is associated with a group of 12 multimedia content and user-supplied tag **720** is associated with a group of 11 multimedia content. The statistics, in certain embodiments, may be rendered as icons representing each multimedia content. The icons may include squares or a graphic representing the file type associated with each of the multimedia content associated with the user-supplied tag or categories.

[0084] Additional information about the structured labeling graphical user interface may be found in the paper published by the inventors, Kulesza et. al., “Structured Labeling to Facilitate Concept Evolution in Machine Learning” CHI 2014 Association for Computing Machine, which is hereby incorporated by reference in its entirety.

[0085] With the benefit of the structured labeling graphical user interface a user may structure subcomponents of a concept in an open-ended manner. For instance, a user could structure the concept “AUTOS” into subcomponents like “CARS,” “TRUCKS,” “ADVERTISEMENTS,” “REVIEWS,” etc. Regardless of the subcomponent defined by the user, the structured labeling graphical user interface provides assistance to ensure that the user consistently organizes items into the defined subcomponents.

[0086] In summary, embodiments of the invention provide for automated or semi-automated grouping of data (to reduce user effort and errors). A server may generate structured labeling graphical user interfaces to assist labelers as they assign labels to multimedia content. The structured labeling graphical user interfaces may facilitate concept evolution.

The server may provide a prompt for a concept in a graphical user interface. The graphical user interface may include at least two categories for the prompt and one or more user-supplied tags for groups of multimedia content within each of the two categories. The server, in turn, renders multimedia content for display to a user in a portion of the graphical user interface. The server may receive user input in the at least two categories or groups corresponding to the user-supplied tags. The server may associate the multimedia content with the at least two categories and the user-supplied tags.

[0087] In at least one embodiment, a database associated with the server may store log data that includes, among other things, user input, categories, and association between the multimedia content and the categories. The database provides a mechanism for labelers to gather statistics about the labeling tasks, speed or number of multimedia content associated with each label. The database may also allow labelers to share structured labeling data for multimedia content. The labelers may share the structured labeling data as concept definitions to other users: These definitions may be used as guidelines for other labelers (e.g., as guidelines that teach by showing examples). In alternate embodiments, sharing structured labeling data may enable multiple users to collectively define concepts and associate the multimedia content with the concepts. The database may provide corporate memory (i.e., allowing new users to continue labeling—assigning multimedia content to categories—where previous users left off).

[0088] The embodiments of the invention have been described in relation to particular embodiments, which are intended in all respects to be illustrative rather than restrictive. Alternative embodiments will become apparent to those of ordinary skill in the art to which the embodiments of invention pertain without departing from its scope. From the foregoing, it will be seen that this invention is one well adapted to attain all the ends and objects set forth above, together with other advantages which are obvious and inherent to the system and method. It will be understood that certain features and subcombinations are of utility and may be employed without reference to other features and subcombinations. This is contemplated by and is within the scope of the claims.

The technology claimed is:

1. A computer-implemented method for multimedia content labeling, the method comprising:

rendering multimedia content for display to a user in a portion of a graphical user interface; and
receiving user input to associate one of the at least two categories with the rendered multimedia content;
forming a group of multimedia content within at least one of the at least two categories; and
receiving user-supplied tags to describe a concept of the grouped multimedia content.

2. The method of claim 1, further comprising: providing a prompt for a concept in the graphical user interface, wherein the graphical user interface includes the at least two categories for the prompt such that the at least two categories are selected from yes, no, and could be.

3. The method of claim 1, wherein visual indicators represent a number of multimedia content associated with each group of the at least two categories and are displayed proximate to the at least two categories.

4. The method of claim 3, wherein the visual indicators are thumbnails of the multimedia content or icons.

5. The method of claim 3, wherein the visual indicators are numerical values.

6. The method of claim 3, wherein the visual indicator reveals previews of additional content for the multimedia content in response to a hover command over the visual indicator.

7. The method of claim 1, further comprising: providing a viewable region in the graphical user interface with multimedia content similar to multimedia content associated with the at least two categories.

8. The method of claim 1, wherein the user input is dragging and dropping a visual representation of the multimedia content onto or near a one of the at least two categories or user-supplied tags.

9. The method of claim 1, wherein the user input may include removing, combining, or adding user-supplied tags to the graphical user interface.

10. The method of claim 1, wherein the user input may include removing, combining, or adding groups within the at least two categories to the graphical user interface.

11. The method of claim 1, further comprising: storing the user input, categories, and association between the multimedia content and the categories in a database, wherein storing the categories in the database includes storing a number of multimedia content associated with the groups within the at least two categories and a number of multimedia content associated with the user-supplied tags.

12. The method of claim 11, wherein storing the user input in the database includes storing an amount of time that occurs between each input event received from the user.

13. The method of claim 11, wherein storing the user input includes storing a number of revisions made to user-supplied tags or a number of move, combine, add, or delete inputs received for each of the groups within the at least two categories.

14. The method of claim 1, wherein the user input is a gesture or voice command.

15. One or more computer storage media having computer-executable instructions embodied thereon, that when executed, perform a method of generating a structured labeling graphical user interface, the method comprising:

generating a graphical user interface with at least two portions, a first portion for displaying multimedia content and a second portion for assigning the multimedia content to one of at least two categories;

receiving user input for associating multimedia content displayed in the first portion with one of the at least two categories in the second portion of the graphical user interface; and

generating summaries for multimedia content associated with user-supplied tags that correspond to each of the at least two categories, wherein the user-supplied tags receive input that further describe a concept corresponding to the multimedia content associated with one of the at least two categories or the user-supplied tags.

16. The media of claim 15, further comprising: generating a third portion of the graphical user interface having uncategorized multimedia content that are similar to each other.

17. The media of claim 16, wherein generating summaries for multimedia content further comprises:

forming groups within the at least two categories; and
generating summaries for multimedia content associated with the groups within the at least two categories.

18. The media of claim **15**, wherein the graphical user interface is updated to reflect the association between the selected one of the at least two categories and the multimedia content.

19. The media of claim **15**, wherein the generated summaries are rendered in a visual representation of the user-supplied tags.

20. A computer system for structured multimedia content labeling that facilitates concept evolution, the system comprising:

a device configured to provide a user interface, wherein the user interface includes at least two categories and one or more user-supplied tags for each of the at least two categories, to render multimedia content for display to a user in a portion of the user interface; and to receive user input to associate the rendered multimedia content with the at least two categories or the user-supplied tags; and
a database configured to store the user input, categories, user-supplied tags, and association between the multimedia content and the categories or user-supplied tags.

* * * * *