December 13, 2017

Visual search using automatically annotated images

Alexander Haubold

Follow this and additional works at: http://www.tdcommons.org/dpubs_series

Visual search using automatically annotated images

ABSTRACT

Users interested in obtaining content related to certain concepts sometimes struggle with identifying suitable textual search terms. Users may be unable to formulate a query due to limitations of vocabulary, lack of language skills, lack of knowledge of the specific topic, etc. This disclosure describes techniques to enable a user to conduct a visual search and exploration using images with textual annotations that correspond to objects and concepts in the images. The textual annotations are analogous to hyperlinks, and are used by a search engine to obtain and display search results to the user. The search results can be text results or further curated images, e.g., diagrams or schematics annotated with terms and phrases that represent the objects and concepts shown. A user interface per the described techniques enables a user to continue to drill down on a search topic using visual cues in the search results.

KEYWORDS

- visual search
- image search
- object recognition
- image tags
- imagemap
- image annotation

BACKGROUND

Users interested in obtaining content related to certain concepts sometimes struggle with identifying textual search terms that correspond to the concepts. For example, users may be unable to formulate a query due to limitations of vocabulary, lack of language skills, lack of

knowledge of the specific topic, etc. Such situation can occur, e.g., when users search for content beyond their knowledge, such as complex scientific or technological concepts.

In some instances, users may not be able to formulate a close enough original entry point query to search for a problem that they are trying to solve, because they can only think of a high-level concept that is several levels removed from what they want to search for. In some scenarios, users may prefer exploring content related to a topic visually instead of having to formulate and enter text.

DESCRIPTION

The described techniques enable users to conduct online searches using images, schematics, pictures, or other diagrams ("images") instead of or in addition to providing textual queries. In some cases, the techniques include providing curated images that present contextually related concepts and objects visually. The curated images may be schematics and diagrams that are based on photographic or image material. The contents of these images used for search by a search engine are annotated with respective hyperlinks that include carefully curated and selected textual annotations that correspond to terms and concepts visually depicted in the images.

When the user selects (e.g., by clicking, tapping on a touchscreen, etc.) particular content in the image, the information in the associated hyperlink (e.g., text) with the respective content is used to conduct an online search. The search results are returned in the form of typical textual search results. Further, in some instances, the results also include new images that further visually characterize the text and concepts from the results for the user. This presentation of curated returned images in the results allows the user to continue to explore terms and concepts and to drill down on aspects of returned results without having to formulate a textual query.

Users may simply select from the newly returned images in the search results to continue their online exploration and search.

The hyperlinked annotations function effectively as translators from an image to a textual query. The search is independent of the language of the user or the user interface used for the search since the search engine used to conduct the search is capable of translating the annotations into a user-preferred language.
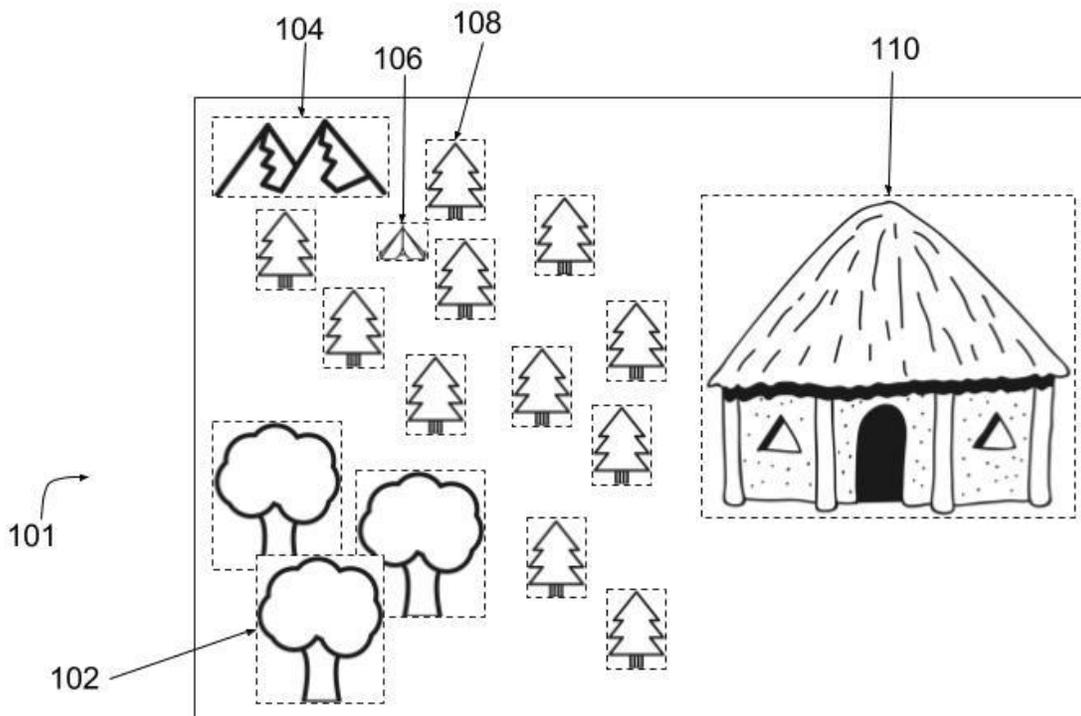

Fig. 1: Visual search using hyperlinked images

Fig. 1 illustrates an example using the described techniques of visual search. A user is presented with an image (e.g., while browsing online, received through a shared album, or message, uploads or captures, etc.) on a computing device. The image may be a schematic, picture, or other type of diagram (101) ("image"). In some cases, the image may be a diagram or schematic curated by a computing device based on photographic or visual material that is designed to visually present contextually related objects and concepts.

Content within a particular image is either manually or programmatically analyzed to identify different objects and concepts present within the image, e.g., as illustrated with labels 102-110. Respective hyperlinks are associated with those objects and concepts. In the example illustrated in Fig. 1, the recognized objects include a first type of tree (102), mountains (104), a tent (106), a second type of tree (108), and a structure (110). The techniques described herein pertain to manually or automatically creating textual annotations for each respective object (102-110) depicted in the image and words or phrases for concepts associated with the objects. When a user selects a particular object within the image, the textual annotations are used by a search engine to retrieve and display search results for information on those objects and concepts.

For example, when the user clicks on tent (106), the user is taken directly to search results for "camping in the woods," because the image of the tent has been annotated with the concept of "camping in the woods." The user does not have to specifically articulate the concept of "camping in the woods" in a textual query, because the image of the tent (106) has been previously annotated to include this concept.

In another example, consider a user that would like to explore more about the particular tree type shown by the tree labeled 108. Instead of having to know the name of the tree (e.g., "pine") and formulate a textual search using the name, the user can simply select the tree labeled 108. When the user clicks on the tree within image, the user is taken to search results for the specific tree species. For example, the user may be presented with typical textual search results on different types and aspects of pine trees and provided with images of different types and aspects of pine trees as well. For example, images may be returned of various pine trees, pine needles, pine cones, pine nuts, etc. These returned images are annotated with textual terms and

phrases corresponding to the species, objects, and concepts. Thus, the user is able continue their search and exploration of the tree (108) without having to formulate further textual inquiries.
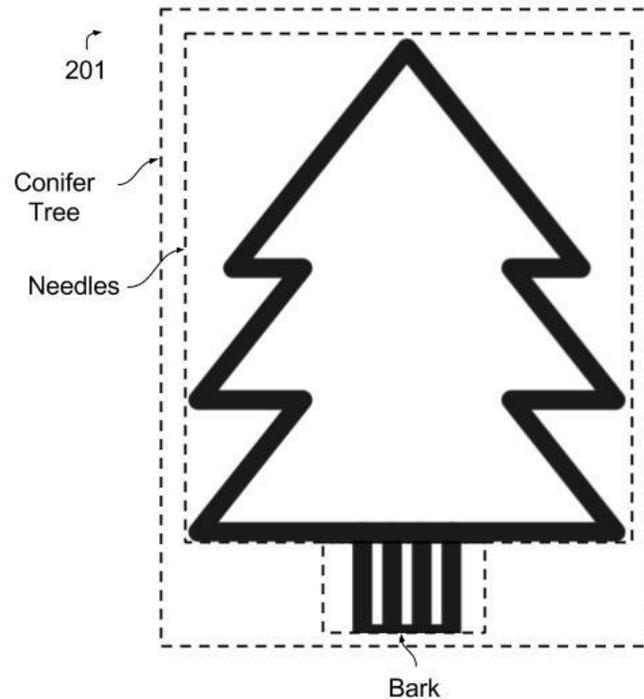

Fig. 2: Visual search using annotated and hyperlinked images

Fig. 2 illustrates an example where a user has selected the particular tree labeled 108 for further exploration. In response to the user selection, a detailed enlarged view of the tree labeled 108 showing further features or objects of the tree such as needles, bark, etc. is presented. This detailed enlarged view of the tree is also programmatically analyzed and curated (e.g., offline, prior to receiving the user selection) to provide annotations for objects shown within the image and concepts, such as the concept of type of tree ("conifer tree"), and objects that are concepts of different parts of the tree ("needles," "bark," etc.). These textual annotations may not be displayed to the user, similar to hyperlinks.

In this example, the user may want to search generally for the concept of items that are attached to the base ("bark") of the tree but may not know enough about this topic area to

formulate the textual queries for "needles," "leaves," "branches," etc. In this case, since the image has been annotated with those terms, the user can simply click on "needles" and is taken to a search results page. The search results page includes textual and visual results on the topic of needles, leaves, branches, etc. to allow the user to continue their understanding and exploration of these concepts and objects.

In another example, the user may conduct an online search that starts with an image of an exploded view of a car motor, that includes the engine block, connecting pipes, various containers for liquids, filters, etc. The engine block image is annotated with textual terms of "cylinders," "sealing rings," "screws," etc. When the user clicks on the engine, textual and visual search results are returned for objects and concepts relating to these concepts. The user may continue to drill down on aspects of the engine that they wish to further explore.

In another example, the user starts a search viewing an image of a forest scene depicting different types of trees, a lake, a creek, etc. The concepts are annotated with text so that when the user clicks on the lake the search engine uses the annotated textual terms to conduct a search and return results. The results include a diagram depicting the life in the lake: fish, algae, polluting plastic bottles, etc.

The terms curated for the objects in the image can be generic or highly contextual and specific. The terms can represent complex textual queries that may not be obvious to the user, but are well suited to the concept the user may be searching for. Generic terms can be used to annotate an image, such as "conifer", "bark", and "needles" of Fig. 2. More contextually specific terms are also possible. Referring again to example of the image of the lake that includes a plastic bottle, generic annotations may include "bottle" or "plastic bottle" and the image may include a more specific conceptual annotation, e.g., "plastic pollution in water bodies."

An implementing user interface can enable the user to select an image in such a way that the selection leads to more search results based on generic annotations or specific contextual annotations. For example, performing a default click (e.g., left click) on a plastic bottle in an image of a lake will return search results for specific annotated concepts such as "plastic pollution in water bodies." If the user wishes to search for generic annotations such as "plastic bottle," a different selection method (e.g., right click) on the plastic bottle can be used and in response, corresponding search results are presented.

Annotations for the curated images used for visual search can be customized based on user input. Otherwise unannotated cues presented in an image may be annotated based on search queries from users for those particular cues. Further, a process of upvoting and downvoting results obtained based on visual searches can be used to modify and create content for future visual searches.

The described techniques can be used in a general online search engine. The techniques can also be deployed in domain-specific search engines such that the visual cues and corresponding annotations are customized to the particular domain.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location

information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to enable a user to conduct a visual search and exploration using images with textual annotations that correspond to objects and concepts in the images. The textual annotations are analogous to hyperlinks, and are used by a search engine to obtain and display search results to the user. The search results can be text results or further curated images, e.g., diagrams or schematics annotated with terms and phrases that represent the objects and concepts shown. A user interface per the described techniques enables a user to continue to drill down on a search topic using visual cues in the search results.