

A Hybrid Method for Image Taxonomy: Using CAPTCHA for Collaborative Knowledge Acquisition

Bruno Norberto da Silva, Ana Cristina Bicharra Garcia

Computer Science Department
Fluminense Federal University
Passo da Patria, 156 – E-453, Niteroi, RJ 24210-240 Brazil
{bsilva,bicharra}@ic.uff.br

Abstract

We present a new method for image taxonomy. Our formulation shows how the combination of automated metadata-based analysis and knowledge-acquisition tools can help build better image search applications. We build on top of existing metadata-based techniques to image taxonomy and combine them to human assistance, in order to guarantee some minimal level of semantic soundness. Knowledge from humans is acquired thru their interaction with CAPTCHAs, which application is shown here in an innovative way. We believe this new application of CAPTCHA is a promising opportunity for knowledge acquisition because it takes advantage of a very special moment: since humans are willing to take a CAPTCHA test to access some resource, it might be possible to take advantage of this moment to extract meaning from human's response to the CAPTCHA test, in a collaborative fashion.

Introduction

The human impossibility of searching alone the enormous volume of information available on the Web has increased the researchers' interests for developing automated techniques to improve information retrieval and recall (Chakrabarti et al, 1999). One way of approaching this search problem is to impose a good semantic classification on the information, moving the difficulties from the retrieval to the inclusion process. There is even a greater challenge when dealing with visual data like pictures and movies. The main issue when browsing for images is the difficulty in identifying its content, thus compromising the precision of the retrieval.

Usually, two basic approaches are applied when dealing with image-content identification (Hyvönen et al, 2003):

1) an implicit-content approach: where images are characterized according to its physical properties like colors, texture and shape, or

2) a metadata-based approach: in which images are previously categorized according to a set of semantic primitives from which a retrieval can latter be executed.

Searching algorithms using implicit-content approaches, i.e. computer vision techniques, have still failed to perform well on generic domains. Consequently, current commercial image-searching applications, like Google's or AltaVista's image service (Google Image Search, AltaVista Image Search), make extensive use of metadata information like the textual content of web sites that contains an image or the file name by which an image is saved in a web server. Unfortunately, there are also drawbacks with the metadata approach. The most appealing drawback lies in the acquisition of image semantics. For instance, the image illustrated in Figure 1 was the first result returned from a search in Google's image database for the word 'car' (at the time this work was done). At first sight, this seems grotesque. However, when analyzed in the context of the entire Web document from which the picture was extracted, it was a reasonable retrieval, but still wrong for most users looking for images of a car.



Figure 1. The first search result from Google's image database for the word 'car'

Image characterization seems to be the issue here. While purely automated image characterization still presents inappropriate results, human categorization would be certainly more precise, but there are challenges to be

considered in this latter scenario: 1) the incentive that must be introduced to people to spend their time describing what they see in images, 2) the number of people that must be involved to accept a categorization and 3) how to reconcile the differences among multiple conflicting categorization for the same image.

In other words, we need:

- some sort of reward (monetary, psychological, or of any other kind) in compensation for the cognitive effort exerted by humans,
- an environment that is accessible to a wide range of people, in order not to make the knowledge captured (too) biased, and
- tolerance to discrepancies, since people with diverse backgrounds are expected to have different views of the same objects. Different descriptions for the same image not only must be expected, but should be encouraged.

A successful instance of a human-assisted image content identification can be found in (von Ahn and Dabbish, 2004), in which a framework is presented in the form of an entertainment game that attempts to incentive people to provide textual descriptions to images (Figure 2). In this game, two players are paired anonymously and are asked to type whatever they guess their anonymous partner is typing. The only thing they have in common is that the game presents both players with the same image. Whenever the two partners agree on a label, they raise their score and are faced with another image. Since they have no communication with their partners, the only reasonable strategy available for agreeing on a label is to provide something related to the image shown, and thus the submitted label might be a good description of the image presented.

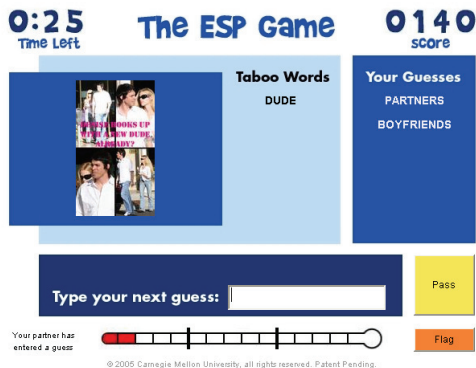


Figure 2. The ESP Game. A semi-automatic knowledge acquisition method is applied in the form of an entertainment game. Users are anonymously paired and encouraged to provide textual descriptions to images. If partners agree on a guess, their score is raised and the common description is saved as a description to the image.

While this approach provides sufficient rewards to achieve a large volume of human participation (the entertainment of the game experience) and supports conflicting descriptions for the same image, there are drawbacks associated with the use of entertainment as a means to incentive human participation, namely the fact that entertainment hardly ever attracts participation of people of diverse nature or with different backgrounds (which might make the acquired knowledge too biased and thus inappropriate for various domains), the need of protection against cheating of various kinds, among others.

In this paper, we present a hybrid method for building image taxonomy that improves image retrieval based on human assistance. Our method differs from previous approaches by integrating metadata-based image search techniques (Smith and Chang, 1997) with CAPTCHA's (von Ahn et al, 2004) model of interaction as an interface with human knowledge. This way, we expect to gather the advantages of both approaches: by allowing for automated ways of collecting information, we make the information acquisition process fast and reasonably efficient, and by incorporating human participation, we refine our knowledge base guaranteeing some level of semantic soundness.

Our method can be described in two phases. In the first stage, an automated search procedure navigates through the Web collecting random labels. Then, an image repository (like AltaVista Image Search) is queried using the previously collected labels. This search result will be used as the raw material for our CAPTCHA construction. Whenever a user tries to reach a page protected by our CAPTCHA, one label from our database is displayed and a random subset of images initially associated with the label is selected to generate a CAPTCHA test.

If the test-taker passes the test, the submitted answer is examined for the extraction of additional semantics to be added to our image ontology. If the CAPTCHA is carefully designed, this new information imbedded in the test's answer could be informative of the content of images present in the ontology.

It is worthy to note that the use of CAPTCHA presents an enormous opportunity for human-assisted knowledge acquisition because

- the incentive for inducing human participation is implied in CAPTCHA's very nature. Since CAPTCHAs are employed to protect some resource, one can reasonably assume that the agent trying to access the functionality is willing to take to test. And by combining CAPTCHAs with collaborative knowledge acquisition, we make use of a very special moment when designer's interests (who want to acquire knowledge) and user's interests (who want to access some resource) are directly aligned,
- since CAPTCHA's use has found applications all over the Web, the range of people that currently take CAPTCHA tests are enormous, thus enabling

any CAPTCHA-based technology to have a large potential audience,

- and, since CAPTCHA's definition presents no restriction on the nature of the test being applied, the test can be designed in order to achieve any desired knowledge acquisition.

In the next Session, we review the CAPTCHA concept, present the modeling of our ontology and describe our method of ontology-building for image retrieval with CAPTCHAs. Session 3 will contain a description of the experiments performed to validate our proposal. A general discussion is made in Session 4 and conclusions and directions for future works are presented in Session 5.

CAPTCHA, Ontology and Collaborative Knowledge Acquisition

CAPTCHA

CAPTCHA stands for Completely Automated Public Turing Test to Tell Computers and Humans Apart (von Ahn et al, 2004). It is an application that generates and grades tests that attempt to distinguish humans from computers. A CAPTCHA working procedure consists on the generation and submission of a test to some agent and the later evaluation of a received response. The agent's capacity of accurately solving the test should be indicative of its nature: whether this agent is either a human or a computational agent. In other words, a CAPTCHA test should pose little or no problem for a human to solve it, but must be impossible for current computers to pass. It has been successfully adopted in a wide range of applications running on the Web, from email accounts management (Yahoo! Mail) to SMS message sending (Oi Internet). One special CAPTCHA instance named Pix (Figure 3) is of special interest to us.



© 2004 Carnegie Mellon University, all rights reserved.

Figure 3. The Pix CAPTCHA. Pix maintains a collection of images in its database, each associated with a set of labels. For each test, Pix selects a label at random, selects 4 images associated with that label and includes them on the test. The agent is asked to identify which label applies to all 4 images.

Pix (The CAPTCHA Project) maintains a large collection of images, each associated with one or more textual labels. A test is generated by selecting a label at random, picking up four images related to that label, and then asking the user to indicate, from a list of all existing labels, which one relates to all four images.

In this CAPTCHA, it is assumed that it is not possible for a computer to identify a common aspect of all four images and then find this feature in the list. The CAPTCHA we have designed to demonstrate our method works roughly as a reverse Pix, and will be presented in further details in Session 3.

An implementation of Pix can be found in the CAPTCHA project home page (The CAPTCHA Project), as well as different instances of other CAPTCHAs.

An Ontology for Image Classification

In order to enable image classification, we model an ontology to better assist the process of image classification. Our modeling is deeply inspired by the semiotic relation between signifier and signified proposed by Saussure (Saussure, 1966). We initially consider two types of objects, Images and Labels, and allow an association between images and labels called Describes, which means that the label is a good indicator of the image content.

As an example, were the image in Figure 4 present in our ontology, one could reasonably expect associations with labels such as heart and love. Unexpected associations might include the labels war, car or bat.



Figure 4. Description for this image content could be heart, love, among others.

With this relation alone, it is possible to provide a number of functionalities for image searching. Additional relations (binary, ternary, etc) could be inserted in the ontology to enhance the expressive power of our method. We discuss this possibility in Section 4.

Knowledge Acquisition with CAPTCHA

Our method is constructed to work on top of image searching services like those extensively applied over the Web. It queries those services for label-image pairings and then presents these pairings to humans to verify or invalidate the association. In order to manage the correctness of an association, we add a Confidence Rank to

each relation present in the ontology, meaning that the higher an association's Confidence Rank is, the more certain the system is about that relation. The Confidence Rank is updated indirectly by human indication, and we extract this indication implicitly from the answer provided by the CAPTCHA test taker. The more humans indicate an association is pertinent, the higher the Confidence Rank of a pairing will be. If the Confidence Rank is above a Certainty threshold (Figure 5), the system will consider that relation to be sound. If the Confidence Rank lies between the Certainty threshold and a Suspicion threshold, the association is suspected to be true but is yet to be confirmed with human assistance. If the Confidence Rank falls below the Suspicion threshold, the association is considered to be false.



Figure 5. The Confidence Rank is used to manage the correctness of an ontology association. The higher a Confidence Rank is, the more likely a label correctly describes an image.

Our method of knowledge acquisition is based on two independent procedures which we call a Search Step and a Resolve Step. The Search Step is responsible for retrieving data from the Web and feeding it into the ontology. The Resolve Step refines the association captured by the Search Step by checking their correctness with human users. Both steps are described in depth below.

The Search Step. The Search Step procedure is performed as in the sequence-diagram depicted in Figure 6. A Searching-agent starts browsing the Web looking for random words, which can be acquired from any Web service.

Later, an image searching service is queried with the vocabulary acquired from the previous stage. Finally, for each query executed in the image search, a pairing between the string queried and each image returned is forwarded to an Ontology-building agent as a *Describes* association. This association means that the label is supposed to be representative of the image content. Before inserting the new objects and associations in the ontology, the Ontology-building agent marks each recently discovered pairing with an initial Confidence Rank. This default Confidence Rank value will depend on the reputation of the image-searching service used by the Searching agent. Since current image searching services are not yet very accurate, this value should indicate that further verifications need to be applied to confirm the soundness of the relation (i.e. the initial Confidence Rank lies

between the Suspicion threshold and the Certainty threshold).

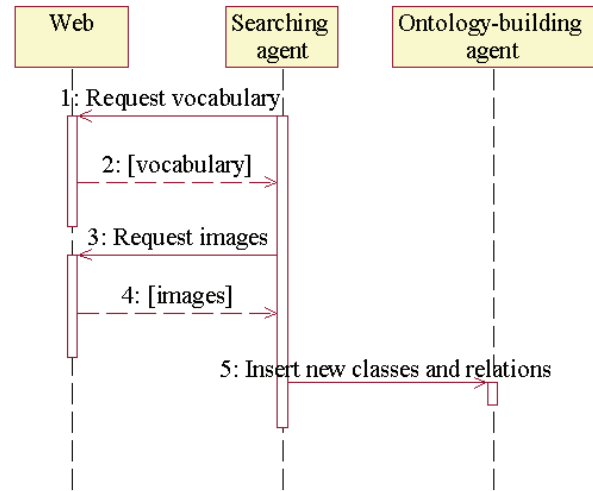


Figure 6. The Search Step sequence diagram. The procedure begins by a Searching Agent retrieving random words from Web services. With the returned vocabulary, a search is executed on image-search directories, like Google's image database. For each query executed in the search, a pairing between the string queried and each image returned is added to the ontology as a *describes* association

The Resolve Step. The Resolve Step aims at detecting faults and discovering new associations between the elements already present in the ontology, as indicated in the sequence-diagram from Figure 7.

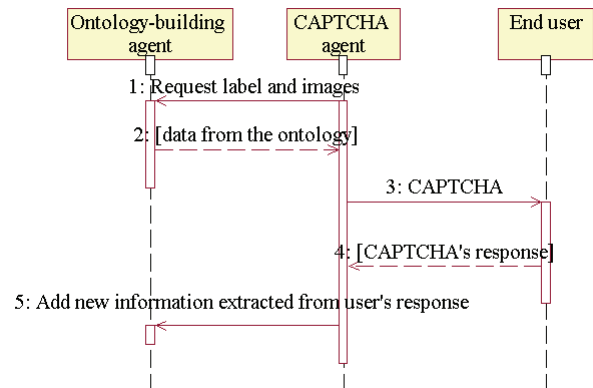


Figure 7. The Resolve Step sequence diagram. A CAPTCHA agent retrieves a textual label and a set of images from the ontology to generate a CAPTCHA test. The test is submitted to the user, who replies with a response to the test. The CAPTCHA agent analyses the response to extract additional semantics and add/update information on the ontology.

The CAPTCHA-agent retrieves from the Ontology-building agent a textual label and a set of images, some that the agents are certain to relate to the textual label (i.e. the Confidence Rank of the relation is above the Certainty threshold), some that the agents are certain not to relate to the label (whose Confidence Rank falls below the Suspicion threshold) and some which relation to the label is yet to be attested (the Confidence Rank lies between the Certainty and Suspicion thresholds). These data will be grouped to form a CAPTCHA to be submitted to an end-user. Finally, depending on the CAPTCHA response, the Confidence Rank of each relation involved in this iteration can be increased or decreased, sharpening the correctness of the ontology.

In Figure 8, we show an instance of a CAPTCHA generated by the CAPTCHA-agent. It works roughly as a reverse Pix and receives as input a label and a collection of images, some related to the label, some that are not related to the label and some which relation is to be verified. The end-user is presented the label and the images formatted in columns. All but one column contain only images whose semantic relation to the current label is known a priori by the CAPTCHA agent (whether the image relates to the label or not).

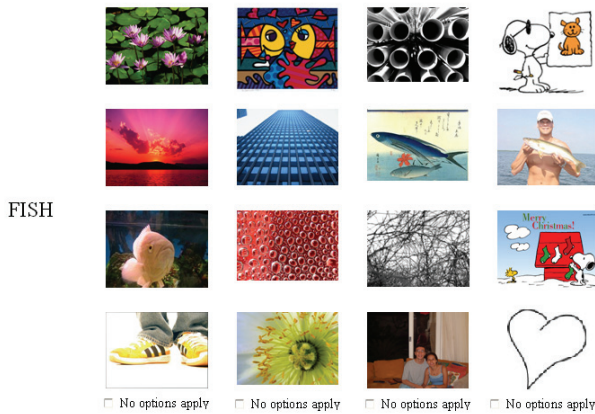


Figure 8. A CAPTCHA generated by the CAPTCHA agent. The user is requested to identify in every column which image relates to the label on the left, but is left unaware that the last column is an experimental column, of which content the CAPTCHA agent is not so sure about. The image selected by the agent in this column will be the source of knowledge acquisition from this test.

One special column, called the experimental column, is the source of knowledge acquisition from the CAPTCHA response. In this column, the CAPTCHA agent inserts those images whose semantic still remains to be verified. The user is then asked to identify, in each column separately, which one image relates to the label presented. Of course, since the CAPTCHA agent must support the generation of a test that certainly has at least one correct answer, and since he has no guarantee that the experimental column will indeed have any image that relates to the current label, the agent is forced to present

the user with an option to indicate that no options apply in any given column. This sets the stage for us to design regular columns (not an experimental one) where the correct answer will be this outside option (no options apply), where all images in the column will not relate to the selected label. This is to avoid any chance for the user to distinguish the experimental column from a regular one.

When reviewing the correctness of the user's response to the test, all columns, except the experimental one, will be taken into account. If the user has identified correctly all associations in the regular columns, the CAPTCHA agent will consider the response acceptable, will assume that the experimental column selection is also correct and inform the Ontology-building agent of the user's response. With this new information, the Ontology-building agent can now update the Confidence Rank of all images presented in the experimental column. The one image that was selected by the user (if any) will have its Confidence Rank related to the current label increased, while the others' will suffer a decrease.

It is important to stress that relations whose Confidence Rank falls below the Suspicion threshold are just as useful in the generation of this CAPTCHA as those whose Confidence Rank reaches above the Certainty threshold. In each column, ideally at most one image should relate to the selected label, in order to protect against random-strategy attacks from computer agents. All regular columns are then formed by no more than one certain relation and a group of certain negative relations.

Random attacks from computational agents could be made less successful by generating tests with more columns or more images per columns. This way, the probability of passing the test using random responses would quickly drop to any acceptable level.

Experiments

In order to evaluate the performance of our method, we implemented a version of the model. A database of images and labels was generated by collecting 600 images from AltaVista's image database (AltaVista Image Search), each associated with a random label (totaling 30 different labels in our database). We asked 10 volunteers to repeatedly take a CAPTCHA test. All of them were Computer Science graduates from the Fluminense Federal University, in Brazil, between 23 and 28 years old. The CAPTCHA was made available over the Internet for free use.

We assigned the values of 1 to the Suspicion threshold and 9 to the Certainty threshold. Whenever in the experimental column, each indication from users that an image relates to a label added 2 units to the Confidence Rank of the relation, while a lack of indication from a test's response subtracted 1 unit from the Confidence Rank. All relations had an initial Confidence Rank of 4 units.

After a period of two weeks, we analyzed the data present in the ontology and could identify original information present in the ontology. The relations associated with the image from Figure 9 are an example of a successful

acquisition of knowledge from our experiments. The original label associated with this image was *John Lennon*. After our experiments, we could observe that such image had established relations with the labels *Rock Star* and *Beatles*.



Figure 9. A successful instance of our knowledge-acquisition method. This image was initially associated with the label *John Lennon*. After our experiments, we could describe it with the labels *Rock Star* and *Beatles*.

One drawback identified in our experiments was the large amount of time required for the method to acquire new information. There is clearly a tradeoff between speed in the knowledge acquisition process and confidence in the acquired knowledge. We are still experimenting different values of the Certainty and Suspicion thresholds, and investigating their effect on the global behavior of our method.







General Discussion

The opportunity raised by CAPTCHAs on cooperative knowledge acquisition is significant. If the tests are properly designed, one could induce users to cooperatively supply required information since both part's interests are directly aligned: users want to pass the CAPTCHA to access some service or perform some task, and the CAPTCHA designer wants to acquire information to enhance Web services.







This formulation is specially interesting because it softens the bottleneck of relying on a single or small group of humans to identify an image's content, since now this effort is equally distributed across all CAPTCHA's users. The administrator of a popular Web service could rapidly grow the information present in her ontology the more users visit her system and take the CAPTCHA test.

Unlike previous approaches to this problem (von Ahn and Dabbish, 2004), we need not design additional incentives or interactions apart from what previously existed with CAPTCHA. Additionally, the incentive embedded in CAPTCHA use is very diverse and general, something which does not occur very easily with entertainment games, which could entertain only very specific kinds of people.

In order to provide richer semantic information about objects, one might model an ontology with relations other than the *Describes* relation we mentioned before. In the general case, our CAPTCHA agent could select a random relation from the ontology and create a test asking the users to indicate how one could make some presented predicate true using the data present in the database. Figure 10 contains an example of such scenario. Here, the user is asked to identify an animal (X1) that is a predator of another (X2). This approach easily generalizes to N-ary relations, where users are asked to identify every parameter in the relation. The parameters in this general case could be of any type (not only images).

IS PREDATOR OF (X1, X2)		IS PREDATOR OF (X1, X2)		IS PREDATOR OF (X1, X2)	
<input type="radio"/> X1 <input type="radio"/> X2		<input checked="" type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2	
<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2	

☐ No options apply

<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2	
<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2		<input type="radio"/> X1 <input type="radio"/> X2	

☐ No options apply

Figure 10. A CAPTCHA test for a binary relation. In this test, users are asked to identify each parameter, from a certain relation from the ontology, that makes the given predicate true. This test easily generalizes to N-ary relations.

Conclusions and Future Research

We presented a novel approach to the problem of ontology-building for image retrieval. Our main contribution is the combination of existing image-search techniques like metadata-based analysis with knowledge acquisition tools. We have also shown how CAPTCHAs can be applied to serve as knowledge-acquisition assistants, something that can be done with no additional cost in current Web applications.

The knowledge present in our ontology could not only serve as database for image searching tools, but also as object of study for client modeling. Since the knowledge present in the ontology reflects directly what humans have informed to the system, system managers could experiment client modeling techniques for better addressing his business interests.

We are currently running additional experiments to validate our model in more complex settings, where the relations present in the ontology are not only simple binary

relations between labels and images, but more complex ternary or N-ary relations between any combinations of data types (strings, numbers, images, etc).

Another interesting direction for future work would be to work on the design of a CAPTCHA test where users would be able to provide our method with new relations between the elements of the ontology. The need to control vocabularies while still satisfying the CAPTCHA security requirement is a problem we expect to attack in future versions of this work.

References

von Ahn, L.; Blum, M.; Hopper, N.; Langford, J.: *The CAPTCHA Project*: <http://www.captcha.net>

von Ahn, L.; Blum M.; Langford J. 2004. *Telling Humans and Computers Apart Automatically: How Lazy Cryptographers do AI*. Communications of the ACM 47 (2): 56-60

von Ahn, L.; Dabbish L.. 2004. *Labeling Images with a Computer Game*. In Proceedings of the ACM SIGCHI conference on Human factors in computing systems, 319-326.

AltaVista Image Search, www.altavista.com/image/default

Chakrabarti, S.; Dom B.; Gibson D.; Kleinberg J.; Kumar S.R; Raghavan P.; Rajagopalan S.; Tomkins A.. 1999. *Hypersearching the Web*. Scientific American 280 (6): 54-60

Google Image Search, <http://images.google.com/>

Hyvönen, E.; Saarela S.; Styrman A.; Viljanen K. 2003. *Ontology-Based Image Retrieval*. In Proceedings of WWW2003, Budapest, Hungary, Poster papers

Oi Internet, <http://www.oi.com.br>

Saussure, F. 1966. *Course in General Linguistics*. New York, NY : McGraw-Hill Book Company.

Smith, J.; Chang S. 1997. *Visually Searching the Web for Content*. IEEE MultiMedia 4 (3): 12-20

Yahoo! Mail, <http://www.yahoo.com>