Exploiting text for extracting image processing resources

Gregory Grefenstette, Fathi Debili, Christian Fluhr, Svitlana Zinger

CEA LIST
92265 Fontenay-aux-Roses, FRANCE
{Gregory.Grefenstette,Christian.Fluhr}@cea.fr,
CNRS UMR 8135 – LLACAN
7, rue Guy-Môquet, 94801 Villejuif CEDEX, France
fathi.debili@wanadoo.fr

Abstract

Much everyday knowledge about physical aspects of objects does not exist as computer data, though such computer-based knowledge will be needed to communicate with next generation voice-commanded personal robots as well in other applications involving visual scene recognition. The largest attempt at manually creating common-sense knowledge, the CYC project, has not yet produced the information needed for these tasks. A new direction is needed, based on an automated approach to knowledge extraction. In this article we present our project to mine web text to find properties of objects that are not currently stored in computer readable form.

1. Introduction

As computer applications are attached to more and more everyday objects, there will be an expectation that the machines that we interact with know many things about themselves and about their environment. For example, a number of prototypes of personal robots have beguin to appear such as Honda's ASIMO and Sony's QRIO). It is expected that the natural means communication with these robots will be voice input (Jikinen, 2003) and that the robot will be able to respond to this input in a natural way, in the current setting in which it is found. These scenario presuppose both that the robot is able to visually process its surrounding, for example, recognizing what room it is, recognizing what objects are found there. These suppositions imply that the robot, and its visual processing mechanism, possess models about everyday objects that allow this recognition. But everyday knowledge about objects (e.g. can you pick it up? would it be found in a kitchen? what color is it?) does not currently exist as computer data. The largest attempt at manually creating common-sense knowledge, the CYC project (Lenat, 1995) begun in 1984, and founded and funded as a company since 1994, has not yet produced the information needed for these tasks. Hand construction of such information is difficult to perform, both in terms of completeness of the information needed, and in terms of deciding what must be modeled and how this modeling should be implemented.

Since the project began, a new resource, the Web has of course appeared, and its volume continues to expand at a great rate. We have estimated[1] that there were 80 billion words of English available through Altavista in 2001 and that this number rose to at least 145 billion by 2004. There is a lot of text, then, available. This article examines whether this volume of text could be exploited for gathering automatically part of the information that will be needed for the visual decoding tasks in applications such as personal robots, though it is hoped the information would be useful for any application involving object recognition in images.

Information extraction from text has enjoyed increasing interest over the past 10 years, as linguistic processing matures and as an ever growing source of text becomes available on the WWW. But information extraction research has mainly been concerned with finding named entities and terminology for information retrieval and classification applications.

In this article we present our project to mine web text to find properties of objects that are not currently stored in computer readable form.

2. Probing for Knowledge

Our methodology is a combination of probing (Sato and Nagao, 1990), text mining (Harabagiu *et al*, 2003), and information extraction techniques (Grishman, 2003). Probing means generating a request that is sent to a Web interface in order to recover a specific type of information. We report here on probing the web for two types of information about objects, determining the usual color of things, and finding the usual locations of things. Both of these pieces of information should be useful for image processing and decoding, the first for helping to identify objects and the second for reducing the number of possible objects to be considered in a given setting, supposing that the setting can be identified by some other means.

2.1. Color of things

In order to find the color of things, we have probed the web by taking all the nouns from a full-form lexicon of English, and prepending one of the following "color" words (purple, red, orange, yellow, green, blue, black,

[1] http://www.infonortics.com/searchengines/sh04/slides/greffen.pdf

beige, ivory, brown, white, gray, grey, silver, copper, gold, golden, bronze, pink, striped) to each noun. We then sent these paired lists of contiguous words to a search engine, for example "brown sky". For each pair, the search engine provided a list of links containing the term, which we did not use, and the total number of pages (this value is usually only an estimation[2]) that contained the term. When we performed this search on the noun "sky" we found, the following page counts at one point last year:

blue sky	2110000 pages
red sky	175000
black sky	62400
grey sky	36100
gray sky	27900
white sky	23800
orange sky	17300
purple sky	16700
yellow sky	13700
green sky	12800, etc.

Likewise, sending off a query for "grass" yielded the responses:

green grass	307000 pages
blue grass	285000
black grass	13400
yellow grass	12600
brown grass	10200
golden grass	5000 …

When this probing technique was applied on a large scale, we found many of the color associations that one might expect for common things.

Consider the following list of items and the page counts of the most frequent colors found with each:

eyes -- blue 1620000, brown 455000, green 401000
tea -- green 1410000, black 317000, white 64800
blood -- red 1350000, white 873000, blue 83600
cat -- black 1310000, white 127000, blue 63000
skies -- blue 439000, grey 28400, gray 23400
rice -- brown 384000, white 261000, golden 27800

This probing technique seems to produce interesting results for items that have typical colors, such as fruit, and animals, and even articles of clothing as can be seen in the following list:

shirt -- white 250000 , blue 138000 , red 121000
shoes -- red 207000 , black 184000 , white 90900
skirt -- black 97700 , blue 24400 , white 20100
socks -- white 121000 , black 37600 , red 22200
tie -- black 408000 , white 38300 , red 24800

But the simple concatenation of color names to nouns brings into play other linguistic phenomena in which the color term is not being used as in a strictly descriptive sense. For example, color words appear frequently appearing in names such as the White House, the Red Cross, Brown University, the Golden Gate, and a number of popular commercial brands such as Red Bull. Other types of problems come from compound nouns in which the first element is a color but which can not be removed from the phrase without changing the sense, in other words, where it is not being used used as a distinctive attribute of the second element in the phrase, but has become an integral part of the concept. Examples of this inseparability are the common compounds *pink slip, yellow jacket*, *red giant*, etc. We discuss these linguistic problems in greater detail in (Grefenstette, 2005).

2.2. Object location

In addition to color, whose identification should help image processing since this is one aspect of images that is easy to recognize by image processing (Stricker and Orengo, 1995), we have begun examining other information that might be useful in image processing. that could be extracted from text.

Object recognition is a difficult problem for image processing. Taking a subset of WordNet (Zinger *et al*, 2005) that corresponds to physical objects, we attempted to extract relations between objects and locations. This information might help a personal robot to better recognize objects by reducing the number of possible objects to be recognized in a given setting (Szummer and Picard, 1999.)

For example, if we know that we are in a certain room, for example a kitchen, the types of objects that might be found there in a normal situation should be constrained by the location. To derive a list of items that are visible in a given room such as kitchen, we could ask people to list these items, maybe using some social interaction interface such as ConceptNet (Liu and Singh, 2004). We decided to explore whether we can use a web probing technique and simple text normalization tools to see if we could identify which objects are most likely to occur in a kitchen.

First we imagined a short list of contiguous phrases that could be searched for using standard search interfaces, and which should find text talking about the kitchen. These probe phrases were

"on the kitchen table"
"on the kitchen counter"
"into the kitchen"
"on the kitchen floor".

Given these seed phrases, our procedure was the following:

1. For each phrase we recovered 1000 URLs from a popular search engine.
2. Each page was fetched and locally stored.
3. The locally stored version of the page was transformed into raw text using the Unix program *lynx* which attempts to produce plain text version of web pages.
4. Each text version of the page was converted into sentences using a simple text tokenizer (Grefenstette, 1999).
5. Any sentence containing the word kitchen was retained, as well the sentence preceding and following this sentence
6. These selected sentences were then uniquely sorted to eliminate doubles.

[2] http://aixtal.blogspot.com/2005/01/web-googles-counts-faked.html

7. Words in these unique sentences were separated by new-lines by replacing every non-alphabetic character by a new line.
8. These words found were then sorted, and counted.
9. Using a lexicon for the English language, all words which were not tagged as nouns in this lexicon were removed from the sorted list.
10. From the remaining list, we removed any word that was not in the subset of WordNet that corresponds to physical objects (Zinger *et al*, 2005).

The results of these ten steps are presented below. The most common objects in the remaining list were the following (preceded by the number of unique sentences):

3998 room
2654 counter
1562 cabinets
1288 sink
1254 furniture
1220 kitchens
1026 wall
941 cabinet
844 cook
787 stove
747 appliances
726 bathroom
717 glass

In order to clean this data a bit further, we added an eleventh step to the ten steps above:

11. We transformed these raw frequencies into a value similar to mutual information by dividing the sentence frequency by the word frequency that we had derived from a previous crawl of the web.

Resorting by this simple version of mutual information, we now find the following items to the top of the list with each word now preceded by this value, then the number of sentences, then the word:

1136.37	2	crockeries
1092.9	2	rattraps
790.91	30650	kitchen
524.591	8	hassock
483.435	102	fibber
392.088	510	stools
358.821	42	pursuer
324.151	2	washtubs
290.773	45	windowsill
248.448	2	calcimine
246.916	1	underskirts
234.88	8	dishcloths
231.611	157	cupboards
230.614	8	potholder
225.564	15	doorframe
224.217	1	haircloth
220.752	2	iceboxes
197.11	3	dishrag
194.484	11	scullery

As is well known with mutual information, this brings rare words to the forefront. If we then eliminate any word appearing in less than 5 different sentences, and take the first two hundred items appearing in the context of our probe phrases, we get the list:

ants appliances apron bakers banquette baseboard baseboards basement bathrobe bedfellow bedroom bedspread blazes blenders bluebells bookcase butcher butlers cabinet cabinets carafe carousels casseroles cavemen cellar cereal cheesecloth chef chefs chopsticks closets coffeepot colander contraptions cookers cookstove corer couch crock crockery crocks cupboard cupboards curtains cutlery cynics dearie dish dishcloths dishes dishpan dishwasher doorbell doorframe doorknob doormat doorways drapes drawers drawers duds duds dustpan earthenware entryway extinguisher faucet faucets fibber flatware flooring flowerpot footman freezer freezers fridge furniture galleys germs glasses grandma grandmas grater graters groceries grout grubby hag hairball hallway hassock hearths hobs hoods housekeepers hutch icebox jackhammer joists juicers kettle kitchen kitchens kitchenware kitties knife knives knobs laddie ladle ladles laminate larder laundry linoleum maggots mess messes mitts mommy nigger nightgown nook nooks pail panelling pantries pantry papayas paring parsnips pedant peeler peelers peninsulas phoebe pinhead planking plinth plumber poops porch potholder potholders pots puke pursuer racks refrigerator refrigerators roaches saucepans sawhorse scraps scullery shelves shred silverfish silvers silverware sink sinks skylights slicer slob snipping soffit spatula spatulas sponges spoons squeezer stairs stairwell stool stools storeroom stove swatter swivel tablecloth teapot teapots toaster toasters towel trivet utensil utensils vanities wainscoting wallflower wash washing whiff whisk whisks windowsill woodenware yams

Such information as in the list above provides a number of items that might not have appeared in a manually constructed list, e.g. *slicer, grater, flowerpot…* but which make sense. This list could further be cleaned by removing items corresponding to animate objects, such as *footman, chef, mommy*, etc. (which have been included in the list of physical objects derived from WordNet). We will add this additional step in the future.

In addition, it would be interesting to distinguish object which appear anywhere in the house, and more specifically in a given room. We will explore this discrimination by producing lists, using the same eleven steps, but different seed patterns. For example, the following types of seeds:

"on the dining room floor"
"on the living room floor"
"on the bedroom floor"
"on the bathroom floor"
"on the hall staircase"
etc.

should produce lists of similar items (i.e. those found in the home) but with differing frequencies that should allow for more accurately placing objects in each room. This technique is very similar to using semantic axes seeds to distinguish vocabulary along semantic axes, as in Grefenstette *et al*. (2006). In that work a number of seeds where defined for each endpoint of an axis corresponding to a different type of emotions. Using web frequencies, as here, to find how often a new word appeared next to each seed word allowed us to place a new word along the axis between the positive end of the emotion or the negative end. In addition, comparing the frequencies with which the new word appeared with the seedwords of each semantic axis allowed us to class the centrality of the word in the semantic class. Here in this article, the seeds (kitchen phrase, or bathroom phrases) serve the same purpose of postionning the words, not along emotional axes, but along typical dimensions corresponding to typical locations.

3. Conclusion

In this article, we have explored using the web to extract real world information that could be useful for image processing. The results give promise that the discovery of certain types of useful information that might be extracted by simple probing and text analysis from the Web. The Web contains so much text (billions of words for most European languages) that it should be possible to continue mining it for new types of lexical resources, corresponding to real-world knowledge. These lexical resources, those needed for specific image treatment tasks, still need to be built and widely distributed. The techniques described here should useful in this construction.

4. Acknowledgments

This research was sponsored in part by the EU Network of Excellence MUSCLE (Multimedia Understanding through Semantics, Computation and Learning, FP6-507752).

5. References

Grefenstette, G. (2005) The Color of Things: Towards the Automatic Acquisition of Information for a Descriptive Dictionary, in *Revue Française de Linguistique Appliquée (RFLA)*, December, pp. 83-94.

Grefenstette, G. (1999) Tokenization in *Syntactic Wordclass Tagging*, Hans van Halteren (ed.), Dordrecht, Kluwer Academic Publishers, chap. 9, pp. 117–133

Grefenstette, G., Yan Qu, David A. Evans, James G. Shanahan. (2006) Validating the Coverage of Lexical Resources for Affect Analysis and Automatically Classifying New Words along Semantic Axes. In Computing Attitude and Affect in Text: Theory and Applications. Editors: Shanahan, James G.; Qu, Yan; Wiebe, Janyce; Springer; ISBN: 1-4020-4026-1

Grishman R.. (2003) Information Extraction. In The Oxford Handbook of Computational Linguistics, Ruslan Mitkov, editor, Oxford University Press, Chapter 30.

Harabagiu, S.M., Moldovan, D.I., Clark, C., Bowden, M., Williams, J. & Bensley, J.. (2003) Answer Mining by Combining Extraction Techniques with Abductive Reasoning. In *Proceedings of TREC'2003* pp. 375-382

Jokinen, K. (2003). Natural Interaction in Spoken Dialogue Systems. *Proceedings of the Workshop Ontologies and Multilinguality in User Interfaces.HCI International 2003*, Crete Greece, June 2003. Vol 4, pp. 730-734..

Lenat, DB (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11): 33-38

Liu, H. and Singh, P.. (2004). ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal 22(4). pp. 211-226. Kluwer Academic Publishers.

Sato S. and Nagao M. (1990) Toward Memory-based Translation. In *Proceedings of COLING-90*, 3, pp. 247-252

Stricker, M. and Orengo, M. (1995) Similarity of color images. In *Storage and Retrieval for Image and Video Databases III, SPIE 2420*, San Jose, CA, Feb. pp 381-392.

Szummer, M. and Picard RW (1998) Indoor-Outdoor Image Classification, *IEEE CAIVD*, Bombay, India, Jan., pp. 42-51

Zinger, S., Millet C., Mathieu B., Grefenstette G.,. Hède, P and Moellic P.-A. (2005) Extracting an Ontology of Portrayable Objects from Wordnet. In *MUSCLE / ImageCLEF workshop on Image and Video retrieval evaluation*, Vienna, Austria, Sept, pp. 17-23.