

Structured vs. unstructured tagging – A case study

Judit Bar-Ilan, Snunith Shoham, Asher Idan, Yitzchak Miller and Aviv Shachak

Department of Information Science, Bar-Ilan University

Ramat-Gan, Israel

972-3-5318351

{barilaj,shohas}@mail.biu.ac.il, asherid@netvision.net.il, milleri@mail.biu.ac.il and shachaa4@biu.013.net.il

ABSTRACT

In this paper we describe and discuss a tagging experiment of images related to Israeli and Jewish cultural heritage. The first group of participants was asked to assign the images tags that describe them, while the second group was asked to provide free-text values to predefined metadata elements. The results show that on the one hand structured tagging provides guidance to the users, but on the other hand different interpretations of the meaning of the elements may worsen the tagging quality instead of improving it. Our recommendation is to experiment with a system where the users provide both the tags and the context of these tags.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *indexing methods*.

General Terms

Experimentation

Keywords

Structured tagging, unstructured tagging, images, cultural heritage

1. INTRODUCTION

Image tagging on the Web has recently become extremely popular (see for example the popular photo sharing and tagging services Flickr (<http://flickr.com/>) and Smugmug (<http://www.smugmug.com/>)). Although we do not challenge the saying, that “a picture is worth a thousand words”, currently the best method to retrieve pictures is based on textual descriptions. Image processing and recognition are very active research fields, but there are no well-developed commercial systems that retrieve images based on image recognition. Yahoo! and Google image searches are also based on the text near the tag in the html files.

The popular tagging systems usually do not impose any limitations on the choice of tags (del.icio.us allows only single word tags, but otherwise there are no limitations).

In addition to free tagging, there are highly structured metadata systems as well. One example is VRA Core Version 3 (<http://www.vraweb.org/vracore3.htm>) is a very detailed metadata element set enabling the creation of records “to describe *works* of visual culture as well as the *images* that document them” [5]. The VRA metadata set defines the elements (fields in the traditional librarian jargon), which should be given values (filled in). The recommendation is to use controlled vocabularies, especially the Getty vocabularies [2]. Not only the use of controlled

vocabularies is recommended, but specific rules apply for assigning values to the specific fields (e.g., date formats and rules for recording personal names).

The VRA metadata set was created mainly for use in museums, thus it is to be filled in by professionals. A simpler set of metadata elements is the Dublin Core [1] which is intended to be used by the general public (especially the simple version of the Dublin Core). The central aim of the developers of the DC (Dublin Core) was to substantially improve “resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects” [6]. The Dublin Core has a simple and a qualified version. Although it recommends the use of different controlled vocabularies for some of the elements, these are only recommendations and all fields can be assigned free-text values.

The basic philosophy of the DC and of the librarian world in general is that field-based resource discovery is much more effective than free text search. An often mentioned example is that field-based search enables one to differentiate between books written *by* Shakespeare and books written *about* Shakespeare. Library catalogs and bibliographic databases are well-known examples of this approach.

As an initial phase of a larger project, we conducted a small experiment to compare field based descriptions of images with free tags of the same images.

2. THE EXPERIMENT

The goal of this experiment was to compare free text tagging with what we call *structured tagging*, i.e. assigning free text keywords (tags) to predefined metadata elements (fields).

Twelve images related to Jewish and Israeli cultural heritage were chosen (see Figure 1). For each image, the Web page from which the image was taken was given, in order to provide some context to the picture. For example the first picture was taken outside the building in which Israel’s independence was declared in 1947. The page, http://www.knesset.gov.il/docs/eng/megilat_eng.htm provides clear context in this case. For some of the other images, the context is much less clear, for example, picture 11 depicts part of the arch of Titus in Rome. The picture was taken from <http://www.biblelight.net/temple.htm>, where the specific picture serves mostly as an illustration. Picture 5 is one of famous Chagall windows at the Hadassah Hospital in Jerusalem. The page itself provides no context; the tagger can either rely on his/her previous knowledge or try to access the parent page, <http://www.md.huji.ac.il/special/chagall/> in order to learn something about the image.






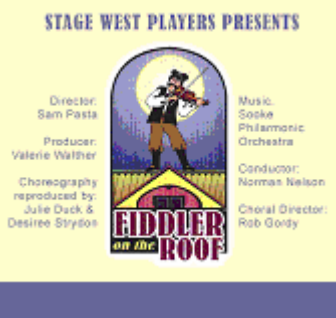






<p>1.</p>  <p>www.knesset.gov.il/docs/eng/megilat_eng.htm</p>	<p>2.</p>  <p>www.hebron.co.il/top/str_Machpela.html</p>	<p>3.</p>  <p>www.daat.ac.il/daat/kitveyet/mabua/36/kasuto.htm</p>
<p>4.</p>  <p>www.rhodesjewishmuseum.org/torah.htm</p>	<p>5.</p>  <p>www.md.huji.ac.il/special/chagall/yehuda.html</p>	<p>6.</p>  <p>stagewest.ca/showinfo.html?fiddlerontheroof</p>
<p>7.</p>  <p>haggadahsrus.com/y.Art10-Kibbutz1955.htm</p>	<p>8.</p>  <p>www.trekker.co.il/israel/i-shaarim-09.htm</p>	<p>9.</p>  <p>www.acs.ucalgary.ca/~elsegal/Shokel/MHCS02_MosaicMusings.html</p>
<p>10.</p>  <p>mop.ort.org.il/rooms/graphics/zionut/1.htm</p>	<p>11.</p>  <p>www.biblelight.net/temple.htm</p>	<p>12.</p>  <p>www.chelm.org/jewish/chags/sukkot/suk2pic.html</p>

Figure 1. The images with the URLs of the pages within which they originally appear

Forty-seven students from the Department of Information Science participated in the experiment. This was a convenience sample of first year students. The students have not received previous training in cataloguing and/or indexing, and thus constituted a sample of inexperienced users. The students were divided into two groups. The first group of twenty students was asked to provide a short description of each picture and a list of free-text keywords to describe the content of the pictures (with encouragement to relate to artistic aspects as well). For each picture they were asked to fill in and submit the appropriate form. In the form there were ten empty textboxes for assigning keywords.

The second group of twenty seven students received the same basic instructions, but they were asked to fill in a structured form. This form contained the following fields:

- General themes
- Symbols
- Personalities
- Description of event
- Location of event
- Time of event
- Object type
- Object creation date
- Creator
- Related links
- Additional information about the image that did not match any of the predefined fields
- Recommending additional fields to improve the form

The students were told that not all the fields are relevant to all the pictures. The fields for the structured form were defined based on a small pretest the authors conducted on 10 pictures (there was some overlap between the two sets of pictures), where each of the authors individually tagged the images.

3. RESULTS

Here we only comment on the taggings where the page in which the image is embedded is in English.

The [first picture](#) is a photo of the crowd standing outside the Museum in Tel-Aviv on May 14, 1948 while inside the Museum David Ben-Gurion read out the Proclamation of Independence and members of the People's Council signed it. The students from the first group (unstructured tags) concentrated on what they saw in the picture itself, and provided a good overall description of what can be seen in the picture. Only 5 out of the 20 students mentioned the date of the event.

The students of the second group provided less comprehensive descriptions under *general theme*. A considerable number of students from the second group (structured tags) added information that can be found only on the page in which the photo appears. Under *symbols* five students listed general symbols related to the event like the Proclamation of Independence and the Israeli flag, neither of which appear in the picture, and ten students out of the 27 mentioned David Ben-Gurion as a

personality, even though he is not in the picture, but is related to the event. On the page in which the URL appears, there is an audio recording of him reading out the Proclamation. Note that only three students (out of 20) mentioned Ben Gurion in the unstructured group. The second group provided accurate date of the event. It seems that the structured group felt obliged to fill in most of the fields, and this is why they included information that appears only on the page and not in the image and/or made associations.

The [fourth picture](#) is a picture of an 800 year Torah scroll from Rhodes that was put on display in Los Angeles in 2003. The page provides extensive information on the history of the scroll and about the different tests to determine the date the scroll was written. The first group (unstructured taggers) provided only superficial information, stating that this is a picture of an ancient Torah Scroll. Only one student mentioned Rhodes and one other student mentioned Los Angeles in the keywords. Spain, where the scroll was originally written was mentioned by almost all the students. The keywords did not say anything about the age of the scrolls; but the age was mentioned by 1/3 of the members of this group in the description field.

Almost all of the structured taggers mentioned that the picture show a Torah scroll. Some of them noted this under *general theme*, while others under *symbols* (the Torah is definitely one of the central symbols of Judaism). Only 1/3 of this group assigned the value Torah Scroll to the *symbols* element. Most of those who chose to fill in the *description of event* field describe the unveiling of the Torah in Los Angeles. For *event location* we received rather surprising results, among them Los Angeles, synagogue, Rhodes and Spain. Similar variability in the results was observed for the *time of event* field: 2003, 1492, 15th century, 12th century. The confusion is understandable, the scroll has a very long history – it was written in the 12th century in Spain, was transferred to Rhodes around 1492 when the Jews were expelled from Spain. After the 2nd World War the scroll was transferred to Argentina (this part of the history was not mentioned by the taggers) and in 2003 it went on display in a Los Angeles synagogue. So what event are we talking about? There are a number of events related with this image. The elements *event location* and *event date* were ambiguous.

We observed a similar problem with [picture 6](#), the Fiddler on the Roof play (structured tagging). Which event are we thinking of? The play (in Canada) in 2005 or the theme of the play, Jews in a small Russian township at the beginning of the 20th century, or perhaps the event is that a fiddler plays his tune on the roof? What is the location: Canada, the roof or the theatre? Some of the taggers in both groups associated the Canadian show with Haim Topol who played the part of Tevye the milkman at Broadway and in the movie. He is not mentioned in the image, but he is clearly associated with the play.

Even though there is almost no context in the page in which [picture 5](#) appears, both the structured and the unstructured descriptions are of rather high quality, probably because the Chagall windows (one for each of the twelve tribes) at the Hadassah hospital in Jerusalem are so well known.

[Picture 7](#) is part of a modern Haggadah (Haggadah is a book telling the story of the exodus of the Jews from Egypt which is read on the first evening of the Passover). Some viewed the Haggadah as the theme of the picture, while other defined it as one of the symbols (similar to the case of the Torah scroll).

Picture 9 is a picture of a mosaic from the 6th century that was discovered in Beit Alpha and depicts the binding of Isaac. How should be the location defined here? Is it Beit Alpha, where the mosaic was found or is it the Moriah mountain where the binding took place? There was no consensus among the participants on this issue, which is not surprising, since both interpretations are appropriate. The “unstructured” taggers described the mosaic in the description field and added keywords related to the event captured by the mosaic.

Picture 11 is a photo of a part of the arch of Titus in Rome. This part of the arch shows the Romans taking away the Menorah (the candlestick) from the Temple in Jerusalem. In this case the large majority of the students defined the Menorah as a *symbol*, while they defined the theme of the photo as the destruction of the Second Temple.

The last picture is a photo of a sukkah (a temporary shelter commemorating the life of the Jews in the wilderness) in which the Jews sit during the Sukkoth holiday. Rather interestingly, unlike the case of the Torah and the Haggadah, almost all the students defined the sukkah both as the *theme* of the picture and as the *symbol*.

4. DISCUSSION AND CONCLUSION

Our findings show that structured tagging usually resulted in more detailed descriptions. However, there are specific problems related to structured tagging. In this experiment we observed two such problems:

- 1) Some of the elements were not well-defined and could be interpreted in several ways (e.g. event and location). One possible way to overcome this problem is to provide more accurate metadata elements. The problem with this approach is that the form becomes complex, and only a small number of elements are relevant for the description of a specific image. A complex form discourages the users from providing descriptions.
- 2) Some values can be assigned to several fields (e.g. symbol and theme in our experiment). Each tagger has his/her own point of view and it will be clear to him/her which option to choose. However these decisions complicate discovery and retrieval. The major advantage of field based retrieval is that it allows the user to focus on specific fields. If the desired information is in a field that was excluded (for example, in case it is clear to the user that the Torah is a *symbol*, and decides to limit the search to the *symbol* field, excluding the *theme* field), not all the relevant information will be retrieved. In this case the user is better off with free text search.

It is hard to make clear recommendations based on such a small experiment. We believe that the use of elements has added-value, since overall, the “structured” group provided higher quality descriptions. Probably, the system should suggest a list of elements from which the user can choose, and allow him/her to add new elements if none of the existing ones are appropriate for his/her needs. This list should serve as guide to the user helping him/her in creating structured descriptions. The list should be only a recommendation, so that the user is not forced to fit the information he/she wants to provide into an inappropriate list of elements.

Currently Rawsugar [3] allows the use of subtags [4]. When subtags are used, the tag itself can be interpreted as the name of the field, and the subtag as its value, like in the examples provided by Rawsugar: *restaurants>Chinese* or *location>Bay Area*. Our recommendation is to go one step further, to provide a different terminology for the element and for its value (which is the tag) and to supply a list of recommended elements in addition to allowing the users to define their own elements. In the above example, *restaurant* and *location* are elements, while *Chinese* and *Bay Area* are tags assigned to these elements. The basic role of the elements is to provide context, i.e., *Chinese* is a cuisine in the above setting, while *language>Chinese* is a linguistic characterization.

The main advantage of free tagging that it does not put any limitations on the users and allows them to tag through association. It allows to characterize the images from apriori unexpected aspects.

5. ACKNOWLEDGEMENT

We thank the students of the Department of Information Science at Bar-Ilan University who participated in the experiment.

6. REFERENCES

- [1] Dublin Core Metadata Initiative. <http://dublincore.org/>
- [2] Getty Vocabularies. http://www.getty.edu/research/conducting_research/vocabularies/
- [3] Rawsugar. <http://www.rawsugar.com/index.faces>
- [4] Rawsugar. http://www.rawsugar.com/pages/tour/tour_2.faces
- [5] VRA Core Version 3. <http://www.vraweb.org/vracore3.htm>
- [6] Weibel, S. L. and Lagoze, C. (1997). An element set to support resource discovery. *International Journal on Digital Libraries*, 1 (1997), 176-186.