Semiotic Tagging: Enriching the Semantics of Tags for Improved Image Retrieval

Frank Nack
University of Amsterdam
Amsterdam, Netherlands
Email: nack@uval.nl

Ansgar Scherp
Leibniz Information Center for Economics
Kiel, Germany
Email: a.scherp@zbw.eu

Chantal Neuhaus
University of Koblenz-Landau
Koblenz, Germany
Email: cneuhaus@uni-koblenz.de

***Abstract*—SemioTag is an approach towards tagging that utilizes the semiotic sign categories icon, index, and symbol as classification structures to be used by users during the annotation and search of images within social media-oriented repositories. We compared the influence of this approach on the tagging and querying behaviour of users, with respect to usability, efficiency, and user experience, between the standard Flickr tagging and querying method and the one used in SemioTag. Our results show that semiotic tagging is considered more tedious and takes about twice the time as standard tagging. However, subjects produced a larger number of tags with semiotic tagging. Finally, querying with semiotic tags is not considered more cumbersome than querying using standard tags. Subjects stated that semiotic-based search provides more reasonable results than search based on standard tagging because it provided more control on the query. Semiotic search turned out to be faster. Overall, the findings clearly indicate to further investigate in the direction of semiotic tagging. We anticipate application of semiotics for particular types of human-centered IR such as explorative search.**

I. Introduction

Massive image databases, such as Flickr[1], Picasa[2], or Istagram[3], created to facilitate the swift exchange of images among family, friends, and peers, differ from already established professional image databases, such as the MAGNUM database[4], or the Getty[5] because they are fully established on user generated content. However, these environments are not under editorial supervision and thus do not rely on classification schemata in form of annotation taxonomies or ontologies to facilitate the access to the content. Instead, Flickr and friends make use of folksonomy tagging (also known as collaborative tagging, social classification, social indexing, and social tagging), a method of collaboratively creating and managing tags to annotate and categorize content [1]. The advantage of tagging is to easily create a vocabulary based on freely chosen keywords instead of a controlled set of terms and structures, as introduced in professional repositories. This highly popular approach also carries serious problems. Typically, there is no contextual information available about the semantics of a tag, no matter if it is a single tag or a bag of tags. Additionally, different people may use drastically different terms to describe the same concept. This lack of semantic distinction can lead to inappropriate selection and thus strongly weakens the access and use of the available content. Research tried to address this semantic problem through provision of tag recommenders or tag analysis tools [2], [3], [4], [5], [6], [7].

In this article, we introduce a semiotics-based approach to overcome the lack of semantics of tags without diminishing their ease of use. Semiotics is here understood as the study of signs, in particular the visual sign system [8], [9], with a focus on the relation between signs and the things to which they refer, their denotata, or meaning. We argue that a semiotics-based methodology for tagging facilitates support for contextualized access to image material and thus improves the quality of dealing with image material. The approach outlined in this paper makes use of the three forms a sign can take within semiotics, namely *icon*, *index* and *symbol*. These forms have been integrated into the annotation and retrieval interfaces of a Flickr-like environment, called SemioTag. We performed a user test with 12 subjects in a repeated measures design in which we compared the usability and efficiency of users under the standard Flickr tagging and querying method against the one used in SemioTag. We also investigated the experiences of the subjects under both conditions. We present the results of these tests and conclude not only with an extensive evaluation of the subjects' behavior and experiences but also with remarks on the validity of semiotic theory.

II. Semiotics: Theory of Signs and Symbols

An image is a form of utterance that provides identifiable semantics. The same image presented in different communication contexts might appear with a modulated semantics because the situation and the resulting point of view on the image creates new levels of meaning. Thus, it is useful to quickly examine the generation of meaning through an image, which forms the basis for the decision which tag to use for tagging or the querying of an image. In the following, we discuss the image and its interpretational context based on semiotic theory [8], [9].

The first access to an image is on its optical level where the viewer tries to identify as many objects as she can in the available time of perception. Each object is mentally

[1]http://www.flickr.com, last visited 11.01.2014
[2]http://picasa.google.com/, last visited 11.01.2014
[3]http://instagram.com/, last visited 11.01.2014
[4]http://inmotion.magnumphotos.com/, last visited 11.01.2014
[5]http://www.gettyimages.com/, last visited 11.01.2014

transformed into a sign, which in semiotics is composed out of the "signifier" (the form which the sign takes) and the "signified" (the concept it represents). Within semiotics, a fundamental division of "signs" is established [9], which should be understood not as a classification of distinct "types of signs" but rather as a differentiation between "modes of relationship", i.e., between sign vehicles and their referents. These three divisional concepts are:

1) Symbol: the signifier does not resemble the signified but the relationship is fundamentally arbitrary or purely conventional so that the relationship must be learnt. Examples are the morse code or national flags.
2) Index: the signifier is not arbitrary but is directly connected in some way (physically or causally) to the signified. This link can be observed or inferred. Examples of indexes are a natural sign as smoke (where there is smoke there is fire), or a measuring device, such as a thermometer.
3) Icon: the signifier is perceived as resembling or imitating the signified, by possessing some of its qualities, such as a portrait or a scaled model.

An image sign system is perceived as resembling what it depicts, thus it is mainly interpreted on an iconic basis. This is the level on which feature-based image understanding is operating on [10]. Eco extended this generalized view by showing that the signification of iconic signs is based on a socially determined reticular system of small semantic systems (codes), such as iconographic, rhetoric, or stylistic codes, and rules for their combination [8].

The meaning making of an image can thus be understood as process of the creation and interpretation of signs, where establishing the relation between the "signifier" and the "signified" represents the process of signification. How the relationships in a sign as well as in a sign system are established is to some extent idiosyncratic per user and often performed unconsciously using familiar systems of conventions. The choice is thus established through syntagmatic (surface) and paradigmatic (choice) structures [8].

Syntagmatic structures are used to analyze an image on its surface. Typical relations are spatial, either in form of structural blocks of perceptual qualifications according to which we recognize objects or recall perceived objects, or in form of figurative, forming conditions of perception such as the relationship between objects and background, contrast in light, and geometrical proportions. Such relations are then used to establish relevance values of signifiers (e.g., an object in the front is more important than another in the back). Paradigmatic analysis, on the other hand, aims to identify the pre-existing sets of signifiers that manifest the content of the image. Here signification is based on the consideration of positive or negative connotations of each signifier in the context of a thematic paradigm (e.g. public/private). Paradigmatic signification establishes relations that manifest the contrast between signifiers that belong to the same set from which those used in the image. As a result of the paradigmatic signification

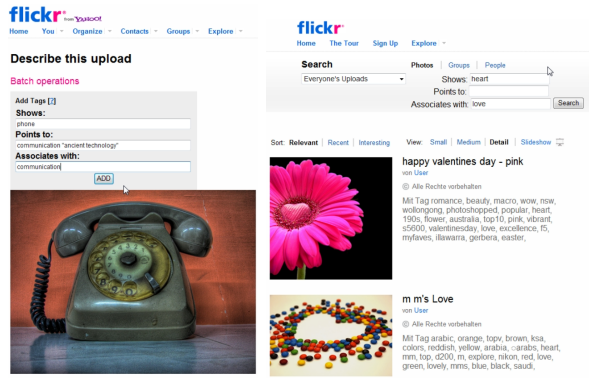

Fig. 1. Interfaces for semiotic Annotation (left) and Search (right) in SemioTag

an image can also work on an indexical level as it establishes association by contiguity (here the image can play, e.g., the role of a visual metaphor, where certain qualities from one sign point to another).

However, codes can only realize their full potential impact if there is an awareness of them and if they can relate to existing knowledge, which is applied on cognitive as well as procedural level [11]. Here lies the essential problem of tagging: It is impossible for a system to determine the divisional concept of a sign merely based on different linguistic representations, as the image will portray specific qualities about the object for which the word is simply inadequate.

A solution to this problem is a taxonomic or ontological approach, where explicit schemata are used to overcome the semantic problem (see professional image repositories). However, those complex hierarchies, which also describe some of Eco's reticular system of small semantic systems, divert from the ease of use that make tags so popular. If the relationship between the signifier, taken from the image, and the signified (the tag) can be classified, then a signification with respect to the role of the tag can be established. Making use of the symbol, index, or icon potential of a tag, one can ignore the particular signification of the tag in the set of "codes" but can utilize its mode for getting a better understanding of how an image is perceived by a user.

III. SemioTag, a "Semiotic Tagging"-System

Based on the outlined semiotic concepts we created SemioTag, a Flickr like environment with semiotics-inspired annotation and search interface, as depicted in Figure 1. SemioTag applies the semiotic idea of assigning modes of relationships between the sign vehicle and its referents. We intentionally restrict ourselves to the three divisional concepts, namely icon, index, and symbol, to reduce the complexity of the annotation as well as querying task, assuming that the representation is strong enough to trigger the most appropriate idiosyncratic relations in the user's conceptual system. Thus, a sign does never exclusively represent one concept but it will emphasize one (see discussion in Section II).

The intention of SemioTag is not to establish a completely new infrastructure (that is why it looks like Flickr) but rather to stay as close as possible to the already established tagging processes for annotation and querying. However, what we intended was to extend the semantic scope of tags. As result, we established a system internal tag representation in form of a triple of booleans, e.g., dog = (symbolic = no, iconic = yes, index = no), where the instantiation of the triple represents the use of a tag for a particular image by a particular user based on a particular syntagmatic and paradigmatic context. Apart from a larger number of tagging fields nothing changed with respect to the annotation or query process. Users can assign as many tags to an image in whatever order, as long as they place a tag in one of the three slots. In the case of using tags for the retrieval of images users can also make use of an unlimited number of tags in whatever order as long as they specify which role a tag should play in that query.

It needs to be mentioned that the two SemioTag interfaces do not use the terms index, icon, and symbol but rather represent the different slots with descriptive action labels, namely "Show" (for icon), "Points to" (for Index), and "Associates with" (for Symbol). In pre-evaluations of the interface we detected that users showed difficulties in conceptualizing the semiotic terms. We experimented with a number of alternatives and the transformation of the label in from of a concept into an "action" was perceived as best match.

The retrieval algorithm of SemioTag mimics Flickr's "most interesting" search algorithm. Thus, a user of SemioTag receives a list of images ordered in a similar way as he would have received from Flickr. Although SemioTag is quite simple, we hypothesized that it poses additional effort on the user. We were interested in comparing the user's effort on tagging and querying between the standard approach used in Flickr and the semiotics-inspired approach in SemioTag. We wanted to know in particular if the semiotic approach at the end resulted in a perceived improvement of retrieval results in a context of search where the searcher has an idea about what to search for but is satisfied by finding a relevant match. Finally, we were interested to investigate how users experience the different tasks under the two conditions. In the following two Sections IV and V, we describe the evaluation design and procedure and the evaluation results and interpretation.

IV. Evaluation Design and Procedure

Twelve MSc students and Phd students of computing science (two female, 10 male) participated in the experiment. Their age was between 24 and 34 (avg.=29.75, sd=2.73). All subjects were familiar with tagging systems in general and had experienced Flickr before. Three subjects actively used Flickr for uploading and tagging content. Half of the subjects used other social media platforms for sharing and tagging content. Their self judgment about their expertise on semiotics on a 5-point-Likert scale (1 = none, 5 = expert) revealed no or very limited knowledge on semiotics (avg.=1.42, sd=0.67). None of the subjects had any knowledge about SemioTag prior to the experiment.

A. Dataset

The novelty of our approach required the generation of a data set that provided images being tagged with the triple tag structure. Based on semiotic theory literature [12], [8], we defined a set of three times 11 tags covering the semiotic sign categories icon, index, and symbol. The tags assigned for each category were:

- Icon : dog, cat, house, apple, skyline, sunset, phone, ferrari, paris, rose, and money.
- Index : heart, smoke, thunder, warmth, time, justice, knowledge, belief, greed, communication, and bank.
- Symbol trust, pandora's box, apocalypse, beauty, love, beelzebub, ganymede, ghost, freedom, god, and capitalism.

For each of the 33 tags, we crawled 50 images from Flickr from the "most interesting" category. We chose this category because in here images were already ranked based on various aspects, such as social media (click-throughs, comments by whom and when, favorite by whom), tag use and number, etc. In that way the retrieved images represented a set that addressed iconic, indexical, and symbolic aspects, stronger than those provided by Flickr's simple or advanced search. When crawling the 50 images for the 33 tags, we have stored the ranking of the images in the result set. Thus, this rank provides the order by which Flickr has provided the most interesting images for a given tag. We use this rank for the retrieval process in our querying tasks during the experiment. In total we retrieved by this 1650 images and their associated tags. The tags were processed by steming and checking their compliance to the English version of Wordnet [13] and Wikipedia[6]. We also removed special characters such as +, -, &, !, and tags that represented automatically generated metadata, like camera model, lenses data, and so forth. We also removed duplicated images. From the remaining 1354 images we took the remaining set of tags (42,587) and asked an expert to classify the tags for each image manually into the semiotic categories icon, index, and symbol using a graphical tool. This manual classification had been conducted on the assumption what the original author might have intended with each tag in the first place. The established set served as the ground truth dataset.

B. Experiment

We conducted a repeated measures design [14], where the subjects had to perform 12 tagging tasks and 8 querying tasks under each condition, i.e., the standard Flickr interface and SemioTag. For the annotation tasks subjects were provided with images instead of asking them to utilize their own. In that way, we could control the distribution of the semiotic categories. For the search tasks, we provided the structure (slots) and the reason for the query. Thus we controlled the condition but the final query had to be designed by the subjects. They also had to decide about the image they felt to be relevant enough to answer the query they have posted. The

[6]http://www.wikipedia.org/, last visited 11.01.2014

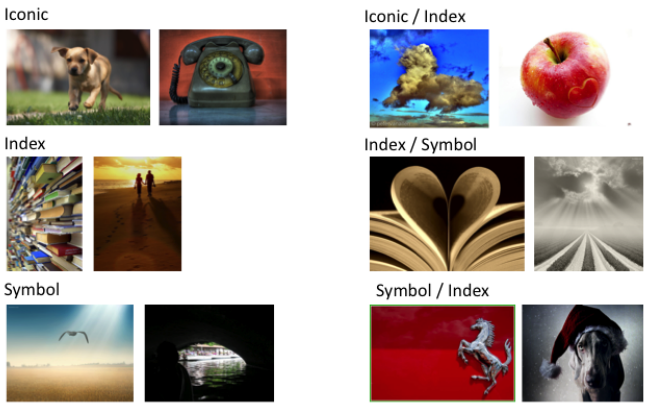

Fig. 2. The Images used for our Annotation Tasks

setup for both conditions followed the Latin-Square design. Subjects performed the tasks individually in the lab, accompanied by the instructor. A pre-evaluation was performed with two subjects to verify the experiment procedure.

C. Procedure

The instructor provided the subjects an introductory sheet that explained the basics of semiotics, in particular the categories icon, index, and symbol. Each of these categories was introduced with images and related tags as example. The subjects were allowed to ask as many questions until being comfortable with the three semiotic concepts. The subjects were also told that he can make use of the introductory sheet at any time during the experiment. The subjects were informed about the general procedure of the experiment and encouraged to verbalize their thoughts ("think aloud"). In addition, the subjects were told that they do not need to conduct the tasks as fast as possible but that they should perform the tasks simply best to their knowledge and abilities.

The tagging tasks were applied on 12 images selected from the ground truth data set. The images were selected based on their semiotic roles. We specifically selected two images for each semiotic role individually and combinations of the roles as depicted in Figure 2. For the tagging tasks, the subjects were asked to provide as many tags as they felt like. To counterbalance any carry-over effect, we randomized under both conditions the order in which the images where presented. Conducting the tagging tasks took between 18.74 to 56.6 minutes (avg.=31.73, sd=12.28). Tasks performed under the original Flickr interface condition made use of the original tag set of 42,587 tags without the triple representation of the three semiotic categories. SemioTag used the ground truth set.

For querying, the participants had to conduct 8 tasks to find images in the dataset. These were structured in 5 different groups:

- **Iconic category**:
  - Search for a picture that depicts a rose.
  - Search for a picture that depicts a phone.
- **Index category**:
  - Search for a picture that depicts capitalism.
  - Search for a picture that depicts justice.
- **Symbolic category**:
  - Search for a picture that depicts beauty in a unconventional way.
  - Search for a picture that depicts human freedom.

To examine the strength of the semiotic search, the subjects were also asked to query mixed categories as follows.

- **Index and iconic**:
  - Search for a picture that depicts love through something.
- **Index, iconic and symbolic**:
  - Search for a picture of an inspiring bird that depicts hope.

Subjects were informed to perform each querying task sequentially and select the image they think best matches the query. They were told that there is no wrong and right answer regarding the image they select. The subjects were also told that the querying task is not conducted on their own annotations of images but on a Flickr sub set. After each set of tagging and querying, the subjects filled out a usability and experience questionnaire in which they were asked to agree on statements covering accuracy, ease of use, and confidence statements on a 5-Point-Likert scale. After a subject had finished all tasks a concluding questionnaire had to be filled in which the subjects were asked to compare the different approaches and provide specific feedback regarding the use of semiotic theory. Each subject received 10 Euro as compensation for their participation.

V. Evaluation Results and Interpretation

The results presented in this section address the differences between annotating and querying behavior of subjects using the original Flickr interface and SemioTag. The statistical test performed on the observed data are the Shapiro-Francia Test to verify if the data is normally distributed. Depending on its outcome, we either applied Repeated-Measure-ANOVA (RM ANOVA) in cases of normal distribution or the Friedman Test for non-normal distribution of data samples. We first investigate the tagging behavior of the subjects. Subsequently, we provide evaluations of the querying behavior. A particular section is devoted to the evaluation of the eye tracking of subjects that sheds additional insights on the tagging and querying behavior observed for both conditions.

A. Tagging

For getting a better understanding of the influence of the semiotic approach on image tagging, we have collected quantitative as well as qualitative data. The quantitative data provides information about the time needed to tag under each condition as well as the amount of tags generated by each subject under each condition. The qualitative data was collected in form of a questionnaire that investigated, e.g., appropriateness of each interface to support the task with respect to accuracy, intuitiveness, workload, and so forth.

Figure 3 shows the time in seconds needed for tagging the 12 images (see data set described in Section IV-A) under the two conditions.Tagging the images under the semiotic condition takes on average 219% more time than with the normal tagging. Applying the Friedman Test shows that all differences but image 11 are significant (T critical = 4.84). Figure 4 shows the number of tags assigned to the 12 images under the two conditions. Semiotic tagging generated on average 9.92 more tags per image. For half of all tasks the increase was significant (Friedman Test, T critical=4.84, RM ANOVA, T critical=2.628).

Having a closer look at the duration of the tagging process it can be noticed that once subjects got used to semiotic tagging the difference stabilized around twice the time against standard tagging. Thus, the novelty aspect, as represented by the highest difference with 552% for image 1 in Figure 3, quickly fades. An explanation for the significantly longer duration of the annotation process in semiotic tagging is that the cognitive process of establishing a relation between tag and sign in the image, as described in [15], [16], [8], [17], becomes more explicit as the semiotic interface enforces the subjects to think and reflect about where to place a tag. As subjects were encouraged to think aloud while performing the annotation tasks it could be observed that most were undecided in which semiotic input field they should insert their tag. One example was the tag "Christmas". In this case, the subject could not decide whether to add it to the index category or to the symbolic category. Another subject struggled with the language and asked about the term of the "water-city in Italy". Once "Venice" was named, he associated this name with additional concepts in the realm of travel and holidays, which only started after he had the term "Venice" established. This is an indicator that semiotics-based annotation of images stimulates mental association but depends on the language proficiency of the individual and the individual ability of the user to apply principles such as similarity, contiguity, and contrast. This is also supported by the observation that in particular the images categorized as indexical or symbolic showed a significant increase of tags under the semiotic condition.

In order to assess the quality of the obtained semiotic tags, we asked the expert who generated the ground truth dataset to also evaluate the tags provided by the subjects with respect to their correctness, anticipating the subjects' tasks. Figure 5 shows the amount of correct and incorrect annotations for the annotation tasks covering the 12 images. The incorrect numbers describe cases where the potential role was not met.

The results demonstrate that applying all three categories only an average of 77% of the annotations were correct. Having a closer look at the tasks in which the lowest values were obtained, it becomes apparent that the subjects had a problem to distinguish between the role of index and symbol. Thus, we performed a second evaluation where we joined both categories into one, resulting into an annotation accuracy of 94% on average. The conclusion we draw from this, however, is that the triple distinction established in semiotics does not correspond with the actual annotation reasoning of our subjects, who are all trained in analytic and reflective thinking. For a final interface this finding means that potentially only two slots should be provided as otherwise a higher annotation accuracy failure is introduced into the annotation process. For the study of semiotics this finding indicates to revisit this aspect of its theory and, based on additional experiments with a larger population, potentially alter the theory.

We also asked the subjects to fill in a questionnaire. Figure 6 presents the results for the annotation tasks in section T1 to T7, where each of the questions had to be answered on a 5-Point-Likert scale [18], from 1 = fully disagree to 5 = fully agree.

The findings underpin the observations gained from the quantitative analysis. The overall stronger positive experience of the standard annotation method can be attributed to the familiarity subjects had with it. The differences between the two conditions in T1 to T4 are not significant and hence it can be stated that semiotic tagging seems to be equally appropriate in tagging images (T1; T4) and can capture the semantics of an image as accurately as normal tagging (T2). Though we had expected that it would at the end rank higher in this respect. It would be interesting to investigate in a long time study if subjects who make use of semiotic tagging change their opinion. The results for T5 to T7 support the observations made during the performance of annotation tasks and the think aloud statements subjects demonstrated struggle and insecureness about the different semiotic concepts. The subjects also reflected on the pros and cons of placing tags in either the index or symbol slot resulting in a view that either could be good and hence a feeling of uncertainty that the final choice is the right one.

B. Querying

We also wanted to gain a better understanding of the influence of a semiotic approach on querying images. The queries were performed on the dataset described in Section IV-A. We were not interested in identifying the precision or recall value of each query, but rather how long it would take a subject to retrieve one image to be considered relevant to the query, where we apply Birger's definition [19] of relevance: "Something (A) is relevant to a task (T) if it increases the likelihood of accomplishing the goal (G)." However, we identified relevant images in the dataset and compared the choices made by the subjects with those we had established as correct results. Based on that, we first analyze the time needed for finding the images under the normal and semiotic condition (see Figure 7). Subsequently, we look into the number of queries executed for each querying task (see Figure 8).

The figures show that besides iconic search (see task 1 and 2), subjects needed roughly two queries to retrieve a relevant image. An exception under both conditions is that at least one query in the indexical as well as symbolic category were considerably higher. This is an indicator that subjects in general had a problem with the index and symbol roles. Overall it can be stated, except for search task 7, that no

		Iconic		Index		Symbol		Iconic/Index		Iconic/Symbc		Index/Symbol	
	Picture #	1	2	3	4	5	6	7	8	9	10	11	12
Normal Tagging	Minimum	4.18	8.61	4.94	9.99	5.66	8.69	5.22	6.42	4.27	8.15	6.05	6.82
	Maximum	33.85	96.09	88.27	64.05	61.29	71.64	49.91	45.20	33.46	93.56	46.40	70.77
	Average	13.37	26.84	22.08	23.87	20.24	27.50	20.64	16.04	15.57	27.74	19.11	21.51
	Standarddev.	10.92	26.62	24.77	17.59	16.99	22.50	16.63	12.36	10.92	23.93	13.66	21.88
Semiotic Tagging	Minimum	39.15	13.23	15.53	20.82	23.57	18.46	16.60	11.07	12.48	15.33	13.36	11.83
	Maximum	106.04	116.60	96.56	148.47	88.51	353.93	69.17	136.26	90.92	88.46	48.60	122.00
	Average	73.82	43.88	41.84	48.09	38.41	69.35	34.51	31.77	32.18	43.53	28.25	43.69
	Standarddev.	27.70	30.20	26.57	39.34	17.38	91.18	15.30	35.22	23.89	22.29	10.45	30.42
	Difference	552%	163%	189%	201%	190%	252%	167%	198%	207%	157%	148%	203%
	Friedman T1	**12.00**	**8.33**	**8.33**	**8.33**	**8.33**	**5.33**	**8.33**	**8.33**	**12.00**	**5.33**	1.33	**12.00**
	Friedman T2	-	**25.00**	**25.00**	**25.00**	**25.00**	**8.80**	**25.00**	**25.00**	-	**8.80**	1.38	-

Fig. 3. Time for Normal Tagging vs. Semiotic Tagging

	Iconic		Index		Symbol		Iconic/Index		Iconic/Symbc		Index/Symbol			
Picture #	1	2	3	4	5	6	7	8	9	10	11	12	Avg.	Stdev.
# Normal Tags	29	28	25	50	42	38	33	38	29	39	34	45	35.83	11
# Semiotic Tags	32	34	37	67	47	50	43	43	44	57	47	48	45.75	10
Difference	3	6	12	17	5	12	10	5	15	18	13	3	9.92	5
Friedman T1	0.14	1.00	**5.44**	4.45	2.77	**4.50**	2.27	1.60	**10.00**	3.60	**7.36**	1.80		
Friedman T2	0.13	1.00	**9.13**	**6.69**	3.13	**6.60**	2.57	1.69	**55.00**	4.71	**17.47**	1.94		
RM-ANOVA-T					1.77	**3.88**								

Fig. 4. Number of Tags in Normal Tagging vs. Semiotics Tagging

	Iconic		Index		Symbol		Iconic/Ind.		Iconic/Symb.		Index/Symb.			
Picture #	1	2	3	4	5	6	7	8	9	10	11	12	Avg.	Stdev.
Correct: Symbolic	5	2	0	0	10	6	8	2	1	9	2	3	4.00	3.52
Iconic	17	13	13	42	30	30	18	24	22	35	18	20	23.50	9.03
Index	6	8	10	9	1	1	11	5	6	6	11	9	6.92	3.42
Incorrect: Symbolic	0	2	0	0	0	8	1	2	2	1	2	3	1.75	2.22
Iconic	1	1	0	3	5	5	3	4.5	5	2	3.5	2	2.92	1.74
Index	0	8	10	12	1	1	2	5.5	8	4	9.5	9	5.83	4.12
P (3 categories)	97%	68%	70%	77%	87%	73%	86%	72%	66%	88%	67%	70%	77%	
P (2 categories)	97%	97%	100%	100%	89%	90%	93%	90%	89%	96%	92%	96%	94%	

Fig. 5. Correctly and Incorrectly Classified Tags and Precision P for Three Semiotic Categories as well as Two Semiotic Categories (if Index and Symbol are Merged)

	Normal		Semiotic		Friedman	
	Avg.	Stdev.	Avg.	Stdev.	T1	T2
T1 Appropriate support in conducting the tagging tasks.	3.08	1.38	2.83	1.34	0.14	0.13
T2 Semantics of the media assets can be accurately captured.	2.58	1.00	2.50	1.00	0.00	0.00
T3 Tagging images from other people was significantly harder.	3.40	1.26	3.70	1.42	0.33	0.31
T4 The tagging approach is suitable for the annotation task.	3.33	0.89	3.00	0.74	1.28	1.32
T5 The tagging approach is intuitive to use.	4.42	0.67	2.67	1.50	**9.00**	**33.00**
T6 Working on the tasks to tag the images was tedious.	2.33	0.90	3.58	0.90	**15.40**	**4.84**
T7 I am confident to have assigned the right tags to the images.	3.17	0.83	2.33	0.98	**5.44**	**9.13**

	Normal		Semiotic		Friedman	
	Avg.	Stdev.	Avg.	Stdev.	T1	T2
Q1 I had control over the system's behavior while querying.	3.17	1.19	3.42	1.16	1.29	1.32
Q2 System provides appropriate support for the querying tasks.	2.83	1.27	3.08	0.90	0.67	0.65
Q3 The search by tags returns precise result sets.	2.17	1.03	2.50	0.90	2.67	3.14
Q4 I have found all media assets that I was expecting.	2.33	1.15	2.25	0.87	0.00	0.00
Q5 Querying by tags is intuitive to use.	3.50	1.00	3.42	0.90	0.50	0.49
Q6 Working on the task to query the images by tags was tedious.	2.67	1.15	3.08	1.08	1.80	1.94

Fig. 6. User Ratings for the Annotation Tasks and Quering Tasks

		Iconic		Index		Symbol		Ic./In.	All	
	Task #	1	2	3	4	5	6	7	8	
Normal Querying	Minimum	11.05	12.39	22.17	10.45	25.61	19.96	32.66	16.14	
	Maximum	77.03	126.25	194.50	461.04	170.88	132.40	187.33	152.12	
	Average	23.93	33.77	63.77	96.53	85.26	71.68	93.13	66.25	
	Standarddev.	18.95	32.69	45.93	133.78	44.97	43.15	52.20	40.88	
Semiotic Querying	Minimum	7.75	17.23	14.48	16.19	24.77	20.99	30.03	8.94	
	Maximum	27.88	54.48	103.03	150.08	188.22	129.70	240.21	138.28	
	Average	14.25	33.02	56.96	51.91	97.63	55.73	80.90	39.31	
	Standarddev.	5.83	11.39	27.86	37.48	57.40	26.15	56.72	35.53	Average
	Difference	60%	98%	89%	54%	115%	78%	87%	59%	80%
	Friedman T1	1.33	0.33	0.33	0.09	1.33	0.09	0.00	0.33	
	Friedman T2	1.38	0.31	0.31	0.08	1.38	0.08	0.00	0.31	

Fig. 7. Time Needed for Querying under both Conditions

		Iconic		Index		Symbol		Ic./In.	All	
	Task #	1	2	3	4	5	6	7	8	
Normal	Average	1.08	1.08	1.91	2.45	1.91	2.36	1.72	2.17	
Querying	Standarddev.	0.29	0.29	1.50	2.93	1.31	1.87	1.35	2.37	
Semiotic	Average	1.08	1.33	1.92	2.55	3.08	2.60	2.25	2.17	
Querying	Standarddev.	0.29	0.65	1.38	2.57	2.75	1.14	1.60	1.90	Average
	Difference	100%	123%	101%	104%	161%	110%	131%	100%	116%
	Friedman T1	0.00	3.00	0.67	1.00	1.00	0.00	4.00	0.14	
	Friedman T2	-	3.67	0.65	1.00	1.00	0.00	**5.50**	0.13	

Fig. 8. Amount of Queries Passed under both Conditions

significant difference can be established between the amount of queries used under both conditions. Despite the fact that subjects performed on average more queries under the semiotic condition they still performed around 20% faster than with the standard interface, though the differences are not significant. Thus, the larger tagging time is counterbalanced by the faster query time.

As semiotic tagging has produced a larger set of tags, it could be argued that the mere richness provides the means for faster retrieval. However, the dataset used for the query tasks was the ground truth set described in Section IV-A that only used tags already available in Flickr. As the subjects did not perform the queries on their own annotations as described in the experiment design in Section IV-B we even had expected to see that subjects would need more time and queries because they made use of tags that were not in the dataset. This was obviously not the case.

In accordance with the tagging procedure, we also asked the subjects to fill in a usability and experience questionnaire at the end of each query task. Figure 6 presents the results for the annotation tasks in section Q1 to Q6, where each of the questions had to be answered on a 5-Point-Likert scale [18], from 1 = fully disagree to 5 = fully agree. The results underpin the observations gained from the quantitative analysis. Though the subjects still found that semiotic querying is more complicated (Q6), we found that semiotic querying provided subjects with a stronger feeling of control on and more appropriate method for querying images with tags (Q1, Q2). More importantly, the subjects felt more certain that they worked with the right tags (Q6), which contrasts strongly against the inferior experience mentioned associated with semiotic tagging. It seems that especially for querying the provided emphasizing structure positively influences the searching behavior of the subjects.

When querying for images, the subjects also started exploring the possibilities of semiotic search. One subject in particular tried to match the search results gained through standard search with those provided by semiotic search. He experimented with a number of iconic terms, such as money and rose, and realized that once they were applied to different slots the overlap altered. This helped him to better understand the underlying notion of semiotic search. Another subject did not enter any of the query terms into the slots but only used synonyms or associations with that query. Two subjects made use of different queries to find a matching image. Thus they established relevance with a change of category and related similarity in the result test.

C. Questionnaire

At the end of the experiment, we asked subjects to fill an experience questionnaire that focused on the annotation and querying experiences gained using the semiotics-based interface. In the questionnaire the subjects had to value seven statements C1 to C7 on a 5-Point-Likert scale [18], from 1 = fully disagree to 5 = fully agree. The results are shown in Figure 9.

The subjects consider a semiotic approach for image tagging provides a more precise representation of what an image might mean (C1). However, having used the semiotic approach for annotating as well as for searching images the subjects still felt that the new approach hindered them in performing the tasks (C2). Moreover, they were hesitant in imagining that a frequent use of semiotics would improve their performance (C3). This hesitance might be rooted in the fact that most subjects had a problem with distinguishing the different semiotic roles a tag can perform (C4–C7). In particular the difference between the role index and symbol made no sense to the subjects, which results in the suggestion to merge both roles into one for the next version of the interface.

The comments in the questionnaire's free text section confirmed the problem of subjectivity in the process of tagging as well as search. Two of the subjects reflected in particular on their problems with the querying tasks "inspiring bird that depicts hope" as they could not associate hope with a bird nor think that a bird is inspiring. In this context, five subjects commented again on their problem with the query task "unconventional beauty" as they could not conceptualize it. Others, as stated earlier, formed immediately an image of an old person in their head before they started searching. This problem of conceptualization was connected by all subjects to the form of the semiotic sign categories, which were considered too abstract. Mainly the categories index and symbol caused problems, even though subjects agreed that both provided the actual means to better capture what an image means rather than what it merely shows. However, the subjects did not believe that both categories or their union (strongly suggested) can overcome the problem of subjectivity inherited in the interpretation or envision of an image but merely soften it.

VI. Conclusion and Future Work

We presented a new approach for tagging images, called semiotic tagging, and its implementation SemioTag. The results of an experiment with 12 subjects showed that semiotic

	Avg.	Stdev.
C1 Semiotic tagging provides a more precise annotation.	3.17	1.03
C2 Tagging in semiotic roles hinders me in using the system.	3.33	0.65
C3 More frequent use of semiotics would improve my performance.	2.92	1.16
C4 Separation into the three semiotic roles is useful.	2.42	0.90
C5 Separation between icon and the other semiotic roles is clear.	3.83	1.47
C6 Separation between the semiotic roles index and symbolic is clear.	1.92	1.00
C7 Semiotic categories index and symbolic should be merged.	4.08	1.24

Fig. 9. Questionnaire Comparing Normal Tagging and Semiotic Tagging (C1–C3) and Specific Questions on Semiotic Tagging (C4–C7)

tagging asks a stronger cognitive load from users in the process of annotating an image, though the difference is not significant. It became apparent that as a side effect the subjects also provided a larger number of tags, which showed a variety of alterations among them in form of synonyms, associations, oppositions, antonyms, and so forth. It can be justified to state that a semiotic-inspired annotation process forces the user to reflect about the relation between tag and image (the relation between signifier and signified, where the tag represents the signified) and hence produces better quality annotations. However, the experiments also showed that, mainly to the inexperience of subjects with semiotics, semiotic tagging also generated a stronger insecureness about the rightness of the relation between tag and chosen category, resulting in the opinion that semiotic tagging is less intuitive to use than traditional tagging.

The experiments addressing the retrieval of images showed that the subjects felt more confident and in control with the semiotic approach compared to traditional tagging, resulting in a substantially faster process of retrieving images relevant to a query. The experiments revealed in particular that providing users with an emphasis structure that allows to value query components speeds up the retrieval process as it allows to better specify the own mental image of the potential target. It can also be stated that the saved time during the query process makes up for the additional time spend on annotations, which needs to be considered as querying of a certain data set is performed more frequently than adding new data.

The approach described in this article is a small but important step towards a better understanding of tagging in particular. The interesting aspect for further research is to better understand the intrinsic relation of ”signifier” and ”signified” in a sign and sign systems, which has an influence on the way how tagging and search will be performed, requesting research in supportive interfaces. Issues such as feedback to queries beyond the image through additional contextual data will be a challenge. However, as already been demonstrated by SemioTag, this reflection supporting complexity can stimulate creativity and hence support processes, such as investigating, evaluating, comparing, and synthesizing material.

Finally, both experiment sets indicate that the distinction manifested in semiotics between the index and symbol category could not be reproduced by the subjects. The findings should stimulate additional experiments that can empirically manifest that this distinction is or is not apparent and required for human communication and meaning making. Our hypothesis at the moment is that this distinction can be abandoned.

References

[1] I. Peters, *Folksonomies. Indexing and retrieval in Web 2.0.* De Gruyter Saur, 2009.

[2] B. Sigurbjörnsson and R. van Zwol, “Flickr tag recommendation based on collective knowledge,” in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW ’08. New York, NY, USA: ACM, 2008, pp. 327–336. [Online]. Available: http://doi.acm.org/10.1145/1367497.1367542

[3] H. Xu, X. Zhou, M. Wang, Y. Xiang, and B. Shi, “Exploring flickr’s related tags for semantic annotation of web images,” in *Proceedings of the ACM International Conference on Image and Video Retrieval*, ser. CIVR ’09. New York, NY, USA: ACM, 2009, pp. 46:1–46:8. [Online]. Available: http://doi.acm.org/10.1145/1646396.1646450

[4] T. Rattenbury and M. Naaman, “Methods for extracting place semantics from flickr tags,” *ACM Trans. Web*, vol. 3, pp. 1:1–1:30, January 2009. [Online]. Available: http://doi.acm.org/10.1145/1462148.1462149

[5] L. Wu, L. Yang, N. Yu, and X.-S. Hua, “Learning to tag,” in *Proceedings of the 18th international conference on World wide web*, ser. WWW ’09. New York, NY, USA: ACM, 2009, pp. 361–370. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526758

[6] L. Xie, A. Natsev, M. Hill, J. R. Smith, and A. Phillips, “The accuracy and value of machine-generated image tags: design and user evaluation of an end-to-end image tagging system,” in *Proceedings of the ACM International Conference on Image and Video Retrieval*, ser. CIVR ’10. New York, NY, USA: ACM, 2010, pp. 58–65. [Online]. Available: http://doi.acm.org/10.1145/1816041.1816052

[7] A. Sun and S. S. Bhowmick, “Quantifying tag representativeness of visual content of social images,” in *Proceedings of the international conference on Multimedia*, ser. MM ’10. New York, NY, USA: ACM, 2010, pp. 471–480. [Online]. Available: http://doi.acm.org/10.1145/1873951.1874029

[8] U. Eco, *Einfuehrung in die Semiotik (Introduction to Semiotics).* Wilhelm Fink Verlag, Munich, 1985.

[9] C. S. Peirce, *The Collected Papers of Charles Sanders Peirce - 1 Principles of Philosophy and 2 Elements of Logic*, C. Hartshorne and P. Weiss, Eds. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1960.

[10] C. G. M. Snoek and A. W. M. Smeulders, “Visual-concept search solved?” *IEEE Computer*, vol. 43, no. 6, pp. 76 – 78, June 2010.

[11] F. Nack and L. Hardman, “Denotative and connotative semantics in hypermedia: proposal for a semiotic-aware architecture,” *New Rev. Hypermedia Multimedia*, vol. 7, no. 1, pp. 7–37, 2002.

[12] D. Chandler, *Semotics: The Basics*, 2nd ed. Routledge, March 2007.

[13] G. A. Miller, “Wordnet: a lexical database for english,” *Commun. ACM*, vol. 38, pp. 39–41, November 1995. [Online]. Available: http://doi.acm.org/10.1145/219717.219748

[14] A. Field and G. Hole, *How to design and report experiments*. Sage publications Ltd., 2003.

[15] R. Arnheim, *Art and Visual Perception: A Psychology of the creative eye.* Faber & Faber., 1956.

[16] S. Heath., Ed., *Image, Music, Text - Essays selected and translated by Stephen Heath.* London: Fontana Press, 1977.

[17] J. Aumont, *The Image*, C. Pajackowska, Ed. BFI Publishing, 1997.

[18] J. Lazar, J. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Wiley, 2009. [Online]. Available: http://books.google.de/books?id=H_r6prUFpc4C

[19] B. Hjørland, “The foundation of the concept of relevance,” *JASIST*, vol. 61, no. 2, pp. 217–237, 2010.