# A CONVERSION PROCESS FROM FLICKR TAGS TO RDF DESCRIPTIONS

Mohamed Zied MAALA
*France Telecom R&D*
*38-40 rue du General Leclerc, 92130 Issy-les-Moulineaux Cedex 9, France*
*zied.maala@orange-ftgroup.com*

Alexandre DELTEIL
*France Telecom R&D*
*38-40 rue du General Leclerc, 92130 Issy-les-Moulineaux Cedex 9, France*
*alexandre.delteil@orange-ftgroup.com*

Ahmed AZOUGH
*France Telecom R&D*
*38-40 rue du General Leclerc, 92130 Issy-les-Moulineaux Cedex 9, France*
*ahmed.azough@orange-ftgroup.com*

**ABSTRACT**

The recent evolution of the Web, now designated by the term Web 2.0, has seen the appearance of a huge number of resources created and annotated by users. However the annotations consist only in simple tags that are gathered in unstructured sets called folksonomies. The use of more complex languages to annotate resources and to define semantics according to the vision of the Semantic Web, would improve the understanding by machines and programs, like search engines, of what is on the Web. Indeed tags expressivity is very low compared to the representation standards of the Semantic Web, like RDF and OWL. But users appear to be still reluctant to annotate resources with RDF, and it should be recognized that Semantic Web, contrary to Web 2.0, is still not a reality of today's Web. One way to take advantage of Semantic Web capabilities right now, without waiting for a change of the annotation usages, would be to be able to generate RDF annotations from tags. As a first step toward this direction, this paper presents a tentative to automatically convert a set of tags into a RDF description in the context of photos on Flickr. Such a method exploits some specificity of tags used on Flickr, some basic natural language processing tools and some semantic resources, in order to relate semantically tags describing a given photo and build a pertinent RDF annotation for this photo.

**KEYWORDS**

Web 2.0, tags, RDF, annotation generation.

## 1. INTRODUCTION

Web 2.0 and Semantic Web are two trends influencing the evolution of the Web since several years. Web 2.0 consists in a greater collective content creation and a larger social interaction between users. A huge number of resources have been created by users and annotated by them. This is a major change compared to the original web, where collective creation was much less developed. Many resource repositories like Wikipedia [1], Del.icio.us [2] and Flickr [3] have appeared and gather millions of user created pages, bookmarks and photos. However up to now the annotations made by the users on these resources consist only in simple tags, that are gathered in unstructured sets called folksonomies and thus do not convey a formally defined semantics. Therefore they do help to improve queries on the Web, but not so much, since resources will be found only if the query syntactically matches a tag they are annotated by.

Contrary to Web 2.0, Semantic Web is still a vision and not yet a reality. It is based on the idea that describing resources with symbolic annotations (using vocabularies defined in formal ontologies) will enable machines and tools to understand their semantics and will improve the pertinence of tasks such as query answering.

This paper focuses in the study of the plateform Flickr [3], a photo sharing website and web services suite. Flickr [3] was developed by Ludicorp, a Vancouver, Canada-based company founded in 2002. Ludicorp launched Flickr in February 2004. In March 2005, Yahoo! Inc. acquired Ludicorp and Flickr. Flickr allows photo submitters to categorize their images by use of keywords "tags" (a form of metadata), which allow searchers to easily find images concerning a certain topic such as place name or subject matter.

Flickr [3] provides rapid access to images tagged with the most popular keywords. Flickr also allows users to categorize their photos into "sets", or groups of photos that fall under the same heading. However, sets are more flexible than the traditional folder-based method of organizing files, as one photo can belong to many sets, or one set, or none at all (the concept is directly analogous to the better known "labels" in Google's Gmail). Flickr's "sets", then, represent a form of categorical metadata rather than a physical hierarchy.

This paper interests more exactly to the study of Flickr tags and present a new method to convert Flickr [3] tags describing a picture into RDF annotations describing it semantically. This method can be viewed as the first step enabling to transform resources described using tags to a semantic description describing the same resources or can be viewed too as a first bridge between the web 2.0 and the semantic web. This method is based on linguistic rules, on natural language treatment, on integrating some human knowledge to be able to provide semantic description for pictures from tags. To the best of our knowledge, few works exist enabling conversion from tags to semantic description. One work [4] exists that "converts" Flickr tags to RDF descriptions, but gives bad results because Flickr tags are transformed into RDF topics in a fully syntactic way without extracting the semantic of the tags. Our method helps a user to understand picture tags and to found relationships between them.

This paper contains five sections. Section 2 presents the different ways of tagging pictures used in Flickr. Section 3 introduces our conversion method from tags to RDF semantic description. Section 4 describes related work and compares our method with existing approaches.

## 2. SURVEY ON WAYS OF TAGGING PICTURES IN FLICKR

Before conceiving a method to generate a RDF description from tags on Flickr, it is useful to know the specificities of photo tag annotations. Therefore we have attempted to analyse the different ways users exploit Flickr annotation capabilities in order to tag photos.

## 2.1 Tagging habits

The following tagging habits can be distinguished:

- Very few tags: unfortunately too many photos contain no tag at all or very few tags (one or two such in figure 1). In this case, it is impossible or very difficult to generate a RDF description.
- Sentence tagging: users can use quotes to enter a full sentence as a tag such as in figure 2 (in case no quotes are used, space is understood by Flickr as a separator between tags).
- Vertical sentence tagging: it is the same case as the previous one, but users forgot to (or intentionally did not) put the sentence between quotes. Thus the sentence can be read vertically, because Flickr has understood each space separated word to be a different tag (such as in figure 3).
- Too many tags: contrary to the previous case, the information attached to the photo is very rich (as in figure 2) and describes many different aspects (content, location ...). The difficulty for

generating a RDF description lies in finding the relevant associations between the tags (for instance which noun is subject of which verb).

- Non-sense tags: these tags correspond to something not understandable for a human being not knowing the annotator universe of thinking such as in figure 2 (for instance the tag *noneof100#2*). It could for instance be a nickname of some people on the photo, or of a location...
- Space free tagging: the users write a sentence by concatenating words in order to put the whole sentence on the same line ; for example in figure 2 a user has written the tag "I love nature". These users may not be aware of the possibility of using quotes.
- Collective tagging: due to the interface Flickr provides (see figure 4), it is possible to tag several photos concurrently. Therefore it sometimes happens that a photo is described with a tag that does not apply directly to it but to a photo that has been uploaded at the same time. The photo the tag applies to belongs to the previous five or next five photos of the current photo in the "photostream" (as six photos can be concurrently tagged).

Fig 1. Use of few tags
*Tags{Hawai, Tourist}*

Fig 2. Use of sentence as tags
*Tags{Paya Lake, Makra top, Kaghan valley, nature, water, I love nature, wow, noneof100#2, top-v111, top-v1111, deleteme, saveme, saveme2, saveme3, deleteme2, deleteme3, saveme4, saveme5, deleme4, saveme6, deleteme4, deleteme5, deleteme6, deleteme7, deleteme8, saveme7, deleteme9, saveme8, deleteme10, saveme9, Most, bravo, Big Fave, Outstanding shots }*

Fig 3. Tag with vertical sentence
*Tags{ Here, some, more, photos, off, Hudson, River, New, Jersey }*

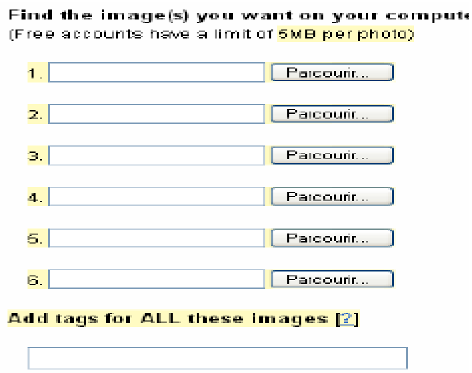


Fig 4. Collective tagging due to pictures upload user interface

Moreover many tags contain typing errors, due perhaps to a too high typing speed or to a lack of knowledge of correct typing. The figure 5 presents a histogram of photo tags number on a sample size of one thousand photos. This figure shows more precisely the distribution of the number of keywords.

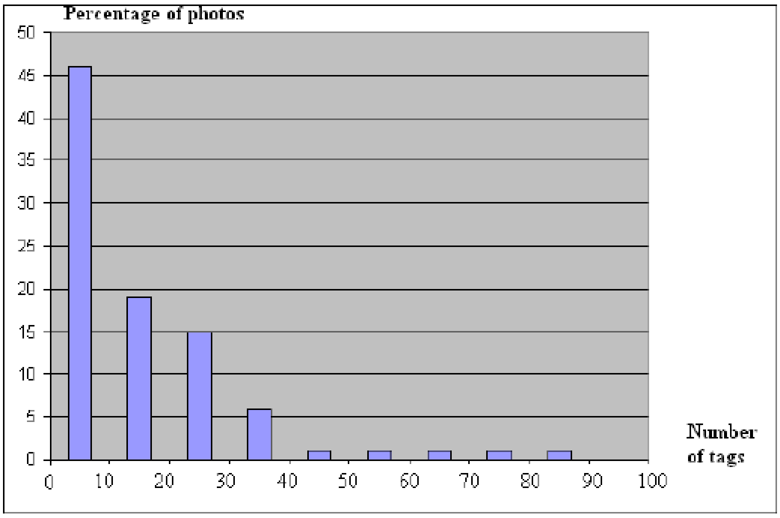


Fig 5. Distribution of the number of keywords

## 2.2 Flickr interface

To know which tags are the most employed by users and their links with pictures, we studied popular tags presented in figure 6. On 145 popular tags, 10% describe a celebration (birthday, Christmas), 13% are related to a date (June, July...), and 62 tags (about 42% of the total) express places, on 20 tags are names of common places (mountain, house, garden,...) and 42 tags (about 28% of the total) are name of countries or towns (Canada, Japan, Paris,...). Among the popular tags, some are related to the camera (Nikon, cameraphone ...).


Fig 6. Popular tags

Table 1 shows percentage of photos by category.

Table 1. Photo percentage by category

| Tag category | | Percentage of photos | |
|---|---|---|---|
| Place | Country | 38 | 71 |
| | Landscape | 42 | |
| | Building | 10 | |
| Time | Year/Season | 20 | 20 |
| | Month/Day | 5 | |
| Event | | 11 | |
| Name | | 35 | |
| Camera | | 17 | |
| Action | | 53 | |
| Non sense | | 49 | |

The next section presents more precisely what the tags describe exactly about the photo.

## 2.3 Tagging content

An analysis of a sample of one thousand photos shows that the tags can be clustered in the following groups:

- Place: the location can be described at very different levels of granularity. At the largest level of granularity, the continent, the country, the region, the city, a mountain range . . . are found frequently. At a smaller level of granularity, description of the building or the immediate natural site the photo has been taken in can be found: a building, a university, a house, a beach… Finally at the smallest level of granularity, there can be a description of a room or a piece of furniture: bed, chair . . .
- Time: the time can also be described at different levels of granularity. The year, the season and the month are the most frequently found. The exact day is much less frequent. Some times of the day are (sunrise, sunset…).
- Event: the holy days (Christmas . . .), the birthdays, the weddings . . .
- Name: people names (Emma, Jean . . .), nicknames. . .
- Camera: many tags indicate the make or the model of the camera (Nokia, Canon . . .), the colors (black & white ...), artistic judgments on the photo . . .

This knowledge of the way people tag photos on Flickr gives an indication on the natural language processing tools and semantic resources that are needed in order to be able to transform a set of tags into an RDF annotation. The process of automatically generating RDF annotations is now described in next section.

# 3. CONVERSION PROCESS OF FLICKR'S TAGS TO RDF ANNOTATION

This section introduces a method to convert tags describing a photo into a RDF annotation. This method can be viewed as a first tentative to transform web 2.0 annotations into semantic web annotations. The problem can also be viewed as transforming a bag of tags into a relational description. This method mainly relies on detecting the category each tag belongs to, among a set of six categories (location, time, event, people, camera, activity). Using this set of categorized tags, it then tries to identify the possible arguments of verbs (verbs are in the category denoting activity) in infinitive or present participle form. This method thus applies only on photos described by tags when some of them are verbs. In the next sections, the components needed in the conversion process are described one by one.

## 3.1 CONVERSION PROCESS COMPONENTS

Automatic conversion from photo tags to RDF annotations is a difficult task. This process essentially requires several components: some basic natural language processing tools (mainly a stemmer), and semantic resources like Wordnet, semantic nets and specialized databases containing knowledge on specific subjects (for instance locations, cameras ...).

### 3.1.1 Wordnet

WordNet [7] is a semantic lexicon for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The purpose is twofold: to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications. The database can also be browsed online. WordNet [7] was created and is being maintained at the Cognitive Science Laboratory of Princeton University under the direction of psychology professor George A. Miller. As of 2006, the database contains about 150000 words organized in over 115000 synsets for a total of 207000 wordsense pairs, 11488 verbs, 22141 adjectives, 4601 adverbs. WordNet [7] distinguishes between nouns, verbs, adjectives and adverbs because they follow different grammatical rules. Every synset contains a group of synonymous words or collocations (a collocation is a sequence of words that go together to form a specific meaning, such as “car pool”); different senses of a word are in different synsets. The meaning of the synsets is further clarified with short defining glosses (Definitions and/or example sentences). For example, the noun vacation has two senses. The first sense of the word vacation is given by a synonym holiday and the definition: leisure time away from work devoted to rest or pleasure. The second sense of the word vacation is given by the definition: the act of making something legally void. Most synsets are connected to other synsets via a number of semantic relations.

These relations are based on the type of word, and include:

- Nouns
  – Hypernyms: Y is a hypernym of X if every X is a (kind of) Y – hyponyms: Y is a hyponym of X if every Y is a (kind of) X
  – Coordinate terms: Y is a coordinate term of X if X and Y share a hypernym
  – Holonym: Y is a holonym of X if X is a part of Y
  – Meronym: Y is a meronym of X if Y is a part of X
- Verbs
  – Hypernym: the verb Y is a hypernym of the verb X if the activity X is a (kind of) Y (travel to movement)
  – Troponym: the verb Y is a troponym of the verb X if the activity Y is doing X in some manner (lisp to talk)
  – Entailment: the verb Y is entailed by X if by doing X you must be doing Y (sleeping by snoring)
  – Coordinate terms: those verbs sharing a common hypernym
- Adjectives
  – Related nouns
  – Participle of verb
- Adverbs
  – root adjectives While semantic relations apply to all members of a synset because they share a meaning but are all mutually synonyms, words can also be connected to other words through lexical relations, including synonyms, antonyms (opposites of each other) and derivationally related, as well. WordNet [7] also provides the polysemy count of a word: the number of synsets that contain the word. If a word participates in several synsets (i.e. has several senses), then typically some senses are much more common than others.

### 3.1.2 Knowledge resources

As it has already been explained in 2.3, most of Flickr photos are described by tags that denote:

- Places: continents, countries, cities, natural environment, objects on which (or in which) people can stand (buildings, furniture ...)
- Time: years, seasons, days...
- Events: Christmas, birthday...
- Names: Emma, Jean, nicknames...
- Cameras : Nokia, Canon, colors...

In order to be able to understand the meaning of these tags and correctly build a RDF annotation, some semantic resources are needed. For each tag category described above, the resources have been either created or crawled from the web and sometimes completed.

- Places: two place resources are used, a database containing geographical locations (for instance Los Angeles is in California which is in the US which is in America) and an ontology of things where people can be (for instance people can be at a table, which can be inside a house, which can be inside a city; or people can be in a car, that can be on a road, that can be in a state, ...).

  – For the first one, we crawled several websites (like for instance Yahoo! Meteo) to obtain lists of cities, with the countries and continents in which they are located.
  – For the second one, we had to complete Wordnet in order to be able to infer which kind of things could be a location for people. This consisted in adding about 200 location relations (meaning "can be located in").
- Time: there are not so many concepts for denoting time; we completed Wordnet and obtained an ontology of about 50 concepts (containing seasons, days, months, moments of the day ...)
- Events: as for time, we completed Wordnet and obtained an ontology of about 50 concepts denoting events (birthday, wedding, vacation, holy days ...)
- Cameras: we gathered a set of makes and models by crawling online shopping websites (for new and used products).

The method presented in this paper tries to convert a set of tags into a RDF annotation. The RDF language is thus presented in the following section.

### 3.1.3 RDF

Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata model but which has come to be used as a general method of modeling knowledge, through a variety of syntax formats. The RDF metadata model is based upon the idea of making statements about resources in the form of subject-predicate-object expressions, called triples in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion "The sky has the color blue" in RDF is as a triple of specially formatted strings: a subject denoting "the sky", a predicate denoting "has the color", and an object denoting "blue".

Below, a RDF ressource description introducing a "Person" whose name is "Emma" is presented:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://www.emma.htm">
<dc:title>Emma</dc:title>
<dc:publisher>PersonalPage</dc:publisher>
<foaf:primaryTopic>
<foaf:Person>
<foaf:name>Emma</foaf:name>
</foaf:Person>
</foaf:primaryTopic>
```

*</rdf:Description>*
*</rdf:RDF>*

## 3.2 CONVERSION PROCESS DESCRIPTION

This section describes the method enabling to generate a RDF description from a set of Flickr tags. Figure 7 shows the main components used in the process. It takes in input all the tags describing a photo and returns in output a RDF description of the photo. The semantic relations in, at, by, event, shot − by, describes, agent and object are introduced to form the resulting RDF annotation. The different steps are then the following (the photo is denoted by r):

- A stemmer enables to transform a tag into its non inflectional form,
- Using the semantic resources, each tag is then categorized in one of the six categories (location, time, event, people, camera, activity),
- All tags grouped in the location category are ordered from the smallest to the largest, say $l_1 \leq l_2 \leq . . . \leq l_n$). The generated triples are: $(r, in, l_1), (l_1, in, l_2). . . (l_{n-1}, in, l_n)$.
- Similarly all tags grouped in the tag category are ordered from the smallest to the largest, say $t_1 \leq t_2 \leq ... \leq t_n$). The generated triples are: $(r, at, t_1), (t_1, at, t_2). . . (t_{n-1}, at, t_n)$.
- For each event e a triple (r, event, e) is created,
- For each camera c a triple (r, shot − by, c) is created,
- For each verb *v* in the activity category, the corresponding signature is retrieved from Wordnet, say x → y.

For each tag ***a*** of type ***x*** and each tag ***y*** of type ***y***, the triple *(r, describes, v)*, *(v, agent, x)* and *(v, object, y)* are added.

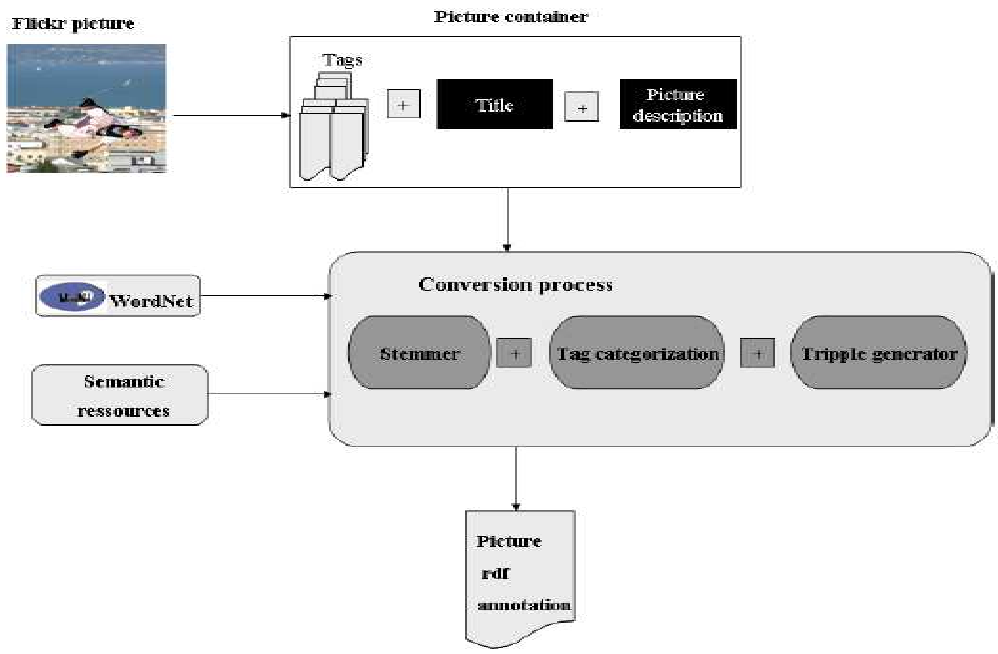


Fig 7. Conversion process components

For instance, tags describing the photo represented in figure 8 will lead to the annotation: *(r, in, NewHartford), (NewHartford, in, Connecticut), (r, describes, ski), (ski, agent, JeffG)*.

Fig 8. Example 1
*Tags{Connecticut, skiing, New Hartford, Jeff G}*

The tags describing the photo represented in figure 7 will lead to the annotation: *(r, in, Austria), (r, event, honeymoon), (r, at, January), (January, at, 2007)*.

Fig 9. Example 2
*Tags{Soll, Austria, Skiing, Honeymoon, January, 2007}*

The tags describing the photo represented in figure 10 will lead to the annotation: *(r, in, queensland), (queensland, in, Australia), (r, describes, fly), (fly, agent, birdie)*.

Fig 10. Example 3
*Tags{pelican, birdie, queensland, flying, Australia, iansand, waterfowl, peopleplacesevents}*

The tags describing the photo represented in figure 11 will lead to the annotation: *(r, in, Liege), (Liege, in, Belgium), (r, describes, drive), (drive, agent, Gaelle)*.

Fig 11. Example 4
*Tags{motion, Belgium, liege, drive, gaelle, starlet}*

The tags not recognized by the different semantic resources (for instance *iansand, peopleplacesevents…* ) are ignored during the conversion process.

## 4. RELATED WORK

A close research problem to ours is that concerning semiautomatic generation of annotations. [6] explain how, based on KA community initiative (Knowledge Annotation initiative of the Knowledge Acquisition community), an ergonomic and knowledge base-supported annotation tool was developed, and how this tool was extended with mechanisms that semi-automatically propose new annotations to the user. Supporting the evolving nature of semantic content, authors describe their idea of evolving Ontologies supporting semantic annotation; they conclude that semantic annotation and ontology engineering must be considered as a cyclic process. Although this work is important, some issues remain unsolved in this paper. Authors mentioned that an integrated system of annotation and ontology construction combining knowledge base-supported, ergonomic annotation, with an environment and methods for ontology engineering and learning from text supporting evolving Ontologies should be build. Furthermore, ergonomic, ontology and semiautomatic suggestion of the system should be evaluated. In addition annotated facts are not reusable since the system didn't support the RDF standard for representing metadata on the web.

Another work in the same domain it the one done by [8]. In this work a framework, S-CREAM, was developed to that allows for creation of metadata and is trainable for a specific domain. It supports the semi-automatic annotation of web pages based on the information extraction component Amilcare. It extracts knowledge structure from web pages through the use of knowledge extraction rules. These rules are the result of a learning-cycle based on already annotated pages. Authors are further investigating how different tools may be brought together, e.g. to allow for the creation of relational metadata in PDF, SVG, or SMIL.

Not very far from this [5] treat the generation of Ontologies. [5] present a comprehensive architecture and generic method for semi-automatic ontology acquisition from given intranet resources. A new approach for supporting the overall process of engineering Ontologies from text is described. Based on a given core ontology extended with domain specific concepts, the resulting ontology is restricted to a specific application using a corpus-based mechanism for ontology pruning. On top of the ontology two approaches supporting the difficult task of determining non-taxonomic conceptual relationships are applied. To complete this work several techniques for evaluating the acquired ontology should be developed. Also it should be elaborated how the results of different learning algorithms will have to be assessed and combined in the multi-strategy learning set newly introduced by the authors.

Some works were done on the Conversion of WordNet to a standard RDF/OWL representation. [9] presents an overview of the work in progress at the W3C to produce a standard conversion of WordNet to the RDF/OWL representation language in use in the Semantic Web community. The paper explains the steps involved in the conversion and details design decisions such as the composition of the class hierarchy and properties, the addition of suitable OWL semantics and the chosen format of the URIs. Some issues remain open like supporting different versions of WordNet in RDF/OWL and defining the relationship between them. Furthermore, the integration of WordNet with sources in other languages is not solved. Most of existing works provide a semi-automatic generation of annotations.

[4] is a tool that converts automatically Flickr tags to RDF. However it does not provide a really semantic description of photos but it rather syntactically translates each tag in a separate RDF triple.

## 5. CONCLUSION AND FUTURE WORK

This paper has presented a conversion process from Flickr's photo tags to RDF annotations, thus leading to a first bridge between Web 2.0 and Semantic Web. Before conceiving this method, the ways people used to tag photos on Flickr were analyzed. It has shown that people mainly employed six categories of tags, each one denoting a certain aspect of the photo: location, time, event, people, camera, and activity. For each one of these categories, semantic resources have been either reused and completed (like Wordnet) or crawled from the web (like camera and location databases). Using these semantic resources, the method presented in this paper tries to identify the category of each tag. It then uses the signatures of verbs (tags of category activity) in Wordnet to associate a verb with its subject and complement and thus to build a RDF triple. Other triples are built by using tags of other categories, for instance by linking the photo with the smallest location as well as a location with a more general location. This method gives its best results for photos containing in their tags verbs, as these tags will provide the RDF relations that are the less common and thus the most interesting. Future work will try to take advantage of the presence of other information (the title and the legend of the photo) to improve the understanding of what the photo is about and to generate a RDF description that is more accurate.

## 6. REFERENCES


[1] http://www.wikipedia.org/.

[2] http://del.icio.us/.

[3] http://www.flickr.com/.

[4] www.kanzaki.com/works/2005/imgdsc/flickr2rdf.

[5] A method for semi-automatic ontology acquisition from a corporate intranet. In Proceedings of EKAW-2000 Workshop "Ontologies and Text", Juan-Les-Pins, France, October 2000, number 1937 in Springer Lecture Notes in Artificial Intelligence (LNAI), 2000.

[6] M. Erdmann, A. Maedche, H. Schnurr, and S. Staab, 2000, From manual to semi-automatic semantic annotation: About ontology-based text annotation tools.

[7] C. Fellbaum, 1998, Wordnet: an electronic lexical database, *Language, Speech and Communication, Cambridge*.

[8] S. Handschuh, S. Staab, and F. Ciravegna, 2002, S-cream — semi-automatic creation of metadata. *In 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*.

[9] M. van Assem, A. Gangemi, and G. Schreiber, 2006, Conversion of WordNet to a standard RDF/OWL representation. *In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy*.