KNOWLEDGE-BASED IMAGE ANNOTATING

Elena Sokolova and Mikhail Boldasov
Russian State University for the Humanities, Miusskaya sq, 6, Moscow, GSP-3, Russia

Keywords: Natural Language Generation, Ontology, Subject Mater Description, Image Description.

Abstract: We present an experiment of image annotating for photographs of one collection on base of ontology. A fragment of ontology, which consists of concepts of visual objects, their features and relations, is constructed using SemTalk2 software tool. Using this ontology there are prepared semantic annotations for photographs from the collection. Semantic annotations are ready for their further automatic processing. We consider them as input for automatic generation of photograph descriptions in a given NL. In this paper we discuss types of visual concepts, structure of Ontology and Image Models and their possible applications.

1 INTRODUCTION

Last decennium finding an image suitable for a particular purpose received a considerable attention. Solution of the task of image retrieval is mapping of textual indexes and annotations to target images. Indexes and annotations are manually prepared, or found in the Web, or summarized from other texts in the Web. This approach is valid for retrieval of named entities especially persons and toponyms (Gornostay, 2009). It can explore statistical salience of objects in the image and in the text and then match mentioned and depicted objects (Deschacht, Moens, 2007). This approach is not valid for art collections or cinema records where no named entities are presented, for example: “Mid shot a man walking between two lanes”. In this case user can be interested in types of depicted objects, their composition and spatial relations. Objects theyself should be described in sources of knowledge – ontologies, and annotating can be characterized as “ontology-based” (Schreiber et al., 2001).

The system that implements ontology-based approach to Natural Language Generation (NLG) of image descriptions can be logically divided into two parts: Recognizer that prepares so called Image Models (IMs) and Generator that prepares image descriptions in a given NL. IMs are considered as primary semantic interpreted results of image recognition. IMs and generated picture descriptions can be used further for sophisticated image retrieval.

According to our understanding, the current state of art in Artificial Intelligence does not allow to construct Recognizer module, which can prepare IMs valid for the further NLG of textual annotations. So we need to model this process manually and prepare IMs by hand. In the following text we use abbreviation “IM(s)” for manually prepared Image Model(s).

In this paper we discuss what are IMs and what concepts can describe “visual” world. One method sketched in (Hollink et al., 2003) is to include in the formalized model of a picture content existing knowledge bases and lexico-semantic databases. Unlike this, our research is more practical oriented to NLG. That means that prepared IMs should be available as an input for our NLG system.. Our investigation shows that existing knowledge bases and ontologies are not sufficient as a basis for IMs.

We tried two methods of text generation: Upper-Model-based (Kruiff et al., 2000) and knowledge-and-transformation-based (Boldasov, Sokolova, 2002), (Boldasov, Sokolova, 2003). The former was in AGILE project supported by EC aimed to multilingual generation of software manuals in three Slavic languages. The latter has begun in InBase project of Russian Academy of Science and currently it is supported by our interest to the task. Working for InBase project we have implemented NLG environment DEMLinG, which concepts were further approved by three toy-Generators, prepared for different domains: query to DB (Boldasov, Sokolova, 2002), annotation of content of DB, and picture description. Our main interest in “ontology-based” image annotating is to investigate the requirements for input for NLG as well as the main

Sokolova E. and Boldasov M. (2009).
KNOWLEDGE-BASED IMAGE ANNOTATING.
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pages 376-379

principles of Text Planning – one of the tasks of NLG.

In section 2 we describe images and corpus of their descriptions, in section 3 the means of subject matter description are discussed, in section 4 we present concepts of "visual" ontology, in section 5 we describe our experiment with SemTalk2 and in Conclusion we resume the experiment results and sketch areas where IMs and ontology could be used.

2 IMAGE MODEL AND CORPUS

Our experimental materials were colored photographs of Prokudin-Gorsky dated from the period of 1900-1917 (www.prokudin-gorsky.ru) and associated textual descriptions, written by students of the Russian State University for Humanities describing just what is depicted in the photographs. Every description consists of approximately from 2 to 15 lines. We consider about 100 photographs and our corpus contains about 250 descriptions, each photograph having two or more descriptions made by different students. In this paper we explore landscapes.

We consider IM as a set of Objects with particular characteristics, and a number of Relations between them. Following (Hollink et al., 2003) IMs are based on predicate information - triads (agent-process-object) and the settings (time, location and artist) (Tam et al., 2001). Using this method for our task, we would have two problems:

1. differently to objects, actions have no area on the surface of a photograph, they are indirect knowledge that ought to be not recognized but inferred from location of objects, postures and gestures of humans and animals;
2. linguistic ontologies and art collection DBs are not developed to present visual information.

Therefore we need a special ontology to present ontological and visual features of objects related by spatial and ontological relations. As a source of this information we use our corpus of photographs. We consider NL descriptions as some result of analysis of visual information in the picture. The descriptions are used to explore the following things:

- what objects and relations were noticed by the author, hence what concepts ought to be presented in our ontology and what objects and relations will form IMs;
- how this information is expressed in the text.

Here is description of the photograph 00957 translated from Russian:

In a country land on a hill between *secular* **spruces and pine trees a little** *snug* **chapel is situated. Shine in the sun its silver cupola, doors are broken open and on the stairs someone attendant nestles** *came from the proximate village.* **The chapel** *is build on the occasion of a particular case:* **its porch exposes to the old pine tree** *in which an image of the Madonna icon turned up to the residents of the proximate village. And* **now the man sitting on stairs** *peers to the pine brunches hoping to see the miraculous icon again.*

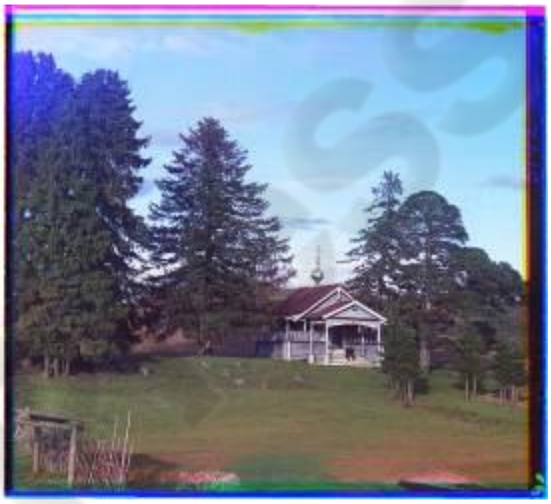

Figure 1: Photo 00957 – Materik i Chapel of the Mother of God and the pine tree on which the icon appeared.

In this description bold are words describing just visual objects and their features and relations which ought to be included into IM. The rest of the text concerns hypothesis and impressions of the author.

3 SUBJECT MATTER DESCRIPTION

Description template that was proposed in (Tam et al., 2001) and used in (Schreiber et al., 2001), (Hollink et al., 2003) consisting of triads and settings is not possible for our corpus since our images are static, and they have usually no action at all. The only type of physical processes that are mentioned in the names of photographs are like this "Rafts sitting on the rocks at the village of Kurya".

We need spatial relations and composition relations of visual nature which are direct interpretation of what we see.

Composition relations are presented by one relation – INCLUDES. The surface of photograph is divided into areas that are formed by boundaries of depicted objects. Area of one object can enclose area of another object, e.g., (SKY ((FLYING-BIRD (WINGS)) SUN)). Objects that are inside the area of another object are "included" in it. For relations of partly included objects, e.g., OBJECTS standing on GROUND, we use our knowledge about what can "stand on" and what can be "localization".

Describing a picture in NL we often divide it into "layers" – groups of objects that are in equal distance from the viewer, e.g., "in the foreground we see a group of people, on the back – a street". The concept LAYER presents these groups of objects. Practically people use from 0 to 2 ("foreground" and "background") layers but in some cases picture description can contain more layers.

Objects whose areas are not intercrossed or that are intercrossed not as "an object standing on the GROUND" are related by **spatial relations**, which are usually bidirectional, e.g.:

- To-the-left(X, Y) / To-the-right(Y, X)
- Near(X, Y) / Near(Y, X)
- Around(Y, X) / In-the-centre(Y, X)

We consider IM as a kind of visual specification with elements of semantic interpretation. Both can be of different level of discriminating – general vs. more detailed description.

4 ONTOLOGY

We define **classes of objects** designed by English words, e.g., HOUSE, TREE, SMOKE-STACK, FIELD. In IMs we use instances of classes, their **visual parameters** and **meanings**, e.g,, COLOR: BLACK, BLUE, etc.; SIZE: SMALLinWIDTH, BIG. etc.; SHAPE: SQUARE, ROND, etc.

We also need hypernyms for the classes. They can be used in the situations of visual haziness or for the second nomination of the same entity in the text, e.g., for classes EDIFICE, CHAPEL, BARN and CABIN we need a **superclass** BUILDING which is related with them by "**is-a"** relation. Hypernyms can be extracted from existing ontologies, e.g., WordNet. But concept descriptions in WordNet are not valid to our purposes since they are mostly functional, e.g.:

- DOOR is-a "movable barrier (a barrier that can be moved to allow passage).

We can also use ontological relations "part holonym – part-meronym". For classes WALL, ROOF, WINDOW and DOOR part holonym class is BUILDING, EDIFICE. For BUILDING among its part meronyms are WALL, ROOF, PORCH. So we need **Classes** and **Superclasses** that correspond to the "is-a" relation, one **Composition relation** "part" that has two terminals - part-holonym and part-meronym, and a number of **Spatial relations**.

The resulted ontology should be reasonable easy to use it for composing IMs. So, it is not a good idea to make it possible e.g. to choose a visual parameter of an object from the whole set of visual parameters of any class, or to choose possible relation from the whole set of relations between any of classes in the ontology. Thus we need to invent a kind of filter that controls that the proper object is supported with a proper set of visual parameters and relations.

This filter presents description of subject matter as information prepared for communication, where every object is presented in some **cognitive perspective (CP). CPs are** containers of visual parameters and participants of relations. Class consists of one or several CPs, e.g. class RIVER consists of SURFACE and MIRROR CPs. Instance of a Class can be assigned in IM manually with an extra CP if it performs not typical role in the picture, e.g. if a SCARF is used as a SKIRT we need to combine in IM these two CPs both.

The ontology can be used in two paradigms: image recognition and NLG of picture descriptions. Here we pay attention only on NLG paradigm. For the recognition process we need some reasoning.

5 EXPERIMENT

Ontology aught to be implemented in OWL (Web Ontology Language) because it is based on paradigm of XML, which is convenient for processing and specially prepared to describe ontologies.

The standard solution to manage OWL descriptions by Altova SemanticWorks failed, because the system is not optimized enough to our task specifics. Another tool SemTalk2 has a user-friendly editor for Semantic Web ontologies. SemTalk2 is based on Visio Diagrams that are used to introduce new objects to the ontology or IM. It supports two types of diagrams: Class Diagrams – for description of ontology and Instance Diagrams – for description of IMs. Maintenance of ontology is supported not only by Diagrams but also by hierarchic View and importing external models is allowed.

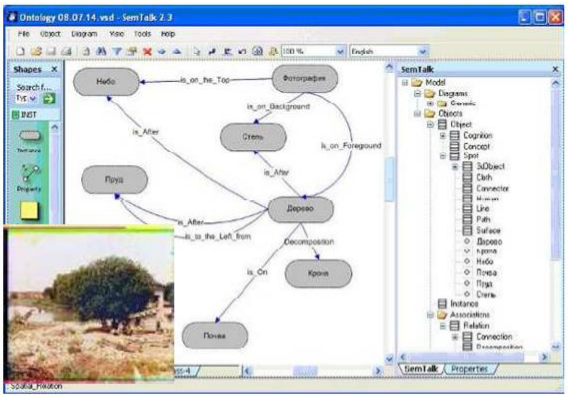

Figure 2: Screenshot of the opened IM in SemTalk2 application, supported with the photo 00039 that is described on the opened IM.

Screenshot in Figure 2 displays an opened IM in SemTalk2 which is prepared based on two student's descriptions of photograph 00039. Ontology that was used for this IM is placed in the right part of the Screenshot. It is possible to drag-and-drop the desired classes from the ontology to the IM.

SemTalk2 satisfies to our task much better than Altova SemanticWorks. But using the free version that is available in Web has the following lacks:

- the application is not enough bug-fixed. When we used it not by scenario that was described in user guide, we got data inconsistency;
- our idea of CPs is not supported;
- it would be better if an individual would be displayed by name of its Class or CP.

6 CONCLUSIONS

We sketched a method of semi-automatic IM creation and discussed issues of interaction between visual data and human knowledge, as well as appropriate theoretical aspects that should be used as the base for ontology and IM. Prepared ontology and IMs can be used:

- in Recognizer providing features and possible relations for objects depicted in the picture since knowledge plays decisive role in the process of the image recognition;
- IMs can be used in the multilingual image retrieval, e.g., in art collections and photograph collections instead of key words;
- in NLG of photo descriptions as input IMs.

Our experiment showed that SemTalk2 is valuable for IM construction in general, but the free version has some lacks for this task.

REFERENCES

Gornostay T., Aker A. Development and implementation of multilingual object type toponym-referenced text corpora for optimizing automatic image description genereation // Proceedings of the Conference on computational linguistics and intellectual technologies DIALOG'2009 - 2009, Bekasovo, Russia May 27-31, 2009. Pages: 580 – 587.

Deschacht K., Moens M-F. Text analysis for automatic image annotation // The 45th Annual meeting of the Association for Computational Linguistics, Prague, June 2007. http://acl.ldc.upenn.edu/P/P07/P07-1126.pdf

Schreiber A.Th., Dubbeldam B., Wielemaker J., Wielinga B.J. Ontology-based photo annotation //IEEE Intelligent systems, 16(3):66-74, May-June 2001. http://www.cs.vu.nl/~guus/papers/Schreiber01a.pdf

Hollink L., Shreiber G., Wielemaker J., Wielinga B. Semantic annotation of image collections // S. Handschuh, M. Koivunen, R. Dieng, and S. Staab, editors, Knowledge Capture 2003. Proceedings Knowledge Markup and Semantic Annotation Workshop, Florida, USA, October 2003. P. 41-48. http://www.cs.vu.nl/~guus/papers/Hollink03b.pdf

Kruijff G-J., Bateman J., Teich E., Sharof S., Sokolova L., Kruijff-Korbayova I., Skoumalova H., Staykova K., Hana J., Hartley T. Multilinguality in a text generation system for three Slavic languages // Proceedings of the 18th conference on Computational linguistics - Volume 1, 2000, Saarbrьcken, Germany July 31 - August 04, 2000. Pages: 474 – 480.

Boldasov M.V., Sokolova E.G. User query understanding by the InBASE system as a source for Multilingual NL generation module // Text, Speech and Dialogue (P. Sojka, I. Kopeček and K. Pala eds.). – Proceedings of the 5th International conference, TSD 2002, Brno, Czech Republic, September 2002. Springer-Verlag Berlin Heidelberg, Germany, 2002. Pages: 33-40.

Boldasov M.V., Sokolova E.G. QGen – generation module for the register restricted InBASE system // Computational linguistics and intelligent text processing (A.Gelbukh ed.). – Proceedings of the 4th International conference, CICLing 2003, Mexico City, Mexico, September 2002. Springer Berlin Heidelberg, Germany, 2003. Pages: 465-476.

Tam, A.M. Leung, C.H.C. Structured Natural-Language description for semantic content retrieval // Journal of the American Society for Information Science and Technology Vol. 52(11), September 2001. Pages: 930 – 937.

Hunter J. Adding multimedia to the semantic Web – building an MPEG-7 ontology // Proceedings of International Semantic Web Working Symposium (SWWS), Stanford, July 30 - August 1, 2001 http://archive.dstc.edu.au/RDU/staff/jane-hunter/semweb/paper.html .