Placing Media Items Using the Xtrieval Framework

Fabian Krippner, Gerald Meier, Jens Hartmann, Robert Knauf
Chemnitz University of Technology
Straße der Nationen 62
09111 Chemnitz, Germany
{fabk, gmei, hajen, knauf}@hrz.tu-chemnitz.de

ABSTRACT
In this paper we describe our approaches and results of evaluating the metadata in tagged user-generated videos as well as their visual features in order to extrapolate geographical relevance. The evaluation was done in the context of the MediaEval 2011 Placing Task in which we had to determine and to assign the best fitting geographical coordinates to each video. Our main goal was to realize this task with a retrieval framework developed by the Chemnitz University using the bag-of-words model to compare parts of metadata. This framework is used for indexing and comparing purposes. Particularly, it incorporates multiple lists of stop words, stemming lists and dictionaries. For enhancement purposes, we also used the GeoNames gazetteer despite noticing that the overall results seem to be slightly better using sole metadata comparisons.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]

Keywords
Geographical Coordinates, Geotagging, Flickr videos, Gazetteer, Bag-of-Words model, Geographical Location

1. INTRODUCTION
The internet is filled with many hosting sites for different kinds of media such as videos, pictures, and music. Tagged with heterogeneous kinds of information, it is sometimes difficult or impossible at all to clearly assign the resource's origin to one specific location on the globe. Regardless of whether the given tags of the videos are sufficient, it is a challenging task to assume a fitting location using a data set of ground truth. For that purpose, the main goal of the *MediaEval 2011 Placing Task* was to address the topic of automatically geotagging videos taken from the Flickr community.

Starting from a set of training videos, geographical coordinates were to be derived as accurately as possible by using three different variations of ground truth data.

The criterion of the gained results' relevance were the distances between the actual coordinates of a video clip's location provided by the particular Flickr user and the ones that have been estimated by the Placing Task participants' algorithms. The exact requirements and conditions are explained in the Placing Task overview paper [1].

MediaEval 2011 Workshop, September 1-2, Pisa, Italy.

2. RELATED WORK
In the field of adding geographical information, "geotags", or GPS coordinates there are many works already addressing the annotation of images and videos as well as the extraction of comparable image features. For videos Kelm et al. give an explanation of three ways on how to place a video on the globe using the metadata and external resources [2]. Serdyukov et al. make use of the textual annotations associated with uploaded images in combination with GeoNames[1] in order to retrieve geographical relevant information through a language based model [3]. Hays and Efros suggest a large image database divided into locations for the purpose of scene matching by comparing images instead of text using the nearest-neighbor method [4].

But most inspiring was the way of comparing metadata and scanning them for geographical relevant information that Perea-Ortega et al. used in their MediaEval 2010 work [5] as well as the metadata- and keyword-focused approach of Choi et al. [6]. This led us to try and realize their approaches using *Xtrieval* framework [7] which has been developed at the Chair Media Informatics of the Chemnitz University. It is based on Lucene[2] and improves it by using the bag-of-words model instead of straight keyword lookups.

Particularly, we benefitted from its incorporated whitespace tokenizer and the modified stop word list. Lucene DataDocuments provided an adequate object type to store all development data sets.

3. EVALUATION OF RESULTS
Making use of the built-in methods of Xtrieval, we started with creating data collections from the development data and the data to be tested. We continued by indexing the development metadata collection and using the test data collection to search on it.

For both data collections we extracted the following metadata fields from the respective XML files of the videos: description, keywords, title, locality, region, country, and user ID. Additionally, we extracted the data fields containing latitude and longitude information from the development data collection. Furthermore, we created a collection based on the Flickr images' metadata.

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

[2] http://lucene.apache.org/java/docs/index.html

Processing both metadata collections (video and image descriptions), we merged the particular data fields into a single field. Here, we left out the fields “latitude”, “longitude”, “userid”, and “docno” and handled them separately later on.

The extracted fields which coincided were merged to one field called “bag”. Thereby, except for the field “userid”, we applied the bag-of-words model which is used for direct comparisons using a search method based on the Lucene searcher. This method works closely with the Lucene Index. It reevaluates the given query by searching for frequently used terms in relevant documents and appending them to the initial search query with the intention of enhancing the subsequent Lucene search.

We derived the score for each hit between development data and test data from the score formula of the Similarity[3] class in Lucene. From each hit set the procedure returned we took the hit with the highest score as the best match between both data sets.

In former tryouts where we used different contents in the “bag” we found our optimum of using all described fields in it. Leaving out even one field led to an overwhelming amount of videos being missed in the results. For example leaving out the field “Keywords” from the XML data led to 1,407 (26.31%) missing hits when querying the test data.

We noticed 278 videos which did not receive a hit. By initiating another search process in which we used the missing videos as the test data collection, we were able to reduce this number by 37 videos through repeating the former search process and adding an additional query for matching user IDs. Subsequently, we added the latter results to the results of the first search.

For the second run which permitted the usage of a gazetteer, we created a new index using the development data and the GeoNames database. By creating a collection over the countries and features we gained a new basis for our search.

Our first search resulted in 23 more hits than without the gazetteer.

Table 1. Results determined by the distances between predicted and actual geographical coordinates

Run	1km	10km	100km	1000km	10000km
Pure Development	9.37%	21.78%	30.67%	44.92%	86.37%
Gazetteer	9.86%	21.49%	29.79%	43.26%	84.16%

Compared to the final results provided by the MediaEval team, the number of determined coordinates in a very close proximity of 1km and less was higher when using the gazetteer than without. For the moment, this answered our initial expectation.

Surprisingly, regarding ranges of more than 1km difference to the true coordinates, the results varied slightly to an extent up to 3.8%, but this time in favor to the search without the additional geographic database.

[3] http://lucene.apache.org/java/2_4_0/api/org/apache/lucene/search/Similarity.html

87.86% of all results were equal in both the searches with and without Gazetteer, including videos exceeding the 10000km threshold. An amount of 63.43% of all videos shared the same coordinates while the rest was divided in two groups: One group (11.11% of the cases) delivered better results using the Gazetteer while the other (13.31%) performed better based only on the development data. This leads us to the conclusion that the use of a Gazetteer could improve the search results by nearly the same amount as the original data. Thereby, further refinement in the selection process is needed for more accurate results.

4. OUTLOOK

The applied bag-of-words model did assign a correct location to ca. 10% of the tested videos. So, for further development we will include different sub-bag correlations in contrast to now, where we found our optimum in a complete “bag”. We will try different “bags” for different data, stacking searches after one another, and figuring out a better system of weighing between the distinct ground truth resources. Thereby, we expect to minimize the deviation of results and to gain benefits of each particular ground truth data set. Furthermore, the application of filter adjustments determining the origin by language and using speech recognition will be a next step as well as using the image feature set, particularly, the Color and Directivity Descriptor, which we didn't get to be fully realized by the time of submission.

5. REFERENCES

[1] A. Rae, V. Murdock, P. Serdyukov, and P. Kelm. Working Notes for the Placing Task at MediaEval. In *Working Notes for the MediaEval 2011 Workshop*, Pisa, Italy, 2011.

[2] P. Kelm, S. Schmiedeke, and T. Sikora. Multi-modal, multi-resource methods for placing flickr videos on the map. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ICMR '11, pages 52:1-52:8, New York, NY, USA, 2011. ACM.

[3] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 484-491, New York, NY, USA, 2009. ACM.

[4] J. Hays and A. Efros. Im2gps: estimating geographic information from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2008, pages 1-8, Anchorage, Alaska, USA, 2008.

[5] J.M. Perea-Ortega, M.A. Garcia-Cumbreras, L.A. Urena-Lopez, and M. Garcia-Vega. SINAI at Placing Task of MediaEval 2010. In *Working Notes Proceedings of the MediaEval 2010 Workshop*, Pisa, Italy, 2010.

[6] J. Choi, A. Janin, and G. Friedland. The 2010 ICSI Video Location Estimation System. In *Working Notes Proceedings of the MediaEval 2010 Workshop*, Pisa, Italy, 2010.

[7] J. Kürsten, T. Wilhelm, and M. Eibl. Extensible Retrieval and Evaluation Framework: Xtrieval. In *Proceedings of the Lernen - Wissen - Adaption Workshop*, LWA 2008, pages 107-110, Würzburg, Germany, 2008.