

CLEF-IP Image Tasks Submission Format

May 30, 2011

1 Prior Art Image Task

1.1 Submission format

The format of the CLEF-IP submissions is similar to the TREC format. A submission consists of a single ASCII text file containing at most 1000 lines per topic, in the standard format used for most trec submissions: five columns separated by whitespace.

Below is an example of part of a submission file:

```
EP-8888888-A2      Q0      EP-0000001      1      1012
EP-8888888-A2      Q0      EP-0000002      2      1011
EP-8888888-A2      Q0      EP-0000002      3      999
...
```

In the trec style submission files:

- the first column is the topic number (as specified in the topic file)
- the second column is the query number within that topic. This is to be set to Q0;
- the third column is the document number of the retrieved document, excluding the kind code; this is the value of the country and doc-number attribute in the patent-document xml tag, concatenated with a dash;
- the fourth column is the rank of the document retrieved;
- the fifth column shows the score (integer or floating point) that generated the ranking. This score must be in decreasing order.

A run is uniquely identified by the participant id, the method used and the task type as follows:

```
participantID_method and/or runID_IMG_PAC_{IMG/TXTIMG}.runP
```

IMG or TXTIMG at the end indicates whether image data alone or image and text data are used in the run.

An example of a valid filename is:

```
TUW_tfidf_run01_IMG_PAC_TXTIMG.runp
```

or

```
TUW_sift_IMG.runp
```

The maximum number of runs per task for a participant is 8. Submissions must be uploaded to an ftp server, access credentials will be communicated at a later time. Together with the run files, and for each of them, we require the participants to upload a concise description of the retrieval methods and document fields used to obtain the respective runs. The naming of the description files should follow the naming of the run files, and should be identical to the run file to which they correspond, but with the extension `.txt`.

1.2 Evaluation measures

We will mainly use the the `trec_eval` implementations of the precision, recall, F1, MAP and nDcg measures. The relevance judgements to be used during the evaluation will be extracted automatically from the patent documents used to create the topics.

2 Image Classification Task

2.1 Submission format

When the 1,000 test images are made available, participants should submit files with classification results (also referred to as runs) in the following format:

```
<image name 1> <score for class 1> <score for class 2> ... <score for class 9>
<image name 2> <score for class 1> <score for class 2> ... <score for class 9>
.
.
.
<image name 1000> <score for class 1> <score for class 2> ... <score for class 9>
```

Each score can be an arbitrary floating point number, where higher numbers denote higher confidence regarding the classification into that image type. Each image is only classified into a single class, so the highest score indicates the class into which the image is classified. However the scores can be used to indicate the confidence of the classification. Each file should contain exactly 1,000 lines.

The lines should be sorted in the ascending numerical order of the filenames, i.e. 0001.tif, 0002.tif, 0003.tif up to 1000.tif. Include the full filename with the extension .tif. Below is an example of the first 2 lines of a potential submission.

```
0001.tif 0.1 0.9 0.05 0.2 0.0 0.0 0.05 0.03 0.0
0002.tif 0.0 0.01 0.85 0.4 0.21 0.0 0.05 0.1 0.1
```

The files should be named according to the following format:

```
<participant identifier>_method and/or runID_IMGCLS_<submission number>.dat
```

For example: `TUW_SVM_run003_IMGCLS_2.dat`

The maximum number of runs per task for a participant is 8. Instructions for submitting run will be communicated at a later time. Together with the run files, and for each of them, we require the participants to upload a concise description of the classification methods used to obtain the respective runs. The naming of the description files should follow the naming of the run files, and should be identical to the run file to which they correspond, but with the extension `.txt`.

2.2 Evaluation measures

Note that it is not permitted to use any additional data for training and setup of the systems. If you need test data for system tuning, you need to split the available training data into a training and validation set.

We will use equal error rate (EER) and Area Under Curve (AUC) of a ROC curve and True Positive Rate (TPR) per class averaged over all classes to evaluate the performance of the individual runs.

The OCTAVE script that will be used for the evaluation of the submissions is available for download (it should also run in MATLAB). An example submission file (`test_sub.txt`) (and an arbitrary ground truth file, `test_gt.txt`) is also included.