Cross-Language and Cross-Media Image Retrieval: An Empirical Study at ImageCLEF2007

Steven C. H. Hoi

School of Computer Engineering
Nanyang Technological University
Singapore 639798
{chhoi}@ntu.edu.sg

Abstract. This paper summarizes our empirical study of cross-language and cross-media image retrieval at the CLEF image retrieval track (ImageCLEF2007). In this year, we participated in the ImageCLEF photo retrieval task, in which the goal of the retrieval task is to search natural photos by some query with both textual and visual information. In this paper, we study the empirical evaluations of our solutions for the image retrieval tasks in three aspects. First of all, we study the application of language models and smoothing strategies for text-based image retrieval, particularly addressing the short text query issue. Secondly, we study the cross-media image retrieval problem using some simple combination strategy. Lastly, we study the cross-language image retrieval problem between English and Chinese. Finally, we summarize our empirical experiences and indicate some future directions.

1 Introduction

Digital image retrieval has attracted a surge of research interests in recent years due to the rapid growth of digital media contents overwhelming over the World Wide Web (WWW). Most existing search engines usually employ text based retrieval methods to search the digital images from WWW. They have yet to solve the retrieval tasks very effectively. Until now, general image retrieval is still a challenging research problem. There are several major obstacles for image retrieval. First of all, Web images may be associated with textual descriptions in different languages. When searching the images with different languages, the retrieval tasks will be suffered critically without further translation processing. Moreover, many Web image may not associate with textual descriptions, which makes the traditional text based retrieval difficult to reach some Web images without text annotations. Finally, even some images are associated with keywords, it can still be difficult for a short text query due to some challenges, such as word ambiguity. In this paper, we study some methodology to attack some of these challenges for a benchmark image retrieval evaluation campaign.

ImageCLEF is the cross-language image retrieval track which is run as part of the Cross Language Evaluation Forum (CLEF) campaign. The goal of this track

is to evaluate retrieval of images described by text captions based on queries in a different language; both text and image based retrieval techniques can be explored. The ImageCLEF provides an annual benchmark evaluation for image retrieval research from 2003 [1]. In this year, there are two types of retrieval tasks in the ImageCLEF, including general photographs retrieval and medical image retreival. In this paper, we discuss our participation in the photo retrieval tasks at ImageCLEF2007.

In this participation, we offer the major contributions in three aspects. First of all, we study an empirical evaluation of language models and smoothing strategies for cross-language image retrieval. Secondly, we conduct an evaluation of cross-media image retrieval, i.e., combining text and visual contents for image retrieval. The last contribution is the empirical evaluation of a methodology for bilingual image retrieval spanning English and Chinese sources.

The rest of this paper is organized as follows. Section 2 reviews some methodology of the TF-IDF retrieval model and the language model for information retrieval. Section 3 presents our implementation for this participation, and outlines our empirical study on the cross-language and cross-media image retrieval. Section 4 set out our conclusions.

2 Review of Language Models and Smoothing Techniques

In this section, we review several existing language models and smoothing techniques to be studied in our experiments. In our participation, we have performed an extensive set of experiments to evaluate the performance of several state-of-the-art language models in application to text-based image retrieval and also examine the influence of several popular smoothing techniques.

More specifically, two kinds of retrieval models are studied: (1) The TF-IDF (Term Frequency-Inverse Document Frequency) retrieval model, and (2) The KL-divergence language model based methods. Three smoothing strategies for Language Models evaluated in our experiments [2] include: (1) the Jelinek-Mercer (JM) method, (2) Bayesian smoothing with Dirichlet priors (DIR), and (3) Absolute discounting (ABS).

2.1 TF-IDF Information Retrieval Model

We study the TF-IDF retrieval model, which is a well-known information retrieval model for text-based retrieval tasks [3]. In general, one can assume that each document and each query can be represented as a term frequency vector $\mathbf{d} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{q} = (y_1, y_2, \ldots, y_n)$ respectively, where n is the number of total terms, x_i and y_i are the frequency (counts) of term t_i in the document vector $\mathbf{d}$ and query vector $\mathbf{q}$, respectively. In a retrieval task, given a document collection $\mathcal{C}$, the inverse-document-frequency (IDF) of a term t is defined by $log(N/n_t)$, where N is the total number of documents in $\mathcal{C}$, and n_t is the number of documents that contain the term t. For the TF-IDF representation, all terms in the query and documents vectors are weighted by the TF-IDF

weighting formula, i.e., $\mathbf{d}' = (tf_d(x_1)idf(t_1), tf_d(x_2)idf(t_2), \ldots, tf_d(x_n)idf(t_n))$ and $\mathbf{q}' = (tf_q(y_1)idf(t_1), tf_q(y_2)idf(t_2), \ldots, tf_q(y_n)idf(t_n))$. For a simple TF-IDF retrieval model, one simply takes $tf_d(x_i) = x_i$. One can also define some other heuristic formula for the TF function. For example, the Okapi retrieval approach is a special case of TF-IDF model by defining the document TF formula as [4]:

$$tf_d(x) = \frac{k_1 x}{x + k_1(1 - b + b\frac{l_d}{l_c})} \tag{1}$$

where k_1 and b are two parameters for the document TF function, l_d and l_c are the lengths of the given document and collection, respectively. Similarly, a query TF function can be defined with parameters k_1 and b as well as l_q representing the average length of queries. For similarity measure of the TF-IDF retrieval model, cosine similarity function is often adopted, which a measure of similarity between two vectors of n dimensions by finding the angle between them.

2.2 Language Modeling for Information Retrieval

Language model, or the statistical language model, employs a probabilistic mechanism to generate text. The earliest serious approach for a statistical language model may be tracked to Claude Shannon [5]. To apply his newly founded information theory to human language applications, Shannon evaluated how well simple n-gram models did at predicting or compressing natural text. In the past, there has been considerable attention paid to using the language modeling techniques for text document retrieval and natural language processing tasks [6].

The KL-divergence Measure. Given two probability mass functions $p(x)$ and $q(x)$, $D(p||q)$, the Kullback-Leibler (KL) divergence (or relative entropy) between p and q is defined as

$$D(p||q) = \sum_x p(x) log \frac{p(x)}{q(x)} \tag{2}$$

One can show that $D(p||q)$ is always non-negative and is zero if and only if $p = q$. Even though it is not a true distance between distributions (because it is not symmetric and does not satisfy the triangle inequality), it is often still useful to think of the KL-divergence as a "distance" between distributions [7].

The KL-divergence based Retrieval Model. In the language modeling approach, we assume a query q is generated by a generative model $p(q|\theta_Q)$, where θ_Q denotes the parameters of the query unigram language model. Similarly, we assume a document d is generated by a generative model $p(q|\theta_D)$, where θ_Q denotes the parameters of the document unigram language model. Let $\hat{\theta}_Q$ and $\hat{\theta}_D$ be the estimated query and document models, respectively. The relevance of

d with respect to q can be measured by the negative KL-divergence function [6]:

$$-D(\hat{\theta}_Q||\hat{\theta}_D) = \sum_w p(w|\hat{\theta}_Q) log p(w|\hat{\theta}_D) + (-\sum_w p(w|\hat{\theta}_Q) log p(w|\hat{\theta}_Q)) \quad (3)$$

In the above formula, the second term on the right-hand side of the formula is a query-dependent constant, i.e., the entropy of the query model $\hat{\theta}_Q$. It can be ignored for the ranking purpose. In general, we consider the smoothing scheme for the estimated document model as follows:

$$p(w|\hat{\theta}_D) = \begin{cases} p_s(w|d) & \text{if word } w \text{ is present} \\ \alpha_d p(w|\mathcal{C}) & \text{otherwise} \end{cases} \quad (4)$$

where $p_s(w|d)$ is the smoothed probability of a word present in the document, $p(w|\mathcal{C})$ is the collection language model, and α_d is a coefficient controlling the probability mass assigned to unseen words, so that all probabilities sum to one [6]. We discuss several smoothing techniques in detail below.

2.3 Three Smoothing Techniques

In the context of language modeling study, the term "smoothing" can be defined as the adjustment of the maximum likelihood estimator of a language model so that it will be more accurate [2]. As we know that a language modeling approach usually estimates $p(w|d)$, a unigram language model based on a given document d, one of the simplest methods for smoothing is based on the maximum likelihood estimate as follows:

$$p_{ml}(w|d) = \frac{c(w;d)}{\sum_w c(w;d)} \quad (5)$$

Unfortunately, the maximum likelihood estimator will often underestimate the probabilities of unseen words in the given document. Hence, it is important to employ smoothing methods that usually discount the probabilities of the words seen in the text and assign the extra probability mass to the unseen words according to some model [2].

Some comprehensive evaluation of smoothing techniques for traditional text retrieval can be found in literature [8, 2]. They have been an important tool to improve the performance of language models in traditional text retrieval. To achieve efficient implementations for large-scale tasks, three representative methods are selected in our scheme, which are popular and relatively efficient. They are discussed in turn below.

The Jelinek-Mercer (JM) Method. This method simply employs a linear interpolation of the maximum likelihood model with the collection model, using a coefficient λ to control the influence:

$$p_\lambda(\omega|d) = (1-\lambda) p_{ml}(\omega|d) + \lambda p(\omega|\mathcal{C}) \quad (6)$$

It is a simple mixture model. A more general Jelinek-Mercer method can be found in [9].

Bayesian Smoothing with Dirichlet Priors (DIR). In general, a language model can be considered as a multinomial distribution, in which the conjugate prior for Bayesian analysis is the Dirichlet distribution with parameters [2] $(\mu p(\omega_1|\mathcal{C}), \mu p(\omega_2|\mathcal{C}), \ldots, \mu p(\omega_n|\mathcal{C}))$. Thus, the smoothing model can be given as:

$$p_\mu(\omega|d) = \frac{c(\omega; d) + \mu p(\omega|\mathcal{C})}{\sum_\omega c(\omega; d) + \mu} \tag{7}$$

Note that μ in the above formula is a DIR parameter that is usually estimated empirically from training sets.

Absolute Discounting Smoothing (ABS). The absolute discounting method subtracts a constant from the counts of seen words for reducing the probabilities of the seen words, meanwhile it increases the probabilities of unseen words by including the collection language model. More specifically, the model can be represented as follows:

$$p_\delta(\omega|d) = \frac{\max(c(\omega; d) - \delta, 0)}{\sum_\omega c(\omega; d)} + \sigma p(\omega|\mathcal{C}) \tag{8}$$

where $\delta \in [0, 1]$ is a discount constant and $\sigma = \delta|d|_\mu/|d|$, so that all probabilities sum to one. Here $|d|_\mu$ is the number of unique terms in document d, and $|d|$ is the total count of words in the document, i.e., $|d| = \sum_\omega c(\omega; d)$.

3 Cross-Language and Cross-Media Image Retrieval

The goal of the photographic retrieval task is to find as many relevant images as possible from an image collection given a multilingual statement describing a user information need. This task intends to simulate the text-based retrieval from photographs with multilingual captions, meanwhile queries for content-based image retrieval will also be offered. In this section, we study techniques to address several open challenges in this retrieval task, including (1) short text query problem, (2) cross-media image retrieval, and (3) cross-language retrieval. In the following part, we first describe the experimental testbed and setup at the ImageCLEF 2007, in which we have participated in the photo retrieval task. We will then conduct the empirical evaluations to address the above challenges and summarize our empirical experiences.

3.1 Experimental Testbed and Setup

The experimental testbed contains 20,000 color photographs with semi-structured captions in English, German and Spanish. For performance evaluations, there are 60 queries, each of them describes the user's information needs by short text in a range of languages including English, Italian, Spanish, French, German, Chinese, Japanese and Russian, and sample images.

For the photographic retrieval task, we have studied the query tasks in English and Chinese (simplified). Both text and visual information are used in our experiments. To evaluate the language models correctly, we employ the *Lemur* toolkit [1]. A list of standard stopwords is used in the parsing step.

To evaluate the influence on the performance of using the different schemes, we have evaluated the methods by trying a variety of different configurations in order to examine every aspects of the solutions. In particular, there groups of performance evaluations will be studied in the subsequent parts.

Table 1. The performance evaluation of Language Models for text-based image retrieval tasks

Run ID	Method	Query	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET	REL
Eng-kl-dir-fb2	KL-DIR	English	English	TEXT	AUTO	FB	0.1660	0.2217	1827	3416
Eng-kl-jm-fb1	KL-JM	English	English	TEXT	AUTO	FB	0.1641	0.2017	1788	3416
Eng-tf-idf-fb3	TF-IDF	English	English	TEXT	AUTO	FB	0.1641	0.2150	1955	3416
Eng-kl-jm-fb2	KL-JM	English	English	TEXT	AUTO	FB	0.1640	0.2033	1870	3416
Eng-kl-abs-fb2	KL-ABS	English	English	TEXT	AUTO	FB	0.1635	0.2017	1757	3416
Eng-okapi-fb2	OKAPI	English	English	TEXT	AUTO	FB	0.1612	0.2333	1674	3416
Eng-kl-abs-fb1	KL-ABS	English	English	TEXT	AUTO	FB	0.1611	0.1950	1700	3416
Eng-kl-dir-fb1	KL-DIR	English	English	TEXT	AUTO	FB	0.1603	0.2117	1682	3416
Eng-kl-abs-fb3	KL-ABS	English	English	TEXT	AUTO	FB	0.1593	0.2000	1797	3416
Eng-kl-dir-fb3	KL-DIR	English	English	TEXT	AUTO	FB	0.1571	0.1867	1823	3416
Eng-kl-jm-fb3	KL-JM	English	English	TEXT	AUTO	FB	0.1566	0.1917	1860	3416
Eng-tf-idf-fb2	TF-IDF	English	English	TEXT	AUTO	FB	0.1560	0.2117	1842	3416
Eng-okapi-fb3	OKAPI	English	English	TEXT	AUTO	FB	0.1540	0.1950	1733	3416
Eng-tf-idf-fb1	TF-IDF	English	English	TEXT	AUTO	FB	0.1540	0.2133	1750	3416
Eng-okapi-fb1	OKAPI	English	English	TEXT	AUTO	FB	0.1492	0.2000	1726	3416
Eng-kl-abs	KL-ABS	English	English	TEXT	AUTO	NOFB	0.1455	0.1883	1570	3416
Eng-okapi	OKAPI	English	English	TEXT	AUTO	NOFB	0.1437	0.1850	1556	3416
Eng-kl-jm	KL-JM	English	English	TEXT	AUTO	NOFB	0.1428	0.1850	1547	3416
Eng-kl-dir	KL-DIR	English	English	TEXT	AUTO	NOFB	0.1419	0.1850	1554	3416
Eng-tf-idf	TF-IDF	English	English	TEXT	AUTO	NOFB	0.1341	0.1900	1539	3416

"TF-IDF" and "OKAPI" are two typical retrieval methods, "KL" denotes Kullback-Leibler divergence based model, "DIR" denotes the smoothing technique using the Dirichlet priors, "ABS" denotes the smoothing using the absolute discounting, and "JM" denotes the Jelinek-Mercer smoothing approach.

3.2 Evaluation of Language Models and Smoothing Techniques.

In our experiments, we study several retrieval methods by language models with different smoothing techniques for the text-based image retrieval tasks. Table 1 shows the results of a number of our submissions with respect to the text based retrieval approaches by Language Models. The listed methods are ranked by the MAP (mean average precision) score. From the results, we can observe that the best approach is the "Eng-kl-dir-fb2" solution, which is based on the KL-divergence language model with the Dirichlet priors smoothing technique. We also found that the retrieval methods by KL-divergence language models do not always outperform the traditional TF-IDF and Okapi approaches, while the language models tend to outperform the TF-IDF and Okapi approaches on average. Further, we found that the retrieval methods with pseudo-relevance feedback (FB) consistently outperform the ones without any feedback. For example, the "Eng-kl-dir" approach is the KL-divergence language model approach using the Dirichlet priors smoothing technique without feedback, which achieved only a MAP score of 0.1419. However, by engaging the relevance feedback, the MAP

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

performance will be importantly improved, such as the "Eng-kl-dir-fb2" solution, which achieved a MAP score of 0.1660. Finally, comparing several different smoothing techniques, there is no a clear evidence that which smoothing technique significantly outperform the others, though the Dirichlet priors smoothing approach achieved the best MAP performance among all runs.

3.3 Cross-Language Image Retrieval

In this part, we study the bilingual image retrieval using Chinese queries and English sources. To this purpose, the first step is to translate the Chinese queries into English. In our experiment, we simply test an online translation tool offered by Google inc[2]. Fig. 1 shows some examples of the translation results.

1 旅社游泳池
CHN2ENG: Hostels Swimming Pool
Original: accommodation with swimming pool

2 具有超过两个塔的教堂
CHN2ENG: have more than two towers of the churches
Original: church with more than two towers

3 在前景的宗教雕像
CHN2ENG: The prospect of religious statues
Original: religious statue in the foreground

4 站立在巴塔哥尼亚山风景前面的团体
CHN2ENG: Patagonia standing in front of the scenic mountain groups
Original: group standing in front of mountain landscape in Patagonia

5 动物游泳
CHN2ENG: Animals swim
Original: animal swimming

6 美国的直路
CHN2ENG: U.S. straight roads
Original: straight road in the USA

7 站在盐田的团体
CHN2ENG: Yantian groups stand
Original: group standing in salt pan

8 接待家庭相片
CHN2ENG: Host family photo
Original: host family posing for a photo

9 的的喀喀湖旁的旅客住处
CHN2ENG: The Titicaca Lake in passenger accommodation
Original: tourist accommodation near Lake Titicaca

10 委内瑞拉旅游景点
CHN2ENG: Venezuela tourist attractions
Original: destinations in Venezuela

Fig. 1. Some examples of Chinese to English query translation in our experiments.

Given the translated results, we conducted the experimental evaluations to examine the retrieval performance. Table 2 shows the experimental results of cross-language retrieval evaluation. From the experimental results, we found that the average retrieval performance of the bilingual retrieval tasks is less than the results of the single language image retrieval as shown in Table 1. For example, for a same retrieval method by the KL-divergence language model with the Dirichlet priors smoothing technique, the scheme "Chn-kl-dir-fb3" achieved only the MAP of 0.1429 in the bilingual retrieval task, while the same approach "Eng-kl-dir-fd3" can achieve the MAP of 0.1571 in the single langauge retrieval

[2] http://www.google.com/language_tools

tasks. Nonetheless, the overall performance of the bilingual approach is quite impressive. In the future work, we will study more advanced translation techniques to improve the results [10].

Table 2. The performance evaluation for cross-language image retrieval tasks between Chinese (simplified) queries and English sources.

Run ID	Method	Query	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET	REL
Chn-tf-idf-fb3	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1574	0.2000	1874	3416
Chn-kl-dir-fb3	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1429	0.1650	1709	3416
Chn-tf-idf-fb2	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1413	0.1783	1790	3416
Chn-kl-abs-fb3	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1406	0.1667	1713	3416
Chn-kl-abs-fb2	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1385	0.1500	1732	3416
Chn-kl-dir-fb2	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1382	0.1600	1763	3416
Chn-kl-jm-fb2	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1380	0.1533	1801	3416
Chn-kl-jm-fb3	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1378	0.1600	1748	3416
Chn-kl-jm-fb1	KL-JM	Chinese S	English	TEXT	AUTO	FB	0.1345	0.1533	1696	3416
Chn-kl-dir-fb1	KL-DIR	Chinese S	English	TEXT	AUTO	FB	0.1333	0.1650	1672	3416
Chn-okapi-fb3	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1312	0.1517	1646	3416
Chn-kl-abs-fb1	KL-ABS	Chinese S	English	TEXT	AUTO	FB	0.1309	0.1417	1675	3416
Chn-tf-idf-fb1	TF-IDF	Chinese S	English	TEXT	AUTO	FB	0.1286	0.1767	1553	3416
Chn-okapi	OKAPI	Chinese S	English	TEXT	AUTO	NOFB	0.1268	0.1417	1404	3416
Chn-kl-dir	KL-DIR	Chinese S	English	TEXT	AUTO	NOFB	0.1265	0.1467	1410	3416
Chn-kl-abs	KL-ABS	Chinese S	English	TEXT	AUTO	NOFB	0.1264	0.1483	1411	3416
Chn-kl-jm	KL-JM	Chinese S	English	TEXT	AUTO	NOFB	0.1252	0.1450	1415	3416
Chn-okapi-fb1	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1237	0.1350	1654	3416
Chn-tf-idf	TF-IDF	Chinese S	English	TEXT	AUTO	NOFB	0.1223	0.1567	1388	3416
Chn-okapi-fb2	OKAPI	Chinese S	English	TEXT	AUTO	FB	0.1177	0.1383	1540	3416

"TF-IDF" and "OKAPI" are two typical retrieval methods, "KL" denotes Kullback-Leibler divergence based model, "DIR" denotes the smoothing technique using the Dirichlet priors, "ABS" denotes the smoothing using the absolute discounting, and "JM" denotes the Jelinek-Mercer smoothing approach.

3.4 Cross-Media Image Retrieval

In this task we study the combination of text and visual information for cross-media image retrieval. We consider a simple combination scheme to combine the information from both the textual and visual modalities. Specifically, for a given query, we first rank the images using the language modeling techniques. We then measure the similarity of the top ranked images with respect to the sample images of the query. Finally, we combine the similarity values from both textual and visual modalities and re-rank the retrieval results based on the overall similarity scores.

In our experiment, three types of low-level visual features are engaged: color, shape, and texture [11, 12]. For color features, we use the grid color moment. Each image is partitioned into 3×3 grids and three types of color moments are extracted for representing color content of each grid. Thus, an 81-dimensional color moment is adopted for the color feature. For shape features, we employ the edge direction histogram. A Canny edge detector is used to acquire the edge images and then the edge direction histogram is computed from the edges. Each histogram is quantized into 36 bins of 10 degrees each. An additional bin is used to count the number of pixels without edge information. Hence, a 37-dimensional edge direction histogram is used for the shape feature. For texture features, we adopt the Gabor feature [13]. Each image is scaled to 64×64. Gabor wavelet transformation is applied on the scaled image with 5 scale levels and 8 orientations, which results in 40 subimages. For each subimage, three moments

are computed: mean, variance, and skewness. Thus, a 120-dimensional feature vector is adopted for the texture feature. In total, a 238-dimensional feature vector is employed to represent each of images in the testbed.

Table 3 shows the cross-media retrieval results. Before examining the experimental results, we potentially expect that the cross-media retrieval approaches are very likely to improve the text-based retrieval approaches. Unfortunately, the observations from the empirical results are somewhat surprising, which are not consistent with our intuition. Two possible reasons may explain this conflict results. One reason may be the wrong results reported from the official evaluations. Another possible reason may be due to some mistake engaged when combining the visual and textual similarity values. (We will examine all of the reasons since our similar approach in ImageCLEF2005 achieved much better results [14].)

In fact, in our current implementation, we do not engage other advanced combination methods. In future work, we will try some nonlinear combination methods. For example, we can train an SVM classifier with the sample images and then apply the classifier to re-rank the top images from the text retrieval results. We will consider this in our future work and expect it will importantly improve current results.

Table 3. The performance evaluation for cross-media image retrieval tasks with queries of both textual and visual information.

Run ID	Query	Method	Source	Modality	RunType	QE/RF	MAP	P10	REL_RET	REL
Visual	Euclidean	Visual	Visual	VISUAL	AUTO	NO	0.0511	0.2067	883	3416
Eng-kl-abs-fb1-tv3	KL-ABS	English	English	MIXED	AUTO	FB	0.0296	0.1333	866	3416
Eng-kl-abs-fb2-tv3	KL-ABS	English	English	MIXED	AUTO	FB	0.0222	0.0900	866	3416
Eng-kl-abs-fb3-tv3	KL-ABS	English	English	MIXED	AUTO	FB	0.0187	0.0633	869	3416
Eng-tf-idf-fb1-tv2	TF-IDF	English	English	MIXED	AUTO	FB	0.0146	0.0383	849	3416
Eng-okapi-fb1-tv2	OKAPI	English	English	MIXED	AUTO	FB	0.0119	0.0100	831	3416
Eng-kl-dir-fb1-tv2	KL-DIR	English	English	MIXED	AUTO	FB	0.0114	0.0233	827	3416
Eng-tf-idf-fb2-tv2	TF-IDF	English	English	MIXED	AUTO	FB	0.0111	0.0133	837	3416
Eng-kl-abs-fb1-tv2	KL-ABS	English	English	MIXED	AUTO	FB	0.0110	0.0150	837	3416
Eng-kl-jm-fb1-tv2	KL-JM	English	English	MIXED	AUTO	FB	0.0101	0.0083	828	3416
Eng-tf-idf-fb3-tv2	TF-IDF	English	English	MIXED	AUTO	FB	0.0101	0.0050	827	3416
Eng-kl-abs-fb2-tv2	KL-ABS	English	English	MIXED	AUTO	FB	0.0099	0.0083	816	3416
Eng-kl-abs-fb3-tv2	KL-ABS	English	English	MIXED	AUTO	FB	0.0099	0.0083	815	3416
Eng-kl-dir-fb2-tv2	KL-DIR	English	English	MIXED	AUTO	FB	0.0099	0.0100	822	3416
Eng-kl-jm-fb3-tv2	KL-JM	English	English	MIXED	AUTO	FB	0.0099	0.0067	800	3416
Eng-okapi-fb3-tv2	OKAPI	English	English	MIXED	AUTO	FB	0.0099	0.0117	819	3416
Eng-kl-dir-fb3-tv2	KL-DIR	English	English	MIXED	AUTO	FB	0.0098	0.0100	825	3416
Eng-kl-jm-fb2-tv2	KL-JM	English	English	MIXED	AUTO	FB	0.0098	0.0083	802	3416
Eng-okapi-fb2-tv2	OKAPI	English	English	MIXED	AUTO	FB	0.0093	0.0133	807	3416
Eng-kl-abs-fb3-tv1	KL-ABS	English	English	MIXED	AUTO	FB	0.0081	0.0083	728	3416
Eng-kl-abs-fb2-tv1	KL-ABS	English	English	MIXED	AUTO	FB	0.0072	0.0100	725	3416
Eng-kl-abs-fb1-tv1	KL-ABS	English	English	MIXED	AUTO	FB	0.0070	0.0067	729	3416

"TF-IDF" and "OKAPI" are two typical retrieval methods, "KL" denotes Kullback-Leibler divergence based model, "DIR" denotes the smoothing technique using the Dirichlet priors, "ABS" denotes the smoothing using the absolute discounting, and "JM" denotes the Jelinek-Mercer smoothing approach.

4 Conclusions

In this paper we reported our empirical study of cross-language and cross-media image retrieval in the ImaegCLEF 2007 campaign. We have conducted three parts of empirical evaluations for three different purposes. One is to evaluate the techniques of language models and smoothing techniques with applications to text-based image retrieval. In this year, we found that the language models

approaches did not achieve significantly promising results as we achieved in the ImageCLEF2005 campaign. The main reason is that the testbed in this year is totally different from the one in 2005. In this year, images are only associated with very short text captions, which makes the text retrieval models difficult to achieve excellent performance. The second major evaluation is the cross-media image retrieval by combining both textual and visual information. In the evaluation, some strange observations were found. We will study the problem in more details in our future work. Finally, we also examined a commercial language translation tool for the cross-language retrieval tasks and found good retrieval results. In future work, we will study more effective techniques to overcome the limitation of current approaches.

References

1. P. Clough, H. Müeller, T. Deselaers, M. Grubinger, T. Lehmann, J. Jensen, and W. Hersh, "The CLEF 2005 cross language image retrieval track," in *Proceedings of the Cross Language Evaluation Forum 2005*. Springer Lecture Notes in Computer science, 2005.
2. C. Zhai and J. Lafferty, "A study of smoothing methods for language models applied to ad hoc information retrieval," in *ACM International SIGIR Conference (SIGIR'01)*, 2001, pp. 334–342.
3. R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley, 1999.
4. S. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford, "Okapi at trec-3," in *The Third Text REtrieval Conference (TREC-3)*. NIST.
5. C. E. Shannon, "Prediction and entropy of printed English," *Bell Sys. Tech. Jour.*, vol. 30, pp. 51–64, 1951.
6. C. Zhai and J. Lafferty, "Model-based feedback in the kl-divergence retrieval model," in *Proc. Tenth International Conference on Information and Knowledge Management (CIKM2001)*, 2001, pp. 403–410.
7. T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
8. D. Hiemstra, "Term-specific smoothing for the language modeling approach to information retrieval: the importance of a query term," in *Proceedings 25th ACM SIGIR conference*, 2002, pp. 35–41.
9. F. Jelinek and R. Mercer, "Interpolated estimation of markov sourceparameters from sparse data," *Pattern Recognition in Practice*, pp. 381–402, 1980.
10. C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.
11. C.-H. Hoi and M. R. Lyu, "A novel log-based relevance feedback technique in content-based image retrieval," in *Proceedings 12th ACM International Conference on Multimedia (MM 2004)*, New York, USA, 2004, pp. 24–31.
12. S. C. H. Hoi, M. R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509–524, 2006.
13. P. Wu, B. Manjunath, S. Newsam, and H. Shin, "A texture descriptor for browsing and similarity retrieval," *Journal of Signal Processing: Image Communication*, vol. 16, no. 1-2, pp. 33–43, Sep 2000. [Online]. Available: http://vision.ece.ucsb.edu/publications/00ICJ.pdf
14. S. C. Hoi, J. Zhu, and M. R. Lyu, "Cuhk at imageclef 2005: Cross-language and cross-media image retrieval."