CUHK Experiments with ImageCLEF 2005[*]

Steven C.H. Hoi, Jianke Zhu and Michael R. Lyu
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{chhoi, jkzhu, lyu}@cse.cuhk.edu.hk

Abstract

This paper describes the empirical studies of cross-language and cross-media retrieval for the ImageCLEF competition in 2005. It reports the empirical summary of the work of CUHK (The Chinese University of Hong Kong) at ImageCLEF 2005. This is the first participation of our group at ImageCLEF. The task we participated this year is the "Bilingual ad hoc retrieval" task. There are three major focuses and contributions in our participation. The first is the empirical evaluations of language models and the smoothing strategies for cross-language image retrieval. The second is the evaluations of cross-media image retrieval, i.e., combining text and visual content for image retrieval. The last one is the evaluation of the bilingual image retrieval between English and Chinese. We provide empirical analysis on the experimental results. From the official testing results of the Bilingual ad hoc retrieval task, we achieve the highest MAP result (0.4135) in the monolingual query among all organizations.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [**Database Managment**]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

Language Models, Text Based Image Retrieval, Multimodal Image Retrieval, Cross-Language Retrieval, Cross-Media Retrieval, Smoothing Strategy

1 Introduction

Visual information retrieval has been an active research topic for many years. Although content-based image retrieval (CBIR) has been received considerable studies in the community [9], there is so far few benchmark image dataset available. The CLEF (Cross Language Evaluation Forum) organization [7] began the ImageCLEF campaign from 2003 for benchmark evaluation of cross-language image retrieval [4]. ImageCLEF 2005 offers four different tasks: bilingual ad hoc retrieval,

[*]The work described in this paper was fully supported by two grants: Innovation and Technology Fund ITS/105/03, and the Research Grants Council Earmarked Grant CUHK4205/04E.

interactive search, medical image retrieval and automatic image annotation task. This is the first participation of our CUHK group (The Chinese University of Hong Kong) at ImageCLEF. The task we participated this year is the “Bilingual ad hoc retrieval”.

In the past decade, traditional information retrieval mainly focused on the document retrieval problems [3]. Along with more and more attentions in multimedia information retrieval in recent years, the cross-language and cross-media retrieval have been put forward as an important research topic in the community [4]. The cross-language image retrieval is to tackle the multimodal information retrieval task by unifying the techniques from traditional information retrieval, natural language processing (NLP), and traditional CBIR solutions.

In this participation, we offer the main contributions in three aspects. The first is the empirical evaluation of language models and the smoothing strategies for cross-language image retrieval. The second is the evaluation of cross-media image retrieval, i.e., combining text and visual content for image retrieval. The last one is the methodology and empirical evaluation of the bilingual image retrieval between English and Chinese.

The rest of this paper is organized as follows. Section 2 introduces the TF-IDF retrieval model and the language model based retrieval methods. Section 3 describes the details of our implementation for this participation, and outlines our empirical study on the cross-language and cross-media retrieval system. Finally section 4 concludes our work.

2 Language Models for Text Based Image Retrieval

In this participation, we conducted extensive experiments to evaluate the performance of Language Models and the influences of different smoothing strategies. More specifically, two kinds of retrieval models are studied in our experiments: (1) The TF-IDF retrieval model (2) The KL-divergence language models based method. The smoothing strategies for Language Models are evaluated in our experiments [11]: (1) Jelinek-Mercer (JM), (2) Dirichlet prior (DIR), (3) Absolute discounting (ABS).

2.1 TF-IDF Similarity Measure for Information Retrieval

We incorporate the Language Models (LM) with the TF-IDF similarity measure[3]. TF-IDF is widely used in information retrieval, which is a way of weighting the relevance of a query to a document. The main idea of TF-IDF is to represent each document by a vector in the size of the overall vocabulary. Each document D_i is then represented as a vector $(w_{i1}, w_{i2}), \cdots, w_{in}$ if n is the size of the vocabulary. The entry $w_{i,j}$ is calculated as:

$$w_{ij} = TF_{ij} \times \log(IDF_j) \tag{1}$$

where TF_{ij} is the term frequency of the j_{th} word in the vocabulary in the document D_i, i.e. the number of occurrences. IDF_j is the inverse document frequency of the j_{th} term, given as

$$IDF_j = \frac{\#documents}{\#documents\ containing\ the\ j_{th}\ term} \tag{2}$$

The similarity between two documents is then defined as the cosine of the angle between the two vectors.

2.2 Language Modeling for Information Retrieval

A statistical language model, or more simply a language model, is a probabilistic mechanism for generating text. The first serious statistical language modeler was Claude Shannon [8]. In exploring the application of his newly founded theory of information to human language, thought of purely as a statistical source, Shannon measured how well simple n-gram models did at predicting, or compressing, natural text. In the past several years there has been significant interest in the use of language modeling methods for a variety of text retrieval and natural language processing tasks [10].

2.2.1 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

$$D(p||q) = \sum_{x} p(x) log \frac{p(x)}{q(x)} \tag{3}$$

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 still often useful to think of the KL-divergence as a "distance" between distributions [5].

2.2.2 The KL-divergence based Retrieval Model

For 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 that 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 language models respectively. The relevance value of d with respect to q can be measured by the following negative KL-divergence function [10]:

$$-D(\hat{\theta}_Q||\hat{\theta}_D) = \sum_{w} p(w|\hat{\theta}_Q) logp(w|\hat{\theta}_D) + (-\sum_{w} p(w|\hat{\theta}_Q) logp(w|\hat{\theta}_Q)) \tag{4}$$

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:

$$p(w|\hat{\theta}_D) \quad = \quad \begin{cases} p_s(w|d) & \text{if word } w \text{ is seen} \\ \alpha_d p(w|\mathcal{C}) & \text{otherwise} \end{cases} \tag{5}$$

where $p_s(w|d)$ is the smoothed probability of a word seen 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 [10]. In the subsequent section, we discuss several smoothing techniques in details.

2.3 Several Smoothing Techniques

A smoothing method may be as simple as adding an extra count to every word, or words of different count are treated differently. In order to solve the problem efficiently, we select three representative methods that are popular and relatively efficient. The three methods are described below.

2.3.1 Jelinek-Mercer (JM)

This method involves a linear interpolation of the maximum likelihood model with the collection model, using a coefficient λ to control the influence of each model.

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

Thus, this is a simple mixture model (but we preserve the name of the more general Jelinek-Mercer method which involves deleted-interpolation estimation of linearly interpolated n-gram models.

2.3.2 Dirichlet prior (DIR)

A language model is a multinomial distribution, for which the conjugate prior for Bayesian analysis is the Dirichlet distribution with parameters $(\mu(\omega_1|C), \mu p(\omega_2|C), \ldots, \mu p(\omega_n|C))$. Thus, the model is given by

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

The Laplace method is a special case of the technique.

2.3.3 Absolute discounting (ABS)

The idea of the absolute discounting method is to lower the probability of seen words by subtracting a constant from their counts. It is similar to the Jelinek-Mercer method, but differs in that it discounts the seen word probability by subtracting a constant instead of multiplying it by $1 - \lambda$. The model is given by

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

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 documents, so that $|d| = \sum_\omega c(\omega; d)$.

Table 1: Summary of three primary smoothing methods used in our submission

Method	$p_s(\omega\|d)$	α_d	parameter
JM	$(1 - \lambda)p_{ml}(\omega\|d) + \lambda p(\omega\|C)$	λ	λ
DIR	$\frac{c(\omega;d)+\mu p(\omega\|C)}{\sum_\omega c(\omega;d)+\mu}$	$\frac{\mu p(\omega\|C)}{\sum_\omega c(\omega;d)+\mu}$	μ
ABS	$p_\delta(\omega\|d) = \frac{\max(c(\omega;d)-\delta,0)}{\sum_\omega c(\omega;d)} + \delta\|d\|_\mu \delta p(\omega\|C)$	$\frac{\delta\|d\|_\mu}{\|d\|}$	δ

The three methods are summarized in Table 1 in terms of $p_s(\omega|d)$ and α_d in the general form. It is easy to see that a larger parameter value means smoothing in all cases. Retrieval using any of the three methods can be very efficiently, when the smoothing parameter is given in advance. It is as efficient as scoring using a TF-IDF model.

3 Cross-Language and Cross-Media Image Retrieval

In this section, we describe the experimental setup and our experimental development at the ImageCLEF 2005. In addition, we analyze the results of our submission.

3.1 Experimental Setup

The bilingual ad hoc retrieval task is to find as many relevant images as possible for each given topic. The St. Andrew collection is used as the benchmark dataset in the campaign. The collection consists of 28,133 images, all of which associate with textual captions written in British English (the target language). The caption consists of 8 fields including *title*, *photographer*, *location*, *date*, and one or more pre-defined categories (all manually assigned by domain experts). In the ImageCLEF 2005 campaign, there are totally 28 queries for each language. For each query, two image samples are given. Figure 1. shows a query example of images, title and narrative texts in the campaign.

<top>

<num> Number: 1 </num>

<title> aircraft on the ground </title>

<narr> Relevant images will show one or more airplanes positioned on the ground. Aircraft do not have to be the focus of the picture, although it should be possible to make out that the picture contains aircraft. Pictures of aircraft flying are not relevant and pictures of any other flying object (e.g. birds) are not relevant. </narr>

</top>

Figure 1: A query example in the ImageCLEF 2005 campaign.

3.2 Overview of Our Development

For the Bilingual ad hoc retrieval task, we studied the query tasks in English and Chinese (simplified). Both text and visual information are used in our experiments. To study the language models, we employ the *Lemur* toolkit [2] in our experiments. A list of standard stopwords is used in the parsing step.

To evaluate the influence on the performance by different schemes, we produced the results by using different configurations. Tables 2 shows the configurations and the experimental results in detail. In total, 36 runs with different configurations are submitted in our submission.

3.3 Analysis on the Experimental Results

In this part, we empirically analyze the experimental results of our submission. The goal of our evaluation is to check whether the language model is effective for cross-language image retrieval and what kinds of smoothing techniques achieve better performance. Moreover, we like to know the performance comparison between the Chinese query and the monolingual query.

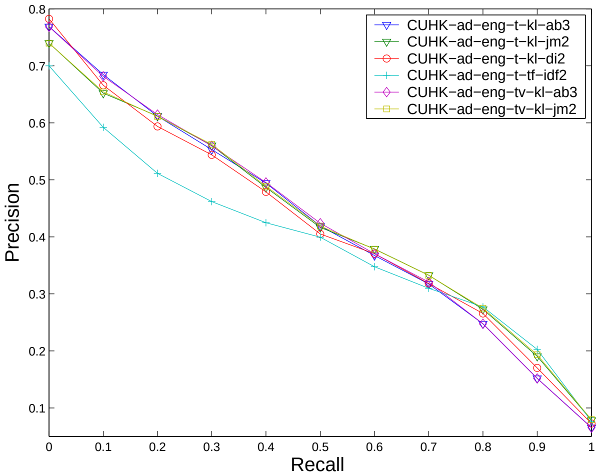

Figure 2: Experimental Result of Precision vs. Recall with Selected Configuration

Table 2: The configurations and testing results of our submission

Run ID	Language	QE	Modality	Method	MAP
CUHK-ad-eng-t-kl-ab1	english	without	text	KL-LM-ABS	0.3887
CUHK-ad-eng-t-kl-ab2	english	with	text	KL-LM-ABS	0.4055
CUHK-ad-eng-t-kl-ab3	english	with	text	KL-LM-ABS	0.4082
CUHK-ad-eng-t-kl-jm1	english	without	text	KL-LM-JM	0.3844
CUHK-ad-eng-t-kl-jm2	english	with	text	KL-LM-JM	0.4115
CUHK-ad-eng-t-kl-di1	english	without	text	KL-LM-DIR	0.382
CUHK-ad-eng-t-kl-di2	english	with	text	KL-LM-DIR	0.3999
CUHK-ad-eng-t-tf-idf1	english	without	text	TF-IDF	0.351
CUHK-ad-eng-t-tf-idf2	english	with	text	TF-IDF	0.3574
CUHK-ad-eng-tn-kl-ab1	english	without	text	KL-LM-ABS	0.3877
CUHK-ad-eng-tn-kl-ab2	english	with	text	KL-LM-ABS	0.3838
CUHK-ad-eng-tn-kl-ab3	english	with	text	KL-LM-ABS	0.4083
CUHK-ad-eng-tn-kl-jm1	english	without	text	KL-LM-JM	0.3762
CUHK-ad-eng-tn-kl-jm2	english	with	text	KL-LM-JM	0.4018
CUHK-ad-eng-tn-kl-di1	english	without	text	KL-LM-DIR	0.3921
CUHK-ad-eng-tn-kl-di2	english	with	text	KL-LM-DIR	0.399
CUHK-ad-eng-tn-tf-idf1	english	without	text	TF-IDF	0.3475
CUHK-ad-eng-tn-tf-idf2	english	with	text	TF-IDF	0.366
CUHK-ad-eng-v	english	without	vis	Moment-DCT	0.0599
CUHK-ad-eng-tv-kl-ab1	english	without	text+vis	KL-LM-ABS	0.3941
CUHK-ad-eng-tv-kl-ab3	english	with	text+vis	KL-LM-ABS	0.4108
CUHK-ad-eng-tv-kl-jm1	english	without	text+vis	KL-LM-JM	0.3878
CUHK-ad-eng-tv-kl-jm2	english	with	text+vis	KL-LM-JM	0.4135
CUHK-ad-eng-tnv-kl-ab2	english	with	text+vis	KL-LM-ABS	0.3864
CUHK-ad-eng-tnv-kl-ab3	english	with	text+vis	KL-LM-ABS	0.4118
CUHK-ad-eng-tnv-kl-jm1	english	without	text+vis	KL-LM-JM	0.3787
CUHK-ad-eng-tnv-kl-jm2	english	with	text+vis	KL-LM-JM	0.4041
CUHK-ad-chn-t-kl-ab1	chinese	without	text	KL-LM-ABS	0.1815
CUHK-ad-chn-t-kl-ab2	chinese	with	text	KL-LM-ABS	0.1842
CUHK-ad-chn-t-kl-jm1	chinese	without	text	KL-LM-JM	0.1821
CUHK-ad-chn-t-kl-jm2	chinese	with	text	KL-LM-JM	0.2027
CUHK-ad-chn-tn-kl-ab1	chinese	without	text	KL-LM-ABS	0.1758
CUHK-ad-chn-tn-kl-ab2	chinese	with	text	KL-LM-ABS	0.1527
CUHK-ad-chn-tn-kl-ab3	chinese	with	text	KL-LM-ABS	0.1834
CUHK-ad-chn-tn-kl-jm1	chinese	without	text	KL-LM-JM	0.1843
CUHK-ad-chn-tn-kl-jm2	chinese	with	text	KL-LM-JM	0.2024

LM denotes Language Model, KL denotes Kullback-Leibler divergence based, DIR denotes the smoothing using the Dirichlet priors, ABS denotes the smoothing using Absolute discounting, JM denotes the Jelinek-Mercer smoothing.

3.3.1 Empirical Analysis of Language Models

Figure 2 and Figure 3 plot the curves of *Precision* vs. *Recall* and the curves of *Precision* vs. *Number of Returned Documents* respectively. From the experimental results in Figure 2 and Figure 3 as well as Table 2, one can observe that the KL-divergence language model outperforms the simple TF-IDF retrieval model importantly (around 5%). In evaluation of the smoothing techniques, we observe that the Jelinek-Mercer smoothing and Absolute Discounting Smoothing yield better results than the Dirichlet prior (DIR).

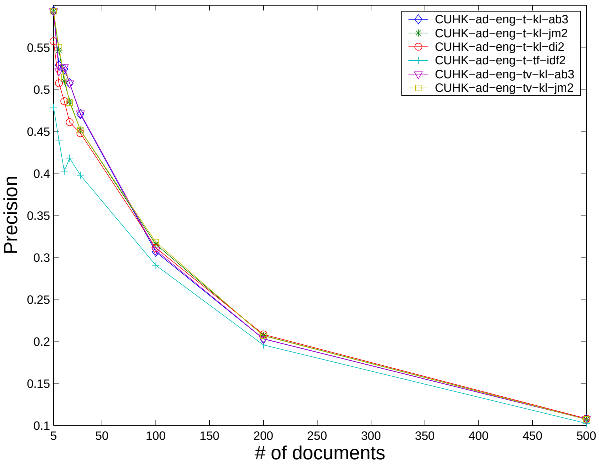

Figure 3: Experimental Result of Precision vs. Number of Returned Documents with Selected Configuration

3.3.2 Cross-Language Retrieval: Chinese-To-English Query Translation

To deal with the Chinese queries for retrieving English documents, we first adopt a Chinese segmentation tool from the Linguistic Data Consortium (LDC) [1], i.e., the "LDC Chinese segmenter" [1], to extract the Chinese words from the given query sentences. The segmentation step is important toward effective query translation. Figure 4 shows the Chinese segmentation results of part queries. We can see that the results can still be improved.

For the bilingual query translation, the second step is to translate the extracted Chinese words into English words using a Chinese-English dictionary. In our experiment, we employ the LDC Chinese-to-English Wordlist [1] for the translations. The final translated queries are obtained by combining the translation results.

From the experimental results shown in Table 2, we can observe that the mean average precision of Chinese-To-English Queries is about the half of the monolingual queries. There are a lot of ways to improve the performance. One is to improve the Chinese segmentation algorithm. Some post-processing tricks may be effective for improving the performance. Moreover, the translation results can be further refined. One can tune better results by adopting some Natural Language Processing techniques [6].

3.3.3 Cross-Media Retrieval: Re-Ranking Scheme with Text and Visual Content

In this participation, we study the combination of text and visual content for cross-media image retrieval. In our development, we suggest the re-ranking scheme in combination with text and visual content. For a given query, we first rank the images by using the language modeling techniques. On the top ranking images, we then re-rank the images by measuring the visual similarity to the query.

[1] It can be downloaded from: http://www.ldc.upenn.edu/Projects/Chinese/seg.zip .

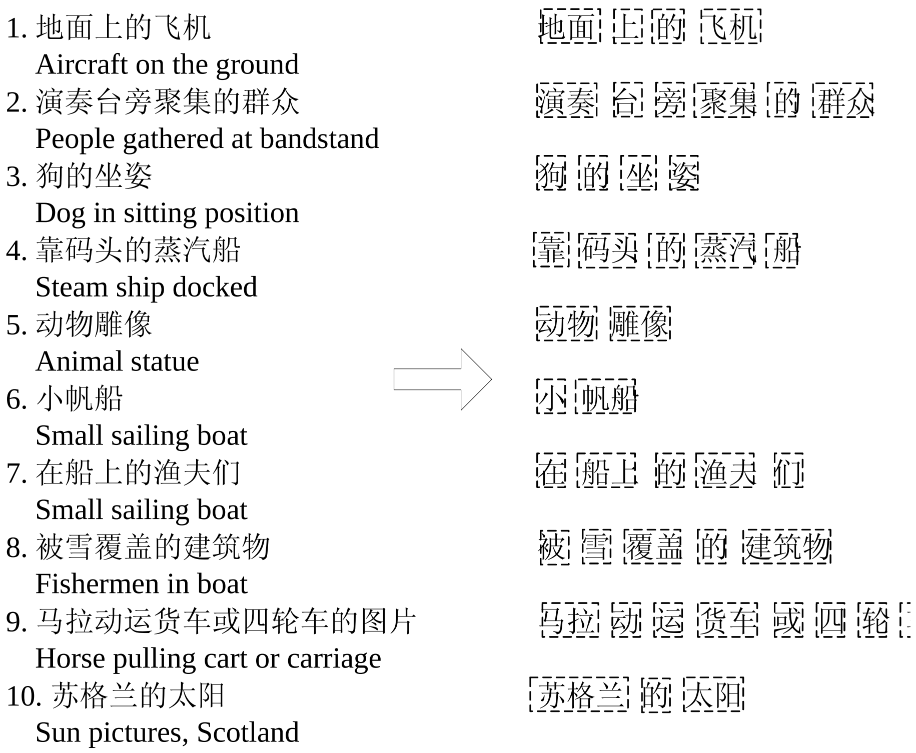

Figure 4: Chinese segmentation results of part Chinese (Simplified) queries

In our experiment, two kinds of visual features are used: texture and color features. For the texture feature, the discrete cosine transform (DCT) is engaged to calculate coefficients that multiply the basis functions of the DCT. Applying the DCT to an image yields a set of coefficients to represent the texture of the image. In our implementation, a block-DCT (block size 8x8) is applied on the normalized input images which generate a 256-dimensional DCT feature. For the color feature, 9-dimensional color moment is extracted for each image. In total, each image is represented by a 265-dimensional feature vector.

As shown in Table 2, the MAP of query results using only the visual information is about 6%, which is much lower than the text information with over 40%. From the experimental results, we can observe the re-ranking scheme only produce a marginal improvement compared with the text only approaches. Some reasons can be explained for the results. One is the engaged visual features not effective enough to discriminate the images. Another possible reason is that the ground truth images in the given query may not be quite different in visual content. It is interesting to study more effective features and learning methods for improving the performance.

3.3.4 Query Expansion for Information Retrieval

From the experimental results in Table 2, we observe that all the queries are greatly enhanced by adopting Query Expansion [2] (QE). The average improvement for all the queries is around 1.71% which accounts %4.12 of the maximum MAP of 41.35%. It is interesting to find that the QE especially benefits a lot for the Jelinek-Mercer smoothing method, the mean gain with QE is about 2.49% which accounts %6.02 of the maximum MAP of 41.35%.

[2] Query expansion refers to adding further terms to a text query (e.g. through PRF or thesaurus) or images to a visual query

4 Conclusions

In this paper, we reported our empirical studies of cross-language and cross-media image retrieval at the ImaegCLEF 2005 campaign. We addressed three major focuses and contributions in our participation. The first is the empirical evaluations of Language Models and the smoothing strategies for Cross-Language image retrieval. The second one is the evaluation of Cross-Media image retrieval, i.e., combining text and visual content for image retrieval. The last one is the evaluation of the Bilingual image retrieval between English and Chinese. We conducted empirical analysis on the experimental results and provided the empirical summary of our participation.

References

[1] Http://www.ldc.upenn.edu/projects/chinese/.

[2] Http://www.lemurproject.org/.

[3] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval.* Addison Wesley, 1999.

[4] P. Clough, H. Mueller, and M. Sanderson. The clef cross language image retrieval track (imageclef) 2004. In *In the Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004) (LNCS).* Springer, 2004.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[6] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing.* The MIT Press, 1999.

[7] J. Savoy. Report on clef-2001 experiments (cross language evaluation forum). In *LNCS 2406*, pages 27–43, 2002.

[8] C. E. Shannon. Prediction and entropy of printed english. *Bell Sys. Tech. Jour.*, 30:51–64, 1951.

[9] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[10] Chengxiang Zhai and John Lafferty. Model-based feedback in the kl-divergence retrieval model. In *In Tenth International Conference on Information and Knowledge Management (CIKM2001)*, pages 403–410, 2001.

[11] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'01)*, pages 334–342, 2001.