

Cheshire at GeoCLEF 2008: Text and Fusion Approaches for GIR

Ray R. Larson
School of Information
University of California, Berkeley, USA
`ray@sims.berkeley.edu`

Abstract

In this paper we will briefly describe the approaches taken by Berkeley for the main GeoCLEF 2008 tasks (Mono and Bilingual retrieval). The approach this year used probabilistic text retrieval based on logistic regression and incorporating blind relevance feedback for all of the runs and in addition we ran a number of tests combining this type of search with OKAPI BM25 searches using a fusion approach. All translation for bilingual tasks was performed using the LEC Power Translator PC-based MT system.

Our results were good overall with Cheshire systems runs appearing in the top 5 participants for each task (German, English and Portuguese both Monolingual and Bilingual) with the highest ranked runs for Monolingual Portuguese and for Bilingual German, English and Portuguese. All of these top-ranked runs used the fusion approach.

However, once again this year we did not attempt to do any specialized geographic processing, because it appears that purely textual approaches to GIR are more effective when only textual topics, lacking explicit geographic coordinate constraints, are used.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval

General Terms

Algorithms, Performance, Measurement

Keywords

Cheshire II, Logistic Regression

1 Introduction

Geographic Information Retrieval (GIR) as it was originally defined was concerned with providing access to georeferenced information resources using a combination of Information Retrieval (IR) and Data Retrieval (DR) (or database) methods[6]. In GeoCLEF the nature of the topics has tended to emphasize the IR aspects of GIR, largely because no explicit georeferencing of documents or topics has been supplied to the experimenter.

Without the explicit georeferencing of documents and/or topics the experimenter (or searcher) is faced with attempting to provide such georeferencing and therefore solving all of the attendant

problems of ambiguity and multiplicity of toponyms and the issues of name polysemy that explicit georeferencing is intended to alleviate. An alternative approach is to, in effect, ignore geographic clues by treating them like any other term in a normal IR search process. Our approach this year has been to take this latter approach, and use only text retrieval methods on the provided topics with no explicit identification or treatment of toponyms.

This paper describes the retrieval algorithms and evaluation results for Berkeley's official submissions for the GeoCLEF 2008 track. All of the submitted runs were automatic without manual intervention in the queries (or translations). We submitted nine Monolingual runs (three German, three English, and three Portuguese) and eighteen Bilingual runs (three runs for each of the three languages to each other language). The runs varied in the topic elements used, and whether or not a fusion approach (described below) was used.

This paper first describes the retrieval algorithms and fusion operations used for our submissions, followed by a discussion of the processing used for the runs. We then examine the results obtained for our officially submitted runs, and finally present conclusions and future directions for GeoCLEF participation.

2 The Retrieval Algorithms and Fusion Operators

Note that this section is virtually identical to one that appears in our Adhoc-TEL and Domain Specific papers, with the addition of the Okapi BM-25 description and fusion operations subsections. The basic form and variables of the *Logistic Regression* (LR) algorithm used for all of our submissions was originally developed by Cooper, et al. [5]. As originally formulated, the LR model of probabilistic IR attempts to estimate the probability of relevance for each document based on a set of statistics about a document collection and a set of queries in combination with a set of weighting coefficients for those statistics. The statistics to be used and the values of the coefficients are obtained from regression analysis of a sample of a collection (or similar test collection) for some set of queries where relevance and non-relevance has been determined. More formally, given a particular query and a particular document in a collection $P(R | Q, D)$ is calculated and the documents or components are presented to the user ranked in order of decreasing values of that probability. To avoid invalid probability values, the usual calculation of $P(R | Q, D)$ uses the "log odds" of relevance given a set of S statistics, s_i , derived from the query and database, such that:

$$\log O(R | Q, D) = b_0 + \sum_{i=1}^S b_i s_i \quad (1)$$

where b_0 is the intercept term and the b_i are the coefficients obtained from the regression analysis of the sample collection and relevance judgements. The final ranking is determined by the conversion of the log odds form to probabilities:

$$P(R | Q, D) = \frac{e^{\log O(R|Q,D)}}{1 + e^{\log O(R|Q,D)}} \quad (2)$$

2.1 TREC2 Logistic Regression Algorithm

For GeoCLEF we used a version the Logistic Regression (LR) algorithm that has been used very successfully in Cross-Language IR by Berkeley researchers for a number of years[3]. The formal definition of the TREC2 Logistic Regression algorithm used is:

$$\begin{aligned} \log O(R|C, Q) &= \log \frac{p(R|C, Q)}{1 - p(R|C, Q)} = \log \frac{p(R|C, Q)}{p(\bar{R}|C, Q)} \\ &= c_0 + c_1 * \frac{1}{\sqrt{|Q_c|} + 1} \sum_{i=1}^{|Q_c|} \frac{qtf_i}{ql + 35} \end{aligned}$$

$$\begin{aligned}
& + c_2 * \frac{1}{\sqrt{|Q_c|} + 1} \sum_{i=1}^{|Q_c|} \log \frac{t f_i}{c l + 80} \\
& - c_3 * \frac{1}{\sqrt{|Q_c|} + 1} \sum_{i=1}^{|Q_c|} \log \frac{c t f_i}{N_t} \\
& + c_4 * |Q_c|
\end{aligned} \tag{3}$$

where C denotes a document component (i.e., an indexed part of a document which may be the entire document) and Q a query, R is a relevance variable,

$p(R|C, Q)$ is the probability that document component C is relevant to query Q ,

$p(\overline{R}|C, Q)$ the probability that document component C is *not relevant* to query Q , which is $1.0 - p(R|C, Q)$

$|Q_c|$ is the number of matching terms between a document component and a query,

$q t f_i$ is the within-query frequency of the i th matching term,

$t f_i$ is the within-document frequency of the i th matching term,

$c t f_i$ is the occurrence frequency in a collection of the i th matching term,

$q l$ is query length (i.e., number of terms in a query like $|Q|$ for non-feedback situations),

$c l$ is component length (i.e., number of terms in a component), and

N_t is collection length (i.e., number of terms in a test collection).

c_k are the k coefficients obtained through the regression analysis.

If stopwords are removed from indexing, then $q l$, $c l$, and N_t are the query length, document length, and collection length, respectively. If the query terms are re-weighted (in feedback, for example), then $q t f_i$ is no longer the original term frequency, but the new weight, and $q l$ is the sum of the new weight values for the query terms. Note that, unlike the document and collection lengths, query length is the “optimized” relative frequency without first taking the log over the matching terms.

The coefficients were determined by fitting the logistic regression model specified in $\log O(R|C, Q)$ to TREC training data using a statistical software package. The coefficients, c_k , used for our official runs are the same as those described by Chen[1]. These were: $c_0 = -3.51$, $c_1 = 37.4$, $c_2 = 0.330$, $c_3 = 0.1937$ and $c_4 = 0.0929$. Further details on the TREC2 version of the Logistic Regression algorithm may be found in Cooper et al. [4].

2.2 Blind Relevance Feedback

In addition to the direct retrieval of documents using the TREC2 logistic regression algorithm described above, we have implemented a form of “blind relevance feedback” as a supplement to the basic algorithm. The algorithm used for blind feedback was originally developed and described by Chen [2]. Blind relevance feedback has become established in the information retrieval community due to its consistent improvement of initial search results as seen in TREC, CLEF and other retrieval evaluations [7]. The blind feedback algorithm is based on the probabilistic term relevance weighting formula developed by Robertson and Sparck Jones [10].

Blind relevance feedback is typically performed in two stages. First, an initial search using the original topic statement is performed, after which a number of terms are selected from some number of the top-ranked documents (which are presumed to be relevant). The selected terms are then weighted and then merged with the initial query to formulate a new query. Finally the reweighted and expanded query is submitted against the same collection to produce a final ranked

list of documents. Obviously there are important choices to be made regarding the number of top-ranked documents to consider, and the number of terms to extract from those documents. For ImageCLEF this year, having no prior data to guide us, we chose to use the top 10 terms from 10 top-ranked documents. The terms were chosen by extracting the document vectors for each of the 10 and computing the Robertson and Sparck Jones term relevance weight for each document. This weight is based on a contingency table where the counts of 4 different conditions for combinations of (assumed) relevance and whether or not the term is, or is not in a document. Table 1 shows this contingency table.

Table 1: Contingency table for term relevance weighting

	Relevant	Not Relevant	
In doc	R_t	$N_t - R_t$	N_t
Not in doc	$R - R_t$	$N - N_t - R + R_t$	$N - N_t$
	R	$N - R$	N

The relevance weight is calculated using the assumption that the first 10 documents are relevant and all others are not. For each term in these documents the following weight is calculated:

$$w_t = \log \frac{\frac{R_t}{R - R_t}}{\frac{N_t - R_t}{N - N_t - R + R_t}} \quad (4)$$

The 10 terms (including those that appeared in the original query) with the highest w_t are selected and added to the original query terms. For the terms not in the original query, the new “term frequency” (qtf_i in main LR equation above) is set to 0.5. Terms that were in the original query, but are not in the top 10 terms are left with their original qtf_i . For terms in the top 10 and in the original query the new qtf_i is set to 1.5 times the original qtf_i for the query. The new query is then processed using the same LR algorithm as shown in Equation 4 and the ranked results returned as the response for that topic.

2.3 Okapi BM-25 Algorithm

The version of the Okapi BM-25 algorithm used in these experiments is based on the description of the algorithm in Robertson [12], and in TREC notebook proceedings [11]. As with the LR algorithm, we have adapted the Okapi BM-25 algorithm to deal with document components :

$$\sum_{j=1}^{|Q_c|} w^{(1)} \frac{(k_1 + 1)tf_j}{K + tf_j} \frac{(k_3 + 1)qtf_j}{k_3 + qtf_j} \quad (5)$$

Where (in addition to the variables already defined):

K is $k_1((1 - b) + b \cdot dl/avcl)$

k_1 , b and k_3 are parameters (1.5, 0.45 and 500, respectively, were used),

$avcl$ is the average component length measured in bytes

$w^{(1)}$ is the Robertson-Sparck Jones weight:

$$w^{(1)} = \log \frac{\left(\frac{r+0.5}{R-r+0.5}\right)}{\left(\frac{n_{t_j}-r+0.5}{N-n_{t_j}-R-r+0.5}\right)}$$

r is the number of relevant components of a given type that contain a given term,

R is the total number of relevant components of a given type for the query.

Our current implementation uses only the *a priori* version (i.e., without relevance information) of the Robertson-Sparck Jones weights, and therefore the $w^{(1)}$ value is effectively just an IDF weighting. The results of searches using our implementation of Okapi BM-25 and the LR algorithm seemed sufficiently different to offer the kind of conditions where data fusion has been shown to be most effective [9], and our overlap analysis of results for each algorithm (described in the evaluation and discussion section) has confirmed this difference and the fit to the conditions for effective fusion of results.

The system used supports searches combining probabilistic and (strict) Boolean elements, as well as operators to support various merging operations for both types of intermediate result sets. However, in GeoCLEF we did not use this capability.

2.4 Fusion Operators

The Cheshire II system used in this evaluation provides a number of operators to combine the intermediate results of a search from different components or indexes. With these operators we have available an entire spectrum of combination methods ranging from strict Boolean operations to fuzzy Boolean and normalized score combinations for probabilistic and Boolean results. These operators are the means available for performing fusion operations between the results for different retrieval algorithms and the search results from different components of a document. We will only describe one of these operators here, because it was the only type used in the GEOCLEF runs reported in this paper.

The MERGE PIVOT operator is used primarily to adjust the probability of relevance for one search result based on matching elements in another search result. It was developed primarily to adjust the probabilities of a search result consisting of sub-elements of a document (such as titles or paragraphs) based on the probability obtained for the same search over the entire document. It is basically a weighted combination of the probabilities based on a “DocPivot” fraction, such that:

$$P_n = DocPivot * P_d + (1 - DocPivot) * P_s \quad (6)$$

where P_d represents the document-level probability of relevance, P_s represents the subelement probability, and P_n represents the resulting new probability estimate.

For all of our fusion experiments this year, the P_s was the estimated probability of relevance for a document obtained by a TREC2 with blind feedback search using the topic title and description (as described above) normalized using MINMAX normalization, and P_d was an OKAPI BM-25 search using the topic title, description, and narrative, also normalized to 0-1 range using MINMAX normalization. The “DocPivot” value used for all of the runs submitted was 0.29.

Note that this is not the first time we have used some fusion approaches in GeoCLEF. In the first GeoCLEF (2005) we also employed fusion approaches in some of our runs, but these did not use the TREC2 with blind feedback algorithm[8], which appears to make an important difference.

3 Approaches for GeoCLEF

In this section we describe the specific approaches taken for our submitted runs for the GeoCLEF tasks. First we describe the indexing and term extraction methods used, and then the search features we used for the submitted runs.

3.1 Indexing and Term Extraction

The Cheshire II system uses the XML structure of the documents to extract selected portions for indexing and retrieval. Any combination of tags can be used to define the index contents.

Table 2: Cheshire II Indexes for GeoCLEF 2006

Name	Description	Content Tags	Used
docno	Document ID	DOCNO	no
pauthor	Author Names	BYLINE, AU	no
headline	Article Title	HEADLINE, TITLE, LEAD, LD, TI	no
topic	Content Words	HEADLINE, TITLE, TI, LEAD BYLINE, TEXT, LD, TX	yes
date	Date of Publication	DATE, WEEK	no
geotext	Validated place names	TEXT, LD, TX	no
geopoint	Validated coordinates for place names	TEXT, LD, TX	no
geobox	Validated bounding boxes for place names	TEXT, LD, TX	no

Table 2 lists the indexes created by the Cheshire II system for the GeoCLEF database and the document elements from which the contents of those indexes were extracted. The “Used” column in Table 2 indicates whether or not a particular index was used in the submitted GeoCLEF runs.

The georeferencing indexing subsystem of Cheshire II was used for the geotext, geopoint, and geobox indexes. This subsystem is intended to extract proper nouns from the text being indexed and then attempts to match them in a digital gazetteer. For GeoCLEF we used a gazetteer derived from the World Gazetteer (<http://www.world-gazetteer.com>) with 224698 entries in both English and German. The indexing subsystem provides three different index types: verified place names (an index of names which matched the gazetteer), point coordinates (latitude and longitude coordinates of the verified place name) and bounding box coordinates (bounding boxes for the matched places from the gazetteer). All three types were created, but we ended up not using any of the geographic indexes in this year’s submissions. Because we do not use complete NLP parsing techniques, the system is unable to distinguish between proper nouns for places from those for individuals. This leads to errors in geographic assignment where, for example, articles about Irving Berlin might be tagged as referring to the city.

Because there was no explicit tagging of location-related terms in the collections used for GeoCLEF, we applied the above approach to the “TEXT”, “LD”, and “TX” elements of the records of the various collections. The part of news articles normally called the “dateline” indicating the location of the news story was not separately tagged in any of the GeoCLEF collections, but often appeared as the first part of the text for the story.

Geographic indexes were not created for the Portuguese sub-collection due to the lack of a suitable gazetteer. We plan for later work to substitute the “GeoNames” database which is much more detailed and provides a more complete geographical hierarchy in its records, along with alternate names in multiple languages.

For all indexing we used language-specific stoplists to exclude function words and very common words from the indexing and searching. The German language runs *did not* use decompounding in the indexing and querying processes to generate simple word forms from compounds. Although we tried again this year to make this work within the Cheshire system, we again lacked the time needed to implement it correctly.

The Snowball stemmer was used by Cheshire for language-specific stemming.

3.2 Search Processing

Searching the GeoCLEF collection using the Cheshire II system involved using TCL scripts to parse the topics and submit the title and description or the title, description, and narrative from the topics. For monolingual search tasks we used the topics in the appropriate language (English, German, and Portuguese), for bilingual tasks the topics were translated from the source language

Table 3: Submitted GeoCLEF Runs

Run Name	Description	Type	MAP
BERKGCMODETD	Monolingual German	TD auto	0.2295 *
BERKGCMODETNDN	Monolingual German	TDN auto	0.2050
BERKMODETDNPIV	Monolingual German	TDN auto fusion	0.2292
BERKGCMOENTD	Monolingual English	TD auto	0.2652
BERKGCMOENTDN	Monolingual English	TDN auto	0.2001
BERKMOENTDNPIV	Monolingual English	TDN auto fusion	0.2685 *
BERKGCMOPTTD	Monolingual Portuguese	TD auto	0.2170
BERKGCMOPTTDN	Monolingual Portuguese	TDN auto	0.1741
BERKMOPTTDNPIV	Monolingual Portuguese	TDN auto fusion	0.2310 *
BERKGCBIENDETD	Bilingual English \Rightarrow German	TD auto	0.2150
BERKGCBIENDETDN	Bilingual English \Rightarrow German	TDN auto	0.1682
BERKBIENDETDNPIV	Bilingual English \Rightarrow German	TDN auto fusion	0.2251 *
BERKGCBIPTDDET	Bilingual Portuguese \Rightarrow German	TD auto	0.1950
BERKGCBIPTDETDN	Bilingual Portuguese \Rightarrow German	TDN auto	0.1108
BERKBIPTDETDNPIV	Bilingual Portuguese \Rightarrow German	TDN auto fusion	0.1912
BERKGCBIDEENTD	Bilingual German \Rightarrow English	TD auto	0.2274
BERKGCBIDEENTDN	Bilingual German \Rightarrow English	TDN auto	0.1894
BERKBIDEENTDNPIV	Bilingual German \Rightarrow English	TDN auto fusion	0.2304 *
BERKGCBIPTESTD	Bilingual Portuguese \Rightarrow English	TD auto	0.1886
BERKGCBIPTESTDN	Bilingual Portuguese \Rightarrow English	TDN auto	0.1540
BERKBIPTENTDNPIV	Bilingual Portuguese \Rightarrow English	TDN auto fusion	0.2101
BERKGCBIDEPPTD	Bilingual German \Rightarrow Portuguese	TD auto	0.1346
BERKGCBIDEPPTDN	Bilingual German \Rightarrow Portuguese	TDN auto	0.1260
BERKBIDEPPTDNPIV	Bilingual German \Rightarrow Portuguese	TDN auto fusion	0.1488
BERKGCBBIENPTTD	Bilingual English \Rightarrow Portuguese	TD auto	0.1913
BERKGCBBIENPTTDN	Bilingual English \Rightarrow Portuguese	TDN auto	0.1762
BERKBIENPTTDNPIV	Bilingual English \Rightarrow Portuguese	TDN auto fusion	0.2074 *

to the target language using the LEC Power Translator PC-based machine translation system. Table 3 shows the runs submitted and the characteristics of each run, including which task it was submitted for, and the topic elements used in searching (this are indicated by using the “T” for the “title” element, “D” for the “description” element and “N” for the “narrative” element. The topic elements used were combined into a single probabilistic query. For those runs including the term “fusion” in the “Type” column, both TREC2 with blind feedback algorithm results (using only topic title and description) and Okapi BM-25 results (using title description and narrative) were combined using the MERGE_PIVOT fusion operator described above with a pivot value of 0.29.

4 Results for Submitted Runs

The summary results (as Mean Average Precision) for the submitted monolingual and bilingual runs for both English, German and Portuguese are also shown in Table 3, the Recall-Precision curves for these runs are also shown in Figures 1 (for monolingual) and 2 (for bilingual). In Figures 1 and 2 the names for the monolingual runs represent the search algorithms and topic elements used for that particular target language, which can easily be compared with full names and descriptions in Table 3 (since each combination has only a single run). The names for Bilingual runs indicate the languages and topic elements used (except for PIV which is the fusion runs that

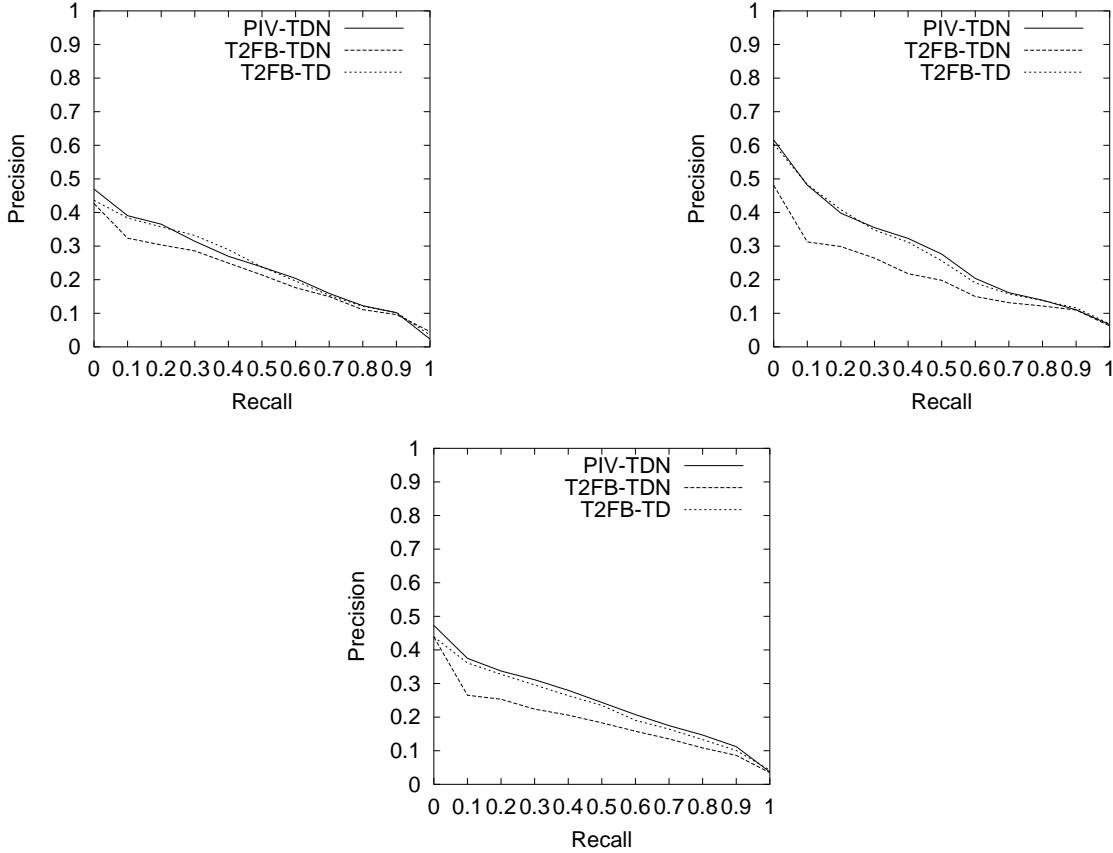


Figure 1: Monolingual Runs – German (top left), English (top right), and Portuguese (lower)

use title, description and narrative).

Table 3 indicates the runs that had the highest overall MAP for the task and language by asterisks next to the run name.

Once again we found some rather interesting results among the official runs. For example, it seems clear that using topic title and description alone is a much better approach with our algorithms than using title description and narrative. However, in most cases the fusion approach either exceeds, or is very close to the performance of the TREC2 with blind feedback search with title and description due to “supporting evidence” for relevance from the Okapi BM-25 algorithm.

Last year the “weak man” in our runs was German, both monolingual and bilingual. At the time we were not sure if that might just be due to decompounding issues. However, it turned out that the *real* cause of last year’s poor performance was that in indexing the database we had completely forgotten all of the SDA German collection, and thus the database we used in 2007 had only about half of the German documents. Needless to say, this was remedied in our submitted runs for this year.

5 Analysis and Conclusions

Because we used a similar processing approach (except for fusion operations) this year as we used for some of our runs submitted for GeoCLEF 2006 and 2007, we build Table 4 examine the differences. Overall, we did see some distinct improvements in the text-based approaches over those used in 2006 and 2007. In Table 4 the MAP for our best runs from 2006, 2007, and 2008 for

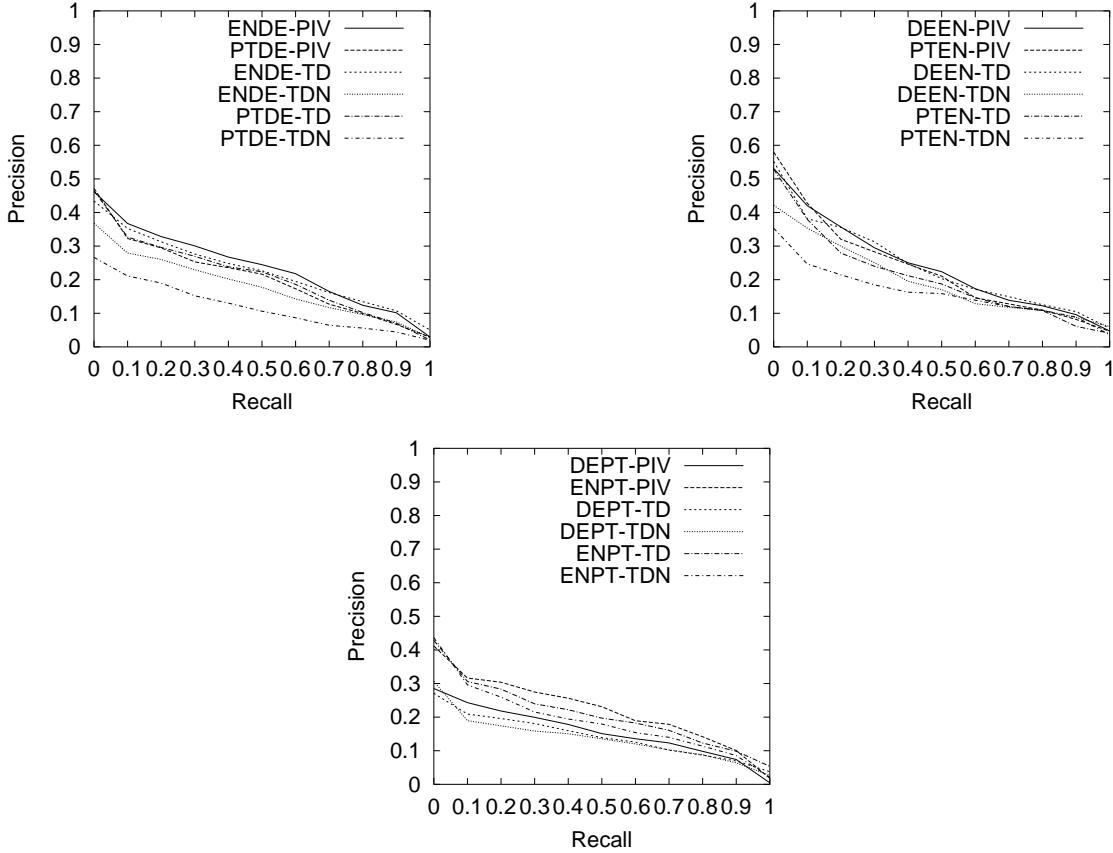


Figure 2: Bilingual Runs – To German (top left), To English (top right) and to Portuguese (lower)

various tasks are shown, along with the percentage difference between each pair of years.

Based on the summary data across participants for GeoCLEF available on the DIRECT system, it is apparent that the text-based and fusion approaches that we used this year are quite effective relative to some other approaches. In all of the six main GeoCLEF tasks there was a cheshire run in the top five participants listed. In Monolingual Portuguese and each Bilingual task (German, English, and Portuguese) one of the cheshire runs was ranked highest in MAP over all participants.

We need to do further testing with the fusion approaches, since the results are sensitive to the “pivot” value used. In addition, since the data show the relative importance of using only the title and description element instead of title, description, and narrative, we need to see if using using only title and description only in both parts of the fusion query further improves, or degrades performance.

The challenge for next year is to reintroduce actual geographic elements to the mix to see if, for example, automatic expansion of toponyms in the topic texts will enhance or degrade performance over the purely textual approach. Since this was done explicitly in many of the topic narratives we, and use of narrative proved counter-productive we suspect that such expansion may be more a source of noise instead of fostering improved results. In previous years it appeared that implicit or explicit toponym inclusion in queries led to better performance when compared to using titles and descriptions alone in retrieval. But given the results this time, some doubt has been cast over that assumption, at least for the algorithms that we have been using.

Although we did not do any explicit geographic processing other than in indexing for this year, we plan to do so in the future, because we still believe that use of geographical knowledge and evidence in topics and documents *should* improve performance over purely text-based methods.

Table 4: Comparison of Berkeley’s best 2006, 2007 and 2008 runs for English, German and Portuguese

TASK	MAP 2006	MAP 2007	MAP 2008	Diff. '06-'07	Diff. '07-'08	Diff. '06-'08
Monolingual English	0.250	0.264	0.268	5.303	1.493	6.716
Monolingual German	0.215	0.139	0.230	-54.676	39.565	6.522
Monolingual Portuguese	0.162	0.174	0.231	6.897	24.675	29.870
Bilingual English⇒German	0.156	0.090	0.225	-73.333	60.00	30.667
Bilingual English⇒Portuguese	0.126	0.201	0.207	37.313	2.899	39.130

However, given the results reported above, this belief is still unsupported in our experiments.

References

- [1] Aitao Chen. Multilingual information retrieval using english and chinese queries. In Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck, editors, *Evaluation of Cross-Language Information Retrieval Systems: Second Workshop of the Cross-Language Evaluation Forum, CLEF-2001, Darmstadt, Germany, September 2001*, pages 44–58. Springer Computer Scinece Series LNCS 2406, 2002.
- [2] Aitao Chen. *Cross-Language Retrieval Experiments at CLEF 2002*, pages 28–48. Springer (LNCS #2785), 2003.
- [3] Aitao Chen and Fredric C. Gey. Multilingual information retrieval using machine translation, relevance feedback and decompounding. *Information Retrieval*, 7:149–182, 2004.
- [4] W. S. Cooper, A. Chen, and F. C. Gey. Full Text Retrieval based on Probabilistic Equations with Coefficients fitted by Logistic Regression. In *Text RETrieval Conference (TREC-2)*, pages 57–66, 1994.
- [5] William S. Cooper, Fredric C. Gey, and Daniel P. Dabney. Probabilistic retrieval based on staged logistic regression. In *15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen, Denmark, June 21-24*, pages 198–210, New York, 1992. ACM.
- [6] Ray R. Larson. Geographic information retrieval and spatial browsing. In Linda Smith and Myke Gluck, editors, *GIS and Libraries: Patrons, Maps and Spatial Information*, pages 81–124. University of Illinois at Urbana-Champaign, GSLIS, Urbana-Champaign, 1996.
- [7] Ray R. Larson. Probabilistic retrieval, component fusion and blind feedback for XML retrieval. In *INEX 2005*, pages 225–239. Springer (Lecture Notes in Computer Science, LNCS 3977), 2006.
- [8] Ray R. Larson, Fredric C. Gey, and Vivien Petras. Berkeley at GeoCLEF: Logistic regression and fusion for geographic information retrieval. In *Cross-Language Evaluation Forum: CLEF 2005*, pages 963–976. Springer (Lecture Notes in Computer Science LNCS 4022), 2006.
- [9] Joon Ho Lee. Analyses of multiple evidence combination. In *SIGIR '97: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 27-31, 1997, Philadelphia*, pages 267–276. ACM, 1997.
- [10] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, pages 129–146, May–June 1976.

- [11] Stephen E. Robertson, Stephen Walker, and Micheline M. Hancock-Beaulieu. OKAPI at TREC-7: ad hoc, filtering, vlc and interactive track. In *Text Retrieval Conference (TREC-7), Nov. 9-1 1998 (Notebook)*, pages 152–164, 1998.
- [12] Stephen E. Robertson and Steven Walker. On relevance weights with little relevance information. In *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 16–24. ACM Press, 1997.