

Exeter at CLEF 2003: Cross-Language Spoken Document Retrieval Experiments

Gareth J. F. Jones* and Adenike Lam-Adesina
Department of Computer Science,
University of Exeter, EX4 4QF, U.K.
{G.J.F.Jones,A.M.Lam-Adesina}@exeter.ac.uk

Abstract. Cross-Language Spoken Document Retrieval (CLSDR) combines both the complexities of retrieval from collections characterized by speech transcription errors and language translation issues between search requests and documents. Thus achieving effective retrieval in this domain is potentially very challenging. For the CLEF 2003 SDR task we adopted a standard query translation strategy using commercial machine translation tools and explored pseudo-relevance feedback using a small contemporaneous collection and a much larger text collection from a different time period.

1 Introduction

Both Cross-Language Information Retrieval (CLIR) and Spoken Document Retrieval (SDR) are affected by limitations in language processing technologies. In the case of the former this relates to translation between the languages of queries and documents, and in the latter to the difficulties encountered in transcription of spoken data. These issues are analyzed in more detail in [1]. Spoken Document Retrieval (CLSDR) combines the difficulties of both CLIR and SDR. Thus retrieval in this domain is very challenging.

For the CLEF 2003 CLSDR task we adopted a query translation strategy and investigated the use of a small contemporaneous text collection and a large text document set from a different period to the test collection as pilot collections to augment the spoken document test set. All query statements were translated from the source language into English using two machine translation tools: Systran Version:3.0 (Sys MT) and Globalink Power Translation Pro Version 6.4 (Pro MT) machine translation (MT) systems. The task is based on the TREC 9 SDR task. A detailed description of the task is found in [2].

The remainder of this paper summarizes our retrieval system and gives results and analysis of our experimental results.

2 System Setup

The basis of the experimental system was the same as that used for our submissions to the monolingual, bilingual and multilingual tasks for CLEF 2003. The system

* now at School of Computing, Dublin City University, Ireland
email: Gareth.Jones@computing.dcu.ie

combines Okapi BM25 term weighting with pseudo relevance feedback (PRF), and standard procedures of stop word removal and Porter stemming, full details are given in [3]. The parameters of the PRF system were set identically to those for the text retrieval system given in [3]. The Okapi parameters $K1$ and b were optimized for the SDR test collection.

3 Merged collections

In our experiments for the CLSDR pilot track held at CLEF 2002 we experimented with the combination of the test collection with a small contemporaneous text document collection for term weight estimation [4]. This method aims to improve retrieval performance for the test set by better estimation of term weights. Our results for CLEF 2002 indicated that the method can give improvements in retrieval performance even when using only a small number of additional documents. Results for ITC-irst however showed that large improvements can be realized if a much larger number of contemporaneous text documents is used [5]. However, this large collection of truly contemporaneous documents was not available to us. This led us to investigate the use of an alternative large text document collection. In this case we used the document set from the TREC-8 and TREC-9 ad hoc retrieval tasks. This consists of around 500,000 text documents taken from 1994, some 4 years earlier than the SDR data set which is from February to June 1998. There is likely to be considerable vocabulary mismatch between these document collections, e.g. names, places, events, and the aim of this experiment was to find out if a pilot collection of this type could still provide improvement in retrieval performance.

In addition, we again used the two small collections of truly contemporaneous text documents. These sources are taken from New York Times Newswire Service (excluding non-NYT sources) and Associated Press Worldstream Service (English content only), totaling about 20,000 news stories, and are taken from exactly the same period as the spoken document test collection.

4 Experimental Results

This section describes our results for the CLEF 2003 CLSDR task. We report baseline and feedback results for five topic languages: English, French, German, Italian and Spanish. Our results include runs for topic translations using both Sys MT and Pro MT systems. Results for each condition are shown in terms of average precision and the total number of relevant documents retrieved for the complete query set.

In the following tables of results the following labeling conventions are adopted for the selection of topic expansion terms and $cfw(i)$ of the test collection:

TCow(i): topic expansion using only the test collection.

CCow(i): topic expansion using the combined spoken and small text collections.

PCow(i): topic expansion using the TREC document pilot collection.

TCcfw(i): $cfw(i)$ values taken from the test collection in the final retrieval run.

CCcfw(i): $cfw(i)$ values taken from the combined spoken and small text collections in the final retrieval run.

PCcfw(i): $cfw(i)$ values taken from the TREC document pilot collection in the final retrieval run.

Initial results are presented for the following methods:

1. Baseline run without feedback
2. Feedback runs using expanded query from the test collection
3. Feedback runs using queries expanded from the pilot collection and term weight estimated from the test collection. Initial query terms are upweighted by multiply by 1.5
4. Same as 3 but initial query terms are upweighted by 3.5

Table 1. Retrieval results for topic translation using Sys MT

| Sys MT | | English | French | German | Italian | Spanish |
|-----------------------------|---------|---------|--------|--------|---------|---------|
| 1. Baseline | Av.P | 0.311 | 0.227 | 0.203 | 0.231 | 0.250 |
| | Rel-Ret | 1587 | 1424 | 1369 | 1531 | 1548 |
| 2. TCow(i), TCcfw(i) | Av.P | 0.382 | 0.281 | 0.270 | 0.279 | 0.292 |
| | % chg. | +22.8% | +23.8% | +33.0% | +20.7% | +16.8% |
| | Rel-Ret | 1795 | 1558 | 1498 | 1638 | 1641 |
| 3. CCow(i), TCcfw(i),1.5 | Av.P | 0.364 | 0.283 | 0.274 | 0.299 | 0.304 |
| | % chg. | +17.0% | +24.7% | +34.9% | +29.4% | +21.6% |
| | Rel-Ret | 1824 | 1618 | 1541 | 1684 | 1720 |
| 4. CCow(i), TCcfw(i),3.5 | Av.p | 0.371 | 0.276 | 0.268 | 0.296 | 0.307 |
| | % chg. | +19.3% | +21.6% | +32.0% | +28.1% | +22.8% |
| | Rel-Ret | 1789 | 1577 | 1524 | 1653 | 1707 |

Table 2. Retrieval results for topic translation using Pro MT

| Pro MT | | English | French | German | Italian | Spanish |
|-----------------------------|---------|---------|--------|--------|---------|---------|
| 1. Baseline | Av.P | 0.311 | 0.189 | 0.188 | 0.234 | 0.235 |
| | Rel-Ret | 1587 | 1356 | 1307 | 1503 | 1564 |
| 2. TCow(i), TCcfw(i) | Av.P | 0.382 | 0.244 | 0.245 | 0.288 | 0.298 |
| | %chg. | +22.8% | +29.1% | +30.3% | +23.1% | +26.8% |
| | Rel-Ret | 1795 | 1533 | 1442 | 1570 | 1715 |
| 3. CCow(i), TCcfw(i),1.5 | Av.P | 0.364 | 0.262 | 0.242 | 0.301 | 0.315 |
| | % chg. | +17.0% | +38.6% | +28.7% | +28.6% | +34.0% |
| | Rel-Ret | 1824 | 1589 | 1431 | 1624 | 1710 |
| 4. CCow(i), TCcfw(i),3.5 | Av.P | 0.371 | 0.256 | 0.229 | 0.293 | 0.308 |
| | % chg. | +19.3% | +35.4 | +21.8% | +25.2% | +31.1% |
| | Rel-Ret | 1789 | 1574 | 1420 | 1602 | 1682 |

Results for our CLSDR runs are shown in Tables 1 and 2. It can be seen that as expected the monolingual English result is the best in all cases with respect to both average precision and number of relevant documents retrieved. CLSDR performance is comparable for the French, Italian and Spanish topic statements with lower results for the German topics with Sys MT. This result is a little surprising for Systran French topic translation which has previously been shown to be more effective than other

topic translations in our CLEF bilingual text retrieval experiments [6]. PRF using only the test collection is observed to be effective for topic expansion in all cases. Results for query expansion using the merged document collection are more mixed. In the case of Italian and Spanish topics this approach clearly outperforms test collection only query expansion. However, there is little difference between the results for these methods when using French and German topics.

We carried out further experiments using the TREC-7 and TREC-8 ad hoc document collection as a pilot searching collection. Results are presented for the following methods:

5. Topics are expanded using the TREC document pilot collection and then further expanded using the combined collection from the earlier experiments, final retrieval run with cfw(i) values from the test collection.
6. As 5 with the final retrieval run on the test collection using cfw(i) from the pilot collection.

Table 3. Retrieval results for topic translation with Sys MT and pilot searching

| Sys MT | | English | French | German | Italian | Spanish |
|-----------------------------|---------|---------|--------|--------|---------|---------|
| 5. PC->Ccow(i), CCcfw(i) | Av.P | 0.341 | 0.257 | 0.248 | 0.282 | 0.260 |
| | % chg. | +9.6% | +13.2% | +22.2% | +22.1% | +4.0% |
| | Rel-Ret | 1667 | 1564 | 1436 | 1635 | 1529 |
| 6. PC->Ccow(i), PCcfw(i) | Av.P | 0.338 | 0.255 | 0.258 | 0.284 | 0.263 |
| | % chg. | +8.7% | +12.3% | +27.1% | +22.9% | +5.2% |
| | Rel-Ret | 1683 | 1584 | 1429 | 1649 | 1591 |

Table 4. Retrieval results for topic translation with Pro MT and pilot searching

| Pro MT | | English | French | German | Italian | Spanish |
|-----------------------------|---------|---------|--------|--------|---------|---------|
| 5. PC->CCow(i), CCcfw(i) | Av.P | 0.341 | 0.252 | 0.232 | 0.274 | 0.259 |
| | % chg. | +9.6% | +33.3% | +23.4% | +17.1% | +10.2% |
| | Rel-Ret | 1667 | 1520 | 1393 | 1609 | 1621 |
| 6. PC->CCow(i), PCcfw(i) | Av.P | 0.338 | 0.258 | 0.235 | 0.270 | 0.258 |
| | % chg. | +8.7% | +36.5% | +25.0% | +15.4% | +9.8% |
| | Rel-Ret | 1683 | 1531 | 1378 | 1621 | 1629 |

From the results in Tables 5 and 6 it can be seen that expansion using the TREC document is less effective than using either test collection only expansion or the combination of the test collection with the small text collection from the same time period.

This result is not altogether surprising since the TREC text document sets is taken from a period some 4 years earlier than the TREC SDR documents. This result indicates that while using large text document sets can be useful in CLSDR as illustrated in [5], these documents must have an appropriate match, presumably relating to vocabulary and topic coverage, to the spoken document collection.

5 Conclusions and Further Work

The results for the CLEF 2003 CLSDR task reported in this paper establish baseline performance figures against which our exploration of techniques for CLSDR can be measured. The experiments reported here show that PRF is effective for this task, as would be expected since it is generally a useful technique for text CLIR and SDR. The effectiveness of large additional text collections for parameter estimation for query expansion in CLSDR has been shown to be dependent on the match of the time periods covered in the collections.

References

1. G.J.F. Jones and M.Federico, CLEF 2002 Cross-Language Spoken Document Retrieval Pilot Track Report, In *Proceedings of the CLEF 2002: Workshop on Cross-Language Information Retrieval and Evaluation*, Rome, September 2002.
2. M.Federico and G.J.F.Jones, CLEF 2003 Cross-Language Spoken Document Retrieval Pilot Track Report, In *Proceedings of the CLEF 2003: Workshop on Cross-Language Information Retrieval and Evaluation*, Trondheim, August 2003.
3. A.M.Lam-Adesina and G.J.F.Jones, Exeter at CLEF 2003: Experiments with Machine Translation for Monolingual, Bilingual and Multilingual Retrieval, In *Proceedings of the CLEF 2003 Workshop on Cross-Language Information Retrieval and Evaluation*, Trondheim, August 2003.
4. G.J.F. Jones and A.M. Lam-Adesina, Exeter at CLEF 2002: Cross-Language Spoken Document Retrieval Experiments, In *Proceedings of the CLEF 2002: Workshop on Cross-Language Information Retrieval and Evaluation*, Rome, September 2002.
5. N.Bertoldi and M.Federico, Cross-Language Spoken Document Retrieval on the TREC SDR Collection. In *Proceedings of the CLEF 2002: Workshop on Cross-Language Information Retrieval and Evaluation*, Rome, September 2002.
6. G.J.F. Jones and A.M. Lam-Adesina. Exeter at CLEF 2001: Experiments with Machine Translation for Bilingual Retrieval. In *Proceedings of the CLEF 2001: Workshop on Cross-Language Information Retrieval and Evaluation*, pages 59-77, Darmstadt, Germany, 2001.