# The University of Amsterdam at CLEF 2004

Sisay Fissaha Adafre        Willem Robert van Hage        Jaap Kamps*
Gustavo Lacerda de Melo        Maarten de Rijke

Informatics Institute, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
E-mail: {sfissaha,wrvhage,kamps,glacerda,mdr}@science.uva.nl

### Abstract

We describe the official runs of our team for the CLEF 2004 ad hoc tasks. We took part in the monolingual task (for Finnish, French, Portuguese, and Russian), in the bilingual task (for Amharic to English, and English to Portuguese), and, finally, in the multilingual task.

## 1 Introduction

In the CLEF 2004 evaluation exercise we participated in all three ad hoc retrieval tasks. We took part in the *monolingual* tasks for four non-English languages, Finnish, French, Portuguese, and Russian. The Portuguese language was new for CLEF 2004. Our participation in the monolingual task was a further continuation of our earlier efforts to monolingual retrieval [11, 5, 6]. Our first aim was to continue our experiments with a number of language-dependent techniques, in particular stemming algorithms for all European languages [14], and compound splitting for the compound rich Finnish language. A second aim was to continue our experiments with language-independent techniques, in particular the use of character n-grams, where we may also index leading and ending character sequences, and retain the original words. Our third aim was to experiment with combinations of runs.

We took part in the *bilingual* task, this year focusing on Amharic into English, and on English to Portuguese. Our bilingual runs were motivated by the following aims. Our first aim was to experiment with a language for which resources are few and far between, Amharic, and to see how far we could get by combining the scarcely available resources. Our second aim was to experiment with the relative effectiveness of a number of translation resources: machine translation [16] versus a parallel corpus [7], and query translation versus collection translation. Our third aim was to evaluate the effectiveness of our monolingual retrieval approaches for imperfectly translated queries, shedding light on the robustness of these approaches.

Finally, we continued our participation for the *multilingual* task, where we experimented with straightforward ways of query translation, using machine translation whenever available, and a translation dictionary otherwise. We also experimented with combination methods using runs made on varying types of indexes.

In Section 2 we describe the FlexIR system as well as the approaches used for each of the tasks in which we participated. Section 3 describes our official retrieval runs for CLEF 2004. In Section 4 we discuss the results we have obtained. Finally, in Section 5, we offer some conclusions regarding our document retrieval efforts.

## 2 System Description

### 2.1 Retrieval Approach

All retrieval runs used FlexIR, an information retrieval system developed at the University of Amsterdam. FlexIR supports many types of preprocessing, scoring, indexing, and retrieval tools. It also supports several retrieval models, including the standard vector space model, and language models. Our default retrieval model is a vector space model using the Lnu.ltc weighting scheme [1] to compute the similarity between a query and a document. For the experiments on which we report in this note, we fixed *slope* at 0.2; the pivot was set to the average number of unique words per document. We also experimented with language models [3]. Here, we used a uniform query term importance weight of 0.15.

---

*Currently at Archives and Information Studies, Faculty of Humanities, University of Amsterdam.

**Text normalization.** We do some limited text normalization by removing punctuation, applying case-folding, and mapping diacritics to the unmarked characters. The Cyrillic characters used in Russian can appear in a variety of font encodings. The collection and topics are encoded using the UTF-8 or Unicode character encoding. We converted the UTF-8 encoding into KOI8 (*Kod Obmena Informatsii*), a 1-byte per character encoding. We did all our processing, such as lower-casing, stopping, stemming, and n-gramming, on documents and queries in this KOI8 encoding. Finally, to ensure proper indexing of the documents using our standard architecture, we converted the resulting documents into the Latin alphabet using the Volapuk transliteration. We processed the Russian queries similar to the documents.

**Morphological Normalization.** We carried out extensive experiments with different forms of morphological normalizations for monolingual retrieval [4]. These include the following:

*Stemming* — For all languages we used a stemming algorithm to map word forms to their underlying stems. Stemming is a language-dependent approach to morphological normalization. We used the family of Snowball stemming algorithms, available for all the languages of the CLEF collections. Snowball is a small string processing language designed for creating stemming algorithms for use in information retrieval [14].

*Decompounding* — For the compound-rich Finnish language, we also apply a decompounding algorithm. We treat all the words occurring in the Finnish collection as potential base words for decompounding, and also use their associated collection frequencies. We ignore words of length less than four as potential compound parts, thus a compound must have at least length eight. As a safeguard against oversplitting, we only regard compound parts that have a higher collection frequency than the compound itself. We retain the original compound words, and add their parts to the documents; the queries are processed similarly.

*n-Gramming* — For all languages, we used character n-gramming to index all character-sequences of a given length that occur in a word. n-Gramming is a language-independent approach to morphological normalization. We used three different ways of forming n-grams of length 4. First, we index pure 4-grams. For example, the word `Information` will be indexed as 4-grams `info nfor form orma rmat mati atio tion`. Second, we index 4-grams with leading and ending 3-grams. For the example this will give `_inf info nfor form orma rmat mati atio tion ion_`. Third, we index 4-grams plus original words. For the example this gives `info nfor form orma rmat mati atio tion information`.

**Stopwords.** Both topics and documents were stopped using the stopword lists from the Snowball stemming algorithms [14]; for Finnish we used the Neuchâtel-stoplist [10]. Additionally, we removed topic specific phrases such as 'Find documents that discuss . . . ' from the queries. We did not use a stop stem or stop n-gram list, but we first used a stop *word* list, and then stemmed/n-grammed the topics and documents.

**Blind Feedback.** Blind feedback was applied to expand the original query with related terms. We experimented with different schemes and settings, depending on the various indexing methods and retrieval models used. For our Lnu.ltc runs term weights were recomputed by using the standard Rocchio method [13], where we considered the top 10 documents to be relevant and the bottom 500 documents to be non-relevant. We allowed at most 20 terms to be added to the original query.

**Combined Runs.** We combined various 'base' runs using either a weighted or unweighted combination methods. The weighted interpolation was produced as follows. First, we normalized the retrieval status values (RSVs), since different runs may have radically different RSVs. For each run we reranked these values in $[0, 1]$ using $RSV'_i = (RSV_i - min_i)/(max_i - min_i)$; this is the Min_Max_Norm considered in [8]. Next, we assigned new weights to the documents using a linear interpolation factor $\lambda$ representing the relative weight of a run: $RSV_{new} = \lambda \cdot RSV_1 + (1 - \lambda) \cdot RSV_2$. For $\lambda = 0.5$ this is similar to the simple (but effective) combSUM function used by Fox and Shaw [2] The interpolation factors $\lambda$ were loosely based on experiments on earlier CLEF data sets. When we combine more than two runs, we give all runs the same relative weight, effectively resulting in the familiar combSUM.

## 3   Runs

We submitted a total of 24 retrieval runs: 12 for the monolingual task, 7 for the bilingual task, and 5 for the multi-lingual task. Below we discuss these runs in some detail.

### 3.1   Monolingual Runs

All our monolingual runs used the title and description fields of the topics. We constructed five different indexes for each of the languages using Words, Stems, 4-Grams, 4-Grams+start/end, and 4-Grams+Words:

- *Words* — no morphological normalization is applied, although for Finnish *Split* indicates that words are decompounded.

- *Stems* — topic and document words are stemmed using the morphological tools described in Section 2. For Finnish, *Split+stem* indicates that compounds are split, where we stem the words and compound parts.

- *n-Grams* — both topic and document words are n-grammed, using the settings discussed in Section 2. We have three different indexes: *4-Grams*; *4-Grams+words* where also the words are retained; and *4-Grams+start/end* with beginning and ending 3-grams.

On all these indexes we made runs using the Lnu.ltc retrieval model; on the Words and on the Stems index we also made runs with a language model, resulting in 7 base runs for French, Portuguese, and Russian. In addition, for the compound rich Finnish language we also applied a decompounding algorithm [4], on words and on stems, from which we produced base runs with both the Lnu.ltc retrieval model and a language model, leading to a total of 11 base runs for Finnish.

All our official submissions were combinations of the base runs just described. For each of the four languages we constructed two combinations of stemmed and n-grammed base runs, as well as a "grand" combination of all base runs. Table 1 provides an overview of the runs that we submitted for the monolingual task. The third column in Table 1 indicates the type of run, and for two-way combinations the interpolation factor λ used is given in the fourth column.

| Run | Language | Type | Factor |
|---|---|---|---|
| UAmsC04FiFi4GiSb | FI | 4-Grams+words;Split+stem | 0.4 |
| UAmsC04FiFi4GiWd | FI | 4-Grams+start/end;Split | 0.4 |
| UAmsC04FiFiAll | FI | Grand combination of 11 runs | - |
| UAmsC04FrFr4GiSb | FR | 4-Grams+words;Stems | 0.6 |
| UAmsC04FrFr4GiWd | FR | 4-Grams+start/end;Words | 0.6 |
| UAmsC04FrFrAll | FR | Grand combination of 7 runs | - |
| UAmsC04PoPo4GiSb | PT | 4-Grams+words;Stems | 0.4 |
| UAmsC04PoPo4GiWd | PT | 4-Grams+start/end;Words | 0.4 |
| UAmsC04PoPoAll | PT | Grand combination of 7 runs | - |
| UAmsC04RuRu4GiSb | RU | 4-Grams+words;Stems | 0.5 |
| UAmsC04RuRu4GiWd | RU | 4-Grams+start/end;Words | 0.5 |
| UAmsC04RuRuAll | RU | Grand combination of 7 runs | - |

Table 1: Overview of the monolingual runs submitted. For combined runs column 3 gives the base runs that were combined, and column 4 gives the interpolation factor λ.

## 3.2 Bilingual Runs

For the bilingual task, we focused on Amharic to English, and English to Portugues. We submitted a total of 7 runs; all of them used the title and description fields of the topics. For our bilingual runs, we experimented with the WorldLingo machine translation [16] for translations into Portuguese, with a parallel corpus for translations into Portuguese, and with a variety of techniques for the Amharic topics, as we will now explain.

### 3.2.1 English to Portuguese

**Machine Translation.**  We used the WorldLingo machine translation [16] for translating the English topics into Portuguese. The translation is actually in Brazilian Portuguese, but the linguistic differences between Portuguese and Brazilian are fairly limited.

**Parallel Corpus.**  We used the sentence-aligned parallel corpus [7], based on the Official Journal of the European Union [15]. We built a Portuguese to English translation dictionary, based on a word alignment in the parallel corpus. Since the word order in English and Portuguese are not very different, we only considered potential alignments with words in the same position, or one or two positions off. We ranked potential translations with a score based on:

- *Cognate matching* — Rewarding similarity in word forms, by looking at the number of leading characters that agree in both languages.

- *Length matching* — Rewarding similarity in word lengths in both languages.

- *Frequency matching* — Rewarding similarity in word frequency in both languages.

To further aid the alignment, we constructed a list of 100 most frequent Portuguese words in the corpus, and manually translated these to English. The alignments of these highly frequent words were resolved before the word alignment phase. We built a Portuguese to English translation dictionary by choosing the most likely translation, where we only include words that score above a threshold. The length of the translation dictionary is 19,554 words. We use the translation dictionary resulting from the parallel corpus for two different purposes. Firstly, we translate the English topics into Portuguese. Secondly, we translate the Portuguese collection into English.

### 3.2.2 Amharic to English

Amharic, which belongs to the Semitic family of languages, is one of the most widely spoken languages in Ethiopia. In Amharic, word formation involves affixation, reduplication, Semitic stem interdigitation, among others. The most characteristic feature of Amharic morphology is root-pattern phenomena. This is especially true of Amharic verbs, which rely heavily on the arrangement of consonants and vowels in order to code different morphosyntactic properties (such as perfect, imperfect, jussive etc.). Consonants, which mostly carry the semantic core of the word, form the root of the verb. Consonants and vowel patterns together constitute the stems, and stems take different types of affixes (prefixes and suffixes) to form the fully inflected words; see [12].

For our bilingual Amharic to English runs, we attempted to show how the (minimal) available resources for Amharic can be used in (Amharic-English) bilingual information retrieval settings. Since English is used on the document side, it is interesting to see how the existing retrieval techniques can be optimized in order to make the best use of the output of the error-prone translation component.

**Resources and Query Translation.**  Our Amharic to English query translation is based mainly on dictionary look up. We used an Amharic-English bilingual dictionary which consists of 15,000 fully inflected words. Due to the morphological complexity of the language, we expected the dictionary to have limited coverage. In order to improve on the coverage, two further dictionaries, root-based and stem-based, were derived from the original dictionary. We also tried to augment the dictionary with a bilingual lexicon extracted from aligned Amharic-English Bible text. However, most of the words are old English words and are also found in the dictionary. The word dictionary also contains commonly used Amharic collocations. Multiword collocations were identified and marked in the topics. For this purpose, we used a list of multiword collocations extracted from an Amharic text corpus. The dictionaries were searched for a translation of Amharic words in the following order: word-dictionary, stem dictionary, root dictionary.

| Total no. of words | Word dictionary | Root dictionary | English spell checker |
|---|---|---|---|
| 1,893 | 813 | 178 | 57 |

Table 2: Coverage of the respective techniques over the words occurring in the Amharic topics.

Leaving aside the ungrammaticality of the output of the above translation, there are a number of problems. One is the problem of unknown words. The words may be Amharic words not included in the dictionary or foreign words. Some foreign words and their transliteration have the same spelling or are nearly identical. To take advantage of this fact, a word is checked using an English spellchecker (Aspell); if found, it is returned as a translation. In some cases, there may be typographical variations between the English word and its transliteration; to address this, the first word among the suggestions will be checked for string similarity. If it falls above some threshold, it is taken as translation. Other unknown words are simply passed over to the English translation. Another problem relates to the selection of the appropriate translation from among the possible translations found in the dictionary. In the absence of frequency information, which allows selecting the right translation, the most frequently used English word is selected as a translation of the corresponding Amharic word. This is achieved by querying the web. The coverage of the translation is 55%. The number of correct translations is still lower. Table 2 gives some idea of the performance of the translation strategy.

For both English and Portuguese we used a similar set of indexes as for the monolingual runs described earlier (Words, Stems, 4-Grams, 4-Grams+start/end, 4-Grams+words); for all of these, Lnu.ltc runs were produced, and for the Word and Stems indexes we also produced a language model run, leading to 7 base runs for the Amharic to English task. Additionally, for the English to Portuguese task we used three types of translation: query translation

using machine translation (WorldLingo), query translation using a parallel corpus (query EU), and collection translation using a parallel corpus (collection EU). This gave rise to a total of 21 base runs for the English to Portuguese task.

Table 3 provides an overview of the runs that we submitted for the bilingual task. The fourth column in Table 3 indicates the type of run.

| Run | Topics | Documents | Type | Factor |
|---|---|---|---|---|
| UAmsC04EnPo4GiSb | EN | PT | 4-Grams+words;Stems (collection EU) | 0.7 |
| UAmsC04EnPo4iSPC | EN | PT | 4-Grams+words;Stems (query EU) | 0.7 |
| UAmsC04EnPo4iSWL | EN | PT | 4-Grams+words;Stems (WorldLingo) | 0.7 |
| UAmsC04EnPoAll | EN | PT | Grand combination of 21 runs | - |
| UAmsC04AmEnWrd | AM | EN | Words | - |
| UAmsC04AmEn4GiSb | AM | EN | 4-Grams+words;Stems | 0.7 |
| UAmsC04AmEnAll | AM | EN | Grand combination of 7 runs | - |

Table 3: Overview of the bilingual runs submitted. For combined runs column 4 gives the base runs that were combined, and column 5 gives the interpolation factor $\lambda$.

## 3.3 Multilingual Runs

We submitted a total of 4 multilingual runs, all using the title and description of the English topic set. The multilingual runs were based on the following mono- and bilingual runs:

- *English to English* – This is just a monolingual run, similarly processed as the other monolingual runs above.

- *English to Finnish* — We translated the English topics into Finnish using the Mediascape on-line dictionary [9]. For words present in the dictionary, we included all possible translations available. For words not present in the dictionary, we simply retained the original English words.

- *English to French* — We translated the English topics into French using the WorldLingo machine translation [16].

- *English to Russian* — Again, we translated the English topics into Russian using the WorldLingo machine translation [16].

Results of the mono- and bilingual runs just described were combined using unweighted combSUM. We also translated topics using another Russian on-line translator. However, the resulting translations were identical those provided by WorldLingo. We submitted a fifth multilingual run, UAmsC04EnMuAll2, including English to Russian results using both translations. This run scored inferior due to the overweighting of the Russian documents.

Table 4 provides an overview of the runs that we submitted for the multilingual task. The fourth column in Table 4 indicates the document sets used.

| Run | Topics | Documents | Type |
|---|---|---|---|
| UAmsC04EnMu4Gr | EN | EN, FI, FR, RU | 4 × 4-Grams+words |
| UAmsC04EnMuWSLM | EN | EN, FI, FR, RU | 8 × Words LM, Stems LM |
| UAmsC04EnMu3Way | EN | EN, FI, FR, RU | 12 × Words, Stems, 4-Grams+start/end |
| UAmsC04EnMuAll | EN | EN, FI, FR, RU | Grand combination of 7 runs per language |

Table 4: Overview of the multilingual runs submitted. Column 4 indicates the base runs used to generate the multilingual run.

## 4 Results

This section summarizes the results of our CLEF 2004 submissions.

## 4.1 Monolingual Results

Table 5 contains the mean average precision (MAP) scores for all the monolingual 'base' runs described in the previous section. The language model experiment clearly indicate the effectiveness of the stemming algorithm. For the vector space model, there is a small loss for Portuguese, but also a gain in performance for the other three languages. The outcome for the n-gram runs is less clear: there is a substantial gain in effectiveness for Finnish, but no or only a moderate gain for the other three languages. When comparing 4-gram with 4-gram+start/end, we see that including leading and ending 3-grams is always effective. Similarly, including words is effective for three of the four languages.

|  | *Finnish* | *French* | *Portuguese* | *Russian* |
|---|---|---|---|---|
| *Words* | 0.3776 | 0.4084 | 0.4032 | 0.3186 |
| *Stems* | 0.4549 | 0.4312 | 0.4023 | 0.3611 |
| *4-Grams* | 0.4949 | 0.3673 | 0.3439 | 0.2783 |
| *4-Grams+start/end* | **0.5264** | 0.3794 | 0.3653 | 0.3212 |
| *4-Grams+words* | 0.4930 | 0.4133 | 0.3723 | 0.3357 |
| *Words LM* | 0.3825 | 0.4059 | 0.4040 | 0.2958 |
| *Stems LM* | 0.4530 | **0.4463** | **0.4269** | **0.3847** |

Table 5: Overview of MAP scores for monolingual base runs. Best scores are in boldface.

Table 6 contains the MAP scores for the Finnish decompounding experiments. Decompouding leads to improvements for both retrieval models; decompounding and stemming only leads to improvements for the language model run. All Finnish n-gram runs in Table 5 outperform all decompounded runs.

|  | *Finnish Words* | | *Finnish Stems* | |
|---|---|---|---|---|
|  | *Words* | *Split* | *Stems* | *Split* |
| *Lnu.ltc* | 0.3776 | 0.4329 | 0.4549 | 0.4414 |
| *LM* | 0.3825 | 0.4021 | 0.4530 | **0.4617** |

Table 6: Overview of MAP scores for Finnish decompounding runs.

Finally, Table 7 lists the MAP scores for our official runs. For these, the grand combination of all base runs always outperforms the combination of a single (non)stemmed run and a single n-grammed run. When comparing with the best scoring base runs in Tables 5, we see that there is only a substantial improvement for Russian. There is a moderate improvement for French and Portuguese. The best Finnish n-gram run even outperforms the grand combination.

|  | *Finnish* | *French* | *Portuguese* | *Russian* |
|---|---|---|---|---|
| *4-Grams+words;(Split+)stem* | 0.4787 | 0.4410 | 0.4110 | 0.4227 |
| *4-Grams+start/end;(Split+)words* | 0.5007 | 0.4092 | 0.4180 | 0.4058 |
| *All base runs* | **0.5203** | **0.4499** | **0.4326** | **0.4412** |

Table 7: Overview of MAP scores for our officially submitted runs. Best scores per language are in boldface.

## 4.2 Bilingual Results

Table 8 shows the mean average precision scores for our base runs. For the resource-poor Amharic to English task, we expected a fairly low performance, somewhere in the 0.12–0.20 range. However, the vector space model run on the Words index is surprisingly effective. Furthermore, n-gramming leads to a loss of performance.

If we compare the different translation methods for the English to Portuguese tasks, and for the plain Words index, we see that, for query translation, the machine translation is more effective than the parallel corpus. This is as expected, since a word by word translation dictionary was derived from the parallel corpus. However, if the parallel corpus is used to translate the collection, we obtain a higher score for the Words index than both query translation methods. Applying a stemming algorithm is helpful for the MAP score for both ways of query translation, although it hurts the score of the collection translation. The use of n-gramming is only effective for query translation with the parallel corpus, where it leads to substantial improvements in the MAP score.

Table 8 also contains the run combinations that were submitted as official runs. The combination of a stemmed and a n-grammed run does generally not lead to improvement. The combination of all base runs leads to the best

| | Amharic to English (query) | English to Portuguese (query EU) | (Wordlingo) | (collection EU) |
|---|---|---|---|---|
| Words | 0.2071* | 0.2641 | 0.3220 | 0.3830 |
| Stems | 0.1961 | 0.3201 | 0.3901 | 0.3281 |
| 4-Grams | 0.1224 | 0.3704 | 0.2134 | 0.2954 |
| 4-Grams+start/end | 0.1300 | 0.3826 | 0.2296 | 0.2856 |
| 4-Grams+words | 0.1467 | 0.3678 | 0.2355 | 0.3203 |
| Words LM | 0.1694 | 0.2511 | 0.3167 | 0.3471 |
| Stems LM | 0.1703 | 0.2993 | 0.3835 | 0.3257 |
| 4-Grams+words;Stems | 0.1915* | 0.2755* | 0.3207* | 0.3850* |
| All base runs | **0.2138***  | | **0.4366***  | |

Table 8: Overview of MAP scores for all bilingual runs. Best scores are in boldface. Officially submitted runs are marked with an asterisk.

performance for Amharic to English, as well as for English to Portuguese. The score for English to Portuguese is particularly impressive, even outperforming our best monolingual score for Portuguese.

## 4.3 Multilingual Results

Table 9 shows our mean average precision scores for all base runs used in the multilingual task. We did not apply decompounding to the Finnish topics. As an aside, we see that for monolingual English, the language model is particularly effective. The results for Finnish, French, and Russian are generally in line with the monolingual results discussed above, be it that the n-gramming approaches are generally more effective on the translated topics.

| | | English to | | |
|---|---|---|---|---|
| | English | Finnish | French | Russian |
| Words | 0.4488 | 0.2057 | 0.3351 | 0.2012 |
| Stems | 0.4885 | **0.2719** | 0.3677 | 0.1478 |
| 4-Grams | 0.3986 | 0.2376 | 0.3585 | 0.2140 |
| 4-Grams+start/end | 0.4369 | 0.2578 | **0.3810** | **0.2623** |
| 4-Grams+words | 0.4387 | 0.2270 | 0.3596 | 0.2595 |
| Words LM | 0.4909 | 0.1913 | 0.3489 | 0.1935 |
| Stems LM | **0.5156** | 0.2303 | 0.3676 | 0.1978 |
| 4-Grams+words | | 0.2333* | | |
| Words LM;Stems LM | | 0.3040* | | |
| Words;Stems;4-Grams+start/end | | 0.3258* | | |
| All | | **0.3427*** | | |

Table 9: Overview of MAP scores for all multilingual runs (bottom half) and of the mono- and bilingual runs used to produce them (top half). Best scores are in boldface. Officially submitted runs are marked with an asterisk.

Table 9 also includes the run combinations submitted as official runs; recall that all these combinations are unweighted. The additional multilingual run having two English to Russian runs for each of the indexes scored lower with an MAP of 0.2520. On the whole, the performance increases with the number of runs included in the combination.

## 5 Conclusions

In this paper we documented the University of Amsterdam's participation in the CLEF 2004 ad hoc retrieval tasks: monolingual, bilingual, and multilingual retrieval. For the monolingual task, we conducted experiments on the effectiveness of morphological normalization approaches and combination methods. Our results shed further light on the effectiveness of language-dependent and language-independent approached to morphological normalization. As to the bilingual task, we experimented with bilingual retrieval in a resource-poor language, Amharic, and examined the relative effectiveness of different translation resources and of query versus collection translation. Our results indicate interesting differences between the bilingual approaches. The effectiveness of combining different translation methods was highlighted by the fact that the best bilingual score outperformed the best monolingual

score. Finally, for the multilingual task, we experimented with straightforward query translations and combination methods, and showed the effectiveness of combining a wide range of base runs.

## Acknowledgments

## References

[1] C. Buckley, A. Singhal, and M. Mitra. New retrieval approaches using SMART: TREC 4. In D.K. Harman, editor, *The Fourth Text REtrieval Conference (TREC-4)*, pages 25–48. National Institute for Standards and Technology. NIST Special Publication 500-236, 1996.

[2] E.A. Fox and J.A. Shaw. Combination of multiple searches. In D.K. Harman, editor, *The Second Text REtrieval Conference (TREC-2)*, pages 243–252. National Institute for Standards and Technology. NIST Special Publication 500-215, 1994.

[3] D. Hiemstra. *Using Language Models for Information Retrieval*. PhD thesis, Center for Telematics and Information Technology, University of Twente, 2001.

[4] V. Hollink, J. Kamps, C. Monz, and M. de Rijke. Monolingual document retrieval for European languages. *Information Retrieval*, 7(1):33–52, 2004.

[5] J. Kamps, C. Monz, and M. de Rijke. Combining evidence for cross-language information retrieval. In C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors, *Advances in Cross-Language Information Retrieval, CLEF 2002*, volume 2785 of *LNCS*. Springer, 2003.

[6] J. Kamps, C. Monz, M. de Rijke, and B. Sigurbjörnsson. Language-dependent and language-independent approaches to cross-lingual text retrieval. In C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors, *Cross-Language Information Retrieval, CLEF 2003*, *Lecture Notes in Computer Science*. Springer, 2004.

[7] P. Koehn. European parliament proceedings parallel corpus 1996-2003, 2004. `http://people.csail.mit.edu/people/koehn/publications/europarl/`.

[8] J.H. Lee. Combining multiple evidence from different properties of weighting schemes. In E.A. Fox, P. Ingwersen, and R. Fidel, editors, *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 180–188. ACM Press, New York NY, USA, 1995.

[9] Mediascape. English-Finnish-English on-line dictionary, 2004. `http://efe.scape.net/`.

[10] Neuchâtel. CLEF resources at the University of Neuchâtel, 2004. `http://www.unine.ch/info/clef`.

[11] C. Monz and M. de Rijke. Shallow morphological analysis in monolingual information retrieval for Dutch, German and Italian. In C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors, *Evaluation of Cross-Language Information Retrieval Systems, CLEF 2001*, volume 2406 of *LNCS*, pages 262–277. Springer, 2002.

[12] A. Nega. *Development of Stemming Algorithm for Amharic Text Retrieval*. PhD thesis, University of Sheffield, 1999.

[13] J.J. Rocchio, Jr. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall Series in Automatic Computation, chapter 14, pages 313–323. Prentice-Hall, Englewood Cliffs NJ, 1971.

[14] Snowball. Stemming algorithms for use in information retrieval, 2004. `http://www.snowball.tartarus.org/`.

[15] European Union. Official Journal of the European Union, 2004. `http://europa.eu.int/eur-lex/`.

[16] Worldlingo. Online translator, 2004. `http://www.worldlingo.com/`.