CLEF 2002 – Overview of Results

Martin Braschler

Eurospider Information Technology AG
Schaffhauserstr. 18, 8006 Zürich, Switzerland
martin.braschler@eurospider.com
Université de Neuchâtel, Institut interfacultaire d'informatique
Pierre-à-Mazel 7, CH-2001 Neuchâtel, Switzerland

Abstract. In its third year, the CLEF campaign has again seen considerable growth on multiple fronts. While the explosive growth in the number of participants has slowed somewhat, the number of actual experiments has grown considerably, as has their complexity (more data to process and more languages to handle). The main tracks of the CLEF 2002 campaign attracted 37 participating groups who submitted nearly 300 different experiments. In this overview, a description of the tracks and tasks, and a summary of the principal research results are given. As for the last two years, we have also examined the multilingual test collection produced as a result of the campaign with respect to the completeness of its relevance assessments, with very favorable findings.

1 Introduction

The third CLEF campaign was held from January to September 2002. Overviews of the results of the two previous campaigns have been published elsewhere (see [4] for CLEF 2000, and [5] for CLEF 2001). CLEF, a successor to the TREC-6-8 cross-language information retrieval (CLIR) tracks [11], had experienced explosive growth in the number of participants in the first two years of its existence. In 2002, the number of participants stabilized somewhat, growing at a slower rate. However, the 2002 campaign saw substantial growth in other notable areas, mainly in the experiments conducted for the campaign – both in terms of the number of experiments submitted, but also in terms of their complexity. This has resulted in the 2002 campaign having what is probably the "richest" set of results produced through CLEF activities so far.

CLEF 2002 offered several different tracks and tasks, and this paper covers the core activities: the multilingual, bilingual, monolingual and domain-specific tracks, which are directly managed by the CLEF consortium. In addition, CLEF 2002 featured an interactive track (expanded from 2001) and a pilot experiment on cross-language speech retrieval. The results of these tracks are discussed elsewhere ([9], [10]). The aim of this paper is to summarize and analyze the main results and research directions, as well as to compare the findings with those of the previous years and
provide a statistical analysis (Sections 2, 3, 4 and 5). We also investigate the CLEF 2002 test collection with respect to the completeness and validity of the relevance assessments (Section 6). The paper closes with a summary and conclusions (Section 7).

2 Tracks and Tasks

The core tracks and tasks offered for the CLEF 2002 campaign were:

- **Multilingual Retrieval.** Retrieval of text documents in any of five languages (English, French, German, Italian, Spanish) using queries formulated in one language (choice of eleven different languages\(^1\); see also Table 4). Result lists contain items from all five document languages. This was the “main” track, and participants were actively encouraged to tackle it.

- **Bilingual Retrieval.** Retrieval of text documents written in a language different from the query language. Participants could choose one of seven possible target languages (Dutch, Finnish, French, German, Italian, Spanish, Swedish), and from a selection of eleven different query languages. English as a target language was permitted for newcomers only.

- **Monolingual Retrieval.** Retrieval of text documents from a collection written in one of seven languages: Dutch, Finnish, French, German, Italian, Spanish or Swedish. For this track, the query language is identical to document language.

- **Domain-specific Retrieval.** Retrieval on a German or a French document collection containing scientific texts. An accompanying thesaurus/controlled vocabulary was available, and queries were provided in three (German, English, Russian) and two (English, French) languages, respectively.

Participants sent their results in the form of ranked lists containing those documents that best match a given query. They submitted one or several experiments ("runs") for an individual task. For reasons of tractability by the campaign organizers, the maximum number of experiments for each task was limited.

In total, 37 groups from 12 different countries participated in one or more of the tracks and tasks that were offered for CLEF 2002 (see Table 1). Table 2 compares the number of participants and experiments to those of earlier TREC CLIR tracks [11] and earlier CLEF campaigns. While the first two CLEF campaigns in 2000 [4] and 2001 [5] were clearly a breakthrough in promoting larger participation with respect to the CLIR track at TREC, the growth in the number of participants has slowed somewhat. CLEF has a high retention rate of participants (Table 1).

A total of 282 experiments were submitted, an increase of more than 40% compared to last year. A breakdown into the individual tasks can be found in Table 3.

All query languages were used for experiments, including the translations of the queries into Chinese, Portuguese, Russian, and Japanese (post-campaign only), which were provided by independent third parties. English, French, German and Spanish were the most popular query languages, with Dutch and Italian closely behind. By far

\(^1\) After the campaign, queries were also made available in a twelfth language, Japanese.