



Project Name	INFILE
Acronyme	Information, Filtering, Evaluation
Call Offer	ANR 2006 ANR -06 MDCA -011
Thematic	Automatic Treatment of Languages
Programs	Mass of Data and Ambient Knowledge
Duration	2 years (2007 - 2008)

Partners :

CEA LIST Leader (http://www-list.cea.fr)
GERIICO (http://www.univ-lille3.fr/fr/recherche/equipes-recherche/geriico/)
ELDA (http://www.elda.org)
For more information see http://www.infile.org

Introduction

INFILE (*INformation, Filtrage, Evaluation*) is a cross-language adaptive filtering evaluation campaign, sponsored by the French National Research Agency, and organized by the CEA LIST (C. Fluhr, H. Dahmani), ELDA (K. Choukri, C. Ayache), and the University of Lille 3 (S. Chaudiron, I. Timimi).

This new campaign extends the last filtering track of TREC 2002 in the following ways:

- INFILE is crosslingual (English, French and Arabic), a corpus of 100,000 comparable newswire stories for each language. There will be two groups of topics one about general news and events, and a second on more technology subjects
- Evaluation will be performed using an automatic interrogation of test systems with a simulated user feedback. Each system will be able to use the feedback at any time to increase performance.

The test systems will provide a Boolean decision for each document according to each filter profile. A curve of the evolution of efficiency will be computed.

This campaign is also open to participants with monolingual systems.

International extension of the INFILE evaluation campaign

- INFILE is a pilot track of the CLEF 2008 campaigns.
- INFILE will benefit of the international CLEF framework.
- NIST will promote the participation to INFILE especially for American participants during the next TREC workshop.

Brief description of the protocol

- General information about the domain of topics) is given to each participant.
- 15 days afterwards, 25 to 50 actual topics are given to participants. Topics are composed of a list of keywords (simple and complex noun phrases) and up to 3 documents illustrating each topic.
- Documents are transmitted by the organizer to an automated interface of each participating system. The interface returns a Boolean response for each topic.
- After reception of this response, the organizer sends a feedback consisting of expected topic assignments for each document submitted.
- Participants may adapt their system at any time using this feedback.

Dry run

A dry run will be organized to ensure that the whole process including the automated submission process runs correctly for each system.

Evaluation

To measure adaptive filtering performance, after every N documents (for example N=20000) precision, recall and measures that were used in the last Filtering track of TREC (2002) will be computed to plot an effectiveness evolution curve. At the end of the test set a mean effectiveness will be computed for each system.

We will also measure the number of documents that each system uniquely and correctly filters (originality measure). Discussions are open to other measures.

Ground truth

Because of the corpus size and the fact that feedback must be sent immediately after receiving the answer, a pooling methodology is not possible. We will use a new corpus not already used in a previous evaluation campaign to avoid distortion between participants.

We have decided to build the corpus and ground truth in the following way:

- Specialists of the domain and professionals of information systems (assessors) will identify documents relevant for the topics.
- A set of non relevant but close-to-topic documents will be specially chosen to ensure some confusion with the topics.
- A large set of documents that are out of scope will be added to prevent manual examination of the corpus. To ensure that this corpus does not contain any relevant document, assessors will use state-of-the-art information retrieval systems on the full data to detect and eliminate such documents.

This will be done before the beginning of the test.

At the end, a limited control (via limited pooling) will be performed on documents not yet judged (in the larger set of documents) but which have been chosen by at least 2 systems for a given topic. This check will permit the organizers to adjust the ground truth and the measured results and to attribute a degree of reliability to the feedback given to the participants. If few modifications are needed, the way each system uses the feedback to increase performance can be considered as representative. If this limited pooling control detects that many modifications are needed, the results of using feedback will be less reliable to judge.

Expected Schedule

The INFILE schedule is being finalized. Here's the current (Nov 2007) draft version:

December 3, 2007	Call for participation – Registration opens
March 18 - April 15, 2008	Dry run campaign – Protocol's validation
June 3 - July 1, 2008	Evaluation campaign
July 15, 2008	Release of preliminary Individual Results
July 16 - 25, 2008	Results analysis – Final Adjudication
August 15, 2008	Submission of papers for Working Notes
September 17-19, 2008	Infile one day Workshop – CLEF workshop