Overview of CLEF 2009 INFILE track

Romaric Besançon*, Stéphane Chaudiron**, Djamel Mostefa+,
Ismaïl Timimi**, Khalid Choukri+, Meriama Laïb*

*CEA LIST
18, route du panorama BP 6 92265
Fontenay aux Roses France

**Université de Lille 3 – GERiiCO
Domaine univ. du Pont de Bois 55-57,
BP 60149 - 59653 Villeneuve d'Ascq cedex France

+ELDA
rue Brillat Savarin
75013 Paris France

romaric.besancon@cea.fr, meriama.laib@cea.fr, stephane.chaudiron@univ-lille3.fr, mostefa@elda.org,
ismail.timimi@univ-lille3.fr, choukri@elda.org

Abstract

The INFILE@CLEF 2009 track is the second run of this track on the evaluation of cross-language adaptive filtering systems. It uses the same corpus as the 2008 track, composed of 300,000 newswires from Agence France Presse (AFP) in three languages: Arabic, English and French, and a set of 50 topics in general and specific domain (scientific and technological information). We proposed this year two tasks : a batch filtering task and an interactive task to test adaptive methods. Results for the two tasks are presented in this paper.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

General Terms

Measurement, Performance, Experimentation, Algorithms

Keywords

Information Filtering, Competitive Intelligence

1 Introduction

The purpose of the INFILE (INformation FILtering Evaluation) track is to evaluate cross-language adaptive filtering systems, i.e. the ability of automated systems to successfully separate relevant and non-relevant documents in an incoming stream of textual information with respect to a given profile, the document and profile being possibly written in different languages.
The INFILE track has first been run as a pilot track in CLEF 2008 campaign [Besançon et al, 2008]. Due to some delays in the organization, the participation in the 2008 was weak (only one participant submitted results), so we decided to propose to rerun the campaign in 2009, using the same document collection and topics.
The INFILE project is funded by the French National Research Agency and co-organized by the CEA LIST, ELDA and the University of Lille3-GERiiCO.
Information filtering in the INFILE track is considered in the context of competitive intelligence: in this context, the evaluation protocol of the campaign has been designed with a particular attention to the context of use of filtering systems by real professional users. Even if the campaign is mainly a technological oriented evaluation process, we adapted the protocol and the metrics, as close as

possible, to how a normal user would proceed, including through some interaction and adaptation of his system.
The INFILE campaign can mainly be seen as a cross-lingual pursuit of the TREC 2002 Adaptive Filtering task [Robertson and Soboroff, 2002] (adaptive filtering track has been run from 2000 to 2002), with a particular interest in the correspondence of the protocol with the ground truth of competitive intelligence (CI) professionals. In this goal, we asked CI professionals to write the topics according to their experience in the domain.
Other related campaigns are the Topic Detection and Tracking (TDT) campaigns from 1998 to 2004 [Fiscus and Wheatley, 2004], but in the TDT campaigns, focus was mainly on topics defined as "events", with a fine granularity level, and often temporally restricted, whereas in INFILE (similar to TREC 2002), topics are of long-term interest and supposed to be stable, which can induce different techniques, even if some studies show that some models can be efficiently trained to have good performance on both tasks [Yang et al., 2005].

2 Description of the tasks

In addition to the adaptive filtering task already proposed in 2008 [Besançon et al, 2008], we introduced in 2009 the possibility to test batch filtering systems.

For both tasks, the document collection consists in a set of newswire articles provided by the Agence France Presse (AFP) and covering recent years, the topic set is composed of two different kinds of profiles, one concerning general news and events, and a second one on scientific and technological subjects.
The filtering process may be crosslingual: English, French and Arabic are available for the documents and topics, and participants may be evaluated on monolingual runs, bilingual runs, or multilingual runs (with several target languages).
The purpose of the information filtering process is to associate documents in an incoming stream to zero, one or several topics: Filtering systems must provide a Boolean decision for each document with respect to each topic.

For the batch filtering task, participants are provided with the whole document collection and must return the list of relevant documents for each topic (since the filtering process supposes a binary decision for each document, the document list does no need to be ranked).

For the adaptive filtering task, the evaluation is performed using an automatic interactive process, with a simulated user feedback: systems are allowed for each document considered relevant to a topic to ask for a feedback on this decision (i.e. ask if the document was indeed relevant for the topic or not), and can modify their behaviour according to the answer. The feedback is allowed only on kept document, there is no relevance feedback possible on discarded documents. In order to simulate the limited patience of the user, a limited number of feedbacks is allowed: this number has been fixed in 2009 to 200 feedbacks (it was 50 in 2008; but most participants considered this insufficient). The adaptive filtering task uses an interactive client-server protocol, that is described in more details in [Besançon et al.,2008].

The batch filtering task has been run from April 2nd (document collections and topics made available to the participants) to June 1st (run submission), and the adaptive filtering task has been run from June 3rd to July 10th.

3 Test collections

3.1 The topics

A set of 50 profiles has been prepared, covering two different categories: the first group (30 topics) deals with general news and events concerning national and international affairs, sports, politics etc and the second one (20 topics) deals with scientific and technological subjects. The scientific topics were developed by CI professionals from INIST[1], ARIST Nord Pas de Calais[2], Digiport[3] and OTO Research[4]. The topics were developed in both English and French. The Arabic version has been translated from English and French by native speakers.

Topics are defined with the following elements: a unique identifier, a title (6 words max.), describing the topic in a few words, a description (20 words max.), corresponding to a sentence-long description, a narrative (60 words max.), corresponding to the description of what should be considered a relevant document and possibly what should not, keywords (up to 5) and an example of relevant text (120 words max.), taken from a document that is not in the collection (typically from the web).

Each record of the structure in the different languages correspond to translations, except for the samples which need to be extracted from real documents. An example of topic in the three languages is presented in Fig. 1.

<top> **<num>**147**</num>** **<title>**Care management of Alzheimer disease**</title>** **<desc>**News in the care management of Alzheimer disease by families, society and politics**</desc>** **<narr>**Relevant documents will highlight differents aspects of Alzheimer disease management: - human involvement of carers : families, health workers - financial means: nursing facilities, diverse grants to carers - political decisions leading to guidelines for optimal management of this great public health problem **</narr>** **<keywords>** **<keyword>**Alzheimer disease**</keyword>** **<keyword>**Dementia **</keyword>** **<keyword>**Care management **</keyword>** **<keyword>**Family support **</keyword>** **<keyword>**Public health**</keyword>** **</keywords>** **<sample>**The AAMR/IASSID practice guidelines, developed by an international workgroup, provide guidance for stage–related care management of Alzheimer's disease, and suggestions for the training and education of carers, peers, clinicians and programme staff. The guidelines suggest a three-step intervention activity process, that includes: (1) recognizing changes; (2) conducting...**</sample>** **</top>**	**<top>** **<num>**147**</num>** **<title>**Prise en charge de la maladie d'Alzheimer**</title>** **<desc>**Actualités dans le domaine de la prise en charge de la maladie d'Alzheimer, tant au niveau des familles, de la société qu'au niveau des choix politiques**</desc>** **<narr>**Les documents pertinents présenteront les divers aspects de la prise en charge de la maladie d'Alzheimer : - moyens humains mis en jeu : familles, personnels de santé - moyens financiers : structures d'accueil, aides diverses aux malades et aux aidants - décisions politiques avec établissement de recommandations permettant d'encadrer de façon optimale ce problème majeur de santé publique **</narr>** **<keywords>** **<keyword>**Maladie d'Alzheimer**</keyword>** **<keyword>**Démence **</keyword>** **<keyword>**Prise en charge **</keyword>** **<keyword>**Aide aux familles **</keyword>** **<keyword>**Santé publique **</keyword>** **</keywords>** **<sample>**Un an après l'entrée en vigueur du plan ministériel, un rapport de l'OPEPS rendu public le 12 juillet 2005 dresse un bilan assez sévère de la prise en charge de la maladie d'Alzheimer et des maladies apparentées. Selon l'OPEPS*, la politique de prévention des facteurs de risque est insuffisante, ... **</sample>** **</top>**	**<top>** **<num>**147**</num>** **<title>**العناية بمرض الزهايمر**</title>** **<desc>**الأحداث المتعلقة بالعناية بمرض الزهايمر، على مستوى الأسر والمجتمع وأيضا على مستوى الاختيارات السياسية.**</desc>** **<narr>**الوثائق التي تتعلق بالعناية بمرض الزهايمر من مختلف الجوانب : - الإمكانات البشرية المستخدمة : الأسر، موضفو الصحة، - الموارد المالية : بنيات الإستقبال، المساعدات المختلفة للمرضى والمساعدين، - القرارات السياسية : التعليمات الصادرة من أجل وضع إطار أمثل لهذا المشكل الكبير في الصحة العمومية.**</narr>** **<keywords>** **<keyword>**الصحة العمومية**</keyword>** **<keyword>**مساعدة الأسر**</keyword>** **<keyword>**عناية**</keyword>** **<keyword>**الجنون**</keyword>** **<keyword>**مرض الزهايمر**</keyword>** **</keywords>** **<sample>**... الوضع عبر الهاتف كلما اقتضت الحاجة ذلك. وكانت دراسة سابقة قد كشفت أن عدد المصابين بمرض الزهايمر سيتضاعف أربع مرات خلال العقود الأربعة المقبلة، ويصيب واحداً من أصل كل 85 شخصاً على وجه الأرض.وأكدت الدراسة أن هذه الإحصائية المخيفة مرتبطة بشكل رئيسي بارتفاع عدد كبار السن في مختلف دول العالم، الناجم عن تحسن الأنظمة الصحية، وقدرت أنه بحلول العام 2050 فإن أعداد أولئك المرضى ستقفز إلى 62.8 مليون شخص. بحسب الـCNN.**</sample>** **</top>**

Fig. 1 An example of topic for the INFILE track, in the three languages

1 the French Institute for Scientific and Technical Information Center, http://international.inist.fr/
2 Agence Régionale d'Information Stratégique et Technologique, http://www.aristnpdc.org/
3 http://www.digiport.org
4 http://www.otoresearch.fr/

3.2 The document collection

The INFILE corpus is provided by the Agence France Presse (AFP) for research purpose. We used newswire articles in 3 languages: Arabic, English and French[5] and a 3 years period (2004-2006) which represents a collection of about one and half million newswires for around 10 GB, from which 100,000 documents of each language have been selected to be used for the INFILE filtering test. News articles are encoded in XML format and follow the News Markup Language (NewsML) specifications[6]. An example of document in English is given in Fig. 2. All fields are available to the systems and can be used in the filtering process (including keywords, categorization...).

```
<NewsML Version="1.1">
 <NewsEnvelope>
  <TransmissionId>807</TransmissionId>
  <DateAndTime>20050615T212137Z</DateAndTime>[...]
 </NewsEnvelope>
 <NewsItem>
  <Identification>
   <NewsIdentifier>
    <ProviderId>afp.com</ProviderId>
    <DateId>20050615</DateId>
    <NewsItemId>TX-SGE-DPE59</NewsItemId>
    <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
    <PublicIdentifier>urn:newsml:afp.com:20050615:TX-SGE-DPE59:1</PublicIdentifier>
   </NewsIdentifier>
   <NameLabel>Mideast-unrest-Israel-Palestinians</NameLabel>
  </Identification>
  <NewsManagement>[...]</NewsManagement>
  <NewsComponent>
   <TopicSet FormalName="NewsTopics">
    <Topic Duid="topic1"><TopicType FormalName="SlugKeyword"/><Description>Mideast</Description></Topic>
    <Topic Duid="topic2"><TopicType FormalName="SlugKeyword"/><Description>unrest</Description></Topic>
    <Topic Duid="topic3"><TopicType FormalName="SlugKeyword"/><Description>Israel</Description></Topic>
    <Topic Duid="topic4"><TopicType FormalName="SlugKeyword"/><Description>Palestinians</Description></Topic>
   </TopicSet>
   <NewsLines>
    <SlugLine>Mideast-unrest-Israel-Palestinians</SlugLine>
    <HeadLine>Israel says teenage would-be suicide bombers held</HeadLine>
   </NewsLines>
   <AdministrativeMetadata>[...]</AdministrativeMetadata>
   <DescriptiveMetadata>
    <Language FormalName="en"/>
    <SubjectCode><Subject FormalName="11999000"/></SubjectCode>
    <SubjectCode><Subject FormalName="INT" Vocabulary="urn:newsml:afp.com:20011001:AFPCatCodes:1"/></SubjectCode>
    <Location>
     <Property FormalName="Country" Value="ISR"/>
     <Property FormalName="City" Value="JERUS"/>
    </Location>
    </DescriptiveMetadata>
   <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="NITF3.1-body.content"/>
    <Characteristics><Property FormalName="Words" Value="89"/></Characteristics>
    <DataContent>
      <p>JERUSALEM, June 15 (AFP) - The Israeli security service said Wednesday it had arrested four Palestinian teenage boys who were preparing to carry out suicide bombings.Shin Beth said the four, aged 16 and 17, belonged to the Fatah movement. It said they planned to hit targets in Israel or Israeli troops.</p>
      <p>Four other young adults, also accused of Fatah membership, were picked up in Nablus in the north of the West Bank some weeks ago.</p>
      <p>Shin Beth said the network was financed by the Shiite Lebanese Hezbollah group.</p>
      <p>ms/sj/gk</p>
    </DataContent>
   </ContentItem>
  </NewsComponent>
 </NewsItem>
</NewsML>
```

Fig. 2 Exemple of document in the INFILE collection

5 Newswires in different languages are not translations from a language to another (it is not an aligned corpus): the same information is generally rewritten to match the interest of the audience in the corresponding country.

6 NewsML is an XML standard designed to provide a media-independent, structural framework for multi-media news. NewsML was developed by the International Press Telecommunications Council. see http://www.newsml.org/

Since we need to provide a real-time simulated feedback to the participants, we need to have the identification of relevant documents prior to the campaign, as in [Soboroff and Robertson, 2002]. The method used to build the collection of documents with the knowledge of the relevant documents is presented in details in [Besançon et al.,2008]. A summary of this method is given here.

We used a set of 4 search engines (Lucene[7], Indri[8], Zettair[9] and the search engine developed at CEA-LIST) to index the complete collection of 1.4 million documents. Each search engine has been queried using different fields of the topics, which provides us with a pool of runs. We first selected the first 10 retrieved documents of each run, and these documents were assessed manually. We then iterate using a *Mixture of Experts* model, computing a score for each run according to the current assessment and using this score to weight the choice of the next documents to assess. The final document collection is then built by taking all documents that are relevant to at least one topic (core relevant corpus), all documents that have been assessed and judged not relevant (difficult corpus: documents are not relevant, but share something in common with at least one topic, since they have been retrieved by at least one search engine), and a set of documents taken randomly in the rest of the collection (filler corpus, with documents that have not been retrieved by any search engines for any topic, which should limit the number of relevant documents in the corpus that have not been assessed).

Statistics on the number of assessed documents and relevant documents is presented in Table 1. The repartition of relevant documents across topics is presented in Fig3.

	eng	**fre**	**ara**
number of documents assessed	7312	7886	5124
number of relevant documents	1597	2421	1195
avg number of relevant docs / topic	31,94	48,42	23,9
std deviation on number of relevant docs / topic	28,45	47,82	23,08
[min,max] number of relevant docs / topics	[0,107]	[0,202]	[0,101]

Table 1 Statistics on the number of assessed documents and the number of relevant documents, in each language

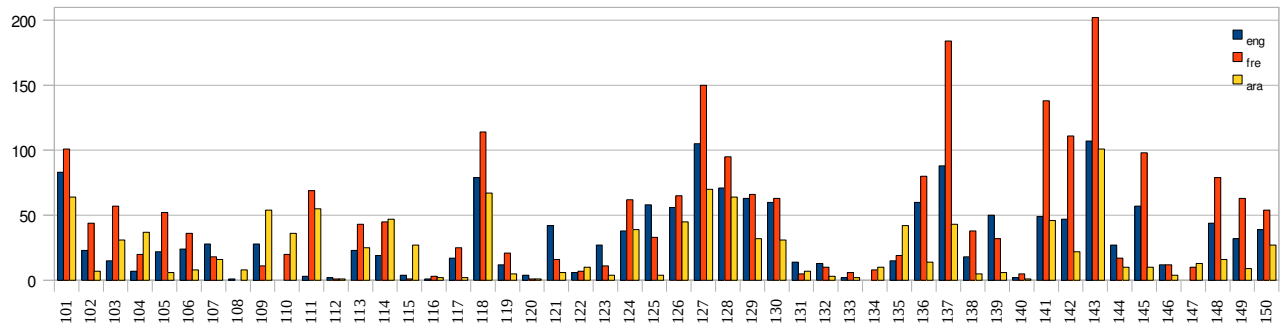

Fig. 3 Number of relevant documents for each topic, in each language

4 Metrics

The results returned by the participants are binary decisions on the association of a document with a profile. The results, for a given profile, can then be summarized in a contingency table of the form:

7 http://lucene.apache.org
8 http://www.lemurproject.org/indri
9 http://www.seg.rmit.edu.au/zettair

	Relevant	Not Relevant
Retrieved	a	b
Not Retrieved	c	d

On these data, a set of standard evaluation measures is computed:

- Precision, defined as $P=\frac{a}{a+b}$
- Recall , defined as $R=\frac{a}{a+c}$
- F-measure, which is a standard combination of precision and recall [Van Rijsbergen, 1979]

  $$F=\frac{1}{\alpha\frac{1}{P}+(1-\alpha)\frac{1}{R}}$$

  depending on a parameter α , and defined as

  We used the standard value $\alpha=0.5$, which gives the same importance to precision and recall (F-measure is then the harmonic mean of the two values).

Following the TREC Filtering tracks [Hull and Roberston, 1999, Robertson and Soboroff, 2002] and the TDT 2004 Adaptive tracking task [Fiscus and Wheatley, 2004], we also consider the linear utility, defined as

$$u=w_1\times a-w_2\times b$$

where w_1 is the importance given to a relevant document retrieved and w_2 is the cost of an non relevant document retrieved.

Linear utility is bounded positively (to 1 for a perfect filtering), but unbounded negatively (negative values depend on the number of relevant documents for a profile). Hence, the average value on all profiles would give too much importance to the few profiles on which a systems would perform

$$u_n=\frac{\max(\frac{u}{u_{max}},u_{min})-u_{min}}{1-u_{min}}$$

poorly. To be able to average the value, the measure is scaled as follows:

where u_{max} is the maximum value of the utility and u_{min} a parameter considered to be the minimum utility value under which a user would not even consider the following documents for the profile. In the INFILE campaign, we used the values $w_1=1$, $w_2=0.5$, $u_{min}=-0.5$ (same as in TREC 2002).

We considered last year the *detection cost* measure (from the Topic Detection and Tracking campaigns [NIST, 1998]), but we do not present this score in this paper (we found that detection cost values were often low and not really discriminant between participants).

To compute average scores, the values are first computed for each profile and then averaged. In order to measure the adaptivity of the systems in the adaptive filtering track, the measures are also computed at different times in the process, each 10,000 documents, and an evolution curve of the different values across time is presented.

Additionally, we use the two following measures, introduced last year in INFILE: the first one is an originality measure, defined as a comparative measure corresponding to the number of relevant documents the system uniquely retrieves (among participants). It gives more importance to systems

that use innovative and promising technologies that retrieve "difficult" documents.
The second one is an anticipation measure, designed to give more interest to systems that can find the first document in a given profile. This measure is motivated in CI by the interest of being at the cutting edge of a domain, and not missing the first information to be reactive. It is measured by the inverse rank of the first relevant document detected (in the list of the documents), averaged on all profiles. The measure is similar to the mean reciprocal rank (MRR) used for instance in Question Answering Evaluation [Voorhees, 1999], but is not computed on the ranked list of retrieved documents but on the chronological list of the relevant documents.

5 Overview of the results

On the 9 participants registered for both tasks, 5 submitted results : 3 participants submitted results for the batch filtering task (a total of 9 runs), 2 for the interactive filtering task (3 runs). Participants were different for the two tasks. Table 1 present the participant list.

team name	institute	country
IMAG	Institut Informatique et Mathématiques Appliquées de Grenoble	France
SINAI	University of Jaen	Spain
UAIC	Universitatea Alexandru Ioan Cuza of IASI	Romania
HossurTech	société CADEGE	France
UOWD	University of Wollongong (Comp.Sci & Engineering)	Dubai

Table 1 Participant list

Concerning the languages, 6 runs out of 9 are monolingual English for the batch filtering task, 3 are multilingual from English to French/English. For the interactive task, one run is monolingual English, one is monolingual French, and one is bilingual French to English. Table 2 summarizes the total number of runs for each language pair. No participant submitted runs with Arabic as source or target language.

nb runs			
	english	French	Arabic
English	10	3	0
French	1	1	0
Arabic	0	0	0

Table 2 Repartition of runs according to the source and target languages

The runs and their characteristics are presented in Table 3.

team	run	task	source	target	topic fields	document fields
IMAG	IMAG_1	batch	eng	eng	all	all
IMAG	IMAG_2	batch	eng	eng	all	all
IMAG	IMAG_3	batch	eng	eng	all	all
UAIC	uaic_1	batch	eng	eng	num, title, desc, narr, keywords, sample	DateID, NewsItemID, Slugline, Headline, DataContent, Country, City, FileName
UAIC	uaic_2	batch	eng	eng-fre	num, title, desc, narr, keywords, sample	DateID, NewsItemID, Slugline, Headline, DataContent, Country, City, FileName
UAIC	uaic_3	batch	eng	eng-fre	num, title, desc, narr, keywords, sample	DateID, NewsItemID, Slugline, Headline, DataContent, Country, City, FileName
UAIC	uaic_4	batch	eng	eng-fre	num, title, desc, narr, keywords, sample	Headline, DataContent, FileName
SINAI	topics_1	batch	eng	eng		
SINAI	googlenews_2	batch	eng	eng		
HossurTech	hossur-tech-001	adaptive	fre	eng	all	
HossurTech	hossur-tech-004	adaptive	fre	fre	all	
UOWD	base	adaptive	eng	eng	title,desc	DataContent

Table 3 The runs, by team and by run name, and their characteristics

Evaluation scores for the runs in the batch filtering task are presented in Table 4, gathered by the target language (multilingual runs appears in several groups, in order to present the individual scores on each target language). Best result is obtained on monolignual English, but for the only participant that tried multilingual runs, the results obtained for the different target languages (English and French) are comparable.

monolingual english

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
IMAG	IMAG_1	1597	413	0,26	0,30	0,21	0,21	0,43
UAIC	uaic_4	1597	1267	0,09	0,66	0,13	0,05	0,73
UAIC	uaic_1	1597	1331	0,06	0,69	0,09	0,03	0,75
UAIC	uaic_2	1597	1331	0,06	0,69	0,09	0,03	0,75
UAIC	uaic_3	1597	1507	0,06	0,82	0,09	0,03	0,86
IMAG	IMAG_2	1597	109	0,13	0,09	0,07	0,16	0,22
IMAG	IMAG_3	1597	66	0,16	0,06	0,07	0,22	0,14
SINAI	topics_1	1597	940	0,02	0,50	0,04	0,00	0,57
SINAI	googlenews_2	1597	196	0,01	0,08	0,01	0,13	0,10

crosslingual english → french

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
UAIC	uaic_4	2421	1120	0,09	0,44	0,12	0,05	0,58
UAIC	uaic_3	2421	1905	0,06	0,75	0,10	0,03	0,83
UAIC	uaic_2	2421	1614	0,06	0,67	0,09	0,02	0,76

multilingual english → english/french

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
UAIC	uaic_4	4018	2387	0,07	0,56	0,11	0,02	0,72
UAIC	uaic_3	4018	3412	0,05	0,81	0,08	0,02	0,85
UAIC	uaic_2	4018	2945	0,05	0,70	0,07	0,02	0,80

Table 4 Scores for batch filtering runs, sorted by F-score

Scores for the runs in the adaptive filtering task are presented in Table 5. The scores are worse than the scores obtained on the batch filtering results, but the language pairs and the participants are not the same. We also note than both batch and adaptive results for the INFILE 2009 campaign are worse than the results obtained for the adaptive task in the INFILE 2008 edition.

monolingual english

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
UOWD	base	1597	20	0,00	0,01	0,01	0,03	0,05

monolingual french

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
HossurTech	hossur-tech-004	2421	790	0,05	0,31	0,06	0,05	0,53

crosslingual french → english

team	run	num_rel	num_rel_ret	precision	recall	F-score	Utility	anticipation
HossurTech	hossur-tech-001	1597	819	0,10	0,45	0,10	0,07	0,59

Table 5 Scores for adaptive filtering runs

Results for originality measure are presented in Table 6. The upper part of the table present

originality scores for every run that has the same target language (i.e. the number of relevant documents that this particular run uniquely retrieves). Since this global comparison may not be fair for participants who submitted several runs, which are presumably variants of the same technique and will share most of the relevant retrieved documents, we present in the lower part of the table the originality scores using only one run for each participant (we chose the run with the best recall score). We see here that participant with lower F-scores can have a better originality score. However, due to the small number of participants, the relevance of the originality score is arguable in this context, since it seems to be strongly linked to the difference of the recall score.

originality on all runs

target lang=eng			target lang=fre		
team	**run**	**originality**	**team**	**run**	**originality**
UAIC	uaic_3	39	HossurTech	hossur-tech-004	177
HossurTech	hossur-tech-001	18	UAIC	uaic_3	82
SINAI	googlenews_2	15	UAIC	uaic_2	0
SINAI	topics_1	9	UAIC	uaic_4	0
UAIC	uaic_4	4			
IMAG	IMAG_1	1			
UAIC	uaic_1	0			
IMAG	IMAG_3	0			
UOWD	base	0			
UAIC	uaic_2	0			
IMAG	IMAG_2	0			

originality on best run

target lang=eng			target lang=fre		
team	**run**	**originality**	**team**	**run**	**originality**
UAIC	uaic_3	267	UAIC	hossur-tech-004	1292
HossurTech	hossur-tech-001	20	HossurTech	uaic_3	177
SINAI	topics_1	9			
IMAG	IMAG_1	4			
UOWD	base	0			

Table 6 Originality scores

6 Conclusion

The INFILE campaign has been organized for the second time this year in CLEF, to evaluate adaptive filtering systems in a cross-language environment. The document and topic collection were the same as the 2008 edition of the INFILE@CLEF track. Two tasks have been proposed: a batch filtering task and an adaptive filtering task, that used an original setup to simulate the incoming of newswires documents, and the interaction of a user through a simulated feedback. We had this year more participants than last year and more results to analyze. However, the innovative crosslingual aspect of the task has still not really been explored, since most runs were monolingual English and no participant used the Arabic topics or documents. The lack of participation for the adaptive task is also disappointing since it does not provide enough data to compare batch techniques to adaptive techniques and does not allow to conclude on the interest of the use of the used feedback on the documents.

References

[Besancon et al, 2008] Besancon R., Chaudiron S., Mostefa D., Hamon O., Timimi I. and Choukri K. (2008) Overview of CLEF 2008 INFILE Pilot Track, Overview of CLEF 2008 INFILE Pilot Track.

[Fiscus and Wheatley, 2004] Fiscus, J. and Wheatley, B. (2004). Overview of the tdt 2004 evaluation and results. In TDT'02. NIST.

[Hull and Roberston, 1999] Hull, D. and Roberston, S. (1999). The trec-8 filtering track final report. In Proceedings of the Eighth Text REtrieval Conference (TREC-8). NIST.

[NIST, 1998] NIST (1998). The topic detection and tracking phase 2 (tdt2) evaluation plan. http://www.nist.gov/speech/tests/tdt/1998/doc/tdt2.eval.plan.98.v3.7.pdf.

[Robertson and Soboroff, 2002] Robertson, S. and Soboroff, I. (2002). The trec 2002 filtering track report. In Proceedings of The Eleventh Text Retrieval Conference (TREC 2002). NIST.

[Soboroff and Robertson, 2002] Soboroff, I. and Robertson, S. (2002). Building a filtering test collection for trec 2002. In Proceedings of The Eleventh Text Retrieval Conference (TREC 2002). NIST.

[Van Rijsbergen, 1979] Van Rijsbergen, C. (1979). Information Retrieval. Butterworths, London.

[Voorhees, 1999] Voorhees, E. (1999). The trec-8 question answering track report. In Proceedings of the Eighth Text REtrieval Conference (TREC-8). NIST.

[Yang et al., 2005] Yang, Y., Yoo, S., Zhang, J., and Kisiel, B. (2005). Robustness of adaptive filtering methods in a cross-benchmark evaluation. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 98–105, Salvador, Brazil.