

## Proposal

### Evaluation of Cross-Language Spoken Document Retrieval Systems

*Marcello Federico*  
ITC-irst - Centro per la Ricerca Scientifica e Tecnologica  
38050 Povo (Trento) - Italy  
federico@itc.it

CLEF Workshop, 3-4 September 2001, Darmstadt

## Outline

- Motivations
- TREC's legacy
- Objectives
- Implementation

CLEF Workshop, 3-4 September 2001, Darmstadt

## Motivations

- Digital audio archives, and in general video and film archives, represent a very relevant sector in the digital library landscape.
- Automatic speech recognition technology has recently proven to be effective for indexing the audio trace of TV news programs.
- Positive experience by TREC with four evaluations of American monolingual Spoken Document Retrieval (SDR).

It would be interesting to extend the TREC experience to other European languages and to the cross-language IR framework.

CLEF Workshop, 3-4 September 2001, Darmstadt

## TREC's Legacy

TREC has organized four SDR evaluations, during years 1997-2000, which showed (to me) that:

- IR performance does not significantly worsen when error prone automatic transcripts are used in place of perfect hand made ones;
- few labs were able or interested in performing the complete track, i.e. automatic transcription of the test collection plus IR;
- unpleasant legal issues arose with the usage of the data.

CLEF Workshop, 3-4 September 2001, Darmstadt

## TREC's Legacy (cont'd)

The above considerations would suggest that :

- future evaluations on SDR could be successful even if only focused on IR issues, e.g. based on "provided" automatic transcripts;
- evaluation data should be distributed for free and under reasonable usage conditions.

CLEF Workshop, 3-4 September 2001, Darmstadt

## TREC's Legacy (cont'd)

TREC SDR evaluations have left an important legacy of data and tools (available on the Internet under mild conditions) which could be exploited for future evaluations, i.e.:

- automatic transcripts of 570 hours of American broadcast news;
- a significant set of queries with relevance assessments;
- evaluation software for SDR.

CLEF Workshop, 3-4 September 2001, Darmstadt

### Objectives

The here proposed evaluation has the following short and medium term objectives:

- to evaluate effectiveness of current CLIR systems, usually tested on clean-texts, on noisy automatic transcripts of spoken docs;
- to invite participants from the CLIR community, with no direct interest in ASR technology, to work on problems specific to SDR, e.g. retrieval with unknown story boundaries.

CLEF Workshop, 3-4 September 2001, Darmstadt

### Implementation

A first experimental evaluation on cross-language SDR could be organized in parallel with CLEF 2002, as follows:

- use TREC SDR 2000 automatic transcripts as the target collection;
- translate SDR 2000 queries into several European languages;
- use the available relevant assessment.

Development data could be prepared as well, e.g. from SDR 1999 data.

CLEF Workshop, 3-4 September 2001, Darmstadt

### Implementation (cont'd)

Some open issues:

- a. literal vs. task-oriented translations, e.g. news on demand service;
- b. single vs. multiple translations, e.g. each query by 5-10 people;
- c. written vs. spoken queries, e.g. recorded from the telephone;
- d. known story boundaries cond. vs. unknown story boundaries cond.;
- e. others ?

CLEF Workshop, 3-4 September 2001, Darmstadt

### Thanks to

- John Garofalo (NIST, USA)
- Donna Harman (NIST, USA)
- Gareth Jones (U. Essex, UK)
- Doug Oard (U. Maryland, USA)
- Carol Peters (CNR Pisa, Italy)

for providing useful information, comments, and ideas!

CLEF Workshop, 3-4 September 2001, Darmstadt