

Outline of a Multimedia Task

Franciska de Jong

University of Twente

September 15, 2004

Proposal

- Corpus development for heterogeneous content:
 - Domain: news
 - Formats: text (newspaper articles)
A/V (television news broadcasts)
- Task development:
 - initially strong links with existing tasks
 - increasing complexity for future CLEFs

Multi/Cross-lingual aspects

- Assumption: choice for news implies choice for a comparable corpus (many topics are relevant in many countries/linguistic regions) that can be searched from any language present
- General added value for video search: objects/ events in the footage not mentioned in the commentary in language x can be found via language y
- Additional features (long term perspective):
 - for some tasks video metadata can serve as interlingual representation

Motivation

- Continued focus on known task: news item retrieval
- Content acquisition allows shift to more recent content (e.g., year 2001) which may be attractive for new languages
- Complementarity to TRECVID: explore the contribution of textual material to user needs for multimedia search
- Variation wrt existing CLEF-tasks can be sought in various directions:
 - Cross-media search (find format x, with query in format y)
 - Multi-document retrieval (clustering?) combined with reasoning to decide on redundancy/complementarity
 - Multilingual/multidoc retrieval as basis for summarization
 - QA for multimedia

Task description

Suggestions (in arbitrary order):

- 1a. Given a broadcast news item+transcript in language x, find newspaper articles that give background information in language x and y
- 1b. Truly cross-media: as 1a, but without use of transcripts/subtitles
- 1b. As 1a/b, but asking for varying opinions on an issue (too difficult to judge?)
- 2a. Given a newspaper item, find corresponding video fragments
- 2b. As 2a, but with additional constraints on the occurrence of shots of specific faces, buildings, etc.

Documents/information needs/topic development

- Clearance for Dutch content is available for :
 - Public broadcast news shows for 2001
 - Corpus of subtitles (since 2000)
 - Corpus of autocue files (since 2001)
- Relevant information needs
 - Production of documentaries
 - Historical research (9/11, development of cattle diseases, child murder dossiers, ...)
 - Business intelligence (profile watch)
- Topic development
 - Depending on availability of corpora for other collections and task chosen

Assessments/Evaluation Measures

- Requires more thought.
- Parallel with hard track/passage retrieval?

Questions/Info

- Get in touch with Franciska de Jong
- f.m.g.dejong@ewi.utwente.nl