Presentation of AMIS

K. Smaïli
AMIS: Access Multilingual Information opinionS

- Starting date: December 2015
- Duration: 36 months
- The consortium is composed of partners from three countries:
  - France: University of Lorraine (LORIA), University of Avignon (LIA)
  - Poland: University of Science and Technology Kraków (AGH)
  - Spain: University of DEUSTO (Bilbao)
AGH University of Science and Technology, Kraków - Poland

AGH has skills on video content summarization

AGH leads several WP:

- Definition of the requirements and data video collection
- Video summarization and video content analysis
- Automatic Evaluation of the different components
- Dissemination
DEUSTO University Bilbao Spain

DEUSTO has skills in designing protocols of evaluation for people with special needs.

DEUSTO composed by psychologists, engineers and linguists is in charge of:

- End-user Evaluation
- Collecting social network data
- Protocol of tests and evaluation
LIA of University of Avignon (UA) has skills on speech, audio and language processing and, more specifically, on automatic summarization (text and audio).

LIA is in charge of:

- Text and audio Summarization
- Coverage of an event on social network
University of Lorraine - LORIA France

LORIA - university of Lorraine has skills on speech recognition and machine translation

LORIA is the coordinator of AMIS and is responsible of few WP:

- Speech Recognition (MULTISPEECH)
- Machine Translation and Opinion mining (SMarT)
- Overlaid text extraction (QGAR)
The key challenge and potential impact

- With the growth of information on internet, a new issue arises: How to access to a maximum of information?
- High educated people do not speak more than two or three languages, while the majority speaks only one, which makes this huge amount of information inaccessible.
- How to make the main idea presented in a video in a foreign language accessible and easy to understand by everyone?
- Accessing to information in foreign languages would permit to access to the other side of a story.
- Due to political, socio-cultural or religion reasons, divergence of opinions may exist within two medias from two different sources.
Main objectives

Objective 1: Understanding the main idea of a video by summarizing
- Input: A video in Arabic or French
- Output: A text and speech Synthesis in English

Objective 2: Cross-lingual opinion Analysis
- Input: Arabic video will be compared to French or English video
- Output: A review concerning the degree of divergence between the two videos
Objective 1: Understanding by summarizing

The global architecture allowing to reach the first objective necessitates several components which collaborate.
Available data and components necessary for the first version

Figure: Rate of realization
The video Database

Data on conflicting topics for studying the cross-lingual opinions in Arabic, English and French.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Arabic channels</th>
<th>English channels</th>
<th>French channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syria</td>
<td>Real Madrid - FC Barcelona</td>
<td>Animal rights</td>
<td></td>
</tr>
<tr>
<td>Women’s rights</td>
<td>Homosexual marriage</td>
<td>Drug liberalization</td>
<td></td>
</tr>
<tr>
<td>Death sentence</td>
<td>Occupied territories</td>
<td>Trump</td>
<td></td>
</tr>
</tbody>
</table>

Arabic channels

English channels

French channels

AMIS

Presentation of AMIS

March 21, 2017
Video Downloader

- We developed a toolkit to download the necessary videos

- The user can choose the channel, a time interval and the keywords for the queries.
Some statistics

- 100 hours of videos have been collected for each of the three languages.
- For each video in a language, videos in the 2 other languages correspond to the same topic.
Arabic Automatic Speech Recognition System

Development of an Arabic automatic speech recognition system with the following data:

**Acoustic corpus (hours)**

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nemlar</td>
<td>33</td>
<td>3</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>NetDC</td>
<td>19</td>
<td>3</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>52</strong></td>
<td><strong>6</strong></td>
<td><strong>5</strong></td>
<td><strong>63</strong></td>
</tr>
</tbody>
</table>

**Text corpus**

- GigaWord corpus was collected from 9 sources of information with a total of $10^9$ words.
- 315K words from the transcription of the acoustic training data
- A vocabulary of 95K words with an average of 5.07 pronunciations for each entry
Evaluation of the Arabic ASR

<table>
<thead>
<tr>
<th>Model</th>
<th>DEV</th>
<th>TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>13.54</td>
<td>14.42</td>
</tr>
<tr>
<td>Correct W2Vec</td>
<td>13.03</td>
<td>14.14</td>
</tr>
<tr>
<td>Rescoring</td>
<td>13.07</td>
<td>14.02</td>
</tr>
<tr>
<td>Rescoring + Correct W2Vec</td>
<td>12.26</td>
<td>13.45</td>
</tr>
</tbody>
</table>

- Tuning and Test on 31K words.
- Lattice rescoring with a 4 gram LM.
- Correction of words containing $\$ \text{(Word2Vec - CBOW)}$
Testing the ASR on the video database

Automatic transcriptions (Arabic English and French) have been performed on 300 hours on the collected videos thanks to the 3 developed ASR systems.
The data necessary for AMIS have been collected:
- Videos
- Social Network data
- Text data have been harvested from the Web completed with bought data from LDC

The components of the architecture are under progress and will be ready for M18.

The next step consists in improving the synergy between the different components.