Enabling the IIHS Vision, Part 1

Brandon Muramatsu
Andrew McKinney
Peter Wilkins—Our colleague at MIT at 0° C

January 2010
2 Demos For January 2010

- **SpokenMedia**
  - Video/audio transcription, enabling translation
  - Process and tools
  - “Access to high-quality learning must be open to all”

- **Open IIHS Experience**
  - Course/activity design; student interaction
  - “Make curriculum openly available”
“The IIHS Website is our commitment to a different way of looking at things.”

– Aromar Revi
5 January 2010
“The Institution will fail or scale based on language.”

– Aromar Revi
5 January 2010
Our Goals with this Demo

- Demonstrate transcripts and translations of IIHS videos
- Describe the process and our experiences
  - Transcribe -> Edit -> Translate -> Present
Four step process (simple)
First we used the SpokenMedia to do automatic transcription.
Next we did hand edit and translation steps.
Then we created a player for the presentation of the video and transcripts…
Challenges of Planning, Critical Challenges facing Cities of Global South, IIHS, and Challenges within the IIHS Curriculum Framework

I think one central challenge of planning is legitimacy of government as a key institution that should be involved in bringing the change. The legitimacy of the government is under attack in many different parts of the world so that it has to be re-established as a major planning institution. I think the second challenge for planning at least in the US or the west is that the market is...
First we used the SpokenMedia to do automatic transcription.
Lecture Transcription

• Jim Glass and his group have years of research experience for spoken languages

• Lectures are a different type of spoken language
  • Much of the speech recognition research has focused on real time transcription of news broadcasts, or interactive voice response systems (telephone)
  • Broadcast news has something like 300 unique words in an hour long broadcast
  • Broadcast news is well structured, prepared copy (in studio via teleprompters), clear transitions between speakers, etc.
  • Lectures are conversational and spontaneous
  • Can use highly specialized vocabularies, engineering, physical sciences, mathematics
Spoken Lecture Project

• Processor, browser, workflow
• Prototyped with lecture & seminar video
  – MIT OCW (~300 hours, lectures)
  – MIT World (~80 hours, seminar speakers)

Supported with iCampus MIT/Microsoft Alliance funding
How does it work?

• Audio
  • System only needs audio (waveform), extracts from video
• Domain Model (base is generic domain model)
  • System needs to know what words it can expect to find in the audio
  • Syllabus, lecture notes, index from text book, research papers
  • Build library of domains
  • Separate sub-process for text for domain model
• Speaker model (base is generic speaker model)
  • If multiple lectures by the same author, best to create a speaker model
  • Separate sub-process for speaker model
• Process—with audio, domain and speaker models
• Output
  • Time coded transcript (standard formats)
  • Links media and transcript
• Applications
  • Search/retrieval
  • Player
We only used part of the process due to time constraints.

Audio separation, speech processing, time-coded transcript, and then presentation through a SpokenMedia player.
Next we did hand edit and translation steps.
For this demo, we did computer-based automatic transcription, sent a file to IIHS for “editing” that consisted of performing a hand transcription (due to the format we sent, and the low accuracy of the automatic transcription in this case), a time alignment (though I actually feel that it's “off” or “slow”) and then a hand translation by IIHS.

<table>
<thead>
<tr>
<th>Automatic Transcription</th>
<th>Hand Transcription</th>
<th>Time Adjusted</th>
<th>Translated Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>I</td>
<td>मेरे बयान मे</td>
</tr>
<tr>
<td>think</td>
<td>think</td>
<td>think</td>
<td>की एक मुख्य पृष्ठी है</td>
</tr>
<tr>
<td>once</td>
<td>one</td>
<td>one</td>
<td>नेपाल की एक मुख्य पृष्ठी है</td>
</tr>
<tr>
<td>and</td>
<td>central</td>
<td></td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>challenge</td>
<td>central</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>challenger</td>
<td>planning</td>
<td>challenge of</td>
<td></td>
</tr>
<tr>
<td>planning</td>
<td>is</td>
<td>planning</td>
<td></td>
</tr>
<tr>
<td>nice</td>
<td>legitimacy</td>
<td>is</td>
<td></td>
</tr>
<tr>
<td>legitimacy</td>
<td>of</td>
<td>legitimacy of</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>government</td>
<td>government</td>
<td>सरकार की एक गणतंत्र शक्ति के रूप में केवल</td>
</tr>
<tr>
<td>government</td>
<td>as</td>
<td>as</td>
<td></td>
</tr>
</tbody>
</table>

Edit & Translate: Accuracy
Recognizer Accuracy

• Base accuracy is approximately 50% (generic domain and speaker models)
• Increase accuracy with speaker model up to 80-85%, and specific domain model
  • This approach is good for courses with multiple lectures by the same speaker
  • Domain models get more useful as more relevant text documents are indexed (keyword/noun phrase extraction)
• Initial results indicate that doing one 99% accurate (by hand/manual) transcript can help immensely for additional lectures by the same speaker
  • Better use of limited resources
• Search accuracy is closer to 90%, searches tend to be for unique words which the processor is better at recognizing
Then we created a player for the presentation of the video and transcripts…
I think one central challenge of planning is legitimacy of government as a key institution that should be involved in bringing the change. The legitimacy of the government is under attack in many different parts of the world so that it has to be reestablished as a major planning institution. I think the second challenge for planning at least in the US or the west is that the market is...
Check it out for yourself

- Demo site:
  
  http://oki-dev.mit.edu/spokenmedia

  all the videos from IIHS website…it's not just Bish!
Thank You!

Brandon Muramatsu, mura@mit.edu
Andrew McKinney, mckinney@mit.edu