Collaborative Caption Editing System

IBM Research – Tokyo
IBM Japan
Enhancing the Quality of Caption Editing System

Speaker: Ali Sobhi
sobhi@us.ibm.com

Authors: Ali Sobhi
Reiko Nagatsuma
Takashi Saitoh

Session ID: WEB-027

Wednesday February 29, 2012
3:10 PM – 4:10 PM

Grand Hyatt – Elizabeth F
1: Auto Speech Recognition

Too hard to understand if there is no caption

2: Transcribe all text

<<original>>

abc...

<<after recognition>>

xxxxx...

Language model problem

Hard work to transcribe for 1 person

• Hard to set up the equipment
• Not too easy to input
• Loss of concentration
Upload the files

CCES Server

separate the task

assign short task

merge all tasks

finally

watch video with caption
Overview: Collaborative Caption Editing System (CCES)

Flow of creating captions:
- Register contents
- Create Caption
- Final adjustments
- Captioned content

Pre-processing:
- Segmentation video or audio data

Caption Editing:
- Create caption for each segment by novice users

Confirmation:
- Confirm and fix the caption data
- Confirmation by experienced users

Adjustment:
- Format the caption text
- Final adjustment by expert

For persons with hearing impairments

Edit captions using Web browsers.

Creating captions effectively by crowd-sourcing

- Novice users can contribute
  - Accumulate knowhow from work
  - Improve their skills through collaboration with experts
- Experts can concentrate on confirming and adjusting
Confirm the caption immediately

Reporting function

Repeat the phrase automatically
User Pilot: IBM Centennial Volunteer Program

- Target content: Lecture video from Waseda University.

- Total participants: 143 (Nov. 12th – Dec. 23rd)
  - 5 participants are working for confirmation task

- Participants who login more than 3 days: 80
  - Average work items/login: 16 items (16 items x ~30 seconds each = ~8 minutes)
  - Average operation time: 45 min

- Active time
  - Morning: before 9:00 am
  - Lunch time: noon - 1:00 pm
  - Evening: after 7:00 pm - 11:00 pm

and holiday, Saturday
User Pilot: IBM Centennial Volunteer Program

- Average operation time (educational content)
  - Input captions: 163 sec / each content (30 sec)
  - Confirmation: 82 sec / each content (30 sec)

- Real operation time (25 min content)
  - Input captions: 151 min
  - Confirmation: 66 min
  - Adjustment: 48 min
  - Total: 257 min (~4.2 hours - an expert spends ~8 hours to caption 15 min. of content)

- Comments from users are the following:
  - I am pretty impressed with this...
  - Response of the system is pretty good, and user interface is very useful.
  - It is good for novice PC user because:
    - No need for special equipment (only browser)
    - Repeat one phrase automatically
    - Available to input easily with short segmentation
Learning effectiveness (sample data of 3 users)

- Person A: 231 sec
- Person B: 210 sec
- Person C: 194 sec