Homework 3: Dialog

◆ Part 1
  • Call TellMe and get two sets of driving directions
  • Call CMU’s Let’s Go
  • Call Amtrak

◆ Part 2
  • Build your own pizza ordering systems
  • Register with Tell Me Studio
  • Use VoiceXML to build a system

◆ Results are due 17th November 3:30pm
Speech Processing 15-492/18-492

Spoken Dialog Systems
Beyond VoiceXML:
the *Olympus Spoken Dialog Framework*
Write (several) vxml “pages” and resources

- Your dialog application control
- Provide grammar for understanding
- Define what your system says

Generally just use provided ASR/TTS

Great for basic form-filling applications

- What if your application can’t be made into a form-filling one?
Olympus Spoken Dialog Framework

- **A general dialog system architecture**
- **Modular, open source framework**
  - Provides components needed to build SDS
    - ASR/TTS, Language Understanding/Generation, Dialog Management, etc.
  - Can replace components with other options
    - e.g., use a different ASR engine
  - Tied together via Galaxy message-passing communication infrastructure

- [http://wiki.speech.cs.cmu.edu/olympus](http://wiki.speech.cs.cmu.edu/olympus)
Example Olympus Systems

- **Let’s Go! (bus information)**
  - **TeamTalk (robot interaction)**
    - [http://wiki.speech.cs.cmu.edu/teamtalk/](http://wiki.speech.cs.cmu.edu/teamtalk/)
  - **Vera**
    - [http://www.speech.cs.cmu.edu/~awb/vera.wmv](http://www.speech.cs.cmu.edu/~awb/vera.wmv)
- **Many others**
Organization of Olympus Systems

- **Core components**
  - Generic, useful in multiple different systems

- **Application components**
  - System-specific, useful for a single application
Olympus Core Directory Structure

- Source code for all system-independent Galaxy servers
- Scripts to compile Olympus
- External dependencies
- Tools and scripts for LM training, log mining…
- Binaries
- Generic system configuration includes
- System-independent resources (ASR and VAD acoustic models)
System Directory Structure

- Source code for system-specific Galaxy servers
- System configurations
- Dialog logs
- System-specific binaries
- System documentation
- System-specific resources (grammars, language models, ...)

Diagram:
```
- Tutorial2
  - .svn
  - Agents
  - Bin
  - Configurations
  - Documents
  - logs
  - Resources
```
Pipeline Architecture in Olympus

Recognition AUDIOSERVER

Interpretation PHOENIX

Interaction Mgr APOLLO

Dialog Manager RAVENCLAW

Synthesis KALLIOPE

Generation ROSETTA

Recog. Engine (SPHINX)

Knowledge Source

Backend

Phone / Desktop
The Olympus Architecture

- Slot-filling templates
- Allows for random variations
- Grammar-based
- Robust parser
- Plan-based
- Interface between real world and dialog manager
- Manages timing/turn-taking
- Allows multiple recognition engines (SAPI, Swift, Flite)
- Does playback
- Interface to external engines (SAPI, Swift, Flite)
- Suitable for channel/domain
- Allows multiple recognition engines
- Controls dialog • Plan-based
- Slot-filling templates • Allows for random variations
- Robust parser • Interface to external engines
- Interface between real world and dialog manager
- Manages timing/turn-taking
- Allows multiple recognition engines (SAPI, Swift, Flite)
- Does playback
Grammar

- **Used for two things:**
  - Parsing
  - ASR language model if one isn’t available
- **The Phoenix Parser**
  - Context-Free Grammar
  - Robust parser
Phoenix Parser / Grammar

**CFG Grammar**
- Manually-generated domain-specific grammar rules
- Reusable, generic sub-grammars
  - [Yes], [No], [Number], [DateTime], [Help], [Repeat], [Suspend], etc…

```plaintext
DO YOU HAVE SOMETHING A BIT LARGER?
[NeedRoom] (    
    [_i_want] (DO YOU HAVE SOMETHING) )
[RoomSizeSpec] (    
    [room_size_spec] (    
        [rss_larger] (LARGER) ))
```

**Parses all incoming hypotheses and passes all parses along…**

```plaintext
[room_size_spec]
    ([rss_large])
    ([rss_small])
    ([rss_larger])
    ([rss_smaller])
    ([rss_smallest])
    ([rss_largest])
;
[rss_large]
    (large)
    (big)
    (huge)
;
[rss_larger]
    (*the larger)
    (*the bigger)
    (too small)
;
[rss_largest]
    (*the largest)
    (*the biggest)
;
[rss_small]
    (small)
    (little)
;
```
Example Phoenix Grammar

[Place]
  (carnegie mellon university)
  (downtown)
  (robinson town center)
  (the airport)
  (south hills junction)
  (mount oliver)
  (the south side)
  (oakland)
  (bloomfield)
  (polish hill)
  (the strip district)
  (the north side)

;
Confidence Annotation - Helios

- **Builds accurate confidence scores using features from 3 sources of knowledge:**
  - Speech recognition
  - Language understanding
  - Dialog management

- **Selects hypothesis with maximum confidence score**