What is a dialogue system?

- McTear (2002 p.92) (Spoken Dialogue System)
  - "Computer systems that use spoken language to interact with users to accomplish a task"
  - "Computer systems with which humans interact on a turn-by-turn basis and in which spoken natural language plays an important part in the communication" (citing Fraser, 1997)

- Flycht-Eriksson (2001) (Dialogue System)
  - "a computer system that interacts with users utilizing connected natural language dialogue, where the use of natural language need not consist of predefined commands"

What is a dialogue system?

- Modalities
  - text or speech, and possibly any number of accompanying modalities

- Complexity of interaction
  - adequate response to user initiative,
  - extended conversation,
  - topical coherence,
  - goal-directedness,
  - meta-communication,
  - multi-task, multi-domain, multi-device, multi-party, ...

Common system categories

- Interactive voice response systems
- Question-answering systems
- Interactive guides
- Natural-language interfaces
  - to databases, information seeking
  - to expert systems (problem-solving or collaborative systems)
- Spoken dialogue systems
- Conversational agents
- Conversational systems / Chatbots
- Dialogue support (translation) systems
Interactive guides

Question-answering systems

History in brief

- In the beginning …
  - The Turing Test
  - Conversational partners
    - ELIZA (1966), PARRY (1971), …
  - Database interfaces
    - BASEBALL (1963),
    - LUNAR (1972)
  - Commercial systems: INTELLECT, Q&A
History in brief

- Natural-language interfaces 1975-1990
  - Syntax-driven semantic interpretation
  - LUNAR (Woods et al. 1972, Woods1977)
  - CHAT-80,
  - Definite-Clause Grammars (Warren & Pereira, 1982)
  - The Core Language Engine (1992)

History in brief

- AI research 1970-1990
  - SHRDLU (1972)
  - varied types of input (procedural semantics),
  - knowledge-based (blocks world, language)
  - Groce (1977)
  - dialogue structure as based in task structure
  - Allen, Cohen, Perrault, Carberry, Traum, ...
    - Agents act because they have beliefs, desires and intentions,
    - Dialogues are goal-directed,
    - Speech acts contribute to goal fulfillment

History in brief

- Spoken dialogue systems (1990- )
  - Dialogue systems as applications of speech technology
    (ASR, TTS)
  - Emphasis on system building
    - GALAXY (MIT, 1994)
    - Circuit-Fix-It Shop (1994)
    - Philips train timetable information system (1995)
    - Waxholm (1996)
    - TRAINS
    - GoDiS
    - ...
History in brief

The shift to empiricism in CL/AI (1990-)
- Data collection and analysis
- Robust methods for understanding
- Design for usability
- Methods for system development
- Evaluation methods, and
- The deployment of commercial systems
- The launching of the Loebner Prize

The Loebner Prize

Multimodal systems (1980-)
- "Put that there" (Bolt, 1980)
- Multimodal input guidance
- NLMenu (Tennant, 1983)
- Multimodal input interpretation
- CUBRICON (Neal et al., 1989)
- XTRA (Wahlster, 1989)
- Multimodal generation
- COMET (Feiner & McKeown, 1991)
- WIP (Wahlster et al., 1993)
Merits of (practical) dialogue systems

- Natural interaction
- Allows for fragmentary input
- Clarifications
- Complex requests
- Contextual interpretation
- The language can be adapted to the user

Levels of Complexity

- **Strict Policy**
  - User can only specify information relating to current goal/subgoal
  - Context is easier to determine

- **Free Policy**
  - Handle unintended requests or requests that deviate from the task
  - Context more difficult to determine
  - Can lead to confusion/errors

Initiative

- **System-initiative**: system always has control, user only responds to system questions
- **User-initiative**: user always has control, system passively answers user questions
- **Mixed-initiative**: control switches between system and user using fixed rules
- **Variable-initiative**: control switches between system and user dynamically based on participant roles, dialogue history, etc.
Components/Architecture

Knowledge sources

- Dialogue history
  - A record of the dialogue so far
- Task record
  - Representation of the information to be gathered
- World knowledge
  - General commonsense background information
- Domain model
  - Information for the specific domain
- Conversational model
  - Principles of conversational turn-taking
- User model
  - User-specific information

Dialogue management

- Focus handling
  - Contextual interpretation considering user input, dialogue history and domain
- Dialogue control
  - Task control
  - Sub dialogue control
  - Clarifications
Dialogue and task complexity (Allen)

<table>
<thead>
<tr>
<th>Technique used</th>
<th>Example Task</th>
<th>Dialogue phenomena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite state script</td>
<td>Operator phone calls</td>
<td>User answers questions</td>
</tr>
<tr>
<td>Frame based models</td>
<td>Time-table information</td>
<td>User asks questions, simple clarifications by system</td>
</tr>
<tr>
<td>Sets of contexts</td>
<td>Travel booking</td>
<td>User asks questions, simple clarifications by system</td>
</tr>
<tr>
<td>Plan-based models</td>
<td>Design consultant</td>
<td>Dynamically generated topic structures, collaborative negotiation</td>
</tr>
<tr>
<td>Agent-based models</td>
<td>Disaster relief management</td>
<td>Different modalities</td>
</tr>
</tbody>
</table>

Finite State Dialogue Modeling

- System asks a series of questions that the user answers: "What number would you like to call?", "Is this a Linköping number?"
- Initiative always with the system.
- Context is fixed by the question being asked.

Finite State Dialogue Grammar

```
S / Greet
S / Qn
U / Ans
S / Bye
```

Frame Based Dialogue Modeling

- System acquire enough information in order to perform a specific action.
- Single fixed context
- Form filling where the form specifies all relevant information for an action:
  - Monitor the form for completion.
  - From user utterances extract relevant elements.
  - Use empty slots as triggers for questions to the user.
Menus, Forms and Links

S: Welcome to the Travel agency.
What do you want to do?
A: Book a trip.
S: Book a trip.
From where do you want to go?
A: From Stockholm
S: At what time?
A: Three o'clock
S: You want to travel from Stockholm at three o'clock.
A: Abort
S: Aborting. Welcome to the travel agency.

VoiceXML: Forms

```xml
<?xml version="1.0"?>
<vxml version="2.0" xml:lang="se">
<form>
  <field name="to">
    <prompt>Where do you want to go?</prompt>
    <option>stockholm</option>
    <option>göteborg</option>
    <option>malmö</option>
  </field>
  <field name="when" type="time">
    <prompt>At what time?</prompt>
  </field>
  <filled>
    <prompt>You want to travel to <value expr="to"/>
    at <value expr="when"/></prompt>
    <submit next="booking_to.vxml"/>
  </filled>
</form>
</vxml>
```

FIA: Form Interpretation Algorithm

- Go through each field in the form.
- Stop at the first field that is not filled.
- Play the prompt, start the speech recogniser.
- Fill the fields in the form that are specified by the matching grammar.
- If all fields are filled: Leave the form.
  - Otherwise: start again.
- Also possible to control the execution by scripting!