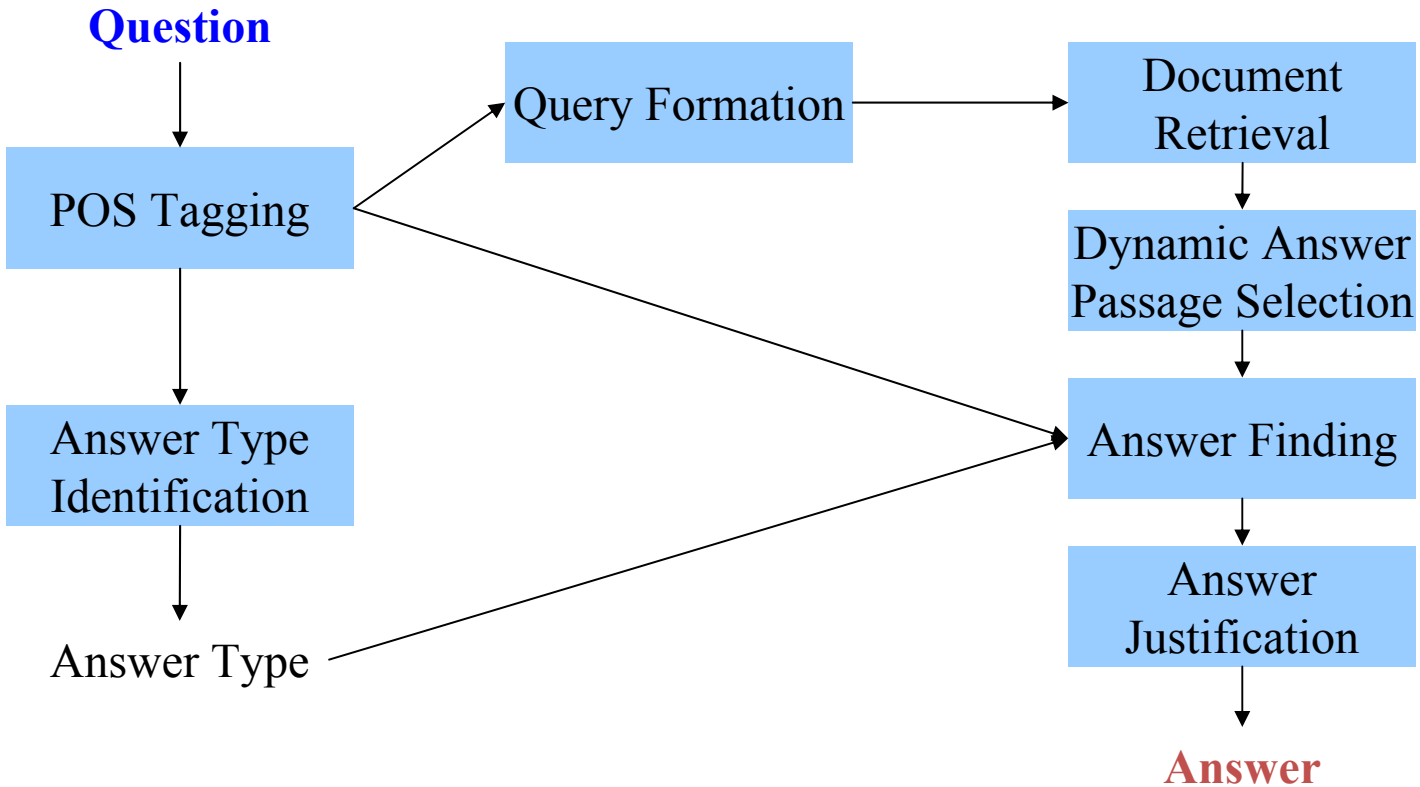
Current Challenges of SLU
(For Information Access Dialog Systems)

Gary Geunbae Lee, Ph.D., Professor
Dept. CSE, POSTECH

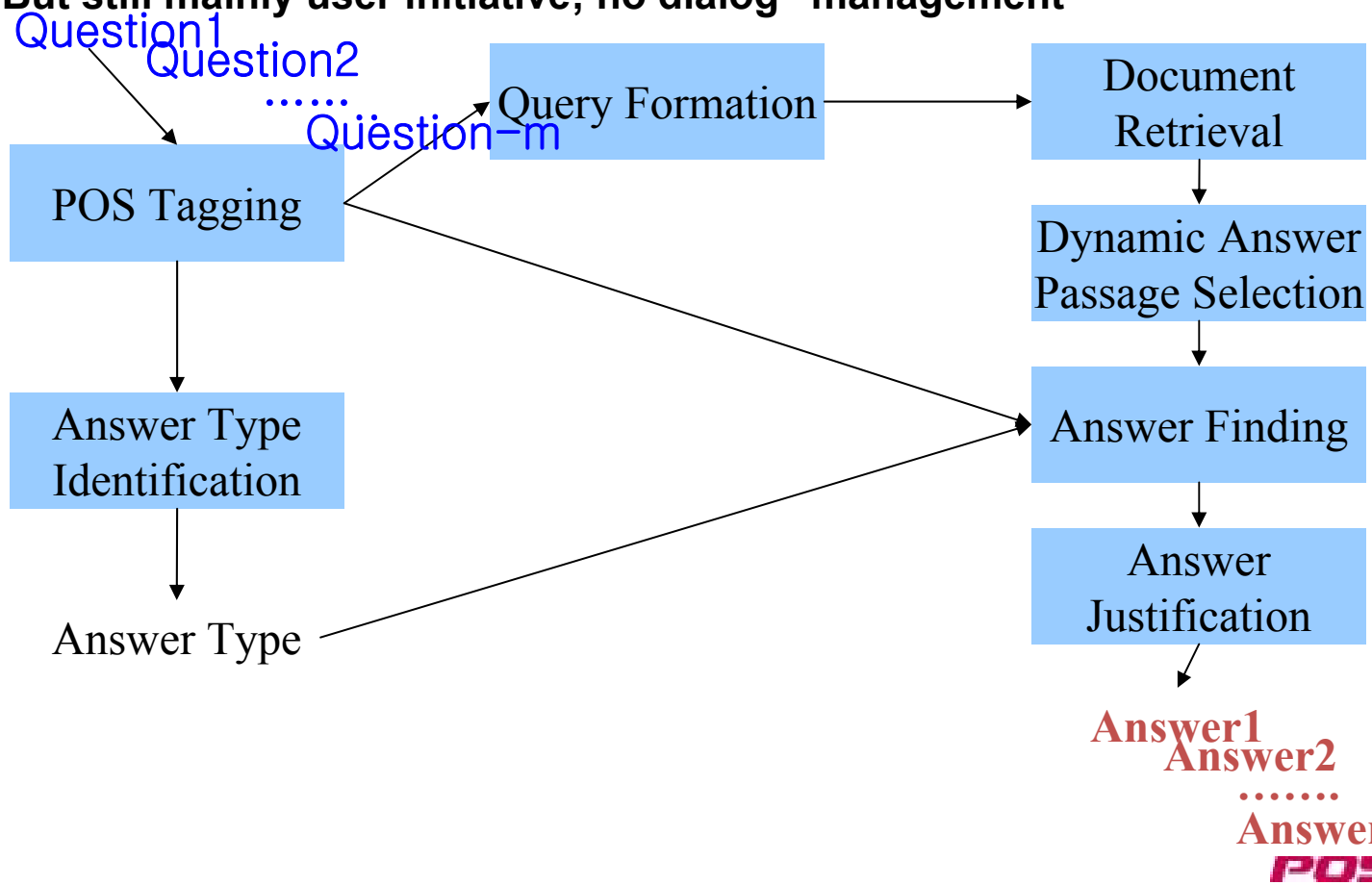
Conventional Question Answering

- Question Answering System
  - SiteQ [Lee et al. 2001; Lee and Lee, 2002]
  - Search answers, not documents (new form of IR)
  - One shot question answering

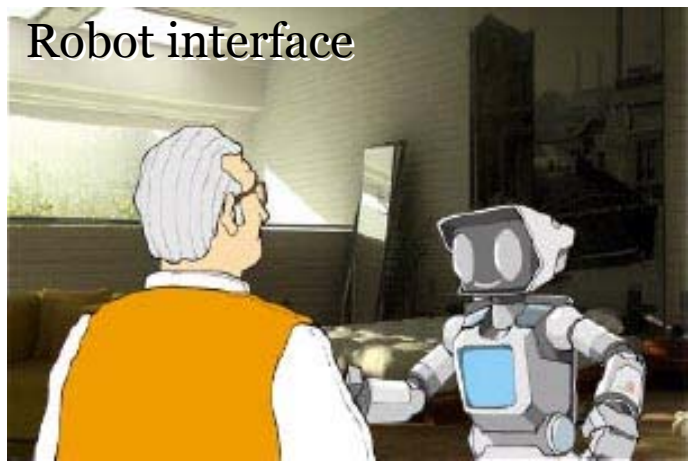


Interactive Question Answering

- **New challenges for Question Answering System [TREC ciQA; HLT-NAACL2006 workshop]**
  - **Series of related questions in a session / Interact with other people**
  - **Should handle anaphora, ellispses and other discourse related problems**
  - **But still mainly user initiative; no dialog "management"**

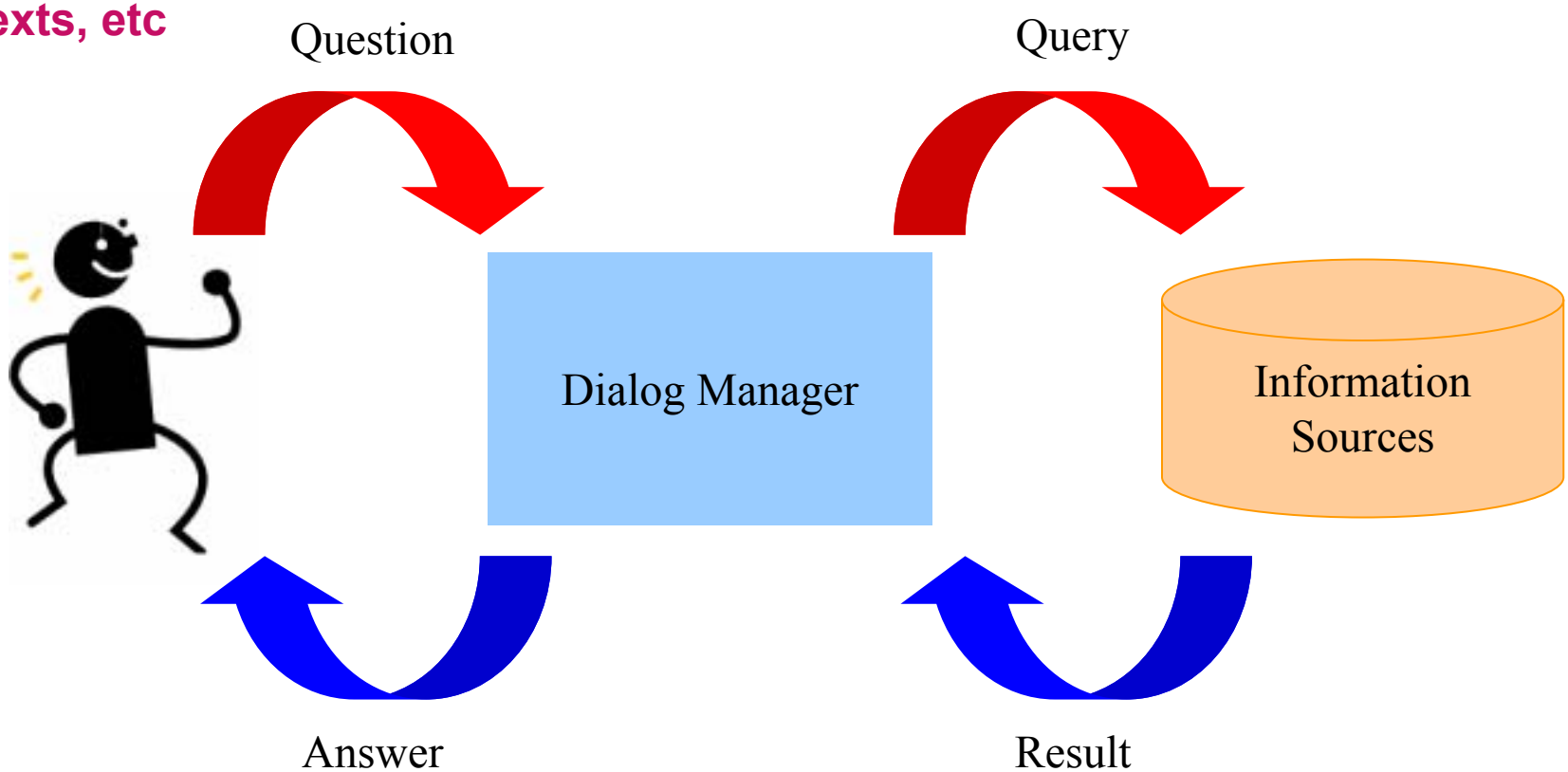


Ubiquitous Dialog Systems for everyday life



Information Access Dialog

- Combination of Ubiquitous Dialog system with Question Answering/Information Retrieval
- “Mixed Initiative Dialog Management” interacts with QA/IR
- Information access from RDB, website, FAQlists, semantic web, Free texts, etc

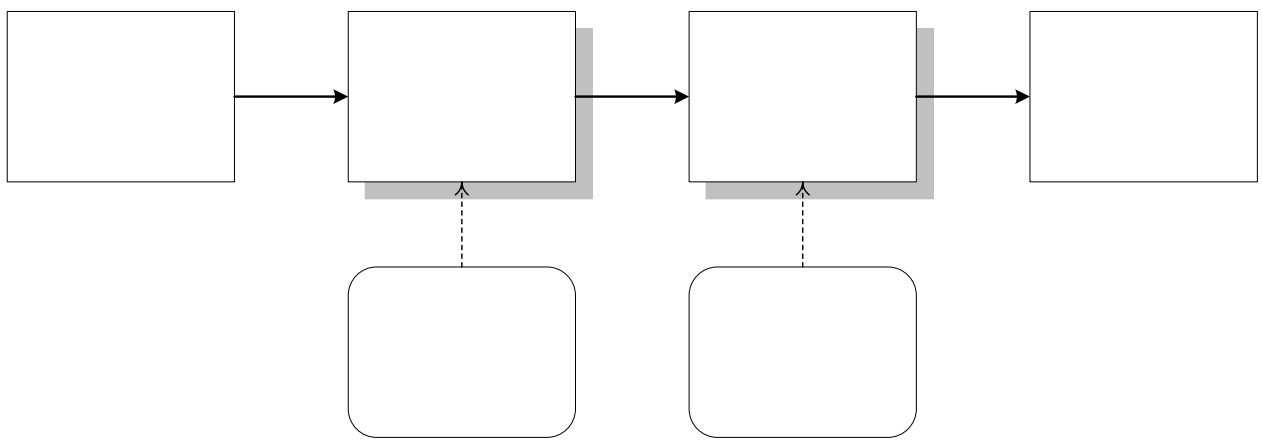


What about Data-driven SLU?

- SLU - map *natural language speech* to *frame structure* encoding of its *meanings* (meaning – DA, domain, named-entity, relations, etc)

```
<frame domain='ATIS'>
  <utt>Show me flights from Denver to New York
    on Nov. 18th</utt>
  <slot type='DA' name='Show_Flight'/>
  <slot type='NE' name='FROM.CITY'>Denver</slot>
  <slot type='NE' name='TO.CITY'>New York</slot>
  <slot type='NE' name='MONTH'>Nov.</slot>
  <slot type='NE' name='DAY_NUMBER'>18th</slot>
</frame>
```

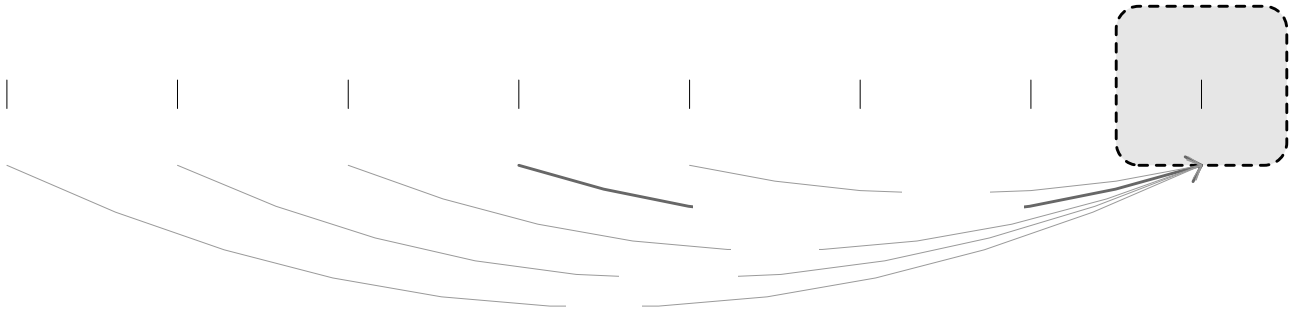
```
<frame domain='EPG'>
  <utt>I want to watch LOST</utt>
  <slot type='DA' name='Search_Program'/>
  <slot type='NE' name='PROGRAM'>LOST
  </slot>
</frame>
```



Problems and Challenges of Data-driven SLU for information access dialog system

- Long distance & non-local features
- More complex syntactic structures
- Combining with inferences and QA
- Automatic acquisition of Knowledge for Dialog
- Joint & efficient decoding → probabilistic graphical models
- Multimodal SLU (integration and reference)

Non-local/Long-distance Features



•Long-distance dependency and Trigger Features

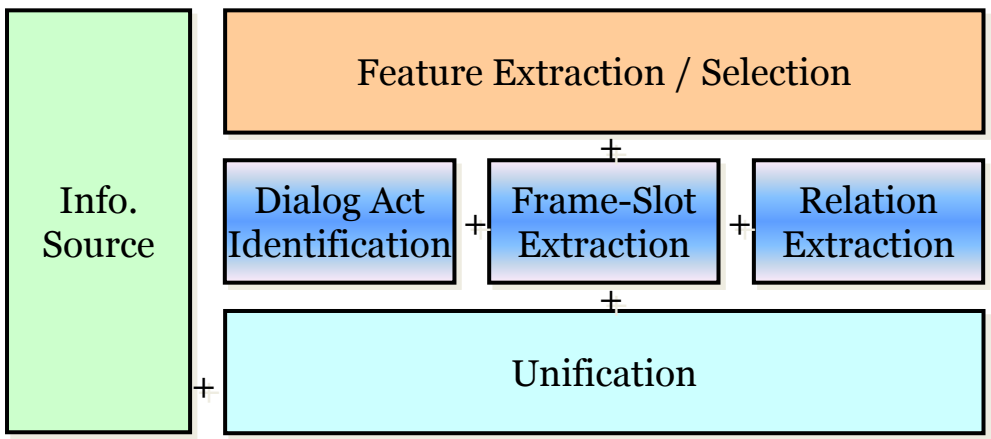
Algorithm 1 Trigger Selection

1: Initialize training data $\mathcal{D}$ with local features and a trigger set $t = \langle\rangle$
2: **while** t is increased **do**
3: Learn a ME classifier on $\mathcal{D}$: $\Lambda \leftarrow \text{TrainME}(\mathcal{D})$
4: Make candidates: $g \leftarrow \text{GenerateTriggers}(\mathcal{D}, \Lambda)$
5: Optimize μ: $\hat{\mu} \leftarrow \text{OptimizeGain}(\mathcal{D}, g, \Lambda)$
6: Select triggers: $g^* \leftarrow \text{SelectTrigger}(g, \hat{\mu})$
7: Update training data: $\mathcal{D} \leftarrow \text{UpdateData}(\mathcal{D}, g^*)$
8: Update a trigger set: $t \leftarrow t \cup g^*$
9: **end while**
10: **return** $\mathcal{D}, t$

•Outline of Trigger Selection Algorithm

More complex syntactic structure

- **Semantic Frame Extraction** (~ *Information Extraction Approach*)
  1) Dialog act / Main action Identification ~ ***Classification***
  2) Frame-Slot Object Extraction ~ ***Named Entity Recognition***
  3) Object-Attribute Attachment ~ ***Relation Extraction***
  – 1) + 2) + 3) ~ **Structure *Unification; Reference analysis***



Overall architecture for semantic analyzer

How to get to DisneyWorld?
Domain: Navigation
Dialog Act: WH-question
Main Action: Search
Object.Location.Destination=DisneyWorld

I like DisneyWorld.
Domain: Chat
Dialog Act: Statement
Main Action: Like
Object.Location=DisneyWorld

Examples of semantic frame structure

Inference and QA for SLU

user	When can I watch the Wayne Rooney's game on TV?

SLU

Wayne Rooney: Person Name

Generate query statement

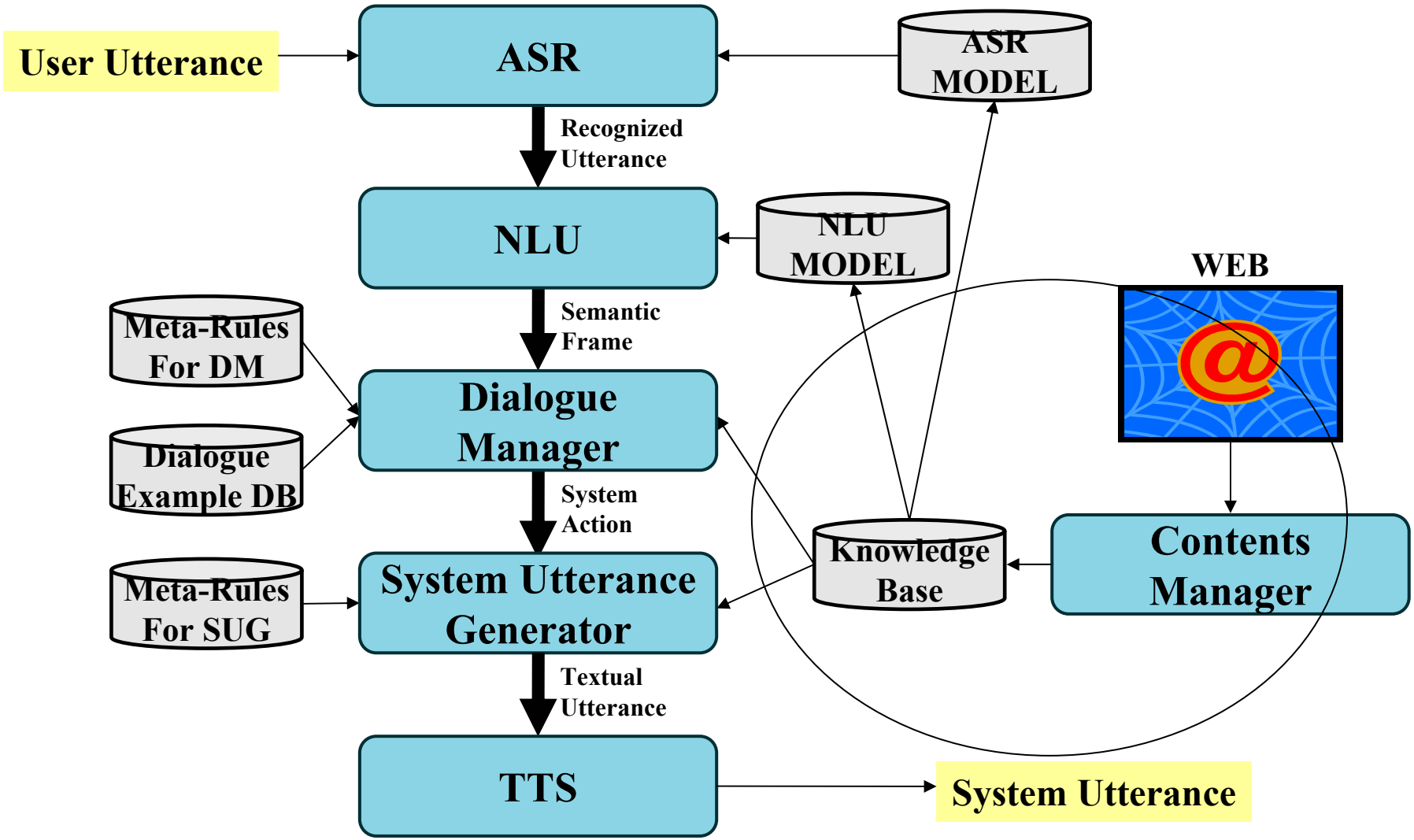
Inference Procedure	Query statement (in SPARQL)

Wayne Rooney is a football player.
Wayne Rooney is a member of Manchester United.
The opponent of Manchester United is Chelsea FC on that game.
	SELECT ?match
	WHERE {
	?match owl:hasMonth owl:Feb .
	?match owl:hasDay owl:d_3 .
	owl:Rooney owl:isMemberOf ?t .
	?match owl:hasTeam ?t .
	}

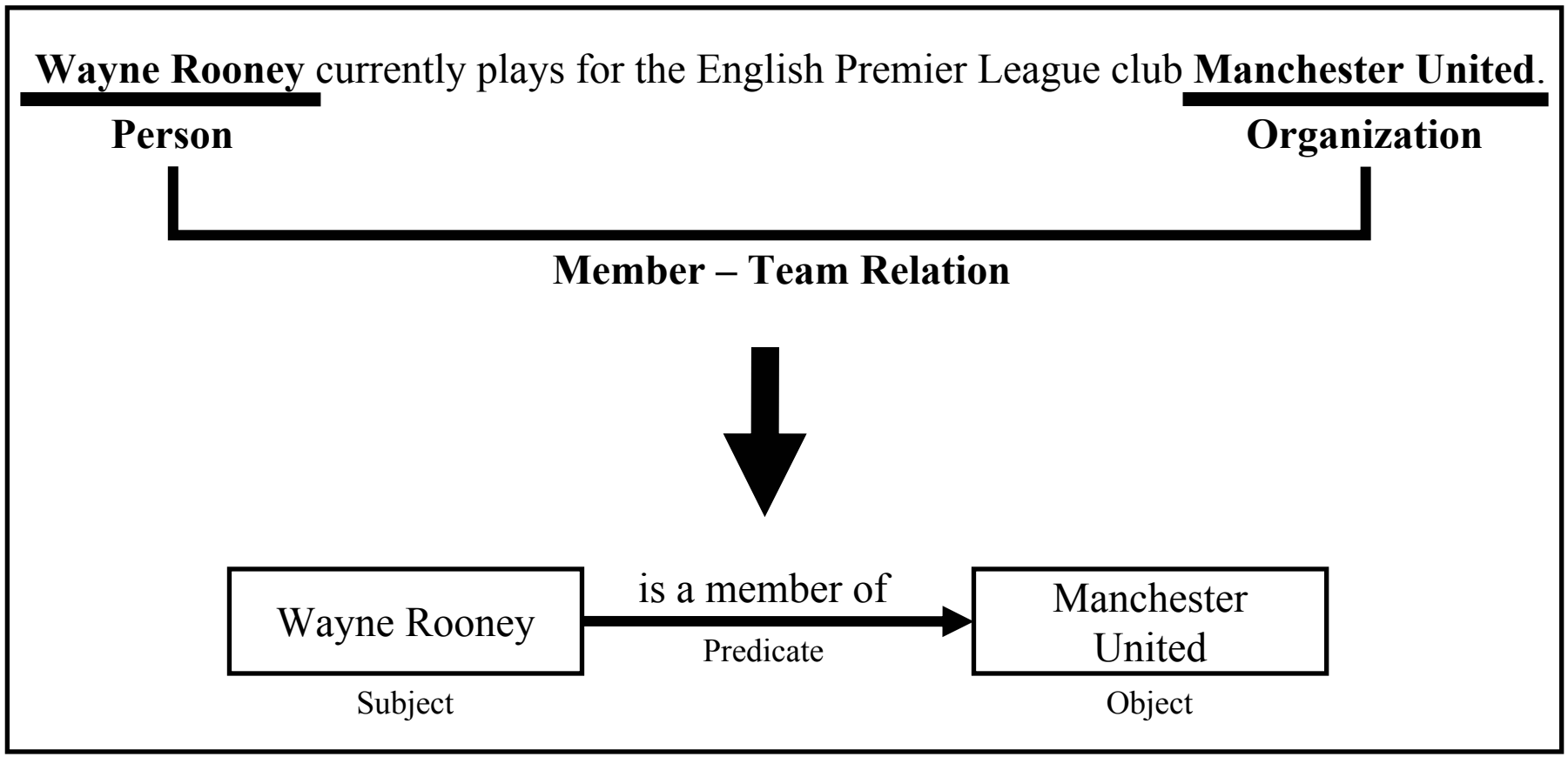
Query result

match	Manchester United vs. Chelsea FC at 21:00, 3rd, Feb
system	The match of Manchester United vs. Chelsea FC will be being telecasted at 21:00 on Saturday.

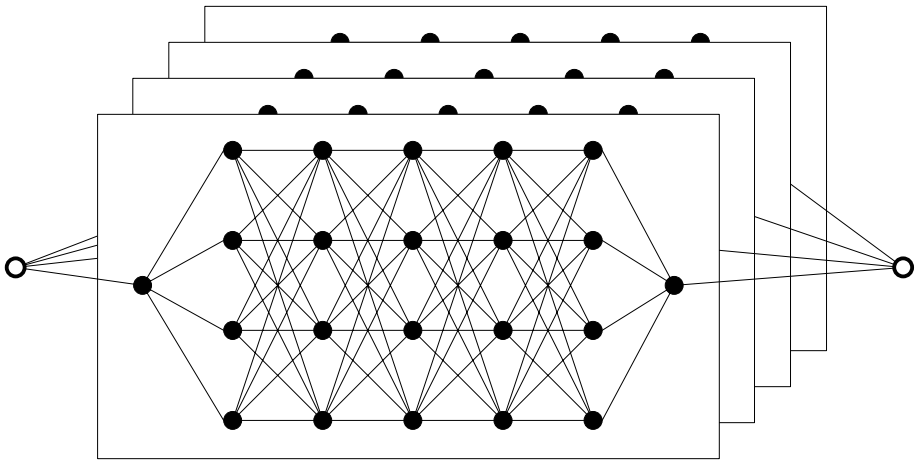
Automatic knowledge acquisition for Dialog systems



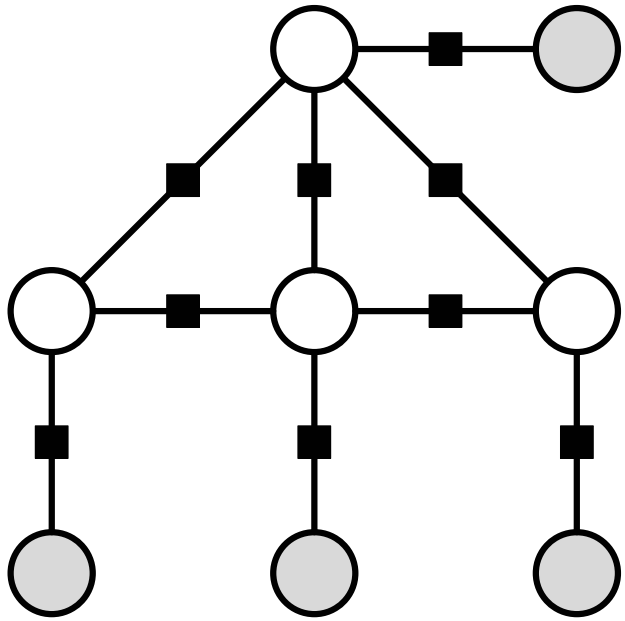
An Example of Content Feeding



Joint /Efficient Decoding



•Search Space



•Factor Graph of Triangular-chain CRF

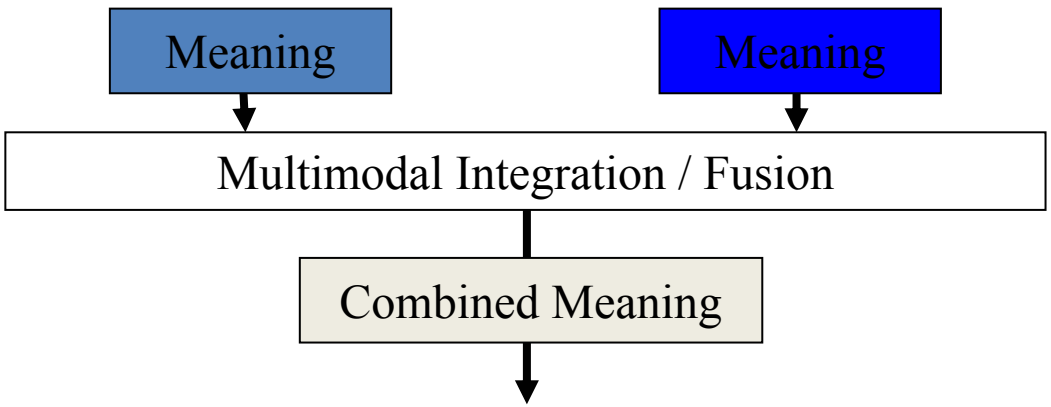
$$p(\mathbf{y}, z|\mathbf{x}; \Theta) = \frac{1}{Z(\mathbf{x})} \exp\left(\phi(\mathbf{x}, \mathbf{y}, z)\right)$$

$$\phi(\mathbf{x}, \mathbf{y}, z) = \sum_t \sum_k \{\lambda_k \underbrace{f_k(y_{t-1}, y_t, z)}_{\text{edge-transition}} + \mu_k \underbrace{g_k(y_t, \mathbf{x})}_{\text{NE-observation}}\} + \sum_k \nu_k \underbrace{h_k(z, \mathbf{x})}_{\text{DA-observation}}$$

$$f_k(y_{t-1}, y_t, z) = \underbrace{f_k^1(y_{t-1}, y_t)}_{\text{y,y dependency}} \cdot \underbrace{f_k^2(y_t, z)}_{\text{y,z dependency}}$$

Multimodal SLU & Reference analysis

- Combining information from multiple input modalities to understand user's intention and attention



  - Multimodal **reference resolution** is a special case of multimodal integration
    - Speech + pen gesture.
    - The anaphoric references are from discourse history or gesture