Provably Reliable QA Interfaces

http://cs.washington.edu/research/nli

Outline

I. Motivation
II. Reliable NLIs
III. Semantic tractability theory
IV. Implementation and experiments

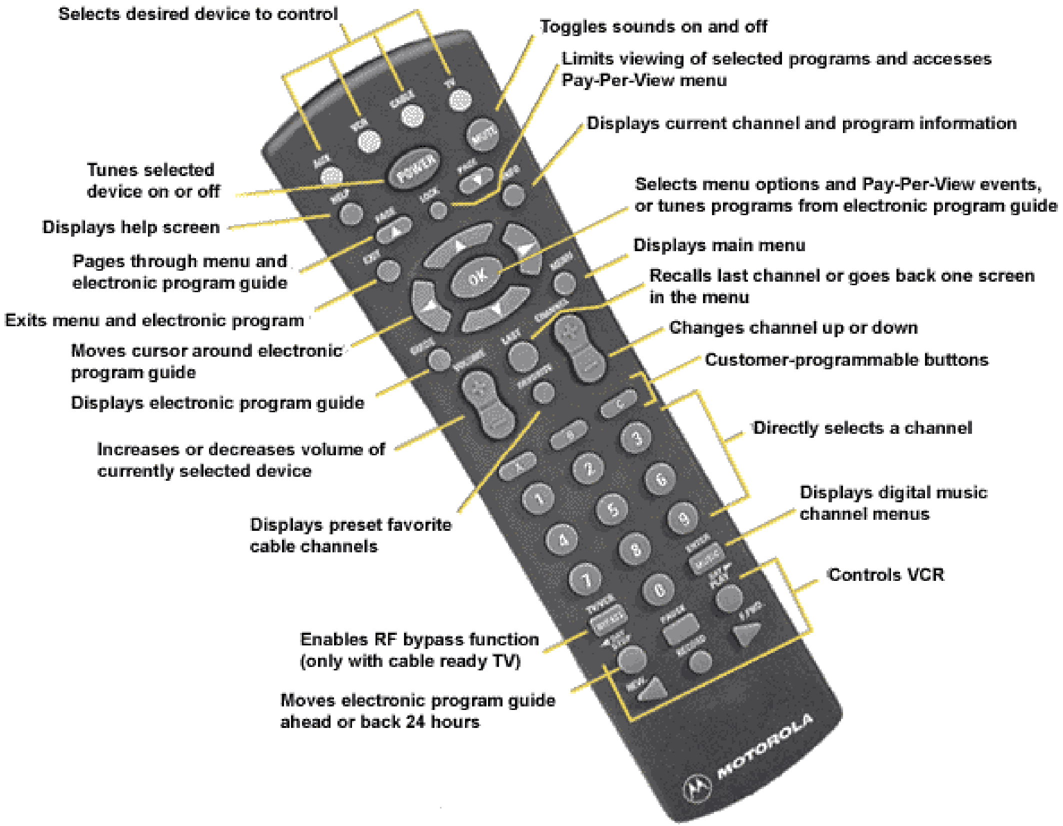
Joint work with Ana-Maria Popescu, Alex Yates, Henry Kautz, and Dan Weld.

I. The Dominant UI Paradigm

- Menu buttons
- Hyperlinks
- Remote controls

Just Click it!

What to click?



Clicking breaks when..

- Click-challenged
  - Hands busy (driving).
  - Disabled.
- Limited Screen/keyboard Real estate
  - Cell phones, Microwave.
  - Ubiquitous computing..
- Complexity: SQL, shell scripts, menu hell.

Alternative Interface Paradigm

Speech + NL Understanding + Agents.

- “Where is Lord of the Rings showing?”
- “Defrost my corn.”
- “Delete all my old messages except the ones from Mom.”

Substantial Research Challenges!

State of the Art

- Commercial Speech Systems allow single word input.
- NLIDBs are unreliable (try Microsoft's English Query).
- Nontrivial Autonomous Agents that respond to complex requests?!

II. Our Focus

- Not speech.
- 1991 – 1997: built softbots.
- Today:

Reliable Natural Language Interfaces

Why Reliable?

Imagine “intelligent” interfaces that..

- Sometimes delete the wrong file.
- Can report incorrect flight times.

AI cannot be an excuse for incompetence (Norman, Schneiderman).

Some Common Objections

- Can't the interface just confirm?
  - Yes, but will users attend?
- Speech understanding isn't reliable.
  - That will gradually change.
  - Still need reliable language module.
- NL understanding is AI-complete!

III. Semantics is hard

- Syntactic, scopal ambiguity
- Word-sense ambiguity
- Time, events, liquids, holes,…
- Shakespeare, Faulkner,…
- Discourse, Pragmatics (speech acts)

We have to think about tractable classes!

Our Basic Hypothesis

There are common situations where Semantic Interpretation is tractable.

Sentence → Target expression

Our Research Strategy

1. Identify easy-to-understand NL sentences.
2. Develop Taxonomic Theory of Semantic Tractability.
3. Build Reliable NLIs.
4. Test NLIs experimentally.

Semantic Tractability

Easy to understand questions:

What are the Chinese restaurants in Seattle?
What Microsoft jobs require 2 years of experience?
What rivers run through Texas?

Semantically tractable questions
are quite common: 77.5% - 97%

Semantic Intractability

Q: What is the second highest mountain in the US?

A: The word 'second' is unknown; please rephrase your query.

Q: What are the states bordering the states bordering the states bordering Montana?

A: huh?

Semantically Tractable Qs

Given a lexicon and a parse, IF

1. Q contains at least one wh-word.
2. There exists a *valid mapping* from Q to a set of database elements.
3. Q maps to a nonrecursive datalog clause.

THEN Q is Semantically Tractable.

Valid Mapping

Words (phrases) → DB elements.

- Lexical constraints: Lord of the Rings.
- Attachment constraints:
  - What is the population of Atlanta?
- Semantic constraints:
  - Attributes need values: Cuisine → Chinese.
  - Values can have implicit attributes.

Guarantees on ST Qs.

Precise = NLIDB implementation.

- Precise can detect ST Qs.
- Precise is *sound* for ST Qs.
- Precise is *complete* for ST Qs.

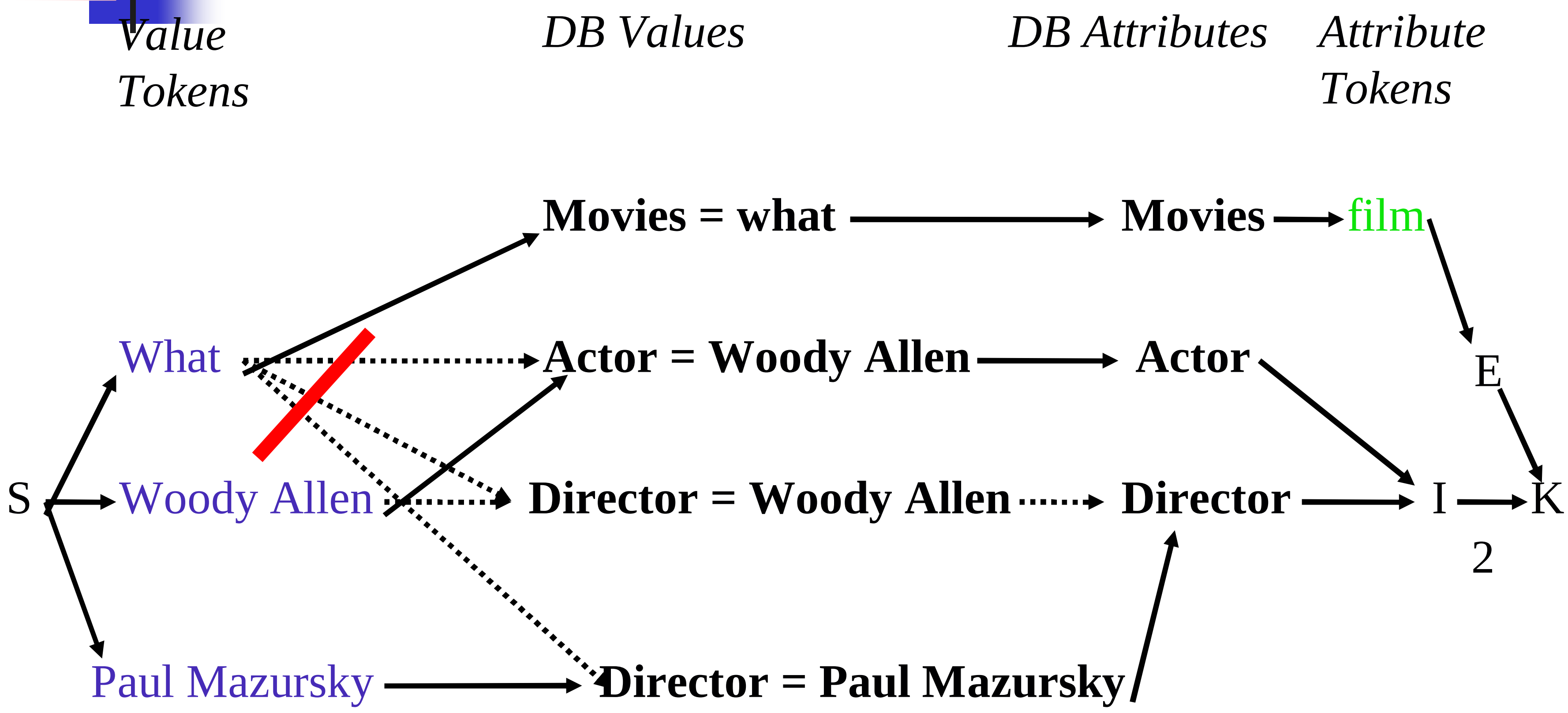
Will Precise capture "user intent?"

IV. Precise Implementation

- Lexicon extracted from DB + Wordnet.
- Parser plug in (Charniak 2000).
- Semantic constraints via graph match.
  - Word1 → DB_EL1
  - Phrase2 → DB_EL2, → DB_EL3

Match is computed via maxflow.

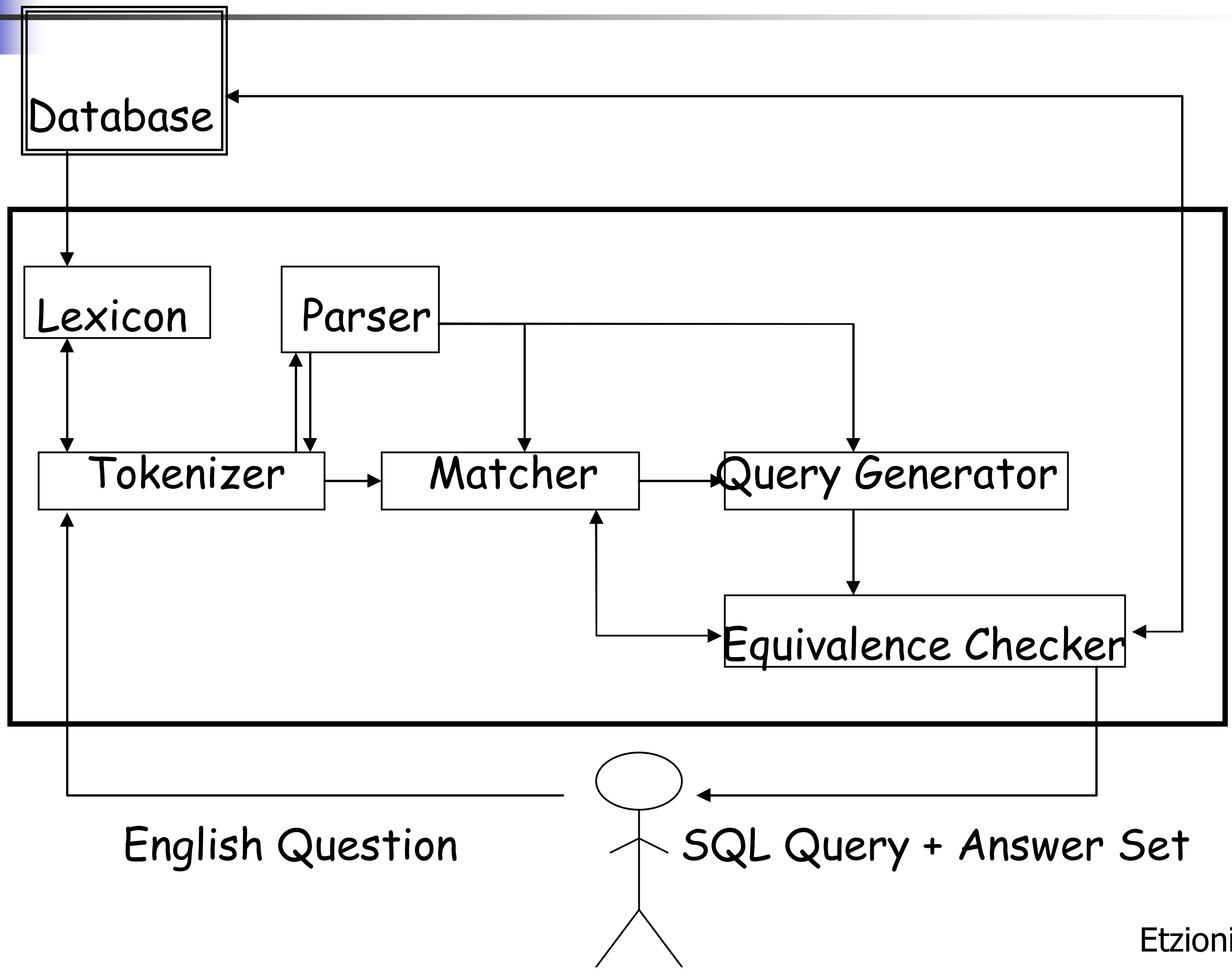
PRECISE graph representation



"What are the Paul Mazursky films with Woody Allen?"

Etzioni, 473

PRECISE Architecture



Ambiguity meets Reliability

- Precise computes all valid mappings.
- Many possibilities → clarifying Q.

- What is the population of New York?
  - The population of New York **City** is…
  - The population of New York **State** is..

Experiments

Systems

Mooney, PRECISE, EnglishQuery

Datasets

Sets of NL questions labeled with SQL queries

Geography (846) Jobs (577) Restaurants (224)

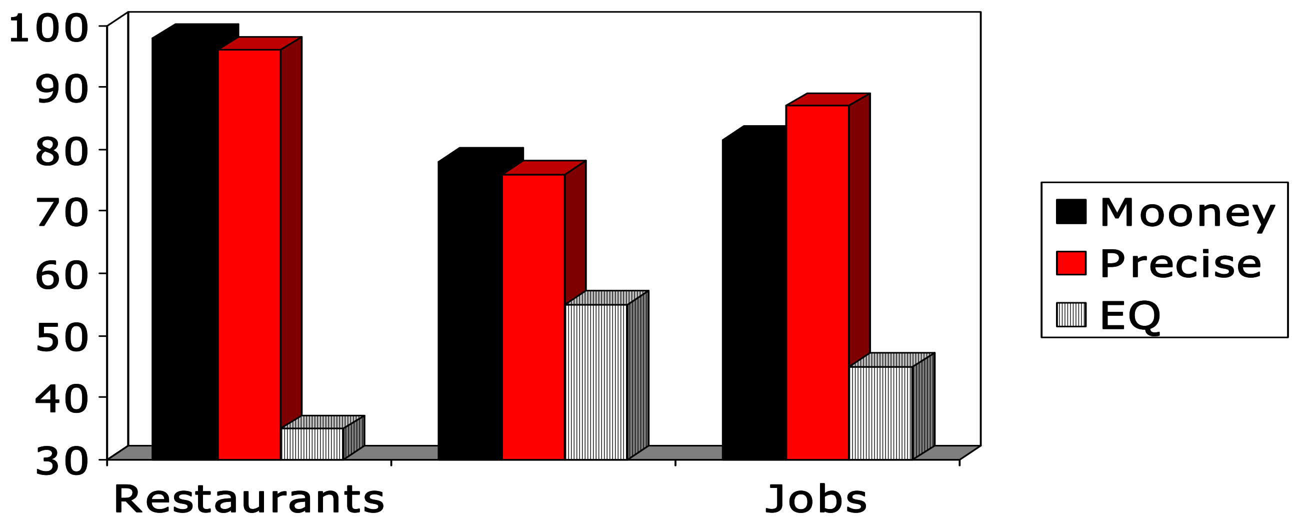
Measure

PRECISE's performance

Prevalence of semantically tractable questions

Fraction Answered

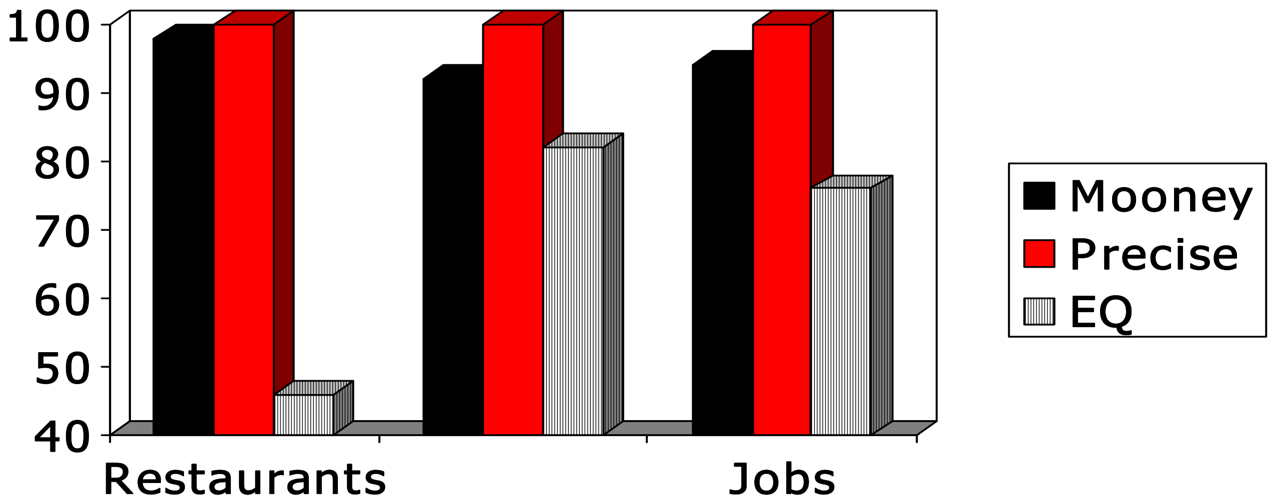
Recall $= Q_{answered}/Q$



Recall > 75 %

Error Rate

Precision = $Q_{correct}/Q_{answered}$

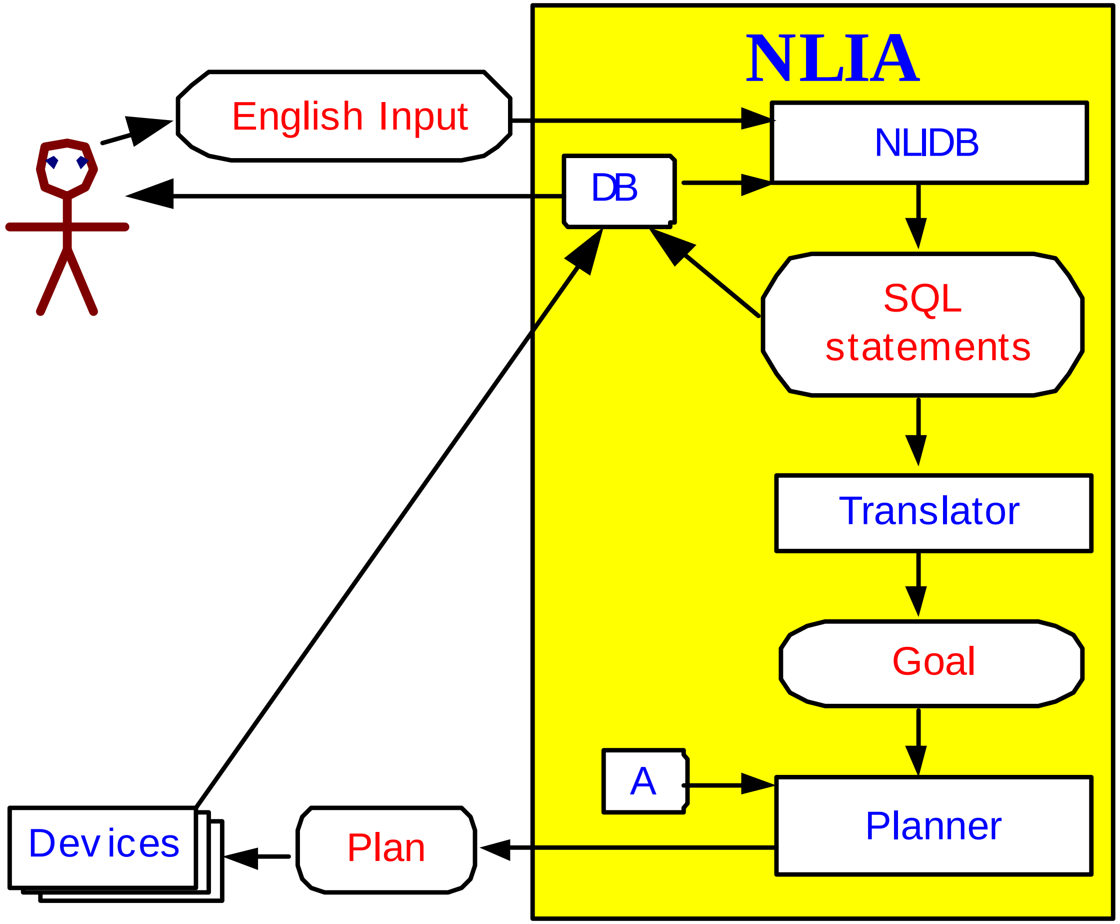


PRECISE made no mistakes on semantically tractable questions

Exact – Case Study

- Exact is an NLI to the Panasonic KX-TC1040W telephone/answering machine
- It uses Precise to formulate goals for the Blackbox planner (Kautz & Selman)
- The database model for the phone has 5 relations (between 2 and 11 attributes each) and the action set includes 37 actions.

Exact – Diagram



Conclusion

- Click-it UI paradigm is fraying.
- Need reliable NLIs.
- Semantically Tractable sentences.
  - Taxonomic Theory.
  - Precise and Exact implementations.
  - Promising, preliminary experiments.