Building Software Agents for Planning Monitoring, and Optimizing Travel

Craig A. Knoblock
University of Southern California

Acknowledgements

- Jose Luis Ambite, USC
- Greg Barish, Fetch Technologies
- Oren Etzioni, University of Washington
- Kristina Lerman, USC
- Martin Michalowski, USC
- Steve Minton, Fetch Technologies
- Ion Muslea, SRI
- Maria Muslea, USC
- Jean Oh, CMU
- Snehal Thakkar, USC
- Rattapoom Tuchinda, USC
- Alexander Yates, University of Washington

Introduction

- Wealth of travel-related data available online
- Web provides unprecedented access to information to end users
- Abundance of computing power available

- We can exploit these three factors to:
  - Support better planning of travel
  - Provide real-time monitoring of travel plans
  - Exploit data mining techniques to minimize problems and cost

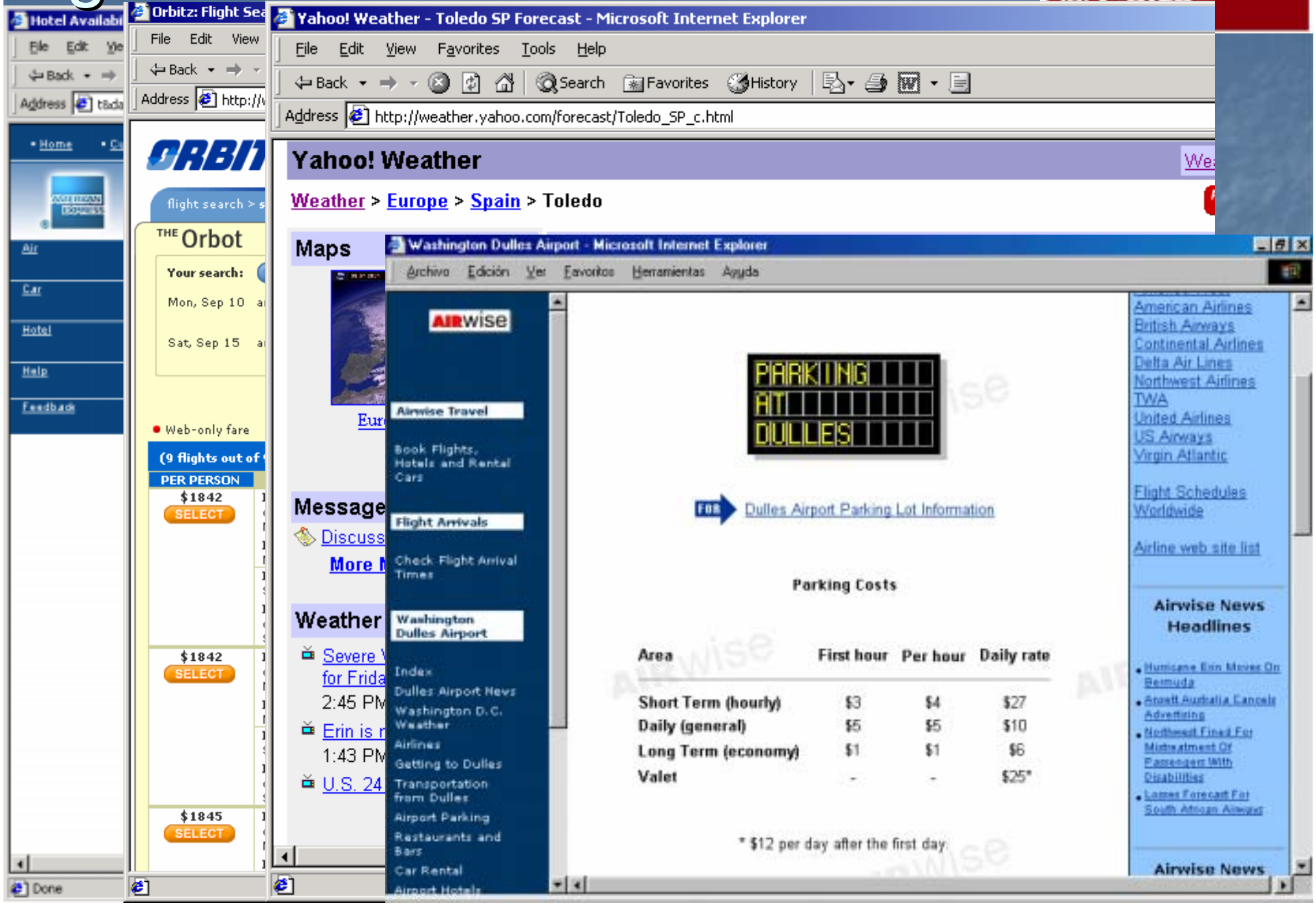
Outline

- Agent Access to Online Sources
- Interactive Planning of a Trip
- Building Agents for Monitoring Travel
- Mining Online Sources to Optimize Travel
- Conclusions

Outline

- Agent Access to Online Sources
- Interactive Planning of a Trip
- Building Agents for Monitoring Travel
- Mining Online Sources to Optimize Travel
- Conclusions

Agent Access to Online Sources



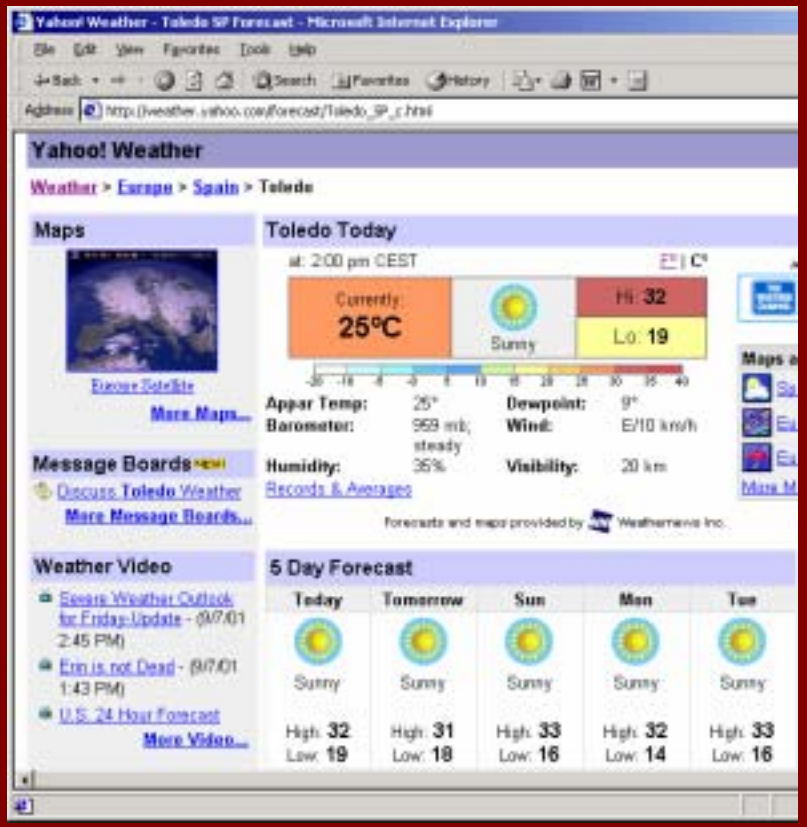
Problem: Information Not in a Usable Format

- Web pages are intended for human consumption
- Web services and XML are designed to solve this problem, but not available for most data
- Need to turn these online sources into ‘agent-enabled’ sources
  - Support database like querying by a software agent
  - Return information in a structured format, such as XML

Wrappers for Live Access to Online Sources

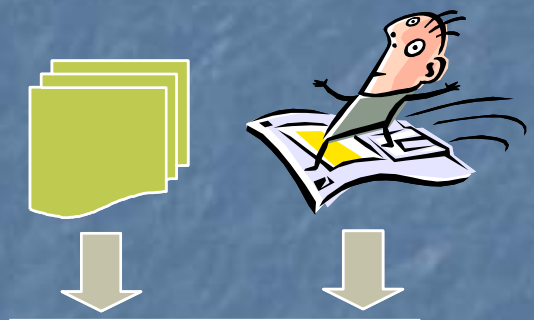
ENTER 04 @ Cairo

Wrapper




```
<YAHOO_WEATHER>
- <ROW>
  <TEMP>25</TEMP>
  <OUTLOOK>Sunny</OUTLOOK>
  <HI>32</HI>
  <LO>19</LO>
  <APPARTEMP>25</ APPARTEMP >
  <HUMIDITY>35%</HUMIDITY>
  <WIND>E/10 km/h</WIND>
  <VISIBILITY>20 km</VISIBILITY>
  <DEWPOINT>9</DEWPOINT>
  <BAROMETER>959 mb</BAROMETER>
  </ROW>
</YAHOO_WEATHER>
```

Learning a Wrapper



GUI

Labeled Pages

EC Tree

Inductive Learning System

Wrapper

Status

- Almost any source on the Web can be turned into an agent-enabled source
  - Time to construct a wrapper ranges from a few minutes to a few hours
  - Tools are easy to learn
- Makes it possible to exploit the huge amount of information available online
- Wrapper learning technology has been licensed to Fetch Technologies, which has a commercial product available

Outline

- Agent Access to Online Sources
- Interactive Planning of a Trip
- Building Agents for Monitoring Travel
- Mining Online Sources to Optimize Travel
- Conclusions

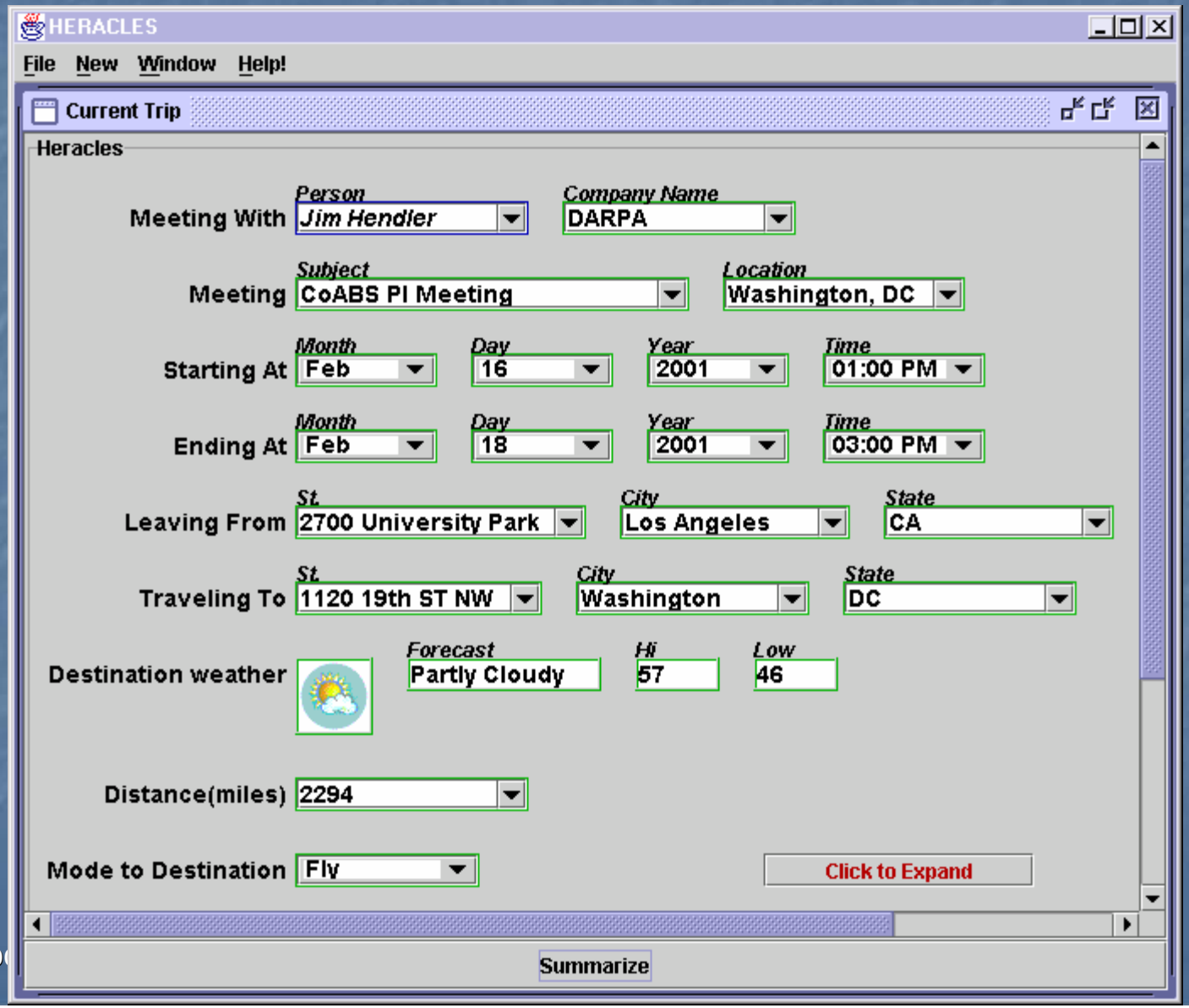
Interactive Trip Planning

- Current systems provide support to select flights, hotels and cars
  - Integrates the planning at the level of dates and locations
- There are many more factors involved in planning a trip
  - Which airports to fly into and out of
  - Whether to drive or take a taxi to the airport
  - How to get form the airport to the destination
  - Proximity of hotel to meeting
  - Etc...
- Ideally a system will
  - Provide all of the data required to make these decisions
  - Provide a way to consider the tradeoffs of the various choices

Heracles Constraint-based Planning

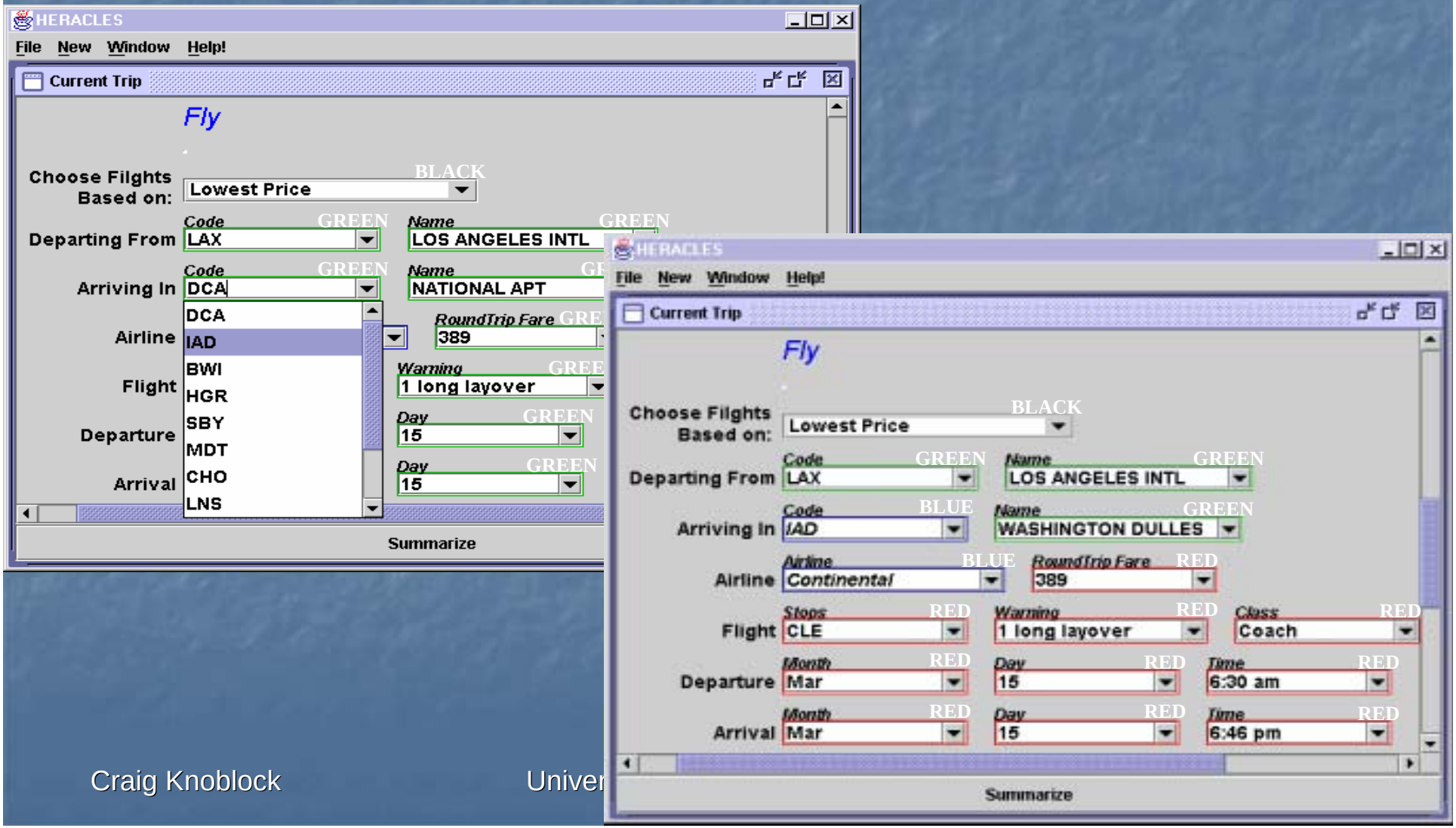
- Framework for building integrated applications
- Extract and integrate data for a given task
  - Live access to online sources using the wrappers
- Constraint-based decides what sources to query and how to integrate the results
  - Tight integration of user choices

Travel Planner



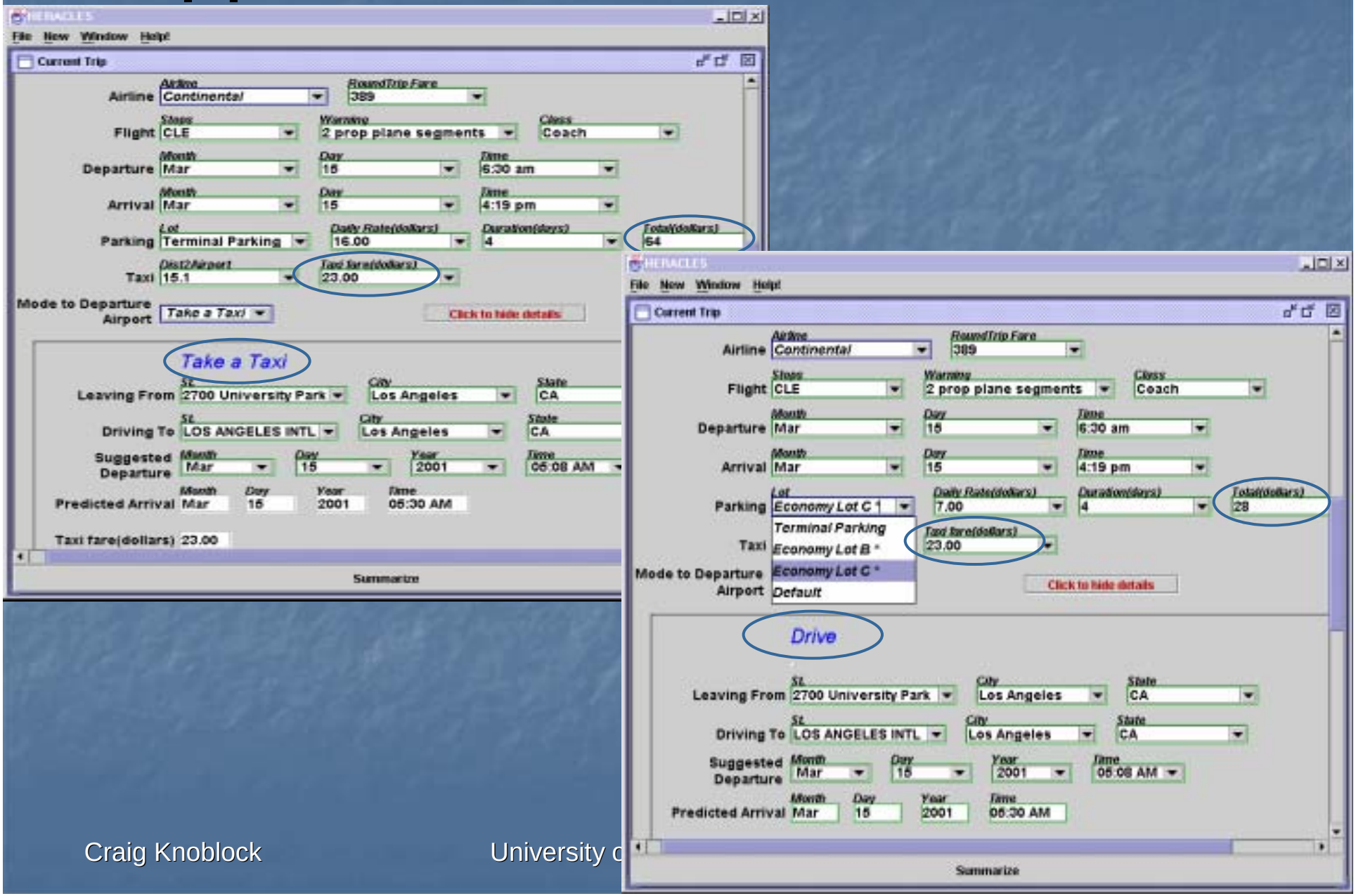
Dynamically Updates Slots as Information Becomes Available

ENTER 04 @ Cairo



Supports Informed Choices

ENTER 04 @ Cairo

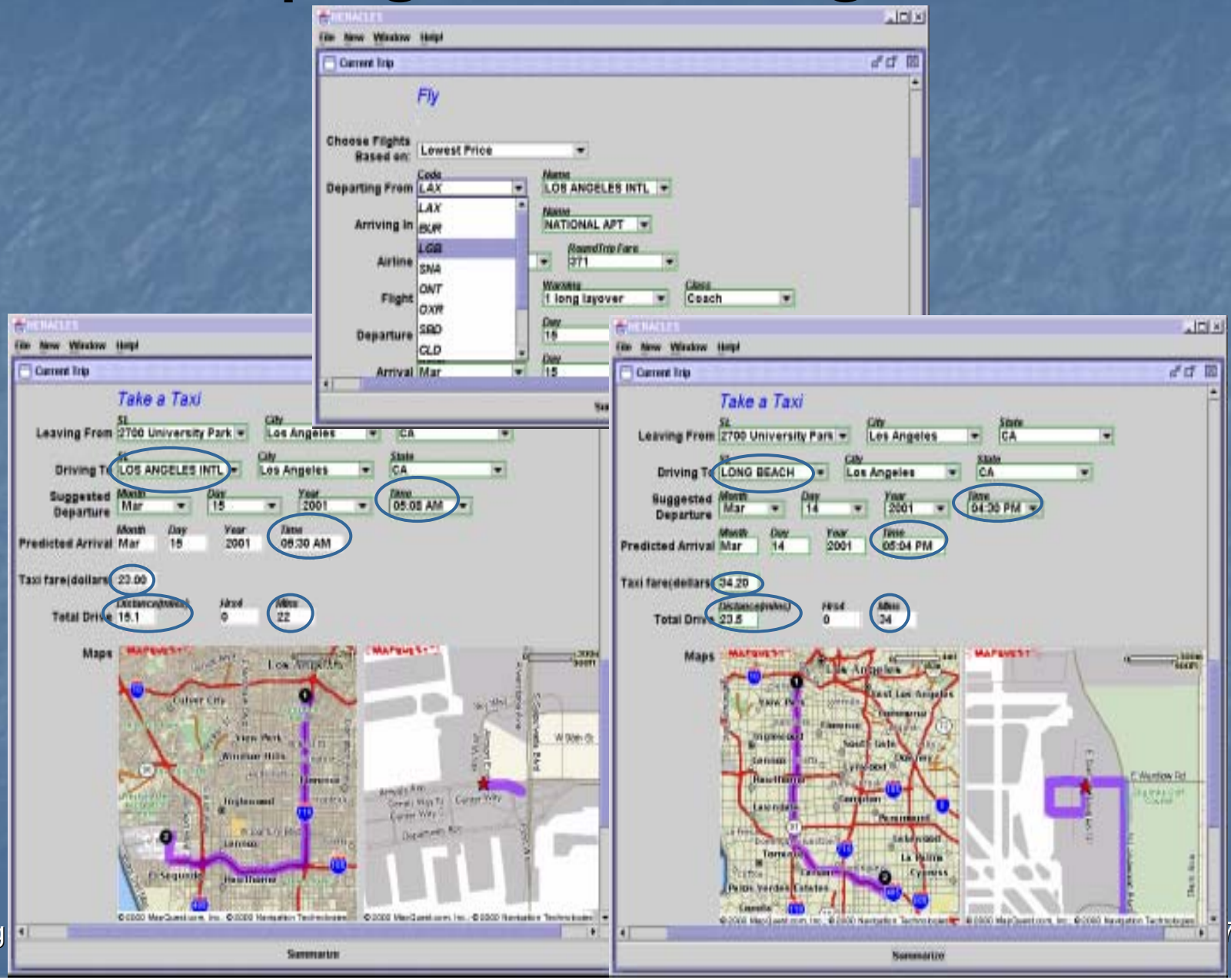


Craig Knoblock

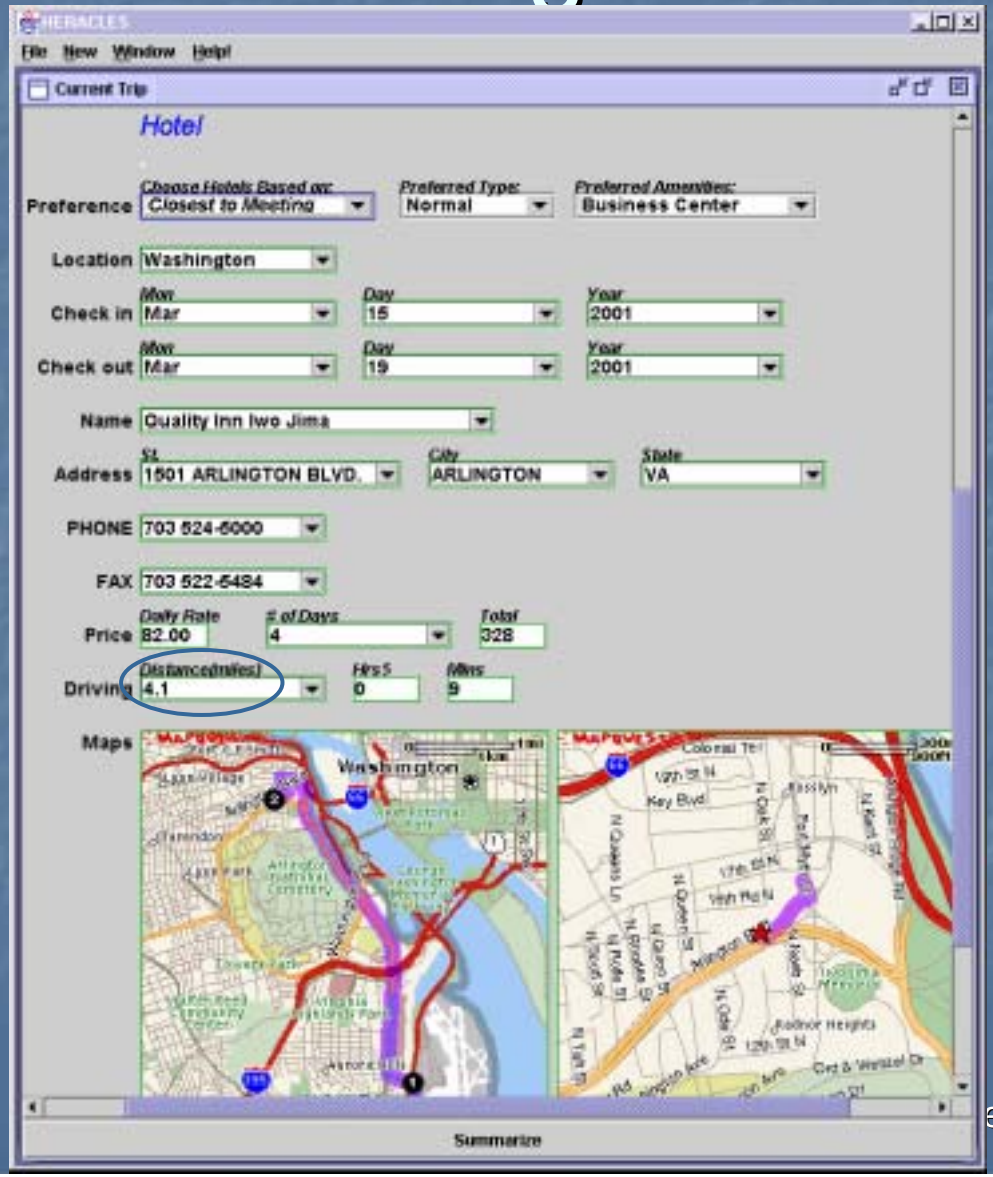
University o

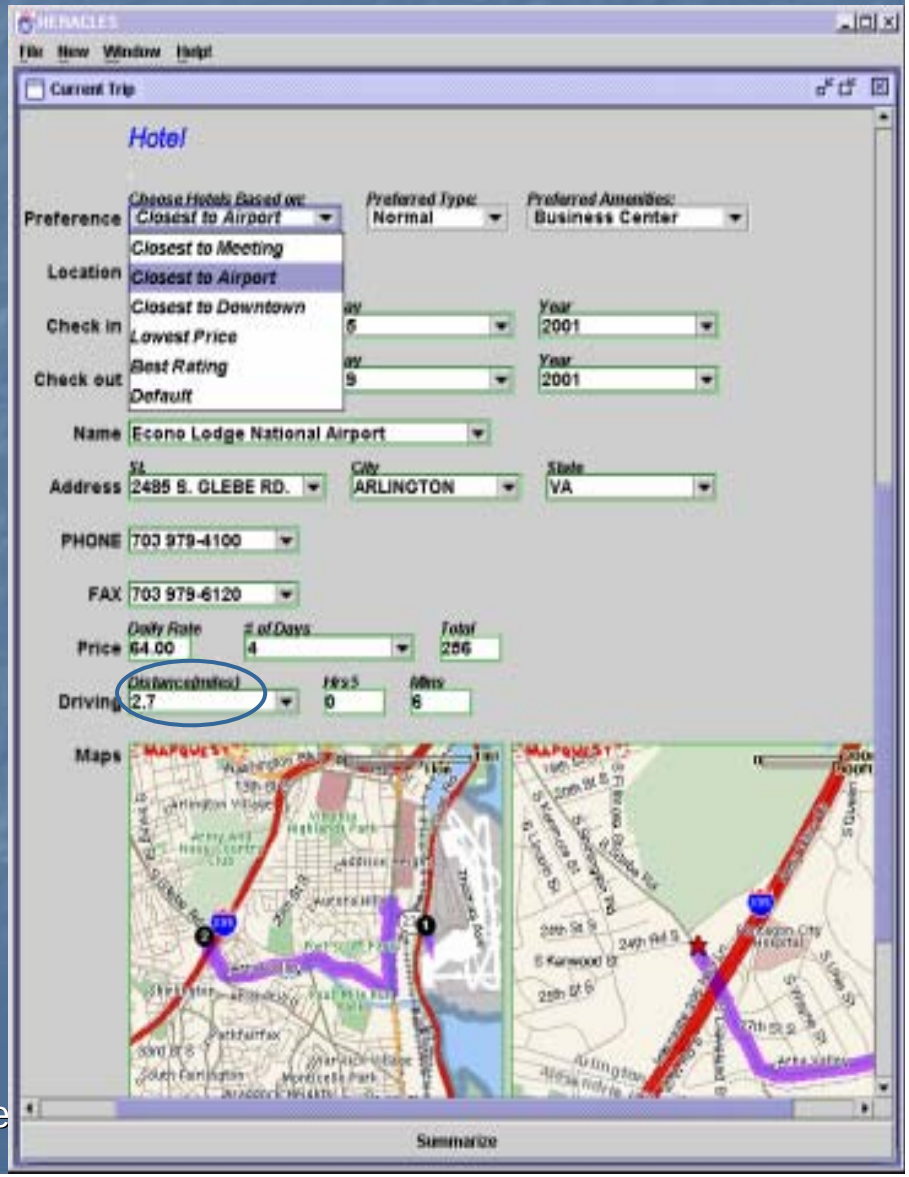
Propagates Changes

ENTER 04 @ Cairo

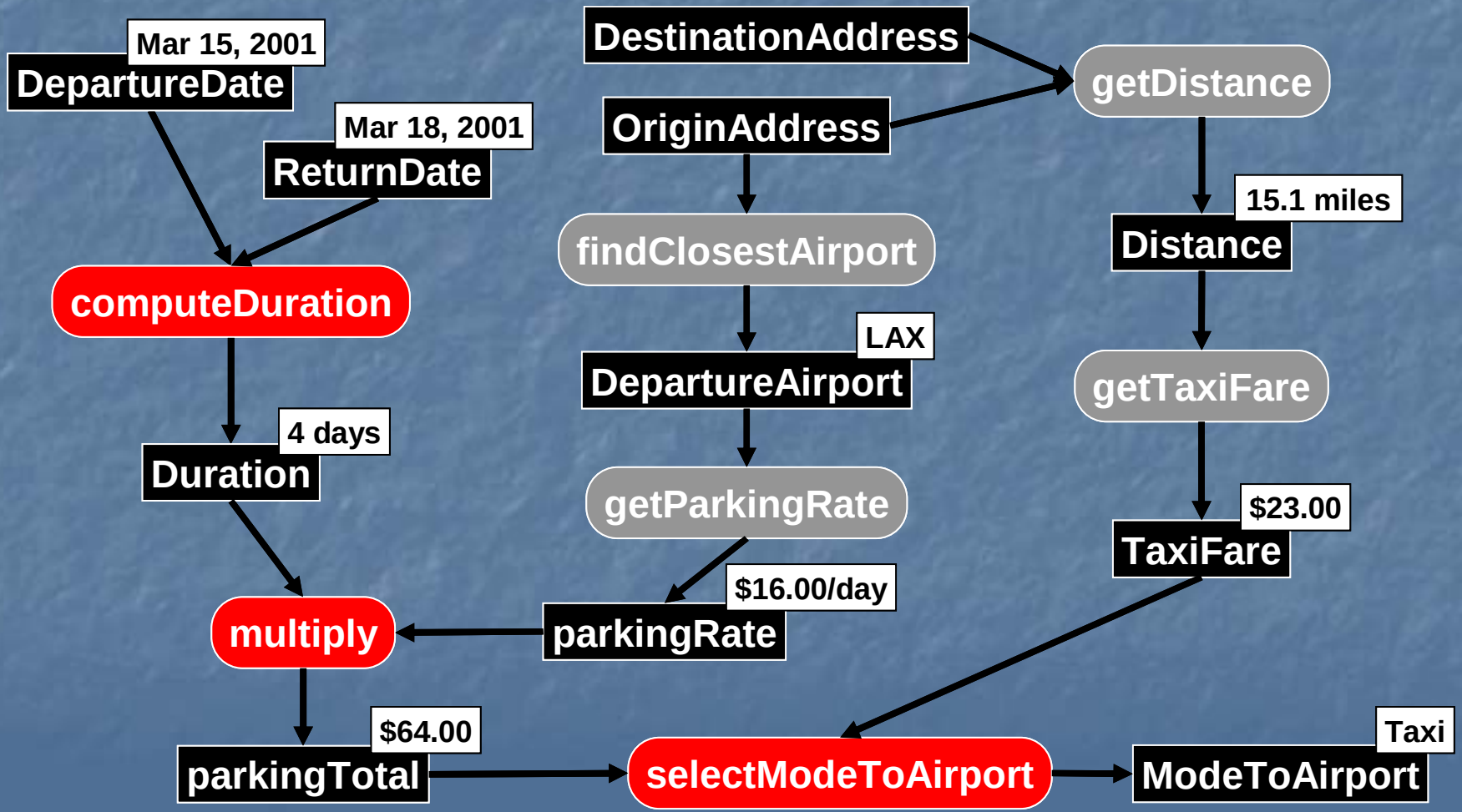


User Can Specify High-Level Preferences





Constraint Network: Drive or Taxi?



ENTER 04 @ Cairo

Summary

- Integration of wide range of data from many different sources
- Tight integration of data using constraints to capture the dependencies
- Supports better decision making
  - Easy to consider costs of specific choices
  - Easy to compare tradeoffs

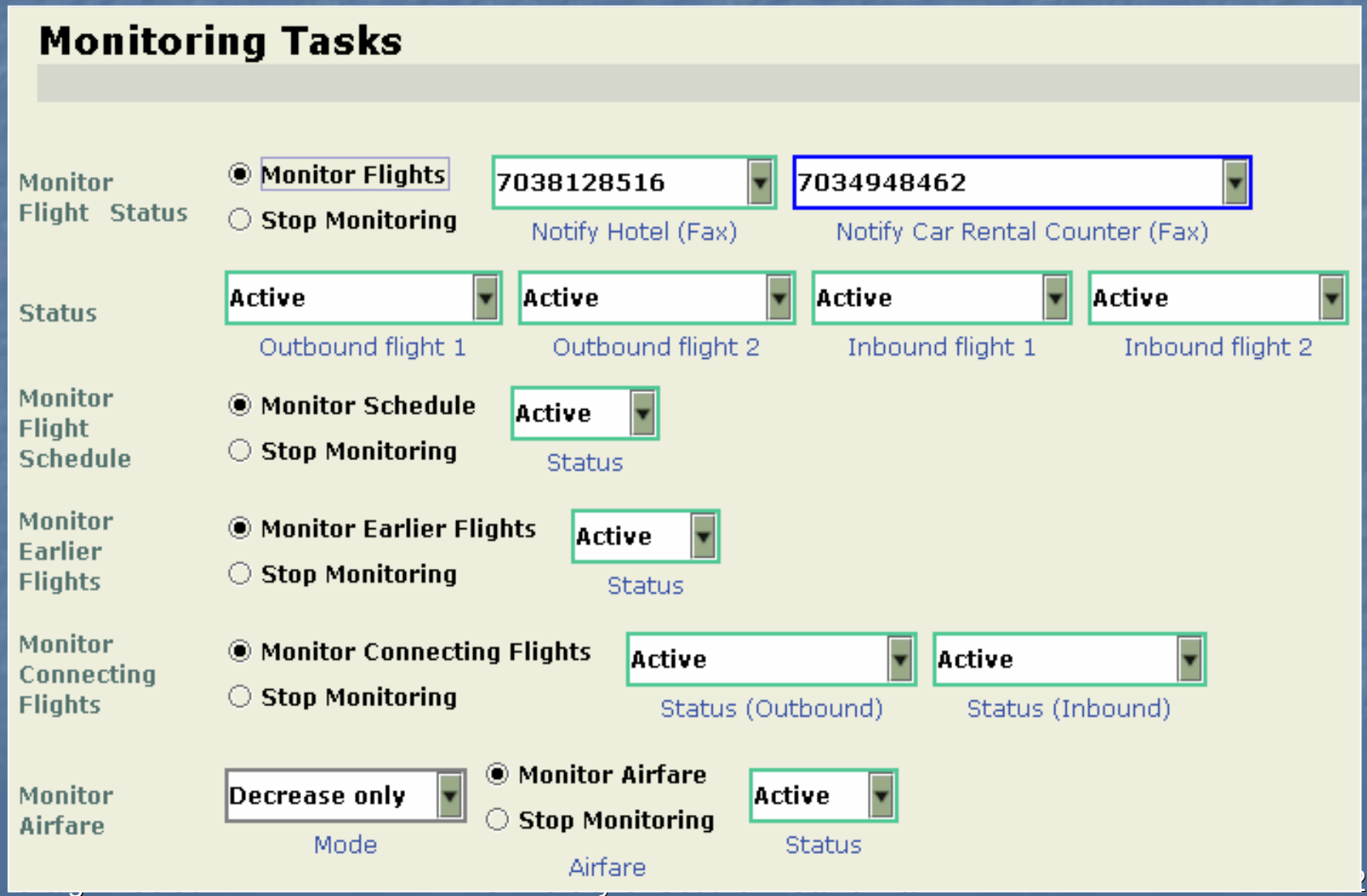
Outline

- Agent Access to Online Sources
- Interactive Planning of a Trip
- Building Agents for Monitoring Travel
- Mining Online Sources to Optimize Travel
- Conclusions

Agents for Monitoring Travel

- Many opportunities and possible problems can arise during travel
- Current environment:
  - Wide access to data
  - Abundance of computer resources
  - Availability of cell phones and portable computers
- Makes it possible to monitor all aspects of a trip
- Create personal assistants that monitor your travel plan to
  - exploit opportunities
  - avoid problems

Automatically Configuring Agents



Agents Deployed to Monitor Travel Itinerary

ENTER 04 @ Cairo

Travel Itinerary

Flight Prices & Schedules

Flight Status

Weather

Restaurants

Actual Messages Sent

ENTER 04 @ Cairo

- Flight-Status Agent:
  - Flight delayed message:

    ```
    Your United Airlines flight 190 has been delayed.
    It was  originally scheduled to depart at 11:45 AM
    and is now scheduled to depart at 12:30 PM.
    The new arrival time is 7:59 PM.
    ```

  - Flight cancelled message:

    ```
    Your Delta Air Lines flight 200 has been cancelled.
    ```

  - Fax to hotel message:

    ```
    Attention: Registration Desk
    I am sending this message on behalf of David
    Pynadath, who has a reservation at your hotel. David
    Pynadath is on United Airlines 190, which is now
    scheduled to arrive at IAD at 7:59 PM. Since the
    flight will be arriving late, I would like to request
    that you indicate this in the reservation so that the
    room is not given away.
    ```

Actual Messages Sent (cont.)

- Airfare Agent: Airfare dropped message

```
The airfare for your American Airlines itinerary
(IAD - LAX) dropped to $281.
```

- Earlier-Flight Agent: Earlier flights message

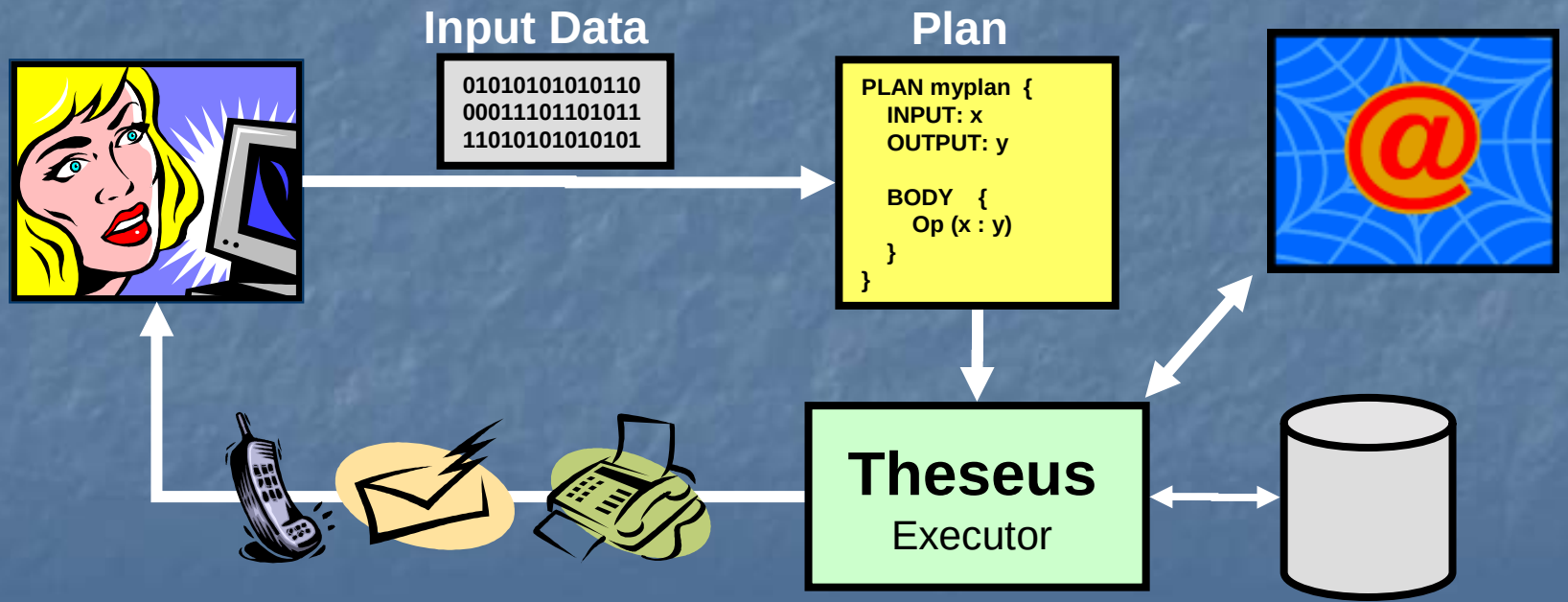
```
The status of your currently scheduled flight is:
# 190 LAX (11:45 AM) - IAD (7:29 PM) 45 minutes Late
If you would like to return earlier, the following
United Airlines flights will arrive earlier than your
scheduled flights:
# 946 LAX (8:31 AM) - IAD (3:35 PM) 11 minutes Late
--------
# 388 LAX (9:25 AM) - DEN (12:25 PM) 10 minutes Late
# 1534 DEN (1:20 PM) - IAD (6:06 PM) On Time
```

Challenges in Building Monitoring Agents

- **Problem**
  - Information gathering may involve accessing and integrating data from many sources
  - Total time to execute these plans may be large
- **Why?**
  - Slow remote sources
  - Unpredictable network latencies
  - Binding patterns
    - Source cannot be queried until a previous query has been answered
  - Result: execution is often I/O-bound

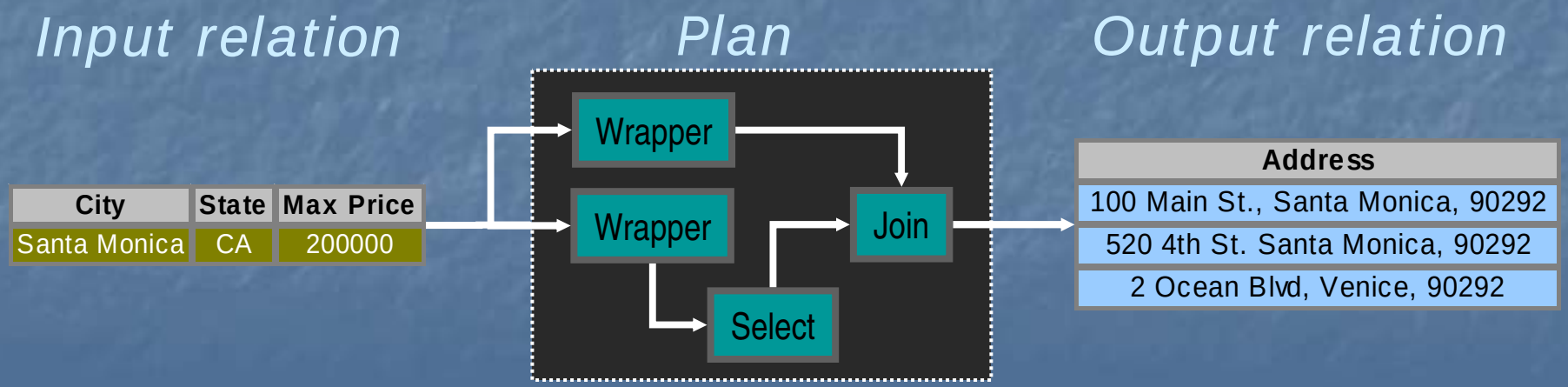
Theseus Agent Execution System

- **Plan language** and **execution system** for Web-based information integration
  - Expressive enough for monitoring a variety of sources
  - Efficient enough for real-time monitoring



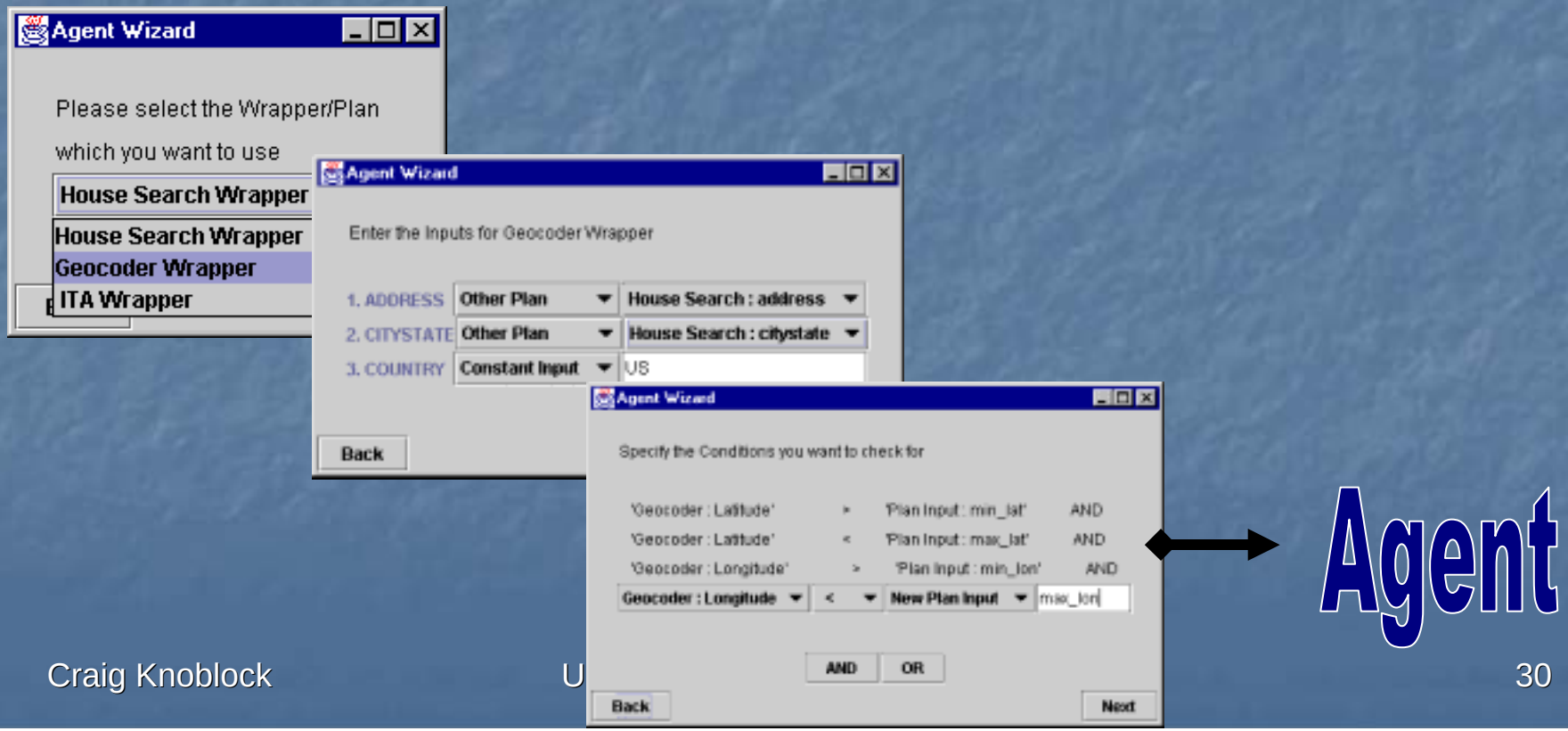
Streaming Dataflow

- Plans consist of a network of operators
  - Examples: **Wrapper**, **Select**, etc.
  - Operators produce and consume data
  - Operators "fire" upon any input data



Current Work

- Challenge: How to build monitoring agents without the need to program them?
- We are developing an agent wizard that leads the user through a series of questions and then builds the required agent



Outline

- Agent Access to Online Sources
- Interactive Planning of a Trip
- Building Agents for Monitoring Travel
- Mining Online Sources to Optimize Travel
- Conclusions

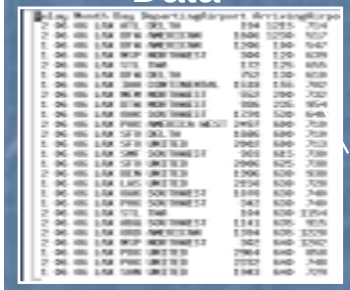
Mining Online Sources to Optimize Travel

- Wealth of online data provides many opportunities for data mining
- Two examples:
  - Predicting flight delays from historical flight delays and weather forecasts
  - Predicting airline prices to minimize cost

Predicting Weather-related Flight Delays

ENTER 04 @ Cairo

Historical Flight Data



Historical Weather Data

Learned Flight Delay Predictor

Agent

Prediction

Airline Predictor Result

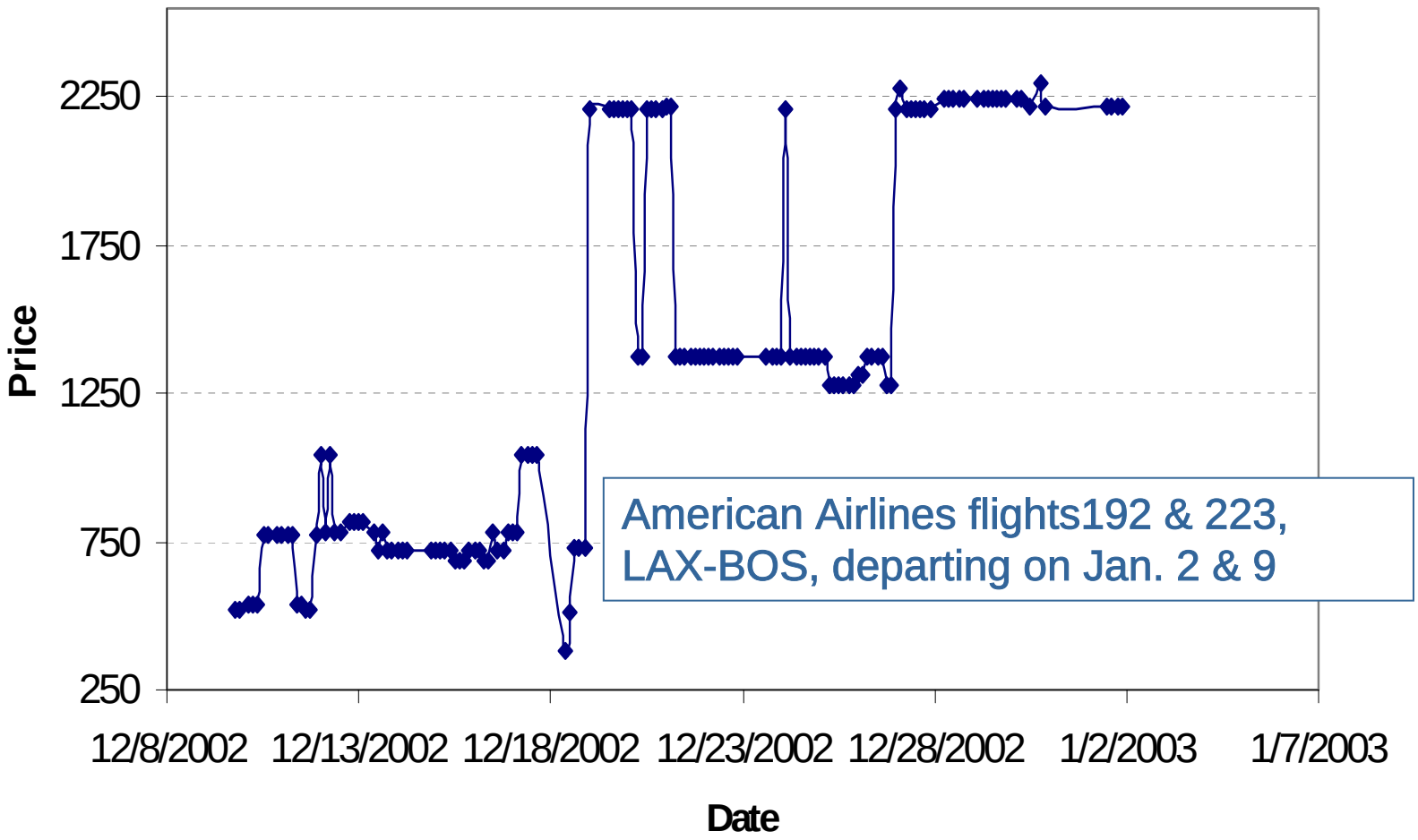
You requested the following

- *AIRLINE* USAIR
- *SEASON* 2
- *DEPART_CITY* ORL–orlando
- *DEPART_HOUR* 15
- *DEPART_DAY* Sat
- *ARRIVE_CITY* ORD–Chicago
- *ARRIVE_HOUR* 17
- *ARRIVE_DAY* Sat

The resulting prediction for your flight

Prediction	*Confidence*
ON TIME	89.26892%

Predicting Airline Prices



ENTER 04 @ Cairo

Hamlet: To Buy or Not to Buy

- Collected airline flight data over several months
- Developed a learning algorithm to predict whether to buy immediately or wait to buy a ticket
- Exploits the fact that airline pricing is done with a relatively static, but unknown algorithm
- Pricing can be learned by considering the pricing on the same flight on previous days

Data Set

- Extracted data from online sources using wrappers
- Collected over 12,000 price observations:
  - Lowest available fare for a one-week roundtrip
  - LAX-BOS and SEA-IAD
  - 6 airlines including American, United, etc.
  - 21 days before each flight, every 3 hours

Learning Algorithm

- Stacking with three base learners:
  1. Rule learning (Ripper) (e.g., R=wait)
  2. Time series
  3. Q-learning (e.g., Q=buy)
- Ripper used as the meta-level learner.
- Output: classifies each decision point as 'buy' or 'wait'.

Experimental Results

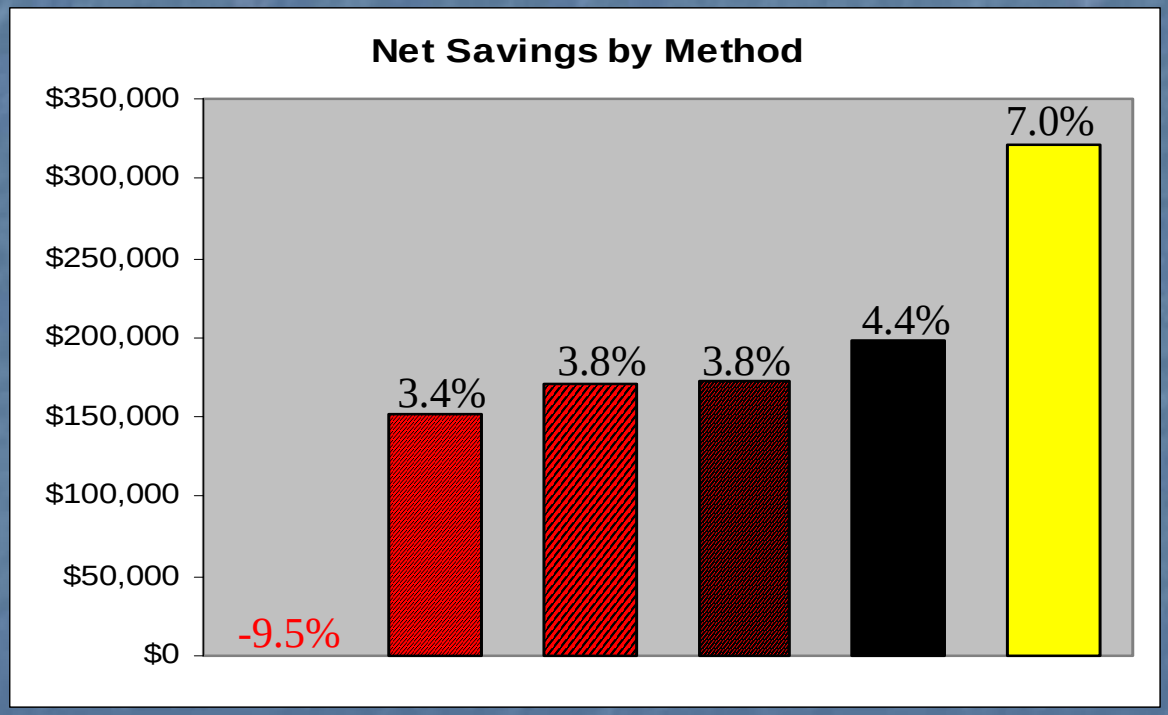
- Real price data; Simulated passengers
- Learner run once per day on “past data”
- Execution: label each purchase point until buy (or sell out)
- Compute savings (or loss)

Savings by Method

- Savings over "buy now".
- Penalty for sell out = upgrade cost.
- Total ticket cost is $4,579,600.

Method	Savings	Losses	Upgrade Cost	% Upgrades	Net Savings	% Savings	% of Optimal
Optimal	$320,572	$0	$0	0%	$320,572	7.0%	100.0%
By hand	$228,318	$35,329	$22,472	0.36%	$170,517	3.8%	53.2%
Ripper	$211,031	$4,689	$33,340	0.45%	$173,002	3.8%	54.0%
Time Series	$269,879	$6,138	$693,105	33.00%	-$429,364	-9.5%	-134.0%
Q-learning	$228,663	$46,873	$29,444	0.49%	$152,364	3.4%	47.5%
Hamlet	**$244,868**	**$8,051**	**$38,743**	**0.42%**	**$198,074**	**4.4%**	**61.8%**

Savings by Method



Upgrade Penalty

Method	Upgrade Cost	% Upgrades
Optimal	$0	0%
By hand	$22,472	0.36%
Ripper	$33,340	0.45%
Time Series	$693,105	33.00%
Q-learning	$29,444	0.49%
Hamlet	**$38,743**	**0.42%**

Savings on "Feasible" Flights

- 24% of the time savings possible

Comparison of Net Savings (as a percent of total ticket price) on Feasible Flights

Method	Net Savings
Optimal	30.6%
By hand	21.8%
Ripper	20.1%
Time Series	25.8%
Q-learning	21.8%
Hamlet	**23.8%**

Conclusions

- The Web provides unprecedented access to data
- Build wrappers to turn these sources into agent-enabled sources
- Combine these sources to build an integrated travel planning system
- Automatically generate a set of agents to monitor all aspects of a travel plan
- Mine the data sources to advise a traveler about prices, chances of delays, etc.
- There are many more uses of this widely available data...

More Information

- Email:
  knoblock@isi.edu
- Papers available from my homepage:
  http://www.isi.edu/~knoblock

ENTER 04 @ Cairo

Backup

Ripper

- Features include price, airline, route, hours-before-takeoff, etc.
- Learned 20-30 rules…

IF hours-before-takeoff ≥ 252 AND price ≥ 2223

AND route = LAX-BOS THEN *wait*

Simple Time Series

- Predict price using a fixed window of *k* price observations weighted by *α*.
- We used a linearly increasing function for *α*

$$p_{t+1} = \frac{\sum_{i=1}^{k} \alpha(i)\, p_{t-k+i}}{\sum_{i=1}^{k} \alpha(i)}$$

Q-learning

Natural fit to problem

$$Q(a,s) = R(a,s) + \gamma \cdot \max_{a'}(Q(a',s'))$$

$$Q(b,s) = -price(s)$$

$$Q(w,s) = \begin{cases} -300000 & \text{if flight sells out after } s. \\ \max(Q(b,s'), Q(w,s')) & \text{otherwise.} \end{cases}$$