Natural Language Processing: Challenges and Opportunities in Intelligent Transportation

Barbara Di Eugenio
Computer Science
University of Illinois at Chicago

Natural Language Processing

Natural Language Processing (NLP) encompasses all **computational** aspects regarding the interpretation **and** the production of **natural** languages such as English, Chinese, Italian or Arabic.

Applications: search, machine translation, dialogue systems, summarization

NLP: why is it a challenge?

I lost my voice. Can you find a pharmacy on my way back from the conference?

NLP: why is it a challenge?

I lost my voice. Can you find a pharmacy on my way back from the conference?

I lost my voice. Can you call the nearest pharmacy with my phone?

NLP: why is it a challenge?

I lost my voice. Can you find a pharmacy on my way back from the conference?

I lost my voice. Can you call the nearest pharmacy with my phone?

I am now at the conference. Can you find me a pharmacy on the way back?

Dialogue Systems Development

1. Collect natural dialogues between humans
2. Transcribe / annotate dialogues
3. Mine the dialogues to build computational models for features of interest
4. Implement models in dialogue interfaces (syntatic / semantic / pragmatic processing)
5. Run systematic evaluation with users

Project 1: Driving Directions (with Michael J. Trolio)

Much effort on understanding best type of interface between driver and intelligent vehicle, e.g. best combination of input / output modalities
[Bernsen & Dybkiær 02],[Geutner et al 02],[Buehler et al 03], [Coletti et al 03], [John et al 04], [Salvucci 05]

Orthogonal issue: of all information that interface **could** communicate to driver, how much **should** it select?

Our system provides driving directions, and enhances them with cues that occurs when two people exchange route information
[Klein 82], [Wunderlich & Reinert 82], [Fraczak et al 98]

Many algorithms can generate text output for arbitrary navigation between two termini – most famously, Google Maps. However, resulting directions lack natural cues

Our approach

Our software system can generate both essential and enhanced directions.
It consists of:

- route generator that uses factual map data to generate route
- natural language (English) front-end

It was evaluated with user study: users given enhanced directions were less likely to make mistakes, i.e., take incorrect turns.

Disclaimers. System is not deployed inside vehicle (not even as simulation).
Language is typed, not spoken.

Essential vs. Enhanced Directions

Essential directions (as from Google, 6/11/2009)

A Soscol Avenue & First Street (Napa, CA)
B Steele Canyon Rd & Rimrock Dr

1.	Head northeast on 1st St toward McKinstry St	0.6 km
2.	Turn left at CA-121/Silverado Trail	2.6 km
3.	Turn right at CA-121/Monticello Rd	19.8 km
4.	Turn left at CA-128/Capell Valley Rd	0.1 km
5.	Turn right at Steele Canyon Rd	9.7 km

Enhanced Directions

...

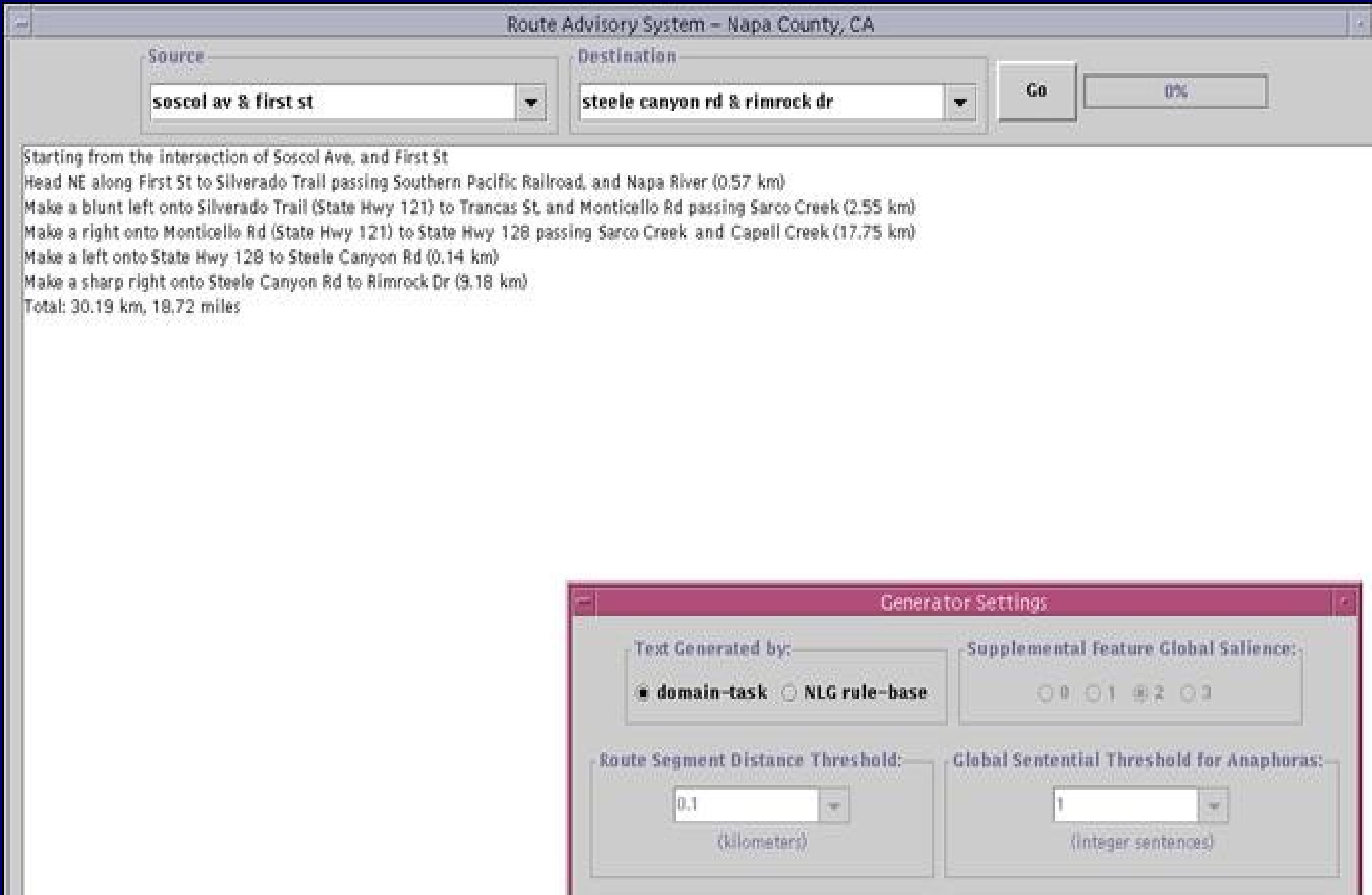
Make a right onto Monticello Rd (State Hwy 121) to State Hwy 128,
first passing Sarco Creek, and then passing Capell Creek
several times (17.75km).
This is the longest stretch of the road, about 18 kilometers.
...

The software system

Geographic domain: Napa County, California.
Raw data: Tiger/Line 95 database (US Census bureau)
(http://www.census.gov/geo/www/tiger/descr_95.html)

Information for each county contained in 17 text files.
Features can be: points (e.g. intersections of streets); lines (e.g., roadways); polygons (e.g. parks and lakes).

Basic route computation: A* algorithm on map graph based on Tiger/Line 95 files


Route Advisory System – Napa County, CA
Source
soscol av & first st
Destination
steele canyon rd & rimrock dr
Go
0%
Starting from the intersection of Soscol Ave. and First St
Head NE along First St to Silverado Trail passing Southern Pacific Railroad, and Napa River (0.57 km)
Make a blunt left onto Silverado Trail (State Hwy 121) to Trancas St, and Monticello Rd passing Sarco Creek (2.55 km)
Make a right onto Monticello Rd (State Hwy 121) to State Hwy 128 passing Sarco Creek and Capell Creek (17.75 km)
Make a left onto State Hwy 128 to Steele Canyon Rd (0.14 km)
Make a sharp right onto Steele Canyon Rd to Rimrock Dr (9.18 km)
Total: 30.19 km, 18.72 miles
Generator Settings
Text Generated by:
domain-task
NLG rule-base
Supplemental Feature Global Salience:
0 1 2 3
Route Segment Distance Threshold:
0.1
(kilometers)
Global Sentential Threshold for Anaphoras:
1
(integer sentences)

Enhanced route descriptions

1. Aggregation: how to group information to be communicated.
   We perform only syntactic aggregation: from

   Turn right on Monticello Road.
   Turn right on Hudson St.
   Turn right on Silverado Trail.

   system produces

   The following three turns are right turns:
   First, onto Monticello Road; second, onto Hudson St.; finally, onto Silverado Trail.

2. Inclusion of supplemental information.
   Establish salience (i.e., importance) of supplemental features; use salience level to decide which features to include and in which form [Fraczak et al. 98].

Supplemental features:

- Transition distances (transition is roadway traveled during one segment of route)

- Landmarks: divided into large (e.g.rivers, railways) and small (e.g., creeks, sloughs)

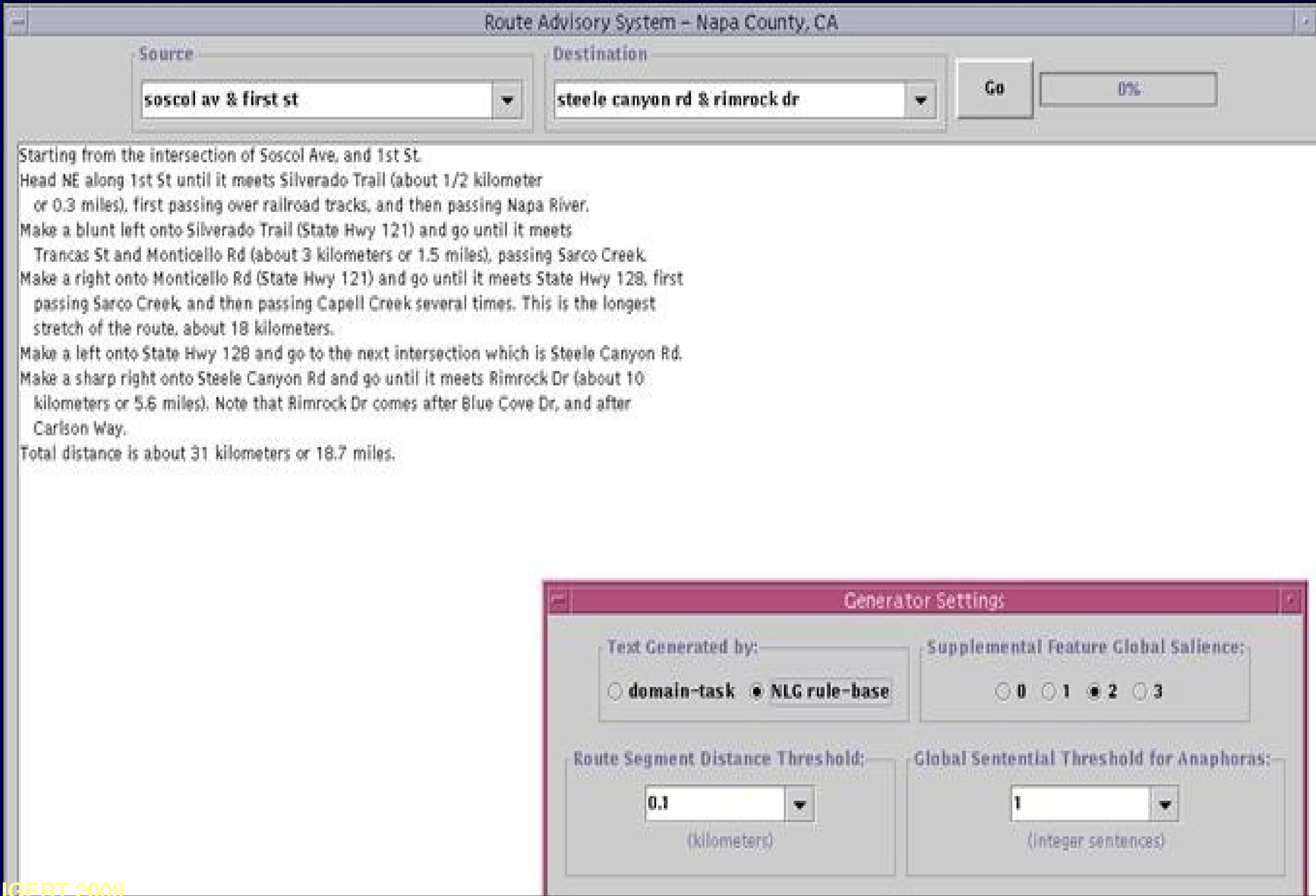
  Make a right onto Monticello Rd (State Hwy 121) to State Hwy 128, passing Sarco Creek

- Signposts: special types of intersections, signal that terminus for current segment is fast approaching

  Make a sharp right onto Steele Canyon Rd and go until it meets Rimrock Dr (about 10 kilometers or 5.6 miles). Note that Rimrock Dr comes after Blue Cove Dr.

- Referential expressions: how to refer to entities in the domain.

  Make a left onto Silverado Trail and go until it meets Trancas St and Monticello Rd (about 3 kilometers of 1.5 miles), passing Sarco Creek. Make a right onto Redwood Rd and go to the next intersection which is State Highway 29, again passing the creek.

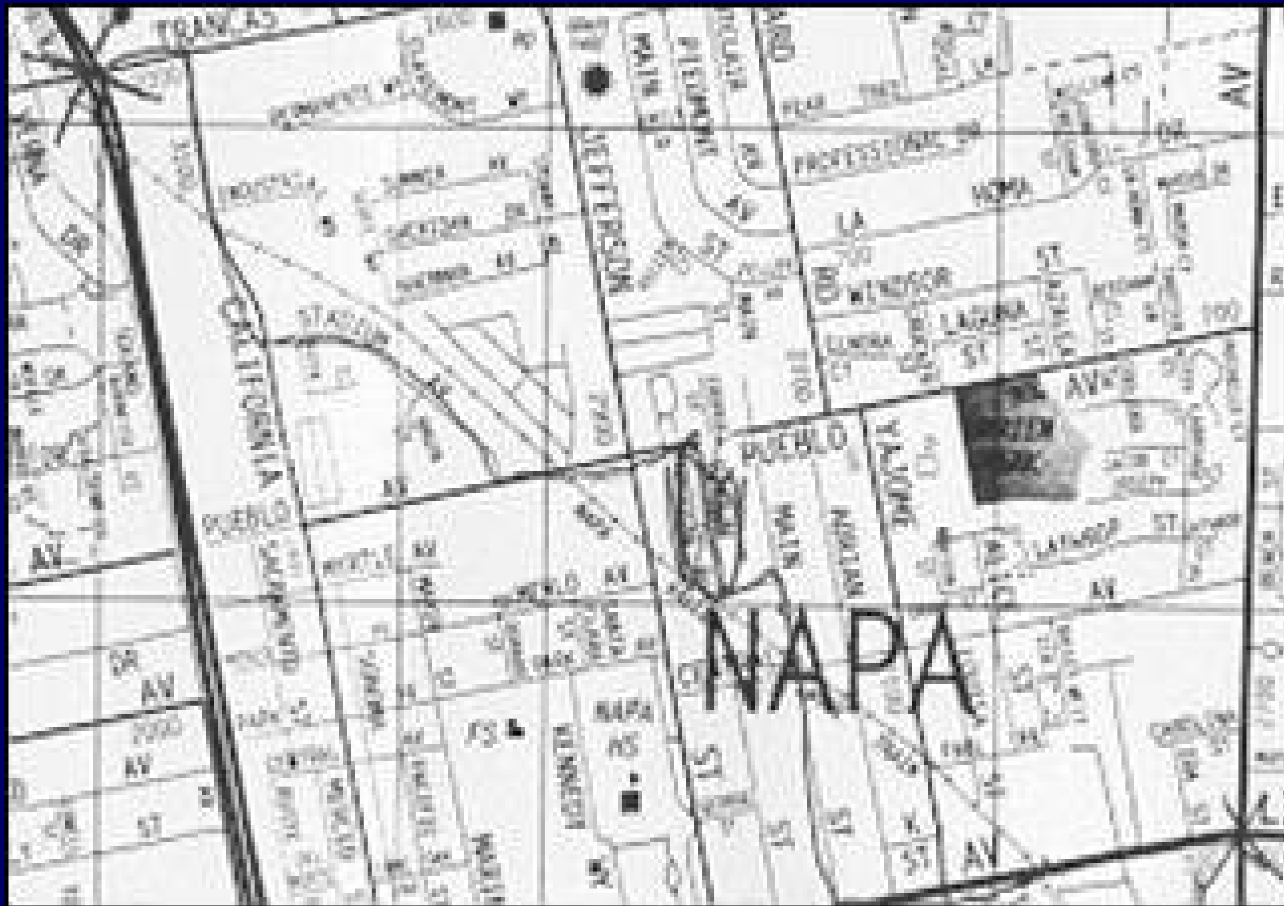

Route Advisory System – Napa County, CA
Source
soscol av & first st
Destination
steele canyon rd & rimrock dr
Go
0%
Starting from the intersection of Soscol Ave, and 1st St.
Head NE along 1st St until it meets Silverado Trail (about 1/2 kilometer or 0.3 miles), first passing over railroad tracks, and then passing Napa River.
Make a blunt left onto Silverado Trail (State Hwy 121) and go until it meets Trancas St and Monticello Rd (about 3 kilometers or 1.5 miles), passing Sarco Creek.
Make a right onto Monticello Rd (State Hwy 121) and go until it meets State Hwy 128, first passing Sarco Creek, and then passing Capell Creek several times. This is the longest stretch of the route, about 18 kilometers.
Make a left onto State Hwy 128 and go to the next intersection which is Steele Canyon Rd.
Make a sharp right onto Steele Canyon Rd and go until it meets Rimrock Dr (about 10 kilometers or 5.6 miles). Note that Rimrock Dr comes after Blue Cove Dr, and after Carlson Way.
Total distance is about 31 kilometers or 18.7 miles.
Generator Settings
Text Generated by:
domain-task
NLG rule-base
Supplemental Feature Global Salience:
0 1 2 3
Route Segment Distance Threshold:
0.1
(kilometers)
Global Sentential Threshold for Anaphoras:
1
(integer sentences)

Experiment

Between-subjects experiment with two groups of 15 subjects each.
Subjects in Group A receive essential directions, subjects in Group B enhanced directions.

Four routes were selected in advance: sample route for subjects to familiarize themselves with system; one urban route, one inter-urban route, and one urban-rural route

Task: outline route on map with pen in hand while reading text
Subjects administered questionnaire post-treatment.


NAPA
CALIFORNIA
JEFFERSON
PUEBLO
PROFESSIONAL DR
WINDSOR
LAGUNA

Results

	Group A (essential)	Group B (enhanced)
Total Time	22’2”	22’2”
Route precision	.95	.91
Route recall	.39	.45
Two Question score	.79	.77
Wrong turns	4.4	2.7

Difference in number of wrong turns is marginally significant (Mann-Whitney test, $U=94$, $p=0.09$)

Generating directions in real time more complex

- Cognitive load on driver: issues of timing and amount of information provided at any given time. Integrate into cognitive model of driver [Salvucci 05]

- Run experiment in which instructions are given orally, and piecemeal, with different levels of saliency

Project 2: Aggregation for Local Search (with Alberto Tretti - supported by Motorola)

Local Search: geographically constrained search against a structured database of local business listings

Find me a cheap Thai restaurant nearby
Where can I find bubble tea in Chicago?

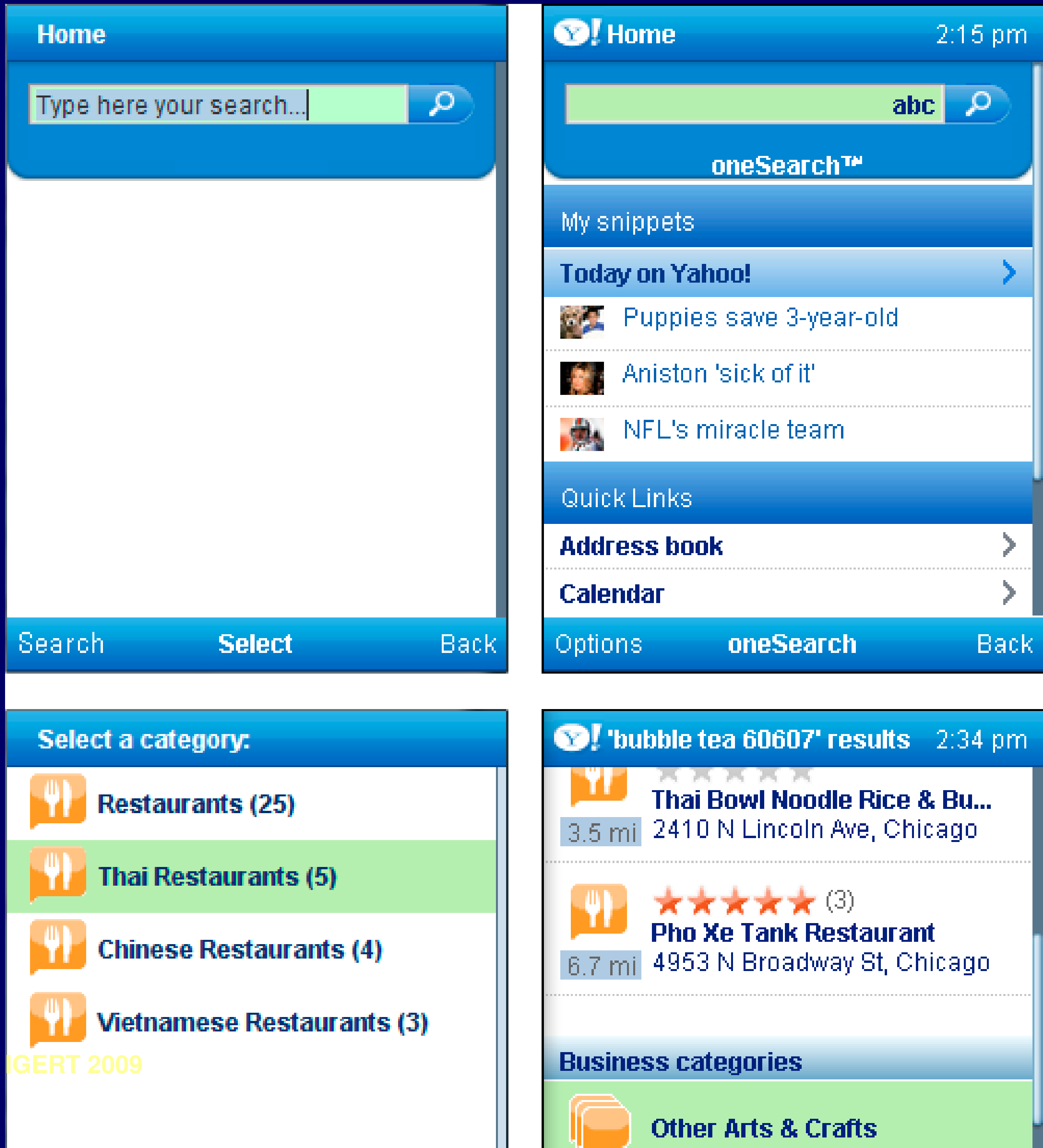
Aggregation: What and why

Aggregate results of local search info meaningful classes and choose appropriate classes to present to user

Search engines (e.g. Yahoo! Local) still return listings, possibly containing hundreds of items:

- Long lists of items difficult to navigate on portable devices
- Text-to-speech not real solution: cognitive limitations in remembering those long lists

Aggregation for Local Search



Hello Local

Exploits Yahoo!Local list of categories (reverse engineered into a tree structure) to group local listings by categories and eliminate those non relevant to the query

Chocolate store, Chicago: 257 results

- 75 Food Manufacturers & Processor
- 59 Grocery Stores,
- 51 B2B Food Wholesalers
- 50 Candy & Sweets
- **27 Cellular Providers**

Algorithm

1. Store Yahoo!s results in tree structure (node=category)
2. Perform greedy search to find *dominant* category(ies)
3. Aggregate and rank homonym results within the category of interest
4. Return the new list of results to the user

Evaluation

Between-subject experiments with 3 groups of 13 subjects each.
Compared 3 systems: Yahoo! Local, Y! Go, and Hello Local.

Subjects performed 7 different tasks, 3 easy and 4 difficult.

Imagine you are downtown Chicago (zip code 60601) and you start to get hungry, but you dont remember the name of that good place your friend recommended. Find one of the following places without using its name: [Chipotle] or [Taco Bell]

Objective measures plus satisfaction questionnaire

Results

User satisfaction quite high, reduction in number of results about 43% on average

Objective measures (time on task, number of queries, position of intended result in final result list)

- users were significantly slower with Hello Local on easy tasks (ie, overhead not worth it)
- on difficult tasks, favorable (but not significant) trends for all measures for Hello Local

Conclusions

- NLP is challenging!
- Sample of two transportation / mobility related projects
- Much more is possible!