VOICE SEARCH ON MOBILE DEVICES

Geoffrey Zweig

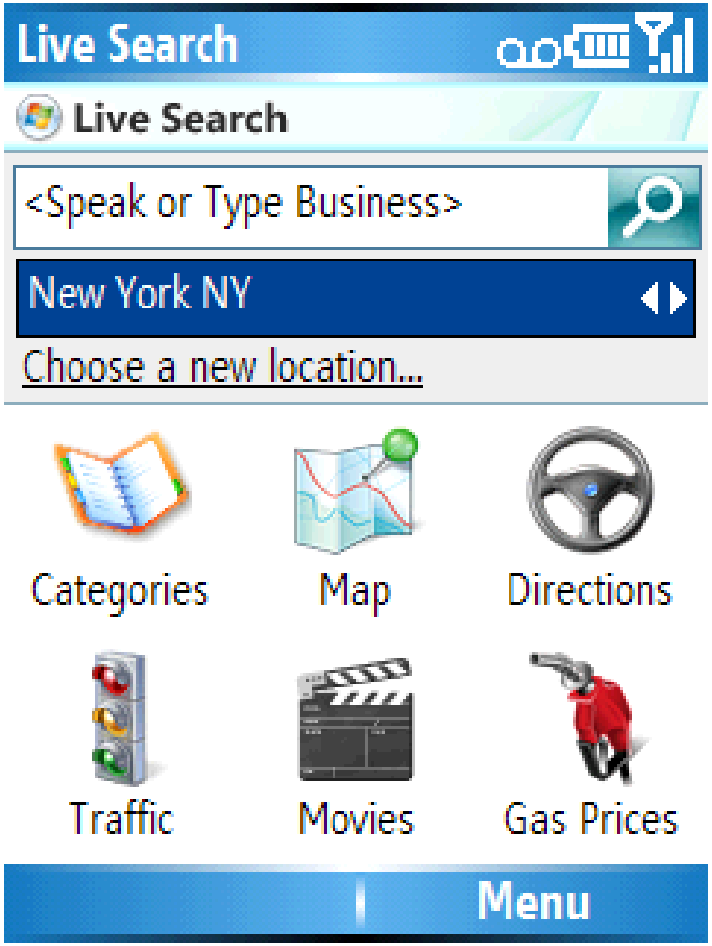
Outline

- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important?
- The Competitive Landscape
- Basic Technology
- Advancing the State-of-the-Art
- Next generation Applications

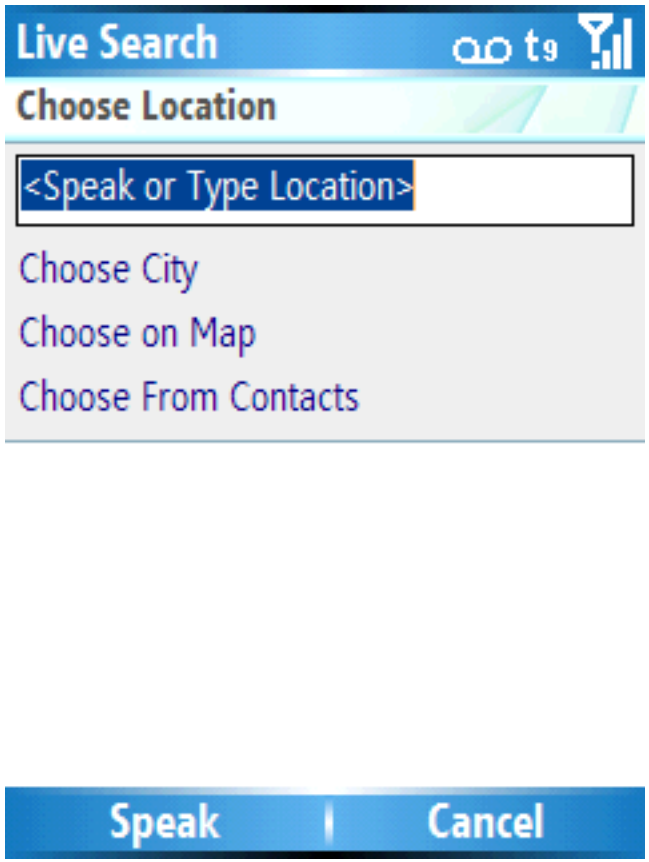
What is Mobile Voice Search?

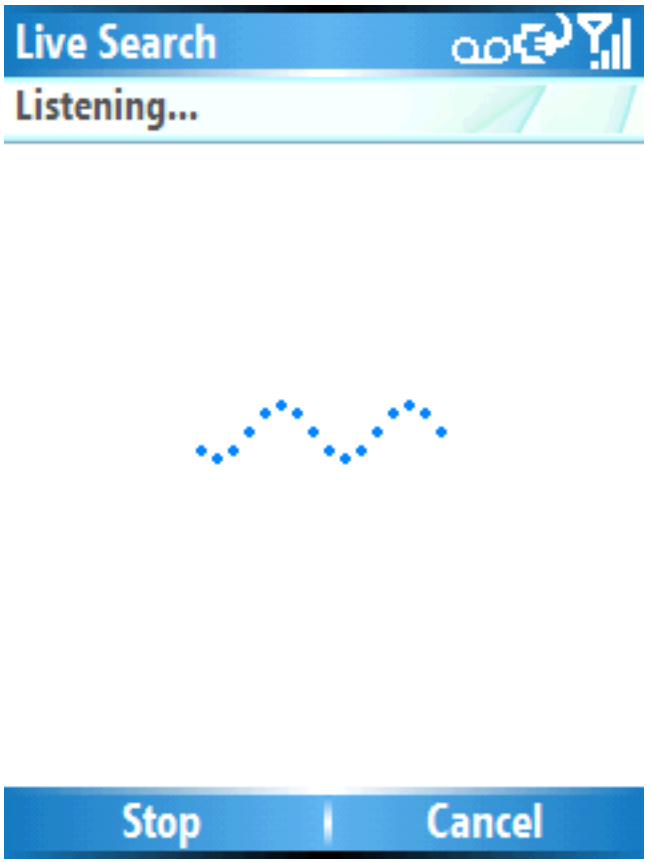
- Getting information when you are on-the-go
- Business-information
  - Phone numbers
  - Addresses
  - Ratings
  - Hours
- Maps & Directions
- Entertainment
  - Movie showtimes
  - Restaurant recommendations

Live Search for Windows Mobile

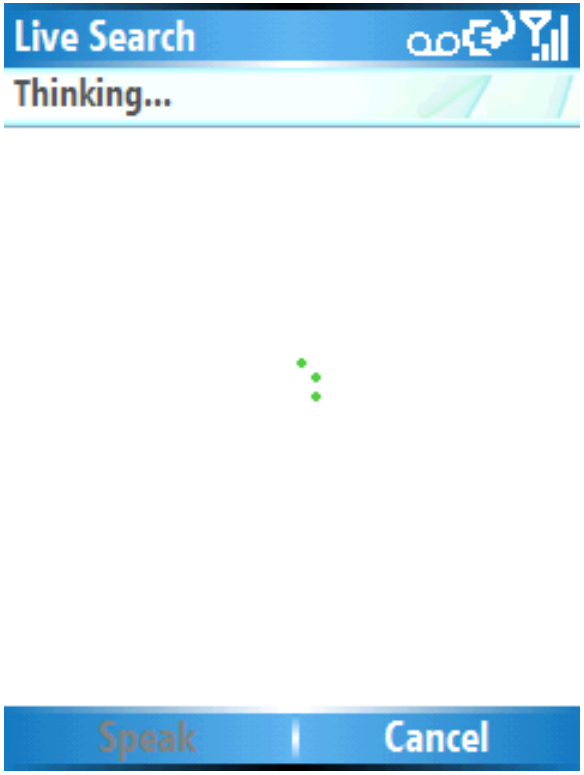


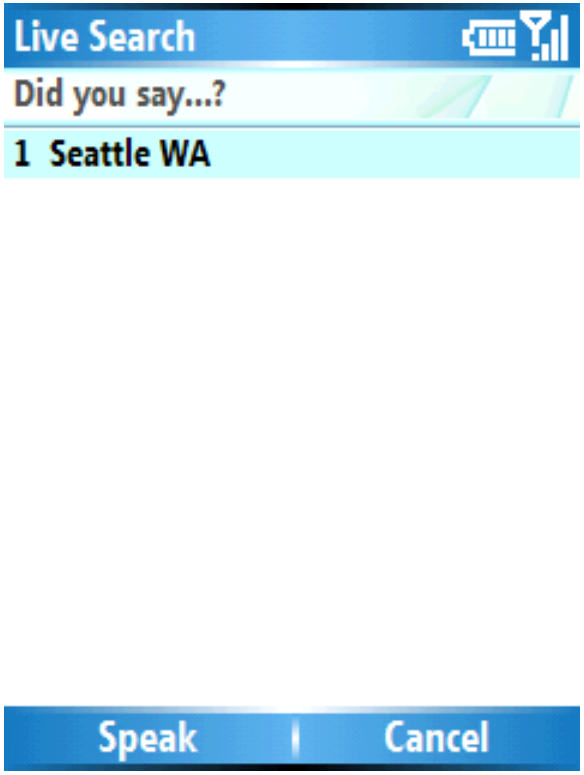
Asking for Seattle



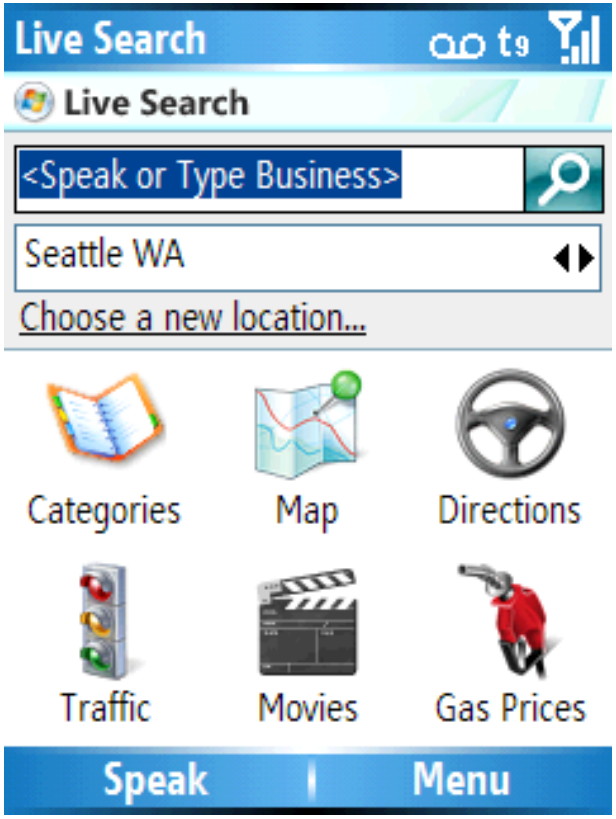


Confirming the Location

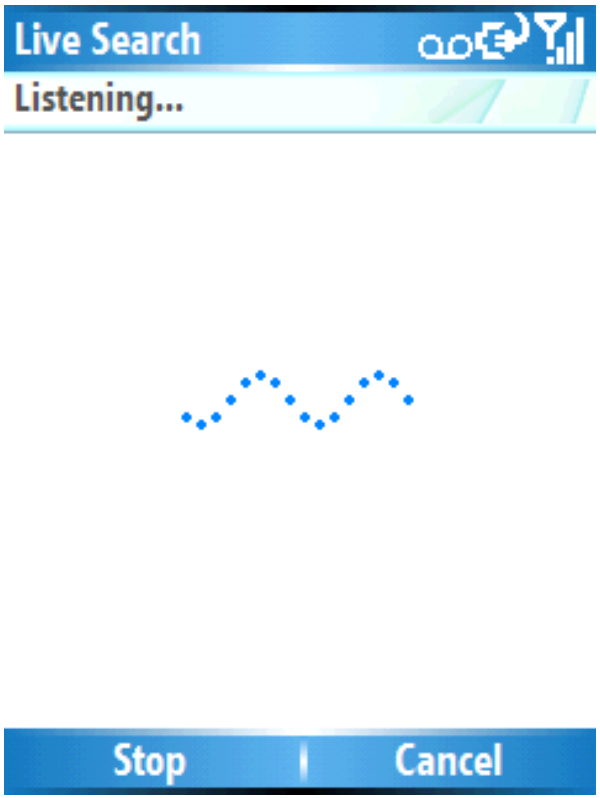


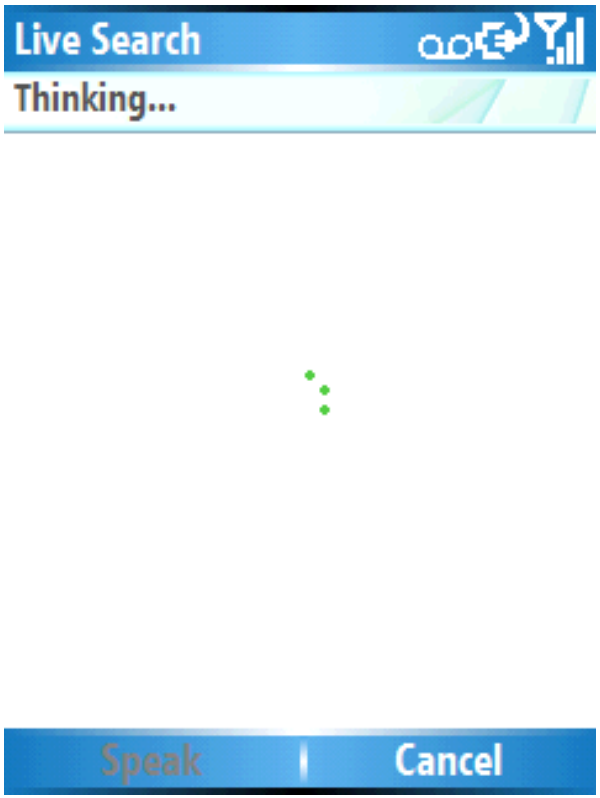


Now we're in Seattle

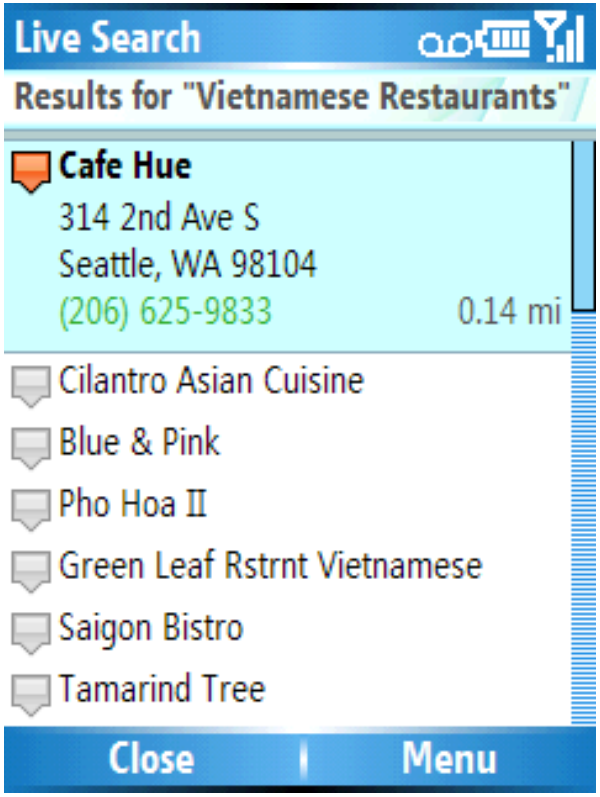


Asking for Vietnamese Restaurants

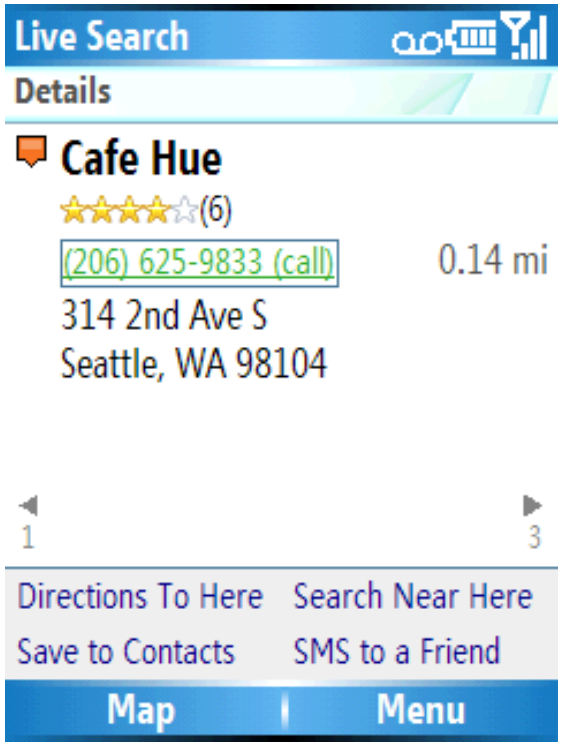




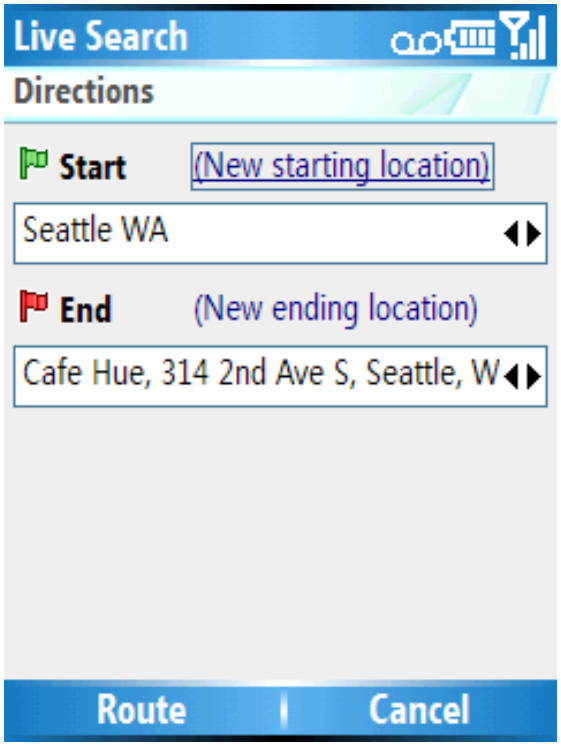
Finding a Vietnamese Restaurant



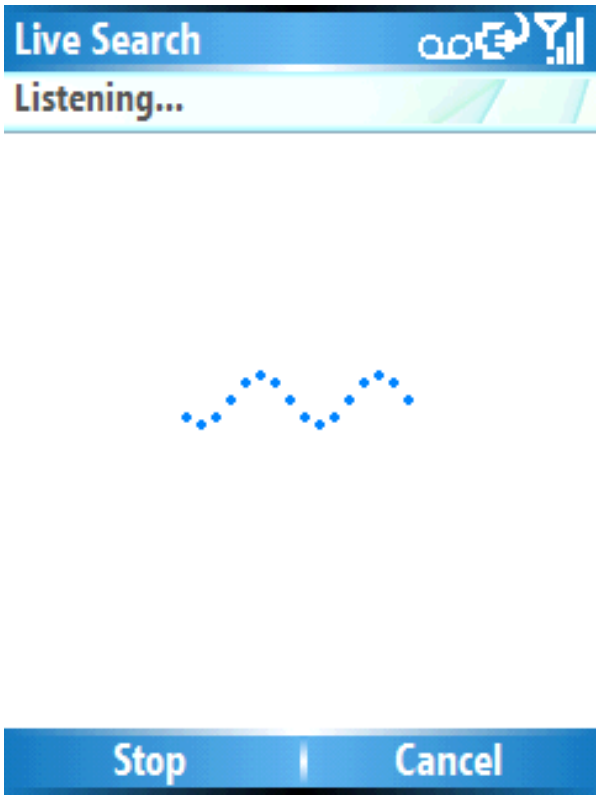
The Details

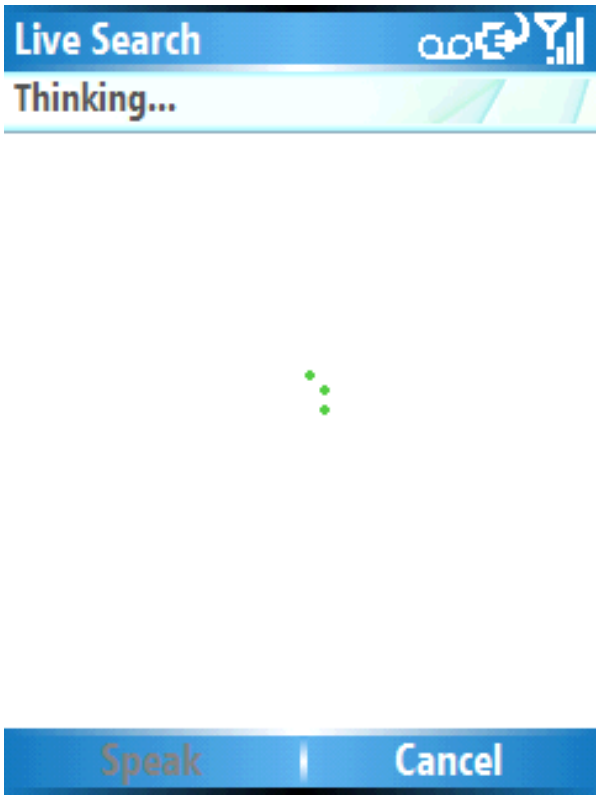

Live Search
Details
Cafe Hue
(6)
(206) 625-9833 (call)
0.14 mi
314 2nd Ave S
Seattle, WA 98104
1
3
Directions To Here
Search Near Here
Save to Contacts
SMS to a Friend
Map
Menu

Let's Get Directions

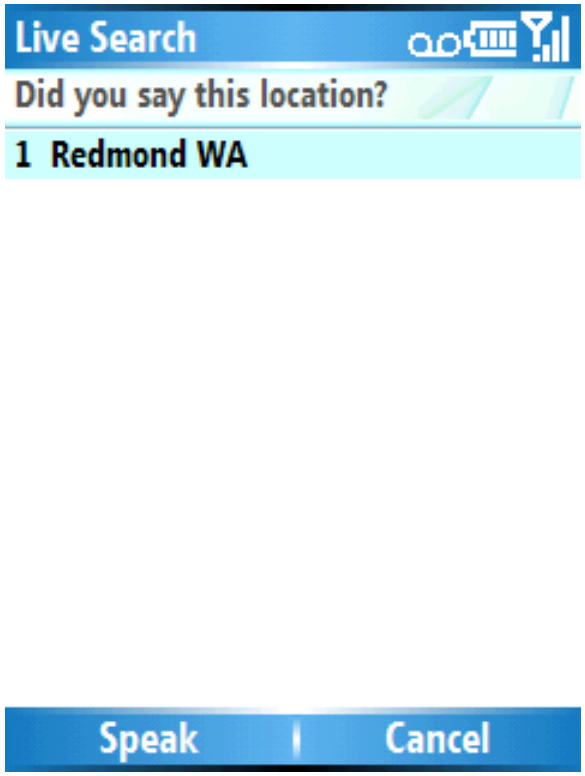


Starting from 8350 159th PL NE Remond, WA

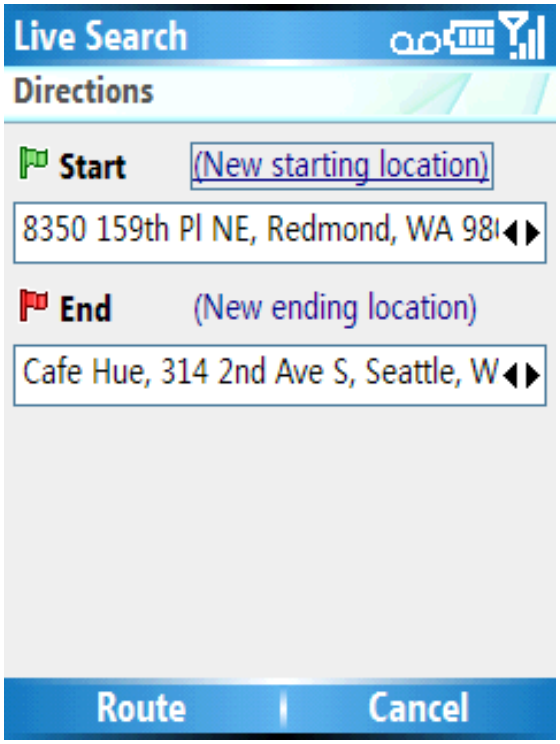


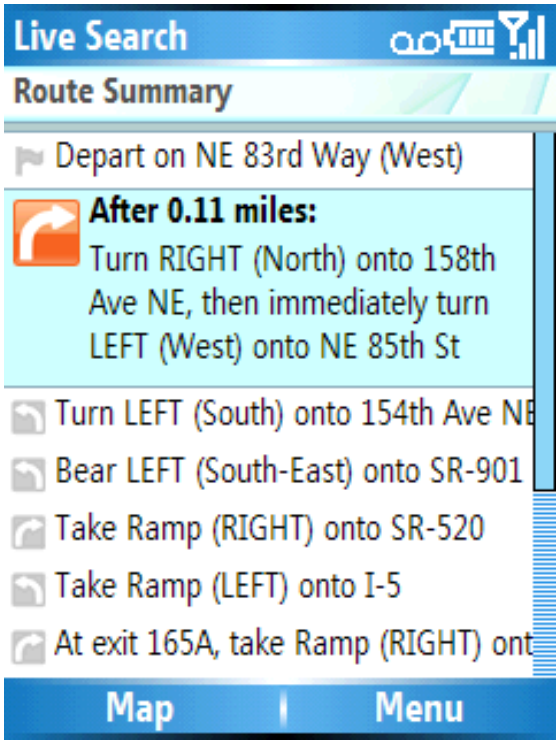


Specifying a Starting Point

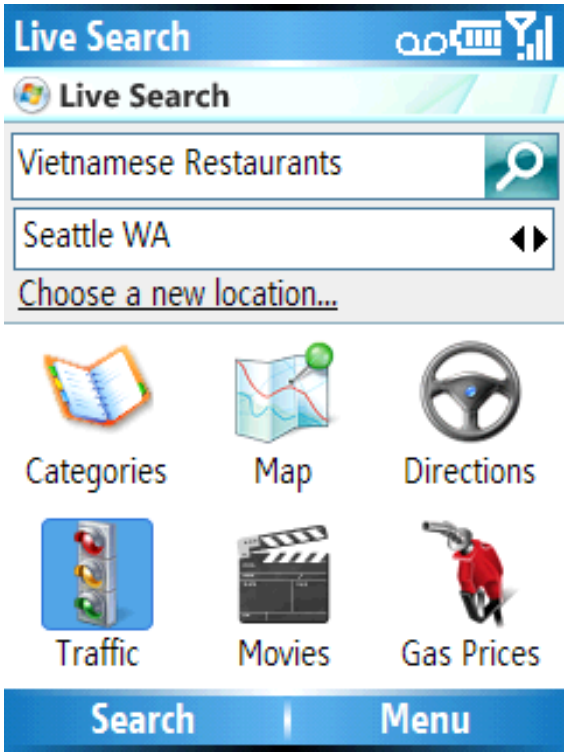


And Now we can Go!

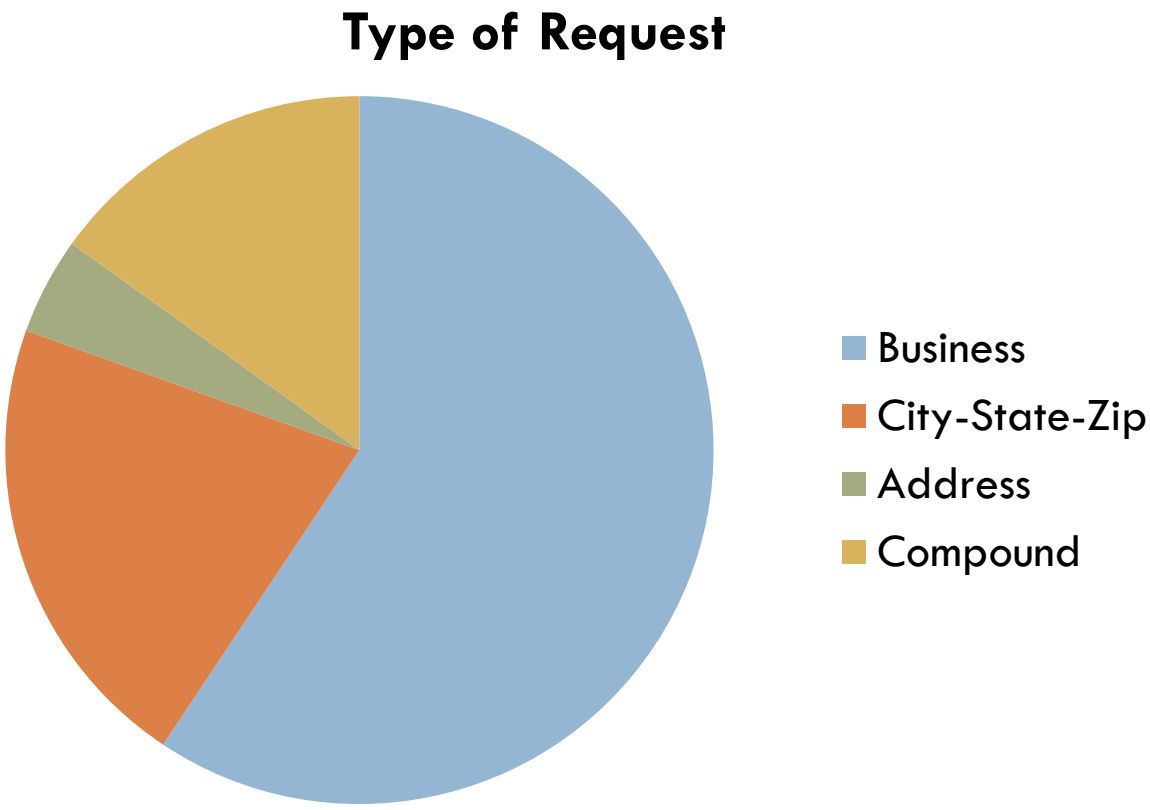




You can even check the traffic



What People Ask For – By Type



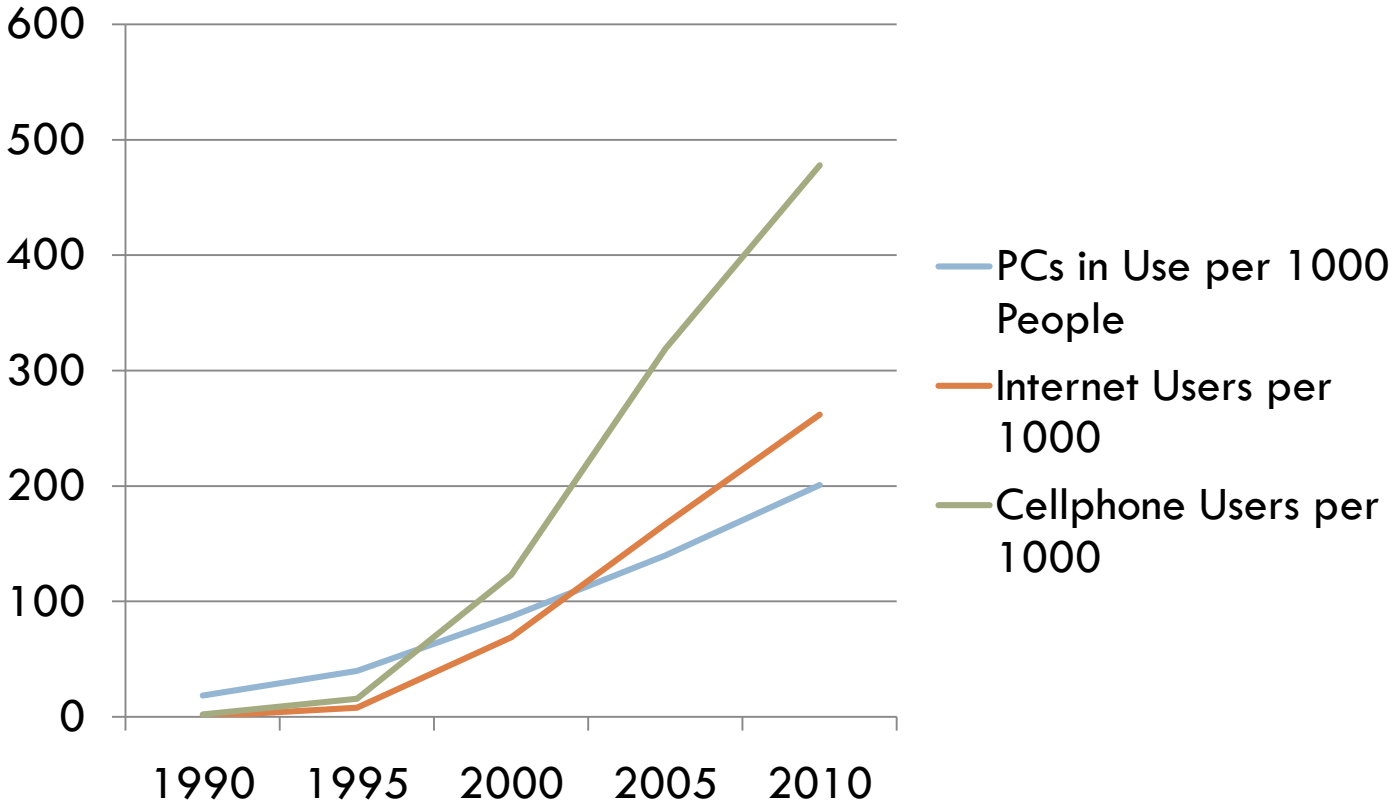
Frequent Requests

Businesses		Cities	
Pizza	(1.5%)	Dallax TX	(0.80%)
Best Buy		Seattle WA	
Starbucks		Chicago IL	
Movies		Redmond WA	
McDonald's		Los Angeles CA	
Wal-Mart		Orlando FL	
Mexican Restaurant		Miami FL	
Pizza Hut		Bellevue WA	
Target		San Diego CA	
Restaurants	(0.73%)	New York, NY	(0.47%)
Perplexity = 8514		Perplexity = 4741	

Outline

- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important?
- The Competitive Landscape
- Basic Technology
- Advancing the State-of-the-Art
- Next generation Applications

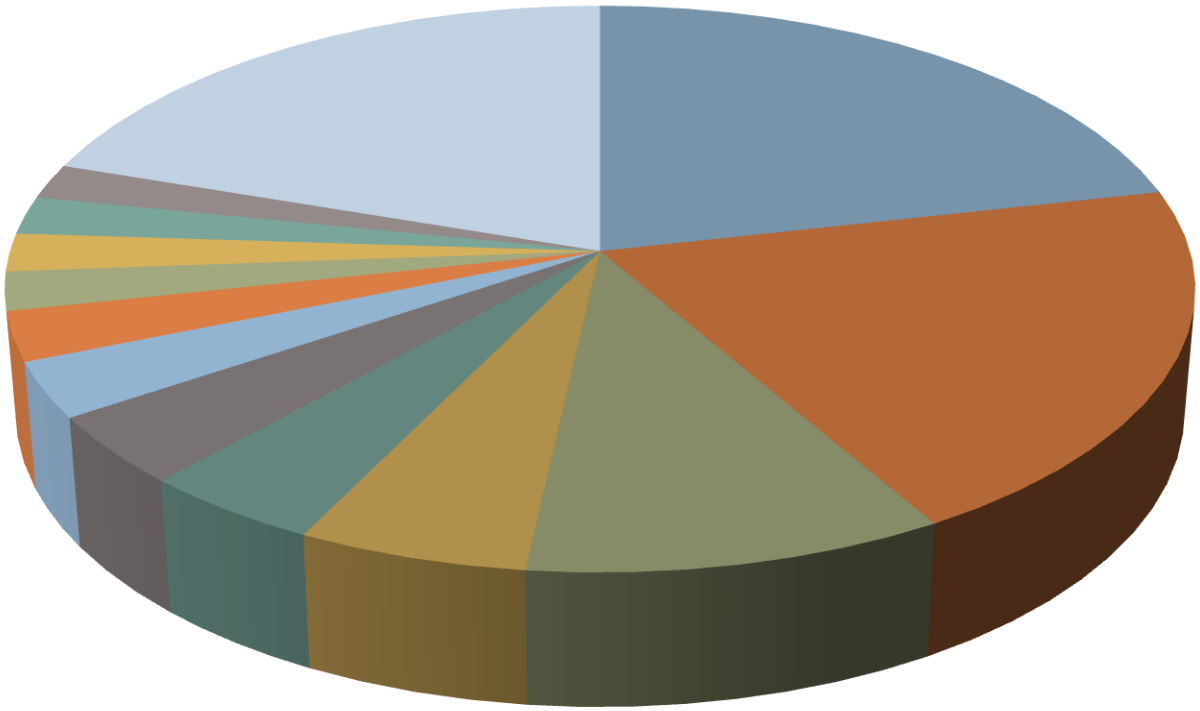
Skyrocketing Cellphone Use



Computer Industry Almanac

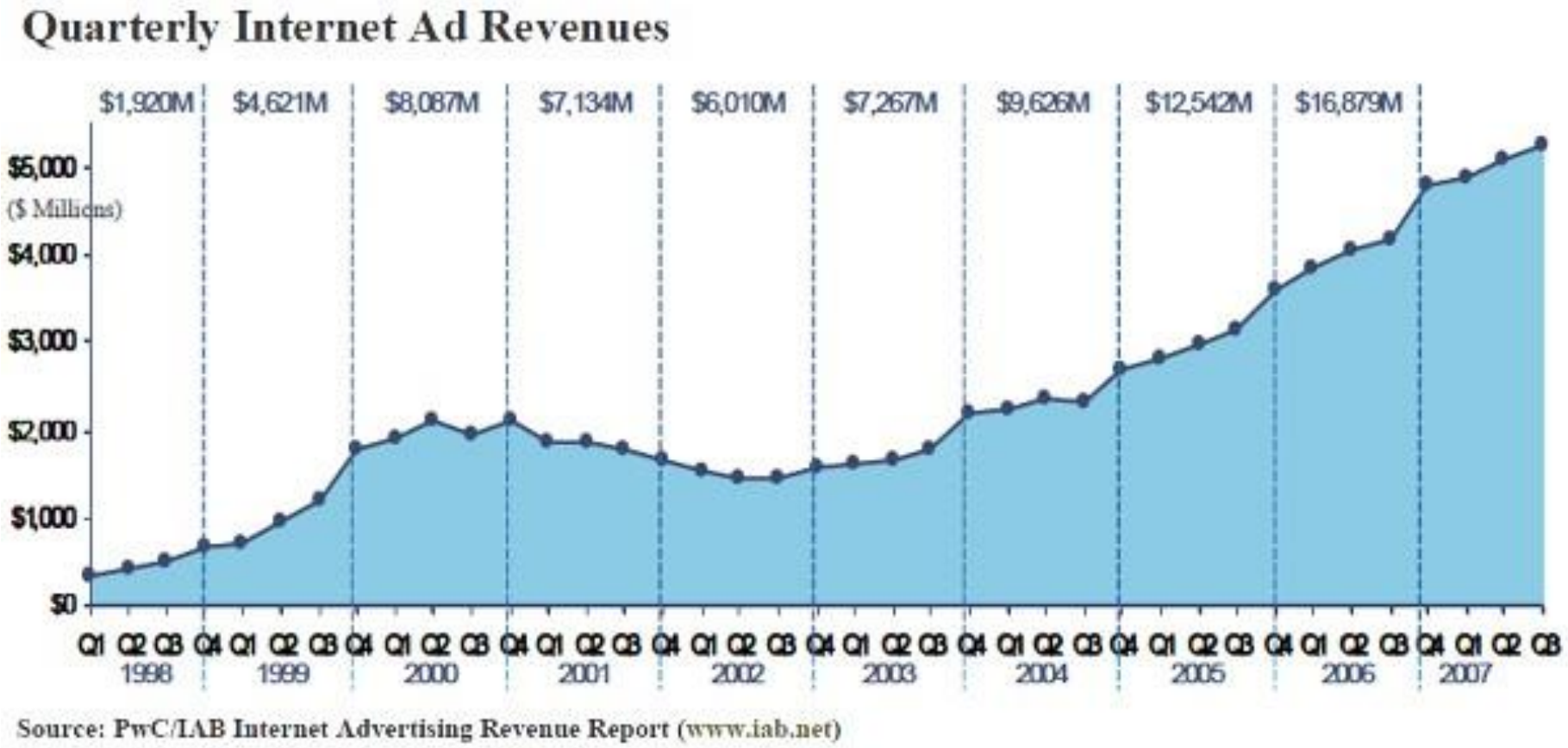
It's a Global Market

Number of Cellphones: ~2.2B in 2005



Infoplease.com

Potentially Big Revenues



Will mobile search be like internet search?

Monetization

- Free 411 services create modest revenue streams
- But multimodal has advantages:
  - You are looking at a screen
  - You can be sms'd and that sticks around
  - Voice provides demographic clues not present in web search – gender, race, age, education
- Many possibilities
  - Standard search-specific advertising
    - You say "Zales Jewelers" system suggests "Tiffany's"
  - Demographically targeted ads
    - Men get different results from women
  - Batched ads sent to email account provided at registration

Outline

- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important?
- The Competitive Landscape
- Basic Technology
- State-of-the-Art
- Next generation Applications

Competitive Landscape: Basic Search

- Live Search for Windows Mobile
  - http://wls.live.com from your phone
  - Businesses, directions, maps, traffic, movies, gas
  - Windows Mobile phones
- Tellme by Mobile
  - http://www.tellme.com/products/TellmeByMobile
  - Businesses, directions, maps
  - Java phones
- V-enable
  - http://www.v-enable.com/directory_assistance.html
  - Businesses, directions, maps, weather
  - Demo only – not currently available

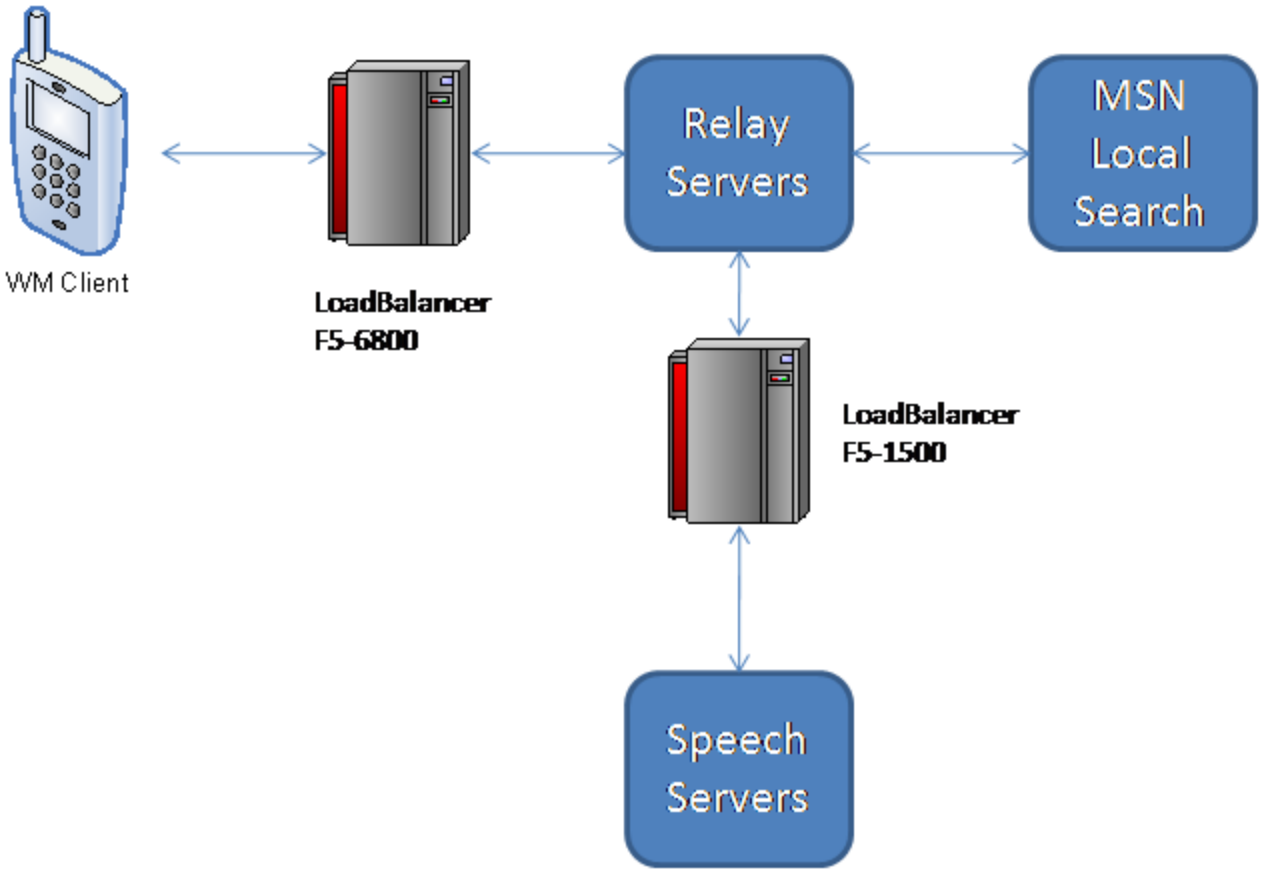
Competitive Landscape: Beyond Search

- Vlingo
  - http://vlingo.com/
  - Businesses, directions, maps, music downloads
  - sms by voice
  - Java phones
- Nuance Voice Control
  - http://www.nuance.com/voicecontrol/
  - Businesses, directions, maps, weather, stocks, sports, movies, web search
  - Send emails, update calendar, go to web pages
  - Blackberry, Treo, Windows Mobile phones

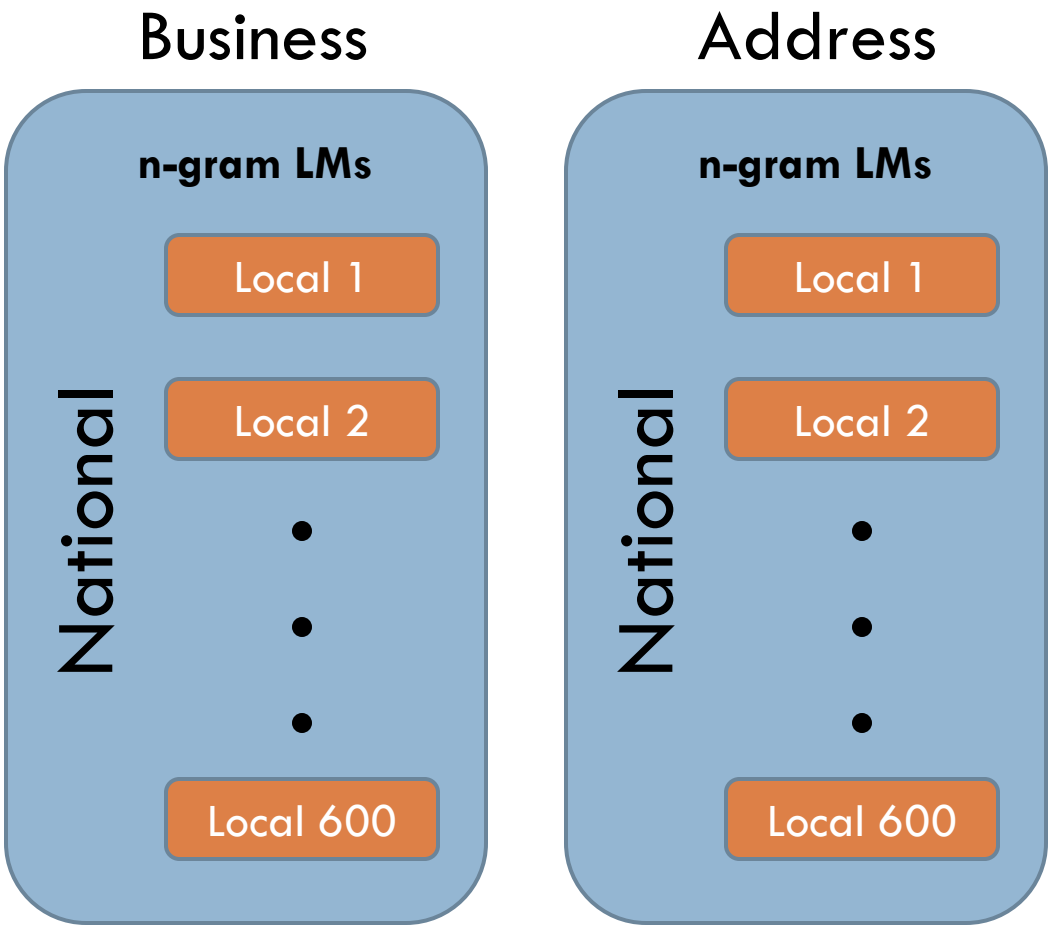
Outline

- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important? -- Trends in Cellphone use
- The Competitive Landscape
- Basic Technology
- Advancing the State-of-the-Art
- Next generation Applications

Client-Server Architecture

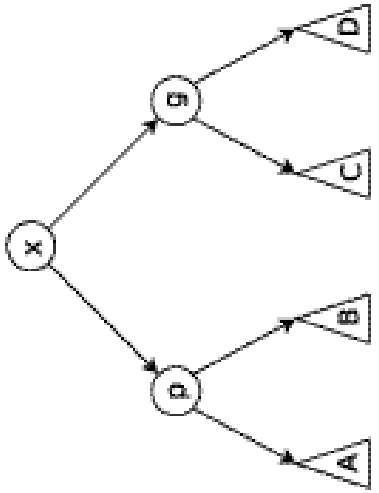


Typical Grammar Setup



City-State-Zip

Enumerated grammar



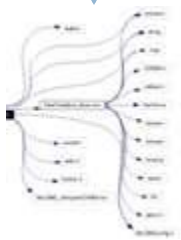
Sample Performance Levels

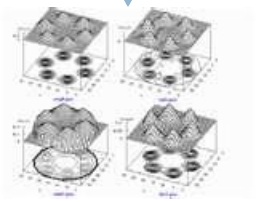
	1-best	N-best	N-best depth	Inter-annotator agreement
Overall	42%	47	3.6	67%

Outline

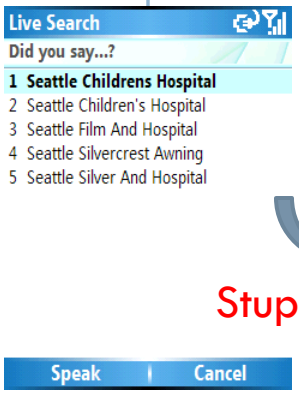
- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important? -- Trends in Cellphone use
- The Competitive Landscape
- Basic Technology
- Advancing the State-of-the-Art
- Next generation Applications

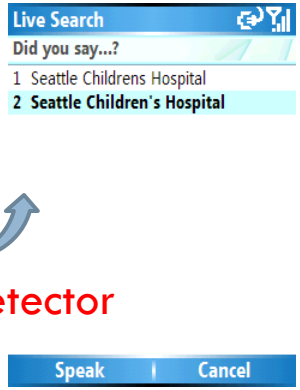
Click-Driven Automated Feedback





Language Model Acoustic Model





Stupid Detector

Automated Feedback Methods

- Data addition
  - What people click on & associated audio
  - Text searches from web
- Discriminative LM training
  - Adjust LM to maximize posterior probability of correct words
  - Need to know competitors – from nbest lists
- Translation-based data generalization
- Maximum likelihood database cleaning
  - Learn error model of the mistakes people make when entering data
  - Recover the likeliest intended entries
- Adaptive N-best postprocessing
  - Remove what history shows is obviously stupid
  - Reorder and augment the rest based on further analysis
- Personalization
  - Per-person / user-profile grammars
  - Per-person speaker-adaptive transforms

Sample Click Data

Entries that frequently co-occur

Clicked	Competitor
McDonald's	Mc Donald
Coffee	Coffey
Mexican Restaurant	Mexican Restrant
Coffee	Copy
Mexican Food	Mexican Foods
Starbucks	Star Box
Starbucks	Starbuck's
Sex	6
Burger King	13

Discriminative LM Training (Xiao Li)

- Idea
  - Increase n-gram probabilities of the true hypothesis
  - Decrease n-gram probabilities of confusable competitors
- The LM is estimated to maximize p(W|O)
- Leveraging click data
  - View clicked item as "truth"
  - View n-best alternatives as "competitors"

N-best alternatives

1. Maine Home
2. Maine School
3. Maine Car
4. Maine
5. Maine Heart
6. **Maine Mall**
7. Maine Homes
8. Mayo
9. Maine Golf
10. Maine Home Care

Rescoring Results

- Experiments:
  - Rescore n-best alternatives using the baseline LM and discriminatively trained LM
  - Inspect if the rescored one-best is the user clicked item

One-best Acc	Train Set	Dev set	Test set
# utterances	150K	1.3K	1.4K
Baseline	71.1%	71.5%	70.5%
Discriminative Training	-	74.8%	72.7%

Fraction of time the clicked item is at the top of the n-best.

Translation LM (Xiao Li, ICASSP-08)

- Goal:
  - "Translate" listing forms to query forms
  - Use translated query forms to augment the training data for LM estimation.
- Example

  listing Kung Ho Cuisine Of China can have
  - *"Kung Ho Chinese Restaurant"*
  - *"Kung Ho Restaurant"*
  - *"Kung Ho"*

Recognition Results

- Experiments
  - Test set: 3K directory-assistance utterances
  - Different LM training sets:

Sentence accuracy	One-best	N-best
Listings	38.6%	48.3%
Listings + transcription	41.5%	51.4%
Listings + transcription + translation	43.1%	52.5%

Maximum Likelihood Database Recovery

Wi: intended words (unknown, e.g. “Starbucks” or “Al’s Quick Mart”)
Wc: Corrupted words in data (observed, e.g. “Starbuck’s” or “Al’s Kwik Mart”)
Want to find the likeliest intended word sequence

$$\arg\max_{wi} P(w_i \mid w_c) = \arg\max_{wi} \frac{P(w_i)P(w_c \mid w_i)}{P(w_c)}$$

$$= \arg\max_{wi} P(w_i)P(w_c \mid w_i)$$

LM built on clean data

Error model

wi	wc	P(wc | wi)
Starbucks	Starbuck’s	0.5
Starbucks	Starbucks	0.5
Quick	Quick	0.3
Quick	Kwik	0.3
Quick	Quik	0.3

Transductive aparatus used to recover the likeliest words

Maximum Likelihood Database Recovery++ (G. Zweig, ICASSP 2008)

W: intended words (unknown)
l_i: intended letters (unknown)
l_c: corrupted letters (observed)
Want to find the likeliest word and letter sequence underlying the observations

$$\arg\max_{w,l_i} P(w, l_i \mid l_c) = \arg\max_{w,l_i} \frac{P(w, l_i)P(l_c \mid w, l_i)}{P(l_c)}$$

$$= \arg\max_{w,l_i} P(w)P(l_i \mid w)P(l_c \mid w, l_i)$$

$$= \arg\max_{w,l_i} P(w)P(l_i \mid w)P(l_c \mid l_i)$$

LM built on clean data

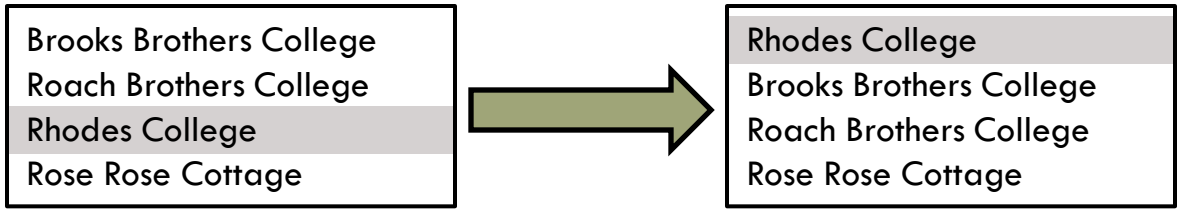
1:1 Spelling probabilities

Error model for typos

Database Recovery Steps

- Learn error model by aligning letters of click-pairs
  - Coffey vs. Coffee
  - Starbuck's vs Starbucks
- Learn language model from current version of database
- Letter-to-word from a list of in-language words
- Run database letters through transductive aparatus to recover words

Feedback-Driven N-best Postprocessing (Dan Bohus)



- Approach
  - Click prediction model

$$P(Click|f) = \frac{e^{\bar{\alpha}\cdot\bar{f}}}{1 + e^{\bar{\alpha}\cdot\bar{f}}}$$

  - Features
    - Recognized words
    - Historical click-through rates
    - Intra n-best comparisons
    - User-specific features
    - Text query log features

- Preliminary Results
  - 23% improvement in average position of clicked item

Outline

- What is Mobile Voice search?
  - An example: Live Search for Windows Mobile
- Why is it important?
- The Competitive Landscape
- The Technology
- Advancing the State-of-the-Art
- Next generation Applications

Next Generation Applications

- Better integration with information sources
  - Unstructured information
    - The web – "www dot langtech dot org"
  - New kinds of structured information
    - Product information
    - Movie reviews
    - Nutrition information – "Do apples have vitamin D?"
- Access to private information
  - "Show me my benefits information on the company website"
  - "Show me the email from Langtech about the banquet"
- Two-way interaction
  - Rating products and businesses

VoiceRate – A Sample NextGen Application

Get a rating or leave a rating?

Get a rating.

Leave a rating.

- Local business, National Business, or Product ?

- A product.

- Which product?

- Fisher Price Kick-Play Bouncer

- This is rated 4.3 out of 5. Here are some of the things people had to say about this product: …

- Local business, National Business, or Product ?
- A product.

- Which product?
- Stanley 9 piece screwdriver set

- Please rate it.
- Excellent. It works great.

- For comparison with others, please rate it on a scale of one to five.
- Five.
- Thank you for using Voice Rate.

VoiceRate Benefits

- User Benefits:
  - Facilitates informed impulse purchases
  - Let's you provide immediate feedback
  - Access to ratings for:
    - 1.1M products (electronics, toys, books, DVDs, etc.)
    - 270k restaurants (local businesses) in 1600 metros
    - 3k national businesses (airlines, car rental companies, etc.)
- Researcher Benefits:
  - Fertile test-bed for many technologies
    - Understanding verbal reviews
    - Summarizing across multiple reviews
    - Making pair-wise comparisons
    - Explaining why people like X better than Y
    - Core ASR
  - Data collection

Provider Benefits

- Sales of Targeted ads
  - Ask about Toro Snowblower; Snapper Snowblowers pays to suggest their product
  - Determine caller demographics by voice – tailor ads
- Sale of market research services
  - When a person *leaves* a review
    - For example, if you call to review a lawnmower, Honda can pay to ask "Did the mower cut the grass evenly?"
  - When a person *gets* a review
    - If I call and ask about the Toro Power Curve Snow-blower, Toro can pay to ask: "To help determine if there are any better products, how important is noise to you in a snowblower?"
- Location-specific ads
  - If you are in a Target store and call about X, that Target can to offer you a deal.

Conclusions

- Mobile Voice Search is a key technology area
  - Impact on a large fraction of the world's population
  - Global in scope
- Multi-modal interfaces are key
  - Speech recognition is necessary because data entry just too hard otherwise
- Click-driven feedback will drive system improvements
- Current applications are just scratching the surface

Thanks to VoiceSearch Collaborators!

- Xiao Li
- Dan Bohus
- Patrick Nguyen
- Julian Odell
- Oliver Scholz
- Alex Acero