

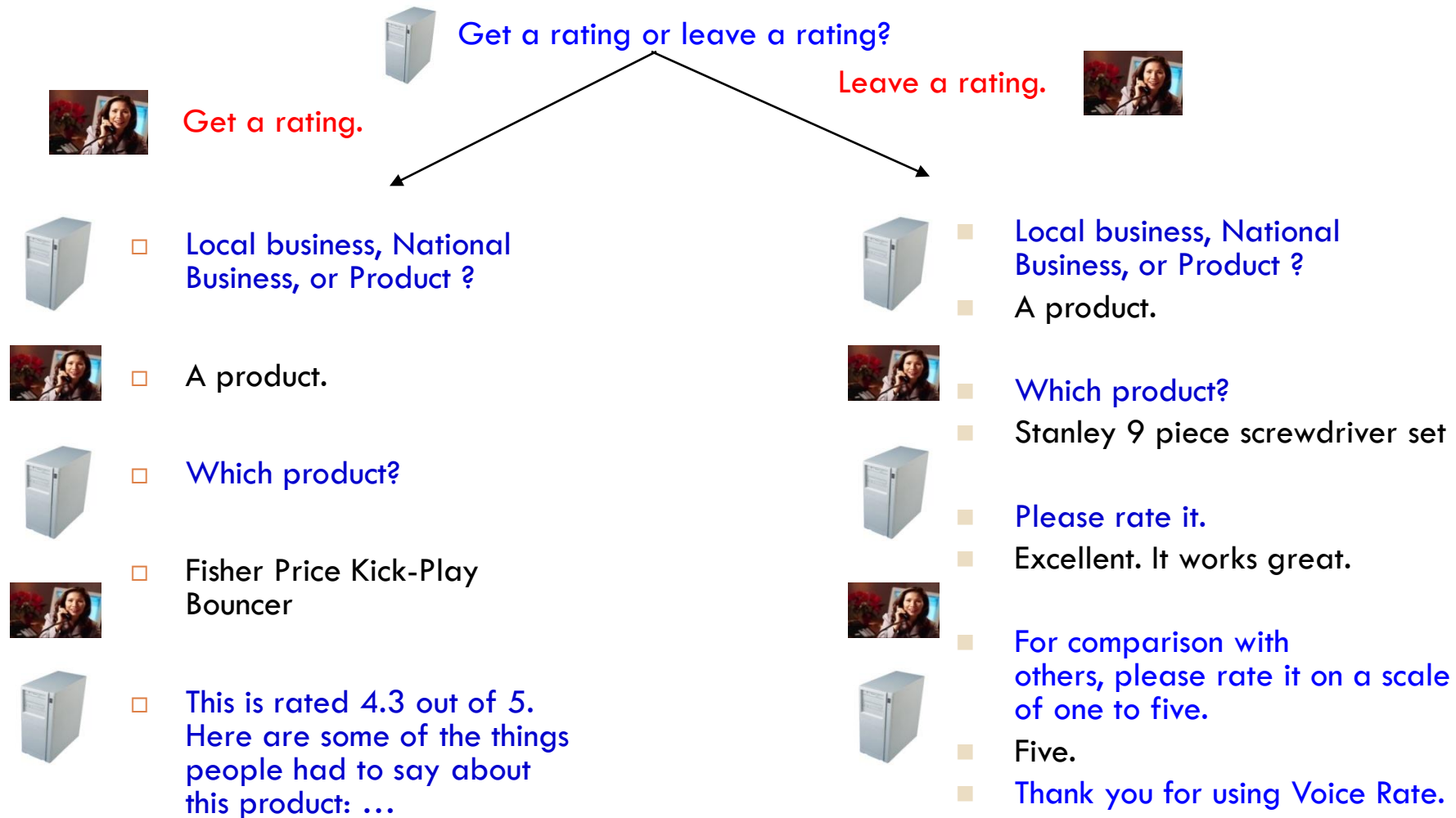
NLP CHALLENGES IN VOICESEARCH

Geoffrey Zweig

Key Points

- Current systems are “plain vanilla,” e.g.:
 - ▣ 1-800-Call411, 1-800-555-TELL, 1-800-Free411
- In more complex domains, we must use NLP to go beyond “here is the phone number”
- Getting it right will enable
 - ▣ Interactions in richer domains - through summarization
 - Restaurant, product reviews
 - Email summarization
 - ▣ Community input – through automatic moderation
 - When people provide information, it is often off-topic or offensive
 - ▣ Better information navigation

VoiceRate – Summarization in a Rich Domain



The Alki Homestead

- Positive Comments:
 - “Nice fluffy biscuits and great service”
 - “It is a must for me to have dinner at the ALKI homestead”
 - “My favorite dishes are the broiled salmon and the prime rib”
- Negative comments:
 - “I have never in my life had such rude service”
 - “The baked potatoes were either overcooked or undercooked”
 - “Crappy service”

How it Works

- Problem: Want to convey information that will help make an informed decision. Verbatim playback of dozens of reviews would be too tedious.
- Idea: pick snippets of text that express extreme opinions and play those.
- Execution:
 - ▣ Train max-ent classifier to predict numerical rating from unigram and bigram features
 - ▣ Select snippets of text with large net max-ent score wrt positive/negative ratings
 - ▣ Snippet is punctuation-delimited text
 - ▣ Nguyen et al. '07

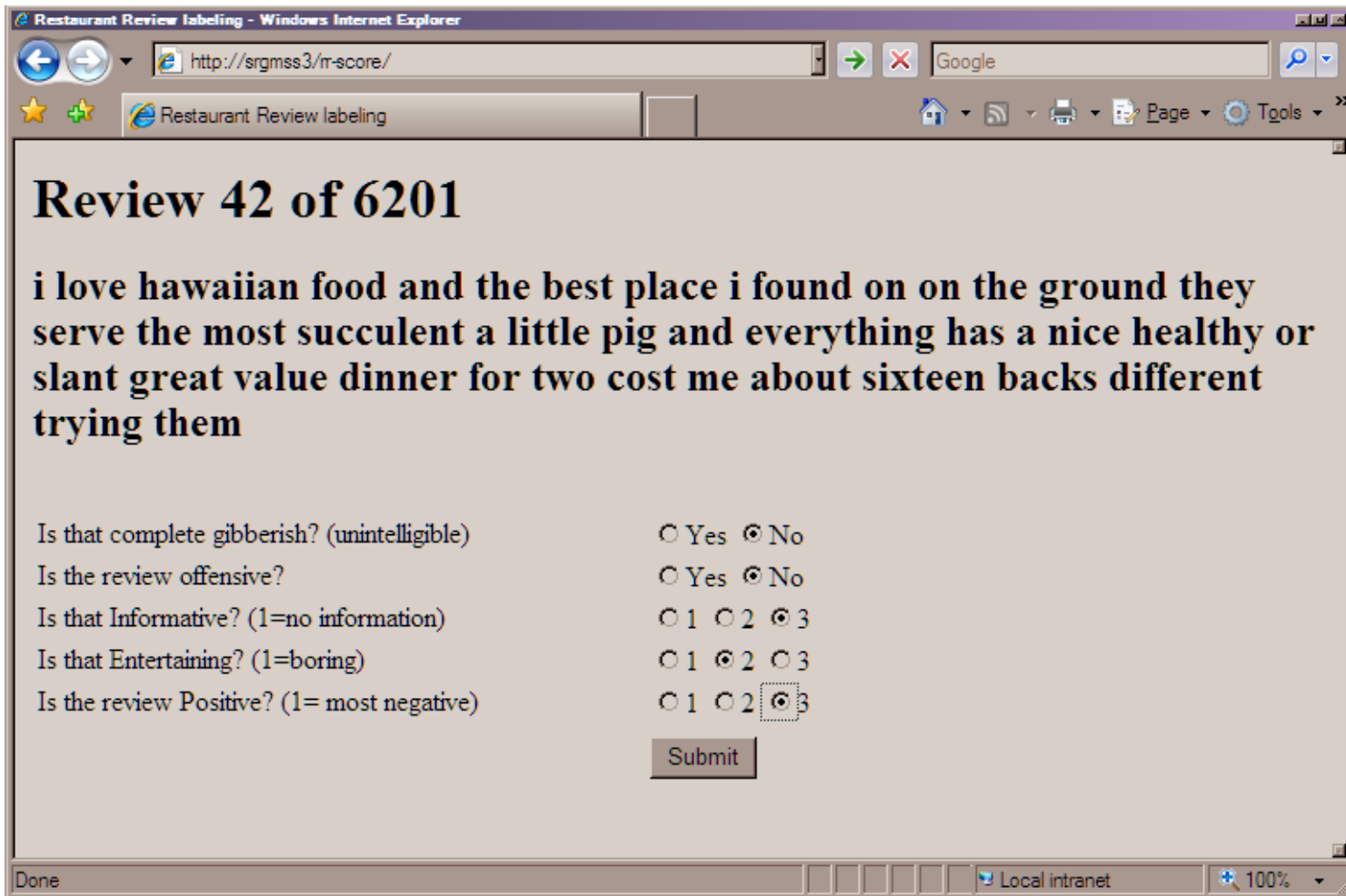
Heavily Weighted Features

N-gram feature	Snippet
Worst pizza	I had some of the worst pizza I've ever been served outside a microwave box
Was excited	I read all of the wonderful reviews and I was excited to try it
Worst restaurant	Absolutely the worst restaurant I've been to in years
Courteous	Courteous and friendly owners
Worst food	Was the worst food experience I have had in years in SF

NLP for Content Moderation in VoiceSearch

- Want to let people contribute to the system by leaving reviews and then playing some. But:
- Donated reviews are often inappropriate
 - ▣ Off-topic
 - ▣ Offensive
 - Cursing, swearing
 - Denigrating immigrants, various sexual orientations, different races, religions and other social groups
 - Sexually explicit
 - ▣ Plain ranting

Data Collection to Build a Scrubber



The screenshot shows a Windows Internet Explorer browser window titled "Restaurant Review labeling - Windows Internet Explorer". The address bar contains "http://srgmss3/m-score/" and the search bar contains "Google". The page content is as follows:

Review 42 of 6201

i love hawaiian food and the best place i found on on the ground they serve the most succulent a little pig and everything has a nice healthy or slant great value dinner for two cost me about sixteen backs different trying them

Is that complete gibberish? (unintelligible) Yes No

Is the review offensive? Yes No

Is that Informative? (1=no information) 1 2 3

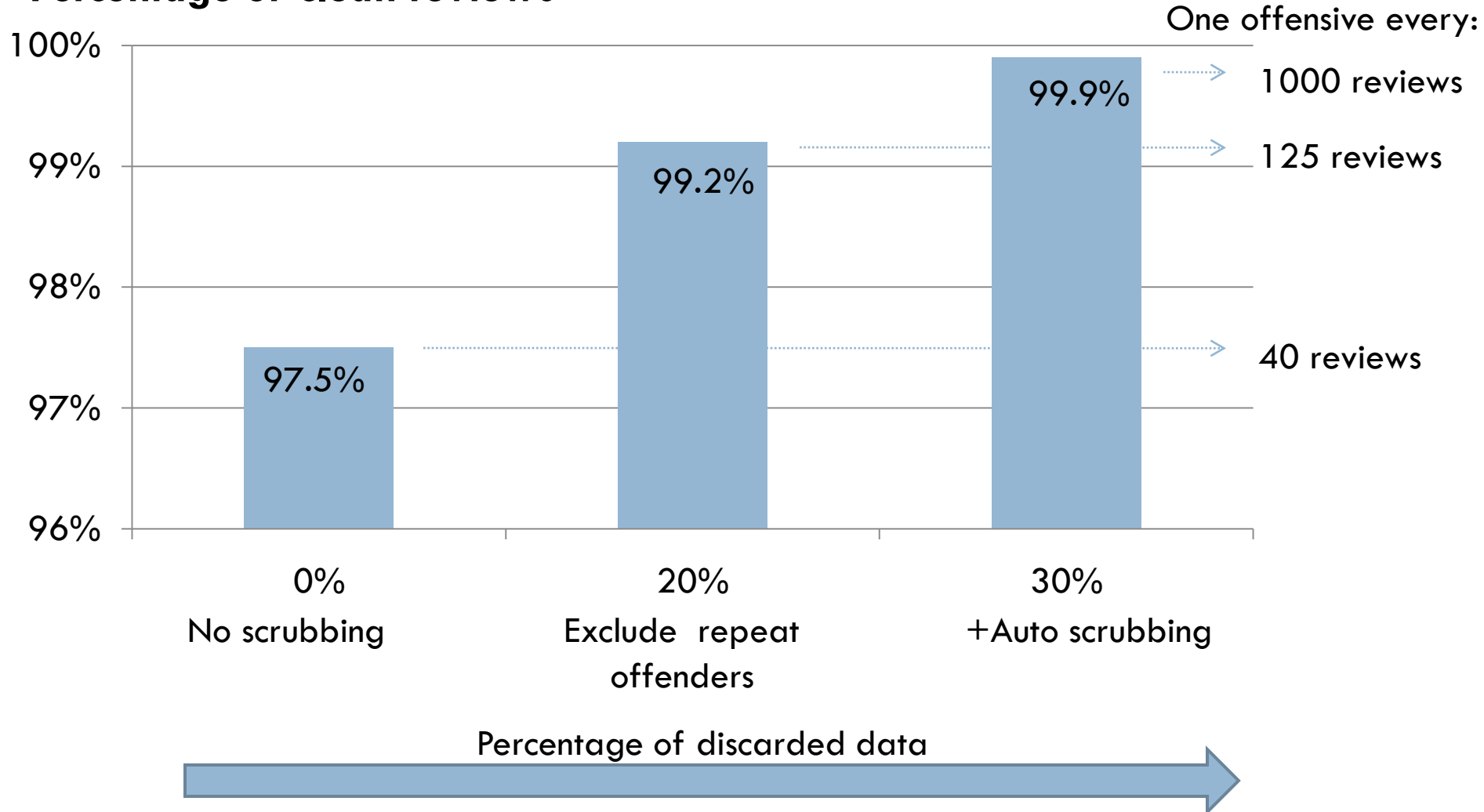
Is that Entertaining? (1=boring) 1 2 3

Is the review Positive? (1= most negative) 1 2 3

The browser status bar at the bottom shows "Done", "Local intranet", and "100%" zoom.

We Can Detect Offensiveness

Percentage of clean reviews



NLP for Navigation

What we have now without NLP

Live Search

Details

Cafe Hue
★★★★☆(6)
[\(206\) 625-9833 \(call\)](tel:(206)625-9833) 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

With NLP

Live Search

Details

Cafe Hue
★★★★☆(6)
[\(206\) 625-9833 \(call\)](tel:(206)625-9833) 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

Super Pho
Nasty Waiters
Great soup!

More Super Pho
Friendlier Waiters
Heavier Asian Cuisine



THANK
YOU