Natural Language Processing: Making Sense of Language with NLTK

Michael Prappas
EMIS 8331
March 9, 2017

New – *Play 'Today's Hits' station on Pandora*

Set an alarm for eight a.m.

Add gelato to my shopping list

New – *How is traffic?*

New – *What's on my calendar today?*

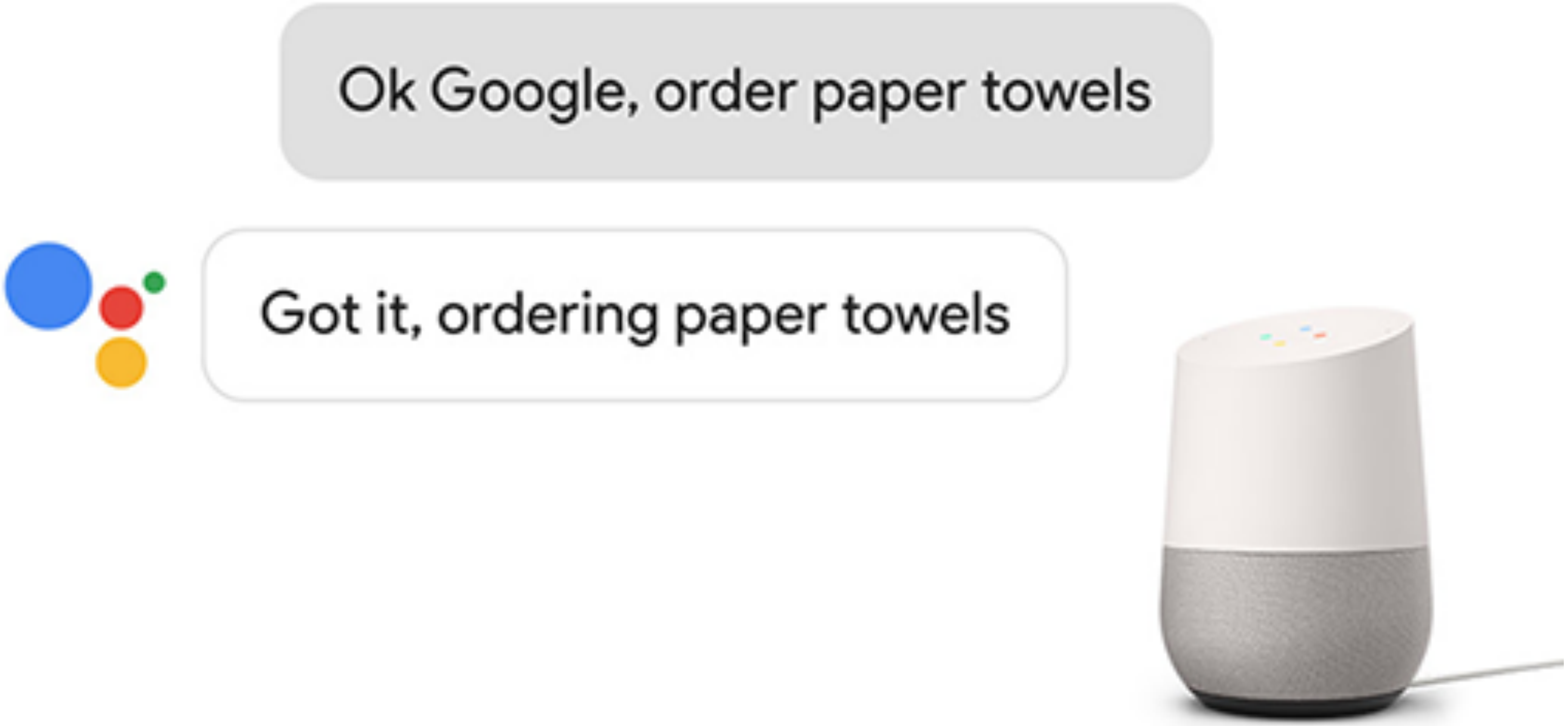
Wikipedia: Abraham Lincoln

New – *When do the Seattle Mariners play next?*

New – *Read my book*

What's the weather in Los Angeles this weekend?

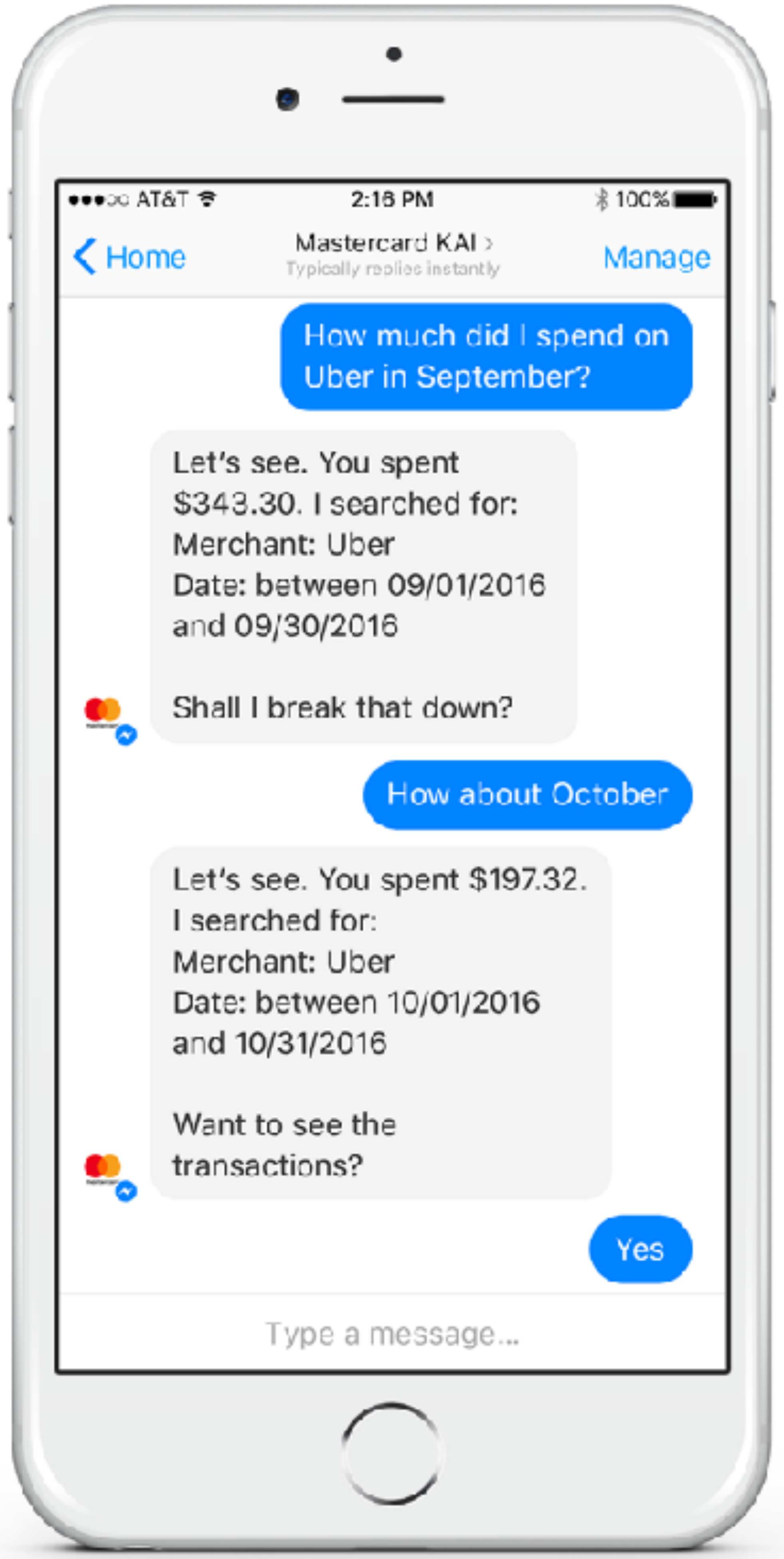
New – *Turn off the lights*



Ok Google, order paper towels

Got it, ordering paper towels

mastercard



How much did I spend on Uber in September?

Let's see. You spent $343.30. I searched for:
Merchant: Uber
Date: between 09/01/2016 and 09/30/2016

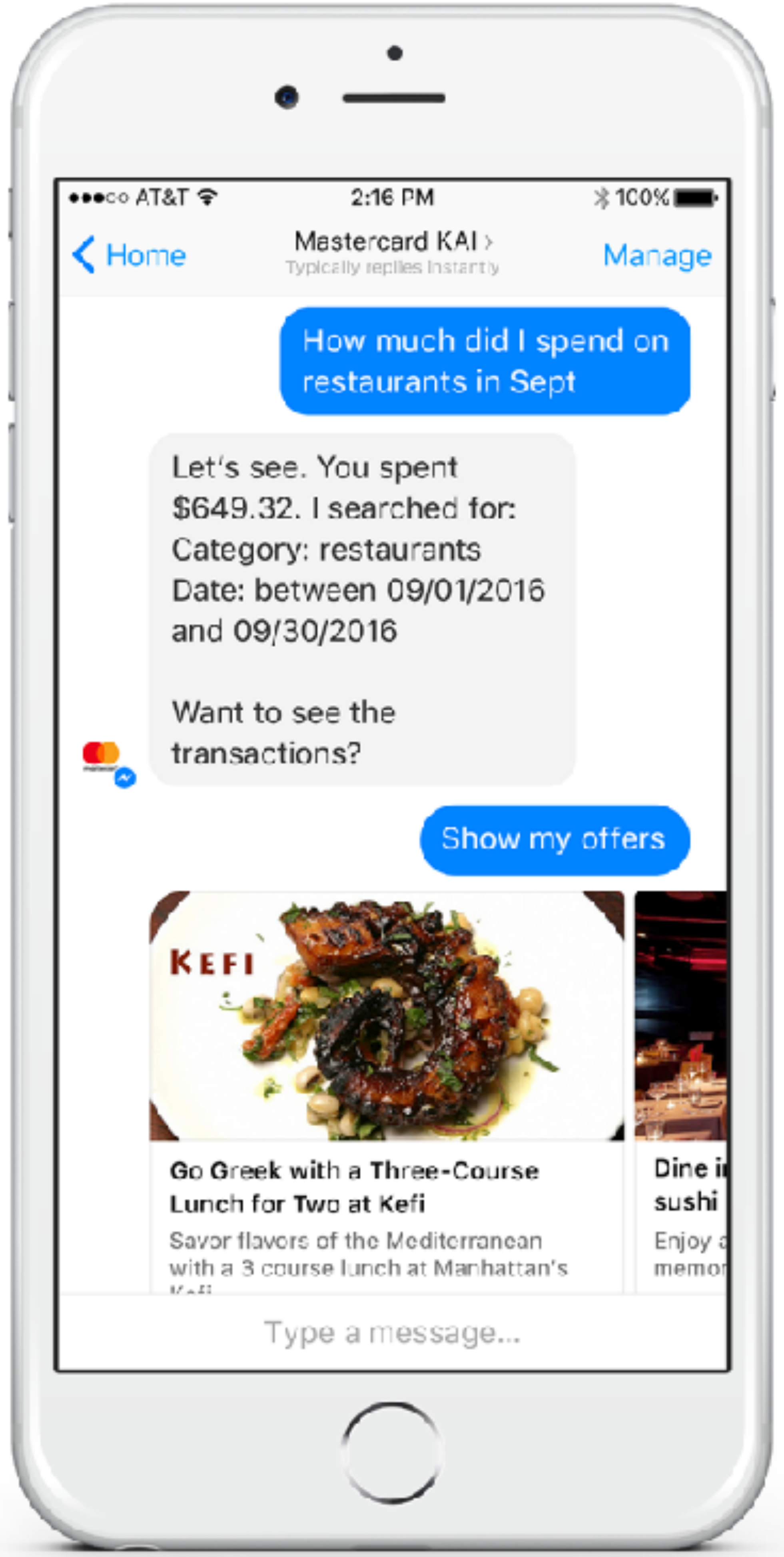
Shall I break that down?

How about October

Let's see. You spent $197.32. I searched for:
Merchant: Uber
Date: between 10/01/2016 and 10/31/2016

Want to see the transactions?

Yes



How much did I spend on restaurants in Sept

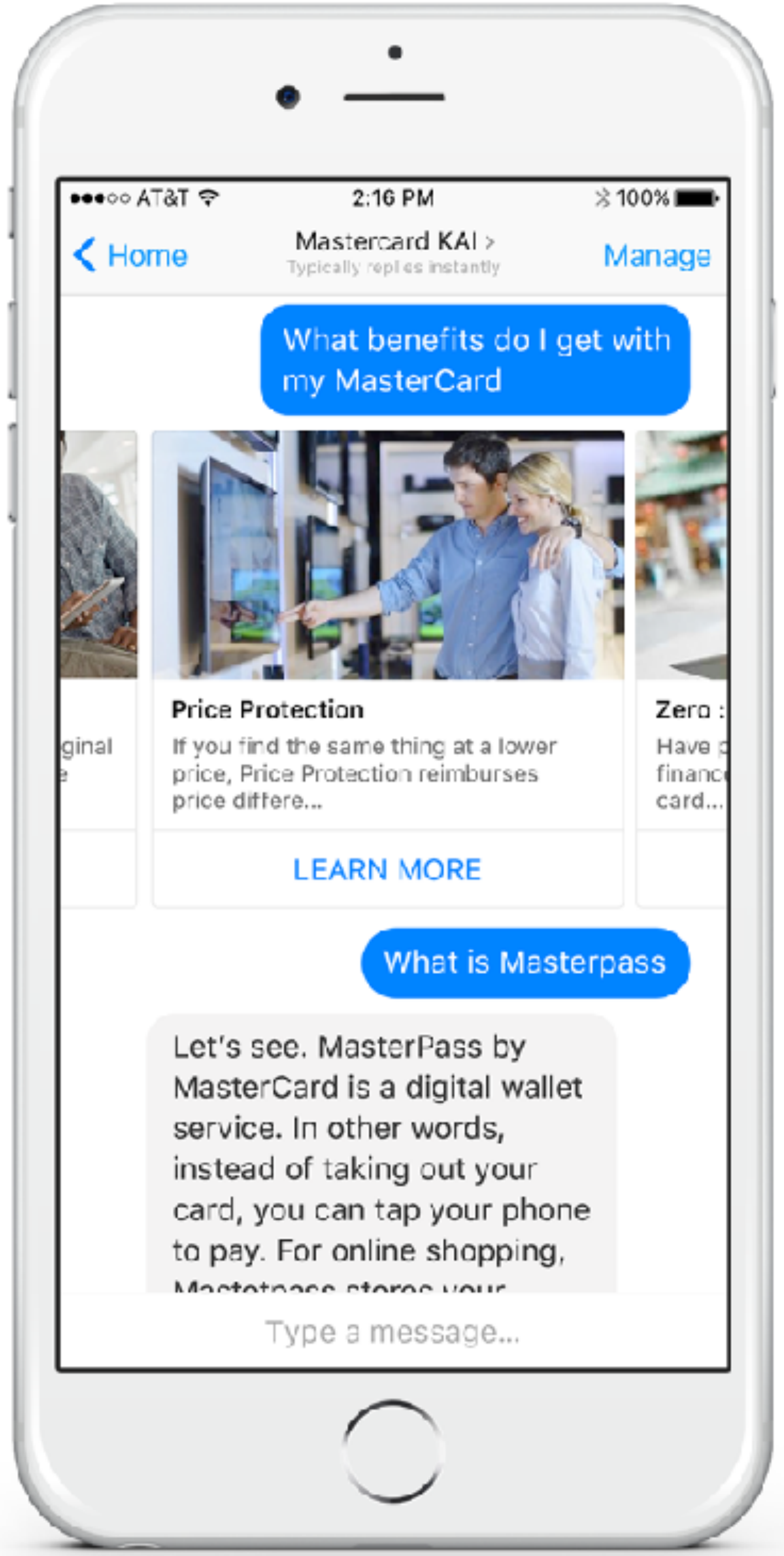
Let's see. You spent $649.32. I searched for:
Category: restaurants
Date: between 09/01/2016 and 09/30/2016

Want to see the transactions?

Show my offers

Go Greek with a Three-Course Lunch for Two at Kefi

Savor flavors of the Mediterranean with a 3 course lunch at Manhattan's Kefi



What benefits do I get with my MasterCard

Price Protection

If you find the same thing at a lower price, Price Protection reimburses price differe...

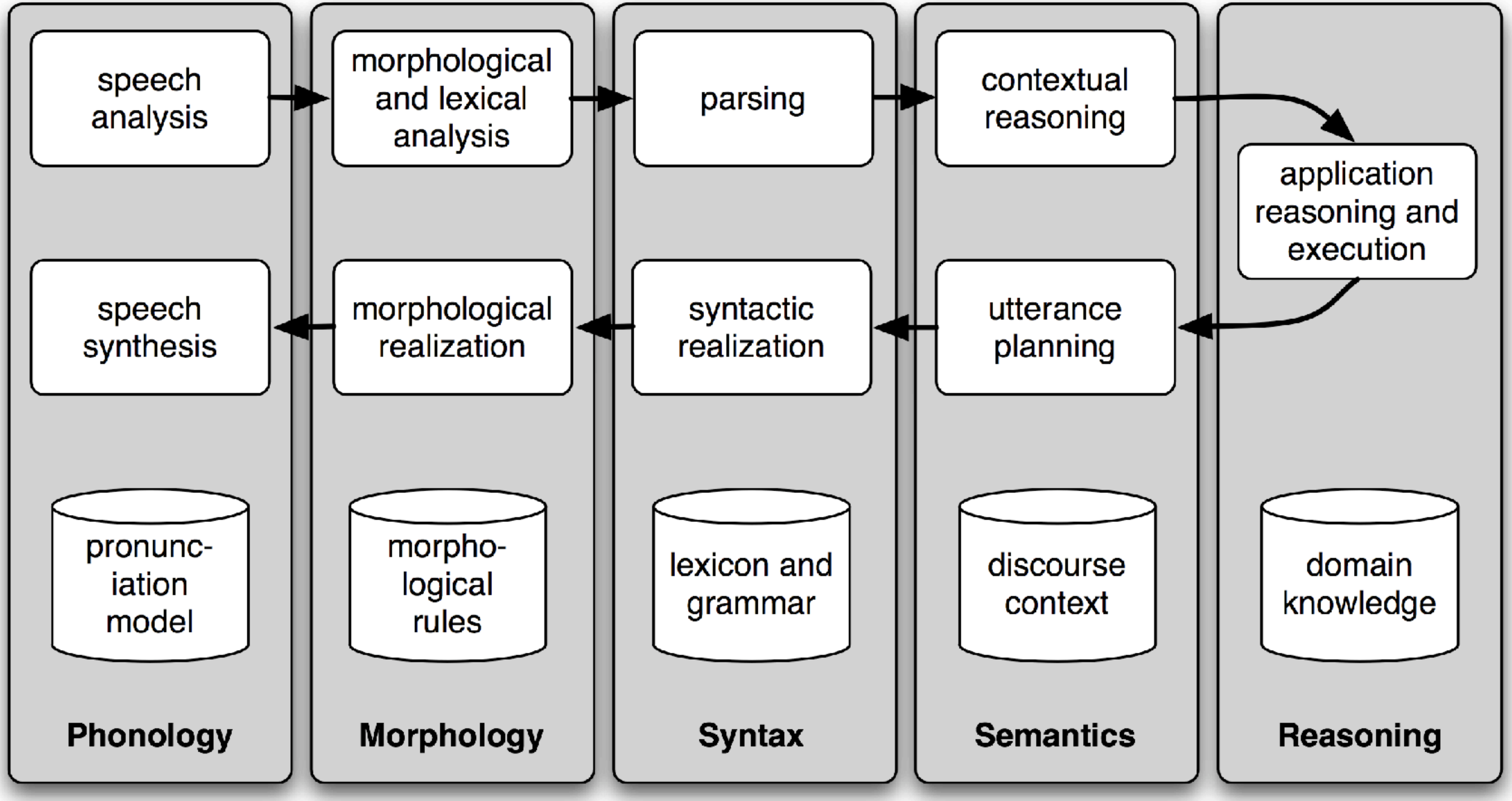
What is Masterpass

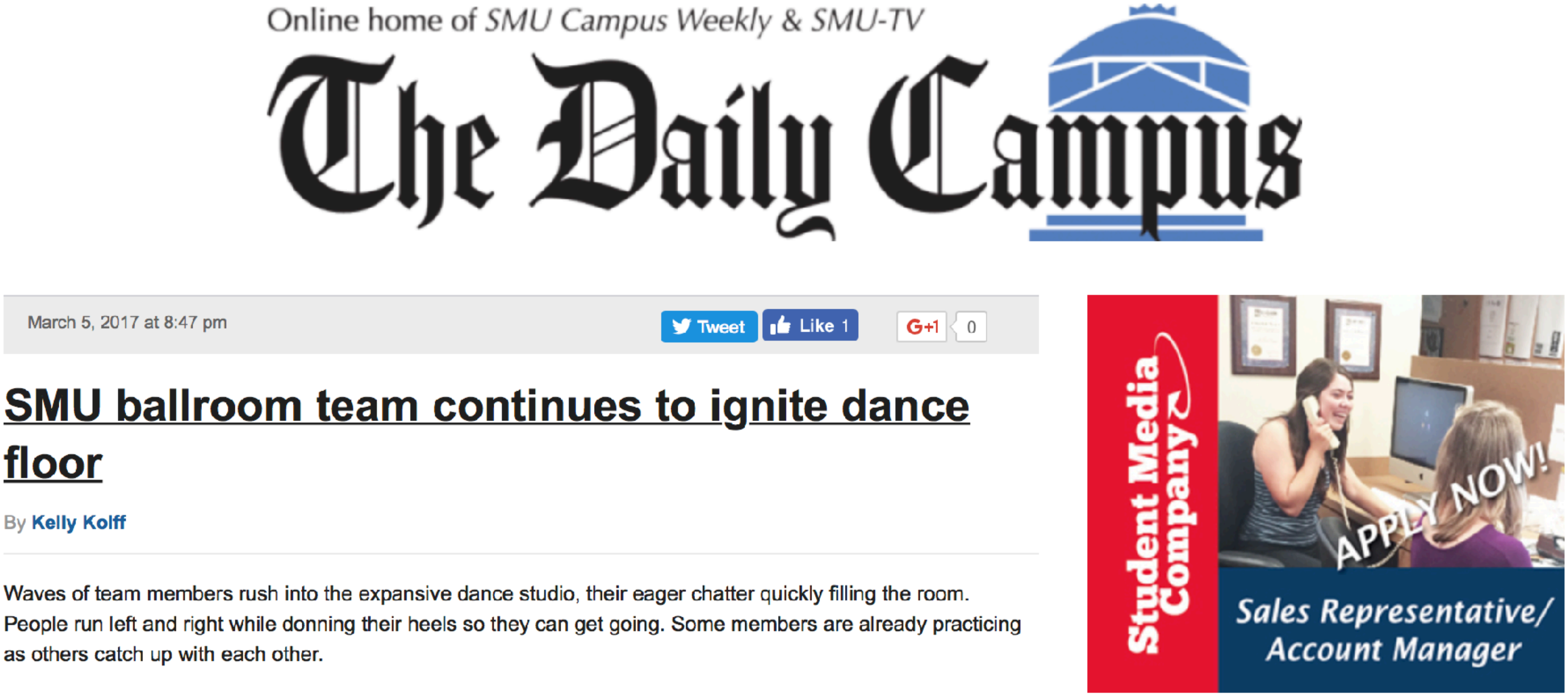
Let's see. MasterPass by MasterCard is a digital wallet service. In other words, instead of taking out your card, you can tap your phone to pay. For online shopping, Masterpass stores your

POWERED BY

kasisto

- About 75% of the world’s data is unstructured: text, video, audio
- Our reliance on NLP is growing
    - Digital assistants
    - Social media
    - Machine translation
    - Predictive typing


speech analysis
morphological and lexical analysis
parsing
contextual reasoning
application reasoning and execution
utterance planning
syntactic realization
morphological realization
speech synthesis
pronunc- iation model
morpho- logical rules
lexicon and grammar
discourse context
domain knowledge
Phonology
Morphology
Syntax
Semantics
Reasoning



Online home of *SMU Campus Weekly & SMU-TV*

The Daily Campus

March 5, 2017 at 8:47 pm

SMU ballroom team continues to ignite dance floor

By **Kelly Kolff**

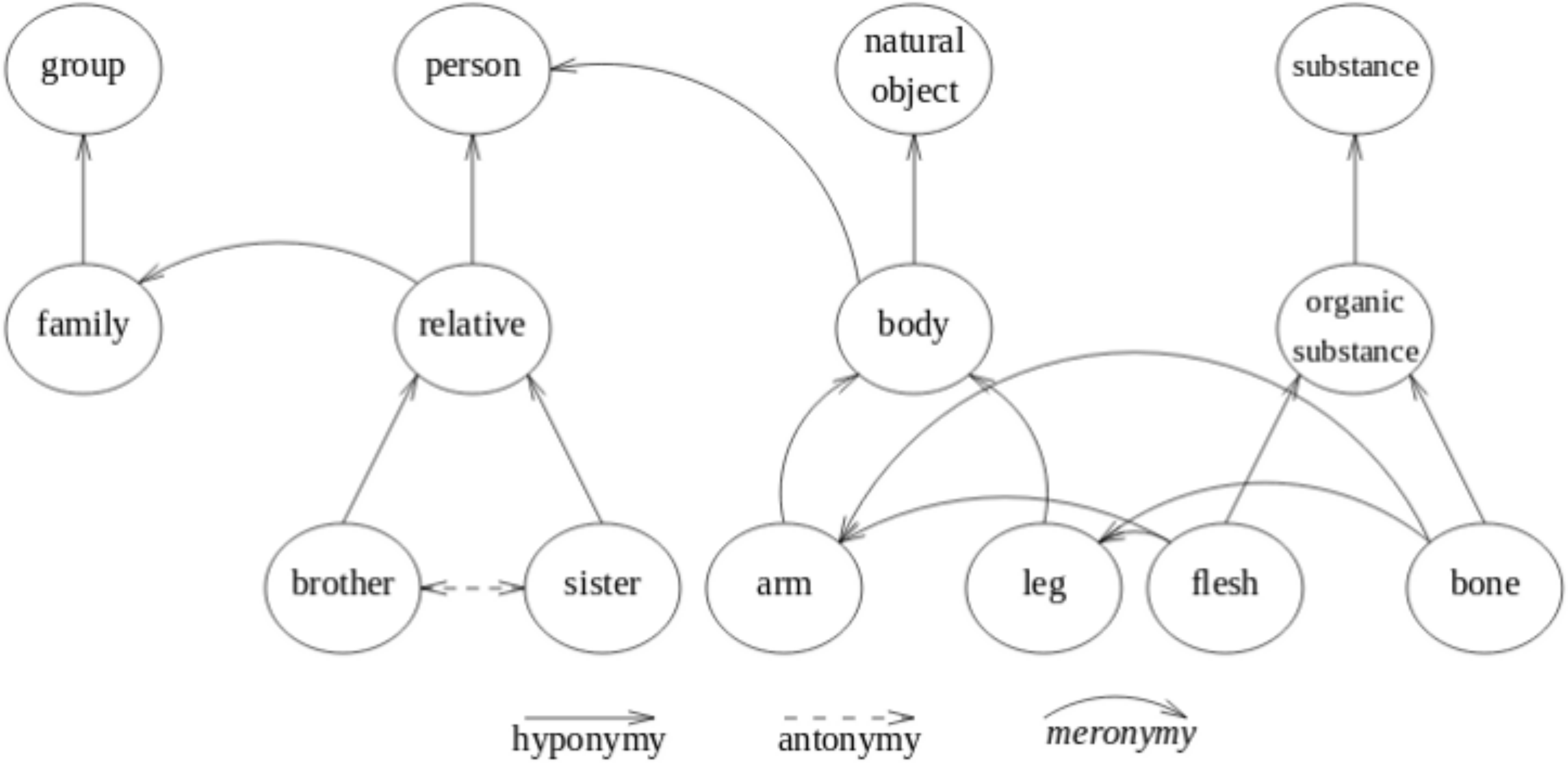
Waves of team members rush into the expansive dance studio, their eager chatter quickly filling the room. People run left and right while donning their heels so they can get going. Some members are already practicing as others catch up with each other.

Example Problems in Natural Language Understanding

- Word sense disambiguation: “He served the dish.”
- Pronoun resolution
    - The thieves stole the paintings.
    - They were subsequently sold.
    - They were subsequently found.
- RTE
    - The new order also will ban travelers from six countries who did not obtain a visa before Jan. 27 from entering the United States for 90 days. The directive no longer includes Iraq, as the original order did, but covers immigration and travel from Iran, Libya, Somalia, Sudan, Syria and Yemen. (Fox News, March 6, 2017)
    - According to the order, a traveler from Syria who does not currently have a visa could enter the United States.

Natural Language Toolkit (NLTK)

- An open-source collection of linguistic corpora and Python packages for common language processing tasks
- Simple, consistent, and extensible
- Best option for pick-up-and-play NLP
- Resources available
    - Large bodies of text of all different kinds and languages
    - Pronouncing dictionary
    - WordNet


group
person
natural object
substance
family
relative
body
organic substance
brother
sister
arm
leg
flesh
bone
hyponymy
antonymy
meronymy

Simple Operations on Text

- Frequency distributions
- Bigrams & collocations

Categorizing Words

- Parts of speech review
- Tokenization
- Working with dictionaries

Automatic Tagging

- Working up in complexity
  - Default
  - Regular expressions
  - Lookup
  - Unigram
  - N-gram
- Where to go for more information?
  - Classification

Language: Ambiguous, Unlimited

- “I shot an elephant in my pajamas.”
- “Fish fish fish.”
    - Fish go fishing for other fish.
- “Fish fish fish fish fish.”
    - The fish that other fish go fishing for also go fishing for fish themselves.

“In the loveliest town of all, where the houses were white and high and the elms trees were green and higher than the houses, where the front yards were wide and pleasant and the back yards were bushy and worth finding out about, where the streets sloped down to the stream and the stream flowed quietly under the bridge, where the lawns ended in orchards and the orchards ended in fields and the fields ended in pastures and the pastures climbed the hill and disappeared over the top toward the wonderful wide sky, in this loveliest of all towns Stuart stopped to get a drink of sarsaparilla.” (*Stuart Little*, E.B. White)

Grammar

- Diagramming a sentence
- Chunking
- Parsing a sentence
- Recursive descent parsing
- Shift-reduce parsing
- Left-corner parsing
- Chart parsing
- Feature-based grammars

References

- *Natural Language Processing with Python*, Bird, Klein, and Loper (2009)
- https://www.coursera.org/learn/natural-language-processing#
- “Jumping NLP Curves: A Review of Natural Language Processing Research”, Cambria and White (2014)
- “NLTK: The Natural Language Toolkit”, Loper and Bird (2002)
- “Parsing as Language Modeling”, Choe and Charniak (2016)
- “Natural Language Processing”, Joshi (1991)
- “Names in WordNet: A Lexical Inheriting System”, Miller (1993)

- Links to various part-of-speech tag keys
    - http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html
    - http://universaldependencies.org/u/pos/
    - http://www.comp.leeds.ac.uk/ccalas/tagsets/brown.html

- Other sources referenced
    - http://www.bbc.com/news/business-26383058
    - http://www.foxnews.com/politics/2017/03/06/trump-signs-new-immigration-order-narrows-scope-travel-ban.html
    - http://www.smudailycampus.com/ae/smu-ballroom-team-continues-to-ignite-dance-floor