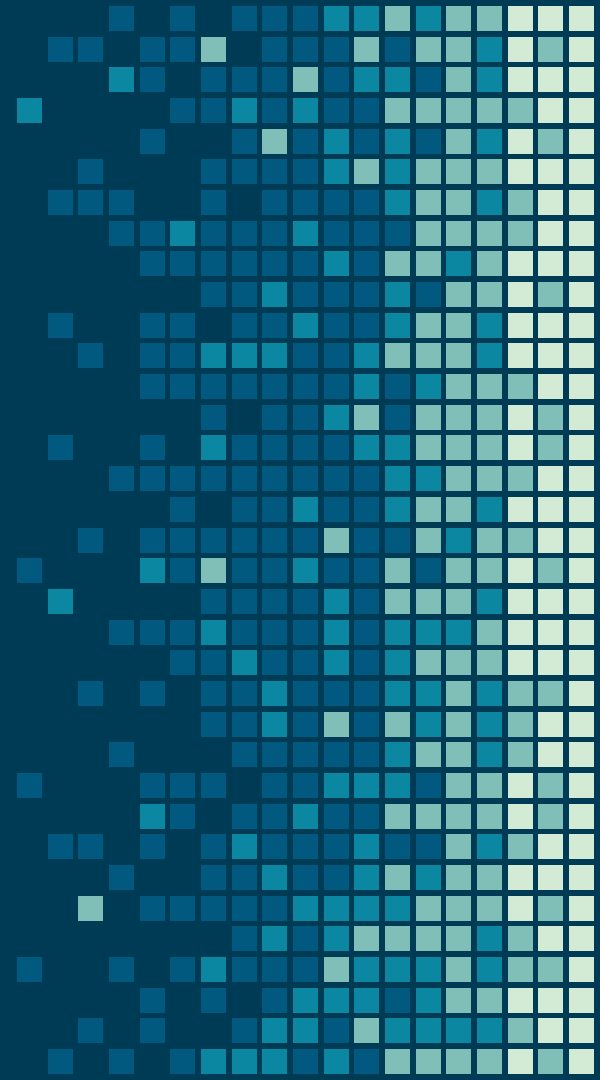


# Feature Based Customer Opinion Mining - A Modern Approach

Undergraduate: Norbert Eke

Supervisor: Dr. Jeffrey Andrews

Collaborator: Dr. Abdallah Mohamed



# 1. Background

Intro to Feature Based Opinion Mining



# Feature based Opinion Mining

- Feature: characteristic or aspect of product
- Opinion: subjective statement, belief
- Opinion mining: detection of patterns among opinions



# Feature based Opinion Mining

Digital\_camera\_1:

Picture quality:

Positive: 253      <individual opinion phrases>

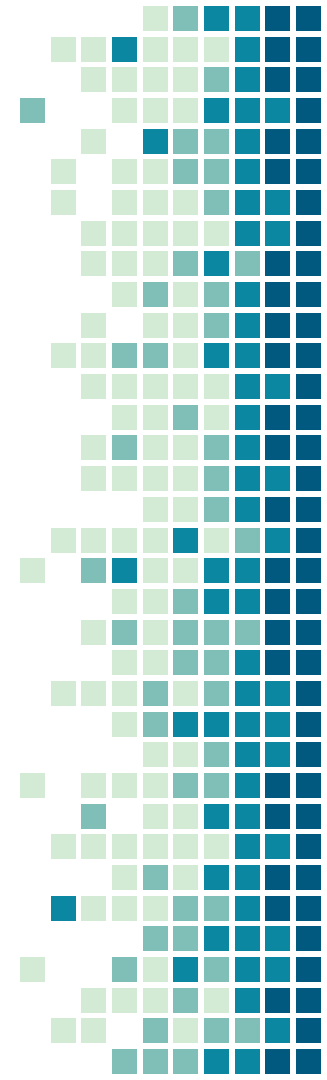
Negative: 6      <individual opinion phrases>

Size:

Positive: 134      <individual opinion phrases>

Negative: 10      <individual opinion phrases>

...



# Why should you care about customer reviews ?

- 91% of consumers regularly or occasionally read online reviews (Bonelli, 2016)
- 84% of people trust online reviews as much as a personal recommendation (Bonelli, 2016)

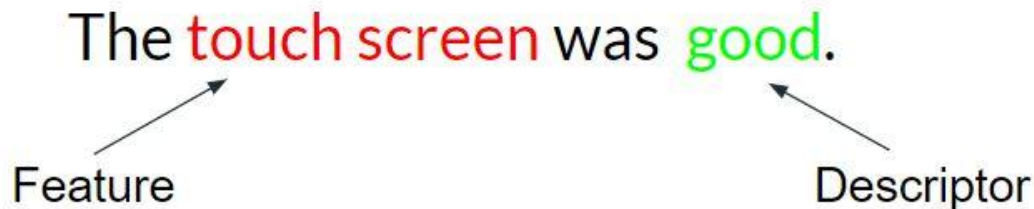


# Research Problem

- Large number of reviews
  - 26,380 reviews posted every minute by Yelp users (Shrestha, 2016)
- No one reads all the reviews
  - 88% of consumers form an opinion by reading up to 10 reviews (Shrestha, 2016)
- Need a way to interpret the content of the reviews, without reading them all

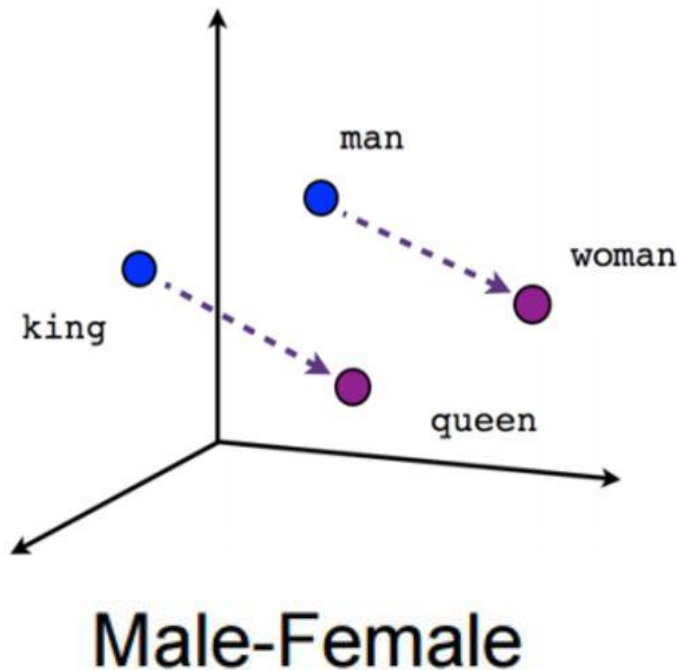


# Opinion Phrases



- Opinion Phrases: <feature, descriptor> word pairs
  - E.g. <screen, blurry> , <display, poor>
- Main research goal: Design a technique to identify opinion phrases

# Word Embeddings



- High dimensional vector representation of each word
- Used to reconstruct linguistic context of words
- Capture semantic similarity between words

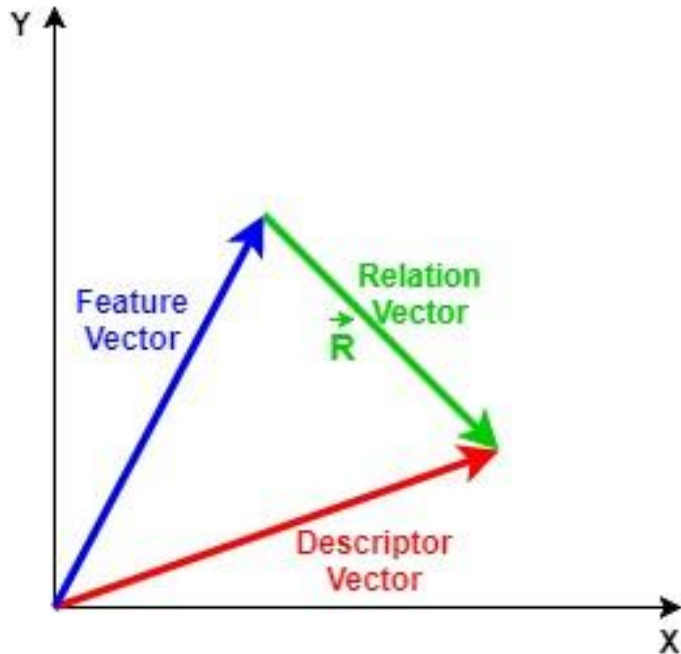


## 2. Techniques Designed

Relation Vectors and Machine Learning

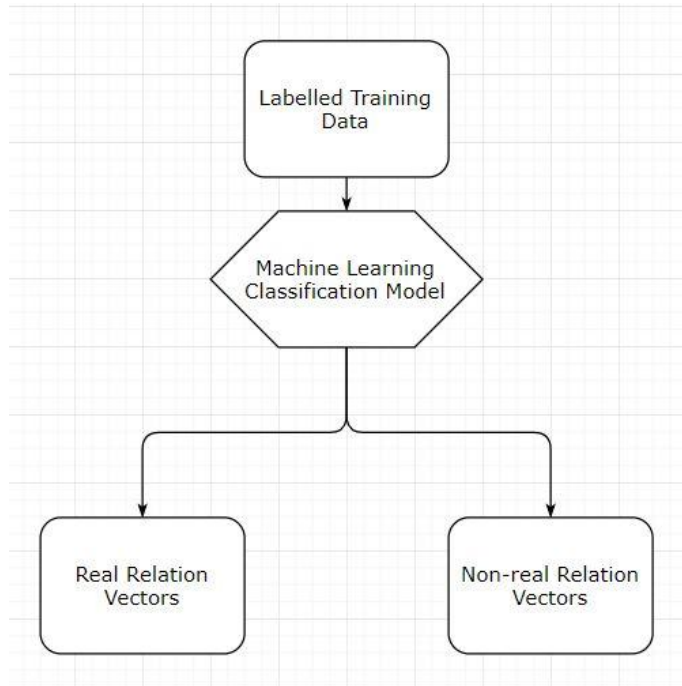


# Feature - Descriptor Relation Vector



- Convert words into vectors using Word Embeddings
- Relation Vector: Relationship between feature and descriptor

# Feature-Descriptor Relation Vector Classification Model



- Training Data: Labelled real and non-real Relation Vectors
  - Real: delicious meal
  - Non-real: table meal
- Model finds real Relation Vectors at around 80 % accuracy

# Natural Language Processing

## Techniques Used

### Feature Extraction

KEEP UP ON YOUR READING WITH AUDIO BOOKS

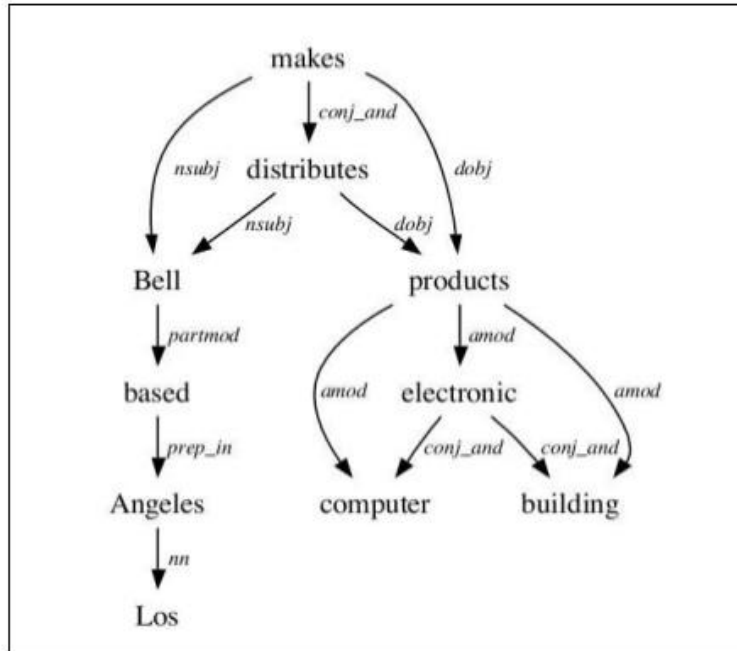
Audio books are highly popular with library patrons in the town

of Springfield, Greene County, MO. "People are mobile

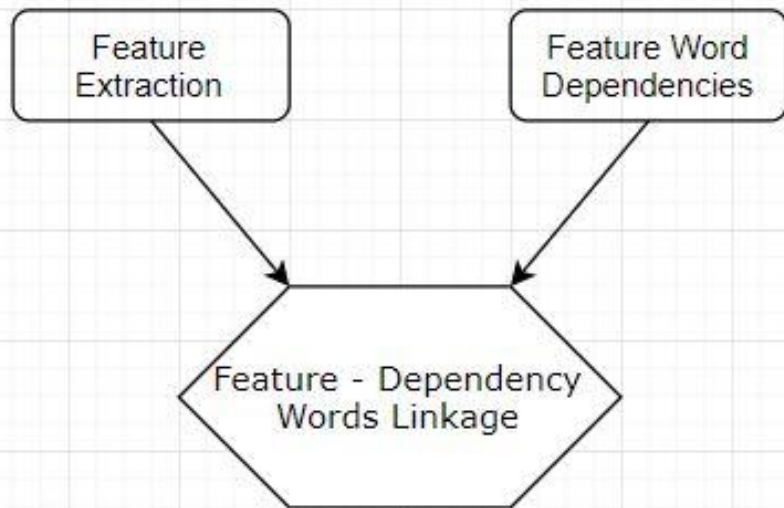
and busier, and audio books fit into that lifestyle" says Gary

Sanchez, who oversees the library's \$2 million budget...

### Dependency Parsing

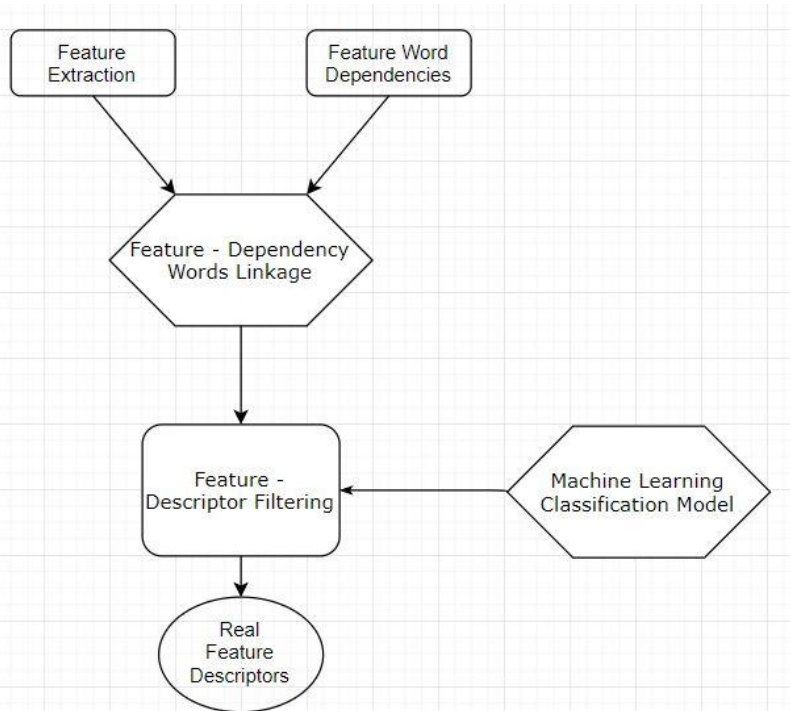


# Feature - Feature Dependency Word Linkage



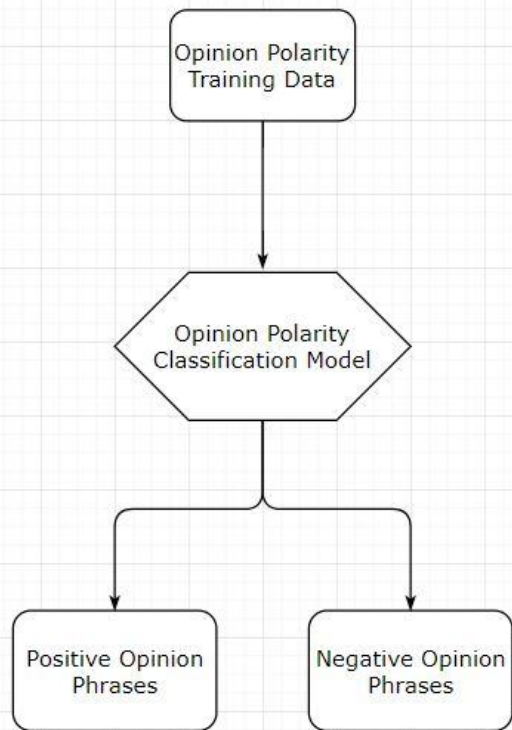
- Linkage: Create a relation vector between each feature and all of its word dependencies
- Linkage finds 75-85 % of the feature descriptor word pairs

# Feature - Descriptor Relation Vector Filtering



- Relation Vector filtering:  
Use the classification model to filter out all the non-real relation vectors produced by the linkage

# Opinion Polarity & Opinion Feature Breakdown



Digital\_camera\_1:

Picture quality:

Positive: 253

<individual opinion phrases>

Negative: 6

<individual opinion phrases>

Size:

Positive: 134

<individual opinion phrases>

Negative: 10

<individual opinion phrases>

...

# Results

- Technique has been tested on labelled benchmark data sets
- Performance measured in percentage accuracy

Techniques	Data set Computer dataset (331 opinions)	Hotel dataset (131 opinions)
Opinion Linkage Version 1	<b>36%</b>	<b>40%</b>
Opinion Filtering Version 1	<b>42%</b>	<b>45.3%</b>
Opinion Linkage Version 4	<b>72.8%</b>	<b>73.8%</b>
Opinion Filtering Version 4	<b>78.1%</b>	<b>76.1%</b>
Opinion Linkage Version 7	<b>85%</b>	<b>93%</b>
Opinion Filtering Version 7	<b>80%</b>	<b>78.4%</b>



# Demo of the technique:

## Features Extracted

review ID	Review	feature extracted
1	Bright , vivid , clean lines , wide viewing_angle , and best of all great price !	lines
1	Bright , vivid , clean lines , wide viewing_angle , and best of all great price !	viewing_angle
1	Bright , vivid , clean lines , wide viewing_angle , and best of all great price !	price
2	The monitor works very well with a sharp and bright display.	monitor
2	The monitor works very well with a sharp and bright display.	display
3	The images are vivid and crisp , the text and fonts are very clear.	fonts
3	The images are vivid and crisp , the text and fonts are very clear .	text
3	The images are vivid and crisp , the text and fonts are very clear .	images

# Demo of the technique: Feature Dependency Linkages

Bright, vivid, clean **lines**, **wide viewing\_angle**, and best of all great price!

The **monitor** works very well with a sharp and bright **display**.

The **images** are vivid and crisp, the **text** and **fonts** are very clear.

## Linkages for viewing\_angle:

viewing\_angle - lines

viewing\_angle - Bright

viewing\_angle - vivid

viewing\_angle - clean

viewing\_angle - wide

viewing\_angle - best

## Linkages for price:

price - best

price - great

## Linkages for monitor:

monitor - works

## Linkages for fonts:

fonts - text

## Linkages for lines:

lines - Bright

lines - vivid

lines - clean

lines - viewing\_angle

lines - wide

lines - best

## Linkages for display:

display - works

display - sharp

display - bright

## Linkages for text:

text - clear

text - fonts

# Demo of the technique: Feature Descriptor Filtering

Bright, vivid, clean **lines**, **wide viewing\_angle**, and best of all great price!

The **monitor** works very well with a sharp and bright **display**.

The **images** are vivid and crisp, the **text** and **fonts** are very clear.

## Linkages for lines:

lines - vivid

lines - clean

lines - wide

lines - best

## Linkages for display:

display - sharp

display - bright

## Linkages for text:

text - clear

## Linkages for viewing\_angle:

viewing\_angle - vivid

viewing\_angle - clean

viewing\_angle - wide

viewing\_angle - best

## Linkages for price:

price - best

price - great

## Linkages for monitor:

monitor - works

# Demo of the technique:

## Feature based Opinion Mining System

Product Name:Computer

**Feature: lines**

Positive opinions:

- clean lines
- best lines
- vivid lines
- wide lines

**Feature: viewing\_angle**

Positive opinions:

- vivid viewing\_angle
- clean viewing\_angle
- wide viewing\_angle
- best viewing\_angle

**Feature: text**

Neutral opinions:

- clear text

Product Name:Computer

**Feature: display**

Positive opinions:

- bright display

Neutral opinions:

- sharp display

**Feature: images**

Neutral opinions:

- vivid images

**Feature: price**

Positive opinions:

- best price
- great price

# References

- Slide 4, 15 - Feature Based Opinion Mining image:  
Hu, Minqing, and Bing Liu. "Mining opinion features in customer reviews." *AAAI*. Vol. 4. No. 4. 2004.
- Statistics from slide 5-6:  
Shrestha, Khusbu . "50 Stats You Need to Know About Online Reviews." Vendasta. N.p., 29 August 2016. Web. 11 June 2017.  
Bonelli, Sherry. "Local Consumer Review Survey 2016 | The Impact Of Online Reviews." BrightLocal, Bright Little Light Ltd, 2016. Accessed 28 Aug. 2017.
- Slide 8 - Word Embeddings image:  
NSS. "Word Vectors." An Intuitive Understanding of Word Embeddings: From Count Vectors to Word2Vec, ANALYTICS VIDHYA , 4 June 2017, Accessed 21 Aug. 2017.
- Slide 12 - Feature Extraction image:  
Bird, StevenJul, et al. "Figure 5.1: Location Detection by Simple Lookup for a News Story." *Extracting Information from Text*, Natural Language Toolkit, 1 July 2015, [www.nltk.org/book/ch07.html#fig-locations](http://www.nltk.org/book/ch07.html#fig-locations). Accessed 21 Aug. 2017.
- Slide 12 - Dependency Parsing image:  
Darari, Fariz. "Dependency Parsing." Slideshare, 30 Jan. 2013, [www.slideshare.net/fadirra/architecture-16258993](http://www.slideshare.net/fadirra/architecture-16258993). Accessed 21 Aug. 2017.

# Acknowledgements

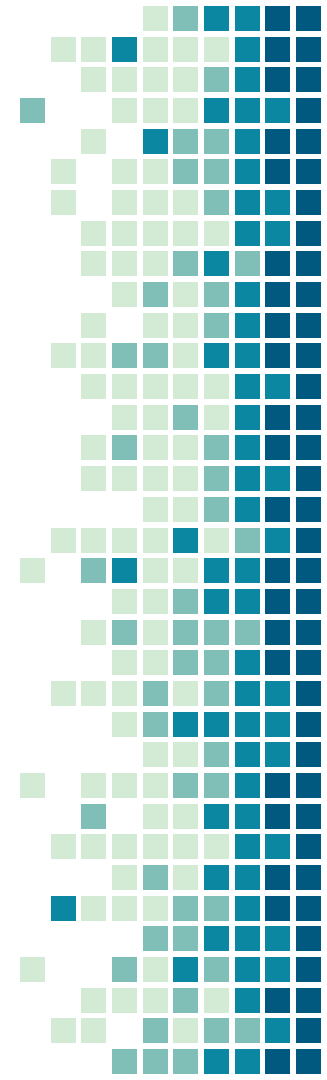
Dr. Jeffrey Andrews, Statistics

Dr. Abdallah Mohamed, Computer Science

Irving K. Barber School of Arts and Sciences

Undergraduate Research Program

Fellow Students and Researchers



# THANK YOU!

## Any questions?

