

## Toward Semantic-based Identification of Fake Online Reviews

Student: Louis Hickman. Advisor: Julia Taylor

### Existing Strategies

Most existing work on fake review detection uses the statistical machine learning methods are so prevalent in contemporary Natural Language Processing (NLP)

#### •Reviewer properties

- If reviewer has only written one review [3], if reviewer has only visited the site once and never visits again [4], or a combination of behavioral and statistical features [5]
- Identifying fake reviewer groups based on shared behaviors [6]

#### •Statistical features

- [2] achieved nearly 90% accuracy using just  $n$ -gram features
- [7] found that deceptive reviews have greater lexical complexity, have a higher ratio of first person pronouns to other words, mention the brand their reviewing more frequently, and differences in the use of positive and negative words
- For a survey of approaches see [8]

### Semantic Possibilities

Few, if any, existing approaches utilize the meaning of words in their analysis of whether an online review is legitimate. Many possibilities exist that could reveal the legitimacy of a review.

#### •Fact-checking

- Can verify the floor they state their room is in exists, that the room type exists, if a named hotel employee exists, etc.

#### •Patterns of Meaning

- Similar to the existing strategies to find patterns of character strings, regularities in word usage
- But what feature to look at for patterns?

### Corpus

•[1] and [2] developed a corpus of known fake reviews by using Amazon Mechanical Turks to write fake reviews on Chicago area hotels. These are compared against real-life online reviews of the same hotels.

### Initial Analysis

Positive Legit Hilton Reviews:

12 correct, 3 neutral, 5 incorrect.

Negative Legit Hilton Reviews:

18 correct, 1 neutral, 1 incorrect.

Positive Deceptive Hilton Reviews:

14 correct, 2 neutral, 4 incorrect.

Negative Deceptive Hilton Reviews:

10 correct, 3 neutral, 7 incorrect.

Positive Legit James Reviews:

10 correct, 2 neutral, 8 incorrect.

Negative Legit James Reviews:

15 correct, 2 neutral, 2 incorrect.

Positive Deceptive James Reviews:

17 correct, 2 neutral, 1 incorrect.

Negative Deceptive James Reviews:

6 correct, 7 neutral, 7 incorrect.

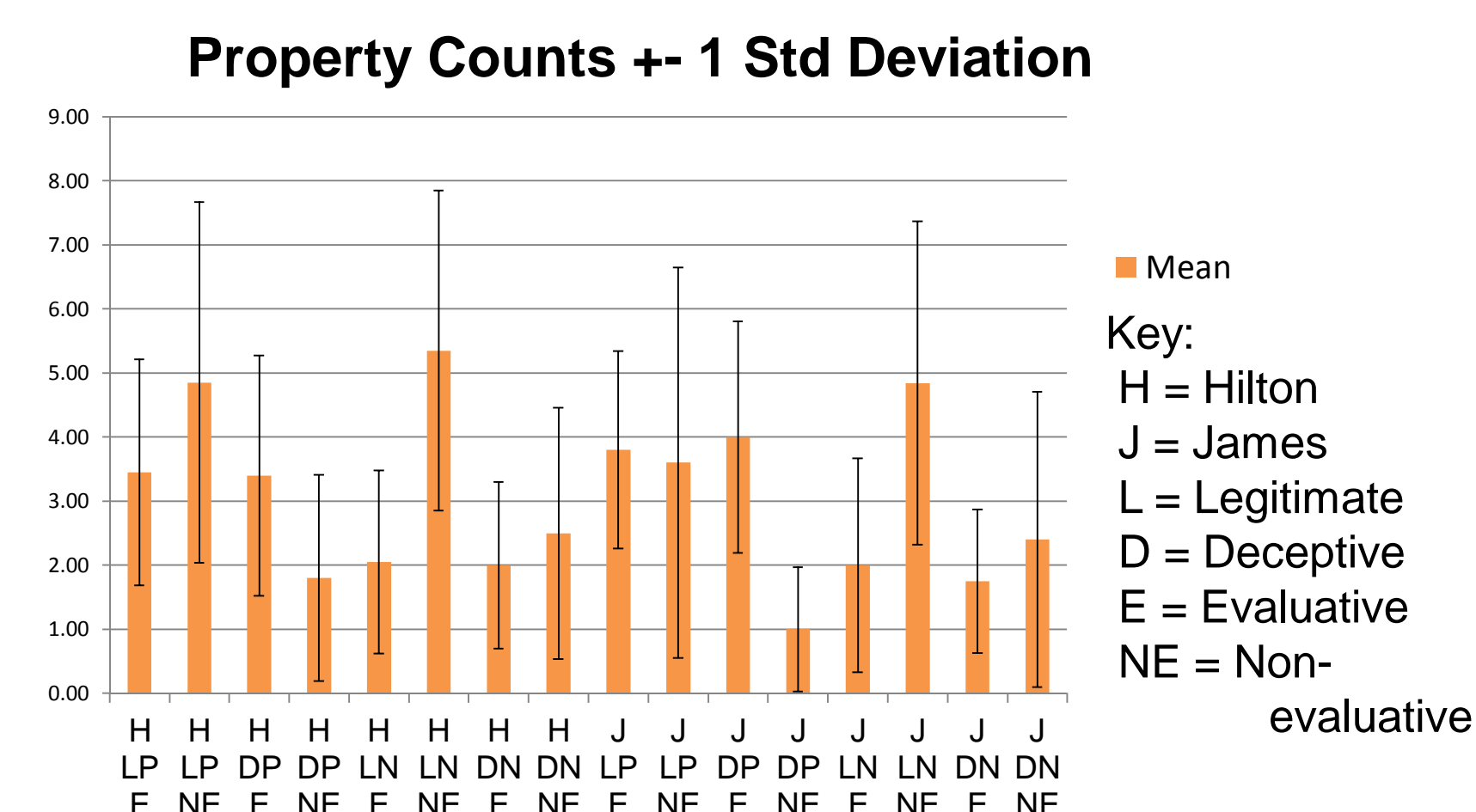
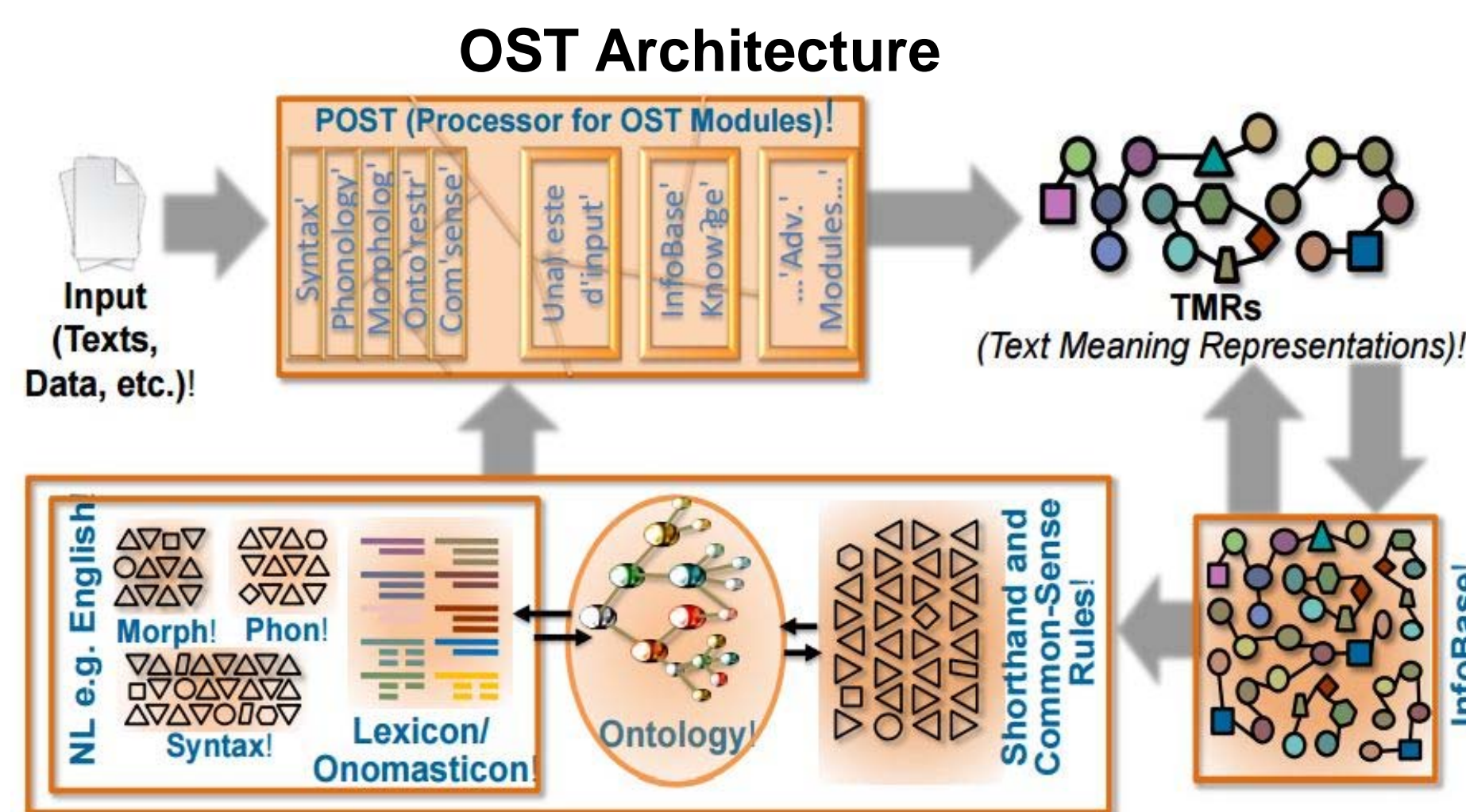
Overall: 102/159 correct (64.2%), 22/159 neutral (13.8%)  
35/159 incorrect (22.0%).

### Proposed Method

Use Ontological Semantics Technology (OST) to test whether legitimate reviews have more non-evaluative properties while deceptive reviews have more evaluative properties.

•**Evaluative properties:** Activated by statements such as, "The bed was nice," "The room was awesome."

•**Non-evaluative properties:** Activated by statements such as, "The room had two double beds," "The room cost \$249 a night."



### Hilton

### Property Counts

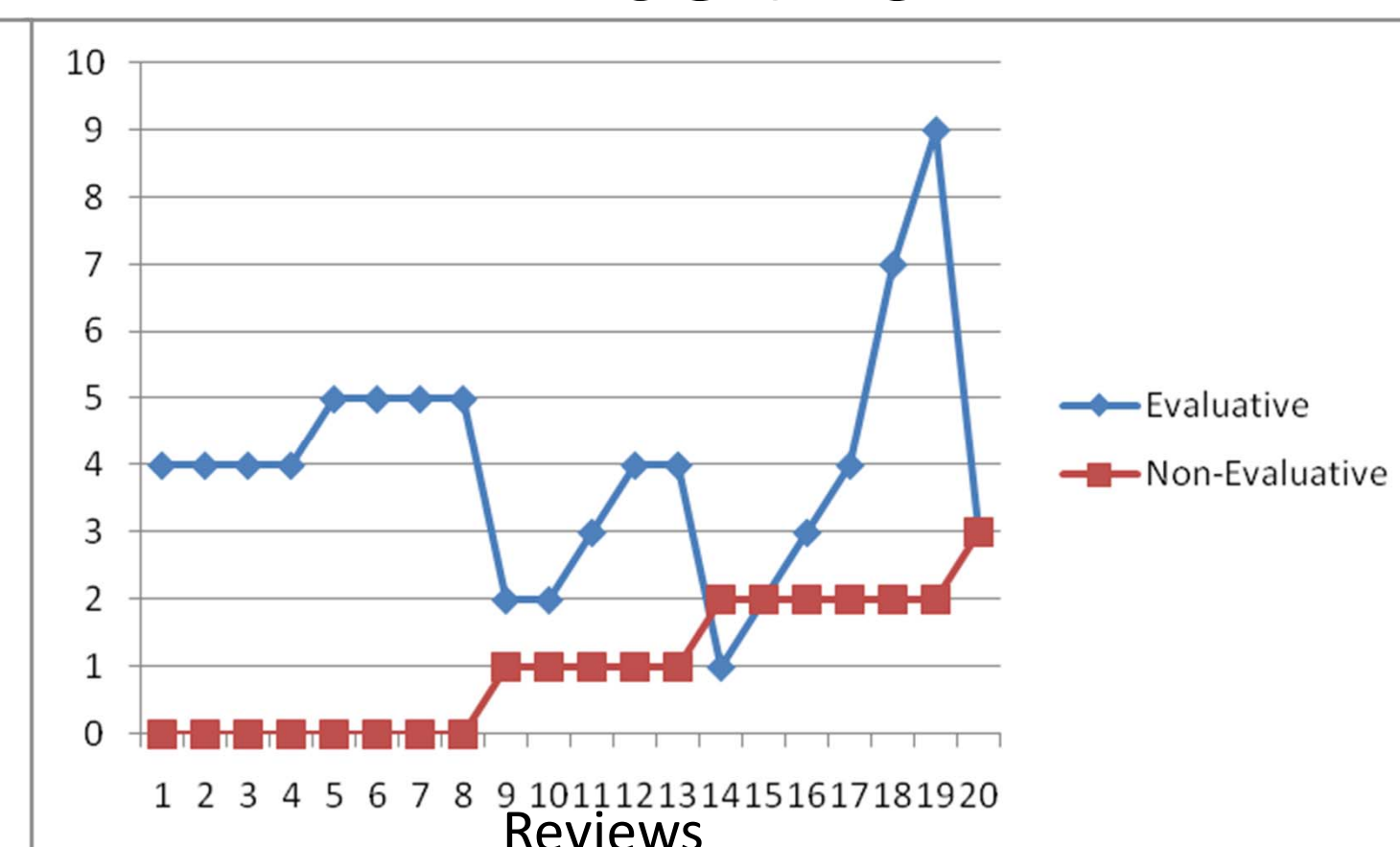
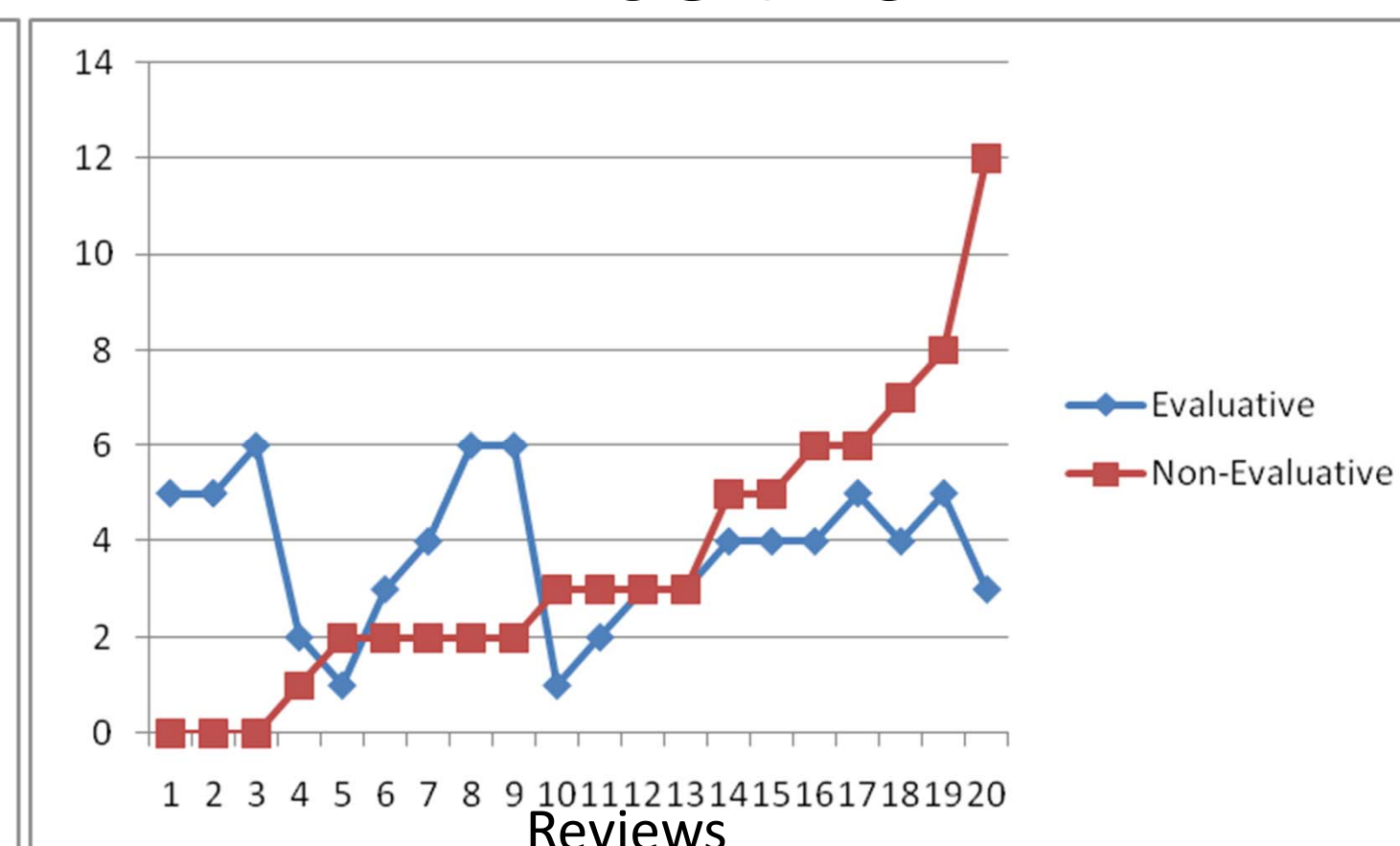
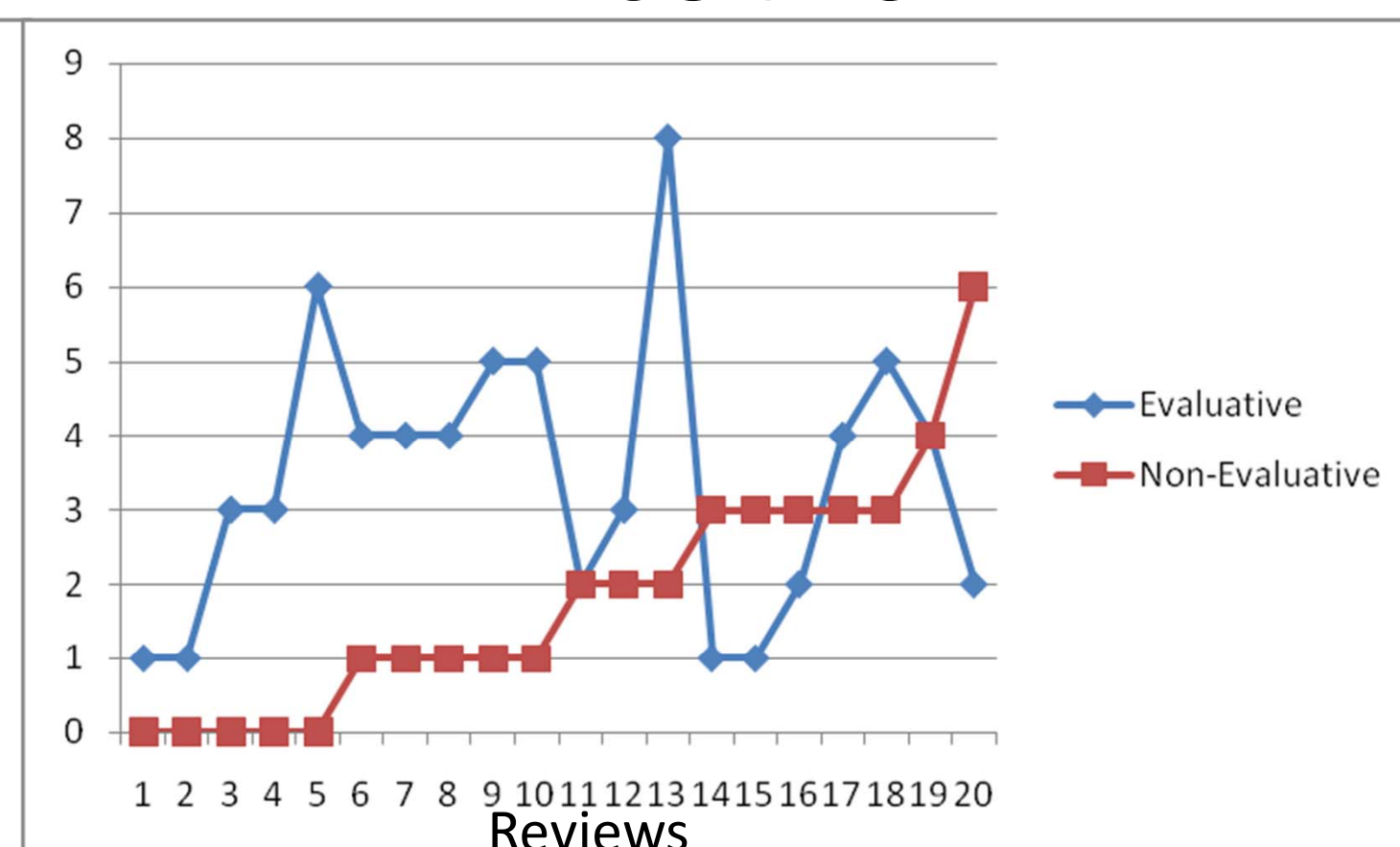
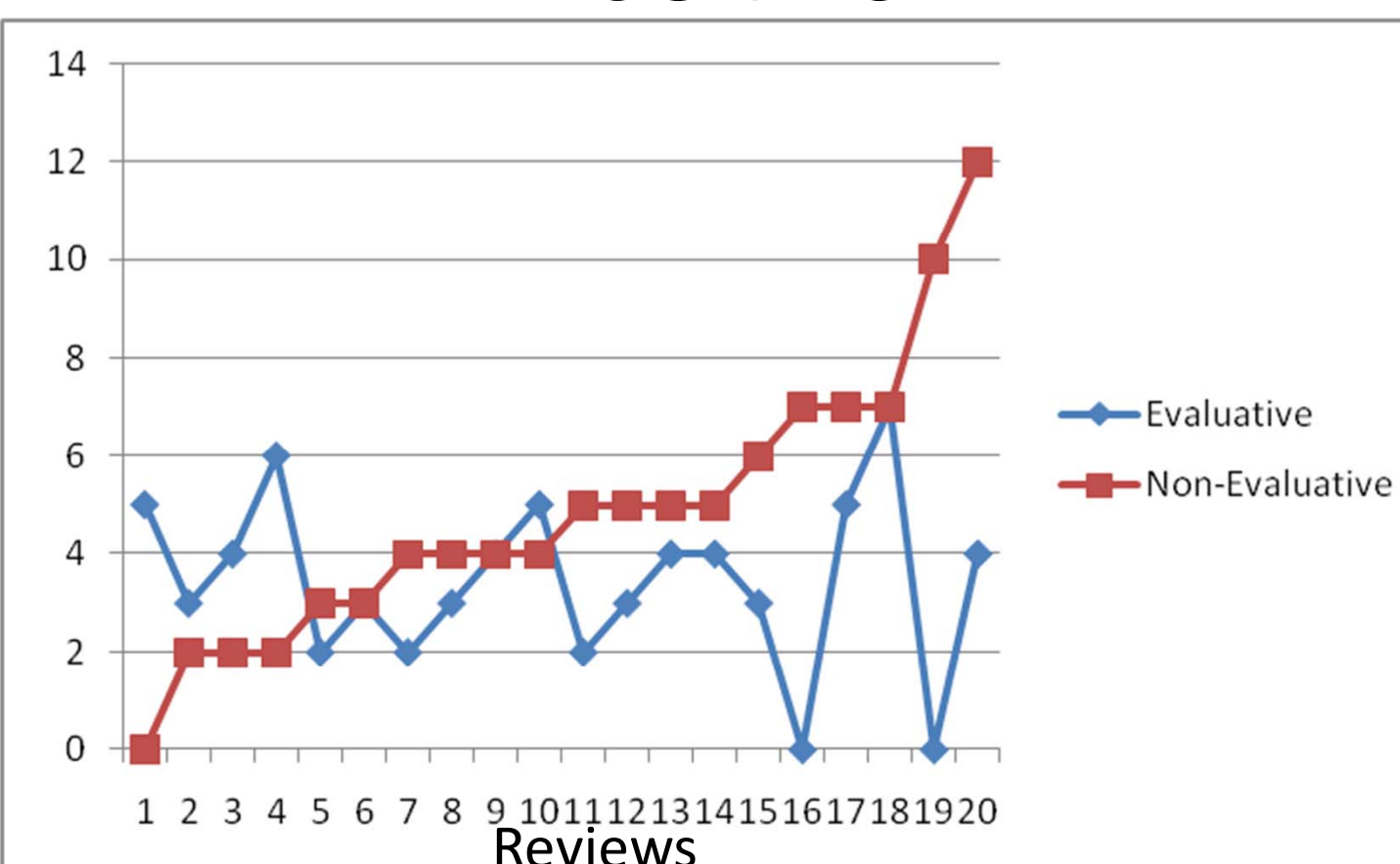
### The James

#### Positive

#### Positive

#### Positive

#### Positive

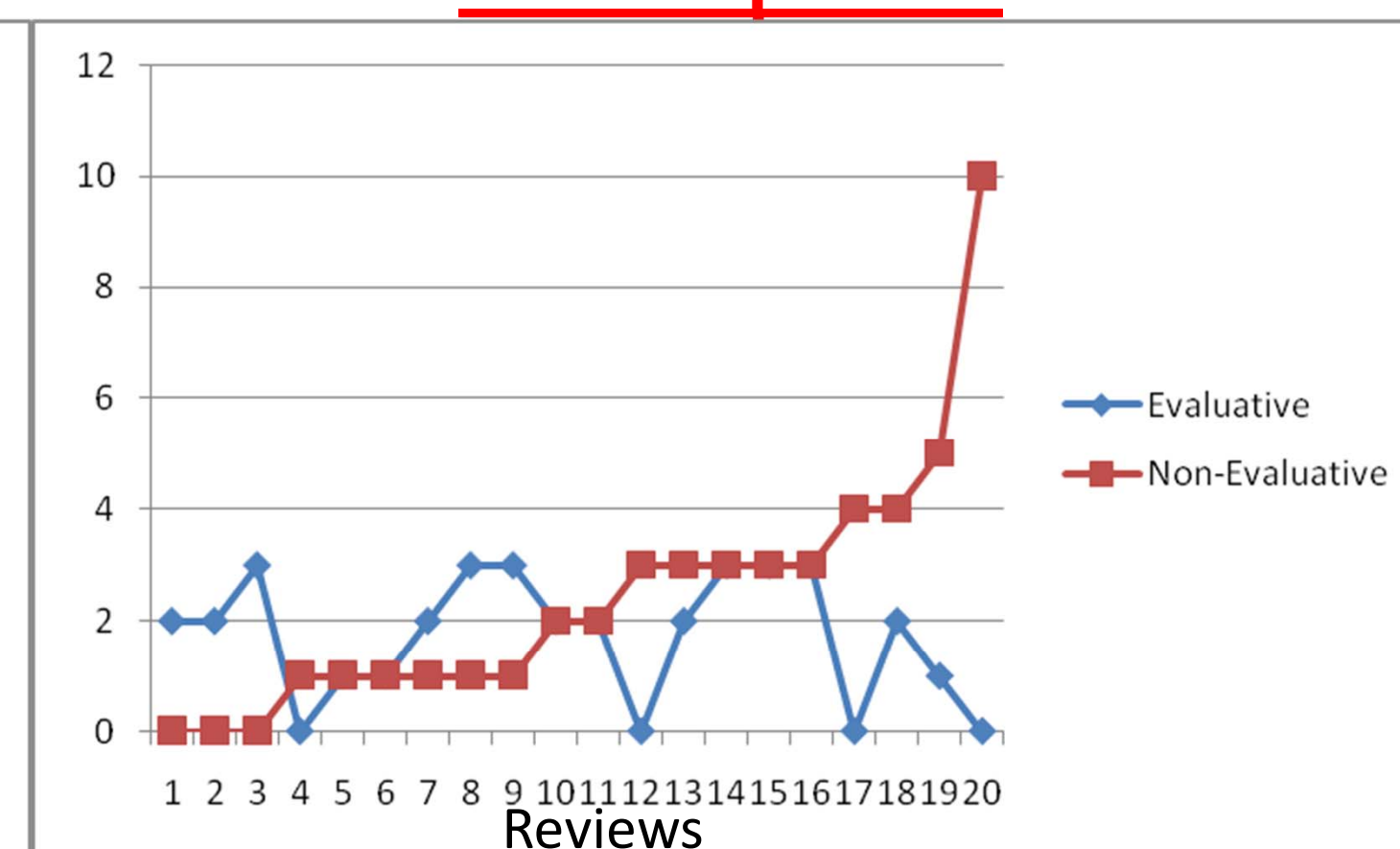
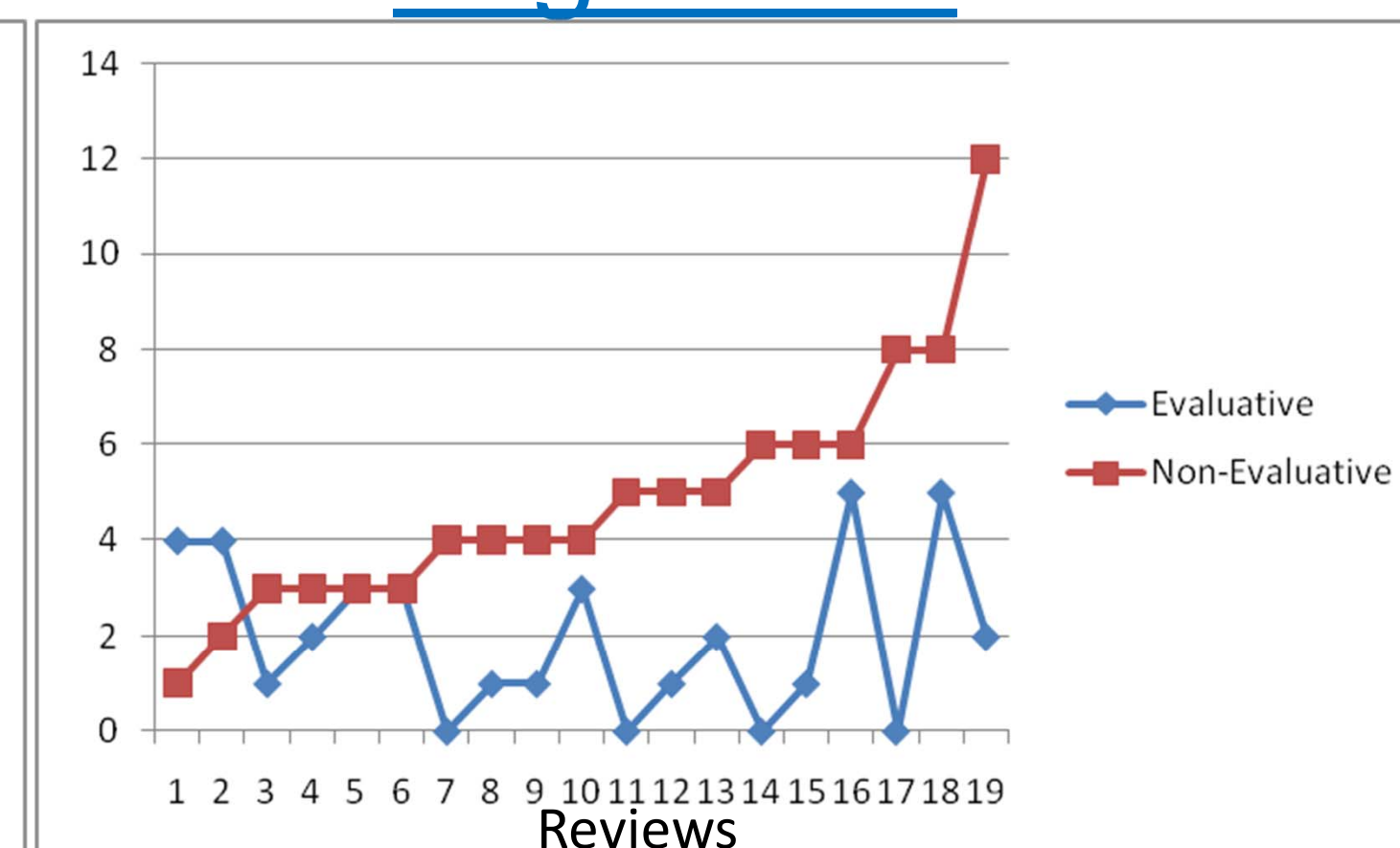
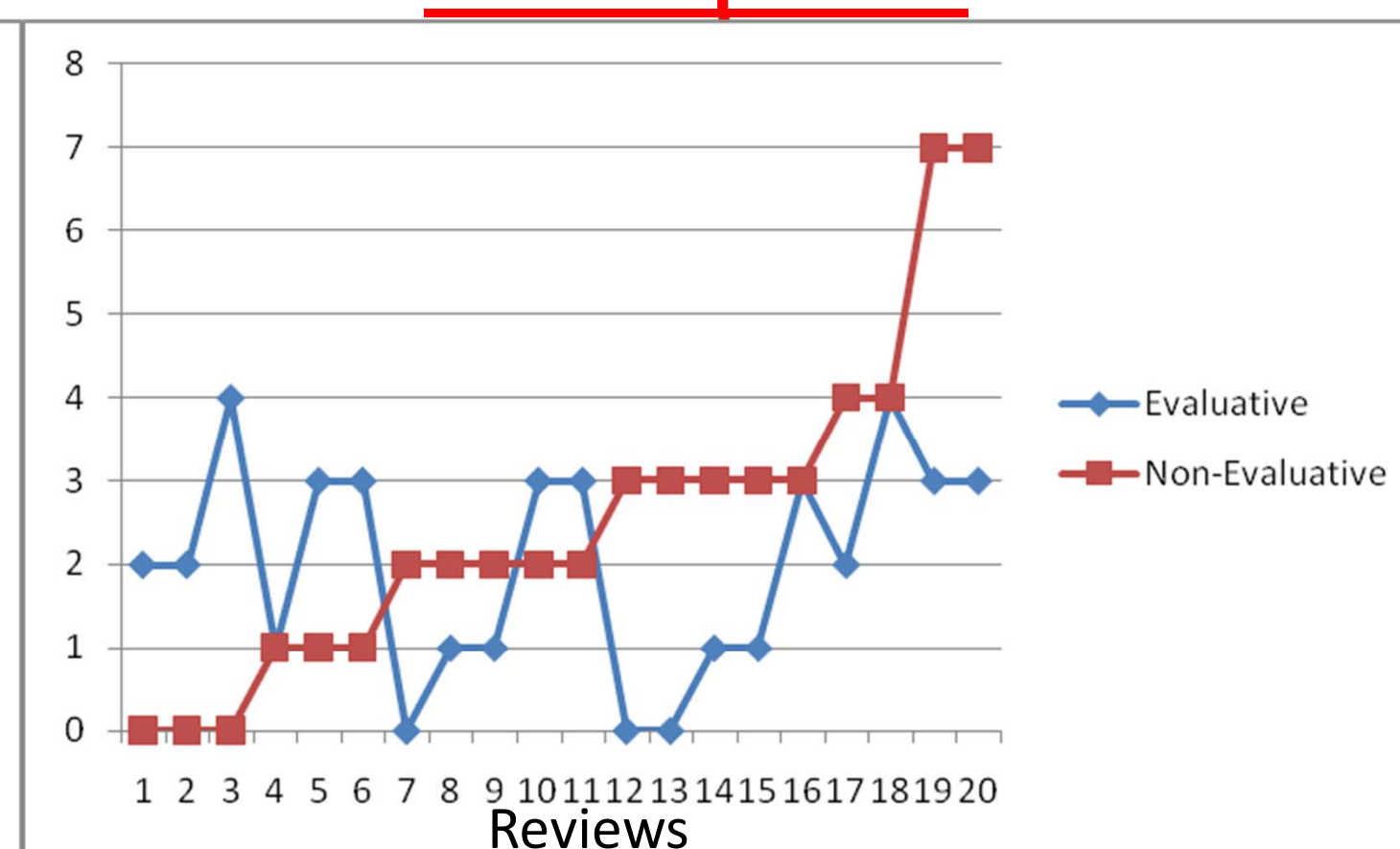
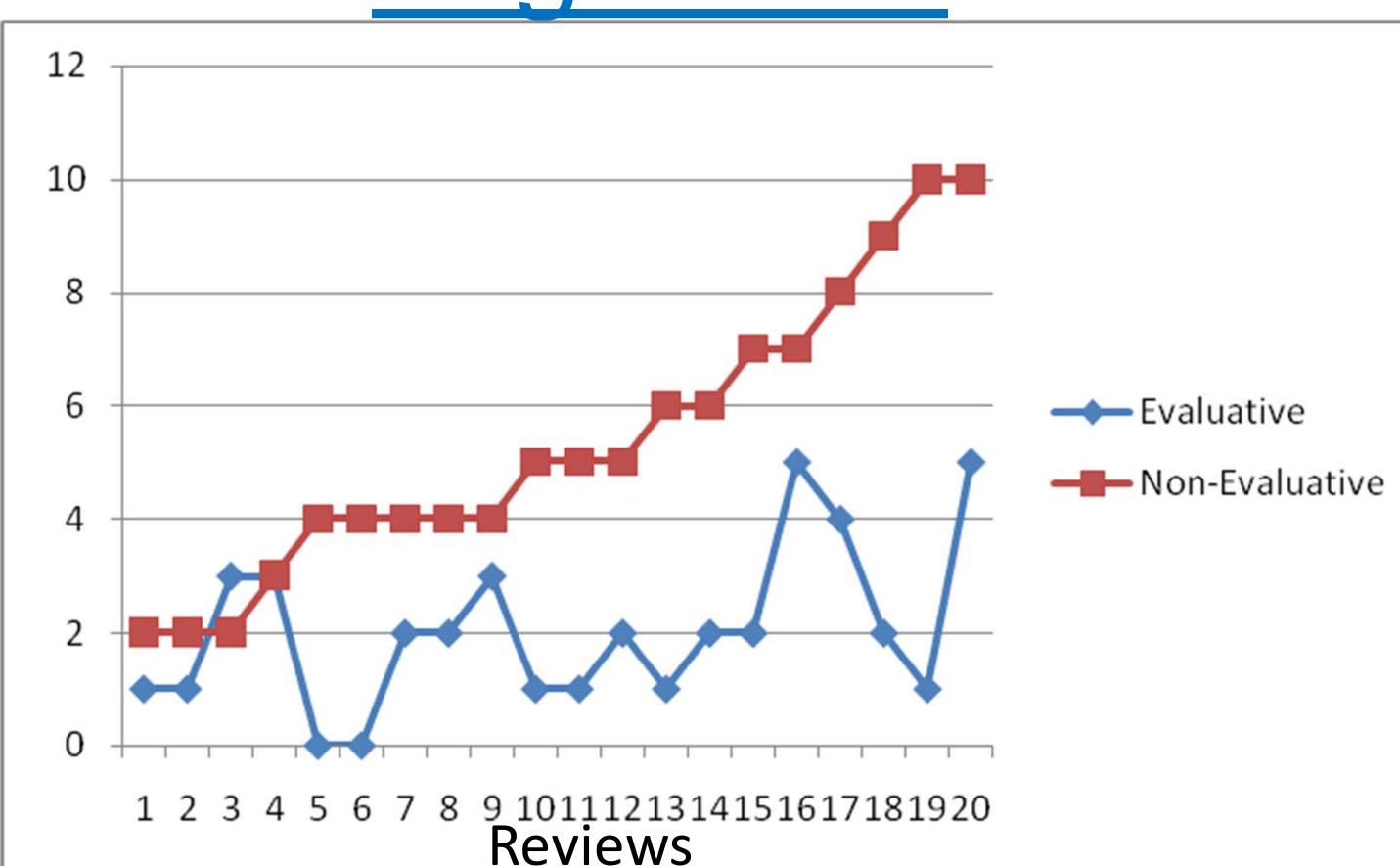


#### Legitimate

#### Deceptive

#### Legitimate

#### Deceptive



#### Negative

#### Negative

#### Negative

#### Negative

### References

- [1] Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. *Proceedings of NAACL-HLT 2013*, Atlanta, Georgia 497-501.
- [2] Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the for Computational Linguistics*, Portland, Oregon, 309-319.
- [3] O'Connor, P. (2008). User-generated content and travel: A case study on Tripadvisor.com. *Proceedings of the Information and Communication Technologies in Tourism*, Austria, 47-58.
- [4] Keates, N. (2007). Deconstructing TripAdvisor. *Wall Street Journal*, 1 June, 4.
- [5] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. University of Illinois at Chicago. (2013). Fake review detection: classification and analysis of real and pseudo reviews. (UIC-CS-2013-03)
- [6] Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. *Proceedings of the 21st International conference on World Wide Web*. New York, 191-200.
- [7] Yoo, K. & Gretzel, U. (2009). Comparison of deceptive and truthful travel reviews. *Proceedings of the Information and Communication Technologies in Tourism*, Amsterdam, 37-47.
- [8] Heydari, A., Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, 42(7), 3634-3642.