

Week 7 Coding Project

“Natural Language Processing Tasks”

2018 Amazon Review Data: Video Games Category

by Joshua Douglas Thomas

October 24, 2025

Professor Bryan MacKenzie

Wilmington University

INTRODUCTION

According to International Business Machines, Natural Language Processing is (IBM, n.d) “a subfield of computer science and artificial intelligence (AI) that uses machine learning to enable computers to understand and communicate with human language.” In this document, findings are presented on various NLP tasks, using 2018 Amazon Reviews, in the Video Games category, that have been culled from (n.d) the Web.

CONCLUSIONS

PARTS-OF-SPEECH TAGGING

According to established theory, Parts-of-Speech Tagging is useful for establishing a viable framework for assigning parts of speech to any given word. A part of speech is a critical element in the structure and meaning of a phrase, or sentence. The English language has more than several parts of speech to choose from when building sentences and conveying meaning.

SENTIMENT ANALYSIS

In this project, the author manually defined lists of words to Exclude, Positive words, and Negative words. Then, this author presented two ways one can infer from sentiment analysis. Sentiment analysis, according to established theory, is a field, or subject, in which a body of text can be inferred to be ‘positive’, ‘negative’, or ‘neutral’.

WORD DEFINITIONS, SYNONYMS, AND ANTONYMS

In this project, this author tested Positive and Negative word lists and presented word definitions, synonyms, and antonyms. Using Python 3 libraries, this information was fetched and stored to a variable specifically formatted for output. The requested information was drawn from a corpus, which, can be thought of as an accepted code base for Python programmers everywhere to use in their projects from which to achieve desired, derived results. According to GeeksForGeeks.com (n.d) “In Natural Language Processing (NLP) a corpus refers to a large and structured collection of text that is used for training, testing or evaluating NLP models. A custom corpus is a specialized collection of text data tailored to meet the specific needs of a particular NLP task or application.”

WORD CLOUD

A word cloud presented the most frequently-occurring word in a body of text in the largest font size, and the least-frequently-occurring word, the smallest. In the one hundred video game reviews curated, the most frequently occurring word was “fun,” so we infer this video game was a “fun” video game to play!

SIMILARITY DETECTION

In a future revision, this author plans to get similarity detection working properly. Similarity detection compares two bodies of text and determines ‘how similar’ they are. Length of the text is also considered.

WORKS CITED

(n.d).

https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/#sample-review.

GeeksforGeeks. (2025, July 11). NLP | Custom corpus.

<https://www.geeksforgeeks.org/nlp/nlp-custom-corpus/>.

IBM. (n.d.). What is artificial intelligence (AI)? I.B.M.

<https://www.ibm.com/think/topics/artificial-intelligence>.