

# Natural Language Processing for UX

Leslie A. McFarlin

Principal UX Architect



# Table of Contents

- [What is Natural Language Processing \(NLP\)?](#)
- [Applying NLP to UX research](#)
  - Topic Modeling
  - Sentiment Analysis
- [Advanced Work](#)
  - Custom Neural Networks
  - Future Applications
- [Important Links](#)

# What Is Natural Language Processing (NLP)?

- AI branch that helps computers understand language similar to how humans do.
- Combines rule-based modeling of language with statistics, machine learning, and deep language models.
- Common NLP tasks include:
  - Speech-to-text
  - Part of speech (POS) tagging
  - Sentiment Analysis
  - Speech generation

# Applying NLP to UX Research

# General Application for UX@{Employer}

- Currently, NLP contributes to two main applications:
  - Topic Modeling
  - Sentiment Analysis
- Topic modeling is a way to highlight themes in texts.
- Sentiment analysis offers insights into the tenor of commentary.
- These methods are applied to text data, either collected from surveys or transcripts generated from audio recordings.
  - Survey data is usually generated into a XLS file from SurveyMonkey.
  - MS Teams automatically generates recordings into a DOC file.

# Preparing for NLP: Data Cleaning

- Data cleaning prepares language content for text analysis.
  - Multi-step process for getting data ready for more in-depth analyses.
  - Time spent in this step can be minimized by a **data quality plan**.
- Data cleaning consists of the following general steps:
  - Removing duplicates, irrelevant data points, and missing data.
    - Definitions of irrelevant data depend on your research question.
    - Missing data for NLP as used by {Employer} UX is not usually a significant issue but looking at patterns of missing data can be insightful.
  - Standardize text inputs to remove capitalization and punctuations, expand contractions, fix typos, and reduce words to stems or lemmas.
  - Remove stop words, words that hold no value for the content (articles, prepositions, etc.)
- **Note:** Always keep a copy of uncleaned data as some basic metrics are useful to derive from it (utterance lengths, word counts, character counts).

# Preparing for NLP: Text Analysis

- Text analysis offers qualitative summaries of the properties of the components within a text (words, sentences, paragraphs).
  - Text analysis can tell you what individual texts look like.
- Provides information useful for topic modeling and can set a larger project up for text analytics.
  - Text analytics example: Classify fleet requests by request category (text analysis), and then calculate average time to resolution by category.
- Text Analysis guide to be developed to help with this.

# Preparing for NLP: Tools

- NLP work done to date is with Jupyter Lab via Anaconda.
- Python's various NLP libraries provide many of the needed functions for performing text analytics.
  - [spaCy](#)
  - [Natural Language Toolkit](#) – used since May 2021.
  - [gensim](#) – used since October 2021.
- Be sure to include libraries for handling files, data, and making visualizations as needed.
  - File handling: os
  - Data handling: pandas, numpy
  - Visualizations: matplotlib, seaborn, plotly



# Topic Modeling

- Statistical modeling that reveals latent topics within a collection of documents.
- Three common methods:
  - Latent Dirichlet Allocation (LDA)- part of a PoC conducted via FAC-001.
  - Latent Semantic Analysis (LSA)
  - Probabilistic Latent Semantic Analysis (pLSA)
- LDA, a method for texts to be grouped according to latent (unseen) factors in the text.
- LSA, analyzes the relationship between sets of documents and the terms they contain.
- pLSA, expanded LSA model that models co-occurrences according to the probabilities that they will appear in texts.

# Sentiment Analysis

- An analysis that determines the valence of a text (positive, negative, or neutral).
- {Employer} can leverage pre-trained algorithms such as BERT or its variation distilBERT.
  - Requires fine tuning the selected algorithm on a data set similar to the text data collected by {Employer}.
  - Multiple PoCs are exploring the use of BERT and its variations.
- Once {Employer} has a large enough data set to train and test on, a custom neural network should be developed to provide truly accurate findings.
  - Currently there are just over 1000 data points of {Employer} -specific text data.
- Sentiment analysis, when combined with other analyses can provide insights into topics or situations and how clients feel about them.

# Advanced Work

# Custom Neural Networks

- Creating a neural network for sentiment analysis would allow for more accurate results as it could be trained on {Employer} -specific data.
- Both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) can be used for sentiment analysis, however RNNs would be more useful for {Employer} .
- Ensuring accuracy for a {Employer} neural network requires the following:
  - Creating a data quality plan to ensure unstructured data meets a standard for use in training and analysis.
  - Maintenance of training data set or sets.
  - Curation of a {Employer} corpus.

# Topics for Advanced Work

- Audio transcription- using existing Python libraries to generate transcripts from audio recordings.
  - Proof of Concept: [FAC-001\\_AudioExperimentation\\_20October2021.ipynb](#)
  - Requires a subscription to text generation service.
- VUI prototyping via AWS console.
  - 8 Mar 2022: Setup guide in progress.

# Important Links

# {Employer} Links

- [UX Team NLP Page](#)
- {Employer} Git hub link: {removed}

Thank you!

