Text Analytics with Azure Text Analytics

Azure AI: Language Service

Introduction to Text Analytics

Text Analytics - ad as an .md file in IntelliJ

https://slidesgpt.com/l/h9Sr

- Analyzes text through a series of natural language processing (NLP) techniques.
- Natural Language Processing (NLP) - a field of AI focused on enabling machines to understand, interpret and generate human language content.
- Text Analytics goes beyond simple word analysis, and involves interpreting the context, syntax and semantics of text
- Azure Text Analytics service utilizes NLP techniques to analyze text.

Various NLP Techniques

Use Input Text:

“Microsoft was founded by Bill Gates and Paul Allen on April 4, 1975.”

1. **Tokenization**:
   - The text is broken down into smaller units called tokens, which can be words, phrases, or other meaningful elements.
     i. ["Microsoft", "was", "founded", "by", "Bill Gates", "and", "Paul Allen", "on", "April 4, 1975"]
2. **Part-of-Speech Tagging**:
   - Each token is tagged with its part of speech (e.g., noun, verb, adjective), which helps in understanding the syntactic structure of the sentence.
     i. Tags words like "Microsoft" (noun), "founded" (verb), "Bill Gates" (proper noun), ..
3. **Named Entity Recognition (NER)**:
   - Identifies and categorizes entities within the text, such as names of people, organizations, locations, dates, and more.
     i. Recognizes “Microsoft” as an organization, “Bill Gates” and “Paul Allen” as persons, and “April 4, 1975” as a date
4. **Sentiment Analysis**:
   - Determines the sentiment expressed in the text, usually as positive, negative, or neutral. This involves analyzing words, phrases, and context to gauge the sentiment.
     i. Determines the sentiment of the text as neutral, positive or negative

Various NLP Techniques

5. **Key Phrase Extraction**:

    - Extracts significant phrases that represent the main topics or themes of the text.
        i. Extracts key phrases like “Microsoft”, “Bill Gates” “Paul Allen”, “April 4, 1975.”

6. **Language Detection**:

    - Detects the language in which the text is written.

7. **Text Normalization**:

    - Standardizes the text by converting it to a consistent format, such as converting to lowercase, removing punctuation, and stemming or lemmatizing words (reducing words to their root form).

8. **Dependency Parsing**:

    - Analyzes the grammatical structure of the sentence to understand relationships between words (e.g., subject, object).

Sentiment Analysis in Text Analytics

Sentiment Analysis plays a crucial role in Text Analytics by uncovering and analyzing the emotional tone expressed in text data. It’s imperative for businesses to understand customer feedback, social media sentiments and market trends, that enable businesses to make informed decisions based on user attitudes.

Various Tools and Libraries for Sentiment Analysis

- Azure AI: Language Service - Sentiment Analytics
- NLTK
- TextBlob
- VADER

Azure Text Analytics

Azure Text Analytics is an API service model:

It leverages a combination of NLP techniques to analyze text data effectively.

It simplifies text analysis processes and enables developers to build intelligent applications that understand and process text data accurately.

Building a Sentiment Text Analytics Python app

Program Description:

This script is designed to be a simple demonstration of using Azure Text Analytics to analyze the sentiment of a set of predefined text documents.

Program authenticates the client using the Azure Text Analytics key and endpoint. It creates and returns an instance of `TextAnalyticsClient`.

Program then performs sentiment analysis on a predefined list of documents. It sends the documents to the Azure Text Analytics service using the provided client and prints out the sentiment (positive, negative, or neutral) along with confidence scores for each document.