**Critical Documentation of the Skills of Max ( Voice Assistant)**

**Natural Language Processing:**

Speech to Text

Text To Speech

Natural Language Understanding

**Example Skill for Document Preparation on Max :**

**Sequence Diagram between the Alice (User) and Max( Voice Assistant)**

Graphical user interface, text, application, email

Description automatically generated

**Natural Language Understanding :**

**Text Pre-processing:**

* **Tokenization** : splitting a phrase, sentence, paragraph, or an entire **text** document into smaller units, such as individual words or terms
* **Lemmatization** : a common normalization technique in text pre-processing. In lemmatization, words are replaced by their root form or words with similar context. Another text normalization technique like this is called **stemming**. This is often done alongside manual tokenization to yield useful tokens.

**Example : Connect -Connect, connected – Connect, Connects- Connect.**

* **Removing stop words:**
* **Vocabulary building :**
* **Encoding and decoding :** Now that we have a vocabulary of words in the dataset, each of the patterns can be encoded into numerical features for modelling, using any of the common text encoding techniques—count vectorizer, term frequency-inverse document frequency (TF-IDF), hashing, etc.

**Modelling: Transformers (Algorithm)**

1. Bidirectional Encoder Representations from Transformers.
2. T5