



Natural Language Understanding Unit

Algorithms and Techniques

Overview

Following is a brief overview for some of the techniques we thought of and read about for understanding the user input and sending a customized message depending on the intent of the input.

Intent Classification and Entity Extraction

Wit.ai

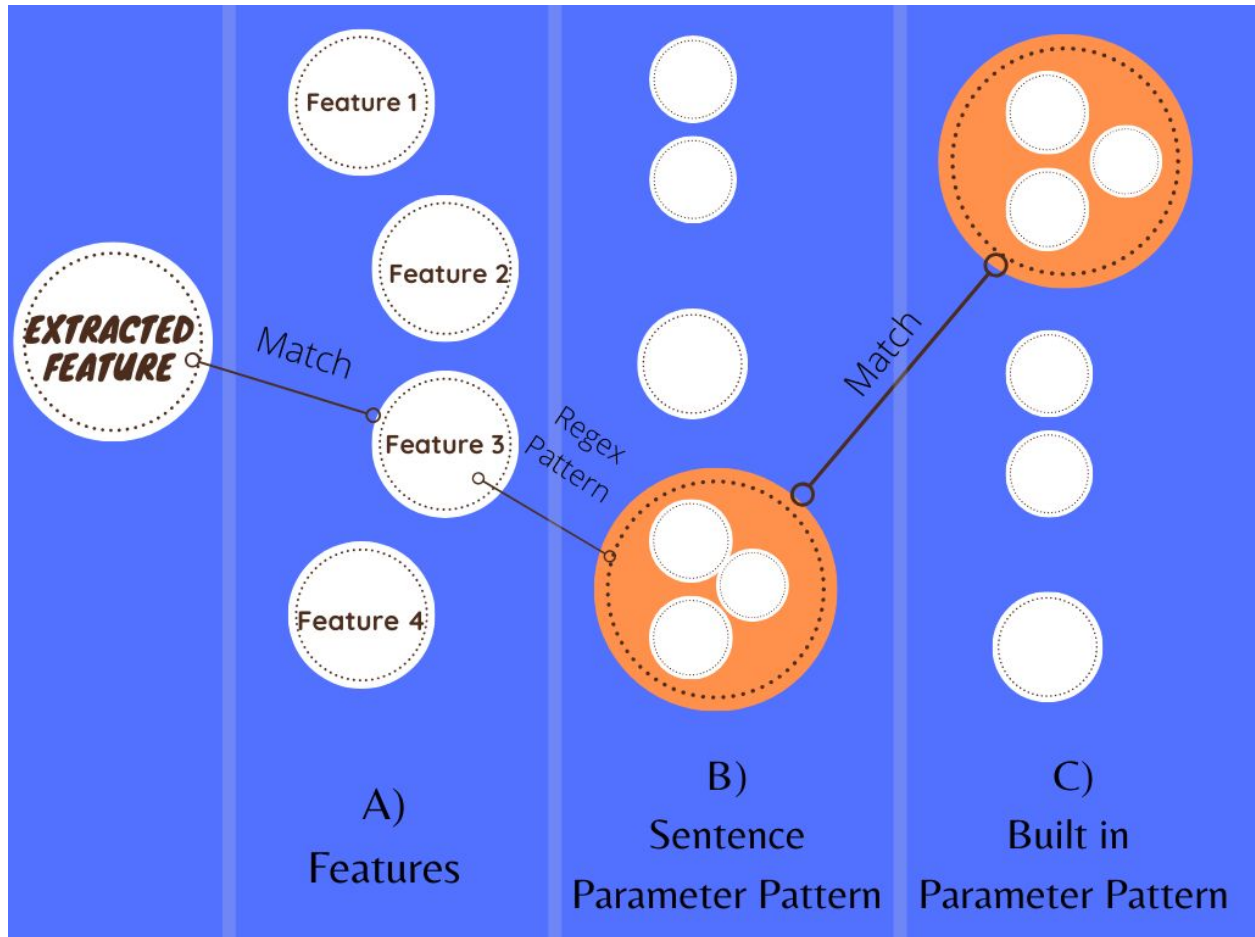
Wit.ai is a natural language interface for applications capable of turning sentences into structured data. It is an interface for using machine learning algorithms to train a model through feeding the model with expressions and specifying the intent and entities to be customized for the application's needs.

Intent Classification takes place using the [duckling library](#). For **Entity Extraction** takes place using a two-layered neural network that is trained to reconstruct linguistic context.

Regex and Node Matching

Using the linguistic some of the rules that were established by Naom Chomsky in his book Syntactic Structures and using NLTK and Regular Expressions; Each sentence is filtered and reduced into a list of Nouns and Verbs then they are matched with a Graph Database containing the keywords for the available features.

Each feature will have a Grammatical Pattern e.g. Noun + Verb + Adjective + Determiner, that has to be matched with the sentence to extract the parameter needed to be passed later to the API module then parsed then sent to the user.



A. Features

All the available features - services - that are matched with the extracted feature.

B. Sentence Parameter Pattern

All the possible Grammatical Patterns of the sentence.

C. Built in Parameter Pattern

Built in patterns for each feature are matched with the most fitting sentence parameter pattern, B.

Progress

The Intent classification and entity extraction is to be finished soon - excluding testing and debugging - as we are currently synchronizing the model with APIs by training the model and adding more expressions and specifying the intent and entities. It is considered to be more accurate or at least the accuracy will constantly increase each time a user sends a message.

Regex and Node Matching is a tricky part, as it requires considerable knowledge of linguistics and Python's Regex. It took a lot of research and reading and after all it might not be as accurate the classification model, even though it is a good topic that combines multiple fields of knowledge, it might take time.