SCQ Einführung in NLP

The respiratory elements are responsible for producing the wind which passes the vocal elements. The vocal elements produce the voice which is in the last step shaped by the articulating elements.

Speech production deals with the cognitive processes which are involved when thoughts are transformed into speech.

Articulatory phonetics deal with speech production. Acoustic phonetics are about aspects of the speech sound. Auditory phonetics address perception and understanding of speech.

In NLP, the amount of training data has a higher influence on the model performance than the machine learning approach.

Diaphasic variations are related to the communication medium. Diatopic variations refer to the linguistic area. Diachronic variations refer to variations of language over time. Diastratic

variations are about variations to social groups.

In pivot machine translation a source language is translated into a target language using a pivot language.

Early systems in NLP were mostly based on rules while nowadays most systems are based on statistics.

Anchors are used to mark a position in a string. Logical ORs are represented by disjunctions. Character classes are used to represent a certain group of characters. Repetitions of preceding elements can be defined by quantifiers.

When working with n-grams, a sequence of two words will be called a bigram. A sequence of three words is called trigram.

Input data for machine learning algorithms have to be in a numerical format. Therefore, unstructured text has to be transferred into a numerical format to embed words in a semantic vector space.

In text preprocessing, the text is split into smaller sub-units using tokenization. The process of removing of sub-units which have no impact for an NLP task is called stop-word removal.

Automatic speech recognition is about converting speech into text while speech synthesis is about generating speech based on a given text.

The speech-to-speech translation pipeline starts with a component for automatic speech recognition. The output of this component is processed by a text-to-text translation component which is followed by a text-to-speech synthesis.

Named entity recognition, relationship extraction and coreference resolution are typical examples for information extraction tasks.

Sentiment analysis and topic identification are both text classification problems. However, sentiment analysis deals with the subjective aspects of a text while topic identification is about the objective aspects.

Chatbots consist of three components: a component for natural language understanding, a dialog management component, and a component for message generation.