

Computerlinguistisches Arbeiten: Protokoll zum Repetitorium der Computerlinguistik. Datenvisualisierung mit Pyplot.

Zuerst wurde der Unterschied zwischen Visualisierungen von Daten für die Bachelorarbeit erwähnt. Es gibt Diagramme und Grafen. Grafen sind Visualisierungen von Daten und Ergebnisse. Diagramme sind Flussdiagramme, Zusammenhänge, Systeme. Zu Datenvisualisierung gehören piecharts / Tortendiagramme, bar charts / Balkendiagramme und Liniendiagramme. Tortendiagramme werden verwendet, um das Verhältnis zwischen gleichwertigen Klassen darzustellen. Balkendiagramme benutzt man, wenn es viele Einzelpunkte gibt, die wichtig sind. Liniendiagramme benötigt man, wenn ein zeitlicher Zusammenhang gibt.

Pyplot ist ein Datenvisualisierungsprogramm. Es enthält die matplotlib Bibliothek, ein python package. Die Datenstruktur ist ein numpy Array. Das ist ein multidimensionales Array (Listen, Matrizen). Es ist eine Liste mit reellen Zahlen, ein Vektor. Mit Hilfe von linspace wird X, ein numpy Array, erzeugt. Es enthält einen Anfangswert, einen Endwert, eine Zahl von Werten (256) und inklusiv einen Endpunkt. Zu jedem Wert wird ein entsprechender Funktionswert ausgerechnet in einem zweiten numpy Array. $[-3.14, 3.03, 3.142] \rightarrow [f(-3.14), f(3.03), f(3.142)]$. Die Funktionen enthalten die dazugehörigen Y Werte.

Als nächstes wurde über Plot Objekt gesprochen. Mit dem Befehl `plt.plot(X,C)` wird ein Plot erstellt, wo für jeden Wert X auf der X-Achse der Wert C auf der C-Achse abgebildet wird. Es wird eine Linie mit dem Befehl `plt.show()` erzeugt. Es werden Plot Formate und ein Speicherungsbehl präsentiert.

Erster Schritt ist die Ausgestaltung von Linien: Liniendicke und farbliche Plotdarstellung. Einheit der Abbildung ist Inch. Es wird angegeben, wie viele Punkte pro Inch man haben will.

Als nächstes geht es um die Rahmen der Plots. Man nimmt Minimum und Maximum auf der X-Achse ($-\pi$, $+\pi$). Man erhöht sie anschließend um 10 Prozent.

Es wurde über Achsenbeschriftung besprochen. Mit entsprechendem Befehl kann man sagen, welche Achsenbeschriftung und wo man Achsenbeschriftung haben will.

Ein zentriertes Koordinatensystem wird wie folgt gemacht. Untere und obere X-Linien werden nach oben verschoben. Die obere Linie wird weggelöscht. Man verschiebt die linke und rechte Begrenzung nach rechts. Die rechte Begrenzung wird gelöscht.

Es wird gezeigt, wie man die Label übergibt. Annotation der Punkte wird mit Hilfe der `annotate()` Funktion gemacht.

Am Schluss wird ein Eindruck gegeben über:

- Scatter Plots
- Dimensionale Reduktion mit t-SNE, um die Daten mit mehr als 2 Dimensionen darstellen.
- Histogrammdarstellung
- Balkendiagrammdarstellung