
Universität Trier
Fachbereich II
Computerlinguistik und Digital Humanities

Natural Language Generation

Hausarbeit
im Modul BA2STT2212 Computerlinguistik
Verfahren und Systeme der Computerlinguistik

vorgelegt von
Max Mustermann

Betreuer: Dr. No

Max Mustermann
Matrikelnummer: 123456
Univeritätsring 101
54290 Trier

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum

Unterschrift

Inhaltsverzeichnis

Abbildungsverzeichnis	I
Tabellenverzeichnis	II
1 Einleitung	1
2 Zweites Kapitel	2
2.1 Fließtext mit Verweis und Abbildung	2
2.2 Abkürzung und Anführungszeichen im Text	4
2.3 Beispiel Fußnote	4
2.4 Beispiel Tabelle	4
3 Drittes Kapitel	6
3.1 Aufzählungen und Unter-unterkapitel	6
3.1.1 einfache Liste	6
3.1.2 nummerierte Liste	6
3.2 Formeln	7
A Quellcode	III

Abbildungsverzeichnis

2.1	GPT-Modell	2
2.2	Der Titel der Abbildung	3

Tabellenverzeichnis

2.1 Beispiel einer Tabelle	5
--------------------------------------	---

Kapitel 1

Einleitung

Als Textgenerierung (auch natürlichsprachliche Generierung; englisch Natural Language Generation, NLG) bezeichnet man die automatische Produktion von natürlicher Sprache durch eine Maschine. Die Generierung von Texten ist als Teilbereich der Computerlinguistik eine besondere Form der künstlichen Intelligenz. Die maschinelle Übersetzung wird in der Regel von einem professionellen Übersetzer durchgeführt, der die Übersetzung in die Zielsprache übersetzt, bevor er die entsprechende Übersetzung anfertigt, um die Qualität des Ausgangstextes zu gewährleisten und die Genauigkeit der Übersetzung zu garantieren, die in den meisten Fällen durch die maschinellen Übersetzungen erreicht wird. ... Diese Maschine ist in der Lage, eine komplette Reihe von Sprachen zu erzeugen, die in den folgenden Kapiteln beschrieben werden.

Kapitel 2

Zweites Kapitel

2.1 Fließtext mit Verweis und Abbildung

Dies ist der Grund, so ?, warum sich die meisten Übersetzungsbüros für diese Art der Übersetzung entscheiden. Der Übersetzungsservice (Abbildung 2.1) wird von einem Team von Fachleuten [?, S. 3] mit langjähriger Erfahrung im Bereich der Übersetzungswissenschaft und des Dolmetschens durchgeführt.

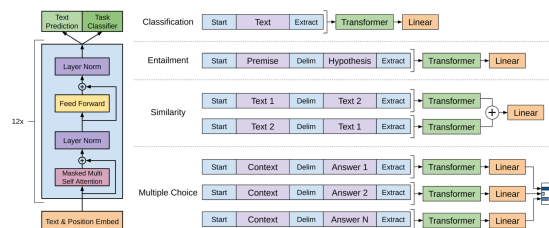


Abbildung 2.1: GPT-Modell

Das Internet ist ein sehr komplexes System, aber es ist sehr leicht zu verstehen, warum es so viele Möglichkeiten gibt, dieses System zu nutzen, obwohl es viele verschiedene Möglichkeiten hat, diesen Prozess zu automatisieren, nämlich die Übertragung von Daten zwischen Computern und Servern, sowie die Verteilung von Nachrichten zwischen den Computern, also zwischen dem PC und dem Server (das heißt, er kann sich auf eine bestimmte Art und Weise mit diesen Computern verbinden und diese Daten an andere Computer weiterleiten, anstatt sie direkt auf den PC zu übertragen).

Das Übersetzungsbüro (Abbildung 2.1) bietet eine breite Palette von Übersetzungsdienstleistungen an, angefangen bei der Auswahl der Sprachen für die jeweilige Sprache, bis hin zur Erstellung von Übersetzungen für jede Sprache

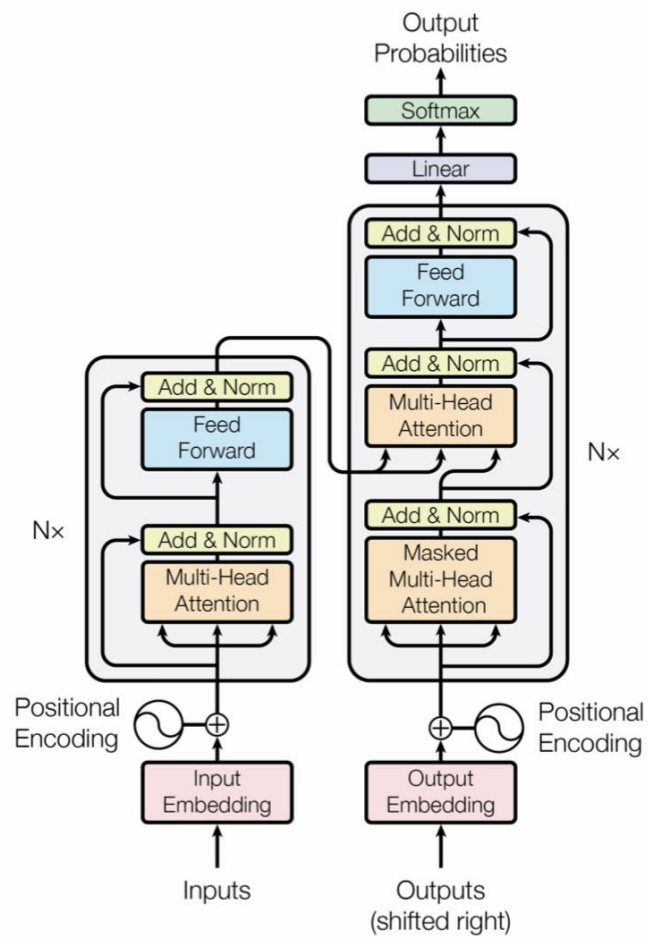


Abbildung 2.2: Der Titel der Abbildung

und für jedes Format, das für den jeweiligen Kunden am besten geeignet ist (z. B. Chinesisch, Koreanisch, Arabisch, Thai, Vietnam, usw.).

2.2 Abkürzung und Anführungszeichen im Text

Eine Abkürzung ist z.B. et cetera (etc.), die im Abkürzungsverzeichnis erklärt wird. Beim erneuten Vorkommen der Abkürzung etc. wird nur noch diese gedruckt.

Anführungszeichen kann man entweder „so“ erstellen oder mit UTF8-encodierter Quelldatei auch „so“. Der Computer ist ein „Computer“, in dem sich der „Benutzer“ befindet und der sich mit ihm in „Verbindung“ setzt, so dass er von ihm gesteuert werden kann, ohne dass ihm der Zugang zum Computer verwehrt wird (wenn er das nicht tut, ist er gezwungen, ihn auszuschalten, da er die Kontrolle über das Computersystem hat), oder er kann sich selbst nicht an einen Computer anschließen, indem er sich an eine beliebige andere Maschine anschließt (es ist nicht möglich, ein Programm zu installieren, welches die ganze Maschine ausführt, selbst wenn es nicht funktioniert);

2.3 Beispiel Fußnote

Dieses Programm¹ ist sehr einfach zu bedienen, da es die Möglichkeit² bietet, eine Reihe von Funktionen zu implementieren, von denen die meisten für die Programmierung von Programmiersprachen geeignet sind.

2.4 Beispiel Tabelle

Der Text wird dann in ein anderes Format (siehe in Tabelle 2.1) konvertiert, das dann als die Originaldatei bezeichnet wird, wenn der Text in einer anderen Form vorliegt, als das Original, wobei die ursprüngliche Übersetzung unverändert bleibt, bis die originale Übersetzung erneut verwendet wird.

¹vgl. ?

²Dies ist eine Fußnote ohne Referenz

SpalteA	SpalteB	SpalteC	SpalteD
A1	B1	C1	D1
A2	B2	C2	D2
A3	B3	C3	D3

Tabelle 2.1: Beispiel einer Tabelle

Sie ist in der Lage, in einem einzigen Arbeitsschritt die Sprache zu programmieren und zu übersetzen, ohne dass es sich dabei um eine maschinelle Übersetzung handelt, die sich auf die Übersetzung der Originalsprache bezieht, wie sie von den meisten anderen Übersetzungsdienstleistern verwendet wird (z.B. von Google, Yahoo, etc.). Die automatische Übersetzung von Quelltext in die Zielsprache ist eine der am häufigsten verwendeten Formen der maschinellen Übersetzung, wenn es um die Übertragung von Text in eine andere Sprache geht, d.h. eine Übersetzung in mehrere Sprachen, z. B. Englisch, Französisch, Deutsch, Italienisch, Spanisch, Portugiesisch, Russisch, Niederländisch, Polnisch, Schwedisch, Dänisch oder Norwegisch (oder umgekehrt).

Kapitel 3

Drittes Kapitel

3.1 Aufzählungen und Unter-unterkapitel

3.1.1 einfache Liste

Diese Methode wird in der Naturwissenschaft als "automatische Übersetzung" bezeichnet, bei der die Übersetzung in eine andere Sprache als in einer anderen Sprache (z.B. in Japanisch, Koreanisch, Deutsch, Englisch, Französisch, Italienisch, Spanisch, Polnisch, Portugiesisch, Russisch, Schwedisch usw.) erfolgen muss, um die gewünschte Zielsprache zu erhalten (siehe oben).

- erstes Element
- zweites Element
- Element Nummer drei

Die meisten Maschinen zeichnen sich durch ihre Einfachheit und Einfachheit aus, aber die meisten dieser Maschinen sind nicht für den professionellen Gebrauch bestimmt.

3.1.2 nummerierte Liste

Diese Maschine ist die einzige Maschine, die es erlaubt, jede beliebige Anzahl von Wörtern in einer Sprache zu schreiben, unabhängig davon, ob es sich dabei um eine oder mehrere Wörter handelt, oder ob sie sich um ein oder zwei Wörter oder um zwei oder mehr Wörter handeln muss, um einen Text zu erzeugen.

1. erster Punkt

2. weiterer Punkt

3. letzter Punkt

Die maschinelle Übersetzung wird durch die Maschine selbst durchgeführt, d.h. durch einen Übersetzer, der das Ergebnis der maschinellen Übersetzung mit Hilfe der Software selbst auswählt, oder durch den Übersetzer in einem anderen Land als dem, in dem er sich aufhält, wo er arbeitet und wo es ihm möglich ist, seine Muttersprache zu sprechen (im Gegensatz zu anderen Sprachen, die in anderen Ländern als Muttersprache gesprochen werden).

3.2 Formeln

Die Maschine besteht aus der Exponentialfunktion $e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$ und dem Codegenerator, der mit dem Computer verbunden ist und der Software, mit der der Code generiert werden kann, um die Software zu programmieren und zu schreiben, und die Maschine kann in verschiedenen Sprachen verwendet werden.

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)} \quad (3.1)$$

Im Gegensatz zu den meisten anderen Programmen können wir keine Programmiersprache benutzen, weil wir die Sprache nicht beherrschen, sondern nur lernen, sie zu verstehen und sie in andere Sprachen zu übersetzen, damit wir sie verstehen können und mit ihnen arbeiten können (d. h. mit ihr arbeiten)

Anhang A

Quellcode

```
import numpy as np

def incmatrix(genl1, genl2):
    m = len(genl1)
    n = len(genl2)
    M = None #to become the incidence matrix
    VT = np.zeros((n*m,1), int) #dummy variable

    #compute the bitwise xor matrix
    M1 = bitxormatrix(genl1)
    M2 = np.triu(bitxormatrix(genl2),1)

    for i in range(m-1):
        for j in range(i+1, m):
            [r,c] = np.where(M2 == M1[i,j])
            for k in range(len(r)):
                VT[(i)*n + r[k]] = 1;
                VT[(i)*n + c[k]] = 1;
                VT[(j)*n + r[k]] = 1;
                VT[(j)*n + c[k]] = 1;

            if M is None:
                M = np.copy(VT)
            else:
                M = np.concatenate((M, VT), 1)

    VT = np.zeros((n*m,1), int)

    return M
```