

Das Erstellen von Wissensgraphen aus unstrukturierten Texten mit Hilfe von großen Sprachmodellen

Houda El Abbassi

A thesis submitted in partial fulfilment of the requirements
for the degree of
Bachelor of Science

Principal Supervisor:
Prof. Dr. Martin Simon

2024

DECLARATION

Signature: _____

2024

Abstract

Die Verwendung von Large Language Models (LLM) wie GPT-3.5 zur Erstellung von Wissensgraphen bietet mehrere Vorteile:

Textverständnis und Generierung: LLMs haben die Fähigkeit, menschenähnliche Sprachmuster zu verstehen und zu generieren. Durch die Verarbeitung großer Mengen von Text können sie Beziehungen zwischen verschiedenen Konzepten erkennen und diese in einem strukturierten Format wiedergeben.

Automatisierung: LLMs ermöglichen eine teilweise oder vollständige Automatisierung des Prozesses zur Erstellung von Wissensgraphen. Das bedeutet, dass Sie große Mengen von unstrukturierten Daten verarbeiten und in einem organisierten Wissensgraphenformat ausgeben können, ohne dass menschliche Expertise für jeden Schritt erforderlich ist.

Skalierbarkeit: Durch die Nutzung von LLMs können Wissensgraphen auf große Mengen von Daten angewendet werden, was die Skalierbarkeit des Systems verbessert. Dies ist besonders nützlich, wenn Sie mit umfangreichen und sich ständig ändernden Datenmengen arbeiten.

Echtzeitaktualisierung: Da LLMs in der Lage sind, neue Informationen zu verarbeiten, können Wissensgraphen schnell aktualisiert werden, wenn sich der zugrunde liegende Wissensstand ändert. Dies ermöglicht eine dynamische und aktuelle Darstellung von Informationen.

Semantische Suche und Analyse: Ein Wissensgraph, der mithilfe von LLMs erstellt wurde, ermöglicht eine semantische Suche, bei der die Bedeutung von Begriffen und Beziehungen verstanden wird. Dies verbessert die Präzision von Suchanfragen

und ermöglicht fortgeschrittene Analysen.

Anwendungen in KI-Systemen: Wissensgraphen, die von LLMs erstellt wurden, können als Grundlage für KI-Systeme dienen, die komplexe Aufgaben wie maschinelles Lernen, Sprachverarbeitung und Entscheidungsfindung durchführen.

Schlüsselbegriffe: Wissensgraphen, Künstliche Intelligenz, NaLLM Framework

Table of Contents

Declaration	i
Abstract	ii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Chapter 1 Einführung	1
1.1 Motivation	1
1.2 Outline	1
1.3 Forschungsfragen /Ziel	1
1.4 Knowledge Graph Construction	1
Chapter 2 Wissensgraphen	3
2.1 Wissensgraphen	3
2.2 Entitäten	3
2.3 Natural Language Generation	3
2.4 Learning to Rank	3
Bibliography	4

List of Tables

List of Figures

Chapter 1

Einführung

Warum Wissensgraphen relevant sind und benutzt werden und dies in Kombi im LLM

1.1 Motivation

1.2 Outline

1.3 Forschungsfragen /Ziel

Ist der erstellte Wissensgraph qualitativ gut im Vergleich zu einem manuell erstellten Wissensgraphen? Wie geht das Framework mit redundanten Daten um oder mit einem nicht wissensreichen Text? Wie sieht der Wissensgraph aus bei Nachtragung? Ist das Erstellen von Wissensgraphen mit Hilfe von großen Sprachmodellen zuverlässig?

1.4 Knowledge Graph Construction

Wissensgraphen Konstruktion erzielt das Extrahieren von strukturierten Daten von unstrukturierten Texten, mit verschiedenen Methoden Named Entity Recognition (NER) (Chiu and Nichols, 2016), Relation Extraction (RE) (Zeng et al., 2015),

Event Extraction (EE) (Chen et al., 2015), Entity Linking (EL) (Shen et al., 2015), and Knowledge Graph Completion (Lin et al., 2015).

NER hat das Ziele verschieden Entitäten zu Identifizieren. RE extrahiert die Beziehung zu den gegebenen Entitäten EE EL versucht zwischen Entitäten eine Beziehung zu finden

Chapter 2

Wissensgraphen

2.1 Wissensgraphen

2.2 Entitäten

Entitäten können realweltliche Objekte, Ereignisse und abstrakte Konzepte darstellen, die einem Knoten entsprechen, und gerichtete Kanten werden als Beziehungen betrachtet. Der Wissensgraph speichert objektive Informationen, die in RDF-ähnlichen Tripeln strukturiert sind. Diese bestehen aus zwei Entitäten und einer Beziehung in Form von (Kopf, Beziehung, Schwanz) oder (Subjekt, Prädikat, Objekt) [8]. In einem Wissensgraphen sind Labels Arten von Beziehungen, die die Fakten verknüpfen können; Kanten (Beziehungen) sind spezifische Fakten, die zwei Knoten (Kopf- und Schwanzentitäten) verbinden.

2.3 Natural Language Generation

2.4 Learning to Rank

bibliography