

# Automatisierte Verarbeitung natürlichsprachlich repräsentierter Sachverhalte zur Identifizierung von Kandidaten für Bezeichner in Datenmodellen

## Eine vergleichende Untersuchung

Sven Christ<sup>1</sup> und Stefan Strecker<sup>1</sup>

**Abstract:** Für das Bestimmen von Kandidaten für Bezeichner von Modellelementen (Entitätstypen, Beziehungstypen, Attributen) aus natürlichsprachlich repräsentierten Sachverhaltsbeschreibungen werden für die Datenmodellierung mit der Modellierungssprache „Entity-Relationship Model“ (ERM) Heuristiken vorgeschlagen, die an Morphologie und Grammatik der natürlichen Sprache orientiert sind. Bereits seit den 1990er Jahren werden diese Heuristiken in Verbindung mit Ansätzen des „Natural Language Processing“ (NLP) eingesetzt, um für das Erstellen von Datenmodellen eine (teil-) automatisierte Modellierungsunterstützung zu realisieren. In diesem Beitrag kontrastieren wir die für das Modellierungswerkzeug TOOL implementierte NLP-basierte Modellierungsunterstützung mit drei Transformer-basierten künstlichen neuronalen Netzen, „Large Language Model“ (LLM), hinsichtlich fünf unterschiedlich komplexen Aufgaben des Identifizierens von Kandidaten für Bezeichner von Modellelementen in einer Variante des ERM. Die vorliegenden, noch vorläufigen Ergebnisse deuten an, dass die verwendeten LLM dem kontrastierten regelbasierten NLP-Ansatz deutlich überlegen sind.

**Keywords:** Datenmodellierung, Entity-Relationship Model, Natural Language Processing, Large Language Model, Identifier Reconstruction

## 1 Problemstellung, Motivation und Forschungsziele

Das Modellierungswerkzeug TOOL [Te19; Te20; TS18] unterstützt die Modellierenden während der Erstellung von Datenmodellen durch automatisiert zur Modellierzeit aus natürlichsprachlich repräsentierten Sachverhaltsbeschreibungen vorgeschlagenen Kandidaten für Bezeichner von Entitätstypen, Beziehungstypen und Attributen [TRS21a; TRS21b]. Damit adressiert TOOL die, nicht nur bei Modellierungsanfängern wiederholt beobachtete Modellierungsschwierigkeit, sinnhafte, zweckangemessene Bezeichner für Modellelemente zu rekonstruieren und zu vergeben („Identifier Reconstruction“) [RSS23].

Die in TOOL realisierte Modellierungsunterstützung basiert auf, an Morphologie und Grammatik der natürlichen Sprache (konkret: Englisch) orientierten Heuristiken für die

---

<sup>1</sup> FernUniversität in Hagen, Lehrstuhl für BWL, insb. Entwicklung von Informationssystemen, Universitätsstraße 41, 58097 Hagen, Deutschland, sven.christ@fernuni-hagen.de; stefan.strecker@fernuni-hagen.de  
This work is licensed under Creative Commons Attribution 4.0 International License <http://creativecommons.org/licenses/by/4.0/>, <https://doi.org/10.18420/modellierung2024-ws-020>

Datenmodellierung mit dem ERM (z. B. [CL89]) und deren Anwendung auf natürlichsprachliche Sachverhaltsbeschreibungen, die durch eine NLP-Implementierung der maschinellen Sprachverarbeitung aus der Computerlinguistik, Stanford CoreNLP [Ma14], aufbereitet wird. Die derzeit vorliegende Implementierung basiert auf der Rekonstruktion ausgewählter Heuristiken in Form von Regeln für eine adaptierte Pipeline der Stanford CoreNLP, anhand derer in erster Linie die Morphologie und Grammatik von natürlichsprachlichen Aussagen in einer Sachverhaltsbeschreibung analysiert und typisiert wird, um aufbauend auf dieser Typisierung etablierte Heuristiken in Regelform zu formulieren. Der in TOOL eingesetzte Ansatz ist insofern zu den regelbasierten NLP-Techniken zu zählen [Ch23].

Ergebnisse einer Evaluation der vorliegenden Implementierung mit zunehmend komplexeren Sachverhaltsbeschreibungen legen die Grenzen des gewählten Ansatzes in Bezug auf elementare Zielkriterien wie die Erkennungsrate („Precision“ und „Recall“) offen und deuten darüber hinaus prinzipielle Einschränkungen etwa bei der Berücksichtigung des satzübergreifenden Kontexts und von impliziten Bedingungen (in einer Sachverhaltsbeschreibung) an. Gleichzeitig stellen die in den letzten Jahren verfügbar gewordenen Sprachverarbeitungsansätze der subsymbolischen Künstlichen Intelligenz (KI), derzeit verbreitet diskutiert unter dem Bezeichner „Large Language Model“ (LLM), die Frage, ob und inwiefern sich diese, nicht an explizit aufbereiteten Heuristik-basierten Regeln orientierten Ansätze für die Aufgabenstellung des Identifizierens von Kandidaten für Bezeichner von Modellelementen in Datenmodellen mit einer Variante des ERM eignen.

In dem vorliegenden Beitrag kontrastieren wir die in TOOL realisierte, NLP-basierte Modellierungsunterstützung mit drei unterschiedlich umfangreich ausgelegten Transformer-basierten LLM, um dieser Frage nachzugehen und erste Einsichten in die gegenwärtige Leistungsfähigkeit der LLM zu erhalten. Dazu führen wir eine vergleichende Studie mit fünf zunehmend komplexeren natürlichsprachlichen Sachverhaltsbeschreibungen durch, die im Anhang aufgeführt sind. Die Ergebnisse dieser vergleichenden Untersuchung schließen an erste Diskussionsbeiträge zur Nutzung von LLM für das konzeptuelle Modellieren an [FFK23] und dienen als Grundlage für die weitere Auseinandersetzung mit der gezielten Nutzung von Ansätzen der subsymbolischen KI für das Entwickeln von zielgruppenadäquater Modellierungsunterstützung.

## 2 Ausgangspunkt

Ausgangspunkt der in TOOL derzeit realisierten Modellierungsunterstützung [TRS21a; TRS21b] bilden Heuristiken, die Morphologie und Grammatik der natürlichen Sprache heranziehen, um Kandidaten für sachgerechte Bezeichner von, in Datenmodellen instanziierten Modellierungskonzepten des ERM zu identifizieren. So weist Chen [Ch83] darauf hin, dass Gattungsnamen („common noun“) Kandidaten für Entitätstypen repräsentieren (z. B. „car“, „beverage“, „tea“). Nachfolgende Arbeiten etwa von Du [Du08] und von Lucassen et al.

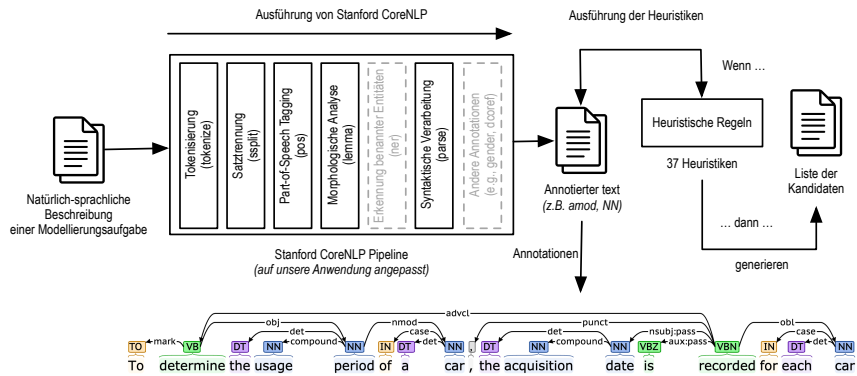


Abb. 1: Für die in TOOL realisierte Modellierungsunterstützung modifizierte „annotation pipeline“ der Stanford CoreNLP zur Generierung einer Liste von Kandidaten für Bezeichner von Entitätstypen, Attributen und Beziehungstypen (Abbildung adaptiert aus [TRS21b]).

[Lu17] untersuchen zunehmend komplexere Heuristiken zur Datenmodellierung, die an Morphologie und Grammatik der natürlichen Sprache orientiert sind (eine Zusammenstellung des gegenwärtigen Stands zeigen wir in [TRS21b]).

Darauf aufbauend implementiert die Modellierungsunterstützung in TOOL eine, an die Zielsetzung des Identifizierens von Kandidaten für Bezeichner von Modellelementen in Datenmodellen angepasste „annotation pipeline“ für die NLP-Bibliothek Stanford CoreNLP (in der Version 3.9.2) (siehe Abb. 1). Als Eingabe erhält die Pipeline einen Text in natürlicher Sprache, der anders als die Eingabe in einem LLM, nicht mit einem Prompt versehen werden muss. Die Ausgabe der von uns modifizierten Pipeline ist jeweils eine Liste der (ggf. falsch) erkannten Kandidaten für Entitätstypen, Beziehungstypen und Attribute anhand eines Bezeichners [TRS21b].

Auch wenn die Erkennungsraten der derzeitigen Implementierung je nach Sachverhaltsbeschreibung akzeptable Werte aufweisen, zeigt die erwähnte Evaluation, dass die Erkennungsraten bei komplexeren Beschreibungen deutlich niedriger ausfallen. Als eine Ursache dafür ist zu identifizieren, dass das in TOOL implementierte Regelwerk basierend auf 37 Heuristiken in seinem Zusammenwirken nicht so genutzt wird, wie es anhand der Aufgabenstellung zu erwarten ist. Die anschließende Detailanalyse der Ursachen führt uns zu der vorläufigen Einschätzung, dass die Komplexität des Zusammenwirkens des Regelwerks Korrekturen nur mit hohem Aufwand zulässt und gleichzeitig nicht abzuschätzen ist, ob dieser Aufwand in einem ökonomisch vertretbaren Verhältnis zu den erzielten Ergebnissen steht. Zudem ist bei komplexeren Sachverhaltsbeschreibungen zu vermuten, dass der gewählte regelbasierte NLP-Ansatz an prinzipielle Grenzen stößt etwa hinsichtlich der Berücksichtigung des satzübergreifenden Kontexts.

Vor diesem Hintergrund entsteht angesichts der diskutierten Leistungsfähigkeit der sogenannten „Large Language Model“ auch für Zwecke der konzeptuellen Modellierung (z. B. [FFK23]) die Motivation für die vorliegende vergleichende Untersuchung, in der wir den regelbasierten NLP-Ansatz in TOOL mit drei unterschiedlich umfangreichen LLM kontrastieren.

### 3 Untersuchungsdesign und Studiendurchführung

Das derzeit diskutierte Einsatzspektrum von LLM umfasst diverse Aufgaben der Mustererkennung, Sprachanalyse und Sprachsynthese sowie Codegenerierung [We22]. Aus diesem Spektrum erscheinen für die genannte Aufgabenstellung solche LLM besonders zielführend, die auf das Verarbeiten und Erzeugen natürlichsprachlicher Aussagen trainiert sind. Dabei ist die Leistungsfähigkeit eines LLM im Sinne der Verarbeitung und Generierung komplexer Sachverhalte maßgeblich von der Anzahl seiner Parameter abhängig [Bi23]. Die Parameter stellen dabei die Gewichte der Verbindungen zwischen den Neuronen des LLM zugrundeliegenden neuronalen Netzes dar und können als Fähigkeit des LLM interpretiert werden, Nuancen und Kontexte in natürlicher Sprache zu erkennen und zu verarbeiten [Ti24]. Um die Auswirkungen der Parameterkonfiguration eines LLM auf die von uns untersuchte Aufgabenstellung zu beurteilen, ziehen wir für die vorliegende vergleichende Studie vortrainierte LLM mit jeweils unterschiedlicher Parameteranzahl heran, wobei alle LLM als sogenannte Instruktionsmodelle trainiert sind, die einen „Prompt“ als Eingabe benötigen, der die Aufgabenstellung und die damit korrespondierende natürlichsprachliche Sachverhaltsbeschreibung enthält: Bei den in dieser Studie eingesetzten LLM handelt es sich um a) Mistral (M1) mit 7 Milliarden Parametern [Ji23], b) Llama 2 (M2) mit 13 Milliarden Parametern [To23] und c) Bard (M3) mit 137 Milliarden Parametern. M1 und M2 sind quelloffen verfügbar und wurden auf der Inferenz-Plattform HUGGINGFACE benutzt. Bard ist ein proprietäres LLM, das eine eigene Oberfläche zur Verfügung stellt [Go]. Diese drei LLM wurden stellvertretend für vergleichbare Modelle gewählt – auch weil sie sich in Vergleichsstudien als leistungsfähig erweisen [Co22].

Wir bewerten die Erkennungsraten anhand der Metriken *Precision* und *Recall*, die in der Bewertung von NLP-Systemen häufig Anwendung finden [De16; NMT05; WL19]. Diese Metriken erlauben es zu beurteilen, inwieweit genau die Bezeichner identifiziert werden, die für die jeweilige Aufgabe zutreffend sind. *Precision* und *Recall* sind definiert als:

$$Precision = \frac{N_{\text{richtig positiv}}}{N_{\text{richtig positiv}} + N_{\text{falsch positiv}}}$$

Formel 1: Wie viele der identifizierten Bezeichner sind tatsächlich Kandidaten für Bezeichner des jeweiligen Modellelements?

Das Untersuchungsdesign stellt jedem untersuchten Ansatz dieselbe natürlichsprachliche Sachverhaltsbeschreibung zur Verfügung, wobei für M1 bis M3 zusätzlich ein Prompt

$$Recall = \frac{N_{\text{richtig positiv}}}{N_{\text{richtig positiv}} + N_{\text{falsch negativ}}}$$

Formel 2: Wie viele der tatsächlichen Kandidaten für Bezeichner des jeweiligen Modellelements sind auch als mögliche Bezeichner identifiziert?

formuliert wird: „Ich gebe dir eine natürlich-sprachliche Beschreibung eines Sachverhalts und du sollst mögliche Bezeichner für Modellelemente eines ER-Modells generieren: Entitätstypen, Attribute, Beziehungstypen und Kardinalitäten: { Aufgabe }“. Die untersuchten fünf Sachverhaltsbeschreibungen und Aufgaben sind im Anhang dargestellt. Die tatsächlich positiven Bezeichner von Modellelementen wurden manuell und einstimmig festgelegt. Tatsächlich positiv sind alle Bezeichner, die in der Sachverhaltsbeschreibung vorkommen und die für das jeweilige Modellelement zutreffend sind. Nahe Synonyme (z.B. die Attribute „Name“ und „BeverageName“) wurden konsolidiert.

## 4 Untersuchungsergebnisse

Tabelle 1 zeigt die Ergebnisse der vergleichenden Untersuchung der derzeit in TOOL realisierten Modellierungsunterstützung (regelbasierter NLP-Ansatz) mit den LLM M1 bis M2 am 10.12.2023 auf der Plattform Huggingface, während M3 am 10.12.2023 auf einer proprietären Plattform von Google ausgeführt wurde. Die Auswertung von Entitätstypen und Attributen sowie Beziehungstypen und Kardinalitäten ist zusammen dargestellt, da das Regelwerk der derzeitigen Implementierung in TOOL beide Konzepte nur gemeinsam identifiziert.

Die Ergebnisse zeigen, dass die Erkennungsraten aller LLM denen des regelbasierten Ansatzes im Durchschnitt überlegen sind und auch die Erkennungsraten der LLM untereinander mit zunehmender Anzahl an Parametern steigen.

Auffällig ist, dass die LLM über alle Aufgaben hinweg durchgängig höhere Recall-Werte als Precision-Werte erzielen. Die eingesetzten LLM scheinen also in der Lage zu sein, eine große Anzahl relevanter Bezeichner zu identifizieren, allerdings generieren sie auch eine beträchtliche Menge an falsch-positiven Bezeichnern, was die Genauigkeit (Precision) reduziert.

Besonders bemerkenswert ist die Überlegenheit des LLM M1 gegenüber dem regelbasierten Ansatz von TOOL hinsichtlich der Evaluationsmetriken. Diese Überlegenheit wird noch augenscheinlicher, wenn man die steigende Anzahl der Parameter der LLM in Betracht zieht, die zu einer weiteren Verbesserung der Evaluationsmetriken führt. Für die gestellten Aufgaben zeigt sich die Tendenz, dass die Erkennungsraten mit einer größeren Anzahl an Parametern steigen.

Interessanterweise zeigen alle LLM, insbesondere auch M3, bei der komplexeren Aufgabe 5 überraschend niedrige Erkennungsraten der Beziehungstypen und Kardinalitäten. Für diese Beobachtung ist eine nahe liegende Erklärung unseres Erachtens nicht zu finden und zu begründen: A5 unterscheidet sich von A1 bis A4 durch die höhere Satzlänge und die mehrfachen Aufzählungen von Substantiven. Ob dies ursächlich für die niedrigen Erkennungsraten ist, lässt sich nicht ohne eine weitergehende Untersuchung schließen. Die Beobachtung zeigt jedoch auf, dass auch LLM nicht uneingeschränkt komplexe Satzkonstruktionen mit hohen Erkennungsraten verarbeiten. Weiterhin ist herauszustellen, dass gerade der regelbasierte Ansatz von TOOL in Aufgabe 5 bei der Erkennung von Entitätstypen und Attributen die besten Ergebnisse erzielt. Zwar sind diese absolut ausgedrückt immer noch niedrig, aber im Vergleich zu den LLM gelingt die Erkennung von Entitätstypen und Attributen durch den regelbasierten Ansatz von TOOL in Aufgabe 5 am besten.

## 5 Fazit und Diskussion

Dieser Beitrag macht Erkenntnisangebote über die Stärken und Schwächen beider Ansätze in der Datenmodellierung. Auffällig ist dabei, dass die LLM insgesamt bessere Ergebnisse liefern als der regelbasierte Ansatz, vor allem bei Aufgaben geringerer bis mittlerer

Tab. 1: Ergebnisse der Evaluation

			TOOL	M1	M2	M3
Aufgabe 1	Entitätstypen & Attribute	Precision	0.17	0.67	0.40	<b>1.00</b>
		Recall	1.00	0.67	0.67	<b>1.00</b>
	Beziehungstypen & Kardinalitäten	Precision	0.00	0.00	0.33	<b>1.00</b>
		Recall	0.00	0.00	<b>1.00</b>	<b>1.00</b>
Aufgabe 2	Entitätstypen & Attribute	Precision	0.20	0.60	0.75	<b>1.00</b>
		Recall	0.33	0.60	0.55	<b>0.80</b>
	Beziehungstypen & Kardinalitäten	Precision	0.00	0.29	0.67	<b>1.00</b>
		Recall	0.00	0.50	<b>1.00</b>	<b>1.00</b>
Aufgabe 3	Entitätstypen & Attribute	Precision	0.23	0.30	0.30	<b>1.00</b>
		Recall	0.60	0.50	0.50	<b>1.00</b>
	Beziehungstypen & Kardinalitäten	Precision	0.00	0.22	0.22	<b>1.00</b>
		Recall	0.00	0.50	0.50	<b>1.00</b>
Aufgabe 4	Entitätstypen & Attribute	Precision	0.25	0.38	<b>0.70</b>	0.50
		Recall	0.43	0.63	<b>0.88</b>	0.75
	Beziehungstypen & Kardinalitäten	Precision	0.00	0.00	0.22	<b>0.50</b>
		Recall	0.00	0.00	<b>1.00</b>	<b>1.00</b>
Aufgabe 5	Entitätstypen & Attribute	Precision	<b>0.18</b>	0.06	0.11	0.00
		Recall	<b>0.22</b>	0.13	0.15	0.00
	Beziehungstypen & Kardinalitäten	Precision	0.00	0.00	0.00	0.00
		Recall	0.00	0.00	0.00	0.00
Durchschnitt			0.18	0.30	0.50	<b>0.73</b>

Komplexität. Jedoch stoßen sie bei komplexeren Sachverhaltsbeschreibungen, insbesondere bei längeren Sätzen mit Aufzählungen, an ihre Grenzen.

Diese Beobachtung legt nahe, dass die Fähigkeit der LLM, komplexe sprachliche Strukturen und feine Nuancen zu erfassen, verbessert werden könnte. Ein möglicher Ansatz hierfür ist die Low-Rank Adaptation of Large Language Models (LoRA). LoRA ermöglicht es, LLM effizient und gezielt für spezifische Anwendungen anzupassen, indem zusätzliche, niedrigdimensionale Anpassungen an den Parametern des Modells vorgenommen werden. Diese Technik könnte genutzt werden, um die LLMs speziell für die Herausforderungen der Datenmodellierung zu optimieren.

Die Überlegenheit der LLM gegenüber dem regelbasierten Ansatz von TOOL zeigt sich vor allem in der Fähigkeit der LLMs, eine größere Anzahl relevanter Bezeichner zu identifizieren. Dies liegt vermutlich an ihrer Fähigkeit, komplexe Sprachmuster und Kontexte zu erkennen. Allerdings führt dies auch dazu, dass die LLM eine höhere Anzahl an falsch-positiven Ergebnissen produzieren, was die Genauigkeit (Precision) beeinträchtigt.

Interessanterweise zeigt sich in der komplexen Aufgabe 5, dass die LLM, einschließlich des umfangreichsten Modells M3, unerwartet niedrige Erkennungsraten aufweisen. Dies deutet darauf hin, dass die Fähigkeit der LLM, komplexe Satzstrukturen und insbesondere lange Aufzählungen effektiv zu verarbeiten, noch begrenzt ist. Hier könnte die Anwendung von LoRA einen wesentlichen Beitrag leisten. Durch die Anpassung der LLM mit LoRA könnten diese spezifisch für die Erfassung und Verarbeitung komplexer, domänenspezifischer Strukturen optimiert werden. Dies würde es den Modellen ermöglichen, nicht nur die allgemeine Sprachstruktur, sondern auch die feineren Details und spezifischen Anforderungen der Datenmodellierung besser zu erfassen.

Die Ergebnisse der Studie legen nahe, dass LLM eine effektive Alternative zu traditionellen, regelbasierten Ansätzen bieten, insbesondere für weniger komplexe Modellierungsaufgaben. Für komplexere Anforderungen zeigen sie jedoch Schwächen, die möglicherweise durch die Einbeziehung von Techniken wie LoRA gemildert oder behoben werden könnten. Durch die Integration solcher Anpassungen könnten LLM nicht nur die allgemeine Sprachstruktur, sondern auch spezifischere, kontextbezogene Details erfassen, was für die Präzision in der Datenmodellierung entscheidend ist.

## Literatur

- [Bi23] Biderman, S.; Schoelkopf, H.; Anthony, Q. G.; Bradley, H.; O'Brien, K.; Hallahan, E.; Khan, M. A.; Purohit, S.; Prashanth, U. S.; Raff, E.; Skowron, A.; Sutawika, L.; Wal, O. V. D.: Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling. In: Proceedings of the 40th International Conference on Machine Learning. International Conference on Machine Learning. PMLR, S. 2397–2430, 2023, URL: <https://proceedings.mlr.press/v202/biderman23a.html>, Stand: 18. 12. 2023.

- [Ch23] Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; Ye, W.; Zhang, Y.; Chang, Y.; Yu, P. S.; Yang, Q.; Xie, X.: A Survey on Evaluation of Large Language Models. Version 8/, 2023.
- [Ch83] Chen, P.P.-S.: English Sentence Structure and Entity-Relationship Diagrams. *Information Sciences* 29/2, S. 127–149, 1983, ISSN: 0020-0255.
- [CL89] Chen, H.-c.; Leung, Y.-s.: Patterns of Lexical Processing in a Nonnative Language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15/2, S. 316–325, 1989, ISSN: 1939-1285, 0278-7393.
- [Co22] Colombo, P.; Noiry, N.; Irurozki, E.; Cléménçon, S.: What Are the Best Systems? New Perspectives on NLP Benchmarking. In (Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; Oh, A., Hrsg.): *Advances in Neural Information Processing Systems*. Bd. 35, Curran Associates, Inc., S. 26915–26932, 2022, URL: [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/ac4920f4085b5662133dd751493946a6-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/ac4920f4085b5662133dd751493946a6-Paper-Conference.pdf).
- [De16] Derczynski, L.: Complementarity, F-score, and NLP Evaluation. In (Calzolari, N.; Choukri, K.; Declerck, T.; Goggi, S.; Grobelnik, M.; Maegaard, B.; Mariani, J.; Mazo, H.; Moreno, A.; Odijk, J.; Piperidis, S., Hrsg.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. LREC 2016. European Language Resources Association (ELRA), Portorož, Slovenia, S. 261–266, 2016.
- [Du08] Du, S.: On the Use of Natural Language Processing for Automated Conceptual Data Modeling. In. 2008, URL: <https://api.semanticscholar.org/CorpusID:60968024>.
- [FFK23] Fill, H.-G.; Fettke, P.; Köpke, J.: Conceptual Modeling and Large Language Models: Impressions From First Experiments With ChatGPT. *Enterprise Modelling and Information Systems Architectures (EMISAJ)* 18/, 3:1–15, 2023, ISSN: 1866-3621.
- [Go] Google: Google Bard - Try Bard, a Generative AI Chat Tool, URL: <https://bard.google.com>, Stand: 05.01.2024.
- [Ji23] Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; Casas, D. de las; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; Lavaud, L. R.; Lachaux, M.-A.; Stock, P.; Scao, T.L.; Lavril, T.; Wang, T.; Lacroix, T.; Sayed, W. E.: Mistral 7B. Version 1/, 2023.
- [Lu17] Lucassen, G.; Robeer, M.; Dalpiaz, F.; van der Werf, J.M.E.; Brinkkemper, S.: Extracting Conceptual Models from User Stories with Visual Narrator. *Requirements Engineering* 22/3, S. 339–358, 2017.



- [Ma14] Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J. R.; Bethard, S.; McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22–27, 2014, Baltimore, MD, USA, System Demonstrations. The Association for Computer Linguistics, S. 55–60, 2014.
- [NMT05] Nakache, D.; Metais, E.; Timsit, J. F.: Evaluation and NLP. In (Andersen, K. V.; Debenham, J.; Wagner, R., Hrsg.): Database and Expert Systems Applications. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, S. 626–632, 2005, ISBN: 978-3-540-31729-6.
- [RSS23] Rosenthal, K.; Strecker, S.; Snoeck, M.: Modeling Difficulties in Creating Conceptual Data Models. *Software and Systems Modeling* 22/3, S. 1005–1030, 2023.
- [Te19] Ternes, B.; Strecker, S.; Rosenthal, K.; Barth, H.: A Browser-Based Modeling Tool for Studying the Learning of Conceptual Modeling Based on a Multi-Modal Data Collection Approach. In (Ludwig, T.; Pipek, V., Hrsg.): Human Practice. Digital Ecologies. Our Future. 14. Internationale Tagung Wirtschaftsinformatik (WI 2019), February 24–27, 2019, Siegen, Germany. University of Siegen, Germany / AISEL, S. 1984–1988, 2019, URL: <https://aisel.aisnet.org/wi2019/specialtrack03/papers/7>.
- [Te20] Ternes, B.; Rosenthal, K.; Barth, H.; Strecker, S.: TOOL - Modeling Observatory & Tool: An Update. In (Michael, J.; Bork, D.; Fill, H.-G.; Fettke, P.; Karagiannis, D.; Köpke, J.; Koschmider, A.; Mayr, H. C.; Rehse, J.-R.; Reimer, U.; Striewe, M.; Tropmann-Frick, M.; Ullrich, M., Hrsg.): Companion Proceedings of Modellierung 2020 Short, Workshop and Tools & Demo Papers Co-Located with Modellierung 2020, Vienna, Austria, February 19–21, 2020. Bd. 2542. CEUR Workshop Proceedings, CEUR-WS.org, S. 198–202, 2020, URL: <https://ceur-ws.org/Vol-2542/Mod20-TuD2.pdf>.
- [Ti24] Titus, L. M.: Does ChatGPT Have Semantic Understanding? A Problem with the Statistics-of-Occurrence Strategy. *Cognitive Systems Research* 83/, S. 101174, 2024, ISSN: 13890417.
- [To23] Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; Bikel, D.; Blecher, L.; Ferrer, C. C.; Chen, M.; Cucurull, G.; Esiobu, D.; Fernandes, J.; Fu, J.; Fu, W.; Fuller, B.; Gao, C.; Goswami, V.; Goyal, N.; Hartshorn, A.; Hosseini, S.; Hou, R.; Inan, H.; Kardaş, M.; Kerkez, V.; Khabsa, M.; Kloumann, I.; Korenev, A.; Koura, P. S.; Lachaux, M.-A.; Lavril, T.; Lee, J.; Liskovich, D.; Lu, Y.; Mao, Y.; Martinet, X.; Mihaylov, T.; Mishra, P.; Molybog, I.; Nie, Y.; Poulton, A.; Reizenstein, J.; Rungta, R.; Saladi, K.; Schelten, A.; Silva, R.; Smith, E. M.; Subramanian, R.; Tan, X. E.; Tang, B.; Taylor, R.; Williams, A.; Kuan, J. X.; Xu, P.; Yan, Z.; Zarov, I.; Zhang, Y.; Fan, A.; Kambadur, M.; Narang, S.; Rodriguez, A.; Stojnic, R.; Edunov, S.; Scialom, T.: Llama 2: Open Foundation

- and Fine-Tuned Chat Models, 2023, arXiv: 2307.09288 [cs], URL: <http://arxiv.org/abs/2307.09288>, Stand: 19. 12. 2023, preprint.
- [TRS21a] Ternes, B.; Rosenthal, K.; Strecker, S.: Automated Assistance for Data Modelers Combining Natural Language Processing and Data Modeling Heuristics: A Prototype Demonstration. In (Lukyanenko, R.; Samuel, B. M.; Sturm, A., Hrsg.): Proceedings of the ER Demos and Posters 2021 Co-Located with 40th International Conference on Conceptual Modeling (ER 2021), St. John's, NL, Canada, October 18–21, 2021. Bd. 2958. CEUR Workshop Proceedings, CEUR-WS.org, S. 25–30, 2021, URL: <https://ceur-ws.org/Vol-2958/paper5.pdf>.
- [TRS21b] Ternes, B.; Rosenthal, K.; Strecker, S.: Automated Assistance for Data Modelers: A Heuristics-Based Natural Language Processing Approach. In (Rowe, F.; Amrani, R. E.; Limayem, M.; Matook, S.; Rosenkranz, C.; Whitley, E. A.; Quammah, A. E., Hrsg.): 29th European Conference on Information Systems: Human Values Crisis in a Digitizing World, ECIS 2021, Marrakech, Morocco, 2021. 2021, URL: [https://aisel.aisnet.org/ecis2021\\_rp/148](https://aisel.aisnet.org/ecis2021_rp/148).
- [TS18] Ternes, B.; Strecker, S.: A Web-Based Modeling Tool for Studying the Learning of Conceptual Modeling. In (Schaefer, I.; Karagiannis, D.; Vogelsang, A.; Méndez, D.; Seidl, C., Hrsg.): Modellierung 2018, 21.-23. Februar 2018, Braunschweig, Germany. Bd. P-280. LNI, Gesellschaft für Informatik e.V., S. 325–328, 2018, URL: <https://dl.gi.de/handle/20.500.12116/14954>.
- [We22] Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; Chi, E. H.; Hashimoto, T.; Vinyals, O.; Liang, P.; Dean, J.; Fedus, W.: Emergent Abilities of Large Language Models, 2022, arXiv: 2206.07682 [cs].
- [WL19] Wang, R.; Li, J.: Bayes Test of Precision, Recall, and F1 Measure for Comparison of Two Natural Language Processing Models. In (Korhonen, A.; Traum, D.; Márquez, L., Hrsg.): Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. ACL 2019. Association for Computational Linguistics, Florence, Italy, S. 4135–4145, 2019.

## Danksagung

Wir möchten Markus Redl für seine Unterstützung bei der Durchführung und Auswertung der LLM danken.

## Anhang

### Quelle von TOOL

TOOL ist nicht öffentlich verfügbar. Ein Zugang kann aber per E-Mail unter folgender Adresse angefragt werden:

lehrstuhl.strecker@fernuni-hagen.de

### Quellen der LLM

- M1: <https://huggingface.co/meta-llama/Mistral-7B-chat-hf>
- M2: <https://huggingface.co/meta-llama/Llama-2-13b-chat-hf>
- M3: <https://bard.google.com>

### Prompt

Ich gebe dir eine natürlich-sprachliche Beschreibung eines Sachverhalts und du sollst mögliche Bezeichner für Modellelemente eines ER-Modells generieren: Entitätstypen, Attribute, Beziehungstypen und Kardinalitäten: { Aufgabe }

#### 1. Coffee shop menu

A coffee shop menu consists of various beverages. Each beverage has a name, price, and list of ingredients. The menu categorizes beverages into coffee, tea, and specialty drinks.

#### Entity types & attributes

1. Beverage
  - BeverageID
  - BeverageName
  - BeverageType (e.g., Coffee, Tea, Specialty Drink)
  - Price
  - IngredientsList

#### Relationship types & cardinalities

1. Category-Beverage
  - One Category has many Beverages

## 2. Customer orders

A coffee shop has a menu with beverages categorized into coffee, tea, and specialty drinks, each with a name, price, and ingredients. Additionally, the coffee shop tracks customer orders, each with an order number, customer name, and the beverages ordered, including the quantity of each beverage and the total price.

### Entity types & attributes

1. OrderBeverage
  - Beverage
  - Quantity
2. Order
  - OrderID
  - CustomerID
  - TotalPrice
  - Beverages
3. Customer
  - CustomerID
  - CustomerName

### Relationship types & cardinalities

1. Customer-Order
  - One Customer places many Orders
2. Order-Beverage
  - One Order includes many Beverages

## 3. Inventory management

A coffee shop has a menu of beverages (coffee, tea, specialty drinks) with names, prices, and ingredients, and a system for tracking customer orders (order number, customer name, beverages ordered, quantity, total price). To manage the coffee shop's inventory, each ingredient (like coffee beans, milk, sugar) has a stock ID, name, and quantity in stock, with tracking of ingredient depletion based on orders and alerts for low stock.

### Entity types & attributes

1. **Ingredient**
  - IngredientID
  - IngredientName
2. **Inventory**
  - IngredientID
  - QuantityInStock

### Relationship types & cardinalities

1. Beverage-Ingredient
  - One Beverage uses many Ingredients
2. Ingredient-Inventory
  - One Ingredient has One Inventory

## 4. Staff scheduling

A coffee shop is offering a menu of beverages (names, prices, ingredients), processing customer orders (order numbers, customer names, beverages, quantities, total prices), and managing inventory (stock ID, name, quantity, stock depletion tracking). The coffee shop has introduced a staff scheduling system. Each staff member has an employee ID, name, and role (barista, manager, cashier). A schedule exists for various shifts, ensuring coverage throughout operating hours.

### Entity types & attributes

1. **Staff**
  - EmployeeID
  - EmployeeName
  - Role
2. **Shift**
  - ShiftID
  - ShiftTime
  - EmployeeIDs

### **Relationship types & cardinalities**

1. Staff-Shift
  - One Staff member is assigned to Many Shifts

### **5. Customer loyalty program**

A coffee shop uses a comprehensive system encompassing a beverage menu (names, prices, ingredients), customer orders (order numbers, customer names, beverages, quantities, total prices), inventory management (stock ID, name, quantity, depletion tracking), and staff scheduling (employee ID, name, role, shift assignments). The coffee shop has also developed a customer loyalty program for regular customers, tracking visits and purchases, with each customer having a loyalty card ID, name, and total number of visits, and offering rewards or discounts based on purchase frequency or volume.

### **Entity types & attributes**

1. **Loyalty Program**
  - LoyaltyCardID
  - CustomerID
  - TotalVisits
  - RewardIDs
2. **Reward**
  - RewardID
  - DiscountAmount

### **Relationship types & cardinalities**

1. Customer-Loyalty Program
  - One Customer is enrolled in One Loyalty Program
2. Loyalty Program-Reward
  - One Loyalty Program offers Many Rewards