



# Community of Practice KIPerWeb

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen



**KIPerWEB**



Forschungsinstitut  
Betriebliche Bildung

- **Update**
  - News & Leaderboard-Update
- **Input**
  - „Wissensgraphen und Ontologien“
- **Diskussion**

# News & Update (16.10.2024)

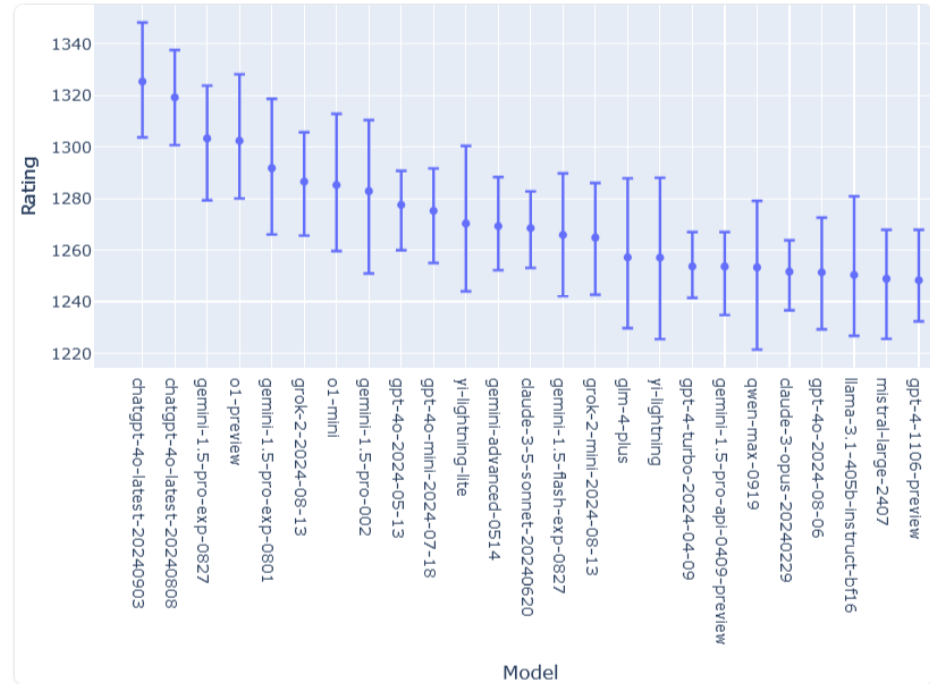


- Nicht im Leaderboard aber laut Benchmarks vor GPT-4o & Claude 3.5 Sonnet: NVIDIA Llama 3.1 Nemotron 70B:

Model	Arena Hard	AlpacaEval	MT-Bench	Mean Response Length
Details	(95% CI)	2 LC (SE)	(GPT-4-Turbo)	(# of Characters for MT-Bench)
<b>Llama-3.1-Nemotron-70B-Instruct</b>	<b>85.0</b> (-1.5, 1.5)	<b>57.6</b> (1.65)	<b>8.98</b>	2199.8
Llama-3.1-70B-Instruct	55.7 (-2.9, 2.7)	38.1 (0.90)	8.22	1728.6
Llama-3.1-405B-Instruct	69.3 (-2.4, 2.2)	39.3 (1.43)	8.49	1664.7
Claude-3-5-Sonnet-20240620	79.2 (-1.9, 1.7)	52.4 (1.47)	8.81	1619.9
GPT-4o-2024-05-13	79.3 (-2.1, 2.0)	57.5 (1.47)	8.74	1752.2

Quelle: <https://huggingface.co/nvidia/Llama-3.1-Nemotron-70B-Instruct-HF>

## Confidence Intervals on model strength (Arena Elo, German)

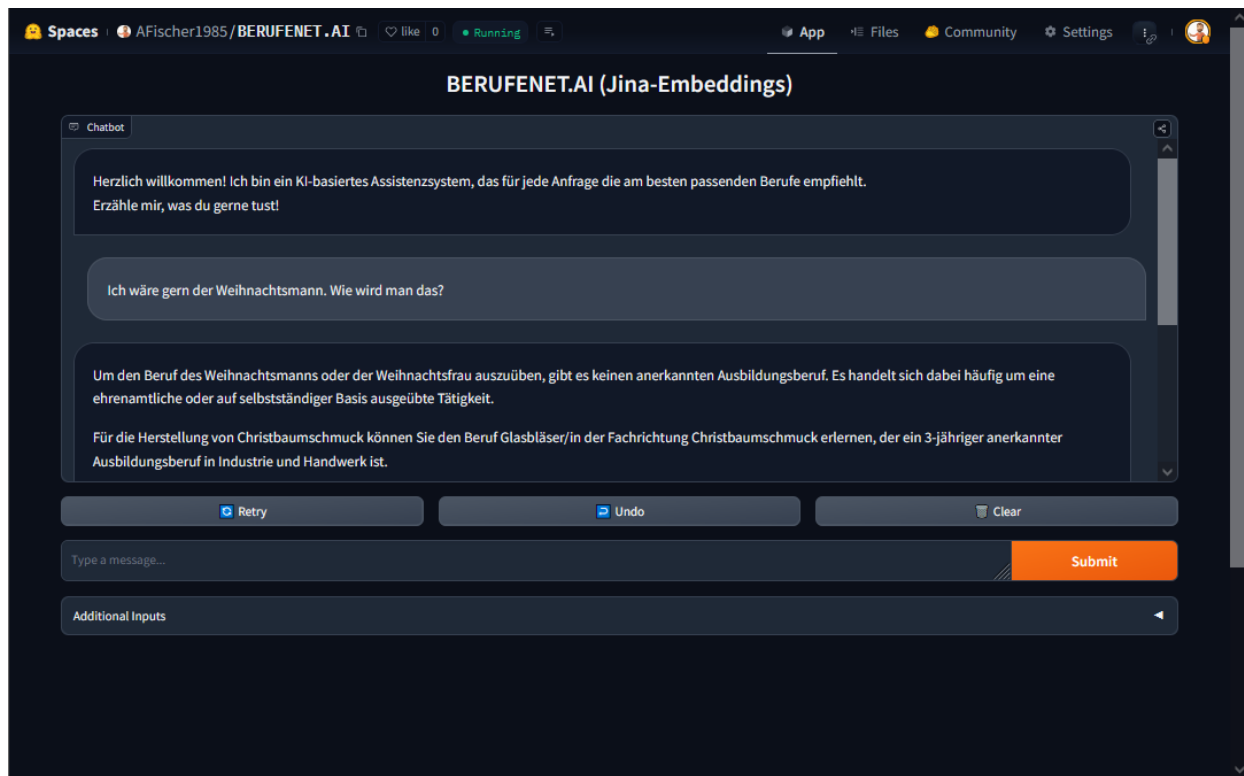


Quelle: <https://chat.lmsys.org/?leaderboard>

# Update: BERUFENET.AI – RAG mit kuratiertem Content



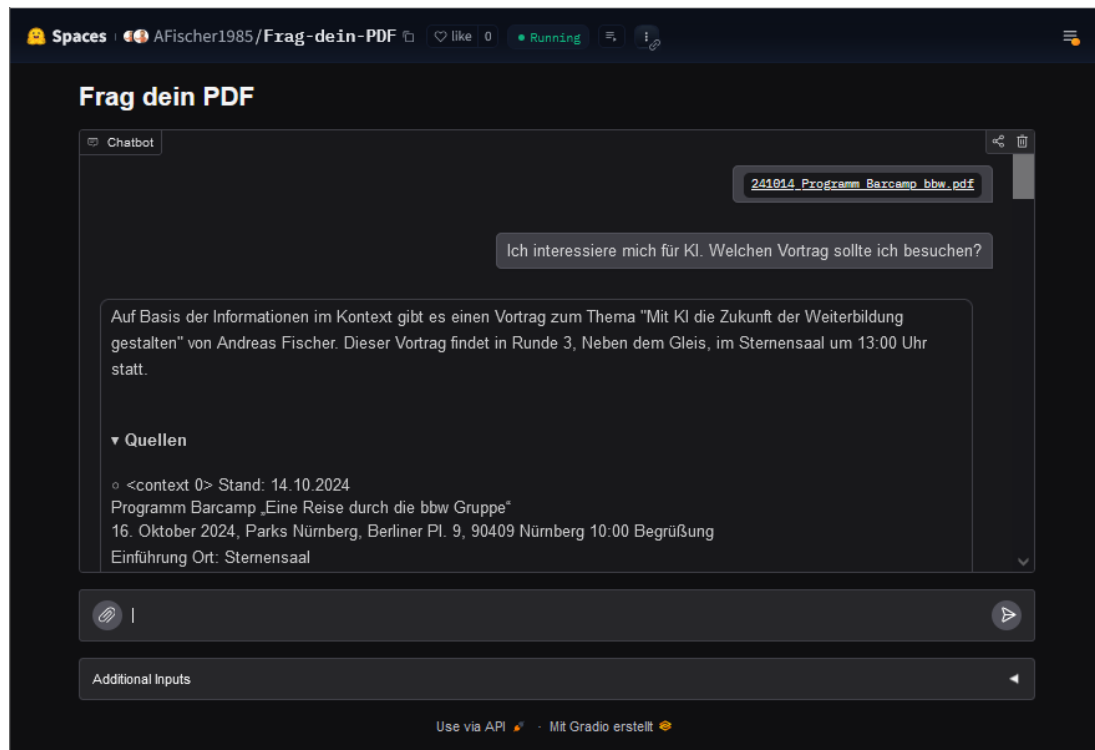
- Antwortet auf Basis der passendsten Berufsbeschreibungen auf BERUFENET
- RAG seit 15.10. auf Basis von Jina-Embeddings-v2-base-de & Mixtral-8x7B-Instruct-v0.1
- GraphRAG ist umgesetzt (aber bis auf Weiteres noch nicht public)



# Update: Frag-dein-PDF – RAG mit User-Content



- Prompt adaptiert von AnythingLLM
- Output mit Quellenangaben
- Klassisches RAG auf Basis von Jina-Embeddings-v2-base-de & Mixtral-8x7B-Instruct-v0.1
- OCR möglich aber oft unerwünscht
- ChromaDB-uploads personalisiert über session-hashes
- Chunking mit Overlap
- Auch hier ist eine GraphRAG-Variante möglich und geplant

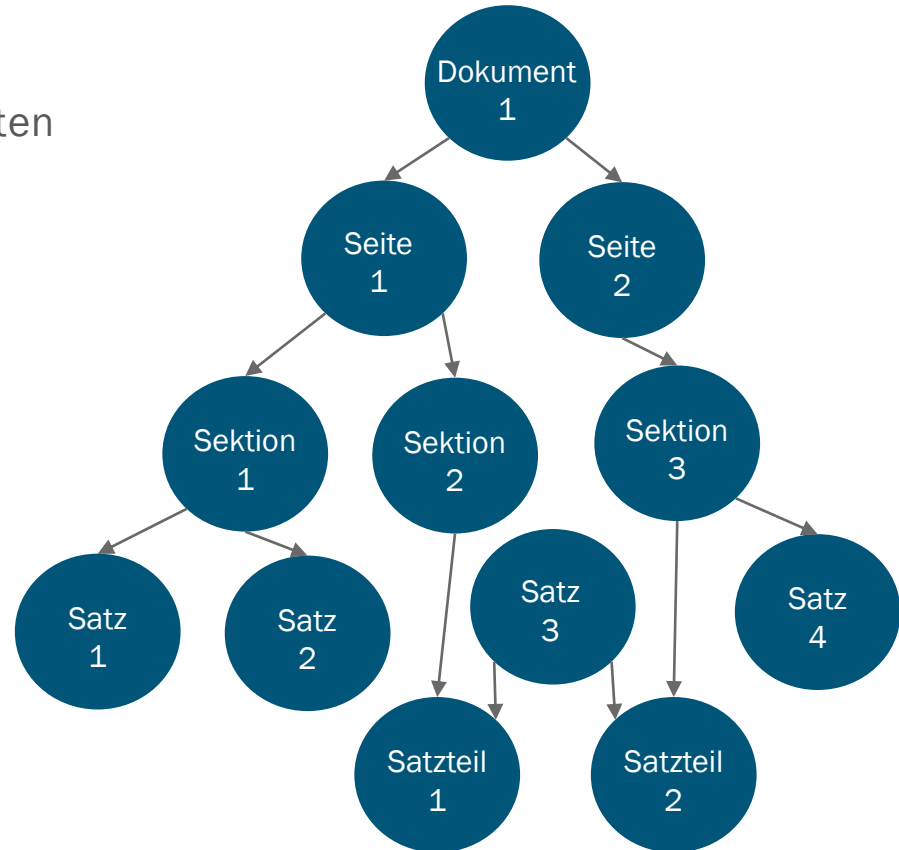


# Überleitung: Dokumente als Graphen

Rechts eine Graph-Repräsentation  
eines semantischen Netzes mit Knoten  
für

- Dokumente
- Seiten
- Sektionen
- Sätze
- Satzbestandteile

Pfeile repräsentieren hier  
simple „contains“-Beziehungen



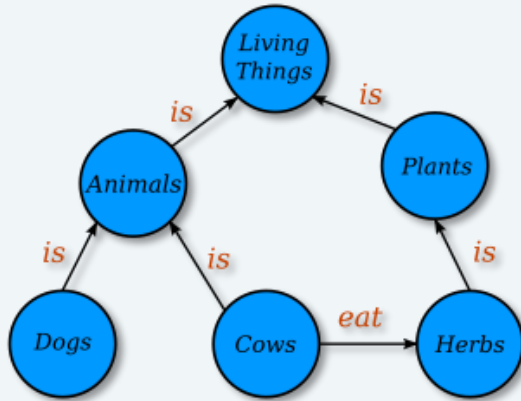
# Fokusthema: Wissensgraphen und Ontologien



- Prompt an Llama-3.1-70B-Instruct: ‚Erstelle die Beschreibung eines Cover-Bildes zur Veranstaltung "Wissensgraphen und Ontologien", die ich als Prompt für eine bildgenerierende KI verwenden kann.‘
- Visualisierung von FLUX.1 [dev]
- **Hinweis:** bei den folgenden Folien handelt es sich um Auszüge aus meinem jüngsten Vortrag zum Thema: *Fischer, A. (2024): Wissensgraphen. Präsentation auf der Sektionenkonferenz der DGS am 25.09.2024*

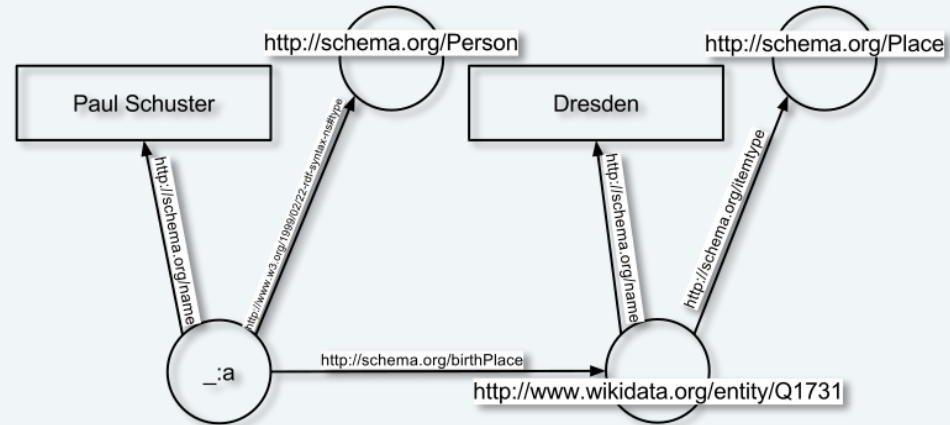


- Wissensgraphen (engl. Knowledge Graphs) gehen auf *Semantische Netze* (Quillian, 1960) sowie das *Semantic Web* (Berners Lee, 2001) zurück und stellen Wissen in Form von Graphen, d.h. über Objekte/Knoten und deren Verbindungen/Kanten, dar (z.B. Google, 2012).



Conceptual Diagram - Example

Quelle: [https://en.wikipedia.org/wiki/Knowledge\\_graph](https://en.wikipedia.org/wiki/Knowledge_graph)



An example of a simple RDF graph

Quelle: [https://de.wikipedia.org/wiki/Semantic\\_Web](https://de.wikipedia.org/wiki/Semantic_Web)



- **Fischer & Dörpinghaus (2024):** Analyse möglicher Bildungspfade auf Basis eines BERUFENET-Wissensgraphen
- **Ortmann, Bönke & Hammer (2023):** Bestimmung der Schnittmenge an jeweils verknüpften Kompetenzen als Ähnlichkeitsmaß für Berufe

- **GraphRAG-Chatbots:**  
Perspektivisch bieten sich Wissensgraphen auch für KI-gestützte Auswertungen und Beratungsangebote an (z.B. Chatbot zur Berufsorientierung/-beratung auf Basis von GraphRAG)

Auf Basis des BERUFENET-Wissensgraphen (vgl. Fischer & Dörpinghaus, 2024) wird aktuell GraphRAG (vgl. CoP-Sitzung 6) über die dynamische Generierung des folgenden System-Prompt implementiert – weitere Attribute lassen sich bedarfsgerecht einbinden. 

Du bist Berufsberater mit einem guten Überblick über alle Berufe in Deutschland.

Berücksichtige bei deiner Antwort auf die letzte Nachricht des aktuellen Dialogs folgende Informationen aus der Datenbank BERUFENET der Bundesagentur für Arbeit:

'KI-Engineer: KI-Engineers entwickeln komplexe intelligente Systeme, die Informationen verarbeiten, Muster erkennen und Entscheidungen treffen können. Dazu nutzen sie beispielsweise künstliche neuronale Netzwerke.'

Folgende Beziehungen zu verwandten Entitäten lassen sich der Datenbank entnehmen:

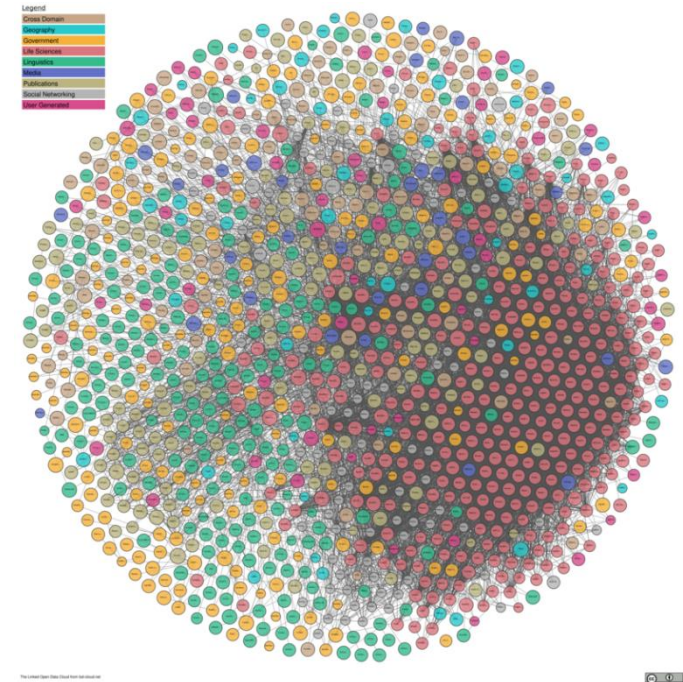
- 'Informatik (weiterführend)' qualifiziert für Tätigkeit als 'KI-Engineer'
- 'Künstliche Intelligenz (grundständig)' qualifiziert für Tätigkeit als 'KI-Engineer'
- 'Ingenieurinformatik, Computational Engineering (grundst.)' qualifiziert für Tätigkeit als 'KI-Engineer'
- 'Softwaretechnik (weiterführend)' qualifiziert für Tätigkeit als 'KI-Engineer'
- 'KI-Engineer' erfordert Kompetenz 'Qualitätsmanagement'
- 'KI-Engineer' erfordert Kompetenz 'Datenschutzrecht'
- 'KI-Engineer' erfordert Kompetenz 'Web-Applikationen (Entwicklung, Programmierung)'
- 'KI-Engineer' erfordert Kompetenz 'Kompetenzgruppe "Programmiersprachen, Programmbibliotheken, Entwicklungsumgebungen"'
- 'KI-Engineer' bietet Aufstiegsmöglichkeit zu 'Informatik (weiterführend)'
- 'KI-Engineer' erfordert Kompetenz 'KI-Plattformen (as a Service)'
- 'KI-Engineer' erfordert Kompetenz 'Informatik'
- 'KI-Engineer' bietet Aufstiegsmöglichkeit zu 'Kybernetik (weiterführend)'
- 'KI-Engineer' erfordert Kompetenz 'Expertensysteme, Künstliche Intelligenz'
- 'KI-Engineer' erfordert Kompetenz 'Machine Learning' (...)

Dr. Andreas Fischer | 09.08.2024 | 4

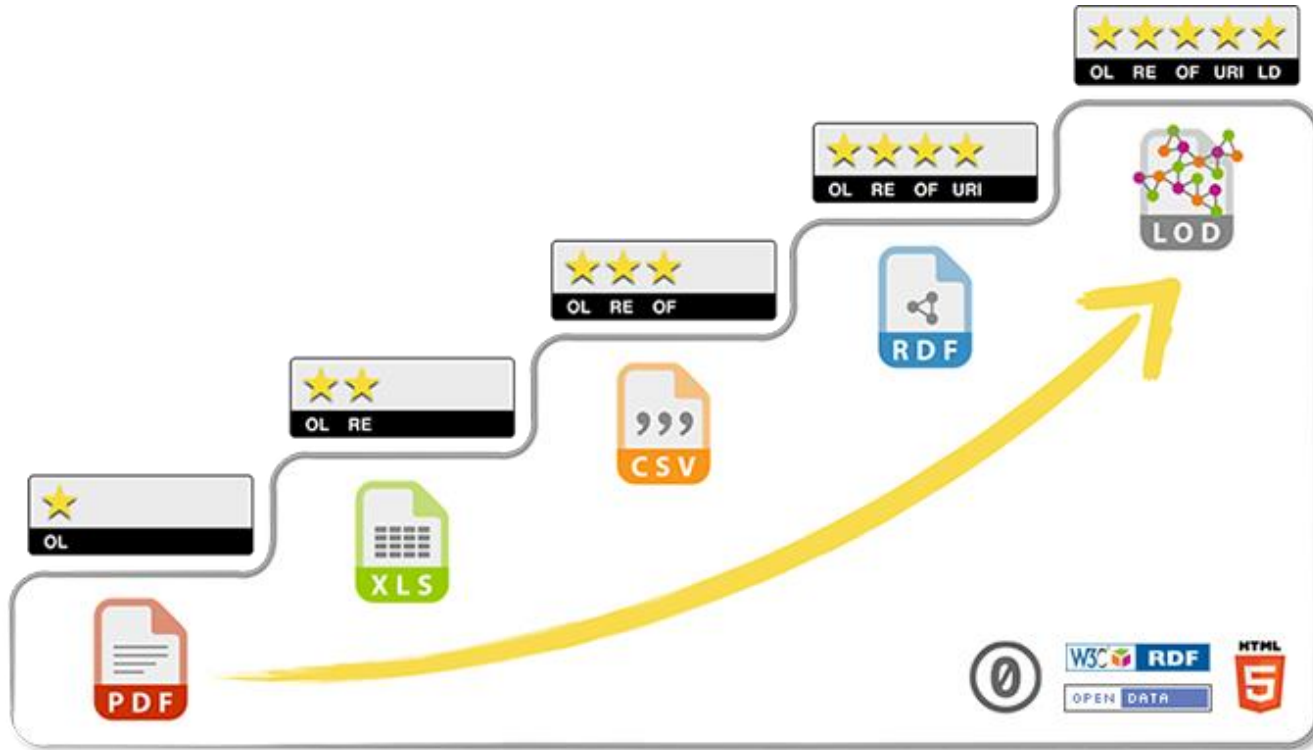
Vgl. Fischer & Dörpinghaus (2024) <https://www.mdpi.com/2673-9585/4/1/3>

# Linked Open Data (LOD) als Wissensgraph

- Linked Open Data Cloud (<https://lod-cloud.net/>)
- bildet als „Giant Global Graph“ (GGG) ein Netzwerk verknüpfter offener Datenbestände, z.B.
  - enzyklopädische Daten („DBpedia“, „Wikidata“, „WordNet“),
  - geografische Daten („OpenStreetMap“, „GeoNames“, „Linked GeoData“)
  - bibliografische Daten („DBLP“, „RDF Book MashUp“)
  - ...
- Oberkategorien: Cross Domain, Geography, Government, Life Sciences, Linguistics, Media, Publications, Social Networking, User Generated
- Vgl. für umfangreiche Graph-basierte Daten mit Fokus auf Statistiken auch Googles „Data Commons“ (<https://datacommons.org/>) seit 2018

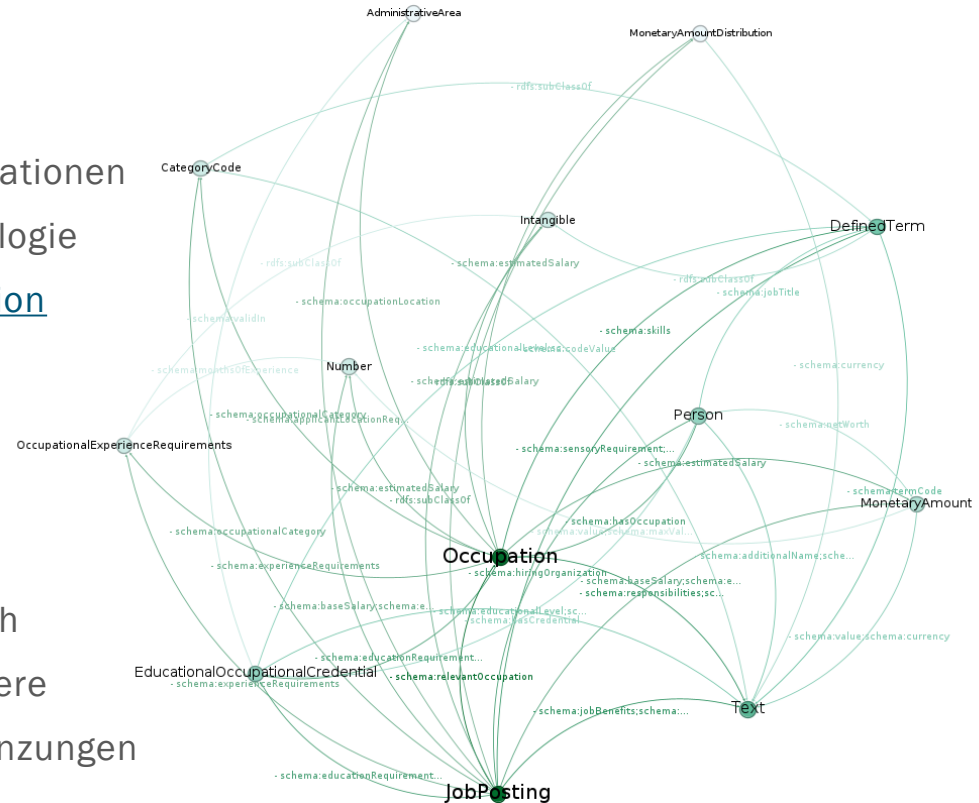


# LOD als 5-Sterne-Variante von Open Data



# „Occupation“ in schema.org

- Mehrwert bieten v.a. verbreitete Ontologien
- Rechts ein Auszug zentraler Konzepte und Relationen zum Konzept „Occupation“ auf Basis der Ontologie von schema.org: <https://schema.org/Occupation>
  - z.B. „rdfs:subClassOf“ Intangible
  - z.B. „schema:skills“ DefinedTerm
  - z.B. Person „schema:hasOccupation“
  - ...
- Beziehungen z.B. zu BERUFENET-Wissensgraph (Fischer & Dörpinghaus, 2024) erlauben bessere Analysen, Validierung und systematische Ergänzungen



- Generell empfiehlt sich die explizite Verknüpfung/Verknüpfbarkeit von Wissensgraphen mit einer formalen, expliziten und verbreiteten Ontologie (wie Schema.org, Basic Formal Ontology BFO, o.ä.) aus unterschiedlichen Gründen:
  1. **Deductive Reasoning:** Ontologie als Basis für Validierung und Erweiterung des Datenbestandes
  2. **Data Integration:** Ontologie als Brücke für die Integration von Daten aus unterschiedlichen Quellen
  3. **Semantic Interoperability:** Ontologie bietet einheitliches Vokabular, an welches multiple Systeme anknüpfen können