POTENZIALE DER ALGORITHMISCHEN NUTZUNG MEDIZINISCHEN WISSENS FÜR DIE VERBESSERUNG DER VERSORGUNG

Fraunhofer

SCAI

Prof. Dr. Martin Hofmann-Apitius
Leiter der Abteilung Bioinformatik
Fraunhofer Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI)

ZI Kongress Versorgungsforschung, Berlin, 7. & 8. September 2022

Fraunhofer Sankt Augustin & Universität Bonn

- Largest research centre for informatics and applied mathematics in Germany
- Approx. 700 employees, about 500 scientists, about 200 students and trainees

- Bonn-Aachen International Center for Information Technology (B-IT)
- International curriculum on Life Science Informatics (LSI)

Fraunhofer SCAI

Wissenschaftliche Problemlösungs-Kompetenz

1. Unstructured information mining
2. Knowledge-graph technologies
3. Models combining data and knowledge
4. Actionable insights and decision support

Development (DEV)

Semantics (SEM)

AI & Data Science (AI-DAS)

Dr. Marc Jacobs

Dr. Alpha Tom Kodamullil

Prof. Dr. Holger Fröhlich

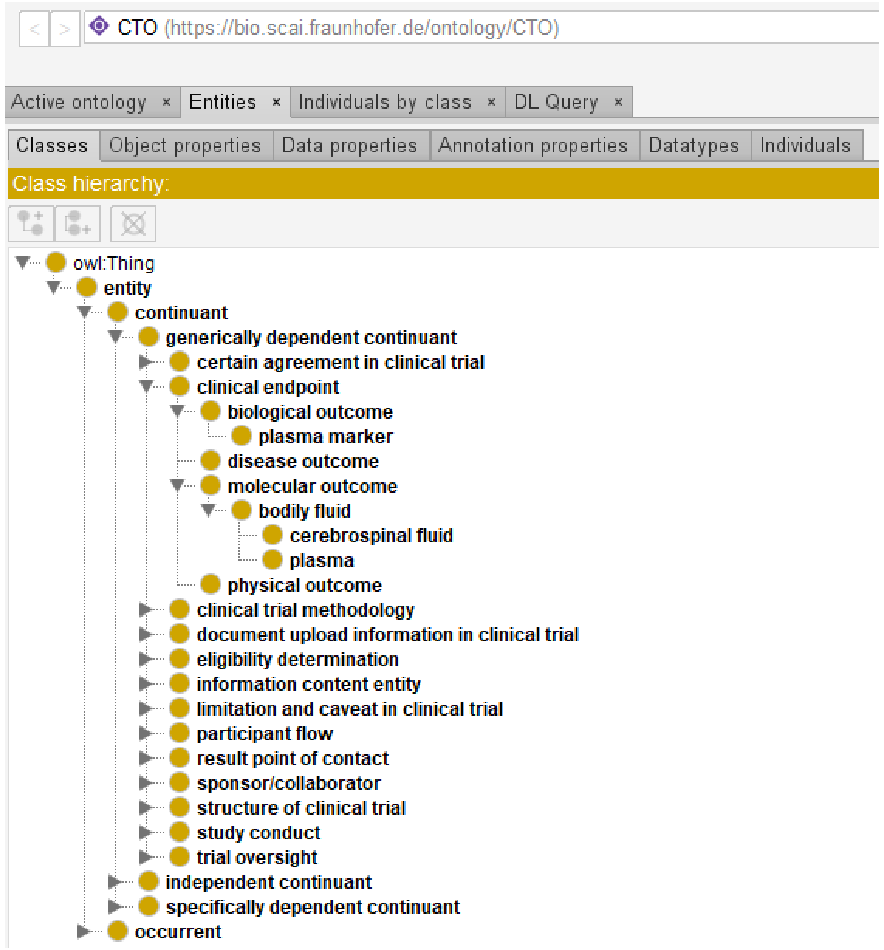
INFORMATIONSEXTRAKTION

From shared semantics to computable knowledge

We are here ....

Formalisierte Semantik für Studiendaten: CTO

Clinical Trial Ontology (CTO)



CTO: a Community-Based Clinical Trial Ontology and its Applications in PubChemRDF and SCAIView

Asiyah Yu LIN[*a,e1], Stephan GEBEL[*b], Qingliang Leon LI[†c], Sumit MADAN[†b], Johannes DARMS[b], Evan BOLTON[c], Martin HOFMANN-APITIUS[b], Yongqun Oliver HE[§d2], Alpha Tom KODAMULLIL[§e3]

[a] *Center for Devices and Radiological Health, FDA, Silver Spring, MD, USA*
[b] *Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, Sankt Augustin, Germany*
[c] *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA*
[d] *University of Michigan Medical School, Ann Arbor, Michigan, USA*
[e] *National Center for Ontological Research, Buffalo, New York, USA*

Download: http://purl.obolibrary.org/obo/cto.owl

Ontologie – Kooperation von SCAI, FDA, NCBI and U. Michiga

COMPUTEBARES WISSEN: VON UNSTRUKTURIERTEN TEXTEN ZU ALGORITHMISCH NUTZBAREN WISSENSGRAPHEN

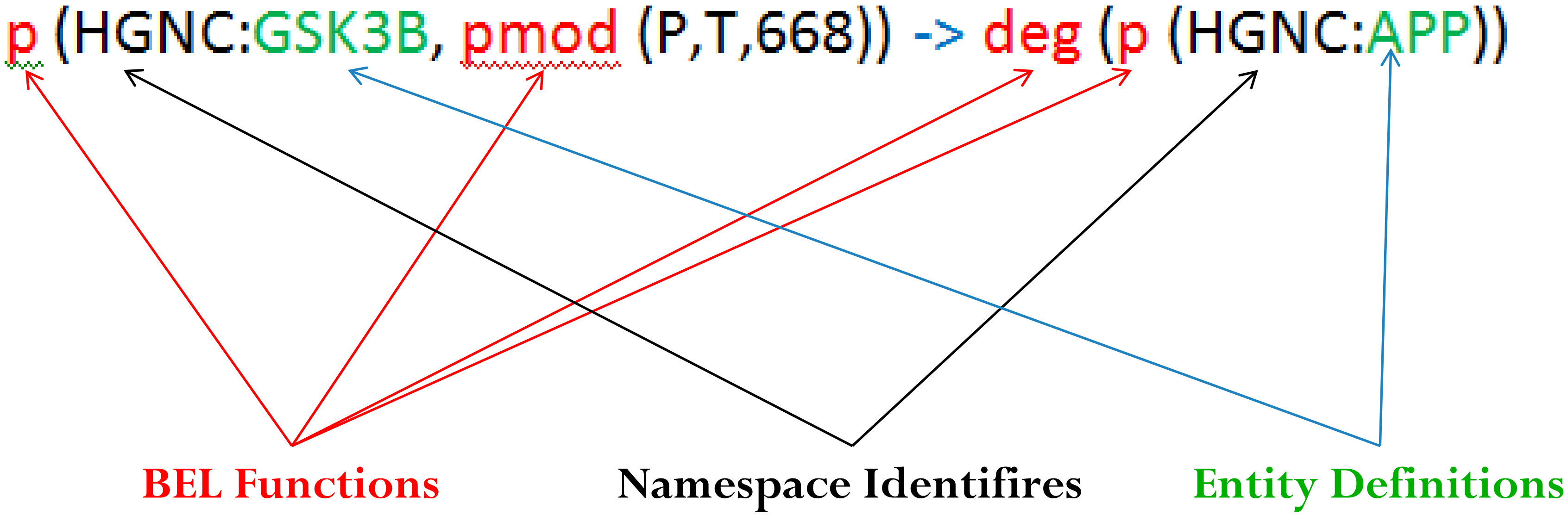
From information extraction to re-usable knowledge graphs

we are here ....

EXTRAKTION VON KAUSALEN UND KORRELATIVEN ZUSAMMENHÄNGEN AUS DER WISSENSCHAFTLICHEN LITERATUR

OpenBEL: Capturing of Knowledge and “Encoding” of Data

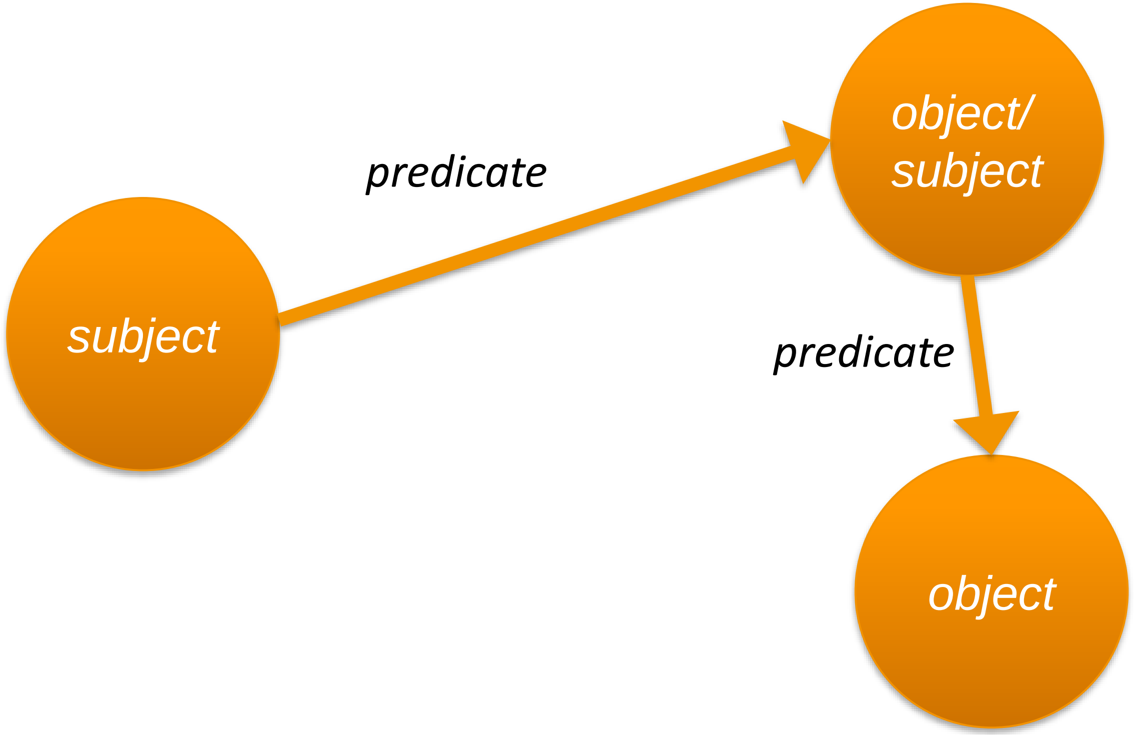
“**Phosphorylation** of **glycogen synthase kinase 3beta** at **Threonine, 668** **increases** the **degradation** of **Amyloid precursor protein**.”



Fraunhofer SCAI

TRIPLES AGGREGIEREN ZU GRAPHEN

Das OBJEKT eines Triples kann das SUBJEKT des nächsten Triples sein

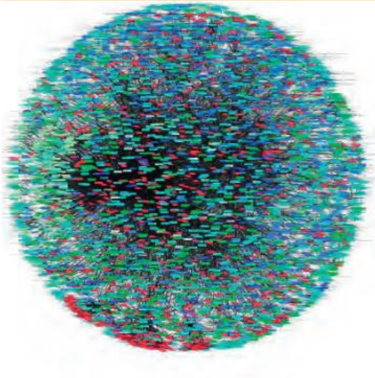


- Subject-predicate-object “triples”
- Object of one triple can be subject of another
- Putting them together makes arbitrarily large knowledge graphs
- Reasoning over causal relationships becomes a graph traversal

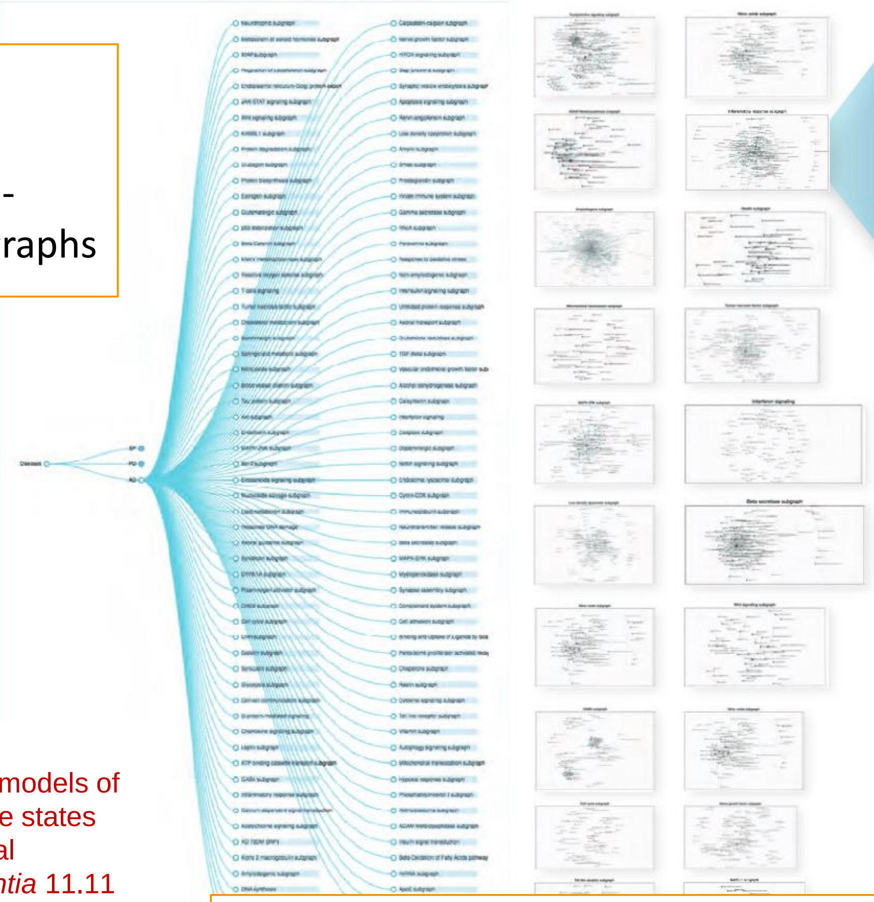
Fraunhofer SCAI

Multimodal Mechanism Models in Neurodegeneration Research

Text Mining extracts cause-and-effect relationships, disease-specific triples form graphs



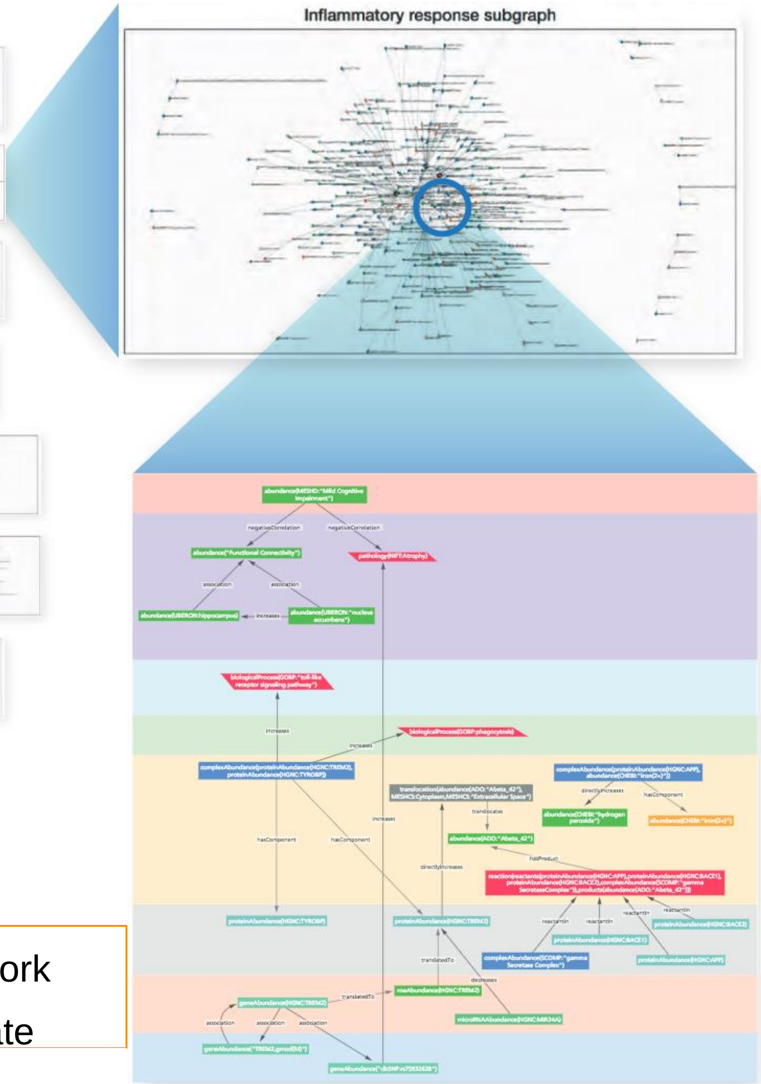
Kodamullil, Alpha Tom, et al. "Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis." *Alzheimer's & Dementia* 11.11 (2015): 1329-1339.



Classifying each relation in the network into the mechanism(s) they participate

Domingo-Fernández, Daniel, et al.

"Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment." Bioinformatics 33.22 (2017): 3679-3681.



Neuroinflammation mechanistic subgraph representing multiple biological levels

Clinical level

(e.g., cognition tests, imaging)

Biological processes

Molecular level (e.g., proteins, chemicals)

Genetic level (e.g., genes, epigenetics, variations)

Cause-and-Effect Models & Applications

- **Identifying early mechanisms in AD aetiology**
  - Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M. *Alzheimers Dement. 2015*
  - Kodamullil AT, Zekri F, Sood M, Hengerer B, Canard L, McHale D, Hofmann-Apitius M. *Nat Rev Drug Discov. 2017*
- **Functional impact of genetic and epigenetic variants**
  - Karki R, Kodamullil AT, … Hofmann-Apitius M. *BMC Bioinformatics. 2019*
  - Naz M, Kodamullil AT, Hofmann-Apitius M. *Brief Bioinformatics. 2016*
- **Analysis of comorbidity**
  - Khanam Irin A, Kodamullil AT, Gündel M, Hofmann-Apitius M. *J Immunol Res. 2015*
- **Repurposing of Drugs**
  - Karki R, Kodamullil AT, Hofmann-Apitius M. *J Alzheimers Dis. 2017*
  - Karki R,… Kodamullil AT, Hofmann-Apitius M. *J Alzheimers Dis. 2020*
- **Comparitive analysis of pathophysiology mechanisms mouse/human**
  - Emon MA, Kodamullil AT, Karki R, Younesi E, Hofmann-Apitius M. *J Alzheimers Dis. 2017*
  - Kodamullil AT, Iyappan A, Karki R, Madan S, Younesi E, Hofmann-Apitius M. *J Alzheimers Dis. 2017*

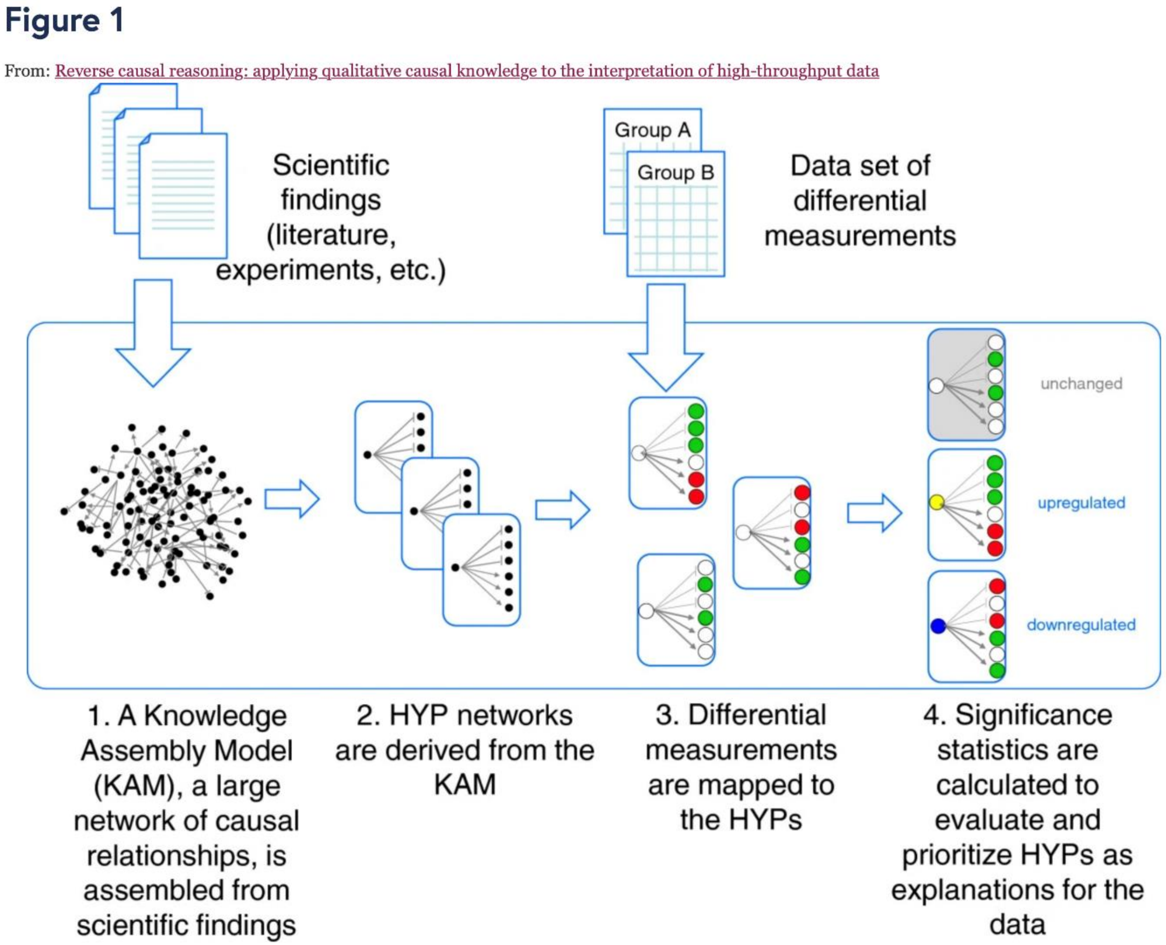
- Alzheimer Disease
- Parkinson Disease
- Epilepsies
- Amyotrophic lateral sclerosis
- Type 2 diabetes
- Schizophrenia
- Bipolar Disorder
- Traumatic Brain Injury
- PTSD
- [COVID-19]

ALGORITHMEN FÜR KNOWLEDGE GRAPHS

From reverse causal reasoning to guilty targets

we are still here ....

REVERSE CAUSAL REASONING:



- Wissensgraph „kodiert“ kausale und korrelative Beziehungen
- Subgraphen repräsentieren Hypothesen zu Mechanismen
- Subgraphen werden auf die Übereinstimmung mit Daten getestet
- Die Signifikanz der Übereinstimmung zwischen Subgraphen und Daten wird berechnet
- Die am besten zu den Daten passenden Subgraphen werden selektioniert

Catlett, N.L., Bargnesi, A.J., Ungerer, S. *et al.* Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. *BMC Bioinformatics* **14**, 340 (2013). https://doi.org/10.1186/1471-2105-14-340

Patient Embeddings in Knowledge Graphs

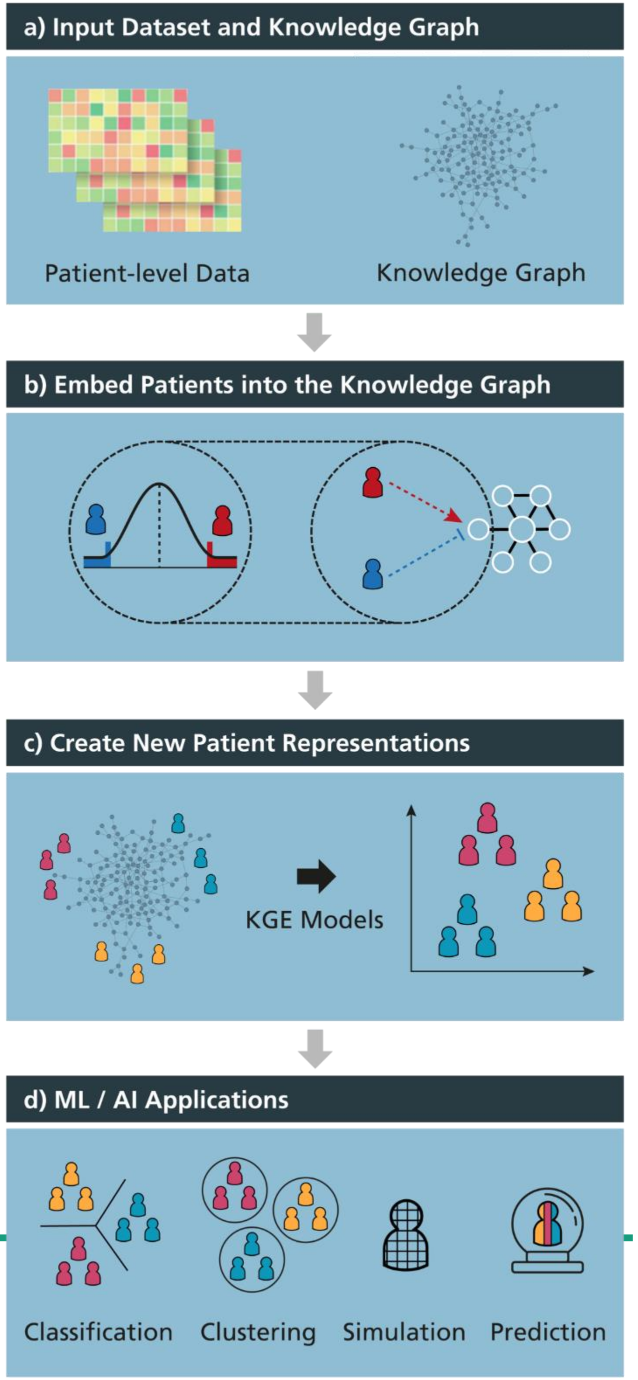
Schematic illustration of the CLEP framework.

(a) CLEP requires two inputs: (i) a patient-level dataset such as multi-omics, and (ii) a KG comprising relations between features measured in the previously mentioned dataset.

(b) Using one of the proposed methods, CLEP incorporates patients into the KG by connecting them to their most distinctive features in the dataset.

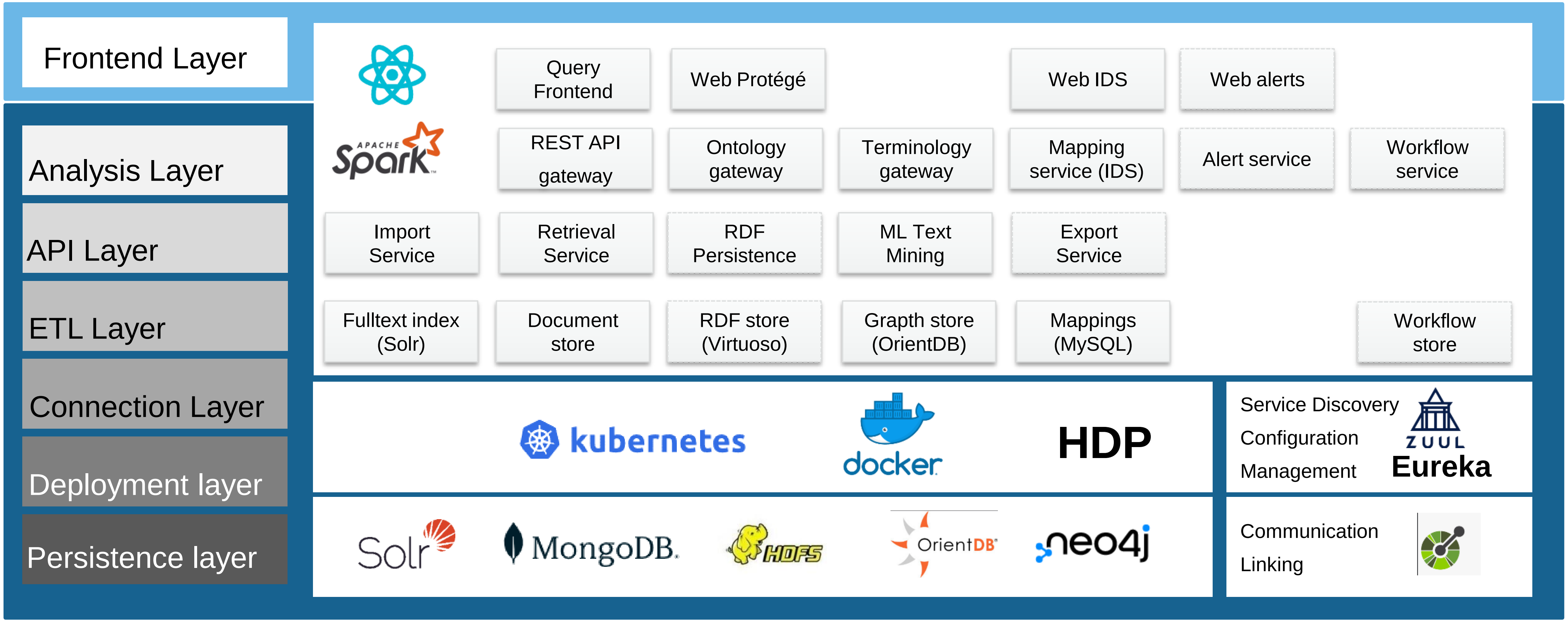
(c) KGEMs are then used to generate new patient representations based on both data- and knowledge-driven features.

(d) These patient representations can subsequently be used for several downstream tasks, such as patient classification and stratification.



Srinivas Bharadhwaj, V., Ali, M., Birkenbihl, C., Mubeen, S., Lehmann, J., Hofmann-Apitius, M., ... & Domingo-Fernández, D. (2021). CLEP: A Hybrid Data-and Knowledge-Driven Framework for Generating Patient Representations. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btab340

Aus dem Maschinenraum: Knowledge Graph Service Infrastruktur



NUTZUNG VON WISSENSGRAPHEN: KG-EMBEDDINGS

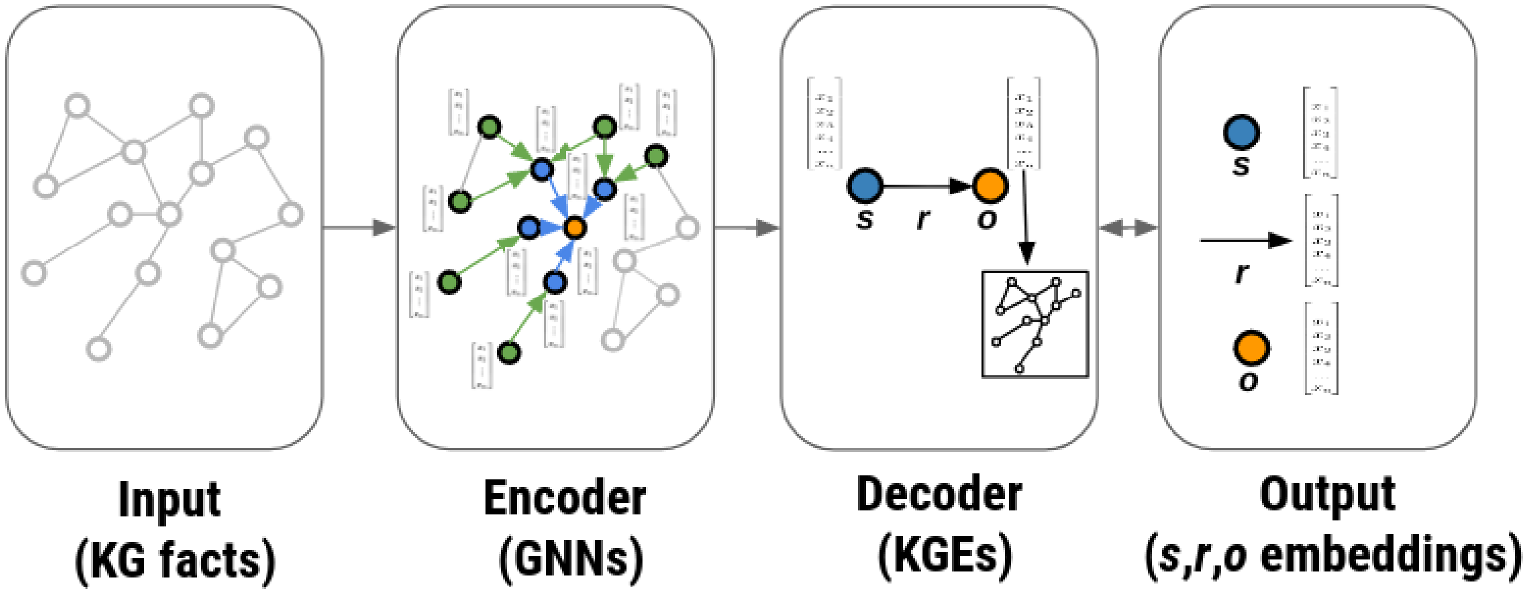
From Understanding Pathophysiology Mechanisms to the Identification of Patient Subgroups

we are here ....

Fraunhofer SCAI

KNOWLEDGE GRAPH EMBEDDINGS

Reduktion der Dimensionalität von komplexen Graphen

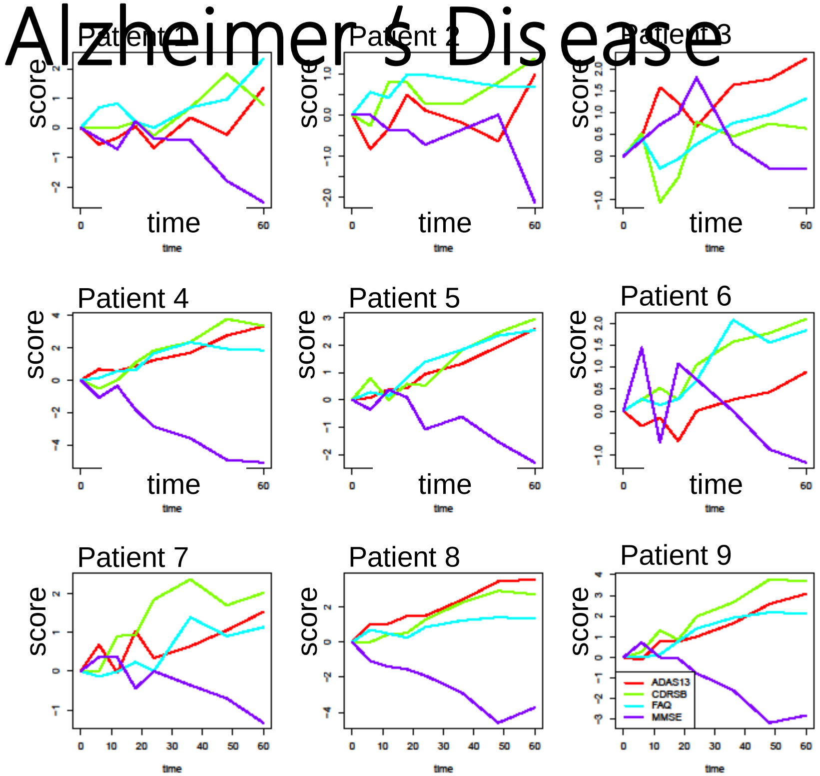


- Wissensgraph wird in einen Vektor umgewandelt
- Die gängigen Algorithmen des „Data Mining“ können sehr gut mit Vektorräumen umgehen.
- Hierdurch wird er „leichter und besser nutzbar“.

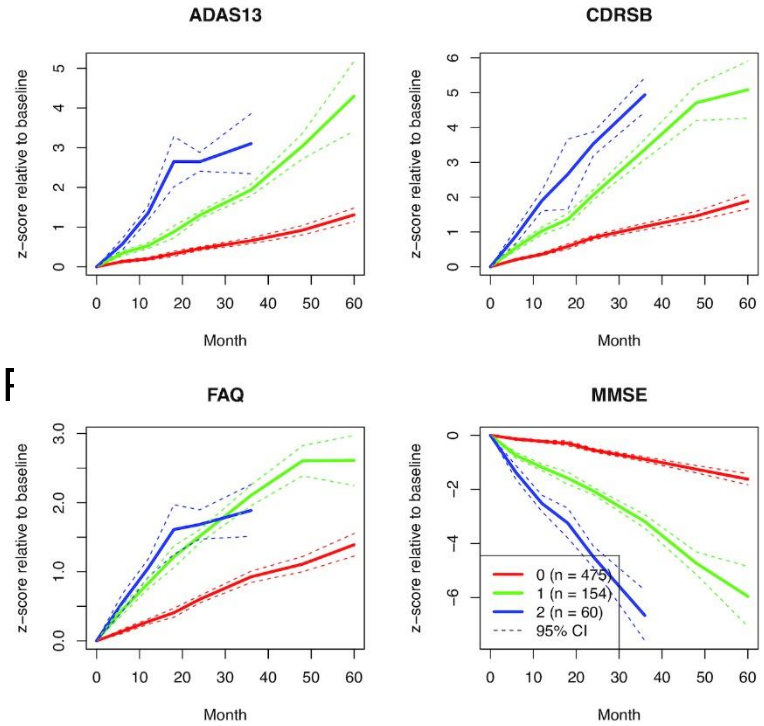
Referenz: https://towardsdatascience.com/improving-knowledge-graph-embeddings-with-graph-neural-networks-8ffb14804199

Fraunhofer SCAI

Stratification of Disease Trajectories in Parkinson's and Alzheimer's Disease



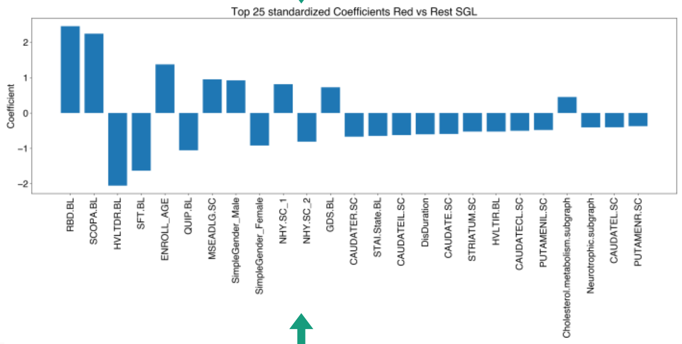
VADER

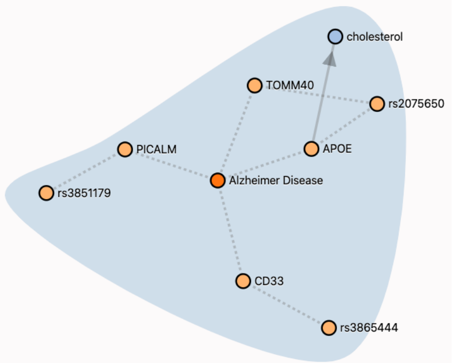


Baseline

- Biomarkers
- Clinical
- Genetics
- Imaging

AI





Predictive molecular mechanisms

Impact: Clinical trial

Johann de Jong, Mohammad Asif Emon, Ping Wu, Reagon Karki, Meemansa Sood, Patrice Godard, Ashar Ahmad, Henri Vrooman, Martin Hofmann-Apitius, Holger Fröhlich, Deep learning for clustering of multivariate clinical patient trajectories with missing values, *GigaScience*, Volume 8, Issue 11, November 2019, giz134, https://doi.org/10.1093/gigascience/giz13

VirtualBrainCloud
Colin Birkenbihl
Fraunhofer SCAI

PERSPEKTIVEN FÜR DIE ZUKUNFT: GUIDELINES, EVIDENZEN AUS DER LITERATUR, REAL-WORLD DATA

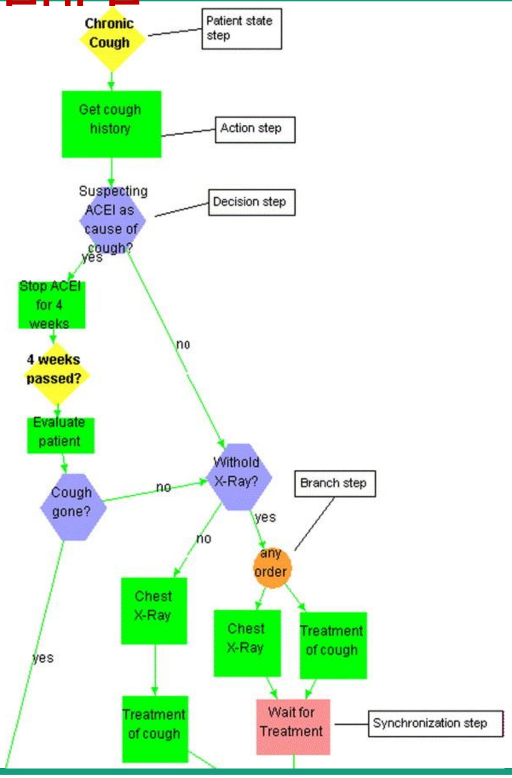
From Clinical Guidelines to Graph-based Analytics & Decision-Support

we are here ....

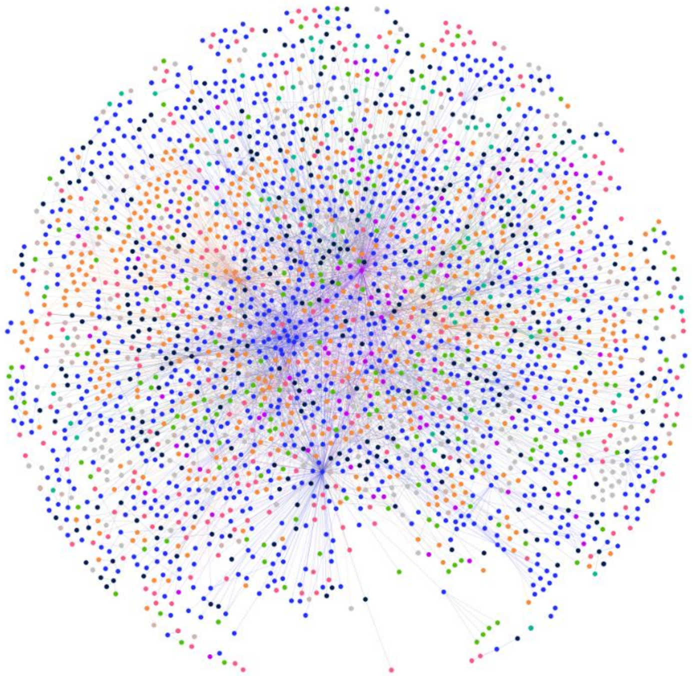
Fraunhofer SCAI

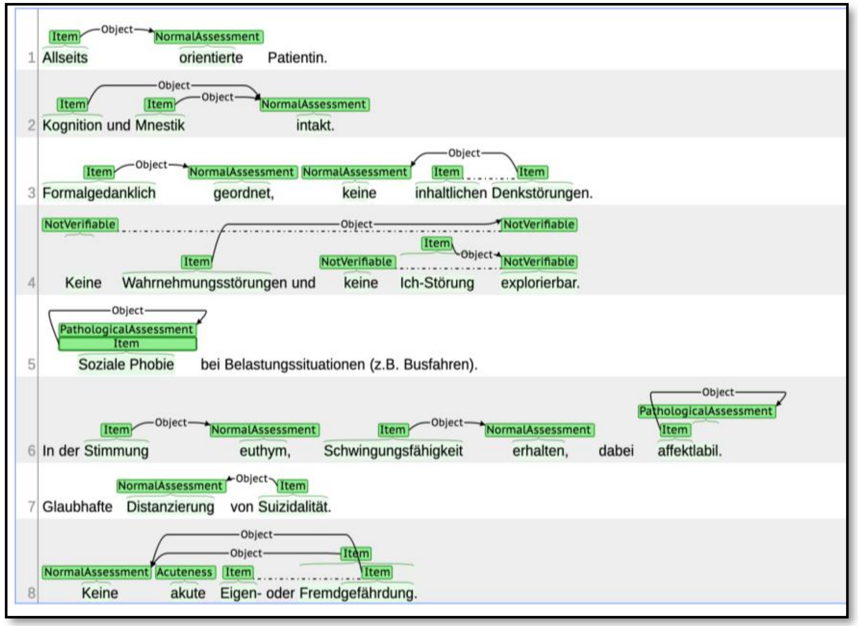
EVIDENZEN AUS DREI WELTEN

Clinical Guidelines – Scientific Studies & Publications – Real-World-Evidence



https://ars.els-cdn.com/content/image/1-s2.0-S1532046404000334-gr2.jpg

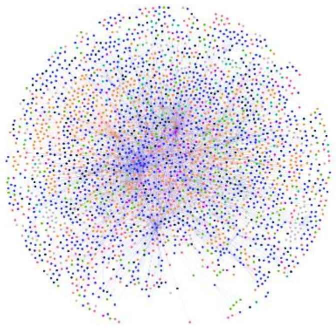




KNOWLEDGE GRAPH EMBEDDINGS FÜR DIE VERSORGUNG (S-FORSCHUNG)

Informationsextraktion, Wissensgraphen und AI-Verfahren für die Versorgungsforschung

Literatur-Wissensgraph

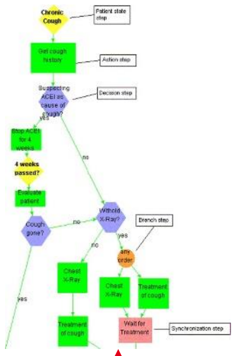


Aktualisierung

Erweiterungen

(co-Morbiditäten, Nebenbedingungen, neue Erkenntnisse / Evidenzen, neue Subgruppen)

Leitlinien-Graph

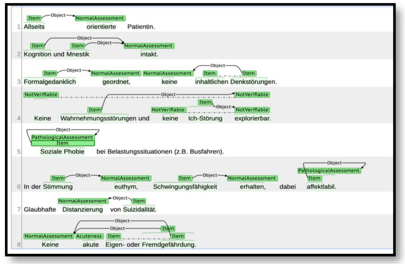


Validierung

Anpassung

(Konkordanz zwischen Leitlinie und realen Behandlungspfaden)

Real-World Evidenzen



Fraunhofer SCAI

DIE WICHTIGEN BOTSCHAFTEN DIESES VORTRAGS

- WissensGraphen sind super geeignet, um Wissen algorithmisch zugänglich zu machen
- Text-2-Graph Verfahren sind inzwischen sehr weit entwickelt und werden ständig besser
- Patienten-Daten und WissensGraphen können gemeinsam genutzt werden, um Patienten-Strata zu identifizieren
- Leitlinien können als Graph-Modelle repräsentiert werden
- Es gibt kein prinzipielles Hindernis für das “Maschinelle Erlernen“ spezifischer, aktueller Leitlinien-Graphen
- Wie bei allen AI-Verfahren ist die Verfügbarkeit von Daten das Haupthindernis
- Die großer Herausforderung für eine durch AI erheblich verbesserte Versorgungsforschung ist die Verfügbarkeit großer Mengen an medizinischen Routine(Versorgungs-)daten.
- Ein selbstständig lernfähiges System, welches Leitlinien intelligent erweitert und anpasst, ist denkbar

Fraunhofer SCAI

THE TEAM @ FRAUNHOFER SCAI

My team with currently 18 nationalities / 3 groups / 12 scientists / 13 PhD students

<u>Information Extraction</u>

Marc Jacobs
Jürgen Klein
Tim Adams
Negin Babaiha
Sumit Madan
Johannes Darms

<u>Knowledge Graphs</u>

Alpha Tom Kodamullil
Vanessa Lage-Rupprecht
Stephan Gebel
Christian Ebeling
Bruce Schultz
Daniel Domingo-Fernandez
Sepehr Golritz-Khatami
Yojana Gadiyja
Andrej Konotopez

<u>Data Science & AI</u>

Holger Fröhlich
Sophie Krix
Manuel Lentzen
Tamara Raschka

Fraunhofer SCAI