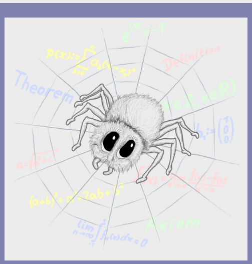
mArachna



Eine OWL-basierte mathematische Wissensbasis

Nicole Natho natho@math.tu-berlin.de
Sabina Jeschke sabina@math.tu-berlin.de
Ruedi Seiler seiler@math.tu-berlin.de
Marc Wilke marc@math.tu-berlin.de

Technische Universität Berlin
Institut der Mathematik

Konzept

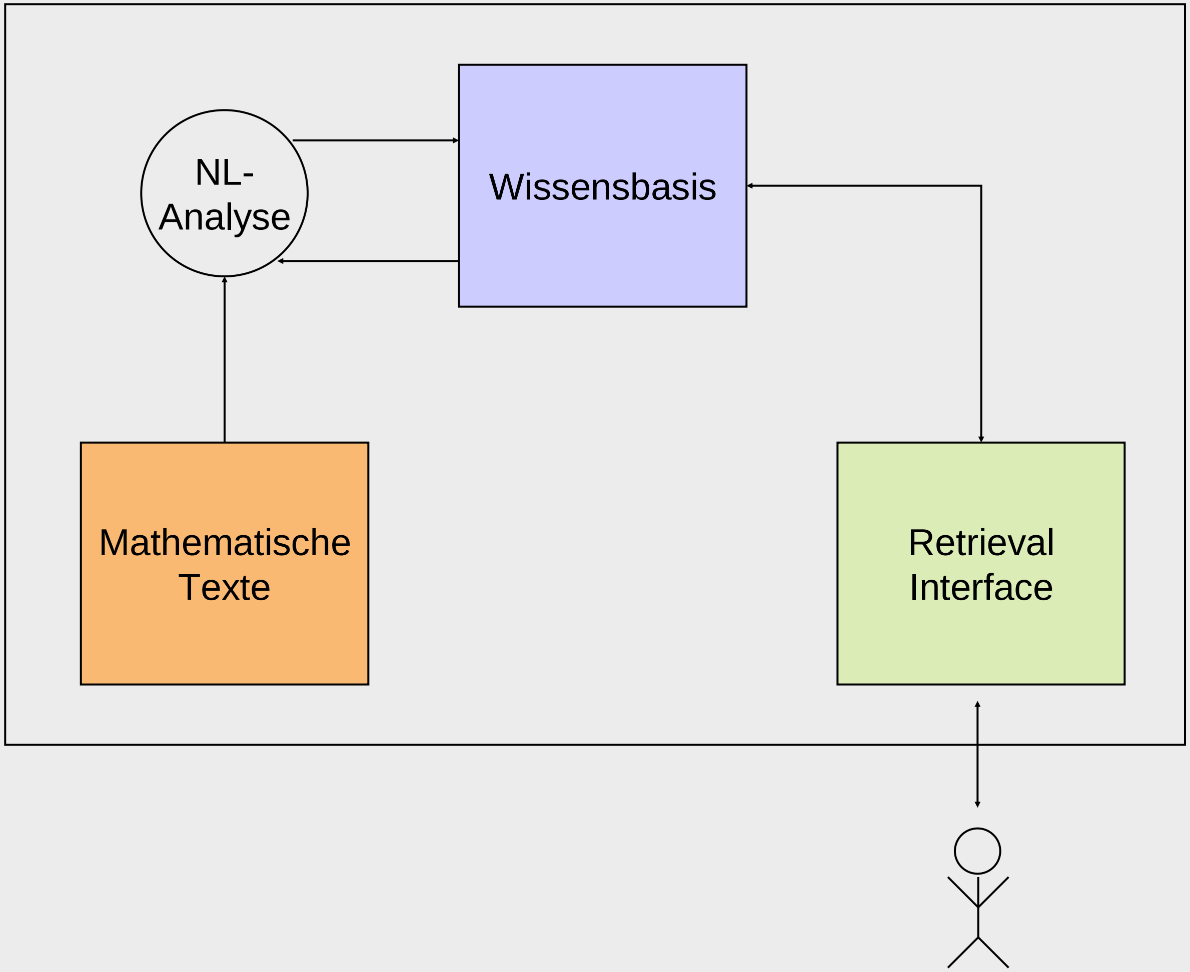
Ziel: Entwicklung einer intelligenten mathematischen Enzyklopädie

Ansatz: Mathematische Texte zu Kursen für Ingenieursstudenten im Grund- oder Bachelorstudium

Aufgabe: Semantische Annotation der mathematischen Texte

Lösung: Automatische Indexierung für Information Retrieval Systeme

Konzept



NL-Analyse
Wissensbasis
Mathematische Texte
Retrieval Interface

Mathematische Sprache
Einfaches Beispiel

Definition (Gruppe):

Sei M eine Menge, $M \neq \emptyset$, und

$$\begin{aligned} * : M \times M &\longrightarrow M \\ \langle a,b \rangle &\mapsto a*b \end{aligned}$$

eine Abbildung.

Das Tupel $\langle M,* \rangle$ heißt <u>Gruppe</u>, wenn gilt:

1. Die Abbildung ist assoziativ: $(a * b) * c = a * (b * c)$ $(a,b,c \in M)$.
2. Für jedes $a,b \in M$ existiert ein $x \in M$ mit $a * x = b$.
3. Für jedes $a,b \in M$ existiert ein $y \in M$ mit $y * b = b$.

aus: Wüst, „Höhere Mathematik für Physiker“

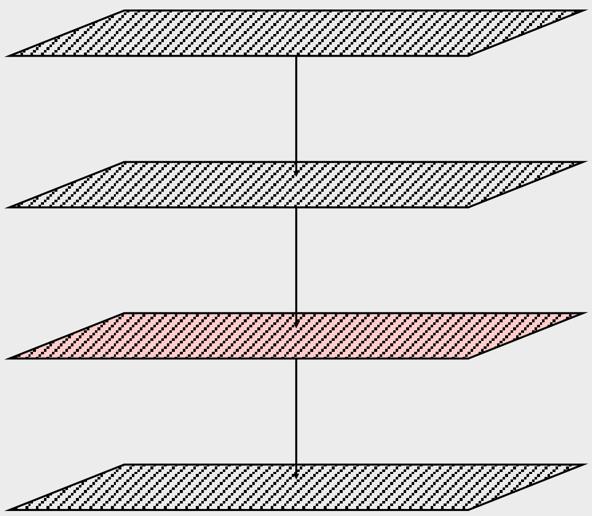
Linguistische Strukturen

Entitätenebene

Binnenstrukturebene

Satzebene

Wort- & Symbolebene



Fokussierung auf Entitäten:
Definitionen, Theoreme, Sätze

Typische Sprachkonstrukte
abhängig vom Entitätentyp

Sei M eine Menge, $M \neq \emptyset$, und $* : M \times M \longrightarrow M$, $\langle a,b \rangle \mapsto a*b$
eine Abbildung. Das Tupel $\langle M,* \rangle$ heißt Gruppe, wenn gilt:
Die Abbildung ist assoziativ: $(a * b) * c = a * (b * c)$ $(a,b,c \in M)$.
...

Linguistische Analyse

⟨M,∗⟩ heißt Gruppe.

↓

TRALE-System

Head-Driven Phrase Structure Grammar

↓

Semantische Analyse

regelbasiert

↓

Resultat: (⟨M,∗⟩, heißt, Gruppe)

Wissensbasis
Konzept

Resultat der semantischen Analyse

Annotierte Tripel/Statements

(⟨M,$*$⟩, heißt, Gruppe)

⇓

Wissensbasis

- Graphenstruktur realisiert mit OWL Full
- OWL Class/Individual: math. Begriffe und Konzepte
- OWL Properties: math. Relationen
- Annotationen: zusätzliche Informationen zum Text

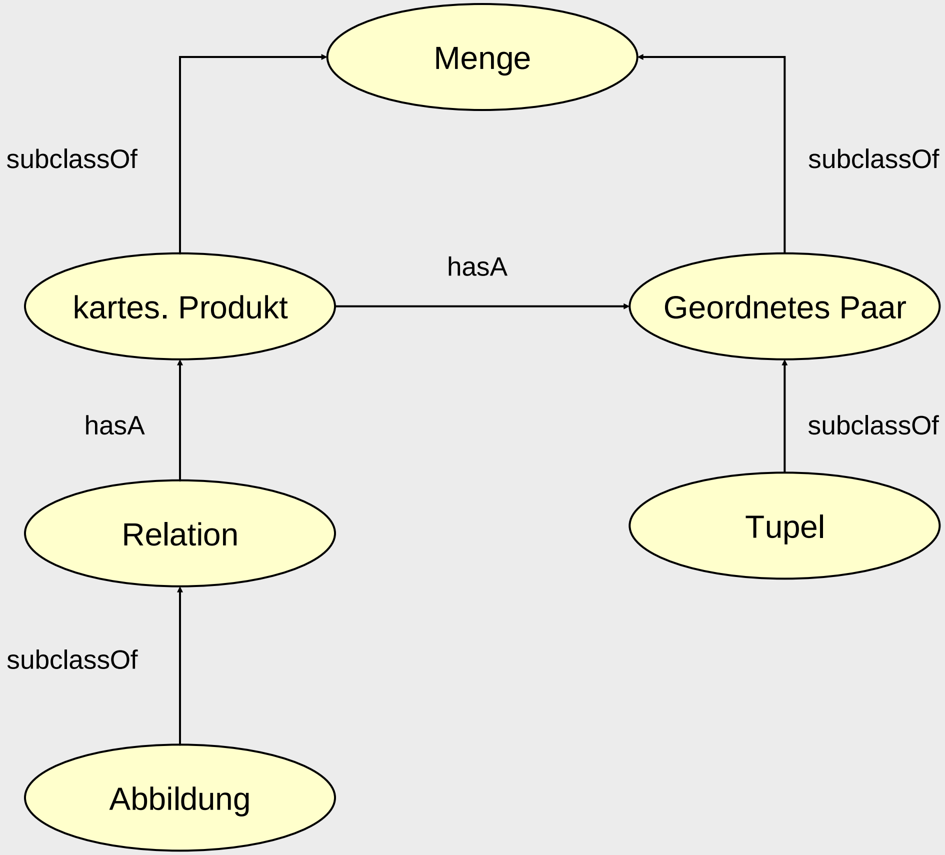
⇓

Resultat

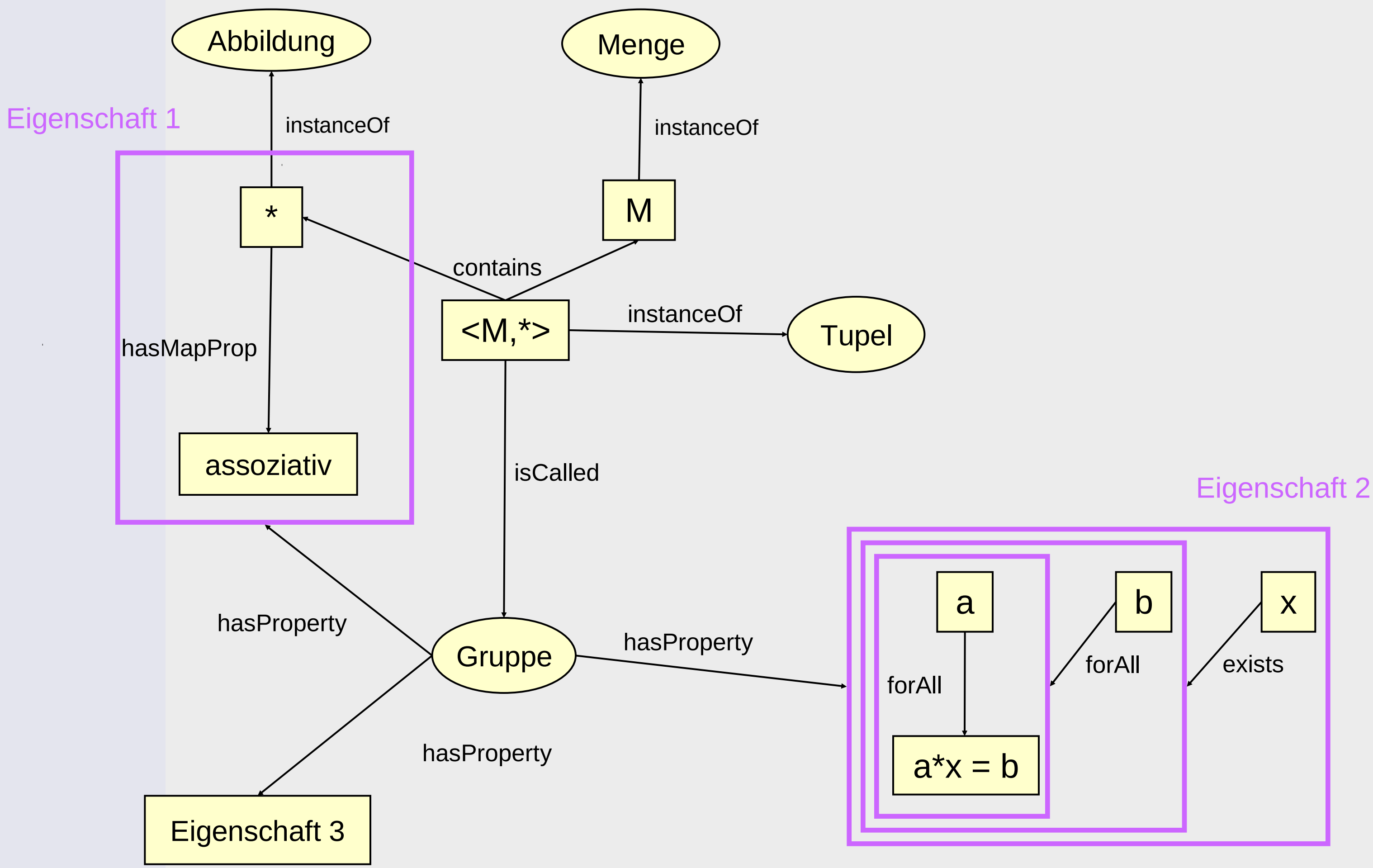
Ontologie des mathematischen Textes

Wissensbasis
Grundlegende Ontologie

Grundlegende Struktur:
Mengenlehre und mathematische Logik



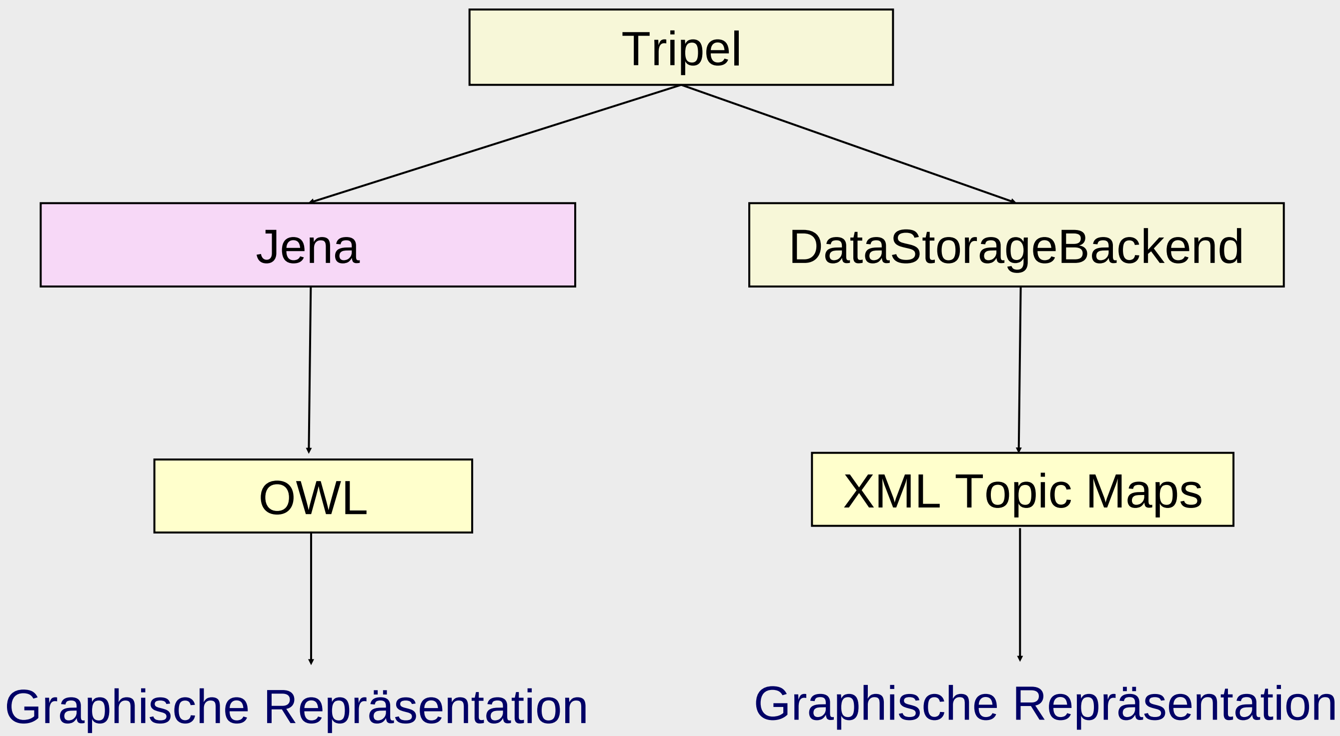
Integration von neuem Wissen



Wissensbasis
Beispiel

```
<owl:Class rdf:ID=“Eigenschaft“>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource=“#hasSubject“ />
         <owl:allValuesFrom rdf:resource=“#MathThing“ />
         <owl:cardinality rdf:datatype=“&xsd;nonNegativeInteger“>1</owl:cardinality>
      </owl:Restriction>
   </rdfs:subClassOf>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource=“#hasProperty“ />
          …
 </owl:Class>
…
<owl:Thing rdf:ID=“Eigenschaft001“ … />
…
<owl:Class rdf:ID=“Gruppe“>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource=“#hasProperty“ />
         <owl:hasValue rdf:resource=“#Eigenschaft001“ />
      </owl:Restriction>
   </rdfs:subClassOf>
…
</owl:Class>
```

Verarbeitung und Visualisierung



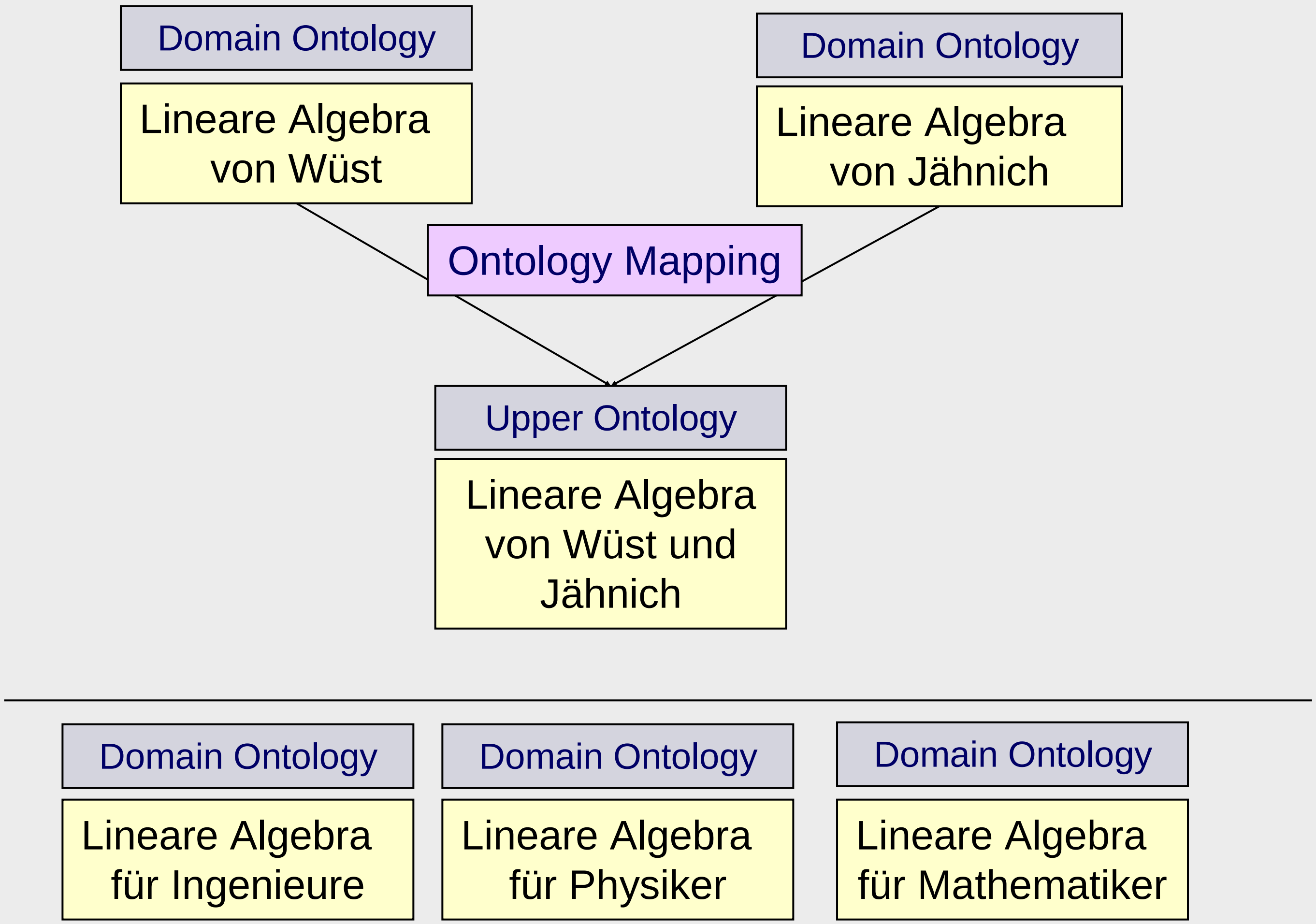
Problem

Problematik bei Erzeugung der Wissensbasis

- undefinierte Begriffe
- fehlende Relationen
- identische Tripel
- widersprüchliche Tripel

Semi-automatischer Ansatz

Mergen von Wissensbasen



Mergen von Wissensbasen

Definition (Gruppe):

Sei M eine Menge …

Das Tupel $\langle M,* \rangle$ heißt *Gruppe,* wenn gilt:

1. Die Abbildung ist assoziativ: $(a * b) * c = a * (b * c)$ $(a,b,c \in M)$.
2. Für jedes $a,b \in M$ existiert ein $x \in M$ mit $a * x = b$.
3. Für jedes $a,b \in M$ existiert ein $y \in M$ mit $y * b = b$.

Alternative Definition!

Definition (Gruppe):

Sei M eine Menge …

Das Tupel $\langle M,* \rangle$ heißt *Gruppe*, wenn gilt:

1. $(a * b) * c = a * (b * c)$ $(a, b, c \in M)$
2. Es gibt ein $e \in M$ mit
   i. $a * e = e * a = a$,
   ii. Es existiert ein $\underline{a} \in M$, so dass $\underline{a} * a = a * \underline{a} = e$ für alle $a \in M$.

Mergen von Wissensbasen

Vorgehen auf Entitätenebene:

1. Identifizierung von äquivalenten Entitäten
2. Weiterverarbeitung von nicht-äquivalenten Entitäten

Vorgehen auf Satzebene:

1. Identifizierung und Zusammenfassung von identischen Classes/Individuals und Properties
2. Weiterverarbeitung der verbliebenen Classes/ Individuals und Properties

Erhoffte Vorteile der mathematischen Wissensbasis:

- identische Begriffe
- identische Relationen
- weniger Reasoning

Ergebnisse und Ausblick

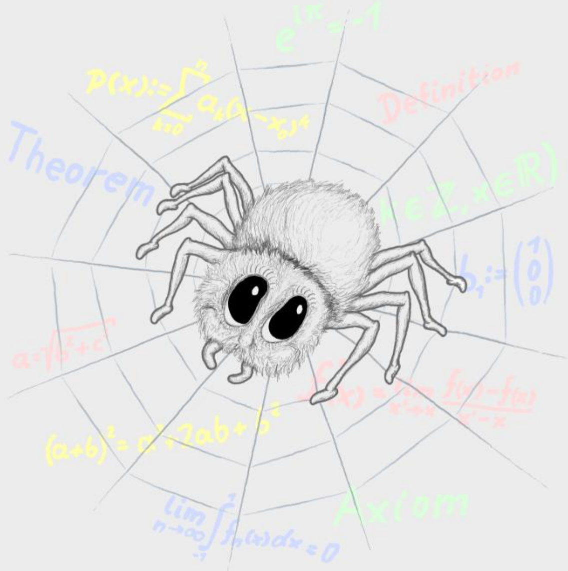
Ergebnisse:

1. Analyse von einfachen mathematischen Texten im Deutschen
2. Erzeugung und Visualisierung von Wissensbasen aus den Texten

Ausblick:

1. Semantische Analyse von mathematischen Formeln im Bezug zum Text
2. Erzeugung und Visualiserung eines vollständigen Kurses der Linearen Algebra
3. Mergen von verschiedenen Kursen zur Linearen Algebra

Ende



Danke für ihre Aufmerksamkeit!

natho@math.tu-berlin.de