

Gründe für das Gespräch:

- weiter an den Themenfeld forschen und darin promovieren.

Was hast du bisher gemacht?

Seit oder bevor der Gründung von der Gruppe FE (3 Jahren) beschäftige ich mich mit Themen rund um Knowledgegraphen.

Damals war die Idee verteilte Informationen zu sammeln um mit deren Hilfe wirtschaftliche Trends zu identifizieren. Bei einem Trend handelt es sich ja um eine Abweichung von einem gegebenen Zustand über die Zeit. Nun wissen wir aber nicht komplett was einen Zustand ausreichend beschreibt und von welchen Faktoren dieser abhängig ist. Wir wissen zum Beispiel, das ein wichtiger Indikator für die Entwicklung eines Marktes, die Anzahl an Anbieter und Nachfrager ist. Zusätzlich wissen wir das häufig Trends aufgrund von Weak Signals entstehen. Weak Signals sind Änderungen, die noch keinen direkten Einfluss auf einem Trend haben, aber dessen Wechselwirkungen mit anderen Entitäten dazu führt, dass sich der Trend ändert. Ein einfaches und offensichtliches Beispiel, sind die Twitter Nachrichten von Elon Musk zu Dodge Coin. Er ist kein Anbieter und evtl. Kein Nachfrager, hat aber einen enormen Einfluss auf den Kurs von Dodge Coin.

Nun hatten wir die Hypothese, das man die direkten Einflussfaktoren auf die Markveränderungen mithilfe von Nachrichtenartikeln verfolgen kann. Die Informationen hieraus haben wir dann extrahiert und mit dem Nachrichtentext annotiert. Dabei haben wir die Annotationen mit DBPedia verknüpft um die Instanzen eindeutig zuordnen zu können und weitere Informationen zu diesen zu erhalten.

Meine Aufgaben hierbei war zunächst die Abfrage von Informationen aus den Knowledgegraphen angereichert um weitere Informationen aus anderen Knowledgegraphen wie DBPedia und Wikidata, und die analyse der Daten. Das hat sich dann über die Zeit mehr in die Weiterentwicklung der Pipeline verschoben. So habe ich da zum Beispiel noch die Identifikation von Fakten mit integriert. Damals habe ich den rel. einfachen Ansatz verfolgt, Relationen in Texten zu identifizieren und diese mit ihren Konzepten aus Wordnet zu verbinden. Somit hatte ich dann die Möglichkeit Klassendefinitionen für Fakte zu definieren. Dabei habe ich einen Fakt als eine Regel betrachtet bei der min. eine Relation mit einem Objekt zusammen vorkommt. Zum Beispiel habe ich damit eine Verkaufsaktion als Regel definiert, bei der es hieß, das min. 2 Unternehmen in den Satz zusammen mit einem Verbkonzept, das verkaufen beschreibt, vorkommen. Der Ansatz hatte mehrere Schwächen, aber war zumindest schon mal ein Vorschlag. SO hatten wir dadurch die Möglichkeit von den Informationen die in dem Text direkt vorkommen weiter zu abstrahieren und neue Erkenntnisse zu erschließen. Allerdings gab es hier zum einen das Problem, das ich nur den Satz als Ausgangsbasis betrachtet habe und zum anderen das ich nicht wirklich betrachten konnte was die Bedeutung des Satzes war. Also zum Beispiel wurde nicht ein "nicht" identifiziert.

Den Ansatz wurde dann aber nicht weiter betrachtet, da dann BERT eine immer größere Rolle gespielt hat. Bei BERT handelt es sich um ein Language Model, dass die statistische Struktur einer Sprache gelernt hat. Dieses Language Model kann dann verwendet werden um zum Beispiel fürs QA den ähnlichsten Textausschnitt zu einer gegebenen Frage zu finden. Damit ist es dann auch möglich die vorher gegebenen Fakten zu identifizieren. Zumindest theoretisch.

Da ich WING studiere, habe ich eine Projektarbeit noch als Teil meines Studiums gemacht. Hierfür habe ich am KtmfK tribologische Modellversuche charakterisiert, sowohl durch den wie diese in der Literatur vorkommen, als auch wie diese in Versuchsaufbauten beschrieben werden, und daraus ein Modell entwickelt. Dieses Modell kann man als eine Art Ontologie betrachten, bei der allerdings noch keine Regeln definiert sind. Das habe bzw. Bin ich gerade drüber dann im zweiten Teil der Arbeit vor/gemacht. Dabei verwende ich einen "Induktiven" Ansatz um automatisch Regeln und Präferenzen zu identifizieren, die für Klassen oder Relationen gelten. Hierbei wird die Instanzen der Versuchsaufbauten in ein ASP überführt und analysiert. Dabei definiert man zunächst wie eine gültige

Regel auszusehen hat. Anschließend werden alle gültigen Regeln und das BK und die Positiv und Negativbeispiele an einen Grounder und einem Solver übergeben. Dieser prüft dann welche kleinstmögliche Kombination an Regeln dazu führt alle Positiv- einzuschließen und alle Negativbeispiele auszuschließen.

Im wesentlichen heißt das, wir können implizites Wissen, das in den Versuchsbeschreibungen drinsteckt, explizit machen, und in eine Ontologie mit integrieren, die dann wiederum dazu genutzt werden kann, die Regel auf alle Individuen anzuwenden.

Für mich offene Fragestellungen sind allerdings noch:

- Was ist der Vorteil gegenüber eines DecisionTrees?
- Was ist der Vorteil gegenüber Faktoranalysen?
- Wie geht man bei komplexen Problemen vor? Wie kann ich hierbei Predicate Invention anwenden?
- Skaliert es?

Ein Punkt bei Ontologienentwicklung ist es ja, das man Sachen aus anderen Ontologien wiederverwenden sollte. Allerdings hab ich häufiger das Gefühl gehabt, das es aus einigen Ontologien immer nur ein oder 2 konzepte sind, die relevant sind. Macht es da überhaupt sinn?

Am Ende hat man einen Flickenteppich, der irgendwie wirr wirkt.

Kann es sein, das viele Ontologien nicht gut sind?

Was möchtest du erforschen?

Ich frage mich ob was man mithilfe von den strukturierten Daten des Semantic Webs als Instanzen und Ontologien als Background und ILP als System zur idenfifikation neuer Regeln oder erkenntnisse anstellen kann. Zum Beispiel bei Systems Dynamics unterstützen kann. Bei System Dynamics handelt es sich ja um die Behandlung von Systemen in einem dynamischen Kontext. Das System wird dabei wie statistisches System beschrieben, also die Elemente und deren Beziehungen (in diesem Fall Wechselwirkungen genannt), allerdings werden auch

Oder mehr ins technische. Da finde ich die Idee von MOSAIK ziemlich interessant. Wenn ich es richtig verstanden habe, ist das Ziel davon Softwarekomponenten unabhängig voneinander arbeiten zu lassen und ein standardisierte Kommunikation zu realisieren, sodass ein dynamischer Austausch zwischen den Komponenten aber auch neuen Komponenten möglich ist.

Da frage ich mich aber, wieso nicht einfach bestehende REST-APIs erweitern?

Warum ich?

- Suchen eines Doktorvaters/mutter

**Was genau versteht du unter einen Knowledgegraphen?**

**Warum Knowledgegraphen im Prozesskontext bzw. Maschinendatenkontext?**

Diese stellen ja an sich nichts anderes wie eine Datenbank dar, deren Optimierungskriterium, das Traversieren von Knoten ist. Allerdings habe ich bei uns den Verdacht, dass häufig die Daten tat. Eher besser in einer relationalen Datenbank passen könnten.

**Was genau verstehst du unter einer Ontologie?**

Vokabular, Liste an Beschreibungen mit deren Regeln, Regelwerk

**Was ist der Unterschied zwischen einer Systembeschreibung und einer Ontologie?**

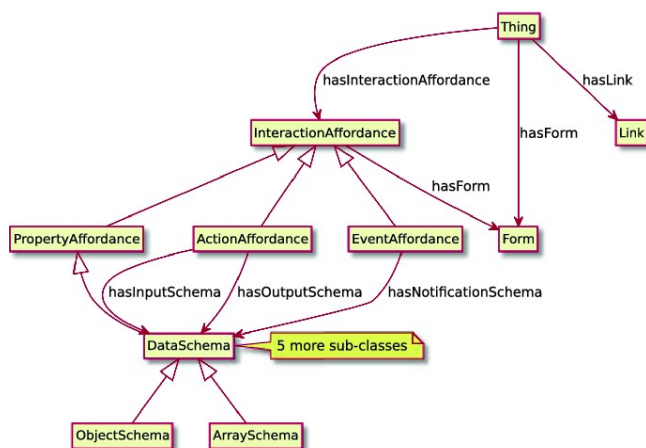
Für mich wirkt beides sehr ähnlich, da beide eine Repräsentation eines Systems sind, und darstellen, wie dieses mit Elementen in dem System interagiert. Allerdings glaube ich, dass bei der Ontologie zusätzlich explizit Regeln, und Axiome definiert werden.

**Komplett anderes Gebiet um Knowledgegraphen einzusetzen?**

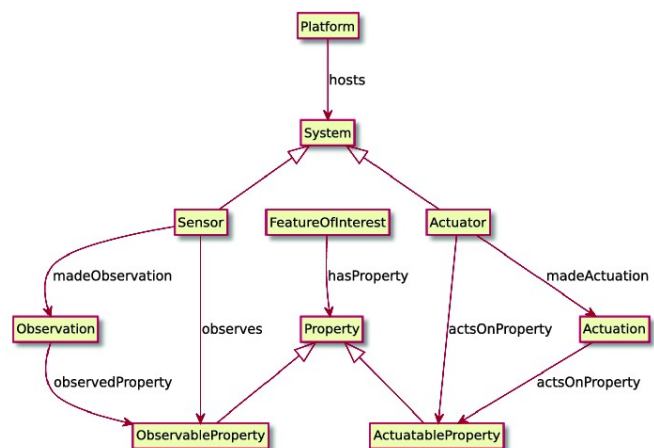
**Was genau ist bei euch der Grund dafür Knowledgegraphen für Maschinendaten einzusetzen? Funktioniert?**

Ich hab das Gefühl, es werden zu viele Ontologien entwickelt, die dasselbe beschreiben. Dazu muss man sich ja nur mal die große Anzahl an Knowledgegraphen betrachten, die es aktuell gibt. Ein Punkt bei Ontologien ist ja eig., dass diese eine geteilte Konzeptionalisierung sein sollten. Das wird auch gemacht, aber die Gruppe ist so klein, dass es sich eig. nicht rendert eine Ontologie zu bilden.

Ontology Deep oder Broad?



(a) Thing Description model



(b) Semantic Sensor Network model



Semantic Web Lab:

Ziel:

Hier geht es darum, maschineninterpretierbare und wiederverwendbare Daten- und Workflowmodelle durch semantische Modellierung zu erhalten und Wissensgraphen für Daten und Workflows in Produktion und Logistik zu gestalten.

Zentrale Fragestellungen:

Wie lassen sich Daten und Workflows im Produktions- und Logistikumfeld modellieren und repräsentieren, sodass diese zum einen maschinell gelesen und interpretiert werden können?

Wie lassen sich Daten und Workflows in ähnlichen Anwendungsszenarien wiederverwenden und bei Bedarf adaptieren?

Link4Pro

Ziel:

Das Forschungsprojekt »Link4Pro – Flexible Vernetzung mehrstufiger Produktionsprozesse zur ressourceneffizienten Produktivitätssteigerung« entwickelt ein neuartiges System zur Erfassung, Analyse und Optimierung von Produktionsketten für kleinere und mittlere Unternehmen (KMU).

Manche Daten kommen sehr hochfrequent im ms bereich, andere hingegen werden einmal zum Start der Maschine oder nur bei Fehlermeldungen eingetragen. Wie handhabt ihr unterschiedliche Formen von Daten?

Interview bei Fraunhofer:

Datenräume:

Das ist auch die Frage unseres übergreifenden Forschungsthemas: Datenintegration. Es geht darum, Daten aus unterschiedlichen Systemen so miteinander zu verbinden, dass diese Daten automatisiert integriert und abgefragt werden können. Dabei spielen formale Sprachen basierend auf mathematischer Logik eine tragende Rolle. Kurz gesagt, wir konzipieren vernetzte Datenräume, indem wir Daten als Wissensgraphen (Knowledge Graphs) repräsentieren und so den Unternehmen einen einheitlichen und flexiblen Zugriff auf alle relevanten Daten erlauben, sowohl unternehmensintern als auch -extern. Ein Beispiel hierfür ist etwa die automatisierte Zusammenführung von physischen Sendungs- mit kaufmännischen Einkaufs- oder Absatzdaten.

→ Welche Daten lest ihr da zum Beispiel ein?

→ Wie verknüpft ihr die Daten?

→ Hat es am Ende tat. Etwas gebracht?

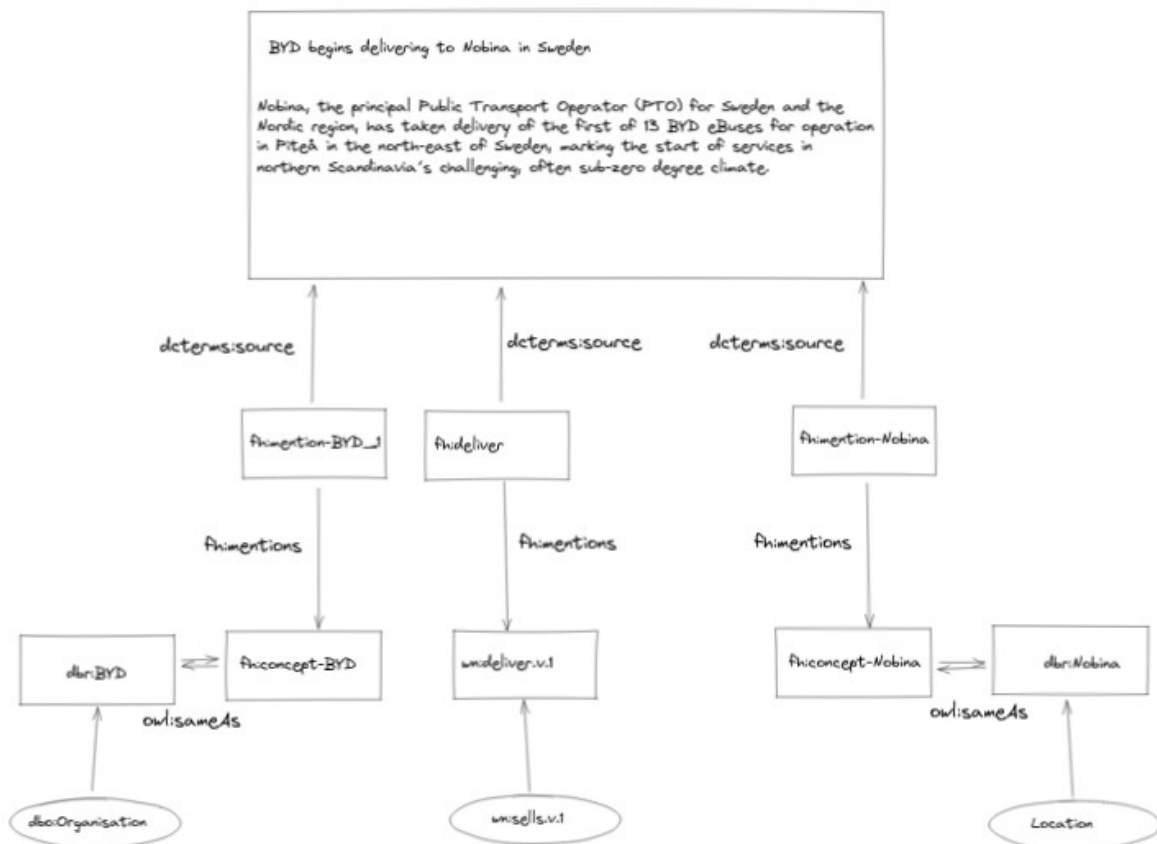
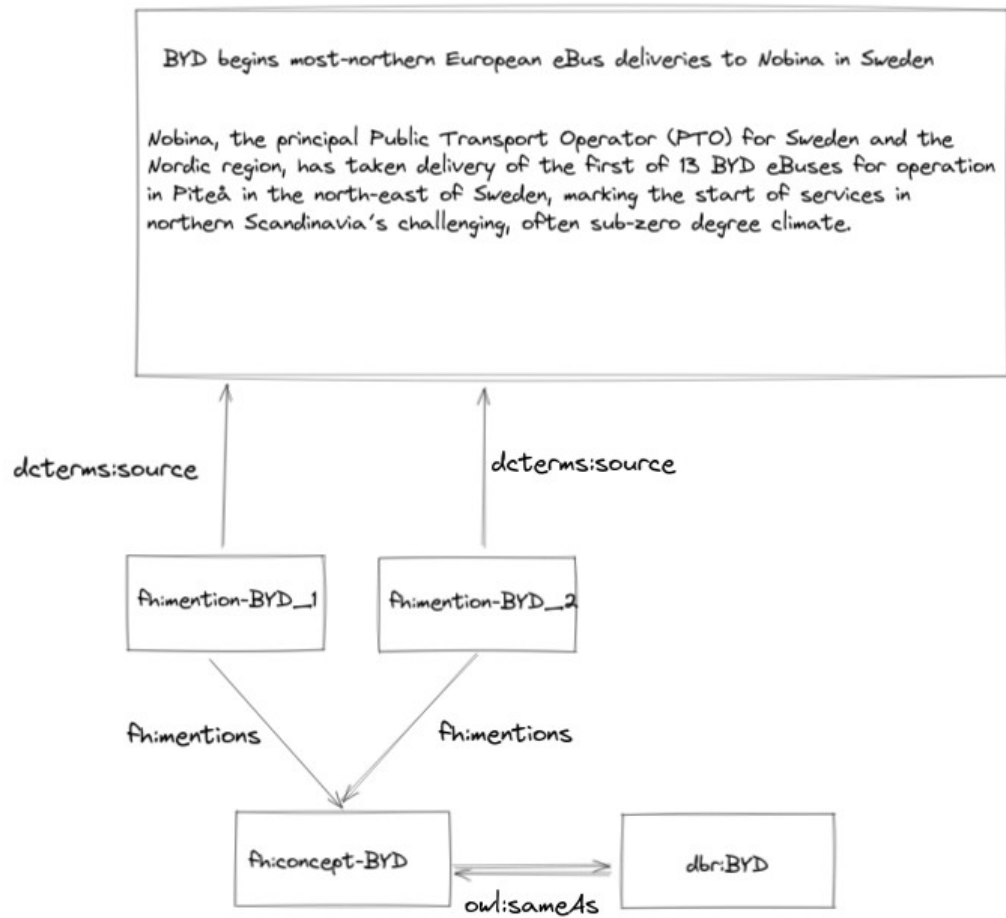
→ Macht es überhaupt sinn die Rohdaten zu integrieren? Große Datenmenge

→ Warum Knowledgegraphen?

→ Wenn es nur um die zusammenfassung der Daten für Analyse zwecke geht aknn man genauso gut auch PowerBI verwenden, das einzelne Datenbanken (auch KG) abfragen kann um an neue Infos zu kommen und diese miteinander zu verknüpfen.

Zweitens – wie kann ich mehr Daten einspielen, bzw. welche Daten lassen sich weiterhin noch nutzbar machen?

In der industriellen Anwendung kommen Daten oft von Sensoren, welche über eine drahtlose Verbindung kommunizieren. Nachdem die Daten verfügbar gemacht wurden und integriert sind, ist der nächste Schritt die Veredelung der Rohdaten hin zu Ereignissen, die für die Prozesse in Unternehmen relevant sind, etwa dem Warenein- oder -ausgang.

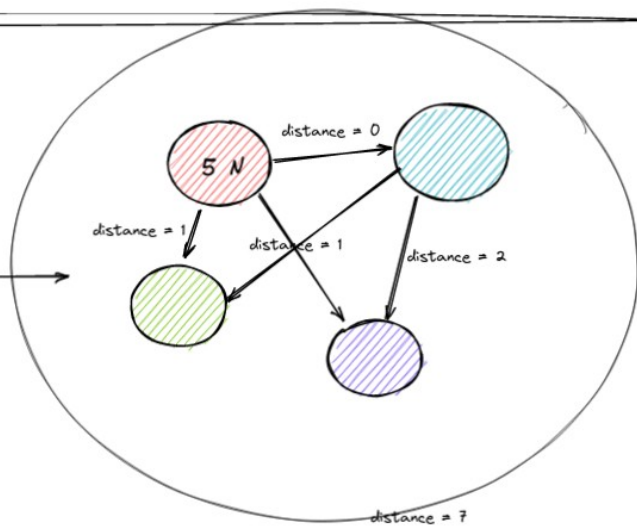
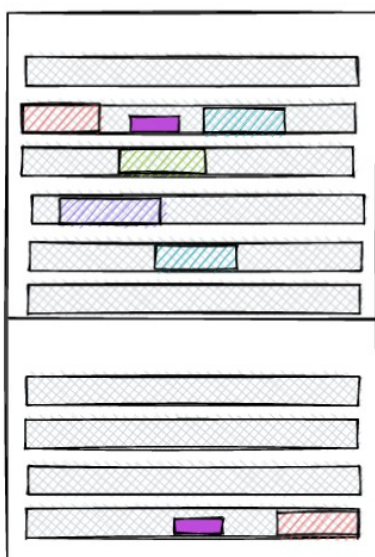


```

"velocity":{
  "question":[
    "What quantity was the velocity has been used during the test?",
    "What quantity was the velocity in meter per second?",
    "What was the velocity?",
    "What quantity was the speed in meter per second?"
  ],
  "expectedAnswerSpace":[
    "([0-9]{0,5}\\.|\\.|[0-9]+ (k|m|)m( )(/|per|)( ))(s|h|min)"
  ],
  "preSearchSpace":[
    " m/s ",
    "m/min",
    "tribometer",
    "velocity",
    "speed"
  ],
  "normalState":"Unspecified",
  "dependend_from_variables":[
    "<Rounds>",
    "<Speed>"
  ],
  "specific_questionType":"VELOCITY",
  "broader_questionType":"OPERATIONAL_PARAMETER",
  "weak_dependency_to":[
  ],
  "strong_dependency_to":[
  ],
  "controlquestions":[
    {
      "to":"TEST_METHOD",
      "question":"What testmethod was used with [ANSWER]?"
    }
  ]
}

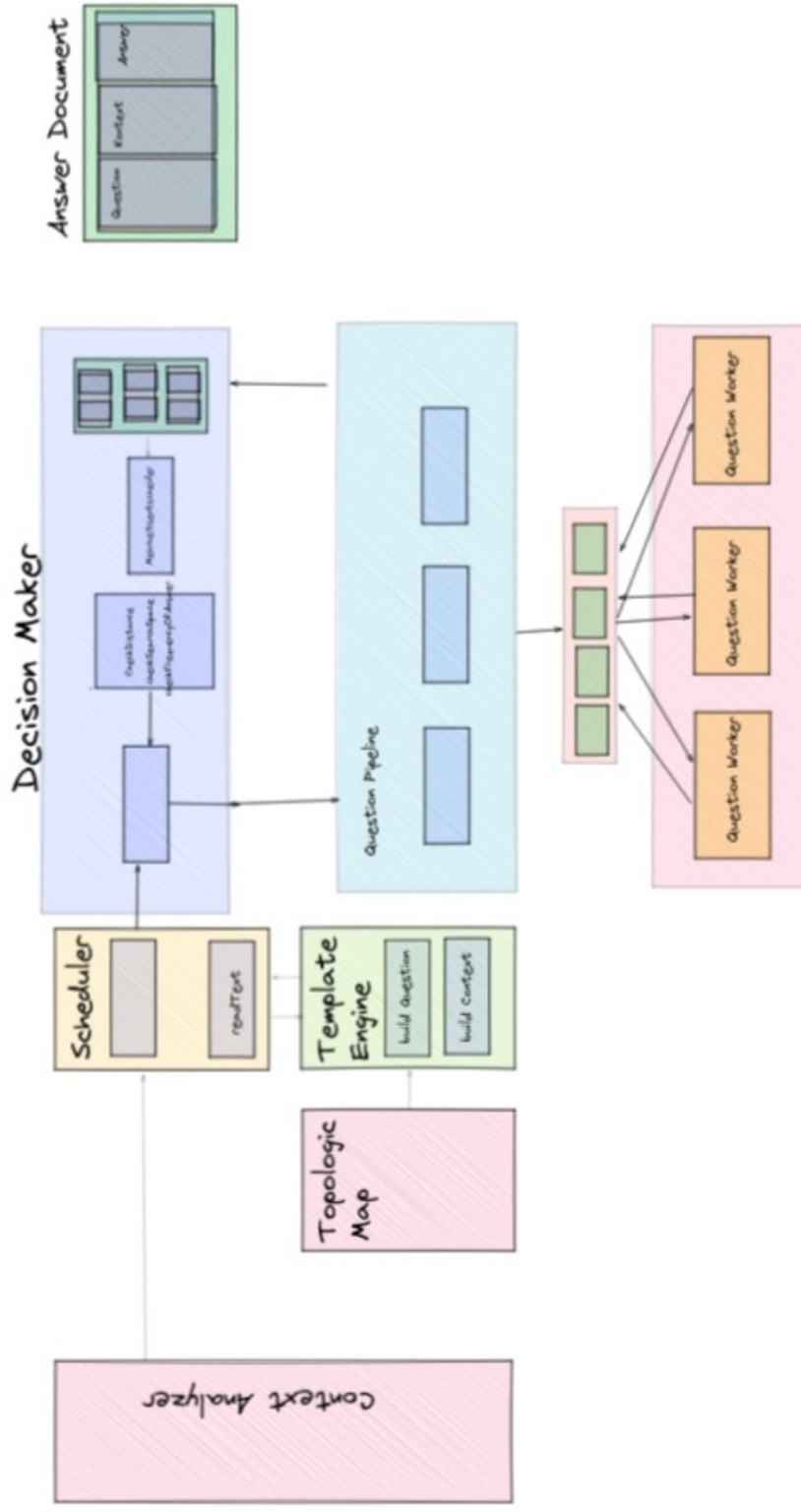
```

## Topologische Karte









#### Examples

```
#pos({heads(c1), tails(c2), heads(c3)},  
      {tails(c1), heads(c2), tails(c3)}).
```

```
#pos({heads(c1), heads(c2), tails(c3)},  
      {tails(c1), tails(c2), heads(c3)}).
```

#### Background Knowledge

```
coin(c1).    coin(c2).    coin(c3).
```

#### Definition of the Searchspace for the Hypothesis

```
#modeh(heads(var(coin))).
```

```
#modeh(tails(var(coin))).
```

```
#modeb(heads(var(coin))).
```

```
#modeb(tails(var(coin))).
```

```
#modeb(coin(var(coin))).
```

```
#modeh(heads(const(coin))).
```

```
#modeh(tails(const(coin))).
```

```
#constant(coin, c1).
```

```
#constant(coin, c2).
```

```
#constant(coin, c3).
```

#### Gelernte Regeln:

```
heads(V1) :- coin(V1), not tails(V1).
```

```
tails(V1) :- coin(V1), not heads(V1).
```

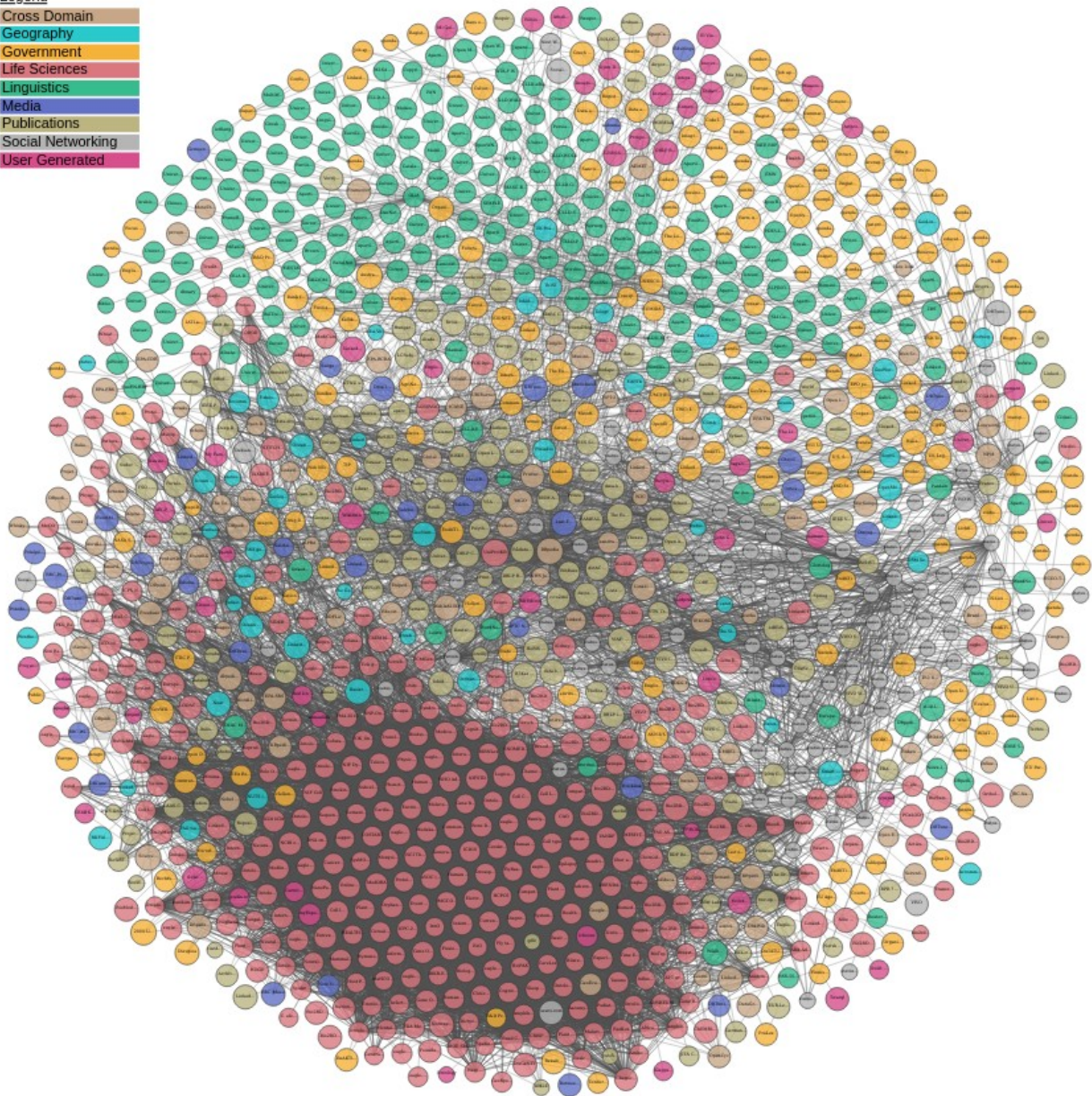
```
#pos({}, {}, {  
  node(1..4).  
  edge(1, 2).  
  edge(2, 3).  
  edge(3, 4).  
  edge(4, 1).  
}).
```

```
#neg({}, {}, {  
  node(1..4).  
  edge(1, 2).  
  edge(2, 1).  
  edge(2, 3).  
  edge(3, 4).  
  edge(4, 2).  
}).
```

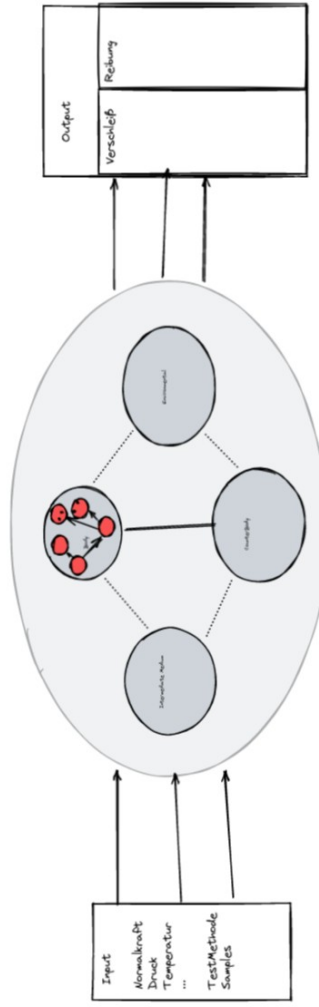
```
0 { in(V0, V1) } 1 :- edge(V0, V1).  
reach(V0) :- in(1, V0).  
reach(V1) :- reach(V0), in(V0, V1).  
:- not reach(V0), node(V0).  
:- V1 != V2, in(V0, V2), in(V0, V1).
```

Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated



<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>1.47 dPa 10 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>
<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>	<p>0.8 dPa 20 %</p> <p>10 mm 0.08 m/s 8 N 1000 m 100 Cr 6 Rm Gs 0.02 m Ball 3,6 kg m/s</p> <p>100 Cr 6 Rm Gs 0.0215 m Average Length Stem Height Stem Height Stem 62 IRC</p> <p>Handleng, Tempel, Aeron-Poligal</p>



bar of the steel/steel-reference also tended to  
: slightly lower with lower pressure and increasing