

# Präsentation

Bachelorarbeit:

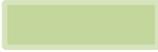
*A keyword translator for Broccoli*

Iradj Solouk



- Einleitung
- Algorithmus
- Evaluation
- Verfeinerung

- Einleitung
  - Motivation
  - Definition
- Algorithmus
- Evaluation
- Verfeinerung

- Manuelle Erstellung einer Query
  - Wissen über Suchelemente nötig
    - Klasse 
    - Instanz 
    - Relation 
    - Text 
  - Inkrementell
  - Vokabular
- Benutzerfreundlichkeit

- Systemkomponente im/am UI



- Input = Benutzereingabe
- Output = (interne) Query/Query Tree
- Kommunikation mit Backend

- Interpretation der Benutzereingabe
  - Kontext
  
- Natürliche Sprachverarbeitung
  - (Stanford)POS tagging
  - (Princeton)WordNet

- Einleitung
- Algorithmus
  - Instance As Root
  - Class As Root
- Evaluation
- Verfeinerung



- Beispiele
  - „Angela Merkel Spouse“
  - „Mississippi River Length“
  
- Angabe der Klasse fehlt
  - Backend

## ■ Vorgehen

Eingabe	„Angela Merkel Spouse“
Instanzen suchen	 „Spouse“
Relation suchen	
Prüfen	

- Eigenschaften
  - kein semantisches Abbilden
  - **Instance As Root** in **Class As Root** konvertieren



- Angabe der Relation kann fehlen
  - „Movie by Stephen Chbosky“
  - „Apollo Astronauts“
- Indikator für Relation
  - „Football player with a **tallness** of  $>1.65$  and who was **born**  $>1965$  and has the forward **position**“

## ■ Routine

- Finde Wurzelklasse(n)
  - morphologisch
  - semantisch

- Finde Subjekt mithilfe von POS tagging
- Stringmatching
  - Wenn ausreichend
    - Klasse gefunden
    - sonst: Verwendung alternativer Ausdrücke (Hypernyme, Synonyme)
- „Movie by Stephen Chbosky“

## ■ Routine

- Finde Wurzelklasse(n)
  - morphologisch
  - semantisch
- Konsumieren von Relationen
  - morphologisch & semantisch

- Füllwörter werden gefiltert
- **Für jedes** übrig gebliebene Eingabewort werden Semantisch verwandte Ausdrücke gesammelt (WordNet)
- Fiktives Beispiel
  - „Tiere welche Mäuse fressen“

# Class As Root



Tier

welche

Mäuse

fressen

- Mäuse
- Maus
- Nagetier
- ...

- Fressen
- Essen
- Verspeisen
- Nahrung aufnehmen
- ...

Tier

Nahrung

Verwandt mit

...

Mäuse

fressen

Mäuse  
Maus  
Nagetier  
...

Fressen  
Essen  
Verspeisen  
Nahrung aufnehmen  
...

- Prüfe Stringmatching dieser Ausdrücke mit den Relationen der Wurzelklasse

Tier

**Nahrung**

Verwandt mit

...

Mäuse

**fressen**

Mäuse  
Maus  
Nagetier  
...

Fressen  
Essen  
Verspeisen  
**Nahrung** aufnehmen  
...

- Prüfe Stringmatching dieser Ausdrücke mit den Relationen der Wurzelklasse
- Merke wahrscheinlichere Relationen
- Umliegende Worte auf Instanz abbilden

Tier

**Nahrung**

Verwandt mit

...

Mäuse

**fressen**

Mäuse  
Maus  
Nagetier  
...

Fressen  
Essen  
Verspeisen  
**Nahrung** aufnehmen  
...

- Prüfe Stringmatching dieser Ausdrücke mit den Relationen der Wurzelklasse
- Merke wahrscheinlichere Relationen
- Umliegende Worte auf Instanz abbilden



- „Football player with a **tallness** of  $>1.65$  and who was **born**  $>1965$  and has the forward **position**“

## ■ Routine

- Finde Wurzelklasse(n)
  - morphologisch
  - semantisch
- Konsumieren von Relationen
  - morphologisch & semantisch
- Instanzerkennung

## ■ Routine

- Finde Wurzelklasse(n)
  - morphologisch
  - semantisch
- Konsumieren von Relationen
  - morphologisch & semantisch
- Instanzerkennung
- Inferenz von Relationen



- Einstufig
  - Für jede gefundene Instanz: finde eine Relation, welche die Wurzelklasse und die Instanz verbinden
- Zweistufig
  - Für jede gefundene Instanz: finde ein paar von Relationen, welche die Wurzelklasse und die Instanz miteinander verbinden
- „Apollo Astronauts“

## ■ Routine



- Finde Wurzelklasse(n)
  - morphologisch
  - semantisch
- Konsumieren von Relationen
  - morphologisch & semantisch
- Instanzerkennung
- Inferenz von Relationen

- Textsuchelement anhängen
- Eigenschaften
  - Mehrere Übersetzungen
  - semantisches Abbilden

- Einleitung
- Algorithmus
- Evaluation
  - Aufbau
  - Ergebnis
- Verfeinerung

- SemSearch Benchmark
  - 50 Queries
- Ground Truth = Manuelle Queries
  - 42 Queries
- Maß: Ähnlichkeit der Hits
  - Optimal
  - Ausreichend
  - Ungenügend

- 42 Queries

Optimal	14
Ausreichend	7
Ungenügend	21

- Positiv übersetzt

„Apollo Astronauts who walked on the moon“

„Boroughs of New York City“

„Did Nicole Kidman have any siblings?“

„Gods who dwelt on Mount Olypmpus“

„Movies starring Erykah Badu“

...

- Ungenügend übersetzt

„Axis powers of World War II“

„Branches of the US military“

„Henry II's brothers and sisters“

„Hijackers in the september 11 attacks“

„Houses of the russian parliament“

...

- Einleitung
- Algorithmus
- Evaluation
- Verfeinerung
- Zusammenfassung

- Natürliche Sprachverarbeitung
  - Subjekt bestimmen
- Komplexe Queries
  - „Compositions composed by someone who had no children“
- Relationen durch semantisches Abbilden bei Instance As Root-Queries finden
  - „Henry II’s brothers and sisters“

- **Ontology Feedback**
  - Abstrakte Klasse einstellen
  - Text Suchelement verwenden um auf eigentliche Klasse zu schliessen
    - „Wonders of the ancient World“

Vielen Dank für Ihre Aufmerksamkeit!