



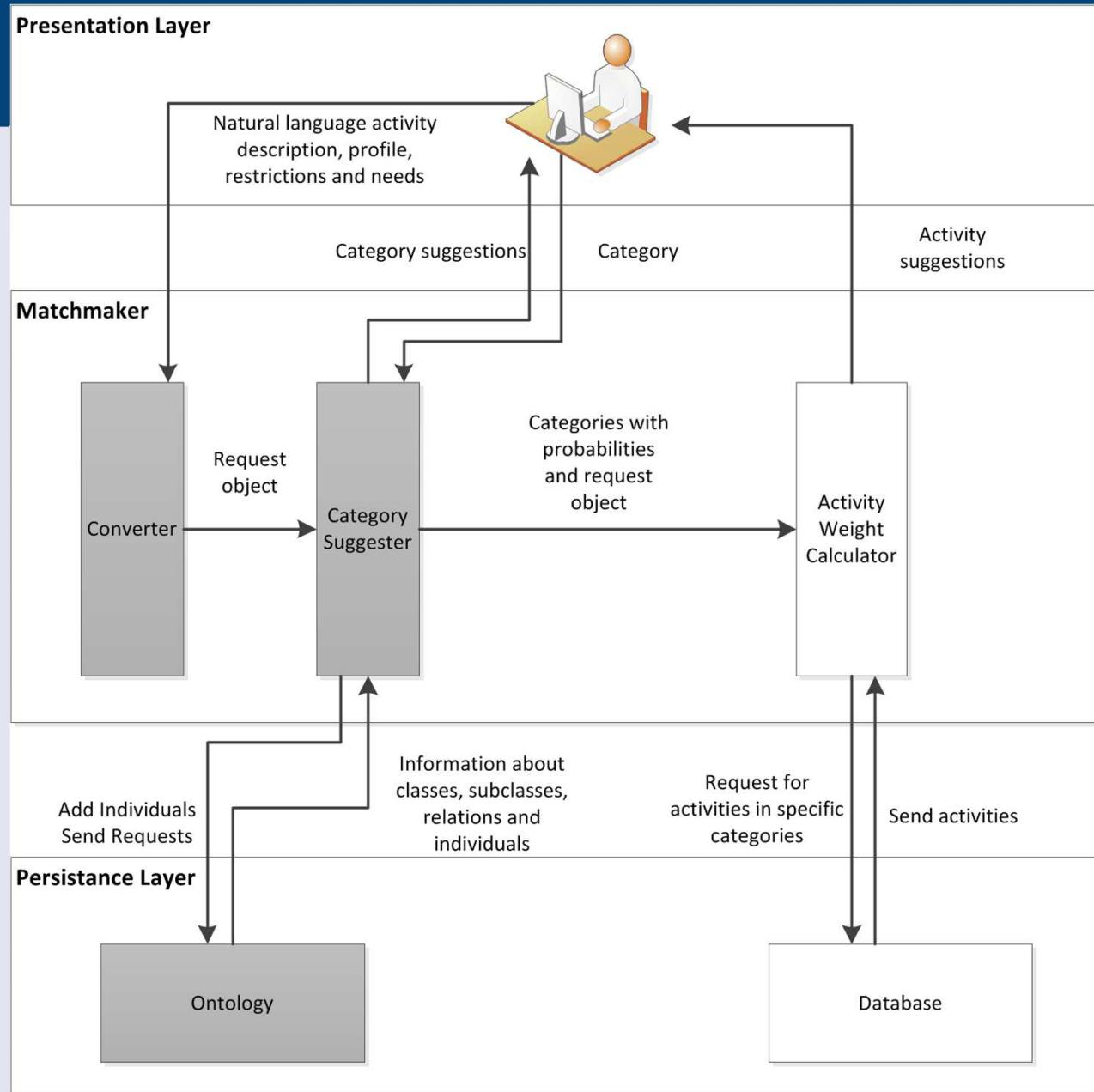
Matching von natürlichsprachigen
Beschreibungen mit Hilfe von
Ontologien

Lukas Berle

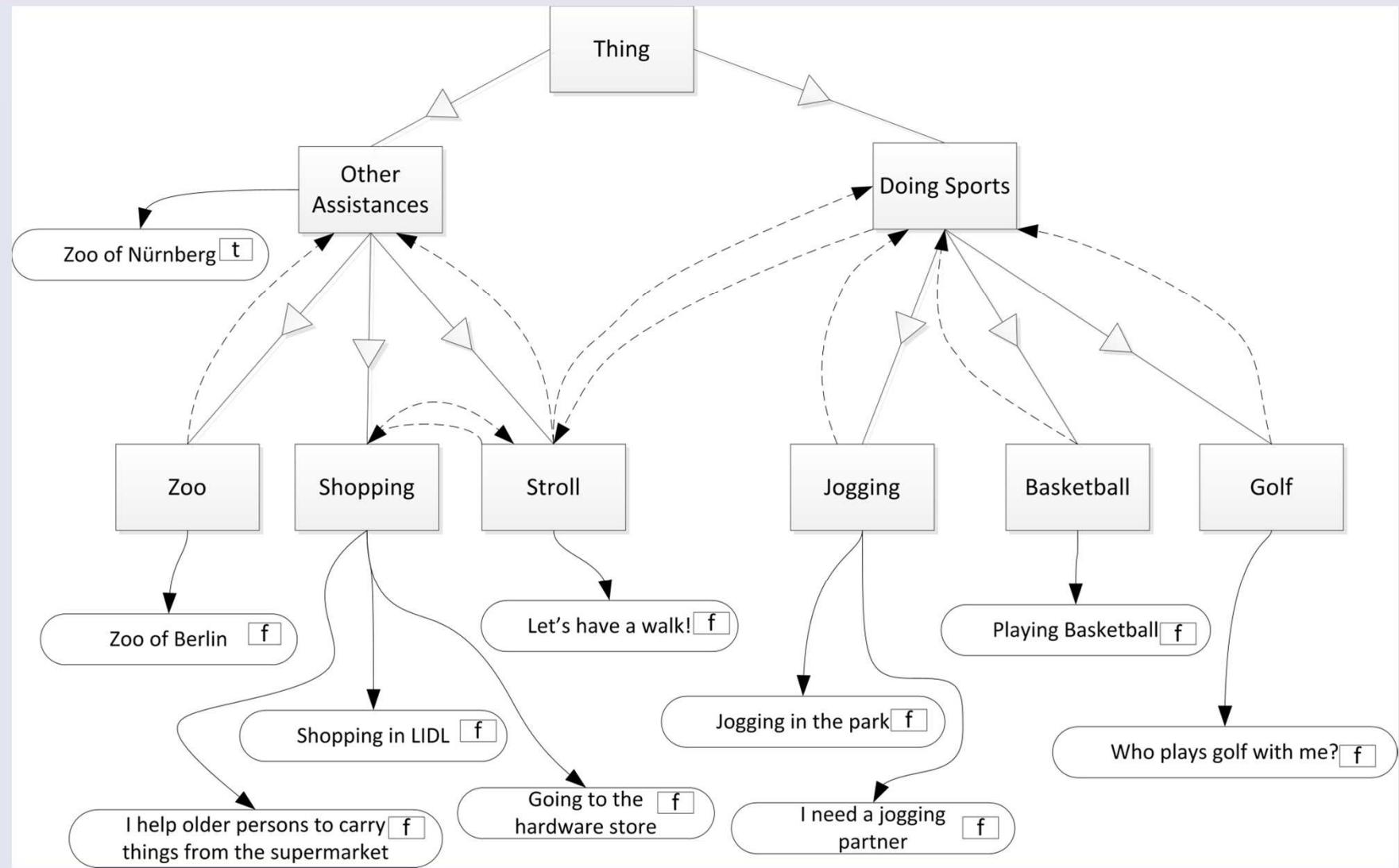
- Ziel und Aufbau des Ansatzes
- Funktionalität
 - Direct Match
 - Individual Based Matching
 - Lernen
- Probleme
 - Irrelevante Wörter
 - Unbalancierte Ontologie
 - „Falsches“ Lernen
- Ausblick
- Programm Demonstration



- User gibt eine natürlichsprachige Beschreibung einer Aktivität in das Suchfeld ein
- Implementierte Software weißt dieser Beschreibung eine oder mehrere Kategorien der Ontologie zu
- Für den Fall, dass mehrere Kategorien zugewiesen wurden, werden Wahrscheinlichkeiten zurückgegeben
- Software „lernt“ durch Usereingaben

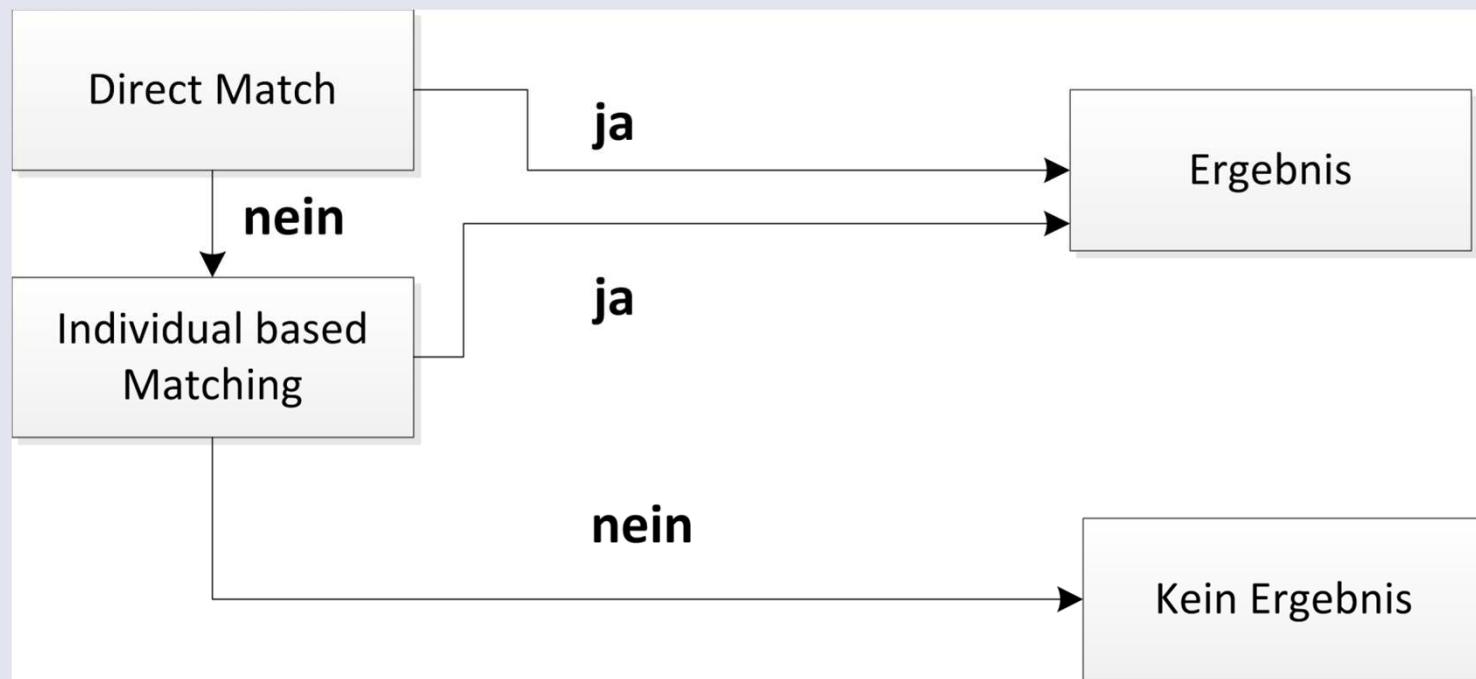


Eine Ontologie

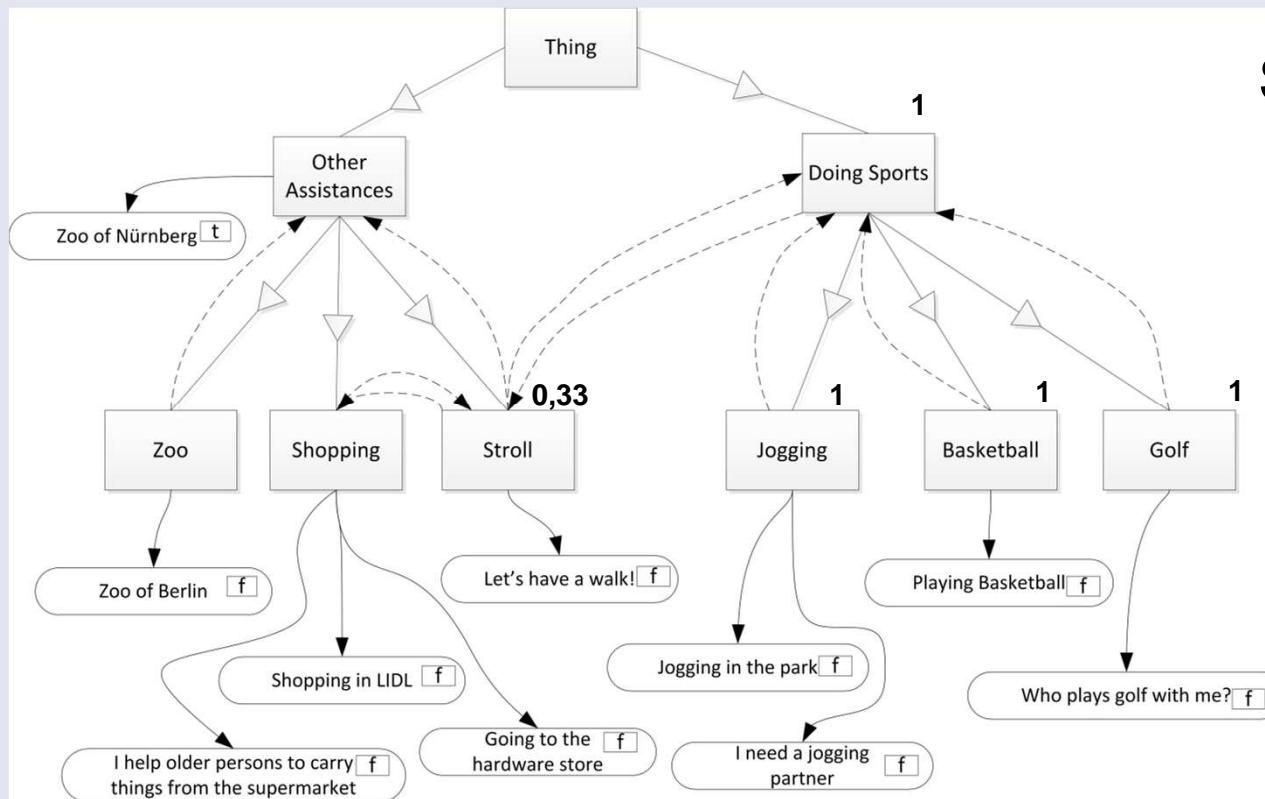


Ansatz

1. Stopwords werden aus Suchbegriff gelöscht
2. Direct Match? Falls nein, Individual Based Matching
3. Rückgabe der Kategorien mit Wahrscheinlichkeiten



Beispiel: Suche nach „Sports“



Similarity Weight: 1/3

Ergebnis nach Normalisierung:
 Doing Sports, Jogging, Basketball und Golf: jeweils **23 %**
 Stroll: **7,6 %**



Beispiel:

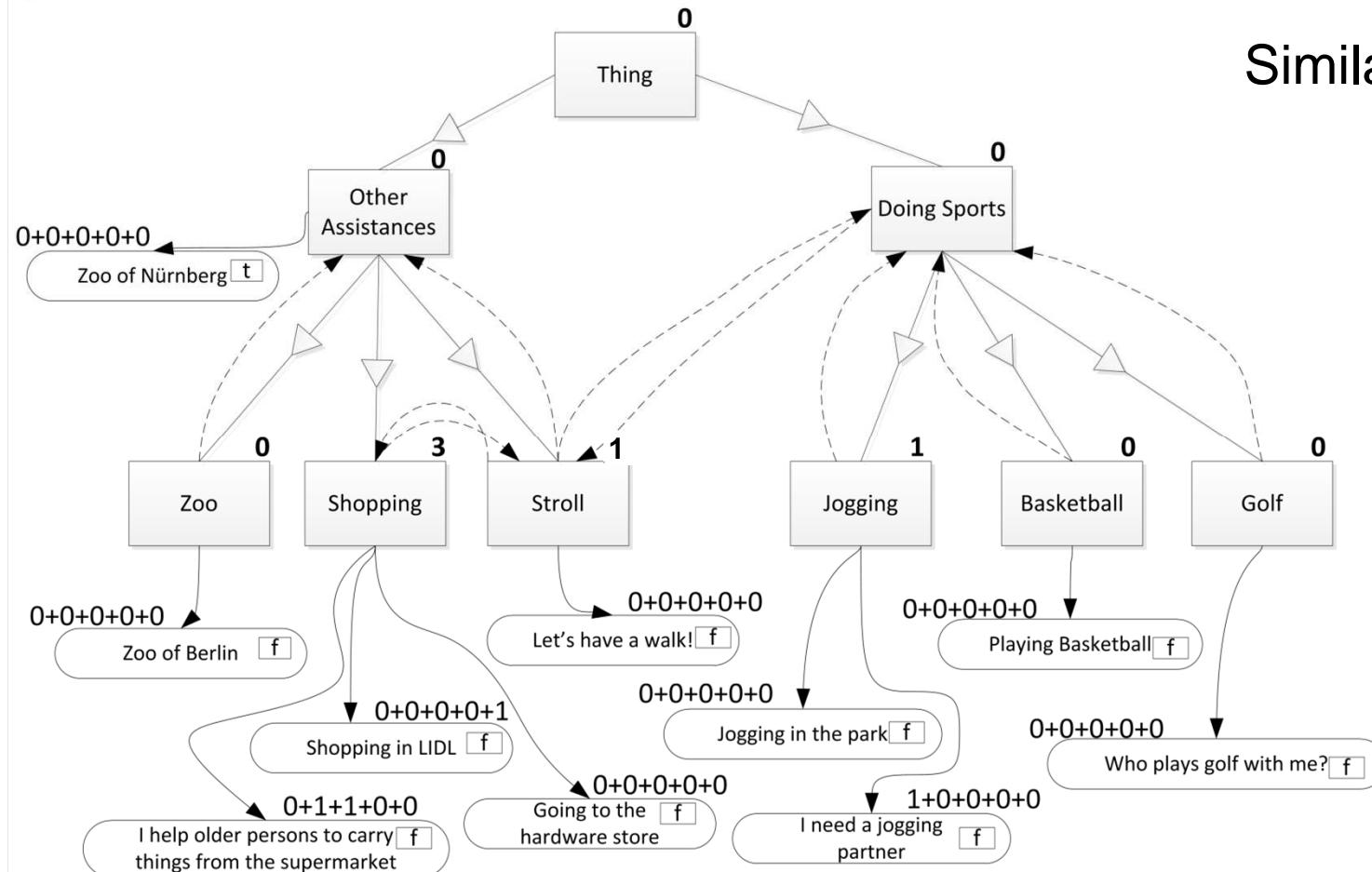
Suche nach „**Need help to carry drinks from LIDL**“

1. Löschung der Stopwords: “Need help carry drinks LIDL”
2. Direct Match → Kein Erfolg (= kein Treffer)
3. Individual Based Matching

Individual Based Matching

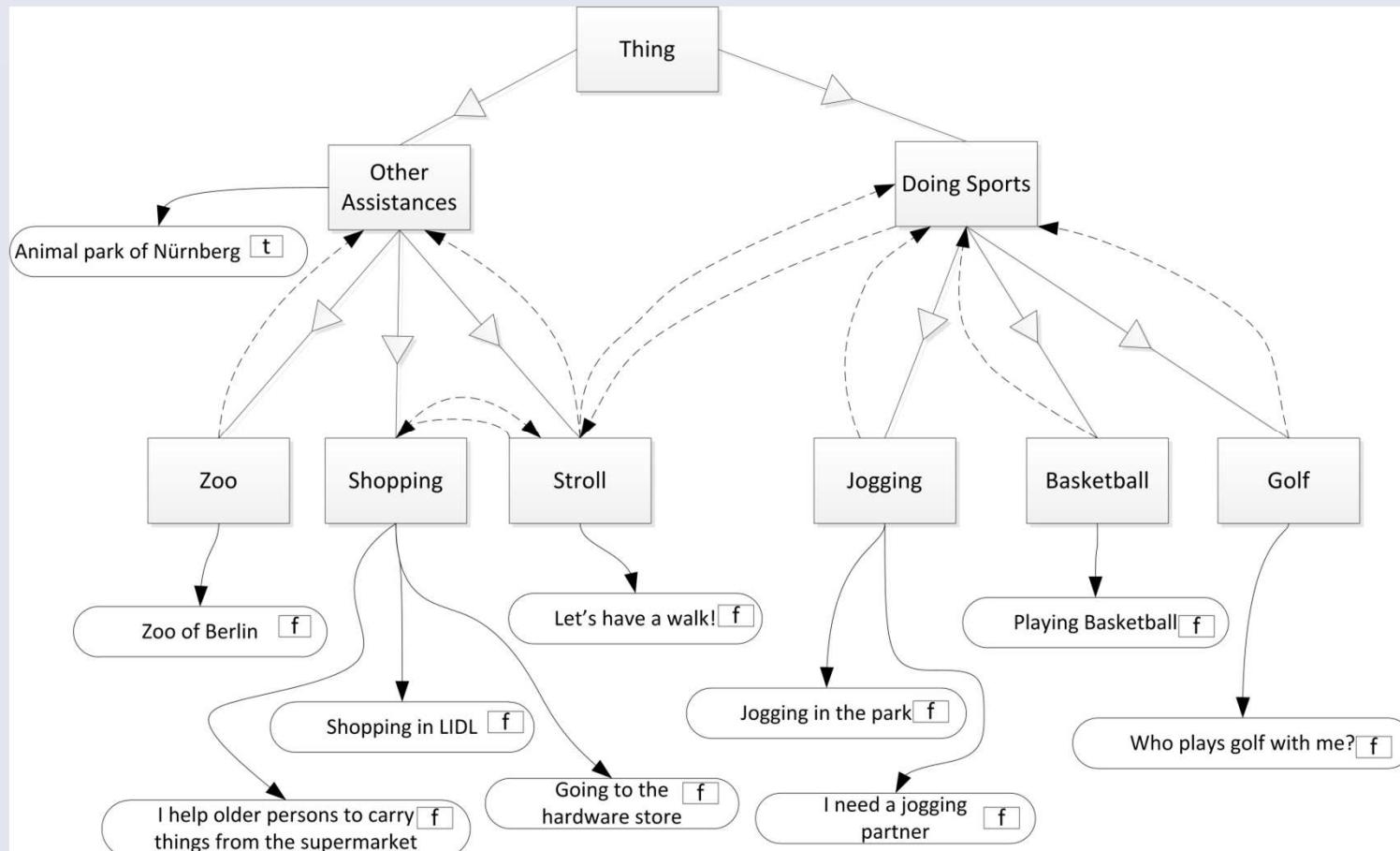
“Need help carry drinks LIDL”

Universität Bamberg



Similarity Weight: 1/3

Normalisiertes Ergebnis: Shopping: 0.6, Stroll: 0.2 und Jogging: 0.2



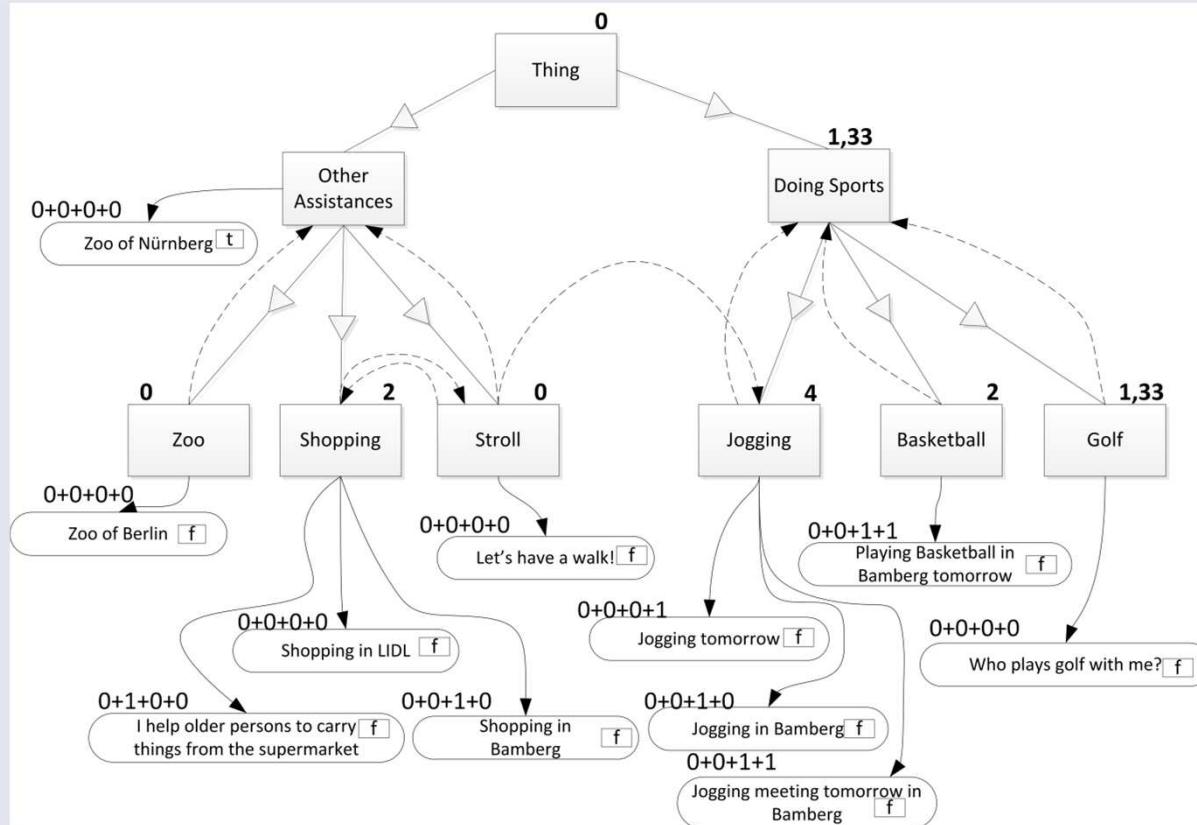
- Irrelevante Wörter in der Ontologie
- Unbalancierte Ontologie
- Falsche Unterordnung eines Individuals

Irrelevante Wörter in der Ontologie

Universität Bamberg



Suche nach: „go supermarket Bamberg tomorrow“



Shopping: 0.19;
Golf: 0.12;

Jogging: 0.38;
Doing Sports: 0.12

Basketball: 0.19;

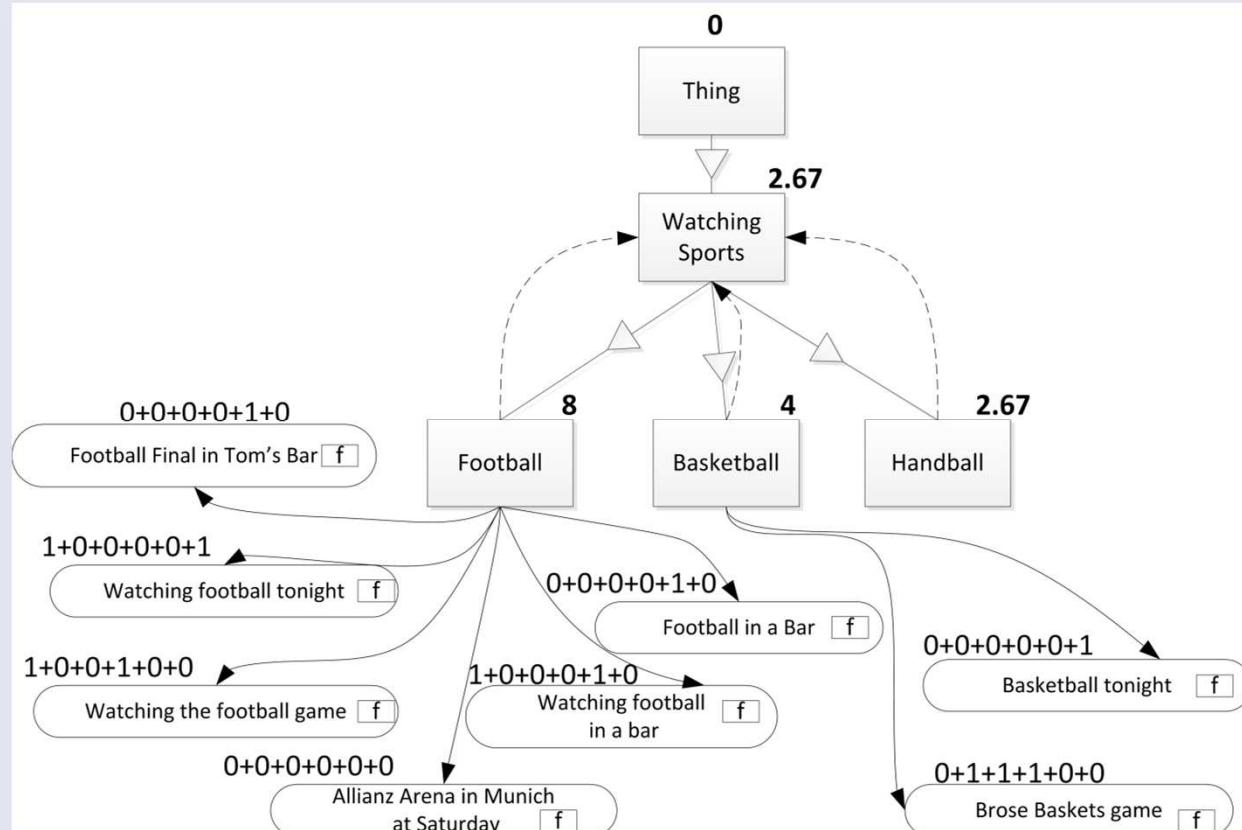
Irrelevante Wörter in der Ontologie

Universität Bamberg



- Erweiterung der Stopword-Liste mit Wörtern, die unter verschiedenen Ontologieklassen vorkommen;
- In diesem Fall: „tomorrow“, „Bamberg“,

Suche nach: „Watching Brose Baskets game bar tonight“



Football: 0.46;

Handball: 0.15;

Basketball: 0.23;

Watching Sports: 0.15



Problem hier: „Watching“ ist wichtig zur Aktiv-Passiv Unterscheidung, allerdings störend bei der Zuordnung, bei welchem Sport der User zuschauen möchte. Populäre Sportarten werden ggü. unpopulären bevorzugt.

Lösungen

- Komplettes Ignorieren von aktiv / passiv
(Wörter wie „watching“ werden automatisch auf Stopword-List gesetzt)
→ einfachste Lösung
- Komplexere Lösung?

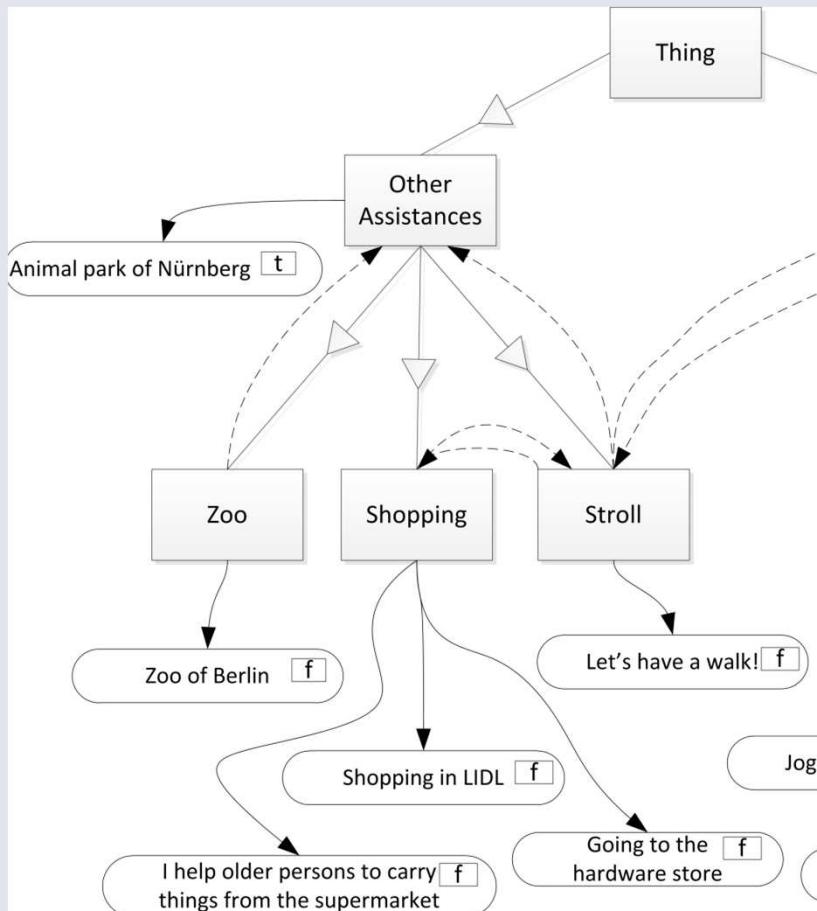
Falsche Unterordnung eines Individuals

Universität Bamberg



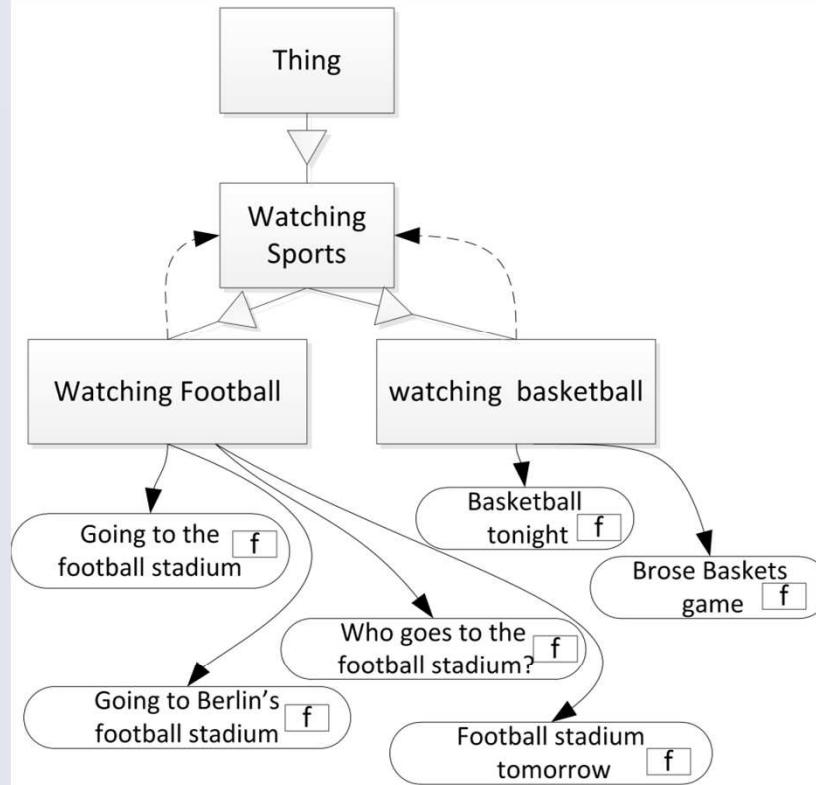
„Fehler beim Lernen“

- Administrator prüft alle Unterordnungen
- Beabsichtigte Manipulationen durch andere Mechanismen (IP-Adresse) aufspüren.

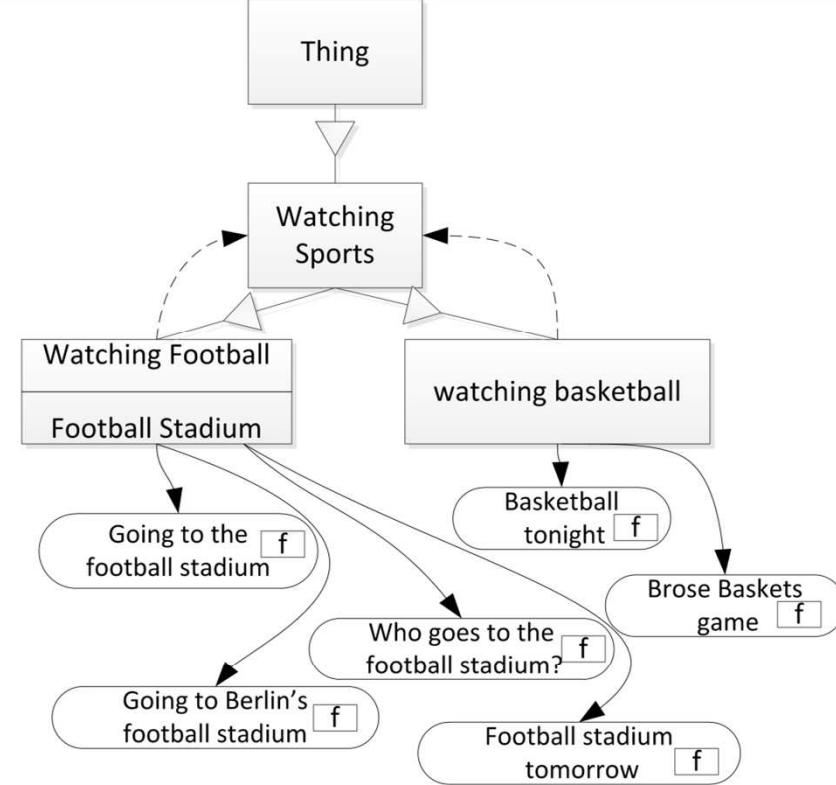


Weiterer Ausblick

Universität Bamberg



Ontology before the transformation



Ontology after the transformation