

SABIO 

**Wissen auf den
Punkt.**

Vorlesungsteil 3: NLP

10.12.2014

evi.com



how tall is the statue of liberty

3. Verstehen der Frage
NLP

which size does the statue of liberty have

how high is the statue of liberty

RECENT QUESTIONS:

How tall is Billy Corgan in feet and inches?

Who flies to Berne Airport?

10.3 stone is how many pounds

5 gallons equals how many quarts

How tall is Adam Brody in feet and inches?

How tall is ben affleck in feet and inches?

Percy Bysshe Shelley cause of death

QUESTIONS LIKE YOURS:

Is the Statue of Liberty in america?

What is the distance between Brussels South Charleroi Airport and the Statue of Liberty?

You asked: how tall is the statue of lib

93 meters is the height of the Statue of Liberty.

the height 93 meters

wikipedia

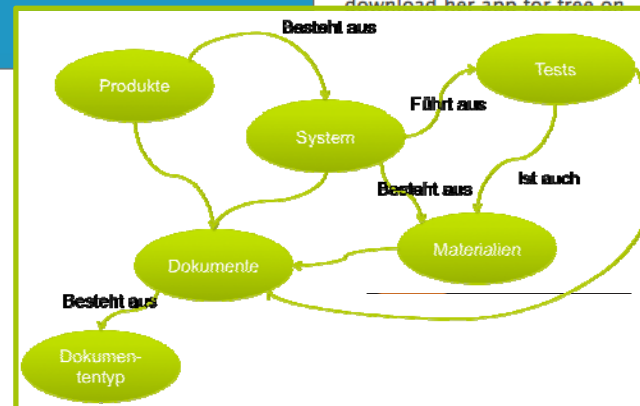
4. Ableiten von Wissens

2. Wissensrepräsentation



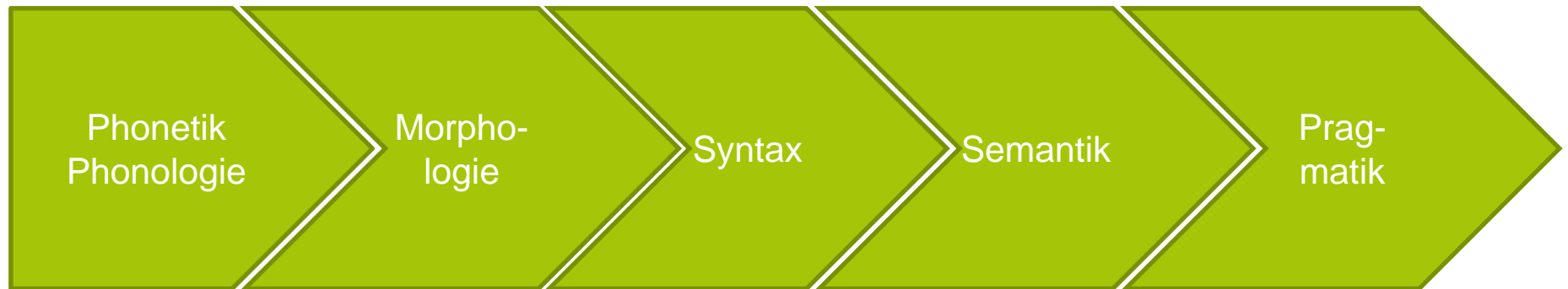
weather, books, music, films, people and places, recipe ideas, shopping and much more. Over the next few months we will be adding all of Evi's power to this site.

to experience all of the power of Evi you can download her app for free on



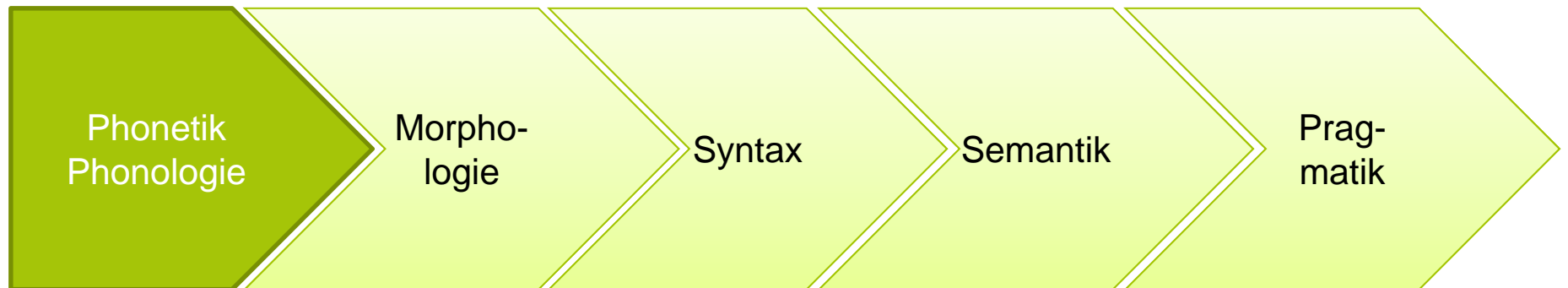
NLP

- Natural Language Processing



NLP

- Natural Language Processing

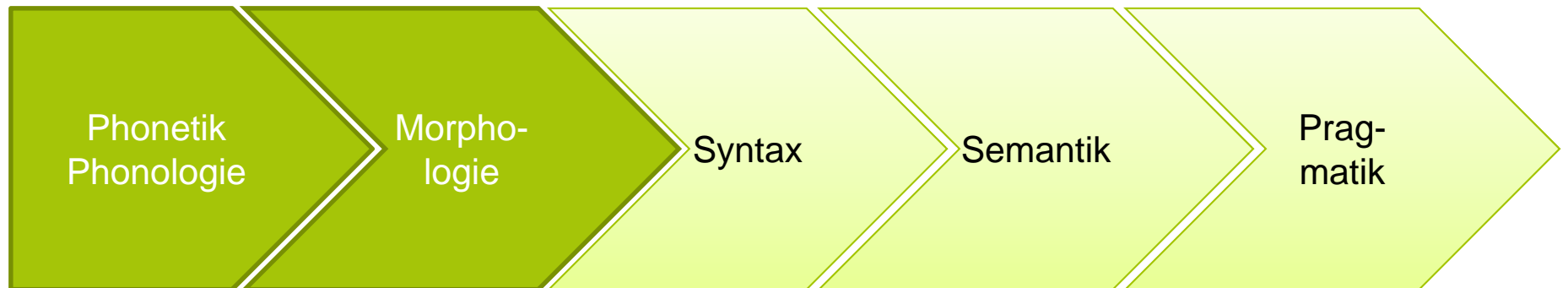


Phonetik und Phonologie

- Ziel: Erkennung von gesprochenen Wörtern
- Analyse der artikulatorischen Merkmale, sowie Lautstruktur
 - Ziel
 - Welche Segmente enthält ein Wort
 - Wie wirkt sich die Struktur auf die Aussprache aus
 - Beispiel: Dieb vs. Diebe
[dieb] vs [diebe]

NLP

- Natural Language Processing



Morphologie

- Am Anfang war das Wort!
- Beschreibt die Bildung und Struktur von Wörtern
- Untersucht wird die lexikalische Wurzel einzelne Wörter haben.
- Beispielsweise in der Deutschen Sprache
 - Das Suffix –e ist der Plural
 - Diebe => Dieb[PL]
 - -> Mehr als ein Dieb

Subprozesse



Subprozesse



Tokenisierung

- Segmentierung der Texte in linguistische Einheiten, wie
 - Wörter, Phrasen, Sätze, Absätze
 - Scheint trivial:
 - Man analysiert lediglich Leerzeichen, Satzzeichen
 - Problem:
 - Japanische Schriftzeichen
 - Dr.
 - Joop!
 - Guess?
 - 12.356.89
 - Abkürzungen

NLP

Tokenisierung
Wie
groß
ist
die
Freiheits-
Statue
in
New
York.

Subprozesse



Stemming

- reduziert den Term zu einer Art "Wortkern", der nicht unbedingt in der Sprache existiert.
 - Freiheit -> frei
 - Freizügig -> freizüg
 - Launisch -> laun

Lemmatisierung

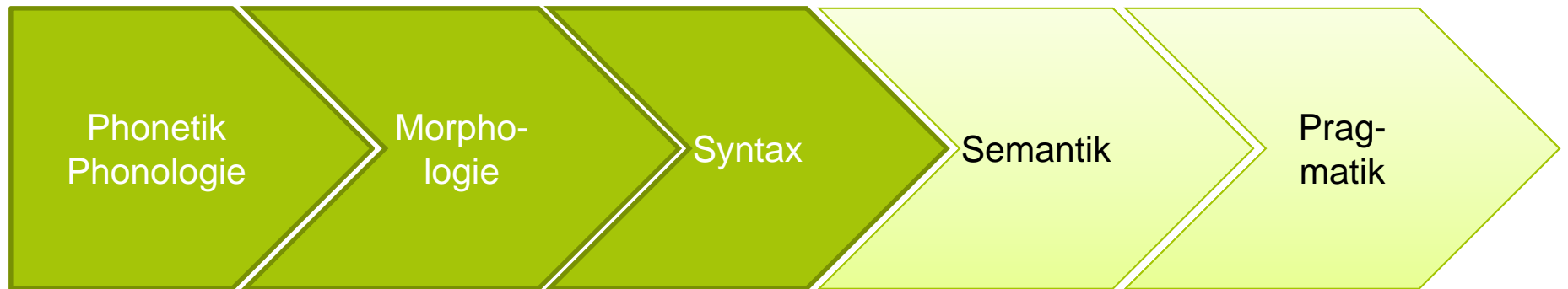
- Reduzierung eines Worts auf seine Grundform (und weitere Information) durch Nachschlagen in einem elektronischen Wörterbuch
- Freiheiten -> Freiheit (en -> 0)
- Umsätze -> Umsatz (ä -> a)
- lässt -> lassen (ä -> a)
- Buches -> Buch (es -> 0)
- buchte -> buchen (te -> en)

NLP

Tokenisierung	Lemma-tisierung
Wie	
groß	
ist	sein
die	
Freiheits-	Freiheit
Statue	
in	
New	
York.	

NLP

- Natural Language Processing



Syntax

- Kombination von elementaren Zeichen (Wörtern) zu zusammengesetzten Zeichen.
 - Segmentierung
 - Zerlegung eines komplexen Ausdrucks in Teile
 - Kategorisierung der komplexen Teilausdrücke
 - Subjekt, Prädikat, Objekt
 - Lineare Abfolge

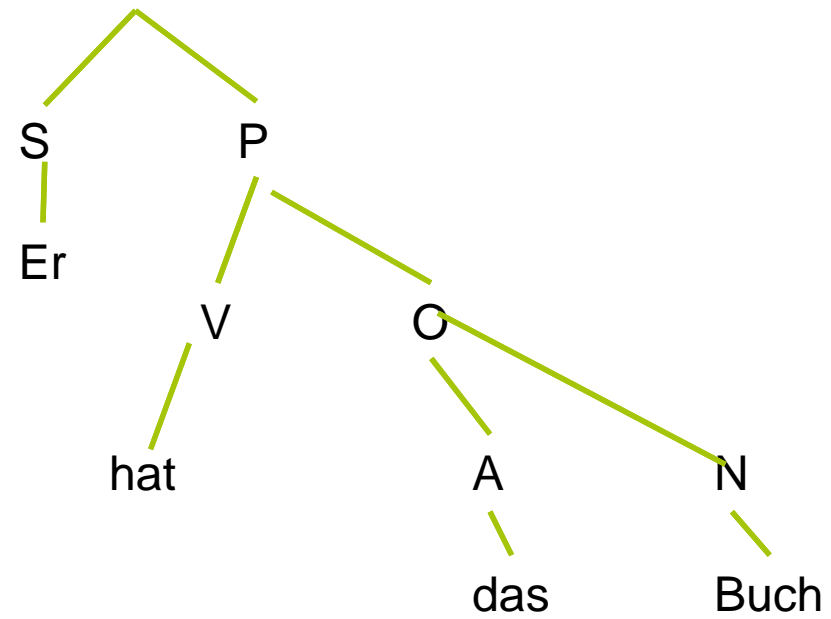
Part of Speech Tagging

- Anreicherung um grammatikalischen Informationen
- Klassifizierung gemäss seiner Wortart
 - Der Laster wiegt schwer. (adverbiales Adjektiv)
 - Er ist schwer. (prädikatives Adjektiv)
 - Der schwere Mann kam herein. (attributiertes Adjektiv)

NLP

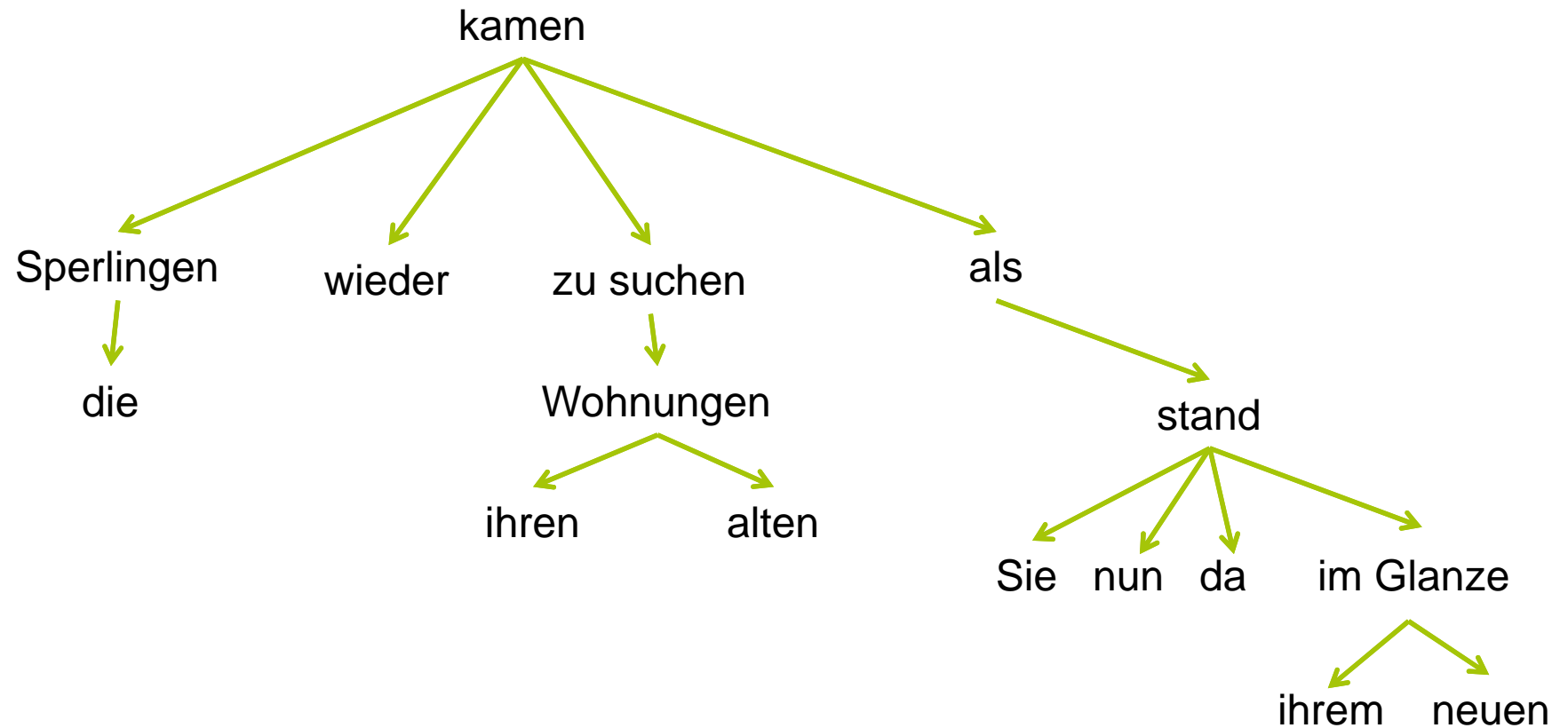
Tokenisierung	Lemma-tisierung	PoS
Wie		Fragewort
groß		Adjektiv
ist	sein	Verb
die		Artikel
Freiheits-	Freiheit	Nomen
Statue		Nomen
in		Präposition
New		Nomen
York		Nomen

Er hat das Buch.



Konstituentenstruktur-Baumgraph nach Dittrich

Als sie nun in ihrem alten Glanze da stand, kamen die Sperlinge wieder, ihre alte Wohnungen zu suchen.



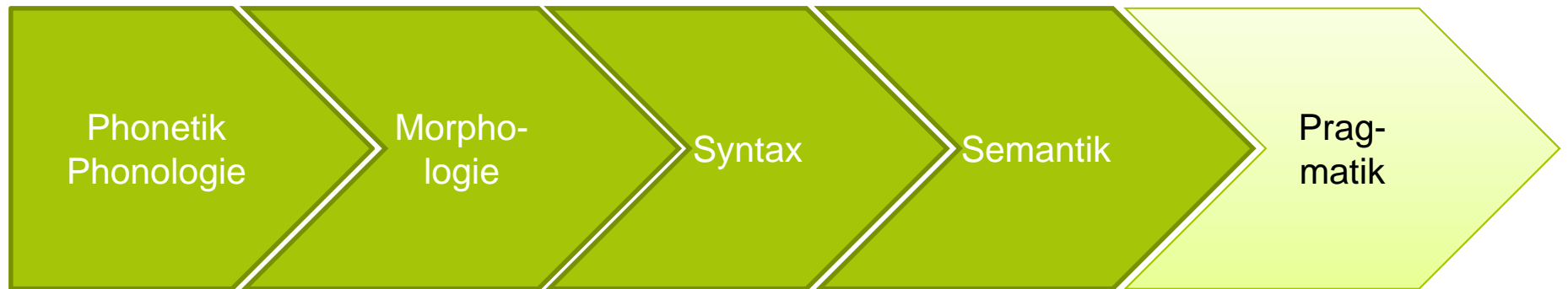
Dependenz-Baumgraph nach Kern

NLP

Tokenisierung	Lemma-tisierung	PoS	Phrase
Wie		Fragewort	
groß		Adjektiv	
ist	sein	Verb	
die		Artikel	
Freiheits-	Freiheit	Nomen	Freiheitsstatue
Statue		Nomen	
in		Präposition	
New		Nomen	New York
York		Nomen	

NLP

- Natural Language Processing



Semantik

- Bedeutungslehre

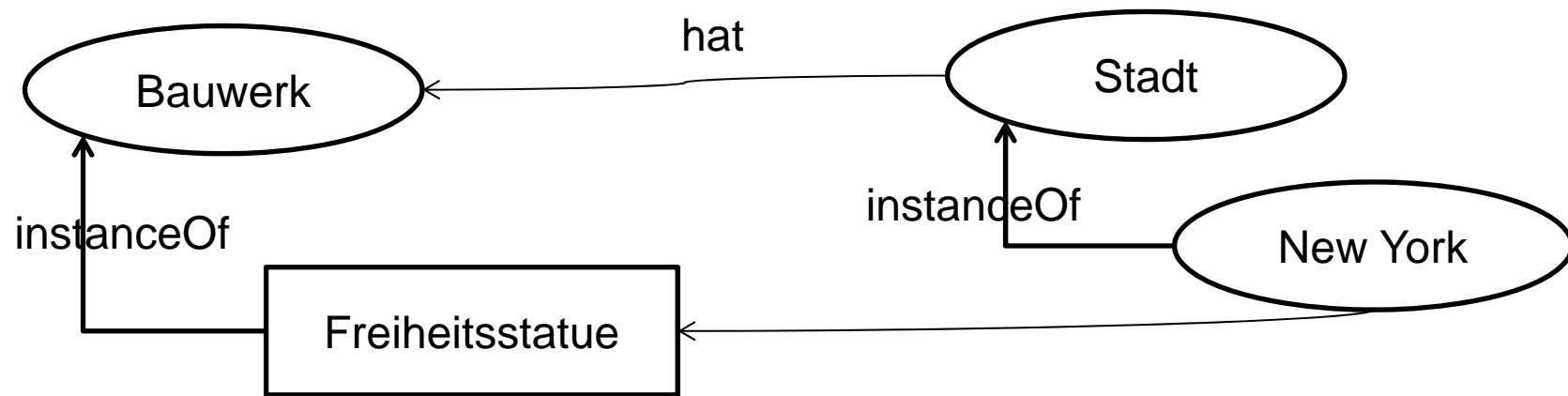


Synonyme

- Zur Vollständigkeit der Suche
 - Hinzufügen von alternativen Begrifflichkeiten
 - Durch Nutzung von Thesauri

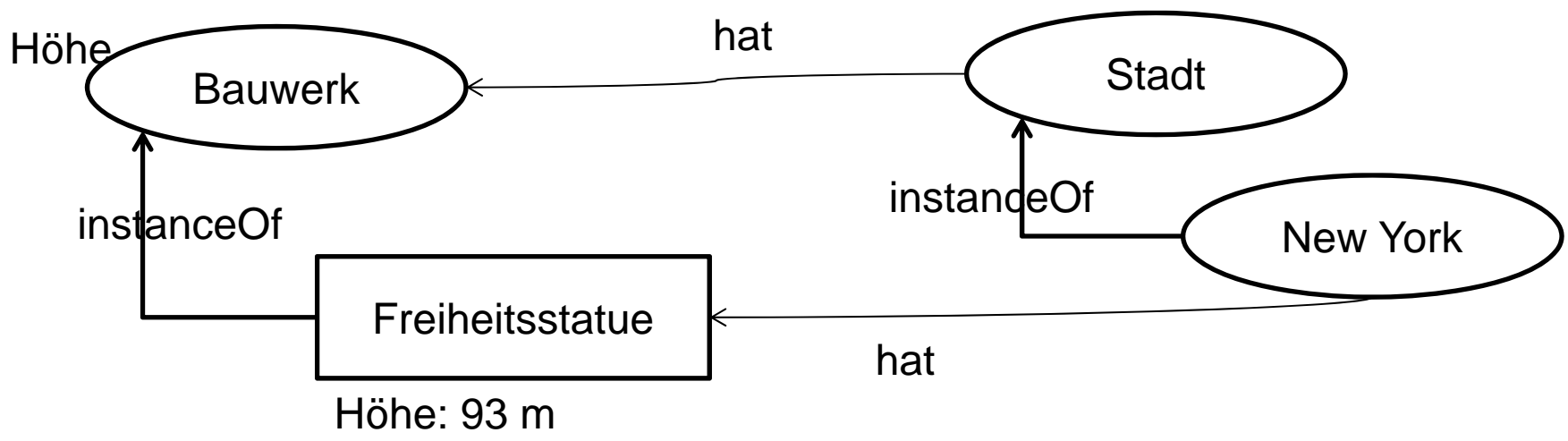
NLP

Tokenisierung	Lemma-tisierung	PoS	Phrase	Synonyme
Wie		Fragewort		
groß		Adjektiv		Breit, hoch, mächtig, riesig
ist	sein	Verb		
die		Artikel		
Freiheits-	Freiheit	Nomen	Freiheitsstatue	
Statue		Nomen		
in		Präposition		
New		Nomen	New York	
York		Nomen		



NLP

Tokenisierung	Lemma-tisierung	PoS	Phrase	Synonyme
Wie		Fragewort		
groß		Adjektiv		Breit, hoch, mächtig, riesig
ist	sein	Verb		
die		Artikel		
Freiheits-	Freiheit	Nomen	Freiheitsstatue	Bauwerk
Statue		Nomen		
in		Präposition		
New		Nomen	New York	Stadt
York		Nomen		



NLP

- Natural Language Processing



Pragmatik

- Erkennung von Beziehung zwischen aufeinander folgender Sätze
- Beispiel:
 - Dialog => Frage -> Antwort
 - Im Diskurs =>
 - Eine Aussage -> ihre Begründung
 - eine Aussage -> ihre Einschränkung

Danke für die
Aufmerksamkeit