

Information Retrieval
Prof. Dr. N. Fuhr

Sascha Kriewel
sascha.kriewel@uni-duisburg.de

Abgabe bis 23. Juni 2004

Übungsblatt Nr. 7

Aufgabe 1: Informationsextraktion – Lexikalische Verarbeitung

- (a) WordNet (<http://www.cogsci.princeton.edu/cgi-bin/webwn>) ist ein online verfügbares lexikalisches Referenzsystem für englische Wörter. Nimm den folgenden Beispielsatz aus einer Reuters-Meldung und füttere WordNet mit einigen Wörtern aus dem Satz. Betrachte und beschreibe welche Art von Information verfügbar ist. Wie könnte man diese Information zur Informationsextraktion nutzen? Siehst Du Probleme bei der Nutzung von WordNet für die Informationsextraktion?

Beispielsatz: „A draft report compiled by an international anti-money laundering group praises Saudi Arabia’s efforts to crack down on the flow of funds from the kingdom’s charities to possible terrorist groups.” (Reuters, Mon Jun 14, 2004 08:25 PM ET)

- (b) Überlege wie man Google, oder einen anderen Websuchdienst, für das Zusammenführen von Teilresultaten nutzen könnte. Tipp: Bestimmung der Häufigkeit der *co-occurrence* von Termen oder Phrasen.

15 Punkte

Aufgabe 2: Informationsextraktion

Es sei die folgende Reuters-Meldung gegeben, sowie ein Schema für die Daten aus Auftragsvergabe-Nachrichten. Analysiere manuell Schritt für Schritt den Text. Gebe für die Schritte die jeweiligen Zwischenergebnisse an. Gebe auch an, welches Zusatzwissen Du gegebenenfalls zur Analyse zur Hilfe ziehen musstest.

Schritte:

- (a) Tokenization
- (b) Morphologische Verarbeitung
- (c) Part of Speech Tagging
- (d) Namenserkennung
- (e) Syntaktische Analyse
- (f) Auflösung von Koreferenzen
- (g) Auffüllen der Schemata
- (h) Zusammenführung („Mischen“) von Ergebnissen

EADS und Dassault entwickeln Aufklärungsdrohnen für Frankreich

Paris (Reuters) - Der europäische Luft- und Raumfahrtkonzern EADS hat vom französischen Verteidigungsministerium einen Auftrag zum Bau von unbemannten Aufklärungsflugzeugen bekommen.

Die EADS sei Hauptauftragnehmer bei dem 300-Millionen-Euro-Projekt, teilte die französische Verteidigungsministerin Michelle Alliot-Marie am Montag in Paris mit. Ebenfalls am Bau beteiligt seien die französischen Rüstungsfirmen Dassault Aviation und Thales.

[...] Das Projekt „EuroMale“ konzentrierte sich auf den Einsatz so genannter Drohnen, die 24 Stunden in der Luft bleiben und Distanzen von bis zu 4500 Kilometer zurücklegen könnten, hiess es. Bis 2008 sollen die Maschinen entwickelt sein.

(Reuters, Montag 14 Juni, 2004 10:23 CET)

Schema für Auftragsvergabe-Nachrichten:

Auftraggeber:

Auftragsempfänger:

Auftragsvolumen:

Auftragsgegenstand:

Auftragszeitraum:

Bekanntgabe:

– Wer?

– Wann?

– Wo?

15 Punkte