

# Vorhersage von ökonomischen Indikatoren aus Textnachrichten

---

## Betreuer:

Artur Andrzejak: mail: [andrzejak@zib.de](mailto:andrzejak@zib.de)  
ZIB, Raum 3152

## Zielgruppe:

Dieses Projekt richtet sich an Studenten im Hauptstudium, die eine Vertiefung im Bereich Data Mining wünschen, und zugleich Interesse an wissenschaftlichen Arbeiten haben. Erfahrung beim Programmieren in Java wird vorausgesetzt.

## Motivation:

Ökonomische Indikatoren wie Index der Konsumentenstimmung, erwartete Preisentwicklung, Umsatz des Handels etc. werden heutzutage entweder durch Umfragen oder komplexe Verfahren der Zeitreihenanalyse vorausgesagt. Allerdings lassen sich erfahrungsgemäß viele dieser Größen aus den Nachrichten in Textform (z.B. aus Internet) zumindest partiell vorhersagen. Die Verfahren der semantischen Textanalyse könnten bei einer solchen Modellierung hilfreich sein, diese sind jedoch sehr komplex. Deshalb wäre es interessant zu untersuchen, inwieweit solche Vorhersagen durch die viel einfachere Analyse der *relativen Häufigkeit von bestimmten Begriffen und Ausdrücken* wie „Inflation“, „erhöhte Insolvenzrate“ etc. gewonnen werden können.

In diesem Projekt soll eine Data Mining Studie durchgeführt werden, die aus den in Internet verfügbaren textuellen Nachrichten einige solcher Indices modelliert und vorhersagt. Dabei sind folgende Teilaufgaben zu behandeln:

- Erstellung eines Java-basierten Frameworks für die automatische Modellierung und Vorhersage von Zeitreihen anhand der Häufigkeitsanalyse von Textausdrücken
- Durchführung einer Data Mining Studie mit der Selektion der relevanten Nachrichtenquellen, der relevanten Begriffe, Auswahl der Modelle, und der Evaluation.

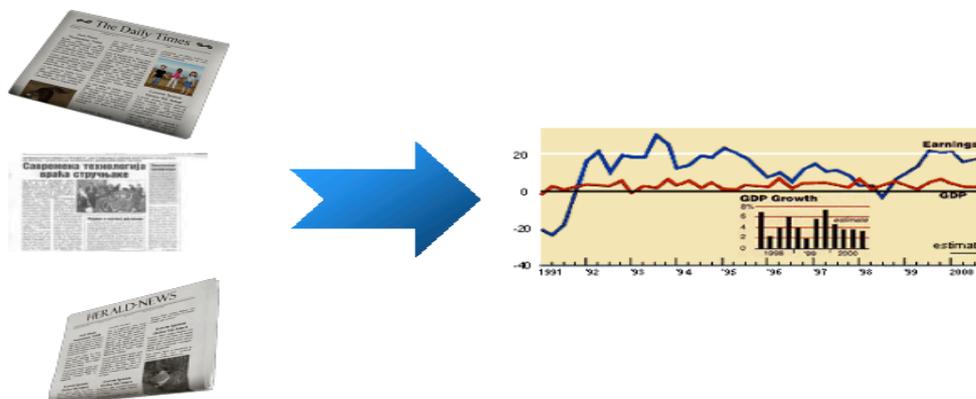


Abbildung 1: Vorhersage von ökonomischen Indikatoren durch Häufigkeitsanalyse von Ausdrücken in Textnachrichten

**Projektziele:**

Das Ziel des Projektes ist es, ein Framework auf Basis der existierenden Software (wie weka, MetaLooper, und andere) für die oben genannte Modellierung/Vorhersage zu erstellen, sowie eine Data Mining Studie durchzuführen. Das Framework enthält folgende Funktionalitäten:

- Download von HTML-Seiten und Extraktion von Text aus diesen Seiten
- Häufigkeitsanalyse von Ausdrücken in den Seiten, inklusive Herausfiltern von irrelevanten Füllwörtern
- Implementierung eines Interfaces zu dem Weka-Framework für Data Mining Algorithmen
- Integration von Attributselektionsmethoden für die Auswahl von signifikanten Quellen und Ausdrücken
- ggf. Implementierung von speziellen Miningalgorithmen.

Die Data Mining Studie soll umfassen: die Vorbereitung der Daten, Selektion der relevanten Quellen/Merkmale, und die Modellierung der Zielgrößen.

*Bei erfolgreichem Abschluss des Projektes wird die Veröffentlichung der Ergebnisse in einem wissenschaftlichen Artikel angestrebt.*

**Durchführung des Projektes:**

Das Projekt soll von einer kleinen Gruppe durchgeführt werden. Start und Ende des Projektes können mit dem Betreuer flexibel vereinbart werden, genauso wie die regelmäßigen Projekttreffen.

Das Projekt beginnt mit einer kleinen Vortragsreihe, in der die Projektteilnehmer kurze Vorträge über verschiedene Grundlagen zu den Themen Selektion der Attribute, Mining von Assoziationsregeln, und andere.