

# A U S H A N G

---

## FREIE UNIVERSITÄT BERLIN Fachbereich Mathematik und Informatik

Promotionsbüro, Arnimallee 14, 14195 Berlin

## DISPUTATION

**Freitag, 27. Juni 2014, 14 Uhr c.t.**

**Ort: SR 006, Takustr. 9, 14195 Berlin**

**Disputation über die Doktorarbeit von**

**Herrn Masood Ghayoomi**

**Thema der Dissertation und der Disputation:  
From HPSG-based Persian Treebanking to Parsing:  
Machine Learning for Data Annotation**

Die Arbeit wurde unter der Betreuung von **Prof. Dr. R. Rojas** durchgeführt.

Abstract: In this talk, we propose using machine learning methods for treebanking from scratch in order to syntactically analyze the Persian language. We first propose a bootstrapping approach to initialize the data annotation process by extracting frequent rules from the sentences. After developing a small data set, we use this data to train a statistical constituency parser. In the next step, we aim at reducing the sparseness of data at the lexical and syntactic construction levels. For the former problem, we propose an unsupervised word clustering approach to provide a more coarse-grained representation of the lexical items. For the latter problem, we propose using active learning to increase the size of the training data by selecting informative samples from the data pool with minimum redundancy. We show how this approach enhances parsing performance with small amount of further annotated.

Die Disputation besteht aus dem o. g. Vortrag, danach der Vorstellung der Dissertation einschließlich jeweils anschließenden Aussprachen.

**Interessierte werden hiermit herzlich eingeladen**

Der Vorsitzende der Promotionskommission  
Prof. Dr. R. Rojas