

## **Protokoll für das Repetitorium vom 29.5.2017**

### **Practical Significance Testing for Experiments in Natural Language Processing**

Im Vortrag dieser Sitzung ging es um das Testen von Systemen, insbesondere um Signifikanztests. Zu Beginn ging Dr. Roth deshalb noch einmal auf die notwendigen Datensätze ein die prinzipiell notwendig sind um ein Programm zu trainieren und zu evaluieren und was ihre Funktion ist. Namentlich sind diese Training-, Development- und Testset. Im Anschluss ging Dr. Roth genauer auf Signifikanztests ein. Hierbei geht es darum festzustellen ob wenn zwei Systeme unterschiedliche Ergebnisse liefern, diese aussagekräftig sind oder ob es sich dabei nur um Zufall handelt. Danach ging er genauer auf das Verfahren ein die Signifikanz zu berechnen. Daraufhin fragte er die Studenten welche Metriken sie in ihren Arbeiten benutzen. Da sich herausstellte dass die meisten mit Accuracy arbeiten erklärte Dr. Roth wie man diese berechnet und zeigte auch eine Python Methode dafür. Danach erklärte er noch kurz wie man Mean Average Precision berechnet und stellte auch dafür eine Python Methode vor. Abschließend wurden noch einige Gefahren beschrieben die bei Signifikanztests entstehen können und wie man sie vermeiden kann.