**FINAL REPORT: EFFECTS OF CORPUS MODIFICATION ON TOPIC MODELING**

**OUTLINE:**

* **ABSTRACT**
  + Research question
  + Findings
  + Contributions
* **1 INTRODUCTION**
  + Research questions (SYSTEM TASK ENVIRONMENT)
    - What effect does modifying the input corpus have on the performance of topic modeling?
    - What variety of corpora are seen in the real-world? (Are our modifications meaningful when compared to real corpora?)
  + Hypotheses (briefly)
  + Findings
* **2 RELATED WORK**
  + Frontier
  + State of existing knowledge
  + State of behaviour of the system (We know it behaves like…)
* **3 EXPERIMENT DESIGN**
  + Corpus modification
  + …
* **4 RESULTS**
  + Independent modification
  + Modification in combination
* **5 CONCLUSIONS**
  + Interpretation of results
  + Contributions
* **ACKNOWLEDGEMENT**
* **REFERENCES**

Our aim is to show that modifying the input vocabulary does affect the possibility to effectively model topics within a corpus. Since it is well-understood that the underlying phenomenon behind the success of topic modeling is word co-occurrence, and we are

However, we don’t just want to see any changes. We want to see interesting changes. Therefore, our goal is to observe that there is a bound on the increase in topic quality that comes from increasing the availability of documents.