

**Data Science 2**

Blatt 5, Abgabe am 20.11.2024 um 12:00

---

Bitte geben Sie eine PDF-Datei ab.

**Aufgabe 1.** *Auf den vorigen Übungsblättern haben Sie Implementierungen von MinHash+LSH geplant und durchgeführt. Rückblickend,*

- 1. was ist Ihnen am schwersten gefallen / was hat gar nicht geklappt?*
- 2. Welche Designentscheidungen haben Sie getroffen und auf welcher Grundlage?*
- 3. Was hat Sie überrascht?*
- 4. Wobei/woraus haben Sie am meisten gelernt?*
- 5. Wie/was können Sie in Zukunft besser planen bzw. was würden Sie im Nachhinein anders angehen?*

*Sie können stichpunktartig antworten oder die Antwort in einen knappen Projektbericht formulieren.*

**Aufgabe 2.** *Betrachten Sie die API Demo*

*<https://stream.wikimedia.org/v2/ui/#/?streams=mediawiki.recentchange>  
und überlegen Sie, was Sie damit anfangen könnten (in 2-3 Sätzen).*

**Aufgabe 3.** *Führen Sie eine explorative Datenanalyse der beiden gerichteten Graphen `web-Google.txt.gz` (Google's web graph toy example) und `wiki-topcats.txt.gz` (2011 top Wikipedia articles) durch. Wir werden später auf diesen Graphen besonders wichtige Knoten berechnen (über Zentralitätsmaße wie den PageRank). Auch diese Seite mit Metadaten könnte hilfreich sein:*

*<https://snap.stanford.edu/data/wiki-topcats.html>*