

## IRLab Plus – Information Retrieval Experimentierkasten

Den Studierenden an der ZHAW stehen für das bessere Verständnis von Information Retrieval-Systemen im Modul «Information Engineering 1» zwei Anwendungen zur Verfügung. Diese können aktuell nur losgelöst voneinander bedient werden, wobei die eine Anwendung aufgrund des technologischen Fortschritts nicht mehr weiterentwickelt werden kann und die zweite nur vorberechnete Ergebnisse liefert. Daher ist es notwendig, eine zeitgerechte und für den Lernerfolg optimierte Anwendung zu erstellen.

Das Ziel dieser Bachelorarbeit ist die Entwicklung einer experimentzentrierten Anwendung, die den Lernprozess der Studierenden unterstützt. Für die Überprüfung dieses Ziels wurde ein reales, wissenschaftliches Experiment durchgeführt. Dazu wird die folgende Forschungsfrage gestellt: Hat das Erweitern der Abfrage mit Synonymen einen positiven Einfluss auf ein Suchresultat?

Um die zu erstellende Anwendung neu zu konzipieren, wurden die Funktionalitäten der bestehenden Anwendungen analysiert und mit zusätzlichen Funktionen angereichert. Spezifisch wurde vor allem der komplette Arbeitsfluss der Studierenden neu modelliert.

Um die Forschungsfrage zu beantworten, wurden auf Basis von 50 Millionen Webseiten mehrere Experimente erstellt. In den einzelnen Experimenten wurden unterschiedliche Arten der Abfrageerweiterung mit Synonymen vorgenommen. Ein Experiment wurde automatisch mit Synonymen und zwei weitere wurden durch manuelle Arbeit mit Synonymen und weiteren kontextsensitiven Begriffen angereichert. Es wurde untersucht, wie die einzelnen Experimente im Vergleich zur Baseline abschneiden.

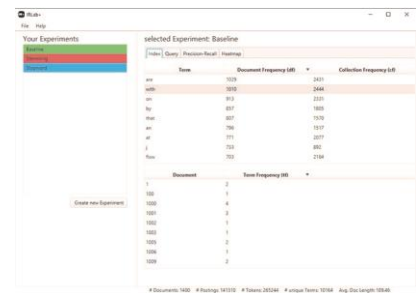
Die Durchführung der Experimente hat gezeigt, dass die neue Anwendung erfolgreich für Experimente eingesetzt werden kann.

Die Untersuchung der Resultate hat ergeben, dass die Abfrageerweiterung mit Synonymen in den meisten Fällen einen negativen Einfluss auf das Gesamtergebnis hat. Obschon für einzelne Abfragen ein besseres Resultat erzielt werden konnte, zeigt dies auf, dass die Synonyme einer Abfrageerweiterung präzise ausgewählt werden müssen. Diese Diskrepanz ist unter anderem auf die Kontextsensitivität einer Sprache sowie auf die Mehrdeutigkeit einzelner Worte zurückzuführen.



Diplomierende  
Pascal Brunner  
Dante Stoppini

Dozent  
Martin Braschler



Das IR-Lab+ ermöglicht die Untersuchung des aufgebauten Index, indem die Anwendenden nach der Dokumenten-Häufigkeit oder der Term-Häufigkeit sortieren können.



Die Heatmap ermöglicht den direkten Vergleich zwischen den unterschiedlichen Experimenten. Beim Klicken auf einen Wert öffnet IRLab+ die darunterliegende Rangliste.