

Protokoll vom 05.11.2008

Als Erstes wurde festgelegt, dass es wichtiger ist, dass die Algorithmen austauschbar sind. Das die Algorithmen optimal arbeiten ist hier zweitrangig, da die Anzahl der Algorithmen relativ groß ist. Wichtiger ist, dass der Benutzer sehen kann, was welcher Algorithmus bewirkt.

Damit kommt auch der grobe Grundriss für das zu erstellende Programm zustande. Das wichtigste ist die Visualisierung des Inverted Index mit Hilfe von Swing. Algorithmen sollen, wie oben schon angedeutet, einfach als Plug-Ins nachzuladen sein. Die Daten, die verarbeitet und dargestellt werden sollen, sowie die Daten, die unter Umständen während der Verarbeitung entstehen sollen in einer beliebigen Datenbank gespeichert werden (hier haben wir uns für MySQL entschieden, aber die Schnittstelle sollte auf jeden Fall leicht zu modifizieren sein, um andere Datenbanken einzubinden).

Grundsätzlich soll das Programm die Positionen der einzelnen Dokumente bei Eingabe eines *Terms* darstellen. Werden mehrere *Terms* eingegeben, soll die Position in den einzelnen Listen dargestellt werden und ein Zusammenhang dargestellt werden zu der Position des Dokumentes für den kombinierten Begriff, zum Beispiel durch Farben. Das Ganze soll nur passieren für eine gewisse Anzahl von Ergebnissen, zum Beispiel die Top10 (evtl. variabel). Ein Vorschlag um solche Informationen darzustellen, war eine Art Tooltip, der angibt, welche Position die Dokumente für die einzelnen Terms hatte und Ähnliches.

Somit ist klar, was die Hauptaufgabe des Programms sein wird: Es soll ein Überblick über einen Index visualisiert werden, um zum Beispiel die Auswirkungen von neuen Dokumenten im Index darzustellen. Oder aber auch wie oben erwähnt die Änderung des Algorithmuses.

Auch soll das Programm nicht nur auf einen bestehenden Index zu greifen, sondern auch einen neuen erzeugen können. Das bedeutet, dass man eine Funktion braucht um einen bestehenden Index einzulesen und darzustellen und eine weitere um eine neue Sammlung zu erzeugen und zu organisieren.

Es wurde auch noch einmal darauf hingewiesen, dass eine Dokumentation für spätere Weiterentwicklungen unablässig ist.

Desweiteren soll als Erstes Englisch als Sprache der zu durchsuchenden Dokumente angenommen werden.

Alles in Allem sollen wir zuerst ein Basisprogramm aufbauen, das leicht zu modifizieren ist, um später neue Ideen z.B. der Visualisierung einzubringen, aber auch um z.B. neue Algorithmen zu testen und einzubringen.