

Die smarte Rechercheplattform



Modul: Softwareprojekt (Master)

Team: Mathias Bredereck (PI), Edmond Kojok (PI), Stephan Rübner (PI)



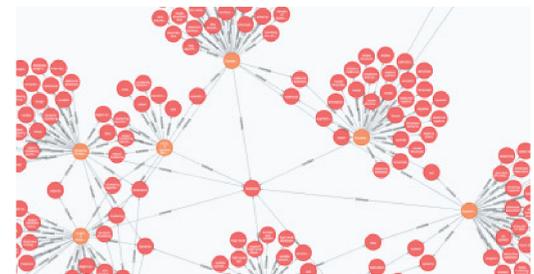
Recherche-Chaos

Problemstellung

- Wissenschaftliches Arbeiten wird stets begleitet von Literaturrecherchen.
- Das Finden passender Literatur gestaltet sich zunehmend als Herausforderung.
- Welche der vielen kostenlosen und kostenpflichtigen Plattformen soll man nutzen?
- Auch das Finden eines geeigneten Themas gestaltet sich meist als Herausforderung.
- Hat man erst ein interessantes Thema, gibt es dann auch passende Literatur?
- Nach welchen relevanten Suchbegriffen kann man suchen?
- Das Ergebnis ist meist eine mehr oder weniger zielgerichtete Recherche.

Idee und Konzept

- Eine Online-Plattform, welche alle Paper **kennt**, nach denen man sucht
- Alle Themen, alle Literatur, alle Plattformen **auf einen Blick**
- Kein Speichern der Paper, sondern **Analyse und Extraktion der Schlüsselwörter** aus dem Volltext der wissenschaftlichen Artikel
- **Interaktives Suchen** von Papern und **interaktives Explorieren** von Themen



Graph-Ansicht in der Neo4j Datenbank

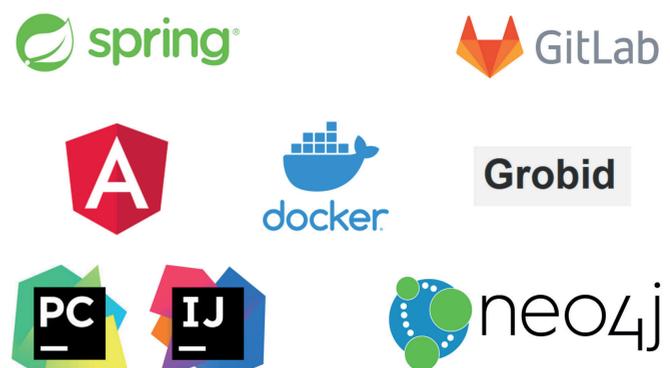
Technische Umsetzung

Die Webanwendung für PaperHub wurde auf dem verbreiteten Java-Framework Spring aufgesetzt. Als Datenbank nutzt PaperHub die graphbasierte Datenbank Neo4j, da die wissenschaftlichen Artikel und weiteren Eigenschaften stark verknüpft sind, und sich so ein „Netzwerk aus Wissen“ gut abbilden lässt. Die Weboberfläche wurde für gute Benutzbarkeit und Erweiterbarkeit in Angular erstellt. Im Hintergrund werden Anbieter von wissenschaftlichen Arbeiten kontinuierlich abgefragt, um die

Datenbasis zu erweitern und aktuelle Artikel und Veröffentlichungen auf Hub anbieten zu können. Hierfür wurde ein erweiterbarer Python-Crawler geschrieben, der sowohl APIs von Anbietern abfragen kann, als auch deren Webseiten crawlt, wenn keine API angeboten wird. Über moderne Techniken wie YAKE und Grobid werden Schlagwörter und weitere Metadaten direkt aus dem Volltext der Artikel ausgelesen und in der Datenbank gespeichert. So wird jede Veröffentlichung durchsuchbar gemacht und die jeweiligen Kernthemen ermittelt.



PaperHub - Artikel im „Netzwerk des Wissens“



Eingesetzte Technologien

Team

Mathias.Bredereck@studmail.w-hs.de
Edmond.Kojok@studmail.w-hs.de
Stephan.Ruebner@studmail.w-hs.de

Betreuung

Prof. Dr. Ulrike Griefahn
Fachgebiet: Programmiersprachen, Programmiermethodik und Web-Technologien