



«VUnit»

«Eidgenössische Stiftungsaufsicht»

Technische Informationen für die Jury

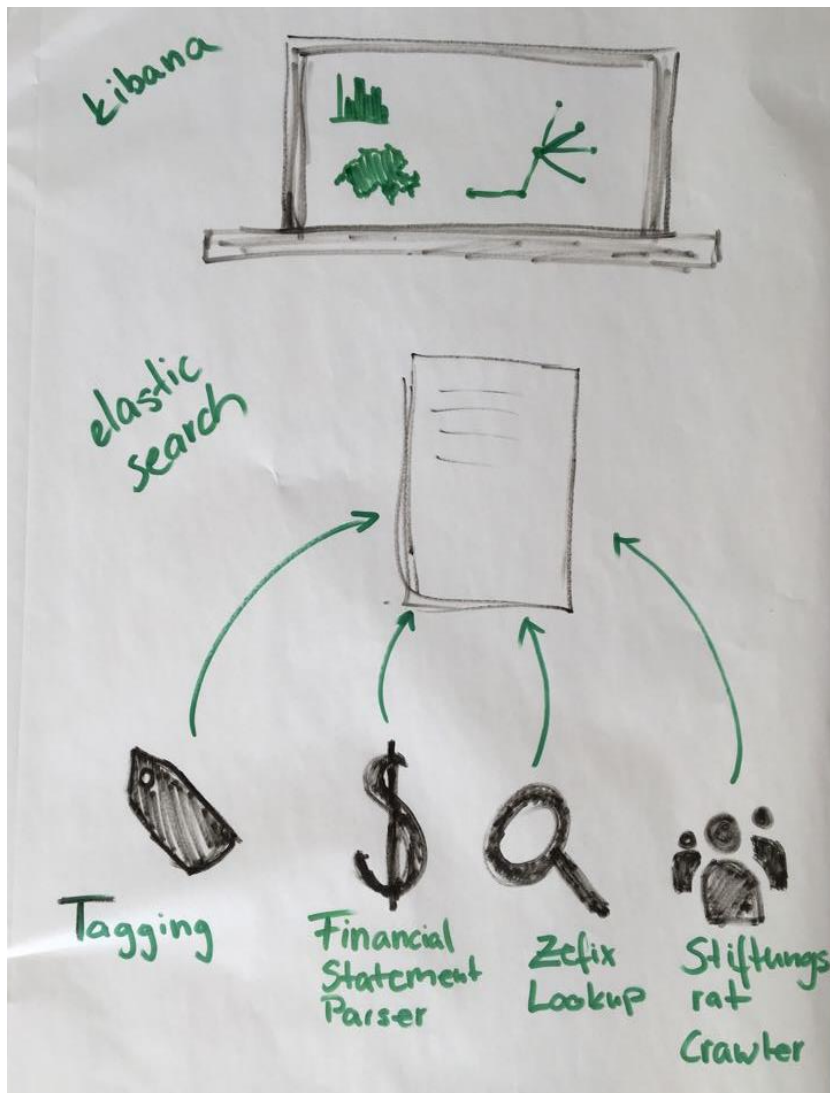


Aktueller Stand des Sourcecodes

<https://github.com/benwasd/baernhaeck18>

Fokus, Technische Grundlage

- Worauf habt ihr euch fokussiert?
 - Bereitgestellte Daten der ~ 4000 Stiftungen mit verschiedenen Techniken anreichern
 - Stiftungsratsmitglieder-Vernetzung mit einem Graph sichtbar machen
 - Geographische Dichte der Stiftungen visualisieren
 - Übersicht über die Zwecke der Stiftungen
- Welche technischen Grundsatzentscheide habt ihr gefällt?



- Kleine Crawler arbeiten parallel auf einem zentralen Datengefäß
- Elastic Search als Datencluster
- Visualisierungen in Kibana

Technischer Aufbau

- Welche Komponenten, Services und Frameworks habt ihr verwendet?
 - .NET Core, .NET Standard und .NET Framework
 - Aspose zum Verarbeiten von PDFs
 - Azure Cognitive Services, Text Analytics API
 - Chromium & CefSharp, zum Crawlen von interaktiven Webseiten
 - Elastic Search Cluster, hostet von Azure
 - Kibana, hostet von Azure
- Wozu und wie werden diese eingesetzt?
 - Bilanzsummen wurden evaluiert indem:
 - Crawlen von Jahresberichten/-rechnungen mittels Google Suchen
 - Parsen der PDF Jahresberichten/-rechnungen (Aspose)
 - Extrahieren nach Total Aktiven
 - Textanalytics zum Erkennen der Sprache der Zweckbeschreibung der Stiftung
 - Aus dem Stiftungszweck «Key Phrase» extrahieren um als Stiftungstags zu verwenden
 - Stiftungsratsmitglieder komplex aus einer Java Server Faces Webapplikation extrahieren, mangels API, Script Execution und daher wurde eine Browserplattform nötig
 - Elastic Search: Speicherung, Indexierung der Daten, Analyse von Keywords
 - Kibana: Visualisierung der Daten in Elastic Search

Implementation

- Gibt es etwas Spezielles, was ihr zur Implementation erwähnen wollt?
 - Schwierigkeiten lagen vor allem bei ZEFIX und Handelsregister Services, sowie Einschränkungen bei diversen APIs (Zeitliche Beschränkung der Abfragen, Blocks durch Google)
- Was ist aus technischer Sicht besonders cool an eurer Lösung?
 - Daten aus strukturierten und unstrukturierten Quellen werden aggregiert und können dann ausgewertet werden, «Big Data»
 - Cloud Lösung

Abgrenzung / Offene Punkte

- Welche Abgrenzungen habt ihr bewusst vorgenommen und damit nicht implementiert? Weshalb?
 - Keine komplette Analyse aller Zahlen in der Bilanz und Erfolgsrechnung, sowie Anhang.
 - Momentan keine neuronalen Netze trainiert, dafür waren zu wenig Daten vorhanden