

Crawling von Datenschutzerklärung-Historien

Zwischenpräsentation

Alexander Prull, Jörn-Henning Daug, Simon Kaleschke

Universität Leipzig

16. Dezember 2016

- 1 tmcrawl
 - Projektbeschreibung
 - Lösungsansatz
 - Softwarearchitektur
- 2 Ergebnisse und Probleme

Gliederung

- 1 tmcrawl
 - Projektbeschreibung
 - Lösungsansatz
 - Softwarearchitektur
- 2 Ergebnisse und Probleme

Projektbeschreibung

Motivation

- Aktuell geltende DSEs analysieren.
- Die Entwicklungsgeschichte von DSE betrachten.
- Trends und Veränderungen beobachten.

Aufgaben

- DSEs (täglich) extrahieren.
- Diese geeignet anzeigen.
- Unterschiede über Zeit darstellen.

DSE = Datenschutzerklärung

Gliederung

1 tmcrawl

- Projektbeschreibung
- **Lösungsansatz**
- Softwarearchitektur

2 Ergebnisse und Probleme

Lösungsansatz

Aufteilung

- Extraktion: Ruby, XPath, SQLite.
- Backend: Java, REST-Services.
- Frontend: HTML5, CSS3, JavaScript, AngularJS, Bootstrap.

Arbeitspakete

- Recherche (100 %)
- Grundgerüst (80 %)
- Feinschliff (0 %)

Gliederung

1 tmcrawl

- Projektbeschreibung
- Lösungsansatz
- Softwarearchitektur

2 Ergebnisse und Probleme

Extraktion

Website

Diff-Tool

Ergebnisse und Probleme