

Aufbau einer Infrastruktur für Information Retrieval- Evaluation

Ein Vortragsvorschlag für den 2.
Studenten-Workshop für
informationswissenschaftliche
Forschung

Nils-Peter Schirrmeister

13.10.2011

„Aufbau einer Infrastruktur für Information Retrieval Evaluationen“¹(im folgenden AIIRE) ist ein vom zfe der Hochschule Darmstadt Projekt, das zum Ziel hat, mithilfe von Open Source Software,-für den Bereich Informationswissenschaft eine Retrieval-Infrastruktur aufzubauen und den Zugriff auf informationswissenschaftlich relevante Testkollektionen zu ermöglichen.

Zukünftige Nutzer lassen sich in zwei Gruppen unterteilen. Einerseits Forscher, die Suchalgorithmen evaluieren wollen, andererseits Studierende, die die Grundlagen der Retrievaltechniken und Evaluation am System erproben wollen.

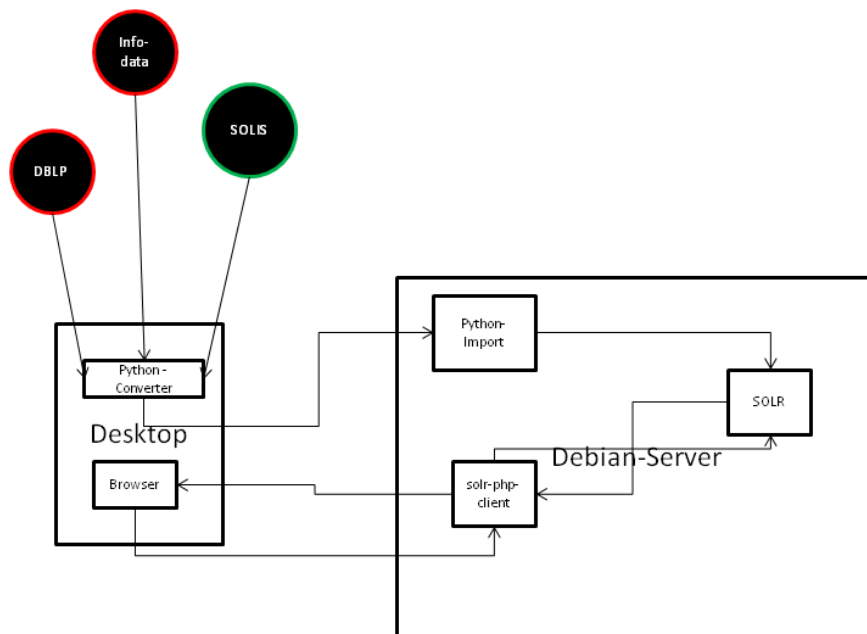



Abbildung 1 AIIRE Aufbau

¹ <http://aiire.media.h-da.de/>

Basis von AIIRE ist zunächst einmal Debian als Betriebssystem. Auf Debian wurde Apache SOLR² installiert. Apache SOLR ist eine Enterprise Search Engine, die unter anderem Volltextsuche und faceted search anbietet. Als nächstes muss das Schema von Apache SOLR definiert werden. Das Schema bestimmt den Aufbau der Datensätze und definiert wie den Dokumenten gesucht werden kann.

Der nächste Schritt besteht darin die Testkollektionen in das neue AIIRE-Schema zu überführen und in Apache SOLR einzuspielen. Die Daten kommen aus 3 Quellen: DBLP³, SOLIS⁴ und infodata⁵. Diese Datenbankinhalte müssen zunächst konvertiert werden. Dies geschieht mithilfe von Skripten, die in Python geschrieben sind. Das Einspielen in Apache SOLR geschieht serverseitig durch ein Zusammenspiel von Python und cURL⁶.



h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Search: *.*

Results 1 - 20 of 419565

Gesucht wurde nach *.*

Aufbau einer Infrastruktur für Information Retrieval-Evaluationen

Beta 2.0

id	iz-solis-90357956
pubyear	2005
collection	iz-solis
doctype	book
timestamp	2011-10-07T19:29:26.27Z
abstract	Aeltere Menschen werden bis heute als potenzielle Opfer sexueller Gewalt kaum wahrgenommen. Seit WissenschaftlerInnen beginnen, sich mit der Thematik zu befassen, wird zunehmend deutlich, dass auch ältere und hochbetagte Frauen von sexueller Gewalt betroffen sein können und dass diese Opfergruppe bislang von Hilfen nur sehr begrenzt erreicht wird. Das Kriminologische Forschungsinstitut Niedersachsen hat mit finanzieller Unterstützung des Niedersächsischen Landesamtes für Zentrale Soziale Aufgaben eine empirische Studie realisiert, deren zentrale Ergebnisse und Schlussfolgerungen hier vorgestellt werden. Im Einzelnen informiert die Broschüre über: Formen sexueller Gewalt, von denen ältere Frauen betroffen sind; Kontexte, in denen solche Gewalt auftritt; Institutionen, die von diesen Gewalterfahrungen bislang Kenntnis erlangen; altersspezifische Probleme bei der Bewältigung der Ergebnisse und dem Aufsuchen von Hilfe; Besonderheiten des Umgangs mit älteren Opfern in beratenden, helfenden und schützenden Institutionen. (Textauszug)
author	Goergen, Thomas Herbst, Sandra Naegete, Barbara Newig, Anja Kennelmeier, Ines Kottenga, Sandra Illid, Nico Pigors, Knea Rabold, Susann
subject	alter Mensch Gewalt soziale Einrichtung Bundesrepublik Deutschland Frau Sexualdelikt Ehe Familie Partnerschaft soziale Institution Opfer Vergewaltigung sexuelle Belästigung
maintitle	'Ich habe gehofft, das wird besser mit den Jahren'
subtitle	sexuelle Gewalterfahrungen älterer Frauen
mainclassification	soziale Probleme 20500
method	deskriptive Studie
language	de
subclassification	Kriminalsoziologie, Rechtssoziologie, Kriminologie Frauenforschung Gerontologie
id	iz-solis-90357957
pubyear	2005

Abbildung 2 Prototypisches Web-Interface AIIRE

Der nächste Schritt ist es die Daten über eine Websuche den zukünftigen Nutzern zur Verfügung zu stellen. Dazu wurde das Framework solr-php-client⁷ genutzt.

- ² <http://lucene.apache.org/solr/>
- ³ <http://www.informatik.uni-trier.de/~ley/db/>
- ⁴ <http://www.gesis.org/unser-angebot/recherchieren/solis/>
- ⁵ <http://www.infodata-edepot.de/>
- ⁶ <http://curl.haxx.se/>
- ⁷ <http://code.google.com/p/solr-php-client/>

Das Projekt steht noch am Anfang. Der Datenimport ist noch nicht zu Gänze abgeschlossen. Das Schema bildet noch nicht alle Felder ab, und das Webinterface erlaubt nur rudimentäre Suchen.

Doch sobald die Infrastruktur fertiggestellt ist, soll AIIRE für den Fachbereich Media(insbesondere Informationswissenschaften) und weitere interessierte Institutionen für Retrieval-Evaluation zur Verfügung stehen.