Das Projekt OAPS:

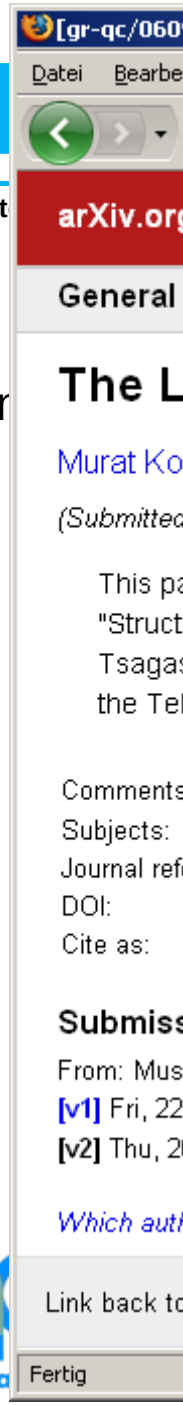
Plagiatsuche für Open Access

Dr.-Ing. Martin Gutbrod

Physikalisch-Technische Bundesanstalt, Braunschweig

Plagiatproblematik

- Aus rund 1000 Einreichungen der IEEE Konferenz Infocom 2006 wurden 12 nach automatischer Plagiatuntersuchung ausgeschlosser
  - Einbau in IEEE Konferenzmanagementsystem EDAS
- arXiv.org: 67 Dokumente 2007 wegen Plagiatismus entfernt
  - Beispiel Bildschirmfoto
- CrossRef: alle Dokumente, die ab Juni 2008 in ihren Dokumentenindex aufgenommen werden, werden per Plagiaterkennungssoftware geprüft
- Kopieren/Einfügen leicht gemacht
- (Fremd-)sprachliche Hürden bei Publikationen
- Kulturell anderer Umgang mit geistigem Eigentum
- Publikationsdruck (z.B. wegen Drittmitteleinwerbung)
- Eigenplagiat: Recyceln eigener Arbeiten verlängert eigene Publikationsliste (Reisen etc.)



Beispiel eines Prüfreports

Herkunftsreport

Digital signiert

Docol©c

Überprüftes Dokument: **test2.rtf**
Überprüft am: **Fri, 5.9.2008 23:58:56 CEST**

Es wurden insgesamt **25** Textstellen überprüft. Davon wurden **18** Textstellen (72,0%) in anderen Dokumenten gefunden. Die kritischen Textstellen wurden in der folgenden Dokumentenvorschau gelb markiert. Die Markierungen können angeklickt werden und zeigen daraufhin maximal 6 gefundene Quellen.

Referenzdokumente

Die folgende Übersicht ist gegliedert nach den Titeln der gefundenen Dokumente. Durch einen Klick auf „x Stellen" werden die speziellen Stellen im Dokument in der Farbe orange hervorgehoben und direkt zur ersten Stelle gescrollt. Ein erneuter Klick auf „x Stellen" setzt die Markierungen wieder zurück.

15 Stellen wurden gefunden in einer Textvorlage mit dem Titel: „***Plagiate: Abschreiben ein Kavaliersdelikt? - KARRIERE-Blog***", zu finden unter:
http://www.news.skripten.at/archives/65-Plagiate-Abschreiben-ein-Kavaliersdelikt.html
13% davon per Mischsuche (2 Stellen)

3 Stellen wurden gefunden in einer Textvorlage mit dem Titel: „***abschreiben /abschreiben***", zu finden unter:
http://www.news.skripten.at/plugin/tag/abschreiben
33% davon per Mischsuche (1 Stelle)

2 Stellen wurden gefunden in einer Textvorlage mit dem Titel: „***Plagiate verhindern!***", zu finden unter:
http://www.plagiate-verhindern.de/

Das Plagiat ist mit Hilfe von "Copy-Paste" ganz klar auf dem Vormarsch

Eine umfangreiche Literatur- und
Franz Schätzlein unter "Plagiate

Was aber ist eigentlich ein Plag

"Ein Plagiat (lat: plagium, Mens
bzw. eines fremden Werkes als ei
unter bestimmten Voraussetzungen strafbar und verpflichtet den Plagiator zum Schadenersatz." (Quelle: de.wikipedia.org)

2 Treffer:

Plagiate: Abschreiben ein Kavaliersdelikt? - KARRIERE-Blog Das alles war aber vergleichsweise harmlos, wenn man sich die Entwicklung des ... Das Plagiat ist mit Hilfe von "Copy-Paste" ganz klar auf dem Vormarsch. ...
http://www.news.skripten.at/archives/65-Plagiate-Abschreiben-ein-Kavaliersdelikt.html

abschreiben anschreiben bewerbung bewerbungsbrief bewerbungsgespräch bewerbungsschreiben ... Das Plagiat ist mit Hilfe von "Copy-Paste" ganz klar auf dem Vormarsch. ...
http://www.news.skripten.at/plugin/tag/abschreiben

Eine Geschichte des Plagiierens hat Josef Karner schon vor einiger Zeit (25.03.2001) auf onlinejournalismus.de unter dem Titel "Der Plagiator" zusammgefaßt, die zeigt, daß das Plagiat keine Erfindung des 21. Jahrhunderts ist, sonder schon seit Jahrhunderten betrieben wird.

open access plagiarism search

„Tag Cloud“ der Beteiligten

Bibliotheksdirektor **DFG**

Jens Brandt: 01.04.2009 **www.oaps.eu**

Oliver Wellnitz: 01.09.2009 **Docoloc**

Siegfried Hackel **Joachim Erich Meier**

Martin Gutbrod Fachbereich Informationstechnologie

Institut für **B**etriebssysteme und **R**echnerverbund

Physikalisch-**T**echnische **B**undesanstalt

Lars Wolf **TU Braunschweig**

Erwarteter Nutzen

Schutz der Autoren vor Plagiatoren

- Ein Autor kann manuell seine Arbeit regelmäßig mit Suchmaschinen exemplarisch prüfen, was wenig effizient ist.
- Autoren können eigene Arbeiten nach OA Veröffentlichung automatisch prüfen lassen.
- Kopieren/Einfügen: Bewusstsein gegen plagiierte Texte schärfen

Verbesserung der Open Access-Veröffentlichungen

- Unterstützung des Begutachtungsprozesses
  - Related Work
  - Aussortierung von Eigenplagiat, Fremdplagiat, Vielfacheinreichungen

OA Repository-Bereiber

- „Integrität“ des Dokumentenbestandes
- Automatisierung der Prüfung von Neuaufnahmen (SOAP/REST-API)
- Untersuchung auf Urheberrechtsverletzungen

Prüfreport für OA-Dokumente

Einfache Interpretation

- Quellen, die aus speziellen Repositorien (bspw. mit DINI-Zertifikat) stammen, besonders kennzeichnen
- Herkömmliche Plagiatsuchwerkzeuge zeigen URL, eventuell Titel des Referenzdokuments und ein kleines Textfragment an
- Erweiterte Metadaten, wie in OAI-PMH definiert und in OAI-Repositorien verfügbar, im Prüfreport mit darstellen
- Bewertung der ähnlichen Dokumente durch Aufnahme von Link- und Zitationsanalysen in den Herkunftsreport

Aufbau eines OA-kompatiblen Plagiatsuchdienst-Harvesters

Einheitlicher Harvester

- Harvesting-Proxy
  - OAI-PMH, jedoch nicht ausschließlich
  - Intelligenz zum Crawlen beliebiger Archive
  - Volltext und URL von besonderem Interesse, Metadaten nice-to-have
- Fehlertolerante Datenextraktion
  - fehlerhafte oder verschlüsselte Dokumente
  - Unterstützung vieler Dateitypen (PDF 1.*, Doc(x), HTML, Tex, PS, ODF, Pixel/OCR etc.)
  - Zeichensatzproblematik
  - OAI-kompatible Metadaten
- Aufbau der Serverinfrastruktur
  - Etwa 200 Millionen Dokumente
  - Hadoop Distributed File System / Lucene

Warum ein weiterer Volltext-Harvester?

Vorhandene Suchmaschinen mit Volltextindex

- **Scirus (Verlag, 450 Millionen Dokumente, API?)**
- **Scientific Commons (Uni Projekt, 12 Mio. Dokumente, max. 3 MB)**
- **BASE (UB, 8 Mio. Dokumente)**
- **Google, Microsoft, Yahoo etc.**

Relevanz

- **Schnittstellen und Abhängigkeiten**
- **Optimiert für Plagiatanalyse**
- **Alle verfügbaren Dokumente aus Bildung, Wissenschaft und Forschung**
- **Metadaten als Ergänzung zu Volltext**
- **Open Access und Closed Access vernetzen und Closed-Access-Inhalte Open Access im Volltext recherchierbar machen („grüngoldner“ Weg)**

Sonstige Arbeitsprogramme

- **Open-Source-Software OPUS um SOAP-Funktionalitäten zur Plagiatsuche erweitern**
- **Installation eines automatischen Verfahrens, um nach einer Veröffentlichungszeit die eigene Arbeit geprüft zu erhalten**
  - **Um festzustellen, ob aus der Arbeit unrechtmäßig abgeschrieben wurde.**
- **Plagiatsuchergebnisse mittels des ArchiSafe langfristig rechtssicher nachweisbar machen: (Digitale Signatur mit Zeitstempel)**

Aggregation und Vernetzung

Möglichkeit einer verteilten Indexgenerierung zur Volltext- und Plagiatsuche?

- Speicherverbrauch eines Volltextindex über ein Repository ist um bis zu einem dreistelligen Faktor kleiner als der Download aller Dokumente
- Die zentrale Berechnung eines Index ist speicher- und prozessorintensiv
  - Verteilte Berechnung direkt auf Seite der Repositories
- Zusammenführung der berechneten Indizes beim zentralen Suchdienst
- Dies lässt eine sehr effiziente Zusammenstellung eines Volltextindex über alle Repositorien erwarten.

Kontakte

Dr.-Ing. Martin **Gutbrod**,
Physikalisch-Technische Bundesanstalt
Tel.: +49-531-592-8410
Email: martin.gutbrod@ptb.de

Prof. Dr. Lars **Wolf**
TU Braunschweig, IBR
Tel.: +49-531-391-3288
Email: wolf@ibr.cs.tu-bs.de

Dr.-Ing. Joachim Erich **Meier**
Physikalisch-Technische Bundesanstalt, Bibliotheksdirektor
Tel.: +49-531-592-8131
Email: Joachim.Meier@ptb.de

Dr. Siegfried **Hackel**
Physikalisch-Technische Bundesanstalt
Tel.: +49-531-592-8400
Email: siegfried.hackel@ptb.de

Dipl.-Inform. Jens **Brandt**,
TU Braunschweig, IBR
Email: brandt@ibr.cs.tu-bs.de
Seit 1. April 2009

Dipl.-Inform. Oliver **Wellnitz**
Physikalisch-Technische Bundesanstalt,
Email: oliver.wellnitz@ptb.de
Ab 1. September 2009

http://www.oaps.eu

Fragen? fragen! ☺

open access plagiarism search