

Dienste zur Entdeckung von Plagiaten und deren (geplanter) Einsatz an der HU

Prof. Dr. Bettina Berendt

Institut für Wirtschaftsinformatik

www.wiwi.hu-berlin.de/~berendt

Medienkommission der HU, 26.6.2006

Agenda

1. Was ist ein Plagiat und warum ist es ein Problem?
2. Was ist ein P.-Entdeckungsdienst?
3. Was ist rechtlich zu beachten?
4. Was ist die aktuelle Lage an der HU?
5. Wie kann es weitergehen?

Was ist ein Plagiat?

- Ungekennzeichnete Verwendung „fremden geistigen Eigentums“
- vulgo:
 - Copy and paste
 - Abschreiben

Warum sind Plagiate ein Problem?

- Verfehlung des Lernziels:
 - Wie und warum funktioniert der wissenschaftliche Erkenntnisprozess?
 - „**Stand on the shoulders of giants**“ (Google Scholar)
 - Wie und warum funktioniert wissenschaftliches Arbeiten?
 - Open source + Zitieren!
- Ungerechtigkeit:
 - Ein Plagiat ist ein Täuschungs(-versuch)

Ausmaß des Problems I

(nur einige Beispiele ...)

- Hausarbeiten-Börsen im World Wide Web und das grassierende Problem des Textraubs aus dem Web
- Siehe z.B. Weber-Wulff (2002) und (Spiegel 2002), welcher einen Plagiatsanteil von 1/3 aller Arbeiten berichtete.
- In einer Befragungsstudie an 23 US-amerikanischen Universitäten (Rutgers University 2003) gaben 38% der ca. 18000 befragten Studierenden an, im vergangenen Jahr mindestens einmal direkt aus dem Internet kopierte Elemente ohne Quellennachweis in ihre Arbeiten aufgenommen zu haben.
- In einer Beobachtungsstudie registrierten Weinstein und Dobkin (2002) eine Rate von 17%.
- Weiteres und produktive Reaktionen: Überblick in www.wiwi.hu-berlin.de/~berendt/NHHL/berendt_NHHL_2003.pdf

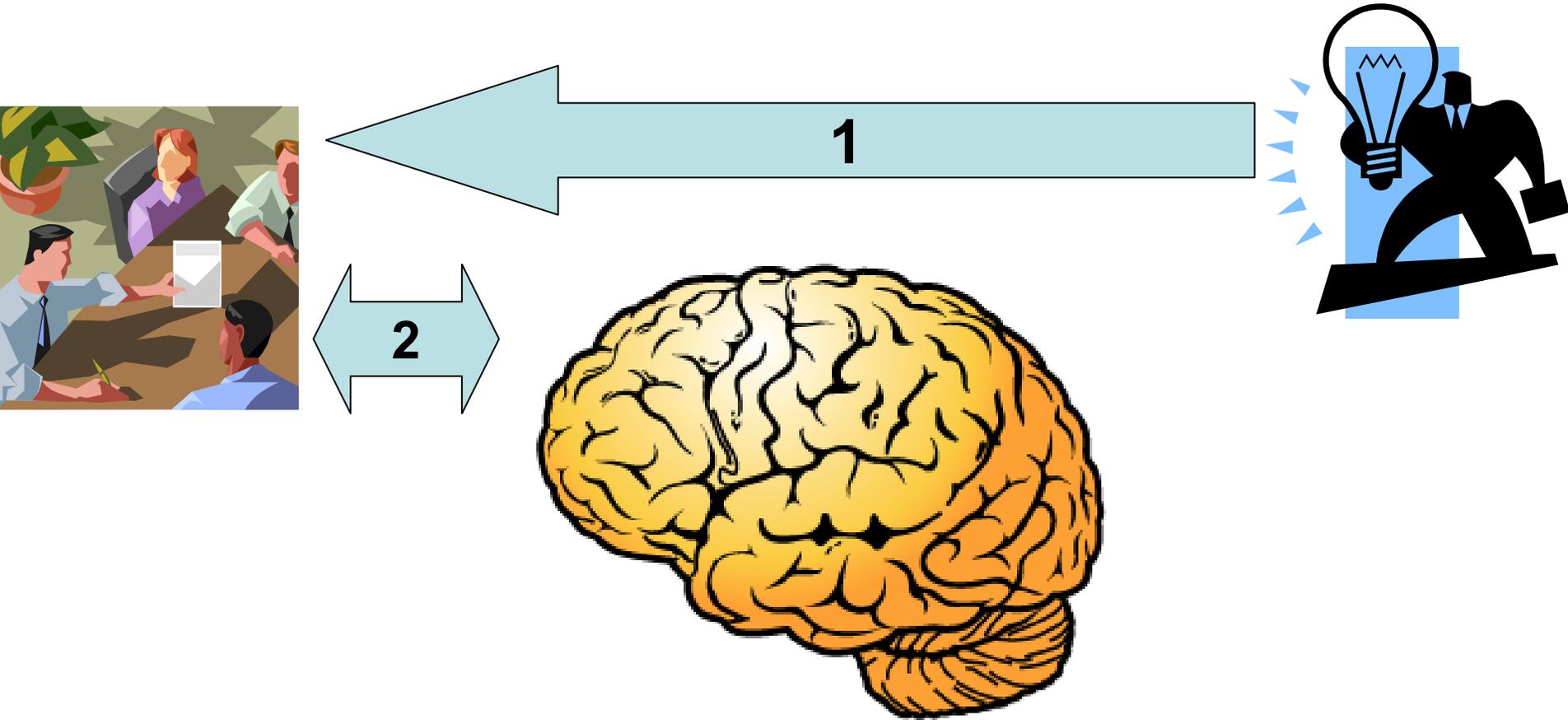
Ausmaß des Problems II

(kleines Meinungsbild an der HU
nach Vorstellung im Multimedia-Kreis)

- Wie hoch schätzen Sie den Anteil der Plagiate bei den Arbeiten Ihrer Studierenden?
 - 56.25% antworteten: „schon bedeutsam (5-15%)“ oder „recht groß (15-25%)“
- Plagiarismus lässt sich durch Ausbildung im wissenschaftlichen Arbeiten (Verständnis, Motivation) und durch geeignete Aufgabenstellungen verringern, aber nicht verhindern

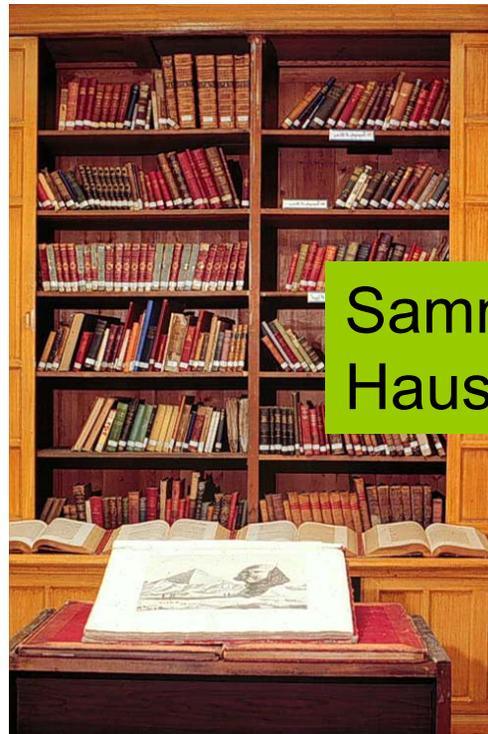
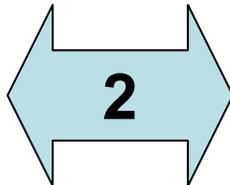
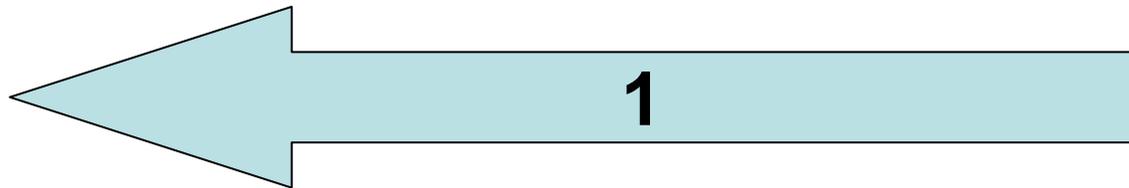
Was ist ein Plagiatssuchdienst?

Plagiatssuchdienst I: Gedächtnis



Plagiatssuchdienst II: Bibliothek

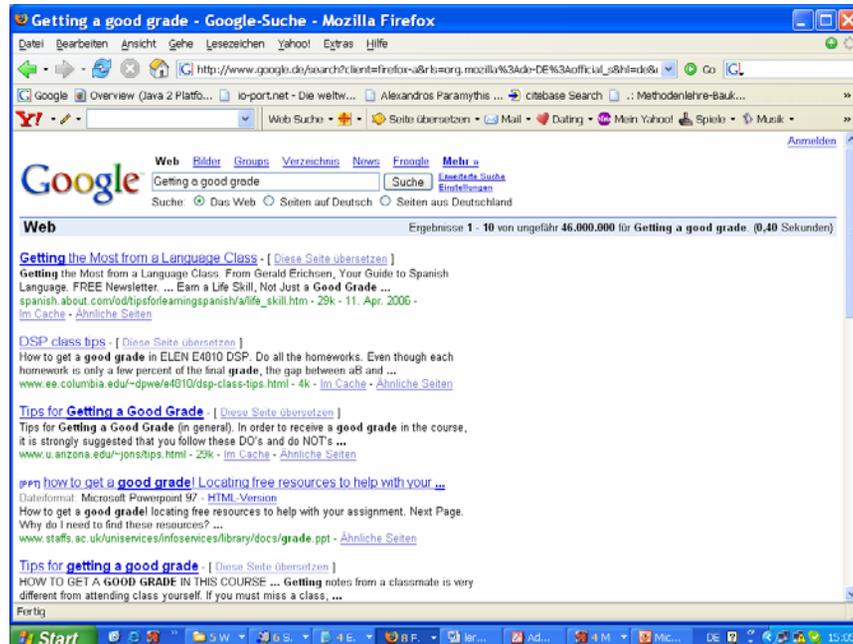
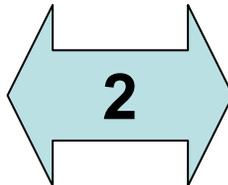
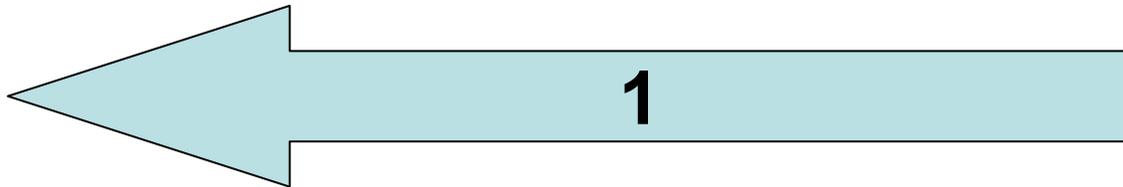
– verwendet von 25% in HU-Umfrage –



Sammlung früherer Hausarbeiten

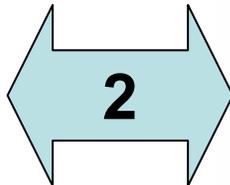
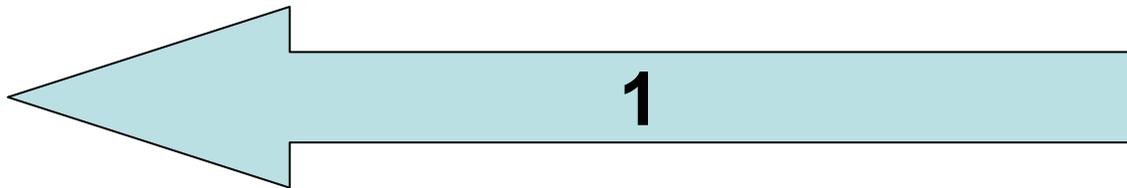
Plagiatsuchdienst III: Internet-Suchmaschine

– verwendet von 81.25% in HU-Umfrage –



Plagiatssuchdienst IV (i.e.S.)

– verwendet von 31.25% in HU-Umfrage –



http://www.turnitin.com - Turnitin - Mozilla Firefox

Turnitin Originality Report
RFID in the Automotive by Xu Lei
Processed on 12-07-05 10:21 CET ID: 21654337 Word Count: 13501

Overall Similarity Index: 17% [exclude quoted](#) [exclude bibliography](#) mode: [show highest matches together](#)
[show highest matches together](#)
[show matches one at a time](#)
[quickview \(classic\) report](#)

1	radio frequency identification (RFID) has been under development for the	56
2	2% match (internet) http://www.cognizantcommunity.co.uk	
3	1% match (internet) http://endtime.com	
4	1% match (internet) http://www.ubipass.com	
5	1% match (internet from 09/16/05) http://www.sunonecertified.ru	
6	1% match (internet) http://www.stevequayle.com	
7	1% match (internet) http://securitytalks.com	
8	1% match (internet) http://www-eu3.semiconductors.com	
9	1% match (internet from 10/14/05)	

1. Introduction One of the next steps in automatic identification technology, past decade and promises to revolutionize supply chain management. More advanced than the barcode system, RFID is already in use in the automotive, publishing, beverage, and other industries. Currently, the global RFID market is estimated to be worth \$1.94 billion² and is driven by the fact that companies are generating and collecting ever increasing amounts of information with efficiency.

Vehicle manufacturers are very proactive in investing in the latest technology to ensure they remain efficient and effective in an extremely competitive market. 1

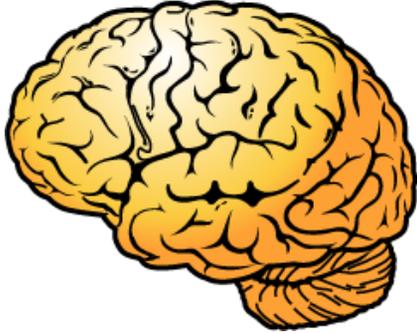
Using new

technology demonstrates their commitment to customer care and in an industry where costs are closely monitored, selecting automatic identification technology, in particular RFID provides the high level of control 1

needed³.
Many applications are already in use, for example, DaimlerChrysler

Fertig

Gemeinsamkeiten



... haben *alle* begrenzte Datenbanken und begrenztes Retrieval!
... benötigen *alle* menschliche Bewertung von Ausmaß, Art und Schweregrad der Plagiiierung

Rechtliche Aspekte (I)

- mit Dank an Herrn Oldenburg und Herrn Kuhring -

- Das Plagiat als Rechtsverstoß
 - keine eigenständige wissenschaftliche Arbeit
 - dafür keine Regelungen im Berliner Hochschulgesetz
 - nur für bestimmte Arbeiten in Prüfungsordnungen geregelt
 - sonst: nur geringfügiger rechtlicher Verstoß

➔ Klärungsbedarf

Rechtliche Aspekte (II)

- Datenschutz
 - Lösung: Arbeiten vor dem Prüfen anonymisieren
- Urheberrecht
 - Lösung: Speicherung der Arbeiten auf fremden Servern (möglichst) verhindern
 - Problem: Verhindert das Vergleichen abgegebener Arbeiten untereinander
- Verhältnismäßigkeit
 - Eingriff nur mit Begründung; Art. 2 GG (Allg. Persönlichkeitsrecht)
 - Haushaltsrecht: Verhältnis Aufwand / Nutzen
 - Lösung: Evaluation
 - Generalverdacht gegen alle Studierenden ist problematisch – muss gerechtfertigt sein
 - Aber: M.E. analog zur Durchführung von Klausuren („Taschen und Jacken an den Rand“) – hier ist der „Generalverdacht“ gerechter als verdachtsbasierte Unterschiede im Verfahren
 - Lt. Herrn Kuhring bestehen hier Unterschiede; Diskussion ist also erforderlich!

Blau heißt:
Diese Lösung steht noch aus!
S. hierzu S. 18ff.

Rechtliche Aspekte (III)

- Geeignetheit
 - Fehldeutungen von Zitaten und (erwünschten) Diskussionen
 - Lösung (s.o.): Eine qualitative Überprüfung der inkriminierten Textstellen ist immer erforderlich – [hierzu Leitlinien zur Hilfestellung?!](#)
 - Trefferquote der Software (vs. Gleichheitsprinzip)
 - [Lösung: Evaluation](#)
 - Software-Angebote zur Verschleierung von Plagiaten
 - Bemerkung: Wettrüsten ist generell nicht vermeidbar; gute Spickzettel anfertigen und sorgfältig umformulieren hat hohen Lernwert!
 - Bei Identität von Arbeiten Urheber nicht feststellbar
 - Bemerkung: M.E. analog zum Problem des Abschreibens bei Klausuren
- Computerbetriebsordnung der HU
 - Entscheidung über dienstliche Nutzung von Online-Diensten
 - Beteiligung DV-Beauftragte; Rahmenentscheidung notwendig

Rechtliche Aspekte (IV)

- S.a. Herrn Kuhrings Stellungnahme
 - zugänglich im Moodle-Plagiarismus-Bereich
<http://lms.cms.hu-berlin.de/moodle/course/view.php?id=2189>

Derzeit an HU in Verwendung

- Turnitin
 - www.turnitin.com
 - ASP, eigene Datenbank
 - Mehrjährige Erfahrungen in Wirtschaftswissenschaftl. Fakultät und anderswo
 - Kostenlose Campus-Lizenz bis Mitte September 2006
 - Instructor-Account zum Testen bei mir: berendt@wiwi.hu-berlin.de
- Plagiarism-Finder
 - <http://www.m4-software.com/>; Shareware:
http://www.freeloadscenter.com/Business/Educational_Tools/Plagiarism_Finder.html
 - Lokal installiert, zerlegt ein Dokument in *viele* Anfragen bei Google
 - Gerrit Oldenburg, Juristische Fakultät
- Ephorus
 - www.ephorus.nl
 - ASP, eigene Datenbank
 - Lt. Hersteller ist eine Integration in andere Software mgl.
 - Carsten Mantel, Institut für Tropenmedizin

Wohin kann der Weg für die HU gehen?

- Welche Variante für softwaregestützte Plagiateerkennung würden Sie bevorzugen?
 - 55% in HU-Umfrage: eine zentrale Lösung
- Möglichkeit 1: Im Anschluss an die jetzige Pilotphase eine Lizenz (z.B. von turnitin) anschaffen
- Möglichkeit 2: Zur Klärung der auf Seiten 13-14 Als „noch ausstehend“ gekennzeichneten Punkte → folgende Schritte (S. 18f.)

Problem Evaluation

- Lückenhafte Aussagen der Dienst-Hersteller zu technischen Parametern (Größe der DB, Anteil von Sprachen, ...)
- Mir bekannte Evaluationen waren sehr klein
 - FHTW-Vergleichsstudie incl. Turnitin und Plagiarism-Finder: 10 Dokumente;
 - Studie von ephorus vs. Metacrawler: 19 Dokumente
- *Recall* UND *precision* sind wichtig
 - Recall = wie viele plagiierte Textstellen werden gefunden
 - Precision: z.B. - werden ungewöhnliche Zitierweisen berücksichtigt, um falsche Positive zu vermeiden
- Eine bottom-up-Evaluation (Dozenten schicken Erfahrungsberichte) ist zu unschlüssig.
- Handlungsempfehlungen, wie ein Dozent gefundene Übereinstimmungen bewerten soll, fehlen.
- Vorteile/Nachteile, technische Machbarkeit, Kosten der Einbettung in andere Software (z.B. Moodle) sind unklar.

Ideen für ein Projekt zur Lösung der angesprochenen Probleme

- *Hintergrund*: persönliches Interesse
 - Lehre: als Dozentin
 - Forschung: Bezug zu meinen Forschungsschwerpunkten Web Mining und studentische Literaturarbeit
- *Ziel*: fundierte Entscheidungsgrundlage für HU
- *Arbeitsschritte*
 - Auswahl zu testender Software
 - Umfrage zu konkreten Aspekten des Bedarfs
 - Definition von Maßen für die Evaluation
 - Erstellen eines hinreichend großen Test-Korpus
 - Quantitative Evaluation von precision und recall
 - Qualitative Evaluation der Usability
 - Ermittlung des Workflow bei Dozenten
 - Einbettung in Ausbildung Medienkompetenz (Studierende, Dozenten)
 - Berechnung der Kosten verschiedener Lösungen
 - Ableitung einer Kaufempfehlung

Vielen Dank für Ihre Aufmerksamkeit!

Mehr Informationen, Anleitungen, Literatur:

www.wiwi.hu-berlin.de/~berendt/Plagiarismus