

Plagiate, und wie (gut) man sie erkennen kann

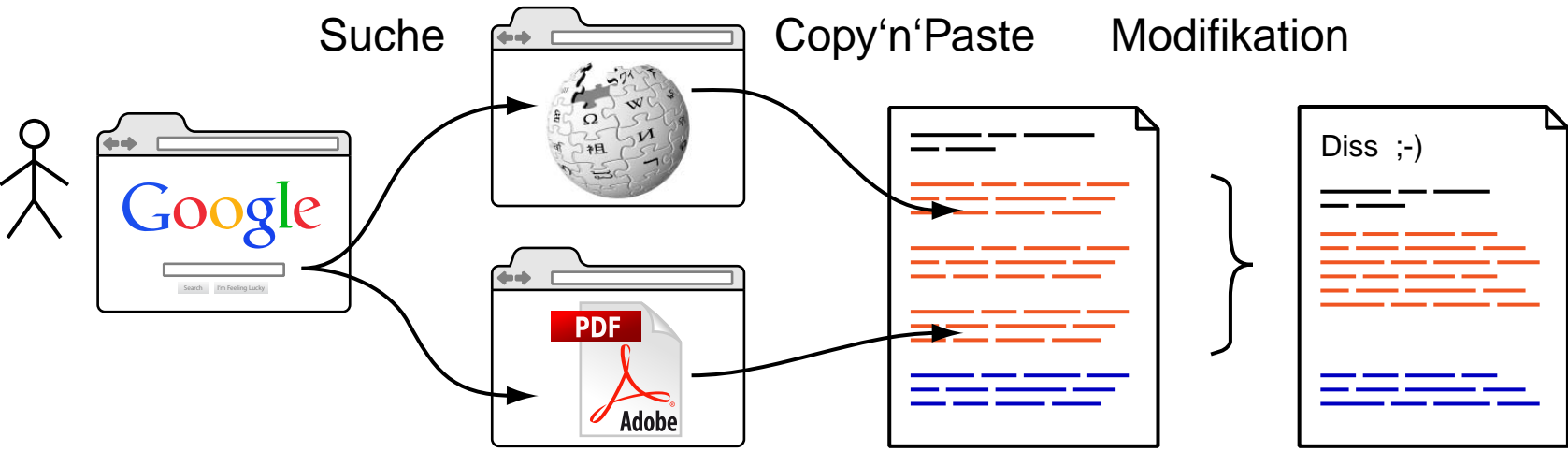
Technologien zur Wiederverwendung von Texten aus dem Web

Martin Potthast

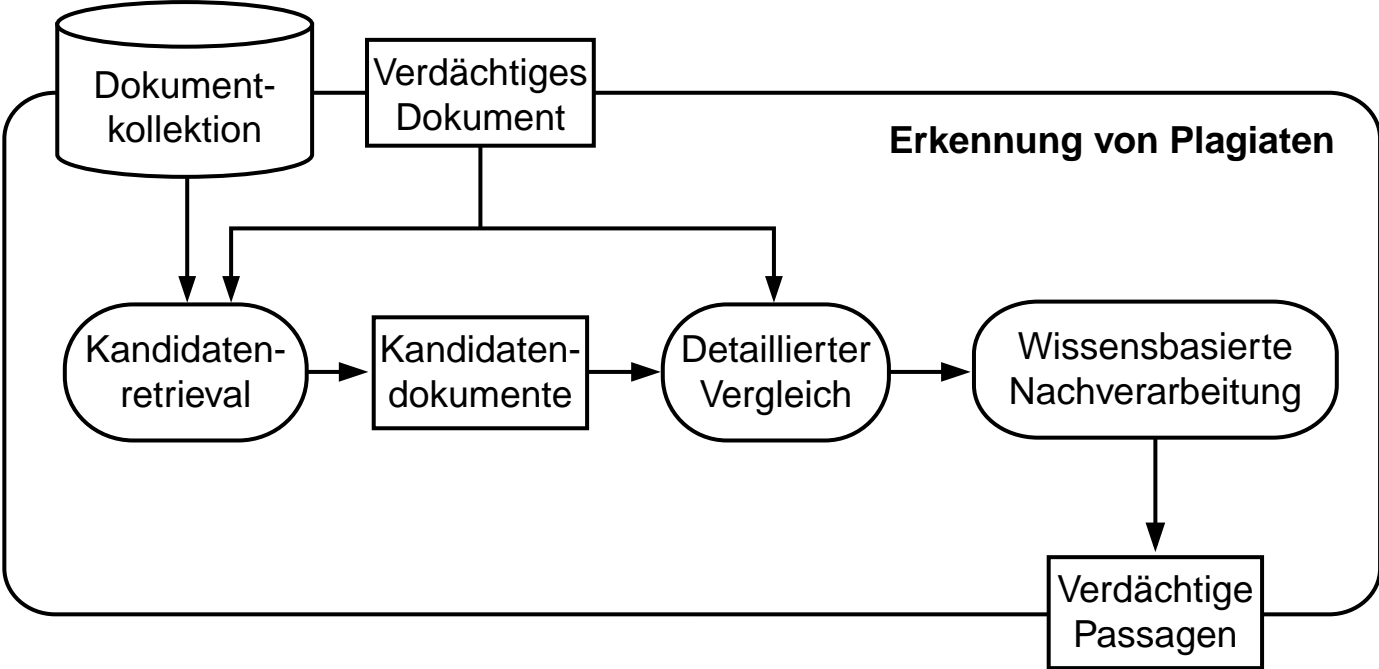
Bauhaus-Universität Weimar

www.webis.de

Wie Menschen plagiiieren

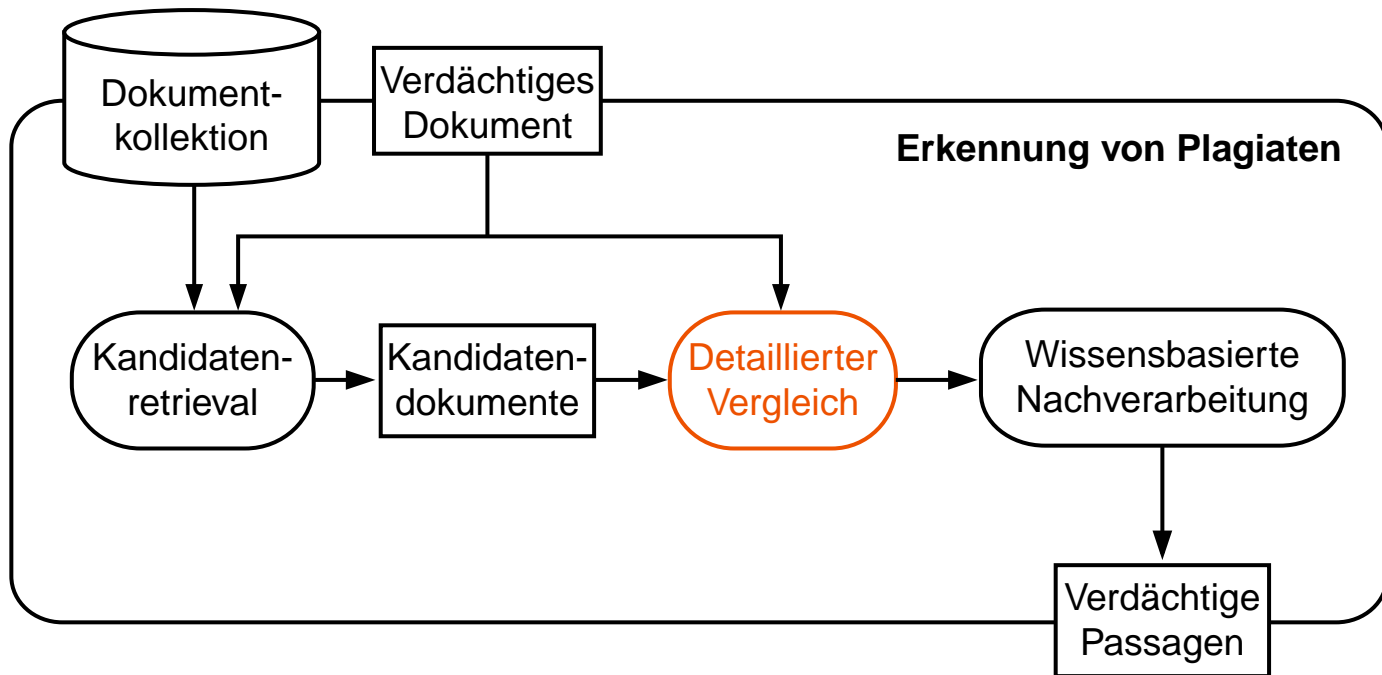


Wie Maschinen Plagiate erkennen



Wie Maschinen Plagiate erkennen

- Wie können sprachübergreifende Plagiate erkannt werden?
- Wie verlässlich sind Algorithmen zur Plagiatserkennung?



Messung sprachübergreifender Textähnlichkeit

Alan Turing was conceived at Chattrapur, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to Maida Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active, and during Turing's childhood years his parents travelled between Hastings, England and India, leaving their two sons to stay with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.



Alan Mathison Turing was born on 23 June 1912. His father was Julius Mathison Turing, member of the civil service in India, and his mother Ethel Sara Turing, the daughter of Edward Waller Stoney. Alan's childhood was spent with his elder brother John, living with a retired Army couple near Hastings, England. His parents returned to India until the end of his father's civil service commission, and visited when they could. Signs of Turing's genius showed early in his life. It is reported that he taught himself reading in less than three weeks.



Messung sprachübergreifender Textähnlichkeit

Alan Turing was conceived at Chatrapur, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to Maida Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active, and during Turing's childhood years his parents travelled between Hastings, England and India, leaving their two sons to stay with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.




Alan Mathison Turing was born on 23 June 1912. His father was Julius Mathison Turing, member of the civil service in India, and his mother Ethel Sara Turing, the daughter of Edward Waller Stoney. Alan's childhood was spent with his elder brother John, living with a retired Army couple near Hastings, England. His parents returned to India until the end of his father's civil service commission, and visited when they could. Signs of Turing's genius showed early in his life. It is reported that he taught himself reading in less than three weeks.



Messung sprachübergreifender Textähnlichkeit


Alan Turing was conceived at Chatrapur, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan turing brought up in England, so they returned to Maida Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active and during Turing's childhood years his parents travelled to Hastings, England and India, leaving their two sons to stay with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.

turing 4
travel 1
teach 0
: 0
army 1
alan 3
active 1



Alan Mathison Turing was born on 23 June 1912. His father was Julius Mathison Turing, member of the civil service in India, and his mother Ethel Sara Turing, the daughter of Edward Valler Stoney. Alan's childhood was spent with his elder brother John, living with a retired Army couple near Hastings, England. His parents returned to India until the end of his father's civil service commission, and visited when they could. Signs of Turing's genius showed early in his life. It is reported that he taught himself reading in less than three weeks.

5
0
1
: 0
1
2
0




Messung sprachübergreifender Textähnlichkeit

Alan Turing was conceived at Chattrapur, Orissa, India. His father was a member of the Indian Civil Ser-


Alan Mathison Turing was born on 23 June 1912. His father was Julius Mathison Turing, member of the

turing travel teach ⋮ army alan active	4 1 0 ⋮ 1 3 1	φ	- Euklidische Distanz - Skalarprodukt - Kosinusähnlichkeit	5 0 1 ⋮ 1 2 0	his the r Stoney. nt with his with a Hastings, rned to ather's and visi- s of arly in his aught
----------------------------------------------------------	---------------------------------	-----------	------------------------------------------------------------------	---------------------------------	----------------------------------------------------------------------------------------------------------------------------------

early in life, Turing showed signs of the genius he was to later prominently display.



nineteen leading in loss with three weeks.



Messung sprachübergreifender Textähnlichkeit

Alan Turing was conceived at Chatrapur, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to Maida Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active, and during Turing's childhood years his parents travelled between Hastings, England and India, leaving their two sons to stay with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.



Turings Vater Julius Mathison Turing, ein britischer Staatsdiener in Chatrapur, Indien, und dessen Frau Ethel Sara wollten, dass ihr Kind in Großbritannien geboren wird. Deshalb kehrten sie nach London-Paddington zurück, wo Alan Turing am 23. Juni 1912 zur Welt kam. Da der Staatsdienst seines Vaters noch nicht beendet war, pendelte dieser während Turings Kindheit zwischen England und Indien. Seine Familie ließ er aus Furcht vor Gefahren in der britischen Kolonie bei Freunden in England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turings.



Messung sprachübergreifender Textähnlichkeit

Alan Turing was conceived at **Chatrapur**, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted **Alan** to be brought up in **England**, so they returned to Maida Vale, **London**, where **Alan Turing** was born on **23 June 1912**. He had an elder brother, John. His father's civil service commission was still active, and during **Turing's** childhood years his parents travelled between Hastings, **England** and India, leaving their two sons to stay with a retired Army couple. Very early in life, **Turing** showed signs of the genius he was to later prominently display.




Turings Vater Julius Mathison **Turing**, ein britischer Staatsdiener in **Chatrapur**, Indien, und dessen Frau Ethel Sara wollten, dass ihr Kind in Großbritannien geboren wird. Deshalb kehrten sie nach **London**-Paddington zurück, wo **Alan Turing** am **23. Juni 1912** zur Welt kam. Da der Staatsdienst seines Vaters noch nicht beendet war, pendelte dieser während **Turings** Kindheit zwischen **England** und Indien. Seine Familie ließ er aus Furcht vor Gefahren in der britischen Kolonie bei Freunden in **England** zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz **Turings**.



Messung sprachübergreifender Textähnlichkeit


Alan Turing was conceived at Chatrapur, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan Turing brought up in England, so they returned to Maida Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active and during Turing's childhood years his parents travelled between England and India leaving their two sons to stay with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.

turing 4
travel 1
two 1
: 0
britisch 0
beendet 0
alan 3



Turings Vater Julius Mathison Turing, ein britischer Staatsdiener in Chatrapur, Indien, und dessen Frau Ethel Sara wollten, dass ihr Kind in Großbritannien geboren wird. Deshalb kehrten sie nach London-Paddington zurück, wo Alan Turing am 23 Juni 1912 zur Welt kam. Da der Staatsdienst seines Vaters noch nicht beendet war, pendelte dieser während Turings Kindheit zwischen England und Indien. Seine Familie ließ er aus Furcht vor Gefahren in der britischen Kolonie bei Freunden in England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turings.


5
0
0
:
2
1
1



Messung sprachübergreifender Textähnlichkeit


Alan Turing was conceived at Chatrapur, Orissa, India. His father was a member of the Indian Civil Ser-

Turings Vater Julius Mathison Turing, ein britischer Staatsdiener in Chatrapur, Indien, und dessen

turing	4		5
travel	1		0
two	1		0
⋮	⋮		⋮
britisch	0		2
beendet	0		1
alan	3		1

ausgenommen
 - Syntaxüberlappung
 - Autom. Übersetzung

early in life, Turing showed signs of the genius he was to later prominently display.



England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turings.



Cross-language Explicit Semantic Analysis (CL-ESA)

Alan Turing was conceived at Chingari, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to the Vale, London, where Alan Turing was born on 23rd June. His father's elder brother, Jim, whose civil service commission was still active, and Turing's childhood years his parents travelled back to England and leaving their two sons to be raised by an Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.



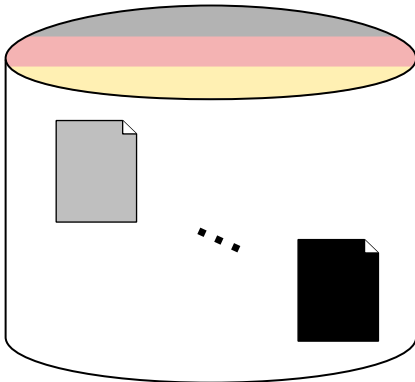
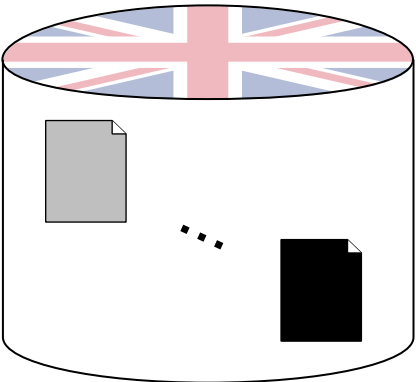
Turing's Vater Julius Mathison Turing, ein britischer Staatsdiener in Chingari, Indien, und dessen Frau Ethel wollten, dass ihr Kind in Großbritannien aufwächst. Deshalb kehrten sie nach London-Pearl zurück, wo Alan Turing am 23. Juni zur Welt kam. Sein älterer Bruder seines Vaters noch aktiv war, während die Kindheit Turing's Kindheit Jahren in England und Indien verbrachte, er wurde in der britischen Armee von seinen Freunden in England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turing's.

Cross-language Explicit Semantic Analysis (CL-ESA)

Alan Turing was conceived at Chingwi, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to the Vale, London, where Alan Turing was born on 23 June 1912. His father's civil service commission was still active, and Turing's childhood years his parents travelled back and forth between England and India leaving their two sons to be with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.



Turing's Vater Julius Mathison Turing, ein britischer Staatsdiener in Chingwi, Indien, und dessen Frau Ethel wollten, dass ihr Kind in Großbritannien aufgezogen wird. Deshalb kehrten sie nach London-Pearl zurück, wo Alan Turing am 23. Juni 1912 zur Welt kam. Seine Vaters dienst seines Vaters noch nicht beendet war, pendelte die Familie und Turing's Kindheit zwischen England und Indien. Die Eltern ließen er aus Furcht, dass er in der britischen Kolonialarmee in der britischen Armee in England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turing's.



Cross-language Explicit Semantic Analysis (CL-ESA)

Alan Turing was conceived at Chingwi, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be born in England, so they returned to London, where they had an elder brother, J. H. His father's civil service commission was still active, and during Turing's childhood visits to his parents travelled back and forth. Turing's father died in 1930, leaving their two sons to be raised by their



Turing's Vater Julius Mathison Turing, ein britischer Staatsdiener in Chingwi, Indien, und dessen Frau Ethel wollten, dass ihr Kind in Großbritannien geboren wird. Deshalb kehrten sie nach London-Perth zurück, wo Alan Turing, dessen ältester Bruder, geboren wurde. Sein Vaters zivile Dienststelle war noch aktiv, und während Turings Kindheit besuchte er seine Eltern in England und Indien. Turing's father died in 1930, leaving their two sons to be raised by their

Log in / create account

Article Discussion Edit View history Search

Reuse

From Wikipedia, the free encyclopedia

To **reuse** is to use an item more than once. This includes conventional reuse where the item is used again for the same function, and new-life reuse where it is used for a different function. In contrast, *recycling* is the breaking down of the used item into raw materials which are used to make new items. By taking useful products and exchanging them, without reprocessing, reuse help save time, money, energy, and resources. In broader economic terms, reuse offers quality products to people and organizations with limited means, while generating jobs and business activity that contribute to the economy.

Historically, financial motivation was one of the main drivers of reuse. In the developing world this driver can lead to very high levels of reuse, however rising wages and consequent consumer demand for the convenience of



3R Concepts

- Waste Disposal Hierarchy
 - Reduce
 - Reuse**
 - Recycle
- Barter
- Dematerialization
- Downcycling
- Dumpster diving
- Ecodesign
- Ethical consumerism
- Freeganism
- Extended producer responsibility
- Industrial ecology
- Industrial metabolism
- Material flow analysis

العربية
বাংলা
Български
Deutsch
Español
Français
한국어
हिन्दी
Bahasa Indonesia

Anmelden / Benutzerkonto erstellen

Artikel Diskussion Versionsgeschichte Suche

Wiederverwendung

Wiederverwendung ist das Prinzip, Aufwand und Material einzusparen, indem ein an einer Stelle nicht mehr benötigter (und damit erneut verfügbar gewordener) Gegenstand an anderer Stelle eingesetzt wird. Durch diese Vorgehensweise erspart man die Vernichtung (auch Zertelung oder Beseitigung) des nicht mehr benötigten Gegenstands und die Erstellung einer neuen Instanz. Im Speziellen kann es sich dabei handeln um:

- Rekonditionierung (Technik), die Aufarbeitung von gebrauchten Produkten
- Retrofit, die Modernisierung oder der Ausbau bestehender Produktionsanlagen
- Recycling, die Verarbeitung von Abfall zu Rohstoffen
- Wiederverwendbarkeit, einmal geschriebene Programmmodule auf universelle Einsetzbarkeit auszuliegen

Die Wiederverwendung kann zusätzlichen Aufwand mit sich bringen: Eine Lagerung ist erforderlich, falls die Wiederverwendung nicht sofort möglich ist. Des Weiteren müssen während der Aufbereitung des Gegenstands eventuell vorhandene Gebrauchsspuren entfernt werden.

Weitere Wortbedeutung [Bearbeiten]

Beamte die das 131er-Gesetz von 1951 betraf, durften ihre Amtsbezeichnung mit dem Zusatz „zur Wiederverwendung (z. Vw.)“ weiterführen.

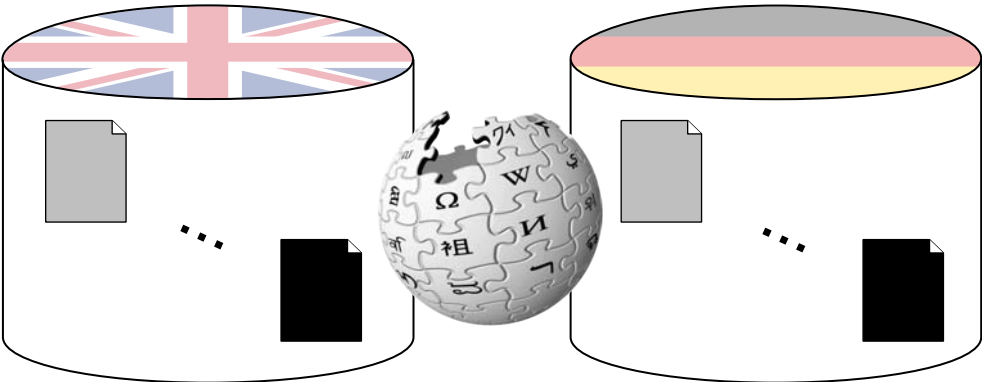
Hauptseite
Über Wikipedia
Themenportale
Von A bis Z
Zufälliger Artikel

- Mitmachen
- Drucken/exportieren
- Werkzeuge
- In anderen Sprachen
- العربية
- বাংলা
- Български
- English**
- Español
- Eesti
- Suomi
- Français
- עברית
- हिन्दी

Cross-language Explicit Semantic Analysis (CL-ESA)

Alan Turing was conceived at Chongwe, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to the Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service commission was still active, and Turing's childhood years his parents travelled back and forth between England and India leaving their two sons to be with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.

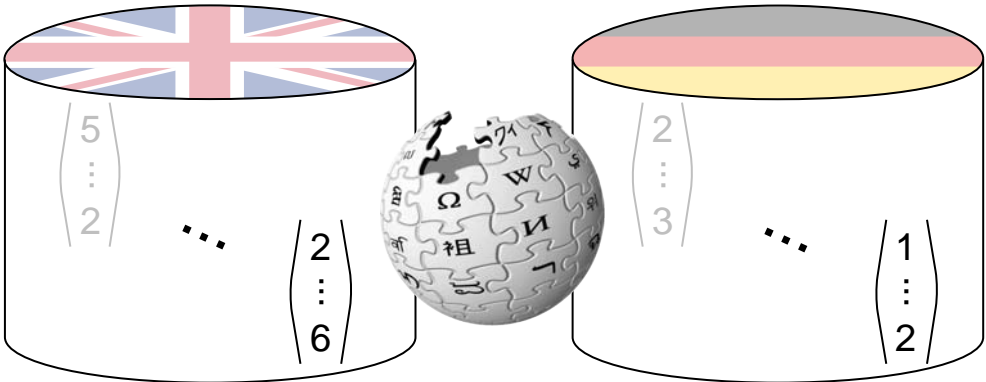
Turing's Vater Julius Mathison Turing, ein britischer Staatsdiener in Chongwe, Indien, und dessen Frau Ethel wollten, dass ihr Kind in Großbritannien aufwachsen wird. Deshalb kehrten sie nach London-Pearl zurück, wo Alan Turing am 23. Juni 1912 zur Welt kam. Er ist das älteste Kind seines Vaters, der noch in Diensten war, während die Eltern und Turing's Kindheit zwischen England und Indien verbrachten. Er ließ er aus Furcht, dass er in der britischen Kolonialarmee in England zurück. Schon in frühester Kindheit zeigte sich die hohe Begabung und Intelligenz Turing's.



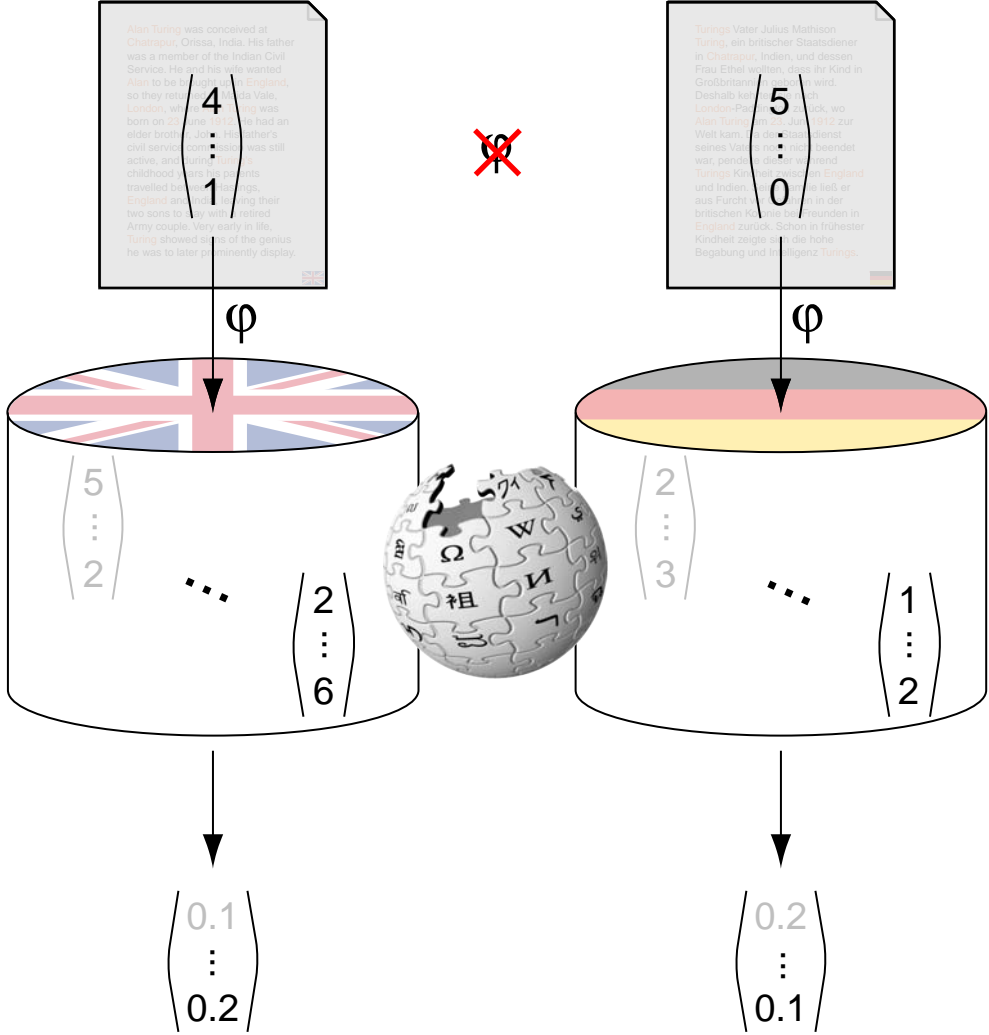
Cross-language Explicit Semantic Analysis (CL-ESA)

Alan Turing was conceived at Chingwi, Orissa, India. His father was a member of the Indian Civil Service. He and his wife wanted Alan to be brought up in England, so they returned to the Vale, London, where Alan Turing was born on 23 June 1912. He had an elder brother, John. His father's civil service career was still active, and Turing's childhood years his parents travelled back and forth between England and India leaving their two sons to be with a retired Army couple. Very early in life, Turing showed signs of the genius he was to later prominently display.

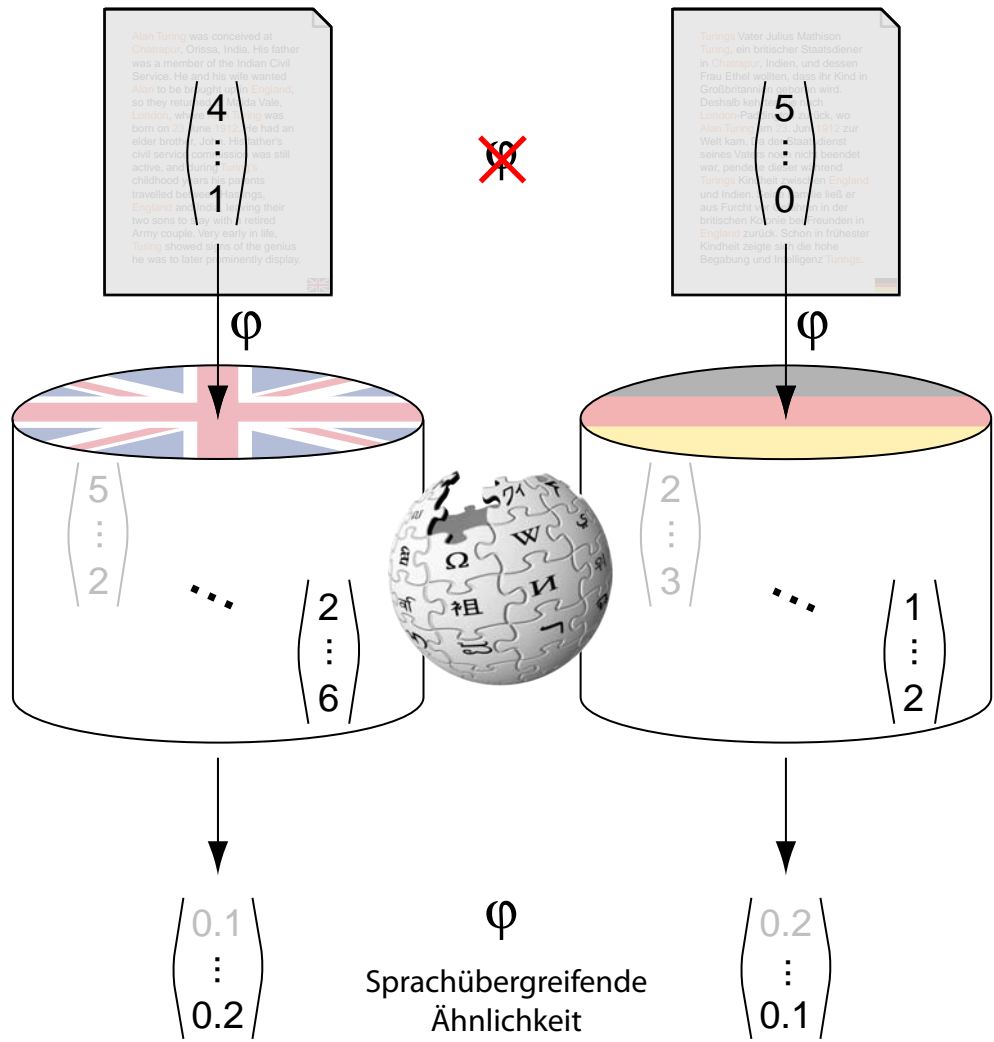
Turing's Vater Julius Mathison Turing, ein britischer Staatsdiener in Chingwi, Indien, und dessen Frau Ethel wollten, dass ihr Kind in Großbritannien aufgezogen wird. Deshalb kehrten sie nach London-Pearl zurück, wo Alan Turing am 23. Juni 1912 zur Welt kam. Er hatte einen älteren Bruder seines Vaters, John. Sein Vaters zivile Dienstleistung war noch aktiv, und Turing's Kindheit wurde von den Eltern zwischen England und Indien hin und her gelassen, während sie ihre beiden Kinder bei einer in der britischen Armee verheirateten Frau aufwachsen ließen. Sehr früh in der Kindheit zeigte sich die hohe Begabung und Intelligenz Turing's.



Cross-language Explicit Semantic Analysis (CL-ESA)



Cross-language Explicit Semantic Analysis (CL-ESA)



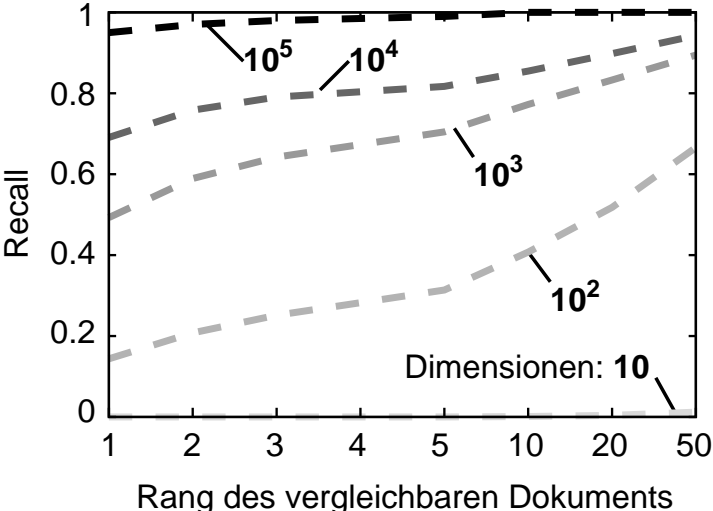
Cross-language Explicit Semantic Analysis (CL-ESA)

Experimente

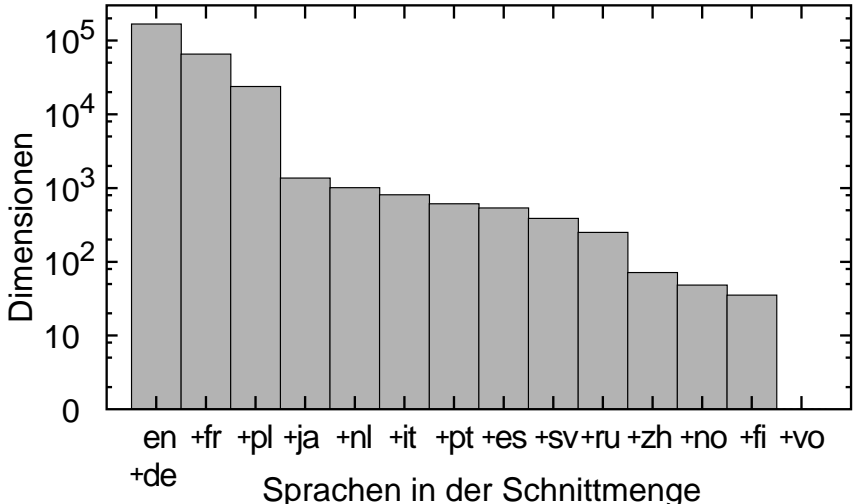
1. Sprachübergreifendes Ranking
2. Bilinguale Rangkorrelation
3. Sprachübergreifende Ähnlichkeitsverteilung
4. Tradeoff Effektivität und Effizienz (Dimensionalität)
5. Multilingualität (Zahl gleichzeitig repräsentierbarer Sprachen)
6. Laufzeit
 - Vergleich mit zwei State-of-the-Art-Modellen
 - Verwendung zweier multilingualer Testkollektionen
 - Vergleich auf 6 Sprachpaarungen
 - > 100 Millionen Ähnlichkeiten berechnet

Cross-language Explicit Semantic Analysis (CL-ESA)

Sprachübergreifendes Ranking



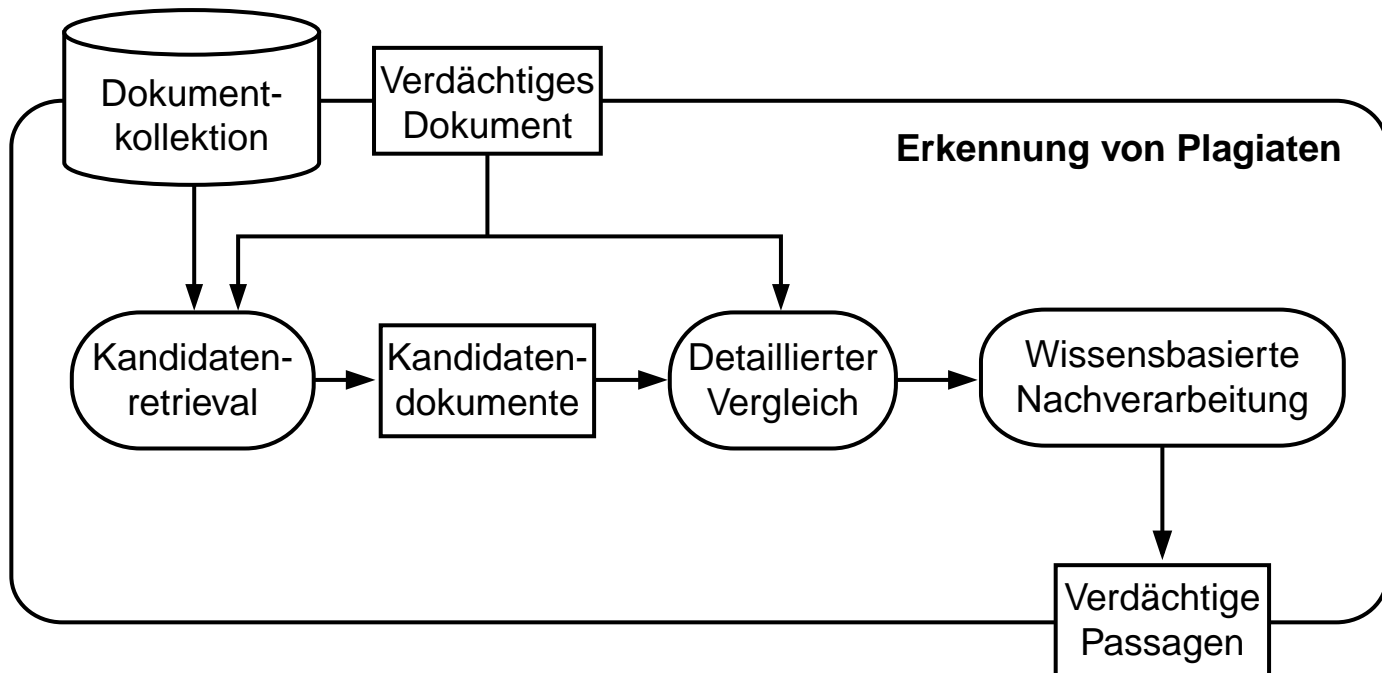
Multilingualität



Wie Maschinen Plagiate erkennen

Wie Maschinen Plagiate erkennen

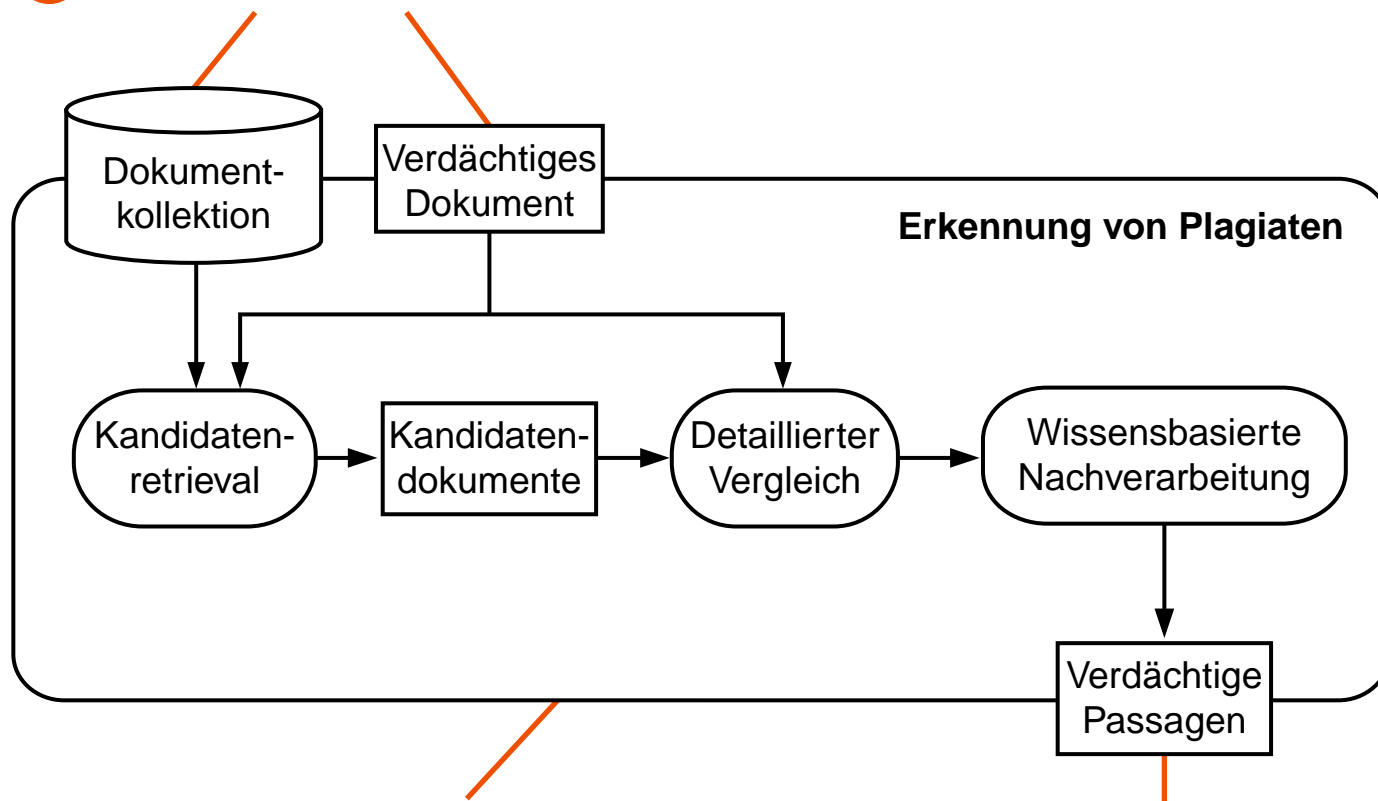
- Wie können sprachübergreifende Plagiate erkannt werden?
- Wie verlässlich sind Algorithmen zur Plagiatserkennung?



Wie Maschinen Plagiate erkennen

- Wie können sprachübergreifende Plagiate erkannt werden?
- Wie verlässlich sind Algorithmen zur Plagiatserkennung?

① Korpus von Plagiaten



③ Alternative Implementierungen

② Erfolgsmaße

Studie zur Evaluierung von Plagiatserkennungsalgorithmen

Evaluierungsaspekt	Text	Code
① <i>Korpusakquise</i>		
Verfügbares Korpus	20%	18%
Eigenes Korpus	80%	82%
<hr/>		
<i>Korpusgröße</i> [# Dokumente]		
[1, 10)	11%	10%
[10, 10 ²)	19%	30%
[10 ² , 10 ³)	38%	33%
[10 ³ , 10 ⁴)	8%	11%
[10 ⁴ , 10 ⁵)	16%	4%
[10 ⁵ , 10 ⁶)	8%	0%

Evaluierungsaspekt	Text	Code
② <i>Erfolgsmaße</i>		
Precision, Recall	43%	18%
Manuell, Ähnlichkeit	35%	69%
Laufzeit	15%	1%
Andere	7%	12%
<hr/>		
③ <i>Vergleichende Experimente</i>		
Keine	46%	51%
Parameter	19%	9%
Andere Algorithmen	35%	40%

- 205 Papiere wurden analysiert
- Ist es möglich, eine Trendwende herbeizuführen?



Internationaler Wettbewerb zur Plagiatserkennung

- ❑ Erste standardisierte Testumgebung
- ❑ Ausgerichtet jährlich seit 2009

- ① Große neue Korpora von Plagiaten (PAN Plagiarism Corpus 2009-2012)
- ② Maßgeschneiderte Erfolgsmaße (Plagdet, Precision, Recall, Granularity)
- ③ 32 teilnehmende Forschergruppen (35 Anmeldungen für 2012)

PAN: Korpuskonstruktion

Ziel:

- ❑ Eine möglichst große, repräsentative Auswahl von plagiierten Dokumenten sowie den verwendeten Quelldokumenten.

Probleme:

- ❑ Reale Plagiatsfälle nicht im großen Maßstab verfügbar
- ❑ Abhängige Variablen unbekannt

Möglichkeiten:

- ❑ Plagiate automatisch generieren (seit 2009 im PAN Plagiatskorpus)
- ❑ Plagiate halbautomatisch generieren (seit 2010 via Mechanical Turk)
- ❑ Plagiate manuell erstellen (seit 2012 via oDesk)
- ❑ Ähnliche Textwiederverwendungen anderer Domänen sammeln
- ❑ Reale Plagiate suchen / sammeln (seit 2011 für deutsche Dissertationen)

→ Crowdsourcing als Schlüssel für große Skalen, Realismus und Diversität

Zusammenfassung & Fazit

Sprachübergreifende Textähnlichkeit

- ❑ CL-ESA benötigt keine Wörterbücher, parallele Korpora, oder Übersetzer
- ❑ Vergleichbare Korpora sind für viele Sprachpaarungen leicht erstellbar

Evaluierung von Plagiatserkennungsverfahren

- ❑ Ergebnisse von PAN zeigen, dass noch viel Luft nach oben ist
- ❑ Das beste, was ein Plagiatserkenner in der Praxis leisten kann ist, das Durchkommen mit plumpen Plagiaten schwieriger zu machen
- ❑ Eigentlich werden nicht Plagiate, sondern bloß verdächtige Textwiederverwendungen erkannt

Wettbewerb zur Plagiatserkennung PAN

- ❑ Webseite: <http://pan.webis.de>
- ❑ Korpora: <http://www.webis.de/research/corpora>
- ❑ Community: <http://groups.google.com/group/pan-workshop-series>