
Autorschaftserkennung im Big-Data-Kontext

Oren Halvani

Fraunhofer Institute for Secure Information Technology (SIT), Darmstadt, Germany
Department of Computer Science, Technische Universität Darmstadt, Germany



ÜBERBLICK

- Motivation
- Autorschafts-Attribution / Verifikation
- Unser Verfahren
- Evaluierung
- Beobachtungen
- Zusammenfassung

MOTIVATION

- In vielen forensischen Szenarien inner-/außerhalb sozialer Netzwerke spielen Identitäten von Autoren eine wichtige Rolle
- Insbesondere gilt es zu klären, ob gegebene Dokumente von einer angeblichen Person geschrieben wurden
- Auf Facebook begegnet man tagtäglich rassistischen, verleumdenden, menschenverachtenden und aggressiven Kommentaren
- Spannend ist hier die Frage, ob solche Kommentare von Fake-Accounts oder von echten Personen ausgehen...

MOTIVATION

- Aber nicht nur in Facebook ist man mit der Klärung von Autorschaften beschäftigt. Manchmal ist es wünschenswert zu wissen, wer hinter Nutzerkommentaren, Blogs, Messenger Chats oder Foren-Postings steckt...

K13 - ONLINE

"Wer Freiheit aufgibt, um Sicherheit zu gewinnen, wird am Ende Beides verlieren!"

Home | User | Forum | Downloads | Links | Texte | Newsarchiv | Votes | Toplist | Suchen

	Beitrag
Skally geschrieben am 24.04.2015 um 04:35:48 Uhr	Boylover Gesucht Hallo, komme aus Berlin und bin 24 Suche jemanden der ebenfalls BL ist und auf Knaben zwischen 12 und 16 Jahren steht oder vielleicht auch jemanden kennt ^^ Danke im voraus ^^
Kurzschrift123 geschrieben am 06.05.2015 um 10:26:11 Uhr	Boylover Gesucht Schick mir mal eine private Kontaktmöglichkeit von dir, Mail Oder Skype bevorzugt
K13online geschrieben am 07.05.2015 um 17:53:25 Uhr	Boylover Gesucht Hallo Ihr Beiden! Wir müssen zum Personenschutz dazu raten, keine persönliche Daten hier ins Forum zu posten. K13-Benutzerservice

[\[Antworten\]](#)

MOTIVATION

24.10.2013 #3

Firdaus ◦
Moderatorin



Religion: دين الحق
Richtung: جماعة الصالحين
Geschlecht: Weiblich
Registriert seit: Dec 2010
Ort: gute Frage
Beiträge: 1.680

Zitat von UmmHandala

Salamu3alaykum. Worauf soll man sich melden. Wenn die Ukht hier angemeldet ist soll sie sich bei mir melden oder über den anonyma Account kannst du weiterleiten barakallahu feeki

وعليكم السلام ورحمة الله وبركاته

War eher für die Brüder gedacht, der Kommentar. Die Schwestern können einfach ne PN schicken in shaa Allah. Werde es ausrichten, Ukhti, barakallahu fiki.

"Ein Großteil von dem, was verursacht, dass Aussagen von dem Erreichen der Wahrheit abweichen ist, das Lob oder den Tadel der Leute zu erwarten, im Austausch zu der Zufriedenheit von Allah oder Seiner Unzufriedenheit." Shaikh Abdul Aziz Al-Tarefe

Menge von kokain intravenös?

Thema abonnieren | I

Geschrieben am: 6.02.2010, 10:57

Hallo leute
ich habe dieses thema erstellt weil ich wissen wollte wie viel kokain+wasser man sich spritzen müsse. ich weiß das es falsch ist aber schreibt mir bitte mal das misch-verhältnis als anfang?
ich habe noch keinerlei erfahrung mit heroine gehabt und hab mir noch keinen druck gegeben ich hab koks bisher nur gesniff und wollte es mal intravenös ausprobieren und ich bin mir im klaren was das für konsequenzen hat nur will man natürlich bei dem ersten mal nix falsch machen also frage ich nach eurem rat.

danke im vorraus 😊👍
mfg

Die mächtigen Macher hinter Schweigers Stiftungsfassade



Selfie-Narziß und Kofferträger der Mächtigen: Ti Schweiger

Erinnern Sie sich noch an den peinlichen Versuch des naiv-gutmenschlichen Filmschauspielers *Ti Schweiger*, ein eigenes Asylantenheim zu gründen? Das ging mit seinen dubiosen Finanzpartnern ziemlich in die Hose. Aber immerhin konnte der Leinwandkasper damit öffentlich signalisieren, dass er als gut erzogener BRD-Promi sich bereitwillig in den Dienst der staatlichen Überfremdungspropaganda stellen will. Er fing sogar an, mit dem dicken *Gabriel* von den Sozis zu kokettieren, obwohl er Politik früher immer ganz blöd fand. Aber die Halbwertszeit von Kinohelden ist ja auch irgendwie begrenzt und wer auch darüber hinaus noch ein Promi bleiben will, muß eben vorsorgen...

DER Presselügenclub

Kontakt zum Presselügenclub
E-Mail: zensurfrei1984@gmail.com

Beliebte Posts

Flüchtlingseloge: Sie bleiben dauerhaft / Anwohner verteidern Asylanten im Jenseitig
Nach wie vor wird die planmäßige Massenüberfremdung Deutschlands mit dem Synonym

MOTIVATION

- Der Wunsch nach automatisierten Lösungen für Ermittler wächst
- Dabei existieren bereits erste Ansätze, die mittlerweile auch schon vor Gericht akzeptiert wurden (diese fungierten als Bestandteile technischer Gutachten)
- Automatisierte Lösungen ersetzen nicht den Ermittler! Sie helfen diesem jedoch den manuellen Aufwand drastisch zu verringern

AUTORSCHAFTS-ATTRIBUTION / VERIFIKATION

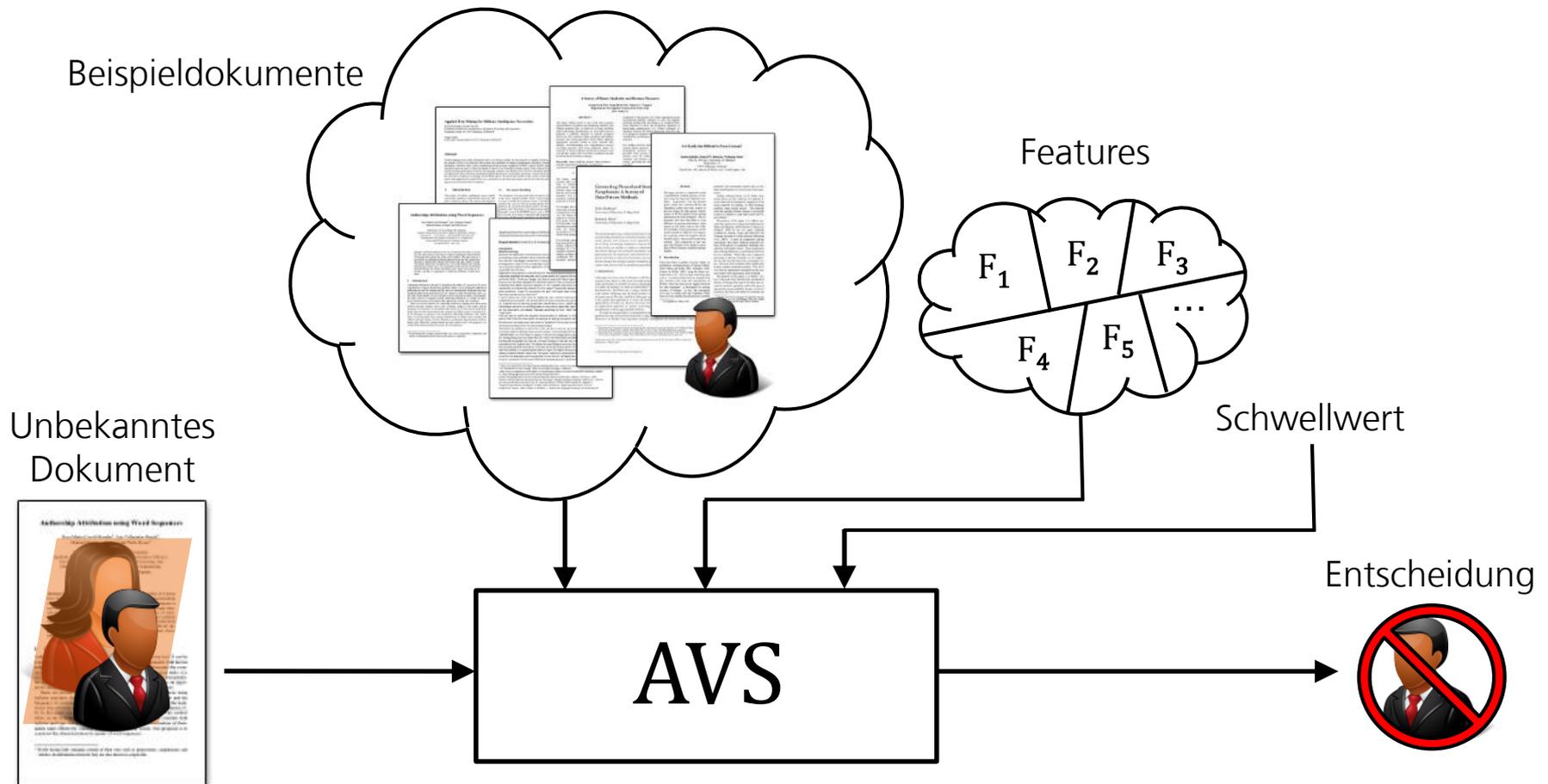
- Die Autorschaftsattribuion (AA) stellt eine Möglichkeit dar, Autoren anonymer Texte zu demaskieren
- Die Autorschaftsverifikation (AV) stellt dagegen eine Möglichkeit dar, angebliche Autoren auf ihre Echtheit hin zu überprüfen
- **Voraussetzung für beide:** Referenztexte von bekannten Autoren (nötig um Schreibstile voneinander unterscheiden zu können)
- Bei AA wird zwischen **Closed-Set** und **Open-Set** unterschieden

↓ ↓
Wahrer Autor befindet sich **garantiert in** / **vielleicht in** der Referenzmenge

AUTORSCHAFTS-ATTRIBUTION / VERIFIKATION

- AV ist deutlich komplizierter zu lösen als AA (per se **Open-Set**)
→ Extrembeispiel: „Ein Autor gegen den Rest der Menschheit“
- AV stellt ein Spezialfall der AA dar, anders gesagt: Ist AV gelöst, so ist damit jedes erdenkliche AA-Problem ebenfalls gelöst
- Nach unserer Erfahrung ist AV in der Forensik viel öfters anzutreffen

UNSER VERFAHREN: SCHEMATISCHES MODELL



UNSER VERFAHREN: FEATURES

- Stilistische Merkmale (kurz: „Features“) sind die Grundlage jeder AVS
- Features werden mit den einfachsten Mitteln (reguläre Ausdrücke, Wörterbücher, etc.) aus der Textoberfläche extrahiert
- Damals kamen komplexe Ansätze aus dem Natural Language Processing zum Einsatz...
- ...diese waren jedoch weder erfolgversprechend noch effizient bzgl. ihrer Laufzeit und spielen heutzutage praktisch keine Rolle mehr

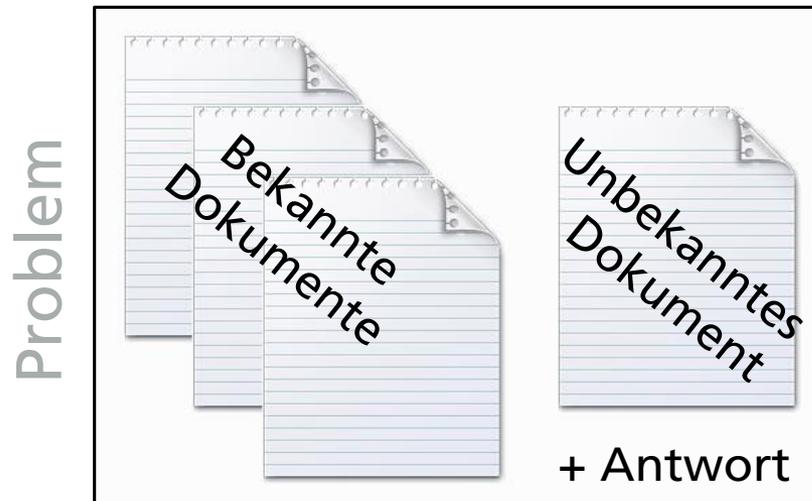
UNSER VERFAHREN: FEATURES

F_i	Name	Feature-Beschreibung	Parameter
F_1	Interpunktionszeichen n -Gramme	Eine Folge von n überlappenden Interpunktionszeichen (Kommas, Bindestriche, etc.)	$n \in \{1, 2, \dots, 10\}$
F_2	Zeichen n -Gramme	Eine Folge von n überlappenden Zeichen	$n \in \{1, 2, \dots, 10\}$
F_3	$n\%$ häufigsten Token	Die $n\%$ häufigsten Tokens (bei $n \approx 30$ sind die Tokens zumeist Funktionswörter).	$n \in \{5, 10, \dots, 50\}$
F_4	Token k -Präfixe	Die ersten k Zeichen eines Tokens	$k \in \{1, 2, 3, 4\}$
F_5	Token k -Suffixe	Die letzten k Zeichen eines Tokens	$k \in \{1, 2, 3, 4\}$
F_6	Token k -Präfix n -Gramme	Die ersten k Zeichen von jedem Token, innerhalb eines Token n -Gramms	$n \in \{2, 3, 4\},$ $k \in \{1, 2, 3, 4\}$
F_7	Token k -Suffix n -Gramme	Die letzten k Zeichen von jedem Token, innerhalb eines Token n -Gramms	$n \in \{2, 3, 4\},$ $k \in \{1, 2, 3, 4\}$
F_8	n -Präfixe- k -Suffix	Die ersten n und die letzten k Zeichen eines Tokens	$n, k \in \{1, 2, 3, 4\}$
F_9	n -Suffixe- k -Präfixe	Die letzten n Zeichen eines Tokens und die ersten k Zeichen eines Folge-Tokens	$n, k \in \{1, 2, 3, 4\}$

Beispiel für Zeichen n -Gramme: **Halvani** \rightarrow (**Hal**, **alv**, **lva**, **van**, **ani**)

UNSER VERFAHREN: TRAININGSKORPORA

- Neben Features benötigt unser AVS zusätzlich für jede Sprache jeweils einen Trainingskorpus...
- Ein Trainingskorpus ist dabei eine einfache Kollektion von sogenannten „Problems“ → Jedes Problem hat dabei die folgende Gestalt:



UNSER VERFAHREN: TRAININGSKORPORA

- Insgesamt setzen wir fünf Trainingskorpora ein (je Sprache = ein Korpus)

1. Niederländisch (NL)
2. Englisch (EN)
3. Griechisch (GR)
4. Spanisch (SP)
5. Deutsch (DE)

- Die Dokumente innerhalb dieser Trainingskorpora erstrecken sich über verschiedene Genres (Foren-Postings, Nachrichtenartikel, Romane, etc.)

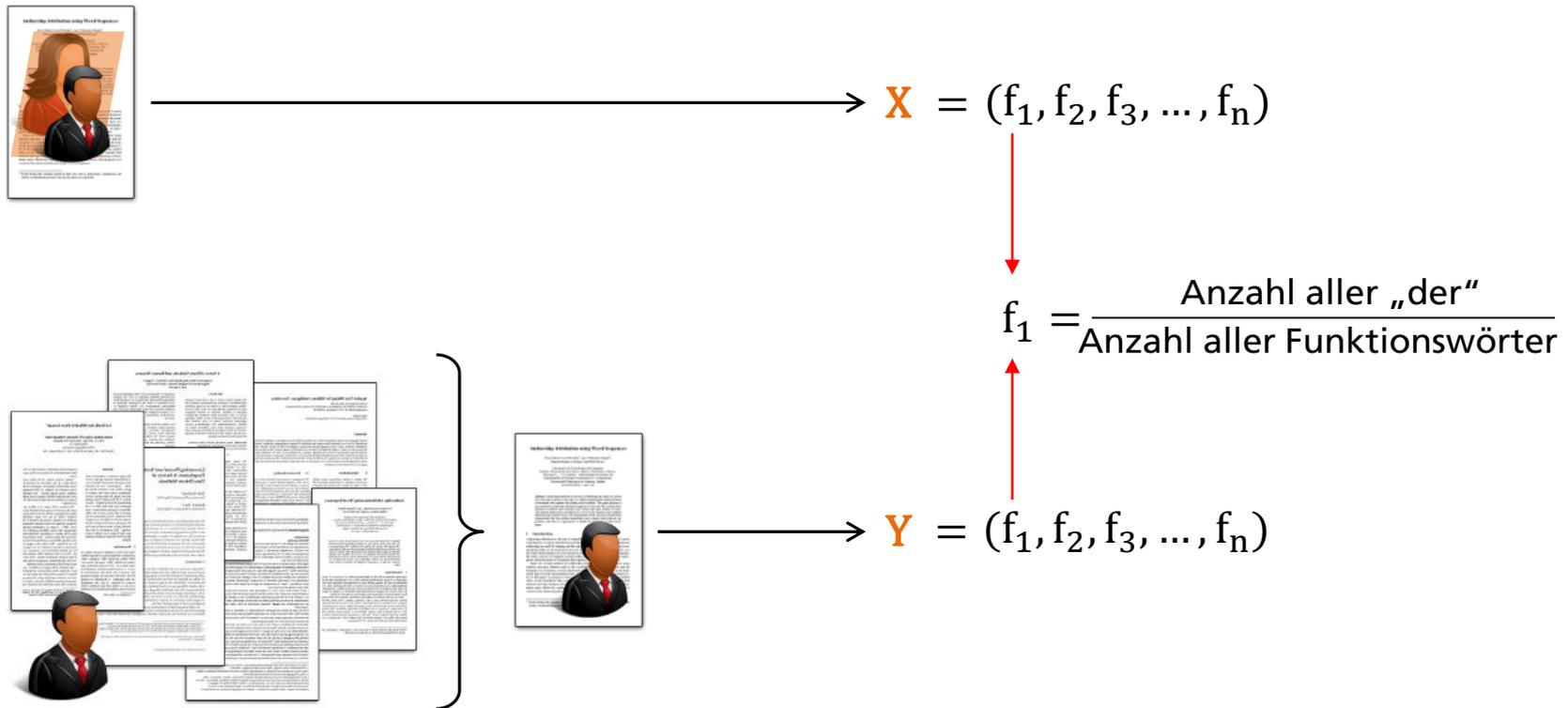
UNSER VERFAHREN: LERNPHASE

- Lernphase unseres AVS:

- (1) Für jedes Problem: Erzeuge korrespondierende Feature-Vektoren und berechne den paarweisen Ähnlichkeitswert
- (2) Für jeden Trainingskorpus: Ermittle den EER-Schwellwert und auf dessen Basis die optimalen Parameter für die einzelne Feature-Kategorien
- (3) Anhand der Parameter: Ermittle das optimale Ensemble (Kombination von Feature-Kategorien)

UNSER VERFAHREN: LERNPHASE → SCHRITT 1.)

- Für jedes Problem: Erzeuge korrespondierende Feature-Vektoren und berechne den paarweisen Ähnlichkeitswert



UNSER VERFAHREN: LERNPHASE → SCHRITT 1.)

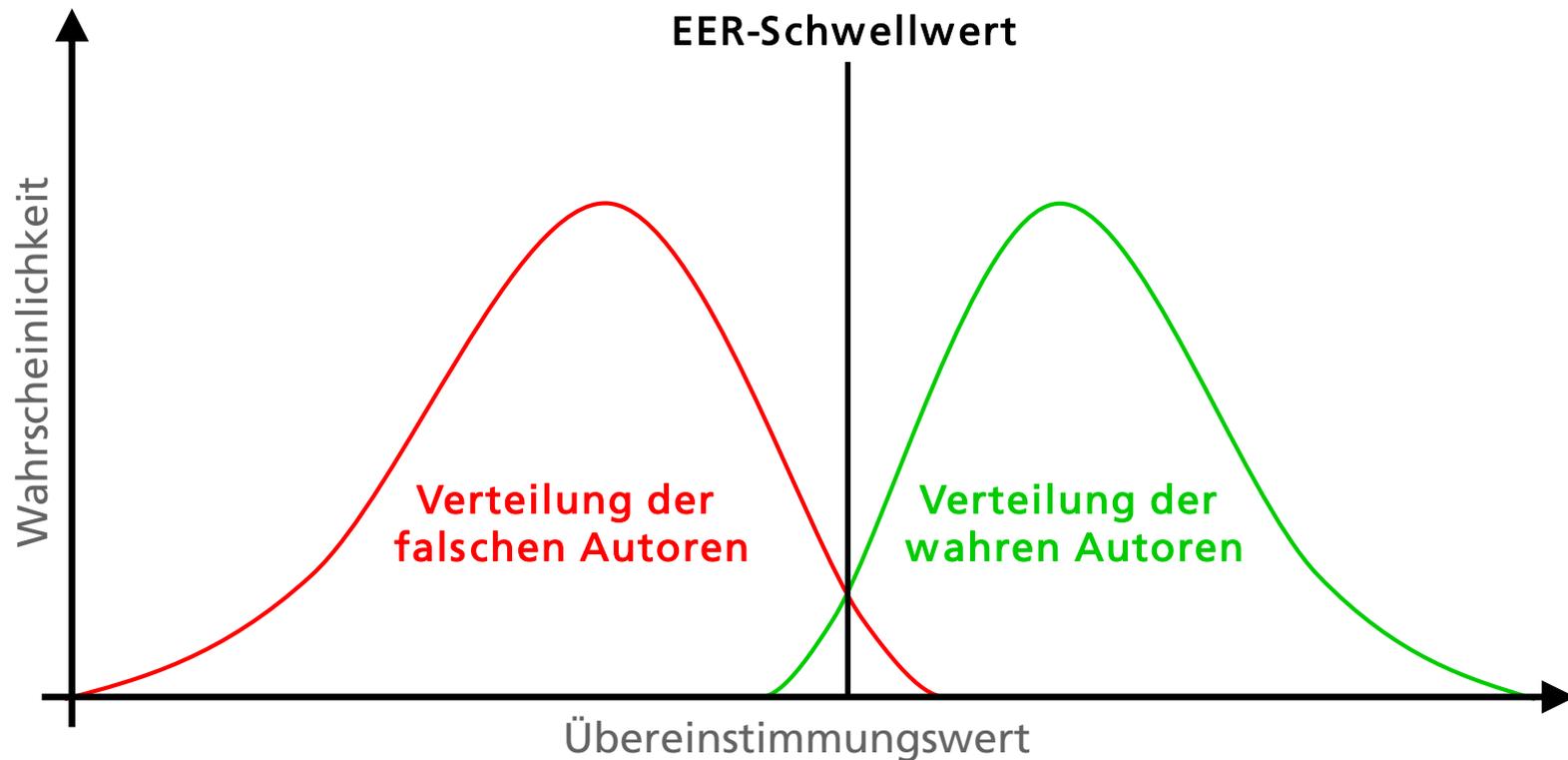
- Für jedes Problem: Erzeuge korrespondierende Feature-Vektoren und berechne den paarweisen Ähnlichkeitswert

$$\text{sim}(X, Y) = \frac{1}{1 + \sum_{i=1}^n (|x_i - y_i|)}$$

- Interpretation: Je mehr sich der resultierende Wert der 1 nähert, desto ähnlicher ist der Schreibstil zwischen **X** und **Y**

UNSER VERFAHREN: LERNPHASE → SCHRITT 2.)

- Für jeden Trainingskorpus: Ermittle den EER-Schwellwert und auf dessen Basis die optimalen Parameter für die einzelne Feature-Kategorien



UNSER VERFAHREN: LERNPHASE → SCHRITT 2.)

- Für jeden Trainingskorpus: Ermittle den EER-Schwellwert und auf dessen Basis die optimalen Parameter für die einzelne Feature-Kategorien
- Alle Problems werden anhand ihrer Ähnlichkeitswerte auf die Vorhersagen Y / N abgebildet. Anhand dieser werden die wahren Antworten gegenvergleichen
- Die Abbildung eines Problems auf dessen Vorhersage erfolgt durch:

$$\text{Classify}(\rho) = \begin{cases} Y & \text{Falls } \text{sim}_\rho > \text{EER-Schwellwert} \\ N & \text{sonst} \end{cases}$$

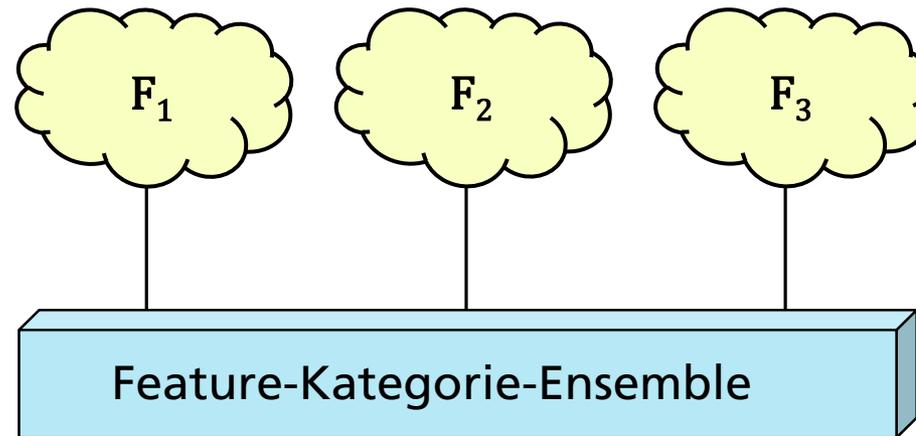
UNSER VERFAHREN: LERNPHASE → SCHRITT 2.)

- Für jeden Trainingskorpus: Ermittle den EER-Schwellwert und auf dessen Basis die optimalen Parameter für die einzelne Feature-Kategorien
- Nachdem Feature-Vektoren, Ähnlichkeitswerte, Abbildungen und EER-Schwellwerte pro Feature-Kategorie und Trainingskorpus berechnet wurden...
- ...gilt es nun optimale (Feature-Kategorie-)Parameter gegenüber dem Accuracy-Maß zu maximieren:

$$\text{Accuracy} = \frac{\text{Anzahl richtiger Antworten}}{\text{Anzahl aller Problems im Korpus}}$$

UNSER VERFAHREN: LERNPHASE → SCHRITT 3.)

- Nachdem optimale Parameter für jede Feature-Kategorie gelernt wurden, kombinieren wir Feature-Kategorien zu sogenannten Ensembles...



- Sämtliche Kombinationen werden berechnet und gegenüber der Accuracy maximiert. Das Ensemble mit der höchsten Accuracy wird anschließend ausgewählt → **Damit ist die Lernphase beendet !**

EVALUIERUNG

- Um das AVS evaluieren zu können, haben wir 28 Testkorpora zusammengestellt (Format analog zu dem der Trainingskorpora)
- Diese enthalten Texte aus den unterschiedlichsten Genres und Themengebiete, verstreut über insgesamt fünf Sprachen
- Anzahl der Problems je Sprache liegt im zwei bis vierstelligem Bereich
- Die Texte sind teilweise stark verrauscht (z.B. bei Social News) bis hochqualitativ (Online-Zeitungsartikel, Offline-Magazine, etc.)

EVALUIERUNG: TESTKORPORA



The Telegraph



🏠 πολιτική οικονομία γνώμες blogs κόσμος science κοινωνία αθλητισμός πολιτισμός ΒΗΜΑtv

TO BHMA

DIE ZEIT



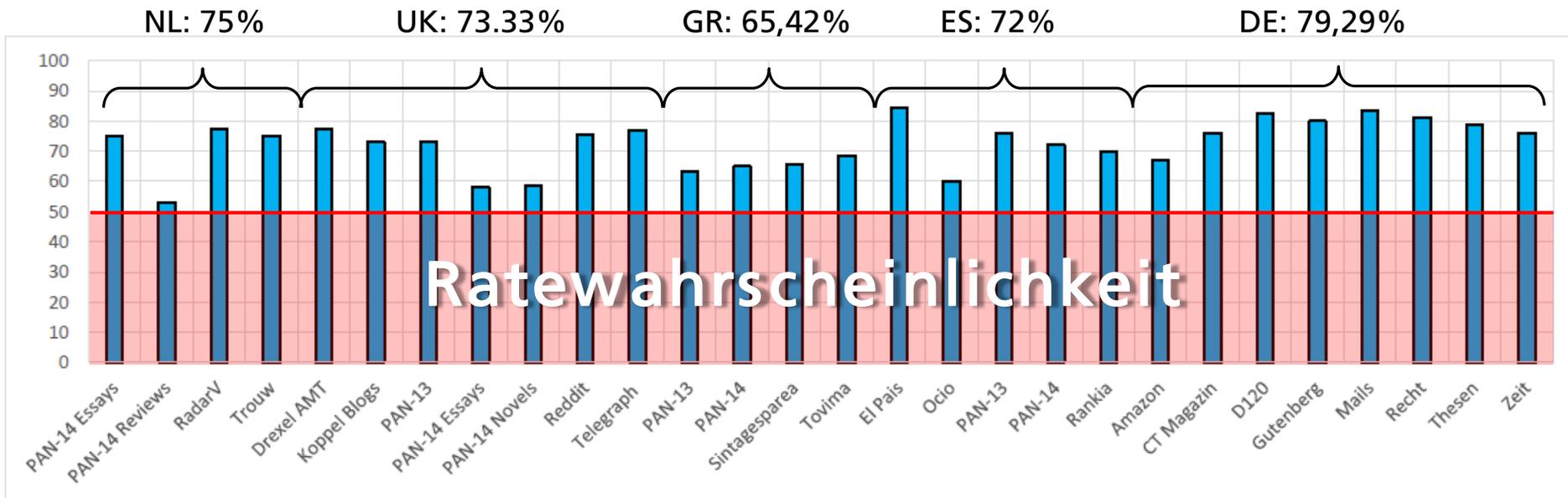
Rankia



EL PAÍS

EVALUIERUNG: RESULTATE

- Die Resultate hinsichtlich der 28 Testkorpora sehen wie folgt aus:



- Im Schnitt liegen die Ergebnisse unseres AVS deutlich über der Ratewahrscheinlichkeit von 50%

EVALUIERUNG: PAN-2015

- Das AVS wurde nicht nur von uns, sondern auch von unabhängigen Experten eines international ausgerichteten Wettbewerbs evaluiert...



PAN.Webis.de

PAN 2015 ◀ ▶

This is the 13th evaluation lab on uncovering plagiarism, authorship, and social software misuse. PAN will be held as part of the [CLEF conference](#) in Toulouse, France, on September 8-11, 2015. Evaluations will commence [from January till June](#). We invite you to take part in any of the three tasks shown below.

[Learn more »](#)

[Register now »](#)

158 already signed up

Plagiarism Detection

Given a document, is it an original?

This task is divided into **source retrieval** and **text alignment**. Source retrieval is about searching for likely sources of a suspicious document. Text alignment is about matching passages of reused text between a pair of documents.

Author Identification

Given a document, who wrote it?

This task focuses on **authorship verification** and methods to answer the question whether two given documents have the same author or no. This question accurately emulates the real-world problem that most forensic linguists face every day.

Author Profiling

Given a document, what're its author's traits?

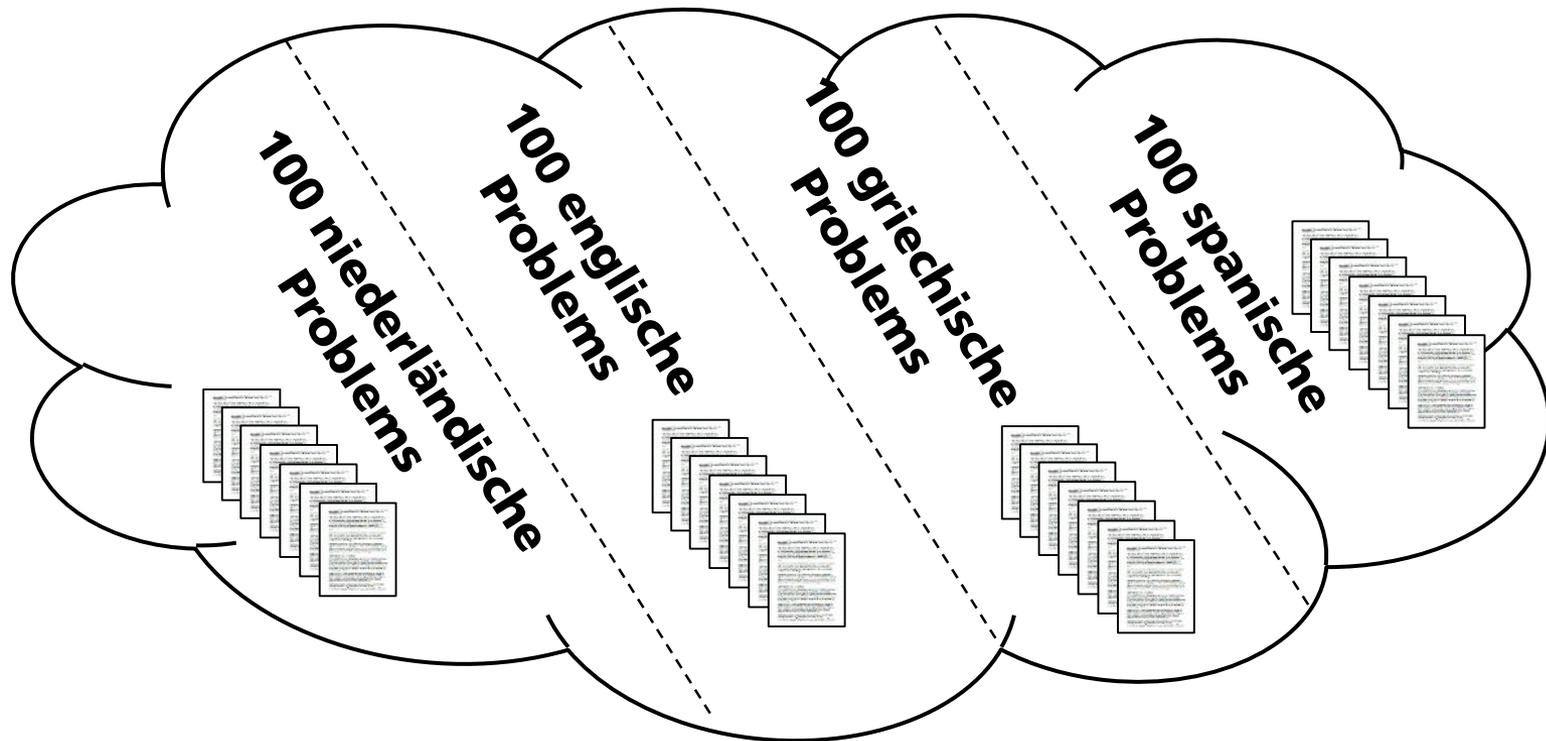
This task is concerned with predicting an author's demographics from her writing. For example, an author's style may reveal her **age, gender, and personality**.



Sponsor

EVALUIERUNG: PAN-2015

- Grundlage der Evaluierung war der öffentlich zugängliche PAN-AI-2015 Trainingskorpus mit insgesamt 400 Problems, verteilt auf vier Sprachen



EVALUIERUNG: PAN-2015

PAN-2015 Corpus

Quelle: PAN15-AI-Overview Slides

	Language	Type	#Problems	#Docs	Avg. known docs per problem	Avg. words per document
Training	Dutch	cross-genre	100	276	1.76	354
	English	cross-topic	100	200	1.00	366
	Greek	cross-topic	100	393	2.93	678
	Spanish	mixed	100	500	4.00	954
Evaluation	Dutch	cross-genre	165	452	1.74	360
	English	cross-topic	500	1000	1.00	536
	Greek	cross-topic	100	380	2.80	756
	Spanish	mixed	100	500	4.00	946
TOTAL			1265	3701	1.93	641

All corpora are balanced (positive/negative problems)

EVALUIERUNG: PAN-2015

- Resultate des PAN2015 Wettbewerbs →
- Grau eingefärbte Zeilen stellen Vergleichsmethoden aus den letzten zwei Jahren dar
- Zu erkennen: Unser AVS verhält sich robust über alle vier Sprachen hinweg, was bei den meisten nicht der Fall ist

Quelle: PAN15-AI-Overview Paper

Rang	Teilnehmer	Sprachen				Schnitt
		NL	EN	GR	SP	
1	Bagnall	0,451	0,614	0,75	0,721	0,628
2	Moreau et al.	0,635	0,453	0,693	0,661	0,606
3	Pacheco et al.	0,624	0,438	0,517	0,663	0,558
4	Huerlimann et al.	0,616	0,412	0,599	0,539	0,538
–	PAN15-ENSEMBLE	0,426	0,468	0,537	0,715	0,532
5	Bartoli et al.	0,518	0,323	0,458	0,773	0,506
6	Gutierrez et al.	0,329	0,513	0,581	0,509	0,478
7	Halvani	0,455	0,458	0,493	0,441	0,462
8	Kocher & Savoy	0,218	0,508	0,631	0,366	0,416
–	PAN14-BASELINE-2	0,191	0,409	0,412	0,683	0,405
9	Maitra et al.	0,518	0,347	0,357	0,352	0,391
10	Castro-Castro et al.	0,247	0,52	0,391	0,329	0,365
–	PAN13-BASELINE	0,242	0,404	0,384	0,367	0,347
11	Gomez-Adorno et al.	0,39	0,281	0,348	0,281	0,323
–	PAN14-BASELINE-1	0,255	0,249	0,198	0,443	0,28
12	Sari & Stevenson	0,381	0,201	-	0,485	0,25
13	Pimas et al.	0,262	0,257	0,23	0,24	0,247
14	Solorzano et al.	0,153	0,259	0,33	0,218	0,235
15	Posadas-Duran et al.	0,132	0,4	-	0,462	0,226
16	Nikolov et al.	0,089	0,258	0,454	0,095	0,201
17	Vartapetian & Gillam	0,262	-	0,212	0,348	0,201
18	Mechti et al.	-	0,247	-	-	0,063

BEOBACHTUNGEN

- Vernünftige Ergebnisse (> 75%) gehen mit steigender Anzahl von Trainingskorpora einher
- Um derartige Ergebnisse zu erzielen braucht es ca. 4–5 kByte (ca. eine DIN-A4 Seite) reinem Textumfang
- Problems mit stark themenfremden Texten und viel Textauschen (z.B. Amazon-Produktreviews) stellen eine Herausforderung dar. Problems mit unterschiedlichen Genres dagegen werden großteils korrekt erkannt
- Als Features sind Zeichen-n-Gramme aktuell ungeschlagen

ZUSAMMENFASSUNG

- **Robust:** Das AVS liefert bzgl. verschiedener Sprachen, Genres oder Themengebiete brauchbare und insbesondere stabile Ergebnisse
- **Praxistauglich:** Die Ergebnisse des AVS wurden in „echten“ forensischen Fällen eingesetzt
- **Portabel:** Einmal gelernte Parameter und EER-Schwellwerte müssen nicht neu trainiert werden (außer es kommen neue Sprachen hinzu)
- **Geringe Laufzeit:** Das AVS macht ausgiebig Gebrauch von Parallelität. Dadurch ist die Abarbeitung eines Problems in < 1 Sekunde möglich
- **Konfigurierbarkeit:** Alles kann ersetzt, modifiziert, kombiniert, ausgetauscht oder erweitert werden (Distanzfunktion, Feature-Kategorien Schwellwertbestimmung, etc.)

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT...



Oren.Halvani@SIT.Fraunhofer.de