# Domänenübergreifende profilbasierte Autorschafts-Attribution

Oren Halvani · Martin Steinebach

Fraunhofer-Institut für Sichere Informationstechnologie
{Oren.Halvani | Martin.Steinebach}@SIT.Fraunhofer.de

## Zusammenfassung

Wir präsentieren in dieser Arbeit einen profilbasierten Ansatz für die automatische Autorschafts-Attribution von anonymen Texten, die aus unterschiedlichen Domänen stammen. Die Autorschafts-Attribution ist dabei eine forensisch linguistische Anwendung, die primär das Schutzziel Authentizität gewährleistet und gleichzeitig das Schutzziel Anonymität aufhebt. Als Ausgangssprache für unsere Untersuchung betrachten wir ausschließlich deutschsprachige Texte. Der Grund dafür ist, dass die meiste Forschungsarbeit auf dem Gebiet der Autorschafts-Attribution im englischsprachigen Raum angesiedelt ist und bisher kaum wissenschaftliche Beiträge für die deutsche Sprache existieren. Unser Ansatz stellt ein zweistufiges Verfahren dar, welches in der ersten Stufe zunächst weniger relevante Autoren aus einer festgelegten Trainingsmenge filtert, um anschließend in der zweiten Stufe die Attribution auf die verbliebene Menge durchzuführen. Wir zeigen in unseren Experimenten, dass der Ansatz eine erfolgreiche automatische Attribution über unterschiedliche Domänen hinweg ermöglichen kann. Dabei testen wir mehrere Parametrisierungen, um eine bestmögliche Performanz zu erzielen.

## 1 Einleitung

In der heutigen Welt existieren unzählige Informationen, die zumeist in elektronischer Form vorliegen. Viele von ihnen entstehen dabei im Internet oder werden dort vertrieben. Zu finden sind sie unter anderem in Foren, Blogs, (Produkt-)Bewertungen, Erfahrungsberichten oder Kommentierungen in sozialen Netzwerken.

Ein Problem, welches die meisten dieser textuellen Informationen anbelangt, ist die fehlende, anonyme oder nichteindeutige Autorschaft. Der Grund dafür ist, dass Autorschaften von Texten zumeist auf Metadaten beruhen, die unterschiedlich ausgeprägt sind. Dazu zählen beispielsweise Dateinamen, Unterschriften oder auch eingebettete Dokumenteigenschaften wie Benutzername, Urheberinformation, E-Mail-Adresse, etc. In der Praxis kommt es dabei häufig vor, dass solche Metadaten entweder nicht vorhanden sind oder bei entsprechender Präsenz leicht editiert bzw. entfernt werden können (siehe z.B. [Bach12]).

In der Forschung haben sich dazu unterschiedliche Disziplinen etabliert, die das Ziel verfolgen, Autorschaften von Texten (unabhängig von assoziierten Metadaten) zu analysieren. Sinnvollerweise werden diese Disziplinen daher unter dem Oberbegriff „Autorschaftsanalyse“ zusammengefasst, wobei die Autorschafts-Attribution die bekannteste Unterdisziplin darstellt. Nach [Stam09] geht es bei dieser Disziplin darum, zu einem gegebenen Dokument eines unbekannten Autors den stilistisch ähnlichsten Autor aus einer festgelegten Trainingsmenge zuordnen zu können. Innerhalb dieser Trainingsmenge existiert für jeden Autor eine bestimmte Anzahl von Beispieltexten, die dessen „Stil-Repertoire“ darstellen.

P. Schartner · J. Taeger (Hrsg.) · D•A•CH Security 2012 · syssec (2012) 174-185.

Eine korrekte Attribution ist dabei genau dann gegeben, wenn es sich bei dem stilistisch ähnlichsten Autor auch gleichzeitig um den wahren Autor des anonymen Dokuments handelt. Hierfür wird die Grundannahme vorausgesetzt, dass die Trainingsmenge Beispieltexte des wahren Autors enthält.

Realisiert werden Attributions-Verfahren in der Regel unter Zuhilfenahme stilometrischer Methoden, deren Zielsetzung es ist, relevante Features (Stilmerkmale) innerhalb der zu untersuchenden Texte ausfindig zu machen. Anhand solcher Features können Autorenstile approximiert und anschließend mit Hilfe von Klassifikatoren z.B. Naive Bayes, Support Vector Machines oder $k-$Nearest Neighbours unterschieden werden. Alternativ dazu können auch Ähnlichkeits- bzw. Distanzfunktionen wie z.B. der Dice-Koeffizient oder die euklidische Distanz eingesetzt werden, um die Attribution durchzuführen. Letztere kommen in dieser Arbeit zur Anwendung.

## 1.1 Notation

Im Verlauf dieser Arbeit werden einige Begriffe des Öfteren wiederverwendet. Aus Gründen der Lesbarkeit werden diese daher wie folgt symbolisiert:

**Tab. 1:** Notation in dieser Arbeit

| Symbol | Erläuterung |
|---|---|
| $\mathcal{A}$, $\varepsilon$ | Bezeichnet einen bekannten $(\mathcal{A})$ bzw. anonymen $(\varepsilon)$ Autor. |
| $\mathcal{D}$ | Bezeichnet einen bekannten $(\mathcal{D}_{\mathcal{A}})$ bzw. anonymes $(\mathcal{D}_{\varepsilon})$ Dokument. |
| $f$ | Bezeichnet ein Feature (z.B. $f_1 =$ „Substantive mit vier aufeinanderfolgenden Vokalen“). |
| $f(\mathcal{D})$ | Bezeichnet die Anwendung eines Features $f$ auf das Dokument $\mathcal{D}$. Es gilt $f(\mathcal{D}) \in \mathbb{R}$. |
| $F$ | Bezeichnet eine Kategorie von Features, es gilt $F = \{ f_1, f_2, \dots \}$ |
| $\mathcal{F}$ | Bezeichnet einen Feature-Vektor. Es gilt $\mathcal{F} = \big( f_1(D), f_2(D), \dots \big)$ |
| $\mathbb{S}_{train}$ | Bezeichnet eine Menge von Trainingsdokumenten. |
| $\mathcal{K}$ | Bezeichnet ein (Dokumenten-)Korpus. |

## 1.2 Bezug zu IT-Sicherheit

Die Autorschafts-Attribution adressiert in erster Linie die Schutzziele Anonymität und Authentizität. Da in der Literatur jedoch keine Einstufung dieser Disziplin in die der IT-Sicherheit gefunden wurde, wird im Folgenden versucht, eine mögliche Einordnung über Umwege zu beschreiben. Nach Heine [Hein10, Seiten: 77–78] kann die Autorschafts-Attribution zunächst als eine Teildisziplin der forensischen Linguistik verstanden werden. Laut [Fobb11] beschäftigt sich diese mit der Analyse von mündlichen und schriftlichen Texten, die Gegenstand einer polizeilichen Ermittlung oder eines gerichtlichen Verfahrens sind. Damit stellt die forensische Linguistik ein Anwendungsgebiet der Forensik dar.

Da in der forensischen Linguistik Methoden und Konzepte zum Einsatz kommen, die aus dem Gebiet der künstlichen Intelligenz stammen (Feature-Selektion, Klassifikation, etc.), kann hier ein Bezug zur Informatik festgestellt werden. Hierdurch lässt sich diese Disziplin anstatt in die allgemeine Form der Forensik in die spezialisierte Form der IT-Forensik kategorisieren, welche Informatik und Forensik in sich vereint. Die IT-Forensik kann wiederum nach [Ploe11, Seite: 8] als ein Teilgebiet der IT-Sicherheit verstanden werden.

Wie oben bereits erwähnt, kann die Verbindung der Autorschafts-Attribution auch direkt über die von ihr adressierten Schutzziele hergestellt werden. Hierbei ist insbesondere das Schutzziel Authentizität zu nennen. Diese hat das Ziel, die Echtheit und Glaubwürdigkeit eines Objekts (1) anhand einer eindeutigen Identität und charakteristischen Eigenschaft (2) überprüfbar zu machen, [Ecke09]. Im Kontext der Autorschafts-Attribution handelt es sich bei (1) um den Text eines Autors, während (2) die Stilmerkmale, die den Schreibstil des Autors repräsentieren, darstellt. Die Authentizität wird dahingehend erfüllt, indem die Identität des Dokuments anhand der Attribution überprüfbar gemacht wird. Wichtig ist hier jedoch zu erwähnen, dass die „eindeutige Identität“ auf ein statistischen Wert beruht und daher nicht hundertprozentig gewährleistet werden kann.

Mit der Erfüllung der Authentizität wird gleichzeitig das Schutzziel Anonymität aufgehoben. Diese hat das Ziel, die Zuordnung von Daten zu bestimmten Personen (bzw. deren Identifizierung) erheblich zu erschweren oder gänzlich unmöglich zu machen [BeAc10]. Da die Autorschafts-Attribution eine Demaskierung von Pseudonymen erlaubt, ist die Anonymität eines Dokuments damit nicht mehr gewährleistet.

# 2 Problemstellung

Die Autorschafts-Attribution stellt ein interdisziplinäres Wissenschaftsfeld dar, welches seit nun mehr als einem Jahrhundert erforscht wird. Beteiligt sind hierbei unter anderem Konzepte und Verfahren aus der Linguistik, Kognitionspsychologie, Informatik oder auch Mathematik.

Trotz langjähriger und vielfältiger Forschungsarbeit auf diesem Gebiet fällt es auf, dass die meisten wissenschaftlichen Beiträge sich nur auf die Untersuchung von englischsprachigen Texten beschränken. Dabei sollte bedacht werden, dass diese Disziplin ein sprachabhängiges Problem darstellt, sodass ein Autorschafts-Attributions System (kurz `AAS`), welches für die Analyse englischer Texte entwickelt wurde, in der Regel nicht auf deutsche Texte angewendet werden kann. Dies liegt vor allem daran, dass in einer `AAS` häufig eingesetzte Sprachkomponenten wie etwa Tokenizer oder Parser nicht sprachübergreifend funktionieren und dadurch Features nicht gewonnen werden können, [KPCT03]. Ohne Features ist wiederum keine Attribution seitens des `AAS` möglich, da sie es sind, die die Autorenstile repräsentieren.

Neben diesem Problem wird ebenso bemerkt, dass zahlreiche wissenschaftliche Beiträge eine domänenspezifische Herangehensweise verfolgen. Diese äußert sich dadurch, dass Features vorgeschlagen werden, die für einzelne Domänen (Romane, Nachrichtentexte, Filmrezensionen, etc.) hohe Erkennungsraten bei der Bestimmung der Autoren erzielen, jedoch bei Anwendung auf andere Domänen wie etwa E-Mails oder wissenschaftliche Ausarbeitungen aufgrund sprachlicher Register oder Genre niedrige Erkennungsraten aufweisen bzw. gänzlich fehlschlagen. Hinzu kommt noch, dass dabei öfters auf eine Erklärung verzichtet wird, aus welchem Grund sich Features für eine bestimmte Domäne nicht eignen.

# 3 Methodik

In dieser Arbeit wird eine domänenübergreifende Autorschafts-Attribution realisiert, die die Punkte aus dem vorherigen Kapitel aufgreift. Dazu wird ein profilbasiertes Verfahren vorgestellt, welches auf insgesamt sieben (deutschsprachige) Korpora unterschiedlicher Domänen angewendet wird. Hierfür werden für jeden Korpus die selben Features verwendet, um dadurch zu vergleichen, ob eine Domänenunabhängigkeit erreicht werden kann.

Das Verfahren basiert dabei auf zwei Stufen, bei der zunächst eine Ähnlichkeitsfunktion verwendet wird, um die Trainingsmenge nach potentiellen Autoren zu filtern. Anschließend wird mit Hilfe einer Distanzfunktion die Attribution auf die verbliebenen Dokumente der potentiellen Autoren durchgeführt. Das Resultat des Verfahrens ist derjenige Autor, dessen Stil am ähnlichsten zu dem des anonymen Dokuments ist.

Der vorgestellte Ansatz lässt sich dank einer eindeutigen mathematischen Formalisierung leicht nachimplementieren und liefert für ein Großteil der getesteten Korpora erfolgsversprechende Ergebnisse. Maßgebend für eine erfolgreiche Attribution der Autorschaften ist dabei die gewählte Parametrisierung, auf die in den Experimenten genauer eingegangen wird.

# 4 Attributions Verfahren

In diesem Kapitel wird der sogenannte „Two-Stage Pairwise-Similarity“ Verfahren von [Halv12] beschrieben. Dazu werden zunächst einige benötigte Komponenten und damit verbundene Verarbeitungsschritte erläutert.

## 4.1 Benötigte Komponenten

Sei $\mathbb{D}$ eine Menge von $r$ Dokumenten und $\mathbb{A}$ eine Menge von $m$ Autoren, die diese Dokumente verfasst haben, gegeben. Ein Tupel $(\mathcal{D}_i, \mathcal{A}_j) \in \mathbb{D} \times \mathbb{A}$ beschreibt die Assoziation zwischen einem Dokument $\mathcal{D}_i$ und dessen Autor $\mathcal{A}_j$. Zur Vereinfachung wird anstelle des Tupels die Notation $\mathcal{D}_{i\mathcal{A}_j}$ verwendet. Damit kann nun eine Trainingsmenge $\mathbb{S}_{train}$ wie folgt definiert werden:

$$\mathbb{S}_{train} = \{\, \mathcal{D}_{1\mathcal{A}_1}, \mathcal{D}_{2\mathcal{A}_1}, \ldots, \mathcal{D}_{1\mathcal{A}_2}, \mathcal{D}_{2\mathcal{A}_2}, \ldots, \mathcal{D}_{r-1\mathcal{A}_m}, \mathcal{D}_{r\mathcal{A}_m} \,\}, \text{ mit } m = |\mathbb{A}| \text{ und } r > m$$

Hierbei repräsentiert jedes $\mathcal{D}_{i\mathcal{A}_j}$ das $i-$te Dokument eines zugehörigen $j-$ten Autors $\mathcal{A}_j$, wobei von jedem $\mathcal{A}_j \in \mathbb{A}$ eine bestimmte Anzahl an Dokumenten existiert.

## 4.2 Verarbeitungsschritte

Wie eingangs erwähnt, handelt es sich bei dem vorgestellten Verfahren um einen profilbasierten Ansatz. Dieser erfordert, dass sämtliche $r$ Dokumente der insgesamt $m$ Autoren zu sogenannten Autor-Profilen konkateniert (aneinandergefügt) werden, [Stam09, Seite: 13]. Realisiert wird dies, indem die Dokumente zunächst nach ihren Autoren wie folgt sortiert werden:

$$\begin{aligned}
\mathbb{D}_{\mathcal{A}1} &= \{\, \mathcal{D}_{1\mathcal{A}_1}, \mathcal{D}_{2\mathcal{A}_1}, \ldots, \,\} \\
\mathbb{D}_{\mathcal{A}2} &= \{\, \mathcal{D}_{1\mathcal{A}_2}, \mathcal{D}_{2\mathcal{A}_2}, \ldots, \,\} \\
&\vdots \\
\mathbb{D}_{\mathcal{A}m} &= \{\, \mathcal{D}_{1\mathcal{A}_m}, \mathcal{D}_{2\mathcal{A}_m}, \ldots, \,\}
\end{aligned}$$

Somit ensteht für jeden Autor $\mathcal{A}_j$ eine Dokumentenmenge $\mathbb{D}_{\mathcal{A}_j}$ die dessen Dokumente enthält. Die Menge $\mathbb{S}_{train}$ kann umgeformt werden in:

$$\mathbb{S}_{train} = \mathbb{D}_{\mathcal{A}_1} \cup \mathbb{D}_{\mathcal{A}_2} \cup \ldots \cup \mathbb{D}_{\mathcal{A}_m}$$

Der nächste Schritt besteht darin, sämtliche $\mathcal{D}_{i\mathcal{A}_j} \in \mathbb{D}_{\mathcal{A}_j}$ zu einem Autor-Profil $\mathcal{D}_{BIG\mathcal{A}_j}$ wie folgt zu konkatenieren:

$$\mathcal{D}_{BIG\mathcal{A}_1} = \mathcal{D}_{1\mathcal{A}_1} \circ \mathcal{D}_{2\mathcal{A}_1} \circ \ldots \circ \mathcal{D}_{\ell_1\mathcal{A}_1} \text{ für } \ell_1 = |\mathbb{D}_{\mathcal{A}_1}|$$
$$\mathcal{D}_{BIG\mathcal{A}_2} = \mathcal{D}_{1\mathcal{A}_2} \circ \mathcal{D}_{2\mathcal{A}_2} \circ \ldots \circ \mathcal{D}_{\ell_2\mathcal{A}_2} \text{ für } \ell_2 = |\mathbb{D}_{\mathcal{A}_2}|$$
$$\vdots$$
$$\mathcal{D}_{BIG\mathcal{A}_m} = \mathcal{D}_{1\mathcal{A}_m} \circ \mathcal{D}_{2\mathcal{A}_m} \circ \ldots \circ \mathcal{D}_{\ell_m\mathcal{A}_m} \text{ für } \ell_m = |\mathbb{D}_{\mathcal{A}_m}|$$

Damit ändert sich die Trainingsmenge $\mathbb{S}_{train}$ in:

$$\mathbb{S}'_{train} = \{\, \mathcal{D}_{BIG\mathcal{A}_1}, \mathcal{D}_{BIG\mathcal{A}_2}, \ldots, \mathcal{D}_{BIG\mathcal{A}_m} \,\}$$

## 4.3 Two-Stage Pairwise-Similarity Ansatz

Ausgehend von $\mathbb{S}'_{train}$ werden nun die zwei Stufen des Two-Stage Pairwise-Similarity Ansatzes wie folgt beschrieben.

### 4.3.1 Erste Stufe

In der ersten Stufe des Verfahrens werden aus dem gegebenen anonymen Dokument $\mathcal{D}_\varepsilon$ sowie jedem Autor-Profil $\mathcal{D}_{BIG\mathcal{A}_j}$ die $k-$häufigsten $n-$Gramme extrahiert. Ein $n-$Gramm stellt dabei einen Ausschnitt einer längeren Zeichenkette dar, welches genau $n$ Einheiten lang ist. Als Einheiten kommen hier unter anderem Buchstaben, Tokens oder Sätze in Frage. Das folgende Beispiel verdeutlicht die Konstruktion von Buchstaben $n-$Gramme, die in der ersten Stufe verwendet werden. Sei das Wort $\omega =$ `Autor` gegeben, dann können aus $\omega$ die folgenden $n-$Gramme gebildet werden:

**Tab. 2:** Beispiel für eine $n-$Gramm Konstruktion

| Größe | Bezeichnung | Ergebniss |
|---|---|---|
| $n = 1$ | Unigramm | `(A),(u),(t),(o),(r)` |
| $n = 2$ | Bigramm | `(Au),(ut),(to),(or)` |
| $n = 3$ | Trigramm | `(Aut),(uto),(tor)` |
| $n = 4$ | Tetragramm | `(Auto),(utor)` |
| $n = 5$ | Pentagramm | `(Autor)` |

Die aus $\mathcal{D}_\varepsilon$ und $\mathcal{D}_{BIG\mathcal{A}_j}$ extrahierten $n-$Gramme werden zunächst in ihre zugehörige Mengen $M_\varepsilon$ und $M_{\mathcal{A}_j}$ abgelegt. Mittels einer mengenbasierten Ähnlichkeitsfunktion:

$$sim : (M_1, M_2) \longrightarrow \{\, s \mid (s \in \mathbb{R}) \wedge (0 \leq s) \wedge (s \leq 1) \,\}$$

werden dann für $M_\varepsilon$ sowie die einzelnen $M_{\mathcal{A}_1}, M_{\mathcal{A}_2}, \ldots, M_{\mathcal{A}_m}$ paarweise Ähnlichkeitswerte $sim_j = sim(M_\varepsilon, M_{\mathcal{A}_j})$ berechnet und mit dem korrespondierenden Autor $\mathcal{A}_j$ in einer Folge gespeichert:

$$Similarities = \Big( (sim_1, \mathcal{A}_1), (sim_2, \mathcal{A}_2), \ldots, (sim_m, \mathcal{A}_m) \Big)$$

Der nächste Schritt besteht darin, diese Folge nach den Ähnlichkeitswerten absteigend zu sortieren, sodass Autoren die in den vorderen Tuplen vorkommen, die zu $\varepsilon$ stilistisch ähnlichsten

Autoren repräsentieren. Anschließend wird $Similarities$ anhand eines Splitparameters `SPLIT` in zwei Teilfolgen $Sub_1, Sub_2$ aufgeteilt, wobei `SPLIT` die prozentuale Größe von $Sub_1$ angibt. Für das Verfahren ist nur $Sub_1$ relevant, sodass $Sub_2$ dagegen verworfen wird. Aus $Sub_1$ werden dann die Namen der Autoren entnommen, die potentielle Kandidaten für $\varepsilon$ darstellen und in der folgenden Menge gespeichert:

$$Candidates = \{\, \mathcal{A}_{c_1},\, \mathcal{A}_{c_2},\, \ldots,\, \mathcal{A}_{c_k} \} \text{ mit } c_i \in \{\, 1,\, 2,\, \ldots,\, |\mathbb{A}| \,\} \text{ und } k = \left\lceil \frac{\texttt{SPLIT} \cdot |\mathbb{A}|}{100} \right\rceil$$

Im letzten Schritt der ersten Stufe werden schließlich sämtliche $\mathcal{A}_{c_i} \in Candidates$ mit ihren Autor-Profilen assoziiert. Das Resultat der ersten Stufe ist somit die folgende Trainingsmenge:

$$\mathbb{S}''_{train} = \{\, \mathcal{D}_{BIG\mathcal{A}_{c_1}},\, \mathcal{D}_{BIG\mathcal{A}_{c_2}},\, \ldots,\, \mathcal{D}_{BIG\mathcal{A}_{c_k}} \,\}$$

### 4.3.2 Zweite Stufe

In der zweiten Stufe werden aus $\mathcal{D}_\varepsilon$ sowie jedem $\mathcal{D}_{BIG\mathcal{A}_j} \in \mathbb{S}''_{train}$ genau $n$ Features entnommen, um damit das anonyme Dokument sowie die Autor-Profile in ihre Feature-Vektor Darstellung $\mathcal{F}_{\mathcal{A}_\varepsilon}$ bzw. $\mathcal{F}_{\mathcal{A}_j}$ zu überführen. Ein Feature-Vektor hat dabei die folgende Gestalt:

$$\mathcal{F} = \big(\, f_1(D),\, f_2(D),\, \ldots, f_n(D) \,\big)$$

Jedes $f_i(D)$ stellt dabei die Anwendung eines Features $f_i$ auf ein und dasselbe Dokument $\mathcal{D}$ dar. Das Resultat dieser Anwendung entspricht einer relativen Häufigkeit, die in der Regel beschreibt, wie häufig $f_i$ in $\mathcal{D}$ vorgekommen ist. Hierbei gilt stets $f_i(D) \in [0; 1]$. Nach der Feature-Vektor Überführung ändert sich die Trainingsmenge in die folgende endgültige Form:

$$\mathbb{S}'''_{train} = \{\, \mathcal{F}_{\mathcal{A}_{c_1}},\, \mathcal{F}_{\mathcal{A}_{c_2}},\, \ldots,\, \mathcal{F}_{\mathcal{A}_{c_k}} \,\}$$

Der nächste Schritt besteht nun darin, mit Hilfe einer beliebigen Distanzfunktion:

$$dist : (\mathcal{F}_1, \mathcal{F}_2) \longrightarrow \big(\, \mathbb{R}^+ \cup \{0\} \,\big)$$

paarweise Stil-Unterscheidungswerte $dist_{c_i} = dist(\mathcal{F}_{\mathcal{A}_\varepsilon}, \mathcal{F}_{\mathcal{A}_{c_i}})$ zwischen den Feature-Vektoren $\mathcal{F}_{\mathcal{A}_\varepsilon}$ und allen $\mathcal{F}_{\mathcal{A}_{c_i}} \in \mathbb{S}'''_{train}$ zu berechnen. Diese werden (analog zu den Ähnlichkeitswerten) mitsamt der dazugehörenden Autoren $\mathcal{A}_{c_i}$ in einer Folge gespeichert:

$$Distances = \Big(\, (dist_{c_1}, \mathcal{A}_{c_1}),\, (dist_{c_2}, \mathcal{A}_{c_2}),\, \ldots,\, (dist_{c_k}, \mathcal{A}_{c_k}) \,\Big)$$

Im letzten Schritt wird $Distances$ anhand der $dist_{c_i}$ aufsteigend sortiert, sodass die Attribution des anonymen Dokuments dadurch erfolgt, dass aus $Distances$ der erste Tupel $(dist_{c_i}, \mathcal{A}_{c_i})$ entnommen wird, der den niedrigsten $dist_{c_i}$ aufweist. Hierbei stellt $\mathcal{A}_{c_i}$ den zu $\varepsilon$ stilistisch ähnlichsten Autor dar, welcher zur Vereinfachung durch $\mathcal{A}_{max}$ symbolisiert wird. Eine Attribution wird insgesamt als korrekt bezeichnet, sofern die folgende Konjunktion gilt:

$$\big(\, \mathcal{A}_{max} = \varepsilon \,\big) \wedge \big(\, \mathcal{A}_{max} = \mathcal{A}_{true} \,\big)$$

Hierbei drückt $\mathcal{A}_{true}$ den wahren Autoren des anonymen Dokuments $\mathcal{D}_\varepsilon$ aus.

# 5 Evaluierung

In diesem Kapitel wird das beschriebene Attributions-Verfahren anhand von drei Experimenten evaluiert. Dazu werden zunächst die benötigten Ressourcen sowie die Parametrisierungen angegeben, mit denen das Verfahren hinsichtlich der Experimente initialisiert wurde. Im Anschluß daran werden die Ergebnisse präsentiert sowie Erkenntnisse dazu erläutert.

## 5.1 Verwendete Ressourcen

Bevor die Experimente durchgeführt werden konnten, mussten zuvor die folgenden Ressourcen vorliegen:

1. **Korpora:** Unter einem Korpus wird innerhalb dieser Arbeit eine Ansammlung von Dokumenten verstanden, die in einer annotierten Form vorliegen. Die Annotation eines Dokuments repräsentiert dabei eine vereinfachte Darstellung von sprachlichen Ebenen (wie etwa Morphologie, Syntax oder Semantik) und umfasst unter anderem Tokens, tokenisierte Sätze, Wortarten, etc. Eine Menge von Korpusse wird als Korpora bezeichnet.
2. **Features:** Unter diesem Begriff werden im Kontext der Autorschaftsanalyse Stilmerkmale verstanden, mit deren Hilfe Autorenstile angenähert werden können. Features werden aus den sprachlichen Ebenen eines Dokuments entnommen und benötigen daher einen entsprechenden Zugang zu diesen, welcher durch die Annotationen realisiert wird.

### 5.1.1 Verwendete Korpora

Als Korpora wurden öffentlich zugängliche Daten wie beispielsweise Forenbeiträge, wissenschaftliche Ausarbeitungen sowie Computermagazine verwendet. Zudem wurden auch nichtöffentliche Daten in Form von E-Mails aus privater und geschäftlicher Korrespondenz für die Korpora benutzt. Die folgende Tabelle listet die Korpora auf und erläutert dabei deren Inhalte:

**Tab. 3:** Verwendete Korpora und deren Kurzbeschreibung

| Korpus | Autoren | Beschreibung der Inhalte |
|---|---|---|
| $\mathcal{K}_{kom}$ | 5 | Abschnitte einer Studienarbeit über digitale Lernspiele |
| $\mathcal{K}_{news}$ | 6 | Persönliche Blogs einiger Tagesschau Redakteure |
| $\mathcal{K}_{thesen}$ | 15 | Master- und Diplomarbeiten von Studenten der TU Darmstadt |
| $\mathcal{K}_{mails}$ | 26 | E-Mails unterschiedlicher Themengebiete (Studium, Arbeit, Freizeit, ...) |
| $\mathcal{K}_{recht}$ | 32 | Forenbeiträge rechtsbezogener Themen (Miet-, Straf-, Scheidungsrecht, ...) |
| $\mathcal{K}_{ct}$ | 50 | Kolumnen mit praxisbezogenen Computerthemen (Software/Hardware) |
| $\mathcal{K}_{d120}$ | 50 | Forenbeiträge von Studenten (Informatik-Fachbereich der TU Darmstadt) |

Neben der Autorenanzahl und der Domänen unterscheiden sich die Korpora zusätzlich durch die Qualität der darin befindlichen Texte. So enthält beispielsweise $\mathcal{K}_{thesen}$ qualitativ hochwertige sachliche Texte, die kaum Rechtschreibfehler aufweisen, während in $\mathcal{K}_{d120}$ überwiegend subjektive Texte vorkommen, die neben Rechtschreibfehlern weitere grammatikalische Eigenarten umfassen. Dazu zählen z.B. unbekannte Wörter, fehlerhafte Wortkombinationen (z.B. bedingt durch Interferenzfehler von Nichtmuttersprachlern) sowie falsche Satzstellungen.

### 5.1.2 Verwendete Features

In den Experimenten wurden insgesamt 13 Feature-Kategorien eingesetzt. Die folgende Tabelle führt diese anhand ihrer Kennung, Kategorie, Anzahl sowie einiger Beispiel-Features auf:

**Tab. 4:** Eingesetzte Feature-Kategorien

| $F_i$ | Feature-Kategorie | Beispiele | $\lvert F_i \rvert$ |
|---|---|---|---|
| $F_1$ | Interpunktionszeichen | (), [], !, ?, ;, :, -, _, ... | 17 |
| $F_2$ | Buchstaben | a-z, ä, ö, ü, ß, A-Z, Ä, Ö, Ü | 59 |
| $F_3$ | Buchstaben Bi-/Trigramme | en, ch, de, sch, ein, ten, ... | 1571 |
| $F_4$ | Funktionswörter | und, oder, als, auch, an, auf, daher, ... | 763 |
| $F_5$ | Wort-Komplexität | Wörter bestimmter Länge, Wörter mit $x$ Vokalen, ... | 50 |
| $F_6$ | Phrasen | Kollokationen, Wort $n-$Gramme, ... | 530 |
| $F_7$ | Wortart Unigramme | Adjektiv, Adverb, Präposition, Interjektion, ... | 54 |
| $F_8$ | Wortart Trigramme | Artikel-Adjektiv-Nomen, Pronomen-Nomen-Artikel, ... | 959 |
| $F_9$ | Satz-Anfänge/Endungen | Satzanfang(Nomen), Satzende(finites Verb), ... | 80 |
| $F_{10}$ | Grammatikalische Fehler | Falsche Verwendung von Genus, Kasus, Kommata, ... | 2292 |
| $F_{11}$ | Anglizismen | Mail, Newsletter, chatten, updaten, einloggen, ... | 1304 |
| $F_{12}$ | Redewendungen | Redensarten, feste Wortverbindungen, ... | 2954 |
| $F_{13}$ | Text-Komplexität | Funktionswort-Dichte, Satz-Mittelfeld Komplexität, ... | 54 |

## 5.2 Experimente

Um die Experimente durchzuführen, wurden die Dokumente in jedem Korpus jeweils in eine Trainings- bzw. Testmenge aufgeteilt. Ausgehend von den beiden Mengen wurde anschließend für jeden $\mathcal{A}_j \in \mathbb{A}$ ein Trainingsprofil $\mathcal{P}_{train}(\mathcal{A}_j)$ sowie ein Testprofil $\mathcal{P}_{test}(\mathcal{A}_j)$ generiert. Hierbei entspricht $\mathcal{P}_{test}(\mathcal{A}_j)$ dem anonymen Dokument $\mathcal{D}_\varepsilon$, während $\mathcal{P}_{train}(\mathcal{A}_j)$ ein Trainingsdokument $\mathcal{D}_{BIG\mathcal{A}_j} \in \mathbb{S}'_{train}$ darstellt. Als Evaluierungsmethode wurde eine „One-Against-All“ Strategie verwendet, in der jedes $\mathcal{P}_{test}(\mathcal{A}_j)$ gegen alle Trainingsprofile, inklusive $\mathcal{P}_{train}(\mathcal{A}_j)$, verglichen wird. Entscheidend sind dabei die Stil-Unterscheidungswerte in der zweiten Stufe des Verfahrens. Falls und nur falls $\mathcal{P}_{train}(\mathcal{A}_j)$ den niedrigsten Stil-Unterscheidungswert zu $\mathcal{P}_{test}(\mathcal{A}_j)$ aufweist, gilt die Attribution als erfolgreich ($\varepsilon = \mathcal{A}_{true}$).

Bei den Experimenten wurden hierbei die folgenden drei Parametrisierungen verwendet, um dadurch ein optimales Attributions-Ergebniss zu finden:

**Tab. 5:** Parametrisierung der drei Experimente

| Parameter | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|
| Stufe $1+2$, Länge von $\mathcal{D}_\varepsilon$ | $\approx 4$ KByte | $\approx 5$ KByte | $\approx 6$ KByte |
| Stufe $1+2$, Feature-Kategorien | Alle | $F_{1-4}, F_7, F_{13}$ | $F_{1-4}, F_7, F_{13}$ |
| Stufe 1, $n-$Gramm Größe | 6 | 5 | 5 |
| Stufe 1, $n-$Gramm Häufigkeit: $k$ | 100 | 120 | 80 |
| Stufe 1, Ähnlichkeitsfunktion: $sim$ | $dice(\ldots)$ | $jaccard(\ldots)$ | $overlap(\ldots)$ |
| Stufe 1, Splitparameter: `SPLIT` | 30% | 50% | 25% |
| Stufe 2, Distanzfunktion: $dist$ | $euclid(\ldots)$ | $euclid(\ldots)$ | $euclid(\ldots)$ |

**Anmerkung:** Eine Auflistung der hier aufgeführten Distanz- bzw. Ähnlichkeitsfunktionen findet sich in [Halv12] (Kapitel „Metriken“).

### 5.2.1 Experiment: 1

Im ersten Experiment wurde für jede Feature-Kategorie $F_i$ eine entsprechende Attribution auf sämtliche Korpora durchgeführt. Das Resultat einer Attribution stellt dabei das sogenannte $Accuracy$-Maß dar, welches wie folgt definiert ist:

$$Accuracy = 100 \cdot \left( \frac{\text{Anzahl aller } \mathcal{A}_{true} \text{ in } \mathcal{K}_i \text{ die als solche vorhergesagt wurden}}{|\mathbb{A}|} \right)$$

Die Ergebnisse dieses Experiments lauten wie folgt:

**Tab. 6:** Erkennungsgenauigkeiten bzgl. aller Feature-Kategorien

| $F_i$ | $\mathcal{K}_{kom}$ | $\mathcal{K}_{news}$ | $\mathcal{K}_{thesen}$ | $\mathcal{K}_{mails}$ | $\mathcal{K}_{recht}$ | $\mathcal{K}_{ct}$ | $\mathcal{K}_{d120}$ | ∅ |
|---|---|---|---|---|---|---|---|---|
| $F_1$ | 100.00% | 88.33% | 53.33% | 61.54% | 31.25% | 24.00% | 52.00% | 57.92% |
| $F_2$ | 80.00% | 66.66% | 60.00% | 38.46% | 50.00% | 50.00% | 56.00% | 57.30% |
| $F_3$ | 80.00% | 66.66% | 93.33% | 57.69% | 28.12% | 72.00% | 82.00% | 68.54% |
| $F_4$ | 80.00% | 66.66% | 73.33% | 57.69% | 28.12% | 38.00% | 78.00% | 60.26% |
| $F_5$ | 80.00% | 33.33% | 66.66% | 42.30% | 18.75% | 28.00% | 42.00% | 44.43% |
| $F_6$ | 80.00% | 100.00% | 53.33% | 38.46% | 9.37% | 26.00% | 42.00% | 49.88% |
| $F_7$ | 80.00% | 83.33% | 60.00% | 61.53% | 28.12% | 32.00% | 64.00% | 58.43% |
| $F_8$ | 80.00% | 83.33% | 46.66% | 42.30% | 18.75% | 24.00% | 30.00% | 46.43% |
| $F_9$ | 100.00% | 83.33% | 66.66% | 42.30% | 15.62% | 20.00% | 46.00% | 53.42% |
| $F_{10}$ | 60.00% | 66.66% | 60.00% | 46.15% | 21.87% | 20.00% | 40.00% | 44.95% |
| $F_{11}$ | 40.00% | 50.00% | 26.66% | 26.92% | 9.37% | 16.00% | 12.00% | 25.85% |
| $F_{12}$ | 60.00% | 100.00% | 80.00% | 53.84% | 25.00% | 40.00% | 66.00% | 52.11% |
| $F_{13}$ | 100.00% | 66.66% | 73.33% | 42.30% | 12.50% | 36.00% | 34.00% | 60.69% |
| ∅ | 78.46% | 73.07% | 62.56% | 47.04% | 22.83% | 32.77% | 49.54% | |

**Erkenntnisse:** Ausgehend von den Spalten in Tabelle 6 fällt zunächst auf, dass das Verfahren für kleinere Korpora (5 bis 15 Autoren) im Durchschnitt die höchsten Ergebnisse erzielen konnte. Bei einer Betrachtung der Zeilen fällt dagegen insbesondere die Feature-Kategorie $F_3$ auf, die im Durchschnitt die höchste Erkennungsgenauigkeit von $68.54\%$ gegenüber allen anderen Feature-Kategorien aufweist. Die Begründung für dieses Ergebniss liegt darin, dass die $n-$Gramme in dieser Feature-Kategorie in der Lage sind, eine Vielfalt von zeichen- als auch wortbasierten Features einzufangen.

Als zeichenbasierte Features konnten so z.B. in einer genaueren Analyse der Testergebnisse morphologische Eigenarten, wie etwa falsche Genus-/Pluralmarkierungen, vergessene Fugenelemente oder unübliche Präfixbildungen, in den einzelnen Autoren-Profile festgestellt werden. Mit Hilfe solcher morphologischer Konstellationen konnten sich Autorenstile, gemessen an anderen Feature-Kategorien, am besten differenzieren lassen. Eine Unterscheidung ist dabei deswegen möglich, weil es unwahrscheinlich ist, dass ein Autor $\mathcal{A}_1$ die selben Fehler wie ein anderer Autor $\mathcal{A}_2$ in einem Text produziert.

Als wortbasierte Features konnten dagegen in den Testergebnissen Funktionswörter wie etwa $\{$ `so, um, in, und, der, ...` $\}$ in den oberen Rängen der $k-$häufigsten $n-$Gramme beobachtet werden. Der Vorteil dieser Wörter ist dabei, dass sie in jedem Text vorkommen, jedoch in einer unterschiedlichen Häufigkeit. Durch diese differenzierten Häufigkeitsverteilungen innerhalb der Autoren-Profile können wiederum Autorenstile annähernd eindeutig charakterisiert

werden. Die Diskriminierungskraft von Funktionswörtern ist auch in der Literatur seit längerem bekannt, zumindest für englischsprachige Texte (siehe z.B. [Stam09]). Anhand des Experiments konnte nun die selbe Aussage auch für deutschsprachige Texte bestätigt werden.

Was die Korpora in Tabelle 6 betrifft, konnte eine interessante Beobachtung bzgl. $\mathcal{K}_{recht}$ gemacht werden, der bei fast jeder Feature-Kategorie schlecht abschneidet. Der Grund dafür liegt vor allem daran, dass die Texte in diesem Korpus sehr einheitlich geschrieben sind, sodass die Autorenstile für nahezu jedes $F_i$ kaum zu unterscheiden waren. Der einheitliche Stil ist dabei hauptsächlich durch den formellen Register bedingt, welcher sich wie folgt bemerkbar macht:

- Erläuterungen von Paragraphen oder Teilen davon aus Gesetzesbücher.
- Fallbeispiele aus abgeschlossenen Gerichtsprozessen.
- Erklärungen zu Vertragsklauseln, AGB's, etc.

Überraschenderweise konnte jedoch festgestellt werden, dass neben dem einheitlichen Stil die juristische Terminologie von $\mathcal{K}_{recht}$ die Ergebnisse in Tabelle 6 nicht beinflusst hat. Diese Aussage wird dadurch bekräftigt, dass eine Attribution anhand von Funktionswörtern und somit von Wörtern, die unabhängig von der Terminologie sind, ebenfalls zu einem schlechten Ergebnis ($28.12\%$) geführt hat.

### 5.2.2 Experiment: 2

In diesem Experiment wurden zunächst die Feature-Kategorien $F_{1-4}, F_7$ und $F_{13}$ ausgewählt, die im Durchschnitt die höchsten Ergebnisse in der Tabelle 6 erzielt haben. Das Ziel war zu untersuchen, ob die Erkennungsgenauigkeiten hinsichtlich dieser Kategorien mit Hilfe einer Feature-Selektion übertroffen werden konnten. Dazu diente ein korrelationsbasiertes Verfahren von [Hall98] als Feature-Selektionsalgorithmus, mit dessen Hilfe die vielversprechendsten Features aus den sechs Kategorien automatisch selektiert wurden. Die Anwendung des Verfahrens anhand der selektierten Features führte zu folgendem Resultat:

**Tab. 7:** Erkennungsgenauigkeiten bei gefilterten Feature-Kategorien

| $F_i$ | $\mathcal{K}_{kom}$ | $\mathcal{K}_{news}$ | $\mathcal{K}_{thesen}$ | $\mathcal{K}_{mails}$ | $\mathcal{K}_{recht}$ | $\mathcal{K}_{ct}$ | $\mathcal{K}_{d120}$ | $\varnothing$ |
|---|---|---|---|---|---|---|---|---|
| $F_1$ | 100.00% | 83.33% | 46.66% | 57.69% | 25.00% | 24.00% | 52.00% | 55.53% |
| $F_2$ | 80.00% | 66.66% | 80.00% | 42.30% | 28.12% | 46.00% | 40.00% | 54.73% |
| $F_3$ | 60.00% | 66.66% | 86.66% | 42.30% | 28.12% | 68.00% | 64.00% | 59.39% |
| $F_4$ | 80.00% | 66.66% | 86.66% | 61.53% | 18.75% | 38.00% | 66.00% | 59.66% |
| $F_7$ | 80.00% | 66.66% | 53.33% | 57.69% | 25.00% | 24.00% | 60.00% | 52.38% |
| $F_{13}$ | 100.00% | 66.66% | 60.00% | 61.53% | 25.00% | 36.00% | 56.00% | 57.88% |

**Erkenntnisse:** Wie hier ersichtlich wird, hat die Feature-Selektion die Ergebnisse im Vergleich zum ersten Experiment größtenteils verschlechtert. Um dies zu begründen, wurden die selektierten Features genauer betrachtet. Dabei stellte sich heraus, dass in fast jeder Kategorie zu viele Features eliminiert worden sind, was in manchen Fällen (z.B. bei $F_4$) zur Folge hatte, dass eine Attribution mit nur $14$ Features durchgeführt wurde, obwohl im Vorfeld $763$ Features zur Verfügung standen. Aber auch die Wahl der Features selbst war in manchen Fällen verwunderlich. So wurden einige irrelevante Features selektiert, die kaum Aussagekraft besaßen, wie etwa das Semikolon in der Kategorie $F_1$, welches in den Beispieltexten nur selten vorkam.

**Anmerkungen:** Neben dem Verfahren von [Hall98] wurden drei weitere Selektionsalgorithmen getestet. Die Ergebnisse waren jedoch schlechter als die in Tabelle 7, sodass diese gar nicht erst aufgeführt wurden. Als Grund für diese schlechten Ergebnisse wird vermutet, dass die Selektionsalgorithmen sämtliche Einzeldokumente der Autoren als Input benötigen, aber das Verfahren die konkatenierte Form dieser Dokumente verwendet, sodass dadurch die Diskriminierungsstärke der Features zu ungenau ermittelt wird.

### 5.2.3 Experiment: 3

In diesem Experiment wurden wieder die Feature-Kategorien $F_{1-4}, F_7$ und $F_{13}$ ausgewählt. Dieses Mal jedoch wurden sie nicht einzeln, sondern stattdessen in einer kombinierten Form angewendet. Da bei sechs Feature-Kategorien insgesamt $64$ Kombinationen möglich sind und für jede Kombination Berechnungen für alle Korpora durchgeführt werden müssten, wurden hier intuitiv nur fünf Kombinationen ausgesucht. Die Ergebnisse dazu lauten wie folgt:

**Tab. 8:** Erkennungsgenauigkeiten bei kombinierten Feature-Kategorien

| Kombination | $\mathcal{K}_{kom}$ | $\mathcal{K}_{news}$ | $\mathcal{K}_{thesen}$ | $\mathcal{K}_{mails}$ | $\mathcal{K}_{recht}$ | $\mathcal{K}_{ct}$ | $\mathcal{K}_{d120}$ | $\varnothing$ |
|---|---|---|---|---|---|---|---|---|
| $F_1, F_2, F_3, F_4$ | $100.00\%$ | $83.33\%$ | $93.33\%$ | $69.23\%$ | $34.37\%$ | $56.00\%$ | $86.00\%$ | $74.61\%$ |
| $F_1, F_3$ | $100.00\%$ | $83.33\%$ | $93.33\%$ | $65.38\%$ | $37.50\%$ | $66.00\%$ | $76.00\%$ | $74.51\%$ |
| $F_1, F_7$ | $100.00\%$ | $83.33\%$ | $93.33\%$ | $69.23\%$ | $28.12\%$ | $58.00\%$ | $68.00\%$ | $71.43\%$ |
| $F_2, F_4, F_7, F_{13}$ | $80.00\%$ | $83.33\%$ | $86.66\%$ | $69.23\%$ | $31.25\%$ | $58.00\%$ | $76.00\%$ | $69.21\%$ |
| $F_2, F_3, F_4$ | $60.00\%$ | $83.33\%$ | $93.33\%$ | $65.38\%$ | $25.00\%$ | $58.00\%$ | $82.00\%$ | $66.72\%$ |

**Erkenntnisse:** Im Vergleich zu den anderen Experimenten sind die Erkennungsgenauigkeiten hier am höchsten. Das beste Ergebnis im Durchschnitt liefert die Kombination $F_1, F_2, F_3, F_4$ mit $74.61\%$. Wird von dem problematischen Korpus $\mathcal{K}_{recht}$ abgesehen, so kommt das Verfahren hier sogar auf $81.32\%$. Neben dieser Kombination liefert die Zusammenführung von $F_1, F_3$ ein fast identisches Ergebniss und benötigt dabei gleichzeitig $822$ Features weniger, was sich wiederum positiv auf die Laufzeit des Verfahrens auswirkt. Somit bietet sich die Kombination von $F_1, F_3$ als ein optimaler Trade-Off für das Verfahren an, sofern die Laufzeit als auch die Erkennungsgenauigkeit die gleiche Priorität haben.

# 6 Zusammenfassung und Ausblick

In dieser Arbeit wurde gezeigt, dass eine domänenübergreifende Autorschafts-Attribution mit Hilfe eines selbstentwickelten profilbasierten Ansatzes realisiert werden kann. Dafür wurden verschiedene Attributions-Szenarien auf sieben Korpora durchgeführt, welche sich durch Register, Genre und andere sprachliche Varietäten voneinander abgrenzen.

Eine erfolgreiche Attribution ist dabei auf eine geeignete Parametrisierung angewiesen, die gleichzeitig viel Spielraum für Einstellungen offen lässt. Hierbei wird angenommen, dass insbesondere die Größe des Ausgangsdokuments als auch die verwendeten Feature-Kategorien für den Erfolg des Verfahrens maßgebend sind. Die Wahl der Ähnlichkeitsfunktion hat dagegen weniger Einfluss auf das Attributions-Ergebniss, da in den durchgeführten Experimenten kaum Unterschiede dazu festgestellt werden konnten. Anders dagegen verhält sich die Wahl der verwendeten Distanzfunktion. Da dieses, in der zweiten Stufe des Verfahrens über den Erfolg bzw. das Scheitern der Attribution entscheidet, sollte sichergestellt werden, dass die

Stil-Unterscheidungswerte untereinander möglichst große Trennschärfen aufweisen. Gute Erfahrungen wurden hierbei vorwiegend mit der euklidischen Distanz gemacht, sodass nur diese in den drei Parametrisierungen verwendet wurde.

Eine optimale Parametrisierung für beliebige Domänen konnte in dieser Arbeit nicht ermittelt werden. Die Ergbnisse zeigen jedoch, dass für den Großteil der getesteten Korpora erfolgsversprechende Attributions-Vorhersagen erbracht werden können. Dies trifft zu, falls dabei zeichenbasierte Features (insbesondere $n-$Gramme) verwendet werden. In weiterführenden Arbeiten könnte daher überprüft werden, ob weitere Feature-Kategorien existieren, die die Performanz von $n-$Gramme überbieten können. Eine andere interessante Fragestellung wäre, ob mit Hilfe komplexerer Techniken, wie etwa der Klassifikation der Beispieldokumente anhand von soziolinguistischen Variablen, die Menge der Autoren intelligenter gefiltert werden kann. Dadurch würde die gefilterte Trainingsmenge nur diejenigen Autoren enthalten, die hinsichtlich soziolinguistischer Variablen wie etwa Alter, Geschlecht, Bildungsniveau oder Muttersprachlichkeit mit dem anonymen Autor übereinstimmen. Anschließend könnte der stilistisch ähnlichste Autor aus dieser Menge bestimmt werden.

## Danksagung

Diese Arbeit wurde unterstützt vom CASED – Center for Advanced Security Research Darmstadt (www.cased.de), gefördert vom Hessischen Ministerium für Wissenschaft und Kunst unter dem LOEWE-Förderprogramm.

## Literatur

[Bach12] D. Bachfeld: Aufklärungsarbeit - Verräterische Metadaten aus Web-Dokumenten extrahieren (2012), .

[BeAc10] M. Bedner, T. Ackermann: Schutzziele der IT-Sicherheit. In: *Datenschutz und Datensicherheit*, 34, 5 (2010), 323–328.

[Ecke09] C. Eckert: IT-Sicherheit - Konzepte, Verfahren, Protokolle. Oldenbourg (2009).

[Fobb11] E. Fobbe: Forensische Linguistik: Eine Einführung. Narr Studienbücher, Narr Francke Attempto Verlag GmbH + Co. KG, Tübingen (2011), .

[Hall98] M. A. Hall: Correlation-based Feature Subset Selection for Machine Learning. Dissertation, University of Waikato, Hamilton, New Zealand (1998).

[Halv12] O. Halvani: Autorschaftsanalyse im Kontext der Attribution, Verifikation und intrinsischer Exploration. Diplomarbeit, Technische Universität Darmstadt / Fraunhofer-Institut für Sichere Informationstechnologie, Darmstadt, Germany (2012).

[Hein10] L. Heine: Linguistics@schools: Abenteuer Sprachwissenschaft; Kooperationsmöglichkeiten zwischen Schule und Hochschule. Peter Lang (2010), .

[KPCT03] V. Keselj, F. Peng, N. Cercone, C. Thomas: N-Gram-based Author Profiles for Authorship Attribution. In: *Computational Linguistics*, 3 (2003), 255–264, .

[Ploe11] M. C. Ploetz: Analyse existierender Studiengänge im Bereich IT-Sicherheit, Bachelor Thesis, Fakultät für Mathematik und Informatik, Fernuniversität Hagen (2011).

[Stam09] E. Stamatatos: A Survey of Modern Authorship Attribution Methods. In: *J. Am. Soc. Inf. Sci. Technol.*, 60, 3 (2009), 538–556, .