



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-18-01

Ähnlichkeitsmessung von ausgewählten Datentypen in Datenbanksystemen zur Berechnung des Grades der Anonymisierung

Jan-Philipp Heinrich · Carsten Neise · Andreas Müller

Februar 2018

Chemnitzer Informatik-Berichte

Ähnlichkeitsmessung von ausgewählten Datentypen in Datenbanksystemen zur Berechnung des Grades der Anonymisierung

Jan-Philipp Heinrich

Technische Universität Chemnitz / profi.com AG business solutions

Carsten Neise

profi.com AG business solutions

Andreas Müller

Technische Universität Chemnitz

Inhaltsverzeichnis

| | | |
|-------|--|----|
| 1 | Motivation..... | 2 |
| 2 | Relevante Datentypen | 2 |
| 3 | Definition des Anonymisierungsgrads θ | 3 |
| 3.1 | Anforderungen an θ | 3 |
| 3.2 | Definition der Algebra für θ | 3 |
| 3.2.1 | Date | 3 |
| 3.2.2 | Boolean..... | 5 |
| 3.2.3 | Integer..... | 6 |
| 3.2.4 | Float..... | 6 |
| 3.2.5 | String | 6 |
| 3.2.6 | Gesamtanonymisierungsgrad θ | 7 |
| 4 | Ergebnisse | 7 |
| 4.1 | Date..... | 8 |
| 4.2 | Integer..... | 8 |
| 4.3 | Float | 8 |
| 4.4 | String | 8 |
| 4.5 | Gesamtanonymisierungsgrad..... | 9 |
| 5 | Resümee und Ausblick | 10 |

1 Motivation

Es soll ein mathematisches Modell zur Berechnung von Abweichungen verschiedener Datentypen auf relationalen Datenbanksystemen eingeführt und getestet werden. Basis dieses Modells sind Ähnlichkeitsmessungen für verschiedene Datentypen.

Hierbei führen wir zunächst eine Betrachtung der relevanten Datentypen für die Arbeit durch. Danach definieren wir für die für diese Arbeit relevanten Datentypen eine Algebra, welche die Grundlage zur Berechnung des Anonymisierungsgrades θ ist.

Das Modell soll zur Messung des Grades der Anonymisierung, vor allem personenbezogener Daten, zwischen Test- und Produktionsdaten angewendet werden. Diese Messung ist im Zuge der Einführung der EU-DSGVO¹ im Mai 2018 sinnvoll, und soll helfen personenbezogene Daten mit einem hohen Ähnlichkeitsgrad zu identifizieren.

2 Relevante Datentypen

Relationale Datenbanken unterstützen eine Menge an verschiedenen Datentypen². Wir beschränken uns im Rahmen dieser Arbeit auf die wesentlichen Datentypen für eine Anwendung im Testdatenbereich, das heißt die Typen die man üblicherweise für Name, Vorname, Telefonnummer, Personalnummer, Geburtsdatum usw. benötigt. Diese Arbeit wurde auf Basis einer MySQL³ Datenbank umgesetzt, die hierfür genutzten Datentypen sind in Tabelle 1 aufgelistet. Eine Vollständigkeit hinsichtlich der existierenden Datentypen wird daher nicht verlangt und würde den Rahmen dieser Arbeit sprengen.

Tabelle 1: Relevante Datentypen in dieser Arbeit in MySQL

| Datentyp | Speicherplatz | Definitionsbereich |
|----------------|----------------------------|---|
| Integer | 4 Bytes | Ganze Zahlen von -2147483647 bis 2147483647 |
| Boolean | 8 Byte | {0, 1} |
| Double | 8 Bytes | Fließkommazahl mit Vorzeichen. Wertebereich von $-(1,79769 \times 10^{308})$ bis $-(2.22507 \times 10^{-308})$, 0 und 2.22507×10^{-308} bis $1,79769 \times 10^{308}$ |
| Date | 3 Bytes | 01.01.1000 bis 31.12.9999. |
| Varchar | Länge des Strings + 1 Byte | Folge von Zeichen |

In einem nächsten Schritt soll für die in Tabelle 1 definierten Datentypen eine Algebra definiert werden auf deren Basis dann der Anonymisierungsgrad θ errechnet wird. θ soll eine Abbildung aus x und x' sein, wobei x repräsentativ für ein Testdatum im Originalbestand und x' ein Testdatum welches anonymisiert bzw. synthetisch erzeugt wurde. Wir definieren die Maskierungsfunktion $f(x) \rightarrow x'$, welche den Originalwert x auf den anonymisierten/synthetisch erzeugten Wert x' abbildet. $\theta(x, x')$ wird dann mit der im folgenden Abschnitt definierten Abbildung $f(x, x')$ errechnet.

¹ <https://de.wikipedia.org/wiki/Datenschutz-Grundverordnung>

² <http://www.schmager.de/mysql.php>

³ <https://www.mysql.com/de/>

3 Definition des Anonymisierungsgrads θ

3.1 Anforderungen an θ

Der Anonymisierungsgrad θ soll folgende Anforderungen erfüllen

- symmetrisch, d.h. $\theta(x, x') = \theta(x', x)$
- θ bildet auf die reellen Zahlen im abgeschlossenen Intervall $[0, 1]$ ab
- $\theta(x, x) = 0$
- $\theta(x, x') = 1$, für x und x' maximal im Sinne des Definitionsbereichs entfernt
- Definitionsbereich von θ sind folgende Datentypen
 - Date
 - Boolean
 - Integer
 - String
 - Float

In den folgenden Abschnitten wird Anwendung von θ auf die in Tabelle aufgelisteten Datentypen definiert.

Folgende Funktionen werden als sinnvoll erachtet zur Bestimmung von θ :

- Lineare Mittelung $f(x, x', d) = \frac{x-x'}{d}$:: d ein Normierungsparameter
- Quadratische Mittelung $f(x, x', d) = \left(\frac{x-x'}{d}\right)^2$:: d ein Normierungsparameter
- Kubische Mittelung $f(x, x', d) = \left(\frac{x-x'}{d}\right)^3$:: d ein Normierungsparameter
- Gauß-Mittelung $f(x, x', \sigma) = e^{-\frac{1}{2}\left(\frac{x-x'}{\sigma}\right)^2}$:: σ ein Normierungsparameter

3.2 Definition der Algebra für θ

3.2.1 Date

Für den Datentyp Date ergibt sich folgender Definitionsbereich

- x, x' : Ganze Zahlen
- x hat die Struktur: DD.MM.YYYY, wobei DD den Tag bezeichnet, MM den Monat bezeichnet, YYYY das Jahr bezeichnet

Wir definieren den Anonymisierungsgrad θ_d wie folgt:

- $\theta_d = \frac{1}{3}(\theta_d^{\text{day}} + \theta_d^{\text{month}} + \theta_d^{\text{year}})$
- θ_d^{year} , siehe Datentyp Integer
- θ_d^{month} :
 - $\max(\Delta(x, x')) = 6$:: punktsymmetrisch
 - $\theta_d^{\text{month}}(\max \Delta(x, x')) = 0$
- θ_d^{day} :
 - $\max(\Delta(x, x')) = 15$:: punktsymmetrisch
 - $\theta_d^{\text{day}}(\max \Delta(x, x')) = 0$

Diskussion

Anhand der obigen Randbedingungen an θ_d^{month} und θ_d^{day} nutzen wir die vier in Abschnitt Anforderungen an θ aufgelisteten Funktionen, und diskutieren anhand der Abweichung den Anonymisierungsgrad θ_d^{month} und θ_d^{day} .

$$\begin{aligned}
 - \theta_d^{\text{day},1} &= \text{mod}_{\text{month}} \frac{|x-x'|}{15} & \theta_d^{\text{month},1} &= \text{mod}_{\text{year}} \frac{|x-x'|}{6} \\
 - \theta_d^{\text{day},2} &= \left(\text{mod}_{\text{month}} \frac{|x-x'|}{15}\right)^2 & \theta_d^{\text{month},2} &= \left(\text{mod}_{\text{year}} \frac{|x-x'|}{6}\right)^2 \\
 - \theta_d^{\text{day},3} &= \left(\text{mod}_{\text{month}} \frac{|x-x'|}{15}\right)^3 & \theta_d^{\text{month},3} &= \left(\text{mod}_{\text{year}} \frac{|x-x'|}{6}\right)^3 \\
 - \theta_d^{\text{day,gauss}} &= 1 - e^{-\frac{1}{2} \left(\frac{\text{mod}_{\text{month}} |x-x'|}{\sigma}\right)^2} & \theta_d^{\text{month,gauss}} &= 1 - e^{-\frac{1}{2} \left(\frac{\text{mod}_{\text{year}} |x-x'|}{\sigma}\right)^2}
 \end{aligned}$$

Zur Berechnung des Anonymisierungsgrades wird die jeweilige Differenz der Tage Modulo 15 und die Differenz der Monate Modulo 6 betrachtet. Anschließend wird das Resultat mit 15 (6) normiert. Die Ergebnisse sind in Abbildung 1 und Abbildung 2 dargestellt.

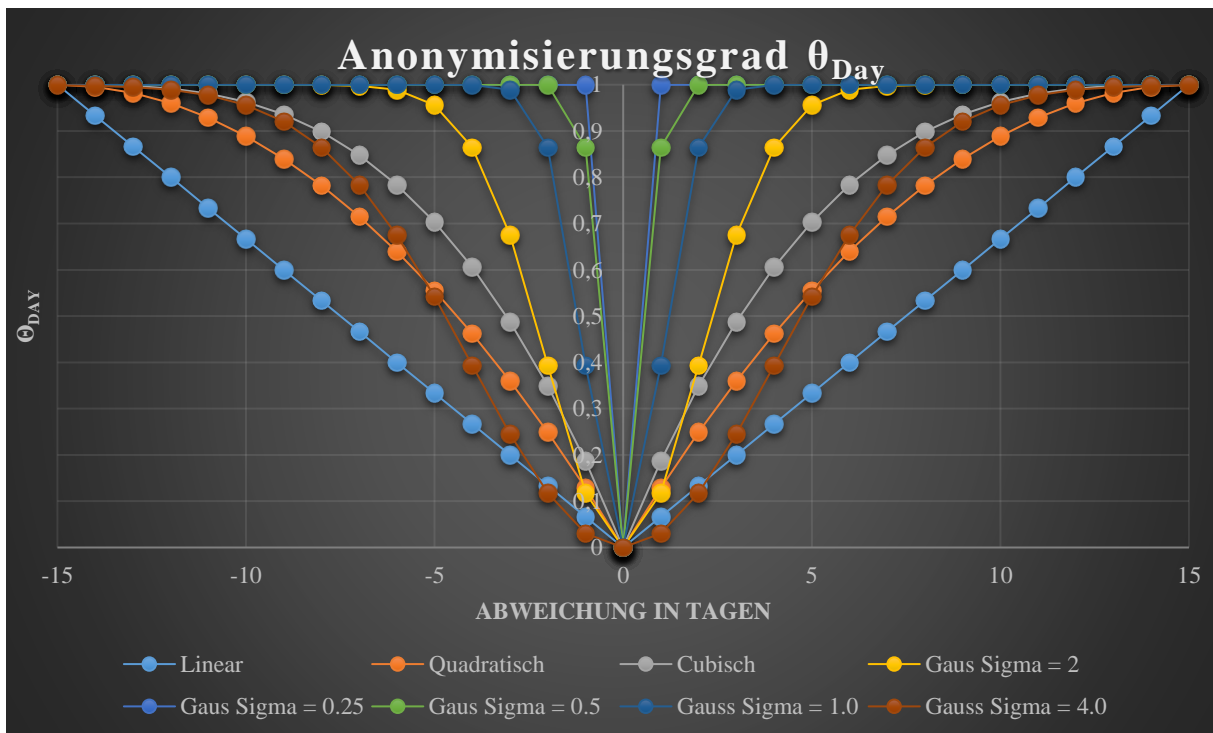


Abbildung 1: ϑ_{day} Abhängigkeit der Differenz in Tagen

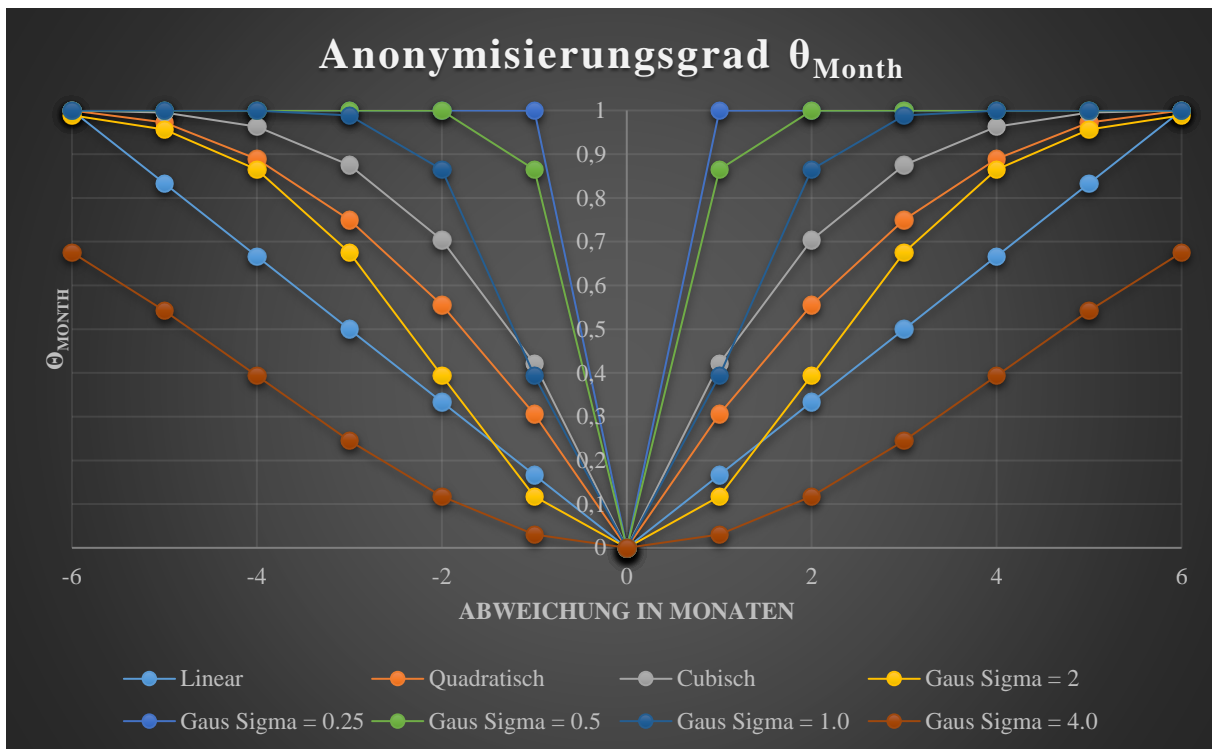


Abbildung 2: ϑ_{month} Abhängigkeit der Differenz in Monaten

Wir sehen, dass die Gauß-Funktion den lokalen Charakter der Änderung am besten abbildet. Wir möchten dies an einem einfachen Beispiel erläutern. Nehmen wir an, das Testdatum im Originalbestand ist der Januar. Wenn das Testdatum im anonymisierten Bestand nun auch der Januar ist, so ist θ_{month} gleich 0. Sei der anonymisierte Monat nun der Februar, dann sehen wir, dass $\theta_{\text{month}} > 0$ ist. Wie stark der Grad der Anonymisierung ist, hängt von der Art der Mittelung und der Einstellung des Parameters σ ab. Für einen anonymisierten Monat März wird θ_{month} nochmals zunehmen, usw. Die größte Verfremdung erhalten wir, wenn wir den Juli wählen. Dies bildet den größtmöglichen Abstand in Monaten dar, nämlich 6. Hier soll θ_{month} gleich 1 sein. Der Grad der Verfremdung ist in der linearen Mittelung gleich für alle Monate. Wir behaupten allerdings, dass eine Anonymisierung auf die Monate Februar oder März sich stark auf den Grad der Verfremdung auswirken soll. Währenddessen ich bei großen Entfernungen zu Monaten Mai, Juni, oder gar Juli kaum noch Auswirkungen auf θ_{month} haben soll. Dies wird am besten durch die Gauß-Funktion dargestellt, weshalb wir diese im Folgenden immer nutzen werden.

Als weiteres sehen wir, dass wir mit dem Parameter σ die Lokalität der Gauß-Funktion anpassen können. Je breiter größer σ gewählt, desto flacher wird die Gauß-Funktion und folglich werden nun auch „entferntere“ Monate relevant, wie im Beispiel oben dann eben auch der Monat Juni. Dies kann für den Datentypen String bspw. genutzt werden um das natürliche Lesen nachzuahmen. Es kann die Reichweite des Einflusses der Anonymität variiert werden, und muss von Anwendungsfall zu Anwendungsfall für die verschiedenen Datentypen im Folgenden eingestellt werden.

3.2.2 Boolean

Für den Datentyp Boolean ergibt sich folgender Definitionsbereich

- $x, x' : \text{True/False}$

Wir definieren den Anonymisierungsgrad θ_b wie folgt:

- $\theta_b = 0$, für $x = x'$
- $\theta_b = 1$, für $x \neq x'$

3.2.3 Integer

Für den Datentyp Integer ergibt sich folgender Definitionsbereich

- x, x' : Ganze Zahlen
- x besitzt N Digits

Wir definieren Anonymisierungsgrad θ_i wie folgt⁴:

- $\theta_i = \frac{1}{2} (\theta_i^l + \theta_i^c)$, wobei θ_i^l der Längenvergleich und θ_i^c den Inhaltsvergleich beinhaltet
- $\theta_i^l = 1 - e^{-\frac{1}{2} \left(\frac{\text{Len}(x) - \text{Len}(x')}{\sigma} \right)^2}$
- $\theta_i^c = \frac{1}{N} \sum_{i=0}^N \left[1 - e^{-\frac{1}{2} \left(\frac{x_i - x'_i}{\sigma} \right)^2} \right]$

3.2.4 Float⁵

Für die Datentyp Float ergibt sich folgender Definitionsbereich

- x, x' : Reelle Zahlen

Wir definieren Anonymisierungsgrad θ_f wie folgt:

- $\theta_f = 1 - e^{-\frac{1}{2} \left(\frac{x - x'}{\sigma} \right)^2}$

3.2.5 String

Für den Datentyp String ergibt sich folgender Definitionsbereich

- x, x' : Alphanumerisch / Character
- x besitzt eine Länge N
- x besitzt einen Zeichensatz mit z Elementen

Wir definieren den Anonymisierungsgrad θ_s wie folgt:

- $\theta_s = \frac{1}{2} (\theta_s^l + \theta_s^c)$, wobei θ_s^l der Längenvergleich und θ_s^c den Inhaltsvergleich beinhaltet
- $\theta_s^l = 1 - e^{-\frac{1}{2} \left(\frac{\text{Len}(x) - \text{Len}(x')}{\sigma} \right)^2}$
- $\theta_s^c = \frac{1}{2} (\theta_{s,c}^o + \theta_{s,c}^d)$, wobei $\theta_{s,c}^o$ den Vergleich bzgl. der Sortierreihenfolge der Characters hat und $\theta_{s,c}^d$ deren Verteilung vergleicht
- $\theta_{s,c}^d = \frac{1}{z} \sum_{i=0}^z \left[1 - e^{-\frac{1}{2} \left(\frac{D(x_i) - D(x'_i)}{\sigma} \right)^2} \right]$, wobei die $D(x_i)$ die Häufigkeit des Characters i in Element x angibt.
- $\theta_{s,c}^o = \frac{1}{z} \sum_{i=0}^z \left[1 - e^{-\frac{1}{2} \left(\frac{\rho_{i,i'}}{\sigma} \right)^2} \right]$, wobei $\rho_{i,i'} = \min(\text{T}(\text{Pos}(x_i), \text{Pos}(x'_i)))$. Folgende Nebenbedingungen müssen hierbei erfüllt sein:
 - o $\text{Pos}(x_i)$ gibt die Position des Characters im String an

⁴ $\text{Len}(N)$ ist definiert als die Anzahl der Digits des Integers

⁵ Zu Abgrenzungszwecken bzgl. Date ist in dieser Arbeit Float genutzt. Entsprechend der in Tabelle 1 definierten Datentypen entspricht dies Double.

- $\min(T(\text{Pos}(x_i), \text{Pos}(x'_i)))$ definiert die minimale Anzahl an Transpositionen⁶ um von der Position des Characters aus dem Original-String zur Position des Characters im anonymisierten String zu kommen
- Wenn der Character in x'_i nicht vorkommt, dann wird $\theta_{s,c}^o$ auf 1 definiert, dh. $\rho_{i,i'} = \square$
- Wenn zwei Characters des gleichen Typs wie im Originalstring vorkommen, dann wird der kürzere Abstand für $\theta_{s,c}^o$ gewählt
- Wenn die Positionen beider Characters in Original und anonymisierten String übereinstimmen ist $\rho_{i,i'} = 0$, und $\theta_{s,c}^o = 0$

Diskussion $\theta_{s,c}^o$

Für $\theta_{s,c}^o$ könnte man auch die bereits bekannte Hamming-Distanz⁷ einführen. Diese ist wesentlich leichtgewichtiger als die von uns eingeführte Größe $\theta_{s,c}^o$. Allerdings gehen bei der Anwendung der Hamming-Distanz wichtige Informationen über Lokalität und Anordnung verloren, so werden bspw. RAMA, MARA und ARAM gleich behandelt im Sinne der Hamming-Distanz. Dies wird in $\theta_{s,c}^o$ mit berücksichtigt.

3.2.6 Gesamtanonymisierungsgrad θ

Für den Anonymisierungsgrad $\theta_{g,g'}$ je Originaldatensatz g und je veränderten Datensatz g' ergibt sich nun als Summe über die im vorherigen definierten θ_l .

$\theta_{g,g'} = \frac{1}{M} \sum_{l=0}^M \theta_l$, für die Repräsentation mit genau einem Vertreter der oberen Entitäten ergibt sich

$$\theta_{g,g'} = \frac{1}{5} (\theta_d + \theta_i + \theta_s + \theta_f + \theta_b)$$

Der Gesamtanonymisierungsgrad θ ist die einfache Summe über g und g'

$$\theta = \frac{1}{N} \frac{1}{M} \sum_{g=0}^N \sum_{g'=0}^M \theta_{g,g'}$$

In unserer Arbeit gehen wir von einem gleichgroßen Datenbestand der Original und Veränderten Datenbank aus, d. h. $N = M$ und es wird kein Subsetting oder ähnliche Techniken angewendet.

4 Ergebnisse⁸

Nachdem im vorherigen Kapitel der Gesamtanonymisierungsgrad eingeführt, wollen wir nun diese Größen prüfen. Hierfür wurde eine Beispieldatenbank eingerichtet, bestehend aus produktiven und verfälschten Testdaten. Die zugrundeliegende Datenbank ist eine MySQL Datenbank. Die Verfälschung der Daten wurde durch das Tool Q-Up Studio⁹ von GFB durchgeführt. Die Daten sollen exemplarisch einer privaten Arztpraxis angehören und somit Patienteninformationen beinhalten, wie bspw. Name, Vorname, Telefonnummer, E-Mail uvm. Zunächst lassen wir unsere Algorithmen auf die im vorherigen Kapitel definierten

⁶ Transposition bezeichnet das Vertauschen zweier Characters. Damit wir eine Art Entfernung für die Characters in einem Wort definieren können, schränken wir das Vertauschen nur auf benachbarte Characters ein.

⁷ *Bell System Technical Journal*, XXIX (2), 1950, S. 147–160.

⁸ Entnommen "Anforderungsanalyse und Entwurf eines Tools zur Validierung von Testdaten gemäß der neuen Datenschutz Grundverordnung", J.-P. Heinrich, TU Chemnitz, 2017

⁹ <https://www.q-up-data.com/language/de/home/>

Anonymisierungsgrade wirken. Wobei wir θ_b hier nicht extra aufführen, da das Ergebnis trivial ist.

4.1 Date

Zu Validierungszwecken wurde $\sigma = 1$ der Einfachheit halber gesetzt, für alle im Folgenden gezeigten Berechnungen. Wir sehen die Ergebnisse für den Datentyp Date in Tabelle 2. Der implementierte Algorithmus ergibt für ein gleiches Datum 0.0. Bei fortschreitender Änderung erhöht sich der Anonymisierungsgrad.

Tabelle 2: Anonymisierungsgrad ϑ_d

| Original | Anonym | Ergebnis |
|------------|------------|----------|
| 24.02.1984 | 24.02.1984 | 0.0 |
| 24.02.1984 | 24.02.1928 | 0.08 |
| 24.02.1984 | 01.12.2012 | 0.7 |

4.2 Integer

Die Ergebnisse für den Datentyp Integer sind in Tabelle 3 dargestellt. Wir sehen, dass für die gleiche Zahl der Algorithmus eine Anonymisierung von 0.0 errechnet. Bei zunehmender Verfälschung wird auch eine erhöhte Anonymisierung errechnet. Als zusätzliche Ergänzung haben wir in den Algorithmus eine Nebenbedingung implementiert, dass bei Umkehr der bestehenden Zahlenfolge die Anonymisierung ebenso den Wert 0.0 ergeben soll.

Tabelle 3: Anonymisierungsgrad ϑ_i

| Original | Anonym | Ergebnis |
|------------|------------|----------|
| 1289334018 | 1289334018 | 0.0 |
| 1289334018 | 1289300000 | 0.67 |
| 1289334018 | 8104339821 | 0.0 |

4.3 Float

Die Ergebnisse für den Datentyp Float sind in Tabelle 4 dargestellt. Wir sehen, dass für eine gleiche Zahl der Algorithmus eine Anonymisierung von 0.0 errechnet. Bei zunehmender Verfälschung erhöht sich der Grad der Anonymisierung.

Tabelle 4: Anonymisierungsgrad ϑ_f

| Original | Anonym | Ergebnis |
|----------|--------|----------|
| 123,45 | 123,45 | 0.0 |
| 123,45 | 12,345 | 0.5 |
| 123,45 | 6,7890 | 0.82 |

4.4 String

Die Ergebnisse für den Datentyp String sind in Tabelle 5 dargestellt. Wir sehen, dass für den gleichen String der Algorithmus eine Anonymisierung von 0.0 errechnet. Wir sehen, dass verschiedene Maskierungsfunktionen einen verschiedenen Einfluss auf den Anonymisierungsgrad haben. So haben Länge, Inhalt und Korrelation eben einen anderen Einfluss, ob das menschlich Gehirn in der Lage ist, eine Beziehung zwischen Originalwert

und verfälschtem herzustellen. Dies wurde beispielhaft in den Algorithmen umgesetzt, muss allerdings gerade in Form der Justierung des Parameters σ oder Einfügen von weiteren Nebenbedingungen erweitert werden.

Tabelle 5: Anonymisierungsgrad ϑ_s

| Original | Anonym | Ergebnis |
|-----------|----------------|----------|
| Annabelle | Annabelle | 0.0 |
| Annabelle |Annabelle | 0.5 |
| Annabelle | Anxxxxlle | 0.05 |
| Annabelle | Belleanna | 0.17 |

4.5 Gesamtanonymisierungsgrad

Um nun das Validierungstool tiefergehend zu testen, wurde die originale Tabelle achtmal vervielfältigt, wobei sukzessive mehr Spalten anonymisiert wurden. Davon ausgehend ist zu erwarten, dass der Gesamtanonymisierungsgrad mit jeder zusätzlich anonymisierten Spalte steigt. In einem zweiten Schritt werden dann die einzelnen Vergleichsalgorithmen mit anschaulichen Werten getestet und diskutiert.

Das Ergebnis ist in Abbildung 1 dargestellt. Sukzessiv werden mehr Spalten der Tabelle anonymisiert (dies spiegelt den Grad der Anonymisierung wider). Bei Grad 1 wurde nur der Vorname anonymisiert, bei Grad 2 Vor- und Nachname und dementsprechend weiter.

Wie erwartet steigt der Grad der Anonymisierung mit jeder anonymisierten Spalte an, was beweist, dass alle Vergleichsalgorithmen arbeiten und im Gesamten bei jedem Schritt der Anonymisierung eine signifikante Verfremdung der Werte stattfindet. Beginnend mit nur anonymisiertem Vornamen im Grad 1 erreicht die Berechnung des Gesamtgrades lediglich einen Wert von 3,12 Prozent. Bei der Betrachtung des 4. Grades der Anonymisierung, bei dem die Spalten Vorname, Nachname, Geburtsdatum und E-Mail anonymisiert wurden, erreicht der Gesamtgrad bereits einen Wert von 14,17 Prozent. Zu diesem Zeitpunkt sind 44 Prozent der Tabellenspalten anonymisiert. Schließlich erreicht der Gesamtanonymisierungsgrad bei der Anonymisierung aller Spalten in Grad 9 einen Wert von 41,87 Prozent.

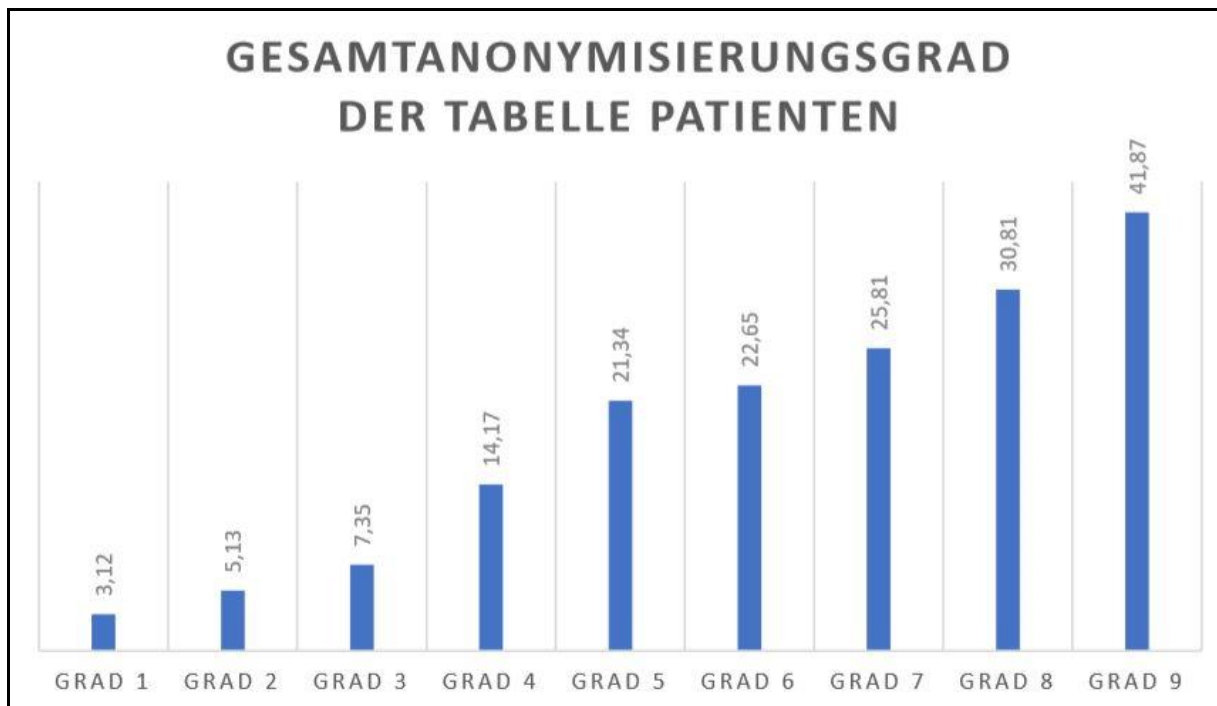


Abbildung 3: Gesamtanonymisierungsgrad 9 als Funktion der anonymisierten Spalten

5 Resümee und Ausblick

Es wurde mit Hilfe eines definierten mathematischen Modells gezeigt, wie man beispielhaft die Anonymisierung und damit Ähnlichkeit zwischen zwei Datenbanken errechnen kann. Mit der Unterstützung des Spring Frameworks¹⁰ und anderer Open Source Technologien wurde ein Tool als Entwurf realisiert und kann fortan als Orientierungshilfe für weitere Forschungen zu diesem Thema dienen. Um als gewinnbringendes Produkt in einem Unternehmen dienen zu können, bedarf es jedoch einer skalierbaren Architektur. Auch die einzelnen Algorithmen zum Vergleich bedürfen einer tiefgreifenden Optimierung hinsichtlich solcher Aspekte wie etwa dem natürlichen Lesen, bei dem gerade der Beginn und das Ende eines Wortes relevant sind.

¹⁰ <https://projects.spring.io/spring-framework/>

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

- CSR-10-01** Maximilian Eibl, Jens Kürsten, Robert Knauf, Marc Ritter , Workshop Audiovisuelle Medien, Mai 2010, Chemnitz
- CSR-10-02** Thomas Reichel, Gudula Rüniger, Daniel Steger, Haibin Xu, IT-Unterstützung zur energiesensitiven Produktentwicklung, Juli 2010, Chemnitz
- CSR-10-03** Björn Krellner, Thomas Reichel, Gudula Rüniger, Marvin Ferber, Sascha Hunold, Thomas Rauber, Jürgen Berndt, Ingo Nobbers, Transformation monolithischer Business-Softwaresysteme in verteilte, workflowbasierte Client-Server-Architekturen, Juli 2010, Chemnitz
- CSR-10-04** Björn Krellner, Gudula Rüniger, Daniel Steger, Anforderungen an ein Datenmodell für energiesensitive Prozessketten von Powertrain-Komponenten, Juli 2010, Chemnitz
- CSR-11-01** David Brunner, Guido Brunnett, Closing feature regions, März 2011, Chemnitz
- CSR-11-02** Tom Kühnert, David Brunner, Guido Brunnett, Betrachtungen zur Skelettextraktion umformtechnischer Bauteile, März 2011, Chemnitz
- CSR-11-03** Uranchimeg Tudevdayva, Wolfram Hardt, A new evaluation model for eLearning programs, Dezember 2011, Chemnitz
- CSR-12-01** Studentensymposium Informatik Chemnitz 2012, Tagungsband zum 1. Studentensymposium Chemnitz vom 4. Juli 2012, Juni 2012, Chemnitz
- CSR-12-02** Tom Kühnert, Stephan Rusdorf, Guido Brunnett, Technischer Bericht zum virtuellen 3D-Stiefeldesign, Juli 2012, Chemnitz
- CSR-12-03** René Bergelt, Matthias Vodel, Wolfram Hardt, Generische Datenerfassung und Aufbereitung im Kontext verteilter, heterogener Sensor-Aktor-Systeme, August 2012, Chemnitz
- CSR-12-04** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Knauf, Jens Kürsten, Albrecht Kurze, Markus Rickert, Marc Ritter , Schlussbericht zum InnoProfile Forschungsvorhaben sachsMedia - Cooperative Producing, Storage, Retrieval and Distribution of Audiovisual Media (FKZ: 03IP608), September 2012, Chemnitz
- CSR-12-05** Anke Tallig, Grenzgänger - Roboter als Mittler zwischen der virtuellen und realen sozialen Welt, Oktober 2012, Chemnitz

Chemnitzer Informatik-Berichte

- CSR-13-01** Navchaa Tserendorj, Uranchimeg Tudevdagva, Ariane Heller, Grenzgänger - Integration of Learning Management System into University-level Teaching and Learning, Januar 2013, Chemnitz
- CSR-13-02** Thomas Reichel, Gudula Rüniger, Multi-Criteria Decision Support for Manufacturing Process Chains, März 2013, Chemnitz
- CSR-13-03** Haibin Xu, Thomas Reichel, Gudula Rüniger, Michael Schwind, Softwaretechnische Verknüpfung der interaktiven Softwareplattform Energy Navigator und der Virtual Reality Control Platform, Juli 2013, Chemnitz
- CSR-13-04** International Summerworkshop Computer Science 2013, Proceedings of International Summerworkshop 17.7. - 19.7.2013, Juli 2013, Chemnitz
- CSR-13-05** Jens Lang, Gudula Rüniger, Paul Stöcker, Dynamische Simulationskopplung von Simulink-Modellen durch einen Functional-Mock-up-Interface- Exportfilter, August 2013, Chemnitz
- CSR-14-01** International Summerschool Computer Science 2014, Proceedings of Summerschool 7.7.-13.7.2014, Juni 2014, Chemnitz
- CSR-15-01** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Herms, Stefan Kahl, Jens Kürsten, Albrecht Kurze, Robert Manthey, Markus Rickert, Marc Ritter, ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL, Januar 2015, Chemnitz
- CSR-15-02** Maximilian Speicher, What is Usability? A Characterization based on ISO 9241-11 and ISO/IEC 25010, Januar 2015, Chemnitz
- CSR-16-01** Maxim Bakaev, Martin Gaedke, Sebastian Heil, Kansei Engineering Experimental Research with University Websites, April 2016, Chemnitz
- CSR-18-02** Jan-Philipp Heinrich, Carsten Neise, Andreas Müller, Ähnlichkeitsmessung von ausgewählten Datentypen in Datenbanksystemen zur Berechnung des Grades der Anonymisierung, Februar 2018, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz