Zusammenfassung

Innerhalb des UniTEC-Empfehlungsdienstes werden Pseudonyme benutzt, um die einzelnen Benutzer zu identifizieren. Die Funktion eines Pseudonyms ist es außerdem, die wahre Identität seines Benutzers zu verschleiern.

Nach einer Einführung in den UniTEC-Empfehlungsdienst und das Identitätsmanagement folgen die Ergebnisse dieser Diplomarbeit. Es wurde untersucht, auf welche Weise die Identität des Benutzers eines Pseudonyms speziell innerhalb des UniTEC-Empfehlungsdienstes durch Angreifer aufgedeckt werden kann. Ebenso wurde untersucht, wie diese Angriffe mit Hilfe technischer Möglichkeiten erschwert bzw. verhindert werden können. Nach einem Fazit und einem Ausblick wird die erstellte Software vorgestellt.

Inhaltsverzeichnis

1. Einführung **7**
1.1. Der UniTEC-Empfehlungsdienst 8
1.1.1. Die Empfehlungen 8
1.1.2. Der Vertrauenswert 10
1.1.3. Statistiken – Das Expertisenmodell 12
1.1.4. Das Kommunikationssubsystem 16
1.1.5. Benutzerschnittstelle 17
1.1.6. Überblick 17

2. Identitätsmanagement innerhalb des UniTEC-Empfehlungsdienstes **19**
2.1. Was ist Identitätsmanagement? 19
2.2. Bestehende Projekte 21
2.2.1. Liberty Alliance 21
2.2.2. Microsoft Passport 21
2.2.3. SUN Network Identity 21
2.3. Anforderungen an das Identitätsmanagement 22
2.3.1. Identifizierbarkeit 22
2.3.2. Pseudonymität 22
2.3.3. Pseudonymität bewahren 23
2.3.4. Pseudonymität absichtlich aufdecken 23
2.4. Pseudonymmanagement 23

3. Pseudonyme **25**
3.1. Pseudonymarten 25
3.1.1. Zusätzliche Komponenten der eigenen Pseudonyme 26
3.1.2. Zusätzliche Komponenten fremder Pseudonyme 26

3.2. Pseudonymmanagement 27

4. Identifizierbarkeit eines Pseudonyms 29

5. Gefahren für die Pseudonymität 33

5.1. Kommunikationssubsystem 34
5.2. Profilerstellung 35
5.3. Empfehlungen 39
5.3.1. Texte 41
5.3.2. Binäre Daten 46
5.4. Externes Weitergeben von Informationen 47
5.5. Warnung des Benutzers 47
5.6. Zusammenfassung 48

6. Absichtliche Aufdeckung der Identität 49

7. Fazit und Ausblick 53

7.1. Fazit 53
7.2. Ausblick 54

A. Software 57

A.1. Architektur 57
A.2. Klassenbeschreibung 59
A.2.1. com.hp.unitec.expertise 59
A.2.2. com.hp.unitec.identity 59
A.2.3. com.hp.unitec.gui 60
A.2.4. com.hp.unitec.mvc 61
A.2.5. com.hp.unitec.pseudonym 63
A.2.6. com.hp.unitec.pseudonym.monitoring 64
A.2.7. com.hp.unitec.pseudonym.monitoring.filter 65
A.2.8. com.hp.unitec.pseudonym.responsibility 65
A.2.9. com.hp.unitec.pseudonym.storage 65
A.2.10. com.hp.unitec.util 66

B. Versuche zur Ähnlichkeit von Texten 67

C. Glossar 71

D. Projektplan **75**
D.1. Einleitung 75
D.1.1. Zweck dieses Dokuments 75
D.1.2. Projektübersicht und Motivation 75
D.2. Beschreibung der Arbeitspakete 77
D.2.1. Definition der Arbeitspakete 77
D.2.2. Abhängigkeit der Arbeitspakete 80
D.3. Zeitplan 81
D.4. Dokumente und Meilensteine 81
D.4.1. Synchronisationspunkte 83
D.5. Anforderung an die Umgebung 83
D.5.1. Verwendete Software 84

Abbildungsverzeichnis **85**

Tabellenverzeichnis **87**

Literaturverzeichnis **89**

Erklärung **91**

1. Einführung

Aufgrund der arbeitsteiligen Welt müssen alle Menschen Produkte einkaufen oder Dienstleistungen in Anspruch nehmen, sei es der Kauf von Lebensmitteln, um die Familie zu versorgen, oder der Gang zum Zahnarzt. Dabei existieren verschiedene Alternativen an Unternehmen, die die Produkte herstellen oder Dienstleistungen anbieten. Nachfolgend wird verdeutlicht, wie es zur Auswahl einer bestimmten Alternativen unter vielen kommt:

1. Der Bedarf eines Produktes oder einer Dienstleistung wird festgestellt und die Entscheidung getroffen, diesen Bedarf zu erfüllen.

2. Es werden Unternehmen gesucht, die den Bedarf erfüllen können. Diese Suche wird von dem zur Verfügung stehenden Informationsangebot (Werbung, Produktinformationen, Empfehlungen von anderen ...) beeinflusst.

3. Eine Entscheidung wird getroffen. Diese basiert auf den im vorherigen Schritt gesammelten Informationen, auf dem Bekanntheitsgrad und Ruf des Unternehmens und auf eigenen Erfahrungen.

4. Das ausgesuchte Produkt wird gekauft bzw. die ausgesuchte Dienstleistung wird in Anspruch genommen.

Mit der stetig gestiegenen Anzahl von Personen, die Zugang zum Internet und dem darin angebotenen Service World Wide Web haben, wurden Schritte der Kaufentscheidung auch auf dieses Medium übertragen. Informationsangebote, die für Schritt 2 benötigt werden, stehen zur Verfügung: Produktionformationen können von Webseiten der Hersteller abgefragt werden, Werbung wird auf populären Webseiten in Form von Bannern eingeblendet, und in speziellen Foren wird über die Qualität von

verschiedenen Produkten in Form von Empfehlungen diskutiert. Auch der Schritt 4 wurde durch Shopping-Systeme im World Wide Web möglich.

Die vorherrschende Form der Empfehlungen in Foren des WWW bildet die Realität nur unzureichend ab. Der Wert einer Empfehlung ist nicht, wie in der realen Welt, an das Vertrauen in den Empfehlenden gekoppelt. Diese Verbindung muss außerhalb des Empfehlungssystems aufgebaut werden (zum Beispiel in Form einer extra erstellten Datenbank). Nach der Inanspruchnahme einer Empfehlung kann sich das Vertrauen in einen Empfehlenden vergrößern oder verringern und entsprechend nachgeführt werden. Da eine solche Vorgehensweise mit einem erheblichen Aufwand verbunden ist, werden vermutlich nur sehr wenige Menschen derartige Verfahren betreiben. Mit dem UniTEC-Empfehlungsdienst wird versucht, das Vertrauen in einen Empfehlenden abzubilden und die Änderung des Vertrauens in eine Person aufgrund der Erfahrungen mit den Empfehlungen zu automatisieren.

1.1. Der UniTEC-Empfehlungsdienst

Der UniTEC-Empfehlungsdienst stellt eine Möglichkeit bereit Empfehlungen zu Unternehmen und ihren Produkten zu erstellen und auszutauschen. Auch ist es möglich, das gewährte Vertrauen bezüglich des Empfehlenden zu verwalten, d. h. die Höhe des Vertrauens kann der Benutzer bei Erhalt einer Empfehlung einsehen und so seine Kaufentscheidung optimieren. Des Weiteren kann die Höhe des Vertrauens verändert werden. Im Normalfall geschieht dies, nachdem der Kauf eines Produktes bzw. die Inanspruchnahme einer Dienstleistung vonstatten gegangen ist und erste eigene Erfahrungen mit dem Produkt bzw. der Dienstleistung gesammelt werden konnten.

1.1.1. Die Empfehlungen

Eine Empfehlung ist ein Ratschlag zu einem bestimmten Sachverhalt, der von einer bestimmten Person gegeben wird. Da Empfehlungen sehr unterschiedlich gestaltet sein können, wurde innerhalb des UniTEC-Empfehlungsdienstes darauf geachtet, dass diese Vielfalt abgebildet werden kann.

Eine Empfehlung wird einer oder mehreren Kategorien zugeordnet. Diese Kategorien erleichtern die weiter unten beschriebene Suche nach Empfehlungen. Durch sie kann das grobe Sachgebiet, in dem eine Empfehlung zu einem bestimmten Produkt oder einer bestimmten Dienstleistung gesucht wird, bestimmt werden.

Die Suche nach Empfehlungen

In der realen Welt werden auf der Suche nach Empfehlungen diejenigen Menschen befragt, denen man in dem betroffenen Sachgebiet hohes Vertrauen entgegenbringt. Eventuell werden zusätzlich Menschen aufgesucht, denen man zwar nicht so sehr vertraut, von denen man aber weiß, dass sie in dem entsprechenden Sachgebiet viel Wissen haben. Diese beiden Vorgänge werden im UniTEC-Empfehlungsdienst ebenfalls abgebildet. Bevor eine Anfrage nach einer Empfehlung ausgeführt werden kann, muss bestimmt werden, an wen die Anfrage gestellt werden soll. Dazu werden aus den weiter unten beschriebenen Vertrauens- und Expertisenmodellen diejenigen Empfehlenden herausgesucht,

1. denen hohes Vertrauen in der entsprechenden Kategorie entgegengebracht wird,

2. die einen hohen Anteil der in der Vergangenheit von dem Anfragenden gestellten Anfragen in der selben Kategorie beantworten konnten, d. h. eine Empfehlung zurückgesendet haben, oder

3. die einen hohen Anteil der Anfragen in der Vergangenheit erfolgreich an einen Dritten weitergeleitet haben. Das heißt, die Anfrage wurde an einen Empfehlenden weitergeleitet, der die Anfrage mit dem Zurücksenden einer Empfehlung beantwortet hat.

Ergänzt wird die Menge dieser Empfehlenden durch eine Anzahl von zufällig ausgewählten Empfehlenden aus der Menge der bekannten Empfehlenden. Dadurch können sich neue Kontakte zu Empfehlenden auftun, die eine hohe Expertise in einer gewissen Kategorie haben, von der man noch nichts wusste. Die Menge aller Empfehlenden, an die die Anfrage weitergegeben wird, wird Nachbarschaft genannt.

In der vorigen Aufzählung wurde schon erwähnt, dass Anfragen weitergeleitet werden können. Dieses Verhalten ist unter anderem dann sinnvoll, wenn ein angefragter Empfehlender keine passende Empfehlung zu der Anfrage parat hat, mit der er hätte antworten können. Dann wird die Anfrage an eine von dem Weiterleitenden erstellte Nachbarschaft weitergegeben, in der Hoffnung, dass ein Empfehlender aus dieser Nachbarschaft die Anfrage beantworten kann.

Weitere Informationen zu Empfehlungen und der Suche nach diesen finden sich in [Terdic, 2003].

1.1.2. Der Vertrauenswert

Der Vertrauenswert ist eine qualitative Abbildung des Vertrauens. Jeder Benutzer des Empfehlungsdienstes hat seine eigene Zuordnung eines Vertrauenswertes zu einem Empfehlenden. Diese Abbildung wird durch das Vertrauensmodell bestimmt, das unten näher beschrieben wird.

Da ein Empfehlender Empfehlungen in unterschiedlichen Kategorien erstellen kann und diese Empfehlungen je nach Wissensstand des Empfehlenden unterschiedliche Qualität aufweisen können (beispielsweise können Empfehlungen eines Empfehlenden in der Kategorie Auto hervorragend und zutreffend sein, jedoch die Empfehlungen in der Kategorie Aktien nur eine geringe Qualität haben), wird jedem Empfehlenden in jeder Kategorie, in der er eine Empfehlung erstellt hat, ein eigener Vertrauenswert zugeordnet.

Der Vertrauenswert ist eine dynamische Komponente, die sich im Laufe der Zeit durch Enttäuschung oder Bestätigung von Vertrauen ändern kann. Eine Änderung in einer gewissen Kategorie kann Einfluss haben auf den Vertrauenswert in anderen Kategorien. Beispielsweise wird eine positive Vertrauenswertänderung in der Kategorie Auto eine positive Vertrauenswertänderung in der Kategorie Kleinwagen des Empfehlenden bewirken, da diese beiden Kategorien thematisch eng beieinander liegen. Die Vertrauenswertänderung in den nicht direkt betroffenen Kategorien wird im Normalfall nicht so hoch ausfallen wie in der Kategorie, die die Vertrauenswertänderung ausgelöst hat.

Das Vertrauensmodell

Das Vertrauensmodell gibt vor, welche Werte der Vertrauenswert annehmen kann, wie sich positive oder negative Erfahrungen mit einer Empfehlung auf den Vertrauenswert auswirken und welche Auswirkungen diese Vertrauenswertänderung auf die anderen Kategorien hat.

Der Vertrauenswert kann einen Wert aus dem Intervall [0,1] annehmen. Hierbei bedeutet der Wert Null, dass dem Empfehlenden in der betroffenen Kategorie nicht vertraut wird. Der Wert Eins entspricht vollem Vertrauen in den Empfehlenden in der entsprechenden Kategorie. Die Zwischenwerte bilden das Spektrum zwischen vollem und keinem Vertrauen ab. Ein Beispiel für die Transformation von Vertrauen in den Vertrauenswert kann Abbildung 1.1 entnommen werden.

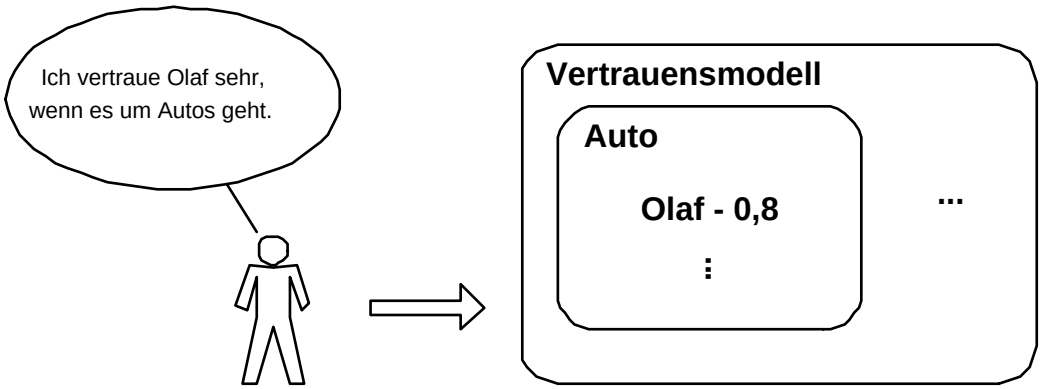


Abbildung 1.1.: Das Vertrauen einer Person in einen andere Person wird im UniTEC-Empfehlungsdienst auf einen Wert zwischen Null und Eins abgebildet.

Eine Empfehlung kann innerhalb des UniTEC-Empfehlungsdienstes von einem Benutzer nur binär eingeschätzt werden. Er kann die Empfehlung als gut oder als schlecht einstufen. Diese Entscheidung fließt in den Vertrauens-Aktualisierungs-Algorithmus (Trust-Update-Algorithm) ein, der aufgrund der Entscheidung und evtl. weiterer Daten (zum Beispiel die vorherigen Einschätzungen der Empfehlungen des Empfehlenden in der entsprechenden Kategorie) berechnet, auf welchen Wert sich der Vertrauenswert ändert. Unterschiedliche Algorithmen wurden von Fabian Aichele untersucht. Die Ergebnisse können [Aichele, 2003] entnommen werden.

Weiter oben wurde erwähnt, dass eine Änderung des Vertrauenswertes auf den Ver-

trauenswert in anderen Kategorien Einfluss haben kann. Um bestimmen zu können, auf welche Kategorien und in welchem Maße eine Änderung Einfluss hat, wird der Zusammenhang zwischen den Kategorien als gewichteter, gerichteter Graph modeliert (vgl. Abbildung 1.2). Auf welche Weise die Änderung des Vertrauenswertes in zusammenhängenden Kategorien erfolgt, wird ebenfalls in der Studienarbeit von Fabian Aichele [Aichele, 2003] dargelegt.

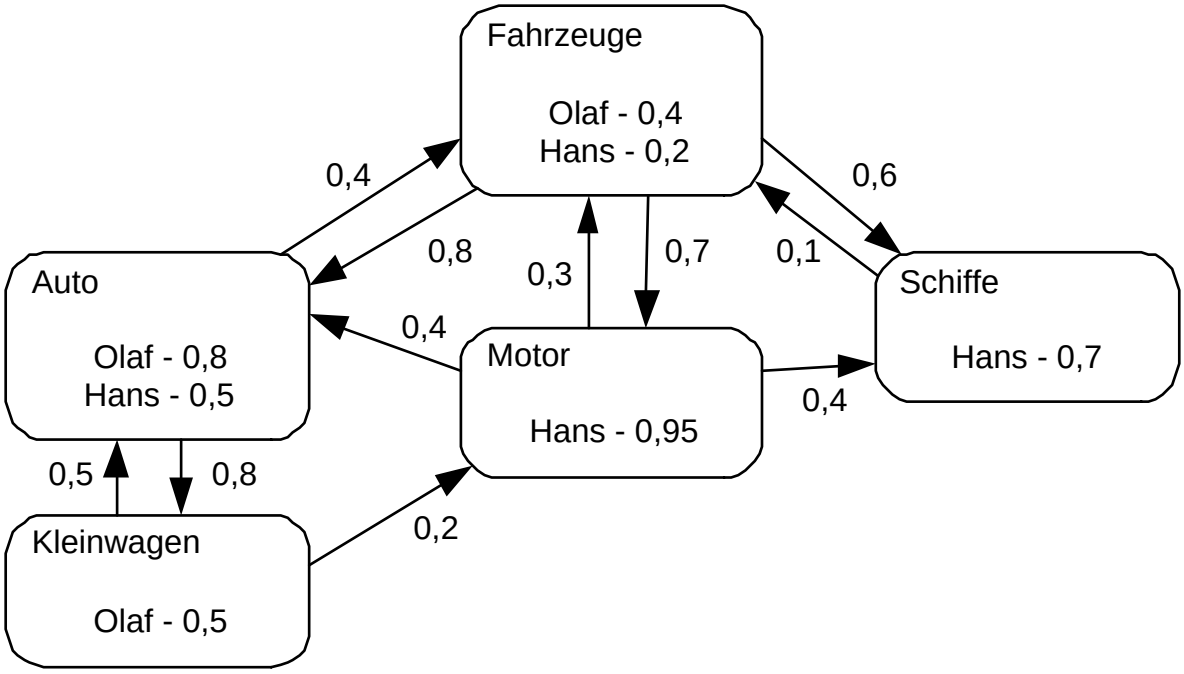


Abbildung 1.2.: Dargestellt ist ein Ausschnitt aus einem Vertrauensmodell. Die einzelnen Kategorien sind in einem gerichteten und gewichteten Graphen organisiert. Die Gewichte geben an, wie stark sich eine Änderung in einer Kategorie auf die benachbarten Kategorien auswirkt.

1.1.3. Statistiken – Das Expertisenmodell

Das Expertisenmodell gibt vor, welche Informationen über einen Kontakt mit einem Empfehlenden gespeichert werden sollen. Diese Informationen dienen dazu, einen Teil der sogenannten Nachbarschaft zu bilden. Gespeichert werden:

- die Anzahl der Empfehlungen, die der Empfehlende zum Zeitpunkt der letzten Aktualisierung in der Kategorie gespeichert hatte,

- die Anzahl der Anfragen, die an den Empfehlenden ausgesandt wurden,

- die Anzahl der Anfragen, die der Empfehlende mit einer Empfehlung beantworten konnte,

- die Anzahl der Anfragen, die der Empfehlende weitergeleitet hat und die von einem nachfolgenden Empfehlenden beantwortet wurden,

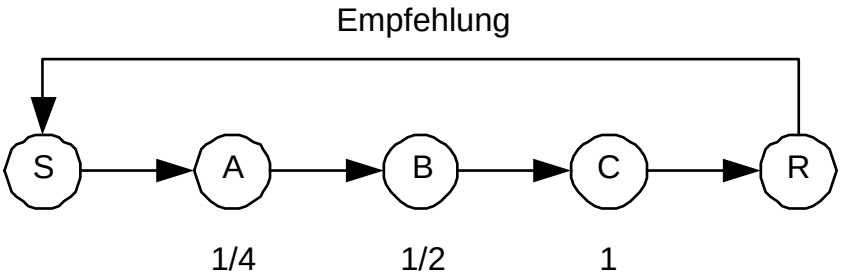
- die durchschnittliche Gewissheit, die die Empfehlungen des Empfehlenden in der Kategorie haben. Die Gewissheit einer Empfehlung gibt an, wie sicher sich der Empfehlende über die Bewertung des Gegenstandes der Empfehlung ist.

Diese Daten werden aufgeschlüsselt nach den einzelnen Kategorien für jeden Empfehlenden, mit dem der Benutzer Kontakt hatte, gespeichert.

Da Empfehlungen mehrfach weitergeleitet werden können, ergibt sich die Frage, für welche Empfehlenden die Anzahl der Anfragen, die der Empfehlende weitergeleitet hat, erhöht wird und um welchen Wert. Zur Auswahl stehen die folgenden Alternativen:

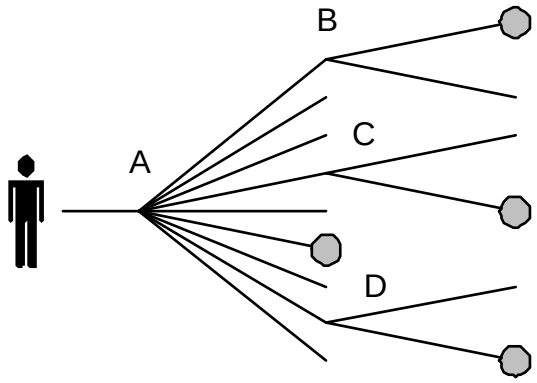
1. Für alle, die an der Weiterleitung der Anfrage bis zum Antwortenden beteiligt waren, wird der Wert um den gleichen Anteil erhöht.

2. Für alle, die an der Weiterleitung der Anfrage bis zum Antwortenden beteiligt waren, wird der Wert entsprechend der Entfernung zum Antwortenden erhöht.



3. Nur für den letzten, der die Anfrage weitergeleitet hat, wird der Wert erhöht. Dieser hat die Anfrage an denjenigen weitergeleitet, der sie beantworten konnte.

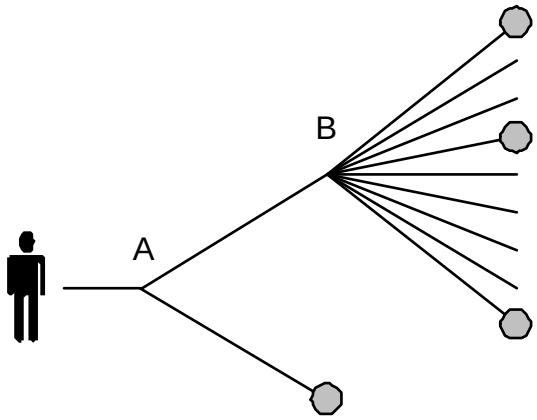
Die Abbildungen 1.3 und 1.4 zeigen zwei unterschiedliche Konfigurationen der Nachbarschaft und der berechneten Hub-Werte[1] bei Anwendung der oben aufgeführten Methoden. Die grauen Kreise bedeuten, dass von diesen Pseudonymen eine Antwort gesendet wird. Ziel ist es, diejenigen Pseudonyme als Hubs in die Nachbarschaft aufzunehmen, die viele Pseudonyme kennen. Dies ist in Abbildung 1.3 das Pseudonym A und in Abbildung 1.4 das Pseudonym B. Anhand der aufgeführten Tabellen ist zu erkennen, dass die Methode 2 das gewünschte Verhalten liefert. Die Methode 1 weist in einem Fall wie in Abbildung 1.3 das beste Verhalten auf (der Knoten A ist im Vergleich zu den anderen Knoten stärker gewichtet), jedoch ergibt die Berechnung in dem anderen Beispiel ein unerwünschtes Verhalten. Die Methode 3 zeigt optimales Verhalten in Abbildung 1.4, im anderen Fall jedoch werden alle Pseudonyme gleich gewichtet.



Pseudonym	Alle gleich	Nach Entfernung	Nur Letzter
A	4	2,5	1
B	1	1	1
C	1	1	1
D	1	1	1

Abbildung 1.3.: Der Sender sendet eine Nachricht an Pseudonym A, das eine sehr große Nachbarschaft für die Kategorie der Anfrage hat. Jedoch antworten nur sehr wenige dieser Nachbarn direkt. Einige (B, C und D) leiten die Anfrage an zwei Pseudonyme weiter, von denen jeweils einer antwortet. Abgebildet ist ebenfalls die Tabelle der nach diesem Vorgang berechneten Hub-Werte.

[1] Ein Hub ist ein Pseudonym, das einen hohen Anteil an erfolgreich weitergeleiteten Anfragen hat. Der Hub-Wert drückt diesen hohen Anteil aus.



Pseudonym	Alle gleich	Nach Entfernung	Nur Letzter
A	4	2,5	1
B	3	3	3

Abbildung 1.4.: Der Sender sendet seine Anfrage an ein Pseudonym mit einer kleinen Nachbarschaft in der Kategorie der Anfrage (Pseudonym A). Dieses sendet die Anfrage weiter an ein Pseudonym, das eine große Nachbarschaft besitzt (Pseudonym B). Ebenfalls dargestellt ist die Tabelle der Hub-Werte, nachdem die Pseudonyme ihre Antworten gesendet haben.

Beitrag zur Nachbarschaft

Aus dem Expertisenmodell werden zwei unterschiedliche Mengen von Empfehlenden für die Nachbarschaft ausgewählt. Zum einen sind das diejenigen Empfehlenden, die einen hohen Anteil der Anfragen mit einer Empfehlung beantwortet haben. Diese werden Authority genannt. Zum anderen sind es diejenigen Empfehlenden, die einen hohen Anteil der Anfragen erfolgreich weitergeleitet haben. Diese werden Hubs genannt.

Aktualisierung des Expertisenmodells

Das Expertisenmodell wird aktualisiert, sobald eine Antwort auf eine Anfrage von einem Pseudonym erhalten wird. Die Anzahl der gesendeten Anfragen an das Pseudonym wird für alle Pseudonyme der Versandkette inkrementiert. Für das antwortende Pseudonym wird zusätzlich der Wert des Zählers erhöht, der angibt, wieviele Anfragen direkt beantwortet wurden (Authority). Für die restlichen Pseudonyme der Versandkette wird der Zähler der erfolgreich weitergeleiteter Anfragen nach dem weiter oben

aufgeführten Algorithmus entsprechend der Entfernung des Pseudonyms vom Antwortenden erhöht. Mit Erhalt der Antwort kann auch die durchschnittliche Gewissheit der Empfehlungen des Empfehlenden in der entsprechenden Kategorie neu berechnet werden.

Die Pseudonyme versenden in regelmäßigen Abständen eine Nachricht über die Anzahl der erstellten Empfehlungen pro Kategorie an eine zufällig ausgewählte Teilmenge der Pseudonyme. Wird eine solche Nachricht empfangen, so können die Einträge innerhalb des Expertisenmodells aktualisiert werden.

1.1.4. Das Kommunikationssubsystem

Damit Empfehlungen ausgetauscht werden können, müssen die Anfragenden und die Empfehlenden kommunizieren können. Dazu dient ein Peer-to-Peer-Netzwerk, das in [Terdic, 2003] ausführlich beschrieben wird.

Charakteristisch für ein Peer-to-Peer-Netzwerk ist, dass alle beteiligten Instanzen gleichberechtigt sind. Das bedeutet, dass jeder Teilnehmer sowohl Dienste anbieten als auch Dienste nutzen kann. Hierbei können beliebige Komponenten eines Rechners (Speicherplatz, Rechenleistung, Dienste ...) innerhalb des Netzwerkes zur Nutzung angeboten oder in Anspruch genommen werden. Typische Beispiele für Peer-to-Peer-Netzwerk-Anwendungen sind Instant Messaging, File Sharing und Grid Computing. Im Falle des UniTEC-Empfehlungsdienstes werden Empfehlungen angeboten und können gesucht werden.

Die später beschriebene Pseudonymität stellt bestimmte Anforderungen an das Kommunikationssubsystem. Dieses muss verschlüsselte Nachrichtenübertragung in einer anonymen Weise bieten. Das heißt, weder der Empfänger noch der Sender noch dritte Personen, die an der Kommunikation beteiligt sind, können eine Verbindung zwischen einer Identifikationsnummer, die der Kommunikation dient, und der Identifikationsnummer der Benutzer des UniTEC-Empfehlungsdienstes erkennen.

1.1.5. Benutzerschnittstelle

Damit der Benutzer den Empfehlungsdienst nutzen kann, ist eine Benutzerschnittstelle notwendig. Der Benutzer kann mit Hilfe dieser unter anderem die Empfehlungen erstellen und nach Empfehlungen suchen.

1.1.6. Überblick

Die beteiligten Komponenten und deren Zusammenhänge sind in Abbilung 1.5 zusammengefasst. Orthogonal zu den oben bereits erwähnten Komponenten ist das Identitätsmanagement, das in dieser Diplomarbeit beschrieben wird. Unter das Identitätsmanagement fallen die Verwaltung und Bereitstellung der eigenen und fremden Pseudonyme und die Wahrung der Pseudonymität. Diese Konzepte werden im weiteren Verlauf näher beschrieben.

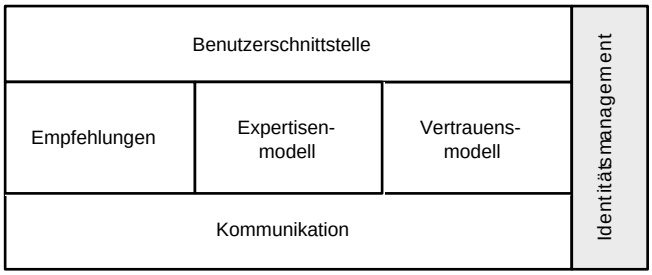


Abbildung 1.5.: Übersicht über den Zusammenhang der beschriebenen Komponenten

2. Identitätsmanagement innerhalb des UniTEC-Empfehlungsdienstes

Die folgenden Abschnitte beschreiben, was unter dem Begriff Identitätsmangement verstanden wird und welche Anforderungen innerhalb des UniTEC-Empfehlungsdienstes an ein Identitätsmanagement gestellt werden.

2.1. Was ist Identitätsmanagement?

Identitätsmanagement ist in der realen Welt ein natürlicher Vorgang, der täglich von jedem durchgeführt wird. Das zentrale Element des Managements der eigenen Identität ist: „Welche Person bekommt wieviel Informationen über meine eigene Identität?“. Die Auswahl der Bestandteile der eigenen Identität, wie zum Beispiel der Name oder die Adresse einer Person, hängt von verschiedenen Faktoren ab. Zum einen von der Situation, in der sich eine Person befindet. Im Privatbereich wird eine Person mehr Information über sich preisgeben als im beruflichen Umfeld. Zum anderen hängt die Menge der preisgegebenen Informationen von der Person ab, die diese Informationen erhalten soll. Welche Rolle nimmt die empfangende Person ein? Einem Arzt gibt man andere Informationen über die eigene Identität preis, als einer Verkäuferin im Supermarkt.

Diesen natürlichen Prozess auf das Internet zu übertragen, erfordert zusätzlichen Aufwand. Die Zahl der Dienste, die Identitätsinformationen fordern, um sie benutzen zu können, steigt zusehends. Ohne technische Hilfsmittel ist es kaum noch möglich, den Überblick zu behalten, welcher Dienst welche Daten von einer Person gespeichert hat. Die Speicherung der Daten und die leichte Zusammenfassbarkeit dieser über mehrere

Dienste hinweg erleichtert es den Dienstanbietern, ein umfassendes Profil zu erstellen. Die Profilbildung ist in der realen Welt normalerweise nicht gegeben. Eine bestimmte Person geht in der Menge von Menschen, die von einer Verkäuferin an einem Tag bedient werden, unter. Ebenfalls ist der Aufwand, den eine Person ohne technische Unterstützung betreiben muss, um ein Profil einer Person zu erstellen, zu hoch, um dies für eine große Zahl von Personen durchzuführen (Telefongespräche, Nachschlagen in Auskunftsverzeichnissen ...).

Jedoch bietet das Internet auch Vorteile gegenüber der realen Welt. Der Benutzer kann auf einfache Weise Teile seiner Identität verschleiern, ohne dass dies sofort überprüfbar wäre. Zum Beispiel können Informationen wie zum Beispiel der Name, die Adresse oder das Alter in Formularen beliebig und der Situation entsprechend gewählt werden – für die Bestellung eines E-Mail-Newsletter wird die reale Postanschrift nicht benötigt und kann durch eine Fantasieadresse ersetzt werden.

Das Aufbauen von Pseudonymen ist inzwischen ein natürlicher Vorgang im Internet. So wird in Chaträumen selten der reale Name verwendet, und auch das Verhalten dort entspricht meist nicht dem Verhalten, das die Person im realen Leben an den Tag legt.

"On the Internet, nobody knows you're a dog."

2.2. Bestehende Projekte

2.2.1. Liberty Alliance

Die Liberty Alliance wurde im September 2001 mit dem Ziel, einen offenen Standard für eine einheitliche Netzwerkidentität zu erstellen, gegründet. Eine einheitliche Netzwerkidentität ist ein Zusammenschluss aller Identitätsinformationen, die ein Benutzer des Internets benötigt, um gewisse Dienste nutzen zu können. Diese Netzwerkidentität soll es dem Benutzer erleichtern, die einzelnen Dienste zu benutzen, indem er die Informationen nur einmal eingeben zu braucht und den Diensten Zugriffsrechte für die benötigten Informationen gewährt. Dabei sind die Themen Sicherheit und Privatsphäre ein großes Anliegen des Zusammenschlusses von ca. 160 Firmen.

Bei der Entwicklung der Standards wird vor allem auf die Interoperabilität und die Annahme der Spezifikationen und Richtlinien durch Firmen geachtet. Ermöglicht werden soll außerdem ein einfach zu benutzendes Single-Sign-On-System, das es dem Benutzer erlaubt, sich einmalig bei einem Dienst anzumelden und dann ohne erneutes Anmelden andere Dienste, die die Spezifikation der Liberty Alliance umgesetzt haben, nutzen zu können.

2.2.2. Microsoft Passport

Microsoft bietet mit Passport einen zentralisierten Identitäts-Management-Dienst an. Der Dienst ermöglicht Single-Sign-On bei den ca. 80 teilnehmenden Partnern von Microsoft. Wenige Identitätsinformationen können hinterlegt werden und an die Partner beim Anmelden weitergegeben werden.

2.2.3. SUN Network Identity

Für die Network Identity von SUN wurden die Spezifikationen der Liberty Alliance implementiert. Ziel der Network Identity ist es, das Verhältnis zwischen einer Firma und seinen Kunden zu intensivieren.

2.3. Anforderungen an das Identitätsmanagement

Aus Sicht des Empfehlungsdienstes sind folgende Anforderungen an das Identitätsmanagement zu stellen:

2.3.1. Identifizierbarkeit

Der Empfehlende muss dauerhaft eindeutig identifizierbar sein. Das identifizierende Merkmal darf sich für einen Benutzer im Laufe der Zeit nicht ändern und darf nicht von zwei Benutzern benutzt werden. Ein einmal vergebenes identifizierendes Merkmal soll nicht wieder vergeben werden, selbst wenn es von dem Benutzer, der dieses innehat, nicht mehr verwendet wird. An das identifizierende Merkmal sind unterschiedliche Daten, wie zum Beispiel Vertrauenswerte, gekoppelt, die bei einer Wiedervergabe des identifizierenden Merkmals negative Auswirkungen auf den neuen Benutzer haben könnten.

2.3.2. Pseudonymität

Informationen über die eigene Identität sind in der heutigen hochvernetzten Gesellschaft ein wichtiges Gut. Vorlieben und Abneigungen von Personen werden aufgezeichnet, zum Beispiel mit Hilfe elektronischer Zahlungseinrichtungen. Es ist denkbar, dass solche Informationen verwendet werden, um Personen gezielt zu bewerben und in ihrer unabhängigen Auswahl zu beeinflussen. Werden diese Informationen zwischen den aufzeichnenden Institutionen zusätzlich ausgetauscht, kann ein noch detailreicheres Bild der Personen erstellt werden, wodurch die Einflüsse auf diese noch genauer abgestimmt werden können. Das bewusst zugespitzte Szenario endet in einer verteilten Datenbank, in der alle Informationen über die Personen gespeichert und abrufbar sind („gläserner Mensch“). Durch einen Empfehlungsdienst werden diese Vorlieben und Abneigungen abgebildet und können so die Grundlage für ein gezieltes Marketing sein.

Um ein allgemein akzeptiertes und dadurch häufig benutztes System zu erstellen, dürfen nur dort Informationen über die wahre Identität aufgezeichnet werden, wo

es sich nicht vermeiden lässt. In allen anderen Fällen sollen Pseudonyme verwendet werden.

Der vollständige Verzicht auf ein identifizierendes Merkmal stellt keine Alternative für den UniTEC-Empfehlungsdienst dar, da die Modellierung von Vertrauen nur vollzogen werden kann, wenn das Maß des Vertrauens an eine „Person“ gebunden werden kann.

2.3.3. Pseudonymität bewahren

Aus den im obigen Abschnitt aufgezählten Gründen werden Pseudonyme zur Identifikation der Benutzer verwendet. Die aufgebaute Pseudonymität soll erhalten bleiben, d. h. es soll zum einen Angreifern schwer gemacht werden, die Identität eines Pseudonyms aufdecken zu können, und zum anderen soll der Benutzer vor Aktionen gewarnt werden, die seine Pseudonymität gefährden könnten.

2.3.4. Pseudonymität absichtlich aufdecken

Vertrauen hat einen engen Zusammenhang zur Identität einer Person. Ist eine Person durch den Namen bekannt, so wird ihr mehr Vertrauen entgegengebracht als einer Person, die einem unbekannt ist. Dies gilt vor allem, wenn damit monetäre Werte zusammenhängen wie beim E-Commerce. Deshalb soll eine Möglichkeit vorgesehen werden, die Identität eines Pseudonyms gegenüber einem anderen aufzudecken. Dies hat so zu geschehen, dass gezielt einzelnen Pseudonymen gegenüber die Identität offengelegt werden kann und dass Dritte diesen Vorgang nicht derart verfolgen können, dass auch sie die Verknüpfung der Identität mit dem Pseudonym herstellen können.

2.4. Pseudonymmanagement

Aus den oben angeführten Anforderungen ergibt sich für den UniTEC-Empfehlungsdienst eine andere Form des Identitätsmanagements. Es wird nicht die Identität eines

Benutzers verwaltet und überwacht, wer welche Informationen bekommen hat, sondern es wird ein Pseudonymmanagement benötigt. Dieses Pseudonymmanagement übenimmt die Verwaltung der Pseudonyme und muss dafür sorgen, dass die Pseudonymität des Benutzers erhalten bleibt.

3. Pseudonyme

Pseudonyme sorgen dafür, dass die wahre Identität des Benutzers verborgen bleibt. Da jeder Benutzer mehrere Pseudonyme haben kann, kann er in verschiedenen Rollen auftreten. In den folgenden Abschnitten wird erläutert, welche Arten von Pseudonymen im UniTEC-Empfehlungsdienst existieren und welche Informationen mit einem Pseudonym verknüpft sind.

3.1. Pseudonymarten

Innerhalb des Empfehlungsdienstes existieren zwei Arten von Pseudonymen. Zum einen diejenigen Pseudonyme, die dem Benutzer zugeordnet sind, die er also selbst erstellt hat, zum anderen diejenigen Pseudonyme, die ihm nicht zugeordnet sind. An diese Pseudonyme kann der Benutzer seine Anfragen nach Empfehlungen senden.

Beide Pseudonymarten haben gemeinsame Attribute. Jedem Pseudonym ist ein öffentlicher Schlüssel zugeordnet, der benutzt werden kann, um mit diesem Pseudonym zu kommunizieren. Dieser Schlüssel hat eine Länge von 1024 bit und wird nach dem RSA-Verfahren erstellt.

Des Weiteren ist allen Pseudonymen eine global eindeutige Identifikationsnummer (ID) zugewiesen. Deren Eigenschaften und der Erzeugungsalgorithmus werden in Kapitel 4 beschrieben.

Als dritte Komponente besitzt jedes Pseudonym einen Namen. Dieser dient dem Benutzer dazu, leichter mit den Pseudonymen umgehen zu können. Der Name ist nur lokal eindeutig. Für eigene Pseudonyme wird dieser vom Benutzer bestimmt, für frem-

de Pseudonyme wird dieser aus einer Liste von möglichen Namen zufällig vergeben, falls der ID des fremden Pseudonyms nicht schon früher ein Name zugeordnet wurde.

Die letzte gemeinsame Komponente ist die Zugehörigkeit zu einer Gruppe. Die Gruppenzugehörigkeit wird zur Zeit nur bei der Zugriffskontrolle auf Empfehlungen durch fremde Pseudonyme verwendet. Dort kann bestimmt werden, welchen fremden Pseudonymen mit einer Empfehlung geantwortet werden darf oder nicht. Für eigene Pseudonyme hat die Gruppenzugehörigkeit zur Zeit keine Bedeutung. Sie kann dazu benutzt werden, Pseudonyme logisch zu ordnen.

3.1.1. Zusätzliche Komponenten der eigenen Pseudonyme

Eigene Pseudoyme haben zusätzlich zu den oben aufgeführten noch zwei weitere Komponenten. Zum einen den zum öffentlichen Schlüssel passenden privaten Schlüssel, um die empfangenen, mit dem öffentlichen Schlüssel verschlüsselten Nachrichten zu entschlüsseln. Zum anderen ein Attribut, das angibt, ob dieses Pseudonym zur Zeit aktiv ist. Ist ein Pseudonym aktiv, so werden Anfragen an das Pseudonym beantwortet, und es können Nachrichten für dieses Pseudonym versandt werden. Des Weiteren werden die für die Kommunikation notwendigen Informationen innerhalb des Peer-to-Peer-Netzwerkes aktuell gehalten. Ist ein Pseudonym nicht aktiv, so ist es nicht erreichbar und kann auch keine Nachrichten versenden.

3.1.2. Zusätzliche Komponenten fremder Pseudonyme

Informationen über fremde Pseudonyme werden gespeichert, sobald mit diesen Kontakt besteht. Neue Kontakte kommen meist aufgrund von Antworten auf weitergeleitete Anfragen zustande. Fremde Pseudonyme haben als einzige weitere Komponente einen Zähler, der angibt, wie gut oder schlecht die Erreichbarkeit eines Pseudonyms ist. Dieser wird bei einem erfolgreichen Kontakt erhöht und bei einem nicht erfolgreichen Kontakt verringert.

3.2. Pseudonymmanagement

Die eigenen und die fremden Pseudonyme werden jeweils nach eigenen Modellen gespeichert. Diese Modelle geben eine einheitliche Grundfunktionalität vor:

- Hinzufügen eines Pseudonyms der entsprechenden Art
- Entfernen eines Pseudonyms
- Erzeugen einer Liste aller gespeicherten Pseudonyme
- Herausgabe eines Pseudonyms mit einer zu übergebenden ID
- Bestimmung der ID eines Pseudonyms mit einem zu übergebenden Namen
- Erzeugen einer Liste aller Gruppen
- Erzeugen einer Liste aller Namen

Es existieren zusätzlich Änderungsmöglichkeiten für die modifizierbaren Attribute der gespeicherten Pseudonyme. Im Falle der eigenen Pseudonyme sind dies der Name, die Gruppenzugehörigkeit und der Aktiv-Status. Im Falle der fremden Pseudonyme sind dies die Gruppenzughörigkeit und die Modifikation des Erreichbarkeitszählers.

Für fremde Pseudonyme besteht darüber hinaus die Möglichkeit, den zufälligen Anteil der Nachbarschaft zu erstellen. Der zufällige Anteil der Nachbarschaft soll dafür sorgen, dass auch Empfehlende mit einem geringen Vertrauenswert oder mit denen bisher kaum Kontakt bestand angefragt werden. Diese Empfehlenden können hochwertige Empfehlungen besitzen, auf die man aufgrund der restlichen Nachbarschaft keinen Zugriff gehabt hätte. Hierzu werden aus der Liste aller Pseudonyme diejenigen herausgefiltert, die schon aufgrund des Vertrauensmodells oder des Expertisenmodells als Nachbarn ausgesucht wurden. Die verbleibende Liste wird, wenn nötig, auf die übergebene Länge gekürzt.

4. Identifizierbarkeit eines Pseudonyms

Damit der Vertrauenswert mit einem Pseudonym verknüpft werden kann, muss dieses eindeutig identifizierbar sein. Dies wird bewerkstelligt, indem man jedem Pseudonym ein Merkmal zuordnet, das einzigartig für jedes Pseudonym ist – die ID. Es gibt mehrere Möglichkeiten, wie die ID generiert werden kann. Im Folgenden werden einige vorgestellt und die geeignetste Alternative ausgewählt.

Eine einfache Möglichkeit, die Einzigartigkeit eines Pseudonyms sicherzustellen, besteht darin, die Vergabe der IDs eine zentrale Stelle durchführen zu lassen. Diese zentrale Stelle verfügt über eine gewisse Menge von möglichen IDs und führt eine Liste, welche IDs verwendet werden. Möchte ein Benutzer ein neues Pseudonym erstellen, so wird er bei der zentralen Vergabestelle nach einer neuen ID anfragen und von dieser eine zugeteilt bekommen.

Eine weitere Möglichkeit ist, die ID lokal zu generieren und anschließend zu überprüfen, ob diese schon von einem anderen Pseudonym benutzt wird. Hierzu gibt es mehrere Alternativen:

- Zufällige Wahl einer ID aus einem vorher festgelegten ID-Raum (Menge aller erlaubten IDs).
- Die ID ist der öffentliche Schlüssel des Pseudonyms.
- Eine ID wird mittels eines Algorithmus aus bestimmten Merkmalen des Pseudonyms generiert.

Eine Gegenüberstellung der verschiedenen Möglichkeiten kann der Tabelle 4.1 entnom-

men werden.

Die zentrale Vergabe kommt aufgrund der Anforderung, dass zentrale Dienste vermieden werden sollen, nicht in Frage. Die zufällige Wahl einer ID aus einem ID-Raum ist ein einfaches Verfahren. Da aber die Kollisionswahrscheinlichkeit vom verwendeten Algorithmus zur Generierung einer Zufallszahl abhängt, wird auch dieser Ansatz vermieden. Der öffentliche Schlüssel des Pseudonyms weist eine große Länge auf und hat damit den Vorteil, dass die Wahrscheinlichkeit einer Kollision sehr gering ist. Jedoch ist die Länge des öffentlichen Schlüssels zu hoch. Die ID wird bei jeder Kommunikation, zum Teil mehrfach, in einer Nachricht verwendet (z. B. ist die ID in einer Empfehlung enthalten). Dies bedeutet eine große Nachrichtenlänge und somit einen erhöhten Zeitbedarf bei der Kommunikation.

Zur Generierung wurde ein Algorithmus gewählt, der den Vorteil der öffentlichen Schlüssel – die Gleichverteiltheit innerhalb des Schlüsselraums – ebenfalls aufweist. Die ID wird mittels des Secure-Hash-Verfahrens SHA-1 (siehe [NIST, 1995]) generiert. Dieses Verfahren begrenzt die ID auf einen Länge von 160 bit. Trotz der geringeren Länge weist die ID sehr geringe Kollisionswahrscheinlichkeit auf (vgl. Abbildung 4.1). Als Eingabe dient der öffentliche Schlüssel des Pseudonyms. Dieser muss bekannt sein, wenn damit kommuniziert werden soll. Die Verbindung ID und öffentlicher Schlüssel ist mittels des Secure-Hash-Verfahrens überprüfbar, indem der Hash des öffentlichen Schlüssel mit der ID verglichen wird. Ein weiterer Vorteil der Secure-Hash-Funktion ist, dass es schwierig ist, aus einer gegebenen ID die Eingabe, also den öffentlichen Schlüssel, zu berechnen. Möchte sich ein Angreifer als ein Pseudonym ausgeben, das er nicht selbst generiert hat, so benötigt er, um kommunizieren zu können, ein Schlüsselpaar. Will er sich als ein bestimmtes Pseudonym ausgeben, benötigt er einen öffentlichen Schlüssel, der bei Anwendung des Secure-Hash-Verfahrens die ID des Pseudonyms ergibt, für das er sich ausgeben möchte. Sollte ihm dies gelingen, so ist die Generierung des passenden privaten Schlüssels aus dem öffentlichen Schlüssel immer noch sehr schwierig.

Aufgrund der geringen Wahrscheinlichkeit einer Kollision wird auf die Überprüfung einer solchen verzichtet.

	Vorteile	**Nachteile**
Zentral	• einfach • keine Kollisionen möglich • beliebige ID-Länge	• single point of failure • sollte in einem Peer-to-Peer-Netzwerk vermieden werden • zusätzliche Kommunikation notwendig
Zufällig	• einfach • beliebige ID-Länge	• Kollisionen mit einer schon bestehenden ID möglich
Öffentlicher Schlüssel	• großer ID-Raum (je nach Schlüssellänge, z. B. 2^{1024}) • öffentliche Schlüssel sind gleichverteilt im Schlüsselraum • selbst bei einer hohen Anzahl von Pseudonymen ist die Kollisionswahrscheinlichkeit sehr gering • ID muss nicht extra generiert werden	• ID ist sehr lang (z. B. 1024 bit) • Kollisionen möglich
Algorithmus	• ID-Länge frei definierbar • Merkmale des Pseudonyms können evtl. aufgrund der ID auf Richtigkeit überprüft werden	• zusätzlicher Aufwand • Kollisionen möglich

Tabelle 4.1.: Übersicht über die Vor- und Nachteile der angesprochenen Verfahren zur Generierung einer ID für die Pseudonyme

Die Wahrscheinlichkeit, dass eine ID schon von einer Person benutzt wird, wenn schon im Vorfeld x IDs vergeben sind, beträgt:

$$P(Kollision) = \frac{x}{2^{160}}$$

Angenommen, jeder Bewohner der Erde (ca. 6 Mrd.) erzeugt 5.000 Pseudonyme, so ergibt sich die Wahrscheinlichkeit einer Kollision zu

$$P(Kollision) = \frac{30 Billionen}{2^{160}} \approx 2,05 * 10^{-35}$$

Abbildung 4.1.: Dargestellt ist die Wahrscheinlichkeit einer Kollision bei der Generierung einer ID der Länge 160 bit mit dem Secure-Hash-Verfahren SHA-1 aus dem öffentlichen Schlüssel des Pseudonyms.

5. Gefahren für die Pseudonymität

In vielen Softwareanwendungen werden Pseudonyme verwendet. Als Beispiele können Chaträume und Diskussionsforen, Auktionshäuser und Instant-Messaging-Systeme im Internet genannt werden. Bei einigen dieser Anwendungen ist es harmlos, wenn die wahre Identität hinter einem Pseudonym herausgefunden wird. Zum Teil ist die Aufdeckung der Pseudonymität sogar erwünscht. So ist es zum Beispiel beim Austauschen von Nachrichten zwischen Freunden mittels eines Instant-Messaging-Systems durchaus beabsichtigt zu wissen, wer hinter welchem Pseudonym steckt. Bei anderen Anwendungen ist es unerwünscht, dass die Verbindung zwischen Identität und Pseudonym bekannt ist. Dies ist auch beim UniTEC-Empfehlungsdienst der Fall. In Kapitel 2 wurden die Gründe aufgezählt, warum für den Empfehlungsdienst Pseudonyme verwendet werden. Damit die Pseudonymität gegenüber Dritten erhalten bleibt, muss untersucht werden, an welchen Stellen Verbindungen zwischen Pseudonym und Identität auf welche Art hergestellt werden können und wie die Aufdeckung der Identität verhindert werden kann.

Die Aufhebung der Pseudonymität ist meist mit erheblichem Aufwand verbunden, da im Normalfall, wenn der Benutzer auf seine Pseudonymität achtet, immer nur sehr geringe Teile der Identität preisgegeben werden. Dass Teile der Identität preisgegeben werden, kann nicht vollständig verhindert werden. Zum Beispiel verrät eine Empfehlung über ein gewisses Gebiet, das der Empfehlende Wissen in dieser Kategorie besitzt.

5.1. Kommunikationssubsystem

Während der Kommunikation kann ein Dritter eine Nachricht abfangen und deren Inhalt lesen. Eine solche Nachricht enthält die ID des Senderpseudonyms und die des Empfängerpseudonyms und im Normalfall zusätzlich die eindeutigen Identifikationsnummeren des Senders und des Empfängers, die zur Kommunikation benötigt werden (zum Beispiel die IP-Adressen bei Verwendung des Internetprotokolls). Diese Identifikationsnummern können dann mit den IDs der Pseudonyme in Verbindung gebracht werden.

Über die Identifikationsnummer der Kommunikation können evtl. weitere Informationen über die an der Kommunikation beteiligten Pseudonyme herausgefunden werden. So ist es möglich, über einen unter der IP-Adresse arbeitenden Webserver Informationen wie etwa den Arbeitgeber oder den Namen einer Person zu ermitteln. Ebenfalls denkbar wäre eine illegale Zusammenarbeit des Angreifers mit Internet Service Providern (ISP), die zu Zwecken der Abrechnung die Zuordnung der Kommunikations-ID zu der Identität speichern.

Weiterhin wäre es möglich, mehrere Pseudonyme eines Benutzers aufgrund der gemeinsamen Kommunikations-ID miteinander in Verbindung zu bringen und somit ein umfassenderes Profil zu erstellen.

Aus diesen Gründen wird die Kommunikation mittels eines asymetrischen Verschlüsselungsverfahren verschlüsselt. Die öffentlichen Schlüssel werden innerhalb des Peer-to-Peer-Netzwerkes in einem Distributed-Hash-Table (DHT) gespeichert (nähere Informationen in [Terdic, 2003]).

Wird nur die Verschlüsselung angewendet, so müssen Sender und Empfänger immer noch die Zuordnung des Pseudonyms zur Kommunikations-ID kennen, um mit dem Gegenüber kommunizieren zu können. Um diesen Umstand zu umgehen, wurde eine Anonymisierungsschicht für das Peer-to-Peer-Netzwerk entwickelt. Die Nachricht gelangt nicht direkt von Sender zu Empfänger, sondern nimmt einen Umweg über eine gewisse Anzahl von so gennanten Mixen. Diese Mixe dienen dazu, den Verlauf einer Nachricht durch das Kommunikationsnetzwerk zu verschleiern. Ein Mix kennt nur den Sender, von dem er die Nachricht bekommen hat, und den nächsten Empfänger, an

den die Nachricht weitergeleitet werden muss. Er kann aber nicht herausfinden, ob der Sender, von dem er die Nachricht bekommen hat, auch wirklich der ursprüngliche Sender der Nachricht ist. Gleiches gilt für den Empfänger der Nachricht. Die Abfolge der Mixe wird zum einen Teil vom Sender und zum anderen Teil vom Empfänger bestimmt. Jedes Pseudonym, das Nachrichten empfangen möchte, speichert seinen Teil der Route im DHT. Möchte nun ein Pseudonym mit einem anderen kommunizieren, so erfragt das Senderpseudonym die Teilroute des Empfängers, fügt seine eigene Teilroute hinzu und sendet die Nachricht ab. Eine ausführliche Dokumentation des anonymen Peer-to-Peer-Netzwerkes kann [Terdic, 2003] entnommen werden.

5.2. Profilerstellung

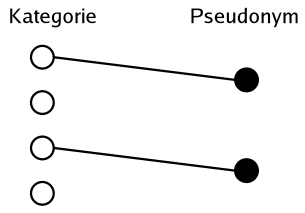
Die Preisgabe eines kleinen Teils der Identität stellt als Einzelereignis betrachtet keine große Gefahr dar. Meist kann nicht direkt auf die Identität hinter einem Pseudonym geschlossen werden. Ein Angreifer, der die Identität eines Pseudonyms aufdecken möchte, kann jedoch viele Anfragen in unterschiedlichen Kategorien an das Pseudonym senden und aufgrund der Antworten ein Profil erstellen, das das Wissen der Person hinter dem Pseudonym mehr oder weniger genau abbildet. Aufgrund dieses Profils kann der Kreis der infrage kommenden Personen eingeengt oder evtl. sogar direkt auf eine Person geschlossen werden.

Da der Benutzer mehrere Pseudonyme verwenden kann, kann er das Gefahrenpotential dieses Angriffes verringern. Auf welche Arten diese Aufteilung der Pseudonyme auf Kategorien und andersherum geschehen kann, wird im Folgenden erläutert.

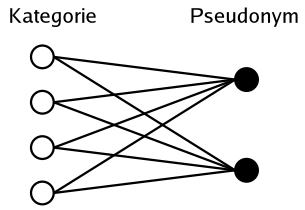
Pseudonyme werden verwendet, um eine Empfehlung zu erstellen und Empfehlungen bei anderen Pseudonymen zu erfragen. Bei der Erstellung der Empfehlungen existieren zwei Alternativen, wieviele Pseudonyme für eine Kategorie Empfehlungen erstellen dürfen. Entweder ist es nur genau einem Pseudonym erlaubt, oder es dürfen beliebig viele Pseudonyme eine Empfehlung in einer gewissen Kategorie erstellen. Bei der zweiten Alternative besteht jedoch die Gefahr, dass aufgrund ähnlicher Empfehlungen Verbindungen zwischen den Pseudonymen hergestellt werden können. Deshalb ist es ratsamer, die Anzahl der Pseudonyme, die in einer gewissen Kategorie Empfehlungen erstellen können, auf eins zu begrenzen.

Die Auswahl der Kategorien, in der ein Pseudonym Empfehlungen erstellen kann, kann ebenfalls auf unterschiedliche Weise gestaltet werden:

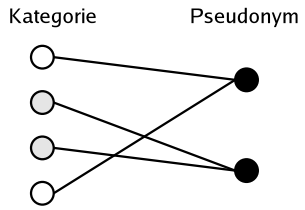
- Ein Pseudonym ist nur für eine Kategorie zuständig.



- Ein Pseudonym kann in allen Kategorien Empfehlungen erstellen.



- Ein Pseudonym kann in semantisch eng verwandten Kategorien Empfehlungen erstellen.



Die Alternative, dass ein Pseudonym nur in einer Kategorie Empfehlungen erstellen kann, ist ungeeignet. Existiert eine große Zahl an Kategorien, so muss auch eine große Zahl an Pseudonymen vorgehalten werden, um in jeder Kategorie eine Empfehlung erstellen zu können. Dies kann zu hohem Speicherplatzverbrauch führen. Ebenfalls fehlt die Möglichkeit, eng verwandte Kategorien durch ein Pseudonym abzudecken und bei Anfragen evtl. aus eng verwandten Kategorien antworten zu können.

Kann ein Pseudonym in allen Kategorien Empfehlungen erstellen, so wird das gesamte Wissen des Benutzers im Laufe der Zeit auf einem Pseudonym abgebildet. Dadurch wird es evtl. möglich, die Identität des Pseudonyms aufzudecken. Verwendet der Benut-

zer mehrere Pseudonyme, können diese durch Empfehlungen in ähnlichen Kategorien miteinander in Verbindung gebracht werden.

Wenn sich die Kategorien, in denen ein Pseudonym Empfehlungen erstellt, auf eng verwandte Kategorien beschränkt, so wird immer nur ein kleiner Teil des Wissens des Benutzers abgebildet. Diese Einteilung muss jedoch vom Benutzer selbst vorgenommen werden. Und selbst dieser wird Schwierigkeiten haben bei der Abgrenzung der Kategorien. Zum Beispiel: Es existieren die drei Kategorien Fenster, Glas und Wein. Es besteht offensichtlich jeweils eine enge Beziehung zwischen Fenster und Glas und Wein und Glas. Jedoch besteht keine Verbindung zwischen Fenster und Wein. An welcher Stelle findet hier die Abgrenzung durch Pseudonyme statt (siehe Abbildung 5.1)?

Abbildung 5.1.: Zwei Gruppierungen der drei Kategorien Fenster, Glas und Wein.

Für die Festlegung, welches Pseudonym in einer Kategorie Anfragen stellen darf, existieren verschiedene Ansätze:

- Es existiert ein spezielles Pseudonym, das nur verwendet wird, um Anfragen zu stellen.
- Ein Pseudonym darf in jeder Kategorie Anfragen erstellen.
- Ein Pseudonym darf nur in den Kategorien Anfragen erstellen, in denen es auch Empfehlungen erstellen darf.
- Ein Pseudonym darf unabhängig von der Erstellung einer Empfehlung in nur einer Teilmenge der Kategorien Anfragen erstellen.

Von der Möglichkeit, nur ein spezielles Pseudonym für die Anfragen zu verwenden, ist abzusehen, da dieses Pseudonym nach einer gewissen Zeit die Interessensgebiete der Identität sehr genau abbildet und es dadurch erleichtert wird, auf die Iden-

tität zu schließen. Bei der letzten Alternative besteht das Problem, dass die Größe der Teilmenge eine wichtige Rolle spielt. Ist diese nicht begrenzt, so ergibt sich die Möglichkeit, dass ein Pseudonym in jeder Kategorie Anfragen stellen darf (dann kann gleich diese Alternative gewählt werden). Ist die Größe jedoch beschränkt, so ergibt sich eine Mindestanzahl an Pseudonymen, um in allen Kategorien Anfragen stellen zu können ($\lceil \frac{AnzahlderKategorien}{MaximaleAnzahlderKategorienproPseudonym} \rceil$). Diese Mindestanzahl kann bei einer ungeschickten Wahl dazu führen, dass eine große Anzahl von Pseudonymen vorgehalten werden muss. Bei der Alternative, dass jedes Pseudonym in jeder Kategorie Anfragen stellen darf, besteht die Gefahr, dass sich im Laufe der Zeit bei allen Pseudonymen die Interessensgebiete des Benutzers abbilden. Dieses Verhalten kann ausgenutzt werden, um bei mehreren Pseudonymen festzustellen, dass sie zu einem Benutzer gehören müssen. Bei der Wahl, dass ein Pseudonym nur in den Kategorien Anfragen stellen darf, in denen es auch Empfehlungen erstellen kann, besteht das Problem, dass für eine Kategorie, für die kein Pseudonym definiert wurde, keine Anfragen erstellt werden können. Das Problem, dass für gewisse Kategorien kein Pseudonym definiert ist, das Anfragen stellen kann, kann folgendermaßen umgangen werden: Ist kein Pseudonym definiert, so wird zufällig aus der Liste der verfügbaren Pseudonyme eines ausgesucht. Dieses ist solange für Anfragen in der Kategorie zuständig, bis der Benutzer ein Pseudonym für das Erstellen von Empfehlungen in dieser Kategorie bestimmt. Ein weiteres Argument dafür, dass für eine Anfrage in einer Kategorie immer das gleiche Pseudonym verwendet wird, ist: In der Anfrage ist das Vertrauen des Senders bzw. des Weiterleitenden in den nächsten Empfänger enthalten (Trust-Chain). Wird nun eine Anfrage in der gleichen Kategorie von zwei unterschiedlichen Pseudonymen an das selbe Pseudonym gesendet und hat sich der Vertrauenswert in den Empfänger noch nicht geändert, so können alle Pseudonyme auf dem weiteren Weg der Anfrage eine Verbindung zwischen den beiden Senderpseudonymen aufgrund des Vertrauenswertes herstellen (vgl. Abbildung 5.2). Auf diese und ähnliche Problemen mit der Anfrage wird in Kapitel 7 eingegangen.

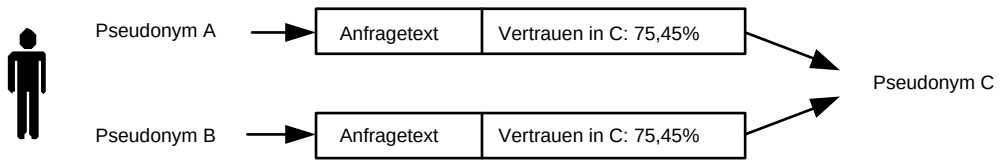


Abbildung 5.2.: Da sich der Vertrauenswert innerhalb des Vertrauensmodell des Senders noch nicht geändert hat, kann Pseudonym C die beiden Pseudonyme A und B miteinander in Verbindung bringen.

Nochmals zusammengefasst (vgl. auch Abbildung 5.3):

- Pro Kategorie darf nur ein Pseudonym Empfehlungen erstellen.
- Ein Pseudonym sollte nur in eng verwandten Kategorien Empfehlungen erstellen.
- Bei Anfragen in einer Kategorie wird das Pseudonym benutzt, mit dem auch Empfehlungen erstellt werden. Sollte keine Zuordnung definiert sein, so wird zufällig ein Pseudonym aus der Menge aller Pseudonyme ausgewählt und solange verwendet, bis eine Zuordnung stattfindet.

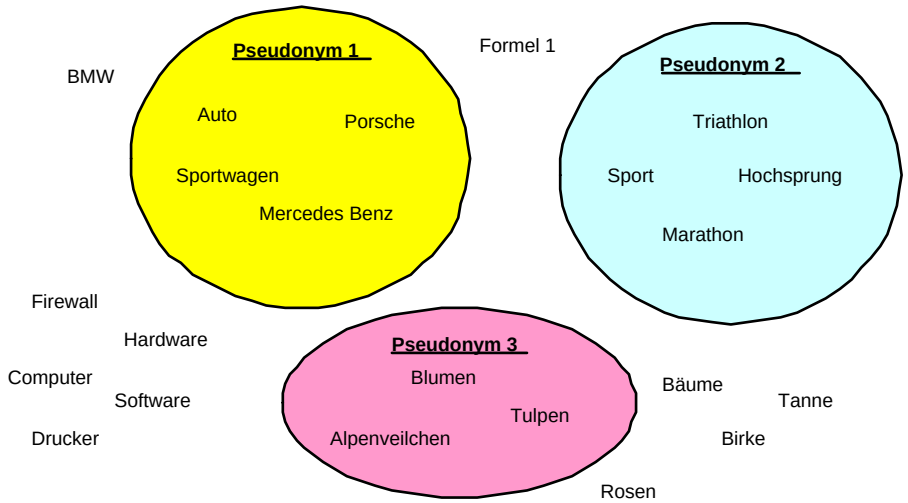


Abbildung 5.3.: Empfohlene Zuordnung von Pseudonymen zu Kategorien

5.3. Empfehlungen

Empfehlungen enthalten persönliche Informationen des Empfehlenden. Dieser kann Texte schreiben, binäre Daten hinzufügen, Zahlen oder Zahlenbereiche definieren und einfache boolsche Werte zu einer Empfehlung hinzufügen. Eine Beispielempfehlung zeigt Abbildung 5.4.

Das größte Gefahrenpotential geht von den Texten und den binären Elementen aus. Diese werden in den folgenden Abschnitten besprochen.

Empfehlung

Kategorie: PDA

Produkt: iPAQ h1935

Firma: Hewlett-Packard

Preis (Dollar): 199

Beschreibung: Der HP iPAQ arbeitet mit einem Samsung Prozesser (203 MHz) und hat eine Bildschirmauflösung von ….

Bild:

Empfehlenswert: ja

Verarbeitungsqualität: sehr gut

Preis/Leistungsverhältnis: gut

Eigene Gewissheit: 80 %

Abbildung 5.4.: Eine Beispielempfehlung

5.3.1. Texte

Die Texte kann der Benutzer frei verfassen und dabei auch Informationen einfügen, die die Pseudonymität des Benutzers verletzen kann.

Informationen der Identität können direkt in den Text einfließen. Beispielsweise kann der Empfehlende seinen eigenen Namen oder seinen Beruf in die Empfehlung schreiben. Um überprüfen zu können, ob solche Informationen in einer Empfehlung vorhanden sind, müssen diese Informationen vorher dem Empfehlungsdienst mitgeteilt werden. Dies wird aus verschiedenen Gründen nicht vorgenommen. Zum einen sinkt durch diesen Vorgang das Vertrauen des Benutzers in den Empfehlungsdienst. Er kann nicht ausschließen, dass der Empfehlungsdienst diese Informationen ohne Erlaubnis an andere Teilnehmer oder eine zentrale Stelle weitersendet und evtl. daraus Profit gewinnt (vgl. Adressenverkauf von großen Firmen). Zum anderen müsste eine Vielzahl von Informationen aufgenommen werden, um einen sicheren Schutz bieten zu können. Angefangen beim Namen des Benutzers, über seine Adresse und seinen Arbeitsplatz, bis zu den Namen seiner Verwandten und Freunde.

In eine ähnliche Richtung gehen typische Empfehlungen, die eine Person regelmäßig gibt und auch innerhalb des Empfehlungsdienstes benutzt. Diese Möglichkeit der Aufdeckung eines Pseudonyms kann nur von Personen ausgenutzt werden, die persönlichen Kontakt zu der Person haben und somit die typischen Empfehlungen selbst kennen.

Ebenfalls kann eine Empfehlung für eine Person typische Rechtschreib- oder Grammatikfehler enthalten, die entweder einem Angreifer in der realen Welt bekannt sind oder die in Empfehlungen unterschiedlicher Pseudonyme vorkommen und so eine Verbindung zwischen diesen beiden bildet.

Ähnliche Texte sowie Sätze, die in Empfehlungen unterschiedlicher Pseudonyme vorkommen, können eine Verbindung zwischen verschiedenen Pseudonymen eines Benutzers herstellen und damit insgesamt mehr Informationen über diesen Benutzer preisgeben. Die Ähnlichkeit zweier Texte kann mit entsprechenden Algorithmen ausgewertet werden.

Diese werden im Folgenden vorgestellt:

Hamming-Distanz Die Hamming-Distanz kann nur von zwei Texten gleicher Länge angegeben werden. Sie wird berechnet, indem buchstabenweise die beiden Texte verglichen werden und bei einem Unterschied die Hamming-Distanz erhöht wird.

```
Die Maus ist grau.
|  | |||||||||  |||
Das Haus ist blau.
```

Abbildung 5.5.: Die beiden Sätze haben jeweils 18 Zeichen. Die Hamming-Distanz beträgt fünf, d. h. man müsste fünf Buchstaben des einen Textes substituieren, um den anderen Text zu erhalten.

Levenshtein-Distanz Die Levenshtein-Distanz vergleicht zwei Texte beliebiger Länge. Die Distanz zwischen diesen beiden Texten ist die Anzahl von Zeichenmodifikationen (löschen, einfügen, substituieren), die durchgeführt werden muss, damit der eine Text in den anderen Text überführt werden kann.

Der Hase läuft schnell.

Der Rasen wächst schnell.

Abbildung 5.6.: Die Levehnstein-Distanz der beiden Sätze beträgt sechs. Im unteren Satz müssen zwei Buchstaben entfernt und vier Buchstaben substituiert werden, um diesen in den oberen Satz zu überführen.

Longest Common Subsequence Dieser Algorithmus findet den längsten gemeinsamen Subtext innerhalb zweier Texte. Ist dieser im Verhältnis zum kürzeren der beiden Texte sehr lang, so kann dies ein Zeichen dafür sein, dass die beiden untersuchten Texte sehr ähnlich sind.

Ein einfacher Algorithmus arbeitet wie folgt:

1. Die beiden Texte seien S und T und haben die Länge m und n.

2. Erstelle eine Matrix der Größe $(m + 1) \times (n + 1)$ und initialisiere die erste Spalte und die erste Zeile mit 0.

3. S(i) sei der i-te Buchstabe des Textes S. Iteriere über m und n. Falls S(i) = T(j), dann ist der Eintrag D(i,j) = D(i-1,j-1) + 1. Ansonsten ist D(i,j) = 0.

4. Die Länge des längsten gemeinsamen Subtextes ist die größte Zahl in der Matrix.

		b	a	a	d	c	a	a	a	a	b	a	a
	0	0	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	0	0	1	0	0	0	0	0	0	0
a	0	0	1	1	0	0	2	1	1	1	0	1	1
a	0	0	1	2	0	0	1	3	2	2	0	1	2
b	0	1	0	0	0	0	0	0	0	0	3	0	0
a	0	0	2	1	0	0	1	1	1	1	0	4	1
a	0	0	1	3	0	0	1	2	2	2	0	1	5
d	0	0	0	0	4	0	0	0	0	0	0	0	0

Abbildung 5.7.: Die Tabelle zeigt das Ergebnis der Berechnung nach dem oben aufgeführtem Algorithmus der Longest Common Subsequence. Verfolgt man die Zeile bzw. die Spalte bis zu den eingetragenen Zeichen, so trifft man auf das Ende des longest common substrings und kann diesen bis zu seinem Beginn verfolgen. Im Beispiel ist die Zeichenfolge aabaa der longest common substring mit der Länge 5.

Greedy-String-Tiling Greedy-String-Tiling (GST) ist der longest common subsequence ähnlich. Es wird sukzessive ein möglichst langer Textauszug, der in beiden Texten vorkommt, gesucht, wobei in einem vorherigen Schritt gefundene Textauszüge nicht mehr berücksichtigt werden. Dies geschieht solange, bis kein weiterer Textauszug gefunden wird, der die Forderung nach einer frei definierbaren Mindestlänge (minimum-match-length) erfüllen kann. Der kürzere der beiden Texte wird pattern genannt (P), der längere Text wird text (T) genannt. Gesucht werden maximale Textauszüge (maximal-match), die sowohl in P als auch in T vorkommen und noch nicht markiert wurden. Diese Textauszüge werden P_p und T_t genannt, wobei p bzw. t die Länge des Textauszuges angibt. Ein Textauszug wird markiert, sobald festgestellt wurde, dass dieser maximal ist. Der Algorithmus arbeitet folgendermaßen (vgl. [Wise, 1993] und das Beispiel in Abbildung 5.8):

```
Repeat
  maxmatch := minimum-match-length
  starting at first unmarked token of P, for each P_p do
    starting at the first unmarked token of T, for each T_t do
```

```
        j := 0
        while P_(p+j) = T_(t+j) AND unmarked(P_(p+j)) AND
                                        unmarked(T_(t+j)) do
          j := j + 1
        if j = maxmatch then add match(p, t, j) to list of matches
        else if j > maxmatch then
          clear list of matches
          add match(p, t, j) to list of matches
          maxmatch := j
    for each match(p, t, maxmatch) in list
      if not occluded then
        for j := 0 to maxmatch - 1 do
          mark_token(P_(p+j))
          mark_token(T_(t+j))
  until maxmatch = minimum-match-length
```

Für die Erkennung von Textähnlichkeiten ist der Greedy-String-Tiling-Algorithmus am besten geeignet, da dieser möglichst viele möglichst lange übereinstimmende Textteile sucht. Um den Algorithmus innerhalb des Empfehlungsdienstes anwenden zu können, wurde er folgendermaßen modifiziert:

- Der Algorithmus vergleicht nicht mehr einzelne Buchstaben, sondern ganze Wörter.
- Aus den Eingabetexten werden Stoppwörter (Wörter, die sehr häufig vorkommen, wie zum Beispiel die Artikel) herausgefiltert.
- Die Wörter werden so weit wie möglich auf ihre Lexeme zurückgeführt (aus Häuser wird Haus). Dieser Vorgang wird stemming genannt.
- Denkbar wäre auch, dass Wörter unterschiedlicher Bedeutung mit Hilfe eines Synonymlexikons so ausgetauscht werden, dass für eine gewisse Bedeutung nur noch ein Wort während der Ähnlichkeitsuntersuchung verwendet wird. Dadurch können auch Ähnlichkeiten zwischen Texten aufgedeckt werden, die nicht rein syntatktisch, sondern auch von der Bedeutung des Textes abhängen.

Es wurden Versuche vorgenommen, bei der ca. 20 Texte mit Hilfe des Greedy String

<u>Vorgaben:</u>
Text T = baadcaaaabaa
Pattern P = caabaad
minimum-match-length = 1

<u>1. Durchgang:</u>
Gefunden wird: match(1, 7, 5)

b	a	a	d	c	a	a	a	a	b	a	a
							x	x	x	x	x

c	a	a	b	a	a	d
	x	x	x	x	x	

<u>2. Durchgang:</u>
Gefunden wird: match(0, 4, 1) und match(6, 3, 1)

b	a	a	d	c	a	a	a	a	b	a	a
			x	x			x	x	x	x	x

c	a	a	b	a	a	d
x	x	x	x	x	x	x

Abbildung 5.8.: Eine Beispielberechunung der Distanz zwischen den beiden Texten „baadcaaaabaa“ und „caabaad“.

Tiling Algorithmus untersucht wurden (vgl. Kapitel B). Diese Texte waren zum Großteil englische Produktbeschreibungen des Online-Auktionshauses Ebay. Von diesen wurde angenommen, dass sie einer Empfehlung recht nahe kommen. Ein Ergebniss ist, dass sich eine Ähnlichkeit der Texte nicht ohne eine tiefere semantische Anlayse erkennen lässt. Hierbei ist jedoch davon auszugehen, dass ein menschlicher Angreifer dem Computer überlegen ist. Der Algorithmus eignet sich allerdings für das Auffinden von Sätzen bzw. Satzteilen, die unverändert von einer Empfehlung in eine andere Empfehlung kopiert wurden. Hierzu wird vorgeschlagen, die minimumMatchLength auf den Wert drei oder vier zu setzen.

5.3.2. Binäre Daten

Eine Verbindung zwischen zwei Pseudonymen kann hergestellt werden, wenn diese die gleichen binären Elemente benutzen, die der Allgemeinheit nicht zugänglich sind. Ein Foto eines Produktes, das ursprünglich auf der Webseite des Herstellers hinterlegt war, stellt keine Gefahr dar, da dieses von jeder Person benutzt werden kann. Das Video vom letzten Sommerurlaub in Italien, das zum einen in einer Empfehlung über das Hotel und zum anderen in einer Empfehlung über das Restaurant im gleichen Ort enthalten ist, ist ein verbindendes Element, wenn die Empfehlungen von unterschiedlichen Pseudonymen gegeben werden. Solche Vorkommnisse, bei denen exakt das gleiche binäre Element in Empfehlungen verschiedener Pseudonyme vorkommt, können leicht aufgedeckt werden.

Bilder können die gleiche Szene aus unterschiedlichen Blickwinkeln betrachten und somit eine Verbindung zwischen Empfehlungen verschiedener Pseudonyme möglich machen, obwohl die binären Elemente nicht exakt gleich sind. Ähnlich verhält es sich mit Videoszenen, die sowohl ebenfalls die Szene aus unterschiedlichen Blickwinkeln betrachten, als auch unterschiedlich nachbearbeitet werden können. Zum Beispiel kann das gleiche Video in der Empfehlung des ersten Pseudonyms mit Musikuntermalung und in der Empfehlung des anderen Pseudonyms ohne Musikuntermalung, aber dafür mit dem Orginalton, enthalten sein.

5.4. Externes Weitergeben von Informationen

Informationen über die Identität eines Pseudonyms können, nachdem sie aufgedeckt wurden (vgl. Kapitel 6), auf nicht überprüfbaren Wegen, weitergegeben werden, z. B. per Telefon oder E-Mail.

5.5. Warnung des Benutzers

Der Benutzer sollte bei entdeckten Elementen einer Empfehlung, die die Pseudonymität gefährden können, gewarnt bzw. die Erstellung der Empfehlung unterbunden werden. Dazu wurde ein dreistufiges Warnsystem erstellt: kein Eingriff notwendig, Warnung des Benutzers und Unterbinden der Erstellung.

Im Rahmen der Rechtschreib- und Grammatikprüfung sollte bei einer geringen Anzahl von Fehlern eine Warnung ausgegeben und die Möglichkeit zur Korrektur gegeben werden. Übersteigt die Zahl der Fehler eine vorher festgelegte Schwelle, so ist das Erstellen der Empfehlung solange zu unterbinden, bis die Fehlerzahl aufgrund der Korrektur durch den Benutzer mindestens unter den Grenzwert gesunken ist.

Im Rahmen der Suche nach gleichen Sätzen bzw. Satzteilen sollte ebenfalls bei einer geringen Zahl von Übereinstimmungen (vorgeschlagen wird maximal zwei Übereinstimmungen), eine Warnung und die Möglichkeit zur Korrektur gegeben werden. Übersteigt die Anzahl der Übereinstimmungen jedoch den Grenzwert, so ist eine Korrektur zu erzwingen.

Strenger ist bei der Verwendung von binären Elementen zu verfahren. Hier sollte bei Auftreten eines binären Elements in zwei Empfehlungen unterschiedlicher Pseudonyme auf jeden Fall auf eine Korrektur bestanden werden.

Die Gefahr beim Bestehen auf eine Korrektur der beiden zuletzt genannten Punkte besteht darin, das der Benutzer entweder den Satz minimal ändert, so dass eine Übereinstimmung durch den Algorithmus nicht mehr erkannt werden kann, der Mensch hingegen, diesen Zusammenhang immer noch einfach entdeckt. Ähnlich verhält es sich bei der Wahl der binären Elemente. Wird eine Korrektur vorgeschrieben, so kann der

Benutzer ein ähnliches Element auswählen, dessen Zusammengehörigkeit durch den Empfehlungsdienst nicht ermittelt werden kann, der Mensch jedoch diesen Zusammenhang einfach erkennen kann.

5.6. Zusammenfassung

Es können nur wenige Gefahren auf Seiten der Empfehlungen auf technischem Wege, d. h. durch die Software des Empfehlungsdienstes, verringert werden. Dies sind:

- Jede Kategorie wird nur von einem Pseudonym abgedeckt,
- Rechtschreibüberprüfung,
- Grammatiküberprüfung,
- Gleiche Sätze in Empfehlungen unterschiedlicher Pseudonyme,
- Gleiche binäre Elemente in Empfehlungen unterschiedlicher Pseudonyme.

Allgemein können folgende Empfehlungen ausgesprochen werden um die Gefahr der Aufdeckung der Identität eines Pseudonyms zu verringern:

- Jedes Pseudonym sollte eng verwandte Kategorien abdecken.
- Vermeidung von Informationen in einer Empfehlung, die die Identität offenlegen könnten (z.B. der eigene Name in einem Text oder ein Urlaubsvideo).
- Vermeidung typischer Empfehlungen, die immer wieder in der realen Welt gegeben werden.
- Die eigene Identität nur Pseudonymen offenbaren, denen schon ein gewisses Grundvertrauen entgegen gebracht wird.

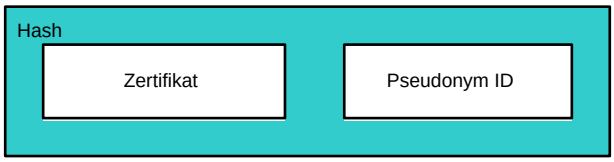
6. Absichtliche Aufdeckung der Identität

Der vorangegangene Abschnitt hat sich damit beschäftigt zu verhindern, das die Verbindung zwischen Pseudonym und Identität aufgedeckt wird. In manchen Situationen kann es jedoch sinnvoll sein, Teile der Identität gegenüber einem anderen preiszugeben. Dies ist zum Beispiel der Fall unter Freunden, die sich gegenseitig so sehr vertrauen, das sie sich sicher sind, dass keiner der anderen diese Information weitergibt.

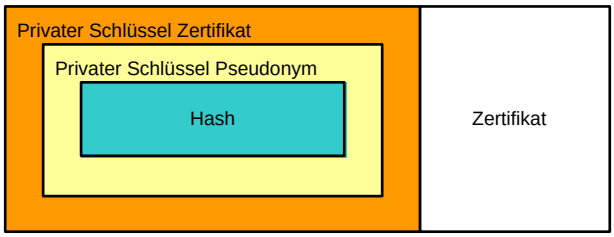
Besitzt ein Benutzer ein Zertifikat, das ihm Teile seiner Identität (z. B. der Name durch X.509-Zertifikate) oder bestimmte Fähigkeiten (z. B. der Beruf oder ein Fortbildungszeugnis) bescheinigt, so kann es sinnvoll sein, diese Zertifikate anderen Pseudonymen zu kommen zu lassen. Diese Pseudonyme können dann aufgrund des entgegengebrachten Vertrauens und des Inhalts des Zertifikats entscheiden, ob sie denjenigen, der ihnen gegenüber etwas preisgegeben hat, in Zukunft anders behandeln möchten. Denkbar wäre hier, dass ebenfalls ein Zertifikat zurückgesendet wird oder das der Vertrauenswert in das Pseudonym in einer Kategorie, die evtl. mit dem Zertifikat in Verbindung steht, manuell verändert wird.

Zur Zeit werden durch den Empfehlungsdienst nur X.509-Zertifikate, die den Namen einer Person an einen öffentlichen Schlüssel binden, unterstützt. Damit der Empfänger überprüfen kann, ob der Sender des Zertifikat auch wirklich der Besitzer dieses ist, wurde das nachfolgend dargestellte Protokoll entwickelt.

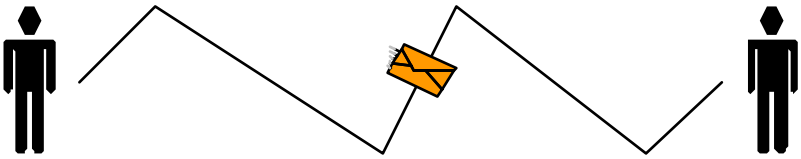
1. Der Sender erstellt einen Hash-Wert des Zertifikats und der ID des Pseudonyms.



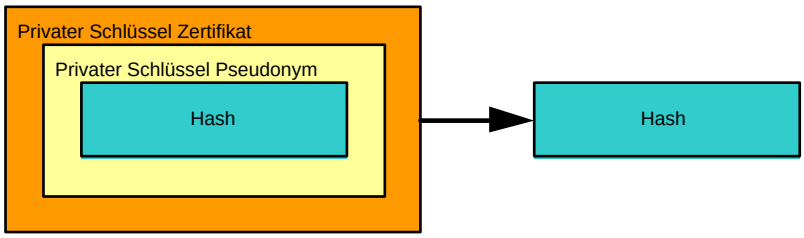
2. Der Sender verschlüsselt den Hash-Wert zuerst mit dem privaten Schlüssel des Pseudonyms und anschließend mit dem privaten Schlüssel des Zertifikats. Zusätzlich wird das Zertifikat an die entstandene Nachricht angehängt.



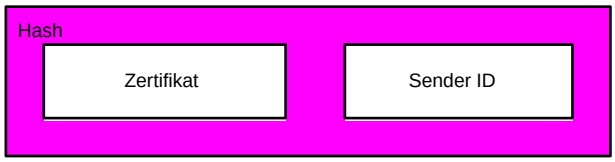
3. Der Sender schickt die verschlüsselte Nachricht zusammen mit dem Zertifikat über das anonyme Peer-to-Peer-Netzwerk an den Empfänger.



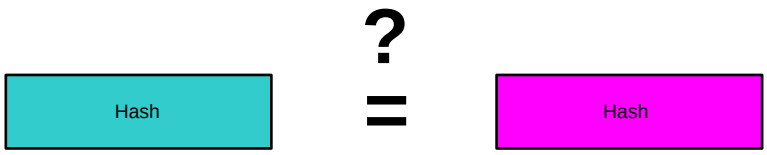
4. Der Empfänger entschlüsselt den verschlüsselten Teil mit Hilfe des öffentlichen Schlüssels des Zertifikat, der in demselben enthalten ist und dem öffentlichen Schlüssel des Pseudonyms, der dem Peer-to-Peer-Netzwerk entnommen werden kann.



5. Abschließend berechnet der Empfänger den Hash-Wert aus Sender-ID und erhaltenem Zertifikat.



6. Stimmen der erhaltene und der berechnete Hash-Wert überein, so muss der Sender sowohl den privaten Schlüssel des Pseudonyms, als auch den privaten Schlüssel des Zertifikats besitzen. Damit ist gezeigt, dass der Sender der Besitzer des Pseudonyms und des Zertifikats sein muss.



Damit der Benutzer nachvollziehen kann, wem gegenüber er seine wahre Identität preisgegeben hat, wird dieser Vorgang dauerhaft lokal mitprotokolliert.

Da die Pseudonymität innerhalb des Empfehlungsdienstes als hohes Gut angesehen wird, sollte dieser Vorgang nur stattfinden, wenn man sich sicher ist, das der Gegenüber, dem man Teile seiner Identität preisgibt, diese Informationen vertraulich behandelt.

7. Fazit und Ausblick

7.1. Fazit

Der UniTEC-Empfehlungsdienst erfordert eine andere Art von Identitätsmanagement, als es z. B. in der Arbeit von Marit Hansen [Köhntopp, 2000] definiert ist. Es geht nicht darum, wer welche Informationen der Identität eine Person hat, sondern wie verhindert werden kann, das Identitätsinformationen preisgegeben werden. Um dies zu bewerkstelligen wurden Pseudonyme in den Empfehlungsdienst eingeführt, die die wahre Identität des Benutzers verschleiern. Weiterhin wurde untersucht auf welche Weise die Pseudonymität eines Benutzers hintergangen und Identitätsinformationen aufgedeckt werden können. Ein Teil der Gefahren kann technisch verhindert werden, wie z. B. das Mithören durch Dritte verhindert werden kann, indem der Kommunikationskanal verschlüsselt wird. Ein anderer Teil der Gefahren kann nicht auf technische Weise verhindert werden. In diese Kategorie zählt die Verbindung die zwischen dem Verhalten des Benutzers in der realen Welt und dem UniTEC-Empfehlungsdienst besteht. Bei einigen aufgedeckten Gefahren kann der Benutzer jedoch unterstützt werden. In diese Kategorie fällt zum Beispiel die Rechtschreibprüfung der Empfehlungen und die Kontrolle der binären Elemente.

Jedoch ist auch anzunehmen, das eine Vielzahl von Hinweisen notwendig ist, um die Pseudonymität eines Benutzers aufdecken zu können. Einzelne Fakten, wie zum Beispiel das Wissen, das der Benutzer einen HP Drucker besitzt und deshalb darüber eine Empfehlung geschrieben hat, treffen für eine Vielzahl von Personen zu. Auch kann aus verschiedenen Merkmalen nicht auf das Wesen einer Person geschlossen werden. Wenn zum Beispiel eine Person über ein Luxusprodukt eine Empfehlung schreibt, ist noch nicht sicher, ob sich diese Person das Objekt leisten kann oder nur aus Interesse über das Objekt eine Empfehlung schreibt.

7.2. Ausblick

Die Unterstützung des Benutzers kann noch weiter ausgebaut werden. Zum Beispiel könnte zusätzlich zur syntaktischen Ähnlichkeit der Texte unterschiedlicher Empfehlungen die semantische Ähnlichkeit von Empfehlungen unterschiedlicher Pseudonyme untersucht werden und entsprechend Warnungen ausgegeben werden.

Der Sprach- und somit auch der Schreibstil verschiedener Person ist unterschiedlich und zum Teil auch sehr signifikant. Eine interessante Möglichkeit wäre sicherlich, die Texte der Empfehlungen so zu gestalten, dass sie für alle Benutzer einem einheitlichen Stil folgen, aber trotzdem nichts von ihrer Aussagekraft verlieren oder den Benutzer in irgendeiner Weise beim Erstellen der Empfehlung behindert.

Eine detailiertere Untersuchung der binären Elemente würde den Benutzer bei der Wahrung seiner Pseudonymität unterstützen. Denkbar wäre hier eine Szenenerkennung von Bildern oder Videodaten. Auch die Analyse von Tonaufnahmen kann dem Benutzer eine Hilfe sein, da er durch diese Maßnahme seine Pseudonymität schützen kann.

Untersuchenswert ist sicher auch die Frage, ob die Pseudonymverwendung und -verwaltung soweit automatisiert werden kann, das der Benutzer damit nicht mehr in Berührung kommt und sich vollständig auf die Aufgabe eines Empfehlungsdienstes – das Erstellen und das Auffinden von Empfehlungen – konzentrieren kann.

In dieser Diplomarbeit wurden hauptsächlich die Empfehlungen als Gefahrenquelle für die Pseudonymität einer Person betrachtet. Der Vorgang des Anfragens kann aber unter Umständen auch für die Aufdeckung der Identität ausgenutzt werden. Das heißt speziell die Verbindung von unterschiedlichen Pseudonymen eines Benutzers könnte aufgrund der Informationen, die eine Anfrage enthält, hergestellt werden. Ein verbindendes Element ist der Anfragetext. Dies soll an einem kurzen Beispiel verdeutlicht werden: Ein Benutzer ist auf der Suche nach einem speziellen Werkzeug, um sein Auto reparieren zu können. Er hat zwei Pseudonyme erstellt. Das eine ist für das Themengebiet Auto zuständig, das andere für das Themengebiet Heimwerken. Er startet eine Anfrage in einer Kategorie, für die das erste Pseudonym zuständig ist. Nach kurzer Zeit überlegt er sich, dass eine Anfrage in einer Kategorie, für die das zweite Pseudonym

zuständig ist, Erfolg versprechender ist und sendet eine Anfrage mit dem gleichen oder sehr ähnlichen Text. Treffen beide Anfragen bei ein und demselben Benutzer ein, so kann dieser die beiden Pseudonyme aufgrund des Anfragetextes einander zuordnen.

Eine weitere Zuordnungsmöglichkeit bieten die in den Anfragen enthaltenen Trust-Chains. Für eng verwandte Kategorien werden die Einträge der jeweiligen Kategorie im Vertrauensmodell ähnlich aussehen. Sowohl die Pseudonyme, für die ein Vertrauenswert einer Kategorie gespeichert wird, als auch der Vertrauenswert dieser Pseudonyme wird in beiden Kategorien ähnlich sein. Wird nun wiederum eine Anfrage in eng verwandten Kategorien, die durch zwei unterschiedliche Pseudonyme abgedeckt werden, versendet, so kann ein Benutzer, der beide Anfragen erhält aufgrund des ähnlichen Vertrauenswertes evtl. Rückschlüsse über die Verbindung der Pseudonyme des Senders schließen. Die letzten beiden Fälle führen zu der Frage, wie die Nachbarschaft gestaltet werden kann, um diese Art von Angriffen zu verhindern.

A. Software

A.1. Architektur

Während der Arbeit an der Diplomarbeit wurde in Zusammenarbeit mit Ralf Terdic und Fabian Aichele zwei mögliche Architekturen der Software erstellt, die sich in der Verwaltung der Benutzer unterscheiden. Die zugrunde liegenden Modelle und das Peer-to-Peer-Netzwerk sind bei beiden Architekturen gleich.

Bei der ersten Alternative ist jedem Benutzer ein Empfehlungsdienst-Agent zugeordnet, der sich um die Weitergabe der Empfehlungen und die Anfragen nach Empfehlungen kümmert (vgl. Abbildung A.1). Jeder Benutzer verwaltet seine Daten unabhängig von anderen Benutzern, die evtl. sogar den gleichen Rechner benutzen können. Hierdurch ist kein zusätzliches aufwändiges Benutzer-Management notwendig und die Daten der Benutzer sind gegeneinander abgeschottet.

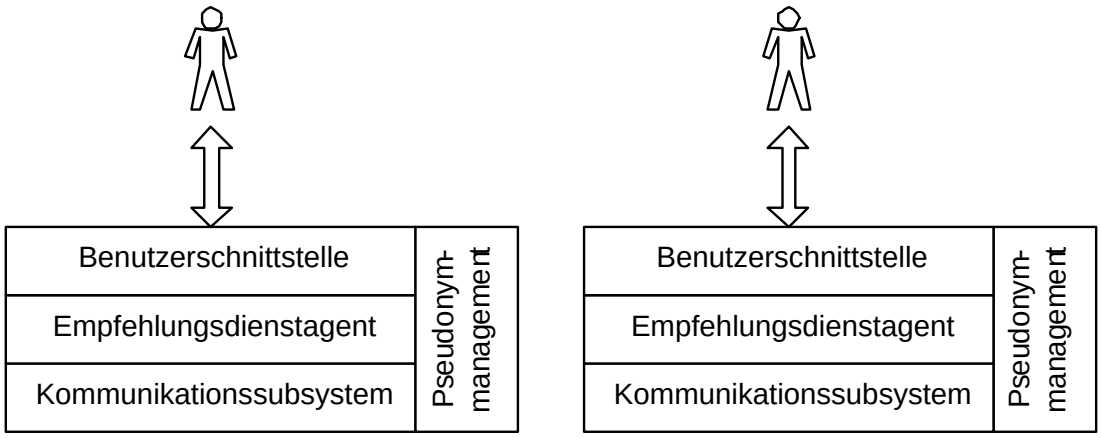


Abbildung A.1.: Die Software kann nur einen Benutzer verwalten.Sind mehrere Benutzer auf einem Rechner aktiv, so muss jeder einen eigenen Empfehlungsdienst benutzen.

Die zweite Alternative sieht vor, das ein Empfehlungsdienst-Agent mehrere Benutzer gleichzeitig von verschiedenen Orten aus bedienen kann (vgl. Abbildung A.2). Hierzu ist eine aufwändige Benutzer-Verwaltung inklusive Authentifikation und Autorisation notwendig. Durch die zusätzliche Kommunikation zwischen den Benutzern und dem Empfehlungsdienst-Agenten kommen zusätzlich Sicherheitsaspekte auf. Die Kommunikationsstrecke zwischen Empfehlungsdienst-Agenten und Benutzer muss ebenfalls gegenüber Dritten gesichert werden. Denkbare ist es, das der Empfehlungsdienst-Agent über eine Webservice für die Benutzer erreichbar ist. Nachteilig hierbei ist, das bei einem Ausfall des Empfehlungsdienst-Agenten unter Umständen viele Benutzer nicht mehr am Peer-to-Peer-Netzwerk teilnehmen können und so ein größerer Teil der Empfehlungen gegenüber der oben vorgestellten Lösung nicht mehr verfügbar wären.

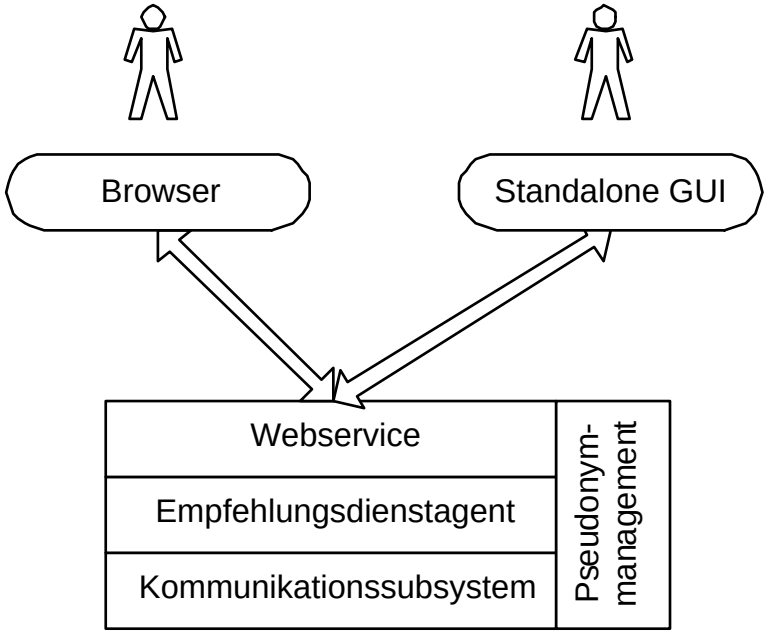


Abbildung A.2.: Ein Empfehlungsdienstagent kann von mehreren Benutzern gleichzeitig benutzt werden.

Innerhalb der beiden Diplomarbeiten wurde aufgrund der begrenzten Zeit und des zu begrenzenden Aufwands die erste Alternative gewählt.

A.2. Klassenbeschreibung

Dieser Abschnitt enthält eine kurze Beschreibung aller Klassen, die innerhalb des Projektes vom Autor erstellt wurden. Eine ausführlichere Beschreibung zusammen mit den Methodenbeschreibungen kann dem API entnommen werden.

A.2.1. com.hp.unitec.expertise

Name	Beschreibung
ExpertiseEntryNot-FoundException	Diese Ausnahme wird geworfen, falls bei der Suche nach einem Eintrag im Expertisen-Modell dieser nicht gefunden werden konnte.
ExpertiseModel	Die Hauptklasse des Expertisen-Modell. Über diese Klasse ist der Zugriff auf die einzelnen Einträge möglich.
ExpertiseModel-Entry	Kapselt einen Eintrag des Expertisen-Modells für ein bestimmtes Pseudonym des Benutzers. Gespeichert wird die ID des fremden Pseudonyms und die weiter oben aufgeführten Statistikdaten.
ExpertiseReply-MessageListener	Aktualisiert die Einträge des Expertisen-Modells, sobald eine Antwort auf eine Anfrage erhalten wird. Dazu muss diese Klasse bei einer Instanz der Klasse MessageDispatcher registriert werden.

A.2.2. com.hp.unitec.identity

Name	Beschreibung
IdentityManager	Nimmt die Nachricht einer Identitäts-Aufdeckung entgegen und verarbeitet sie. Gleichzeitig dient diese Klasse zum Versenden von Nachrichten die der Identitäts-Aufdeckung dienen.
RevelationInfo	Kapselt die Informationen, die benötigt werden, um die eigene Identität aufzudecken.

A.2.3. com.hp.unitec.gui

Name	Beschreibung
AddIdentityDialog	Mit Hilfe dieses Dialog kann ein Zertifikat mit passendem privaten Schlüssel importiert werden. Diese Informationen werden im IdentityManager gespeichert.
ExpertiseController	Ist für die Erstellung des ExpertiseView und die Verknüpfung zum ExpertiseModel zuständig.
ExpertiseView	Stellt die Werte des ExpertiseModel dar.
ForeignPseudonym-Controller	Ist für die Erstellung des ForeigPseudonymView und die Verknüpfung zum ForeignPseudonymModel zuständig.
ForeignPseudonym-View	Stellt die Werte aus dem ForeignPseudonymModel dar.
IdentityController	Ist für die Erstellung des IdentityView und die Verknüpfung zum IdentityManger zuständig
IdentityView	Stellt die Benutzeroberfläche dar, die es ermöglicht, eine Identität preiszugeben.
InitializeDialog	Mit Hilfe dieses Dialoges werden Werte vom Benutzer abgefragt, die zum Starten des Systems notwendig sind.
MenuBarController	Stellt die Verbindung der MenuBar-Komponenten zu den entsprechenden Methoden der anderen Klassen her.
MenuBarView	Zeigt die MenuBar an.
NewPseudonym-Dialog	Mit Hilfe dieses Dialoges können die notwendigen Werte vom Benutzer abgefragt werden, um ein neues Pseudonym zu erstellen.
OwnPseudonym-Controller	Stellt die Verbindung des OwnPseudonymView und des OwnPseudonymModel her.
OwnPseudonymView	Stellt die Daten des OwnPseudonymModels dar.
Responsibility-Controller	Stellt die Verbindung zwischen dem ResponsibilityView und dem Speicher, in dem die Daten für die Zuständigkeit der einzelnen Pseudonyme für die Kategorien gespeichert sind, her.
ResponsibilityView	Stellt die Verantwortlichkeiten der einzelnen Pseudonyme dar.

Name	Beschreibung
RevealDialog	Mit Hilfe dieses Dialogs kann die eigene Identität aufgedeckt werden.
UniTECController	Der Controller, der alle anderen Controller erstellt und kontrolliert.
UniTECView	Bietet einen Container in dem alle anderen Views eingebettet werden können.

A.2.4. com.hp.unitec.mvc

Die Benutzerschnittstelle ist an den Empfehlungsdienst-Agenten mittels des Model-View-Controller-Designpattern angebunden. Dieses Paket liefert die Basisklassen die die Benutzerschnittstelle benutzen kann.

Name	Beschreibung
AbstractController	Implementation des Interfaces Controller. Von dieser abstrakten Klasse können alle weiteren Controller erben.
AbstractModel	Implementation des Interfaces Model. Von dieser abstrakten Klasse können alle weiteren Models erben.
Controller	Definiert die Methoden, die ein Controller nach dem MVC-Pattern mindestens haben muss.
Model	Definiert die Methoden, die ein Model nach dem MVC-Pattern mindestens haben muss.
ModelEvent	Diese Klasse soll benutzt werden, um Änderungen in einem Modell an die entsprechenden Views zu propagieren.
ModelListener	Dieses Interface definiert die Methoden, die eine Klasse implementieren muss, wenn sie an ModelEvents eines Models interessiert ist.
View	Dieses Interface wird von den Klassen implementiert, die Daten dem Benutzer präsentieren.

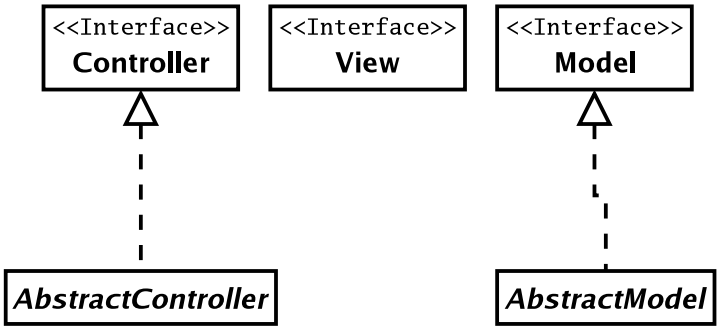


Abbildung A.3.: UML-Diagramm com.hp.unitec.mvc

A.2.5. com.hp.unitec.pseudonym

Name	Beschreibung
ForeignPseudonym	Erbt von Pseudonym. Kapselt die zusätzlichen Daten, die ein fremdes Pseudonym inne hat. Zur Zeit ist dies ein Zähler für die Erreichbarkeit.
ForeignPseudonym-Model	Erbt von PseudonymModel. Speichert die fremden Pseudonyme, mit denen der Benutzer Kontakt hatte. Bietet Methoden die fremden Pseudonyme zu manipulieren.
NicknameAlreadyIn-UseException	Diese Ausnahme wird geworfen, falls ein Benutzer versucht einem eigenen Pseudonym einen Namen zu geben, der schon von einem anderen Pseudonym verwendet wird.
OwnPseudonym	Erbt von Pseudonym. Kapselt die zusätzlichen Daten, die ein eigenes Pseudonym inne hat. Zur Zeit sind dies der private Schlüssel eines eigenen Pseudonyms und der Aktiv-Status.
OwnPseudonym-Model	Erbt von PseudonymModel. Speicher die eigenen Pseudonyme, die der Benutzer erzeugt hat und bietet Methoden, den Namen, die Gruppenzugehörigkeit und den Aktiv-Status zu ändern
Pseudonym	Kapselt die Daten, die die eigenen und die fremden Pseudonyme gemeinsam haben. Zur Zeit sind dies die ID, der öffentliche Schlüssel, der Name und die Gruppenzugehörigkeit.
PseudonymAlready-ExistsException	Diese Ausnahme wird geworfen, falls versucht wird ein Pseudonym zu einem Modell hinzuzufügen, das schon existiert.
PseudonymFactory	Bietet eine Methode ein neues eigenes Pseudonym zu erstellen. Es wird das Schlüsselpaar gemäß der RSA-Spezifikation erstellt und anschließend die ID generiert.
PseudonymModel	Vereint die Methoden, die die Modelle für eigene und fremde Pseudonyme gemeinsam haben. Zum Großteil sind dies Abfragemethoden.
PseudonymNot-FoundException	Diese Ausnahme wird geworfen, falls versucht wird, Informationen über ein Pseudonym abzufragen, das nicht existiert.

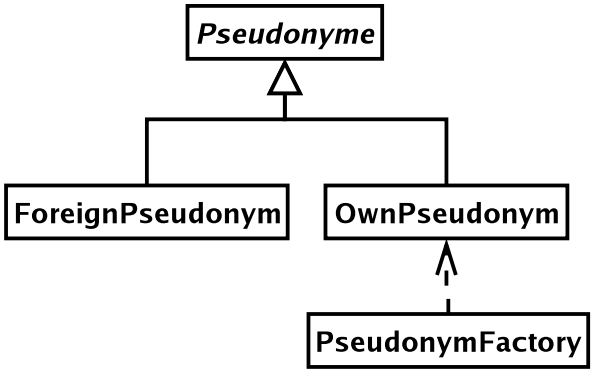


Abbildung A.4.: UML-Diagramm com.hp.unitec.pseudonym

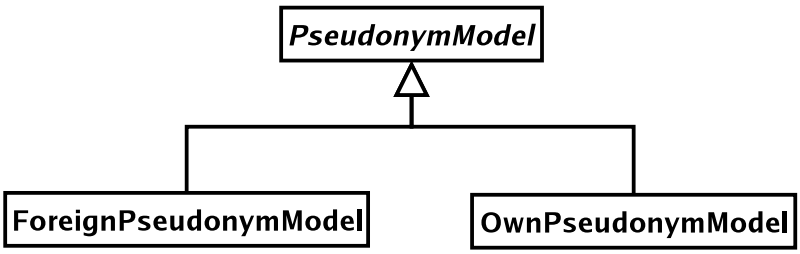


Abbildung A.5.: UML-Diagramm com.hp.unitec.pseudonym

A.2.6. com.hp.unitec.pseudonym.monitoring

Name	Beschreibung
BinaryCheck	Mit Hilfe dieser Klasse kann überprüft werden, ob in einer anderen Recommendation ein binäres Element verwendet wird, das auch in der zu prüfenden Recommendation vorkommt.
RecCheckAlgorithm	Dieses Interface definiert die Methoden, die ein Algorithmus, der eine Recommendation überprüft, implementieren muss.
Recommendation-Checker	Mit Hilfe dieser Klasse können Recommendations überprüft werden. Es werden alle Algorithmen aufgerufen, die im Konstruktor angelegt wurden.
SpellCheck	Eine einfache Rechtschreibprüfung für Empfehlungen.
WordMatchCheck	Überprüft, wieviele Worte eines Textes einer Empfehlung in den Texten anderer Empfehlungen vorkommen.

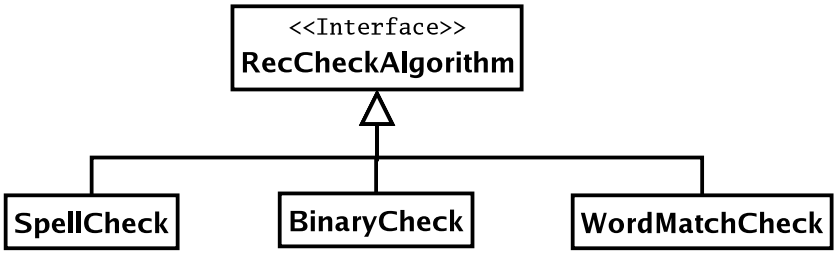


Abbildung A.6.: UML-Diagramm com.hp.unitec.pseudonym.monitoring

A.2.7. com.hp.unitec.pseudonym.monitoring.filter

Name	Beschreibung
BasicWordFilter	Ein einfacher Filter, der nicht filtert, sondern alle Worte durchlässt.
MostCommonWord-Filer	Dieser Filter filtert die Worte einer vorher definierten Wortliste aus.
WordFilter	Das Interface, das jeder Filter implementieren muss.

A.2.8. com.hp.unitec.pseudonym.responsibility

Name	Beschreibung
ResponsibilityEntry	Kapselt die Information, für welche Kategorie ein Pseudonym zuständig ist.
ResponsibilityModel	Hier werden die Einträge gespeichert, welches Pseuodnym für welche Kategorie zuständig ist.

A.2.9. com.hp.unitec.pseudonym.storage

Name	Beschreibung
StorageIterator	Diese Klasse ermöglicht es, durch einen Storage zu iterieren. Dazu wurde das Iterator-Interface implementiert.

A.2.10. com.hp.unitec.util

Name	Beschreibung
CryptoStore	Dies ist eine Hilfsklasse um auf einfache Weise Zertifikate und private Schlüssel in einem Java CryptoStore zu speichern.

B. Versuche zur Ähnlichkeit von Texten

Während der Bearbeitung der Diplomarbeit wurden Versuche mit dem Greedy-String-Tiling-Algorithmus durchgeführt. Diese Versuche umfassten vergleiche von ca. 20 unterschiedlichen Texten. Nachfolgend werden die zwei umfangreichsten Testreihen mit jeweils acht Texten vorgestellt.

Die Ergebnisse der ersten Testreihe können Abbildung B.1 entnommen werden. Bei dieser Testreihe wurden zufällig acht Produktbeschreibungen aus dem Angebot des Online-Auktionshauses Ebay von acht verschiedenen Personen entnommen. Vier der Texte wurden in der Kategorie Brettspiele (Board Games), zwei der Texte in der Kategorie Laserdrucker (Laser Printers) und die restlichen zwei Texte in einer Subkategorie von Laserdrucker, nämlich HP Laserjet monochrom, erstellt. Dargestellt ist die Anzahl der Token, nachdem aus den Texten Stoppwörter herausgefiltert und die übrigen Wörter auf ihre Lexeme reduziert wurden. Es wurden jeweils alle Texte paarweise verglichen. Zum Einen mit einer minimum-match-length von eins, d. h. es genügt, wenn einzelne Worte in den beiden Texten übereinstimmen, um diese zu markieren. Zum Anderen mit einer minimum-match-length von zwei, d. h. es müssen mindestens Wortpaare übereinstimmen, damit eine Markierung der Textstelle vorgenommen wird. Bei gleichen oder ähnlichen Kategorien fällt eine hohe Übereinstimmung der untersuchten Texte auf. Dies war zu erwarten, da i. A. gleiche Worte benutzt werden, um einen ähnlichen Sachverhalt zu beschreiben. Untersucht man die Ergebnisse der Vergleiche von Texten unterschiedlicher Kategorien, so fällt auf, dass eine hohe Übereinstimmung nur bei den Texten vorkommt, die eine stark unterschiedliche Länge aufweisen.

	Category	Token	min-match-length	ebay2		ebay3		ebay4		ebay5		ebay6		ebay7		ebay8	
ebay1	Board Games	124	1	16/79	**20,3%**	14/124	11,3%	14/82	**17,1%**	2/103	1,9%	17/124	13,7%	24/124	**19,4%**	12/124	9,7%
			2	2/79	2,5%	0/124	0,0%	4/82	4,9%	0/103	0,0%	2/124	1,6%	0/124	0,0%	0/124	0,0%
ebay2	Board Games	79	1			25/79	**31,6%**	13/79	**16,5%**	8/79	10,1%	18/79	**22,8%**	21/79	**26,6%**	15/79	**19,0%**
			2			6/79	7,6%	2/79	2,5%	0/79	0,0%	4/79	5,1%	2/79	2,5%	4/79	5,1%
ebay3	Board Games	195	1					12/82	14,6%	9/103	8,7%	16/140	11,4%	31/195	**15,9%**	19/195	9,7%
			2					0/82	0,0%	0/103	0,0%	0/140	0,0%	2/195	1,0%	0/195	0,0%
ebay4	Board Games	82	1							1/82	1,2%	8/82	9,8%	16/82	**19,5%**	7/82	8,5%
			2							0/82	0,0%	2/82	2,4%	0/82	0,0%	0/82	0,0%
ebay5	HP, LaserJet - Monochrome	103	1									22/103	**21,4%**	36/103	**35,0%**	39/103	**37,9%**
			2									5/103	4,9%	15/103	14,6%	12/103	11,7%
ebay6	HP, LaserJet - Monochrome	140	1											46/140	**32,9%**	31/140	**22,1%**
			2											10/140	7,1%	2/140	1,4%
ebay7	Laser Printers	313	1													69/268	**25,7%**
			2													13/268	4,9%
ebay8	Laser Printers	268	1														
			2														

2 * kürzerer Text < längerer Text
1.5 * kürzerer Text < längerer Text
Wert > 15%
gleiche Kategorie
ähnliche Kategorie

Abbildung B.1.: Die acht untersuchten Texte sind Produktbeschreibungen von acht unterschiedlichen Personen in zum Teil unterschiedlichen Kategorien. Diese Produktbeschreibungen wurden zufällig aus dem Angebot des Online-Auktionshauses Ebay ausgewählt.

Die Texte der zweiten Testreihe wurden von einer Person in z. T. unterschiedlichen Kategorien erstellt. In Abbildung B.2 sind die Meßergebnisse dargestellt. Hier fällt auf, dass es eine hohe Übereinstimmung der Texte, die in der gleichen Kategorie erstellt wurden, gibt. Diese Übereinstimmung ist für die Pseudonymität eines Benutzers nicht gefährlich, solange dieser nur ein Pseudonym für gleiche und eng verwandte Kategorien benutzt. Auch hier ergeben sich hohe Übereinstimmungen der Texte, wenn diese deutlich unterschiedlich lang sind. Es fällt jedoch auch auf, das Übereinstimmungen innerhalb von ähnlich langen Texten unterschiedlicher Kategorien auftreten. Nach der Untersuchung der Textstellen, die die Übereinstimmung verursachen, wurde festgestellt, das dies wiederverwendete Danksagungen und ähnliche Floskeln waren (z. B. „Thanks for viewing“).

Abschließend kann aus den Versuchsreihen folgendes Fazit gezogen werden: Der Algorithmus eigenet sich nicht, um semantische Ähnlichkeiten der Texte aufgrund syntaktischer Ähnlichkeiten zu entdecken. Der Algorithmus eignet sich jedoch, um längere gleiche Sätze oder Satzteile zu identifizieren. Diese können mit einer geschickt gewählten minimum-match-length aufgedeckt werden. Der Autor schlägt eine minimum-match-length von drei oder vier vor. Eine größere minimum-match-length könnte dazu führen, dass längere Sätze, die aufgrund der Stoppwortfilterung in der Länge geschrumpft sind, übersehen werden.

	Category	Token	min-match-length	bunny2		bunny3		bunny4		bunny5		bunny6		bunny7		bunny8	
bunny1	Ceramic, Porcelain	37	1	3/37	8,1%	5/37	13,5%	2/17	11,8%	4/37	10,8%	5/37	13,5%	3/37	8,1%	2/37	5,4%
			2	0/37	0,0%	0/37	0,0%	0/17	0,0%	0/37	0,0%	0/37	0,0%	0/37	0,0%	0/37	0,0%
bunny2	Children, Books	65	1			8/65	12,3%	4/17	23,5%	10/65	15,4%	5/65	7,7%	38/65	58,5%	14/64	21,5%
			2			0/65	0,0%	2/17	11,8%	2/65	3,1%	0/65	0,0%	28/65	43,1%	6/65	9,2%
bunny3	Metalware, Brass	79	1					3/17	17,6%	13/79	16,5%	8/73	11,0%	7/70	10,0%	7/68	10,3%
			2					2/17	11,8%	2/79	2,5%	2/73	2,7%	0/70	0,0%	0/68	0,0%
bunny4	Fishing	17	1							4/17	23,5%	4/17	23,5%	4/17	23,5%	1/17	5,9%
			2							2/17	11,8%	2/17	11,8%	0/17	0,0%	0/17	0,0%
bunny5	Shoulder Bags	97	1									9/73	12,3%	13/70	18,6%	7/68	10,3%
			2									2/73	2,7%	4/70	5,7%	0/68	0,0%
bunny6	McDonalds, Other Items	73	1											6/70	8,6%	7/68	10,3%
			2											0/70	0,0%	0/68	0,0%
bunny7	Children, Books	70	1													15/68	22,1%
			2													2/68	2,9%
bunny8	Children, Books	68	1														
			2														

2 * kürzerer Text < längerer Text
1.5 * kürzerer Text < längerer Text
Wert > 15%
gleiche Kategorie

Abbildung B.2.: Die acht untersuchten Texte dieser Versuchsreihe stammen alle von einer Person. Diese sind ebenfalls dem Angebot des Online-Auktionshauses Ebay entnommen.

C. Glossar

Anfrage Eine Anfrage ist eine Nachricht, die versendet wird, wenn ein Benutzer eine Empfehlung in einer bestimmten Kategorie sucht.

Anfragender Ein Anfragender ist ein Pseudonym, das eine Empfehlung in einer bestimmten Kategorie sucht. Zu diesem Zweck erstellt er eine Anfrage und sendet diese an seine berechnete Nachbarschaft.

Authority Ein Pseudonym ist eine Authority, sobald sie einen hohen Anteil der Anfragen, die an es gestellt werden, mit einer Empfehlung beantworten konnte.

Benutzer Der Benutzer ist die Person, die den UniTEC-Empfehlungsdienst benutzt.

Empfehlender Ein Empfehlender ist ein Pseudonym, das eine Empfehlung in einer bestimmten Kategorie erstellt hat. Diese Empfehlung ist mit dem Empfehlenden verknüpft.

Empfehlung Eine Empfehlung ist ein Ratschlag in einer bestimmten Kategorie zu einem bestimmten Produkt oder einer bestimmten Dienstleistung. Die Empfehlungen des UniTEC-Empfehlungsdienstes sind so variabel gestaltet, dass sie Empfehlungen der realen Welt möglichst genau abbilden können.

Empfehlungsdienst Ein Dienst, der es erlaubt Empfehlungen über verschiedene Produkte, Dienstleistungen oder sonstige Dienste zu erstellen, auszutauschen und einzusehen.

Expertisenmodell Das Expertisenmodell gibt vor, welche Statistiken über fremde Pseudonyme aufgeschlüsselt nach Kategorien gespeichert werden sollen. Dazu gehört die Anzahl der Anfragen an das Pseudonym, die Anzahl der Anfragen, die

das Pseudonym direkt beantwortet hat, die Anzahl der Anfragen, die das Pseudonym weitergeleitet hat, die Anzahl der Empfehlungen, die das Pseudonym in einer gewissen Kategorie gespeichert hat und die durchschnittliche Gewissheit des Empfehlenden.

Gewissheit Die Gewissheit ist ein Bestandteil einer Empfehlung. Sie gibt an, wie sicher sich der Empfehlende bei der Erstellung der Empfehlung war.

Hub Ein Hub ist ein Pseudonym, das einen hohen Anteil von Anfragen, die im gestellt wurden, erfolgreich weiterleiten konnte. Erfolgreich weiterleiten bedeutet, dass der Nachbar, an den die Anfrage weitergeleitet wurde, diese mit dem Versand einer Empfehlung beantworten konnte.

Identifikationsnummer, ID Die Identifikationsnummer wird benötigt, um ein Pseudonym eindeutig identifizieren zu können. Eine Identifikationsnummer sollte nie gleichzeitig von unterschiedlichen Pseudonymen benutzt werden.

Nachbar Ein Nachbar ist ein Pseudonym von dem erwartet wird, dass es entweder eine Anfrage beantworten kann oder andere Pseudonyme kennt, die die Anfrage beantworten können.

Nachbarschaft Die Menge von Nachbarn an die eine Anfrage gesendet wird.

Peer Teilnehmer eines Peer-to-Peer-Netzwerkes.

Peer-to-Peer-Netzwerk Ein Peer-to-Peer-Netzwerk ist eine Kommunikationsform, bei der alle Peers gleichberechtigt sind. Das heißt jeder Peer kann Dienste anbieten und Dienste von andern Peers nutzen. Nach Möglichkeit soll auf zentrale Komponenten verzichtet werden.

Pseudonym Ein Pseudonym verschleiert die Identität eines Benutzers. Der Benutzer kann ein Pseudonym benutzen, um eine gewisse Rolle abzubilden.

Pseudonym, eigenes Ein eigenes Pseudonym ist aus Sicht eines Benutzer ein Pseudonym, das er selbst erstellt hat, und mit dem er Empfehlungen erstellen und nach diesen suchen kann.

Pseudonym, fremdes Ein fremdes Pseudonym ist aus Sicht eines Benutzers ein Pseudonym, das er nicht selbst erstellt hat, sondern das er aufgrund einer Anfrage kennengelernt hat. Diesen Pseudonymen kann der Benutzer Anfragen nach Empfehlungen senden.

Pseudonymität Die Pseudonymität ist ein Zustand eines Benutzers, bei dem der Gegenüber die Identität des Benutzers nicht kennt, den Benutzer aber trotzdem identifizieren kann.

Trust-Update-Algorithm siehe Vertrauens-Aktualisierungs-Algorithmus.

UniTEC Abkürzung für Universal Trust-Architecture for Electronic Commerce. Dies ist der Name des Projektes.

Vertrauens-Aktualisierungs-Algorithmus Dieser Algorithmus berechnet die Veränderung eines Vertrauenswert innerhalb des Vertrauens-Modell nachdem eine erhaltene Empfehlung als gut, bzw. schlecht eingestuft wurde.

Vertrauensmodell Das Vertrauensmodell gibt vor, wie der Vertrauenswert zu einem Pseudonym gespeichert werden soll und wie der Vertrauenswert mit Hilfe eines Vertrauens-Aktualisierungs-Algorithmus aktualisiert werden soll.

Vertrauenswert Qualitative Einschätzung des Vertrauens in einen Empfehlenden. Innerhalb des Agenten wird der Vertrauenswert in das Intervall $[0, 1]$ abgebildet.

D. Projektplan

D.1. Einleitung

D.1.1. Zweck dieses Dokuments

Der Projektplan dient dem Bearbeiter und dem Betreuer als Grundlage für die Projektüberwachung und -steuerung. Er legt den Projektablauf aber nicht entgültig fest, sondern soll im Laufe der Arbeit ergänzt und angepasst werden. So ist es z.B. möglich, dass anfänglich spezifizierte Arbeitspakete weiter verfeinert oder Arbeitspakete zugunsten anderer aufgegeben werden.

Der Projektplan kann in Absprache mit dem Betreuer in beiderseitigem Einverständnis in folgenden Fällen angepasst werden:

- bei Verzug,
- bei vorzeitigem Abschluss eines Arbeitspakets oder
- bei Gewinn neuer Erkenntnisse.

D.1.2. Projektübersicht und Motivation

Bei diesem Projekt handelt es sich um eine Diplomarbeit an der Abteilung Verteilte Systeme des Instituts für Parallele und Verteilte Systeme (IPVS) der Universität Stuttgart. Sie findet im Rahmen des Projektes UniTEC (Universal Trust-architecture

for Electronic Commerce), einer Kooperation zwischen der Abteilung Verteilte Systeme und Hewlett-Packard, statt. Die maximale Bearbeitungszeit von sechs Monaten beginnt am 1. Mai 2003 und endet somit am 31. Oktober 2003. Von Seiten der Abteilung Verteilte Systeme steht Dipl.-Inf. Michael Kinateder und von Seiten der Firma Hewlett-Packard Oliver Kopp als Betreuer zur Verfügung.

Verschiedene Studien[1] haben gezeigt, das die Konsumenten dem E-Commerce gegenüber noch Vorbehalte haben. Die natürlichen vertrauensbildenden Maßnahmen (wie zum Beispiel Mund-zu-Mund-Propaganda) fehlen in der digitalen Welt. UniTEC verfolgt das Ziel, das Vertrauenserlebnis der realen Welt in die digitale Welt zu übertragen. Dazu soll eine in vielen Bereichen verwendbare Architektur für die Einschätzung von Vertrauen gegenüber anderen Personen, Dienstleistungen, Produkten u. ä. aufgebaut werden. Hierzu kann unter anderem zwischen den Anwendern die Einschätzung der Vertrauenswürdigkeit anderer Personen, Dienstleistungen, usw. ausgetauscht und das eigene Vertrauen aufgrund dieser Informationen angepasst werden.

Diese Diplomarbeit beschäftigt sich mit einer Komponente dieser Architektur. Es soll ein leistungsfähiges Identitätsmanagement aufgebaut werden. Hierbei spielt die Frage: „Wem gebe ich wie viel Information über meine eigene Identität preis?“ eine zentrale Rolle. In der realen Welt ist Identitätsmanagement ein natürlicher Prozess. Bei direktem Kontakt der Personen liegt die Kontrolle, wie viel Informationen eine Person dem Gegenüber zur Verfügung stellt, bei der Person selbst. Jedoch entstehen auch hier schon erste Probleme, sobald die erste Person die zweite verlässt. Wird die zweite Person mit den vertrauenswürdigen Daten so umgehen, wie es die erste Person von ihr erwartet oder werden die Informationen ohne Zustimmung oder Wissen weitergegeben?

In der digitalen Welt besteht dieses Problem noch viel stärker. Identitäts-Informationen können leicht und schnell kopiert und verteilt werden, ohne das die betroffene Person darüber informiert werden muss. Um diese Möglichkeiten einzuschränken, wird eine Software-Einheit benötigt, die dem Benutzer möglichst viel Kontrolle über seine Identitäts-Informationen gibt. Diese Software soll die Verwaltung sowohl der eigenen Identitäts-Informationen, als auch der Informationen, die man von anderen Personen bekommt, ermöglichen. Offene Fragen sind: Wie kann man weitmöglichst einschränken,

[1] `http://www.cdt.org/privacy/survey/findings/`, `http://www.privacyexchange.org/iss/surveys/surveys.html`

dass Informationen nicht ohne Erlaubnis weitergegeben werden? Wie kann man verhindern, das Informationen während der Kommunikation abgefangen und ausgewertet werden? Vor allem im Hinblick auf das Aufdecken von Pseudonymen aufgrund gesammelter Informationen, die einem gemeinsamen, global eindeutigem Merkmal zugeordnet werden können.

Parallel zu dieser Diplomarbeit bearbeitet Ralf Terdic die Diplomarbeit „Konzeption und Realisierung eines Peer-2-Peer Empfehlungsdienstes“ und Fabian Aichele die Studienarbeit „Realisierung und Evaluierung von Trust-Update-Algorithmen für das UniTEC Reputationssystem “. Da die drei Themen stark miteinander in Verbindung stehen, wird eine enge Zusammenarbeit angestrebt.

In den nachfolgenden Abschnitten werden zunächst die Arbeitspakete identifiziert und beschrieben, dann wird ein Zeitplan zur Durchführung festgelegt. Schließlich erfolgt die Definition der Meilensteine und der dafür zu erstellenden Dokumente.

D.2. Beschreibung der Arbeitspakete

Dieser Abschnitt umfasst die Definition der wesentlichen Arbeitspakete und die Abhängigkeiten zwischen einzelnen dieser Pakete.

D.2.1. Definition der Arbeitspakete

Die im Rahmen der Diplomarbeit durchzuführenden Tätigkeiten lassen sich in folgende, zusammenhängende Arbeitspakete aufgliedern. Für die Bearbeitung sind 127 Arbeitstage vorgesehen.

AP1 - Projektplan: Erstellung eines initialen Projektplanes
Zeitraum: 5.5. - 13.5.
Aufwand: 4 AT
Ergebnisse: initialer Projektplan
Status: abgeschlossen

AP2 - Einarbeitung: Nach dieser Phase soll der Bearbeiter einen Überblick über die Aufgabenstellung (Motivation, Ziele) und der damit zusammenhängenden Themengebiete (Identitätsmanagement, Peer-2-Peer) haben. Dieses Wissen soll mindestens soweit reichen, dass die anschließenden Aufstellung der Möglichkeiten eines Identiätsmanagementsystems ohne Probleme in kurzer Zeit erfolgen kann.
Zeitraum: 5.5. - 23.5.
Aufwand: 11 AT
Ergebnisse: Einführende Kapitel Identitätsmanagement und Peer-2-Peer in erster Version fertig gestellt
Status: abgeschlossen

AP3 - Aufstellung der Charakteristika: In diesem Arbeitspaket wird eine Liste erarbeitet, welche Möglichkeiten ein elektronisches Identiätsmanagementsystem bieten kann. Aspekte, die das Identitätsmanagement nur am Rand betreffen, sollen bewusst aufgenommen werden. Die Auswahl der zu realisierenden Aufgaben folgt im nächsten Arbeitspaket.
Zeitraum: 26.5. - 30.5.
Aufwand: 5 AT
Ergebnisse: Liste mit möglichen Use-cases für den Benutzer
Status: abgeschlossen

AP4 - Spezifikation: Aus der im vorangegangenen Arbeitspaket aufgestellten Liste möglicher Use-cases werden diejenigen heraus gesucht, die das zu erstellende System abdecken soll. Anschließend werden die Anforderungen, die das zugrunde liegende Peer-2-Peer-Netz erfüllen muss, identifiziert. Besonderes Augenmerk wird auf die Sicherheit der Identitäts-Informationen innerhalb der Software, d.h. auf dem Rechner des Anwenders, und außerhalb, d.h. während der Kommunikation zwischen zwei Benutzern des Systems gelegt. Am Ende wird eine Beschreibung der Funktionen, die das System erfüllen muss und derjenigen, die in die Kategorie „Nice-to-have“ fallen, vorhanden sein.
Zeitraum: 2.6. - 13.6.
Aufwand: 9 AT
Ergebnisse: Spezifikation des Identifikationsmanagementsystem und des zugrunde liegenden Peer-2-Peer-Netz
Status: abgeschlossen

AP5 - Evaluation bestehender Systeme: Erfassung bestehender Lösungen und Überprüfung dieser auf ihre Tauglichkeit für das Projekt aufgrund der in der Spezifikation festgehaltenen zu erfüllenden Merkmale. Anschließend Auswahl der Lösungsmöglichkeiten oder Ideen, die Funktionen des in der Spezifikation festgehaltenen Systems erfüllen.
Zeitraum: 16.6. - 24.6.
Aufwand: 6 AT
Ergebnisse: Liste von bestehenden Lösungen, die Funktionen der Spezifikation abdecken und die in das zu erstellende System übernommen werden sollen.
Status: abgeschlossen

AP6 - Erarbeiten der fehlenden Lösungen: In dieser Phase werden die noch fehlenden Funktionen, die nicht durch bestehende Lösungen abgedeckt werden können, erarbeitet und festgehalten. Nach dieser Phase sind die Algorithmen zur Lösung der Aufgabe bekannt.
Zeitraum: 25.6. - 11.7.
Aufwand: 13 AT
Ergebnisse: Kapitel über die Spezifikation und der ausgesuchten Algorithmen
Status: abgeschlossen

AP7 - Systementwurf: Erarbeitung der Klassen mit ihren Aufgaben und die Beziehungen untereinander. Dies wird mit Hilfe der Unified Modelling Language (UML) stattfinden.
Zeitraum: 14.7. - 25.7.
Aufwand: 10 AT
Ergebnisse: UML-Diagramme mit erklärenden Texten
Status: abgeschlossen

AP8 - Implementation, Test und Korrektur: Die im vorangegangenen Arbeitspaket definierte Struktur wird in Code umgesetzt. Gleichzeitig wird die Klassendokumentation mittels Javadoc durchgeführt. Der Code soll durch Codereviews überprüft werden. Die Durchführungstage können wahrscheinlich nur sehr kurzfristig bestimmt werden, da sie von der Auslastung der HP-Mitarbeiter abhängt. Die erstellten Module werden laufend Funktionstests unterzogen. Es ist vor allem auf die Abhängigkeiten zwischen den drei Arbeiten zu achten. Die während der Implementierung und des Tests aufgedeckten Fehler werden korrigiert.

Zeitraum: 28.7. - 3.10.
Aufwand: 48 AT
Ergebnisse:
Status: abgeschlossen

AP9 - Erstellen der Diplomarbeit: Die fehlenden Teile der schriftlichen Ausfertigung, die nicht in einem anderen Arbeitspaket bearbeitet wurden, werden angefertigt. Der Fortschritt wird nach Fertigstellung eines Kapitels zusammen mit dem Betreuer überprüft und gegebenenfalls die Arbeit angepasst.
Zeitraum: 6.10. - 31.10.
Aufwand: 20 AT
Ergebnisse:
Status: abgeschlossen

AP10 - Zwischenvortrag: Der Zwischenvortrag dient dem Präsentieren der bisherigen Ergebnisse und als Übung für den Abschlußvortrag. Hier kann noch auf die Spezifikation des Systems Einfluss genommen werden.
Zeitpunkt: 29.8.
Ergebnisse: Präsentation
Status: abgeschlossen

AP11 - Abschlußvortrag: Die Ergebnisse der halbjährigen Arbeit werden präsentiert.
Zeitpunkt: 8.12.
Ergebnisse: Präsentation
Status: offen

D.2.2. Abhängigkeit der Arbeitspakete

Die Abhängigkeiten zwischen den Arbeitspaketen der Diplomarbeit können Abbildung D.1 entnommen werden. Die Abhängigkeiten des Arbeitspaketes AP11 (Erstellen der Diplomarbeit) sind aufgrund der Übersichtlichkeit nicht im Bild aufgenommen, da dieses von allen anderen Arbeitspaketen abhängig ist.

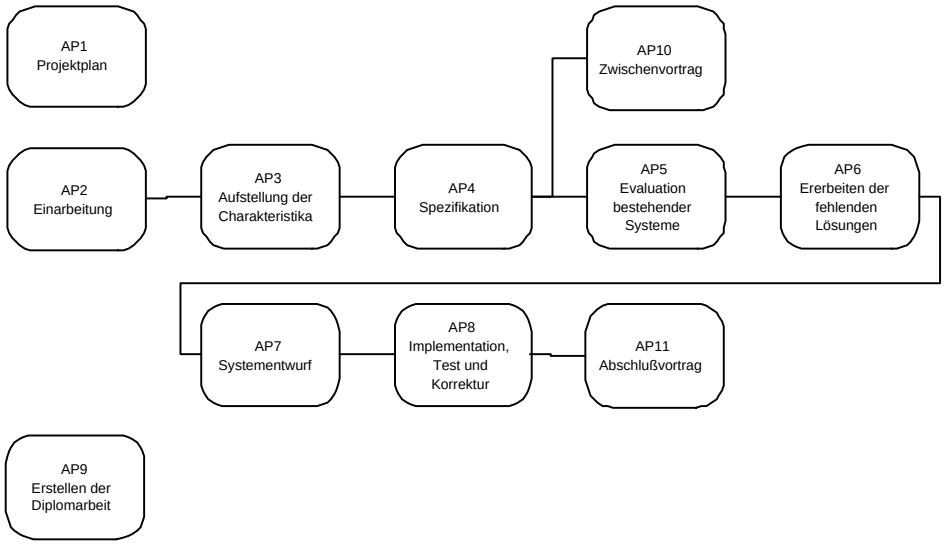


Abbildung D.1.: Die Abhängigkeit der Arbeitspakete

D.3. Zeitplan

Tabelle D.1 zeigt eine Übersicht über die zeitliche Planung der Arbeitspakete und Meilensteine. Der Bearbeitungzeitraum beträgt sechs Monate. Wenn sich der Zeitraum mehrerer Arbeitspakete überlappt, dann werden diese parallel bearbeitet.

D.4. Dokumente und Meilensteine

Als Ergebnis der jeweiligen Arbeitspakete sollen zu den Meilensteinen folgende Dokumente entstehen, die nach diesen einer Versionskontrolle unterliegen und nur in Absprache mit dem Betreuer geändert werden sollen.

Projektplan: Das vorliegende Dokument

Spezifikation: Ausführliche Beschreibung, welche Aufgaben das zu erstellende System leisten soll

Realisierungsbericht: In diesem Dokument wird aufgeführt, mit welchen Algorithmen die in der Spezifikation enthaltenen Anforderungen realisiert werden.

geplanter Zeitraum	Aufwand	Arbeitspakete und Meilensteine	tatsächlicher Zeitraum
5.5. – 13.5.	4 AT	AP1 Projektplan	5.5. – 22.5.
5.5. – 23.5.	11 AT	AP2 Einarbeitung	5.5. – 23.5.
23.5.		**Meilenstein 1**	
26.5. – 30.5.	5 AT	AP3 Aufstellung der Charakteristika	26.5. – 30.5.
2.6. – 13.6.	9 AT	AP4 Spezifikation	2.6. – 13.6.
13.6.		**Meilenstein 2**	
16.6. – 24.6.	6 AT	AP5 Evaluation bestehender Systeme	16.6 – 24.6.
25.6 – 11.7.	13 AT	AP6 Erarbeiten der fehlenden Lösungen	25.6. – 11.7.
11.7.		**Meilenstein 3**	
14.7. – 25.7.	10 AT	AP7 Systementwurf	14.7. – 8.8.
25.7.		**Meilenstein 4**	
28.7. – 3.10.	48 AT	AP8 Implementation, Test und Korrektur	23.7. – 29.9.
3.10.		**Meilenstein 5**	
6.10. – 31.10.	20 AT	AP9 Diplomarbeit	29.9. – 31.10.
31.10		**Meilenstein 6**	
29.8.		AP10 Zwischenvortrag	29.8.
8.12.		AP11 Abschlußvortrag	

Tabelle D.1.: Der Zeitplan des Projekts. (AT = Arbeitstage)

Evaluationsprotokoll: Übersicht der evaluierten Lösungen mit ihren Eigenschaften und Bewertung ihrer Tauglichkeit für das System

Entwurf: Klassenbeschreibungen und ihr Zusammenhang

Code: Der Entwurf umgesetzt in ausführbaren Javacode

Testprotokolle: Beschreibung des Ablauf des Tests und der Ergebnisse

Diplomarbeit: Schriftliche Ausfertigung über die Ergebnisse dieses Projektes

Präsentationen: Die beiden Präsentationen für die Zwischen- und Abschlußvorträge

Zu den Meilensteinen sind die folgenden Dokumente zu erstellen:

23.5.2003 - Meilenstein 1 Projektplan, Erste Version der Einführungskapitel

13.6.2003 - Meilenstein 2 List der Charakteristika eines Identitätsmanagement-Systems, Spezifikation

11.7.2003 - Meilenstein 3 Evaluationsprotokolle, Realisierungsbericht

25.7.2003 - Meilenstein 4 Entwurf

3.10.2003 - Meilenstein 5 Code, Testprotokolle

31.10.2003 - Meilenstein 6 Diplomarbeit

D.4.1. Synchronisationspunkte

Da zwischen den drei Arbeiten große Abhängigkeiten bestehen, sind Synchronisationspunkte, zu denen bestimmte Teilaufgaben bearbeitet sein müssen, zwingend nötig. Diese werden nachfolgend aufgelistet:

30. Mai Festlegen der Anforderungen an das zugrunde liegenden Peer-2-Peer-Netzwerk und Aufteilung der Aufgaben

6. Juni Festlegung der Anforderungen zwischen den Systemen, die in den anderen Arbeiten erstellt werden

14. Juli Festlegung der Schnittstellen zwischen den Projektteilen

15. - 19. September Erster Integrationstest

1. - 2. Oktober Zweiter Integrationstest

29. August und 8. Dezember Vorträge

D.5. Anforderung an die Umgebung

Die Firma Hewlett-Packard stellt dem Bearbeiter einen Arbeitsplatz inklusive einem für die Aufgabe angemessen ausgestatteten Rechner zur Verfügung.

D.5.1. Verwendete Software

- Java Software Developement Kit Standard Edition Version 1.4.1_02
- Eclipse als Entwicklungsumgebung
- CVS als Konfigurationmanagement
- LaTeX zur Erstellung benötigter Dokumente
- Omondo UML Plugin für Eclipse zur Erstellung der UML-Diagramme

Abbildungsverzeichnis

1.1. Abbildung des Vertrauens auf einen Vertrauenswert 11
1.2. Ausschnitt aus einem Vertrauensmodell 12
1.3. Eine Konfiguration der Nachbarschaft mit den zugehörigen Hub-Werten 14
1.4. Eine weitere Konfiguration der Nachbarschaft 15
1.5. Übersicht über den Zusammenhang der beschriebenen Komponenten . 17

4.1. Wahrscheinlickeit einer Kollision bei der ID-Generierung 32

5.1. Zwei Gruppierungen der drei Kategorien Fenster, Glas und Wein. . . . 37
5.2. Verbindung zweier Pseudonyme aufgrund der Trust-Chain 38
5.3. Empfohlene Zuordnung von Pseudonymen zu Kategorien 39
5.4. Eine Beispielempfehlung 40
5.5. Beispiel einer Hamming-Distanz 42
5.6. Beispiel einer Levenshtein Distanz 42
5.7. Beispielberechnung einer longest common subsequence 43
5.8. Beispiel einer Distanz bei Greedy-String-Tiling 45

A.1. Einfache Architektur 57
A.2. Mehrbenutzer-Architektur 58
A.3. UML-Diagramm com.hp.unitec.mvc 62
A.4. UML-Diagramm com.hp.unitec.pseudonym 64
A.5. UML-Diagramm com.hp.unitec.pseudonym 64
A.6. UML-Diagramm com.hp.unitec.pseudonym.monitoring 65

B.1. Ergebnisse der Untersuchung des GST-Algrotihmus I 68
B.2. Ergebnisse der Untersuchung des GST-Algrotihmus II 70

D.1. Die Abhängigkeit der Arbeitspakete 81

Tabellenverzeichnis

4.1. Übersicht Generierung der ID 31

D.1. Zeitplan des Projekts 82

Literaturverzeichnis

[Aichele 2003] Aichele, Fabian: *Realisierung und Evaluierung von Trust-Update-Algorithmen für das UniTEC Reputationssystem*, Stuttgart, Diplomarbeit, 2003

[Köhntopp 2000] Köhntopp, Marit: *Datenschutz von A-Z.* Kap. Identitätsmanagment, Helmut Bäumler, Astrid Breinlinger, Hans-Hermann Schrader, 2000

[Leuf 2002] Leuf, Bo: *Peer to Peer – collaboration and sharing over the Internet.* Addison-Wesley, 2002

[NIST 1995] NIST: *Secure Hash Standard.* http://www.itl.nist.gov/fipspubs/fip180-1.htm. April 1995

[Schoder u. a. 2002] Schoder, Detlef (Hrsg.) ; Fischbach, Kai (Hrsg.) ; Teichmann, Renè (Hrsg.): *Peer-to-Peer Ökonomische, technologische und juristische Perspektiven.* Berlin, u. a., 2002

[Terdic 2003] Terdic, Ralf: *Konzeption und Ralisierung eines Peer-2-Peer Empfehlungsdienstes*, Stuttgart, Diplomarbeit, 2003

[Wise 1993] Wise, Michael J.: *String Similarity via Greedy String Tiling and Running Karp-Rabin Matching.* Dezember 1993. – Department of Computer Science, University of Sydney, Australia

Erklärung

Hiermit versichere ich, diese Arbeit selbständig verfasst und nur die angegebenen Quellen benutzt zu haben.

(Jochen Widmaier)

Stuttgart, den 28. Oktober 2003