

TIMA

The logo features the word "TIMA" in a bold, dark blue, sans-serif font. Two bright blue arrows are integrated into the design: one starts at the top right, extends horizontally to the left, and then curves downwards to point at the letter 'M'; the other starts at the bottom left, extends horizontally to the right, and then curves upwards to point at the letter 'A'. This creates a continuous loop around the text.

Abschlussarbeit

TIMA

Nathanael Philipp, Felix Rauchfuß, Kai Trott

22. September 2015



Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
2	TIMA als Citizen Science Projekt	3
2.1	Spiele	3
2.1.1	Wortselektierungsalgorithmus	3
2.1.2	ExcludeWords	4
2.1.3	Assoziationskette	4
2.2	Newsletter	5
2.3	Datenqualität	5
3	Assoziationsdatenbank und API	7
3.1	Backend und Datenbank	7
3.1.1	Datenmodell	7
3.2	API	9
3.2.1	Nicht autorisierte Anfragen	9
3.2.2	Autorisierte Anfragen	10
3.2.3	OAI-PMH	12
4	Applikationen und Webseite	13
4.1	Webfrontend	13
4.2	Applikation	13
4.2.1	Framework	13
4.2.2	Aufbau	13
4.3	Sicherheit	14
5	Ausblick	16
6	Zusammenfassung	18

1 Einleitung

In Zeiten von schnellen Prozessoren und riesigen Speichermedien sind mithilfe von Text Mining und automatischer Sprachverarbeitung eine Vielzahl von Datenbanken entstanden, die ganze Sprachen aufgrund von Satzbau, Wortkookurrenzen und Wortarten analysieren und speichern. In dieser Menge der Daten fehlt jedoch eine sehr wichtige Eigenschaft von Worten - die Assoziation. Bisher war es nicht möglich, diese bedeutende menschliche Fähigkeit maschinell zu simulieren. TIMA, rekursives Palindrom für "TIMA is my association", oder auf Deutsch: "TIMA ist meine Assoziation", setzt es sich zum Ziel eine Datenbank zu schaffen, bei der diese Verbindungen zwischen Worten abgerufen werden können. Da wie beschrieben bisher keine automatische Methode dazu existiert, setzt TIMA auf eine Menge freiwilliger Nutzer, die ihre Assoziationen zu Worten eingeben.

Um eine relevante Menge an Daten sammeln zu können, wird TIMA als Citizen Science Projekt aufgezogen. Ein derartiges Projekt hat eine Reihe besonderer Ansprüche und um ihnen gerecht zu werden, sind eine ganze Reihe Vorkehrungen zu treffen, die in dieser Arbeit betrachtet werden sollen.

Zuerst wird näher betrachtet, warum das Erstellen einer Assoziationsdatenbank überhaupt sinnvoll ist und welche Funktionen für Nutzer ansprechend wären. Danach werden einige Gedanken zur Gestaltung als Citizen Science Projekt geäußert. Besonders wird dabei auf den Schutz der gesammelten Daten eingegangen und der wichtigen Frage, wie man Nutzer motiviert, am Projekt teilzunehmen. Nach den technischen Details zur Implementierung einzelner Bestandteile: der Datenbank, der API, der Webseite und einer ersten App, die die Möglichkeiten der API anschaulich demonstriert, folgt ein Abschnitt zum Ausblick auf zukünftige Projekte, die entweder zur Datenbank beitragen können oder sie nutzen.

1.1 Motivation

Die Möglichkeiten einer Assoziationsdatenbank sind vermutlich in erster Linie in Bereichen der automatischen Sprachverarbeitung angesiedelt, dort jedoch beinahe in jedem Teilgebiet nutzbar.

Ein großes Problem aller Automaten ist ihre Reaktivität. Sie sind stets auf Schlüsselbegriffe angewiesen, die der Nutzer eingibt, beziehungsweise spricht. So ist für Suchanfragen jeglicher Art, ob nun Internetsuche, Eingabe in das Navigationsge-

1 Einleitung

rät oder Sprachbefehle neuartiger Steuerungen für mobile Endgeräte, wie zum Beispiel Siri, sehr schwer die korrekte Reaktion zu liefern, wenn der Nutzer von fest programmierter Terminologie abweicht. Sucht ein Autofahrer statt einer Tankstelle nach Benzin, wird er eventuell keine Antwort bekommen, obwohl einem Menschen intuitiv klar ist, wonach der Fahrer sucht. Eine Maschine kann diese Schlüsse jedoch nicht ziehen und daher müsste man als Programmierer jede einzelne dieser Möglichkeiten bedenken und implementieren. Selbst für ein Navigationsgerät mit relativ eingeschränktem Handlungsspielraum ist dies schon sehr aufwendig, für eine Internetsuchmaschine oder Anwendungen im Bereich des Computational Advertisings jedoch quasi unmöglich. Die Varianz an Suchbegriffen ist einfach zu groß.

Die Problematik einer solchen Datenbank ist jedoch, dass sie sich nicht automatisch erstellen lässt. Schon per Definition ist eine Assoziation eine vom Menschen gezogene Verbindung zwischen zwei Sachverhalten. Über Kookurrenzen lassen sich über Umwege ähnliche Ergebnisse erzielen. Echte Assoziationen, wie sie Menschen ziehen, werden jedoch nur ein Bruchteil der Ergebnisse darstellen. Will man schlechte Ergebnisse vermeiden, ist es unumgänglich, die Assoziationsdatenbank per Hand von Menschen füllen zu lassen. Dass dies auf gewöhnlichem Weg ein sehr großer, auch finanzieller, Aufwand wäre, zeigt sich alleine daran, dass es bisher keine derartige Datenbank gibt, obwohl ein Nutzen, vor allem im Bereich Internetwerbung, nicht von der Hand zu weisen ist. Daher wollen wir hier den Citizen Science Ansatz benutzen, um eine derartige Datenbank zu realisieren.

Ein weiterer Vorteil, das Projekt mit Citizen Science Ansatz zu bearbeiten, bietet die größere Streuung von Assoziationen. Wenn eine einzelner Nutzer eine Assoziation zu einem bestimmten Wort eingeben soll, wird diese sehr oft die gleiche, oder zumindest eine sehr ähnliche sein. Wenn eine große Menge Personen Assoziationen eingibt, werden die Datenbank mit einer größeren Auswahl von Zusammenhängen erfüllt. Stammen diese unterschiedlichen Personen auch noch aus sehr differenzierten Hintergründen, lokal und mit verschiedenen Interessen, so werden die Assoziationen sehr vielfältig. Ein Elektrotechniker wird sicherlich mit dem Begriff Halbleiter etwas anderes verbinden als ein Grundschullehrer. Ein Jugendlicher, der in einer Dorf am Meer groß geworden ist, wird vermutlich einen anderen Bezug zu Fisch haben, als ein Gleichaltriger aus einer Gebirgsstadt.

In den nachfolgenden Kapiteln wird erklärt, wie wir TIMA als Citizen Science Projekt umgesetzt haben.

2 TIMA als Citizen Science Projekt

In einem Citizen Science Projekt sind zusätzliche Schwerpunkte zu beachten als in einem gewöhnlichen Informatikprojekt. Die Nutzer sind keine bezahlten oder interessierten Angestellten, sondern zielt auf die breite Masse. Nur mit einer großen Anzahl Nutzer lassen sich die Menge Daten sammeln, die die Assoziationsdatenbank sinnvoll macht. So nutzen wir die Intelligenz der Masse, die wichtigsten Assoziationen zu einem bestimmten Wort zu finden. Es ergeben sich zwei Themen, die in den meisten anderen Projekten nicht so gewichtig sind.

In einem Citizen Science Projekt muss die **Datenqualität** überprüft werden. Da fast jeder Nutzer ein Laie ist, gibt er eventuell nicht immer ideale Werte ein. Dadurch, dass Nutzer freie Werte eintragen können, ist das Projekt auch gegen Vandalismus besonders anfällig. Zuletzt muss bedacht werden, dass die Nutzer in keiner Form bezahlt werden. Daher muss sich TIMA damit auseinander setzen, wie es Menschen dazu motiviert, etwas zum Projekt beizutragen. Dabei fokussieren wir uns auf **Spiele**, um einen Anreiz zum Mitwirken zu geben.

2.1 Spiele

Wir nutzen einen spielerischen Ansatz für das Eintragen von Assoziationen, um Menschen zur Mitarbeit bei TIMA zu motivieren.

In der ersten Spielvariante wurde dem Nutzer ein Wort aus der Datenbank übergeben und dieser sollte seine Assoziation dazu geben. Danach wurde gegebenenfalls das neue Wort in die Datenbank übertragen und die Assoziation gespeichert. Zuletzt erhielt ein registrierter Nutzer eine bestimmte Menge an Punkten. Auf der Webseite bieten wir das Einsehen der momentan besten Spieler an. Wir erhoffen uns durch diesen Wettbewerbsgedanken weitere Motivation für unsere Nutzer.

2.1.1 Wortselektierungsalgorithmus

In diesem Abschnitt wird die Funktionsweise des Wortselektierungsalgorithmus genauer beschrieben. In der Datenbank ist es wünschenswert, dass jedes Wort mindestens eine Assoziation hat und zusätzlich ein ungefähres Gleichgewicht an Assoziationen verteilt zwischen allen Worten herrscht. Der Wortselektierungsalgorithmus ist zuständig für das Auswählen der Worte, für die eine Assoziation gegeben werden

2 TIMA als Citizen Science Projekt

soll, damit diese Eigenschaften erreicht werden. Folgende Kriterien müssen erfüllt werden:

- Jeder Benutzer soll möglichst zu jedem Wort mindestens eine Assoziation geben.
- Wörter, die wenig assoziiert wurden, entweder insgesamt oder von einem einzelnen Benutzer, sollten für diesen Nutzer bevorzugt werden.
- Wörter sollen ausgeschlossen werden können, um z.B. zu verhindern, dass das selbe Wort mehrmals hintereinander kommt.

Der Algorithmus, der diese Anforderungen erfüllt, sieht folgendermaßen aus:

Listing 2.1: Wortselektierungsalgorithmus

```
1 w = [alle Wörter einer Sprache]
2 w = w - [alle auszuschließenden Wörter]
3 if ( Anonymer Benutzer )
4     w = [15 Wörter mit niedrigstem Häufigkeit aus w]
5 else
6     w = w - [alle auszuschließenden Wörter des Benutzers]
7     w = [15 Wörter mit niedrigstem Auftreten in der
          AssociationHistory des Benutzers1 aus w]
8 return [zufälliges Wort aus w]
```

2.1.2 ExcludeWords

Falls ein Benutzer zu einem Wort keine Assoziation einfällt, hat er die Möglichkeit das Wort zu überspringen. Damit er nun nicht in kurzer Zeit wiederholt danach gefragt wird (vgl. Abschnitt 2.1.1), wird das Wort auf eine Ausschlussliste gesetzt. Einträge die älter als 7 Tage sind, werden automatisch von der Liste gelöscht. Danach wird der Benutzer erneut nach seiner Assoziation zu diesem Wort gefragt.

2.1.3 Assoziationskette

Dieses Spiel sollte für einen Nutzer deutlich interessanter sein, als zu beinahe kontextfreien Worten hintereinander Assoziationen einzutragen. Voraussetzung für dieses Spiel ist jedoch eine bereits gefüllte Assoziationsdatenbank. Das Spiel Assoziationskette funktioniert folgendermaßen:

Die Spieler assoziieren nacheinander auf die jeweilige Assoziation ihres Vorgängers. Dadurch bildet sich eine Kette mehrerer Assoziationen bis zu bestimmten Abbruchbedingungen. Ein mögliches Ende stellt das Assoziieren eines schonmal verwendeten

2 TIMA als Citizen Science Projekt

Wortes dar und die damit verbundene Schließung eines Kreises in der Kette. Das Spiel ist ebenfalls beendet, sollte einem Spieler keine weitere Assoziation zeitnah einfallen. Bei einem Computergegner wird so lange wie möglich versucht die Abschlusskriterien nicht zu erfüllen, damit der Spieler selbst in den meisten Fällen dafür verantwortlich ist.

Alle gegebenen Assoziationen des Spielers werden hierbei mit Punkten belohnt und für die Datenbank verwendet. Die Gesamtlänge der Kette wird nochmal extra als Bonus berechnet, hat jedoch keine Auswirkungen auf die Datensätze in der Datenbank.

Das Spiel Assoziationskette steht als Variante mit einem Computergegner momentan sowohl auf der Webseite, als auch als App zur Verfügung. Der Computer wählt dabei seine Assoziation zufällig aus den 10 am meisten gewählten Assoziationen aus der Datenbank zu dem gegebenen Wort.

2.2 Newsletter

Ein weiteres Feature über das TIMA verfügt, ist der Newsletter. Falls ein Benutzer möchte, kann er auf der Webseite Wörter auswählen, die ihn interessieren und wöchentlich darüber einen Newsletter erhalten, in dem zu jedem Wort die Assoziationen mit deren Häufigkeit enthalten sind. Durch die regelmäßige Erinnerung könnte sich ein Nutzer motiviert fühlen, erneut an einem Spiel für TIMA teilzunehmen.

2.3 Datenqualität

Um die Datenqualität in der Datenbank zu sichern, setzen wir auf drei Dinge:

1. Rechtschreibprüfung
2. Mengenrelevanz
3. Nutzermanagement

Rechtschreibung

Die naheliegendste Überprüfung, um die Datenqualität zu gewährleisten, ist eine Rechtschreibkontrolle. Dabei wird ein eingegebenes Wort mit einer Datenbank abgeglichen, die aus mehreren Wortschätzen der entsprechenden Sprache zusammengefügt ist. Momentan ist dies nur für Deutsch umgesetzt und besteht aus folgenden Datenbanken:

Befindet sich ein eingegebenes Wort nicht in dieser Datenbank, wird der Nutzer vor der entgeltlichen Abgabe gefragt, ob er diese Eingabe wirklich tätigen möchte.

2 TIMA als Citizen Science Projekt

Bestätigt er dies, bekommt er für diese Assoziation wie gehabt Punkte, jedoch wird das Wort nicht in die Datenbank übertragen. Wir gehen davon aus, dass Worte, die nicht in der Verbindung dieser schon für sich genommen umfassenden Datenbanken zu finden sind, in der Assoziationsdatenbank vernachlässigbar sind.

Mengenrelevanz

Mit einer Rechtschreibprüfung lässt sich natürlich nicht verhindern, dass ein Nutzer falsche oder unrelevante Daten eingibt. Ob eine Assoziation falsch ist, kann man natürlich von außerhalb nicht entscheiden. Jeder Mensch hat eigene Assoziationen zu bestimmten Worten.

Wir betrachten eine Assoziation als richtig, wenn sie im Allgemeinen nachvollziehbar ist. Gibt ein Nutzer jetzt eine Assoziation ein, die diesem Anspruch nicht genügt, ob mutwillig oder nicht, kann man davon ausgehen, dass andere Menschen diese Assoziation nicht geben. In der Datenbank wird eine falsche Assoziation also statistisch irrelevant.

Nutzermanagement

Ein weiteres mächtiges Werkzeug zur Vermeidung von Vandalismus ist die Nutzerkontrolle. In der Datenbank werden alle Assoziationen gespeichert, die ein Nutzer jemals gegeben hat. Sollte ein Nutzer auffällig werden, zum Beispiel, weil er innerhalb sehr kurzer Zeit sehr viele Punkte in Spielen erreicht, kann er überprüft werden. Sollte bei einer solchen Überprüfung auffallen, dass seine Assoziationen sehr oft nicht dem allgemeinen Verständnis entsprechen, können diese Eintragungen aus der Datenbank gelöscht werden. Dieser Vorgang ist momentan nicht automatisiert.

3 Assoziationsdatenbank und API

Das Ziel von TIMA ist das Erstellen einer Datenbank, in denen Assoziation gespeichert werden. Daher liegt ein Schwerpunkt unserer Arbeit darin, diese zu Erstellen und zu Befüllen. Die Datenbank ist direkt verknüpft mit einem Webfrontend, das durch die Bereitstellung einer umfassenden API der Hauptanlaufpunkt für die Benutzer und die Apps ist.

In diesem Kapitel werden das Backend der Webseite und die Datenbank beschrieben. Dabei wird genauer auf die Designentscheidungen eingegangen, die zum Aufbau der einzelnen Datenbankbestandteile geführt haben.

3.1 Backend und Datenbank

Für das Backend der Website haben wir uns für Django als grundlegende Bibliothek entschieden. Bei Django handelt es sich um ein in Python geschriebenes Webframework, das dem Model-View-Controller-Schema folgt. Django bietet unter anderem einen sehr komplexen objektrelationalen Mapper, der es ermöglicht auch komplexe Objektstrukturen abzubilden ohne die verwendete Datenbank explizit zu kennen. Neben allen notwendigen Funktionen gewährleistet Django zusätzlich also gute Wiederverwendbarkeit und wurde deshalb für das Erstellen des Backends genutzt.

3.1.1 Datenmodell

In Abbildung 3.1 ist das komplette Datenmodell von TIMA dargestellt. Das Modul `associations.models` spielt dabei die Schlüsselrolle. Hier werden die grundlegenden Daten für die Assoziationsdatenbank gespeichert: die Worte und deren Verknüpfungen.

Das Modell `Word` speichert einzelne Wörter und das Modell `Association` die Assoziationen zwischen diesen Wörtern. Für jedes Wort wird gespeichert, wie oft für dieses nach einer Assoziation gefragt wurde. In ähnlicher Weise besitzt auch jede Assoziation in der Datenbank eine Häufigkeit, die angibt wie oft die Assoziation von Nutzern eingegeben wurde. Um eine Unterscheidung zwischen verschiedenen Sprachen zu ermöglichen, repräsentiert das Modell `Language` die verfügbaren Sprachen. Existiert ein Wort in mehreren Sprachen oder wird in verschiedenen Sprachen genutzt, hat es für jede Sprache einen eigenen Eintrag.

3 Assoziationsdatenbank und API

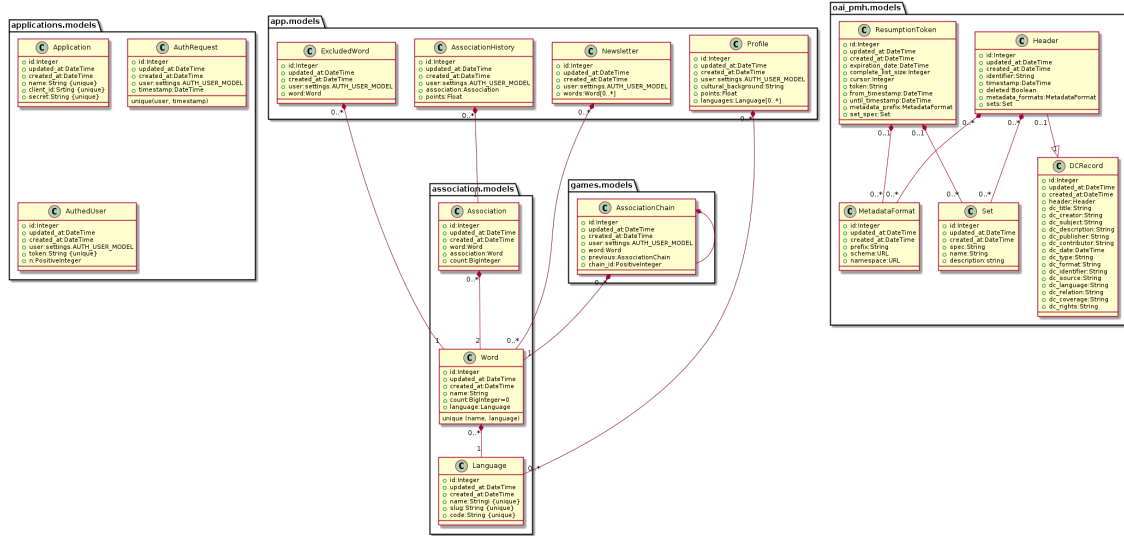


Abbildung 3.1: UML des TIMA Datenmodells

Das Modul `games.models` enthält Modelle die für die verschiedenen Spiele wichtig sind. Dies ist im Moment nur AssoziationsKette (vgl. Abschnitt 2.1), hierfür werden in dem Modul `AssociationChain` die letzte beziehungsweise aktuelle Assoziationskette eines Benutzer gespeichert. Diese wird beim Start eines neuen Spieles gelöscht.

Um grundlegende Funktionen des Benutzermanagements zu ermöglichen, wurden die Modelle des Moduls `app.models` eingeführt. Das Modell `Profile` speichert grundlegende Informationen zu jedem Nutzer, zum Beispiel die Punktzahl und die Sprachen, für die ein Benutzer assoziiert hat. Diese Daten werden in einem Ranglistensystem genutzt, das die Nutzer motivieren soll, sich gegenseitig zu messen. In dem Modell `AssociationHistory` wird die gesamte Assoziationsgeschichte eines Benutzers gespeichert, mit den jeweils für eine Assoziation erhaltenen Punkte. Somit können im Falle eines Misbrauchs die gegebenen Assoziationen aus der Datenbank gelöscht werden und dem Nutzer die Punkte entzogen werden. Das Modell `ExcludeWord` enthält für jeden Benutzer die Wörter, die er innerhalb der letzten sieben Tage übersprungen hat (vgl. Abschnitt 2.1.2). Das letzte Modell in diesem Modul speichert für jeden Benutzer welche Worte er in seinem Newsletter empfangen möchte.

Für die Kommunikation zwischen App und Backend, insbesondere der Autorisierung der Schreibzugriffe auf die Datenbank (vgl. Abschnitt 3.2) dient das Modul `applications.models`. Das Modell `Application` speichert die Apps, die Autori-

3 Assoziationsdatenbank und API

siert sind, mit den nötigen Daten für die Autorisierung (vgl. Abschnitt 3.2.2). Die beiden anderen Modelle dieses Moduls **AuthRequest** und **AuthedUser** speichern die nötigen Information für einen Benutzer der sich authentifizieren möchte oder sich bereits authentifiziert hat. Durch dieses Modul wird also gewährleistet, dass nur Nutzer auf die Datenbank zugreifen können, die eine von uns autorisierte Anwendung nutzt.

Das letzte Modul und die beinhalteten Modell sind für das OAI-PMH erforderlich siehe dafür (vgl. Abschnitt 3.2.3).

3.2 API

Damit verschiedene Apps mit der TIMA Datenbank kommunizieren können, haben wir uns entschieden eine umfangreiche API zu implementieren. Diese lässt sich grob in drei Teile gliedern. Zum einen gibt es die Anfragen, die keiner Autorisierung bedürfen, zweitens jene die einer Autorisierung erfordern und drittens eine OAI-PMH Schnittstelle.

Eine komplette Dokumentation der API ist in der Datei `API.md`¹ im git zu finden.

Im folgenden Abschnitt werden die einzelnen API Anfragen erläutert, zuerst die autorisierungsfreien, dann jene, die eine Autorisierung benötigen. Zum Schluss wird dann noch ein Abschnitt zu OAI-PMH folgen.

3.2.1 Nicht autorisierte Anfragen

Die API-Anfragen, die keiner Autorisierung bedürfen sind allgemeine Anfragen, an die Assoziationsdatenbank, die auch über die Webseite ohne eine Anmeldung erfolgen können.

Rangliste Eine dieser Anfragen ist die nach der Rangliste. Es werden keine weiteren Angaben benötigt und als Antwort kommt ein JSON-Object, das eine Liste der Benutzer enthält, mit den gleichen Daten wie sie über die Webseite einsehbar sind.

Statistik Ebenso ist die Statistik über die API abfragbar. Diese enthält aktuelle Zahlen über Nutzerzahl, Wortmenge und Assoziationen in der Datenbank.

Sprachen Es kann eine Liste aller Sprachen in TIMA angefordert werden, hier ist neben dem Namen, der Sprach-Code in der Antwort enthalten, der bei vielen anderen Anfragen als Parameter angegeben werden muss.

¹<https://github.com/Tima-Is-My-Association/TIMA/blob/master/API.md>

Wörter Um entweder ein einzelnes Wort oder eine Liste von Wörtern ist diese Anfrage bestimmt. Es können optional Wort-IDs, Sprache oder ein Limit für die Anzahl der Assoziationen pro Wort angegeben werden. Das JSON-Objekt der Antwort enthält unter anderem zu jedem Wort einen Link zur Website des Wortes, ein Link zu dieser Anfrage mit der Auswahl auf das einzelne Wort und der OAI-PMH identifier des Wortes.

3.2.2 Autorisierte Anfragen

Autorisierte Anfragen sind notwendig, um Schreibzugriff auf die Datenbank zuzulassen. Außerdem ist eine Authentisierung notwendig, um Nutzer der App oder der Webseite eindeutig identifizieren zu können. Für die Rangliste ist das unumgänglich. Aus diesem Grund war es erforderlich, dass einige API Anfragen einer Autorisierung bedürfen.

Um dies zu realisieren haben wir uns zunächst bestehende Frameworks wie zum Beispiel OAuth2 angeschaut und getestet in wie weit diese unseren Anforderungen genügen. Dies hat allerdings zu keinem zufriedenstellenden Ergebnis geführt, weswegen wir entschieden haben dies selbstständig zu implementieren.

Die grundlegenden Anforderungen die wir dabei hatten sind wie folgt:

1. Sichere Authentisierung einer App
2. Sichere Authentisierung eines Benutzers
3. Sicherstellen, dass spätere Anfragen von einem autorisierten Benutzer kommen

In Abbildung 3.2 ist der Authentisierungsprozess schematisch dargestellt. Der Client ist dabei eine App, über die sich ein Nutzer authentisieren möchte. Die App verfügt über eine `client_id` und über ein `secret`, beides von TIMA vergebene eindeutige zufällige Strings. Der Authentisierungsprozess läuft wie folgt ab:

1. Eine App sendet eine Anfrage an TIMA mit dem `username` des Benutzers und der `client_id`. TIMA prüft diese beiden Werte auf Existenz und antwortet entweder mit **200** (HTTP Response Code) und dem aktuellen Zeitstempel oder mit **404**.
2. Als nächstes sendet die App die eigentliche Authentisierungsanfrage. Mit `username` und `password` des Benutzers, `client_id` der App, dem Zeitstempel der Antwort der letzten Anfrage und einem `token` das aus dem `secret` der App und dem Zeitstempel generiert wird (SHA512).
3. TIMA antwortet wenn die Authentisierung erfolgreich war mit **200** und den folgenden drei Werten:

3 Assoziationsdatenbank und API

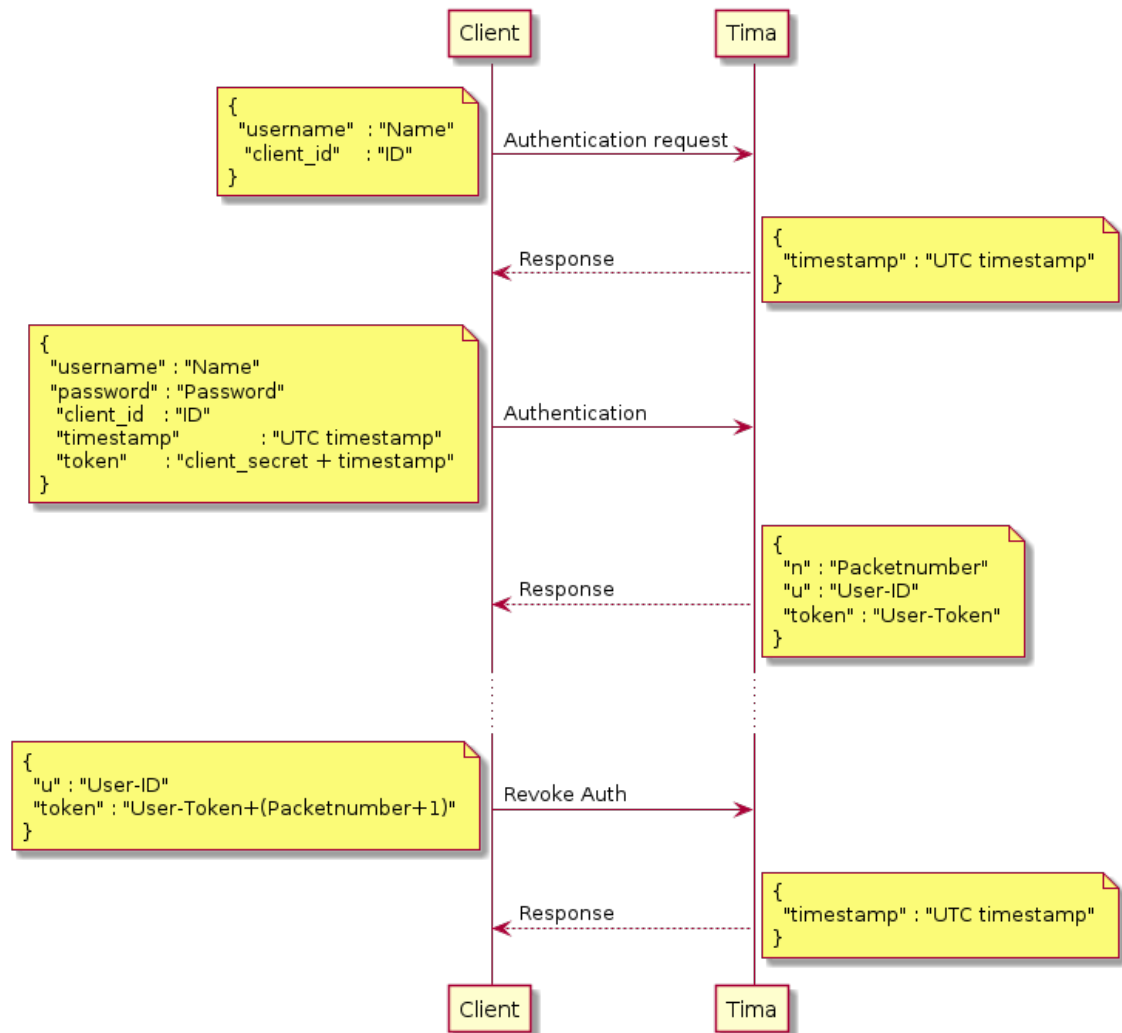


Abbildung 3.2: Authentisierungsprozess

n Paketnummer - jede Anfrage einer App muss diese um eins nach oben zählen. Als Wertebereich ist uint32 zubenutzen.

u eine eindeutige Benutzer-ID, die bei jeder Anfrage mit zusenden ist

token ein zufälliger String, der bei jeder Anfrage zusammen mit n, der Paketnummer, in einem SHA512 Hash zu senden ist

Aus Kompatibilitätsgründen läuft die Paketnummer im wertebereich uit32, also bis maximal 2147483646, und fängt nach erreichen der Maximalzahl wieder bei 0 an. Dies führt jedoch bei momentanen Nutzerzahlen zu keinen Problemen und wird mit sehr hoher Wahrscheinlich auch in Zukunft kein Thema werden.

3.2.3 OAI-PMH

Bei OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) handelt es sich um ein auf XML basierendes Protokoll zum Sammeln von Metadaten. Es wird dabei unterschieden zwischen Data Providern und Service Providern. Ein Data Provider betreibt ein oder mehrere Repositories, die OAI-PMH unterstützen um Metadaten bereitzustellen. Service Provider sammeln die Metadaten der Data Providern und bieten Mehrwertdienste an.

Wir haben das OAI-PMH Protokoll implementiert um Metadaten zu den gesammelten Assoziationen bereitzustellen. Wir stellen dabei Metadaten zu den Wörtern und im geringen Maße zu den Benutzern bereit.

4 Applikationen und Webseite

Die Applikationen und die Webseite sollen die Verwendung vom TIMA für möglichst viele Endnutzer möglich machen.

4.1 Webfrontend

Das Webfrontend ist die Hauptanlaufstelle für Benutzer. Hierüber kann er sowohl anonym als auch angemeldet Assoziationen eingeben, Wörter und deren Assoziationen ansehen und weitere Funktionen wie die Rangliste und andere Statistiken aufrufen.

Das Webfrontend basiert auf Django, wurde zusätzlich zu HTML mit Bootstrap und JQuery erstellt, sowie zur Visualisierung D3.

4.2 Applikation

Die Applikationen sollen die Verwendung von TIM ohne Webbrowser ermöglichen. Durch Applikationen wird besonders für mobile Geräte die Nutzung vereinfacht. In einer ersten Variante der App wurden grundlegende Funktionen der Webseite nachgebaut. Dies diente auch der Entwicklung der API, um etwaige Fehler im Protokoll oder Verbesserungen dessen aufzuzeigen und zu beheben.

4.2.1 Framework

Als Framework wurde sich für Qt5 entschieden aufgrund der weitreichenden Unterstützung des Frameworks auf verschiedenen Endgeräten. Hier ist besonders darauf hinzuweisen, dass Qt5 sowohl auf Android als auch auf iOS läuft und so nicht die gleiche App für beide Betriebssysteme geschrieben werden muss.

4.2.2 Aufbau

Die innere Logik wird durch einen Zustandsautomaten dargestellt um Mehrfachanfragen zu vermeiden und eine einfache Fehlerkorrektur zu ermöglichen. In Abbildung

4.1 wird der Zusammenhang der einzelnen Zustände angezeigt. Das wechseln der Zustände wird ausschließlich über die Signale geregelt, die mit Qt implementiert sind.

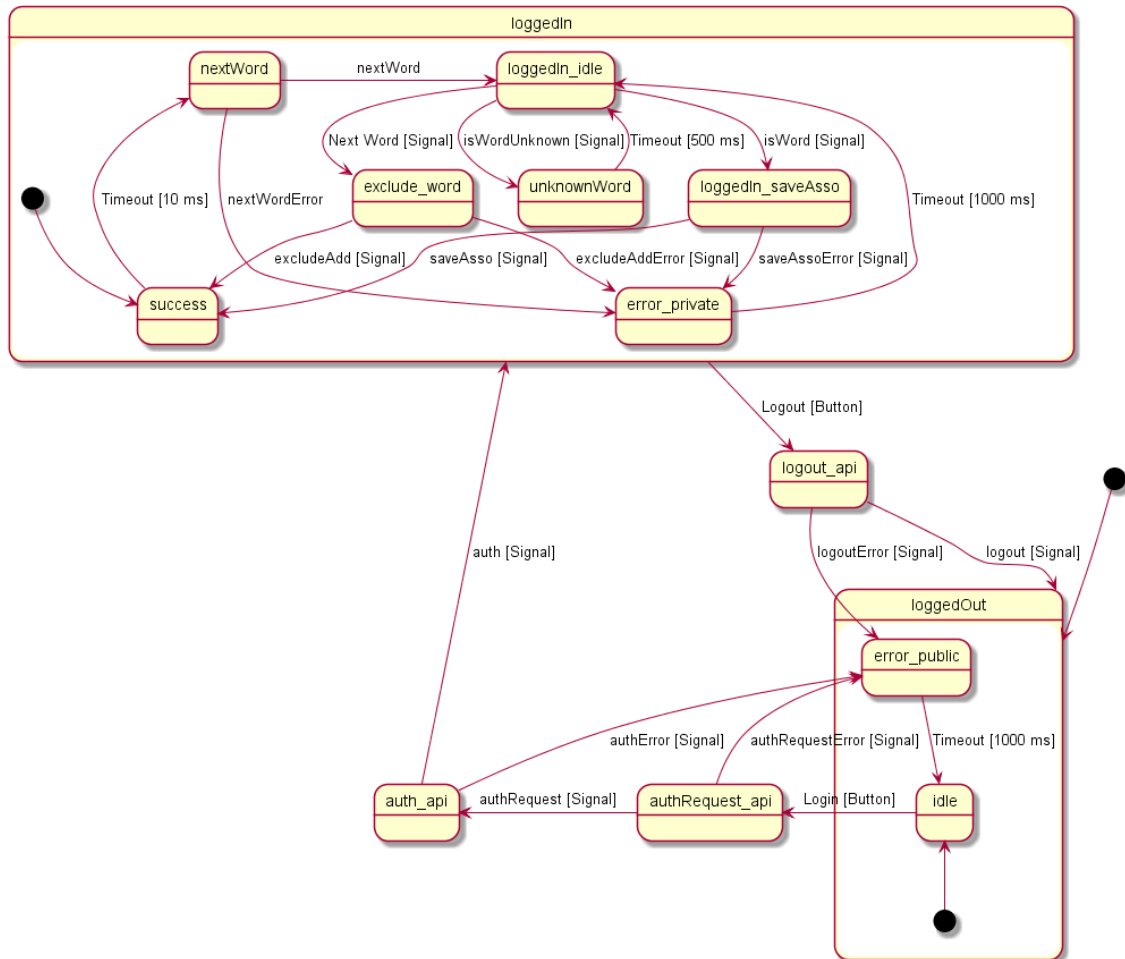


Abbildung 4.1: UML State Diagramm des Applikationszustands-
automaten

4.3 Sicherheit

Die Sicherheit hat bei der Entwicklung eine große Rolle gespielt. Jede App und die Webseite nutzt die in 3.2.2 vorgestellte Autorisierungsmethode. Dies dient dazu, lediglich von TIMA akzeptierte Applikationen Schreibrechte für die Datenbank zu geben. Das Auslesen der Informationen bleibt davon jedoch unangetastet, sofern es

4 Applikationen und Webseite

sich nicht um benutzerspezifische Daten handelt, und ist nach wie vor für jeden offen.

5 Ausblick

Die bisher verrichtete Arbeit bietet sehr gute Voraussetzungen für zukünftige Projekte, die entweder mit den Assoziationen arbeiten, oder die Assoziationsdatenbank erweitern möchten. Durch strikte Trennung von Front- und Backend ist einfaches Austauschen beider Bestandteile jederzeit möglich. Um die Datenbank weiter zu füllen, waren einige weitere Spiele beziehungsweise Funktionen für bestehende Spiele geplant, die jedoch aus Zeitgründen nicht weiter verfolgt wurden.

Assoziationskette Momentan fehlt der Assoziationskette die Möglichkeit, kooperativ mit anderen menschlichen Gegnern zu spielen. Diese könnte in Zukunft implementiert werden. Eine interessante Spielart könnte dabei sein, zufälligen Mitspielern zugewiesen zu werden. Dies könnte auch ein Computergegner. Nach dem Spiel mit einem zufälligen Spieler den Spieler zu fragen, ob er mit einem Menschen oder einem Computer gespielt hat, ist eine interessante Abwandlung des Turingtests, der auch über die Qualität unserer Assoziationsdatenbank entscheiden kann.

Familienduell Das Spiel Familienduell ist momentan nicht implementiert, bietet jedoch großen Unterhaltungswert. Der Name soll hierbei nur die Ausrichtung des Spieles näher bringen, denn wie bei der Fernsehserie², werden dem Benutzer verschiedene verdeckte Antworten auf eine Frage gezeigt und für jede richtig gegebene Antwort erhält der Spieler Punkte. Die Fragen sind jedoch im Unterschied zum Fernsehen, ausschließlich die meist genannten Assoziationen zu einem bestimmten Wort. Zusätzlich sollte eine Zeitbegrenzung eingehalten werden, da ja alle Korrekten Antworten auf TIMA nachgeschaut werden können und so einem möglichen Betrug entgegen gewirkt werden kann. Je nach Schwierigkeitsgrad kann ein Zeitbonus für korrekte Assoziationen gegeben werden und ein Malus bei falschen Antworten. Die Einwirkung auf das Spielvergnügen müsste entsprechend getestet werden.

Jede gegebene Antwort sollte für das füllen der Datenbank verwendet werden, obgleich es eine gesuchte Lösung war oder nicht. Punkte sollte ein Spieler jedoch nur für richtige Lösungen und einen Bonus wenn er alle Lösungen findet erhalten.

Andere Sprachen Wünschenswert wäre auch eine bessere Umsetzung aller Komponenten in andere Sprachen. Zum aktuellen Zeitpunkt ist eine Rechtschreibprüfung

²siehe <https://de.wikipedia.org/wiki/Familien-Duell>

5 Ausblick

nur für Deutsch eingefügt und sowohl die Webseite als auch die App sind nur auf Deutsch und Englisch verfügbar.

Analysen der Daten Mit den vorhandenen Daten lassen sich sicherlich auch viele Analysen durchführen. Zum Beispiel wäre ein Vergleich zwischen Kookurrenzen und Assoziationen möglich. Doch nicht nur im Bereich der automatischen Sprachverarbeitung lassen sich sicherlich interessante Beobachtungen feststellen. Wie sich der Hintergrund des Nutzers auf seine Assoziationen auswirken, ist ein spannender Faktor.

Wir wünschen uns in der Zukunft viele Projekte, die TIMA unterstützen, weiter aufbauen und natürlich nutzen.

6 Zusammenfassung

Das Projekt TIMA besteht inzwischen aus der Datenbank mit dazugehöriger API. Diese setzt für Autorisierung auf ein eigenes Protokoll. Um Nutzer für das Projekt zu gewinnen, setzen wir vor allem auf Spiele und Wettbewerbcharakter durch Ranglisten. Wir haben eine Applikation implementiert, die das Spiel Assoziationskette auf mobilen Geräten ermöglicht. Die Webseite von TIMA bietet dieses Spiel ebenfalls an und zusätzlich noch verschiedene andere Funktionen, wie die Statistik über Nutzer, Sprachen, Worte Assoziationen und die Rangliste der besten Assoziierer.

Momentan befinden sich über 3000 Worte mit mehr als 4000 Assoziationen in der Datenbank. Verteilt sind diese auf vier verschiedene Sprachen, Deutsch, Englisch, Spanisch und Farsi. Die am stärksten vertretene Sprache ist mit großem Abstand Deutsch. Dies rührt sicherlich daher, dass TIMA bisher keine große Verbreitung gefunden hat. Die bisherigen zwölf angemeldeten Nutzer stammen vor allem aus dem Bekanntenkreis der Programmierer. Für die Zukunft gibt es viele verschiedene Anwendungsmöglichkeiten für TIMA, die den Nutzerkreis deutlich erweitern können.