



DHBW Mannheim

Team „Norbert“

Norbert - Your StudyBuddy

Softwareentwurf

13. April 2016

Projektleitung:

Projektmitglieder:

Arwed Mett

Dominic Steinhauser, Tobias Dorra,
Simon Oswald, Philipp Pütz

Inhaltsverzeichnis

1	Grundlegende Komponenten und deren Funktionsweise	1
1.1	Grundlegender Aufbau aus Funktionssicht	3
1.2	Datenhaltung	5
1.3	Interne Prozesse	6
2	Frontend	8
2.1	Ziele	8
2.2	Views	8
2.3	Newsfeed	9
2.4	Profilverwaltung und Einstellungen	9
2.5	Komponenten der Anwendung	10
2.6	Trennung von Daten und Anzeigelogik	10
3	RESTful API	12
3.1	Übersicht der Funktionen	12
3.2	Authentifizierung	14
3.3	Einträge	14
3.4	Notifikationen	14
4	Backend	16
4.1	Übersicht	16
4.2	RESTfull API	17
4.3	Import von Informationen aus externen Systemen	17
4.4	Datenanalyse	17
5	Datenbank	19
5.1	Datenbankschema	19
6	Anbindung externer Dienste - Dropbox	23
6.1	Autorisierungsprozess (OAuth)	23
6.2	Dropbox Endpoints	24
7	Code Conventions	26
8	Exception Policy	28

1 Grundlegende Komponenten und deren Funktionsweise

Norbert hat zur Aufgabe Informationen / Daten innerhalb eines Kurses zu teilen. Jeder Benutzer hat einen Newsfeed in dem diese Informationen angezeigt werden. Die Informationen werden von Newsfeeds der Benutzer und externen Diensten ermittelt. Anschließend muss es eine zentrale Logik geben, die diese „Daten“ klassifiziert und den Benutzern neue Daten vorschlägt. (Siehe Abbildung 1.1)

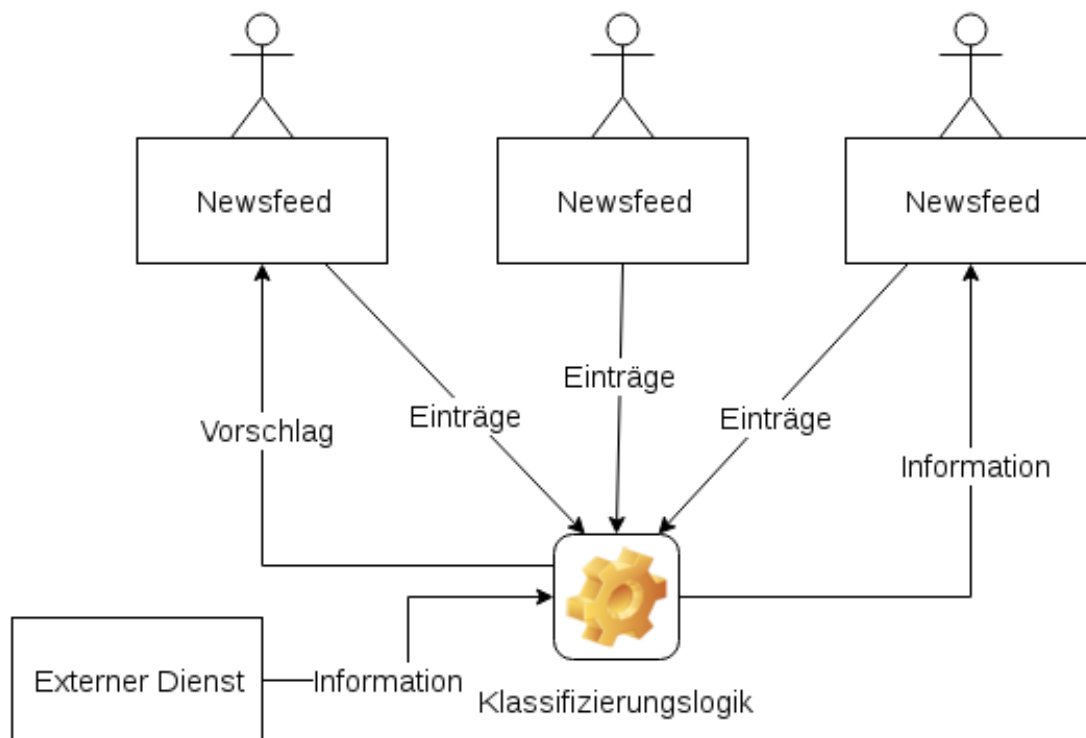


Abbildung 1.1: Funktionsweise des Newsfeeds

Eine grundlegende Anforderung dafür ist das sammeln von Daten. Diese können durch Studenten oder externe Dienste angelegt werden.

Diese Daten werden wenn sie von einem Benutzer erstellt wurden als Einträge bezeichnet und wenn sie von einem externen Dienst angelegt wurden als Informationen bezeichnet.

Bei Verwendung der Anwendung werden dem Benutzer seine Einträge angezeigt.

Jetzt hat der Benutzer die Möglichkeit neue Einträge zu erstellen und hinzuzufügen, oder er bekommt durch die Anwendung neue Einträge vorgeschlagen, welche im folgenden als Vorschlag bezeichnet werden.

Die Daten von einem externen Dienst werden als Information gespeichert und den Benutzern angezeigt.

Es gibt also insgesamt 3 verschiedene Arten von Daten mit denen ein Benutzer interagiert welche in Tabelle 1.2 zu sehen sind.

Eintrag	Vorschlag	Information
Ein Eintrag dient zum notieren von Daten. Er kann mehrere Eigenschaften haben, wie z.B. eine Erinnerung. Ein Eintrag kann von einem Benutzer bearbeitet werden.	Die Anwendung analysiert ständig die Einträge aller Benutzer. Wenn erkannt wird, dass sich ein Benutzer für einen Eintrag eines anderen Benutzers interessieren könnte wird ihm dieser als „Vorschlag“ vorgeschlagen.	Wenn durch einen externen Dienst Daten produziert werden, wie z.B. eine E-Mail an den Kursverteiler versendet wird, werden diese Daten als „Information“ allen Benutzern vorgeschlagen. Der Benutzer hat dann die Möglichkeit diese Informationen zu ausblenden, wenn er sich nicht dafür interessiert.

Tabelle 1.2: Elemente des Newsfeeds

Da alle Benutzer nicht immer eine Verbindung zu der Anwendung haben, jedoch Daten zwischen den Benutzern ausgetauscht werden, wird für die Architektur eine Client-Server Architektur gewählt. Dadurch können Daten zentral auf einem Server gespeichert bzw. klassifiziert werden und anschließend an die Clients verteilt werden. Diese Architektur hat auch den Vorteil, dass die Daten zentral ausgewertet werden können.

Folgende Funktionalitäten müssen von der Architektur zwingend erfüllt werden, damit Norbert korrekt funktionieren kann:

1. Der Newsfeed muss auf dem Client angezeigt werden können.
2. Norbert muss die gesamten Daten (Einträge, Vorschläge, E-Mails. etc.) zentral speichern.
3. Die für den Newsfeeds relevanten Daten müssen von Norbert an die Clients geliefert werden können.

4. Der Client muss Einträge anlegen/bearbeiten können und auf Vorschläge reagieren können, bzw. seine Aktionen Norbert mitteilen können.
5. Norbert muss ToDo-Vorschläge liefern können.

1.1 Grundlegender Aufbau aus Funktionssicht

Die in Grafik 1.2 gezeigte Architektur würde die oben genannten Kriterien erfüllen.

1. Über einen Webbrowser kann der Newsfeed auf den Clients angezeigt werden.
2. Norbert hat die gesamten Daten in getrennten Dateien/Ordern zur Verfügung.
3. Über die Kommunikationsschnittstelle kann Norbert die relevanten Daten an den Client liefern.
4. Über die Kommunikationsschnittstelle können die Aktionen der Clients Norbert mitgeteilt werden.
5. Die Vorschlagslogik berechnet anhand der vorhandenen Daten Vorschläge, welche über die Kommunikationsschnittstelle an den Client geliefert werden.

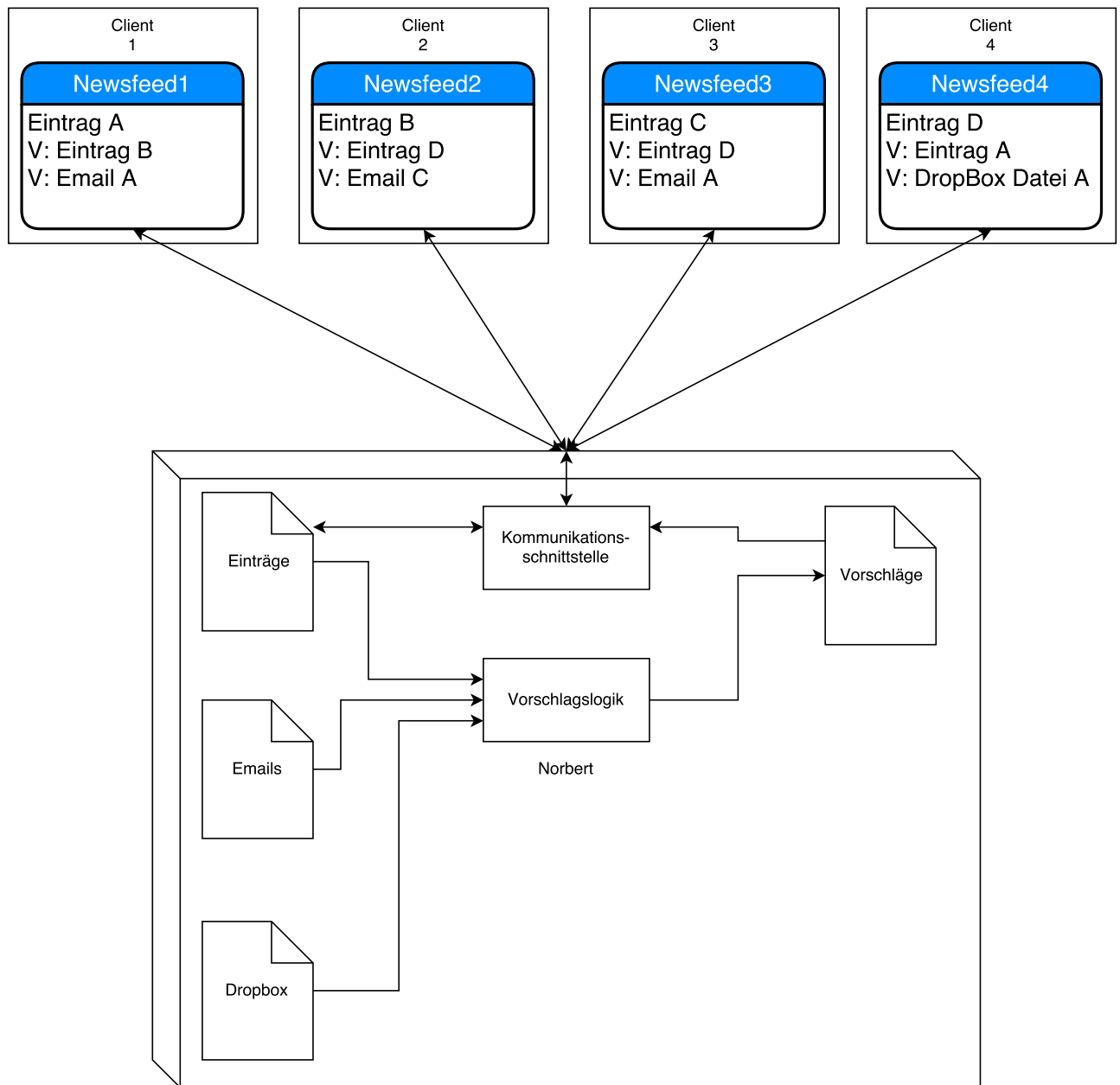


Abbildung 1.2: Grundlegende Funktionalität

Mit dieser Architektur treten aber mehrere Probleme auf:

1. Das manuelle Verwalten von Daten in separaten Dateien ist sehr aufwendig und extremst fehleranfällig.
2. Der Ablauf interner Prozesse ist nicht definiert und es ist unklar, wie dieser gesteuert wird

1.2 Datenhaltung

Das erste Problem kann durch die Verwendung eines Datenbanksystems gelöst werden. Durch eine Datenbank können die Daten zentral verfügbar gemacht werden. Außerdem ist der Zugriff über ein Datenbanksystem weniger fehleranfällig und in der Regel auch effizienter, da die meisten Datenbanksysteme bereits von Haus aus Performanceoptimierungen wie Puffer etc. mitbringen.

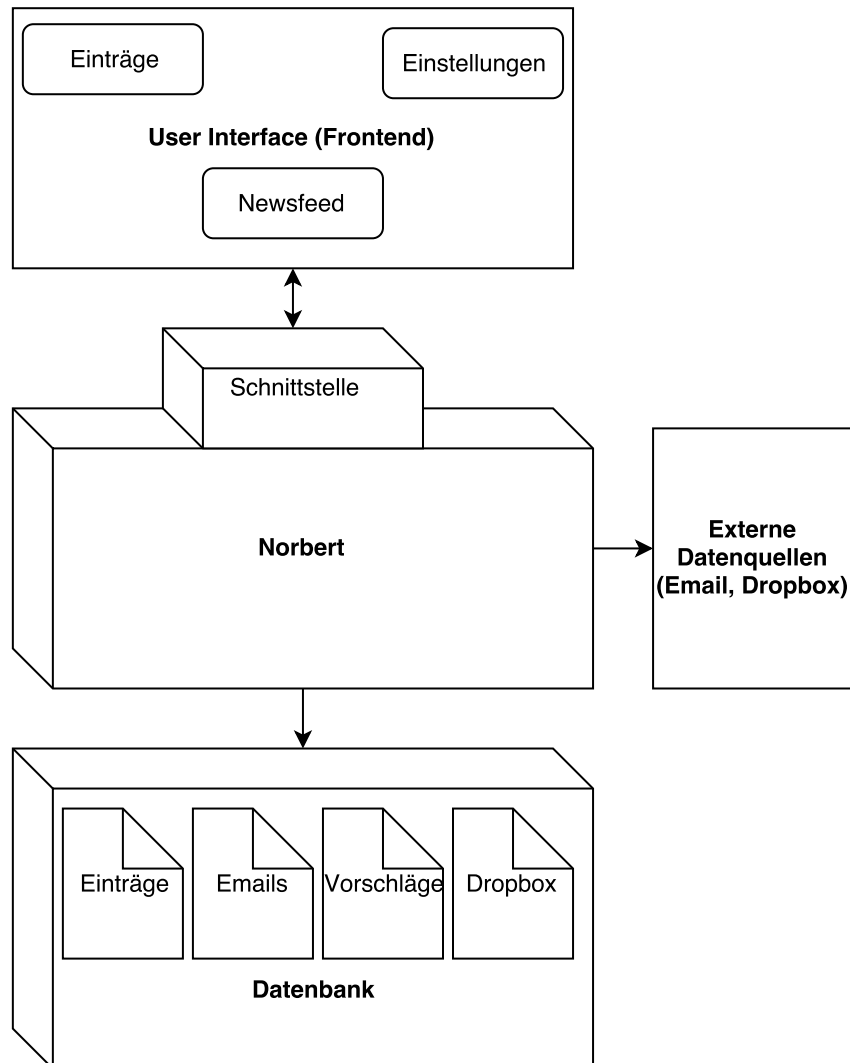


Abbildung 1.3: Kapselung der Daten in einer Datenbank

1.3 Interne Prozesse

Das zweite Problem kann durch eine Komponente zur Aufgabenverwaltung gelöst werden. Anstatt alle Aufgaben in Echtzeit zu bearbeiten, werden bestimmte Aufgaben, die besonders performancelastig sein können, zyklisch bearbeitet. Die Aufgabenverwaltung lässt dieses Aufgaben zyklisch durch den Core, also die zentrale Logik- und Rechenkomponente, abarbeiten. Das Erstellen/Bearbeiten von Einträgen, das Liefern von Daten an den Client, sowie das Akzeptieren von Vorschlägen erfolgt weiterhin in Echtzeit. Vorschläge werden allerdings nur noch in bestimmten Intervallen berechnet, außerdem erfolgt die Indizierung der angebundenen Dropbox zyklisch. Andere Aufgaben, die nicht zwingend in Echtzeit ausgeführt werden müssen, können auch zyklisch ausgeführt werden. Die in Abbildung 1.4 gezeigte Architektur erfüllt die genannten Anforderungen und löst die in Abbildung 1.2 auftretenden Probleme. Die genauere Funktionsweise einzelner Komponenten wird in den folgenden Kapiteln beschrieben.

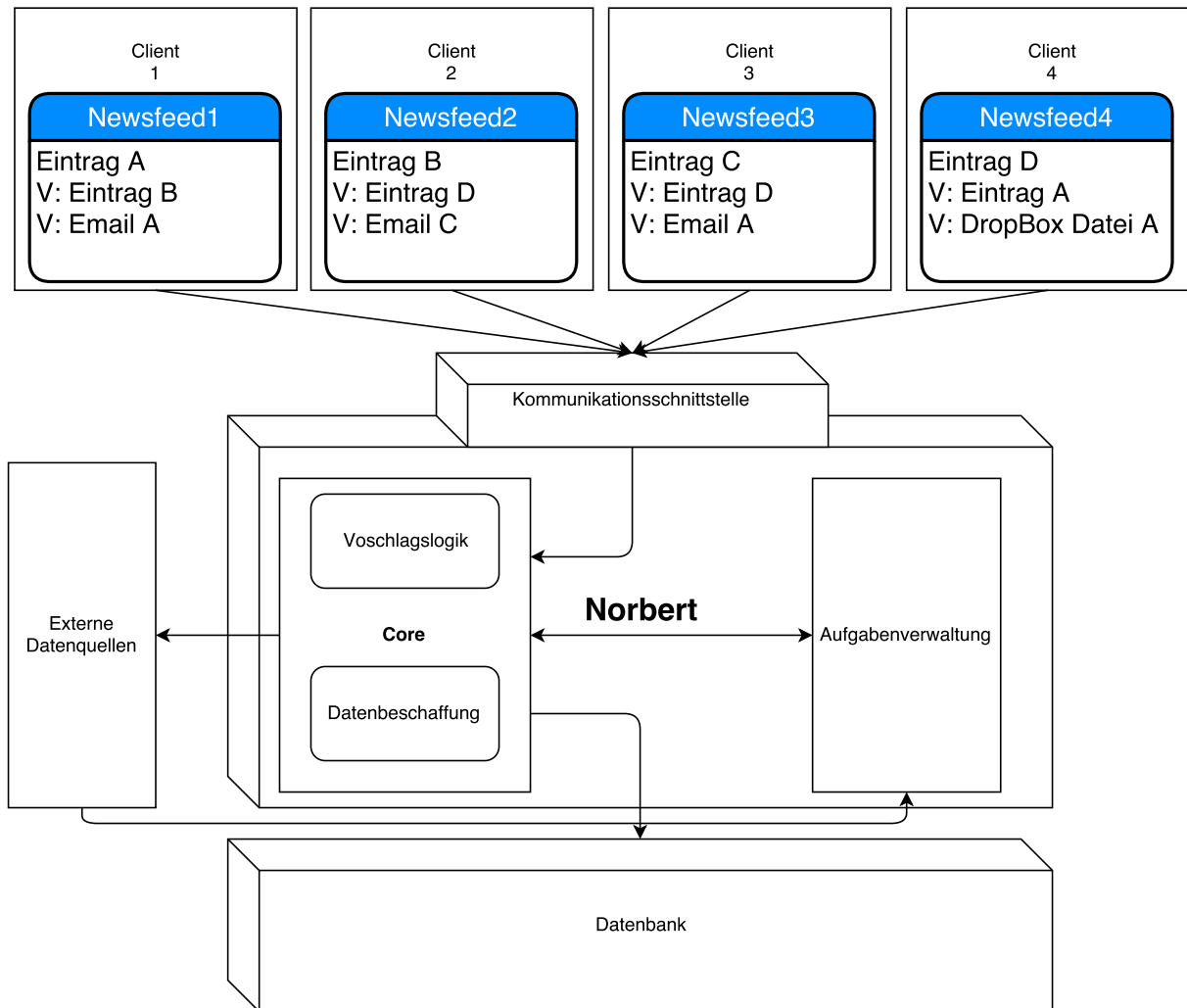


Abbildung 1.4: Kapselung der internen Prozesse

2 Frontend

2.1 Ziele

ID-Kürzel	Beschreibung
FZ-10	Einzelne Teile des Frontends können getrennt voneinander entwickelt werden.
FZ-20	Es gibt eine Trennung von Daten- und Anzeigelogik.
FZ-30	Das Frontend ist leicht um neue Funktionen erweiterbar.

Tabelle 2.1: Ziele der Frontend-Architektur

2.2 Views

Da die Anwendung auf mehreren Plattformen laufen soll wird die Applikation als HTML5 App programmiert. Dabei besteht das Frontend aus folgenden Views:

ID-Kürzel	Beschreibung	Route
F-10	Newsfeed	/
F-20	User Login	/login
F-30	User Registrieren	/register
F-40	Profilverwaltung und Einstellungen	/profile
F-50	Detailansicht von Einträgen	/
F-60	Adminansicht	/administration
F-70	View zur Anzeige von Suchergebnissen	/search-results

Tabelle 2.2: Einsatzbereiche

2.3 Newsfeed

Der Newsfeed ist eine Kollektion aus Einträgen, Informationen und Vorschlägen. Diese werden nach Relevanz sortiert. Dabei kann das Frontend die Einträge abfragen. Der Server ist verantwortlich für die Sortierung der Einträge.

2.3.1 Bearbeiten des Newsfeeds

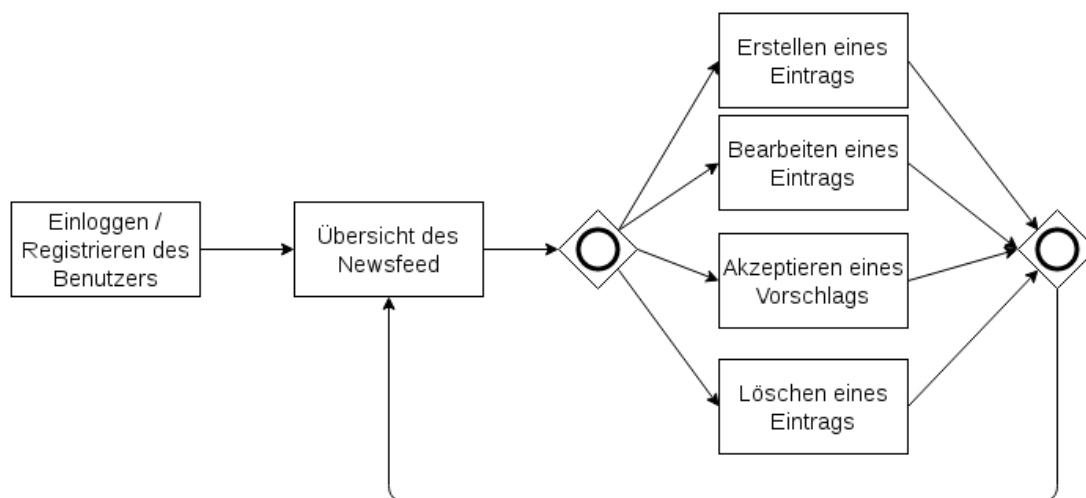


Abbildung 2.1: Ablauf von Dialogen

Ein Benutzer muss sich immer Authentifizieren bevor er seinen Newsfeed sieht. Anschließend erhält er eine Übersicht seiner Einträge. Jetzt hat er die Möglichkeit Einträge anzulegen, verändern, löschen oder einen Vorschlag in seinen Newsfeed aufzunehmen. Wenn ein Eintrag erstellt oder bearbeitet wird öffnet sich ein Dialog in dem der entsprechende Eintrag bearbeitet werden kann. Anschließend wird der neue Newsfeed dargestellt.

2.4 Profilverwaltung und Einstellungen

In der Profilverwaltung und den Einstellungen kann der Benutzer sein Passwort, Username oder E-Mail ändern. Außerdem kann eingestellt werden, ob der Benutzer an Einträge erinnert werden möchte oder nicht (Notifikationen).

2.5 Komponenten der Anwendung

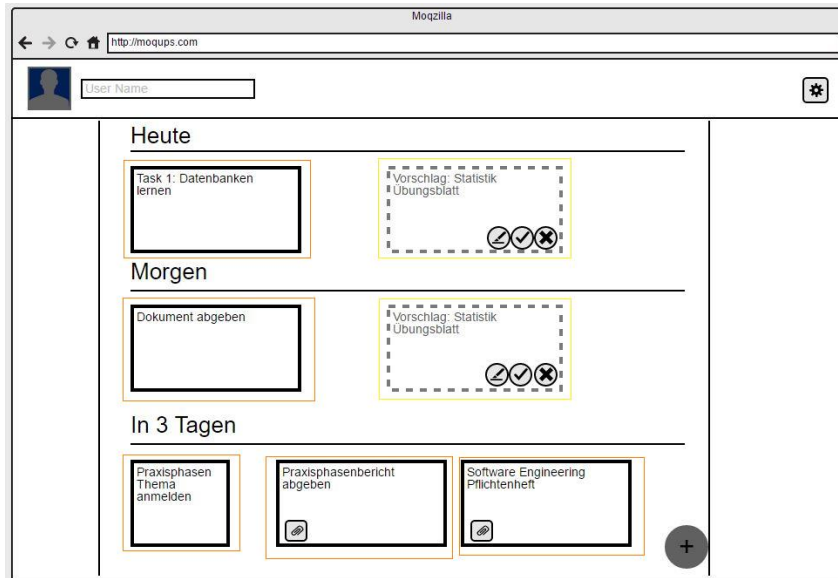


Abbildung 2.2: Komponenten des Newsfeeds

Wie in Abbildung 2.5 zu sehen ist besteht die Anwendung aus mehreren Einträgen und Vorschlägen. Diese UI-Elemente können als Komponenten angesehen werden. Deshalb soll das UI auf React.js basieren, da dort jedes UI-Element eine Komponente ist, die einen Zustand hat und mehrere Aktionen ausführen kann. Der Zustand beinhaltet immer die anzuzeigenden Daten. Wird eine Aktion wie z.B. ein Klick-Event ausgelöst ändert sich der Zustand und die Komponente wird neu gerendert.¹ Dies erleichtert die Entwicklung der Anwendung im Team und erhöht die Wiederverwendung von UI-Komponenten.

2.6 Trennung von Daten und Anzeigelogik

Eines der Ziele ist es Daten und Anzeigelogik getrennt von einander zu entwickeln. Deshalb können nicht alle Daten in den UI-Komponenten gespeichert werden. Möchten z.B. mehrere Komponenten auf die selben Daten zugreifen wird ein extra Lager (Store) für diese Daten benötigt.

Dazu wird Flux als ein Architektur Pattern eingesetzt. Die Grundbestandteile dieser Architektur sind Komponenten, Storages, Actions und ein Dispatcher. Dabei dienen die Komponenten zur Interaktion mit dem User, stellen somit also den „View“ der Anwendung dar. Die Komponenten können Actions auslösen, welche einfache Funktionen sind die ausgeführt werden.

¹Siehe <https://facebook.github.io/react/>

Diese Actions senden ein Objekt mit einer ID an den Dispatcher, welcher die Objekte an alle Stores verteilt. Ein Store ist ein Ort zum lagern von Daten. Jeder Store stellt eine Methode „handleAction“ zur Verfügung. Diese wird von dem Dispatcher aufgerufen und das Objekt der Action übergeben. Anschließend kann der Store aufgrund der in der Action definierten ID seine Daten aktualisieren. Jetzt können sich Komponenten bei dem Store registrieren um über Änderungen informiert zu werden. Geschieht dies wird der View der Komponente erneuert.²

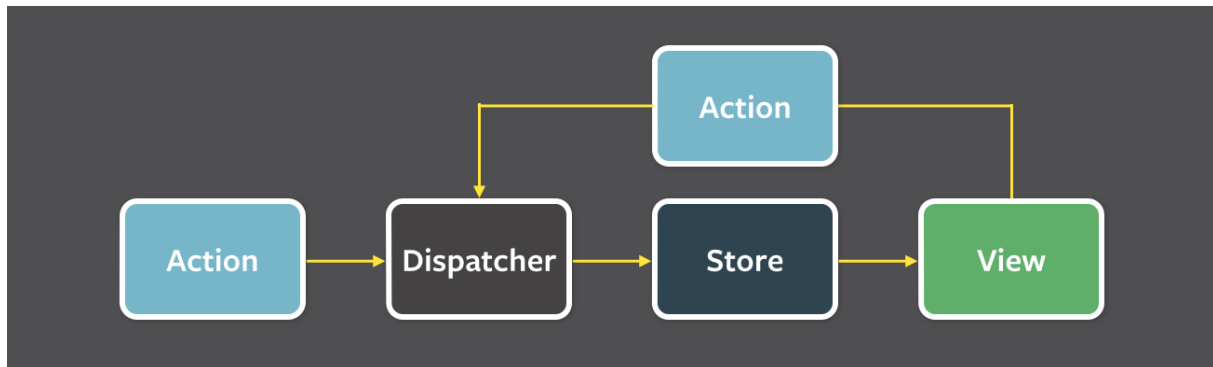


Abbildung 2.3: Flux

<https://facebook.github.io/flux/docs/actions-and-the-dispatcher.html#content>

Diese Architektur unterstützt das Teilen eines Zustandes mehrerer Komponenten, da sich mehrere Komponenten auf einen Storage registrieren und Änderungen an dem Zustand der Anwendung immer global über eine Action ausgeführt werden. Somit können alle Stores ihre Daten bei spezifischen Aktionen aktualisieren. Dadurch wird vermieden, dass sobald sich ein Eintrag ändert eine extra Methode geschrieben werden muss welche den Newsfeed ändert.

²Vgl. <https://facebook.github.io/flux/>

3 RESTful API

Test Die RESTful API basiert auf JSON und ist die zentrale Schnittstelle zwischen den Client Anwendungen (frontend) und dem Backend. Sie basiert auf HTTP-Requests, weshalb alle Anfragen vom Client gesteuert werden.

3.1 Übersicht der Funktionen

Im folgenden werden alle Funktionen der RESTful Schnittstelle aufgelistet.

ID	Methode	Pfad	Beschreibung
R-10 Userverwaltung			
R-10.1	POST	/users/	Erstelle einen Benutzer. Es müssen ein „username“ und ein „password“ übergeben werden.
R-10.2	PUT	/users/:userId	Bearbeiten eines users.
R-10.3	DELETE	/users/:userId	Lösche einen User.
R-10.4	GET	/users/login	Authentifizierung durch session. „username“ und „password“ müssen übergeben werden.
R-10.5	GET	/users/logout	Beendigung der Session.
R-20 Newsfeed			

Tabelle 3.1: RESTful Pfade

ID	Methode	Pfad	Beschreibung
R-20.1	GET	/newsfeed/:userid	Hole alle Einträge für den Nutzer basierten Newsfeed. Zurückgegeben wird ein Array aus Einträgen. Der Parameter start definiert das Start-Datum und der Parameter count definiert wieviele Newsfeedelemente zurückgegeben werden sollen. Der Parameter count kann auch negativ sein.
R-20.2	POST	/entry/	Füge einen Eintrag zum Newsfeed hinzu. Der Parameter heißt „entry“.
R-20.3	GET	/entry/:id	Gibt alle Daten zu dem Eintrag mit der id wieder.
R-20.4	DELETE	/entry/:id	Löschen eines Eintrags.
R-20.5	PUT	/entry/:id	Updaten eines Eintrags.
R-20.6	GET	/newsfeed/recommendation/:userid	Hole alle Empfehlungen für einen Benutzer.
R-20.7	DELETE	/newsfeed/recommendation/:id	Löschen eines Vorschlags.
R-20.8	DELETE	/newsfeed/information/:id	Löschen einer Information.
R-30 Notifikationen			
R-30.1	GET	/notifications/:userid/	Holen aller Notifikationen.
R-40 Suche			

Tabelle 3.1: RESTful Pfade

ID	Methode	Pfad	Beschreibung
R-40.1	GET	/search/	Es werden alle Einträge anhand des Parameters „query“ durchsucht. Das Ergebnis ist eine Liste von Einträge und Informationen zurückgegeben.

Tabelle 3.1: RESTful Pfade

3.2 Authentifizierung

Ein Benutzer darf lediglich die zu ihm gehörenden Einträge bearbeiten. Bei allen Funktionen außer dem Login und der Registrierung muss der Benutzer authentifiziert sein. Die Authentifizierung wird durch Sessions realisiert.

3.3 Einträge

Einträge werden durch die RESTful API als JSON-Objekt zurückgegeben.

Sie können folgende Struktur haben:

```
1 {  
2   "title": "Eintrag",  
3   "tags": ["tag1", "tag2"],  
4   "components": [...(Objects)]  
5 }
```

3.4 Notifikationen

Eine Notifikation ist ein JSON-Objekt mit folgender Struktur:

```
1 {  
2   "date": int (UNIX timestamp),  
3   "entryId": int  
4 }
```


Mit dieser Information kann das Frontend die Notifikation zu einem bestimmten Zeitpunkt aufrufen. Über die „entryId“ kann der passende Eintrag angefordert werden.

4 Backend

Damit das Frontend auf zentral verwaltete Daten zugreifen kann, wird serverseitig die entsprechende Logik benötigt. Diese bildet das Backend und hat die folgenden Aufgaben:

1. Ausliefern des Frontends an den Benutzer.
2. Zentrale Verwaltung aller Einträge, Informationen, Vorschläge und Benutzer.
3. Importieren und Aufbereiten von Informationen aus externen Systemen wie Dropbox, dem E-Mail Verteiler oder dem Kurskalender.
4. Verwaltung von Dateien, die Teil eines Eintrags oder einer Information sind.
5. Generieren von Vorschlägen basierend auf den Interessen der Nutzer.
6. Durchsuchbar machen von Einträgen und Informationen.
7. Anbieten von Diensten zur Datenabfrage und -Manipulation über die RESTfull API.

4.1 Übersicht

Ein großer Teil der Aufgaben, die das Backend zu erfüllen hat, beinhaltet die Kommunikation mit anderen Systemen wie zum Beispiel Dropbox oder dem Kurskalender. Wenn Anfragen an solche externe Systeme gestellt werden, vergeht teilweise sehr viel Zeit, bis diese antworten. Um diese Wartezeiten effektiv nutzen zu können, soll beim Entwurf des Backends verstärkt auf asynchrone Techniken gesetzt werden. Daher wird für die Entwicklung des Backends die Programmiersprache Javascript verwendet, die im Bereich der asynchronen Programmierung ihre Stärken hat.

Das Backend kann grundsätzlich in drei große Teile zerlegt werden:

1. **Die Kommunikationsschnittstelle** ist für die Kommunikation mit den Clients verantwortlich. Sie muss das Frontend in Form von HTML-, CSS- und Javascript-Dateien ausliefern. Außerdem muss die RESTfull API bereitgestellt werden, über die das Frontend mit dem Backend kommuniziert.
2. **Der Core** enthält die Implementation der zentralen Geschäftslogik. Er muss die anfallenden Daten verwalten, Vorschläge generieren und die Suchfunktion bereitstellen. Der Core speichert seine anfallenden Daten in einer Datenbank.

3. **Die Anbindung externer Dienste** integriert Dropbox, den Kurskalender und den Emailverteiler des Kurses in das System.

4.2 RESTfull API

Als Interface zwischen dem Frontend und dem Backend dient eine RESTfull API. Ihr Aufbau wurde bereits im vorhergehenden Kapitel spezifiziert. Um bei HTTP Requests an die einzelnen Routen die richtigen Aktionen tätigen zu können, wird ein Router benötigt. Hierfür kommt die Bibliothek Express.js¹ zum Einsatz. Bei jedem eingehenden HTTP Request wird je nach URL eine passende Funktion ausgeführt. Diese kann dann auf den Core zugreifen, um die entsprechenden Aktionen zu tätigen oder Daten abzufragen. Anschließend wird die HTTP Response generiert und die Anfrage ist beendet.

4.3 Import von Informationen aus externen Systemen

Die externen Systeme wie Dropbox unterliegen nicht unserer Kontrolle. Auf mögliche Änderungen von APIs oder gar deren Deaktivierung haben wir keinen Einfluss. Daher sollten wir die Möglichkeit haben, auf solche Änderungen möglichst einfach zu reagieren: Die Anbindung jedes externen Systems sollte austauschbar sein. Außerdem sollte sie nicht über das gesamte Backend verteilt, sondern an einer einzigen Stelle im Code erfolgen. Die Anbindung externer Systeme ist daher strikt von der zentralen Geschäftslogik getrennt.

Jedes System, aus dem Daten importiert werden kann, ist ein „Information Provider“. Damit der Core alle Information Provider gleich behandeln kann, müssen sie alle ein gemeinsames, fest definiertes Interface implementieren. Dadurch werden die einzelnen Information Provider austauschbar. Jeder Information Provider muss die folgenden Funktionen implementieren:

void syncInformations(data_handle): Weist den Information Provider an, seine Informationen im System von Norbert mit seiner Datenquelle zu synchronisieren. Über das Objekt data_handle bekommt der Information Provider dafür Lese- und Schreibzugriff auf die nötigen Daten.

4.4 Datenanalyse

Zur Bereitstellung der Suchfunktion sowie zur generierung von Einträgen müssen die Daten zunächst analysiert werden. Er besteht aus den folgenden Schritten:

¹expressjs.com

1. **Vorverarbeitung** Es wird der Text aus den Einträgen und Informationen extrahiert. Dieser wird in seine einzelnen Worte zerlegt. Wörter ohne Bedeutung wie „ist“, „und“ oder „ich“ werden entfernt. Die Wörter werden normalisiert, indem die Wortendungen abgeschnitten werden. Jedem eigenständigen Wort wird eine eindeutige ID zugewiesen, die im Folgenden verwendet wird, um Wörter zu repräsentieren, um den Speicherverbrauch zu minimieren.
2. **Berechnung des TF-IDF Maßes** Das TF-IDF² Maß gibt für ein Wort und ein Eintrag/-Information zurück, wie relevant das Wort für diesen Eintrag ist. Dieser Wert muss für jeden Eintrag und jedes Wort berechnet werden.
3. **Clustering der Einträge/Informationen** Die Einträge werden basierend auf den berechneten TF-IDF-Werten mithilfe eines GMM (Gaussian Mixture Model) zu Clustern gruppiert. Jedes Cluster fasst dabei Einträge zusammen, die ähnlich zueinander sind. Das Ergebnis dieses Verfahrens ist eine Liste mit Clustern, sowie für jeden Eintrag und jedes Cluster eine Wahrscheinlichkeit, dass der Eintrag zu diesem Cluster gehört.
4. **Analyse der Nutzerinteressen** Über die vorhandenen Einträge jedes Benutzers wird ermittelt, wie sehr sich der Benutzer für welches der generierten Cluster interessiert. Das potenzielle Interesse des Benutzers an einem Eintrag ermittelt sich aus der Wahrscheinlichkeit, dass der Eintrag in Clustern befindet, die für den Benutzer interessant sind. Zusätzlich wird noch ein zeitlicher Faktor mit eingerechnet, der aktuelle Einträge als interessanter einstuft.

Der Datenanalyseprozess kann je nach Anzahl der Einträge und Informationen im System recht lange dauern. Daher wird er in regelmäßigen Abständen vom Scheduler im Hintergrund ausgeführt, sodass er den Serverprozess nicht blockiert.

Als Vorschläge ergeben sich die Einträge, die als besonders interessant für den Benutzer eingestuft wurden, sich aber noch nicht in seinem Newsfeed befinden.

Um die Suche umzusetzen, müssen die Einträge und Informationen nur nach der Summe der berechneten TF-IDF-Werte für die Suchterme sortiert werden.

Da Informationen nur über die Suche gefunden werden müssen, nicht aber als Vorschläge angezeigt werden können, müssen diese nur Schritte 1 und 2 durchlaufen. Einträge müssen jedoch den vollständigen Prozess durchlaufen.

²<https://de.wikipedia.org/wiki/Tf-idf-Ma%C3%9F>

5 Datenbank

Um die anfallenden Daten zentral speichern zu können, wird eine Datenbank benötigt. Als Datenbankmanagementsystem wird MongoDB verwendet. Die Gründe dafür sind, dass der dokumentenorientierte Ansatz von MongoDB sehr gut mit Javascript harmoniert und das Abspeichern von Entries sehr einfach macht.

MongoDB speichert die Daten in verschiedenen Collections. Diese nehmen eine ähnliche Rolle ein wie die Tabellen bei relationalen Datenbankmanagementsystemen. Eine Collection ist eine Sammlung von Objekten, die einen beliebigen Aufbau haben dürfen.

5.1 Datenbankschema

Die folgenden Collections werden von Norbert für die Geschäftslogikdaten verwendet:

1. **Users:** Speichert die Benutzer von Norbert und ihre Zugangsdaten.
2. **Entries:** Speichert die Entries, die von den Benutzern erzeugt wurden.
3. **Information:** Speichert die Informationen, die aus externen Systemen importiert wurden.

Users Jedes Objekt in der Users-Collection repräsentiert einen Benutzer von Norbert. Ein solches Objekt ist nach dem folgenden Schema aufgebaut:

```
1 {  
2   "id": 42,           # Eindeutige ID  
3  
4   "username": "benni", # Der Benutzername,  
5                       # der zur Anmeldung genutzt wird.  
6  
7   "name": "Ben Utzer", # Der volle Name des Benutzers  
8  
9   "password_hash": "xdjc", # Das gehashte Passwort  
10  "password_salt": "cjdx"  # und der Salt, der zum hashen des Passworts  
11                          # verwendet wurde.  
12 }
```

Entries Jeder Entry wird als ein Objekt in der Entries-Collection repräsentiert. Ein solches Objekt besteht aus einem Titel, diversen Metadaten und einer Liste mit Components. Ein Entry-Objekt ist nach dem folgenden Schema aufgebaut:

```
1 {
2   "id": 123,           # Eindeutige ID
3
4   "created_at": 1234,  # Unix-Timestamp des Erstellzeitpunktes
5                       # des Entries.
6
7   "dirty": false,      # Das dirty-Flag gibt an, ob sich
8                       # das Entry geändert hat.
9                       # In diesem Fall muss es beim nächsten
10                      # Durchlauf des Schedulers neu
11                      # verarbeitet werden (Suchindex, Vorschläge!).
12
13   "owned_by": [42, 43], # Ein Array mit den IDs der Benutzer,
14                      # denen der Eintrag gehört.
15                      # Alle diese Benutzer sehen den Eintrag in
16                      # ihrem Newsfeed und können ihn bearbeiten.
17
18   "hidden_for": [42]    # Eine Teilmenge des "owned-by" Feldes,
19                      # die angibt, welche Benutzer den Entry
20                      # in ihrem Newsfeed ausgeblendet haben.
21                      # Über die Suche ist er dann immernoch
22                      # auffindbar.
23
24   "private": false,     # Wenn der Eintrag als privat markiert
25                      # wurde, kann er von anderen Benutzern
26                      # nicht über die Suche gefunden werden
27                      # und wird niemandem vorgeschlagen.
28
29   "equality_group": 123, # Bei neuen Entries wird der Wert
30                      # "equality_group" gleich der id des
31                      # Entries gesetzt. Wird der Entry beim
32                      # akzeptieren einer Recommendation geklont,
33                      # so hat der neue Eintrag zwar eine neue id,
34                      # behält jedoch die equality_group.
35                      # Die so entstehenden Äquivalenzklassen an
36                      # ähnlichen Entries werden dafür genutzt,
37                      # in der Suche und bei den Vorschlägen
38                      # duplikate zu vermeiden.
39 }
```

```

40  "title": "Überschrift",    # Der Titel des Entries
41
42  "tags": ["C", "Übung"],    # Ein Array mit den Tags, die der
43                               # Benutzer für den Entry vergeben hat.
44
45  "components": [           # Ein Array mit den Komponenten des Entry
46    { /* Component 1 */ },
47    { /* Component 2 */ },
48    { /* Component 3 */ }
49  ]
50 }

```

Jedes Objekt im components-Array ist dabei nach dem folgenden Schema aufgebaut:

```

1  {
2    "type": "text",          # Der Typ der Komponente
3
4    "generated": false       # Gibt an, ob die Komponente von Norbert
5                               # aus dem Inhalt extrahiert wurde.
6
7    "data": {                # Der eigentliche Inhalt der Komponente.
8      /* ... */              # der Aufbau des data-Objekts ist abhängig
9    }                        # von dem verwendeten Komponenten-Typ.
10 }

```

Information Jede Information wird als ein Objekt in dieser Collection gespeichert. Informationen sind grundsätzlich sehr ähnlich zu Entries aufgebaut. Ein Information-Objekt ist nach dem folgenden Schema aufgebaut:

```

1  {
2    "id": 123,               # Eindeutige ID
3
4    "created_at": 1234,      # Unix-Timestamp des Erstellzeitpunktes
5                               # der Information.
6
7    "dirty": false,         # Das dirty-flag gibt an, ob sich
8                               # die Information vom externen Service
9                               # geändert hat.
10                               # In diesem Fall muss sie beim nächsten
11                               # Durchlauf des Schedulers neu
12                               # verarbeitet werden (Suchindex!).
13
14    "hidden_for": [42, 12]   # Ein array mit den IDs der Benutzer, die

```

```
15         # die Information aus ihrem Newsfeed
16         # "gelöscht" haben. Bei diesen wird
17         # die Information nicht mehr angezeigt.
18         # Sie kann jedoch trotzdem noch über die
19         # Suche gefunden werden.
20
21     "title": "Überschrift",    # Der Titel der Information
22
23     "components": [          # Ein Array mit den Komponenten der Information
24         { /* Component 1 */ },
25         { /* Component 2 */ },
26         { /* Component 3 */ }
27     ]
28 }
```

Die Komponen-Liste der Information hat das selbe Format, wie die Komponenten-Liste eines Entries.

6 Anbindung externer Dienste - Dropbox

Eines der Ziele von Norbert - Your StudyBuddy ist es Wissensmanagement zu betreiben und somit Wissen zu sammeln und weiterzugeben. Da die einzelnen Anwender (DHBW-Kurse) verschiedene Cloudlösungen zur Speicherung von kursinternen Informationen sowieso schon besitzen, bietet Norbert ein allgemeines Interface, welches erlaubt die verschiedenen externe Dienste anzubinden. Da Informationen und Wissen von Studierenden meistens in Dropbox-Ordner abgelegt werden, wird in diesem Kapitel die technische Anbindung von Dropbox an Norbert beschrieben. Dabei wird zunächst erst allgemein auf den standardisierten Autorisierungsprozess „OAuth“ eingegangen und daraufhin die verschiedenen Dropbox Endpoints mit ihren Schnittstellen vorgestellt.

6.1 Autorisierungsprozess (OAuth)

OAuth ist ein offenes Protokoll, dass eine standardisierte und sichere API-Autorisierung für mobile Endgeräte, Webanwendungen und Desktop-Applikationen ermöglicht. Der Benutzer kann über den OAuth-Autorisierungsprozess (Abbildung 6.1) einer Anwendung (in diesem Fall Norbert) Zugriff auf gespeicherte Daten erlauben. Dabei loggt sich der Benutzer über die Applikation (Norbert) die Zugriff benötigt, auf dem Dienst, der die freizugebenden Daten speichert, ein (hier Dropbox) und erlaubt der Applikation den Zugriff. Da nicht jede Severinstanz von Norbert über eine Redirect-URL, die nach einem erfolgreichen Login aufgerufen wird, für den OAuth-Prozess verfügt und nicht jeder Severadministrator in der Dropbox Developer Console Norbert - Your StudyBuddy eintragen möchte, wird statt einem Redirect ein Autorisationscode angezeigt, den der Nutzer in Norbert eingeben kann. Norbert kann dann mit diesem Autorisationscode ein wiederverwendbares Access-Token anfragen. Dieses Access-Token dient zur Authentifizierung des Benutzers, sodass kein Benutzername und kein Passwort ausgetauscht oder gespeichert werden muss.

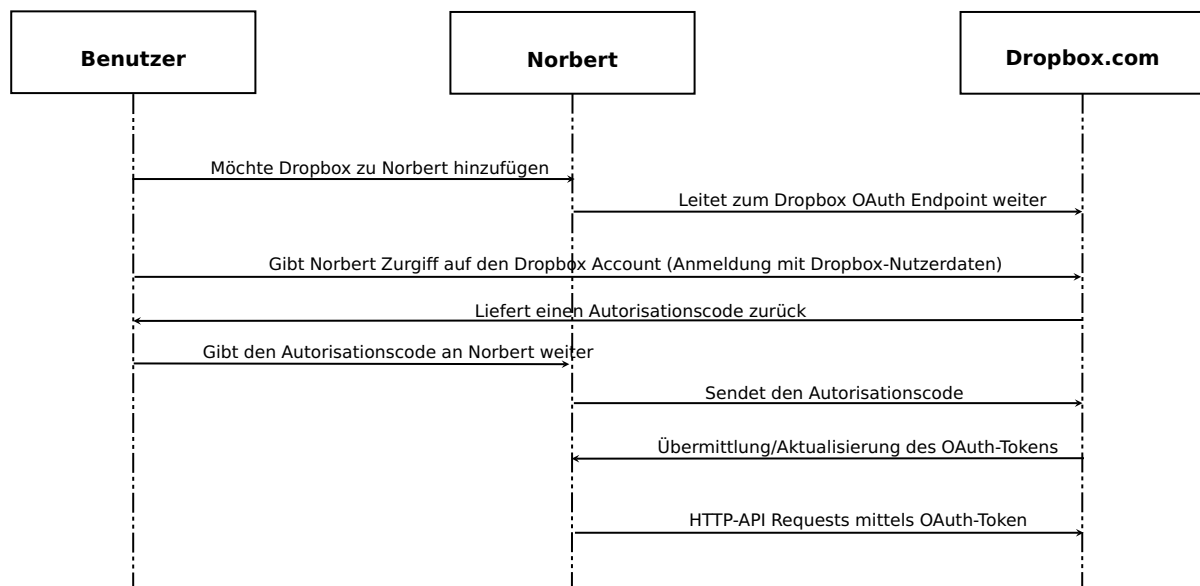


Abbildung 6.1: OAuth-Autorisierungsprozess zwischen Norbert und Dropbox.com

6.2 Dropbox Endpoints

Die Dropbox HTTP-API besteht aus mehreren Endpoints (Abbildung 6.2), die jeweils unterschiedliche Funktionen bereitstellen. Alle API-Aufrufe benötigen zur Autorisierung das individuelle Access-Token des Benutzers.

1. *dropbox.com*: Über diesen Endpoint/Webseite wird die OAuth-Autorisierung durchgeführt. Es wird nur einmalig zur Generierung des Access-Tokens ein Kontakt zu diesem Endpoint hergestellt.
2. *api.dropboxapi.com (RPC-Endpoint)*: Über die Domain *api.dropboxapi.com* können RPCs (remote procedure calls) ausgeführt werden. Dabei nimmt dieser Endpoint JSON-Strings im HTTP-Body entgegen und führt je nach aufgerufener URL verschiedene Funktionen aus. Ein RPC-Funktion ist beispielsweise `list_folder`¹, die als Parameter einen Pfad erwartet und dann alle im entsprechenden Ordner sich befindenden Dateien auflistet. Das Ergebnis der Funktion wird daraufhin wieder im JSON-Format zurückgesendet.
3. *content.dropboxapi.com (Download/Upload-Endpoint)*: Wie der Name schon sagt, können über den Download/Upload-Endpoint Dateien heruntergeladen oder hochgeladen werden. Die benötigten Meta-Daten (Pfad, Datei ID etc.) werden dabei im HTTP-Header als JSON oder URL-Argument übergeben. Funktionsergebnisse werden im JSON-Format im HTTP-Response Header übergeben und Dateien im HTTP-Body übersendet.

¹https://api.dropboxapi.com/2/files/list_folder

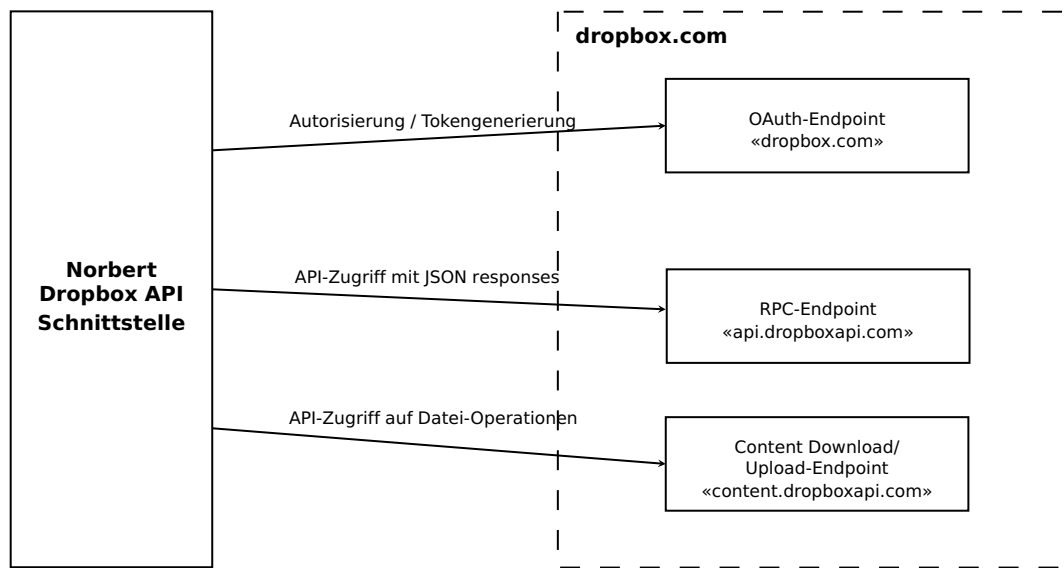


Abbildung 6.2: API-Zugriffe über die Dropbox-Endpoints

7 Code Conventions

Im folgenden werden die Code Conventions genannt, die während des Projekts eingehalten werden sollen, um einheitlichen Code zu erzeugen.

ID-Kürzel	Beschreibung	Begründung
C-10	englische Variablennamen	Es gibt sehr viel mehr englische Dokumentationen zu den einzelnen Programmiersprachen. Damit nicht immer ein Umdenken bzw. Übersetzen stattfinden muss, ist es einfacher die englische Sprache zu verwenden.
C-20	englische Kommentare	Damit der ganze Quelltext in einer Sprache verfasst ist, wird auch hier Englisch als Sprache verwendet.
C-30	Module	Voneinander unabhängige Programmteile (z.B. Klasse, Funktionen) werden getrennt, um die Wartbarkeit zu verbessern und eine bessere Übersicht zu gewährleisten.
C-40	Lokale Variablen	Falls immer es möglich ist, sollen lokale Variablen (Schlüsselwort: var) verwendet werden. Dies vermeidet Änderungen von außerhalb am Wert der Variable.

Tabelle 7.1: Code Conventions

ID-Kürzel	Beschreibung	Begründung
C-50	Wiederverwendbarkeit	Falls Funktionen wiederverwendet werden können, wird dies getan, um Redundanz zu vermeiden und die Wartbarkeit des Codes zu steigern.

Tabelle 7.1: Code Conventions

8 Exception Policy

Das folgende Kapitel beschreibt, welche Fehlerarten, während des Systembetriebs auftreten können.

ID-Kürzel	Beschreibung	Begründung
E-10	Falsche Benutzereingabe	Dem User wird mitgeteilt, dass seine Eingabe falsch ist und fordert diesen nach einer nochmaligen Eingabe auf.
E-20	Falscher Request an REST-API	Der Anwendung HTTP-Statuscode wird zurückgegeben.
E-40	Softwarefehler	Der Fehler wird in einer Log-Datei vermerkt.

Tabelle 8.1: Exception Policy