

px: PEAX Command Line Client

Wirtschaftsprojekt, Herbstsemester 2019

Patrick Bucher

26. Oktober 2019

Abstract

Die Software px ist ein Kommandozeilenprogramm, das den Zugriff auf die API von PEAX für Benutzer mit einem technischen Hintergrund vereinfachen soll. PEAX ist ein digitaler Posteingang, womit der Endanwender seine Post über verschiedene Kanäle empfangen und in einem komfortablen Web-Portal verwalten kann. Mit px soll der direkte Zugriff auf die RESTful API, d.h. ohne Verwendung des Webportals, vereinfacht werden, indem die Handhabung von Tokens (Access Token, Refresh Token) abstrahiert wird, und dem Benutzer eine Reihe intuitiv bedienbarer Befehle für das Ansteuern verschiedener End-points zur Verfügung gestellt wird.

Inhaltsverzeichnis

1	Problemstellung	4
1.1	Analyse des Projektauftrags	4
1.1.1	Endpoints	4
1.1.2	HTTP-Methoden	4
1.1.3	HTTP Status-Codes	5
1.1.4	OAuth 2.0	6
1.1.5	Umgebungen	7
1.1.6	Arten von Parametern	8
1.1.7	Benutzer	8
1.1.8	Betriebssysteme	9
1.1.9	Shells	9
1.2	Ausgangslage und Vorleistungen	9
1.3	Risikoanalyse	11
2	Stand der Praxis	12
2.1	Ansprechen der PEAX API	12
2.2	Kommandozeilenprogramme	12
3	Ideen und Konzepte	13
4	Methoden	14
4.1	Vorgehen	14
4.2	Teststrategie	14
4.2.1	Q1: automatisiert	15
4.2.2	Q2: automatisiert und manuell	15
4.2.3	Q3: manuell	17
4.2.4	Q4: Tools	17
5	Realisierung	19
5.1	Zwei-Faktor-Authentifizierung	19
5.2	Token-Store	19
6	Evaluation und Validierung	20
6.1	Rückmeldungen von Entwicklern	20
6.1.1	Sprint I	20

Inhaltsverzeichnis

7 Ausblick	22
8 Anhang	23
8.1 Systemspezifikation	23
8.1.1 Systemkontext	23
8.2 Technologie-Evaluation	24
8.2.1 Programmiersprache	24
8.2.2 Go	26
8.2.3 Rust	27
8.2.4 Entscheidung	27
8.3 Libraries	28
9 Einstiegsbeispiel	29
Literatur	31
Abbildungsverzeichnis	33
Tabellenverzeichnis	34
Verzeichnis der Codebeispiele	35

1 Problemstellung

Die Problemstellung setzt sich einerseits aus dem gestellten Projektauftrag (siehe Anhang) und andererseits aus der damit implizierten Umgebung (Systeme, Technologien, Benutze, etc.) zusammen (siehe dazu auch Abschnitt 8.1.1 *Systemkontext*, Seite 23).

1.1 Analyse des Projektauftrags

Der Projektauftrag beschreibt einen Command Line Client für eine RESTful-API. In diesem Zusammenhang gibt es Aspekte aus folgenden Bereichen zu analysieren:

Technologie das Protokoll HTTP und der Authentifizierungsmechanismus OAuth 2.0

Server-Umgebung die Umgebungen, die eine PEAX API anbieten

Client-Umgebung die Benutzer, ihr Betriebssystem und ihre Kommandozeile

1.1.1 Endpoints

Eine RESTful-API besteht aus einer Reihe sogenannter *Endpoints*, d.h Pfade zu Ressourcen, die abgefragt und/oder manipuliert werden können. Aus Platzgründen soll hier nicht auf einzelne Endpoints eingegangen werden. Beispielhaft erwähnenswert sind aber etwa der Token-Endpoint, bei welchem der Benutzer im Austausch seiner Credentials (Benutzername, Passwort und optionaler zweiter Faktor) ein *Token Pair* holen kann; und der Document-Endpoint, auf welchen Dokumente hochgeladen werden können.

1.1.2 HTTP-Methoden

Ein Endpoint kann über eine oder mehrere HTTP-Methoden angesprochen werden (R. Fielding & Reschke, 2014, Abschnitt 4.3). Im Kontext der PEAX API sind folgende Methoden relevant:

- GET: Erfragt eine Repräsentation einer bestimmten Ressource; greift nur schreibend auf diese zu.
- HEAD: Analog zu GET, es wird jedoch nur der Header und nicht der Body der Ressource angefragt.
- POST: Übermittelt eine Ressource zur Speicherung oder Manipulation einer bestehenden Ressource.

- PUT: Ersetzt eine bestehende Ressource durch den mitgeschickten Payload.
- DELETE: Löscht eine Ressource.
- OPTIONS: Beschreibt die Kommunikationsoptionen für eine bestimmte Ressource, wird etwa für CORS Pre-Flight Requests eingesetzt (Mozilla Developer Network, o. J.).
- PATCH Führt eine partielle Modifikation auf eine bestimmte Ressource aus. (Dusseault & Snell, 2010) Die Modifikation wird in der Form *JavaScript Object Notation (JSON) Patch* durchgeführt (P. Bryan, 2013).

1.1.3 HTTP Status-Codes

Eine Antwort auf eine HTTP-Anfrage enthält jeweils einen Status-Code (R. Fielding & Reschke, 2014, Abschnitt 6). Bei der PEAX API werden u.a. folgende Status-Codes häufig verwendet:

- 200 OK: Die Anfrage hat funktioniert.
- 201 Created: Die Anfrage hat funktioniert, und dabei wurde eine neue Ressource erzeugt.
- 204 No Content: Die Anfrage konnte ausgeführt werden, liefert aber keinen Inhalt zurück (etwa in einer Suche mit einem Begriff, zu dem keine Ressource gefunden werden kann).
- 204 Partial Content: Der zurückgelieferte Payload repräsentiert nur einen Teil der gefundenen Informationen. Wird etwa beim Paging eingesetzt.
- 400 Bad Request: Die Anfrage wurde fehlerhaft gestellt (ungültige oder fehlende Feldwerte).
- 401 Unauthorized: Der Benutzer ist nicht autorisiert, d.h. nicht eingeloggt im weitesten Sinne.
- 403 Forbidden: Der Benutzer ist zwar eingeloggt, hat aber keine Berechtigung mit der gewählten Methode auf die jeweilige Resource zuzugreifen.
- 404 Not Found: Die Resource wurde nicht gefunden; deutet auf eine fehlerhafte URL hin.

- 405 Method Not Allowed: Die Resource unterstützt die gewählte Methode nicht.
- 415 Unsupported Media Type: Das Format des mitgelieferten Payloads wird nicht unterstützt. In der PEAX API sind dies etwa Dokumentformate, die beim Hochladen nicht erlaubt sind (z.B. .exe-Dateien).
- 500 Internal Server Error: Obwohl die Anfrage korrekt formuliert und angenommen worden ist, kam es bei der Verarbeitung derselben zu einem serverseitigem Fehler.¹
- 380 Unknown: Dieser Status ist nicht Teil der HTTP-Spezifikation, wird aber nach einem Login-Versuch verwendet, wenn eine Zwei-Faktor-Authentifizierung (SMS, One Time Password) verlangt wird, und ist somit für die vorliegende Arbeit von hoher Relevanz.

1.1.4 OAuth 2.0

Im Hinblick auf das Wirtschaftsprojekt hat sich der Autor dieser Arbeit bereits im Vorsemester im Rahmen des Moduls *Computer Science Hot Topics* (INFKOL) mit dem Thema OAuth 2.0 befasst (Bucher & Christensen, 2019). Detaillierte Informationen zu OAuth 2.0 können diesem Paper und den dort zitierten Primärquellen entnommen werden.

An dieser Stelle sollen nur die Grundlagen beschrieben werden, die dann im Umsetzungsteil (siehe Abschnitt 5 *Realisierung*, Seite 19) bei Bedarf genauer eingegangen wird.

Bei einem Login-Vorgang mit OAuth 2.0 sendet der Benutzer seine Credentials (Benutzername, Passwort, optionaler zweiter Faktor wie SMS-Code) an den Token-Endpoint eines *Identity Providers* (IDP). Stimmen diese Angaben mit den Informationen auf dem IDP überein, erhält der Benutzer ein *Token Pair* bestehend aus *Access Token* und *Refresh Token*. Der Access Token dient zum Zugriff auf eine geschützte Ressource und ist in der Regel kurzlebig. (Bei PEAX läuft er nach fünf Minuten ab.) Mithilfe des Refresh Tokens kann sich der Benutzer einen neuen Access Token vom IDP holen. Der Refresh Token ist darum langlebiger (30 Minuten bei PEAX).

¹In der PEAX API kommt es gelegentlich zu solchen Fehlern, die stattdessen mit dem Status 400 Bad Request und einer aussagekräftigen Fehlermeldung beantwortet werden müssten. Wird z.B. bei der Einlieferung von Dokument-Metadaten eine syntaktisch fehlerhafte IBAN mitgegeben, tritt der Fehler erst bei der internen Verarbeitung, und nicht schon bei der Validierung der Anfrage auf. Hier besteht Handlungsbedarf aufseiten der Backend-Entwicklung.

Es ist die Handhabung dieser Tokens (sehr lange, Base64-codierte Strings), die das Ansteuern der PEAX API mit Programmen wie `curl` und `POSTMAN` so mühsam machen.²

1.1.5 Umgebungen

Bei PEAX gibt es verschiedene Umgebungen oder «Stufen», welche den ganzen PEAX-Stack (Datenbank, Backend, Frontend) für einen bestimmten Zweck zur Verfügung stellen:

local Die Entwickler können den PEAX-Stack zum Entwickeln und Testen lokal ausführen.

dev Dies ist das Entwicklungssystem, worauf die Entwickler ihre Änderungen deployen, sobald diese vom jeweiligen Feature-Branch in den develop-Branch gemerged wurden. Diese Umgebung ist tendenziell sehr aktuell, aber dafür auch instabil.

test Auf diese Stufe werden Änderungen übertragen, die auf Stufe dev erfolgreich getestet werden konnten. Diese Umgebung wird vom Product Owner zur Abnahme von User Stories verwendet, ist in der Regel eher stabil und repräsentiert nach jedem Sprint einen potenziell releasefähigen Stand.

stage Diese Stufe dient für die Regressionstests. Hier wird nach jedem Sprint der letzte Stand von test übertragen. Bei einem Release wird der Stand von hier verwendet. Diese Umgebung ist stabil und jeweils maximal zwei Wochen alt.

prod Von stage werden die Änderungen mehrmals pro Jahr (Ziel: einmal pro Monat) auf die Produktivumgebung übertragen. Dies ist die einzige Umgebung, auf der produktive Kundendaten abgelegt werden. Datenschutz und Sicherheit spielen auf dieser Umgebung eine besonders hohe Rolle.

devpatch Dies ist die Entwicklungsumgebung für den Hotfix-Pfad. Nach einem Release wird der aktuelle Stand von prod auf diese Stufe deployed. Bis zum nächsten Release können hier dringende Fehlerkorrekturen vorgenommen werden.

testpatch Dies ist die Testumgebung für den Hotfix-Pfad. Dringende Fehlerkorrekturen werden von devpatch auf diese Stufe übernommen und hier abgenommen. Die Änderungen werden von hier aus direkt auf prod deployed.

²Eine beispielhafte Analyse ergab, dass ein Access Token 1604 Zeichen lang ist.

prototype Hierbei handelt es sich um eine Umgebung, die sporadisch für Prototypen und Demos verwendet wird.

perf Diese Umgebung wurde vor dem grossen v3-Release im Frühling für Performance-Tests verwendet und ist seither nur sporadisch in Betrieb.

1.1.6 Arten von Parametern

Ein HTTP-Request hat verschiedene Parameter: Dies sind einerseits Header-Parameter, wie z.B. Content-Type, womit der MIME-Type der Request-Bodys festgelegt wird, oder Accept, womit dem Server mitgeteilt wird, welcher MIME-Type der Response-Body haben soll (R. T. Fielding et al., 1999).

Auch in der Authentifizierung und Autorisierung spielen Request-Header eine wichtige Rolle, zumal Access Tokens über per Authorization-Header an den Server übermittelt werden (Hilt, Camarillo & Rosenberg, 2012, Kapitel 7.1).

Andererseits gibt es auch Query-Parameter, welche direkt an die URL angehängt werden. Letztere werden oft für die Navigation im Portal verwendet, zumal bei GET-Requests kein Request-Body übermittelt werden kann.

1.1.7 Benutzer

Es gibt verschiedene Gruppen von Benutzern, die px gewinnbringend einsetzen können:

Backend-Entwickler Diese entwickeln, erweitern und korrigieren die RESTful-API, die das Backend von PEAX ausmachen. Von ihnen kann px einerseits für schnelle Tests und das Erstellen von Testdaten verwendet werden, andererseits kann px auch dabei hilfreich sein, die API (gerade Datenstrukturen) explorativ kennenzulernen.

Frontend-Entwickler Ist die Spezifikation eines Endpoints unvollständig, unklar oder gar fehlerhaft, kann darauf kein funktionierendes Frontend aufsetzen. Hier kann px dabei hilfreich sein, das tatsächliche Verhalten des Backends zu überprüfen, und die Struktur der zurückgelieferten Payloads zu betrachten.

Tester Die manuellen Regressionstests finden direkt auf dem Portal bzw. auf der App statt. Oftmals wäre es hilfreich, Testdaten grösseren Umfangs für einen neu registrierten Benutzer zu erstellen. Mithilfe von px können hierzu einfache Skripts zur Verfügung gestellt werden. (Die Skripts werden tendenziell eher von Entwicklern zur Verfügung gestellt, aber die Tester können diese nach Instruktion selbständig ausführen.

1.1.8 Betriebssysteme

Zu Beginn des Projekts (September 2019) waren auf den persönlichen Computern der Entwickler macOS und Windows im Einsatz. Mittlerweile (Stand Oktober 2019) wurden alle Windows-Rechner durch Geräte mit macOS ausgetauscht. Mit den Testern gibt es dennoch einige potenzielle Benutzer (siehe Abschnitt 1.1.7 *Benutzer*, Seite 8), die px immer noch auf Windows einsetzen würden.

Auf zahlreichen virtuellen Maschinen der PEAX-Infrastruktur (etwa für Datenbanken) läuft Linux als Betriebssystem. Hier könnte px für verschiedene Service-Tasks (Monitoring, Alerting) eingesetzt werden.

Es sind somit die Betriebssysteme macOS, Windows und Linux für die Ausführung von px relevant. Was die Architektur betrifft, kommen derzeit nur Intel-Prozessoren mit 64-Bit-Architektur (x86_64) zum Einsatz.

1.1.9 Shells

Verschiedene Betriebssysteme haben verschiedene Shells. Im UNIX-Bereich gibt es auch zahlreiche Shells mit unterschiedlichen Merkmalen, die parallel zueinander installiert werden können.

Bash ist nicht nur die Standard-Shell vieler Linux-Distributionen, sondern kommt bei Entwicklern auch auf Windows als *git Bash* zum Einsatz. Auf macOS gehört Bash ebenfalls zum Lieferumfang, wobei die mächtigere zsh seit macOS Catalina standardmässig zum Einsatz kommt (Apple Support, 2019). Andere populäre UNIX-Shells wie Fish, Ksh und Tcsh haben zwar unterschiedliche Merkmale, jedoch den POSIX-Standard als kleinsten gemeinsamen Nenner (American National Standards Institute, 1993).

Auf Windows spielen zudem die PowerShell sowie `cmd.exe` eine Rolle.

1.2 Ausgangslage und Vorleistungen

Das Projekt px wurde bereits am 11. Juni 2019 auf dem GitLab von PEAX erstellt³. Als erstes wurde eine CI-Pipeline bestehend aus den Schritten 'build' und 'test' erstellt. Die Pipeline wurde mittels eines Dummy-Tests überprüft, der einmal erfolgreich durchlaufen und einmal scheitern sollte, um einen Positiv- und einen Negativtest durchführen zu können.

Es wurde eine Hallo-Welt-Programm im `cmd`-Unterverzeichnis (Donovan & Kernighan, 2015, p. 293) erstellt, welches dazu diente, die Kompilierung für verschiedene Platt-

³<https://gitlab.peax.ch/peax3/px>

formen zu testen. Obwohl Go-Programme mittels ‘go build‘ kompiliert werden können und keine weitere Build-Konfiguration benötigen, wurde ein Makefile erstellt, das ausführbare Programme für verschiedene Plattformen im build-Unterverzeichnis erstellt, also z.B. build/windows/px.exe für Windows oder build/linux/px für Linux.

Das Makefile wurde später um ein release-Target erweitert, womit die kompilierten Artefakte jeweils in eine Zip-Datei verpackt werden, die den aktuellen Versionstag (z.B. v0.0.3⁴) im Dateinamen enthält.

Das Target coverage führt die Testfälle durch, misst die Testabdeckung und generiert eine HTML-Ausgabe des getesteten Codes. Rote Zeilen sind nicht durch einen Testfall abgedeckt, grüne Zeilen hingegen schon. (Donovan & Kernighan, 2015, Kapitel 11.3)

Weiter sind bis am 31. Juli 2019 folgende Features implementiert worden:

px help zeigt eine einfache Hilfeseite auf der Kommandozeile an.

px login führt einen Loginversuch mit den angegebenen Credentials durch. Benutzername und Passwort können entweder als Kommandozeilenparameter oder mittels interaktiver Eingabe (stdin) entgegengenommen werden. Im letzteren Fall wird das eingegebene Passwort nicht angezeigt, was mit einem externen SSH-Terminal-Modul⁵ erreicht wird. Bei einem erfolgreichen Login-Versuch werden access_token und refresh_token aus dem Response-Payload gelesen und im \$HOME-Verzeichnis des jeweiligen Betriebssystem-Benutzers in eine JSON-Datei namens .px-tokens abgespeichert.

px logout löscht ein Token-Paar für eine bestimmte Umgebung. Pro Umgebung kann es zu jedem Zeitpunkt nur ein aktives «Login», d.h. Token-Paar geben. Es besteht auch die Möglichkeit, sämtliche Tokens auf einmal zu löschen. Hierbei wird \$HOME/.px-tokens nicht gelöscht, sondern nur das Property tokens geleert. (Die Datei enthält ein Initialisierungsdatum, das nicht verlorengehen soll.)

px upload lädt eine Datei (z.B. PDF) auf das PEAX-Portal hoch. Diese Funktionalität wurde eingebaut, um die Funktionsweise von px vor dem Ideation-Gremium zu demonstrieren⁶.

Für die Evaluierung des Prototypen werden zudem die Anzahl Aufrufe und das Datum des letzten Aufrufs von px in eine JSON-Datei \$HOME/.px-usage geschrieben.

⁴<https://semver.org/>

⁵<https://godoc.org/golang.org/x/crypto/ssh/terminal#Terminal.ReadPassword>

⁶Die hochgeladene Datei erschien Sekunden später im Web-Portal, was die Anwesenden von der Funktionsweise überzeugte

I Problemstellung

Insgesamt wurden ca. 20 Arbeitsstunden in den Prototyp investiert. Ein grosser Teil des Codes kann für die Weiterentwicklung übernommen werden, muss jedoch umstrukturiert werden. So ist zuviel Logik im Hauptmodul `cmd/px.go`, die zwecks Wiederverwendbarkeit in das Library-Modul `px` überführt werden soll.

1.3 Risikoanalyse

TODO:

Projektrisiko: fehlende Adaption, Grösse der API

Sicherheitsrisiko: Payment-Schnittstelle, Token-Verwahrung

technisches Risiko: fehlende Usability, Implementierung scheitert

2 Stand der Praxis

2.1 Ansprechen der PEAX API

TODO: Problematik mit Postman, cUrl, httpie beschreiben

2.2 Kommandozeilenprogramme

TODO: docker, oc und git beschreiben

3 Ideen und Konzepte

TODO: swiss army knife (Donovan & Kernighan, 2015, p. 290) für CLI

TODO: token storage (sicher und unsicher)

TODO: automatisches Retry

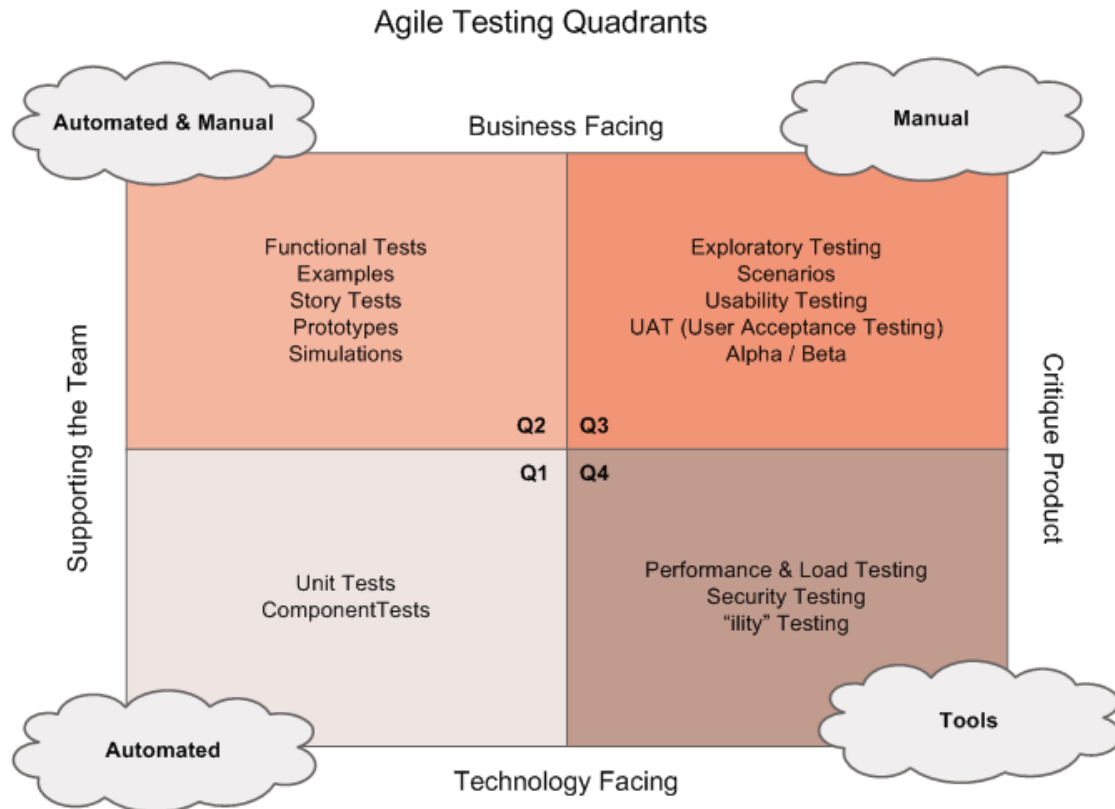


Abbildung 1: *Agile Testing Quadrants* nach Lisa Crispin (<https://lisacrispin.com/2011/11/08/using-the-agile-testing-quadrants/>)

4 Methoden

4.1 Vorgehen

TODO: agiles Entwicklungsprojekt, Ideation, Sprint-Evaluate-Sprint

4.2 Teststrategie

Wie im Vorsemester im Modul *Software Testing* eingeübt, sollen die *Agile Test Quadrants* (Abbildung 1, Seite 14) als Grundlage zur Erarbeitung einer Teststrategie dienen (Crispin & Gregory, 2008, p. 242).

Für die einzelnen Quadranten bieten sich für das vorliegende Projekte folgende Arten von Tests an:

4.2.1 Q1: automatisiert

Im ersten Quadranten geht es um Tests, die vollautomatisch ausgeführt werden können. Diese Tests sollen in einer CI-Umgebung nach jedem Push und vor jedem Merge durchlaufen. Scheitert ein Test, wird der Entwickler notifiziert. Ein Merge-Request soll nicht ausgeführt werden können, wenn Testfälle im Feature-Branch scheitern, sodass die Tests im Master-Branch immer durchlaufen.

Einzelne Funktionen können durch *Unit Tests* abgedeckt werden. Da die Sichtbarkeitsregeln in Go anders geregelt sind als in Java, und ein Unit Test jeweils zum gleichen Modul wie der zu testende Code gehört, können auch interne Funktionen getestet werden, und nicht nur die vom Modul exportierte Schnittstelle (Donovan & Kernighan, 2015, p. 311). Dies erlaubt ein feingranulareres Testing auf Stufe Unit Test.

Die Komponententests sind dann als Tests einzelner Module und somit als Black-Box-Tests zu verstehen, wobei die exportierte (d.h. öffentliche) Schnittstelle angesprochen wird. Auf Mocking soll hierbei verzichtet werden, da das Schreiben *Test Doubles* (Mocks, Spies, Fakes) Kenntnisse der Implementierung und nicht nur der Schnittstelle erfordert. Ändert sich die Implementierung bei gleichbleibendem (und weiterhin erfülltem) Schnittstellenvertrag, sollte auch ein Komponententest weiterhin funktionieren. Dies ist beim Einsatz von Mocks oft nicht gegeben. Eric Elliot bezeichnet Mocking gar als ein *Code Smell*, das die Struktur des Codes verkompliziert, wo doch *Test Driven Development* dabei helfen sollte, Code zu vereinfachen (Elliot, 2019, p. 205),

Ziel der Unit- und Komponententests ist nicht eine möglichst hohe Codeabdeckung, sondern ein optimales Verhältnis von Aufwand und Ertrag: Zeigt es sich, dass für einzelne Tests mit viel Aufwand eine umfassende Umgebung (im weitesten Sinne) aufgebaut werden muss, soll stattdessen geprüft werden, ob der Code nicht besser mittels Tests des zweiten Quadranten getestet werden soll.

4.2.2 Q2: automatisiert und manuell

Im zweiten Quadranten geht es um Tests auf der funktionalen Ebene, die teils automatisch, teils manuell ausgeführt werden.

Da px nicht nur interaktiv, sondern auch in Skripten verwendet werden soll, können komplette Workflows ebenfalls vollautomatisiert durchgetestet werden. Ein folgendes Skript könnte etwa folgenden Ablauf beschreiben:

- Einloggen auf einen System mit speziellen Test-Zugangsdaten

- Hochladen mehrerer Dokumente inklusive Metadaten mit Abspeicherung der dabei generierten Dokument-UUIDs
- Tagging der hochgeladenen Dokumente
- Herunterladen der zuvor hochgeladenen Dokumente und Vergleich mit dem ursprünglich hochgeladenen Dokument (etwa per SHA-2-Prüfsumme)
- Abfragen der zuvor mitgeschickten Metadaten und Tags; Prüfung derselben gegenüber den Ausgangsdaten

Ein solcher Test muss zwangsläufig gegen eine laufende Umgebung durchgeführt werden. (Ähnliche, aber wesentlich einfachere Abläufe werden bereits mit Uptrends durchgeführt, um die Verfügbarkeit der produktiven Umgebung automatisch zu überprüfen.) Hierbei besteht die Gefahr, dass Testfälle aufgrund eines Fehlers in der entsprechenden Umgebung scheitern, und nicht aufgrund der am Code von px vorgenommenen Änderungen. Im schlimmsten Fall müsste ein Merge-Vorgang auf den Master-Branch aufgrund einer nicht funktionierenden Umgebung verzögert werden. Eine pragmatische Lösung wäre es, wenn diese Tests gegen die produktive Umgebung ausgeführt würden. Diese Umgebung ist hoch verfügbar, und bei den seltenen Ausfällen derselben könnte auch notfalls mit einem Merge-Vorgang auf den Master-Branch zugewartet werden.⁷ Die Zugangsdaten für ein entsprechendes Testkonto können innerhalb der CI-Umgebung als verschlüsselte Variablen und lokal als Umgebungsvariablen abgelegt werden.

Das Schreiben von Skripts ist meistens ein Prozess, dem üblicherweise mehrere manuelle Durchgänge der auszuführenden Arbeitsschritte vorangeht. Skripts werden häufig dann geschrieben, wenn ein manueller Vorgang die Finger stärker beansprucht als den Kopf, und aufgrund der nachgebenden Achtsamkeit die Gefahr für Flüchtigkeitsfehler besteht. So soll es auch im vorliegenden Projekt gehandhabt werden: Wird ein neues Feature eingebaut, d.h. eine neue *User Story* umgesetzt, und sich das Testing desselben als aufwändig herausstellt, soll ein Testskript geschrieben werden, das dann sogleich in die CI-Pipeline eingebaut werden kann. Manuelle Tests sollen so möglichst bald und unkompliziert in automatisierte überführt werden.

⁷Der Autor dieser Zeilen ist u.a. auch für die Verfügbarkeit des Produktivsystems zuständig. Ist diese nicht gegeben, werden Entwicklungsarbeiten erfahrungsgemäss unterbrochen, bis das System wieder vollumfänglich verfügbar ist.

4.2.3 Q3: manuell

Nach jedem Sprint erhalten die Entwickler bei PEAX Zugang zum aktuellen Stand der Software mit einem Changelog. Sie haben nun eine Woche Zeit, sich mit den neuen Features vertraut zu machen, und auszuprobieren, ob sich die Software wie gewünscht verhält.

Hier geht es weniger um die Korrektheit gemäss Spezifikation (Akzeptanzkriterien in den User Stories), welche bereits durch Tests in Q1 und Q2 gewährleistet werden sollte, sondern um *User Acceptance Tests*. Hiermit wird geprüft, ob das Inkrement die Bedürfnisse der Benutzer erfüllt und ihren Zielen dient. «Another kind of acceptance testing is *user acceptance testing*. Commonly used in Agile environments, user acceptance testing (UAT) checks that the system meets the goals of the users and operates in a manner acceptable to them (Laboon, 2017, p. 85).

Auch soll im Rahmen dieser Tests die *Usability* des Inkrements überprüft werden. Stossen mehrere Tester auf die gleichen Probleme? Wurde ein Feature (z.B. ein Subcommand) schlecht benannt, sodass dessen Semantik unklar ist? Ist die angebotene Hilfe-Funktion zu einem Befehl unklar oder schlecht formuliert?

Die Tests in Q3 sind jeweils in der ersten Wochenhälfte nach einem Sprint durchzuführen, sodass die Rückmeldungen für die Planung des nächsten Sprints, der in der darauffolgenden Woche startet, berücksichtigt werden kann.

4.2.4 Q4: Tools

Für die Qualitätssicherung können verschiedene Werkzeuge zum Einsatz kommen:

Benchmarks : Bieten sich bei der Implementierung einer performancekritischen Funktion mehrere Varianten an, ist die bessere Variante mithilfe von Benchmarks zu bestimmen. Die integrierte Benchmark-Funktion des go-Tools⁸, die eine Funktion so oft laufen lässt, bis eine statistisch relevante Aussage über deren Performance gemacht werden kann, ist hierzu völlig ausreichend (Donovan & Kernighan, 2015, p. 321).

Profiling : Im Profiling geht es darum, die kritischen Stellen im Code im Bezug auf Rechenzeit, Speicherverbrauch und blockierende Operationen zu ermitteln, um Aufgrund dieser Erkenntnis effektive Optimierungen am Code vornehmen zu können

⁸go test -bench=[pattern]

(Flaschenhalsoptimierung). Hierzu bietet das go-Tool⁹ wiederum sehr mächtige Werkzeuge out-of-the-box an (Donovan & Kernighan, 2015, p. 324).

Quellcodeanalyse : Kompilierbarer und korrekter Quellcode ist nicht automatisch auch guter Quellcode im Bezug auf Klarheit, Einfachheit, Eleganz und Best Practices. Beispielsweise sollen nach Möglichkeit keine veralteten und unsicheren APIs verwendet werden. Exportierte Funktionen, d.h. die öffentliche Schnittstelle eines Moduls, muss dokumentiert sein. Hierzu gibt es einerseits das Tool `go vet`¹⁰, das zum Lieferumfang von Go gehört, und potenzielle Fehler im Code meldet. Das Zusatztool `golint`¹¹ meldet stilistische Unschönheiten im Code.

Da es bei diesen Tools nicht um kategorische Qualitätskriterien (richtig oder falsch), sondern eher um kontinuierliche (schnell/langsam, hoher/tiefer Speicherverbrauch, hohe/tiefer Quellcodequalität) handelt, die einer subjektiven Interpretation bedürfen, sollen diese Tools nicht Teil der CI-Pipeline sein, sondern kontinuierlich (`go vet` und `golint`) und bei konkretem Bedarf (Benchmarking, Profiling) im Entwicklungsprozess eingesetzt werden. Um eine gewisse Quellcodeanalyse gewährleisten zu können, ist die Ausgabe von `go vet` und `golint` am Ende eines jeden Sprints summarisch zu dokumentieren.¹²

⁹`go test -cpuprofile/-memprofile/-blockprofile=[Ausgabedatei]`

¹⁰Vet examines Go source code and reports suspicious constructs, such as Printf calls whose arguments do not align with the format string.

¹¹<https://github.com/golang/lint>

¹²Da die Ausformulierung von Rechtfertigungen oftmals anstrengender ist als die notwendigen Korrekturen am Code vorzunehmen, ist diese Massnahme als Anreiz zu verstehen, die erhaltenen Warnungen zu behandeln statt zu ignorieren.

5 Realisierung

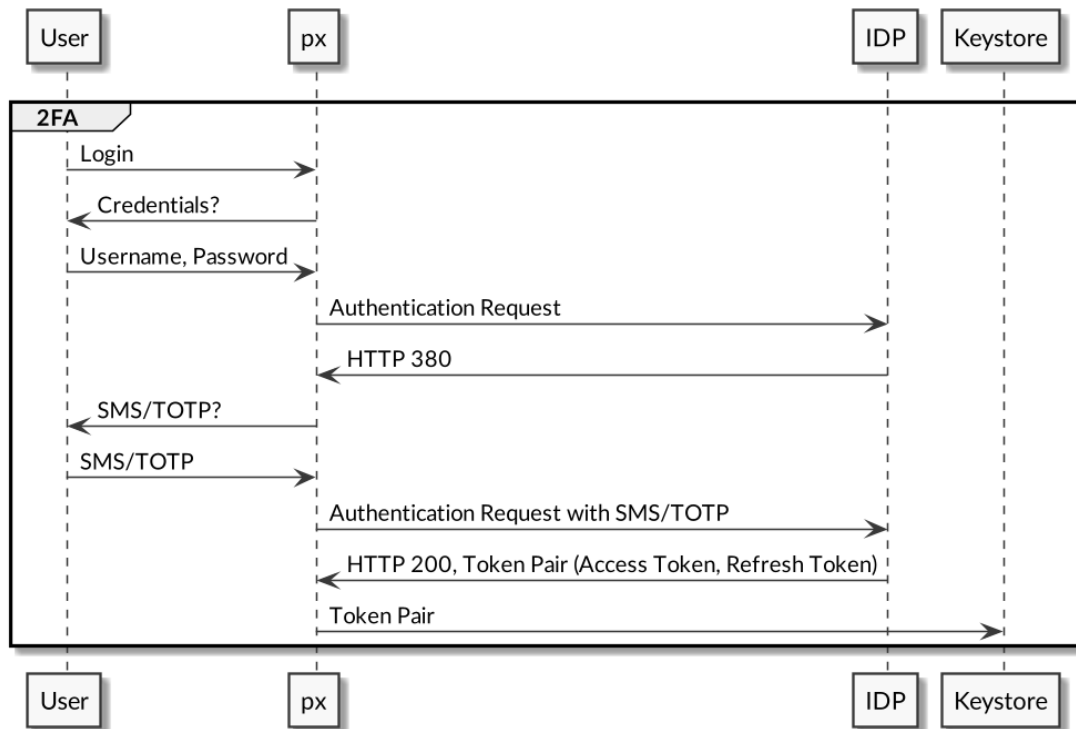


Abbildung 2: Sequenzdiagramm: Der Ablauf der Zwei-Faktor-Authentifizierung

5 Realisierung

TODO: Libraries, Umgebung, Architektur

5.1 Zwei-Faktor-Authentifizierung

5.2 Token-Store

TODO: sicher und unsicher, Datenstruktur

6 Evaluation und Validierung

6.1 Rückmeldungen von Entwicklern

6.1.1 Sprint 1

- MICHAEL BUHOLZER wünscht sich Erfolgs- und Vollzugsmeldungen nach dem Login oder dem Upload eines Dokuments.
 - stdout sollte grundsätzlich «sauber» bleiben, d.h. frei von unnötigen Ausgaben, die ein nachgelagertes Programm wieder herausfiltern müsste. Eine wichtige Maxime von UNIX-Programmen lautet: *«Expect the output of every program to become the input to another, as yet unknown, program.»* (McIlroy Doug, Pinson, E.N, Tague, B.A., 1978, p. 3). Siehe dazu auch *Rule of Silence* (Raymond, 2004, p. 20) und *Silence is Golden* (Gancarz, 1995, p. III).
 - stderr wird nicht nur als Ausgabekanal für Fehlermeldungen verwendet, sondern für Meldungen allgemein. Für Vollzugsmeldungen wäre stderr vorzuziehen.
 - Da stderr in px bisher grundsätzlich für Fehlermeldungen verwendet wird, sollen Erfolgsmeldungen über ein zusätzliches Flag `-verbose/-v` aktiviert werden müssen.
 - Bei anderen Anwendungsfällen signalisiert die Ausgabe des Payloads auf stdout den Erfolg der Operation. Beim Dokument-Upload besteht dieser beispielsweise aus der generierten UUID des hochgeladenen Dokuments.
- PATRICK ROOS sieht die Möglichkeit, px auch zur Handhabung der *Vault Secrets*¹³ (Verschlüsselung und Entschlüsselung von Benutzernamen, Passwörtern etc. zu verwenden).
 - Im Arbeitsalltag von PEAX stellt die Handhabung von Vault Secrets tatsächlich eine teils mühsame und langwierige Aufgabe dar. Hier besteht durchaus Automatisierungsbedarf.
 - px ist als «skriptbare» Anwendung für die PEAX API konzipiert und so potenziell für jeden PEAX-Anwender einsetzbar.
 - Die Verwaltung und Verwendung von Vault Secrets ist hingegen eine Aufgabe im DevOps-Bereich und betrifft nur interne Entwickler bei PEAX.

¹³https://docs.ansible.com/ansible/latest/user_guide/vault.html

6 Evaluation und Validierung

- Eine der obersten Maximen von UNIX lautet: *«Make each program do one thing well. To do a new job, build afresh rather than complicate old programs by adding new features.»* (McIlroy Doug, Pinson, E.N, Tague, B.A., 1978, p. 3) Die Verwaltung von Vault Secrets und das Ansprechen der PEAX API sind klar zwei verschiedene Sachen und somit nicht «one thing». Die genannte Idee muss also anderweitig weiterverfolgt werden.

7 **Ausblick**

TODO: OpenSource, auf dem Laufenden halten, unterschiedliche API-Versionen, wie weiter?

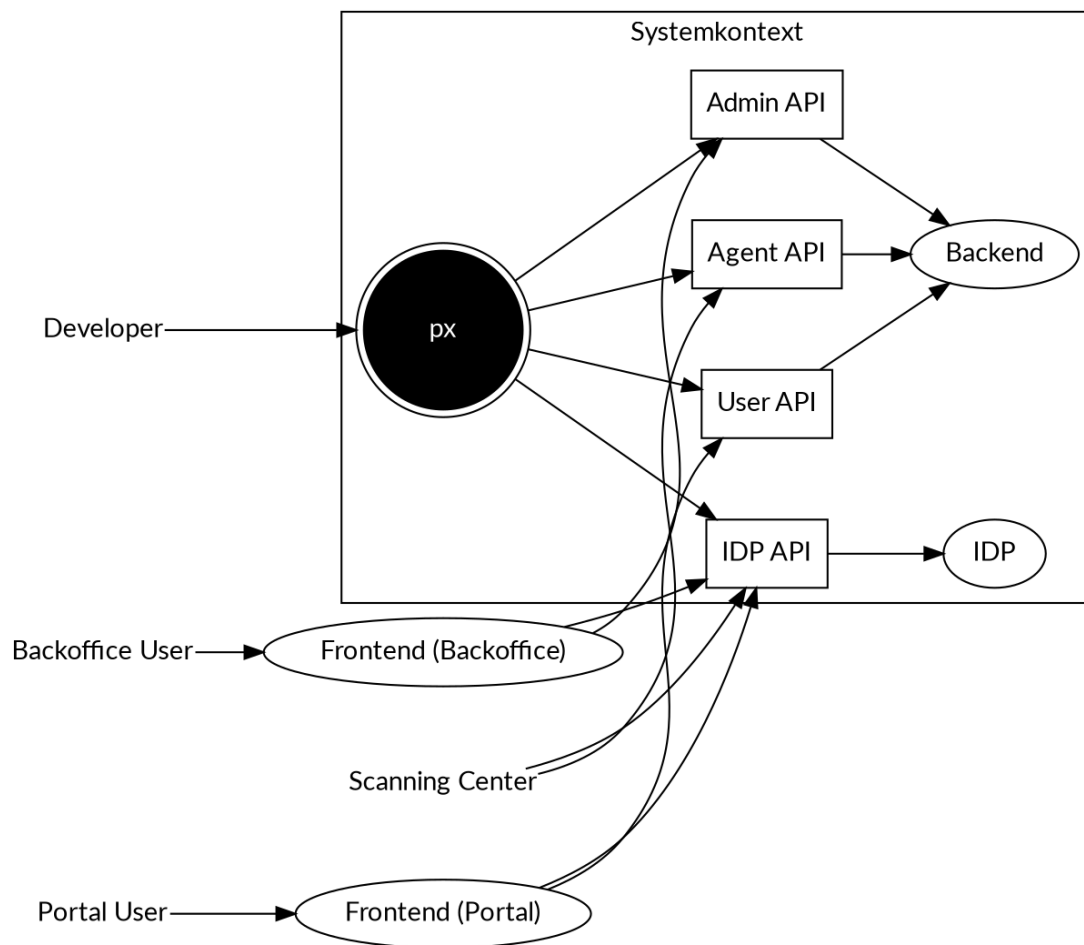


Abbildung 3: Kontextdiagramm: px als der Gegenstand der Arbeit innerhalb des Systemkontexts

8 Anhang

8.1 Systemspezifikation

8.1.1 Systemkontext

Im Kontextdiagramm (Abbildung 3, Seite 23) wird die zu entwickelnde Komponente px im Systemkontext von PEAX dargestellt. Andere Komponenten sind als Ellipsen, Schnittstellen als Rechtecke dargestellt. Ausserhalb vom Systemkontexts befindet sich die irrelevante Umgebung (d.h. die Frontend-Anwendungen). Die im Kontextdiagramm verwendeten Begriffe haben folgende Bedeutungen:

- px: Der PEAX Command Line Client (Gegenstand der vorliegenden Arbeit)
- Developer: Ein Softwareentwickler (im weitesten Sinne) bei PEAX, der px verwendet.
- Backoffice User: Ein PEAX-Angestellter mit administrativen Befugnissen (Benutzerverwaltung).
- Portal User: Ein Benutzer des PEAX-Portals (Kunde).
- Scanning Center: Zulieferfirma, welche die umgeleitete Papierpost der PEAX-Kunden erhält, diese einscannet und dem betreffenden Kunden ins Portal stellt.
- Frontend (Backoffice): Ein Web-GUI für administrative Tätigkeiten zum internen Gebrauch.
- Frontend (Portal): Ein Web-GUI für die Kunden von PEAX (das eigentliche Portal).
- Admin API: RESTful API für administrative Aufgaben
- Agent API: RESTful API zum Einliefern von Dokumenten über Zulieferer
- User API: RESTful API für die Operationen der Kunden
- IDP API: RESTful API für das Token-Management (OAuth 2.0/OpenID Connect)
- Backend: Serverseitige Software mit Businesslogik und Datenspeicher
- IDP: Identity Provider, API-übergreifende Benutzer- und Zugangsverwaltung (AuthN/AuthZ)

8.2 Technologie-Evaluation

- Vorgabe: PEAX API (RESTful)

8.2.1 Programmiersprache

Aus der Aufgabenstellung und dem Umfeld bei PEAX ergeben sich folgende nicht-funktionale Anforderungen an die zu erstellende Software:

Installation Die Software soll sich einfach installieren lassen.

Umgebung Es dürfen keine besonderen Anforderungen an die Umgebung gestellt werden, auf der px läuft.

Plattformen Die Software soll auf allen gängigen, d.h. bei PEAX eingesetzten, Betriebssystemen (Windows, mac OS, Linux) lauffähig sein.

Einheitlichkeit Der Client soll überall die gleiche Befehlssyntax haben.

Performance Ein Command Line Client soll in Skripten verwendet werden können, wodurch das Programm sehr oft in kurzem Zeitraum aufgestartet werden muss.

Java, das bei PEAX im Backend-Bereich zum Einsatz kommt, erfordert die lokale Installation einer JRE in der richtigen Version, was bei Frontend-Entwicklern nicht gegeben ist. Ausserdem werden Wrapper-Skripts benötigt (`java -jar px.jar` ist nicht praktikabel).

Python, Ruby, Perl und andere Skriptsprachen benötigen ebenfalls einen vorinstallierten Interpreter in der richtigen Version.

Zwar gibt es mit Mono eine Variante von .Net, die überall lauffähig ist, hier werden aber wiederum eine Laufzeitumgebung bzw. vorinstallierte Libraries benötigt.

Für die Problemstellung am besten geeignet sind kompilierte Sprachen (C, C++, Go, Rust, Nim). Mit einer statischen Kompilierung lässt sich das ganze Programm in eine einzige Binärdatei überführen, welches denkbar einfach zu installieren ist (Kopieren nach einem der Verzeichnisse innerhalb von \$PATH).

Für JavaScript, das bei PEAX im Frontend zum Einsatz kommt, gibt es mit QuickJS¹⁴ seit kurzem die Möglichkeit, JavaScript zu Binärdateien zu kompilieren. Dies funktioniert aber nicht auf allen Plattformen, ausserdem ist QuickJS noch experimentell und noch nicht für den produktiven Einsatz geeignet.

Um ein Projekt vom gegebenen Umfang innerhalb eines Semesters umsetzen zu können, sind Vorkenntnisse in der einzusetzenden Programmiersprache zwar nicht zwingend, können das Risiko des Scheiterns aber erheblich senken. Gerade bei der Abschätzung von Aufwänden ist Vertrautheit mit den einzusetzenden Werkzeugen sehr hilfreich.

Was (statisch) kompilierte Programmiersprachen betrifft, konnte der Autor dieser Arbeit bereits Erfahrungen mit C, Go und Rust sammeln. Das manuelle Speichermanagement in C (u.a. auch bei Strings) ist einerseits ein grosses Risiko (Buffer Overflows, Segmentation Faults), und wirkt sich andererseits negativ auf das Entwicklungstempo aus. In die engere Auswahl kommen somit Go und Rust.

¹⁴<https://bellard.org/quickjs/>

Im Folgenden werden die gemachten Erfahrungen und die dabei empfundenen Vor- und Nachteile mit den Programmiersprachen Go und Rust einander gegenübergestellt.

8.2.2 Go

Mit Go konnte der Autor dieser Arbeit bereits einiges an Erfahrung sammeln. So wurde neben dem Prototyp zu px bereits die Testat-Aufgabe im Modul Software Testing¹⁵, ein Thumbnailer¹⁶ sowie zahlreiche Utilities¹⁷ (viele darunter als HTTP-Clients) in Go entwickelt. Dabei wurden folgende Vor- und Nachteile ermittelt:

- Vorteile
 - aufgrund weniger Keywords und Features einfach zu lernen
 - hervorragendes Tooling out-of-the-box
 - Cross-Compilation ohne Zusatztools auf alle unterstützte Plattformen möglich
 - schnelle Kompilierung
 - umfassende Standard-Library, die u.a. ein hervorragendes HTTP-Package beinhaltet
 - persönlich bereits viel (positive) Erfahrungen damit gesammelt
 - wird bereits für andere bei PEAX gebräuchliche CLI-Tools eingesetzt (oc, docker)
 - fügt sich sehr gut in die UNIX-Philosophie ein (Tooling, Libraries)
 - Einfaches Interface für nebenläufige Programmierung (Goroutines und Channels)
 - geringer Memory-Verbrauch bei relativ hoher Performance¹⁸
- Nachteile
 - fehlende Features wie Generics und filter/map/reduce
 - Binaries fallen relativ gross aus¹⁹
 - Error-Handling aufwändig und teils repetitiv

¹⁵<https://github.com/patrickbucher/getting-to-philosophy>

¹⁶<https://github.com/patrickbucher/thumbnailer>

¹⁷<https://github.com/patrickbucher/go-scratch>

¹⁸<https://benchmarksgame-team.pages.debian.net/benchmarksgame/fastest/go.html>

¹⁹https://golang.org/doc/faq#Why_is_my_trivial_program_such_a_large_binary

8.2.3 Rust

Der Autor dieser Arbeit konnte sich bereits letztes Jahr im Rahmen des Moduls *Programming Concepts and Paradigms* an der HSLU Informatik mit Rust befassen (Arnold & Bucher, 2018, p. 12). Nach selbständiger Beschäftigung mit dieser Programmiersprache im Sommer können (teils ergänzend) folgende Vor- und Nachteile genannt werden:

- Vorteile
 - viele moderne Features (Generics, filter/map/reduce)
 - hervorragendes Typsystem
 - gutes und ausgereiftes Tooling
 - weder manuelles Memory-Management noch Garbage Collector nötig
 - Pattern Matching führt zu sehr solidem Code
 - gegenüber Go schlankere Binaries
 - kommt bereits in der Form einiger CLI-Tools persönlich zum Einsatz (rg, bat, hexyl, battop)
 - erstklassige Performance (im Bereich von C/C++) bei geringem Speicherverbrauch²⁰
- Nachteile
 - hohe Einstiegshürde und lange Einarbeitungszeit
 - Cross-Compilation benötigt Zusatztools
 - noch keine praxisnahe Erfahrung damit gesammelt
 - aufgrund schlanker Standard Library auf viele Dependencies angewiesen

8.2.4 Entscheidung

Rust hat gegenüber Go einige unbestreitbare Vorzüge (Memory Management, Typsystem, Ausdrucksstärke, Eliminierung ganzer Fehlerklassen, Performance, schlankere Binärdateien). Bezogen auf das umzusetzende Projekt fallen jedoch einige davon kaum ins Gewicht (etwa Performance und Zero-Cost Abstractions). Hier fallen die Vorzüge von Go (umfassende Standard Library, Cross-Compilation) wesentlich stärker ins Gewicht.

²⁰<https://benchmarksgame-team.pages.debian.net/benchmarksgame/fastest/rust-go.html>

Gerade die absichtlich schlank gehaltene Standard Library von Rust, die etwa zur Generierung von Zufallszahlen bereits externe Abhängigkeiten erfordert²¹, dürfte sich im vorliegenden Projektrahmen negativ auswirken, zumal die Evaluation verschiedener Libraries einen sehr hohen Zusatzaufwand zur Folge haben.

Da Go bereits bei der Entwicklung des Prototypen von px erfolgreich zum Einsatz kam, und einige Projektaspekte (grundlegende CI-Pipeline, Makefile für Cross-Compilation und Packaging) bereits damit implementiert werden konnten, soll Go für das vorliegende Projekt den Vorzug erhalten.

Eine spätere Neuimplementierung von px in Rust wäre ein technisch durchaus interessantes, wenn auch praktisch wenig dringendes – als Fallstudie aber durchaus lohnendes – Unterfangen.

8.3 Libraries

TODO: JWT, Keystore, Password Input

²¹<https://doc.rust-lang.org/book/ch02-00-guessing-game-tutorial.html#using-a-crate-to-get-more-functionality>

9 Einstiegsbeispiel

```

package main

import (
    "fmt"
    "log"
    "os"

    "gopl.io/ch05/links"
)

// breadthFirst calls f for each item in the worklist.
// Any items returned by f are added to the worklist.
// f is called at most once for each item.
func breadthFirst(f func(item string) []string, worklist []string) {
    seen := make(map[string]bool)
    for len(worklist) > 0 {
        items := worklist
        worklist = nil
        for _, item := range items {
            if !seen[item] {
                seen[item] = true
                worklist = append(worklist, f(item)...)
            }
        }
    }
}

func crawl(url string) []string {
    fmt.Println(url)
    list, err := links.Extract(url)
    if err != nil {
        log.Print(err)
    }
    return list
}

func main() {

```

9 *Einstiegsbeispiel*

```
// Crawl the web breadth-first,  
// starting from the command-line arguments.  
breadthFirst(crawl, os.Args[1:])  
}
```

Codebeispiel 1: Some Go Code

Literatur

- American National Standards Institute. (1993). *IEEE standard for information technology: Portable Operating System Interface (POSIX) : part 2, shell and utilities*. IEEE Computer Society Press.
- Apple Support. (2019). *Use zsh as the default shell on your Mac*. <https://support.apple.com/en-us/HT208050>. (Abgerufen am 26.10.2019)
- Arnold, L. & Bucher, P. (2018, December). *Rust* (Bericht). (<https://github.com/patrickbucher/pcp-project/raw/master/Rust-Arnold-Bucher/paper.pdf>)
- Bucher, P. & Christensen, C. (2019, May). *OAuth 2* (Bericht). (https://github.com/patrickbucher/inf-stud-hslu/raw/master/infkol/thesis/OAuth2_Bucher-Christensen.pdf)
- Crispin, L. & Gregory, J. (2008). *Agile Testing*. Addison-Wesley.
- Donovan, A. A. A. & Kernighan, B. W. (2015). *The Go Programming Language*. Addison-Wesley.
- Dusseault, L. & Snell, J. (2010, March). *PATCH Method for HTTP* (RFC Nr. 5789). RFC Editor. Internet Requests for Comments. Zugriff auf <http://www.rfc-editor.org/rfc/rfc5789.txt> (<http://www.rfc-editor.org/rfc/rfc5789.txt>)
- Elliot, E. (2019). *Composing Software*. Leanpub. <https://leanpub.com/composingsoftware>.
- Fielding, R. & Reschke, J. (2014, June). *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content* (RFC Nr. 7231). RFC Editor. Internet Requests for Comments. Zugriff auf <http://www.rfc-editor.org/rfc/rfc7231.txt> (<http://www.rfc-editor.org/rfc/rfc7231.txt>)
- Fielding, R. T., Gettys, J., Mogul, J. C., Nielsen, H. F., Masinter, L., Leach, P. J. & Berners-Lee, T. (1999, June). *Hypertext transfer protocol – http/1.1* (RFC Nr. 2616). RFC Editor. Internet Requests for Comments. Zugriff auf <http://www.rfc-editor.org/rfc/rfc2616.txt> (<http://www.rfc-editor.org/rfc/rfc2616.txt>)
- Gancarz, M. (1995). *The UNIX Philosophy*. Digital Press.
- Hilt, V., Camarillo, G. & Rosenberg, J. (2012, December). *A Framework for Session Initiation Protocol (SIP) Session Policies* (RFC Nr. 6794). RFC Editor. Internet Requests for Comments. Zugriff auf <https://tools.ietf.org/html/rfc6794> (<https://tools.ietf.org/html/rfc6794>)
- Laboon, B. (2017). *A Friendly Introduction to Software Testing*.
- McIlroy Doug, Pinson, E.N, Tague, B.A. (1978). *UNIX Time-Sharing System* (Bericht). Bell System Technical Journal 57. (<https://archive.org/details/bstj57-6-1899>)

Literatur

- Mozilla Developer Network. (o. J.). *Cross-Origin Resource Sharing (CORS)*. <https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS>. (Abgerufen am 26.10.2019)
- P. Bryan, M. N. (2013, April). *JavaScript Object Notation (JSON) Patch* (RFC Nr. 6902). RFC Editor. Internet Requests for Comments. Zugriff auf <https://tools.ietf.org/html/rfc6902> (<https://tools.ietf.org/html/rfc6902>)
- Raymond, E. S. (2004). *The Art of UNIX Programming*. Addison-Wesley.

Abbildungsverzeichnis

I	<i>Agile Testing Quadrants</i> nach Lisa Crispin (https://lisacrispin.com/2011/11/08/using-the-agile-testing-quadrants/)	I4
2	Sequenzdiagramm: Der Ablauf der Zwei-Faktor-Authentifizierung	19
3	Kontextdiagramm: px als der Gegenstand der Arbeit innerhalb des Systemkontexts	23

Tabellenverzeichnis

Verzeichnis der Codebeispiele

I	Some Go Code	29
---	------------------------	----