

## BOT-12: Rasa 2.0 Update

Um die neusten Funktionen und Fixes von Rasa zu benutzen, ist ein Update von Version 1.6 auf 2.0 notwendig.

Zuständigkeit	Robert Halwaß
Initiale Schätzung	3 Tage
Programmiersprachen	* Python
Frameworks/Libraries	* Rasa
Services	* NLU
Abhängigkeiten	keine
Anforderungen	* Kompatibilität mit bestehenden NLU-Trainingsdaten erhalten * Mögliche JSON- und Markdown-Dateien in YAML-Dateien umwandeln
Tasks	* BOT-124: docker-compose.yml anpassen * BOT-61: Chatito Kompatibilität testen * BOT-62: Config Anpassen * BOT-63: Duckling updaten * BOT-64: Modell neu trainieren * BOT-116: Performance zwischen Version 1.6 und 2.0 vergleichen

## BOT-30: Chatbot Library: Vereinheitlichung der Kommunikation von Javascript Chatbots mit dem Gateway

Problem: Derzeit muss jede Anwendung, die den BHT-Bot als Chatbot implementieren möchte selbst implementieren, wie die Kommunikation zwischen Anwendung und Gateway aussieht, als auch die Schnittstellen Parameter in Anfrage und Antwort.

Um diese Implementationsredundanz zu verhindern, wird die Kommunikation und Typdefinitionen in einer zentralen Javascript Bibliothek zusammengefasst.

Dies ermöglicht auch weitere geplante Funktionalitäten (wie der asymmetrische Kommunikationskanal zur Requestunabhängigen Server → Client Kommunikation ) zentral entwickelt und mittels Dependency Management schnell in die ChatClients überführt werden.

Initiale Schätzung 1

Technologien	* Javascript * Typescript
Abhängigkeiten	keine
Anforderungen	* Die Library lässt sich in Node und Browser Javascript einbinden \\* Die Library nutzt semantische Versionierung zur Ermöglichung von Non-Breaking-Updates \\* Die fertige Library lässt sich via Dependency-Management (npm/yarn/webpack) userseitig einbinden und updaten \\* Die Library enthält typisierte (typescript) Entitäten für Common Request und Response Format(e) \\* Die Library enthält Unit-Tests für essentielle Funktionen und Typen \\* Die Library ist dokumentiert, sowohl was Nutzung, als auch Contribution angeht \\* Die Library verbessert die Collaboration mittels Linting-Regeln und Workflow-Scripten

Tasks	<ul style="list-style-type: none"><li>* BOT-33 Library Usage Dokumentieren</li><li>* BOT-34 Library in Discord Bot integrieren</li><li>* BOT-35 Library in Telegram Bot integrieren</li><li>* BOT-36 Library in Website integrieren</li><li>* BOT-31 Common Funktionalität / Use Cases identifizieren</li><li>* BOT-32 Typescript Library für Bot erstellen</li></ul>
-------	---

#### BOT-37: Discord Integration des BHT-Bot

Discord ist eine weit verbreitete Kommunikationsplattform, auf der Nutzer sich in „Servern“ vernetzen und dort meist thematisch organisiert kommunizieren können. Die Projektgruppe des WS2020 ist selbst Teil der Zielgruppe des Discord-Messengers, wodurch sich diese Plattform besonders eignet um einen weiteren Chat-Service (neben Telegram) an den BHT-Bot anzubinden. Eine Implementation des Chatbots innerhalb der Discord-Struktur steigert somit zum Einen die Verbreitung (Möglichkeit) des BHT-Bot und bietet gleichzeitig eine gute Möglichkeit Debug-Bot-Instanzen im präferierten Messenger zu betreiben.

Initiale Schätzung	2
Technologien	<ul style="list-style-type: none"><li>* Javascript</li><li>* Docker</li></ul>
Abhängigkeiten	<ul style="list-style-type: none"><li>* BOT-30</li></ul>
Anforderungen	<ul style="list-style-type: none"><li>* Der Discourse Bot benutzt die zu entwickelnde zentralisierte Library zur Gateway Kommunikation um Coderedundanz mit dem Telegram Bot zu verhindern</li><li>* Der Discourse Bot kann (direkte) Nutzer-Nachrichten mittels Gateway verarbeiten und antwortet dem User entsprechend</li><li>* Wenn keine Verbindung mit dem Gateway besteht oder Fehler bei Anfragen auftreten reagiert der Chatbot durch Präsentation einer hilfreichen Fehlermeldung</li><li>* Der Discourse Bot wird äquivalent zum Telegram Bot in das BHT-Bot Universum mittels Docker + compose integriert</li><li>* Der Discourse Bot ist nicht Teil des BHT Bot, er wird als paralleler, unabhängiger Service betrieben</li><li>* Alle Credentials und Urls/Ports werden aus dem Environment bezogen, es gibt keine hard-coded Referenzen zu Strukturen des BHT-Bot Gateways</li></ul>
Tasks	<ul style="list-style-type: none"><li>* BOT-38 NodeJS Chatbot erstellen</li><li>* BOT-39 Docker Container + Compose für Container erstellen</li><li>* BOT-40 Bot Usage dokumentieren</li><li>* BOT-41 Bot Account anlegen für release (<a href="https://discord.com/developers/applications">https://discord.com/developers/applications</a>)</li><li>* BOT-42 Bot Container in Beuth-Docker-Netzwerk einbinden (release)</li></ul>

#### BOT-43: Erstellung eines Common-Frameworks für (Content-)Services

Services im BHT-Bot kommunizieren alle über REST-Schnittstellen. Diese sind alle als Express-Anwendung mit JSON-Kommunikation implementiert, was für viel Code-Redundanz führt. Gleichzeitig benutzt kein Service strukturierte Darstellungen der Schnittstellen, wie Anfragen und Antworten zum Gateway, User Daten oder Rasa-Intents.

Ein spezieller, repetitiver, Unterfall der Microservices sind solche, die Content bereitstellen. Diese erhalten alle Nachrichten vom Gateway und senden ihre Antworten auch dort wieder zurück.

Request- und Response sind durch das Gateway definiert, die resultierende Struktur ist entsprechend für alle Content-Services prinzipiell identisch.

Zur Vermeidung von Code-Redundanzen und Erleichterung des „Kick-Off“ eines neuen Content-Services sollen die Common Funktionen und Entitäten in ein Framework gegossen werden

Initiale Schätzung	1
Technologien	<ul style="list-style-type: none"> <li>* Javascript</li> <li>* Typescript</li> <li>* Dockerfile</li> </ul>
Abhängigkeiten	* BOT-37
Anforderungen	<ul style="list-style-type: none"> <li>* Das Framework implementiert eine NodeJS Express REST-API, äquivalent zu den existierenden Content-Services</li> <li>* Das Framework lässt sich in NodeJS Anwendungen via Dependency-Management einbinden (npm/yarn)</li> <li>* Das Framework abstrahiert den (Express) Server und dessen Routing, so dass ein Content-Service nur noch die Response implementieren muss</li> <li>* Das Framework implementiert die Schnittstelle zum Datenbankservice um a) Userdaten zu speichern/abzurufen und b) eigene Daten zu speichern und abzurufen</li> <li>* Das Framework füllt die „debug-history“ der Requests so, dass ein Service dieses Feature zwangsweise implementiert / nutzt</li> <li>* Requests, Responses, User, Rasa-Intents und ggf. weitere Entitäten werden durch das Framework als typisierte (typescript) Objekte definiert</li> <li>* Das Framework wird in allen bestehenden und geplanten Content-Services implementiert: Wetter, Mensa, Reminder</li> <li>* Die Nutzung des Frameworks ist verständlich dokumentiert</li> <li>* Die Contribution wird mittels Linting, Dokumentation und Build-Skripts erleichtert</li> <li>* Das Framework nutzt Types aus der (noch zu entwickelnde) BHT-Bot Library um Redundanzen zwischen Client-Bibliothek und Service-Framework zu vermeiden</li> </ul>
Tasks	<ul style="list-style-type: none"> <li>* BOT-44 Common Code, Features, Entitäten identifizieren → Use Case ableiten</li> <li>* BOT-45 Typisiertes Javascript Framework erstellen</li> <li>* BOT-46 Framework einbinden in Weather Service</li> <li>* BOT-47 Framework einbinden in Mensa Service</li> <li>* BOT-48 Framework einbinden in Reminder Service</li> <li>* BOT-108 Framework dokumentieren</li> <li>* BOT-117 Request History implementieren - Nutzung enforce</li> </ul>

BOT-49: User-Messenger-Service: Nachricht proaktiv, requestunabhängig an Clients senden

Der BHT-Bot kann bisher nur passiv auf Anfragen warten und diese dann beantworten. Zur Implementierung von asymmetrischer bzw. request-unabhängiger Kommunikation benötigt der Bot einen neuen Service, der als Schnittstelle für diese Art von Kommunikation dient.

Initiale Schätzung 2.5

Technologien	<ul style="list-style-type: none"><li>* Javascript</li><li>* Websockets</li><li>* Docker</li></ul>
Abhängigkeiten	<ul style="list-style-type: none"><li>* BOT-30</li></ul>
Anforderungen	<ul style="list-style-type: none"><li>* Der User-Messenger-Service kann eine Nachricht an einen User (via Client-Unabhängiger User-ID) senden</li><li>* Wenn ein User mehrere Clients benutzt, wird die Nachricht an alle Clients gesendet</li><li>* Wenn ein User (über längere Zeit) nicht erreichbar ist wird die Kennung entfernt</li><li>* Wenn eine Nachricht nicht an einen User gesendet werden kann bekommt der auslösende Service diese Information unterscheidbar zwischen a) Der Nutzer ist gerade nicht erreichbar b) Der Nutzer ist dauerhaft nicht erreichbar (gelöscht) c) Der Dienst ist generell unhealthy</li><li>* Der Service wird als Docker Image via docker-compose in die BHT-Bot Infrastruktur integriert. Er ist Teil des BHT-Bot Repositories</li><li>* Die Funktionalität des Services wird auf Clientseite in die Common-Chatbot-Library (BOT-30)</li><li>* Alle bestehenden ChatBot-Services werden an den Messenger Service angebunden (Telegram, Discord)</li></ul>
Tasks	<ul style="list-style-type: none"><li>* BOT-50 Websocket Registry für ChatBotClients</li><li>* BOT-51 REST-Service für Nachrichtenversand</li><li>* BOT-52 Implementation der Websocket-Registrierung in Common-Library für Chatbots</li><li>* BOT-53 Implementierung der Common-Library-Websocket-Registrierung in TelegramBot</li><li>* BOT-54 Implementierung der Common-Library-Websocket-Registrierung in DiscordBot</li><li>* BOT-56 Dokumentation Usage Service</li><li>* BOT-110 Deployment / Release</li></ul>

BOT-55: Erinnerungs-Service: Behandelt „erinnere mich“ Befehle und erinnert bei Fälligkeit autonom

Erinnerungen scheulen zu können ist ein typischer, weil praktischer, Anwendungsfall in beliebten (Business) Kommunikations-Services wie Slack oder Mattermost. In beiden Fällen wird diese Funktionalität durch die hauseigenen Reminder-Bots zur Verfügung gestellt.

Um die Featuredichte des BHT-Bot zu erhöhen wird ein Reminder-Service erstellt, durch den identische Funktionalität wie bei genannten Diensten zur Verfügung stellt. Durch die Multi-Messenger-Fähigkeit des BHT-Bot wird dieses Feature somit auch für User angeboten, deren Kommunikations-Plattform keinen eigenen Reminder-Bot anbietet.

BeispielAnfragen:

- \* Erwinnere mich am 22.10. an die Klausur in Mathe
- \* Erwinnere mich jeden Donnerstag um 18 Uhr an den Ballettkurs
- \* Erwinnere mich jedes Jahr am 01.01 an den Geburtstag meiner Mutter.
- \* Erwinnere mich in 10 Tagen das Probeabonnement zu kündigen
- \* <Erwinnere> <Zeitpunkt/Zeitspanne/Interval> <Thema>

Initiale Schätzung

3

Technologien	<ul style="list-style-type: none"> <li>* Javascript</li> <li>* Docker</li> <li>* Rasa</li> <li>* MongoDB</li> <li>* Cronjob</li> </ul>
Abhängigkeiten	<ul style="list-style-type: none"> <li>* BOT-43</li> <li>* BOT-30</li> <li>* BOT-12</li> <li>* BOT-49</li> </ul>
Anforderungen	<ul style="list-style-type: none"> <li>* Erfolgreiche „erinnere“-Anfragen werden vom Dienst durch Bestätigung der erkannten und persistierten Daten beantwortet oder</li> <li>* Fehlerhafte „erinnere“-Anfragen werden durch ein Mini-Usage-Tutorial beantwortet, damit der User seine Anfrage korrigieren kann</li> <li>* Erinnerungen werden auf user-ebene (clientunabhängig) gespeichert, so dass ein User die gleichen Erinnerungen in allen genutzten Clients zur Verfügung hat</li> <li>* Erinnerungen werden bei Fälligkeit einmalig (an alle clients des users) ausgespielt</li> <li>* Wiederkehrende Erinnerungen werden ausgespielt und anschließend an Hand des Intervals neu terminiert</li> <li>* Der Nutzer kann Erinnerungen löschen</li> <li>* Der Nutzer kann seine Erinnerungen anzeigen lassen</li> <li>* Der Service wird als Docker Container via docker-compose verwaltet und in das BHT-Repository integriert</li> <li>* Der Service nutzt das (noch zu schaffende) Content-Service-Framework</li> <li>* Wenn der Reminder-Service nach einem Ausfall wieder aktiv wird erinnert er nicht (einzeln) an alle Termine, die zwischenzeitig fällig waren</li> <li>* Der Service ist in Hilfe-Texten des (Chat) BHT-Bots integriert</li> </ul>
Tasks	<ul style="list-style-type: none"> <li>* BOT-57 Rasa Anbindung „Erinnere“-Direktive</li> <li>* BOT-58 Service Endpoint speichert Reminder und Antwortet auf Probleme</li> <li>* BOT-59 Scheduler/Cronjob prüft und sendet regelmäßig fällige Erinnerungen</li> <li>* BOT-60 Dokumentation Service Usage</li> <li>* BOT-109 Deployment / Release</li> <li>* BOT-112 Hilfe-Texte in (Chat) BHT-Bot help-command / willkommensnachricht</li> </ul>

BOT-82: Termin-Scraper, der automatisch Erinnerungen aus öffentlichen Quellen bezieht

Es gibt Termine, an die wollen alle Studierenden typischerweise erinnert werden, als auch Termine, von denen die Universität möchte, dass die Studierende sich daran erinnern. Typische Beispiele hierfür sind Rückmeldefristen und (Beuth-eigene) Feiertage.

Durch einen Webscraper sollen solche Termine aus (öffentlichen) Quellen automatisch bezogen und dann an alle Studierenden ausgespielt werden.

Auch wenn dieses Feature für die meisten Studierenden interessant sein dürfte, muss es eine Möglichkeit zum Opt-Out geben. Ggf. ist sogar angezeigt, dass ein Opt-In erfolgt, was allerdings den Nutzen des Features, vor allem aus Universitätssicht, mindern würde.

Initiale Schätzung	1
Technologien	<ul style="list-style-type: none"><li>* Javascript</li><li>* Docker</li><li>* HTML-DOM</li></ul>
Abhängigkeiten	<ul style="list-style-type: none"><li>* BOT-55</li></ul>
Anforderungen	<ul style="list-style-type: none"><li>* Relevante Termine werden regelmäßig, automatisch bezogen und als Erinnerung gespeichert</li><li>* User können "globale Erinnerungen" aktivieren oder deaktivieren</li><li>* Das Feature ist resistent gegen Änderungen an den Domains oder deren Struktur → Fallen gescrapete Dienste länger aus wird dies reported</li><li>* Das Feature wird als nicht-eigenständig in den Reminder Service integriert</li><li>* Das Feature ist bzgl. Opt-in oder Opt-out in den Hilfe-Texten/Begrüßungsnachrichten des BHT-Bot dokumentiert</li></ul>
Tasks	<ul style="list-style-type: none"><li>* BOT-83 Prüfen ob Opt-In (nötig ist) oder Opt-Out (möglich ist)</li><li>* BOT-84 Quellen mit relevanten Terminen zusammentragen - auf Scrapebarkeit achten</li><li>* BOT-85 Scraping (mittels Scapring-Service) implementieren</li><li>* BOT-86 Termine aus Scraping in Erinnerungs Datenbank überführen</li><li>* BOT-87 Rasa Direktive Opt-In oder Opt-Out implementieren</li><li>* BOT-88 Error-Reporting implementieren, bei Missslungenen Scraping (über gewisse Zeit hinweg)</li><li>* BOT-114 Opt-In oder Opt-Out in Hilfe-Texten / Willkommensnachricht dokumentieren</li></ul>

#### BOT-89: Moodle iCal import als Erinnerungen

Moodle ist die zentrale Online-Lernplattform der Beuth Hochschule. Kurse, die in Moodle verwaltet werden erhalten in der Regel Abgabetermine für Aufgaben, die während des Semesters fällig werden. Oftmals stehen diese Abgabetermine bereits zu Beginn des Semesters in Moodle fest und können dort in einer Kalenderansicht betrachtet werden.

Moodle bietet außerdem eine Funktion zum Export der Eintragungen im eigenen Kalender:

<https://lms.beuth-hochschule.de/calendar/export.php>

Der User kann hier einen Link erzeugen, über den eine iCal Datei bezogen werden kann. Diese Datei enthält die Semestertermine und kann entsprechend in Erinnerungen des Bots umgewandelt werden

Initiale Schätzung	1
Technologien	<ul style="list-style-type: none"><li>* Javascript</li><li>* iCal</li><li>* Rasa</li></ul>
Abhängigkeiten	<ul style="list-style-type: none"><li>* BOT-55</li></ul>

Anforderungen	<ul style="list-style-type: none"> <li>* Der User kann dem Bot seinen Moodle-Kalender-Export-Link senden um einen Import auszulösen</li> <li>* Die iCal Datei hinter dem Export-Link wird in Erinnerungsinträge des Erinnerungsservice umgewandelt</li> <li>* Erinnerungen an Abgabetermine erfolgen zu Beginn des Tages / zeitlich versetzt vor der Fälligkeit der Abgabe, nicht zum im Kalender angegebenen Zeitpunkt</li> <li>* Das Moodle-Import Feature wird nicht-eigenständig in den Reminder-Service integriert</li> <li>* Das Feature (und seine Nutzung) ist in der BHT-Bot Hilfe gelistet</li> </ul>
---------------	---

Tasks	<ul style="list-style-type: none"> <li>* BOT-90 Import Moodle Rasa Direktive</li> <li>* BOT-91 Moodle iCal Download via zentralem Scaper Service</li> <li>* BOT-92 iCal Parsing und Persistierung als (sinnvolle) Erinnerung</li> <li>* BOT-111 Feature Release</li> <li>* BOT-113 Hilfe-Texte in (Chat)BHT-Bot help-command / willkommensnachricht</li> </ul>
-------	--

### BOT-13: Speech To Text (STT)

Es soll ermöglicht werden, dass Benutzern neben Textnachrichten auch mittels Sprachnachrichten mit dem BeuthBot kommunizieren können. Dabei sollen die Sprachnachrichten mittels eines neuen Services in Text übersetzt werden und dann wie andere Textnachrichten verarbeitet werden. Hierzu sollen 3 bekannte STT-Frameworks (Kaldi, Mozilla Voice STT und Wav2Letter) getestet und verglichen werden. Basierend darauf soll eine Entscheidung getroffen werden, welches Framework schlussendlich in der Production-Environment verwendet werden soll. Das Framework wird dann in Form eines neuen Micro-Services in den BeuthBot integriert.

Zuständigkeit	Robert Halwaß
Initiale Schätzung	3 Tage
Programmiersprachen	<ul style="list-style-type: none"> <li>* Python (Mozilla Voice STT)</li> <li>* C++ (Kaldi, WAV2Letter)</li> </ul>
Frameworks/Libraries	<ul style="list-style-type: none"> <li>* Kaldi</li> <li>* Mozilla Voice STT</li> <li>* WAV2Letter</li> </ul>
Services	* Speech
Abhängigkeiten	* BOT-43: Erstellung eines Common-Frameworks für (Content-)Services
Anforderungen	<ul style="list-style-type: none"> <li>* Die Übersetzung soll mittels neuronaler Netze geschehen</li> <li>* Nur Sprachnachrichten auf Deutsch sollen übersetzt werden</li> <li>* Das verwendete Framework muss OpenSource sein und Lokal auf dem BeuthBot-Server ausführbar sein</li> <li>* Es soll keine Model-Adaption durchgeführt werden</li> </ul>

### Tasks

**BOT-23: Komponente zur Umwandlung von Text in Sprache (TTS)**

Neben der bereits vorhandenen Funktion Textnachrichten vom Beuthbot zu erhalten, sollen Nutzer die Möglichkeit bekommen ebenfalls Sprachnachrichten zu empfangen. Hierfür wird eine Komponente zur Konvertierung von Text in Sprache (Eng: „Text-To-Speech (TTS)“) benötigt. Dieses Feature soll dem Nutzer in künftigen, dem Beuthbot hinzugefügten, Messenger-Diensten zur Verfügung stehen. Zur Umsetzung soll optimalerweise von einer Library Gebrauch gemacht werden, welche den Anforderungen gerecht wird.

Initiale Schätzung	1
Technologien	Javascript
Abhängigkeiten	keine
Anforderungen	<ul style="list-style-type: none"><li>*Support für die deutsche Sprache</li><li>*Sprachnachrichten lassen sich als MP3- und WAV-Dateien exportieren</li><li>*Das Feature ist bzgl. Opt-in oder Opt-out in den Hilfe-Texten/Begrüßungsnachrichten des Beuthbots dokumentiert</li></ul>
Nice to Haves	Sprachgeschwindigkeit und Stimme des Sprechers sind konfigurierbar
Tasks	<ul style="list-style-type: none"><li>* BOT-24 Recherche nach geeignetem Tool (TTS)</li><li>* BOT-25 Eigene Implementierung (TTS)</li><li>* BOT-98 Integration in Beuthbot</li></ul>

**BOT-75: Begrüßungsnachricht**

Neue Benutzer des BeuthBots sollen mit einer Begrüßungsnachricht empfangen werden. Diese Nachricht soll die Features des BeuthBots vorstellen und einen Shortcut nennen, welchen die Benutzer verwenden können, wenn sie Hilfe benötigen. Mit dem Shortcut listet der BeuthBot nochmals all seine Features auf.

Initiale Schätzung	1
Technologien	* JavaScript
Abhängigkeiten	Keine
Anforderungen	<ul style="list-style-type: none"><li>* Die Begrüßungsnachricht erscheint nur für neue Benutzer</li><li>* Das System sollte einen Shortcut zur Wiedervorstellung der Features bereitstellen, falls Benutzer Hilfe brauchen</li><li>* Das System muss in der Lage sein, auf die Hilfeanfrage des Benutzers mit Hilfe des Shortcuts innerhalb von 1,5 Sekunden zu antworten</li><li>* Das System sollte gut dokumentiert sein</li><li>* Das System sollte leicht zu verstehen sein</li></ul>
Tasks	<ul style="list-style-type: none"><li>* BOT-93 Client</li><li>* BOT-94 Server</li></ul>

**BOT-74: Webseite**



Um eine komplette Übersicht für alle genutzten BeuthBot-Ressourcen zu haben, soll eine Webseite zur Präsentation dieser Ressourcen erstellt werden. Zu den Ressourcen zählen das Ziemer's Wiki, Telegram, Discord, Github und die Implementation des Chatbots.

Initiale Schätzung	3
Technologien	<ul style="list-style-type: none"> <li>* TypeScript</li> <li>* JavaScript</li> </ul>
Abhängigkeiten	<ul style="list-style-type: none"> <li>* BOT-10</li> <li>* BOT-30</li> </ul>
Anforderungen	<ul style="list-style-type: none"> <li>* Jeder Ressource wird ein Menüpunkt gewidmet, welcher Infos &amp; einen Link zu der jeweiligen Ressource enthält</li> <li>* Anschauliches &amp; Einheitliches Design der Webseite</li> <li>* leicht austauschbare Komponenten</li> <li>* Das System sollte eine Ansicht innerhalb von 3 Sekunden laden</li> <li>* Das System sollte gut dokumentiert sein</li> <li>* Das System sollte leicht zu verstehen sein</li> </ul>
Tasks	<ul style="list-style-type: none"> <li>* BOT-76 Webseite Einrichten</li> <li>* BOT-77 Infos zum Wiki</li> <li>* BOT-78 Infos zu Telegram</li> <li>* BOT-79 Infos zu Discord</li> <li>* BOT-80 Infos zu GitHub</li> <li>* BOT-81 Implementation Chatbot</li> </ul>

#### BOT-11: Universeller Scraper & Download

Der Beuthbot soll einen „universellen“ Web-Scraper beinhalten, der als Grundlage für künftige Features dienen soll, die für konkrete Scraping-Funktionalitäten vorgesehen sind. Aufgrund der hohen Diversität an Datenstrukturen unterschiedlicher Webseiten, soll dieser möglichst abstrakte Funktionalitäten zur Extrahierung von Datensätzen bieten.

Initiale Schätzung	1
Technologien	Javascript
Abhängigkeiten	keine
Anforderungen	<ul style="list-style-type: none"> <li>* Import von HTML- und XML-Dateien</li> <li>* Daten lassen sich im JSON-Format exportieren</li> <li>* Datensätze sind per HTML-Tags und CSS-Selektoren extrahierbar</li> <li>* Dateien einer Webseite lassen sich downloaden</li> </ul>
Tasks	<ul style="list-style-type: none"> <li>* BOT-26 Recherche nach geeignetster Methode (HTML-JSON)</li> </ul>

**Nutzungshinweis:** Auf dieses vorliegende Schulungs- oder Beratungsdokument (ggf.) erlangt der Mandant vertragsgemäß ein nicht ausschließliches, dauerhaftes, unbeschränktes, unwiderrufliches und nicht übertragbares Nutzungsrecht. Eine hierüber hinausgehende, nicht zuvor durch *datenschutz-maximum* bewilligte Nutzung ist verboten und wird urheberrechtlich verfolgt.