

News Client für das BACnet

Wie strukturiert man einen dezentralen News Client, der auch ohne eine aktive Internetverbindung neue Nachrichten beziehen kann?

Maximilian Barth, Philipp Weber, Simon Laube

Prof. Dr. Christian Tschudin

Internet and Security

Abgabetermin: 18.07.2021



Universität
Basel

Einleitung

Im Rahmen der Vorlesung 'Internet and Security' hat sich unsere Gruppe - bestehend aus Maximilian Barth, Philipp Weber und Simon Laube - dazu entschlossen, einen News Client zu programmieren, der nicht von einem zentralen Server abhängt. Unsere Motivation war es, dass News auch in Gebieten zugänglich gemacht werden können, die entweder keinen Internetanschluss besitzen, oder in denen aus politischen Gründen gewisse News Server zensiert sind. Alle drei waren inspiriert von dieser Idee und wir konnten uns deshalb schnell auf dieses Thema einigen.

Das Ziel war es, dass jeder Nutzer, der Zugang zum Internet und zu den News Servern hat, die Artikel per Knopfdruck herunterladen kann. Diese soll er nun jedem beliebigen Nutzer direkt über Bluetooth oder über ein lokales Netzwerk weitergeben können, ohne dass der Server noch ein weiteres mal angefragt werden muss. Zusätzlich soll es möglich sein, Feeds zu erstellen, denen beliebige Artikel angehängt werden können. Via BACnet können diese Feeds von weiteren Usern eingesehen werden. Wir entschieden uns als erste News Quelle für die Website von SRF.

Das ganze soll in ein Programm mit schlichtem, benutzerfreundlichen Design verpackt werden, welches Funktionen wie Filtern nach Kategorie oder Erscheinungsdatum zulässt.

Hintergrund

Wichtig für unser Projekt waren JSON, HTML und die Idee des Scrapers.

JSON (Javascript Object Notation) ist ein kompaktes Datenformat für den Datenaustausch zwischen Anwendungen. Es ist einfach zu benutzen, man muss also keine neue Sprache lernen, um .json Dateien schreiben zu können. Aus diesem Grund haben wir uns dazu entschieden, dieses zu benutzen, um unsere Artikel abzuspeichern.

HTML bedeutet Hyper Text Markup Language, ist zu Deutsch also eine sogenannte Auszeichnungssprache, die dazu verwendet wird, Webinhalte zu strukturieren. Mit Hilfe dieser kann man also festlegen, wo auf der Webseite z.B. ein Bild hin soll, welcher Text ein Titel ist, welcher ein Untertitel, etc. HTML-Dokumente bilden die Grundlage des World Wide Web.

Scraper werden dazu verwendet, Daten von Webseiten zu holen. Es gibt einige Libraries, um diese selber zu programmieren und zu konfigurieren, es gibt aber auch einige andere Tools, mit denen man scrapen kann. Beim Web Scraping wird nach dem HTML-File einer Webseite gesucht und dieses dann extrahiert. Der Benutzer hat nun das ganze HTML-File und kann damit den Inhalt dieser Webseite vollständig oder auch nur bestimmte Inhalte replizieren. Natürlich ist das Web Scraping nicht immer legal. Wenn zum Beispiel von einer News-Seite, für deren Benutzung man bezahlen muss, die Artikelinhalte heruntergeladen und veröffentlicht werden, ist das nicht legal.

Um rechtlichen Problemen aus dem Weg zu gehen, müssen natürlich die Nutzungsbedingungen eingehalten werden. Da es sich in unserem Fall um eine empirische Forschung handelt, befinden wir uns hier allerdings rechtlich noch in einer Grauzone.

Implementation

Wir haben das ganze Projekt in vier Teile aufgeteilt, den Scraper, die Logik, den Transfer von Daten und dem UI.

Den Scraper benutzen wir, um die Artikel von einer News Webseite zu holen. Das funktioniert so, dass dieser zuerst auf einen RSS-Feed zugreift. Dort sind Kurzbeschreibungen neuer Artikel mit den jeweiligen URLs aufgeführt. Wir können danach auf das HTML-File der Webseite jener Artikel, die man unter den jeweiligen URLs findet, zugreifen. Von diesen wird dann der Inhalt des Artikels, den wir für unseren News Client benötigen, herausgefiltert. Der Text wird dann in der Logik weiterverarbeitet und für das UI zugänglich gemacht.

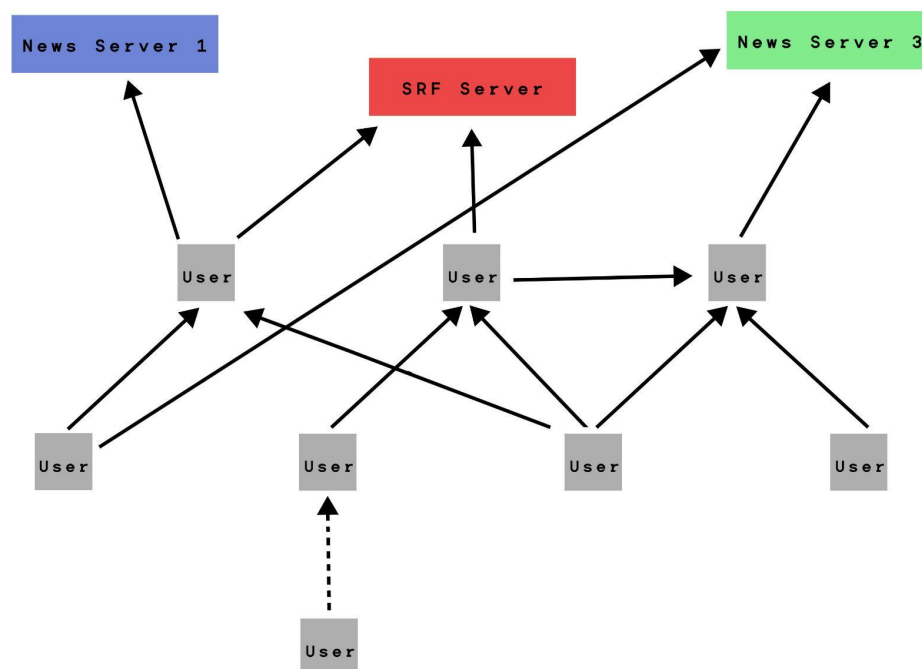
Der Logik Teil ist eine Schnittstelle zwischen dem Scraper, der UI und dem Betriebssystem. Er ist dafür verantwortlich, dass Dateien (Artikel, Key-Files, zip-Files) betriebssystem-abhängig am richtigen Ort gespeichert werden. Eine Klasse 'Article' wird von ihm bereitgestellt, um aus dem Internet geladene Artikel in json oder html zu konvertieren. Die json Formatierung wird verwendet, um Artikel zu speichern und html wird benötigt um diese mit der Library 'PyQt5' auf dem Bildschirm darzustellen.

Bei der Übertragung von Artikeln via Bluetooth oder via das lokale Netzwerk ist es die Aufgabe des Logik Teils, nur die neuen Artikel für den Versand zu sammeln, komprimieren und beim Empfang wieder richtig zu entpacken.

Für den Transfer der json Dateien, also den Artikeln, benutzen wir Python-Sockets. Ein Server kann einem Client, der sich mit ihm verbindet, neue Artikel weitergeben. Wir haben zwei Übertragungsarten implementiert: Bluetooth und Übertragung über das lokale Netzwerk.

Für die Bluetooth-Übertragung brauchen wir die MAC-Adresse des Servers. Bei dessen Start wird ihm seine MAC-Adresse auf dem Bildschirm angezeigt. Diese kann nun vom Client eingegeben werden, um sich mit dem Server zu verbinden. Sobald die Verbindung steht, werden die Daten übertragen.

Für die Local Network Übertragung suchen wir mittels ARP-Request nach Geräten im selben Netzwerk. Diese zeigen wir dem Client an. Er kann danach das gewünschte Gerät, auf welchem ein BACNews-Server läuft, auswählen und von diesem neue Artikel herunterladen. Falls das Gerät, mit dem sich der Client verbinden möchte, nicht in der Liste auftaucht, kann er die IP-Adresse von Hand eingeben.



Resultate

Im folgenden Abschnitt werden die verschiedenen Teile der Applikation vorgestellt und auf deren Funktionalität eingegangen. Grob lässt sich das Programm in vier Abschnitte einteilen:

“read”

Der Reader bildet das Herzstück des Programms. Hier werden heruntergeladene Artikel dargestellt. Es besteht auch die Möglichkeit, die Artikel nach Datum (*today, week, all*) und / oder nach Kategorien (*Sport, Wirtschaft, Ausland, ...*) zu filtern. Zusätzlich wird mittels eines Punktes angezeigt, ob ein Artikel bereits geöffnet wurde. Auf der rechten Seite des dargestellten Textes wird den User*innen die Möglichkeit angeboten, den Artikel zu bookmarken, ihn in einen ihrer BACnet Feeds anzuhängen oder den Artikel in einem neuen Fenster zu öffnen. Falls noch kein Benutzername oder Feed vorhanden ist, wird auf das passende Fenster gewechselt.

“BACnet”

In diesem Abschnitt des Programms können Benutzer*innen mit dem BACnet interagieren. Beim ersten Öffnen des Fensters wird man gebeten, einen Benutzernamen einzugeben. Dieser Vorgang muss nicht wiederholt werden, solange die lokale Datenbank und die Schlüssel nicht gelöscht werden. Das Layout ist sehr ähnlich wie in der “reading section”: Auf der linken Seite des Fensters befindet sich eine Liste von Artikeln, die im ausgesuchten Feed enthalten sind und auf der Rechten wird der ausgewählte Text dargestellt. In der Combo-Box über der Liste stehen alle erstellten und importierten Feeds zur Verfügung. Zusätzlich wird auch der Benutzername der Person angezeigt, die den dazugehörigen Feed erstellt hat. Um einen neuen Feed zu kreieren, muss lediglich der “*create new feed*” Button gedrückt werden. Artikel können in der “reading section” einem Feed angehängt werden und werden daraufhin im BACnet Fenster dargestellt. Die importierten Artikel werden absichtlich nicht in der “reading section” dargestellt, da es theoretisch möglich ist, dass Benutzer die json Files ihrer lokalen Artikel verändern. Durch die strikte Trennung der importierten Artikel von den Lokalen ist es den Benutzer*innen des Programms möglich auf Basis des Usernames und Vertrauens sicherzustellen, dass keine Falschnachrichten gelesen werden.

“archive”

Falls User*innen einen Artikel in der “reading section” bookmarken, wird er im Archiv dargestellt. Dies ermöglicht es alte Artikel, die man vielleicht noch nicht lesen konnte oder man für besonders lesenswert hielt, schnell zu finden. Des Weiteren stehen die gewohnten Funktionen, wie einen Artikel einem BACnet Feed anzuhängen, auch hier zur Verfügung.

“get new articles”

Dieser Abschnitt der Applikation ist für das Herunterladen und Teilen neuer Artikel zuständig. Im Wesentlichen gibt es zwei Tabs:

Import

Es gibt vier verschiedene Wege, wie neue Artikel heruntergeladen werden können:

1. SRF

Hier werden neue Artikel direkt von der SRF Website heruntergeladen und lokal als json Dateien gespeichert. Die neuen Artikel werden direkt im Reader dargestellt und können einem BACnet Feed angehängt werden. Falls der Download fehlschlägt, wird eine Fehlnachricht angezeigt.

2. BACnet

Nach dem Klicken des Buttons werden die User*innen aufgefordert ein Verzeichnis auszuwählen. Dort wird nach .pcap Dateien gesucht und diese werden, wenn vorhanden, importiert. Neue Feeds stehen daraufhin im BACnet Fenster des Programms zur Verfügung.

3. local network

Um diese Funktion nutzen zu können, muss ein Server bereits im gleichen LAN gestartet worden sein. Ein solcher kann initiiert werden, in dem man auf *“export”* wechselt und dort *“local network”* auswählt. Im LAN gefundene Geräte werden in einer Tabelle angezeigt. Es ist möglich, eine Verbindung mit einem Server aufzubauen, in dem man den Eintrag mit der korrespondierenden IP Adresse auswählt und auf *“connect”* klickt. Falls ein Gerät nicht angezeigt wird, kann dessen IP Adresse auch manuell eingegeben werden. Dazu muss nur der *“manually input IP”* Knopf betätigt werden. Auch hier wird im Falle, dass die Übertragung schief läuft eine Fehlnachricht angezeigt. Falls ein Duplikat eines Artikels übertragen wird, wird dieses nicht gespeichert.

4. bluetooth

Die Bluetooth Übertragung funktioniert gleich wie die *“local network”* Variante. Es muss schlicht die MAC Adresse des Servers eingegeben werden, um eine Verbindung aufzubauen. Leider ist dieses Feature auf MacOS nicht verfügbar (der Button wird daher auf MacOS nicht dargestellt). Es war uns nicht möglich, dieses Feature ausreichend auf Linux zu testen.

Export

1. BAC-Net

Diese Funktion ist sehr ähnlich zur *“import”* Version. Die User*innen wählen ein Verzeichnis aus und alle von dem / der User*in erstellten Feeds werden dorthin als pcap Dateien exportiert.

2. local network

Hier wird ein LAN Server gestartet und die dazugehörige IP Adresse im UI dargestellt. Diese können Benutzer*innen in ihrem Programm eingeben um eine Verbindung aufzubauen.

3. bluetooth

Dieser Modus funktioniert gleich wie die LAN Variante. Anstatt einer IP Adresse wird natürlich eine MAC Adresse dargestellt.

Das letzte Element des UIs, das noch nicht erwähnt wurde, ist der *“dark”* bzw. *“light”* Button. Dieser befindet sich oben rechts im Menü. Durch diesen Knopf kann zwischen der dunklen und hellen Version der Applikation gewechselt werden. Dabei wurden die Farben so gewählt, dass es für die Augen angenehm ist, lange Texte zu lesen.

Schlussfolgerung

Positiv war die Aufteilung der Aufgaben. Da wir das Projekt in vier Teile geteilt haben, ordneten wir zuerst jedem einen Teil zu und wer fertig war, konnte am vierten Teil arbeiten. So war es möglich, den Code schön zu trennen, sodass nie zwei am gleichen File gearbeitet haben und es zu keinen Merge-Konflikten kam. Mittels geeigneter Schnittstellen konnten wir das Projekt danach einfach zusammensetzen.

Weniger gut war unsere Planung mit dem BACnet. Uns war von Anfang an nicht ganz klar, wie das BACnet genau integriert werden soll, bzw. wie es überhaupt funktioniert. Wir haben es also aufgeschoben und zuerst den Teil des Projektes gemacht, den wir besser verstanden haben. Gegen Ende hin mussten wir dann den ganzen BACnet Teil integrieren, was dann auch zeitlich etwas knapp wurde. Hier hätten wir uns von Anfang an besser im Klaren sein sollen, wie wir das BACnet genau nutzen möchten und wie man es am besten integriert.

Wir stellten fest, dass der Scraper, auch wenn er einen Tag vorher perfekt funktioniert hat, am nächsten vielleicht nicht mehr funktioniert. Das Problem hierbei ist, dass auf den News Seiten die HTML-Files manchmal umgeschrieben werden. So war zum Beispiel bei SRF der für uns wichtige Artikel Inhalt plötzlich in einer sogenannten 'section', anstatt wie vorhin in einem 'div'. Dies ist zwar nicht schwer zu beheben, heisst aber, dass der Scraper immer wieder angepasst werden muss.

Ausblick

Der News Client könnte in Zukunft mit dem Scrapen weiterer News Seiten ergänzt werden.

Artikel, welche von dem Autor modifiziert wurden, sollen beim Herunterladen, die alte Version ersetzen und nicht als Duplikat ignoriert werden.

Bluetooth funktioniert momentan nur auf Windows und soll auch auf anderen Plattformen laufen.

Artikel können zusätzlich mit (komprimierten) Bildern ergänzt werden.

Eine Möglichkeit, Artikel zu verifizieren, damit diese nicht verfälscht weitergegeben werden können, soll geschaffen werden.

Referenzen

Benutzte Libraries:

- BeautifulSoup4
- PyQt5
- QtAwesome
- requests
- cbor2
- PyNaCl
- SQLAlchemy
- netifaces

Dependencies aus dem letzten Semester:

- Gruppe 04: logMerge
- Gruppe 07: logStore
- Gruppe 14: feedCtrl

Wir haben einige Klassen und Funktionen dieser Gruppen benutzt, um das BACnet in unserem Projekt zu implementieren.