

Zadanie programistyczne nr 4

z Sieci komputerowych

1 Opis zadania

Napisz program `webserver` będący prostym serwerem HTTP, wyświetlającym strony z zadanego katalogu.

Rozpocznij od następujących czynności.

1. Ze strony wykładu pobierz plik `webpages.tar.gz` zawierający strony WWW i rozpakuj go do wybranego katalogu.¹
2. Umieść w pliku `/etc/hosts` następującą zawartość.

```
127.0.0.1 localhost
127.0.1.1 virbian
127.0.1.1 virtual-domain.example.com
```

Od tego momentu odwołania do domen `virbian` i `virtual-domain.example.com` kierowane będą do lokalnego komputera.

Serwer powinien akceptować dwa parametry wywołania: `port` i `katalog`. Pierwszy z nich jest numerem portu, na którym serwer będzie oczekiwać na przychodzące połączenia, drugi zaś katalogiem zawierającym strony WWW (np. te pobrane). Program powinien obsługiwać błędne dane wejściowe, typu nieistniejący katalog, zgłaszając odpowiedni komunikat.

Otwarcie w przeglądarce WWW (działającej na tym samym komputerze lub na tej samej maszynie wirtualnej) adresu `http://nazwa_domeny:port/strona.html` powinno spowodować wyświetlenie zawartości strony `katalog/nazwa_domeny/strona.html`.

1.1 Implementacja

Nie trzeba implementować pełnego protokołu HTTP. Twój serwer powinien natomiast obsługiwać żądania GET. Wystarczy obsługiwać pierwszy wiersz takiego zapytania (czyli pole zawierające adres strony), pole `Host` i pole `Connection` (patrz niżej). W szczególności nie trzeba obsługiwać warunkowych żądań GET, np. pól `If-Modified-Since`.

Twój serwer powinien odsyłać następujące informacje: kod odpowiedzi, pole `Content-Type`, pole `Content-Length` i inne potrzebne dane (przykładowo pole `Location` dla odpowiedzi 301).

1. Zwracane powinny być następujące kody odpowiedzi (w razie potrzeby można zaimplementować też inne zdefiniowane w standardzie HTTP).
 - ▶ 200 OK: w przypadku powodzenia;
 - ▶ 301 Moved Permanently: jeśli przeglądarka chce pobrać obiekt, który jest katalogiem należy przekierować ją do strony `index.html`, która znajduje się w tym katalogu;
 - ▶ 403 Forbidden: jeśli przeglądarka będzie chciała pobrać adres prowadzący do pliku leżącego poza katalogiem danej domeny;

¹Nie zaleca się używania istniejących w starszych wersjach Virbiana stron WWW z katalogu `/var/local/webpages/`; nie są one w pełni zgodne z niniejszą instrukcją.

- ▶ 404 Not Found: jeśli przeglądarka będzie chciała pobrać nieistniejący plik;
- ▶ 501 Not Implemented: jeśli przeglądarka wyśle dane niezrozumiałe dla serwera.

Również w przypadku odpowiedzi o kodach różnych od 200 Twój serwer powinien wysyłać zawartość HTML, stanowiącą prosty opis błędu, który wystąpił. Zawartość ta zostanie wyświetlona przez przeglądarkę.

2. Pole `Content-Type` powinno być ustalane na podstawie rozszerzenia pliku. Poprawnie powinny być obsługiwane pliki `txt`, `html`, `css`, `jpg`, `jpeg`, `png` i `pdf`. Pozostałe pliki mogą mieć typ `application/octet-stream`). W przypadku plików `txt` i `html` można założyć, że są one zapisane w UTF-8 i nie trzeba tego sprawdzać. Przykładowo w przypadku strony HTML zawartość pola `Content-Type` może być równa `text/html; charset=utf-8`.

Twój serwer nie powinien rozłączać się po obsłużeniu pojedynczego zapytania, lecz utrzymywać połączenie z nieaktywnym klientem przez pewien zdefiniowany czas (np. 250-1000 milisekund).² Wyjątkiem jest obsługa zapytania, w którego nagłówku znajduje się pole `Connection: close`. W takim przypadku należy zamknąć połączenie zaraz po obsłużeniu zapytania. W tym samym połączeniu Twój serwer powinien potrafić obsłużyć wiele zapytań (np. o stronę WWW i znajdujące się na niej obrazki). Twój program nie musi obsługiwać połączeń od wielu klientów jednocześnie.

Pamiętaj, że serwerowi HTTP nie wolno wysłać zawartości pliku leżącego poza katalogiem ze stronami WWW danej domeny. Twój serwer powinien być odporny na złośliwego użytkownika zdalnego, który wyśle śmieci zamiast poprawnego żądania HTTP. Twój serwer może wyświetlać komunikaty diagnostyczne (np. informacje o żądaniach HTTP i odpowiedziach) na standardowe wyjście lub wyjście błędu, ale powinny być one zwięzłe i przejrzyste.

2 Uwagi techniczne

Ta sekcja jest taka sama dla wszystkich zadań programistycznych.

Pliki. Sposób utworzenia napisu oznaczanego poniżej jako *imie_nazwisko*: Swoje (pierwsze) imię oraz nazwisko zapisz wyłącznie małymi literami, zastępując litery ze znakami diakrytycznymi przez ich łacińskie odpowiedniki (usuń ogonki). Pomiedzy imię i nazwisko wstaw znak podkreślenia.

Prowadzącemu ćwiczeniopracownię należy dostarczyć plik *imie_nazwisko.tar.xz* z archiwum (w formacie `tar`, spakowane programem `xz`) zawierającym pojedynczy katalog o nazwie *imie_nazwisko* z następującymi plikami.

- ▶ Kod źródłowy w C, C++ lub Rust, czyli pliki `*.c`, `*.h`, `*.cpp`, `*.h` lub `*.rs`. Każdy plik źródłowy na początku powinien zawierać w komentarzu imię, nazwisko i numer indeksu autora lub autorki.
- ▶ Plik `Makefile` pozwalający na kompilację programu po uruchomieniu `make`. Takie wymaganie obowiązuje również w przypadku kodów źródłowych w Rust (wywołanie `make` może uruchamiać `cargo` lub `rustc`).
- ▶ Ewentualnie plik `README.txt` lub `README.md`.

W katalogu tym **nie** powinno być żadnych innych plików, w szczególności skompilowanego programu, obiektów `*.o`, czy plików źródłowych nienależących do projektu.

²Uwaga: niektóre przeglądarki otwierają więcej niż jedno połączenie do serwera. Jeśli Twój program będzie obsługiwać tylko jedno z tych połączeń naraz, zbyt długie oczekiwanie na zamknięcie połączenia spowolni cały proces. W poważniejszych zastosowaniach należałoby obsługiwać więcej niż jedno połączenie naraz; tutaj nie jest to wymagane.

Kompilacja. Kompilacja i uruchamianie przeprowadzone zostaną w 64-bitowym środowisku Linux; w szczególności kody powinny dać się uruchomić w maszynie wirtualnej Virbian bez instalacji dodatkowych pakietów.

- ▶ Kompilacja w przypadku C ma wykorzystywać standard C17 lub C18 z ewentualnymi rozszerzeniami GNU (opcja kompilatora `-std=c17`, `-std=gnu17`, `-std=c18` lub `-std=gnu18`).
- ▶ Kompilacja w przypadku C++ ma wykorzystywać standard C++14, C++17 lub C++20 z ewentualnymi rozszerzeniami GNU (opcja kompilatora `-std=c++14`, ..., `-std=c++20`, `-std=gnu++14`, ..., `-std=gnu++20`).
- ▶ Kompilacja w przypadku Rust ma wykorzystywać edycję 2018, 2021 lub 2024 (opcja kompilatora `-edition 2018`, ..., `-edition 2024`).

Kompilacja w przypadku C i C++ powinna korzystać z opcji `-Wall` i `-Wextra`. Podczas kompilacji w dowolnym języku nie powinny pojawiać się ostrzeżenia.

3 Sposób oceniania programów

Poniższe uwagi służą ujednoczeniu oceniania w poszczególnych grupach. Napisane są jako polecenia dla prowadzących, ale studenci powinni **koniecznie** się z nimi zapoznać, gdyż prowadzący będą się ich trzymać przy sprawdzaniu.

- ▶ Programy będą testowane na zajęciach w obecności autora lub autorki programu.
- ▶ Program niekompilujący się otrzymuje 0 punktów, nawet jeśli ładnie wygląda.
- ▶ Program uruchamiany jest w różnych warunkach opisanych poniżej i otrzymuje za te uruchomienia od 0 do 10 punktów.
- ▶ Następnie obliczane są ewentualne punkty ujemne zgodnie z listą podaną poniżej.
- ▶ Oceniamy z dokładnością do 0,5 punktu. Jeśli ostateczna liczba punktów wyjdzie ujemna, wstawiamy zero. (Ostatnia uwaga nie dotyczy przypadków plagiatów lub niesamodzielnych programów).

Testowanie: punkty dodatnie. Do testów należy uruchomić maszynę wirtualną i przygotować na niej środowisko opisane w tym zadaniu, tj. umieścić w pliku `/etc/hosts` odpowiednie wpisy, a następnie pobrać i rozpakować plik `webpages.tar.xz`.

Uruchomić serwer nasłuchujący na porcie 8888. Testować należy przeglądarką WWW uruchomioną na tej samej maszynie wirtualnej. Należy każdorazowo sprawdzać (np. w narzędziach deweloperskich uruchamianych klawiszem F12), czy wysyłane przez serwer komunikaty są zgodne ze specyfikacją, np. czy pole `Content-Type` jest poprawnie ustawiane.

Należy wykonać następujące testy.

3 pkt. Wejść na stronę `http://virbian:8888/`. Sprawdzić, czy serwer przekieruje nas za pomocą komunikatu 301 do strony `http://virbian/index.html` i czy wyświetlana później strona wygląda następująco:

Strona główna (virbian)

Sekcja 1

Ikonka routera na licencji public domain.



Sekcja 2

Kliknij poniższy przycisk, aby przejść do podstrony z ciekawymi linkami.

Podstrona z ciekawymi linkami

Sekcja 3

Kliknij poniższy przycisk, aby przejść do podstrony page.html z domeny virtual-domain.example.com.

Przejdź do virtual-domain

Ten dokument jest przerobionym przykładem ze strony <https://getbootstrap.com/>

Jeśli pobranie zajmuje dłużej niż kilka sekund, należy przydzielić maksymalnie 1 punkt. Jeśli strona nie przypomina powyższego obrazka, nie sprawdzamy dalej.

- 1 pkt.** Kliknąć przycisk *Podstrona z ciekawymi linkami* i sprawdzić, czy strona `page.html` wyświetla się prawidłowo.
- 2 pkt.** Sprawdzić, czy kliknięcie przycisków *Dokument tekstowy*, *Dokument PDF* i *Plik binarny* działa poprawnie. W przypadku dwóch pierwszych przeglądarka powinna otworzyć plik tekstowy i PDF-a (lub poprosić o uruchomienie przeglądarki PDF-ów). Plik binarny powinien mieć typ `application/octet-stream`; przeglądarka powinna zaproponować nam jego zapisanie. Na końcu należy powrócić na stronę główną.
- 1 pkt.** Kliknąć przycisk *Przejdź do virtual-domain*. Powinna wyświetlić się zawartość strony `page.html` z domeny `virtual-domain.example.com`.
- 1 pkt.** Wejść na stronę `http://virtual-domain.example.com:8888/`. Serwer powinien przekierować nas komunikatem 301 do strony `http://virtual-domain.example.com/index.html`, a następnie zwrócić komunikat 404 ze względu na brak pliku `index.html`.
- 1 pkt.** Sprawdzić, czy zapytanie o stronę `http://virtual-domain.example.com:8888/../virbian/index.html` jest poprawnie obsługiwane, tj. nie powoduje wyświetlania żądanej strony. Uwaga: to zapytanie należy wykonać w trybie wsadowym, np. za pomocą polecenia


```
curl -v --path-as-is \
  http://virtual-domain.example.com:8888/../virbian/index.html
```

 bo większość przeglądarek usunie ciąg `../` z pola adresu.
- 1 pkt.** Połączyć się z serwerem za pomocą programu `telnet` i wysłać mu śmieci zakończone pustym wierszem. Serwer powinien odpowiedzieć komunikatem 501. Sprawdzić, czy serwer potem jest nadal w stanie odpowiadać na żądania.

Testowanie: punkty ujemne. Punkty ujemne przewidziane są za następujące usterki.

- 2 pkt.** Zamykanie połączenia po każdym żądaniu bez czekania na upływanie zadanego czasu (np. 1 sekundy).
- 1 pkt.** Brak poprawności sprawdzania argumentów wywołania programu.

- do -3 pkt.** Zła/nieczytelna struktura programu: brak modularności i podziału na funkcjonalne części, niekonsekwentne wcięcia, powtórzenia kodu.
- 2 pkt.** Aktywne czekanie zamiast zasypiania do momentu otrzymania pakietu.
- 1 pkt.** Brak sprawdzania poprawności wywołania funkcji systemowych, takich jak `recvfrom()`, `write()` czy `bind()`.
- 1 pkt.** Nietrzymanie się specyfikacji wejścia i wyjścia. Przykładowo: wyświetlanie nadmiarowych informacji diagnostycznych lub inna niż w specyfikacji obsługa parametrów.
- 1 pkt.** Zły plik `Makefile` lub jego brak.
- ▶ Program powinien się kompilować poleceniem `make`.
 - ▶ Poszczególne pliki `*.c` i `*.cpp` powinny kompilować się do obiektów tymczasowych `*.o` a następnie powinny być konsolidowane do wykonywalnego programu.
 - ▶ Polecenie `make clean` powinno czyścić katalog z tymczasowych obiektów (plików `*.o`), zaś polecenie `make distclean` powinno usuwać te obiekty i wykonywalny program pozostawiając tylko pliki źródłowe.
- 3 pkt.** Kara za wysłanie programu po terminie; opóźnienie nie może być większe niż 1 tydzień.

Materiały do kursu znajdują się w systemie SKOS: <https://skos.ii.uni.wroc.pl/>.

Marcin Bieńkowski