

Projekt na získanie zápočtu z predmetu Počítačové siete LS 2019/20

6. apríla 2021

1 Získanie zápočtu

1.1 Formálne aspekty

- Odovzdávať sa bude prostredníctvom githubu, sprístupnením repozitára menom `mboxer` užívateľovi `gjenca`
- Repozitár bude obsahovať iba jeden súbor so zdrojovým kódom, nič iné.
- Termín odovzdania projektu je 23.4.2021 23:59:59.
- Po tomto termíne budú projekty vyhodnotené do troch dní, známka bude zapísaná do AIS a študentom budú oznámené prípadné požadované zmeny pre zvýšenie počtu bodov.
- V prípade, že niekto bude mať záujem o zvýšenie počtu bodov, bude vypísaný ďalší termín dokedy budú môcť záujemcovia odovzdať druhú verziu programu.

1.2 Kritériá hodnotenia

- Posudzovať sa bude korektnosť implementácie v zmysle špecifikácie protokolu.
- Posudzovať sa bude aj zdrojový kód: pokiaľ niekto odovzdá príliš škaredý kód, dostane horšiu známku.
- Používajte, (adekvátne účelu) prostriedky Pythonu: funkcie, slovníky, triedy, zabudované dátové typy.

2 Zadanie

Napíšte forkujúci sa TCP server, ktorý bude vykonávať službu zapisovania a čítania textových správ do schránok. Každá schránka je reprezentovaná nejakým podadresárom aktuálneho adresára procesu, správy sú zapisované do týchto adresárov, každá správa je práve jeden súbor. Názvy nových súborov/schránok sú pomocou hašovacej funkcie generované z ich obsahu, tým je zabezpečené, že dva procesy (klienti) nebudú súčasne zapisovať do toho istého súboru.

3 Protokol

3.1 Základné pravidlá

- Protokol bude pre texty používať kódovanie UTF-8, koniec riadku je <LF> (ASCII kód 10), ako v Unixe.
- Bude postavený nad TCP.
- Klient bude serveru posilať *požiadavky* (anglicky *request*) a server po každej požiadavke pošle naspäť *odpoveď* (anglicky *response*).

3.2 Požiadavka

- Požiadavka bude vždy obsahovať v prvom riadku *metódu* (anglicky *method*), t.j. jedno slovo z množiny slov `READ`, `LS`, `WRITE`.
- Nasledovať bude bližšia špecifikácia požiadavky, pomocou niekoľkých *hlavičiek* (headers). Hlavičky nemusia byť uvedené (môže ich byť 0). Každá hlavička je na jednom riadku. Hlavička sa skladá z *identifikátora* a *hodnoty*, sú oddelené dvojbodkou. Identifikátor musí byť ASCII reťazec bez bielych znakov a nesmie obsahovať znak `:` (dvojbodka).¹ Príklad hlavičky:

```
Mailbox:my_own_mailbox
```

- Po hlavičkách bude nasledovať vždy práve jeden prázdny riadok.
- Po prázdnom riadku môže (ale nemusí) nasledovať *obsah požiadavky*. To, či požiadavka má obsah a aký je dlhý je dané metódou a obsahom hlavičiek. **POZOR!** Obsah požiadavky nemusí byť textový, môžu to byť binárne dáta.

¹Vid metódy `str.isascii` a `str.isspace` v dokumentácii Pythonu

3.3 Odpoveď

- Odpoveď bude mať v prvom riadku *stav* (anglicky status).
- Stav obsahuje ako prvé slovo trojciferné číslo, nasledované práve jednou medzerou a popisom. Príklad stavu:

```
200 Bad request
```

- Nasledujú hlavičky, v rovnakom tvare ako pri požiadavke.
- Potom bude vždy práve jeden prázdny riadok.
- Nasledovať obsah odpovede (alebo nič), dĺžka vyplýva z metódy a hlavičiek. Obsah odpovede môže byť binárny, nemusí byť textový.
- Spojenie za normálnych okolností ukončí klient, zavretím socketu.

3.4 Metódy

Teraz ideme špecifikovať po jednotlivých metódach typy rôznych požiadaviek, možných odpovedí a tomu zodpovedajúcich efektov a prípadne obsahov odpovedí. Ak nie je v tabuľke v poslednom stĺpci uvedený žiaden obsah odpovede, obsah odpovede je vtedy prázdny.

3.4.1 WRITE

Hlavičky:

- **Názov hlavičky** Mailbox
Popis hlavičky Názov schránky/adresára kam treba zapísať obsah.
Nesmie obsahovať znak /.
- **Názov hlavičky** Content-length
Popis hlavičky Dĺžka obsahu požiadavky. Celé nezáporné číslo.

Správanie sa v rôznych situáciách:

- **Situácia** V hlavičkách sú neprípustné hodnoty, alebo niektorá hlavička chýba.
Odpoveď (status) 200 Bad request
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Žiaden efekt.
- **Situácia** Mailbox/adresár neexistuje.
Odpoveď (status) 203 No such mailbox
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Žiaden efekt.
- **Situácia** Normálny stav
Odpoveď (status) 100 OK
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Prečíta sa toľko bytov (nie znakov!) zo socketu, ako špecifikuje hodnota hlavičky **Content-length**. Vytvorí sa nová správa/súbor v mailboxe s názvom, ako špecifikuje hodnota hlavičky **Mailbox**. Identifikátor správy/názov súboru je MD5 hash celého obsahu.

3.4.2 Pomôcky

Pre určenie názvu súboru použite `hashlib.md5`.

```
>>> import hashlib
>>> m=hashlib.md5()
>>> m.update(b'prvy riadok')
>>> m.update(b'aj newline\n')
>>> m.update(b'nieco ine')
>>> m.hexdigest()
'c19c1322a38085fc0ae2ea60a87c1429'
```

3.4.3 READ

Hlavičky:

- **Názov hlavičky** Mailbox
Popis hlavičky Názov schránky/adresára odkiaľ treba prečítať obsah. Nesmie obsahovať znak /.
- **Názov hlavičky** Message
Popis hlavičky Názov správy v schránke odkiaľ treba prečítať obsah. Nesmie obsahovať znak /.

Správanie sa v rôznych situáciách:

- **Situácia** V hlavičkách sú neprípustné hodnoty, alebo niektorá hlavička chýba.
Odpoveď (status) 200 Bad request
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Žiaden efekt.
- **Situácia** Správa (alebo mailbox) neexistuje (`FileNotFoundError`).
Odpoveď (status) 201 No such message
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Žiaden efekt.
- **Situácia** Správa (alebo mailbox) sa nedá prečítať z iných dôvodov (`OSError`).
Odpoveď (status) 202 Read error
Odpoveď (hlavičky) Žiadne hlavičky odpovede.
Odpoveď (obsah) Prázdny obsah odpovede.
Efekt Žiaden efekt.
- **Situácia** Normálny stav
Odpoveď (status) 100 OK
Odpoveď (hlavičky) Content-length: *dĺžka správy v bytoch*
Odpoveď (obsah) Obsah správy/súboru.
Efekt Prečíta správu zo súboru, zapíše do socketu.

3.4.4 Pomôcky

Pre otváranie súboru pre čítanie použite `with/open` idióm, aby ste nemuseli myslieť na to, že ho treba zavrieť.

```
>>> with open('xxx/3749f52bb326ae96782b42dc0a97b4c1') as f:  
...     content=f.read()
```

3.4.5 LS

Hlavičky:

- **Názov hlavičky** Mailbox

Popis hlavičky Názov schránky/adresára z ktorého chceme vypísať názvy správ/súborov. Nesmie obsahovať znak /.

Správanie sa v rôznych situáciách:

- **Situácia** V hlavičkách sú neprípustné hodnoty, alebo niektorá hlavička chýba.

Odpoveď (status) 200 Bad request

Odpoveď (hlavičky) Žiadne hlavičky odpovede.

Odpoveď (obsah) Prázdny obsah odpovede.

Efekt Žiaden efekt.

- **Situácia** Mailbox/adresár neexistuje.

Odpoveď (status) 203 No such mailbox

Odpoveď (hlavičky) Žiadne hlavičky odpovede.

Odpoveď (obsah) Prázdny obsah odpovede.

Efekt Žiaden efekt.

- **Situácia** Normálny stav

Odpoveď (status) 100 OK

Odpoveď (hlavičky) Number-of-messages: n , kde n je počet správ/súborov (riadkov odpovede).

Odpoveď (obsah) Názvy všetkých správ/súborov v adresári/schránke, v kódovaní UTF-8, na každom riadku jedna správa.

Efekt Žiaden efekt.

3.4.6 Pomôcky

Zoznam súborov adresári zistíte pomocou `os.listdir`.

```
>>> import os
>>> os.listdir("my_mailbox")
['d41d8cd98f00b204e9800998ecf8427e',
 'dc3d1a12cf64c45e3075f6f49fcf2186',
 '16c720ea0c26367dff45a2668e6f5e6e']
```

3.4.7 Neznáma metóda

Ak klient pošle na začiatku požiadavky čokoľvek iné, ako je riadok obsahujúci známu metódu, server vráti status `204 Unknown method` s prázdnyimi hlavičkami a obsahom a *uzavrie spojenie*.

3.5 Príklad komunikácie klient-server

C→S sú označené dáta zasielané klientom serveru, S→C naopak.

```
C->S: WRITE
C->S: Mailbox:xxx
C->S: Content-length:11
C->S:
C->S: 0123456789
S->C: 100 OK
S->C:
C->S: LS
C->S: Mailbox:xxx
C->S:
S->C: 100 OK
S->C: Number-of-messages:2
S->C:
S->C: some-other-file
S->C: 3749f52bb326ae96782b42dc0a97b4c1
C->S: READ
C->S: Mailbox:xxx
C->S: Message:3749f52bb326ae96782b42dc0a97b4c1
C->S:
S->C: 100 OK
S->C: Content-length:11
S->C:
S->C: 0123456789
C closes connection
```