

Jednoduchý paralelní stahovač webu

Úkolem je vytvořit utilitku podobnou programu `wget`, která stáhne kompletní web ze zadané adresy (startovní URL) a uloží všechny soubory na disk do cílového adresáře. Stahování souborů musí být prováděno paralelně v několika vláknech. Program také musí nahrazovat URL ve stažených souborech tak, aby místo startovní URL obsahovaly cílový adresář – například `<a href="http://www.seznam.cz/blabla.html">` se změni na `<a href="/home/username/web/blabla.html">`.

Vstup:

- Startovní URL
- Cílový adresář pro uložení souborů
- Počet vláken ($N > 0$)

Výstup

- Kopie webu uložená na disku (v adresáři)
 - Všechny soubory dosažitelné ze startovní URL přes tagy `<a>` a `<img>`
 - Musí být zachována adresářová struktura (hierarchie webových stránek)

Další požadavky na řešení

- Nesmí se stahovat žádný soubor, jehož URL nemá jako prefix startovní URL
- Pracovat musí všech N vláken
 - Když existují nezpracované URL, tak žádné vlákno nesmí spát nebo čekat
- Běh programu se ukončí ve chvíli, kdy všechna vlákna spí (čekají na práci) a neexistuje žádné nezpracované URL
- Žádné URL se nesmí zpracovávat vícekrát

Uživatelské rozhraní bude mít podobu příkazové řádky, která bude akceptovat tyto tři argumenty: startovní URL, cílový adresář, počet vláken.

Pro základní testování můžete použít moji homepage (<http://dsrg.mff.cuni.cz/~parizek/>).
Dejte pozor na stahování příliš velkých webů (překročení kvóty, zablokování účtu).