

Čtečka novinek ve formátu Atom s podporou TLS

Síťové aplikace a správa sítí

Obsah

Obsah	2
Uvedení do problematiky	3
Návrh aplikace	3
Třída Program	3
Třída Feed	3
Třída Parser	3
Třída Entry	4
Implementace	4
Chybějící informace	4
Duplicitní vstupní parametry	4
Formát adresy zdroje	4
Více autorů	4
Informace o programu	4
Použití	5
Překlad	5
Konvence spuštění	5
Parametry spuštění	5
Parametr URL	5
Přepínač -f	5
Přepínač -c	5
Přepínač -C	5
Přepínač -T	5
Přepínač -a	5
Přepínač -u	5
Příklady spuštění	6
Zdroje	6

Uvedení do problematiky

RSS a Atom jsou formáty XML souborů využívané ve velkém pro odběr novinek daného serveru.

Formát RSS se v uplynulých letech hodně měnil a každá větší verze sebou nesla úplně nové názvy uzlů. Tento program se tudíž musí přizpůsobit hodně variantám vstupu. Dalším úskalím na které jsem během řešení narazil je dodržování daných standardů, proto jsem program psal tak, aby se vypořádal se všemi možnými uzly, nehledě na použitou verzi RSS. Mnohdy se může stát, že některé uzly, které očekáváme se v souboru vůbec nenachází, i s tímto se program musí umět vypořádat. Taktéž se může stát, že odpověď serveru na dotazovaný soubor nebude úspěšná, v takovém případě se zdroj ignoruje.

Program ovšem umí načítat při jednom spuštění více než jeden zdroj, tudíž v takové situaci nesmí ukončit svou činnost, ale pouze uživatele o této události informovat a přejít k dalším zdrojům.

Návrh aplikace

Vzhledem k možnosti vyvíjet projekt v jazyce C++ jsem si pro jednotlivé části programu vytvořil třídy, které je zaštitují.

Třída Program

Tato třída tvoří jen jednu instanci, která je vytvořena ihned po spuštění programu. Jejím hlavním účelem je uchovávat parametry spuštění a dát celý program do pohybu. K tomu není třeba žádný vnější zásah, kromě vytvoření dané instance.

Při vytvoření instance jsou nejprve načteny, zpracovány a uloženy parametry spuštění programu. Pokud je v parametrech objevena chyba, program je ukončen s návratovým kódem 1.

V případě úspěšného zpracování parametrů je pro každou vstupní URL vytvořena instance třídy *Feed*.

Třída Feed

Pro každou vstupní URL je tvořena právě jedna instance této třídy. Při svém vytvoření zpracuje danou URL a určí z nich informace nutné k navázání spojení se serverem.

Po zavolání funkce *read()* se program spojí se serverem a vyžádá si od něj data. Pokud jsou v pořádku přijata, vytvoří se instance třídy *Parser*, která je zpracuje.

V případě úspěšného zpracování je obsah této instance převeden na řetězec a vypsán na standardní výstup.

Třída Parser

Tato třída obstarává zpracování samotného XML souboru získaného ze serveru. Při vytvoření instance této třídy je určen formát obsahu souboru a poté je započato jeho zpracování. V souboru se vyhledává titulek zdroje a jednotlivé články. Pro každý nalezený článek je vytvořena instance třídy *Entry* a uložena do seznamu článků daného zdroje.

Třída Entry

Třída zpracovávající a uchovávající informace o jednom článku. Při jejím vytvoření jsou vyhledány údaje o titulku článku, datum vytvoření, URL článku a jeho autorech.

Při zavolání funkce *toString()* je vrácen řetězec obsahující informace na základě vstupních parametrů programu.

Implementace

Chybějící informace

Z důvodů zachování konzistence výstupů jsem se rozhodl vypisovat i chybějící informace, jejich skutečný obsah je nahrazen zkratkou *N/A (Not available)*.

Ukázkový výstup tedy může být například řádek "*Autor: N/A*" nebo název zdroje "**** N/A ****".

Duplicitní vstupní parametry

Program počítá s možností zadání duplicitních parametrů. Jediná chybná kombinace vstupních parametrů je využití přepínače *-f* spolu se specifikováním URL.

Chování programu při duplicitě přepínačů *-T*, *-a* nebo *-u* je naprosto stejné, jako kdyby se v parametrech daný přepínač vyskytl pouze jednou.

Při duplicitě parametru URL je brán v potaz poslední výskyt tohoto parametru.

Kdežto u přepínačů *-c*, *-C* nebo *-f* jsou zpracovány všechny jejich výskyty.

Formát adresy zdroje

Program vyžaduje aby každá vstupní URL adresa obsahovala prefix *http://* či *https://* z důvodů jednoznačného určení, zda využít či nevyužít šifrovaný přenos dat. V opačném případě je uživatel o nenalezení tohoto prefixu informován a daný zdroj je přeskočen.

Také se předpokládá, že v souboru feedfile se bude na jednom řádku vyskytovat pouze jedna jediná URL adresa.

Více autorů

Jelikož Atom povoluje výskyt více uzlů obsahující informace o autorech pro jeden článek, Je pro každého z nich na nový řádek vypsáno "*Autor: <jméno>*". Pokud není uvedeno jméno daného autora, na výstup není vypsáno nic. V případě, že je použit parametr *-a*, ale článek jméno autora neuvádí, je na výstup pro daný článek jedenkrát vypsán řádek "*Autor: N/A*".

Informace o programu

Program umožňující čtení novinek ve formátu RSS a Atom. Umožňuje číst nadpisy, URL adresy, autory a datum daných článků. Umístění novinek může být jedna URL specifikována parametrem spuštění nebo i více URL v případě použití souboru a parametru *-f*. Program také umožňuje využít šifrované připojení a specifikovat vlastní certifikáty.

Použití

Překlad

Pro překlad programu stačí využít příkaz *make* v kořenovém adresáři, který zde sestaví spustitelný soubor jménem *feedreader* obsahující výsledný program.

Konvence spuštění

Samotné spuštění a všechny jeho parametry se řídí následující konvencí:
feedreader <URL | -f <feedfile>> [-c <certfile>] [-C <certaddr>] [-T] [-a] [-u]

Parametry spuštění

Parametr URL

Tento parametr udává cílovou adresu zdroje, musí být uvozen prefixem *http://* či *https://*

Přepínač -f

Tento přepínač má v argumentu cestu k souboru obsahujícím zdroje k přečtení.

Přepínač -c

Tento přepínač argumentem specifikuje cestu ke konkrétnímu souboru s certifikáty, které budou porovnány v případě šifrovaného spojení.

Přepínač -C

Tento přepínač v argumentu udává cestu ke složce obsahující certifikáty využívané pro overení v případě využití šifrovaného spojení.

Přepínač -T

Díky tomuto přepínači bude u každého záznamu uvedena informace o datu jeho vytvoření/upravení.

Přepínač -a

Tento přepínač bude mít za následek uvedení informace o autorovi či autech daného záznamu.

Přepínač -u

Tento přepínač uvede u každého záznamu informaci o URL záznamu.

Příklady spuštění

`feedreader https://tools.ietf.org/dailydose/dailydose_atom.xml`

`feedreader https://tools.ietf.org/dailydose/dailydose_atom.xml -u`

`feedreader https://tools.ietf.org/dailydose/dailydose_atom.xml -u -T`

`feedreader https://tools.ietf.org/dailydose/dailydose_atom.xml -C /dev/null`

Zdroje

1. NOTTINGHAM, M. *RFC 4287 - The Atom Syndication Format* [online]. 2005-12 [cit. 2018-11-19]. Dostupné z <https://tools.ietf.org/html/rfc1035>
2. RSS-DEV WORKING GROUP *RDF Site Summary (RSS) 1.0* [online]. 2008-16 [cit. 2018-11-19]. Dostupné z <http://web.resource.org/rss/1.0/spec>
3. RSS ADVISORY BOARD *RSS 2.0 Specification* [online]. 2009-03 [cit. 2018-11-19]. Dostupné z <http://www.rssboard.org/rss-specification>
4. OPENSSL SOFTWARE FOUNDATION *Documentation* [online]. 20018 [cit. 2018-11-19]. Dostupné z <https://www.openssl.org/docs/>