



Dokumentácia k projektu ISA

## **Analýza BitTorrent služby KickassTorrents**

Autor: Daniel Klimaj, [xklima22@stud.fit.vutbr.cz](mailto:xklima22@stud.fit.vutbr.cz)  
Fakulta informačních technologií  
Vysoké učení technické v Brně

21.11.2015

# Obsah

<u>Analýza BitTorrent služby KickassTorrents.....</u>	<u>1</u>
<u>1 Úvod.....</u>	<u>3</u>
<u>2 Základné pojmy.....</u>	<u>3</u>
<u>2.1 Bencode.....</u>	<u>3</u>
<u>2.2 BitTorrent.....</u>	<u>3</u>
<u>2.3 Štruktúra súborov .torrent.....</u>	<u>4</u>
<u>3 Návrh riešenia.....</u>	<u>4</u>
<u>4 Implementácia.....</u>	<u>4</u>
<u>4.1 Moduly.....</u>	<u>4</u>
<u>4.2 Použité knižnice alebo moduly.....</u>	<u>5</u>
<u>4.2 Použitie aplikácie.....</u>	<u>6</u>
<u>5 Záver.....</u>	<u>6</u>
<u>Referencie.....</u>	<u>7</u>
<u>Metriky kódu.....</u>	<u>7</u>

# 1 Úvod

Tento dokument slúži ako dokumentácia k projektu do predmetu Sítové aplikácie a správa sítí. Na úvod v skratke vysvetľuje protokol BitTorrent a štruktúru súborov .torrent. Ďalej potom popisuje návrh a implementáciu aplikácie pre získavanie IP adries zariadení zdieľajúcich torrent na serveri KickassTorrents (<https://kat.cr/>).

## 2 Základné pojmy

Pre implementáciu aplikácie pracujúcej s protokolom BitTorrent a súbormi torrent je potrebné mať nejaké základné znalosti o týchto technológiách. Táto časť sa zaoberá protokolom BitTorrent, štruktúrou súborov typu torrent a bencode.

### 2.1 Bencode

**Bencode** je kódovanie používané v protokole BitTorrent na prenos voľne štruktúrovaných dát. Podporuje reťazce, celé čísla, zoznamy a asociatívne polia (slovníky). Označenie benkódované dáta znamená dáta zakódované pomocou bencode.

### 2.2 BitTorrent

**BitTorrent** je peer-to-peer (P2P) protokol určený ku zdieľaniu súborov medzi viacerými peerami v nespoľahlivých sieťach.

**Peer** je ľubovoľný BitTorrent klient účastníci sa zdieľania súboru.

**Tracker** je HTTP/HTTPS služba, ktorá odpovedá na požiadavku HTTP GET. Požiadavka obsahuje niekoľko parametrov. Pre účely projektu postačia nasledovné:

- *info\_hash* URL-enkódovaný 20 bajtový SHA1 hash info sekcie torrentu
- *peer\_id* 20 bajtový URL-enkódovaný identifikátor klienta, môže to byť ľubovoľná hodnota
- *port* číslo portu, na ktorom klient odpočúva komunikáciu
- *uploaded* množstvo uploadnutých dát, jednotka nie je pevne daná, ale bežne sa používa byte
- *downloaded* množstvo stiahnutých dátového
- *left* množstvo dát, ktoré je ešte potrebné stiahnuť
- *compact* nastavním compact=1 sa peerlist nahradí 6 bajtovým reťazcom na 1 peer, prvé 4 byty určujú IP adresu peeru a zvyšné 2 byty portu peeru

Odpoveď trackeru prichádza vo formáte text/plain a okrem hlavičky obsahuje benkódovaný slovník, z ktorého je pre účely projektu podstatný jediný kľúč –peers. Kľúč peers obsahuje 6-bajtové reťazce obsahujúce IP adresy a porty peerov (viď. compact v popise parametrov požiadavky).

## 2.3 Štruktúra súborov .torrent

Súbor **.torrent** je benkódový slovník obsahujúci rôzne kľúče, najpodstatnejšie sú:

- *info* formát závisí od toho, či sa zdieľa jeden alebo viac súborov, obsahuje ďalší slovník
- *announce* URL trackeru
- *announce-list* zoznam URL trackerov

## 3 Návrh riešenia

Aplikácia stiahne RSS feed zo zadanej adresy. Po overení, že RSS feed pochádza zo serveru <http://kat.cr/> spracuje RSS a získa z neho elementy *item*. Tieto získané dáta zapíše v čitateľnej forme do výstupného súboru. Z prvého elementu *item* získa URL pre stiahnutie torrentu. Tento torrent stiahne a zapíše do súboru s rovnakým menom ako je uvedené v jeho elemente *item*. Z bedekódovaného torrentu sa získa peerlist, z ktorého sa vyberú HTTP trackeri, a vypočíta sa *info\_hash*. Na každý zo získaných HTTP trackerov je odoslaný HTTP GET request a je získaný peerlist, prípadne zhlásená chyba pri komunikácii. Pri kontaktovaní trackeru môže dôjsť k viacerým chybám – timeout, chyba pri pripájaní sa k socketu, chyba pri odosielaní požiadavky, nesprávny formát prijatých dát, vrátený iný stavový kód ako 200 OK, apod.. Všetky problémy sú reportované na `stderr`.

## 4 Implementácia

Aplikácia je napísaná v jazyku Python 2.7.6. Pre prácu s RSS, torrentmi a trackermi sú napísané triedy `RssDocument`, `RssNode`, `Torrent` a `Tracker`. Na sťahovanie RSS a torrentov sa používa neštandardná knižnica `requests` a pre prácu s bencode zas knižnica `bencode`. Aplikácia je vytvorená pre linuxové OS, predovšetkým Ubuntu, na ktorom bola testovaná. Na Unixových OS alebo vo Windows nebola testovaná.

Hlavná logika programu je spracovaná v súbore `antipirat.py`, ktorý spracúva parametre príkazového riadku a podľa nich potom volá požadovanú funkciu alebo vypíše chybu a program sa ukončí v prípade neplatného parametru.

### 4.1 Moduly

Modul `rss.py` definuje triedu **Rss**. Trieda `Rss` je nadtriedou tried `RssDocument` a `RssNode`. Obsahuje jedinou metódu na vyhľadávanie XML podelementov elementu podľa tagu.

Modul `rss_document.py` definuje triedu **RssDocument** pre prácu s RSS feedmi. Využíva `xml.etree` pre parsovanie XML a umožňuje zistiť, či RSS dokument pochádza zo serveru <http://kat.cr/>, získanie prvého alebo všetkých elementov `<item>` obsahujúcich informácie o zdieľanom filme.

Modul `rss_node.py` definuje triedu **RssNode**. Táto trieda sprístupňuje prácu s dátami obsiahnutých v elementoch `<item>` RSS dokumentu, kontrolu validity získaných dát a spracovanie do čitateľnej formy.

Modul `log.py` definuje triedu **Log** s troma statickými metódami na výpis chýb (`Log.error()`), dôležitých informácií (`Log.info()`) alebo ladiacich informácií (`Log.debug()`). Výpis ladiacich informácií treba povoliť nastavením `Log.LOGLEVEL` na `LogLevel.Debug`.

Modul `torrent.py` definuje triedy **Torrent**, ktorá najprv overuje existenciu torrent súboru a následne ho bendeškóduje a získa z neho info hash a announce list (zoznam trackerov). Na výpočet info hashu využíva `urllib.quote_plus()` a `hashlib.sha1().digest()`.

Modul `tracker.py` definuje triedu **Tracker**. Trieda Tracker umožňuje pripojenie k trackeru a získanie peerlistu. V prípade úspešného pripojenia k trackeru (`connect()`) je možné získať peerlist. Pri tom však môže dôjsť k celej rade chýb, preto je získavanie peerlistu implementované tak, aby sa v prípade chyby opakovalo. Počet opakovaní je nastaviteľný pomocou konštanty `TRIES`. Podobným spôsobom je možné nastaviť konštanty `SLEEP_TIME` (doba medzi opakovaním pokusu), `RECV_SIZE` (veľkosť pre `socket.recv()`), `RECV_TIMEOUT` (doba čakania pred zavolaním `socket.recv()`, ak sú prijaté dáta väčšie ako `RECV_SIZE`) a `CONNECT_TIMEOUT` (doba, po ktorú sa pokúša o pripojenie k socketu). Po úspešnom získaní odpovede od serveru zisťuje status kód odpovede, z ktorej sú všetky kódy mimo 200 OK považované za chybu. V prípade 200 OK a úspešného bendeškódovania obsahu odpovede prevádza peerlist do čitateľnej podoby. Ak sa ani po určenom počte opakovaní nepodariť získať peerlist, tak je vrátený prázdny zoznam.

## 4.2 Použité knižnice alebo moduly

V aplikácii sú využité nasledovné štandardné knižnice/moduly:

- `xml.etree`
- `hashlib`
- `re`
- `os`
- `sys`
- `struct`
- `socket`
- `time`
- `urllib`

Použité knižnice/moduly tretích strán:

- bencode - <https://pypi.python.org/pypi/BitTorrent-bencode>
- requests - <http://docs.python-requests.org/en/latest/>

## 4.2 Použitie aplikácie

*./antipirat (-r url|-i filename|-t filename) [-a url]*

<i>-r [--rss] URL</i>	-- URL adresa RSS feedu
<i>-i [--input-announcement] filename</i>	-- uložený RSS feed v súbore filename
<i>-a [--tracker-announce-url] URL</i>	-- podvržené URL trackeru pre získanie peerlistu
<i>-t [--torrent-file] filename</i>	-- stiahnutý torrent, nepoužíva sa RSS
<i>-h [--help]</i>	-- výpis nápovedy

## 5 Záver

Program monitoruje zdieľané filmy na serveri KickassTorrents. Umožňuje sťahovanie RSS feedov a torrentov, ich analýzu a prípadné získanie peerlistu po kontaktovaní trackerov získaných z torrent súborov.

## Referencie

[1] *Bencode* [online], posledná aktualizácia 2015-08-26 16:57 [cit. 2015-11-21], Wikipedia.org. Dostupné z WWW: <<https://en.wikipedia.org/wiki/Bencode>>

[2] *BitTorrent Specification* [online], posledná aktualizácia 2015-11-14 03:07 [cit. 2015-11-21], Theory.org. Dostupné z WWW: <<https://wiki.theory.org/BitTorrentSpecification>>

[3] *Torrent File* [online], posledná aktualizácia 2015-11-06 21:04 [cit. 2015-11-21], Wikipedia.org. Dostupné z WWW: <[https://en.wikipedia.org/wiki/Torrent\\_file](https://en.wikipedia.org/wiki/Torrent_file)>

## Metriky kódu

**Počet súborov:** 7

**Počet riadkov kódu:** 921 (vrátane komentárov a prázdnych riadkov)

**Veľkosť spustiteľného programu:** 11,4 kB (symlink antipirat.py)