

Čtečka novinek ve formátu Atom s podporou TLS

DOKUMENTACE

Autor: Martin Omacht

Login: xomach00

Obsah

[2 Úvod 2](#_Toc530348172)

[3 Čtečka novinek 2](#_Toc530348173)

[3.1 RSS 2](#_Toc530348174)

[3.1.1 RSS 0.91 2](#_Toc530348175)

[3.1.2 RSS 2.0 2](#_Toc530348176)

[3.1.3 RSS 1.0 3](#_Toc530348177)

[3.2 Atom 3](#_Toc530348178)

[4 TLS 4](#_Toc530348179)

[5 Program feedreader 4](#_Toc530348180)

[5.1 Základní informace 4](#_Toc530348181)

[5.1.1 Podporované formáty zdrojů 4](#_Toc530348182)

[5.1.2 Podporované komunikační protokoly 4](#_Toc530348183)

[6 Reference 6](#_Toc530348184)

# Úvod

Tento dokument slouží jako dokumentace k programu feedreader. Tento program funguje jako čtečka novinek ve formátu Atom nebo RSS s podporou TLS. Dokumentace zároveň poskytuje základní informace o formátu Atom a RSS a také o protokolu TLS/SSL.

# Čtečka novinek

Než se pustíme do implementačních detailů, je potřeba si říct něco o tom, co je to čtečka novinek.

Čtečka novinek je program, který stahuje aktuality z webových zdrojů (angl. feed). Tyto zdroje se většinou vyskytují na stránkách s často měnícím se obsahem (např. zpravodajské servery)(1). Nejčastější formáty zdroje jsou Atom a RSS, které využívají XML formát.

## RSS

Zkratka RSS má několik výkladů(1):

* Rich Site Summary (RSS 0.91 od firmy Netscape)
* Resource Description Framework Site Summary (RSS 1.0 – tvůrcem je W3C[[1]](#footnote-1))
* Really Simple Syndication (RSS 2.0 – spravováno Berkman Klein Center for Internet & Society[[2]](#footnote-2))

Formát verzí, které tato čtečka podporuje, a důležité elementy formátu, si rozebereme v následujících sekcích.

### RSS 0.91

Kořenovým elementem musí být <rss> s povinným atributem version, který specifikuje verzi používaného RSS (v tomto případě 0.91). V něm se nachází jediný <channel> element, který obsahuje informace o kanálu a jeho obsahu(2). Tento element pak může obsahovat řadu povinných a nepovinných elementů. Pro účely této čtečky je důležitý povinný element <title>, který udává název kanálu.

Jednotlivé položky zdroje jsou uvedeny v elementech <item>, které se vyskytují taktéž v elementu <channel>. Těchto položek může zdroj obsahovat jakékoliv množství. V elementu <item> jsou pak povinné elementy <title> a <link>. Tyto elementy obsahují titulek a odkaz na položku. Informace o autorovi nebo poslední aktualizace položky tento standard nepodporuje.

Čtečka podporuje i verzi 0.92, která pouze přidává několik volitelných elementů pro element <item>.

### RSS 2.0

Tato verze je založena na verzi 0.91. Jelikož je verze 2.0 zpětně kompatibilní s verzí 0.91, základní struktura elementů je stejná. Pro tuto čtečku však verze 2.0 přidává důležité elementy pro element <item>:

* <author> - email autora položky
* <guid> - řetězec, který unikátně identifikuje položku. Pokud neobsahuje atribut isPermaLink=“false“, čtečka obsah tohoto elementu použije jako URL odkaz na položku.
* <pubDate> - datum a čas publikace položky

Standard také změnil všechny dílčí elementy prvku <item> na volitelné, za podmínky, že bude přítomný alespoň <title> nebo <description>. (3)

### RSS 1.0

Verze 1.0 je založená na RSS 0.9 a zachovává zpětnou kompatibilitu, ale není kompatibilní s verzemi 0.91 a 2.0, které nepoužívají RDF (Resource Description Framework).

Kořenovým prvkem je element <rdf:RDF>, kde prefix jmenného prostoru rdf je asociován schématu syntaxe RDF (může být použit i jiný prefix, ale pro zpětnou kompatibilitu je doporučené používat rdf). Stejně jako ve výše popsaných verzích RSS tento element obsahuje prvek <channel> popisující samotný kanál. Je zde například povinný element <title>, který obsahuje titulek kanálu.

Na rozdíl od RSS verzí 0.91 a 2.0, elementy <item> nejsou umístěné v elementu <channel>, ale jsou přímo v kořenovém prvku. Elementy <title> a <link> jsou stejné jako v ostatních případech. Elementy pro autora a poslední aktualizaci položky standard nepodporuje, ale jeho oficiální modul Dublin Core[[3]](#footnote-3) povoluje elementy <dc:creator> a <dc:date> pro tento účel (prefix jmenného prostoru dc je asociován modulu Dublin Core). (4)

## Atom

Atom je další standard pro publikování syndikovaného webového obsahu, který se snaží vyhnout limitacím a chybám standardu RSS (5).

Kořenový element je <feed>. Ten obsahuje mimo jiné název zdroje v povinném elementu <title>. Dále může obsahovat jakýkoliv počet elementů <entry>, které reprezentují jednotlivé položky zdroje.

Titulek položky zdroje se opět nachází v povinném elementu <title>. Poslední aktualizace reprezentuje element <updated>. Co je ale složitější, je autor položky. Ten se vyskytuje v elementu <author>, ve kterém se pak nachází povinně element <name> se jménem autora. Elementů <author> může být uvedeno více nebo žádný a v takovém případě se použije <author> z elementu <source>. Pokud není k dispozici autor ani v <source>, jako autor se považuje autor uvedený v elementu <feed>. Odkaz na webovou stránku položky je obsahem atributu href v elementu <link>, který buď nemá atribut rel nebo atribut rel má hodnotu alternate. (6)

# TLS

TLS (Transport Layer Security) protokol poskytuje zabezpečené připojení mezi dvěma komunikujícími koncovými uzly. TLS vychází z dřívějších protokolů SSL, ve kterých však byly odhaleny bezpečnostní díry, a tak je dnes tento protokol již zastaralý.

Zabezpečené připojení přes TLS poskytuje následující vlastnosti:

* Autentizace – serverová část připojení je vždy autentizovaná; klientská část je volitelně autentizovaná.
* Důvěrnost – odeslaná data jsou vždy viditelná pouze koncovým uzlům.
* Integrita – odeslaná data nemohou být bez detekce modifikována útočníky.

Tyto vlastnosti jsou dodrženy i pokud útočník má k dispozici kompletní kontrolu nad sítí.

TLS se skládá z dvou primárních komponent:

* Handshake protokol – autentizuje komunikující strany, dohodne kryptografické módy a parametry a zavede klíčovací materiál.
* Record protokol – použije dohodnuté parametry pro zabezpečení komunikace mezi koncovými uzly.

Během fáze handshake server odešle certifikát klientovi, který ho ověří proti souboru důvěryhodných certifikátů (autentizace serveru). Klient odesílá certifikát pouze pokud si ho server vyžádá (autentizace klienta). (7)

# Program feedreader

Program feedreader je čtečka novinek ve formátu Atom a RSS s podporou TLS. Uživatelským rozhraním je příkazový řádek. Program stáhne na základě parametrů uvedené zdroje, které mohou být ve formátu Atom nebo RSS, a vypíše uživatelem požadované informace na standardní výstup.

## Základní informace

Výstup se skládá z názvu zdroje (kanálu) a jednotlivých položek (novinek). Parametry lze přidat vypisování autora položky, odkazu na položku nebo čas poslední úpravy položky (viz 5.2). Pomocí souboru předaným parametrem -f lze specifikovat více zdrojů.

### Podporované formáty zdrojů

Program je testovaných na následujících verzích formátu zdrojů:

* RSS 0.91
* RSS 2.0
* RSS 1.0
* Atom

Avšak díky kompatibilitě některých formátu a benevolenci programu, by neměl být problémy s těmito verzemi:

* RSS 0.9
* RSS 0.92
* RSS 2.0.1-2.0.11

### Podporované komunikační protokoly

Program podporuje protokol HTTP i HTTPS pro stahování zdrojů.

## Použití

# Reference

1. **Přispěvatelé Wikipedie.** RSS. *Wikipedie.* [Online] Wikipedie: Otevřená encyklopedie, 4. 10. 2017. [Citace: 17. 11. 2018.] https://cs.wikipedia.org/w/index.php?title=RSS&oldid=15394320.

2. **UserLand Software.** RSS 0.91 Specification. *RSS Advisory Board.* [Online] 9. 6. 2000. [Citace: 18. 11. 2018.] http://www.rssboard.org/rss-0-9-1 (anglicky).

3. —. RSS 2.0 Specification. *RSS Advisory Board.* [Online] 14. 7. 2003. [Citace: 18. 11. 2018.] http://www.rssboard.org/rss-2-0 (anglicky).

4. **RSS-DEV Working Group.** RDF Site Summary (RSS) 1.0. *web.resource.org.* [Online] 9. 6. 2008. [Citace: 18. 11. 2018.] http://web.resource.org/rss/1.0/spec (anglicky).

5. **Easuwaran, Sathish.** RSS vs Atom. *Saksoft.* [Online] Saksoft Limited, 7. 11. 2015. [Citace: 18. 11. 2018.] https://www.saksoft.com/rss-vs-atom/ (anglicky).

6. **Nottingham, M. a Sayre, R.** RFC 4287 - The Atom Syndication Format. *IETF Tools.* [Online] 17. 12. 2005. [Citace: 18. 11. 2018.] https://tools.ietf.org/html/rfc4287 (anglicky).

7. **Rescorla, E.** RFC 8446 - The Transport Layer Security (TLS) Protocol Version 1.3. *IEFT Tools.* [Online] 10. 8. 2018. [Citace: 18. 11. 2018.] https://tools.ietf.org/html/rfc8446 (anglicky).

1. <https://www.w3.org/> [↑](#footnote-ref-1)
2. <https://en.wikipedia.org/wiki/Berkman_Klein_Center_for_Internet_%26_Society> [↑](#footnote-ref-2)
3. <http://web.resource.org/rss/1.0/modules/dc/> [↑](#footnote-ref-3)