

České vysoké učení technické v Praze
Fakulta jaderná a fyzikálně inženýrská

Katedra softwarového inženýrství
Obor: Aplikace softwarového inženýrství



Porovnání účinnosti komprese dat ve formátech XML a JSON

Comparison of the effectiveness of data compression in XML and JSON format

DIPLOMOVÁ PRÁCE

Vypracoval: Bc. Tomáš Smola
Vedoucí práce: Ing. Tomáš Liška, Ph.D.
Rok: 2015

Před svázáním místo téhle stránky

 s podpisem děkana (bude to jediný oboustranný list ve Vaší práci) !!!!

Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v příloženém seznamu.

V Praze dne

.....
Bc. Tomáš Smola

Poděkování

Děkuji Ing. Tomáši Liškovi, Ph.D. za vedení mé diplomové práce a za podnětné návrhy, které ho obohatily.

Bc. Tomáš Smola

Název práce:

Porovnání účinnosti komprese dat ve formátech XML a JSON

Autor: Bc. Tomáš Smola

Obor: Aplikace softwarového inženýrství

Druh práce: Diplomová práce

Vedoucí práce: Ing. Tomáš Liška, Ph.D.

Katedra softwarového inženýrství, Fakulta jaderná a fyzikálně
inženýrská, České vysoké učení technické v Praze

Konzultant: —

Abstrakt: Abstrakt

Klíčová slova: Klíčová slova

Title:

Comparison of the effectiveness of data compression in XML and JSON format

Author: Bc. Tomáš Smola

Abstract: Abstract

Key words: Key words

Obsah

Úvod	1
1 Obecné seznámení s formáty XML a JSON	2
1.1 XML	2
1.1.1 Charakteristika	2
1.1.2 Syntaktická analýza	2
1.1.3 Parsování	2
1.1.4 Výhody a nevýhody	2
1.2 JSON	2
1.2.1 Charakteristika	2
1.2.2 Syntaktická analýza	2
1.2.3 Parsování	4
1.2.4 Výhody a nevýhody	4
2 Komprese dat	5
3 Popis existujících kompresních algoritmů	6
4 Přehled existujících implementací kompresních algoritmů pro efektivní uchovávání dat ve formátu XML a JSON	7
5 Vlastní implementace vybraných kompresních algoritmů	8
6 Porovnání účinnosti komprese dat ve formátu XML a JSON	9
Závěr	10
Seznam použitých zdrojů	11
Přílohy	12

Úvod

Kapitola 1

Obecné seznámení s formáty XML a JSON

1.1 XML

1.1.1 Charakteristika

1.1.2 Syntaktická analýza

1.1.3 Parsování

1.1.4 Výhody a nevýhody

1.2 JSON

JSON neboli JavaScript Object Notation je odlehčený způsob zápisu (formátování) dat. Tento textový formát je nezávislý na počítačové platformě a je čitelný pro člověka. JSON je založen na dvou univerzálních datových strukturách: kolekce dvojic klíč/hodnota a seřazený seznam hodnot, které podporují v nějaké formě asi všechny známé moderní programovací jazyky. Díky těmto vlastnostem se JSON stal velmi oblíbeným formátem pro vzájemnou výměnu dat.

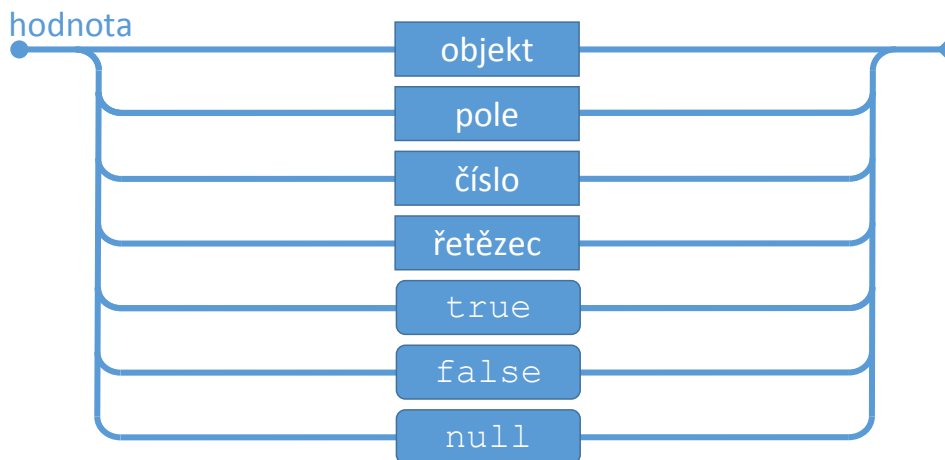
1.2.1 Charakteristika

1.2.2 Syntaktická analýza

Jak již bylo zmíněno, je JSON textový formát a je tedy posloupností tokenů tvořených z Unicode znaků. Sada tokenů obsahuje šest strukturálních tokenů: [(levá hranatá závorka), { (levá složená závorka),] (pravá hranatá závorka), } (pravá složená závorka), : (dvojtečka) a , (čárka); dále obsahuje znakové řetězce, čísla a tři doslovné tokeny: `true`, `false` a `null`.

Hodnoty

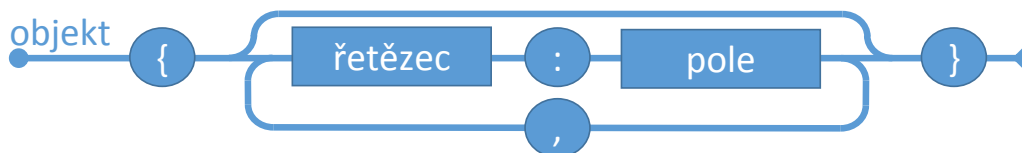
Za hodnotu je v JSONu považován objekt, pole, číslo, řetězec, `true`, `false`, nebo `null`.



Obrázek 1.1: Struktura hodnoty

Objekty

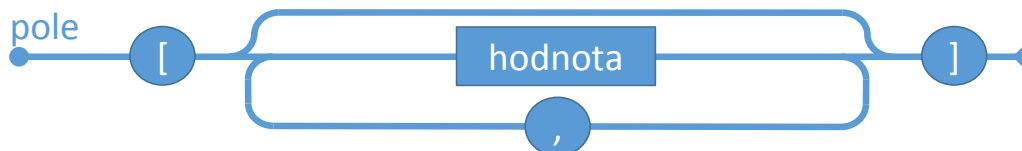
Objekt je reprezentován dvojicí složených závorek, uvnitř kterých je žádná nebo více dvojic klíč/hodnota přičemž klíč je řetězec. Klíč a hodnota jsou odděleny dvojtečkou a jednotlivé dvojice odděluje čárka.



Obrázek 1.2: Struktura objektu

Pole

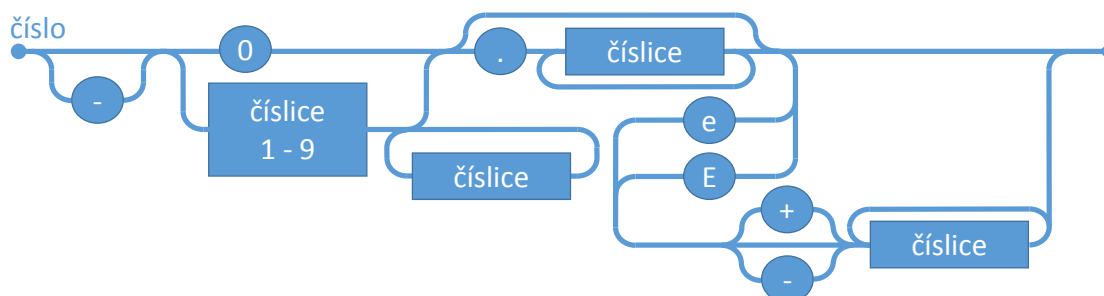
Pole je složeno z dvojice hranatých závorek, mezi kterými může být nula nebo více seřazených hodnot, které jsou odděleny čárkou.



Obrázek 1.3: Struktura pole

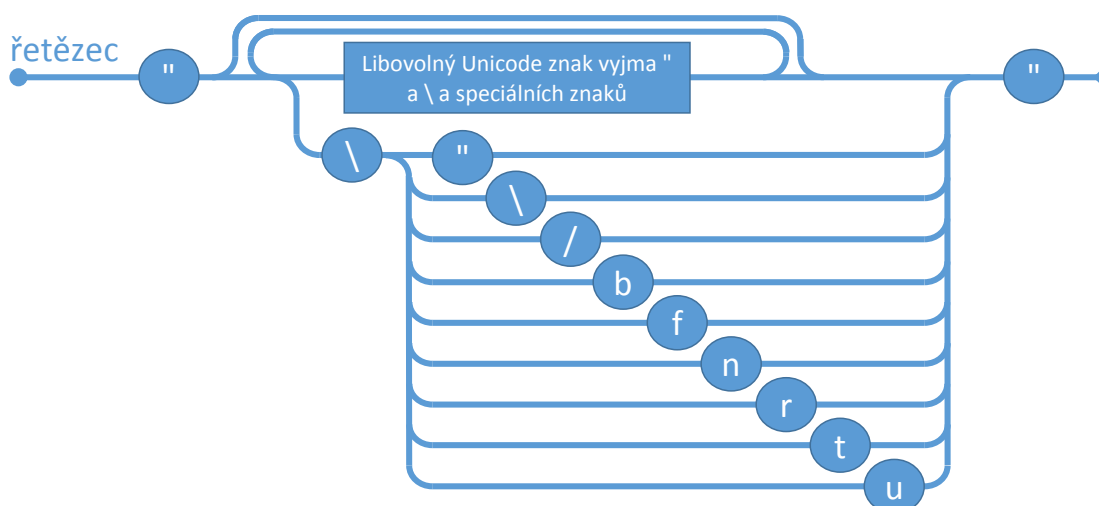
Číslo

Číslo jsou v desítkové soustavě (tedy číslice 0 – 9) záporná čísla jsou uvozena znaménkem - (mínus), desetinná část je oddělena znaménkem . (tečka). Je možný i takzvaný vědecký zápis čísel s použitím symbolů e (malé e) nebo E (velké e) a volitelně lze použít u exponentu znaménka + (plus) nebo - (mínus).



Obrázek 1.4: Struktura čísla

Řetězec



Obrázek 1.5: Struktura řetězce

1.2.3 Parsování

1.2.4 Výhody a nevýhody

Kapitola 2

Komprese dat

Kapitola 3

Popis existujících kompresních algoritmů

Kapitola 4

Přehled existujících implementací kompresních algoritmů pro efektivní uchovávání dat ve formátu XML a JSON

Kapitola 5

Vlastní implementace vybraných kompresních algoritmů

Kapitola 6

Porovnání účinnosti komprese dat ve formátu XML a JSON

Závěr

Seznam použitých zdrojů

Přílohy