

# XML (XML, definice souboru XML, export do XML, import z XML, parsování XML, JSON)

## XML

Extensible Markup Language je univerzální značkovací jazyk pro uspořádání dat – nejspíše nejvíce rozšířený společně s JSON, oproti kterému je však jeho struktura větší. Umožňuje definovat, jak má vypadat (definičními DTD soubory nebo soubory XML Schema), co musí nebo může obsahovat apod. Musí mít tzv. *kořenový element* (např. <clovek></clovek>), v tom jsou pak další vnořené elementy, jejichž obsahem jsou data samotná, např. <krestnijmeno>Tomáš</krestnijmeno>.

## Definice souboru XML

Abychom omezili množinu možných výsledných podob dokumentů XML, používají se takzvané definiční soubory, např. DTD (.dtd) nebo XML Schema (.xsd).

### DTD

DTD je již starý způsob a nemá tolik možností jako XML Schema. Nezapisuje se v XML, zato jeho použití pro malé XML dokumenty je velice jednoduché. V dnešní době se používá prakticky pouze pro definování jazyka HTML.

### XSD

XML Schema je oproti DTD rozšiřitelný v budoucnu (např. možnost vytvoření nových datových typů) a například také podporuje datové typy. Protože je XSD založené na XML, je možné XSD schema procházet XML parserem či jej validovat pomocí jiného XSD schématu.

## Export do XML

Export do XML je proces, kdy je vygenerován XML soubor s nějakými daty (např. z databáze nebo z dat, které zadal uživatel např. při vyplňování webového formuláře). Tento pak může být použit pro zpracování takových dat jinou aplikací či třeba uchováván jako záloha.

## Import z XML

Import z XML je pak proces, kdy je soubor XML někam nahrán/načten a následně (obvykle) parsován, čímž jsou z něj data „vytažena“ a mohou být např. zobrazena nebo uložena do databáze – např. při importu zálohy databáze, z níž se data ztratila.

## Parsování XML

Parsování je proces, kdy z XML souboru data načteme/načítáme a zpracováváme (například vypisujeme nebo postupně analyzujeme). K souboru XML, který chceme parsovat, můžeme přistupovat (číst ho) naráz, nebo proudově. V prvním případě se celý soubor načte do operační paměti, kde po celou dobu parsování zabírá místo. Jde o jednodušší a rychlejší způsob parsování.

Při proudovém čtení je XML soubor čten po částech a je postupně zpracováván. Nezabírá takové místo v operační paměti (vždy je v ní uložena jen daná aktuálně zpracovávaná část).

Např. v jazyce PHP je nejjednodušší XML soubor převést do objektu a pak k němu také objektově přistupovat (např. `$clovek->bydliste->mesto`). Jde o zpracování XML naráz a používá se k němu funkce `simplexml_load_file($file)`.

## JSON

JavaScript Object Notation – soubor dat používaný zejména v jazyce JavaScript (ve své podstatě jde o objekt v JS). Jeho velikost (při stejných datech) bývá poloviční oproti XML, protože není třeba uzavírat žádné elementy. Jeho struktura může vypadat např. takto:

```
{jmeno: 'Tomáš', prijmeni: 'Černý', bydliste: {ulice: 'Červeného', cisloDomu: '2', mesto: 'Brno'}, vek: 18}
```