

Chybná hlavička HTTP Content-Type

Otevřená data se vystavují na webu a jsou přístupná typicky přes protokol HTTP(S). Ten má svá pravidla, která je pro správné fungování webu nutné dodržovat. Jedno z pravidel se týká indikace toho, jaký je formát zdroje (stránky, souboru) je přenášen. Webový server tedy v odpovědi na požadavek klienta na stažení souboru posílá nejen obsah samotný, ale i metainformace, tzv. HTTP hlavičky. Ty si můžeme zobrazit pomocí standardního nástroje curl: `curl -I <URL>`, nebo ve vývojářské konzoli běžného webového prohlížeče. V jedné z nejdůležitějších hlaviček server klientovi sděluje, v jakém formátu jsou přenášena data. Jedná se o HTTP hlavičku Content-Type a hodnotou v této hlavičce je tzv. Media Type, nebo také **MIME Type**, který například pro CSV soubory je `text/csv`. Je-li tato hlavička nastavena špatně, může dojít k chybné interpretaci přenášených dat. Hodnota v této hlavičce by tedy měla odpovídat jak přenášenému obsahu, tak také hodnotě uvedené v metadatech distribuce datové sady. Toto se týká každého zdroje na webu, v kontextu otevřených dat tedy zejména souborů s daty, datových schémat, podmínek užití a dokumentací datových sad. Důležitou roli zde hraje taky správně **implementovaná komprese**.

Seznam nejpoužívanějších datových typů a odpovídajících MIME typů

Úplný seznam MIME typů spravuje [IANA](#) - Autorita pro přidělování čísel (identifikátorů) na Internetu.

Formát dat	MIME-typ
CSV soubor	<code>text/csv</code>
Schéma pro CSV dle CSV on the Web	<code>application/csvm+json</code>
Schéma pro CSV dle Table Schema	<code>application/json</code>
JSON soubor	<code>application/json</code>
Schéma pro JSON soubor	<code>application/json</code>
XML soubor s převážně strukturovaným obsahem	<code>application/xml</code>
XML soubor s převážně textovým obsahem	<code>text/xml</code>
Schéma pro XML dle XML Schema	<code>application/xml</code>
RDF soubor v serializaci Turtle	<code>text/turtle</code>
RDF soubor v serializaci TriG	<code>application/trig</code>
RDF soubor v serializaci N-Triples	<code>application/n-triples</code>
RDF soubor v serializaci N-Quads	<code>text/turtle</code>
RDF soubor v serializaci JSON-LD	<code>application/ld+json</code>
RDF soubor v serializaci RDF/XML	<code>application/rdf+xml</code>

Seznam nejčastějších chybně používaných MIME typů

Zde uvádíme MIME typy, se kterými se nejčastěji setkáváme u chybně poskytovatných otevřených dat.

Formát dat	MIME-typ
HTML stránka	<code>text/html</code>

Formát dat	MIME-typ
Text	text/plain
Proud osmic (bajtů)	application/octet-stream

Konfigurace webového serveru pro správnou indikaci MIME typů

Aby byl MIME typ správně indikován, je obvykle třeba korektně nastavit webový server ([nginx](#), [Apache](#), [Microsoft IIS](#), ...). Pokud jde o staticky poskytované soubory, používá se obvykle tabulka párující přípony souborů a MIME typy. Je třeba dát pozor na to, že soubory se stejnou příponou mohou odpovídat různým MIME typům. Například obecný JSON soubor má příponu .json stejně jako schéma pro CSV, MIME typy se ale liší. Je pak třeba vhodně strukturovat konfiguraci serveru, případně umístění souborů v souborovém systému serveru. Pokud se jedná o dynamicky poskytovaná data, hlavičku Content-Type nastavuje služba, která data generuje.

MIME typy a kódování

Otevřená data se obvykle reprezentují v textových formátech. Znaky v těchto souborech pak z pravidla používají kódování UTF-8, ale u starších XML souborů se můžeme v českém prostředí setkat s kódováními ISO-8859-2 nebo Windows-1250. Je proto vhodné HTTP hlavičku Content-Type rozšířit o informaci indikující použité kódování, celá hlavička pak vypadá například takto: Content-Type: text/csv; charset=utf-8. Například pro webový server nginx je třeba [použít v konfiguraci](#) charset utf-8; a ujistit se, že v hodnotě charset-types je uvedeno text/csv.

From:
<https://opendata.gov.cz/> - **Otevřená data**

Permanent link:
<https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:http-content-type>

Last update: **2020/06/03 09:37**

