

3B OTVORENOST ZAPISA PODATAKA

- Be liberal in what you accept and conservative in what you send

CSV

- comma separate value
- svaki zapis u svom retku, odvojen sa CRLF
- prazna mjesta se ne smiju ignorirati
- može imati redak zaglavlja

XML

- SGML
 - Preteča xml-a, html-a
 - Standard generalized markup language- prva pojava meta jezika koji opisuje druge jezike temeljene na oznakama
 - Ključne komponente- elementi, atributi, tipovi podataka, DTD- document type definition
 - Element `<e> </e>`, sadrži attribute i sadržaj
 - DTD sadrži formalnu gramatiku
- Extensible markup language
- Orijentiran prema strojnoj obradi- uvjetuje veće zahtjeve pri radu sa sadržajem dokumenta
- **Proširivost**-oznake se definiraju po potrebi
- **Odjeljivanje podataka od prezentacije**-opisuju podatke, ne izgled
- **Validacija**-stroga pravila
- **Internacionalizacija**-izvorno koristi UTF8
- **Prenosivost**-obična text datoteka
- **Rasprostranjenost**-neovisan o platformi, dobro podržan u većini jezika
- Ne propisuje oznake nego pravila definiranja
- Oznake definiraju hijerarhijske odnose
- Jezik za opis podataka-konačan broj oznaka, struktura podataka, semantika
- Tipovi dokumenta se definiraju DTD-om ili XML Schemom
- Semantika pojedinih oznaka se definira implicitno
- Normiranje po domenama primjene
 - Horizontalna-između aplikacija u istoj domeni, normiranje onemogućuje ovisnost o proizvođaču aplikacije
 - Vertikalno- u svim domenama ili neovisno o domeni
- `<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>`
- CDATA- blokovi koje parser ignorira
- Bijeli znakovi- razmak, tab, CR, LF- važni i nevažni
- Namespace-prefiksi za svaki dokument zasebno
- `<ex:eindex xmlns:ex="http://www.fer.hr/eindex/ver1.0">`

- Dobro oblikovan dokument- parseri odbijaju obradu loših
 - Struktura dokumenta
 - Imenovanje elemenata i atributa
 - Gniježđenje elemenata
- Definicija tipa dokumenta-koje oznake se mogu koristiti, sve moguće hijerarhijske strukture
- Ne određuje eksplicitno značenje oznaka i struktura
- Definicija tipa dokumenta efektivno definira jezik oznaka- riječi, tvorba rečenica, implicitno značenje, jezik za definiranje opisnih jezika, metajezik
- Ako XML ima tip, parser koristi definiciju tipa za provjeru riječnika, vrijednosti elemenata i atributa te strukture
- **VALJAN XML zadovoljava provjeru tipa dokumenta, dobro oblikovan**
- **XML DTD- document type definition**
- Skup deklaracija o oznakama koje definiraju tip XML-dokumenta po određenim pravilima
- Unutarnji `<!DOCTYPE korijenski-element [DTD definicija]>` , unutar xml dokumenta
- Vanjski `<!DOCTYPE korijenski-element SYSTEM URI>` `<!DOCTYPE korijenski-element PUBLIC FPI URI>` ,u zasebnoj datoteci
 - Privatna i javna vanjska definicija
 - FPI- formal public identifier `norma//odgovoran//tip-dokumenta//jezik`
- Provjerava-elemente, attribute, entitete
- `<!ELEMENT ime-elementa (sadržaj)>`
 - + min 1
 - * 0-n
 - ? 0-1
 - , slijed
 - | xor
 - () grupiranje
- `<!ATTLIST element atribut tip default-vrijednost>`
- Element čiji se atribut definira mora biti definiran, ima atributa jedinstveno
- **NEDOSTATCI**
 - nema provjere valjanosti podataka kod jednostavnog sadržaja
 - nužnost provjere nameće korištenje atributa umjesto elemenata za pohranu podataka
 - slaba provjera tipova podataka kod atributa
 - nema namespace
 - nije xml, nije dobro oblikovan dokument
- **XML SCHEMA**
- Ispravlja nedostatke
 - Definiranje različitih tipova sadržaja elemenata i atributa
 - jednostavnije i preciznije mogućnosti definiranja strukture

- definiranje novih tipova i zapisa podataka
 - namespace
 - xml
- dokument dobro oblikovan i valjan
- elementi `<xsd:element name="ime" type="tip" ...>`
- Atributi `<xsd:attributetype="xsd:string"`
- Tipovi
 - `<complexType` elementi i/ili atributi
 - `<simpleType` string, broj,uri, datum
- Definiranje novih ograničavanjem jednostavnih- dodavanje atributa jednostavnim, navođenjem elemenata koji čine složeni tip, proširenje ili ograničavanje definiranih složenih tipova
- DTD za jednostavnije,kraće dokumente, bitna provjera strukture, mogu se uređivati ručno
- XML Schema-složeno, bitno provjera strukture i tipova vrijednosti elemenata i atributa, duža definicija, treba alate
- Parsiranje- provjera dobre oblikovanosti, neki i provjera valjanosti
 - Linijsko SAX- ne stvara reprezentaciju, temeljen na događajima
 - Potpuno DOM- stvara objektni model dokumenta, stablo
 - Document object model
 - *is a **platform and language- neutral** interface that allows programs and scripts to dynamically **access** and **update** the content, structure, and style of a document*
 - standardni skup razreda i standardne funkcije
- Uređivanje-objektni model- uređivanje sadržaja i strukture, stvaranje modela novog dokumenta, izvoz modela u formatu xml
- Node, nodeList,namedNodeMap
- SAX brži, manje memorije, jednostavniji
- DOM kod manipulacije sa cijelom strukturom, ako se xml često koristi
- XSL – extensible stylesheet language
- Skup preporuka za definiranje transformacija i prezentacije xml dokumenta
- XPATH- pristup i lociranje dijela xml dokumenta
- XSLT- transformations, xml-> druge vrste, stylesheet
- XSL FO

JSON

- Jednostavan, otvoren, neovisan o platformi i jeziku

- Razmjena ili zapis jednostavnih struktura podataka
- Utf8
- Aplikacije weba
- Izravno mapiranje u podatke i strukture JS-a
- Podrška parsiranju na platformama koje podržavaju JS
- API- parse, stringify
- Datatypes
 - Osnovni
 - String
 - Number
 - Boolean
 - Null
 - Složeni
 - Array
 - Object
- Listovi i grane
- Dobra oblikovanost-parser-sintaksa i struktura složenih tipova
- Valjanost-vanjska aplikacija- valjana struktura i vrijednosti
- **Nedostatci**
 - Mali broj definiranih tipova – datum!
 - Ugnježđivanje kraćih binarnih zapisa
 - Podrazumijevana preciznost brojeva
 - Korištenje null u definiciji stringa
- Json Schema- opisivanje i validiranje- type, properties, items

4A OTVORENI PODATCI

- **Open data** and content can be **freely used, modified, and shared** by **anyone** for **any purpose**
- Pravno, tehnički, financijski- licenca, format, besplatni
- **POTPUNI**
- **PRIMARNI**
- **PRAVOVREMENI**
- **PRISTUPAČNI**
- **STROJNO ČITLJIVI**
- **NEDISKRIMINIRAJUĆI**
- **POD OTVORENOM LICENCOM**
- **NEVLASNIČKI**
- Online, besplatni
- Trajni
- Provjereni

- Presumpcija otvorenosti
- Dokumentirani
- Sigurni za otvaranje
- Uz doprinos javnosti
- FAIR data
 - Findable
 - Accesible
 - Interoperable
 - Reusable
- Po strukturi- tabularni, hijerarhijski, mrežni, pod. Kocke
- Po namjeni- zapisi, grafovi, po poretku
- Izrada
- Odabir licence
- Dostupnost i pristupačnost
- Vidljivost
- Open data portals

4B OTVORENI PODATCI

- METAPODATCI-podatci koji detaljno opisuju skupo podataka, strukturu i pojedine dijelove
- Strojno čitljivi
- Lakše otkrivanje i pristupačnost podataka
- Lakše tumačenje
- Lakše povezivanje, interoperabilnost i razmjena podataka
- Lakša analiza kvalitete podataka

RJEČNICI

DC

- Jednostavan model metapodataka, osnovni standard 15 elemenata

DCAT

- Catalog, resource, dataset, distribution, dataservice, catalogrecord

SCHEMA.ORG

- Nezavisna specifikacija rječnika

MICRODATA

- Nema tipove podataka osim text i url
- Nema dobru podršku za internacionalizaciju
- Primarno namijenjen za web stranice

- Za skupove podataka najprikladniji JSON-LD

RDFA- resource description framework in attributes

JSON LD- linked data

- Prikladan za označavanje skupova podataka

5 zvjezdica otvorenih podataka

- Dostupnost na webu u bilo kojem obliku pod otvorenom licencom
- Dostupnost računalno čitljivih strukturiranih podataka
- Otvoreni formati zapisa
- Otvorene norme w3c-a za označavanje pojmova
- Poveži s drugim podacima

4C POVEZANI OTVORENI PODATCI

- Uporaba uri-ja za imenovanje
 - Upotreba http uri-ja da bi se osobe i korisnički agenti mogli upućivati na stvari
 - Pružanje korisnih informacija prilikom pristupanja uri-ju
 - Prilikom objave na webu uključivanje linkova na druge stvari
-
- Rdf- resource description framework
 - Opis resursa u obliku grafa
 - Subjekt, predikat, objekt