

# XML

Informaatika II

Ermo Täks

# Mis on XML?

- Dokumendi märgenduskeel struktuurse info defineerimiseks.
- Tegemist ei ole:
  - Programmeerimiskeelega
  - Andmebaasiga
  - veebidisainiga

# Mis on märgenduskeel?

- Tavalist Wordi dokumenti kirjutades tehakse tegelikult kirjutamisest märksa rohkem ehk tehakse otsuseid ka teksti kujunduse kohta:
  - *Kas Times New Roman font või Geneva?*
  - *Must või punane värv?*
  - *Kas tähe suurus on 12 või 36 punkti?*
  - *Kas reavahetus toimub 45 tähemärgi või 60 tähemärgi järel?*
  - *Kas kolmanda peatüki järel tuleb horisontaalne joon?*
  - *Kas pildid on süvistatud dokumenti või mitte?*
  - *jne*

# Mis on märgenduskeel?

- Enamikes tekstiprotsessorites on info teksti välimuse kohta peidetud
- Märgenduskeel on kood, mille vahendusel arvuti saab teada peidetud infot tekstidokumendi kohta.

# Mis on märgenduskeel? HTML Näide

```
<!DOCTYPE html>
<html>
  <body>
    <h1>Minu esimene veebileht!</h1>
    <p>
      <ul>
        <li> Kas mu tekst on kaldkirjas?</li>
        <li> Kas tekst on alaindeks või ülaindeks</li>
        <li> Kas reavahetus toimub 45 tähemärgi või 60 tähemärgi
järele?</li>
        <li> Kas kolmanda peatüki järel tuleb horisontaalne joon? </li>
        <li> Kas pildid on süvistatud dokumenti või mitte? </li>
        <li> jne </li>
      </ul>
    </p>
  </body>
</html>
```

# Harjutus

- Avage HTML redaktor [siit](#)
- Vajutage nuppu “Try it Yourself!”
- Kopeerige eelmise lehe sisu vasakusse veebiaknasse
- Vajutage ülaservas nuppu “Submit code”
- Vaadake tulemust
- Lisage tekstile “Kas mu tekst on kaldkirjas” ümber algusmärgis `<i>` ja lõpumärgis `</i>`
- Paigutage järgmises lauses elemendid `<sub>` `</sub>` ja `<sup>` `</sup>` õigetesse kohtadesse!

# Sisu määramine

- Arvutitega saab määratleda märksa rohkem kui vaid teksti välimust
- Seda kasutatakse ka teksti tegeliku sisu kirjeldamiseks

# Sisu määramine. Semantika

- Nimi Ermo Täks
  - Arvuti jaoks suvaline tähtede kombinatsioon: e, r, m, o, ,t, ä, k, s
  - Lugeja jaoks
    - praktikumi läbiviija
  - Veel tähendusi:
    - Sugupuus Rita ja Valloti poeg
    - Riigikantselei kodulehel endine IT arendusjuhti
    - TTÜ kodulehel praegune lektor
  - Taolist lisainfot ei võimalda HTML keel kirjeldada



# Struktuurne info

- Luues dokumendi sisu, loome dokumendi struktuuri
- Strkutureeritud info on midagi enam kui vaid tekst ekraanil
- Pannes nime struktuursesse taustsüsteemi, saab arvuti teada märksa rohkem kui vaid tähekombinatsioon: e, r, m, o, ,t, ä, k, s
- Arvut saab teada, et info on Ermo Täksi kui
  - endise Riigikantselei IT arendusjuhi
  - Ermo Täksi kui Rita ja Valloti poja kohta.

# Struktuurne info

- Tavaelus on palju näiteid sellise struktuurse info kohta: Aadressiraamat
  - Iga sisend on defineeritud,
  - iga sissekande osa on defineeritud:
    - Nimi
    - Aadress
    - Linn
    - Telefoni number
    - Jne

# Struktuurne info

- Seadus on rangelt struktureeritud:
  - Pealkiri või seaduse nimi
  - preambula
  - peatükk,
  - paragrahv,
  - lõige,
  - punkt.

# Mis see XML siis on?

- XML on märgenduskeel
  - struktuurse info defineerimiseks
- Seda kasutavad arvutid, et määratleda
  - dokumendi struktuuri ja/või
  - peidetud informatsiooni

# Mis see XML siis on?

- Tegemist on metakeelega ehk keelega, mille abil luuakse teisi keeli.
- Luues näiteks aadressiraamatut, võib vaja minna järgmisi positsioone:
  - `<sisend>`
  - `<nimi>`
  - `<aadress>`
  - `<linn>`
  - `<e-aadress>`

# Mis see XML siis on?

- Aadressiraaamatu sisendiks võib sellisel juhul olla järgmine kirje
  - `<sisend id="1">`
  - `<nimi>Ermo Täks</nimi>`
  - `<aadress>Linnu tee 16</aadress>`
  - `<linn>Tallinn</linn>`
  - `<e-aadress>ermo.taks@ttu.ee</e-aadress>`

# Miks XML hea on?

- Eristab andmed HTML-is
  - võimaldab hoida andmed eraldi HTML failist, mis teeb veebilehe sisu muutmise lihtsaks
- Lihtsustab andmete vahetamist
  - Nn. vahekeel erinevas formaadis hoitavate andmete vahel
  - Dokument võib olla .docx või .doc; .pdf, .html vms laiendiga
- Lihtsustab andmete edastamist
  - Vähendab kulutusi andmete edastamisel üle interneti erinevate keskkondade ja tarkvarade vahel
- Teeb andmed kättesaadavamaks
  - Lihtsustab andmetele ligipääsemist, eelkõige rakendustele, aga ka inimestele

# Definitsioon

- XML on W3C soovitatud üldotstarbeline markeerimise keel, mille eesmärgiks on
  - struktureeritud info jagamine
  - eelkõige veebipõhistes rakendustes.
- XML on laiendatav keel
  - on võimalik defineerida oma elemente
- XML-i võib kasutada spetsiifiliste märgiste loomiseks
- Võib kasutada internetis juba olemasolevaid märgendikeeli



# XML-i kontrollimine

- XML on dokumenteeritud keel
- XML spetsifikatsiooni haldab [W3C](#)
- W3C on akronüüm, mis viitab World Wide Web Consortium-ile.
- Tegemist on grupiga, mis määratleb internetikeskkonna toeks kasutatavate tehnoloogiate standardeid

# XML-i standard

- Värskeim XML standard on kättesaadav [siit](#)
- Kirjeldab XML-i täielikult
- Võib olla keeruline lugemiseks ja arusaamiseks
- Kursusel käsitleme erinevaid osasid XML dokumendi paremaks mõistmiseks

# XML mõisted

- Tähed (characters)
  - Üks tekstiühik, nagu täht, number, tühik, taabeldus jt.
- DTD
  - Dokumendi tüübis (Document Type Definition), XML dokumendi grammatika
  - Paikneb tavaliselt õige (valid) dokumendi päises või osutab kohale, kus asub dokumenditüübis

# XML mõisted

- Sisend
  - XML dokumendi säilitusüksus
  - Iga XML dokument koosneb ühest või enamast sisendist
  - Näiteks järgmine HTML märgend `<html></html>` kirjeldab ühe HTML sisendi
- XML
  - Märgendikeel
- Kommentaarid
  - XML kasutatakse kommentaaride märkimiseks `<-- -->`

# XML mõisted

- XML dokument
  - XML spetsifikatsiooni kohaselt hästi koostatud (valideeritud ja struktuurselt korrektne) dokument
  - Grammatiliselt õige ("well formed") ja kehtiv ("valid") pole päriselt üks ja seesama.
  - Võib olla grammatiliselt õige XML dokument mis pole kehtiv ja vastupidi.
  - Hästi koostatud dokumendis on kontrollitud:
    - kas seal kasutatakse õigete nimetustega elemente,
    - kas dokumendi struktuur on õige
    - ega dokumendis ei sisaldu muid häireid

# XML dokumendi põhiosad

- XML dokument algab juur-elementiga
  - Arvuti on rumal
  - Arvuti vajab juhendeid
  - XML deklaratsioon annab vajaliku juhendi:
    - `<?xml version="1.0">`
  - Esmane ja olulisim dokumendi osa, kohustuslik igale XML dokumendile
  - Sisuliselt ütleb veebibrauserile “Vaata, siit tuleb XML!” ja annab edasise töötlemise juhendid

# XML dokumendi põhiosad

- Esimene tekstirida on XML deklaratsioon milles määratakse XML versioon:
  - `<?xml`
    - *Tähendab: olen xml dokument*
  - `version="1.0"`
    - *Tähendab: kasutan XML esimest versiooni dokumendi kirjeldamiseks*

# XML dokumendi põhiosad

- *encoding*="ISO-8859-1"
  - Tähendab: kasutan tähemärke, mis kuuluvad hulka: Latin-1 / West European character set
- *Standalone*="yes"?
  - Tähendab: kas on vaja lisaks kasutada teisi XML dokumente, et parsida XML dokumenti. Näiteks kas on eraldi DTD faili kasutatud jne.
- XML deklaratsioon tervikuna:  
**<?xml version="1.0" encoding="ISO-8859-1" Standalone="yes"?>**



# XML dokumendi põhiosad

- Järgmine rida (<sisend>), kirjeldab dokumendi **juur-elementi**.
- Järgmised neli rida on juur-elementi **tütar-elementid**.
- Viimane rida tähistab juur-elementi lõppu

# XML dokumendi põhiosad

- XML dokument algab dokument juur-elementidega, jätkub temaga suguluses olevate elementidega:

```
<juur>
```

```
  <tütar>
```

```
    <tütretütar>.....</tütretütar>
```

```
  </tütar>
```

```
</juur>
```

# Põhireeglid

- Kõik elemendid sisaldavad teksti ja atribuute.
- **Kõik XML elemendi peavad olema suletud lõpumärgisega (</...>).**
  - <aadress>See on aadress</aadress>

# Põhireeglid

- XML on tõstutundlik s.t. et `<teema>` pole sama mis `<Teema>`.
  - `<Teema>See on valesti</teema>`
  - `<teema>See on õigesti</teema>`
- XML'is peavad elemendid olema õieti pesastatud:
  - `<b><i>Selles reas on pesastatud valesti</b></i>`
  - `<b><i>Selles reas on pesastatud õigesti</i></b>`

# Põhireeglid

- XML'is peavad atribuutide väärtused olema jutumärkides:
  - `<teema daatum="10/02/2012">`
- Kommentaare XML'is kirjutatakse samamoodi nagu HTML'is
  - `<!--See on kommentaar-->`

# Harjutus. Grammatilise õigekirja kontroll

- Avage veebileht
- Leidke ja eemaldage vead järgmisest tekstist:

```
<?xml version="1.0" ?>
```

```
<Kontaktandmed>
```

```
  <sisend id=1>
```

```
    <nimi>Ermo Täks>/nimi>
```

```
    <aadress>Linnu tee 16</adress>
```

```
    <linn>Tallinn<linn>
```

```
    <e-aadress>ermo.taks@ttu.ee</eaadress>
```

```
  </sisend>
```

```
</kontaktandmed >
```

# Ülesanne

- Koostada iseseivalt korrektse struktuuriga kehtiv XML dokument järgmise sisu põhjal

„VABARIIGIPRESIDENT

OTSUS

Nr 183

Kadriorus 12. novembril 2012

Kohtunikuks nimetamine

Nimetan teise astme kohtunikuks 2. jaanuarist 2013:

Tiina Pappel

Alus: Eesti Vabariigi põhiseaduse § 78 p 13 ja § 150 lg 3,  
kohtute seaduse § 55 lg 1.

Toomas Hendrik Ilves“