Tartu Ülikool

Arvutiteaduse Instituut

XML LEKSIKON

Tarkvarasüsteemide seminari töö

Seminaristid:

Kaimar Karu [KK]
Kristin Kraav-Villems [KKV]
Gunnar Kudrjavets [GK]
Gunnar Nellis [GN]
Harry-Anton Talvik [HT]
Tiina Tamme [TT]
Vassili Tsvetkov [VT]

Seminari juhendaja: **Jüri Kiho [JK]**

Tartu 2000

Saateks

Käesolev leksikon on valminud *Tarkvarasüsteemide* seminari tööna 1999/2000 õppeaasta kevadsemestril. Seminarist osavõtjatel tuli esineda ettekandega ettenähtud teemal, tuvastada antud teemas käsitletud peamised mõisted ja osaleda neid kirjeldavate artiklite koostamisel.

Kuna eestikeelne terminoloogia on veel välja kujunemata, siis rakendasime paindlikku süsteemi, mille kohaselt on põhitekstis eestikeelsete terminite asemel kasutatud (ingliskeelseid) makrosid. Mingi termini asendamisel uuega tuleb lihtsalt teha muudatus makrode tabelis ja kompileerida uuesti põhitekst.

Vormistuslikud vahendid on kokkuvõtlikult esitatud järgmises "artiklis":

märksõna, *ingliskeelne vaste* – antud märksõnale (mis võib olla ka sõnaühend) vastavat mõistet selgitav artikkel; *kaldkirjas* on esitatud siin muuhulgas määratletud mingi teine mõiste (kui viimase kohta ongi omaette artikkel, siis sisaldab see vaid viidet antud märksõnale); kujul →mingi teine märksõna viidatakse käesoleva artikliga lähedalt seotud mingile teisele märksõnale; XML konstruktsioonide konkreetsed elemendid esitatakse `kirjutusmasinakirjas`, meta-elemendid aga *kursiivis*, mittekohustuslikud elemendid võetakse nurksulgudesse, meta-termin kujul *ühikud* tähendab tühemikega eraldatud ühikuid, metasümbol $\flat$ aga mittekohustuslikku tühemikku; vajadusel esitatakse lõpus märksõnu, mille artikleid on soovitatav lisaks vaadata, ning antud märksõna sünonüüme; artikli järel on nurksulgudes märgitud autor (esimesena) ja retsensent või kaasautor. Vt. ka *täiendavad märksõnad.* Sünon. *termin.* [AU/RE]

Käesolev koopia on üksnes leksikoni tööversioon, mis kahtlemata vajab parandamist ja täiendamist.

Seminari juhendaja

Seminari töö alusmaterjalid:

1. S. North, P. Hermans. Teach yourself XML in 21 days. Sams, 1999, 580
2. Extensible Markup Language 1.0 `http://www.w3.org/TR/REC-xml`

Seminari ettekannete materjalid:
`http://madli.ut.ee/~tiina/xml/seminar.html`

asendustekst, *entity value* – →üldolem.

atribuudi väärtus, *attribute value* – →atribuut.

atribuut, *attribute* – →elemendiga seostatud täiendav info kujul

atribuudi_nimi♭`=`♭`"`*atribuudi väärtus*`"`

kus *atribuudi väärtus* on sõne, mis ei sisalda sümboleid '<', '`"`' ja '`&`' (üksiksümbolina), kuid võib sisaldada üldolemi viita ja sümboliviita. [JK/GN]

atribuutloend, *attribute-list* – →elemendis nime järel toodud →atribuutide loetelu (loetelu liikmete eraldajaks on tühemik). [JK/GN]

atribuutloendi kirjeldus, *attribute-list declaration* – elemendi kõigi atribuutide kirjeldus DTD-s kujul

<`!ATTLIST` *elemendi_nimi atribuudikirjeldused*♭>

atribuudikirjeldus esitatakse kujul

atribuudi_nimi atribuuditüüp lisanõue

võimalik lisanõue:

`#REQUIRED`

– atribuut on elemendis kohustulik,

`#IMPLIED`

– atribuut on vaikeväärtuseta,

atribuudi väärtus

– atribuudi vaikeväärtus,

`#FIXED` *atribuudi väärtus*

– atribuudi ainuvõimalik väärtus;

võimalik atribuuditüüp:

`CDATA`
`ID`
`IDREF`
`IDREFS`
`ENTITY`
`ENTITIES`
`NMTOKEN`
`NMTOKENS`
`NOTATION (`♭*valik notatsiooninimesid*♭`)`
`(`♭*valik nimesid*♭`)`

valikus (kahel viimasel juhul) on nimede eraldajaks ♭`|`♭. [JK/GN]

avaliteraal, *public literal* – välisressursi üldinfot sisaldav →välisosuti osa, tavaliselt kujul

"*tüüp//omanik//kirjeldus//keel*"

kus tüübiks on sümbol pluss (`+`), kui omanik on registreeritud ISO 9070 järgi, muidu miinus (`-`); keel antakse kahetähelise keeletähisena ISO 639 järgi. [GN/JK]

dokument, *document* – tekst kujul

[<`?xml` *versioon*[*kodeeringu kirjeldus*][*iseseisvusdeklaratsioon*]♭`?`>]♢
DTD♢
juurelement♢

märgiga ♢ tähistatud kohtades võib esineda suvaline arv →töötluskäske, →kommentaare ja tühemikke; *juurelemendiks* on →element. Vt. ka *versioon, kodeeringu kirjeldus, iseseisvusdeklaratsioon, DTD.* [JK/KKV]

DTD, *DTD (Document Type Definition)* – →dokumendi juurelemendile eelnev (välimine ja/või sisemine) kirjelduste osa, milles kirjeldatakse

– dokumendis kasutatavate →elementide struktuur ja elementidevahelised seosed ning elementide atribuudid,
– →olemid,
– →notatsioonid;

üldkuju:

<`!DOCTYPE` *juurelemendi_nimi*
[*välisviit_välis-alamhulgale*]♭
[*sise-alamhulk*]♭
>

Vt. ka *välis-alamhulk, sise-alamhulk.* [JK/KK]

elemendikirjeldus, *element declaration* – elemendi andmeosa kirjeldus DTD-s kujul

<`!ELEMENT` *elemendi_nimi sisukirjeldus*♭>

võimalikud lihtsamad sisukirjeldused:

`EMPTY`

– elemendi andmeosa on tühi;

`ANY`

– elemendi andmeosa on suvalise struktuuriga;

`(`♭`#PCDATA`[*alamelementide_nimede_loetelu*]♭`)`

– elemendi andmeosaks on kas sõne või üks loetletud alamelementidest; alamelementide nimede loetelus iga alamelemendi nimele eelneb ♭|♭(tähenduses "või");
alamelementide keerukamat struktuuri esindab regulaaravaldisena antud sisukirjeldus

    (♭*rühm*♭)[*kordus*]

kus kordust väljendab: '?' (0 või 1 korda), '*' (0 või enam korda), või '+' (1 või enam korda); rühm koosneb ühest või mitmest liikmest, iga liige võib olla varustatud veel kordusega, liikmeks on kas alamelemendi nimi või omakorda

    (♭*rühm*♭)

mitmeliikmelise rühma sees kasutatakse liikmete eraldajana parajasti ühte järgmisest kahest eraldajast, valikurühma korral

    ♭|♭

jadarühma korral

    ♭,♭

[JK/HT]

element, *element* – terviklik andmeosa XML dokumendis, esitatakse kujul

    <*elemendi nimi*[atribuudid]♭>[*sisu-osised*]</*elemendi nimi*>

kus *sisu-osiseks* on kas element (*alamelement*), →sümbolandmed, →üldolemi viit, →sümboliviit, →sõnesektsioon, →töötluskäsk või →kommentaar; *tühi element* ei sisalda ühtegi sisu-osist, sellise elemendi võib esitada ka kujul

    <*elemendi nimi*[atribuudid]♭/>

elemendi *andmeosa* moodustavad alamelemendid, sümbolandmed ja sõnesektsioonid. Vt. ka *atribuut*. [JK/HT]

HTML, *HTML (Hypertext Markup Language)* – →SGMLi rakendusena loodud hüperteksti →märgistuskeel, mis on mõeldud dokumentide publitseerimiseks veebis; HTMLi märgistus on orienteeritud eeskätt dokumendi kujundusele (esitusele), mitte sisule; võimaldab dokumendile anda teatud formaalse struktuuri: jagada dokument lõikudeks, vormistada tabeleid ja loetelusid, tuua välja pealkirjad ning alapealkirjad, hüperlinkide abil viidata teistele veebidokumentidele, lisada dokumendile multimeedia-objekte. [TT/JK]

iseseisvusdeklaratsioon, *standalone document declaration* – →dokumendi autonoomsuse e. täieliku komplekteerituse tunnus kujul

`standalone♭=♭"yes"`

või

`standalone♭=♭"no"`

esimene neist sätestab, et dokumendis puuduvad välised kirjeldused (kas DTD välis-alamhulgast või siis DTD sise-alamhulgast viidatud välise parameeter-olemi kaudu); teine neist (kehtib ka vaikimisi) sätestab, et selliseid väliseid kirjeldusi võib kasutusel olla. [JK/KKV]

juurelement, *root element* – →dokument.

kodeeringu kirjeldus, *encoding declaration* – →dokumendi või →välis-alamhulga alguses esinev kooditabeli määrang kujul

`encoding♭=♭"`*kooditabeli nimi*`"`

kohustuslik, kui kooditabeli nimeks ei ole UTF-8 või UTF-16; lubatud on ka

`encoding♭=♭-`

mis on samaväärne kodeeringu kirjelduse puudumisega. [JK/KKV]

kommentaar, *comment* – töötlemisel ignoreeritav dokumendi osa kujul

<`--`*ignoreeritav_sõne*`--`>

ignoreeritav sõne ei tohi sisaldada sümbolipaari "`--`". [JK/TT]

märgis, *tag* – →elemendi alguses ja lõpus esinev või tühja elementi esindav, sümbolitega '<' ja '>' piiratud sõne. Sünon. *märge*, *märgend*, *marker*, *silt*. [JK/TT]

märgistuskeel, *markup language* – arvutitekstide esitamise keel, mille korral teksti struktuur ja struktuuriosade omadused näidatakse teatavate →märgiste abil. [JK/TT]

nimesümbol, *name character* – täht, number, punkt, koolon, sidekriips, alakriips (või ka mõni muu →XMLi standardis lubatud sümboleist). [GN/JK]

nimeüksus, *name token* – suvaline mittetühi →nimesümbolite järjend. [GN/JK]

nimi, *name* – →nimeüksus, mis algab kas tähe, alakriipsu või kooloniga. [JK]

notatsioonikirjeldus, *notation declaration* – →notatsioon.

notatsioon, *notation* – välisressursi tähistus; kirjeldatakse DTD-s *notatsioonikirjeldusega*

<!NOTATION *notatsiooni_nimi välisosuti*♭>

millega antud nimi kuulutatakse välisosutiga näidatud ressursi tähiseks; näiteks võib välisressursiks olla programm (alljärgnevas ”lview.exe”), mis töötleb niisugust meediatüüpi, mida XML rakendus ise töödelda ei oska (tavaliselt audio-, pildi- või videofail):

```
<!NOTATION gif
   SYSTEM "file:///C:\Program Files\LView\lview.exe"
>
```

notatsiooni nime (antud juhul ”gif”) saab siis näidata välise parsimata üldolemi kirjelduses. Vt. ka *üldolem.* [JK/GK]

olemikirjeldus, *entity declaration* – →üldolemi või →parameeter-olemi kirjeldus DTD-s. [JK/VT]

olemiviit, *entity reference* – viit →sümbol-olemile, →üldolemile või →parameeter-olemile.

olem, *entity* – kaudselt, vaid olemiviida kaudu dokumenti lülitatav komponent; dokumendi töötlemisel olemiviit tavaliselt asendatakse vastava olemiga. Vt. ka *sümbol-olem, üldolem, parameeter-olem.* [JK/VT]

parameeter-olemi viit, *parameter entity reference* – viit →parameeter-olemile.

parameeter-olem, *parameter entity* – komponent, mis lülitatakse dokumenti olemiviidaga kujul

%*oleminimi*;

peab olema kirjeldatud DTD-s kas

<!ENTITY % *oleminimi asendustekst*♭>

või

<!ENTITY % *oleminimi välisviit*♭>

esimesel juhul on tegemist *sisemise parameeter-olemiga* (olemiks on asendustekst), teisel juhul – *välimise parameeter-olemiga* (olemiks on välisviidaga osutatud →välis-alamhulk); parameeter-olemi olemiviit võib esineda küll kõikjal, kuid selle asendamine olemiga toimub ainult DTD-s (mujal ei ole sümbolil '%' mingit eritähendust); juhul, kui olemiviit ei kuulu asendusteksti, siis viida asendamisel olemiga täiendatakse olemit algusest ja lõpust veel ühe tühikuga. [JK/VT]

SGML, *SGML (Standard Generalized Markup Language)* – üldistatud standardne (ISO 8879:1986) →märgistuskeel, mis on mõeldud arvutitekstide struktuuri ja sisu kirjeldamiseks riistvarast, operatsioonisüsteemist ja tootjast sõltumatu dokumendina; võimaldab →sisupõhist märgistust; kasutatav ka metakeelena eriotstarbeliste ja -kujuliste märgistuskeelte(SGMLi rakenduste) kirjeldamiseks, nt DocBook, MathML, →HTML, →XML. [TT/JK]

sise-alamhulk, *internal subset* – →DTD-s toodud, nurksulgudega ümbritsetud kirjelduste osa

[*kirjeldus-osised*]

kus *kirjeldus-osiseks* võib olla →elemendikirjeldus, →atribuutloendi kirjeldus, →olemikirjeldus, →notatsioonikirjeldus, →töötluskäsk, →kommentaar, →parameeter-olemi viit või ka tühemik. [JK/KK]

sisemine parameeter-olem, *internal parameter entity* – →parameeter-olem.

sisemine üldolem, *internal general entity* – →üldolem.

sõnesektsioon, *CDATA section* – sõne esitus kujul

<![CDATA[*sõne*]]>

(sisaldab nurksulupaari sisaldavat nurksulupaari);
sõnes ei tohi olla sümbolikolmikut ']]>'; kasutatakse selliste sõnede esitamiseks, mis sisaldavad sümboleid '<' ja/või '&'. [JK/HT]

sümbolandmed, *character data* – sõne, mis ei sisalda sümboleid '<' ja '&' ning sümbolikolmikut ']]>'. [JK/HT]

sümboliviit, *character reference* – viit →sümbol-olemile.

sümbol-olem, *character entity* – sümbol, mis lülitatakse dokumenti olemiviidaga kujul

&#*kood*;

kus koodiks on sümboli kas kümnendkood või siis kuueteistkümnendkood prefiksiga 'x' (*Unicode* kodeeringus); DTD-s on selline olemiviit lubatud vaid asendusteksti sees; ei vaja kirjeldamist DTD-s. [JK/VT]

süsteemliteraal, *system literal* – →välisosuti osa, mis näitab välisressursi asukohta; esitusvormiks on

"*URI*"

(URI = *Universal Resource Identifier*). [GN/JK]

tingimuslik sektsioon, *conditional section* – tunnusega (INCLUDE või IGNORE) varustatud →välis-alamhulga kirjeldus-osis kujul

<![♭INCLUDE♭[*välis-alamhulga_kirjeldus-osised*]]>

või

<![♭IGNORE♭[*ignoreeritav tekst*]]>

(sisaldab nurksulupaari sisaldavat nurksulupaari);
ignoreeritavas tekstis peavad piirajad <![ja]]> moodustama korrektse pesastuse; tunnuse INCLUDE korral antud kirjeldus-osised lülitatakse DTD-sse, tunnusega IGNORE varustatud sektsioon ignoreeritakse; tüüpiline rakendus: mõned objektid (nt. elemendid) kirjeldatakse kahe alternatiivse kirjeldus-osiste rühmana, millest mõlemast moodustatakse tingimuslik sektsioon, kummagi tunnus näidatakse oma parameeter-olemi viida abil (nt. %mustand; ja %puhtand;), seega tegelikku alternatiivi valikut saab siis reguleerida vastavate parameeter-olemite eelneva ümbermääratlemisega,

```
<!ENTITY % mustand 'INCLUDE'>
<!ENTITY % puhtand 'IGNORE'>
```

või

```
<!ENTITY % mustand 'IGNORE'>
<!ENTITY % puhtand 'INCLUDE'>
```

[JK/KK]

tühi element, *empty element* – →element.

töötluskäsk, *processing instruction* – kindlale töötlejale adresseeritud sõnumit (programmi, skripti või muud infot) sisaldav dokumendi osa kujul

<?*töötleja_nimi*[*sõnum*]?>

sõnumiks olev sõne ei tohi sisaldada sümbolipaari "?>"; töötleja nimi "xml" on reserveeritud kui eritähenduslik; näide:

```
<?realaudio
   version="5.0" frequency="5.5kHz" bitrate="16Kbps"
?>
```

[JK/GK]

versioon, *version info* – XMLi versiooni numbrit sätestav märkus →dokumendi või →välis-alamhulga alguses kujul

`version♭=♭"1.0"`

(käesoleval ajal on olemas vaid versioon numbriga 1.0). [JK/KKV]

väline parameeter-olem, *external parameter entity* – →parameeter-olem.

väline parsimata üldolem, *external unparsed general entity* – →üldolem. Sünon. *väline binaarolem.*

väline parsitud üldolem, *external parsed general entity* – →üldolem.

välis-alamhulk, *external subset* – väline, failina antud kirjelduste osa kujul

[`<?xml` [*versioon*] *kodeeringu kirjeldus*♭`?>`]

kirjeldus-osised

kus *kirjeldus-osistena* on lubatud kõik →sise-alamhulga kirjeldus-osised ning lisaks veel →tingimuslik sektsioon; kaasatakse →DTD-sse välisviida abil. [JK/KK]

välisolem, *external entity* – väline parameeter-olem või väline parsitud üldolem või väline parsimata üldolem. Vt. ka *üldolem, parameeter-olem.* [JK/VT]

välisosuti, *external reference* – osuti välisressursile, üks järgmistest:

`SYSTEM` *süsteemliteraal*

`PUBLIC` *avaliteraal süsteemliteraal*

`PUBLIC` *avaliteraal*

välisosutit, mis on antud kas esimesel või teisel kujul, nimetatakse *välisviidaks.* [JK/GK]

välisviit, *external identifier* – →välisosuti.

XML, *XML (Extensible Markup Language)* – laiendatav →märgistuskeel, →SGMLi alamhulk; metakeel, mille baasil saab luua ka uusi märgistuskeeli erinevate dokumenditüüpide jaoks; orienteeritud veebirakendustes kasutamiseks, üldisem kui varemloodud →HTML; on kujunemas uueks iseseisvaks standardiks; põhilised kasutusvaldkonnad: e-kaubandus jm andmevahetusega seotud rakendused. Vt. ka *http://www.w3.org/TR/REC-xml.* [TT/JK]

üldolemi viit, *general entity reference* – viit →üldolemile.

üldolem, *general entity* – komponent, mis lülitatakse dokumenti olemiviidaga kujul

`&`*oleminimi*`;`

peab olema kirjeldatud DTD-s kas

<`!ENTITY` *oleminimi asendustekst*♭>

või

<`!ENTITY` *oleminimi välisviit*♭>

või

<`!ENTITY` *oleminimi välisviit* `NDATA` *notatsiooninimi*♭>

esimesel juhul on tegemist *sisemise üldolemiga*, teisel juhul *välimise parsitud üldolemiga*; kolmandal juhul *välimise parsimata üldolemiga*;
sisemise üldolemi viit võib esineda elemendis (sisu-osisena), atribuudi väärtuses ja asendustekstis; viimasel juhul küll viida asendust olemiga ei toimu;
välimiseks parsitud üldolemiks on

[<`?xml` [*versioon*] *kodeeringu kirjeldus*♭`?`>]

sisu-osised

selle viit võib esineda elemendis (sisu-osisena) ja ka DTD-s, kuid seal ainult asendusteksti sees; viimasel juhul küll viida asendust olemiga ei toimu;
välimise parsimata üldolemi viit võib esineda ainult atribuudi väärtuse rollis, selle asendamine toimub vastavalt →notatsioonile. Sünon. *olem*. [JK/VT]

Inglise-eesti märksõnastik

Poolpaksus kirjas märksõnale vastab artikkel leksikonis.

attribute – **atribuut**
attribute-list – **atribuutloend**
attribute-list declaration – **atribuutloendi kirjeldus**
attribute value – **atribuudi väärtus**
CDATA section – **sõnesektsioon**
character data – **sümbolandmed**
character entity – **sümbol-olem**
character reference – **sümboliviit**
character set – märgistik
comment – **kommentaar**
conditional section – **tingimuslik sektsioon**
content model – sisumudel
content-oriented markup – sisupõhine märgistus
declaration – kirjeldus
document – **dokument**
DTD (Document Type Definition) – **DTD**
element declaration – **elemendikirjeldus**
element – **element**
empty element – **tühi element**
encoding declaration – **kodeeringu kirjeldus**
entity declaration – **olemikirjeldus**
entity – **olem**
entity reference – **olemiviit**
entity value – **asendustekst**
external entity – **välisolem**
external identifier – **välisviit**
external parameter entity – **väline parameeter-olem**
external parsed general entity – **väline parsitud üldolem**
external reference – **välisosuti**
external subset – **välis-alamhulk**
external unparsed general entity – **väline parsimata üldolem**
general entity reference – **üldolemi viit**
general entity – **üldolem**

HTML (Hypertext Markup Language) – **HTML**
internal general entity – **sisemine üldolem**
internal parameter entity – **sisemine parameeter-olem**
internal subset – **sise-alamhulk**
markup declaration – märgistuse kirjeldus
markup language – **märgistuskeel**
name character – **nimesümbol**
name – **nimi**
name token – **nimeüksus**
notation declaration – **notatsioonikirjeldus**
notation – **notatsioon**
occurrence indicator – esinemisindikaator
parameter entity – **parameeter-olem**
parameter entity reference – **parameeter-olemi viit**
processing instruction – **töötluskäsk**
public literal – **avaliteraal**
reference – viit
root element – **juurelement**
SGML (Standard Generalized Markup Language) – **SGML**
standalone document declaration – **iseseisvusdeklaratsioon**
system literal – **süsteemliteraal**
tag – **märgis**
version info – **versioon**
XML (Extensible Markup Language) – **XML**

PÕHIMÕISTETE KASUTUSLIK HIERARHIA

- *dokument*
  - *DTD*
    - *välis-alamhulk*
    - *välis-alamhulk*
    - *sise-alamhulk*
      - *kirjeldus-osis*
      - *elemendikirjeldus*
      - *atribuutloendi kirjeldus*
      - *olemikirjeldus*
      - *notatsioonikirjeldus*
      - *parameeter-olemi viit*
      - *töötluskäsk*
      - *kommentaar*
      - *kirjeldus-osis*
    - *sise-alamhulk*
  - *DTD*
  - *element*
    - *atribuutloend*
    - *atribuutloend*
    - *sisu-osis*
    - *element*
    - *sümbolandmed*
    - *olemiviit*
    - *sõnesektsioon*
    - *töötluskäsk*
    - *kommentaar*
    - *sisu-osis*
  - *element*
- *dokument*