



Interneto technologijos

XPath

XML dokumento dalių adresavimas



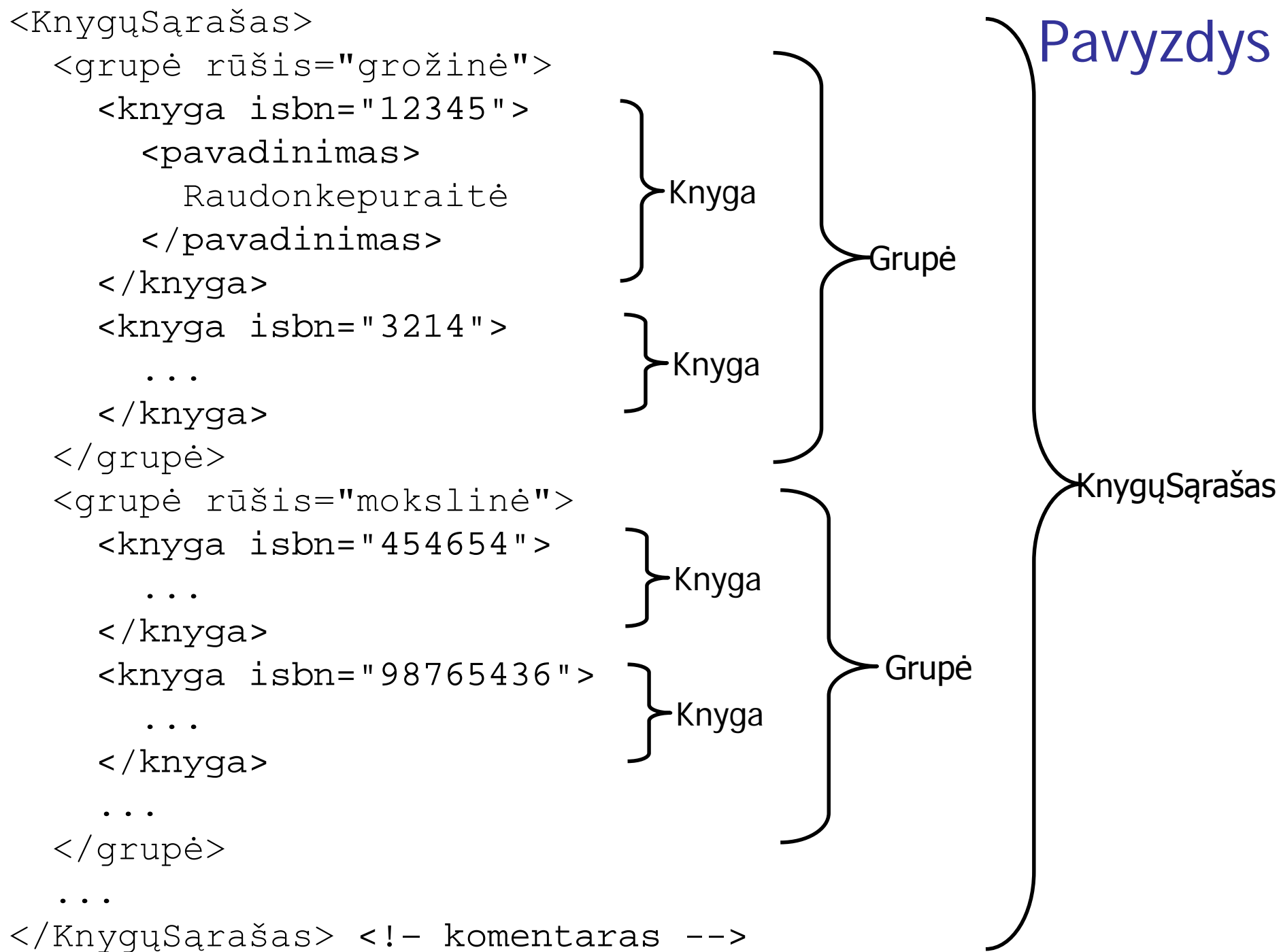
XPath standartas

- XPath 1.0 yra W3C standartas, skirtas XML dokumentų dalių adresavimui
 - <http://www.w3.org/TR/xpath>
 - Apie šią versiją bus kalbama per paskaitas
- 2007m. sausio 23 baigta ruošti versija 2.0
 - <http://www.w3.org/TR/xpath20>
 - praplečia 1.0 versiją naujais duomenų tipais (XMLSchema paprastais tipais), naujomis funkcijomis, ir pan.
 - Nėra suderinamas su XPath 1.0!



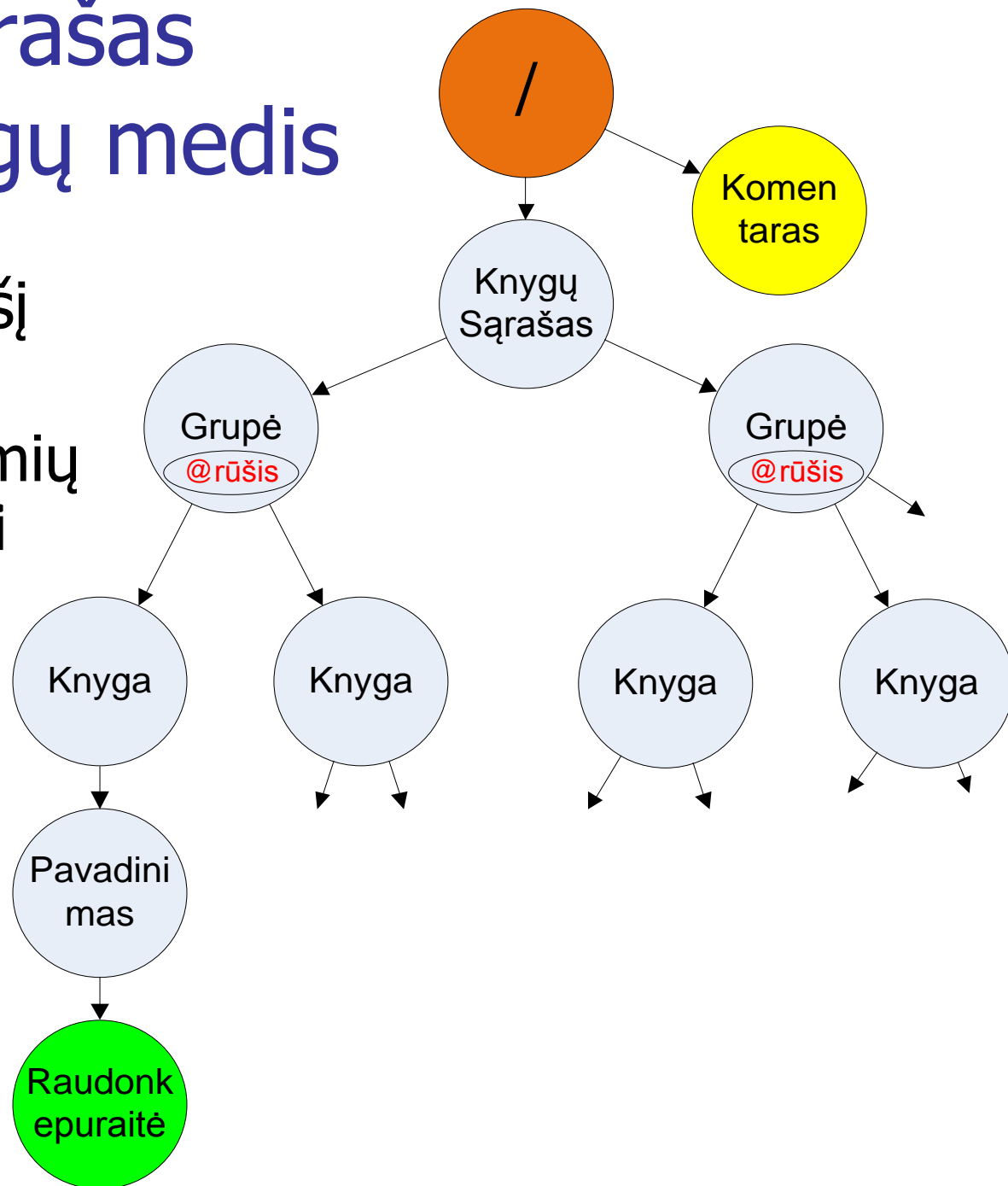
Loginė XML dokumento struktūra

- XPath modeliuoja XML dokumentą kaip **mazgų medį**
- Yra šios mazgų rūšys:
 - šakninis mazgas (/)
 - Tai ne šakninė XML žymė! XPath įveda papildomą mazgą, kurio vienas iš vaikų yra XML šakninė žymė. Kiti šio mazgo vaikai gali būti komentarai ir apdorojimo instrukcijos
 - žymės (pvz.: knyga)
 - tekstas ("Raudonkepuraite")
 - atributai (pvz.: rūšis, isbn)
 - komentarai (<!-- komentaras -->)
 - kiti mazgai (apdorojimo instrukcijos, vardų sritys, ...)
 - juos numato XML specifikacija, šiame kurse neanalizuosime



Knygų sąrašas kaip mazgų medis

- Rodyklės rodo ryšį "tėvas-vaikas"
- Atributai nėra žymių vaikai (!), bet turi žymes kaip tėvus

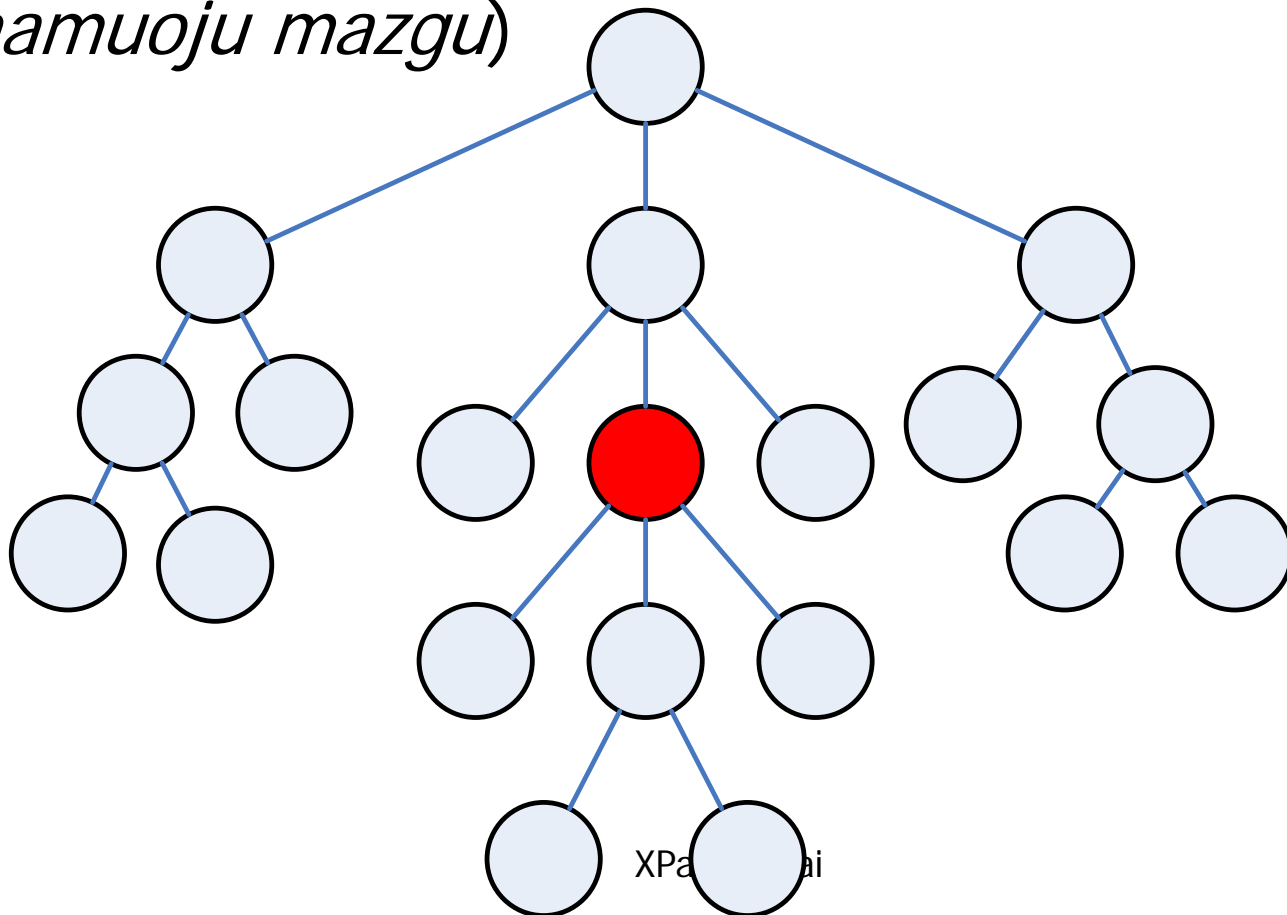




Pavyzdys

- Užklausa lietuvių kalba:
 - Kokia yra knygų sąrašo antros grupės penktos knygos ISBN atributo reikšmė?
- Ta pati užklausa XPath kalba:
 - `/knyguSarasas/grupe[2]/knyga[5]/@isbn`

- Sakykime, kad pasirinkome XML dokumento mazgų medyje vieną mazgą (jį vadinsime *einamuoju mazgu*)





Ašys

- `child` : vaikiniai mazgai (išskyrus atributus!)
- `attribute` : einamojo mazgo atributai, jei einamasis mazgas yra žymė
- `parent` : tėvinis mazgas (galima taikyti ir atributui – bus gražinta jį turinti žymė)
- `descendant` : mazgai-įpėdiniai (vaikai, anūkai, ..., išskyrus atributus)
- `ancestor` : mazgai-protėviai (tėvai, seneliai, ...)
- `following-sibling` : mazgai-broliai, einant į priekį nuo einamojo mazgo; netaikoma atributams
- `preceding-sibling` : mazgai-broliai, einant atgal nuo einamojo mazgo; netaikoma atributams
- `following` : visi mazgai, einant į priekį nuo einamojo mazgo, **išskyrus einamojo mazgo įpėdinius** ir išskyrus visus atributus
- `preceding` : visi mazgai, einant atgal nuo einamojo mazgo, **išskyrus einamojo mazgo protėvius** ir išskyrus visus atributus
- `self` : einamasis mazgas
- `descendant-or-self` : einamasis mazgas ir visi įpėdiniai
- `ancestor-or-self` : einamasis mazgas ir visi protėviai

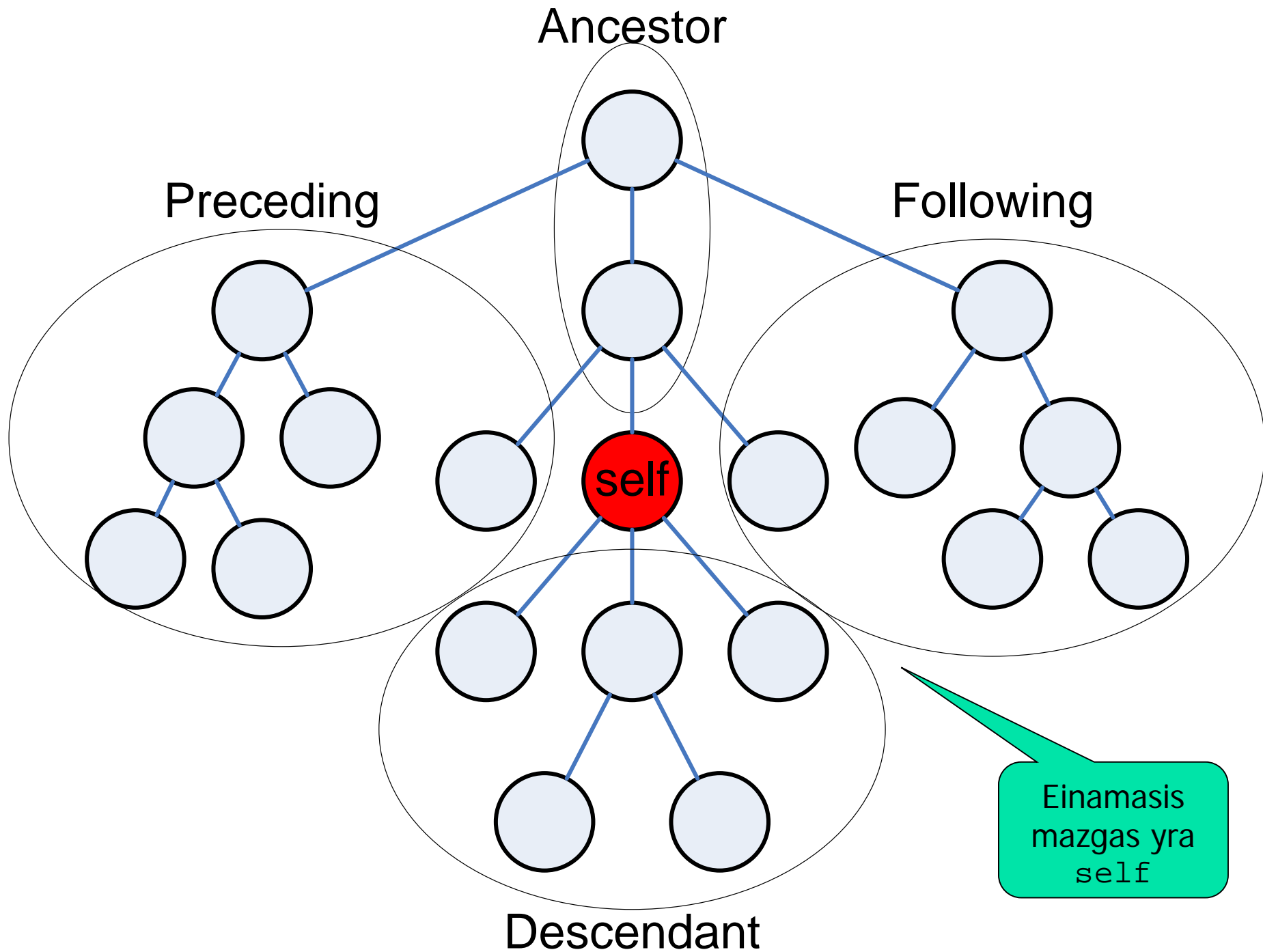


Ašys – pastabos

- Pasirinkus bet kokį XML dokumento einamąjį mazgą, ašys:

ancestor,
descendant,
following,
preceding ir
self

suskaido visų mazgų aibę į *penkis nepersikertančius poaibius*, kurie kartu apima visą dokumentą



Pavyzdys

```
<KnyguSarasas>
  <grupė rūšis="grožinė">
    <knyga isbn="12345">
      <pavadinimas>
        Raudonkepurnaitė
      </pavadinimas>
    </knyga>
    <knyga isbn="3214">
      ...
    </knyga>
  </grupė>
  <grupė rūšis="mokslinė">
    <knyga isbn="454654">
      ...
    </knyga>
    <knyga isbn="98765436">
      ...
    </knyga>
    ...
  </grupė>
  ...
</KnyguSarasas> <!-- komentaras -->
```



XPath kelių pavyzdžiai

(kiekvienas kelias grąžina mazgų aibę, galbūt tuščią)

- `child::knyga` – išrenka einamojo mazgo vaikus žymes vardu `knyga`
- `child::*` - išrenka einamojo mazgo visas vaikus žymes (bet tik žymes, o ne kitas mazgų rūšis)
- `child::text()` – išrenka einamojo mazgo *tekstinius* vaikus
- `child::node()` – išrenka einamojo mazgo *visų rūšių* vaikus mazgus
- `attribute::isbn` – išrenka einamojo mazgo (žymės) *atributą* vardu `isbn` (kaip matome, einama ne `child`, o `attribute` ašimi)
- `attribute::*` - išrenka visus einamojo mazgo atributus
- `descendant::knyga` – išrenka visas žymes "knyga", kurios yra einamojo mazgo įpėdiniai (vaikai, anūikai, proanūikai, ir t.t.)
- `ancestor::*` – išrenka visas žymes, kurios yra einamojo mazgo protėviai (tėvas, senelis, prosenelis, ir t.t.)



XPath kelių pavyzdžiai

- `ancestor-or-self::knyga` – išrenka visas žymes, kurios turi vardą `knyga` ir yra einamojo mazgo protėviai (tėvas, senelis, prosenelis, ir t.t.), įtraukiant ir einamąjį mazgą (jei jis yra žymė ir jo vardas yra "knyga")
- `self::grupė` – išrenka einamąjį mazgą, jei jis yra žymė vardu "grupė", priešingu atveju nieko neišrenka (grąžina tuščią aibę)
- `child::knyga/descendant::aprašymas` – (du žingsniai) išrenka visas žymes `aprašymas`, kurios yra žymių `knyga` įpėdiniai, kurios yra einamojo mazgo vaikai (rezultatas yra žymių `aprašymas` aibė, **žymės knyga į rezultatą neįtraukiamos!**)
- `child::* / child::knyga` – išrenka visas žymes vardu `knyga`, kurios yra einamojo mazgo anūkai
- `/` - išrenka šakninį mazgą (**ne šakninę XML dokumento žymę!**)
- `/child::*` - išrenka šakninę XML dokumento žymę
- `/descendant::knyga` – suformuoja visų dokumento žymių vardu "knyga" aibę



Kelių su predikatais pavyzdžiai

- `child::knyga[position()=1]` – išrenka pirmą einamojo mazgo vaikinę žymę knyga
- `child::knyga[position()=last()]` – išrenka paskutinę einamojo mazgo vaikinę žymę knyga
- `child::para[position()=last()-1]` – išrenka priešpaskutinę einamojo mazgo vaikinę žymę knyga
- `following-sibling::knyga[position()=1]` – išrenka einamojo mazgo pirmą žymę-brolį vardu knyga, kuri XML dokumente yra toliau nei einamasis mazgas
- `/descendant::knyga[position()=42]` – išrenka 42-ą žymę dokumente vardu knyga
- `child::knyga[attribute::isbn="0123456789"]` – išrenka visas einamojo mazgo vaikinės žymės vardu knyga, kurios turi atributą `isbn` su reikšme "0123456789"



Sutrumpinta notacija

- `child` ašis yra ašis pagal nutylėjimą (ją galima praleisti)
 - knyga atitinka `child::knyga`
 - * atitinka `child::*`
 - `knyga/autorius` atitinka `child::knyga/child::autorius`
- attribute ašies sutrumpinimas: `@`
 - `@isbn` atitinka `attribute::isbn`
- `//` atitinka `/descendant-or-self::node()`
 - `//knyga` atitinka `/descendant-or-self::node()/child::knyga`
 - `knyga//autorius` atitinka `child::knyga/descendant-or-self::node()/child::autorius`
- `.` atitinka `self::node()`
 - `./knyga` atitinka `self::node()/descendant-or-self::node()/child::knyga`
- `..` atitinka `parent::node()`
 - `../title` atitinka `parent::node()/child::title`



Sutrumpinta notacija - pastabos

- $m//n$ gražina mazgų aibę, kuri yra tokių aibių sąjunga:
 - m/n ,
 - $m/* /n$, (`child::m/child::* /child::n`)
 - $m/* /* /n$,
 - $m/* /* ...* /* /n$, ir t.t.
- $//n$ analogiškai yra sąjunga tokių aibių:
 - $/n$ (šaknies **vaikai** n)
 - $/* /n$ (`/child::* /child::n` – šaknies **anūkai** n)
 - $/* /* /n$ (šaknies **proanūkiai** n)
 - $/* /* ...* /* /n$, ir t.t. (šaknies **įpėdiniai** n)