



Interneto technologijos

XPath

Formalioji XPath pusė



XPath kelias

- Neformaliai: kelias susideda iš *žingsnių*, žingsnis susideda iš *ašies* (nusako kryptį), *mazgo testo* (filtras) ir nebūtinų *predikatų* (filtrų)

```
Kelias ::= SantykinisKelias | AbsoliutusKelias
AbsoliutusKelias ::= '/' SantykinisKelias?
SantykinisKelias ::= Žingsnis |
                    SantykinisKelias '/' Žingsnis
```

```
Žingsnis ::= AšiesVardas '::' MazgoTestas Predikatas*
```

```
MazgoTestas ::= '*' | Vardas | text() | node() | ...
```

```
Predikatas ::= '[' Išraiška ']'
```

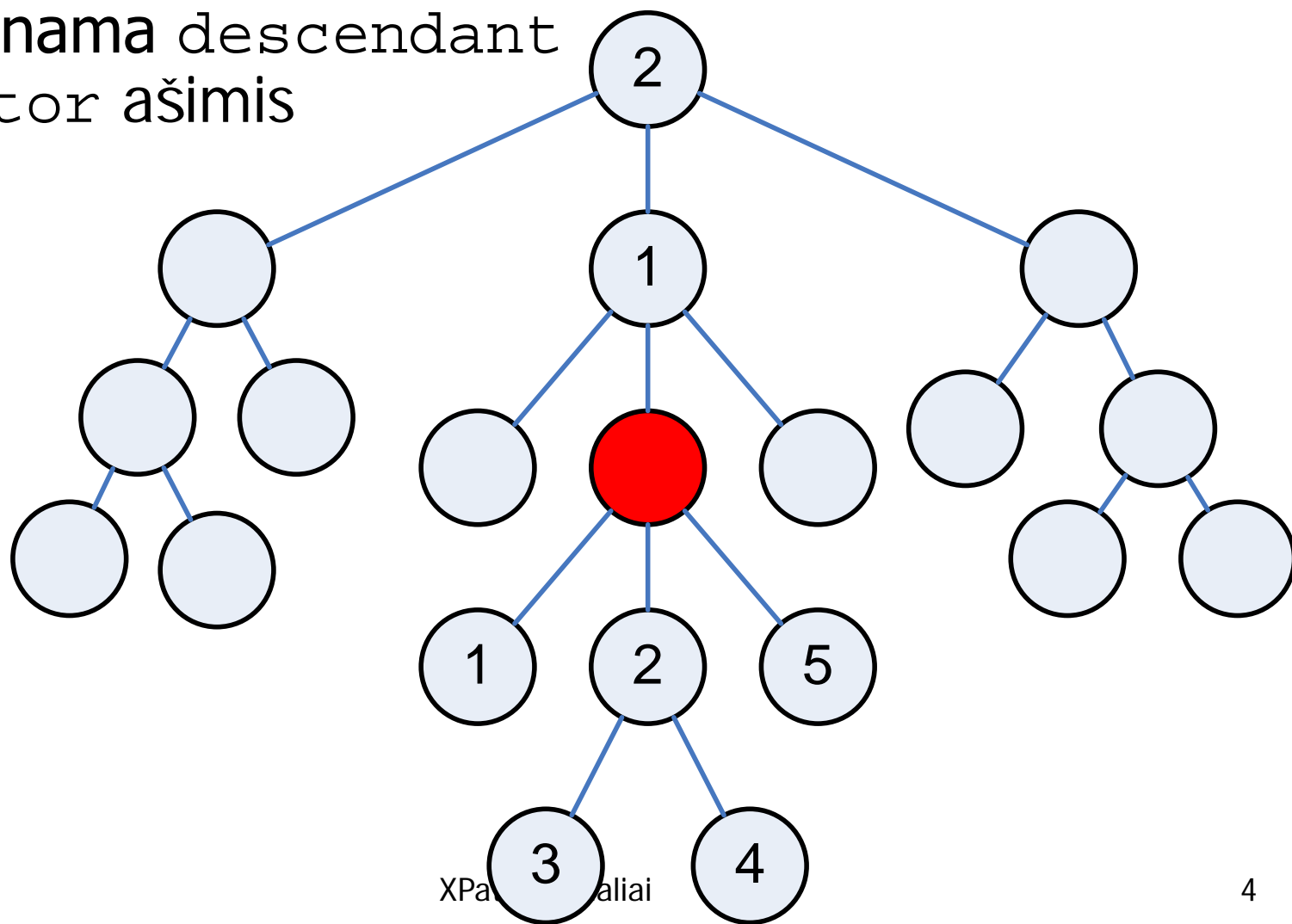


Ašių kryptys ir mazgų pozicija mazgų aibėje

- Kiekviena ašis yra arba einanti į priekį (*forward*), arba einanti atgal (*reverse*)
 - *ancestor*, *ancestor-or-self*, *preceding* ir *preceding-sibling* ašys eina atgal; visos kitos ašys eina į priekį.
 - *self* ašiai kryptis nesvarbi (niekur neiname 😊)
- Kiekvienas mazgas mazgų aibėje turi eiliškumo numerį (poziciją)
- Jei ašis yra einanti į priekį, tai didesni eiliškumo numeriai yra suteikiami mazgams, kurie yra toliau nuo dokumento pradžios.
- Jei ašis yra einanti atgal, tai didesni eiliškumo numeriai yra suteikiami mazgams, kurie yra arčiau dokumento pradžios.
- Pirmas eiliškumo numeris yra 1.

Ašių kryptys ir mazgų pozicija mazgų aibėje

- Čia yra einama descendant ir ancestor ašimis

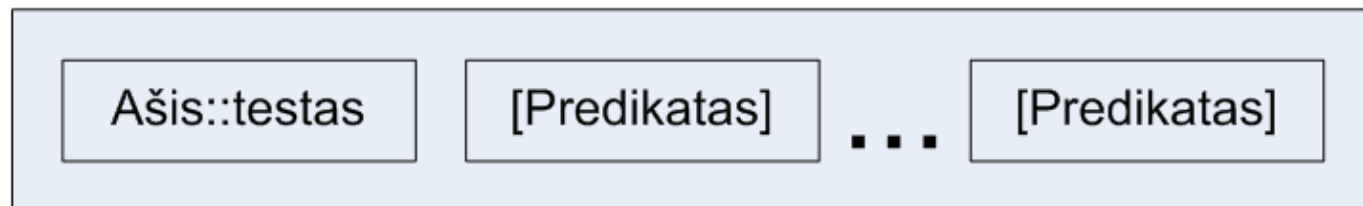


XPath kelias ir žingsniai

Kelias: Žingsnis / Žingsnis / ... / Žingsnis

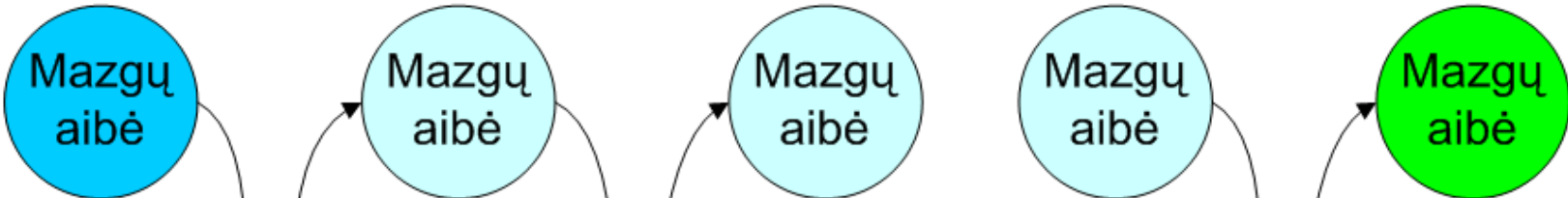


Žingsnis:



Pradinė mazgų aibė. Susideda arba iš einamojo mazgo, arba iš šakninio mazgo (jei kelias prasideda / simboliu)

Rezultatas



Kelias: Žingsnis / Žingsnis / ... / Žingsnis



Žingsnis:

Ašis::testas

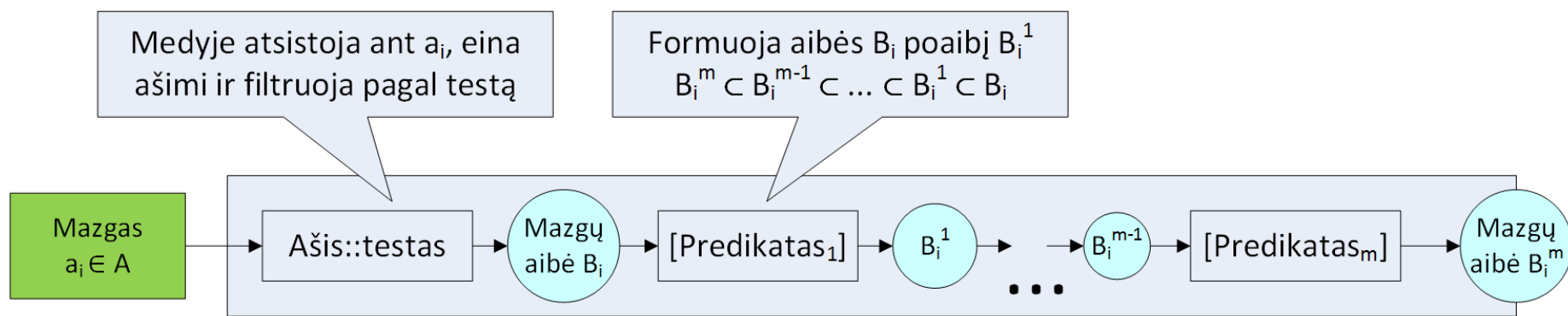
[Predikatas]

■ ■ ■

[Predikatas]

Operacijos su mazgų aibėmis žingsnyje (1)

- Žingsniui įėjime paduodama aibė A (susidedanti iš mazgų)
- Kiekvienam šitos aibės mazgui a_i :



- Kiekvienam mazgui a_i buvo gauta aibė B_i^m
- Galutinė aibė B (žingsnio rezultatas) gaunama atliekant visų šitų aibių B_i^m sąjungą (vienodi mazgai nedubliuojami):

$$B = B_1^m \cup B_2^m \cup \dots \cup B_n^m$$



Operacijos su mazgų aibėmis žingsnyje – predikatai (2)

- Predikatas, atsižvelgdamas į ašį, filtruoja mazgų aibę ir sukuria jos poaibį (kaip rezultatą)
 - Kiekvienam mazgų aibės mazgui yra skaičiuojama predikato išraiška (toliau *kontekstas* ir *aibė* sinonimai),
 - su tuo mazgu kaip *kontekstiniu mazgu* (galim gauti su `self` arba `.`),
 - su mazgų skaičiumi mazgų aibėje kaip *konteksto dydžiu* (grąžina funkcija `last()`),
 - ir su mazgo pozicija mazgų aibėje (atsižvelgiant į ašį) kaip *konteksto pozicija* (grąžina funkcija `position()`)
 - jei šiam mazgui išraiška grąžina *true*, tai jis įtraukiamas į rezultato aibę, jei *false* – neįtraukiamas.
- Jei išraiška grąžina:
 - skaičių, tai tas skaičius bus konvertuotas į *true*, jei jis lygus *konteksto pozicijai*, ir *false* priešingu atveju
 - taigi `para[3]` yra tas pats kaip ir `para[position()=3]`
 - mazgų aibę arba eilutę (string), tai jie konvertuojami į loginę reikšmę iškviečiant funkciją `boolean`.

Predikatų pavyzdžiai

- `child::knyga[attribute::isbn="012"][position()=5]`
iš visų einamojo mazgo vaikinių žymių išrenka tas, kurios turi vardą "knyga", tada iš jų išrenka tas, kurios turi atributą "isbn" su reikšme "0123456789", tada iš jų išrenka penktą žymę
- `child::knyga[position()=5][attribute::isbn="012"]`
iš visų einamojo mazgo vaikinių žymių išrenka tas, kurios turi vardą knyga, tada iš jų išrenka penktą žymę, tada, jei ji turi atributą isbn su reikšme "0123456789", grąžina šią žymę
- `child::knyga[child::pavadinimas='XML Schema']`
išrenka einamojo mazgo vaikinės žymės vardu knyga, kurios turi vaikinės žymės vardu "pavadinimas", kurių turinys (ne vardas!) yra tekstas "XML Schema"
- `child::*[self::knyga or self::žurnalas]`
išrenka visas einamojo mazgo vaikinės žymės, kurių vardas yra arba knyga, arba žurnalas



Predikatų pavyzdžiai

- `//knyga[../@rūšis="grožinė"]`
grąžins visas knygas, kurių tėvinė žymė turi atributą "rūšis" su reikšme "grožinė"
- `//knyga[
autorius=//knyga[@isbn="12345"]/autorius
]`
grąžins visas knygas, kurių autorius toksai pats kaip ir knygos su isbn kodu "12345" autorius

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>
```



"Spąstai" ☺

- Pabandykit pasakyti, ką gražina šitas XPath kelias

- `//knyga[1]`

- Kyla noras interpretuoti kaip
`/descendant::knyga[1]`

Gražina pirmą XML dokumento knygą (aibė bus arba tuščia – knygų išvis nėra, arba iš vieno mazgo – rastos knygos)

- Teisinga interpretacija (perrašius ilga notacija):
`/descendant-or-self::node()/child::knyga[1]`

Gražina visas knygas, kurios yra kokių nors mazgų pirmi vaikai (aibė gali būti iš daugiau nei vienos knygos)



"Spąstai" ☺

- Su "KnygųSąrašo" dokumentu:
 - `//knyga[1]` grąžins pirmą ir trečią knygas (skaičiuojant nuo šakninės žymės), t.y., aibę iš dviejų elementų