熱血講義

프리렉의 열혈강의 시리즈

Python 파이썬

Python

❖ 21-3

: XPath
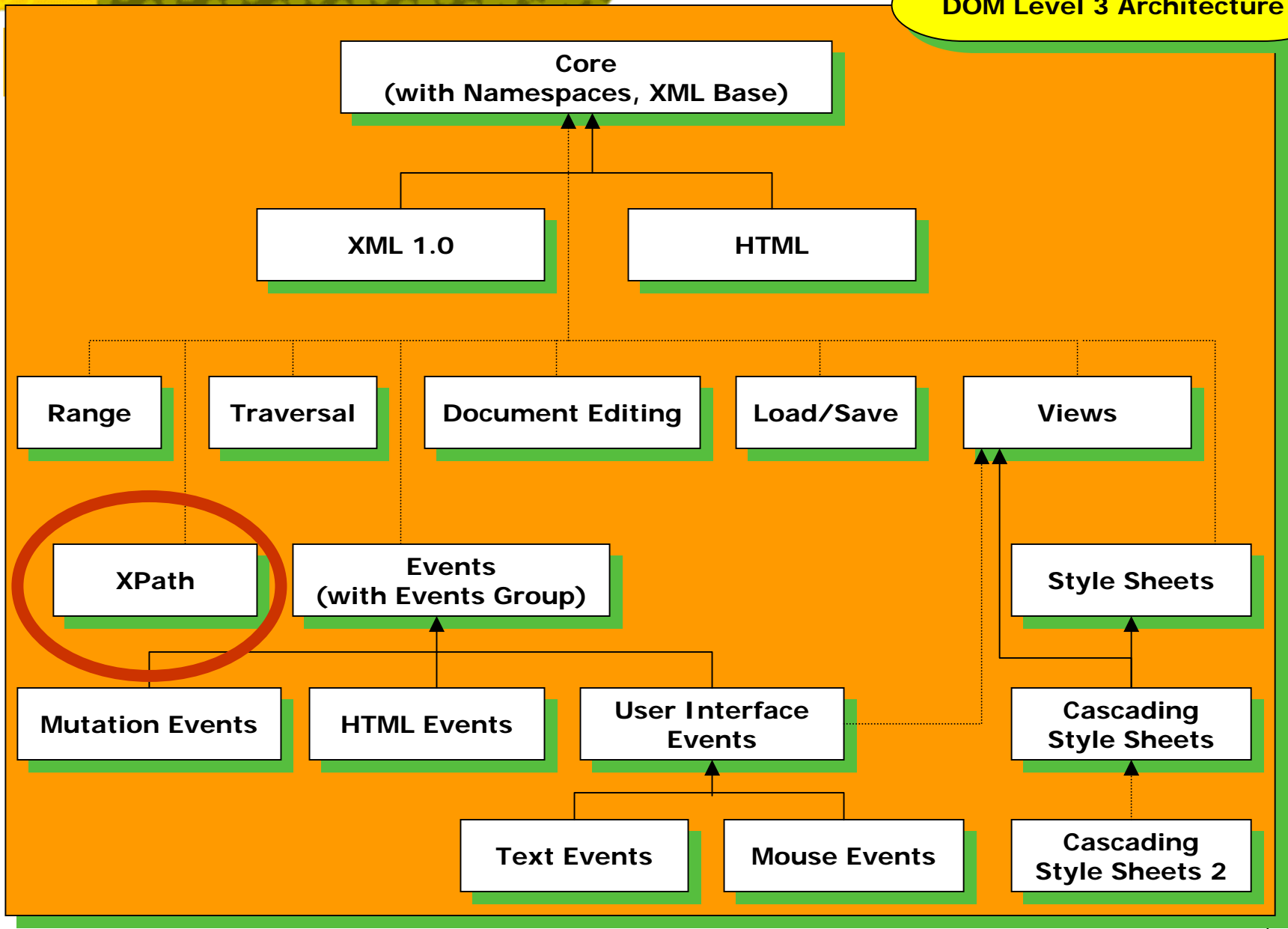
(gslee @mail.kw.ac.kr)

❖

1. ..
2. SAX
3. DOM
4. Traversal
5. XPath
6. XSLT

XPath

- **XML Path Language**
- **DOM   3**
- **http://www.w3.org/TR/xpath20/**
- **XML   /**
  - ➢   (   )
  - ➢
- **XSLT, XQuery, XPointer**
  - ➢ **http://www.w3.org/TR/xslt20/**
  - ➢ **http://www.w3.org/TR/xquery/**
  - ➢ **http://www.w3.org/TR/xptr/**
- **4XPath**
  - ➢ **http://4suite.org**

4

파이썬 (Python)

Core
(with Namespaces, XML Base)

XML 1.0   **HTML**

Range   **Traversal**   **Document Editing**   **Load/Save**   **Views**

XPath   **Events**
(with Events Group)   **Style Sheets**

Mutation Events   **HTML Events**   **User Interface Events**   **Cascading Style Sheets**

Text Events   **Mouse Events**   **Cascading Style Sheets 2**

→ **Inherits**   ┄┄► **Depends on**

- **XPath**
  - ➤ **(node set)**
  - ➤ **: unicode string**
  - ➤ **: floating point number**
  - ➤ **: true or false**

Context(   )

- **Context**
  - ➤ **XPath**
  - ➤ **XSLT, XPointer**

- **Context**
  - ➤ **Context node –**
  - ➤ **Context size –**
  - ➤ **Context position –**
  - ➤ **Variable bindings –**

  - ➤ **Defined namespaces –**

Location path(   )

- ## Location path(   )
  - ➢   XPath
  - ➢
  - ➢

  - ➢

(sample04.xml)

```xml
<?xml version="1.0" ?>
<userlist>
    <user sex="male">
        <name>gslee</name>
        <email>gslee@mail.kw.ac.kr</email>
    </user>
    <user sex="female">
        <name>spam</name>
        <email>spam@mail.kw.ac.kr</email>
    </user>
    <user sex="male">
        <name>HongGilDong</name>
        <email>home@mail.kw.ac.kr</email>
    </user>
    <user sex="female">
        <name>Sunny</name>
        <email>sunny@mail.kw.ac.kr</email>
    </user>
</userlist>
```

Path expression(   )

- '/'   '//'   (step expression)

-       '/'

- '//'

-       (step)
  - ForwardAxis :: NodeTest
  - ReverseAxis :: NodeTest
  - AbbreviatedForwardStep
  - AbbreviatedReverseStep

Axes

-

    ForwardAxis
    - child
    - descendant
    - attribute
    - self
    - descendant-or-self
    - following-sibling
    - following
    - namespace

    ReverseAxis
    - parent
    - ancestor
    - preceding-sibling
    - preceding
    - ancestor-or-self

11

Node test

- **Axes**

-

  - **'*'**

-

  - **processing-instruction()**
  - **comment()**
  - **text()**
  - **node()**

Axes::NoteTest

-
  - descendant::email
  - ancestor::*
  - preceding-sibling::processing-instruction()
  - attribute::*
  - following-sibling::*/descendant::name/child::text()
  - child::ADDRBOOK/child::ENTRY/child::PHONENUM/child::text()

Abbreviations(   )

-

child::	
attribute::	**@**
self::	**.**
parent::	**..**
/descendant-or-self::node()	**//**

following-sibling::*/descendant::PHONENUM/text()

ADDRESSBOOK/ENTRY/PHONENUM/text()

Predicates(   )

- [..]

-

-

-          ,

-

   child::user[attribute::sex="male"]/child::email[1]

   user[@sex="male"]/email[1]

Samples

- ## Sample XPath
  - ➢ **'user/name'**
  - ➢ **'user/name/text()'**
  - ➢ **'user[2]/name/text()'**
  - ➢ **'user[@sex="male"]'**
  - ➢ **'//email'**
  - ➢ **'/userlist//email'**
  - ➢ **'user/email/../name'**

-
  - **+, -, *, div, idiv, mod**

-
  - **<, >, <=, >=, =, !=**
  - **lt, gt, le, ge, eq, ne**

-
  - **is, isnot**
  - ***//book[isbn="1234"] is //book[call="q.abc 890"]***

-
  - **and, or**

-
  - **<<, >>**

position()	
last()	/
count(<node set>)	<node set>
local-name(<node set>)	
concat(<string>, .., <string>)	
contains(<string>,<string>)	` `
not(<boolean>)	` ` ` ` , ` ` ` `
sum(<node-set>)	

user[last()]/email[contains(., 'pymail')]

● PyXML 　　4Suite < 0.12

```
# xpath01.py
# 4XPath from PyXML
# or 4XPath from 4Suite < version 0.12
from xml.xpath import Evaluate
from xml.dom.ext.reader.Sax2 import Reader
from xml.dom.ext import PrettyPrint


reader = Reader()  # Reader          .
dom = reader.fromUri('sample04.xml')  # URI      .


xpath0 = 'user/name'
nodeList = Evaluate(xpath0, dom.documentElement)
for node in nodeList:
        PrettyPrint(node)
```

<email>gslee@mail.kw.ac.kr</email>

<email>spam@mail.kw.ac.kr</email>

<email>home@mail.kw.ac.kr</email>

<email>sunny@mail.kw.ac.kr</email>

4Suite >= 0.12

```
# xpath02.py
# 4XPath from 4Suite >= version 0.12
from Ft.Xml.XPath import Evaluate
from xml.dom.ext.reader.Sax2 import Reader
from xml.dom.ext import PrettyPrint

reader = Reader()  # Reader              .
dom = reader.fromUri('sample04.xml')  # URI        .

xpath0 = 'user/email'
nodeList = Evaluate(xpath0, dom.documentElement)
for node in nodeList:
        PrettyPrint(node)
```

PyXML   4Suite < 0.12

```
# xpath03.py
# 4XPath from 4Suite >= version 0.12
from xml.xpath import Compile
from xml.xpath.Context import Context
from xml.dom.ext.reader.Sax2 import Reader
from xml.dom.ext import PrettyPrint

reader = Reader()  # Reader            .
dom = reader.fromUri('sample04.xml')  # URI      .

expression = Compile('user/email')
context = Context(dom.documentElement)
nodeList = expression.evaluate(context)
for node in nodeList:
        PrettyPrint(node)
```

4Suite >= 0.12

```
# xpath04.py
# 4XPath from 4Suite >= version 0.12
from Ft.Xml.XPath import Compile
from Ft.Xml.XPath.Context import Context
from xml.dom.ext.reader.Sax2 import Reader
from xml.dom.ext import PrettyPrint

reader = Reader()  # Reader                     .
dom = reader.fromUri('sample04.xml')  # URI
        .


expression = Compile('user/email')
context = Context(dom.documentElement)
nodeList = expression.evaluate(context)
for node in nodeList:
        PrettyPrint(node)
```