Parsing XML

XPath Basics & CWE Example

Arnau Sangrà[1]

[1]Software Developer
Ackcent CyberSecurity

Data Driven Security, April 2016

Table of Contents

XPath

A syntax for defining parts of an XML document that uses *path expressions* to navigate in XML documents.

- XPath contains a library of standard functions (+100)
- XPath is a major element in XSLT Standard
- XPath is a W3C recommendation

Commands

Expression	Description
`nodename`	Selects all nodes with the name "nodename"
`/`	Selects from the root node
`//`	Selects nodes in the document from the current node that match the selection no matter where they are
`.`	Selects the current node
`..`	Selects the parent of the current node
`@`	Selects attributes

Commands II

PathExpression	Description
`/Weakness/Description[1]`	Selects the first Description element that is the child of the Weakness element.
`/Weakness/Description[last()]`	Selects the last Description element that is the child of the Weakness element
`/Weakness/Description[last()-1]`	Selects the last but one Description element that is the child of the Weakness element
`//Weakness[@Status='Incomplete']`	Selects all the Weakness elements that have a "Status" attribute, with a value of "Incomplete"
`/Weaknesses/Weakness[@year>2009]/Description`	Selects all the Description elements of the Weakness elements of the Weaknesses element that have a year attribute greater than 2009

Understanding the XML Schema

Before parsing the XML, start by analysing the XSD (XML schema) for root nodes and nested childrens.

1. Download the XML Schema
2. Analyse it!
3. Download the dictionary of CWEs.
4. Inspect the XML with a *good* text editor (+10MB of text)
5. Start parsing the data

CWE XML

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <Weakness_Catalog Catalog_Date="2015-12-07" Catalog_Name="VIEW
3    <Views xmlns:capec="http://capec.mitre.org/capec-2" />
4     (...)
5    </Views>
6    <Categories xmlns:capec="http://capec.mitre.org/capec-2" />
7     (...)
8    </Categories>
9    <Weaknesses xmlns:capec="http://capec.mitre.org/capec-2" />
10    (...)
11   </Weaknesses>
12   <Compound_Elements xmlns:capec="http://capec.mitre.org/capec-
13    (...)
14   </Compound_Elements>
15 </Weakness_Catalog>
```

Inside Weakness Node

```
<Weakness ID="777" Name="Regex . . "
Status="Incomplete" Weakness_Abstraction="Variant">
  <Description>
    <Description_Summary>The software uses a regular expressio
  </Description_Summary>
    <Extended_Description>
      <Text>(...)</Text>
    </Extended_Description>
  </Description>
  <Relationships>
    <Relationship>
      <Relationship_Views>
        <Relationship_View_ID Ordinal="Primary">1000</Relation
        <Relationship_View_ID Ordinal="Primary">699</Relations
      </Relationship_Views>
      <Relationship_Target_Form>Weakness</Relationship_Target_
      <Relationship_Nature>ChildOf</Relationship_Nature>
      <Relationship_Target_ID>625</Relationship_Target_ID>
      <!--Permissive Regular Expression-->
   </Relationship>
  </Relationships>
```

Example 1

CWE Title

Get the title of a CWE by its ID.

```
1 GetCWETitle <- function(doc, cwe = "100") {
2   xpath <- paste("//Weakness[@ID = '",
3                  cwe,
4                  "']/@Name",
5                  sep = "")
6   return(unlist(XML::xpathApply(doc, xpath))[["Name"]])
7 }
```

Example 2

CWE Childrens

Get the CWE associated with a CWE ID.

```
1 GetCWEChildrenNodes <- function(doc, cwe = "100") {
2 xpath <- paste("//Weakness[Relationships/Relationship/Relationship
3                     cwe,
4                     "' and Relationships/Relationship/Relationship_Natu
5                     " and Relationships/Relationship/Relationship_Targe
6                     sep = "")
7 return(XML::xpathApply(doc, xpath))
8 }
```

Example 3

IDs of Children CWE

Get the IDS of all CWE associated with a CWE ID.

```
1  GetCWEChildrenIDs <- function(doc, cwe = "100") {
2    xpath <- paste("//Weakness[Relationships/Relationship/Relationsh
3                   cwe,
4                   "' and ",
5                   " Relationships/Relationship/Relationship_Nature =
6                   " and ",
7                   "Relationships/Relationship/Relationship_Target_Fo
8                   sep = "")
9    return(as.character(XML::xpathApply(doc, xpath)))
10 }
```

Example 4

IDs of Children CWE

Get the IDs of the CWEs associated with CWE ID.

```
1 GetAllCWEChildrenIDs <- function(doc, cwe = "100") {
2   childs <- GetCWEChildrenIDs(doc, cwe)
3   if (identical(childs, character(0))) {
4     return(cwe)
5   } else {
6     return(unique(c(cwe, unlist(lapply(childs, function(x) GetAllC
7   }
8 }
```

Create new parsers

Add new parsers to *net-security* package.

Ideas

- CVE
- CWE
- CPE
- ...