**Модулі обробки XML**

Загальна інформація

Переважно XML використовують для роботи з аутсорс-датою. Замість інтегрування даних в HTML їх записують в окремий XML файл. XML записує дату у форматі “простий текст”, тому сховище не є залежне від платформи яка його використовує, а дані можуть бути легко імпортовані/експортовані.

XML в Python

Інтерфейси Python для обробки XML об’єднані в один XML пакет.

ВАЖЛИВО!!!  
XML модулі не є абсолютно захищені проти неправильно збудованих даних або даних збудованих щоб нести шкоду користувачу(деталі нижче)!

Є кілька базових модулів для роботи з XML, а саме:

1. [xml.etree.ElementTree](https://docs.python.org/3/library/xml.etree.elementtree.html#module-xml.etree.ElementTree): ElementTree API, простий і не громісткий XML процесор
2. [xml.dom](https://docs.python.org/3/library/xml.dom.html#module-xml.dom): - означення DOM AP
3. [xml.dom.minidom](https://docs.python.org/3/library/xml.dom.minidom.html#module-xml.dom.minidom): мінімальна імплементація DOM
4. [xml.dom.pulldom](https://docs.python.org/3/library/xml.dom.pulldom.html#module-xml.dom.pulldom): модуль підтримки для будування часткових DOM дерев
5. [xml.sax](https://docs.python.org/3/library/xml.sax.html#module-xml.sax): базові SAX2 класи і функції
6. [xml.parsers.expat](https://docs.python.org/3/library/pyexpat.html#module-xml.parsers.expat): модуль для “парсення”(зчитування) даних

Вразливість XML

XML модулі не є абсолютно захищені проти даних збудованих щоб нести шкоду користувачу, тому зловмисники можуть використати деякі додаткові можливості(фічі) щоб провести сервісні атаки, отримати доступ до локальних файлів та згенерувати мережеві з’єднання до інших пристроїв.

Ось табличка вразливих модулів:

| **kind** | **sax** | **etree** | **minidom** | **pulldom** | **xmlrpc** |
| --- | --- | --- | --- | --- | --- |
| billion laughs | **Vulnerable** | **Vulnerable** | **Vulnerable** | **Vulnerable** | **Vulnerable** |
| quadratic blowup | **Vulnerable** | **Vulnerable** | **Vulnerable** | **Vulnerable** | **Vulnerable** |
| external entity expansion | **Vulnerable** | Safe (1) | Safe (2) | **Vulnerable** | Safe (3) |
| [DTD](https://en.wikipedia.org/wiki/Document_type_definition) retrieval | **Vulnerable** | Safe | Safe | **Vulnerable** | Safe |
| decompression bomb | Safe | Safe | Safe | Safe | **Vulnerable** |

Приклад роботи з XML

Код:

from xml.dom import minidom

xmldoc = minidom.parse('items.xml')

itemlist = xmldoc.getElementsByTagName('item')

print(len(itemlist))

print(itemlist[0].attributes['name'].value)

for s in itemlist:

print(s.attributes['name'].value)

Вивід:

4

item1

item1

item2

item3

item4