LEARNING

lxml

#lxml

Table of Contents

About

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: lxml

It is an unofficial and free lxml ebook created for educational purposes. All the content is extracted from Stack Overflow Documentation, which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official lxml.

The content is released under Creative Commons BY-SA, and the list of contributors to each chapter are provided in the credits section at the end of this book. Images may be copyright of their respective owners unless otherwise specified. All trademarks and registered trademarks are the property of their respective company owners.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

Chapter 1: Getting started with lxml

Remarks

This section provides an overview of what lxml is, and why a developer might want to use it.

It should also mention any large subjects within lxml, and link out to the related topics. Since the Documentation for lxml is new, you may need to create initial versions of those related topics.

Examples

Installation or Setup

Detailed instructions on getting lxml set up or installed.

Why we need lxml and how to use it ?

First, why do we need lxml ?

lxml.etree is a generic API for XML and HTML handling. It aims for ElementTree compatibility and supports the entire XML infoset. It is well suited for both mixed content and data centric XML. Its generality makes it the best choice for most applications.

The lxml library, is an extension of the old libxml2 and libxsit and it has some major benefits:

1. Very easy python API
2. Well documented
3. No need to deal with memory management
4. No need to worry for segmentation fault

It is also provides a very natural way to deal with any XML data format. Data is automatically converted to Python data types and can be manipulated with normal Python operators

Great! now how can I use it ?

On Linux machines you can install the lxml library using apt-get:

```
sudo apt-get install python-lxml
```

To import and use the library:

```
from lxml import etree
```

To parse the xml file, you can use:

```
    try:
```

```
        parser = ET.XMLParser(remove_comments=False, remove_blank_text=True)
        tree = ET.parse(file, parser=parser)
    except (Exception):
        print ('Failed to open file %s' % file, exc_info=True)
    return tree
```

lxml install

Installing lxml is very easy, had become an easy jobs since Python 2.7.9 (because it comes with an utility which helps developers to download install dependency in an easy manner like Maven for Java) at first you have to run the command then start coding.

```
pip install lxml
```

The second way is to install using easy_install. More details instruction might be found here

Credits

S. No	Chapters	Contributors
1	Getting started with lxml	Community, Iurie Cojocari, omertalmi