Training: Python Academy
High Performance XML with Python

TRAINING GOALS:

Since its early beginnings in 1998, the eXtensible Markup Language, XML, has grown into a standard markup language family for portable data formats. The major document formats, such as the Open Document Format (ODF) known from OpenOffice, or Microsoft's so-called OpenXML format, are based on XML, just like many application level networking protocols such as XML-RPC, SOAP or Jabber/XMPP. Many interfaces of business applications use either standardized, proprietary or ad-hoc XML formats, and their configuration files are often written in XML, too. And clearly, XML has left its fingerprint on the web through RSS/Atom feeds, Ajax interfaces and configurable browser GUIs (XHTML/XUL).

The support of XML in programming languages has constantly improved over the last decade. Today, developers can grab very efficient tools from their tool box that substantially simplify XML handling. Not surprisingly, the Python programming language has some very powerful tools available that often even beat their main contenders from the Java world in terms of performance, and easily in terms of usability.

The objective of this course if to get an understanding of important XML technologies, and to learn how to use the available tools by example.

CONSPECT:

- Introduction to XML
  - XML and the XML Infoset.
  - XML Namespaces.
  - Dealing with XML formats.
- Fast XML processing
  - Parsing and serializing XML files.
  - Extracting information from XML documents (tree navigation, XPath, CSS selectors).
  - Processing and transforming XML documents in main memory.
  - Generating XML documents.
  - Stream processing of large XML files that do not fit into main memory.
- Advanced topics
  - Creating proprietary XML formats.
  - Validating XML formats with schema languages (e.g. RelaxNG, Schematron).
  - Binding XML documents to Python objects (lxml.objectify).

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- Creating application specific XML APIs with lxml.
- Introduction to stylesheet transformations (XSLT processing).
- Note that the advanced topics are subject to time constraints. A choice will be made based on the interest of the participants.

Initially, the course will build up a common understanding of XML (specifically the XML Infoset) and some of its applications. The main theme then deals with efficient processing of XML (and a bit of HTML) in Python.

The presented tool set includes the ElementTree library that comes with Python since version 2.5, and the freely available lxml library that combines a compatible Python API with a large set of additional XML features.

REQUIREMENTS:

- Knowledge of **Python** language.

Difficulty level



CERTIFICATE:

The participants will obtain certificates signed by Python Academy.

TRAINER:

Authorized Python Academy Trainer.

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007