Хранилища данных

№ урока: 2 **Курс:** Python Advanced

Средства обучения: Python3.6, PyCharm

Обзор, цель и назначение урока

Изучить следующие форматы данных: CSV, XML, JSON. Дать базовые знания данных форматов и изучить стандартную библиотеку языка Python. Данные форматы используются для хранения и обмена данными между компонентами сети или программами. Изучить особенности данных форматов и провести сравнение для того, чтобы студент смог выбрать определенный формат для решения конкретных задач.
Изучить основы работы с библиотекой sqlite3 и использования данной СУБД в качестве хранилища данных. Рассмотреть особенности данной библиотеки с практическим уклоном.

Изучив материал данного занятия, учащийся сможет:

- Иметь полное понимание данных форматов
- Создавать и обрабатывать форматы данных CSV, XML, JSON
- Использовать стандартные библиотеки языка Python для работы с данными форматами
- Осуществлять поиск данных в формате XML используя язык XPATH
- Позволит использовать SQLite в своих задачах.
- Позволит создавать свои собственные типы данных и использовать их в хранилище.
- Создавать пользовательские агрегатные и обычные функции, расширяя стандартные возможности SQL.
- Безопасно работать с хранилищем, изучит такое понятие как SQL-инъекции и как обезопасить свое приложение.

Содержание урока

1. Определение формата CSV и его особенностей
2. Формат CSV средствами Python.
3. Определение формата XML и его особенностей.
4. Описание структуры DOM и тегов.
5. Формат XML в Python.
6. Библиотека lxml.
7. Определение формата JSON и его особенностей.
8. Формат JSON в Python.
9. Основные понятия и особенности СУДБ SQLite.
10. Библиотека sqlite3 в Python.
11. Создание собственных типов, функций и агрегаций.
12. SQL-инъекции.

Резюме

- CSV является удобным форматом для обмена данных в табличном виде и не подходит в случае, если необходимо выгрузить неструктурированные данные. Основными составляющими CSV формата являются строка заголовка, строки с данными и разделители столбцов (по умолчанию запятая). Существует возможность сконфигурировать символы разделения столбцов и добавить экранирование значений спец. символами, например,

CyberBionic systematics
CyberBionic Systematics ® 2019
19 Eugene Sverstyuk Str., 5 floor
Kyiv, Ukraine
Tel. 0 800 750 312
E-mail: edu@cbsystematics.com
Site: www.edu.cbsystematics.com | itvdn.com

двойными кавычками. Стандартная библиотека языка содержит модуль `csv` для работы с данным форматом.

- XML – это текстовый формат. По своей сути э то DOM документ с определенной структурой. Данные содержатся в тегах, и теги имеют собственные имена. Данный формат очень похож на HTML- они оба содержат теги, но XML предназначен исключительно для обмена и хранения данных. В стандартной библиотеке языку имеется модуль `xml` для работы с данным форматом. Особенностью данного формата является возможность поиска по документу, используя специальный язык поиска XPATH.
- JSON форма также является текстовым форматом. JSON напрямую связан с языком JavaScript и представляет определение структур объектов. Формат поддерживает небольшое подмножество типов данных, которое упрощает проверку данных на стороне языка программирования. В качестве составных типов данных JSON поддерживает объекты и списки. Стандартная библиотека языка Python содержит модуль `json`, который предоставляет удобные интерфейсы для работы с данным форматом.
- СУБД SQLite представляет из себя хранилище, которое состоит из одного файла. Данная СУБД не имеет рабочих процессов, запущенных и ожидающих соединений, как например MySQL или PostgreSQL.
- SQLite поддерживает типы данных для кортежей значений, но количество типов ограничено, а также существует такое понятие, как аффинированность типов. Аффинированность используется для совместимости SQLite с различными SQL СУБД, используя следующие типы данных – Integer, Text, None, Real, Numeric.
- С стандартной библиотеки языка Python имеется модуль sqlite3 который предназначен для работы с данной СУБД. Он позволяет исполнять SQL-запросы, которые описываются в виде обычных строк, имеющих тип str. Также есть возможность создавать пользовательские агрегатные и обычные функции, а также собственные типы, регистрируя их через специальный интерфейс библиотеки sqlite3.
- В мире SQL СУБД существует такой вид угрозы или атаки, называемый SQL-инъекцией. Данный вид угрозы позволяет встраивать вредоносный SQL-код в запросы, в случае если мы не используем экранирование символом при вставке переменных в SQL запросы учитывая типы их значений. Библиотека sqlite3 предоставляет возможности для предотвращения таких атак.

Закрепление материала

- Опишите основные особенности формата CSV?
- В каких случаях стоит использовать данный?
- С помощью какого инструмента в модуле csv языка Python можно построчно читать данных из файла напрямую в структуру `dict`?
- Какие особенности присущи для формата XML?
- Как называется язык поиска для формата XML?
- Какое ограничение существует для корневых элементов XML документа?
- Что такое DOM?
- Что означает аббревиатура JSON?
- На какой тип данных языка Python похож формат JSON?
- Назовите две функции модуля `json`, позволяющие читать и записывать JSON в файл?
- Если необходимо передавать структуры с вложенными объектами, какой формат CSV/XML/JSON вы бы рекомендовали использовать и почему?
- Что такое СУБД SQLite и что из себя представляет данное хранилище?
- Какие аффинные типы оно поддерживает и зачем нужен данный механизм-аффинирование типов?
- Как добавить собственный тип данных и что нужно создать/зарегистрировать, используя модуль sqlite3 в Python, чтобы данные типы воспринимались в процессе выборки и вставки данных в SQLite хранилище?

CyberBionic Systematics ® 2019
19 Eugene Sverstyuk Str., 5 floor
Kyiv, Ukraine

Tel. 0 800 750 312
E-mail: edu@cbsystematics.com
Site: www.edu.cbsystematics.com | itvdn.com

- Существует ли способ создать SQLite хранилище в памяти?
- Для чего используется константа PARSE_DECLTYPES?

Дополнительное задание

Задание 1
Создайте функцию, которая будет создавать CSV файл на основе данных, введенных пользователем через консоль. Файл должен содержать следующие колонки: имена, фамилии, даты рождений и город проживания. Реализовать возможности перезаписи данного файла, добавления новых строк в существующий файл, построчного чтения из файла и конвертацию всего содержимого в форматы XML и JSON.

Задание 2
Создайте таблицу «материалы» из следующих полей: идентификатор, вес, высота и доп. характеристики материала. Поле доп. характеристики материала должно хранить в себе массив, каждый элемент которого является кортежем из двух значений, первое – название характеристики, а второе – её значение.

Самостоятельная деятельность учащегося

Задание 1
Создайте простые словари и сконвертируйте их в JSON. Сохраните JSON в файл и попробуйте загрузить данные из файла.

Задание 2
Создайте XML файл с вложенными элементами и воспользуйтесь языком поиска XPATH. Попробуйте осуществить поиск содержимого по созданному документу XML, усложняя свои запросы и добавляя новые элементы, если потребуется.

Задание 3
Поработайте с созданием собственных диалектов, произвольно выбирая правила для CSV файлов. Зарегистрируйте созданные диалекты и поработайте, используя их, с созданием/чтением файлом.

Задание 4
Для таблицы «материала» из дополнительного задания создайте пользовательскую агрегатную функцию, которая считает среднее значение весов всех материалов результирующей выборки и округляет данной значение до целого.

Задание 5
Для таблицы «материала» из дополнительного задания создайте пользовательскую функцию, которая принимает неограниченное количество полей и возвращает их конкатенацию.

Рекомендуемые ресурсы

Официальный сайт Python (3.6) - CSV
https://docs.python.org/3.6/library/csv.html
Официальный сайт Python (3.6) - XML
https://docs.python.org/3.6/library/xml.etree.elementtree.html
Официальный сайт Python (3.6) - JSON
https://docs.python.org/3.6/library/json.html
Официальный сайт Python (3.6) - SQLite
https://docs.python.org/3.6/library/sqlite3.html

CyberBionic Systematics ® 2019
19 Eugene Sverstyuk Str., 5 floor
Kyiv, Ukraine

Tel. 0 800 750 312
E-mail: edu@cbsystematics.com
Site: www.edu.cbsystematics.com | itvdn.com