**SOLUTIONS**

**1.**

Make sure that pages that don't exist return an HTTP status code 404. Disallow URLs that search engines shouldn't crawl. Add the nofollow attribute to links that search engines shouldn't crawl. Avoid the dynamic inserting of content.

**2.**

Performance benchmarking:

It’s no good just guessing or assuming where the inefficiencies lie in your code.

Django tools:

django-debug-toolbar is a very handy tool that provides insights into what your code is doing and how much time it spends doing it. In particular it can show you all the SQL queries your page is generating, and how long each one has taken.

Third-party panels are also available for the toolbar, that can (for example) report on cache performance and template rendering times.

Third-party services

There are a number of free services that will analyze and report on the performance of your site’s pages from the perspective of a remote HTTP client, in effect simulating the experience of an actual user.

Get things right from the start

Some work in optimization involves tackling performance shortcomings, but some of the work can be built-in to what you’d do anyway, as part of the good practices you should adopt even before you start thinking about improving performance.

Work at the appropriate level

Django offers many different ways of approaching things, but just because it’s possible to do something in a certain way doesn’t mean that it’s the most appropriate way to do it. For example, you might find that you could calculate the same thing - the number of items in a collection, perhaps - in a QuerySet, in Python, or in a template.

**3**.

The GET and POST are two different types of HTTP requests. GET is used for viewing something, without changing it, while POST is used for changing something. For example, a search page should use GET to get data while a form that changes your password should use POST. Essentially GET is used to retrieve remote data, and POST is used to insert/update remote data.

**4.**

When working with Python applications, it's always a good idea to sandbox your development with a virtual environment. It helps prevent version collisions between libraries you need in your application and libraries you might already have installed on your machine. It also makes it easy to install dependencies within a virtual env using the requirements.txt file. Lastly, it makes sharing your development environment with other developers a snap.

Envato Tuts+ has two excellent videos on how to install virtualenv and virtualenvwrapper. Take a few minutes to walk through those videos to get virtualenv and virtualenvwrapper installed on your machine. If you've already got them installed, then skip the next section.

**5.**

Here are some tips to help you decide whether to use multithreading or multiprocessing:

Use multithreading to make user interaction (UI) programs responsive. These programs have to wait for the user to interact with them, so using threads provides enough computing power.

Use multithreading to create I/O-bound or network bound applications. Threads can provide you with all the computing power you need to access web servers and download content from the internet

Use multiprocessing to create computation-intensive programs. Multiprocessing can help you analyze large volumes of data quickly.

Use multiprocessing to develop programs that are CPU intensive. Multiprocessing can help you speed up processes and provide reliable solutions for programs that involve several CPU tasks.

**6.**

Yes, multithreading scraping increases the process speed significantly. The web crawler will utilize multiple threads.It will be able to crawl all the particular web pages of a website.It will be able to report back any 2XX and 4XX links.It will take in the domain name from the command line.It will avoid the cyclic traversal of links.

**7.**

HTTPS is HTTP with encryption. The only difference between the two protocols is that HTTPS uses TLS (SSL) to encrypt normal HTTP requests and responses. As a result, HTTPS is far more secure than HTTP. A website that uses HTTP has http:// in its URL, while a website that uses HTTPS has https://.

**Special part of the Scrapy framework**

**1.**

Scratch is a block-based visual programming language that can be used to make interactive games and animations. It has a huge online community where you can share, discuss and view scratch programs. The objective behind the development of Scratch is to help children(mostly) to think creatively, work collaboratively and solve problems systematically.

**2.**

Scratch is a high-level block-based visual programming language and website aimed primarily at children as an educational tool for programming, with a target audience of ages 8 to 16.[6] Users on the site, called Scratchers, can create projects on the website using a block-like interface.

The advantages of using scratch are listed below:

\*Scratch allows for young people to integrate creativity in storytelling, games, and animation. Kids can collaborate on projects through the use of Scratch, and share their projects online.

\*Scratch allows students to develop 21st century skills through the use of technology.

\*Scratch can be used by people of all ages including students from elementary- high school ages, and adults in a variety of settings.

\*Scratch is used in over 150 counties and available in over 40 languages. This is great for teachers that are working with bilingual or ESL students.

\*Scratch can be used across curricula and students and teachers can create and share resources via scratch.

**3.**

Scratch framework includes four modules which are designed to process the semi-structured data, users profile data, projects data and programming knowledge points, respectively. For webpages, we design a template-based wrapper method to extract triples from the semi-structured data. As for users profile data, we improve DeepDive, which is a useful tool to extract information but with the problem of wrong labeling, to extract knowledge triples by the proposed Secondary Labeling Algorithm. For projects data, we propose an advanced keywords extraction method (S-TextRank) to extract keywords triples.