

# Web Crawler

---

## GROUP MEMBERS

Ahmed Jamil - 153169

Fahad Ali Khan - 153177

## CLASS

BSCS 5<sup>th</sup>

## SUBJECT

Software Engineering

## SUBMITTED TO

Sir, Ahmad Mohsin

## **INTRODUCTION:**

This project aims to provide a robust system that can be used to monitor a specific portion of a web page and keep you updated about its changes and further information. With this help one can target many website and see only what's important. This application is for those who want to keep themselves updated about new day to day information but can easily get lost in dozens of websites to keep track of.

## **PROBLEM STATEMENT:**

The main problem while surfing the internet is that we get lost in the wide sea of website. Only if we could have a system that can keep track of all those websites in one place. This project aims to provide just that.

## **OBJECTIVE AND SCOPE:**

Core objective is to provide a general purpose website monitor system that one can use to track many websites and keep themselves update without the hassle to skim through the tiresome phase of searching the website. This application can be used by people such as bloggers or business tech insiders.

## **MOTIVATION:**

The motivation to do this project came when we have to manually monitor a website called 'Crunchyroll' for guest passes posted in comments (by premium users) which are codes that are used to unlock premium content. This was a tiresome and time wasting process. This application can help solve such issues though this one being trivial and unimportant but can be used for other important news and content as well.

## **RESOURCE REQUIRED:**

Hardware Resources:

1. Cloud Server
2. Standard PC or Laptop
3. Internet Access

Software Resources:

1. Python Interpreter
2. Microsoft Visual Studio with Python support
3. Azure or Firebase subscription

## **REFERENCES:**

1. <http://www.divms.uiowa.edu/~psriniva/Papers/crawlingFinal.pdf>
2. Christopher Olston and Marc Najork "Web Crawling" in Foundations and Trends in Information Retrieval Vol. 4, No. 3 (2010) 175–246 c 2010 C. Olston and M. Najork ([http://infolab.stanford.edu/~olston/publications/crawling\\_survey.pdf](http://infolab.stanford.edu/~olston/publications/crawling_survey.pdf))