



# Web Grabber in Cloud Computing: An Effective Approach

Sonia Malik<sup>\*1</sup>, Ms. Sapna Aggarwal<sup>\*2</sup>

<sup>1</sup>M.Tech, Jind Institute of Engineering & Technology, KUK, Haryana (India) [soniyamalik90@gmail.com](mailto:soniyamalik90@gmail.com)

<sup>2</sup>Assistant Professor, Jind Institute of Engineering & Technology, KUK, Haryana (India)  
[er.sapna.aggarwal@gmail.com](mailto:er.sapna.aggarwal@gmail.com)

## Abstract:

*The proposed project is a high-speed, multi-threading website download and viewing program. By making multiple simultaneous server requests, the program can quickly download entire website or part of a site including HTML, graphics, Java Applets, sound and other user definable files, and saves all the files in the hard drive in their native format.*

*After downloading, all links within the website are reconstructed creating a complete hard drive copy of the site that you can view at your own pace without being connected to the Internet. Additionally project should duplicate the original directory structure of a site making it easy to download and transfer a site to another server. In the event a person loses his Internet connection while downloading a site, the Resume Session allows him to pick up a session. If he wishes to update a previously downloaded site, the Update Session feature allows him to revisit a site using new search parameters to make sure he has the most current files. The program will have an embedded quick-view browser window in the main program screen. The user can launch his default browser for viewing any downloaded projects.*

**General Terms:** *Grabber, Search engine in Cloud, Security in Search engine.*

**Keywords:** *Search in cloud, secured search engines, Inter cloud communication in cloud search engines.*

## 1. Introduction:-

The main purpose for preparing this document is to give a general insight into the analysis and requirements of the existing system or situation and for determining the operating characteristics of the system. The main objective of the System is that down load the page or entire site easily with the images and other files as pages for offline viewing.

### Features:

Directory tree on the local machine is created automatically to match the downloaded site.  
Individual downloads can be stopped without canceling the entire download sequence.

Site configuration can be programmed and stored, accessible later from a menu.

The Web browser can be accessed from the Web Grabber application.

ISSN 2454-308X



## Scope

This Document plays a vital role in the development life cycle (SDLC) .As it describes the complete requirement of the system. It is meant for use by the developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

Web grabber or Web Crawler allows users to create a project for particular site with a similar structure of the pages (Such as online stores, shopping sites, e-commerce sites etc) and extract data from it to be stored in the database of the current project. The extracted data can be exported to a variety of formats including Microsoft Excel (CSV), Access, TXT, HTML, and XML, SQL script or MySQL script. This variety of export formats allows you to process and analyze data in your customary format.

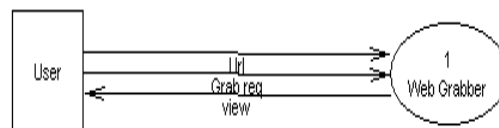
## 2. Data Flow Diagram:-

### Context Level DFD:

In the Context Level the whole system is shown as a single process.

- No data stores are shown.
- Inputs to the overall system are shown together with data sources (as External entities).
- Outputs from the overall system are shown together with their destinations (as External entities).

[[1,1]

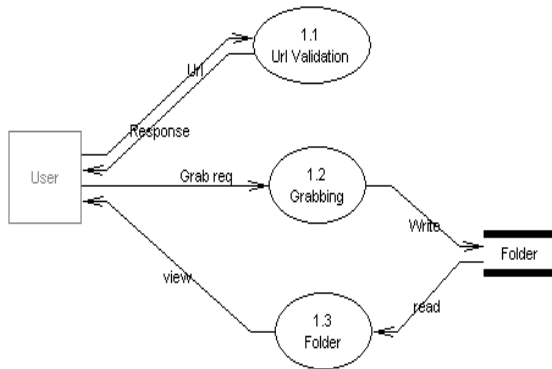


### Top Level DFD:



The Top Level DFD gives the overview of the whole system identifying the major system processes and data flow. This level focuses on the single process that is drawn in the context diagram by 'Zooming in' on its contents and illustrates what it does in more detail.

[1,1]



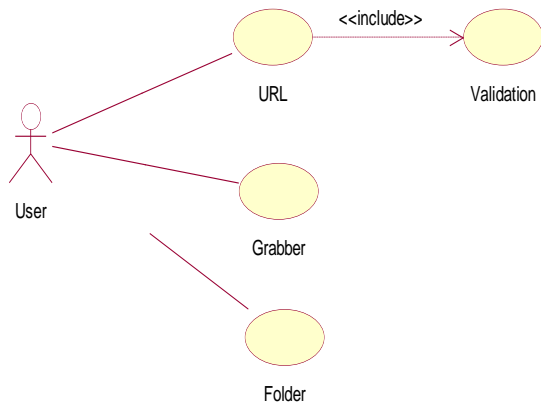
### Use Case Diagram:

- A use case diagram is a diagram that shows a set of use cases and actors and relationships.

#### Contents

- Use case commonly contain
  - Use cases
  - Actors

Dependency, generalization and association relationships



### 3. TECHNOLOGIES USED:-

#### JAVA AWT

Java programming language class library provides a user interface toolkit called the Abstract Windowing Toolkit, or the AWT. The AWT is both powerful and flexible.

Because the Java programming language is platform-independent, the AWT must also be platform-independent. The AWT was designed to provide a common set of tools for graphical user interface design that work on a variety of platforms. The user interface elements provided by the AWT are implemented using each platform's native GUI toolkit, thereby preserving the look and feel of each platform.

#### Java.awt.event.

An action or occurrence, often generated by the user, to which the program might respond—for example, key presses, button clicks, or mouse movements. Events do not adhere to this traditional approach because they occur outside of program control. When an event happens, the application is notified, causing the execution of a piece of code.

In Java, events are generated by objects. An event is represented by a *java.util.EventObject* subclass that carries event information. There are subclasses for each kind of event.



#### Java Applets

An applet is a special kind of Java program that a browser enabled with Java technology can download from the internet and run. An applet is typically embedded inside a web-page and runs in the context of the browser. An applet must be a subclass of the *java.applet.Applet* class, which provides the standard interface between the applet and the browser environment.

#### 4. Benefits:-

- Web grabber will help business men extract and collect the market figures, product pricing data, or real estate data.
- Web grabber will help book lovers extract the information about books, including their titles, authors, descriptions, ISBNs, images and prices from online book sellers.
- Web grabber will help the journalists extract the important news and articles from news sites.]

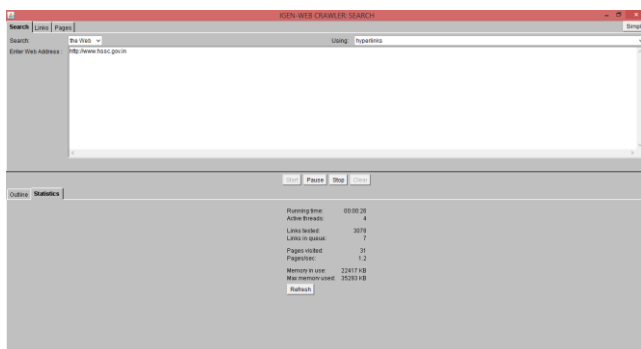
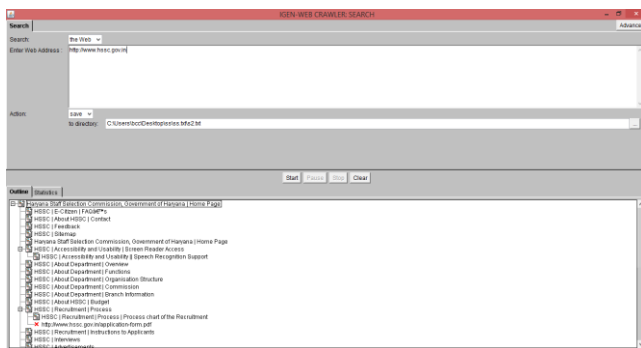
- Web grabber is useful for the all types of people for extracting the data with images, charts and other things.
- Web grabber is useful for the people seeking a job extract, job postings from online job websites, he can find a job faster and minimum inconveniences.

## 5. Related Work:-

### Data Security and Storage

- Data remanence
- Data security mitigation
- Provider data and its security

## 6. Results:-



## 7. Conclusion:-

In this paper, you can know what web grabber is, how it works, and what are the advantages available in this system. This system is very useful for all types of people for extracting the data, files, images etc on any format from any website. You can grab single page or total site using this software. In Web grabber you can download with a code of that page. This is advantageous while designing a project, when he wants to go through the similar features project. Web grabber software installation is easy on any type of system and it will take small amount of time to install in your system.

Only you have to specify the URL, and this smart utility explores the entire web site and downloads all the pictures and files automatically, that meet your specifications to a specified folder. This will save lots of amount when you want to download large amount of data in the web site.

## References:-

1. Akassh A. Mishra, Chinmay Kamat (April 2011),”Migration of search engine process into the cloud”
2. Armbrust.M.Fox. A, Griffith. R, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, (Feb 2009) “Above the clouds: A berkeley view of cloud computing,” University of California, Berkeley.
3. Boneh .D. G. D. Crescenzo, R. Ostrovsky, and G. Persiano, (2004), “Public key encryption with keyword search,” in Proc. of EUROCRYPT’04, volume 3027 of LNCS.
4. Y.-C. Chang and M. Mitzenmacher, (ACNS 2005), “Privacy preserving keyword searches on remote encrypted data”.
5. Chien Chin Chen and Meng Chang Chen, (SIGIR 2012) “TSCAN: A Content Anatomy Approach to Temporal Topic Summarization”.
6. Cong Wang, Ning Cao, Kui Ren, Wenjing Lou “Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data.
7. Song.D,Agner and Perrig. A, (IEEE 2000), “Practical techniques for searches on encrypted data,” in Proc. of IEEE Symposium on Security and Privacy’00.
8. Java AWT, By John Zukowski
9. System Analysis and Design, By Leonard Fertuck
10. Database Management System, By Rajesh Narang
11. Java Applets, By David Bishop