The Hong Kong Polytechnic University

D-Miner
(Version 1.0)
User Manual

Section 1: D-Miner Overview 4
Introduction 4
Benefit from D-Miner 4
Information required by D-Miner 4
System Requirement 4
Section 2: Start Using D-Miner 5
Login the D-Miner 6
Collect Web Data from Social Network Sites 6
Facebook Web Filter 7
Twitter Web Filter 9
Yahoo! Knowledge Web Filter 12
Submit Web Data to the Server 13
My Web Data 14
Upload Web Data 16
Request analysis on uploaded Web Data 17
My Analysis 17
Start Analysis 20
Section 3: Dictionary File 22

Figure 1: The Procedure of Using D-Miner 5
Figure 2: Menu of Web Filter 6
Figure 3: Accept the Facebook Web Filter 7
Figure 4: Allow the collection of Information 7
Figure 5: The control panel of the Facebook Web Filter 8
Figure 6: Progress Bar 9
Figure 7: The control panel of Twitter Web Filter 10
Figure 8: The control panel of Yahoo! Knowledge Web Filter 12
Figure 9: Items of Web Data Submission 13
Figure 10: The information displayed in “My Web Data” 14
Figure 11: The control panel of “Upload Web Data” 16
Figure 12: The Items of Data Submission 17
Figure 13: "My Analysis" Page 18
Figure 14: The Web Data Selector and the Control Panel 20

Section 1: D-Miner Overview

Introduction

D-Miner is a one-stop portal that is capable of collecting web data from social network sites such as **Facebook**, **Twitter** and **Yahoo! Knowledge** and discovering information according to users' preferences.

Benefit from D-Miner

There is much information in social network sites, and it is not possible to review the posts one by one for finding out those related to particular topic(s). D-Miner can automatically retrieve and store data in different social network sites for users and provide a matching engine afterwards.

Information required by D-Miner

Some social network sites require users to login their sites before D-Miner could collect data. However, the authentication process is just carried out between users and the corresponding social network sites. D-Miner will not store your username or password.

System Requirement

D-Miner is a web application. To use the D-Miner, all you need is a web browser with stable Internet connection and a login account. Below is detail information of the web browser required by D-Miner:

Web Browser	Version
Chrome	*11.0 or above* is Recommended
Firefox	*3.6 or above* is Recommended
Internet Explorer (IE)	*7.0 or above* is Recommended

Section 2: Start Using D-Miner

D-Miner is easy to operate. Users could get the analysis result by following the five steps stated below:

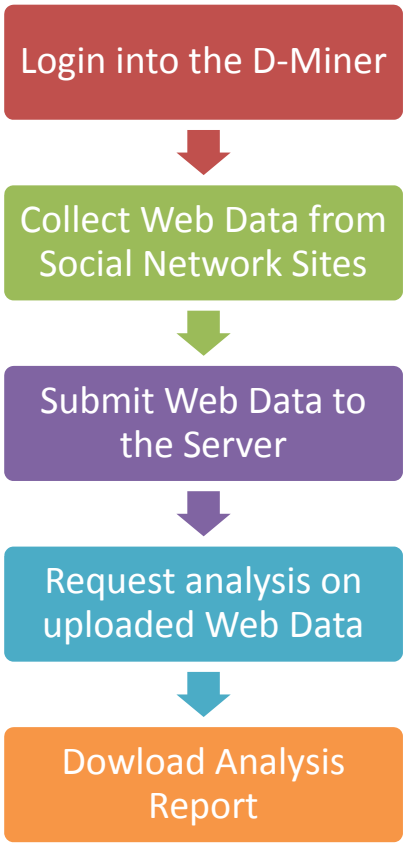


Figure 1: The Procedure of Using D-Miner

Login the D-Miner

To login D-Miner, you have to visit our portal first. The URL of portal is: **http://wms.comp.polyu.edu.hk/moodle**. Click the login button that is located on the right hand side of the menu bar. Then input username and password for authentication. Since **cookie** is required by D-Miner, you have to enable cookie in your browser.

Collect Web Data from Social Network Sites

D-Miner provides different Web Filters to collect data from social network sites for further analysis, and each Web Filter is designated for one social network site. In the current version, three Web Filters for ***Facebook***, ***Twitter*** and ***Yahoo! Knowledge*** are available.

In the menu of the left hand side, there is a section called "Web Filter". All filters are listed inside and named by the social network sites they are designed for. Each Web Filter has its control panel for users to decide filter's parameters. Select the web filter inside the menu for entering the corresponding control panels.

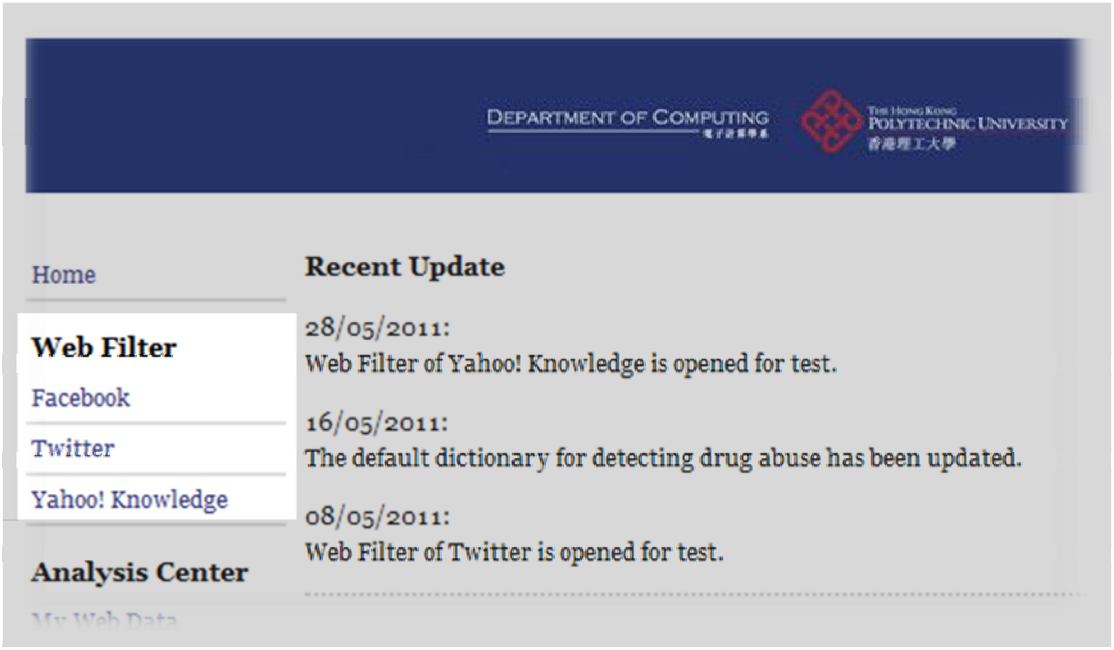


Figure 2: Menu of Web Filter

Here, we will introduce the functionalities of the three available filters and demonstrate how to use them step by step.

Facebook Web Filter

This filter is capable of collecting the information of your Facebook friends and the corresponding Facebook groups joined by them. You have to login your Facebook account before you could access the control panel of the Facebook filter. If it is the first time for you to use our filter, you have to accept the web filter and allow it to collect information for you, as shown in Figure 3 and 4.

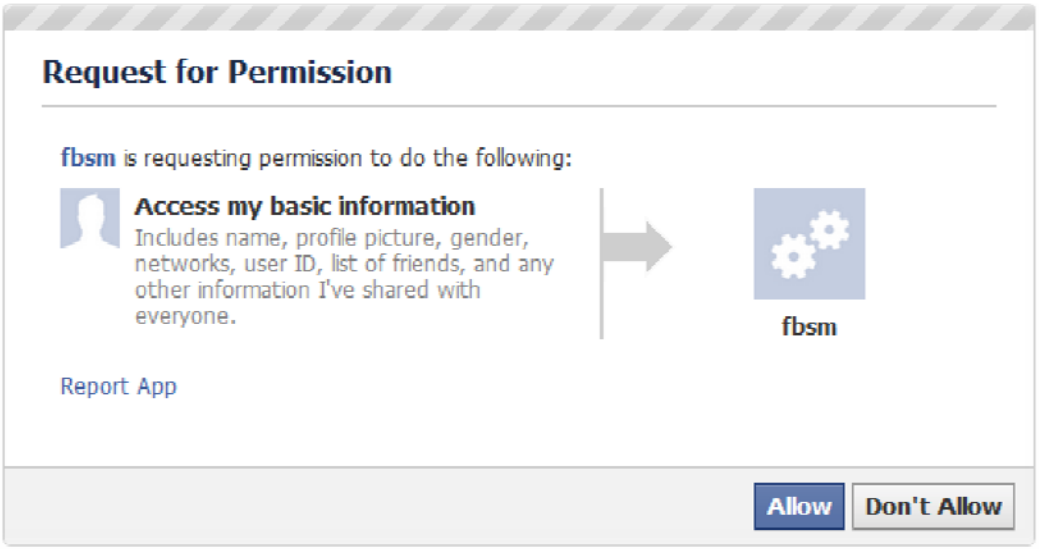


Figure 3: Accept the Facebook Web Filter

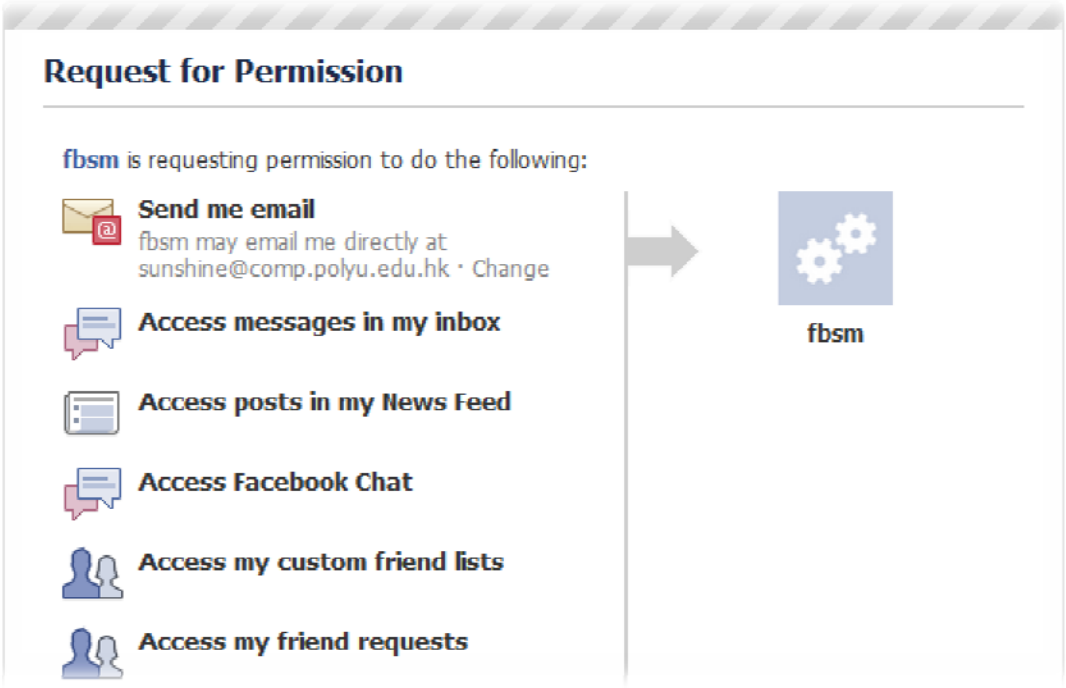


Figure 4: Allow the collection of Information

To start the Web Filter of Facebook, you have to provide two criteria for data collection. They are ***Mode of Collecting Data*** and ***Types of data to be collected***.

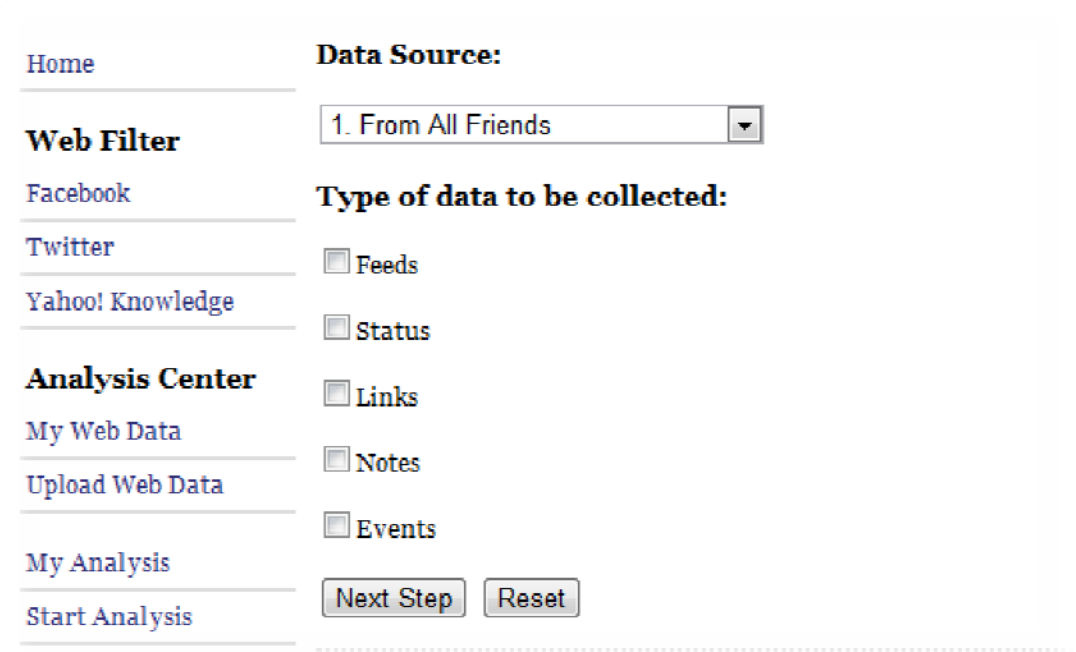


Figure 5: The control panel of the Facebook Web Filter

The "Data Source" determines the information sources to be collected by the filter. There are 4 modes of data collection.

- 1. From All Friends
- 2. From Selected Friends
- 3. From All Friends' Groups
- 4. From Selected Friends' Groups

The "Type of Data to be collected" specifies which sort(s) of data in Facebook are targets to be collected. They include:

- Feed
- Status
- Link
- Note
- Event

Feed

A message (not more than 1,000 characters) posted by the friends of wall owner or Facebook applications.

Status

A short message (not more than 420 characters) posted by the wall owner.

Link

A webpage shared by the wall owner. Common examples are newspaper and video clip.

Note

A passage written by the wall owner

Event
The activities raised or joined by your friends

After you have selected the data source and the types of data to be collected, click "Next Step" to continue. Depending on your selected data source, you need to specify further about how, how many and whose data should be extracted. For example, you need to select the number of friends and number of data per friend to be extracted in mode 1 and 3. And under mode 2 and 4, you need to tick the boxes beside the friends you want to select and number of data per friend to be extracted.

After you have started the data collection, please wait for the process to complete. The progress of data collection is shown via the progress bar.

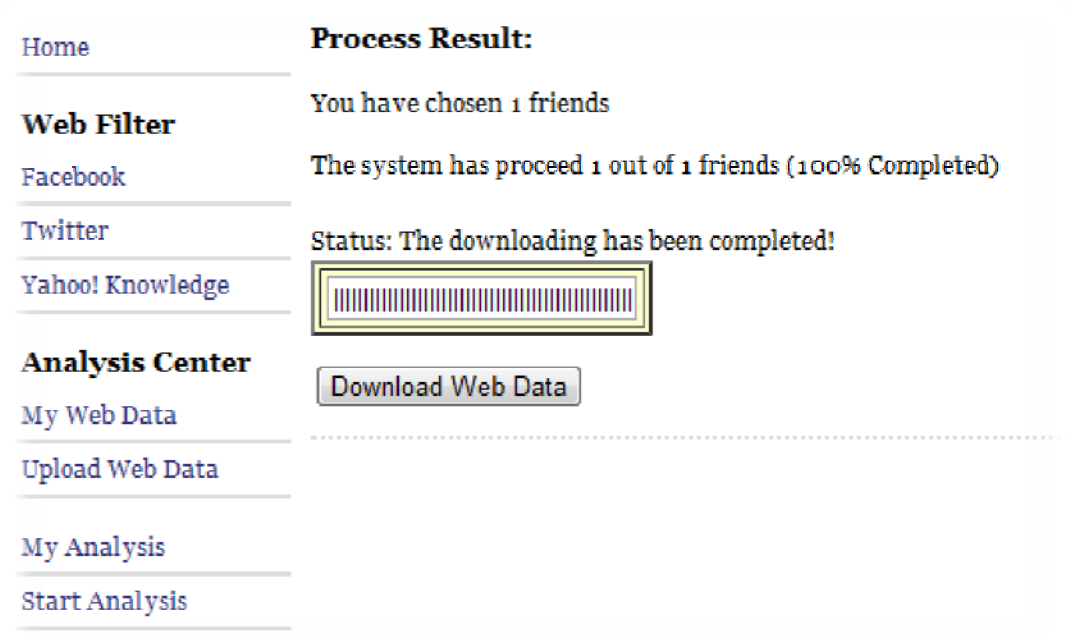


Figure 6: Progress Bar

Twitter Web Filter

This filter is capable of collecting the information of the friends which you have "followed" in Twitter. You have to login your Twitter account before you could access the control panel of the filter. To start the Web Filter of Twitter, you have to provide three criteria for data collection. They are "***From now back to"***, "***Time Zone"*** and "***Searching Range"*** as shown in Figure 7.

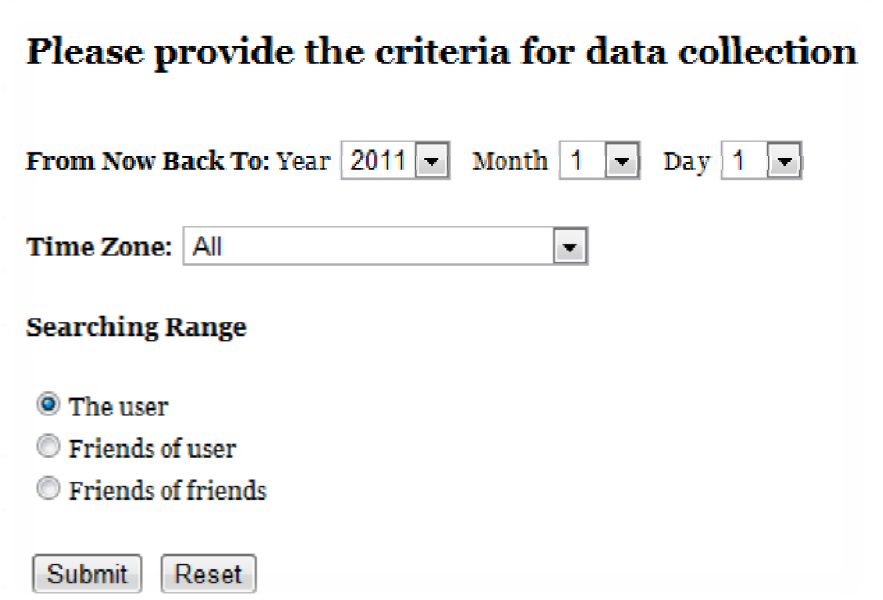


Figure 7: The control panel of Twitter Web Filter

From Now Back To

It limits the number of messages to be collected according to their tweeting days.

Time Zone

It limits the number of messages to be collected according to the time zone that the tweeter located in.

Searching Range

IT determines the information source to be focused in the search. There are three available options, and they are:

- The User
- Friends of User
- Friends of Friends

The user

The system will target on the messages tweeted or re-tweeted by you and their corresponding comments.

Friends of user

The system will target on the messages tweeted or re-tweeted by your friends and their corresponding comments.

Friends of friends

The system will target on the messages tweeted or re-tweeted by the friends of your friends and their corresponding comments.

After you have set all the criteria, click “Start” button to initiate the data collection. Depending on the size of data to be retrieved, the data collection process may take several minutes. After the collection is over, you may click “***Download Web Data***” to save the result.

Yahoo! Knowledge Web Filter

This filter is capable of collecting the questions raised in Yahoo! Knowledge (Hong Kong). Different from the above filters, you are not required to login "Yahoo!" for accessing the control penal of the Yahoo! Knowledge filter. To start the Web Filter of Yahoo! Knowledge, you have to provide two criteria for data collection as shown in Figure 8. They are "***the maximum number of questions to be collected***" and "***Collect topics within***".

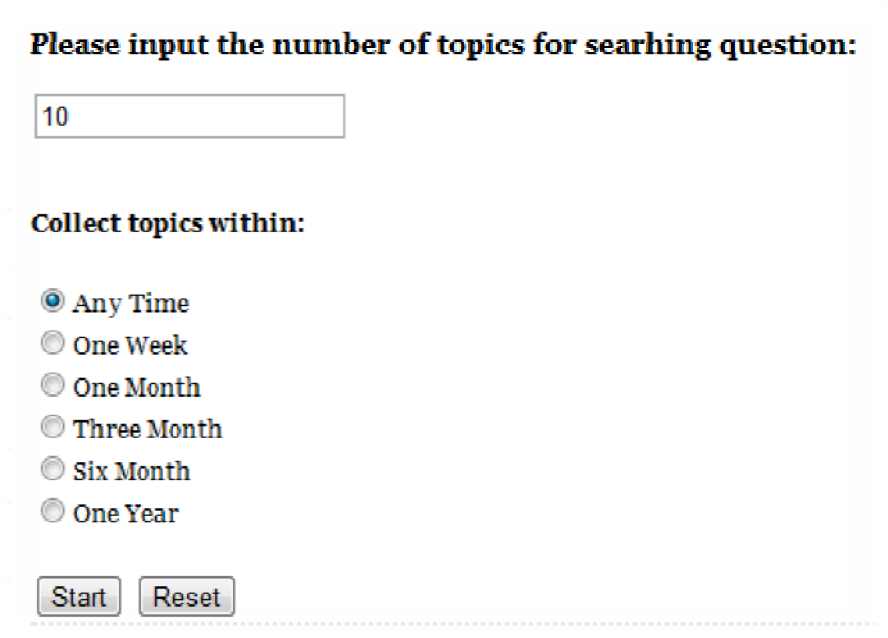


Figure 8: The control panel of Yahoo! Knowledge Web Filter

Maximum number of questions to be collected

It limits the number of questions to be collected from Yahoo! Knowledge (Hong Kong).

Collect topics within

It limits the number of questions to be collected according a past period.

After you have set all the criteria, click "Start" button to initiate the data collection. Depending on the size of data to be retrieved, the data collection process may take several minutes. When it is over, you may click "***Download Web Data***" to save the result.

Submit Web Data to the Server

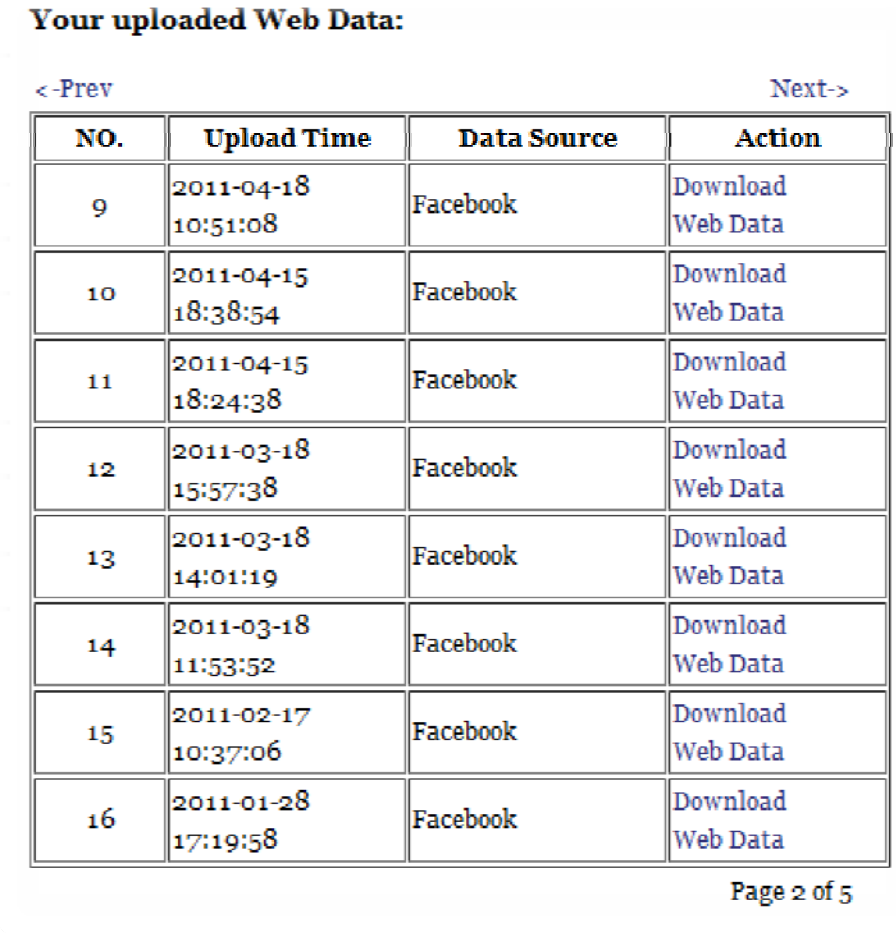
After you have downloaded the web data through the web filters, you are suggested to open the file for verifying whether the data inside is suitable for analysis or not (*However, editing is not recommended*). Once the content is approved, you may send it to the server.

In the menu of the left hand side, there is a section called “Analysis Center”. There are two options related to web data submission. They are “My Web Data” that displays your submitted data and “Upload Web Data” that enables users to submit the web data to the server as shown in Figure 9.

Figure 9: Items of Web Data Submission

My Web Data

In "My Web Data" page, the data is displayed in form of table, and it is shown as Figure 10.



Your uploaded Web Data:

<-Prev Next->

NO.	Upload Time	Data Source	Action
9	2011-04-18 10:51:08	Facebook	Download Web Data
10	2011-04-15 18:38:54	Facebook	Download Web Data
11	2011-04-15 18:24:38	Facebook	Download Web Data
12	2011-03-18 15:57:38	Facebook	Download Web Data
13	2011-03-18 14:01:19	Facebook	Download Web Data
14	2011-03-18 11:53:52	Facebook	Download Web Data
15	2011-02-17 10:37:06	Facebook	Download Web Data
16	2011-01-28 17:19:58	Facebook	Download Web Data

Page 2 of 5

Figure 10: The information displayed in "My Web Data"

For each web data set, 4 types of information are listed out. They are ***"NO."***, ***"Upload Time"***, ***"Data Source"*** and ***"Action"***.

NO.

It represents the sequence of the web data. A smaller number means the more recent the data is.

Upload Time

It indicates the date and time that the data submitted to the server.

Data Source

It indicates the social network site that the data is extracted from.

Action

It enables user to download the corresponding web data by clicking the "Download Web Data" link.

Each page displays at most 8 web data sets. You could click "***<-Prev***" and "***Next->***" buttons to browse all data sets that are available.

Upload Web Data

The “Upload Web Data” page enables you to submit the web data to the server, and you could add 5 files in the page as shown in Figure 11. If the quota is not sufficient to you, you could click the link “More…” and the quota will be increased from 5 to 10. To upload web data, you could click the “Browse” button and select the previously downloaded web data in your computer.

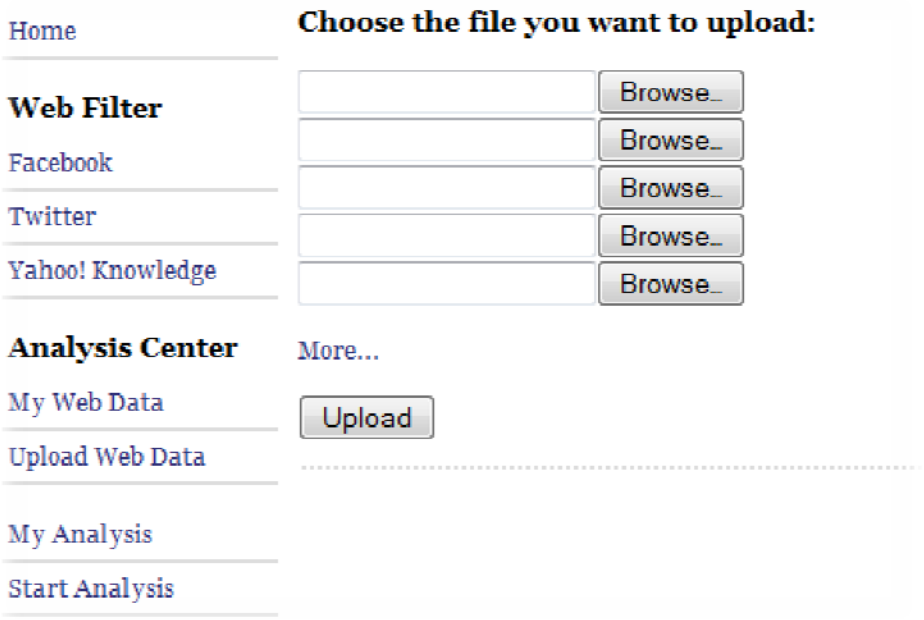


Figure 11: The control panel of “Upload Web Data”

After you have added your desired files, click “***upload***” to process.

Request analysis on uploaded Web Data

After you have submitted your web data to the server, the data will be analyzed. However, such analysis will not be carried out automatically. You have to send a "request of analysis" to initiate the data analysis. After your request is received, the analysis will be carried out.

Similar to the submission web data to the server, the sending of request analysis is also carried out in the "Analysis Center" which is located in the menu of the left hand side. There are two options, "My Analysis" that displays your submitted "requests of analysis" and "Start Analysis" that enables you to send out "requests of analysis" as shown in Figure 12.

Figure 12: The Items of Data Submission

My Analysis

In the "My Analysis" page, the submitted analyses are listed in form of table.

Your submitted analysis:

Next->

Request Time	Request Description	Dictionary Applied	Status	Action
2011-06-09 15:41:34	Yahoo_Knowledge_0609	Standard Drug Dictionary	Pending	Not Available
2011-05-31 17:09:00	Test_For_Yahoo_Again_Again	Standard Drug Dictionary	Completed	Download Report
2011-05-31 17:03:41	Test_For_Yahoo_Again	Standard Drug Dictionary	Completed	Download Report
2011-05-31 10:55:46	Test_For_Yahoo	Standard Drug Dictionary	Completed	Download Report
2011-05-05 10:45:06	Test for Twitter File	Standard Drug Dictionary	Completed	Download Report
2011-04-15 18:39:51	H.L.Li	Standard Drug Dictionary	Completed	Download Report
2011-03-18 15:58:32	HKPC Workshop Testing	Standard Drug Dictionary	Completed	Download Report
2011-03-18 14:01:39	HKPC Workshop Test 02	Standard Drug Dictionary	Completed	Download Report

Page 1 of 2

Figure 13: "My Analysis" Page

For each submitted analysis, 5 types of information are listed out. They are ***"Request Time"***, ***"Request Description"***, ***"Dictionary Applied"***, ***"Status"*** and ***"Action"***.

Request Time
It indicates the date and time that the analysis is submitted to the server.

Request Description
It is a name input by the D-miner user for representing the analysis.

Dictionary Applied
It indicates the keyword dictionary used in the analysis.

Status
It indicates the current status of the analysis. "***Pending***" means the analysis has not been initiated yet and "***Completed***" means the analysis has been finished and the report is ready.

Action
It enables user to download the reports of the corresponding analyses after the analysis is completed.

Each page could display at most 8 submitted analyses. You could click “<-Prev” and “Next->” buttons to browse all analyses those are submitted by you.

Start Analysis

In the "Start Analysis" page, there are two major components, which are the selector of web data for analysis and the control panel for launching analysis.

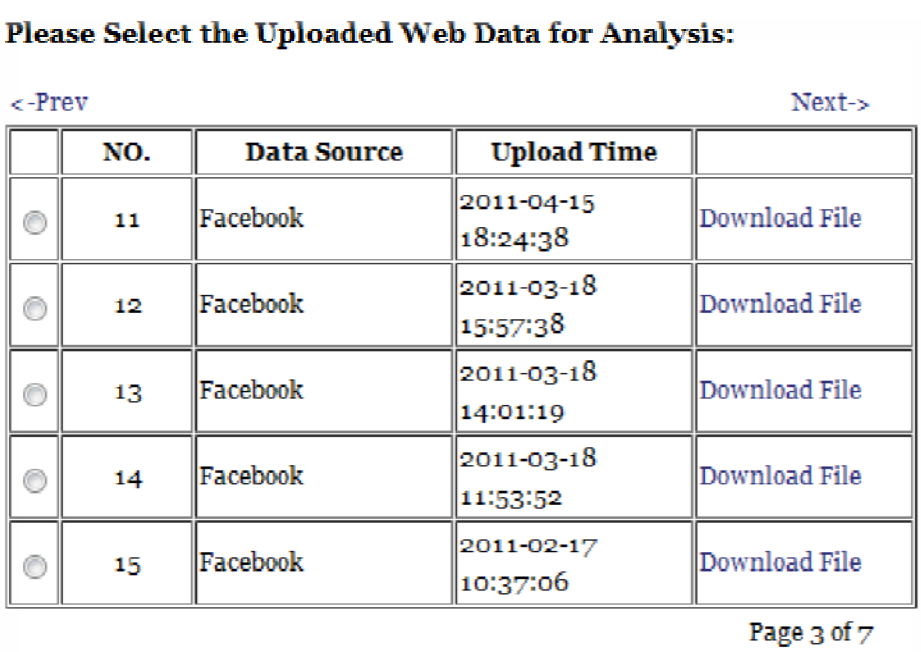


Figure 14: The Web Data Selector and the Control Panel

The web data selector determines the web data to be analyzed by the D-miner and the information contained in the selector is similar to the table in "My Web Data" page, and a radio button is placed in front of each web data for your selection. Each page could display at most 5 submitted analyses. You could click "<-Prev" and "Next->" buttons to browse all available web data.

The control panel for launching an analysis enables you to decide the features of the analysis. They include:

- Description
- Analysis Type
- Dictionary File

Description

A name for users to identify their analyses.

Analysis Type

It indicates the matching algorithm to be used in the analysis. In the current version, "***Keyword Matching***" is available.

Dictionary File

It specifies the interested topic(s) to be discovered in the analysis. You may submit your own dictionary to us. The setup of a new dictionary will be discussed in detail in next section.

After you have set all the criteria, click "Proceed" button to submit the request. D-miner will bring you back to the "My Analysis" page and you will see that the submitted request will be placed at the beginning with "***Pending***" state. The state will be updated to "***Completed***" after the analysis is completed.

Section 3: Dictionary File

The dictionary file is important in the analysis. It is the basis of the matching algorithm. In order to simplify it for our users, it has a simple structure. It is composed of "***Keywords***" and "***Supporting Words***".

Keywords
The words you are interested in *(E.g.: ticket).*

Supporting Words
The words that usually appear with the keywords *(E.g.: Football, Scalped).*

The keywords and the supporting words worked together for performing the desired analysis. The more the keywords are given, the wider the scope of different coverage; the more of the supporting words are given, the deeper of the analyses are.

In D-Miner, a dictionary for discovering drug abuse will be assigned to all users. If you would like to have your own dictionary, you may prepare a list of "***Keywords***" associated with "***Supporting Words***" and submit them to us via sunshine@comp.polyu.edu.hk