

# TwitterVane

## *User Manual*

*01 March 2013*





Introduction.....	3
About TwitterVane.....	3
About this document.....	3
Where to find more information.....	3
System Overview.....	4
Purpose and scope.....	4
How does it work?.....	4
Home Page.....	6
Collections.....	8
Reports.....	15
Criteria driven reports.....	15
Browse Reports.....	17
Streamed Tweets.....	23
Help.....	24



## Introduction

### About TwitterVane

TwitterVane is a tool for leveraging the power of the crowd to select websites for web archiving. It enables curators to define web collections and identify trending Tweets and associated URLs that are relevant to those collections.

### About this document

This document is the TwitterVane User Manual. It describes how to use TwitterVane through its web browser interface. It assumes your system administrator has already set up the system.

### Where to find more information

The public resources for TwitterVane, including the source code, binaries and documentation, are available through the TwitterVane Open Source GitHub project:

<https://github.com/ukwa/twittervane>



### Purpose and scope

The application has been designed to stream Tweets directly from Twitter and store them for analysis. The analysis consists of expanding any shortened URLs associated with the Tweets using the Bitly expander service and associating the Tweets with the Web Collections that have been defined.

### How does it work?

TwitterVane is divided into three major components

#### **TweetView**

*TweetView* is the curator web interface where users define the collections and search terms that will be used to collect Tweets. A series of reports is also provided to identify popular Tweets and top URLs by collection.

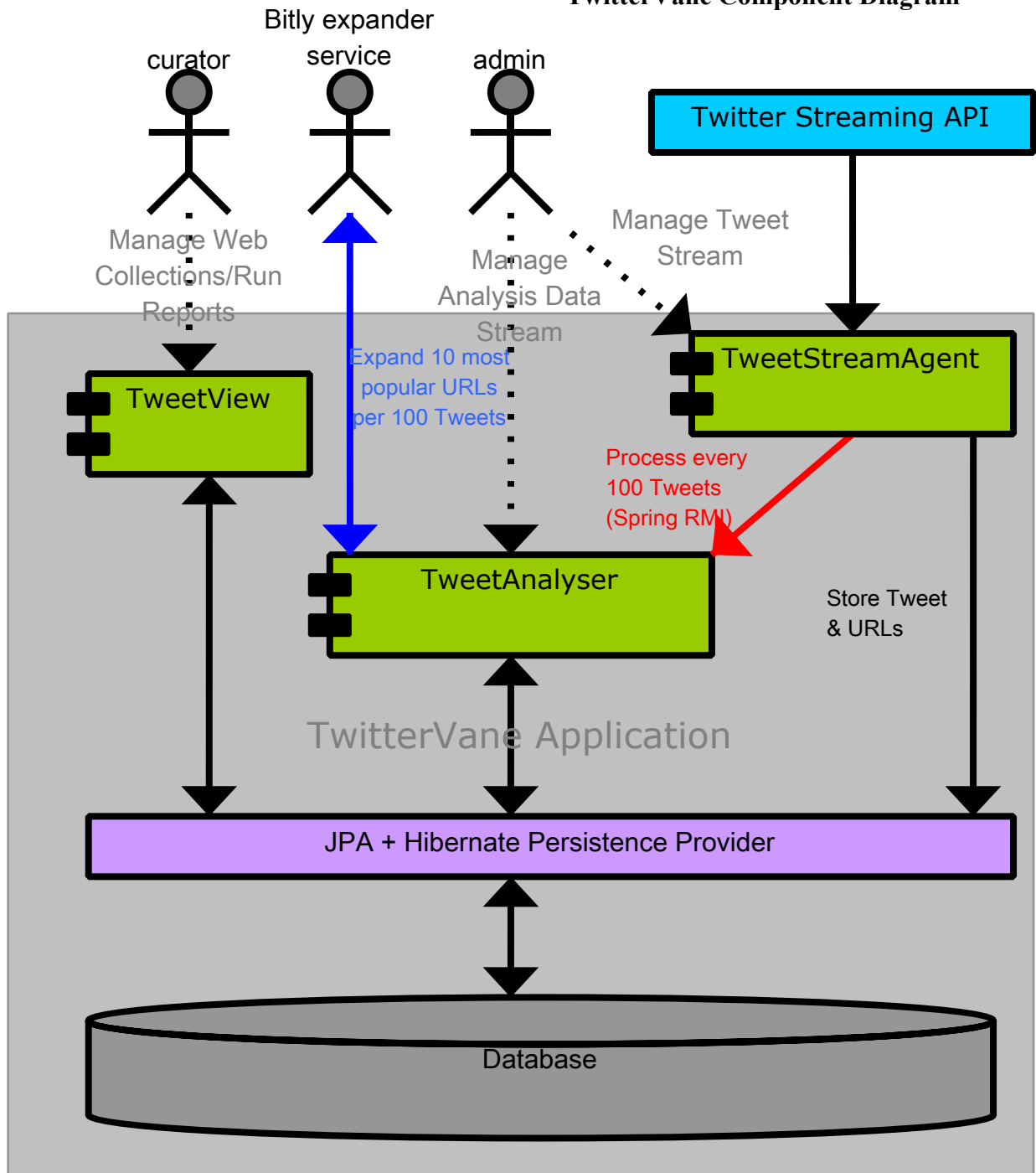
#### **TweetStreamAgent**

*TweetStreamAgent* streams Tweets directly from Twitter and stores them in the TwitterVane database for further analysis. It uses the search terms that have been defined by curators through *TweetView* to filter the Tweet stream. This is a management web interface that enables users to control and configure the Tweet stream.

#### **TweetAnalyser**

*TweetAnalyser* runs periodically, after a specific number of Tweets have been received from the Tweet stream, and expands any shortened URLs using Bitly. Any Tweet URLs are then resolved to the web collection using the collection's search terms.

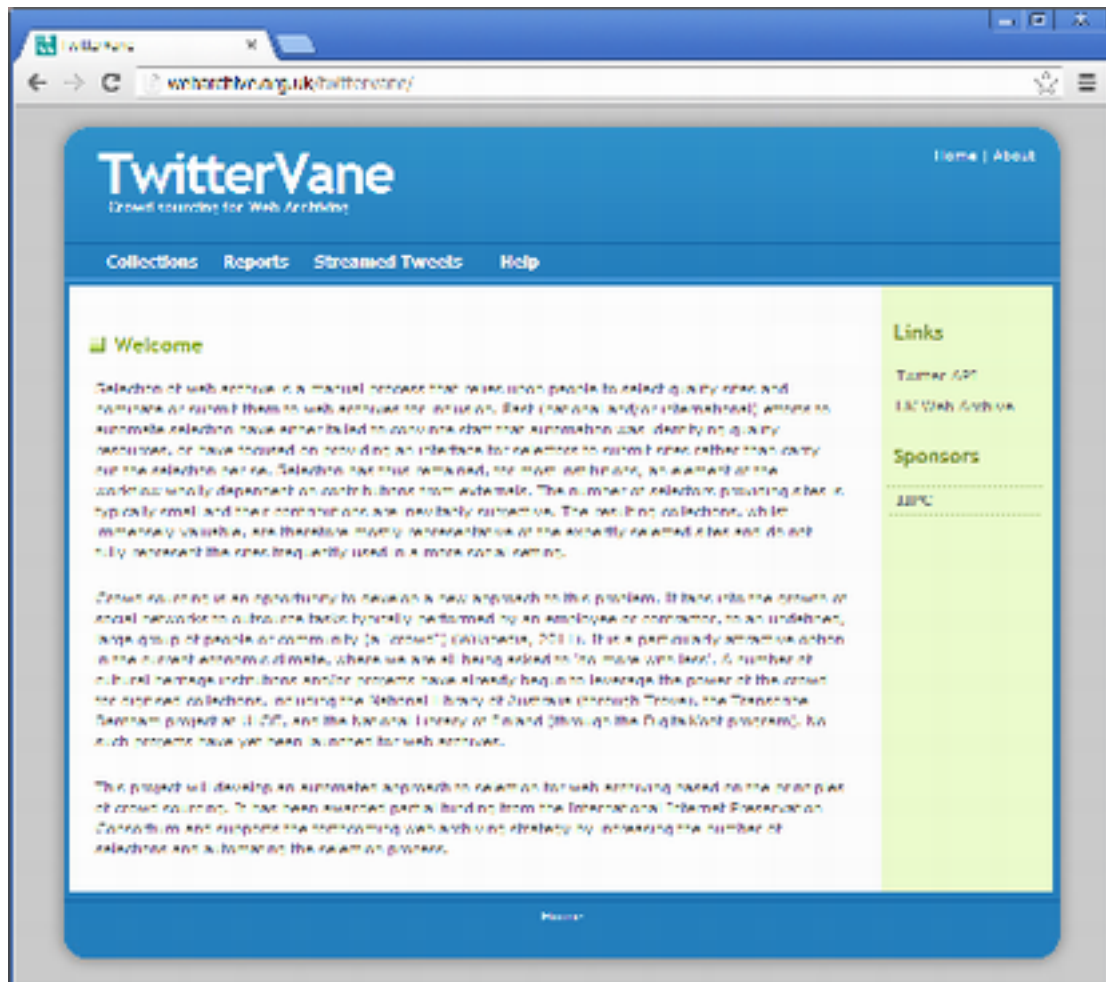
**TwitterVane Component Diagram**





## Home Page

The home page for TwitterVane is pictured below:



The top navigation menu items give access to the functionality to create and manage web collections, and view reports:

**Collections** – create, view and manage web collections and associated search terms

**Reports** – view and run the criteria based reports and ‘browse reports’

**Streamed Tweets** – view the latest Tweets received from the Twitter

**Help** – view the online help and the TwitterVane User Manual (ie: this document)

The left hand navigation menu contains TwitterVane's related links.



The **Collections** page displays the Web Collections that are currently defined on the system, the search terms associated with the collection, and the start and end dates within which Tweets should be collected:

The screenshot shows the TwitterVane web interface. The header includes the TwitterVane logo and navigation links: Home, About, Collections, Reports, Streamed Tweets, and Help. The main content area is titled 'Collections' and contains a table with the following data:

Collection	Start Date	End Date	Search Terms	
Volkswagen	February 14, 2013	February 20, 2013	volkswagen vw vwvans vwvs	Delete
my	February 18, 2013	March 5, 2013	my myvans myvs myvts	Delete
Volkswagen	February 25, 2013	March 5, 2013	Volkswagen	Delete

Below the table is a section titled 'Add New Collection' with the following fields:

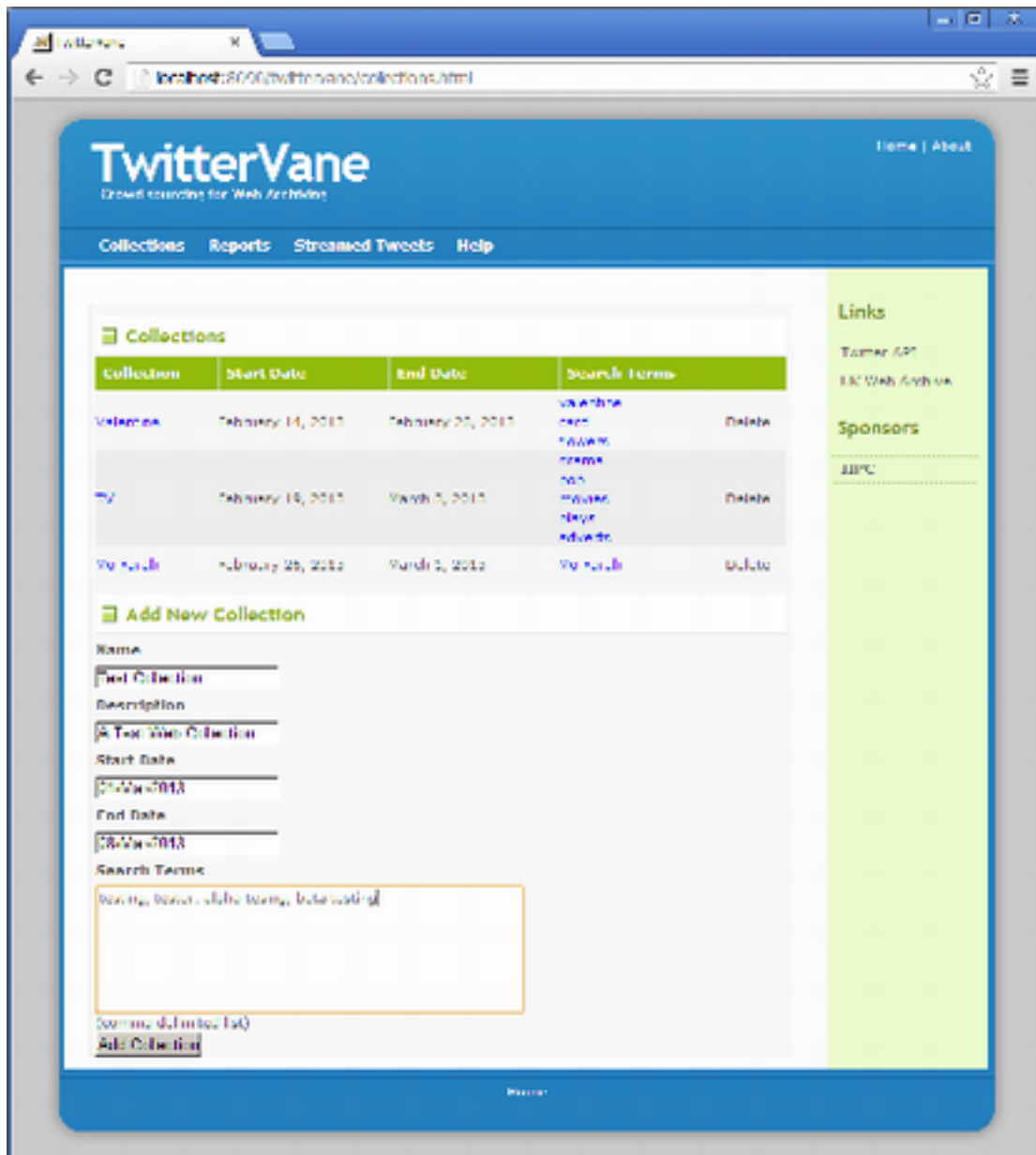
- Name
- Description
- Start Date
- End Date
- Search Terms

At the bottom of the form, there is a text input field containing '(comma delimited list)' and a button labeled 'Add Collection'.

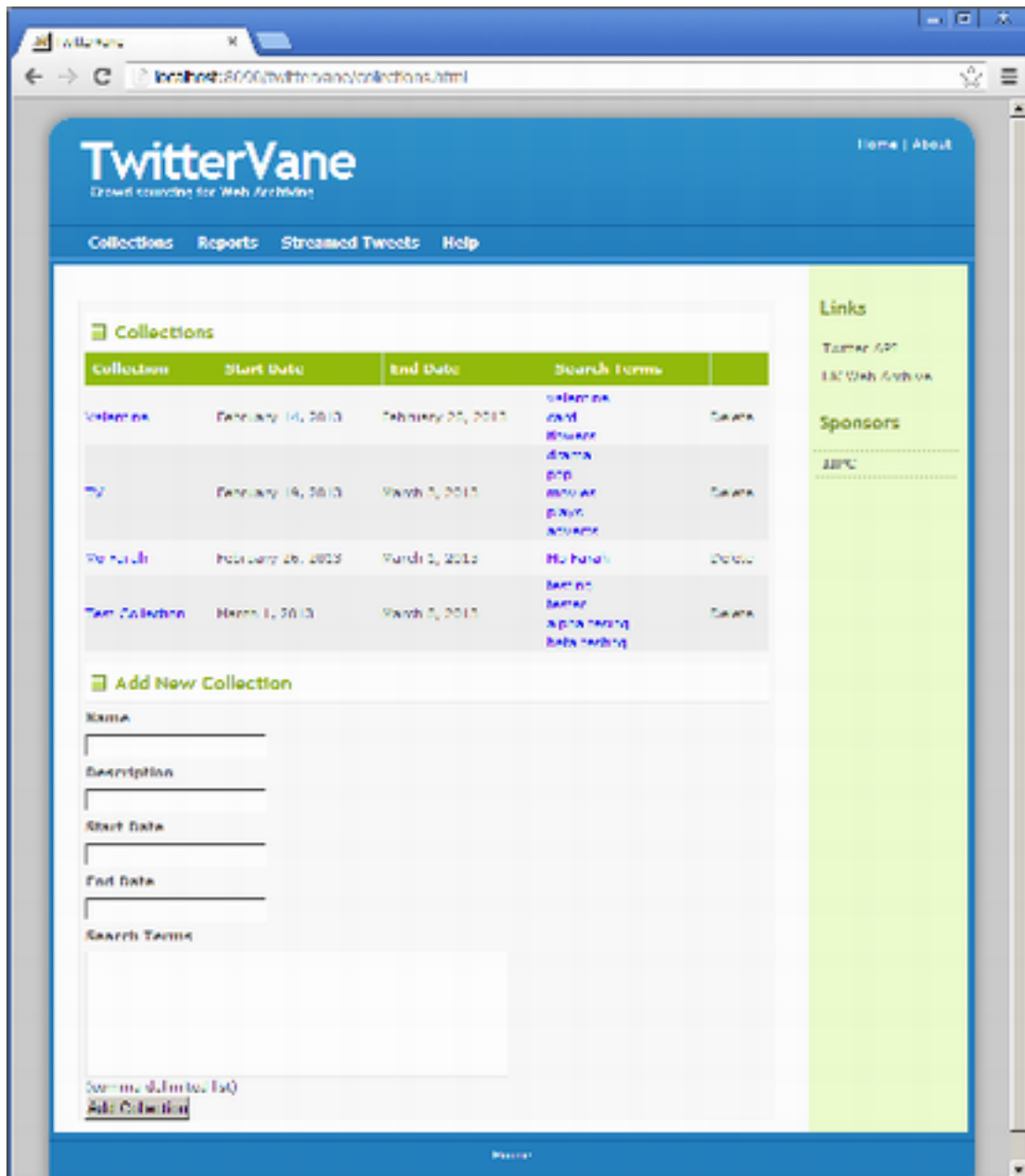
On the right side of the page, there are sections for 'Links' (Twitter API, IBC Web Archive) and 'Sponsors' (APC).

To create a new collection, fill out the name, description, start date, end date and search terms (search terms as a comma-separated list) in the fields at the bottom of the screen and click 'Add'.



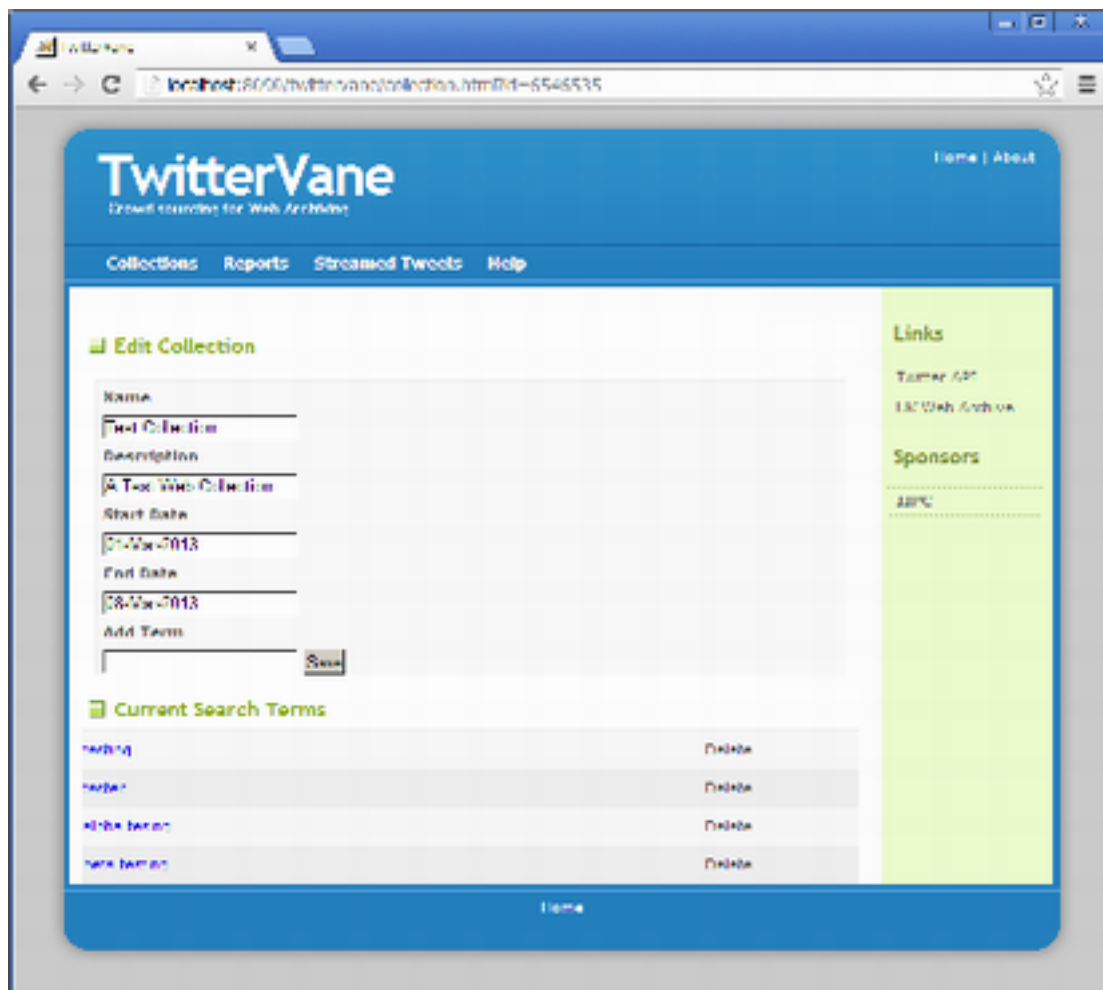


The collection then appears in the list:

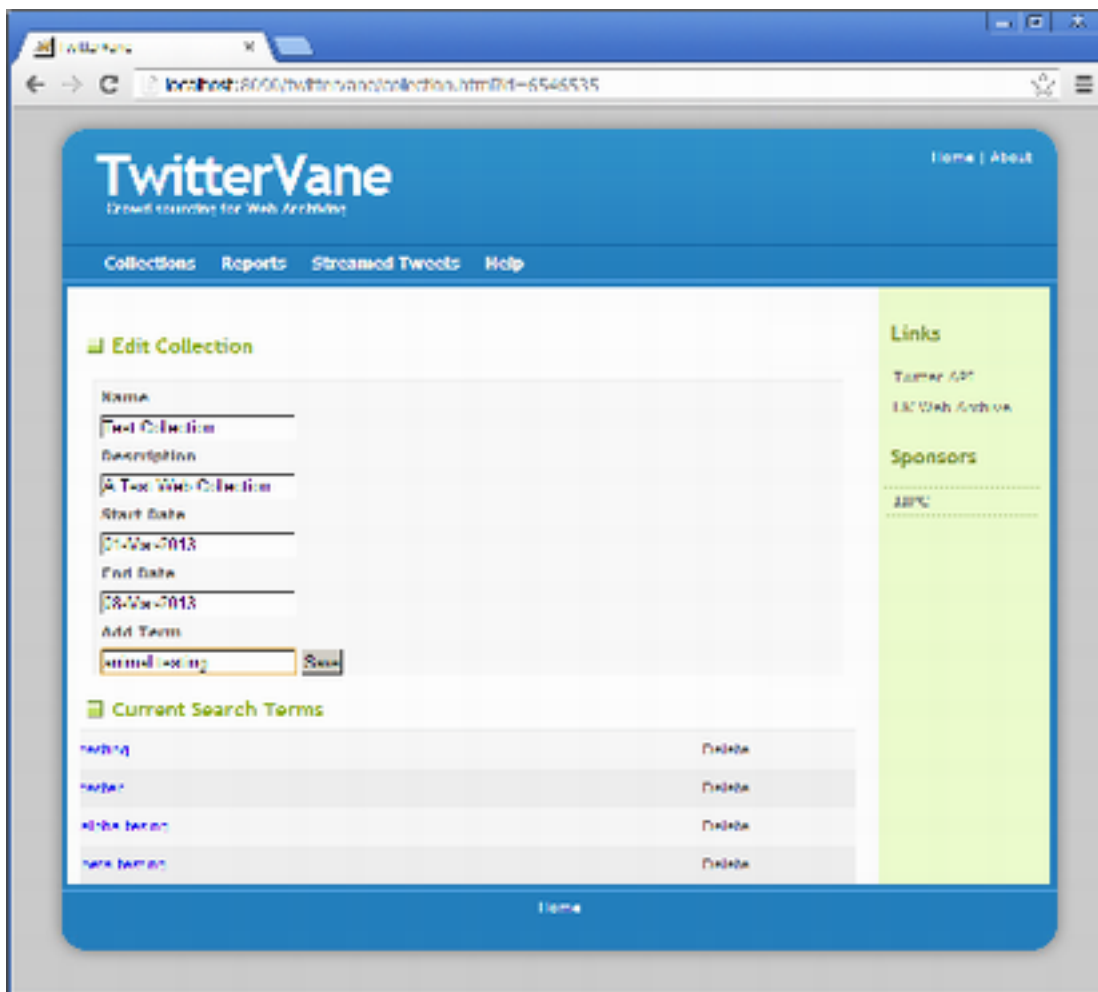


To manage a collection and associated search terms, click on the collection name:

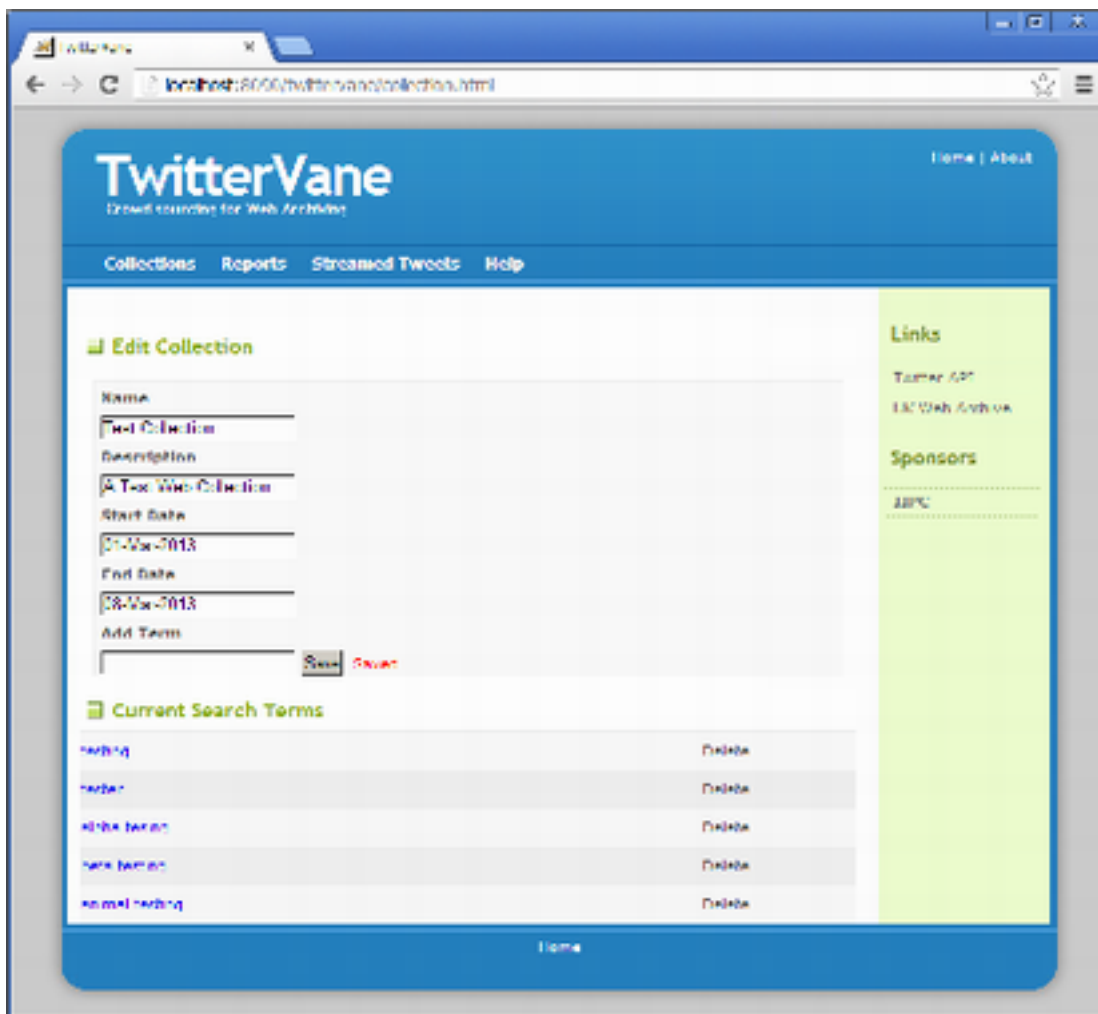
The **Edit Collection** page is then displayed where the collection details can be modified:



Enter the new search term into the 'Add Term' text box at the bottom of the page and click 'Save':

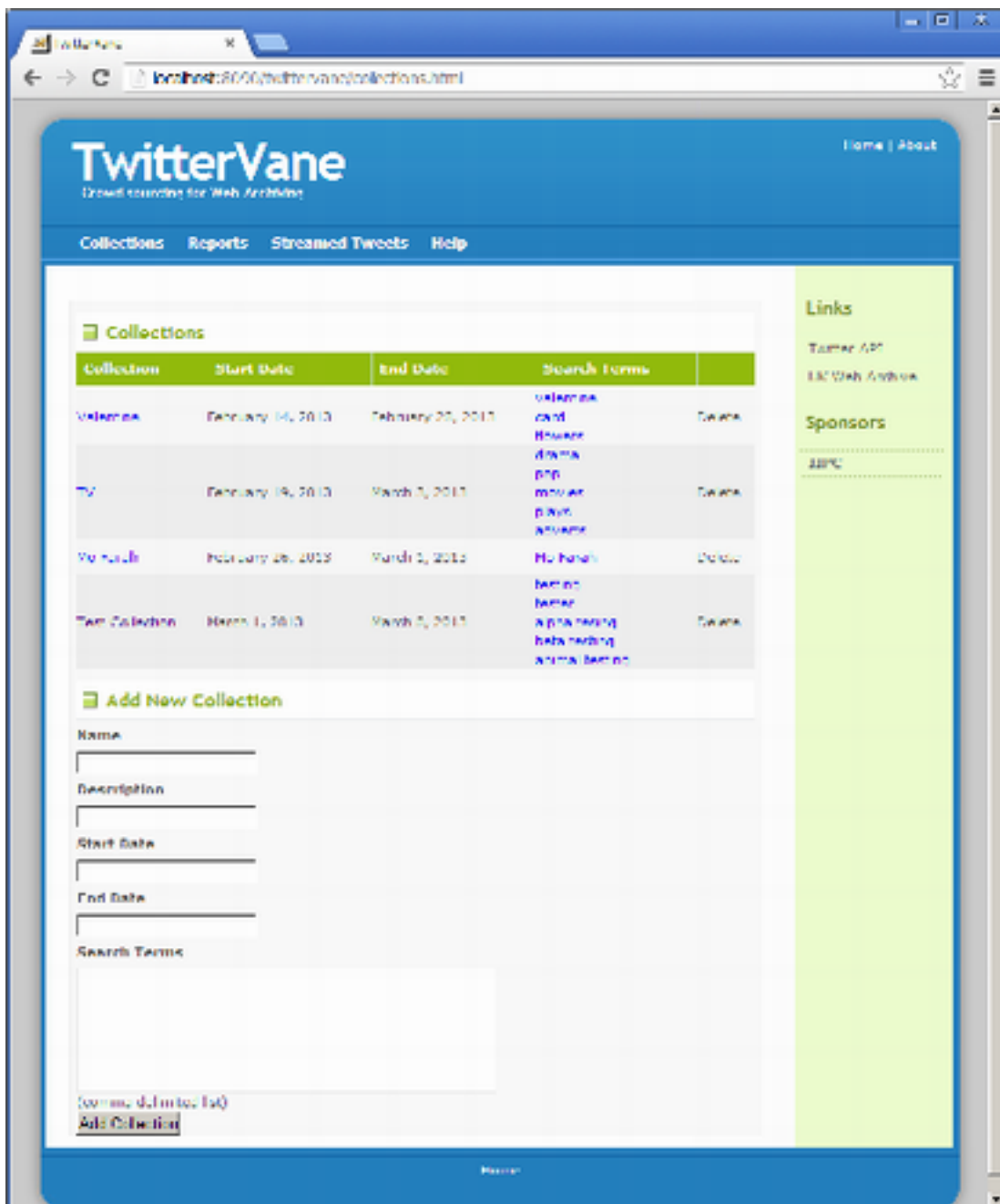


The search term is then listed at the bottom of the page:



To delete a search term, use the delete link on the right of the search term.

To delete a collection, click the collections menu item at the top of the screen and use the delete link on the right of the collection on the **Collections** screen:



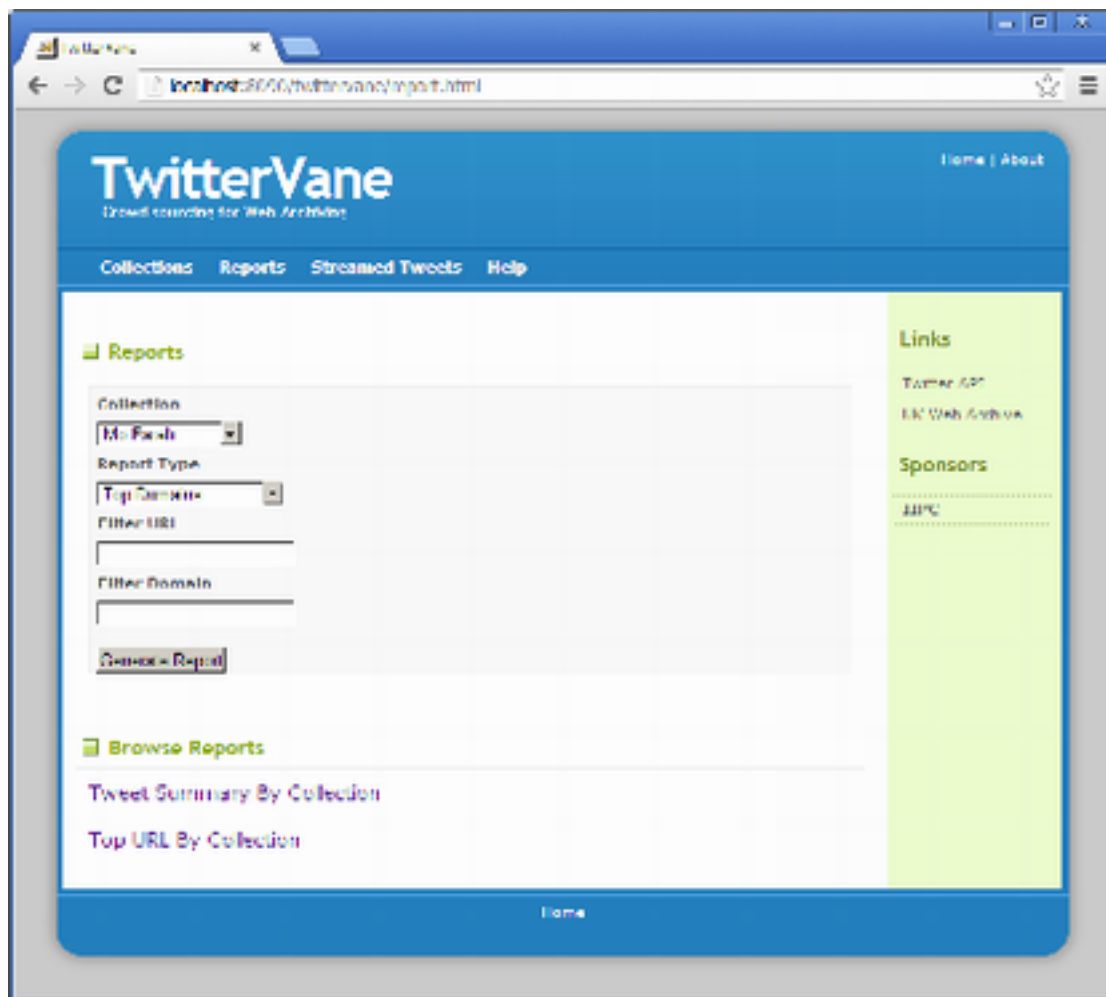
**WARNING:** Deleting a web collection will remove all associated search terms, Tweets and URLs from the database.



## Reports

There are two types of report available in TwitterVane:

1. Criteria driven reports
2. 'Browse Reports'



## Criteria driven reports

TwitterVane has four criteria reports defined:

1. **Top Domains** – returns the first 100 domains containing the greatest number of URLs for the selected Web Collection. The results are sorted on the number of URLs in descending order

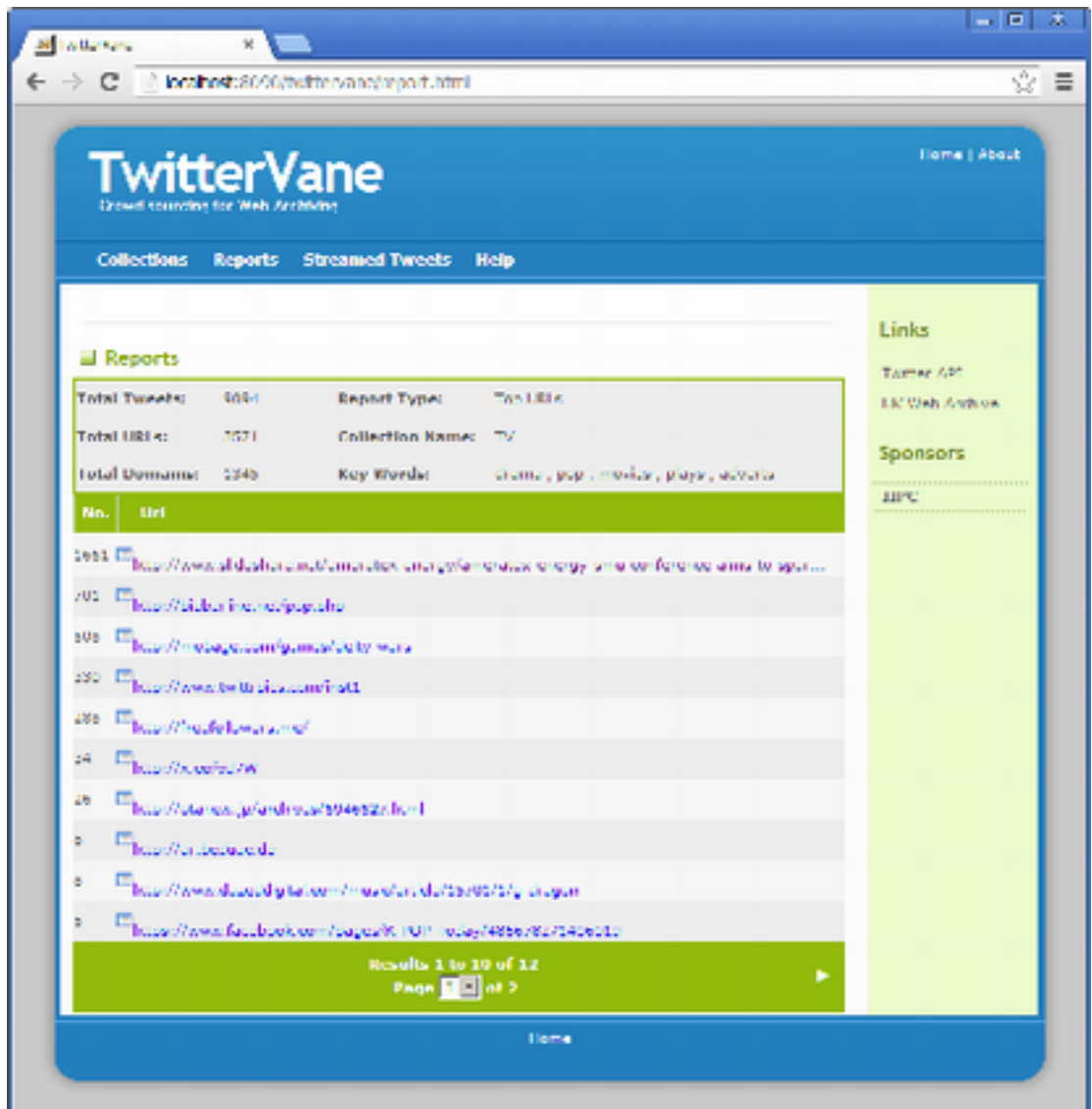
The screenshot shows the TwitterVane web application interface. The header includes the TwitterVane logo and navigation links: Home | About. Below the header is a menu bar with links: Collections, Reports, Streamed Tweets, and Help. The main content area is titled 'Reports' and displays a summary of a report for a collection named 'TV'. The summary includes: Total Tweets: 1041, Total URLs: 1071, Total Domains: 1345, Report Type: Top Domains, Collection Name: TV, and Key Words: drama, page, movie, play, device. Below the summary is a table of domains with columns 'No.', 'Domain', and 'Count'. The table lists the top 10 domains. To the right of the table is a sidebar with sections for Links (Twitter API, UK Web Archive) and Sponsors (JRC).

No.	Domain	Count
1029	lulu	1029
101	flora	101
142	devi	142
435	youtube	435
225	instagram	225
240	blogspot	240
197	volly	197
187	bristolmail	187
182	askfm	182
181	google	181

Results 1 to 10 of 1345  
Page 1 of 135

2. **Top URLs**- returns the top 100 expanded URLs for the selected Web Collection.



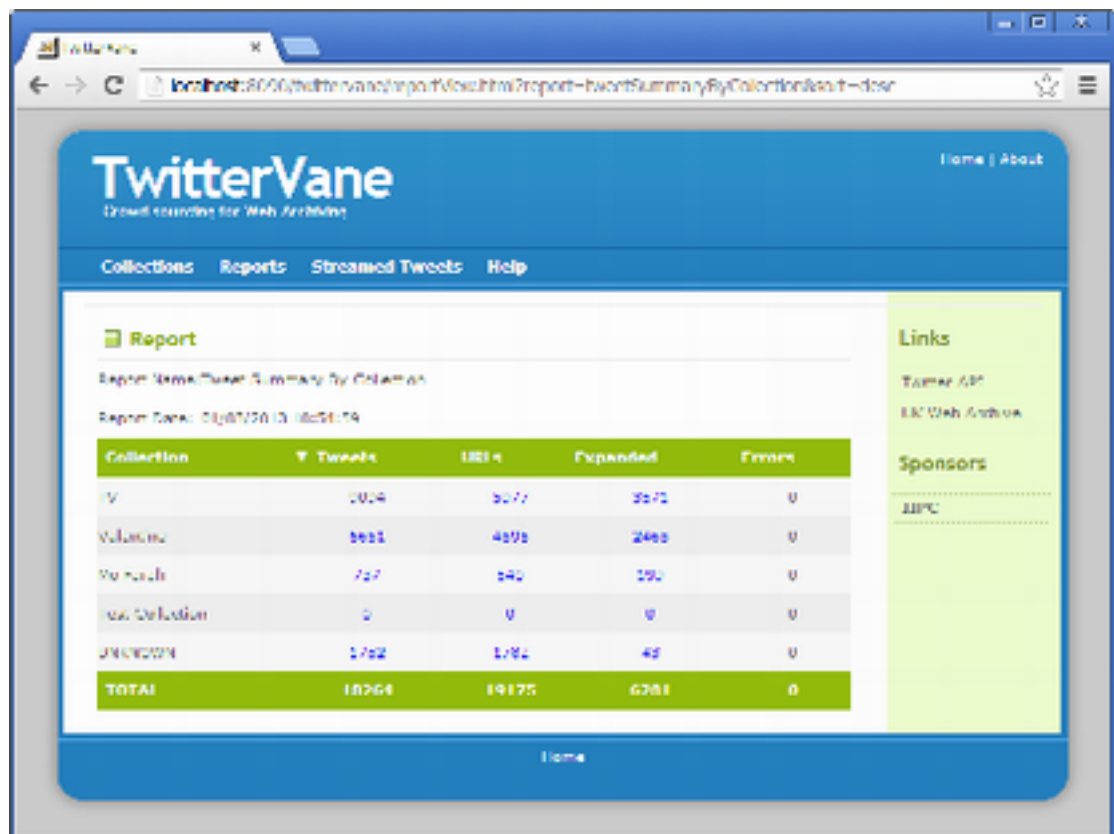


3. **Top URL by Retweet** - returns the most popular expanded URL for the selected Web Collection. Popularity is re-tweet count of the Tweet received from Twitter.

## Browse Reports

There are two 'browse reports' defined for TwitterVane. The reports contain sortable columns and the sort order can be changed by clicking on the column header.

1. **Tweet Summary By Collection** - displays the number of Tweets, URLs, expanded URLs and errors for each collection.



Clicking the Tweets figure will display the number of Tweets by time period for the collection. The time periods are represented graphically for the past week, month and year.



Clicking on the figure under the URLs column for the collection will display all the **unexpanded** URLs for that collection.

**TwitterVane**  
Crowd-sourcing for Web Analytics

Home | About

Collections Reports Streamed Tweets Help

**Report**

Report Name: URLs To Criminals

Report Date: 6/10/2013 10:56:05

Criminal Name: To

URL	Twitter	Tweet
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...
<a href="http://www.4mat.com/...">http://www.4mat.com/...</a>	4mat	RT @bubuyayay: Unfolding drama. The sugar mommy website www.4mat.com (http://www.4mat.com) receiving 90% of its traffic from...

Results 1 to 10 of 5977  
Page 1 of 598

Home

**Links**

Twitter API  
URLs To Criminals

**Sponsors**

HTTPC

Clicking on the figure under the Expanded column for the collection will display all the **expanded** URLs for that collection.



2. **Top URL By Collection** - displays the top URL, Tweets and Re-Tweets for each collection. The top URL is the most frequently occurring URL in that collection.

The screenshot shows a web browser window displaying the TwitterVane application. The browser's address bar shows the URL: `localhost:8080/twittervane/report/view.html?report=topUriByCollection&sort=desc`. The application has a blue header with the TwitterVane logo and the tagline "Crowd sourcing for Web Archiving". Navigation links include "Home" and "About". A secondary navigation bar contains "Collections", "Reports", "Streamed Tweets", and "Help".

The main content area is titled "Report" and displays the following information:

- Report Name:** Top URI By Collection
- Report Date:** 6/25/2013 11:06:19

A table follows, showing the top URIs by collection. The table has four columns: URI, Collection, Tweets, and Estimate.

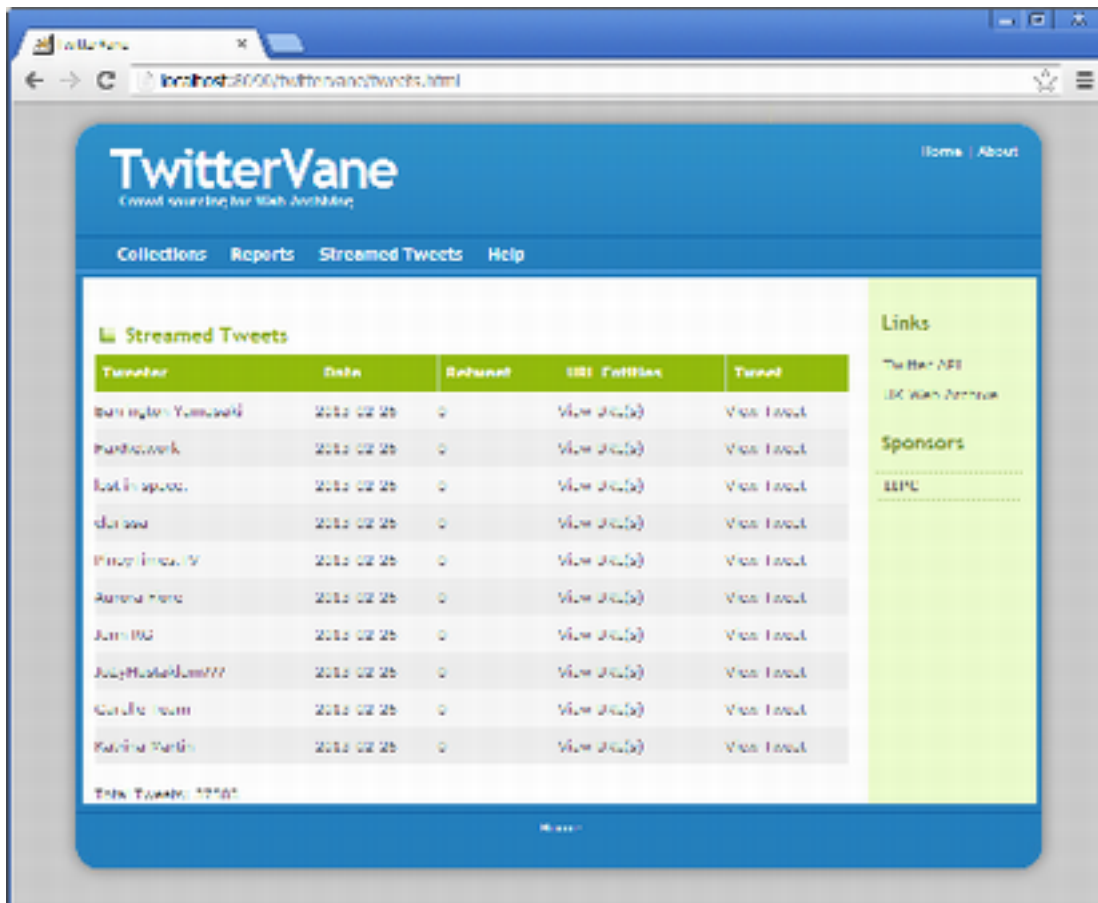
URI	Collection	Tweets	Estimate
<a href="http://www.fox.com">http://www.fox.com</a>	Search Engine	240	0
<a href="http://dubai.fox.com/pug.fox">http://dubai.fox.com/pug.fox</a>	TV	187	0
<a href="http://www.fox.com/pug.fox">http://www.fox.com/pug.fox</a>	My Search	157	0
	Total Collection	0	0
<a href="http://dubai.fox.com/pug.fox">http://dubai.fox.com/pug.fox</a>	UNKNOWN	50	0
<b>TOTAL</b>		<b>714</b>	<b>0</b>

On the right side of the report, there are two sections: "Links" and "Sponsors". The "Links" section contains "Twitter API" and "UK Web Archive". The "Sponsors" section contains "LISC".

At the bottom of the application, there is a "Home" link.



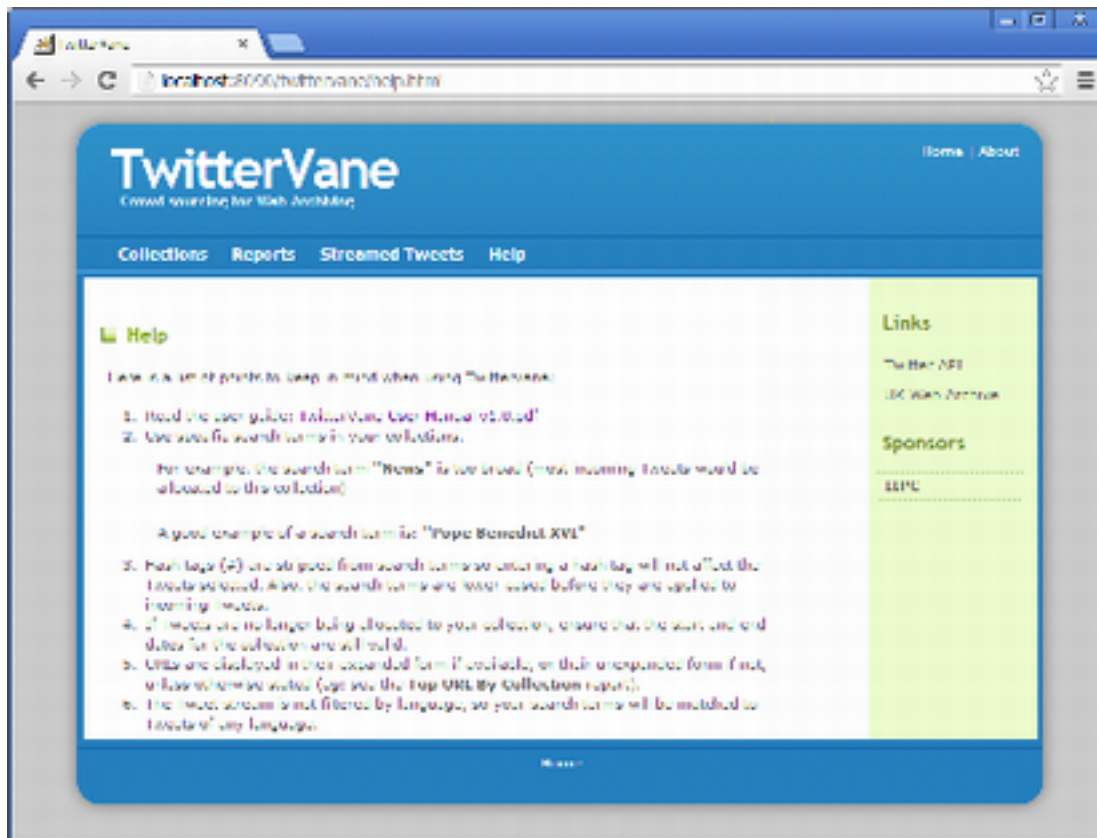
The **Stream Tweets** menu item displays the latest Tweets received from Twitter:



The Tweets displayed are those matching any of the search terms for any of the web collections that have been defined. This view is useful for checking that the Tweet stream is receiving Tweets and to assess how quickly Tweets are being received.



The **Help** menu item displays the online help for TwitterVane:



There is a link to the PDF version of this user manual within the first item of the online help.