TwitterVane

User Manual

05 February 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....7
Reports....14
Criteria driven reports....14
Browse Reports....17

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 though 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

System Overview

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

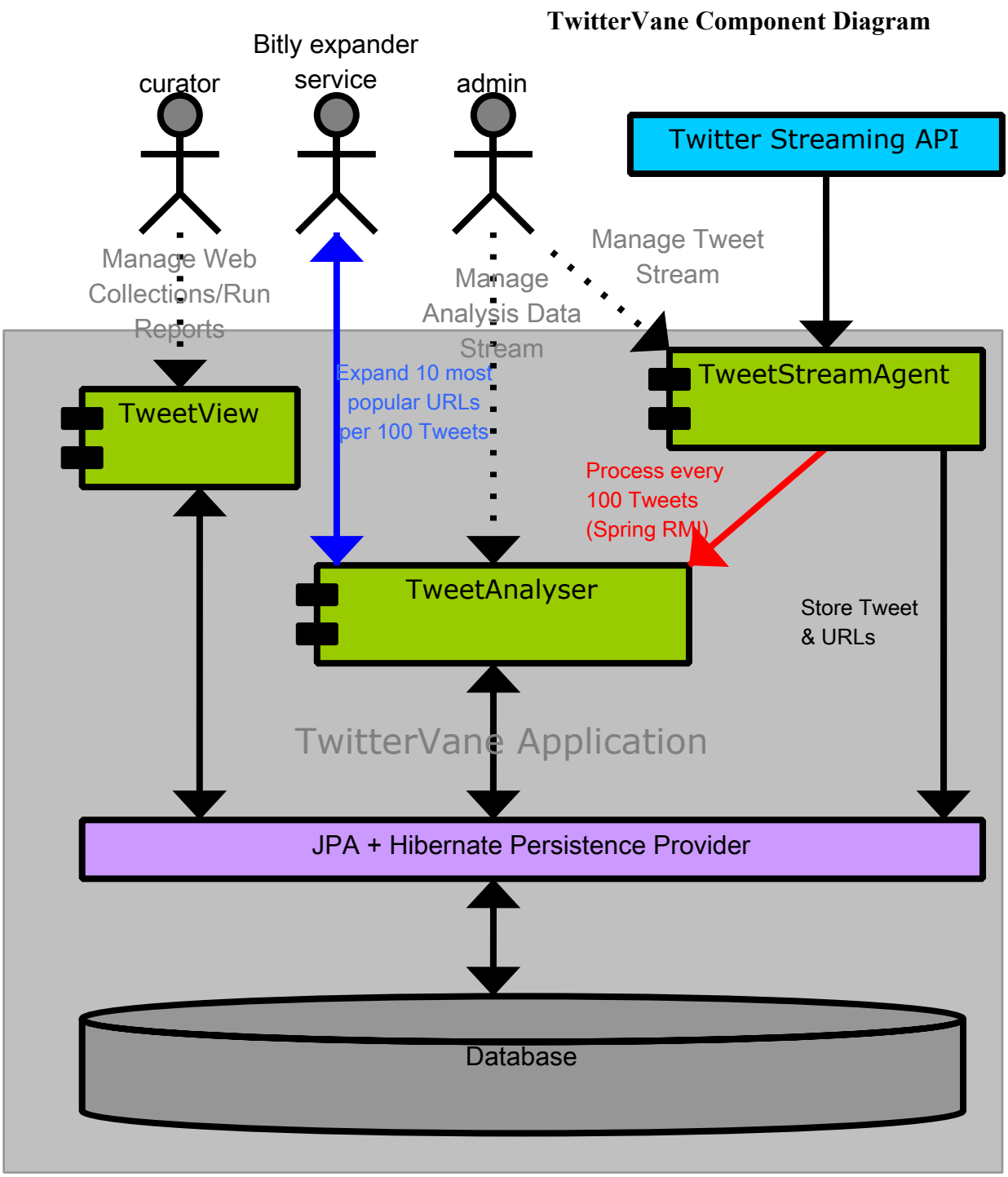
TweetView is the curator web interface where users define the collections and search terms that will be used to collect Tweets. A series of reports is also provided to identify popular Tweets and top URLs by collection.

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.

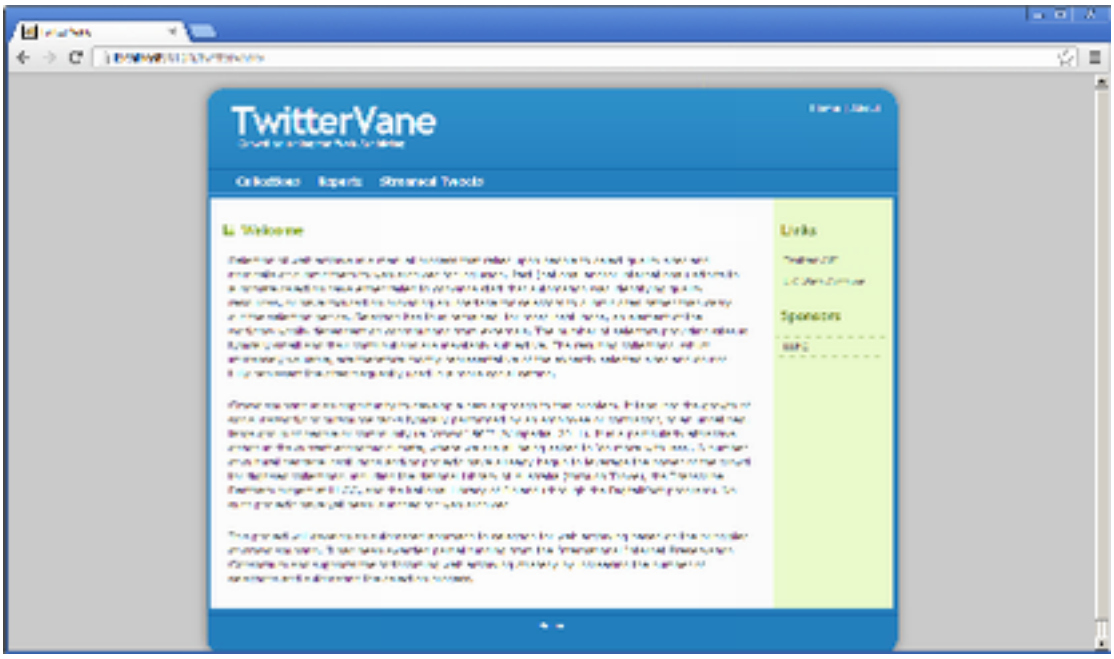
TweetAnalyser

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
curator
Bitly expander service
admin
Twitter Streaming API
Manage Web Collections/Run Reports
Manage Analysis Data Stream
Manage Tweet Stream
TweetView
Expand 10 most popular URLs per 100 Tweets
TweetStreamAgent
Process every 100 Tweets (Spring RMI)
TweetAnalyser
Store Tweet & URLs
TwitterVane Application
JPA + Hibernate Persistence Provider
Database

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:

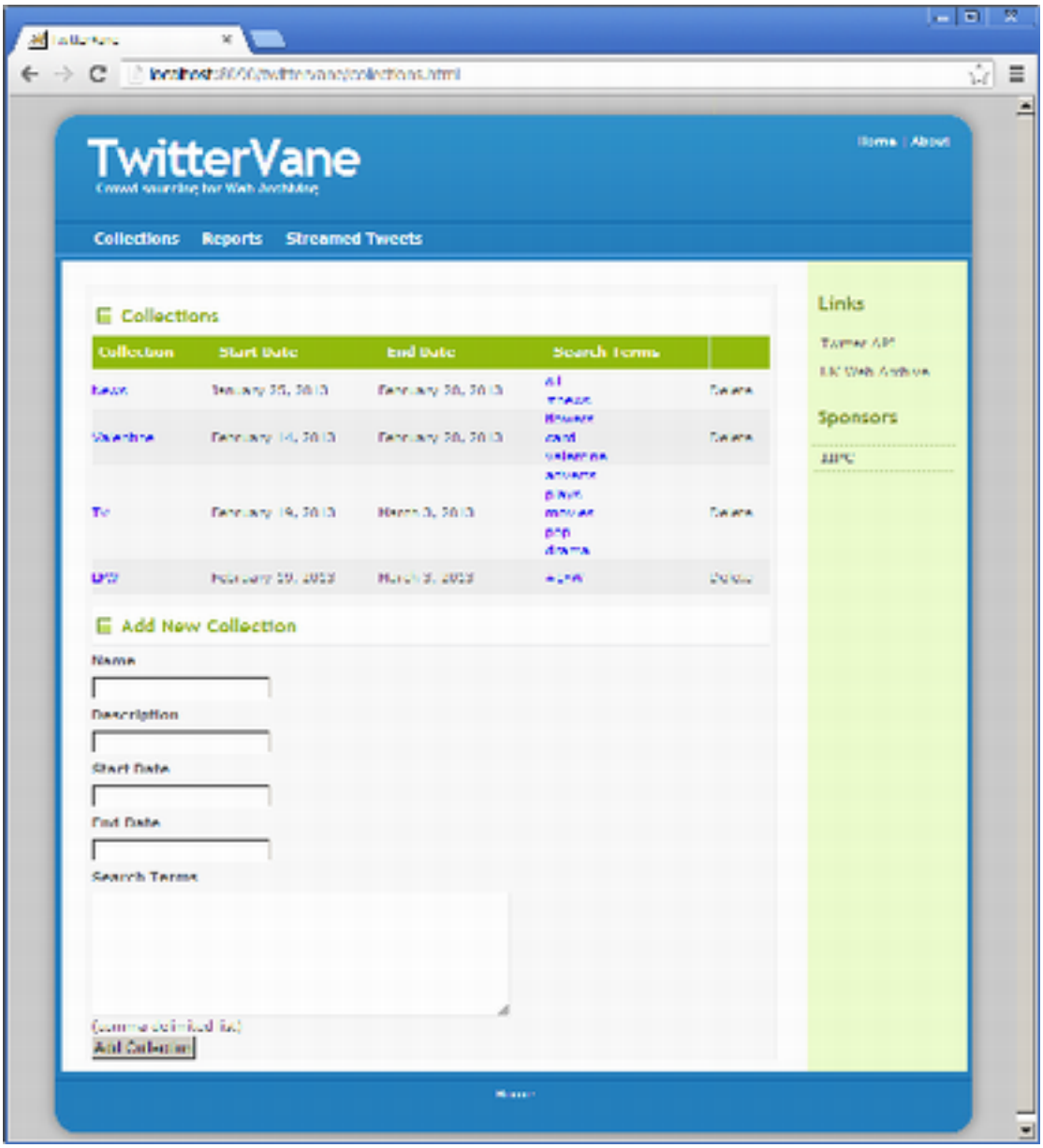
Collections – create, view and manage web collections and associated search terms

Reports – view and run the criteria based reports and 'browse reports'

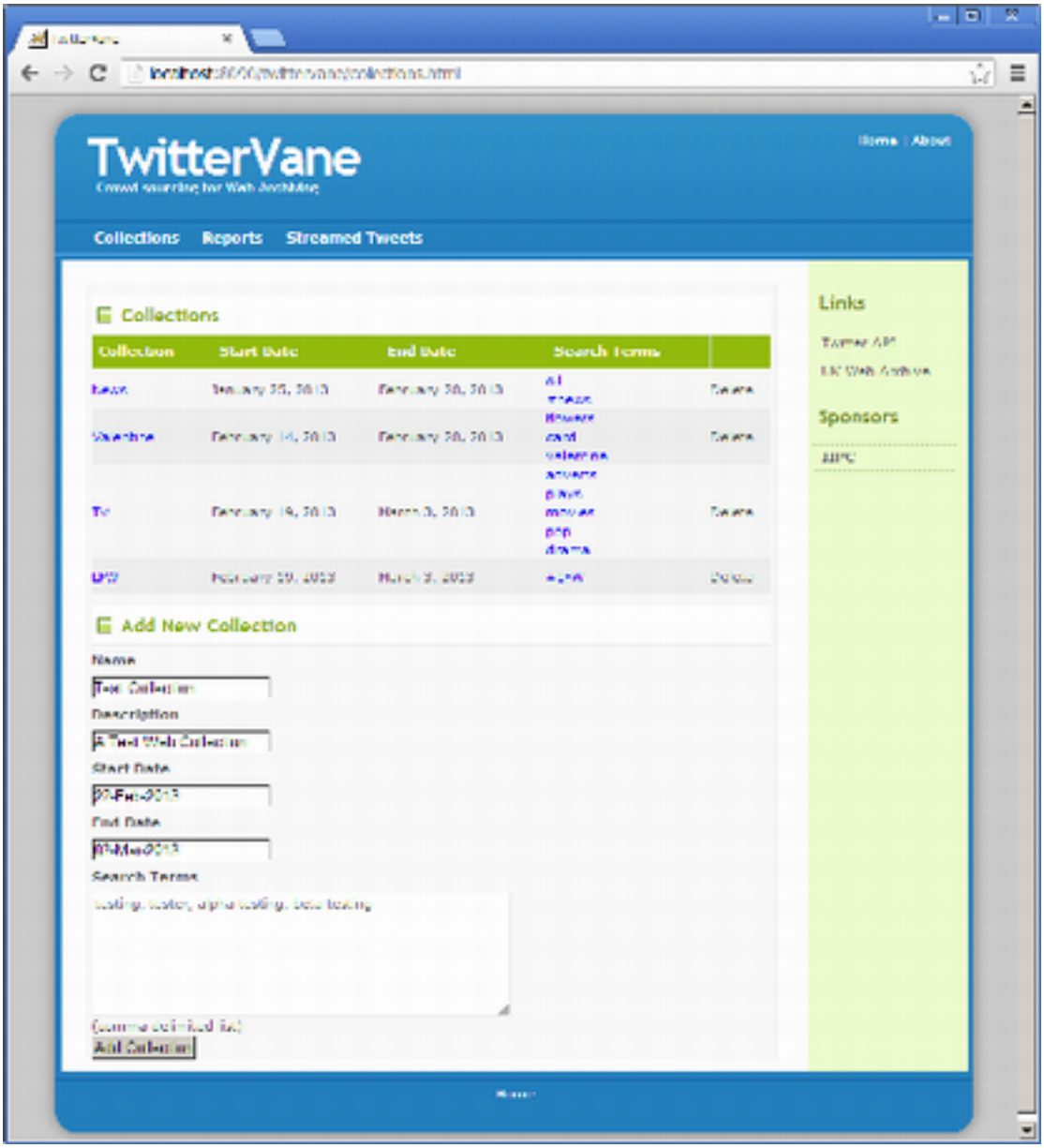
Streamed Tweets – view the latest Tweets received from the Twitter

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

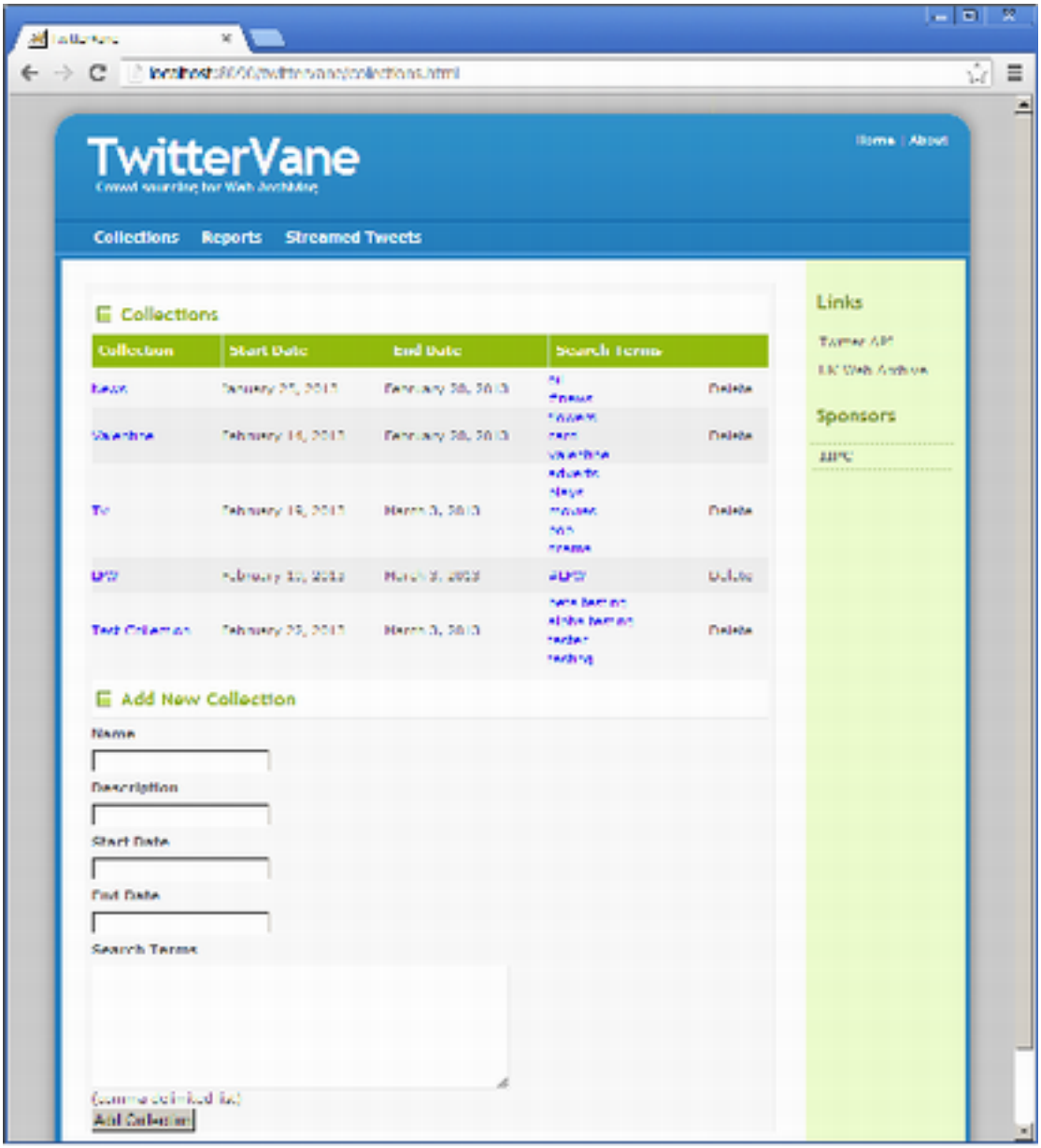
The **Collections** page displays the Web Collections that are currently defined on the system, the search terms associated with the collection, and the start and end dates within which Tweets should be collected:



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’.

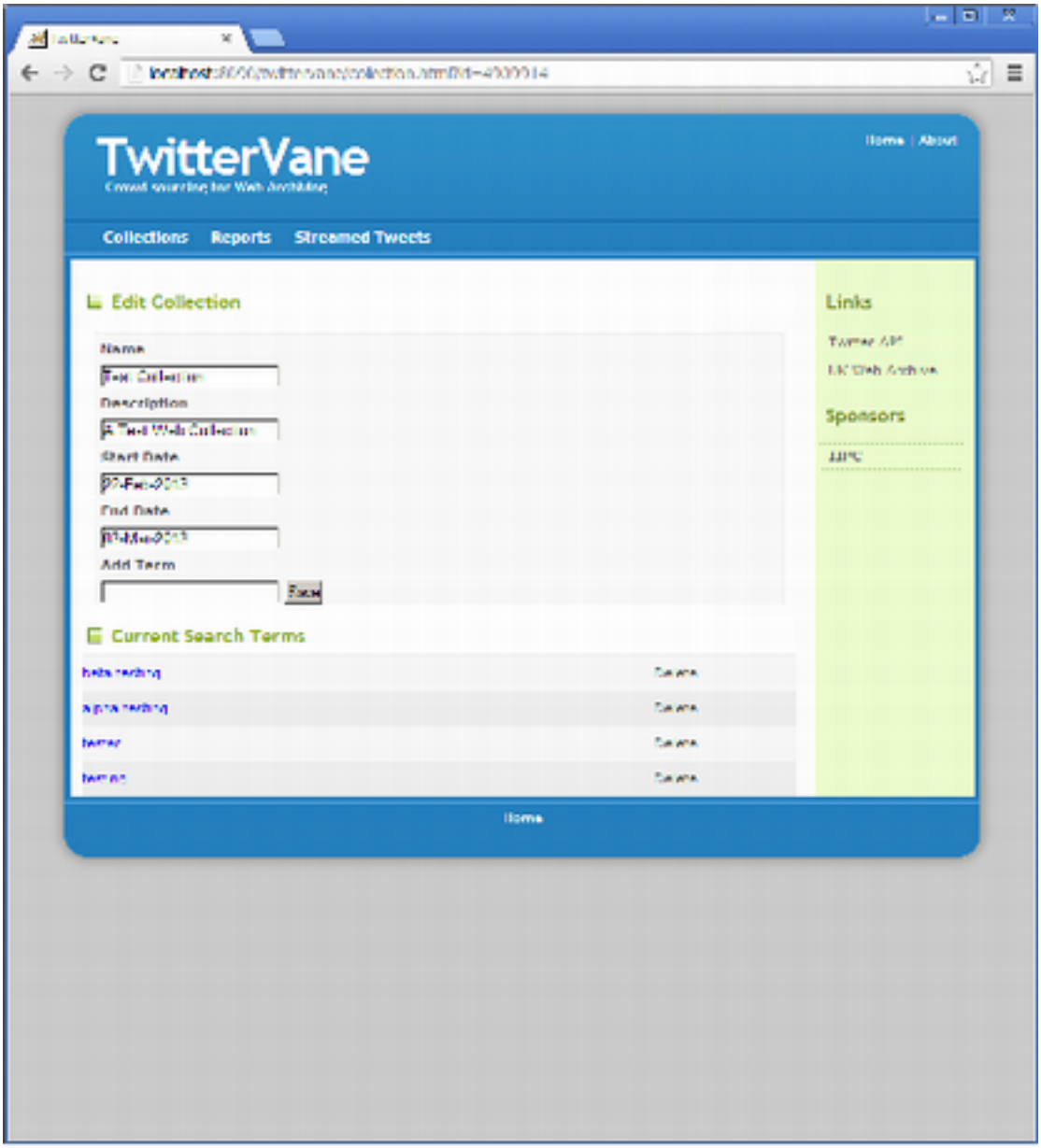


The collection then appears in the list:

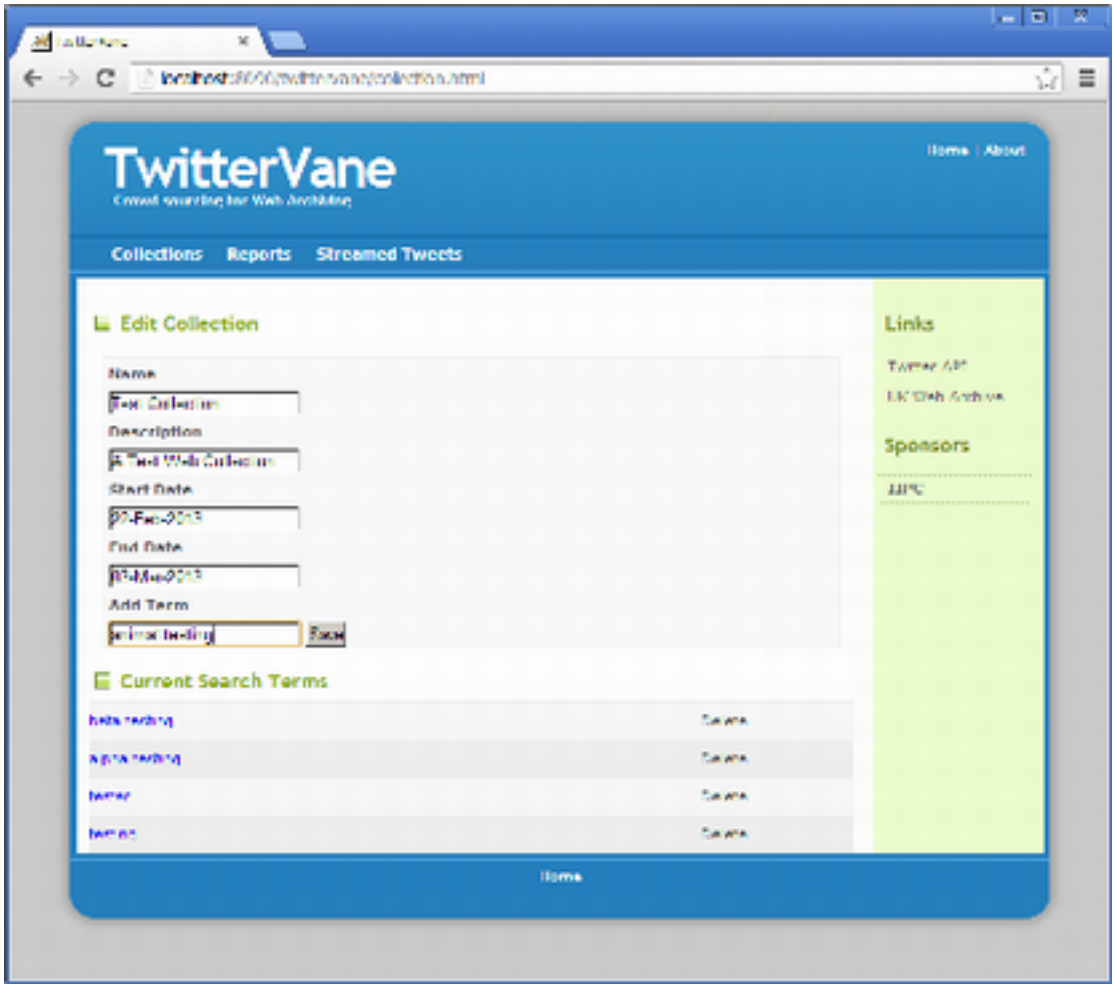


To manage a collection and associated search terms, click on the collection name:

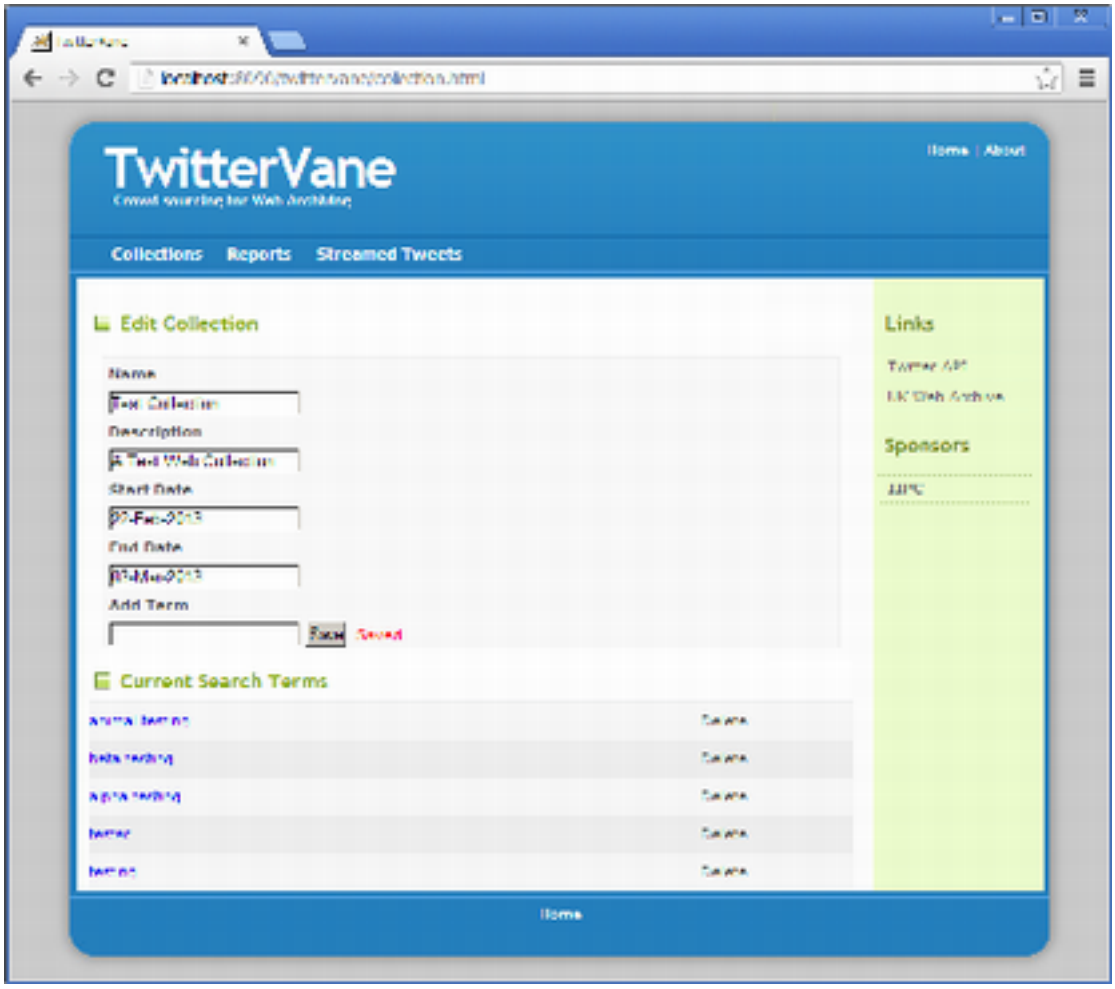
The **Edit Collection** page is then displayed where the collection details can be modified:



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

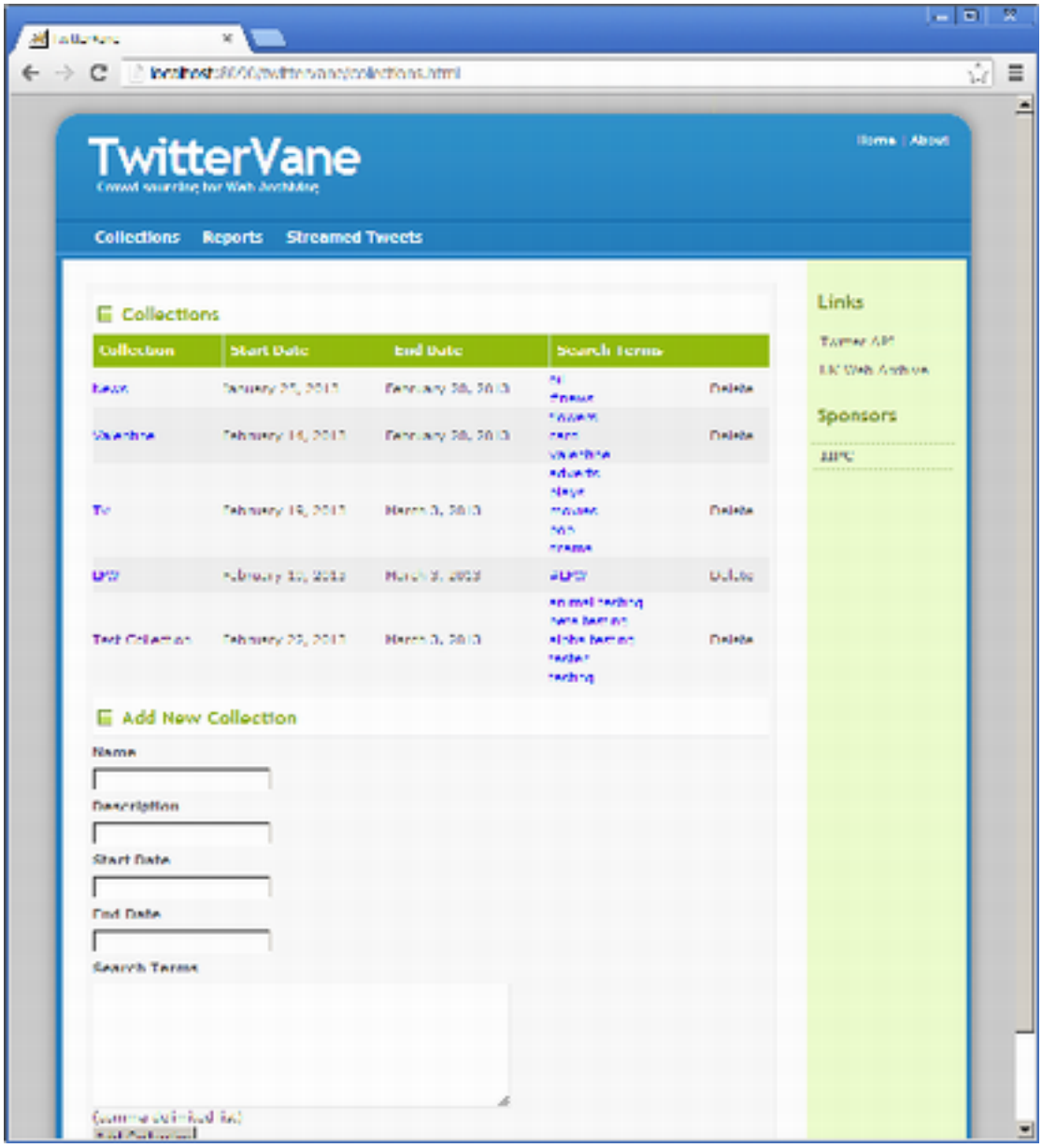


The search term is then listed at the bottom of the page:



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:

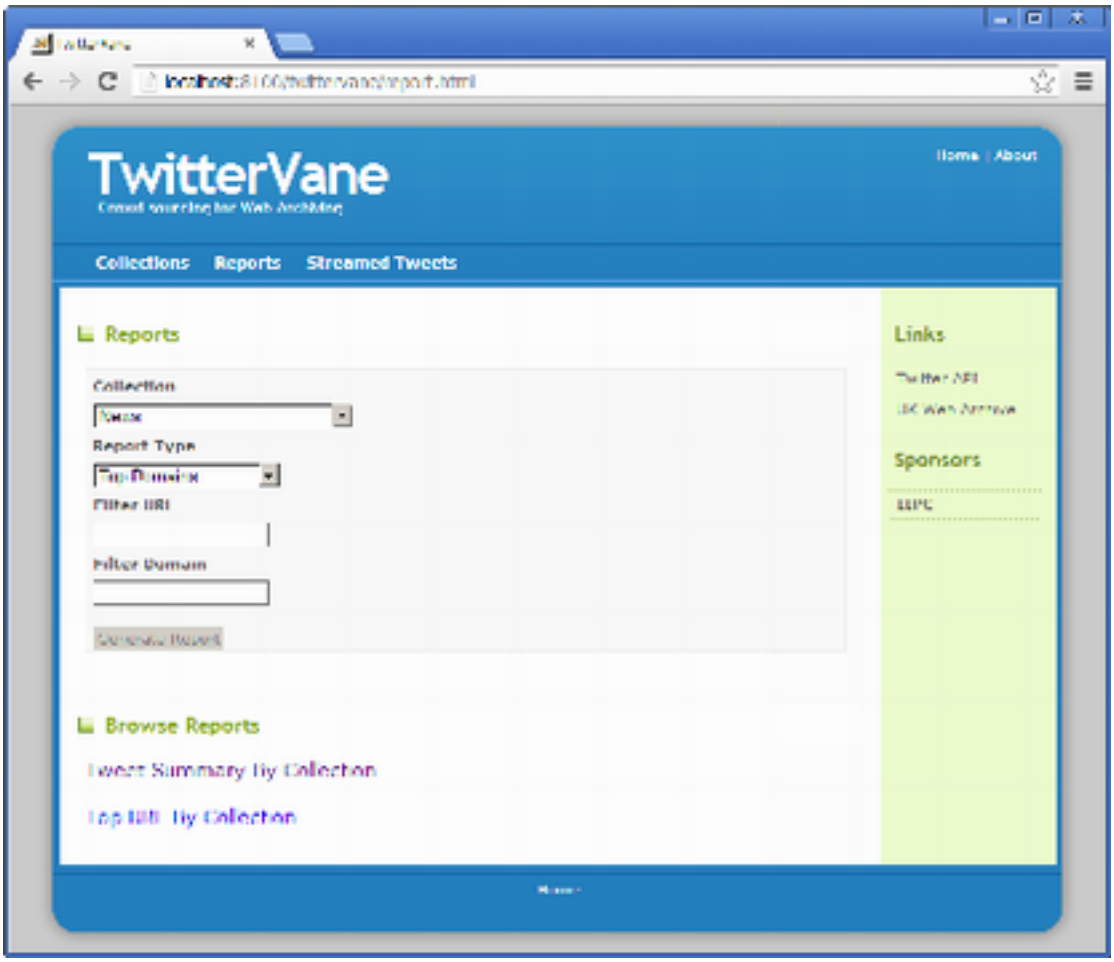


WARNING: Deleting a web collection will remove all associated search terms, Tweets and URLs from the database.

Reports

There are two types of report available in TwitterVane:

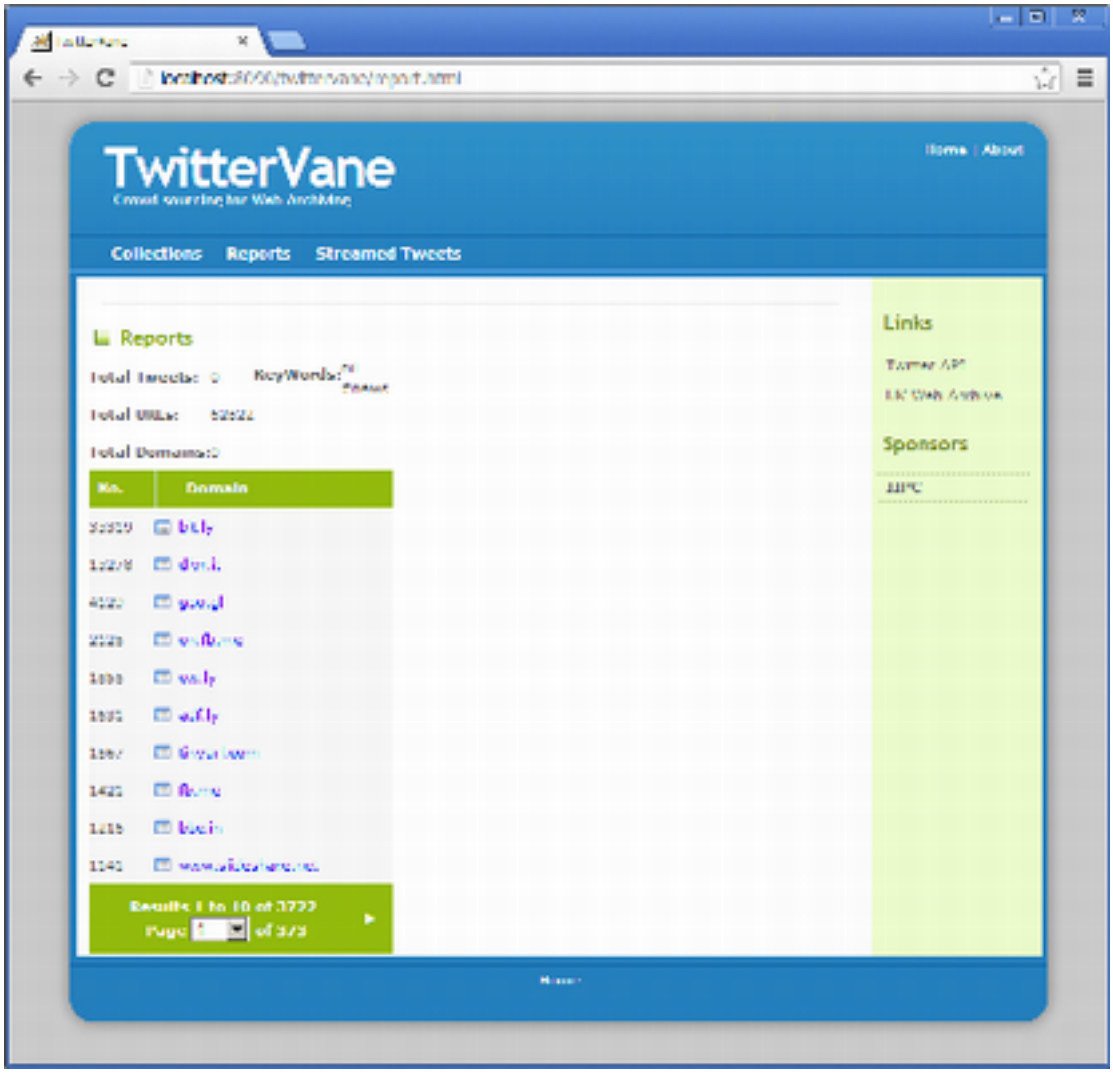
1. Criteria driven reports
2. 'Browse Reports'



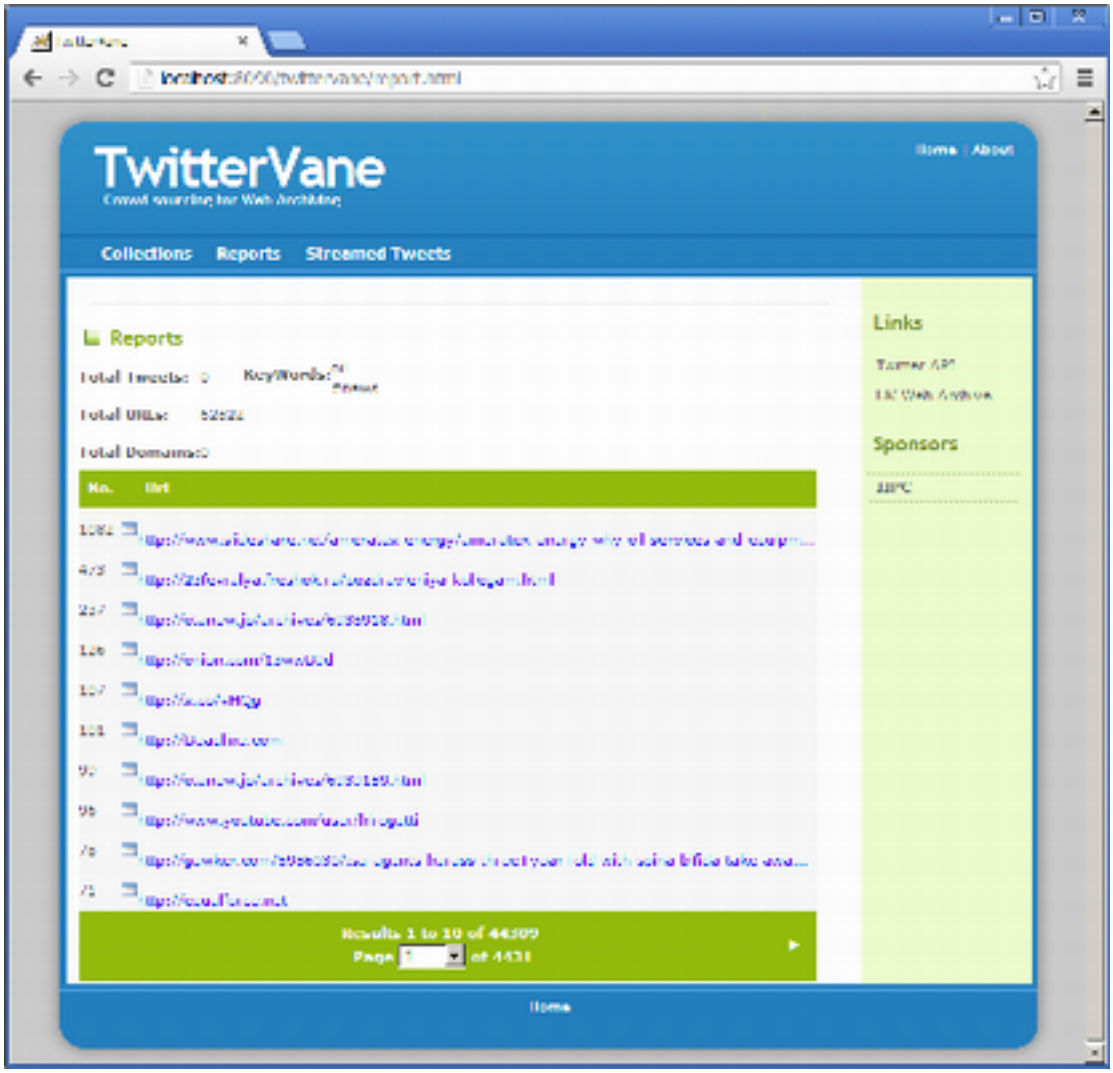
Criteria driven reports

TwitterVane has four criteria reports defined:

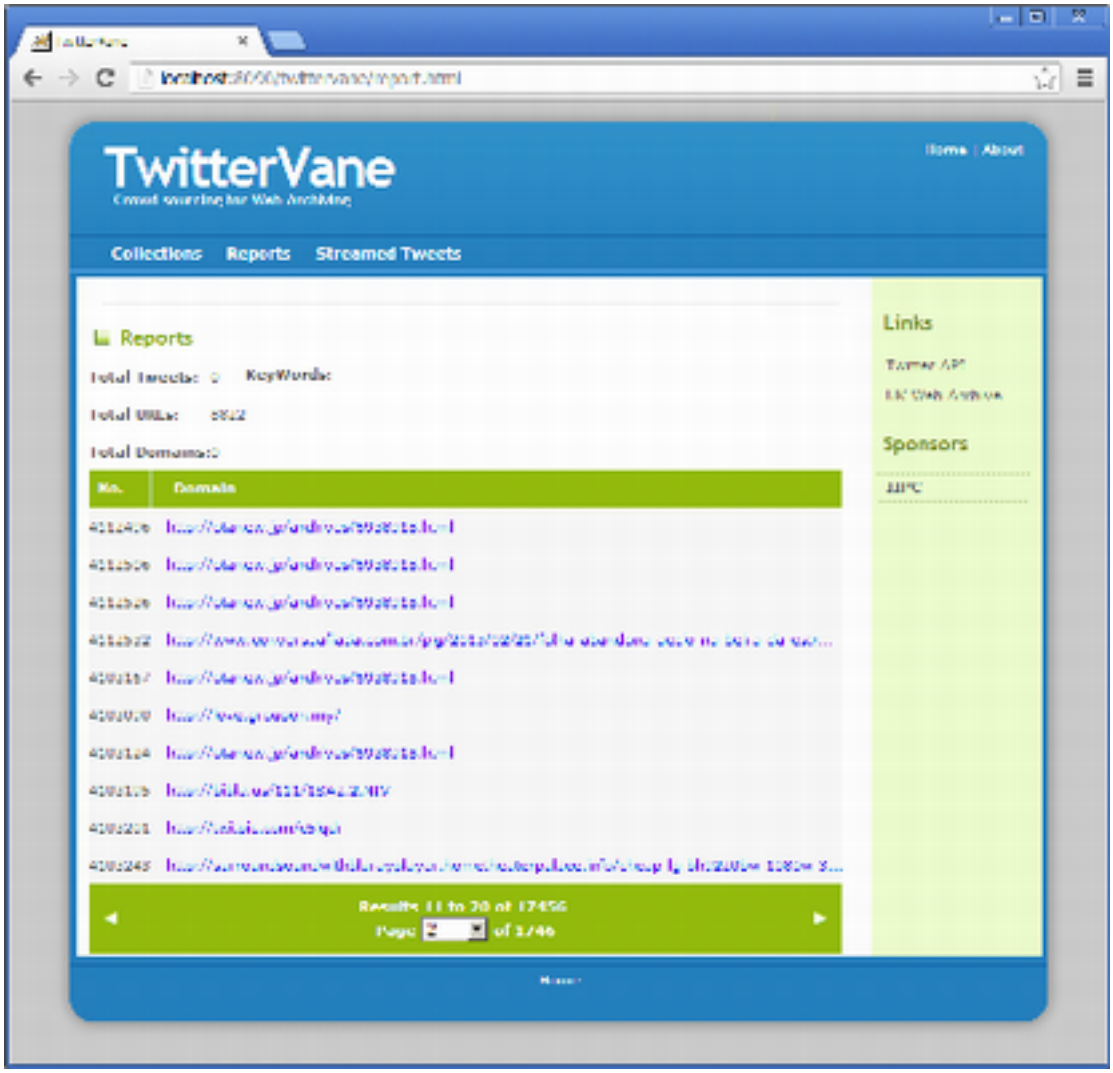
1. **Top Domains** – returns the first 100 domains containing the greatest number of URLs for the selected Web Collection. The results are sorted on the number of URLs in descending order



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



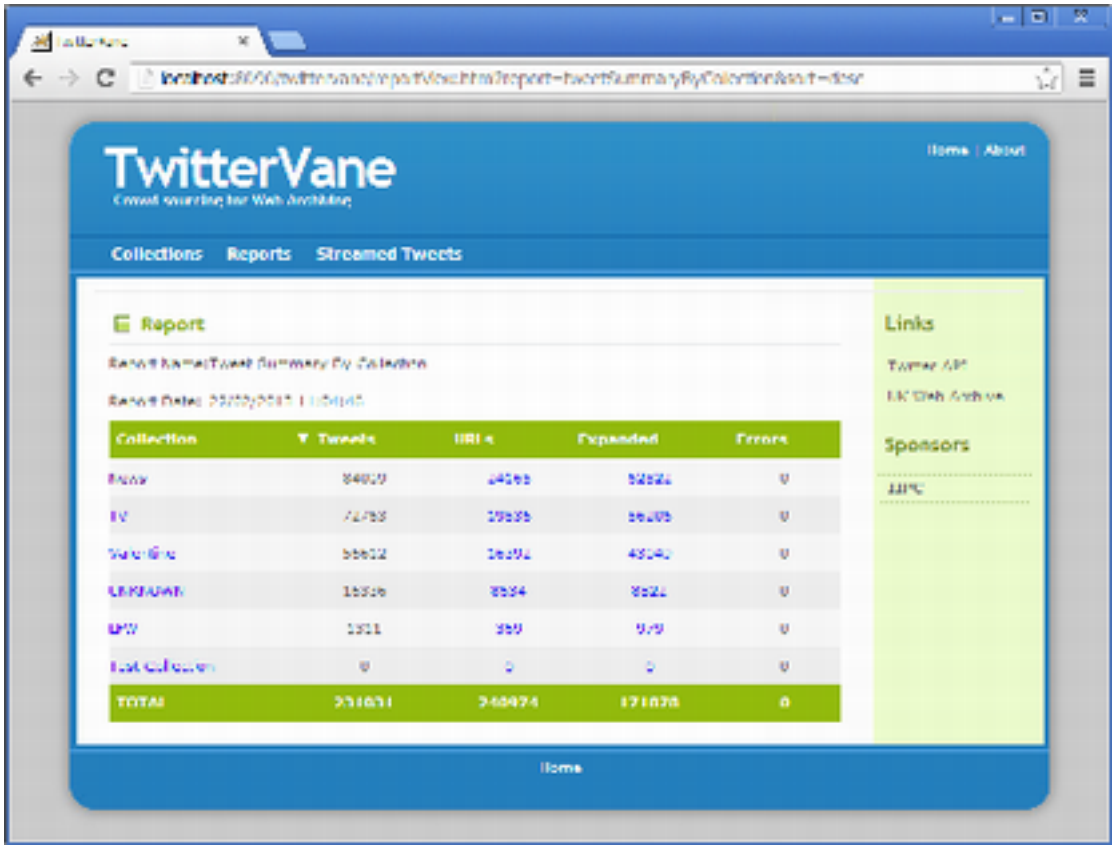
3. **Top URL by Retweet** - returns the most popular expanded URL for the selected Web Collection. Popularity is re-tweet count of the Tweet received from Twitter.
4. **Failed Analysis** - returns all URLs that encountered errors when analysed.



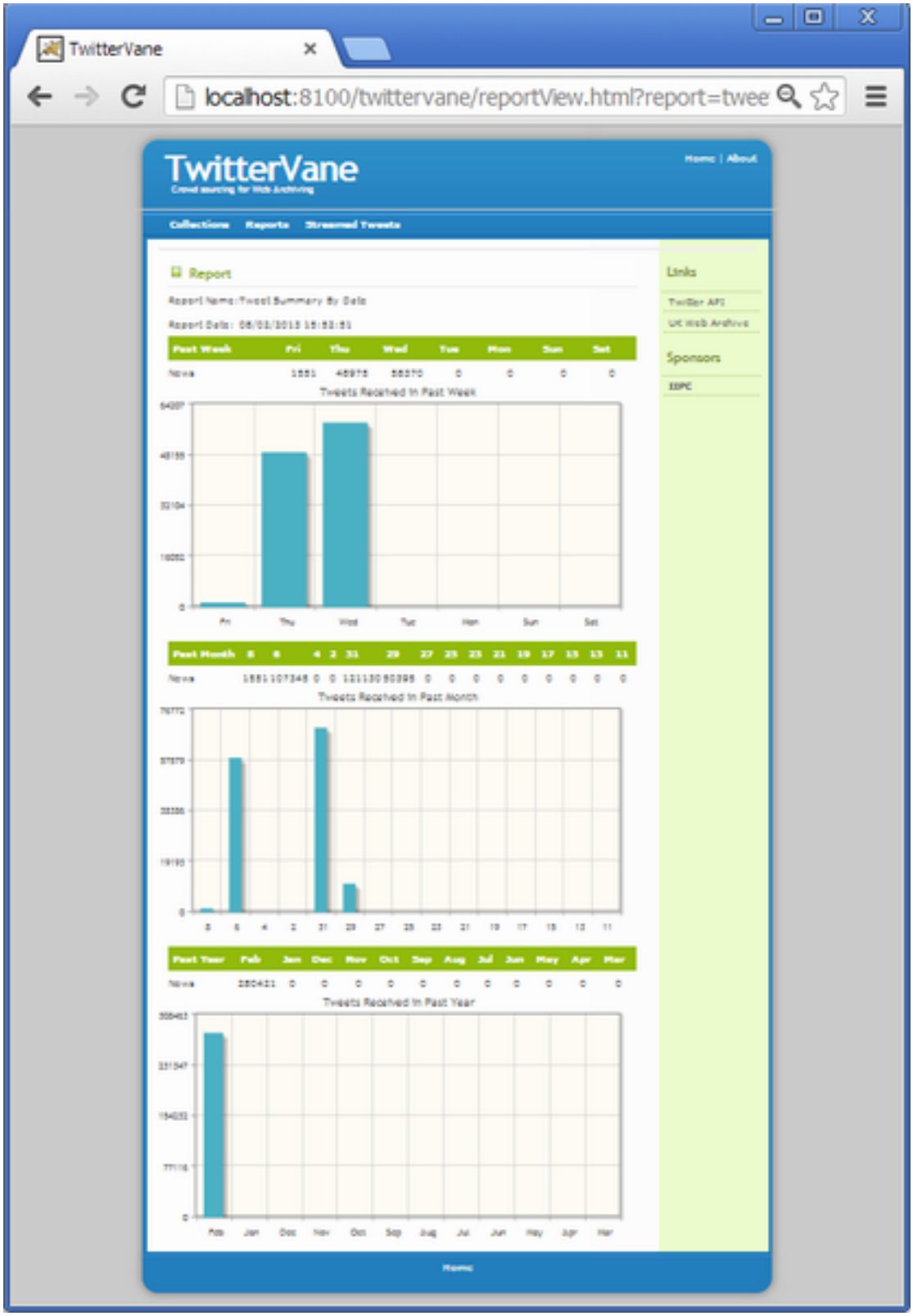
Browse Reports

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 collection name 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.

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.

TwitterVane
Home | About
Collections
Reports
Streamed Tweets
Report
Links
Twitter API
UK Web Archive
Sponsors
URL
Collection
Tweets
Retweets
Home
TOTAL