15400 Calhoun Drive, Suite 400
Rockville, Maryland, 20855
(301) 294-5200
http://www.i-a-i.com

Intelligent Automation Incorporated

Enhancements for a Dynamic Data Warehousing and Mining System for Large-scale HSCB Data

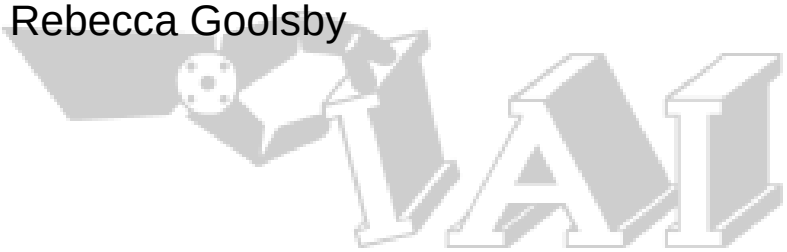
Progress Report No. 3

Reporting Period: May 21, 2016 – June 20, 2016

Contract No. N00014-16-P-3014

Sponsored by
ONR, Arlington VA
C⊙TR/TP⊙C: Dr. Rebecca Goolsby

Prepared by
Onur Savas, Ph.D.



DISTRIBUTION A

Approved for public release; distribution is unlimited.

Progress Report No. 3

Enhancements for a Dynamic Data Warehousing and Mining System Large-Scale HSCB Data

Submitted in accordance with requirements of
Contract #N00014-16-P-3014

Performance period: May 21, 2016 to June 20, 2016
(PI: Dr. Onur Savas, 301.294.4241, osavas@i-a-i.com)

1 Work Performed within This Reporting Period

In this reporting period, we performed the following tasks.

- **Developed automated YouTube data collection capabilities.** We have developed automated YouTube video metadata collection capabilities and released it part of Scraawl. In particular, we have developed two APIs for continuous (Streaming) and recent YouTube data searches, and designed a user-friendly UI for these searches using YouTube Data API v3 [1].

- **Released Scraawl 1.15.**

1.1 YouTube Data Collection

1.1.1 Scraawl YouTube Data Collection API Development

The first API allows for streaming searches, i.e., adds a query for continuous YouTube post collection. The second API allows searching on recent YouTube videos, and the resulting data will be saved in the configured database. Both APIs make use of the (i) `Search:list` Data API functionality, which returns a collection of search results that match the query parameters specified in the API request, and (ii) `Videos: list` Data API functionality, which returns a list of videos that match the API request parameters. Some of the major parameters used in the Scraawl streaming and recent API searches is shown in Table 1.

Table 1: Scraawl API Major YouTube Data Collection Parameters.

Parameter	Explanation
query.q	The `q` parameter specifies the query term to search for. Two words separated by spaces can be treated as AND. Your request can also use the Boolean NOT (-) and OR (|) operators to exclude videos or to find videos that are associated with one of several search terms.
query.channelId	The `channelId` parameter indicates that the API response should only contain resources created by the channel.
query.location & query.locationRadius	The `location` parameter, in conjunction with the `locationRadius` parameter, defines a circular geographic area and also restricts a search to videos that specify, in their metadata, a geographic location that falls within that area.
query.publishedBefore	The `publishedBefore` parameter indicates that the API response should only contain resources created before the specified time.
query.publishedAfter	The `publishedBefore` parameter indicates that the API response should only contain resources created before the specified time.

1.1.2 Scraawl UI Development for YouTube Searches

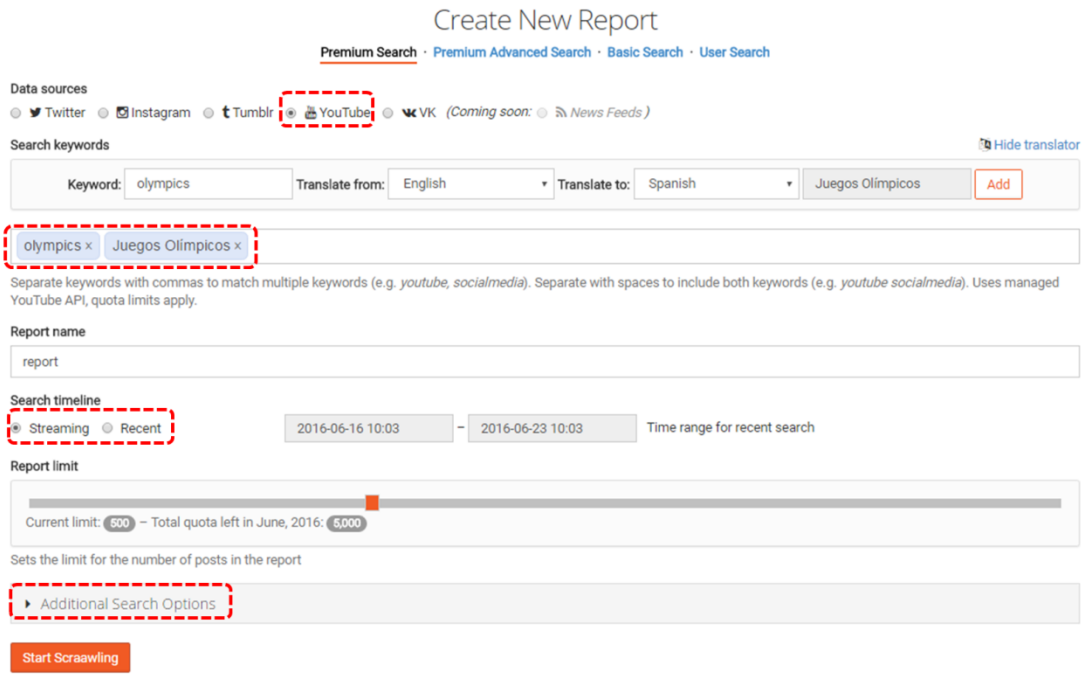


Figure 1: UI for YouTube Searches.

We have also developed a UI to use the above APIs seamlessly. A representative UI is shown in Figure 1. The UI has the same look and feel with other social media searches,

and can be accessed from “Create New Report” view under Scraawl. The UI allows to specify keywords with each box “AND”ed, and the translation capability using Google translate is integrated. The user can also choose between “Streaming” and “Recent” searches, which in turn calls one of the above APIs explained in Section 1.1.1. When “Recent” search is selected, the user has the option to select a time range. When “Streaming” search is selected, data collection will continue until a pre-specified time-out or the data collection limit is reached. Similar to other data feeds, we also allow the user to draw circular bubbles on the world map to restrict their searches to certain region(s) under “Additional Search Options.”

2 Current Problems

None.

3 Work to be Performed in the Next Reporting Period

In the next report period, we will focus on the following tasks:

- We will mature Scraawl basic statistics for YouTube data.
- We will deliver Scraawl 1.16.

References

[1] YouTube Data API, https://developers.google.com/youtube/v3/.