CS 751: Assignment 2

Rohit Lambi

Spring 2015

Contents

1 Q1

1. Choose 100 URIs from A1

2. Generate WARC files of those URIs using: - wget - WARCreate - Heritrix (standalone or via WAIL) - webrecorder.io

3. Describe the resulting WARC files: quantitatively compare and contrast the results of the WARC files of the same URI as generated by different tools - choose interesting examples

4. Demonstrate playback of 2-3 WARCs in the (Wayback Machine (via WAIL or stand-alone) or pywb) and (webrecorder.io) - https://github.com/iipc/openwayback - https://github.com/ikreymer/pywb

1.1 Description

1. The WARC files contain all the request headers, response headers and the html to render the recorded webpage.

2. The WARC file size generated by wget is the least amongst all the tools.

3. For few URLs, WARCreate did not generate WARC file despite of trying multiple times.

4. I replayed different WARC files using webrecorder.io and included the screenshots in this report

5. I was having problem with pywb and hence was not able to replay the WARC files on it.
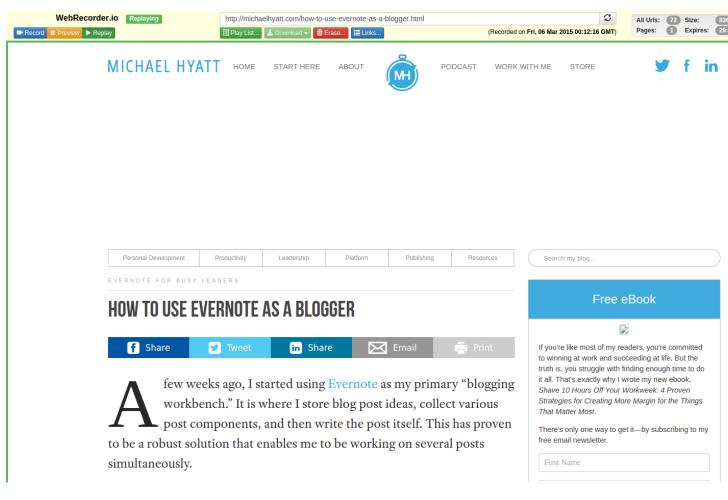
Replay of Different WARC files on webrecorder.io
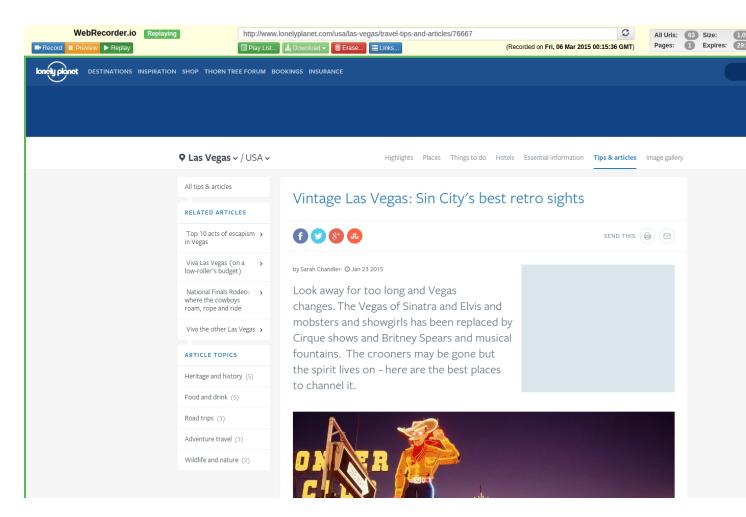

Figure 1:

Home   Services   About us   How it Works   FAQs   Contacts   Login   Experts   Questions

Live Chat Help

My Assignmenthelp
+61-2-6100-3843

Assignment Help| Dissertation Help | Essay Help | Online Writing Services My Assignment Help
.com's Outstanding Features: »Plagiairsed Free » Unmatched Quality » Exact Referenced »
Affordable Prices » Ontime delivery

You are here: Home / Assignment Help / Psychology Assignment Help / 10 Tips to Manage Exam Stress

10 Tips to Manage Exam Stress

January 17, 2015 by admin

Leave a Comment

Filed under Psychology Assignment Help

Pressure of scoring high and doing well in examination can get the better of many students. They complain of a host of physical and mental problems while preparing for their examinations. Such problems range from the benign to the serious and can have disastrous consequences for the students if not dealt properly. Nonetheless, we can deal with it with a firm hand and loving support from our near and dear ones.

Place New Order

Email

Subject

Deadline   5 Mar 2015

Pages   -   1   +
250 words

Continue

Educational Links

⭐ Assignment Help
⭐ Assignment Help UK
⭐ Expert Assignment Writers
⭐ Question Samples

Follow Us!

Figure 2:

lonely planet   DESTINATIONS   INSPIRATION   SHOP   THORN TREE FORUM   BOOKINGS   INSURANCE

📍 **Las Vegas** ∨ / USA ∨   Highlights   Places   Things to do   Hotels   Essential information   **Tips & articles**   Image gallery

All tips & articles

RELATED ARTICLES

Top 10 acts of escapism in Vegas >

Viva Las Vegas (on a low-roller's budget) >

National Finals Rodeo: where the cowboys roam, rope and ride >

Viva the other Las Vegas >

ARTICLE TOPICS

Heritage and history (5)

Food and drink (5)

Road trips (3)

Adventure travel (3)

Wildlife and nature (2)

Vintage Las Vegas: Sin City's best retro sights

SEND THIS

by Sarah Chandler· ⊙ Jan 23 2015

Look away for too long and Vegas changes. The Vegas of Sinatra and Elvis and mobsters and showgirls has been replaced by Cirque shows and Britney Spears and musical fountains. The crooners may be gone but the spirit lives on – here are the best places to channel it.

Figure 3:

2 Q2

1. Ingest the 100 URIs from their resulting WARC files into a SOLR instance - see the code + tutorial at: https://github.com/ukwa/webarchive-discovery

2. Demonstrate several functioning queries on the files (a full front-end is not required) - describe the configuration choices you made in setting up SOLR and processing the documents

2.1 Description

1. I configured a local SOLR instance and indexed 100 WARC files in it.

2. I used default configurations for SOLR.

3. I performed various queries on SOLR and included the resulting screenshots in this report.

Functioning queries on SOLR

Figure 4: Search for 'football'

Figure 5: Search for 'football news'
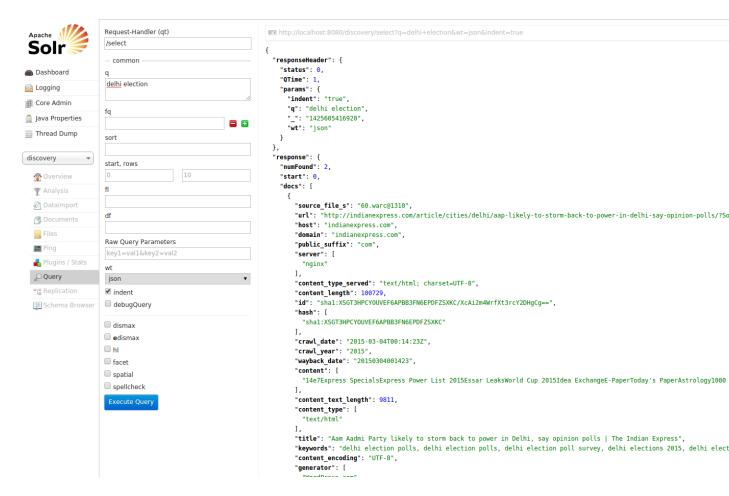
Figure 6: Search for 'medical'

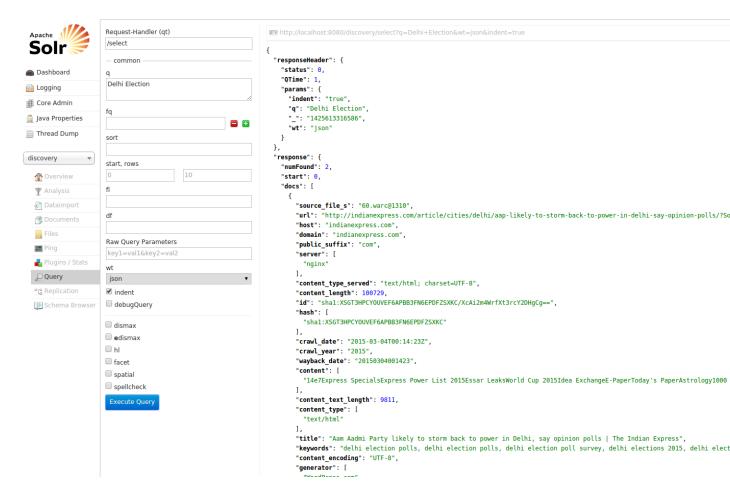Figure 7: Search for 'delhi election'

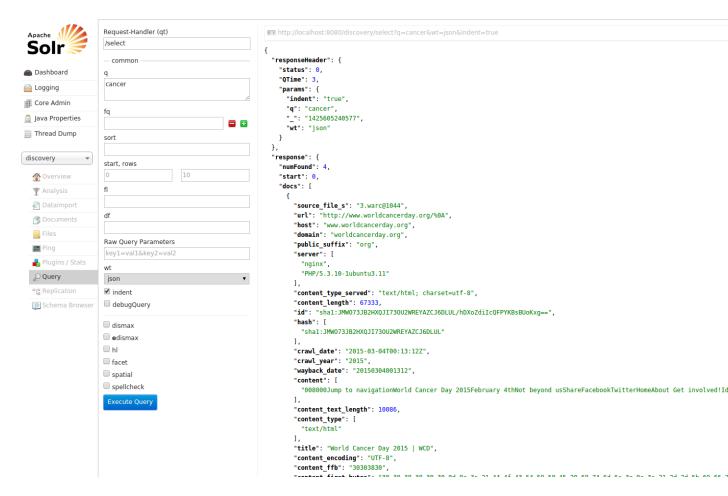Figure 8: Search for 'Delhi Election' using different case than previous query

Figure 9: Search for 'cancer'

3 References

[1] OpenWayback - https://github.com/iipc/openwayback

[2] Python WayBack for web archive replay, web proxy - https://github.com/ikreymer/pywb

[3] WARC and ARC indexing and discovery tool - https://github.com/ukwa/webarchive-discovery