

CS 751: Assignment 2

Rohit Lambi

Spring 2015

Contents

1	Q1	2
	1.1 Description	2
2	Q2	6
	2.1 Description	6
3	References	13

1 Q1

1. Choose 100 URIs from A1
2. Generate WARC files of those URIs using: - wget - WARCcreate - 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, WARCcreate 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

WebRecorder.io **Replaying** <http://michaelhyatt.com/how-to-use-evernote-as-a-blogger.html> (Recorded on Fri, 06 Mar 2015 00:12:16 GMT)

Record Preview Replay Play List... Download Erase... Links...

MICHAEL HYATT HOME START HERE ABOUT MH PODCAST WORK WITH ME STORE

Personal Development Productivity Leadership Platform Publishing Resources

Search my blog...

HOW TO USE EVERNOTE AS A BLOGGER

Share Tweet Share Email Print

A few weeks ago, I started using [Evernote](#) as my primary “blogging workbench.” It is where I store blog post ideas, collect various post components, and then write the post itself. This has proven to be a robust solution that enables me to be working on several posts simultaneously.

Free eBook

If you're like most of my readers, you're committed to winning at work and succeeding at life. But the truth is, you struggle with finding enough time to do it all. That's exactly why I wrote my new ebook, *Shave 10 Hours Off Your Workweek: 4 Proven Strategies for Creating More Margin for the Things That Matter Most*.

There's only one way to get it—by subscribing to my free email newsletter.

First Name

Figure 1:

WebRecorder.io

Replaying

[Record](#)
[Preview](#)
[Replay](#)

[Play List...](#)
[Download](#)
[Erase...](#)
[Links...](#)

<http://myassignmenthelp.com/blog/10-tips-to-manage-exam-stress>

(Recorded on Fri, 06 Mar 2015 00:11:37 GMT)

All Urls: 58

Size: 2.1

Pages: 1

Expires: 22

Home

Services

About us

How it Works

FAQs

Contacts

Login

Experts

Questions

Live Chat Help

MyAssignmenthelp

+61-2-6100-3843

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

1. Make a timetable
2. Sleep well, eat right
3. See your priorities straight
4. Listen to music
5. Don't mug up
6. Take a break
7. Set realistic goals
8. Avoid annoying people
9. Practice relaxing techniques
10. Talk to a shrink

Place New Order

Email
 Subject
 Deadline 5 Mar 2015
 Pages 250 words
[Continue](#)

Educational Links

- [Assignment Help](#)
- [Assignment Help UK](#)
- [Expert Assignment Writers](#)
- [Question Samples](#)

Follow Us!

Figure 2:



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

The screenshot displays the Apache Solr Admin UI interface. On the left is a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and a dropdown menu for 'discovery' containing Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (selected), Replication, and Schema Browser.

The main panel is titled 'Request-Handler (qt)' and contains the following fields and controls:


- /select**: The request handler.
- common**: A dropdown menu.
- q**: The query, set to 'football'.
- fq**: Filter query field.
- sort**: Sort field.
- start, rows**: Range from 0 to 10.
- fl**: Fields to list.
- df**: Default field.
- Raw Query Parameters**: A text area containing 'key1=val1&key2=val2'.
- wt**: Web toolkit, set to 'json'.
- indent**: Checked checkbox.
- debugQuery**: Unchecked checkbox.
- Execute Query**: A blue button.

Below these fields are several unchecked checkboxes: dismax, edismax, hl, facet, spatial, and spellcheck.

On the right, the URL bar shows the request: `http://localhost:8080/discovery/select?q=football&wt=json&indent=true`. The response is a JSON object:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "indent": "true",
      "q": "football",
      "_": "1425605144045",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 7,
    "start": 0,
    "docs": [
      {
        "source_file_s": "93.warc@1309",
        "url": "http://www.telegraph.co.uk/sport/football/football-transfers/11385515/Football-Transfer-Window-2015-The-Big-Moves-Telegraph",
        "content_type_ext": "html0a",
        "host": "www.telegraph.co.uk",
        "domain": "telegraph.co.uk",
        "public_suffix": "co.uk",
        "content_type_served": "text/html;charset=UTF-8",
        "content_length": 144612,
        "id": "sha1:AI72QYJHTDQGCLTGPC5GNKLY5ERR7K6A/r4x85kS0SdVlE6pTcUDNUA==",
        "hash": [
          "sha1:AI72QYJHTDQGCLTGPC5GNKLY5ERR7K6A"
        ],
        "crawl_date": "2015-03-04T00:14:58Z",
        "crawl_year": "2015",
        "wayback_date": "20150304001458",
        "content": [
          "00009066Accessibility linksSkip to navigationAdvertisementTelegraph.co.ukSearch - enhanced by OpenTe"
        ],
        "content_text_length": 7708,
        "content_type": [
          "text/html"
        ],
        "title": "Video: Football Transfer Window 2015 - The Big Moves - Telegraph",
        "keywords": "transfer window, the big moves, animation, watch, video, football, transfer news,transfer gossip, winter",
        "content_encoding": "UTF-8",
        "content_ffb": "30303030",
        "content_first_bytes": "30 30 30 30 39 30 36 36 0d 0a 3c 21 2d 2d 5b 69 66 20 49 45 20 5d 3e 0a 3c 21 64 6f 63 74 79 7",
        "links_hosts": [
          "disqus.com"
        ]
      }
    ]
  }
}
```

Figure 4: Search for 'football'



- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- discovery
- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats
- Query**
- Replication
- Schema Browser

Request-Handler (qt)

/select

common

q

football news

fq

sort

start, rows

010

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

☒ indent
☐ debugQuery

☐ dismax
☐ edismax
☐ hl
☐ facet
☐ spatial
☐ spellcheck

Execute Query


http://localhost:8080/discovery/select?q=football+news&wt=json&indent=true

```

{
  "responseHeader": {
    "status": 0,
    "QTime": 2,
    "params": {
      "indent": "true",
      "q": "football news",
      "_": "1425611952832",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 6,
    "start": 0,
    "docs": [
      {
        "source_file_s": "93.warc@1309",
        "url": "http://www.telegraph.co.uk/sport/football/football-transfers/11385515/Football-Transfer-Window-2015-The-Big-Moves-Telegraph",
        "content_type_ext": "html0a",
        "host": "www.telegraph.co.uk",
        "domain": "telegraph.co.uk",
        "public_suffix": "co.uk",
        "content_type_served": "text/html;charset=UTF-8",
        "content_length": 144612,
        "id": "sha1:AI72QYJHTDQGCLTGPC5GNKLY5ERR7K6A/r4x85kS0SdVLE6pTcUDNUA==",
        "hash": [
          "sha1:AI72QYJHTDQGCLTGPC5GNKLY5ERR7K6A"
        ],
        "crawl_date": "2015-03-04T00:14:58Z",
        "crawl_year": "2015",
        "wayback_date": "20150304001458",
        "content": [
          "00009066Accessibility linksSkip to articleSkip to navigationAdvertisementTelegraph.co.ukSearch - enhanced by OpenText"
        ],
        "content_text_length": 7708,
        "content_type": [
          "text/html"
        ],
        "title": "Video: Football Transfer Window 2015 - The Big Moves - Telegraph",
        "keywords": "transfer window, the big moves, animation, watch, video, football, transfer news,transfer gossip, winter",
        "content_encoding": "UTF-8",
        "content_ffb": "30303030",
        "content_first_bytes": "30 30 30 30 39 30 36 36 0d 0a 3c 21 2d 2d 5b 69 66 20 49 45 20 5d 3e 0a 3c 21 64 6f 63 74 79 74",
        "links_hosts": [
          "discovery"
        ]
      }
    ]
  }
}

```

Figure 5: Search for 'football news'



Dashboard
Logging
Core Admin
Java Properties
Thread Dump
discovery
Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema Browser

Request-Handler (qt)
/select

— common —

q
medical

fq

sort

start, rows
0 10

fl

df

Raw Query Parameters
key1=val1&key2=val2

wt
json

☒ indent
☐ debugQuery

☐ dismax
☐ edismax
☐ hl
☐ facet
☐ spatial
☐ spellcheck

Execute Query


http://localhost:8080/discovery/select?q=medical&wt=json&indent=true


```

{
  "responseHeader": {
    "status": 0,
    "QTime": 17,
    "params": {
      "indent": "true",
      "q": "medical",
      "_": "1425605300615",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 5,
    "start": 0,
    "docs": [
      {
        "source_file_s": "13.warc@4289",
        "url": "http://www.graphicmedicine.org/2015-conference-announcement-and-call-for-papers/",
        "host": "www.graphicmedicine.org",
        "domain": "graphicmedicine.org",
        "public_suffix": "org",
        "server": [
          "nginx/1.6.2"
        ],
        "content_type_served": "text/html; charset=UTF-8",
        "content_length": 41214,
        "id": "sha1:WCP5LSM2PRS8GPZX2UQ53ZLCX2PI3GVX/kfoX7gME+ywWIkjusd+FBA==",
        "hash": [
          "sha1:WCP5LSM2PRS8GPZX2UQ53ZLCX2PI3GVX"
        ],
        "crawl_date": "2015-03-04T00:13:22Z",
        "crawl_year": "2015",
        "wayback_date": "20150304001322",
        "content": [
          "5977712e2e6c1a3763c3ec34f4e4882288d98f2d9b1b1bd90aebca7c1bGraphic Medicine2975HomeBlogPodcastComic ReviewsEditor's"
        ],
        "content_text_length": 5613,
        "content_type": [
          "text/html"
        ],
        "title": "4c2015 Conference Announcement and Call for Papers! | Graphic Medicine",
        "content_encoding": "UTF-8",
        "generator": [
          "WordPress 4.0.1"
        ],

```

Figure 6: Search for 'medical'



Dashboard
Logging
Core Admin
Java Properties
Thread Dump
discovery
Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema Browser

Request-Handler (qt)
/select

q
delhi election

fq

sort

start, rows
0 10

fl

df

Raw Query Parameters
key1=val1&key2=val2

wt
json

☒ indent
☐ debugQuery

☐ dismax
☐ edismax
☐ hl
☐ facet
☐ spatial
☐ spellcheck

Execute Query


http://localhost:8080/discovery/select?q=delhi+election&wt=json&indent=true

```

{
  "responseHeader": {
    "status": 0,
    "QTime": 1,
    "params": {
      "indent": "true",
      "q": "delhi election",
      "_": "1425605416920",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 2,
    "start": 0,
    "docs": [
      {
        "source_file_s": "60.warc@1310",
        "url": "http://indianexpress.com/article/cities/delhi/aap-likely-to-storm-back-to-power-in-delhi-say-opinion-polls/750",
        "host": "indianexpress.com",
        "domain": "indianexpress.com",
        "public_suffix": "com",
        "server": [
          "nginx"
        ],
        "content_type_served": "text/html; charset=UTF-8",
        "content_length": 100729,
        "id": "sha1:XSGT3HPCYOUVEF6APBB3FN6EPDFZSXKC/XcA12m4WrfXt3rcY2DHgCg==",
        "hash": [
          "sha1:XSGT3HPCYOUVEF6APBB3FN6EPDFZSXKC"
        ],
        "crawl_date": "2015-03-04T00:14:23Z",
        "crawl_year": "2015",
        "wayback_date": "20150304001423",
        "content": [
          "14e7Express SpecialsExpress Power List 2015Essar LeaksWorld Cup 2015Idea ExchangeE-PaperToday's PaperAstrology1000"
        ],
        "content_text_length": 9811,
        "content_type": [
          "text/html"
        ],
        "title": "Aam Aadmi Party likely to storm back to power in Delhi, say opinion polls | The Indian Express",
        "keywords": "delhi election polls, delhi election polls, delhi election poll survey, delhi elections 2015, delhi elect",
        "content_encoding": "UTF-8",
        "generator": [
          "WordPress.com"
        ]
      }
    ]
  }
}

```

Figure 7: Search for 'delhi election'



Dashboard
Logging
Core Admin
Java Properties
Thread Dump
discovery
Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema Browser

Request-Handler (qt)
/select
common
q
Delhi Election
fq
sort
start, rows
0 10
fl
df
Raw Query Parameters
key1=val1&key2=val2
wt
json
indent
debugQuery
dismax
edismax
hl
facet
spatial
spellcheck
Execute Query

http://localhost:8080/discovery/select?q=Delhi+Election&wt=json&indent=true
{
 "responseHeader": {
 "status": 0,
 "QTime": 1,
 "params": {
 "indent": "true",
 "q": "Delhi Election",
 "_: "1425613316586",
 "wt": "json"
 }
 },
 "response": {
 "numFound": 2,
 "start": 0,
 "docs": [
 {
 "source_file_s": "60.warc@1310",
 "url": "http://indianexpress.com/article/cities/delhi/aap-likely-to-storm-back-to-power-in-delhi-say-opinion-polls/750",
 "host": "indianexpress.com",
 "domain": "indianexpress.com",
 "public_suffix": "com",
 "server": [
 "nginx"
],
 "content_type_served": "text/html; charset=UTF-8",
 "content_length": 100729,
 "id": "sha1:XSGT3HPCYUVEF6APBB3FN6EPDFZSXKC/XcA12m4WrfXt3rcY2DHgCg==",
 "hash": [
 "sha1:XSGT3HPCYUVEF6APBB3FN6EPDFZSXKC"
],
 "crawl_date": "2015-03-04T00:14:23Z",
 "crawl_year": "2015",
 "wayback_date": "20150304001423",
 "content": [
 "14e7Express SpecialsExpress Power List 2015Essar LeaksWorld Cup 2015Idea ExchangeE-PaperToday's PaperAstrology1000"
],
 "content_text_length": 9811,
 "content_type": [
 "text/html"
],
 "title": "Aam Aadmi Party likely to storm back to power in Delhi, say opinion polls | The Indian Express",
 "keywords": "delhi election polls, delhi election polls, delhi election poll survey, delhi elections 2015, delhi elect",
 "content_encoding": "UTF-8",
 "generator": [
 "WordPress.com"
]
 }
]
 }
}

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

[illegible]

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>