Harvester Automation for Metadata Search Web Application

Part Two: Technologies

Mentors: Nathan Hook, Saquib Aziz Khan, Eric Nienhouse, Christy Grant
Table of Contents

• Introduction
  • The Metadata Search Application, initial state
  • Motivation of the Project

• Implementation and Technologies
  • Automation of the Harvester Service
  • Spring Boot, JGit, Webhook, Ngrok, Docker

• Results

• Future Work

• Acknowledgements

• References
The Metadata Search Application: Initial State

Clone repositories → Harvester Service → Send to Solr → Send to Search App

Welcome to Metadata Search Engine!
This is search engine for accessing data from NCAR/UCAR.

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Motivation of the Project

- Manually clone all repositories
- Manually pull repositories to get updates
- Restart application to start harvesting
Motivation of the Project

Harvester Limitations
Motivation of the Project
Motivation of the Project

Harvester

Limitations
Motivation of the Project

Harvester

Limitations

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Motivation of the Project
Motivation of the Project

Control

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Motivation of the Project

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7

Control

Automation
Motivation of the Project

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7

Control
Automation
Speed
Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Command line

The default interactive shell is now zsh. To update your account to use zsh, please run `chsh -s /bin/zsh`.

```
$ git clone https://github.com/smanalai/simple-spring-app-data
Cloning into 'simple-spring-app-data'...
remote: Enumerating objects: 84, done.
remote: Counting objects: 100% (84/84), done.
remote: Compressing objects: 100% (80/80), done.
remote: Total 84 (delta 37), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (84/84), done.
```

Source Code

```java
public void cloneRepository(){
    try {
        Git.cloneRepository()
            .setURI(String.valueOf(this.uri))
            .setDirectory(new File(this.directory))
            .call();
    } catch (GitAPIException e) {
        throw new GitRepositoryException("Failed to clone repository", e);
    }
    fireCloneEvent();
}
```
Implementation and Technologies: Phase One

Harvester Application

Update Request

Harvester Service

Update
Clean & Reharvest

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase One

Harvester Application

Update
Clean & Reharvest

Update Request

Harvester Service

Get Updates

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase One

Harvester Application

Update Request

Harvester Service

Get Updates

Update Solr

Solr

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase One

Harvester Application
- Update
- Clean & Reharvest

Update Request

Harvester Service

Get Updates

Solr

Update Solr

Update Search App

Welcome to Metadata Search Engine!
This is a search engine to access digital assets from NCAR/UCAR.

Search

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase Two
Implementation and Technologies: Phase Two
Implementation and Technologies: Phase Two

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase Two

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Ngrok.com

webhook

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
<table>
<thead>
<tr>
<th>Session Status</th>
<th>online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account</td>
<td>Sama Manalai (Plan: Free)</td>
</tr>
<tr>
<td>Version</td>
<td>2.3.40</td>
</tr>
<tr>
<td>Region</td>
<td>United States (us)</td>
</tr>
<tr>
<td>Web Interface</td>
<td><a href="http://127.0.0.1:4040">http://127.0.0.1:4040</a></td>
</tr>
<tr>
<td>Connections</td>
<td>ttl</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>HTTP Requests</td>
<td>-------</td>
</tr>
<tr>
<td>POST /webhook</td>
<td>200</td>
</tr>
<tr>
<td>POST /webhook</td>
<td>200</td>
</tr>
</tbody>
</table>
Implementation and Technologies: Phase Three

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase Three

[Diagram showing the flow of information from Github to a webhook trigger, leading to Github updates.]
Implementation and Technologies: Phase Three

Diagram:
- Github Updated
- Trigger Webhook
- Github
- webhooks
- Send POST Request
- ngrok.com

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase Three

1. Github Updated
2. Trigger Webhook
3. Send POST Request
4. Receive JSON
5. ngrok.com
6. Harvester Service

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Implementation and Technologies: Phase Three

Sama Manalai / Terry Yuan
SIParCS 2021: Project 7
Results
Results
Results
Results
Acknowledgements

Thank You To...

**SIParCS Mentors:**
- Nathan Hook,
- Saquib Aziz Khan,
- Eric Nienhouse,
- Christy Grant

**SIParCS Program Leads**
- AJ Lauer,
- Virginia Do,
- Jerry Cycone,
- Max Cordes Galbraith

...and everyone else for making this program happen.
References

https://www.canva.com/
https://www.javatpoint.com/spring-boot-architecture
https://www.twilio.com/blog/2015/09/6-awesome-reasons-to-use-ngrok-when-testing-webhooks.html
https://sendgrid.com/blog/whats-webhook/
https://spring.io/projects/spring-boot