Heritrix for Aspire

Welcome to Heritrix for Aspire, a central location for all information on crawling and processing content using the Aspire Heritrix connector and associated components.

THIS ITEM IS BEING DEPRECATED.

See Aspider Web Crawler

Introduction

About Heritrix for Aspire, how it works, features,

Prerequisites

Supported Heritrix versions, user access requirements, other requirements.

How to Configure

Step by step tutorial to crawl your first Heritrix.

Custom Configuration

If you need to configure authentication or any other advanced features

Using HTTPS

General instructions for crawling Heritrix with https.

FAQ & Troubleshooting

Questions and answers, including troubleshooting techniques.

For Developers

Questions and answers, components and troubleshooting techniques for developers.