Browser Based Harvesting with Umbra

We are currently (summer 2018) working on the integration of browser-based harvesting in NetarchiveSuite. While there are many different approaches to that might be adopted, we have chosen to start with Internet Archives Umbra system as this appears to be the simplest to deploy and run. We will use it purely as a link-identifier, with all harvesting and archiving continuing to take place in NetarchiveSuite and Heritrix.

- Detailed Architecture
- Docker-based Test System
- Overall Architecture for Browser-based Harvesting