BNF sanity test

Describes the minimal sanity test as performed at BNF verifying the integrity of the NAS codebase.

Configuration and test files

Deployment configuration files and scripts: standalone_derby.tar.gz and standalone_postgres.tar.gz

List of test domains: AFNIC_subset_6k_clean.txt

Test runs

The test is ran first on the Derby configuration, then on the PostgreSQL configuration.

Test steps

1) Deploy the application and start it
2) Open the browser and load up the interface: http://<your host>:8074/HarvestDefinition
3) Switch to french language
4) Go to "Définitions / Créer un nouveau domaine" ("Definitions / Create Domain") and import the list of test domains. All 6000 domains should be imported.
5) Go to "Définitions / Rechercher un domaine" ("Definitions / Find Domain(s)") and check that all 6000 domains are present (use * wildcard).
6a) Go to "Définitions / Collectes ciblées / Collectes ciblées : créer un nouvel ensemble à collecter" ("Definitions / selective harvests / Create new selective harvest definition"), and create a new selective harvest.
6b) Name the harvest as you like
6c) Add the following domains: "lemonde.fr, bnf.fr, jeuxvideo.com". Confirm creation when prompted.
6d) Edit each of the three domains, and create a new configuration based on the "frontpages_plus_1_level" template, with a maximum budget of 1000 URI and unlimited weight (input -1). Be sure to name all three configurations the same.
6e) Save the harvest, go back editing it and select the newly created configurations for the domains.
6f) Activate the harvest
7a) Go to "Statut de la collecte" ("Harvest status") and press the "Afficher" ("Show") button. Check that a job is created and sent to the harvester.
7b) Go to "Statut de la collecte / Jobs en cours" ("Harvest status / Running Jobs"), and check that the job appears in the table, and that progress information is properly generated.
7c) Click on the job ID, then click to display the Heritrix console (admin / adminPassword), then pause the job.
7d) Go back to "Statut de la collecte / Jobs en cours" and check that the job is reported as paused.
7e) Click again on the job ID, and check that the contents of the table "Liste des files d'attente à traiter" is coherent with the Heritrix frontier report (same queues).