Article mis à jour le : 2024/11/25 22:30 / Imprimé le 2025/12/16 04:18

Data Scraping

Récupérer des données du web public pour whatever raison

wget

wget -p -k https://web.archive.org/web/20130328173258/http://videogramo.8bitpeoples.com/wget -p -k http://francoisegamma.computersclub.org/

-p : télécharger tous les fichiers nécessaires (css, images, etc.)-k : convertir les liens pour que la page fonctionne hors ligne

httrack

pour créer des miroirs locaux de sites

Beaucoup d'autres possiblités avec wget

http://www.httrack.com/page/2/fr/index.html (existe en paquet debian : webhttrack)

Article extrait de : http://lesporteslogiques.net/wiki/ - WIKI Les Portes Logiques

Adresse: http://lesporteslogiques.net/wiki/ressource/data_scraping

Article mis à jour: 2024/11/25 22:30