BULK REVIEWER

Tim Walsh
Bibliothécaire en conservation numérique, Bibliothèque de l'Université Concordia
2018 Summer Fellow, Harvard Library Innovation Lab
tim.walsh@concordia.ca

Déterminer, examiner et supprimer les dossiers de nature délicate

- Images et répertoires sur disque
- Quel que soit le format du fichier
- Alimenté par bulk_extractor
- Examiner les résultats, rejeter les faux positifs, produire des rapports et exporter des fichiers dans l'application prototype pour navigateur
- Créé à l'aide de Django, Vue.js, bulk_extractor, DFXML et Docker

Numéros de sécurité sociale

Numéros de carte de crédit

Numéros de téléphone

Adresses courriel

Historique internet

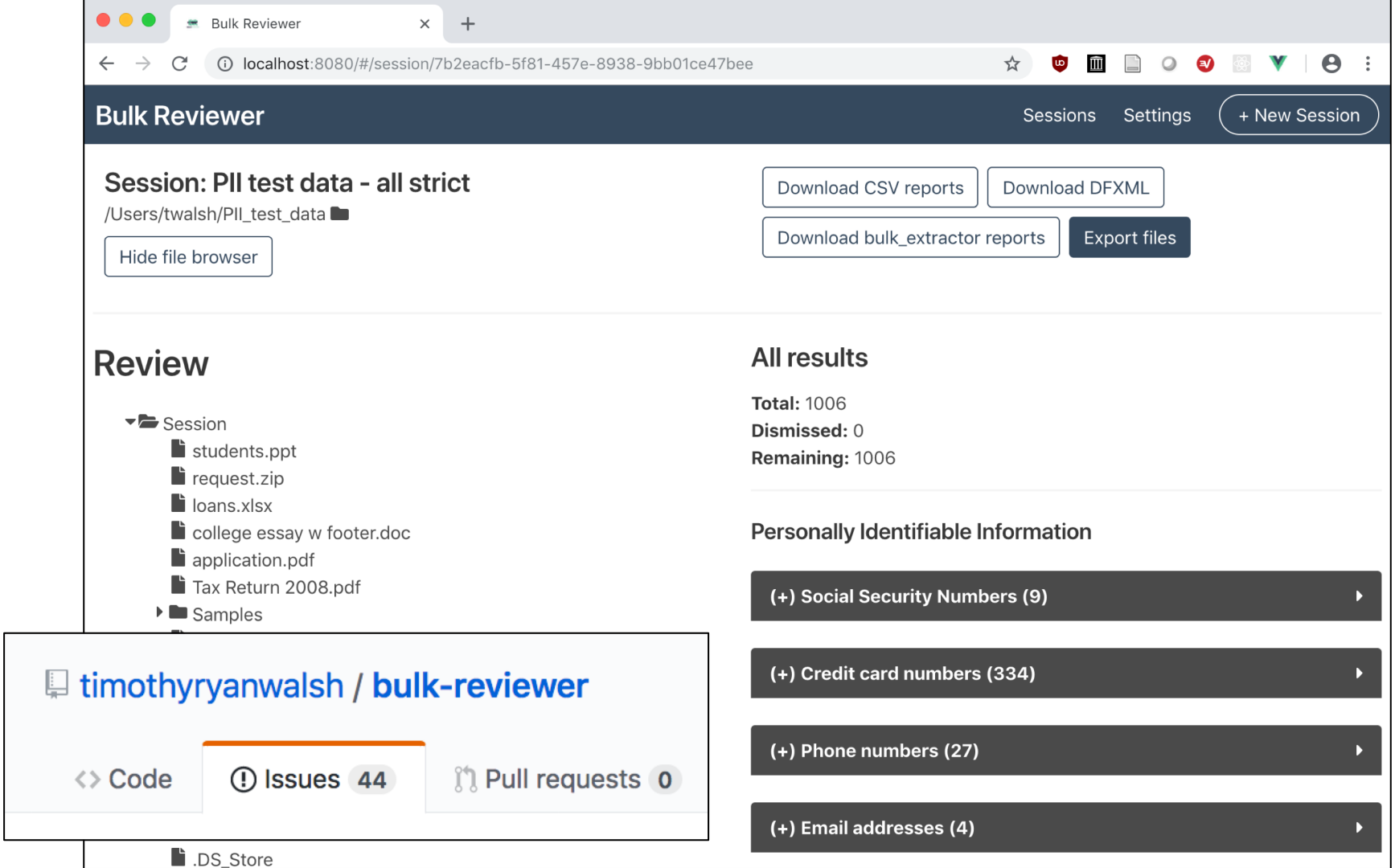
Métadonnées EXIF

Données de géolocalisation (GPS)

Termes de recherche personnalisés

Library Innovation Lab

Langdell Hall. Harvard University.


Bulk Reviewer
localhost:8080/#/session/7b2eacfb-5f81-457e-8938-9bb01ce47bee
Bulk Reviewer
Sessions
Settings
+ New Session
Session: PII test data - all strict
/Users/twalsh/PII_test_data
Hide file browser
Download CSV reports
Download DFXML
Download bulk_extractor reports
Export files
Review
Session
students.ppt
request.zip
loans.xlsx
college essay w footer.doc
application.pdf
Tax Return 2008.pdf
Samples
All results
Total: 1006
Dismissed: 0
Remaining: 1006
Personally Identifiable Information
(+) Social Security Numbers (9)
(+) Credit card numbers (334)
(+) Phone numbers (27)
(+) Email addresses (4)
timothyryanwalsh / bulk-reviewer
Code
Issues 44
Pull requests 0
.DS_Store

Plan de développement

Améliorations à l'outil actuel (alpha) :

- Ajouter de **nouveaux instruments de numérisation** : Facebook, Outlook, identité
- L'adapter au **contexte canadien** : Numérisation du NAS, données d'essai, internationalisation, traduction
- **ROC préalable** des PDF et des images de format standard
- Extensibilité et amélioration de la sécurité
- Améliorations de l'expérience, de l'interface et de l'accessibilité

Améliorations de la prochaine génération (bêta) :

- **Apprentissage automatique** (évaluation des risques fondée sur le contexte)

Je vous remercie!

Vous souhaitez participer? Communiquez avec moi!

tim.walsh@concordia.ca
@bitarchivist
https://github.com/timothyryanwalsh/bulk-reviewer