

# Numérisation : communautés locales, partenariats et objets numériques haute fidélité

Loren Fantin  
Executive Director, ODW

Art Rhyno  
Systems Librarian, U. of Windsor



OurDigitalWorld



# York University, Clara Thomas Special Collections & Archives: Journaux italo-canadiens, 1929-1946



Full text: [Help](#)

Non vi sono questioni di partito 0 di tendenze. Ognuno pub occupare il suo poggio onorevole in questo schieramento di forze per il raggiungimento della Vittoria, Ognuno pub recare il suo contributo di pensiero e di azione.

-- -- x . . . . . A\_T A---- D:..... 'rrn'n In

; Noi domandiamo agli italiani: Vale la pena lottare per questi scopi? .  
C; Yale la pena combattere per questo Canada, di cui siamo parte integrale, dove godiamo un benessere sconosciuto in altri Paesi, dove ei is concesso il dono preziosissimo della liberty e disill'indipendenpy?

Percib salutiamo con animo lieto il sorgere dei Comitati unitari italo-canadesi per la Vittoria delle Nazioni Unite, ed invitiamo tutti gl'italiani del Canada a parteciparvi, senza restriizioni mentali e senza falsi pudori.

These free Italians are anxious to demonstrate to the world that their race is not a solid unit in support of Mussolini and his policies. They would contribute active assistance to the restoration of a democratic and more wholesome regime in Italy. A South American capital wisely was chosen as the place for starting the movement. Among the Italian communities, which are very large in Argentina and exist in most countries of South America, Mussolini's agents have been untiringly busy with propa-

Vale la pena aiutare i nostri fratelli che mordono il freno sotto , giogo fascismo e nazista? V

Anche coloro che hanno ancora la mente ottenebrata da false ideologie, dovranno convincersi presto o tardi che "ttOtt' vi is salvezza all'infuori della partecipazione

# U of Alberta: Holodomor Research & Education Consortium



## HOLODOMOR PHOTO DIRECTORY

**Search the collections**

Find

[Search](#) [Clear](#)

### Search Options

- [Browse by Subject](#)
- [Advanced Search](#)

### Directory Context

- [About the Directory](#)
- [The Holodomor's Forbidden Photographs](#)
- [Holodomor Basic Facts](#)
- [Historical Map](#)

### Photographer Biographies

- [Alexander Wienerberger](#)
- [Nikolai Bokan](#)
- [James Abbe](#)
- [Whiting Williams](#)

### Featured collections

-  Alexander Wienerberger: Innitzer Album
-  Alexander Wienerberger: Beyond the Innitzer album
-  James Abbe Collection
-  Nikolai Bokan Collection
-  Whiting Williams Collection
-  Select 1920s Famine Photos from Ukraine and Russia

### Gallery



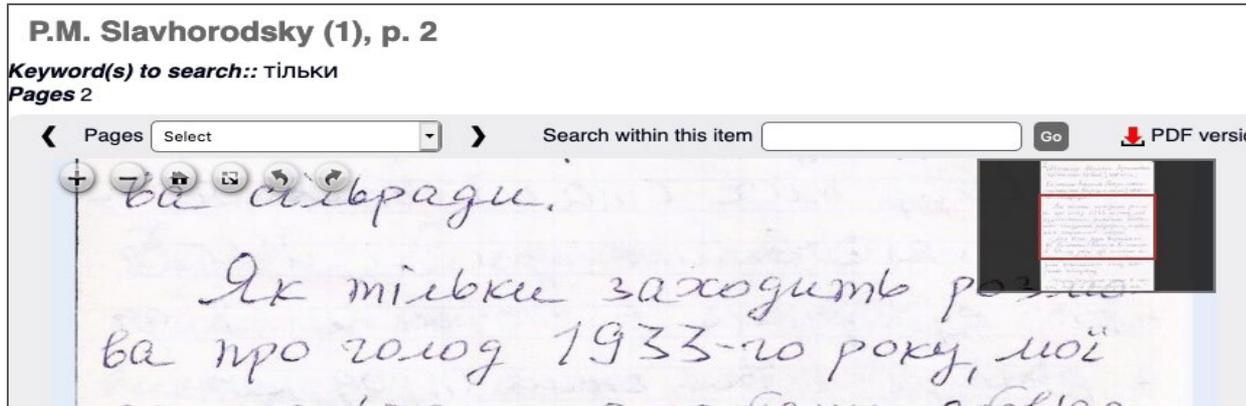
**A male factory worker in his temporary shelter on the outskirts of Kharkiv**

In what appears to be an open, uncultivated field, a man sits in his temporary dwelling. The tiny shelter is slightly raised off the ground and made ...

### Project Team

[Acknowledgements](#)

# Holodomor Research & Education Consortium



Автомобіль «Гуго-Вольф». 1911 г.  
Москва

Кіїв-54  
бул. Зкалова, 52,  
кв. 38.  
Маньку В.А.

258666

Черкаська обл.  
Лисянський р-н,  
с. Бужанка.  
Славгородський П.М.

166272

252054

Пашаго видки: зображення святих міста Київщини

Create electronic postcard

- [Original Letter \(PDF\)](#)
- [Transcription \(PDF\)](#)
- [Translation \(PDF\)](#)

№ 166/272

258 666  
Черкаська обл.  
Лисянський р-н.  
с. Бужанка.  
Славгородський П.М.

Шановні Володимир Антонович і Ліда Борисівна!

Надсилаю вам для книги-меморіалу «33-й: голод» список прізвищ тих сталінських яничарів, які в моєму селі Жаб'янка одбирали в людей хліб до зернини і всі інші продукти харчування, свідомо прирікаючи своїх односельчан на голодну смерть. Ось вони:

Поляруш Мирін Гнатович;  
Дідик Федір (прізвисько Веньгір);  
Дідик Зосим Федорович (прізвисько Розум), 1906 р. народж.  
Дідик Гаврило (прізвисько Шешелей);  
Машир Микола Арсеньович (прізвисько Сотка), завгосп;  
Білокін Василь Гаврилович (прізвисько Вирішальний), голова сільради.

# RCDR : outil de transcription

Transcribing Canada's Documentary Heritage |  
Transcription du patrimoine documentaire du Canada



Canadian Research  
Knowledge Network  
Réseau canadien  
de documentation  
pour la recherche

**WAR DIARY** Army Form C. 2118.

Instructions regarding War Diaries and Intelligence Summaries are contained in F.S. Regs., Part II, and the Staff Manual respectively. Title pages will be prepared in manuscript.

(Erase heading not required.)

Place	Date	Hour	Summary of Events and Information	Remarks and references to Appendices
Field	1916 May	6-6 A.M.	818 men building, improving and repairing trenches 172 - carrying material 84 - wiring Sections in Camp constructing dam and generally employed. Casualties No. 150257 Spr Hart W.H. G.S. wound in Neck.	
-	17-18	-	706 men improving and repairing trenches 220 - carrying and unloading material 92 - wiring 24 - building & repairing dug-outs and 6 men on Bridges. Sections in Camp constructing Dug-outs corner Vlaamtinghe & Keriststraat Rds. Dam and general work in Camp. Casualties: No. 150257 Spr Hart W.H. Died from Wounds.	
-	18-19	-	530 men building, improving and repairing trenches 223 - carrying and unloading material 25 - wiring 28 - working on Dressing Station 18 - on Bridges 12 - at Dump 5 hours 2 - Reconnaissance 12 men on Cellars at Hooze. Sections in Camp, working on Dam and building Dug-outs corner Vlaamtinghe & Keriststraat Rds. Casualties. No. 125344 Spr Alexander A.S. Sick. Sent to Hospital	

833 Wk. W556/1454 700000 515 D.D.&L. A.D.S.S./Form C 2118.

Continue to Next

Transcription:

SHOW HELP

WAR DIARY or INTELLIGENCE SUMMARY (Erase heading not required.)

Instructions regarding War Diaries and Intelligence Summaries are contained in F.S. Regs., Part II, and the Staff Manual respectively. Title pages will be prepared in manuscript.

Army Form C. 2118.

Place Date Hour Summary of Events and Information  
Remarks and references to Appendices

1916  
Field May 16-17th AM 6-6  
818 men building, improving and repairing trenches  
172 - carrying material  
84 - wiring  
Section: in Camp constructing dam and generally employed.  
Casualties No 150257, Spr. Hart, W.H. G S wound in Neck.  
[[illegible]]

Field May 17-18th AM 6-6  
706 men improving and repairing trenches  
220 - carrying and unloading material  
92 - wiring  
24 - building & repairing dug-outs and 6 men on Bridges.  
Sections in Camp constructing Dug-outs and 6 men on Bridges.  
Sections in Camp constructing Dam and general work in Camp.  
Casualties. No. 150257 Spr. Hart, W.H. Died from Wounds.  
[[illegible]]

Field May 18-19th AM 6-6  
530 men building, improving and repairing trenches  
223 - carrying and unloading material  
25 - wiring  
28 - working on Dressing Station  
18 - on Dug-Outs  
12 - at Dump 5 hours  
2 - Reconnaissance 12 men on Cellars at Hooze.  
Sections in Camp, working on Dam and building Dug-outs

Department of Militia and Defence, World War 1 War Diaries,  
8th Field Company, Canadian Engineers, 3rd Canadian  
Division, (R.G. 9, III D 1, Volume 360).

PUBLIC ARCHIVES

# Tendances dans les logiciels de reconnaissance/analyse de texte

**Kaggle Notebook: Topic Modeling: Finding Related Articles**

```
In [17]: # most frequent words
word_count = pd.DataFrame({'word': vectorizer.get_feature_names(), 'count': n
a_vectorized.sum(axis=0)})[0])

word_count.sort_values('count', ascending=False).set_index('word')[ :20].s
t', ascending=True).plot(kind='barh')
```

Out[17]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f4b42374358>

word	count
cell	~950000
virus	~850000
infection	~750000
study	~650000
protein	~550000
patient	~450000
disease	~350000
result	~250000
viral	~150000
show	~100000
include	~80000
case	~70000
high	~60000
human	~50000
increase	~40000
subset	~30000
test	~20000
report	~15000
group	~10000
sample	~5000

**Distant Reader: Risk factors from the CORD dataset**

Introduction | Size & Scope | Readability | Word Frequencies

Introduction

Reader "study carrel", a set of structured data intended to help the student, its corpus. of more than 100,000 (item(s)).

**LIBRARY LABS**

**Newspaper Navigator**

Explore the visual and textual content within the *Chronicle America* digitized newspaper collection in new ways using machine learning.

Dataset and Search Application now live!

**The Newspaper Navigator Search Application**

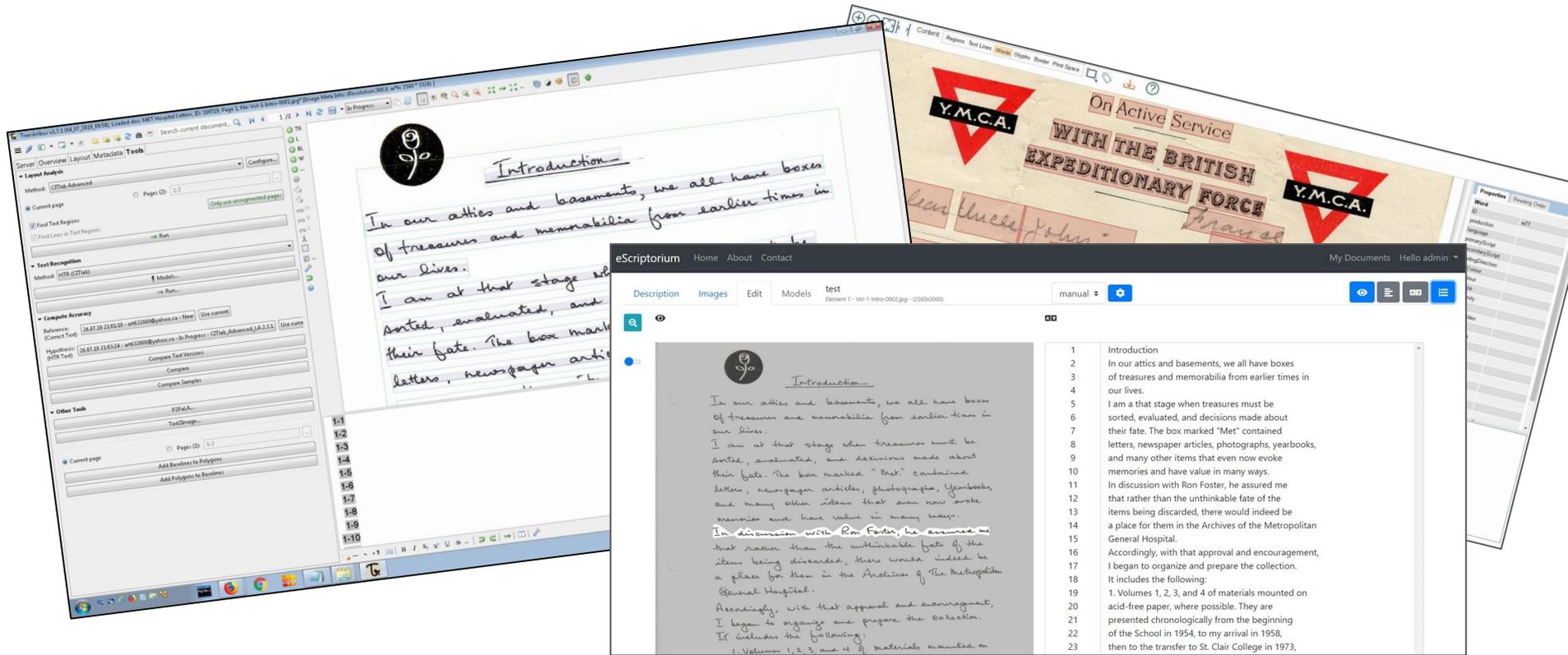
This experimental web application allows you to browse over 1.56 million images extracted from the *Chronicle America* database of digitized historic newspapers using machine learning. Now, you can use this tool to search the images by visual similarity by training your own machine learning classifiers!

- Des projets publics et des ensembles de données voient le jour avec des logiciels plus accessibles et plus puissants (par exemple : [Kaggle](#), [Distant Reader](#), [LoC NN](#))

# Augmentation des options de ROC

- Infonuagique
  - Joueurs importants dans le domaine de l'infonuagique offrent désormais des services de ROC (par exemple: [Microsoft Azure Computer Vision](#), [Google Cloud Vision API](#), [Amazon Textract](#))
  - Systèmes bien établis de ROC offrent des options en infonuagique (par exemple: [ABBYY Cloud OCR SDK](#), [Tesseract](#) déployé derrière les services d'infonuagique, incluant [AWS Lambda](#))
- Plus de transparence et d'exactitude désirées
  - Voir [Collective Compilation of Thematics](#) de [Eldorado Impresso Workshop](#)

# Évolution des systèmes d'écriture



- Évolution en HTR
- Reconnaissance globale/plateformes de formation plus accessibles (par exemple : projets [Transkribus](#), [Aletheia](#), [eScripta](#))

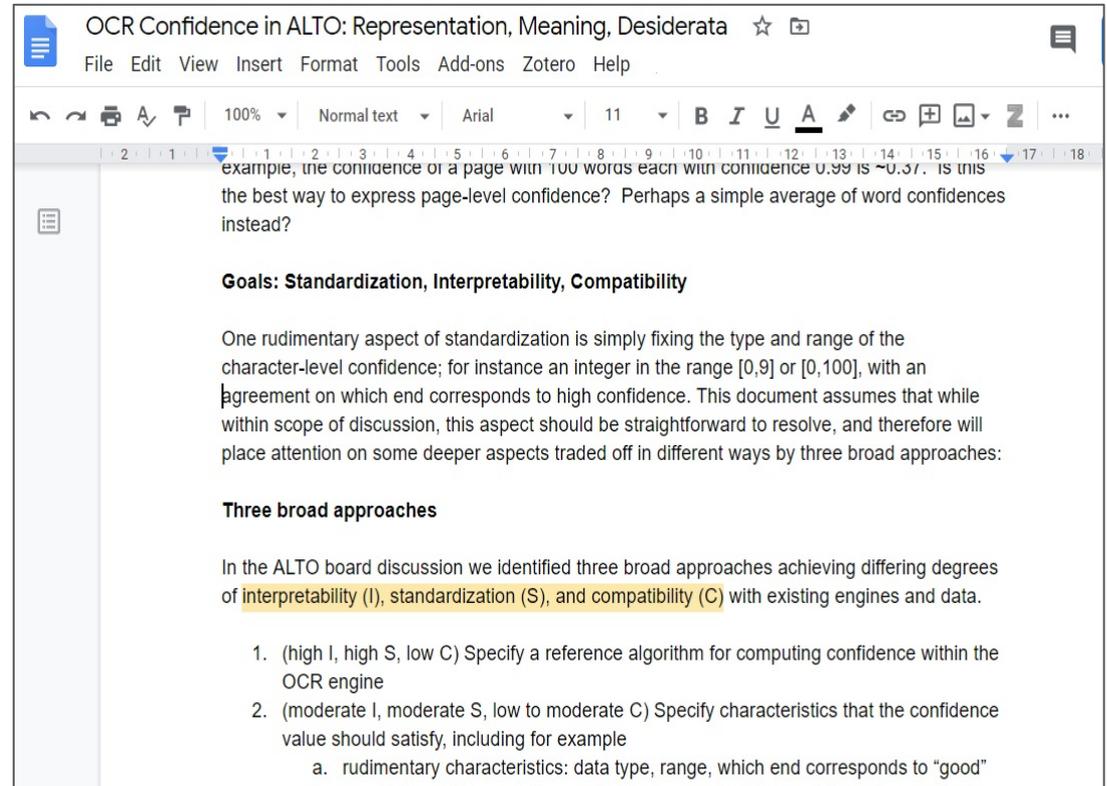
# Profondeur et contrôle accru de l'encodage



altotml / schema

<> Code    ! Issues 18

ALTO support for encoding OCR segmentation ambiguity #63



OCR Confidence in ALTO: Representation, Meaning, Desiderata

File Edit View Insert Format Tools Add-ons Zotero Help

100% Normal text Arial 11 B I U A

example, the confidence of a page with 100 words each with confidence 0.99 is ~0.37. Is this the best way to express page-level confidence? Perhaps a simple average of word confidences instead?

**Goals: Standardization, Interpretability, Compatibility**

One rudimentary aspect of standardization is simply fixing the type and range of the character-level confidence; for instance an integer in the range [0,9] or [0,100], with an agreement on which end corresponds to high confidence. This document assumes that while within scope of discussion, this aspect should be straightforward to resolve, and therefore will place attention on some deeper aspects traded off in different ways by three broad approaches:

**Three broad approaches**

In the ALTO board discussion we identified three broad approaches achieving differing degrees of **interpretability (I), standardization (S), and compatibility (C)** with existing engines and data.

1. (high I, high S, low C) Specify a reference algorithm for computing confidence within the OCR engine
2. (moderate I, moderate S, low to moderate C) Specify characteristics that the confidence value should satisfy, including for example
  - a. rudimentary characteristics: data type, range, which end corresponds to "good"

- Intérêt accru pour les systèmes [ALTO](#)
- Syntaxe plus commune pour les textes composés ou manuscrits

# Partenariat avec University of Windsor pour un programme de journaux communautaires



- Entente formelle en 2018
- Numérisation réalisée au sein de l'établissement avec le [prototype de scanner ODW](#)
- Poursuite des travaux avec IA entrepris en 2007

# Formats de microforme



- Défis et possibilités
- Travail à faire un niveau des fiches
- Vaut l'effort : importance historique des journaux communautaires

# Initiative LAM pour les journaux communautaires

- À la fin 2017, 36 journaux d'Ontario ont fermé
- Appel aux éditeurs pour obtenir leurs archives de journaux
- Protocoles et ententes pour « sauver » physiquement les journaux

Archives de l'Ontario

Associations d'archives en Ontario

Bibliothèque et Archives Canada

Association des bibliothèques de l'Ontario

OurDigitalWorld

Toronto Public Library

Loren Fantin

[lfantin@ourdigitalworld.org](mailto:lfantin@ourdigitalworld.org)

@ODWorg

Art Rhyno

[arthyno@uwindSOR.ca](mailto:arthyno@uwindSOR.ca)

**Thank you!**

**Merci!**



OurDigitalWorld