

NLP Assistant

Computational text analysis as a helper
for manual metadata creation

September 30, 2019

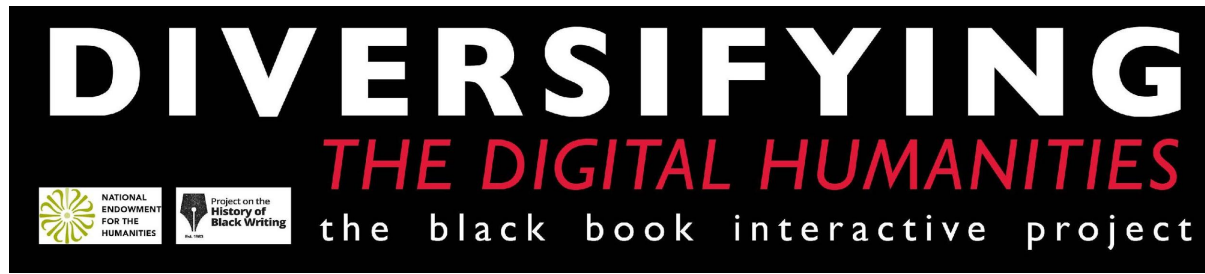
Erin Wolfe

Metadata Librarian

University of Kansas

Libraries

edw@ku.edu



“Of the thousands of African American (AA) novels written by writers from underrepresented communities, most are unknown and understudied.”

“Thus, despite the extensive recovery of AA writing and its growth as part of contemporary literature, its relationship to the Digital Humanities is characterized by a critical invisibility.”

Data collection by BBIP

Bibliographic

- * Publishing process
- * Type of publisher
- * Geographic location of publisher

Author information

- * Education
- * Occupation
- * Birthplace

Literary analysis

- * Themes, genres, subjects
- * Stylistic or narrative features
- * Information about protagonist
- * Geographical setting

Historical context

- * Book reviews
- * Literary Predecessors
Successors
- * Other works by author

- 53 fields in all



WIKIDATA

- * Date of birth
- * Place of birth
- * Date of death
- * Cause of death
- * Place of burial
- * Manner of death
- * Field of work
- * Occupation
- * Religion
- * Education
- * Ethnic group

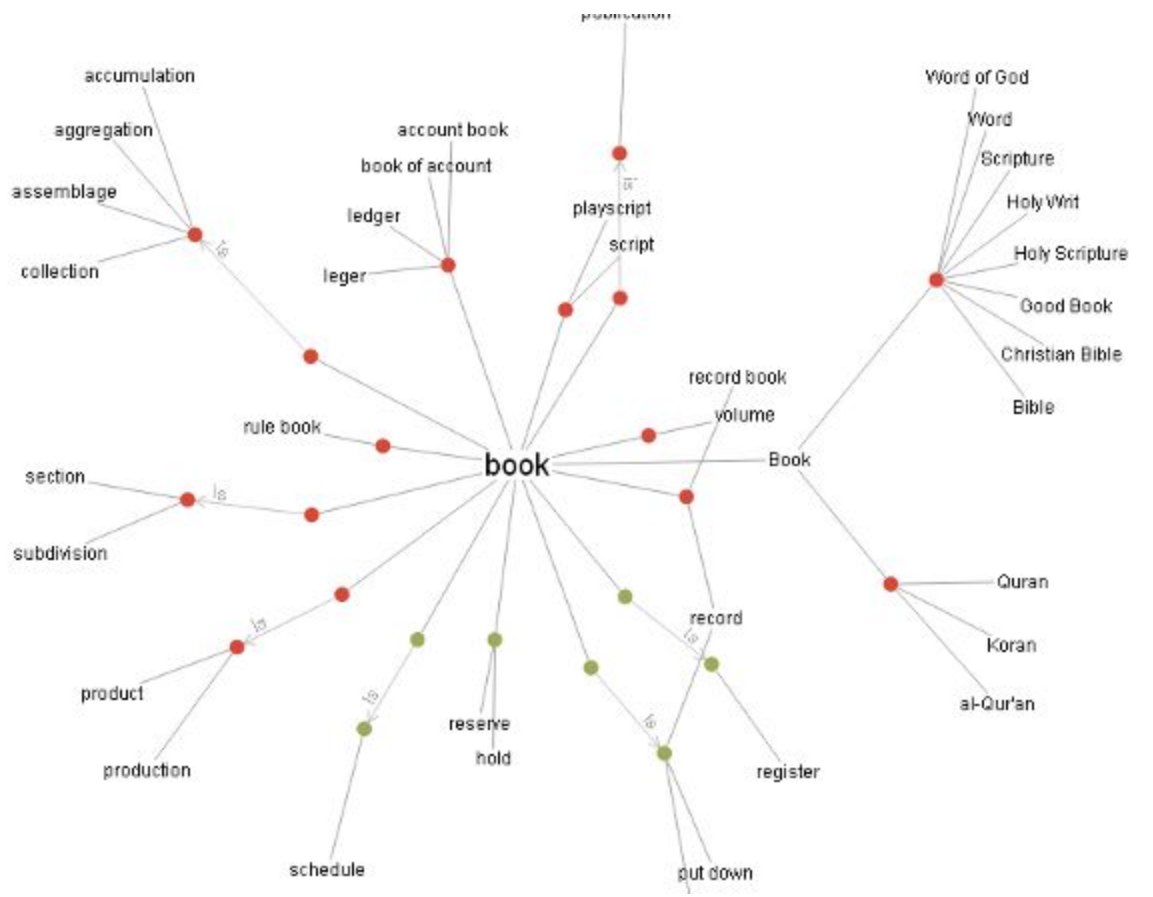
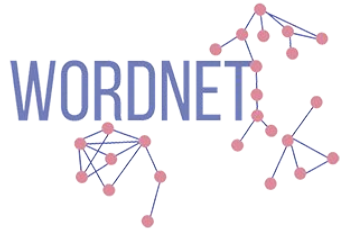


Image : <https://openscience.com/wordnet-open-access-data-in-linguistics/>

FIGHT

DEFINITIONS and SYNSETS

- a hostile meeting of opposing military forces in the course of a war
- an aggressive willingness to compete
- the act of fighting; any contest or struggle
- a boxing or wrestling match
- make a strenuous or labored effort

- And more...

FIGHT

DEFINITIONS and SYNSETS

- a hostile meeting of opposing military forces in the course of a war

battle.n.01

- an aggressive willingness to compete

competitiveness.n.01

- the act of fighting; any contest or struggle

fight.n.02

- a boxing or wrestling match

fight.n.05

- make a strenuous or labored effort

fight.v.03

- And more...

FIGHT

DEFINITIONS and SYNSETS

- a hostile meeting of opposing military forces in the course of a war

battle.n.01

- an aggressive willingness to compete

competitiveness.n.01

- the act of fighting; any contest or struggle

fight.n.02

- a boxing or wrestling match

fight.n.05

- make a strenuous or labored effort

fight.v.03

- And more...

Sense

fight.n.02

- the act of fighting; any contest or struggle

Hypernym

conflict.n.01

Sense

fight.n.02

- the act of fighting; any contest or struggle

Synonyms

fighting; combat; scrap

Hyponyms

affray.n.02

battering.n.01

beating.n.01

brawl.n.02

brush.n.06

close-quarter_fighting.n.01

dogfight.n.04

duel.n.01

fencing.n.03

fistfight.n.01

gunfight.n.01

hassle.n.02

in-fighting.n.01

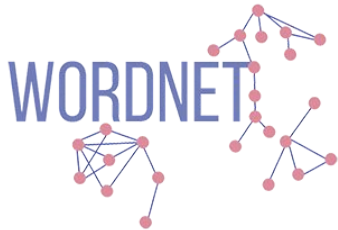
knife_fight.n.01

rumble.n.03

set-to.n.01

shock.n.02

single_combat.n.01



base list of 53 terms



review synsets for each term



filter out unwanted senses



gather hyponyms



list of 1,879 topic-related synsets

fight.n.02

Definition: the act of fighting; any contest or struggle

word	sentence
combat	It was a sheer contest of wills , as intense as physical combat .
combat	Our eyes locked in combat .
fighting	It 's defeated me , but you must prepare yourself to go on fighting .

fight.n.05

Definition: a boxing or wrestling match

word	sentence
fight	Deputy sheriffs and policemen do n't know much about organized recreation ; all they usually do when celebrating is get drunk and pound each other on the back , exchanging loud insults which under ordinary circumstances would result in a fight .
fights	A number of the men got drunk , and there were numerous fights .
fight	When my father learned about this he told me to stand up and fight , but I refused .

Named Entity Recognition (NER)

Named Entity Recognition (NER)

'Marco went to Rome to visit the Catholic Church headquarters.'

```
[('Marco', 'PERSON'),  
 ('went', 'O'),  
 ('to', 'O'),  
 ('Rome', 'LOCATION'),  
 ('to', 'O'),  
 ('visit', 'O'),  
 ('the', 'O'),  
 ('Catholic', 'ORGANIZATION'),  
 ('Church', 'ORGANIZATION'),  
 ('headquarters.', 'O')]
```

Named Entity Recognition (NER)

('Seattle', 'LOCATION')
('San', 'LOCATION')
('Francisco', 'LOCATION')
('Chicago', 'LOCATION')
('New', 'LOCATION')
('York', 'LOCATION')
('San', 'LOCATION')
('Francisco', 'LOCATION')
('Lenox', 'LOCATION')
('Avenue', 'LOCATION')
('New', 'LOCATION')
('York', 'LOCATION')
('Africa', 'LOCATION')

Harlem	LOCATION
Vienna	CITY
England	COUNTRY
America	COUNTRY
New	STATE_OR_PROVINCE
York	STATE_OR_PROVINCE
America	COUNTRY
Ill	STATE_OR_PROVINCE
Greenwich	CITY

Named Entity Recognition (NER)
2005 Automatic Content Extraction (ACE)

Geo-political entity

A place with a population, physical or political boundaries, etc.

* E.g. Chicago, New York, Africa

Location

A physical place without a political entity

* E.g. the beach, the country, the North

Facility

A man-made structure or place with a functional use

* E.g. a barn, my apartment, the kitchen

Named Entity Recognition (NER)

2005 Automatic Content Extraction (ACE)

Geo-political entity

A place with a population, physical or political boundaries, etc.

- Chicago (100)
- Seattle (75)
- New York (72)
- America (64)
- the city (39)
- the United States (25)
- Paris (25)
- Mississippi (20)
- town (18)
- there (18)
- Washington (17)
- Pittsburgh (14)

Named Entity Recognition (NER)

LitBank

Geo-political entity

A place with a population, physical or political boundaries, etc.

- Chicago (100)
- Seattle (75)
- New York (72)
- America (64)
- the city (39)
- the United States (25)
- Paris (25)
- Mississippi (20)
- town (18)
- there (18)
- Washington (17)
- Pittsburgh (14)
- that rapidly growing portion of the city
- other job areas where they competed with white Americans
- any other area in the city
- some port in southeast Alaska which was not icebound
- a country I could belong to without reservation
- a mandated area under the old League of Nations but now under the control of the Union of South Africa
- a suburb of the city called Green Lake
- this isolated island of color where people died from tuberculosis seven times faster than did whites
- a country that did not want me, where I was always a stranger

Thank you

Erin Wolfe
edw@ku.edu