



Universitatea
Politehnica
București



Facultatea de
Automatică și
Calculatoare



Catedra de
Calculatoare

Identificarea cuvintelor cheie din evenimentele istorice

Sesiunea de licențe - Iulie 2014

Autor

Ioana Andronescu

ioana.andronescu@gmail.com

Conducător științific

Ș.L. Dr. Ing. Costin-Gabriel Chiru



- Introducere
- Abordare
- Metode de detectare a peak-urilor
- Rezultate
- Concluzii
- Întrebări



- perspectivă asupra evenimentelor
- conexiuni automate între evenimente



Listă de evenimente

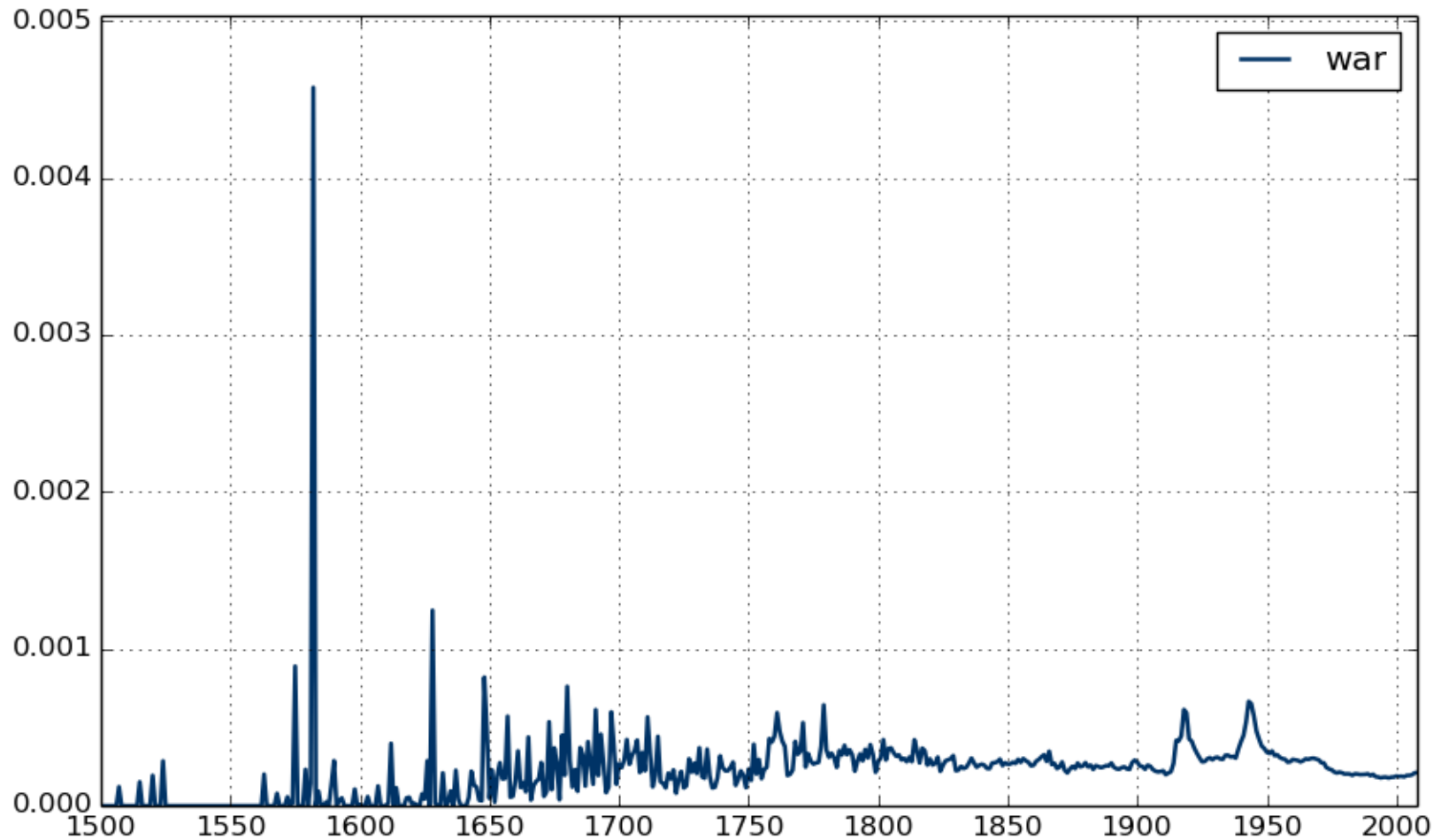
- American Civil War, 1861-1865
- Conquests by the Empire of Japan, 1894-1945
- Deluge, 1655-1660
- Dungan Revolt, 1862-1877, 1895-1896
- French Wars of Religion, 1562-1598
- Iran–Iraq War, 1980-1988
- Korean War, 1950-1953
- Qing Dynasty Conquest of the Ming Dynasty, 1616-1662
- Mexican Revolution, 1911-1920
- Napoleonic Wars, 1803-1815
- Nigerian Civil War, 1967-1970
- Paraguayan War, 1864-1870
- Russian Civil War, 1917-1921
- Second Sudanese War, 1983-2005
- Shaka, 1816-1828
- Soviet War in Afghanistan, 1980-1988
- Spanish Civil War, 1936-1939
- Taiping Rebellion, 1851-1864
- Thirty Years' War, 1618-1848
- Vietnam War, 1955-1975
- World War I, 1914-1918
- World War II, 1939-1945



1. extragerea articolelor de pe Wikipedia
2. căutarea cuvintelor în Google Books Ngrams
3. asocierea relevanței istorice
4. găsirea cuvintelor cheie din fiecare articol
 - cuvinte comune celor 6 metode dintr-un articol
 - cuvinte unice dintr-un articol
5. găsirea cuvintelor care definesc un eveniment “war”



Serii de timp





Metode de detectare a peak-urilor

1. Fereastră glisantă

- Funcția 1: medie dintre diferențele față de vecinii din stânga și din dreapta
- Funcția 2: diferența față de mediile vecinilor din stânga și din dreapta
- Funcția 3: maximul dintre diferența față de vecinii din stânga și din dreapta

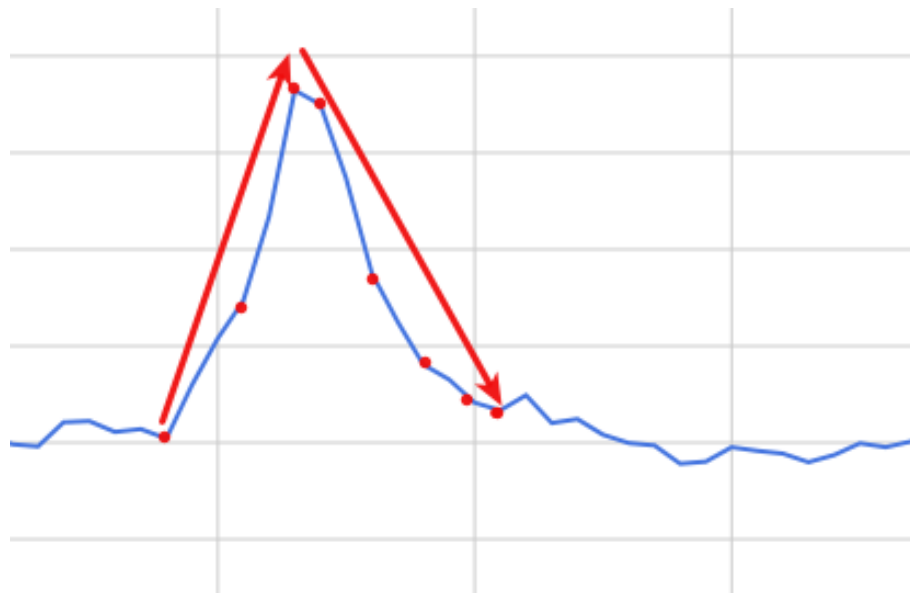




Metode de detectare a peak-urilor

2. Dublă schimbare

- creștere dublă sau scădere dublă

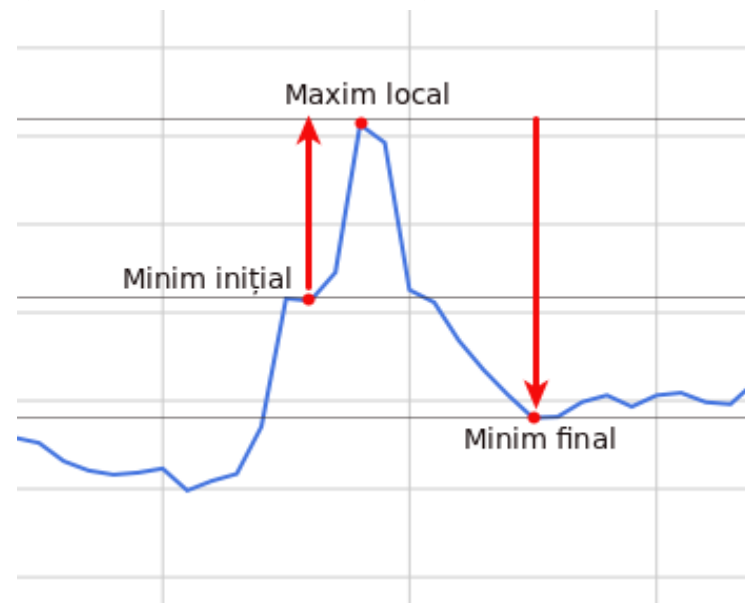




Metode de detectare a peak-urilor

3. Diferență de nivel

- Funcție 1: sumă între distanța de la maxim la punctul minim final și distanța de la maxim la punctul minim inițial
- Funcție 2: funcția 1 împărțită la diferența dintre maximul și minimul local





Comparații metode

Method	Divergence
level LD1	62.07
level LD2	73.20
window SW1	59.55
window SW2	59.55
window SW3	69.92
double change	64.28

Divergența Kullback-Leibler pentru “American Civil War”



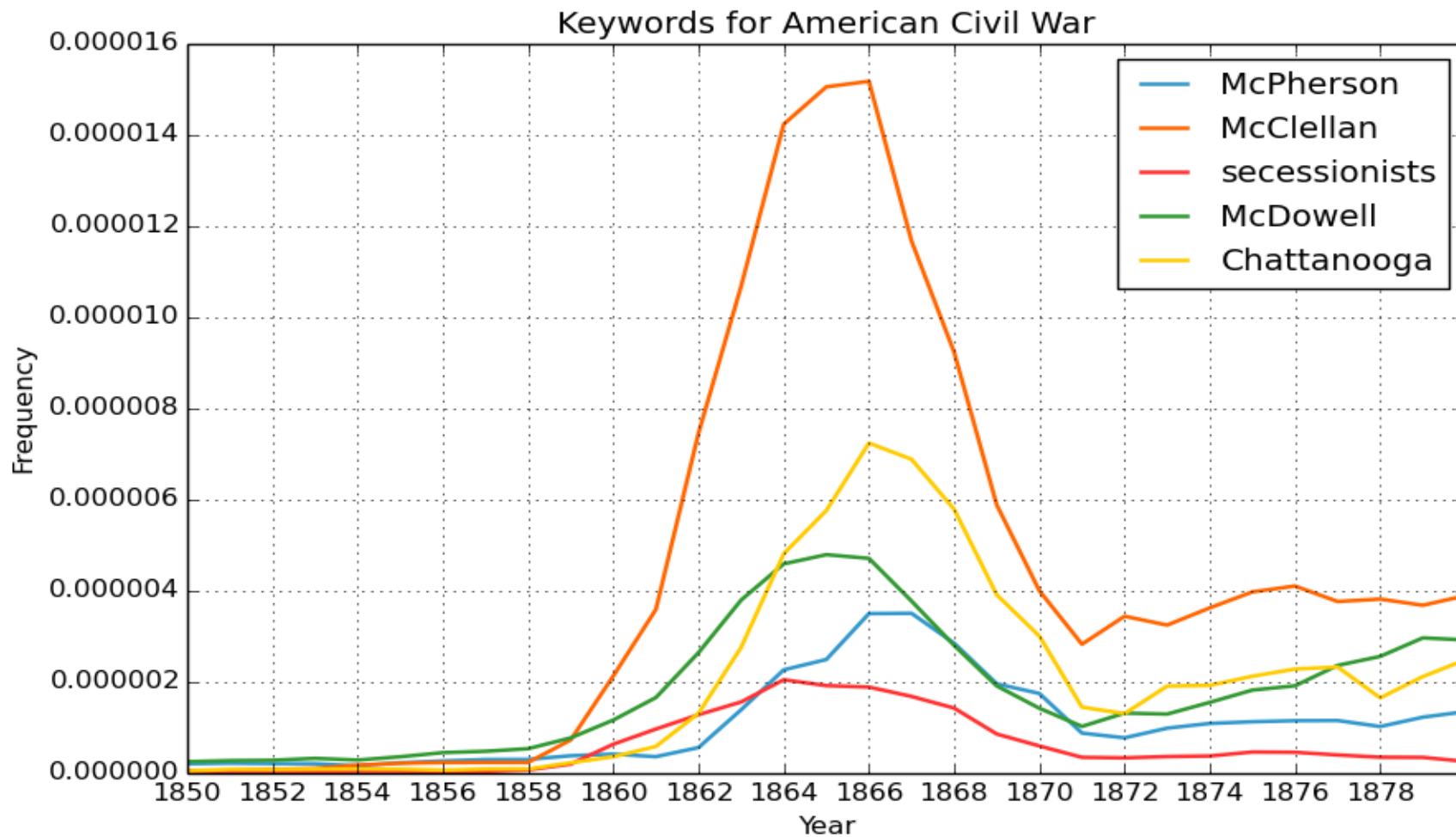
Comparații metode

Method	Divergence
level LD 1	61.05
level LD 2	69.10
sliding window SW 1	66.83
sliding window SW 2	66.78
sliding window SW 3	70.47
double change	64.02

Divergența Kullback-Leibler pentru toate evenimentele



Rezultate





LD1	LD2	SW1	SW2	SW3	D
McDowell,10	McDowell,10	McClellan,9	McClellan,9	Chattanooga,6	Chattanooga,10
Chattanooga,10	Chattanooga,10	McDowell,7	McDowell,7	secessionists,6	McPherson,6
secessionists,10	secessionists,10	secessionists,6	secessionists,6	McDowell,3	McDowell,4
McPherson,10	McPherson,10	McPherson,6	McPherson,6	McClellan,3	McClellan,4
McClellan,9	McClellan,10	Chattanooga,2	Chattanooga,2	McPherson,1	secessionists,2

Cele mai relevante cuvinte pentru “American Civil War” conform celor 6 metode



Keyword	Historical Relevance
Aguiyi	9
Passagem	9
cuius	9
Catholicism	9
Edo	9
Tangistanis	9
Bundy	9
Karelian	9
Jesuit	9
Sufragio	9

Cuvinte care definesc evenimentul “war”



- metodă nouă, funcțională de identificare a peak-urilor
- corpus-ul Google Books Ngrams pentru identificarea cuvintelor cheie



- determinarea delay-ului la care se produc anumite evenimente
- folosirea bigramelor, trigramelor

