

Info Session: Automated Processing of Citation Data

Rachael K. Hinkle

University at Buffalo, SUNY

January 5, 2018

Some Preliminaries

- ▶ You can download these slides and all relevant files at:
 - ▶ <http://rachaelkhinkle.com/research.html>
- ▶ To run the examples on your own you will need:
 - ▶ Python and Jupyter (both included in Anaconda)
 - ▶ Install Anaconda (www.anaconda.com)
 - ▶ Jupyter Notebook is a handy way to run Python code
 - ▶ Go to Terminal/Command Line and type:
`jupyter notebook`
 - ▶ Open the file named “HinklePyDemoSPSA2018.ipynb”

General Steps

1. Read in each separate .txt file.
2. Split each document into a chunk for each citation.
3. Track relevant metadata for each citation.
4. Parse citation chunk to extract desired info.
5. Export one row per citation to .csv file with all variables.

Input: Raw Text File

```

LA-019.2001.La.LEXIS.231ahep.txt -
Copyright 2014 SHEPARD'S(R) - 8 Citing references
State v. Profit, La. 00-1174, 778 So. 2d 1127, 2001 La. LEXIS 231 (La. Jan. 29,
2001)

Restrictions: Restrict between years 1900 - 2014
FOCUS(TM) Terms: No FOCUS terms
Print Format: FULL
Citing Ref. Signal: Hidden

SHEPARD'S SUMMARY

Restricted Shepard's Summary: Restrict between years 1900 - 2014
No subsequent
appellate history.
Citing References:
  Positive      Followed (2)
Analyses:

LexisNexis      HN1 (1), HN2 (3), HN3 (5), HN4 (2)
Headnotes:

Unrestricted Shepard's Summary
No subsequent
appellate history.
Citing References:
  Positive      Followed (2)
Sig Analyses:
nal
:
Pos
iti
ve

Other Sources: Treatises (1)

LexisNexis      HN1 (1), HN2 (3), HN3 (5), HN4 (2)
Headnotes:

PRIOR HISTORY ( 0 citing references )

(CITATION YOU ENTERED):
State v. Profit, La. 00-1174, 778 So. 2d 1127, 2001 La. LEXIS 231 (La.
Jan. 29, 2001)

CITING DECISIONS ( 8 citing decisions )

```

```

LA-019.2001.La.LEXIS.231shep.txt -
CITING DECISIONS ( 8 citing decisions )
LOUISIANA SUPREME COURT

1. Cited by:
State v. Williams, 65 So. 3d 153, 2011 La. LEXIS 1569 (La. 2011)
LexisNexis Headnotes HN3

65 So. 3d 153 p.153

LOUISIANA COURT OF APPEALS

2. Cited by:
State v. Cuevas, 57 So. 3d 605, 2010 La. App. Unpub. LEXIS 731 (La.App.
1 Cir. 2010) LexisNexis Headnotes HN2, HN4

57 So. 3d 605 p.605

3. Cited by:
State v. Shelby, 57 So. 3d 605, 2010 La. App. Unpub. LEXIS 715 (La.App.
1 Cir. 2010) LexisNexis Headnotes HN2, HN4

57 So. 3d 605 p.605

4. Followed by:
State v. Major, 26 So. 3d 289, 2009 La. App. LEXIS 1983 (La.App. 5 Cir.
2009)

26 So. 3d 289 p.295

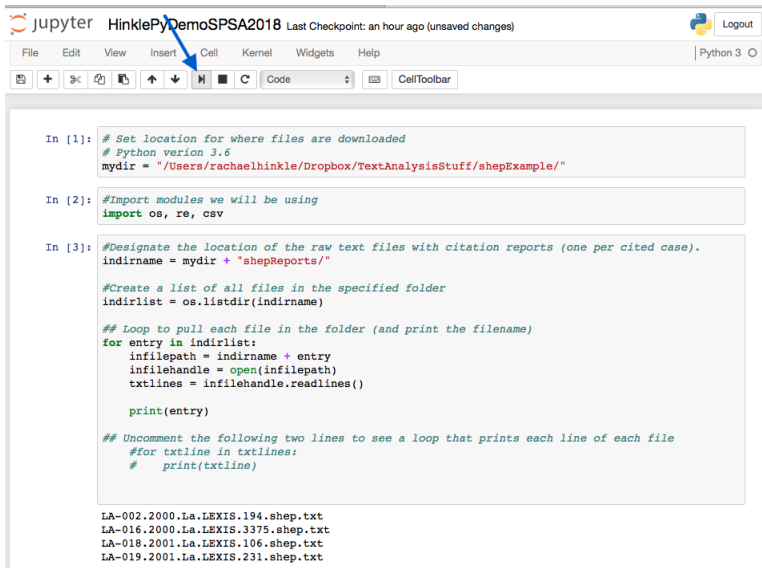
5. Cited by:
State v. Moore, 20 So. 3d 1137, 2009 La. App. LEXIS 1547 (La.App. 2 Cir.
2009) LexisNexis Headnotes HN3

```

Output: Spreadsheet with Citations

	A	B	C	D	E	F
	citedLexCite	CiteNum	CitingCase	CitingCourt	CitingYear	Treatment
2	2000 La LEXIS 194	1	738 So. 2d 207	PRIOR HISTORY	1999	
3	2000 La LEXIS 194	2	745 So. 2d 622	PRIOR HISTORY	1999	Writ granted by
4	2000 La LEXIS 194	3	110 So. 3d 557	LOUISIANA SUPREME COURT	2013	Cited by
5	2000 La LEXIS 194	4	104 So. 3d 401	LOUISIANA SUPREME COURT	2012	Cited by
6	2000 La LEXIS 194	5	809 So. 2d 923	LOUISIANA SUPREME COURT	2002	Cited by
7	2000 La LEXIS 194	6	754 So. 2d 927	LOUISIANA SUPREME COURT	2000	Cited by
8	2000 La LEXIS 194	7	19 So. 3d 594	LOUISIANA COURT OF APPEALS	2009	Cited by
9	2000 La LEXIS 194	8	19 So. 3d 590	LOUISIANA COURT OF APPEALS	2009	Cited by
10	2000 La LEXIS 194	9	989 So. 2d 199	LOUISIANA COURT OF APPEALS	2008	Cited by
11	2000 La LEXIS 194	10	980 So. 2d 769	LOUISIANA COURT OF APPEALS	2008	Cited in Concurring Opinion at, Cited
12	2000 La LEXIS 194	11	950 So. 2d 135	LOUISIANA COURT OF APPEALS	2007	Cited by
13	2000 La LEXIS 194	12	917 So. 2d 667	LOUISIANA COURT OF APPEALS	2005	Cited by
14	2000 La LEXIS 194	13	876 So. 2d 263	LOUISIANA COURT OF APPEALS	2004	Cited by
15	2000 La LEXIS 194	14	811 So. 2d 1	LOUISIANA COURT OF APPEALS	2001	Cited by
16	2000 La LEXIS 194	15	772 So. 2d 225	LOUISIANA COURT OF APPEALS	2000	Followed by, Cited by
17	2000 La LEXIS 194	16	2012 U.S. Dist. LEXIS 3295	4TH CIRCUIT - U.S. DISTRICT COURTS	2012	Cited by
18	2000 La LEXIS 194	17	2003 Cal. App. Unpub. LEXIS 11995	UNPUBLISHED CALIFORNIA APPELLATE DECI	2003	Cited by
19	2000 La LEXIS 194	18	898 So. 2d 1031	FLORIDA DISTRICT COURT OF APPEALS	2005	Cited by
20	2000 La LEXIS 194	19	201 Md. App. 23	MARYLAND COURT OF SPECIAL APPEALS	2011	Cited by
21	2000 La LEXIS 194	20	2008 MD Sp. App. Ct. Briefs LEXIS 2035	BRIEFS	2008	
22	2000 La LEXIS 3375	1	716 So. 2d 893	PRIOR HISTORY	1998	
23	2000 La LEXIS 3375	2	719 So. 2d 65	PRIOR HISTORY	1998	Reconsideration denied by
24	2000 La LEXIS 3375	3	740 So. 2d 200	PRIOR HISTORY	1999	Subsequent appeal at
25	2000 La LEXIS 3375	4	761 So. 2d 555	PRIOR HISTORY	2000	Writ granted by
26	2000 La LEXIS 3375	5	2013 La. LEXIS 2773	LOUISIANA SUPREME COURT	2013	Cited by
27	2000 La LEXIS 3375	6	842 So. 2d 330	LOUISIANA SUPREME COURT	2003	Cited by
28	2000 La LEXIS 3375	7	817 So. 2d 11	LOUISIANA SUPREME COURT	2002	Cited by
29	2000 La LEXIS 3375	8	807 So. 2d 218	LOUISIANA SUPREME COURT	2002	Cited by

Setup: Set directory, Load modules, Read in raw data



The screenshot shows a Jupyter Notebook interface with the title "HinklePyDemoSPSA2018" and a "Logout" button. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with various icons. A blue arrow points to the "Run" button in the toolbar. The notebook contains three input cells with the following code:

```
In [1]: # Set location for where files are downloaded
# Python version 3.6
mydir = "/Users/rachaelhinkle/Dropbox/TextAnalysisStuff/shepExample/"

In [2]: #Import modules we will be using
import os, re, csv

In [3]: #Designate the location of the raw text files with citation reports (one per cited case).
indirname = mydir + "shepReports/"

#Create a list of all files in the specified folder
indirlist = os.listdir(indirname)

## Loop to pull each file in the folder (and print the filename)
for entry in indirlist:
    infilepath = indirname + entry
    infilehandle = open(infilepath)
    txtlines = infilehandle.readlines()

    print(entry)

## Uncomment the following two lines to see a loop that prints each line of each file
#for txtline in txtlines:
#    print(txtline)
```

The output of the third cell shows the following file names:

```
LA-002.2000.La.LEXIS.194.shep.txt
LA-016.2000.La.LEXIS.3375.shep.txt
LA-018.2001.La.LEXIS.106.shep.txt
LA-019.2001.La.LEXIS.231.shep.txt
```

Create empty .csv file and name variables

```
#Create a .csv file and write the first row of the file with the desired variable names
fout = open(mydir + "citationData.csv", "w")
outfilehandle = csv.writer(fout, delimiter=",")

localrow = []
localrow.append("citedLexCite")
localrow.append("CiteNum")
localrow.append("CitingCase")
localrow.append("CitingCourt")
localrow.append("CitingYear")
localrow.append("Treatment")
localrow.append("RawText")
outfilehandle.writerow(localrow)
```

- ▶ Extract precedent citation (from filename)

```
for entry in indirlist:
    infilepath = indirname + entry
    infilehandle = open(infilepath)
    txtlines = infilehandle.readlines()

    #Need citation to the cited case: Here I extract it from the filename itself
    citedLex = re.sub("\.shep\.txt", "", entry)
    citedLex = re.sub("^[^\.]+\.", "", citedLex)
    citedLex = re.sub("\.", " ", citedLex)
    print(citedLex)
```

- ▶ Extract citing court (or other source)

```
# Extract the type of citing court (e.g., circled in blue in slides)
# This variable stays the same for each line until there is a new section
if re.search("^[A-Z0-9+][A-Z\.\']+", txtlines[i]) and not re.search("[0-9]{4} ", txtlines[i]):
    court_section = txtlines[i]
    court_section = re.sub("\([A-Za-z0-9\s]+\)", "", court_section)
    court_section = court_section.strip()
    court_section = re.sub(" ", "", court_section)
```


Extract Citation Blocks

```

#Finding the line containing CITE NUMBER and TREATMENT
if live_line and (re.search("[0-9]+\.",txtlines[i].strip()) or re.search("[[0-9]]{4} ", txtlines[i])):
    block = ""
    match = re.search("[0-9]\d+\.",txtlines[i])

    #create a single line for each citation entry to help with parsing
    #Code adds lines together until it reaches an empty line (or the end of the file)
    j=i
    while j<num_lines and txtlines[j].strip() != "" and not re.search("First Ref",txtlines[j]):
        block = block + " " + (txtlines[j]).strip()
        j=j+1

    block = re.sub(" \.", " ", block)
    block = re.sub("\s+", " ", block)
    block = block.strip()
    raw_text = block

```

What citation blocks look like (one per line)

```

2000 La LEXIS 3375
1. State v. James, La. 98-0428, 716 So. 2d 893, 1998 La. LEXIS 1020 (La. Mar. 27, 1998)
2. Reconsideration denied by: State v. James, La. 98-0428, 719 So. 2d 65, 1998 La. LEXIS 1732 (La. May 15, 1998)
3. Subsequent appeal at: State v. James, La. App. 98-2348, 740 So. 2d 200, 1999 La. App. LEXIS 2121 (La.App. 1 Cir. June 25, 1999)
4. Writ granted by: State v. James, La. 1999-3304, 761 So. 2d 555, 2000 La. LEXIS 1437 (La. May 12, 2000)
5. Cited by: State v. Duhe, 2013 La. LEXIS 2773 (La. Dec. 10, 2013) LexisNexis Headnotes HN1
6. Cited by: State v. Thompson, La. 2002-0333, 842 So. 2d 330, 2003 La. LEXIS 1090 (La. Apr. 9, 2003) LexisNexis Headnotes HN1
7. Cited by: State v. Benoit, La. 2001-2712, 817 So. 2d 11, 2002 La. LEXIS 1515 (La. May 14, 2002) LexisNexis Headnotes HN1
8. Cited by: State v. Lipscomb, La. 00-2836, 807 So. 2d 218, 2002 La. LEXIS 137 (La. Jan. 25, 2002) LexisNexis Headnotes HN1
9. Cited by: State v. Lagarde, La. 2000-1238, 778 So. 2d 585, 2001 La. LEXIS 431 (La. Feb. 9, 2001)
10. Cited by: State v. Carter, 2013 La. App. LEXIS 2799 (La.App. 4 Cir. Dec. 19, 2013)

```

Write row to .csv file for each citation

```
localrow = []
localrow.append(citedLex)
localrow.append(cite_number)
localrow.append(citing_case)
localrow.append(court_section)
localrow.append(citing_case_year)
localrow.append(treatment)
localrow.append(raw_text)

outfilehandle.writerow(localrow)

i=i+1

infilehandle.close()
fout.close()
```

Thank You.