



Text Data Mining: A Comparison of Sources

Ilana Stonebraker
Indiana University-Bloomington



Examples of Data Mining Projects @IU



COVID-19

Politicians who get sick with COVID-19 and affiliation



Wedding Vows

Psychology professor



Sociology

“Bowling Alone” and other studies



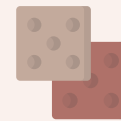
More Covid-19

Where people get their news



Post 9/11

Sentiment analysis of anger versus sadness



Folklore

Not sure about this one but they were asking



All of these
products have
emerged since
2020.

What's out there?

RIP: Nexis RestAPI

Factiva Data Feed	Pre-2020 way to get WSJ and other Dow Jones. One user
Nexis Data Lab	One account, 5-6 can shared account. capped 100,000 records
Proquest TDM Studio	WSJ, Washington Post, New York Times. 5 seats with online viz option
Constellate	Ithaka's product. Good for scholarly
Gale Digital Scholar Lab	Gale Primary Sources, No Code
Hathitrust Data Capsules *	* Only available for your Hathitrust collection

Key Takeaways

01

Journals = Dataset

02

What you get vs.
“see the edges”

03

Emerging field of
products

04

Jupyter Notebooks
is becoming the
platform

Ilana Stonebraker

Head, Business/SPEA Library

Indiana University- Bloomington

is1@iu.edu

