



ProQuest
TDM Studio



Analyze. Visualize. Connect. Discover.

Take a look inside TDM Studio, the powerful and versatile text and data mining solution from ProQuest.

Elevate the Value of Library Content with Text Analysis

In this age of information, tracking trends, patterns, and connections among millions of documents is becoming a key aspect of research. With TDM Studio, it is possible to efficiently access and analyze millions of documents across thousands of providers with consistent XML.

Extract new value from the content the library already subscribes to, save tens of thousands of dollars on additional rights clearances, and save the researcher months of time by making content available for text and data mining through TDM Studio. By elevating the value of your library content, you can support your university to achieve its research goals more efficiently.

TDM Studio offers two pathways to perform text and data mining:



For expert researchers using coding for text analysis:

The **Workbench** provides programmatic access to millions of ProQuest content using R or Python.



For faculty and students that want to visualize data quickly without coding:

Visualizations offer a way to interrogate ProQuest content and easily discover insights across thousands of documents.



about.proquest.com

To talk to the sales department, contact us at
1-800-779-0137 or sales@proquest.com.



Part of **Clarivate**

Two paths to new discoveries:



Workbenches in TDM Studio: For Researchers with Coding Experience

Designed for researchers who use the coding languages of R or Python.

TDM Studio Dashboard

Create projects and manage an unprecedented amount of content.

- Create datasets of up to 2 million documents each by searching publication titles or ProQuest database names
- Manage up to 10 datasets simultaneously for a total of 20 million documents
- Access the development environment, a Jupyter Notebook

| NAME | DESCRIPTION | QUERY | LOCATION | STATUS | DOCUMENT COUNT | DELETE |
|------------------------------|--|--------|---|-----------|----------------|--------|
| Dataset for company research | Examines how often company theories are mentioned in the Washington Post between 1910-1950 | Show + | tdm-site-data/130/corpus/datasetforcompanyresearch/ | Completed | 94,822 | |
| Demerita | | Show + | tdm-site-data/130/corpus/demerita/ | Completed | 2,095 | |
| COVID Daily Telegraph | | Show + | tdm-site-data/130/corpus/COVID/telegaph/ | Completed | 8,522 | |
| COVID June 4 | | Show + | tdm-site-data/130/corpus/COVID/June4/ | Completed | 129,052 | |
| WLU OFF | | Show + | tdm-site-data/130/corpus/WLUOFF/ | Completed | 47,622 | |
| COVID Jun 11 | Newspapers about COVID | Show + | tdm-site-data/130/corpus/COVID/June11/ | Completed | 36,358 | |
| TDM town hall | Cambridge 23 June | Show + | tdm-site-data/130/corpus/TDMtownhall/ | Completed | 112,847 | |

Create New Dataset

Customize and target the content set and avoid irrelevant data.

- Select from ProQuest subscriptions – all rights-cleared for TDM
- Curate a dataset across disciplines from current content or historical archives
- Benefit from the consistent data structure provided across content types – including current and historical newspapers, dissertations, scholarly journals and primary sources
- Filter and refine content to target the dataset specifically for your research topic, eliminating the need to process unrelated documents
- Specify topics by entering keywords and applying full Boolean search syntax
- Target specific date ranges, source types, and document types

88 | Create New Dataset

Selected Publications: 93,858 documents

- The Times of India (1947-16)
- The Times of India (1946) [TOP Stories]
- The Guardian News (1918)
- New York Times (1946-12) [TOP Stories]
- New York Times (1945)
- The Daily Telegraph (1948)
- Sunday Morning Herald (1916)

Date Published: [] [] []

Next: Review Dataset

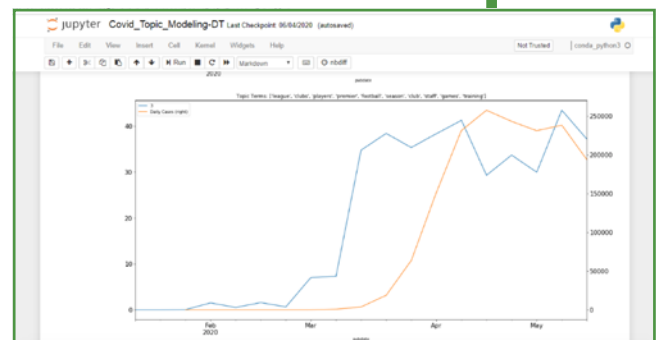
Jupyter Notebook

A familiar integrated development environment.

- Cloud-based development environment that can be accessed from anywhere – on or off campus
- Preconfigured with standard data science libraries in R and Python
- Leverage templated scripts to get started or create your own
- Upload your content and easily incorporate rights-cleared datasets – including open-access content and social media content

```
In [2]: # Data manipulation and modeling libraries
import pandas as pd
import sklearn
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import NMF
import matplotlib
import random

# Reading files and xml parsing (skipped)
import glob as glb
import xml.etree.ElementTree as ET
```





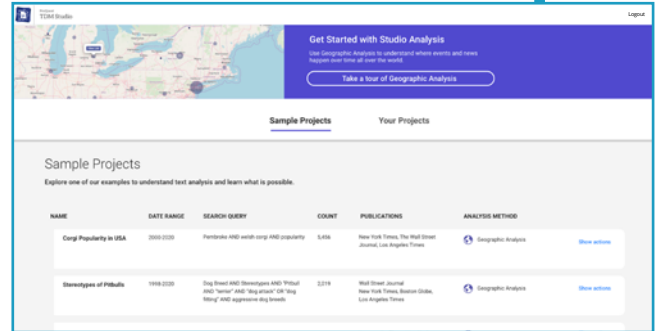
Data Visualizations in TDM Studio: For Discoveries Without Coding

Designed for users of all levels to quickly spot trends and generate insights.

Visualizations Dashboard

Create, access and manage projects.

- Launch several types of visualizations to explore supported data analysis methodologies
- Manage and access up to ten projects of 10,000 documents each
- Use for exploring research topics or teaching data science methodologies in classrooms across disciplines

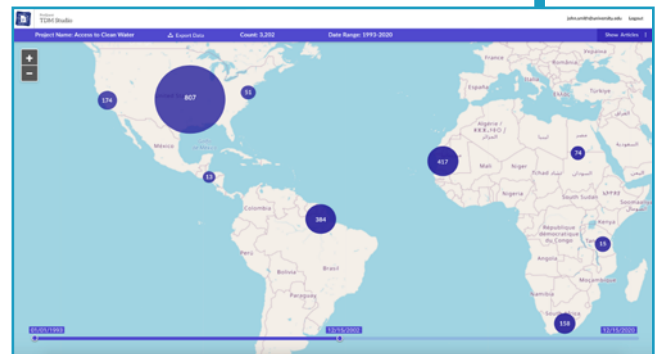


Powerful Visualizations in Minutes

Conduct research or teach data analysis with or without coding skills. These data visualizations provide the capability to enhance data literacy in the classroom, as part of a lab or workshop, or for individual research.

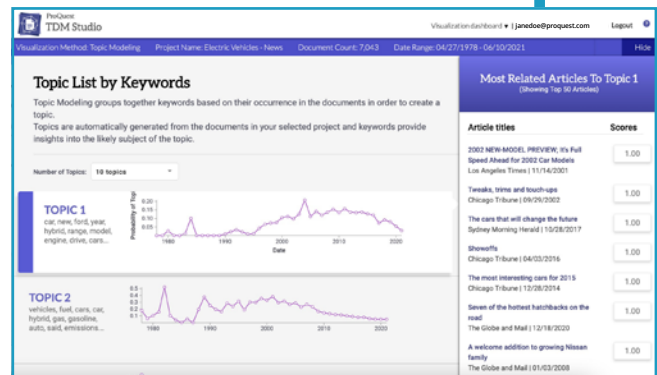
Geographic Analysis

- Map display shows dispersion of locations mentioned in the selected content
- Interactive display allows the exploration of patterns over time – or click on a cluster to explore documents associated with a specific locale



Topic Modeling

- Uncover topics and trends based on topic modeling approaches in the selected document dataset
- Gain a better sense of a topic, and find areas that may be overlooked. Topic modeling can help identify research topics, track trends, or conduct literature reviews

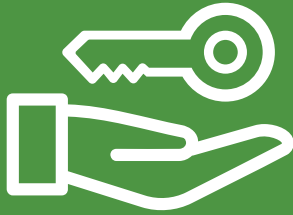


Sentiment Analysis

- Track trends in emotional language such as Anger, Disgust, Fear, Sadness, Happiness, Love, and Surprise over time and across publications
- Explore aggregate sentiment within your dataset, or drill down to explore individual documents
- Export the underlying data to use in your own analysis. Exported data includes document level emotion scores, aggregate scores, and article metadata

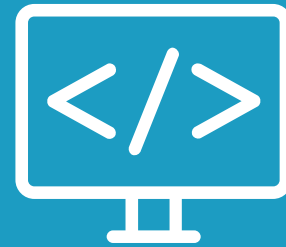


A text and data mining solution for research at all levels and across disciplines



Sought-After Content

TDM Studio offers access to rights-cleared content from ProQuest subscriptions and purchases. This content is applicable across disciplines and can be applied to many different research questions.



Flexibility

By combining an intuitive dataset creation capability, Jupyter Notebook access, and data visualizations, TDM Studio is able to support data literacy as well as research across disciplines for researchers with or without coding skills.



Efficiency

Significantly decrease the time to analysis, quickly create datasets, and either visualize the content relationships or access content in a consistent data schema across ProQuest sources.



Research, Teaching & Learning

Enhance research, teaching and learning outcomes across campus with a solution designed for varied skill levels in research methodologies and approaches.

Explore TDM Studio today.
Visit www.proquest.com/go/tdm-studio



about.proquest.com

To talk to the sales department, contact us at **1-800-779-0137** or sales@proquest.com.