Status Report

WINMAXXERS

June 24, 2026

Review of Goals

24hrs

DONE

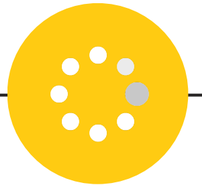
ONGOING

01

FINISH SCRAPING API

02

BUILD GRAPHS



03

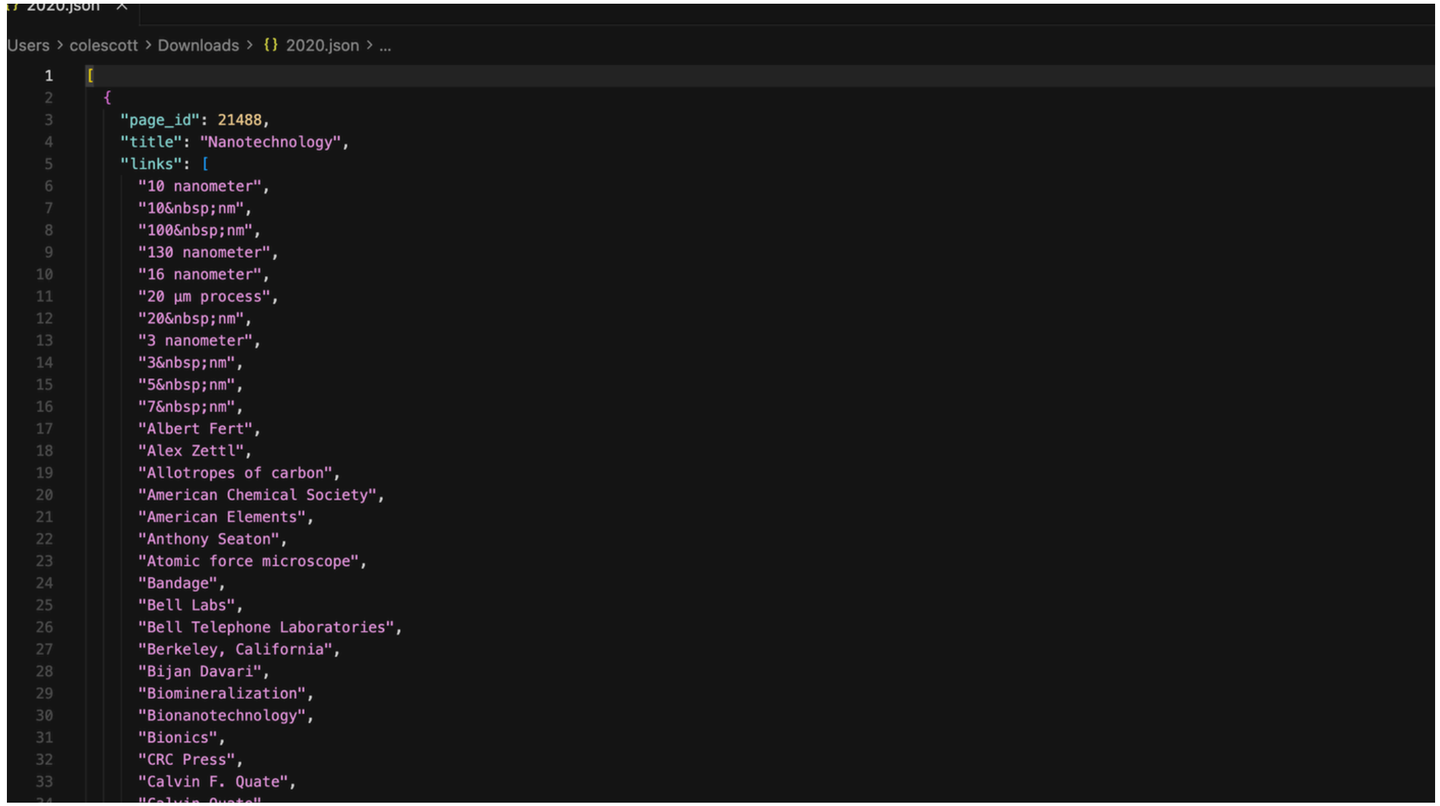
TRAIN MODEL

Goal # 1

SCRAPING THE API

- GENERAL STATUS

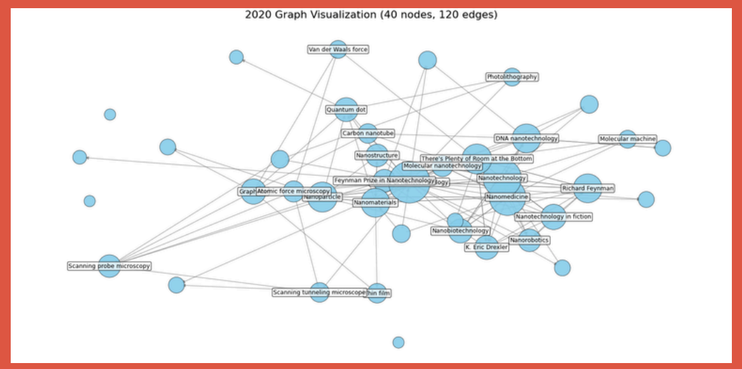
JSON data ready



Goal # 2

GENERATE GRAPHS AND TABLE

GENERAL STATUS



RELAITION FOR 2020 DATA

PROGRESS CHECK

- Completed at 3:45

OUR STEPS

- 1.Load into Pandas DataFrame
- 2. Clean data
- 3. Calculate graph features
- 4. Create feature table
- 5. Export CSV

MAIN FEATURES

- Centrality metrics — betweenness, closeness, eigenvector centrality
- Temporal degree change — growth rate of in/out/total degree over time
- Community evolution — whether a node is switching communities or becoming a bridge
- Assortativity features — whether it connects to high-degree or low-degree nodes
- Local clustering coefficient — how densely its neighbors connect

2020_features

page_id	title	in_degree	out_degree	total_degree	pagerank	betweenness_centrality	closeness_centrality	eigenvector_centrality	katz_centrality	hub_score	authority_score	clustering_coefficient	core_number	avg_neighbor_degree	avg_neighbor_pagerank	community_size	bridging_score
21488	Nanotechnology	146	41	187	0.029994709610346700	0.11876665353622000	0.18014001382574600	0	0.04982183756091740	0.01242747756834900	0.08087310600433570	0.01373908577298410	7	6.390243902439030	0.00100404765329094	135	0.00063173689008627
60786392	Nanotechnology in warfare	1	3	4	0.000243193853587188	2.93906590754324E-05	0.08100913074164	0	0.02657059213655690	0.00501328751855562	0.00031081922847317	0.33333333333333	4	19.333333333333	0.01012246921897600	135	5.87813181508648E-06
14431229	Outline of nanotechnology	1	89	90	0.000217555215358110	0.00010468477870892	0.00069832402234690	0	0.03856566081245430	0.03114859859042	0.00079278413336840	0.05268414481897630	7	6.325842696629210	0.00210368941795821	77	1.15038218361459E-06
7064233	History of nanotechnology	3	17	20	0.000882527192836928	0.00213814573334754	0.13646637383257	0	0.02883830301199660	0.00533936671543704	0.00574935144635320	0.21578947368421100	7	4.823529411764710	0.00421979337495085	123	0.00010181646349274
49615253	SD optical data storage	1	0	1	0.000524171841758717	0.0	0.00125698324022346	0	0.02601143477640850	0.0	-1.09939084831599E-19	0.0	1	0.0	0.00110136419783535	4	0.0

FEATURE CHART IS CSV FORMAT

Goal # 3

TRAIN MODEL

GENERAL STATUS

PROGRESS CHECK

- Came to conclusion on what model (XGBOOST) we are going to train and what would want our output to look like

NEXT STEPS

- Combine all yearly CSV files (2020–2026) into one dataset with a year column
- Create a “future PageRank” value by shifting each topic’s PageRank forward one year
- Use all graph-based features (like in-degree, betweenness, community_size, etc.) as inputs
- Train an XGBoost regression model to learn how these features influence next year’s PageRank
- Feed the model the 2026 data to predict 2027 PageRank
- Rank topics by predicted PageRank to identify emerging nanotechnology topics
- Visualize or summarize the top predicted topics for your final results

ACTION ITEMS

- Build model

Timeline

Combine CSV 1

2 Train Model

Validate Model 3

4 Notebook Polish & Create Presentation

Major Challenges

EASIEST

HARDEST

01

GENERATING THE GRAPHS

02

SELECTING CORRECT FEATURES

03

FINDING WHAT MODEL TO TRAIN

Thank you!

Do you have any questions?