



CHAMPION



CONQUER



VICTOR

VICTORY

SUCCESS

TRIUMPH

WINNER

WinMaxxers



- Team Members Names
 - Quintin Ampofo
 - Cole Scott
 - Damone Washington
 - Darell Barnes
- Future Event Prediction



https://drive.google.com/file/d/1RSo9XUcA_7guj58Akhidq8lYweRKV9hm/view?usp=sharing

<https://hackhpc.github.io/admi26>

Progress Check-in

The big picture: What will be the emerging concept or technology in the field of NanoTechnology?

Current Project Status

- Master list set completed using Category:Nanotechnology (depth-2 traversal)
- Final dataset size: ~1622 pages (nodes)
- Snapshot generator pipeline implemented and currently running (2020–2024 temporal extraction)

Progress So Far

- Wikipedia category crawler + master list generation completed
- Historical snapshot system built:
 - year-specific revision retrieval
 - wikitext link extraction
 - category extraction per page
- Persistence + retry logic added for stability under API rate limits

Interesting Finding

- Switched from **link-based depth expansion** to **subcategory-based expansion**
- This keeps the graph domain-consistent while still allowing structured growth
- Prevents noisy cross-domain expansion (e.g., physics, chemistry, general Wikipedia drift)

```
Saved (753 pages)
[754/1622] Processing: Surface magnon polariton
Saved (754 pages)
[755/1622] Processing: Surface plasmon
Saved (755 pages)
[756/1622] Processing: Surface plasmon polariton
Saved (756 pages)
[757/1622] Processing: Surface plasmon resonance microscopy
Saved (757 pages)
[758/1622] Processing: Surface-enhanced Raman spectroscopy
Saved (758 pages)
```

```
data-prep
├── generate_snapshot_da... U
├── {} master_pages.json U
├── WikiScape.py U
└── {} 2026.json U
```

Progress Check-in

Technology / Resources

- Wikipedia MediaWiki API
- Python (requests, regex, JSON pipeline)
- Custom retry system + rate limiting + checkpoint saving

Next 24 Hours

- Complete snapshot generation for all years (2020–2024)
- Validate failed pages and rerun missing data
- Begin graph construction from snapshots
- Start feature engineering (PageRank, degree, communities)

Bottlenecks / Risks

- API rate limiting → long runtime (~20k+ requests total)
- Some revision lookups intermittently failing
- Full pipeline is compute/time heavy but stable due to checkpointing

