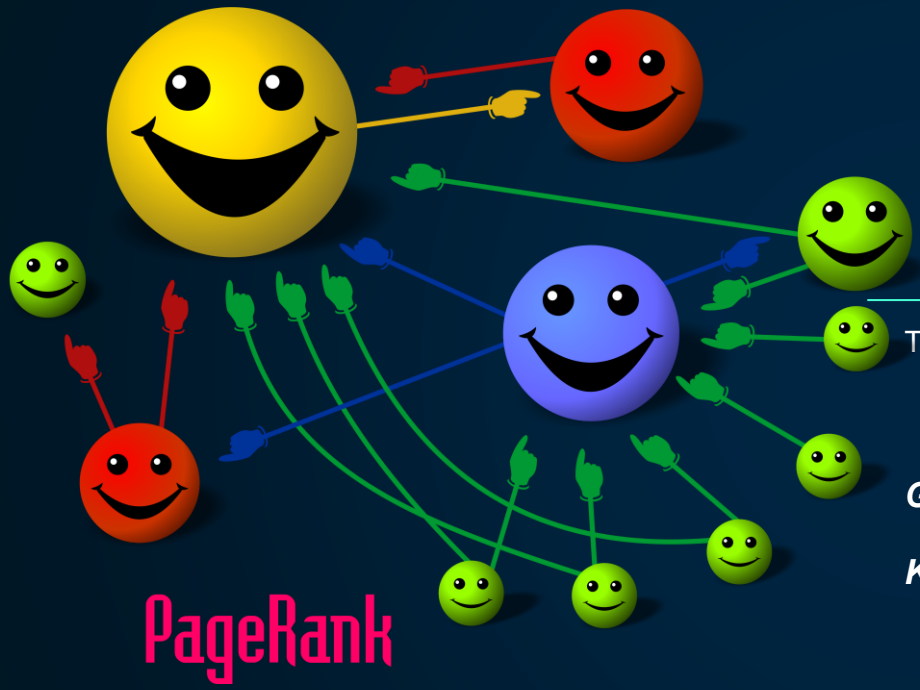

# TOMFoolery Hackathon 2025 |PageRank

29.11.25




# About PageRank

The web can be seen as a **directed graph**

- **Nodes** = webpages
- **Edges** = hyperlinks (A → B means “A links to B”)

***Goal:*** assign an “importance” score to each page using only links

***Key idea:***

- A page is important if **many pages link to it**
- Links from **important pages count more** than links from unimportant pages = *recursively*

https://generic.wordpress.soton.ac.uk/britizen/2017/05/05/exploring-the-pagerank-algorithm/

# Algorithm

$$r^{(k+1)} = \alpha \cdot P^T r^{(k)} + (1-\alpha) \cdot \frac{1}{N} 1$$


### *Input*

- Crawled link graph (JSON): {page_url: [linked_url1, linked_url2, ...]}
- Parameters: damping α=0.85, tolerance tol, max iterations max_iter

### *Idea (Random Surfer)*

- With prob. α: follow a random outgoing link
- With prob. 1−α: jump to a random page
- PageRank = long-run probability of being on each page

### *Computation (Power Iteration)*

- Initialize scores equally: r(0)=[1/N,…,1/N]
- ***Repeat until stable:***
  - **Link spread:** each page splits its score evenly over its outgoing links
  - **Dangling pages:** if no outgoing links, spread score to all pages
  - **Damping:** new_score = (alpha * score_from_links) + ((1 - alpha) * equal_share_to_all_pages)
- Stop when the total change between old and new scores is smaller than tol

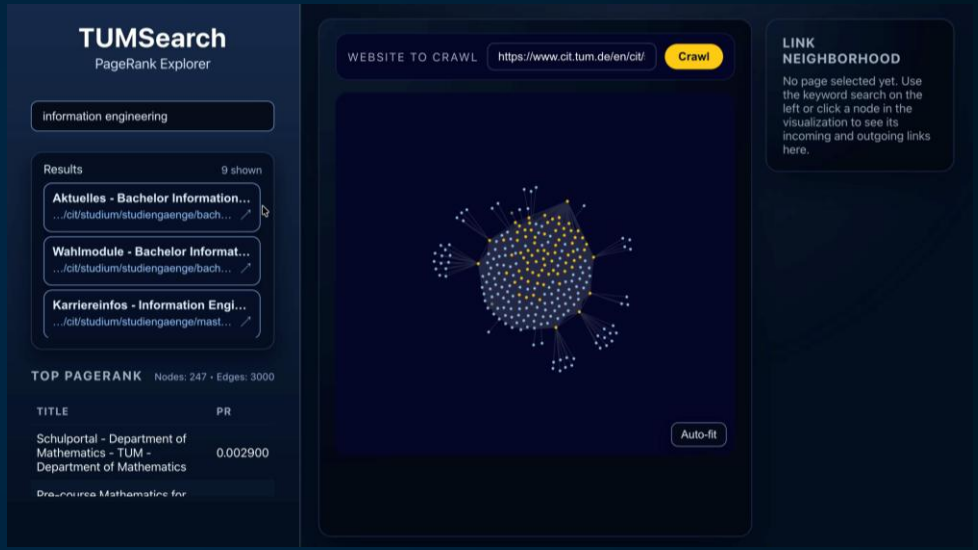
### *Output*

- Score per page (sums to ~1), sorted → **Top 10 most “important” pages**

# Our Implementation

***Overview:***
Interactive web app that crawls websites, computes PageRank, and visualizes the internal link network. Uses a force-directed graph for intuitive exploration of pages and authority.

***Functions:***
**Website Crawler:** Asynchronous crawling, internal links extraction, page title detection, filters non-HTML content.
**PageRank Computation:** Builds directed hyperlink graph, computes PageRank, highlights high-authority pages.
**Interactive Visualization:** Node size & color show PageRank, hover for metadata, click to explore links, smooth layout.
**Keyword Search:** Search pages by title/URL and jump directly to nodes.



# Relevance

***Cybersecurity & Web / IT Security Hygiene***

***Internal Admin & Knowledge / Document Management***

***Website Usability & Student Portal Experience***

***Academic / Research Resource Discovery***

# THE TEAM

Isabelle Berg

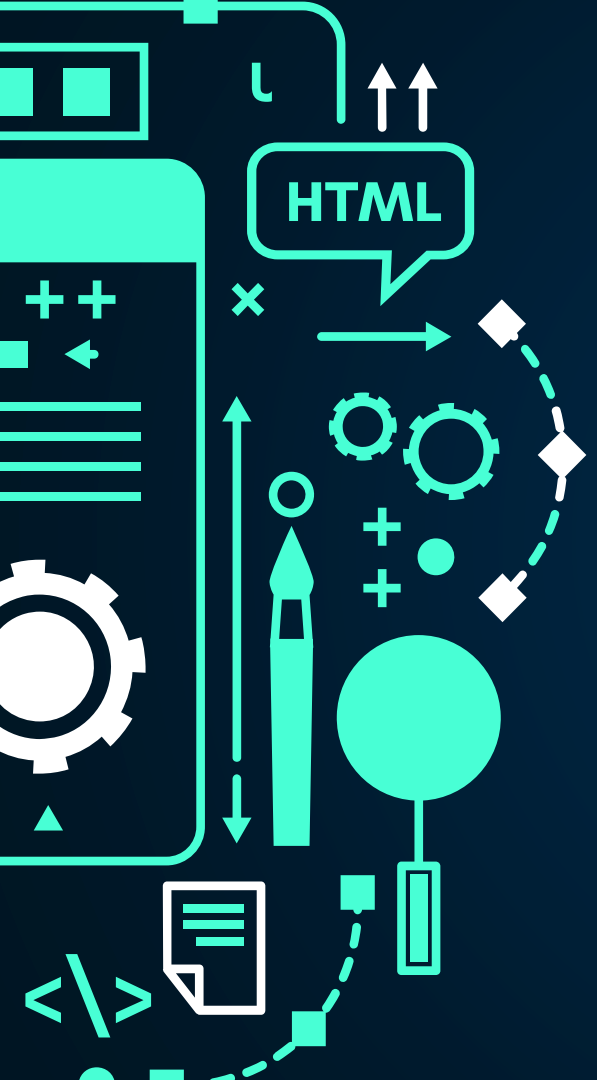
Mariia Andrusiv

Nicolas Popken

Tim Hufnagel

Furgan Siyahov

Antonio Jose Jalandoni




# Thank you for listening!