



An Efficient Text Analysis Tool for Streamlining Literature Reviews

Oliver Bonham-Carter, Ph.D.

Dept of Computer and Information Science
Meadville, PA 16335

<https://cis.allegheny.edu/>

FLAIRS 2025
21th-23st May 2025
Daytona Beach, Florida, USA

Academic Research

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction

Troubles In Research

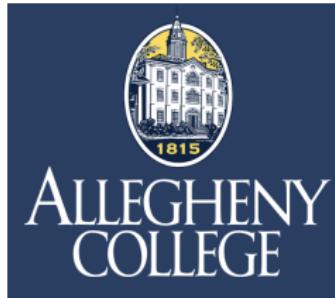
Visualizations

Methods

Results

Conclusions

Reach out!



Senior Academic Research Projects

- **Third Year** complete a seminar course to provide research fundamentals
- **Forth Year** complete original academic research project in major (e.g., *comping*)
- **Faculty** members mentor senior projects

Success in Academic Research

Generally said, ...

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

Visualizations

Methods

Results

Conclusions

Reach out!



How do I know when I am doing things right?

- *Success results when the research question meets one's skill*

Slow Starts in Beginner Projects

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations
Methods
Results
Conclusions
Reach out!

Starting a research project can be overwhelming for novice researchers, often due to the difficulty in narrowing down a clear research question amid a **flood of information**. This often leads to frustration and delays in getting started.

Challenges in Research

- Beginning the Project itself!
- Formulating clear hypotheses (Knox *et al.*)
- Searching for credible sources (Alisic *et al.*)
- Overcoming Imposter Syndrome (Ossovski *et al.*)

Challenges in Beginning Research

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

Visualizations

Methods

Results

Conclusions

Reach out!

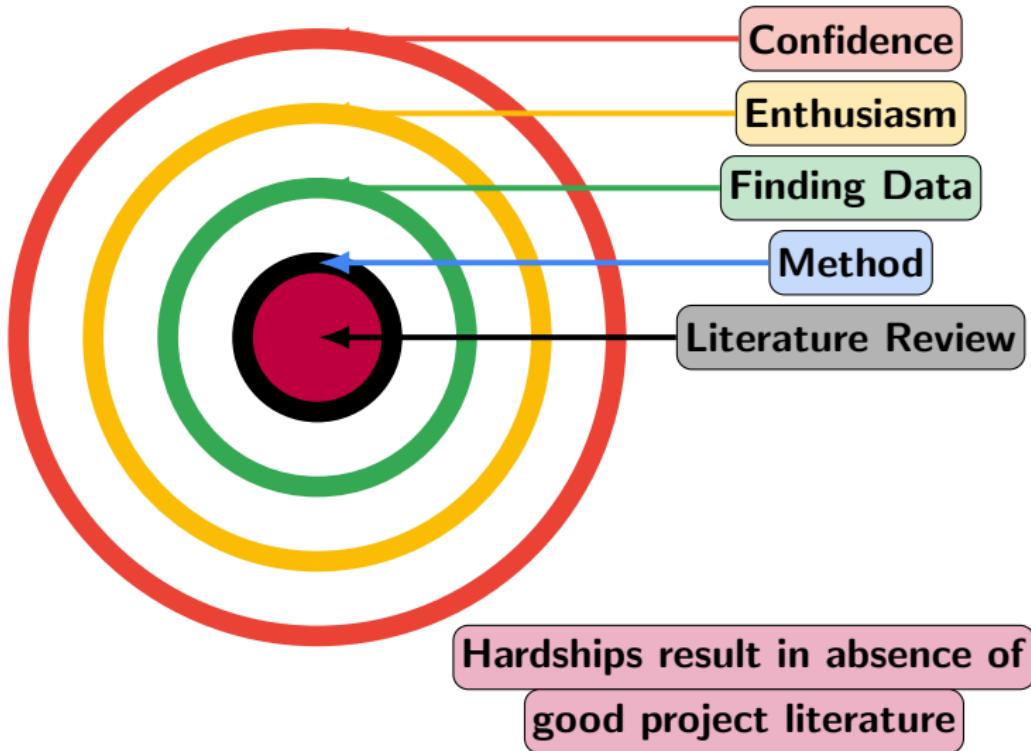


The Cause of These Problems?

- Difficulty in discerning relevant literature for idea
- Literature reviews are insufficient
 - Difficulty in identifying an *feasible* research topic
 - Difficulty in developing a research question
 - Lack of enthusiasm and inspiration

My Experience; One Cause of Troubles

After an area of interest has been established



Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

Visualizations

Methods

Results

Conclusions

Reach out!

Which Books Contain the Best Ideas??

Finding an idea across a lots of books is hard!

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

Visualizations

Methods

Results

Conclusions

Reach out!



- There are often too many results to be helpful to initial steps to create a literature review

Which Article Contains the Best Ideas??

Finding an idea across a lots of articles is also hard!

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

Visualizations

Methods

Results

Conclusions

Reach out!

Articles	About 16,100 results (0.14 sec)
Any time	Society of surgical Oncology Consensus statement: assessing the evidence for and utility of gene expression profiling of primary cutaneous melanoma
Since 2025	, ... of Surgical Oncology Gene Expression Profiling ... - Annals of Surgical ..., 2025 - Springer
Since 2024	Introduction Gene expression profiling (GEP) of primary cutaneous melanoma aims to offer prognostic and predictive information to guide clinical care. Despite limited evidence of ...
Since 2021	☆ Save 99 Cite Cited by 7 Related articles All 4 versions
Custom range...	
Sort by relevance	Glycosylation gene expression profiles enable prognosis prediction for colorectal cancer
Sort by date	R Li, S He, T Qin, Y Ma, K Xu, S Liu, W Zhan - Scientific Reports, 2025 - nature.com
Any type	This study developed a prognostic model for patients with colon adenocarcinoma (COAD) based on glycosylation-associated genes. By analyzing TCGA-COAD data, 110 key genes ...
Review articles	☆ Save 99 Cite Cited by 2 Related articles All 5 versions
<input type="checkbox"/> include patents	
<input checked="" type="checkbox"/> include citations	[HTML] Development, characteristics, and gene expression profiles of a chromosomal deletion line of Brassica juncea (AABB, 2n= 36) with apetalous feature derived ...
<input type="checkbox"/> Create alert	M Hu, Q Hu, J Yu, L Zhou, D Zuo, M Cai, L Gu... - Plant Physiology and ..., 2025 - Elsevier
	... Twenty-four differentially expressed genes which may be associated with the regulation of ABJ's apetalous traits were identified by gene expression profiling. To our knowledge, there ...
	☆ Save 99 Cite Cited by 2 Related articles All 4 versions

- Note: 16K results just for the year 2025

We May Ask

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations
Methods
Results
Conclusions
Reach out!

**How do we address
this informational
overload?**

Visual Learning

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

Benefits

- **Visual Learning:** Instruction based on visual cues and inputs from graphics, less dependence on reading text.
- Visual learning is often more productive for young people than non-visual learning

Published Work

- Visuals enhance cognitive engagement [Tong, Yang, and Chen2024]
- Diagrams, charts, and videos improve memory retention [Ahmad and Abd Halim2024]
- Interactive tools boost motivation and focus

BeagleTA: Text Analysis

Introducing our tool to bring visualizations into literature reviews searches

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!



BeagleTA

- A literature parsing system to help locate articles of specific topics
- Output is **visual** and **interactive** to help find relevant articles according to keywords

Local Corpus

Non-commercial article data

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

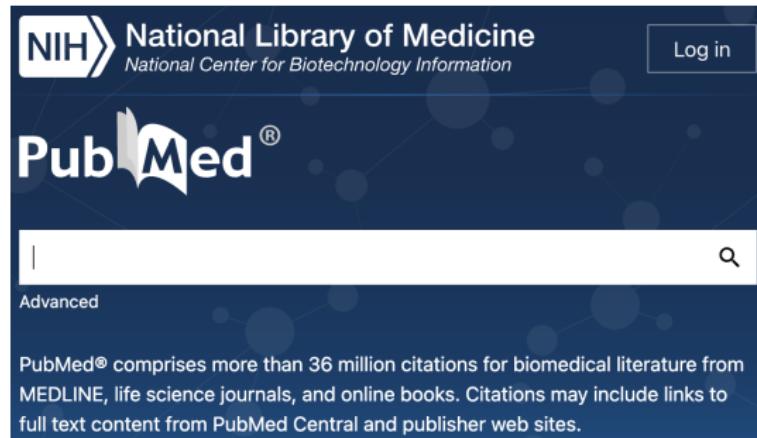
Visualizations

Methods

Results

Conclusions

Reach out!



- PubMed: Millions of articles in STEM science and other fields
- https://ftp.ncbi.nlm.nih.gov/pub/pmc/oa_bulk/oa_noncomm/

Bundled Tools

Because some analyses are involved!

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

Three Components



Parser: To isolate articles based on a *bag of words* approach

Cleaner: Used to remove noise and irrelevant (*zero-value*) results

Selector: Used to make random selections to test general trend theory across large datasets.

- Supporting different types of article analyses
- Corpus quality may be studied for its content

Workflow of Parser

The initial step

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research

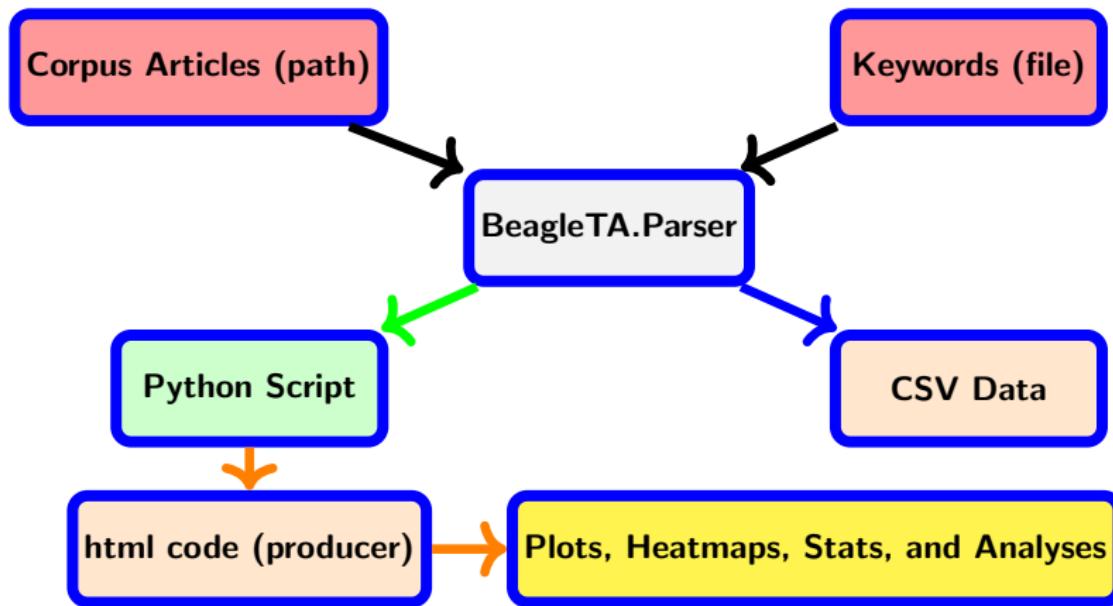
Visualizations

Methods

Results

Conclusions

Reach out!



Setup For Usage: I

Python code generated for interactive plots (html)

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

Create a Virtual Environment

```
python3 -m venv ~/Desktop/venv  
source ~/Desktop/venv/bin/activate  
pip install plotly matplotlib numpy pandas  
networkx seaborn scikit-learn pyvis
```

Virtual Environment for Python Script

- A virtual environment for running Python code to create plots, heatmaps, stats and analyses docs (html)
- The script is generated automatically

Setup for Usage II

Written in Rust For Speed

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

Launching BeagleTA's Parser

```
cargo run --bin 1_beagleTA_parser  
examples/kws.txt  
examples/corpus_short
```

Written in Rust

- Rust must be used to compile the code
- Here we demonstrate on example data include in GitHub repository

The Screenshot of Running Parser

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

```
venv) obonhamcarter@elivers-MacBook-Pro BeagleTA % cargo run --bin 1_beagleTA_parser examples/kws.txt examples/corpus_short
    Finished 'dev' profile [unoptimized + debuginfo] target(s) in 0.29s
    Running `target/debug/1_beagleTA_parser examples/kws.txt examples/corpus_short`  
  
Package name: 'BeagleTA'.  
Package version: '1.5.11'.  
Package edition: '2021'.  
  
Usage: cargo run <keyword_file.txt> <corpus_dir>  
Found 63 files in the directory 'examples/corpus_short'.  
Key words : ["2024", "ethic", "bio", "machine", "mitochondria"]  
  
# Run commands to create virtual environment with Python for the following commands ...  
python3 -m venv ~/Desktop/venv  
source ~/Desktop/venv/bin/activate  
pip install plotly matplotlib numpy pandas networkx seaborn scikit-learn pyvis  
  
# Run commands to create keyword analysis and other visualizations ...  
cd 0_out && python3 visualizations.py && cd ..  
cd 0_out && python3 interactive_network_1.py --datafile output.csv --threshold 6 && cd ..  
cd 0_out && python3 interactive_network_2.py --datafile output.csv && cd ..  
cd 0_out && python3 complete_network.py -f output.csv -t 6 -nc blue -ec green && cd ..
```

Custom Generated Python Script

- After parsing articles for keywords, suggested commands are presented to further the analysis

We May Ask

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations
Methods
Results
Conclusions
Reach out!

**What are the
visuals?**



Articles Found From Keywords

Analysis of keyword performance

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-CARTER,
Ph.D

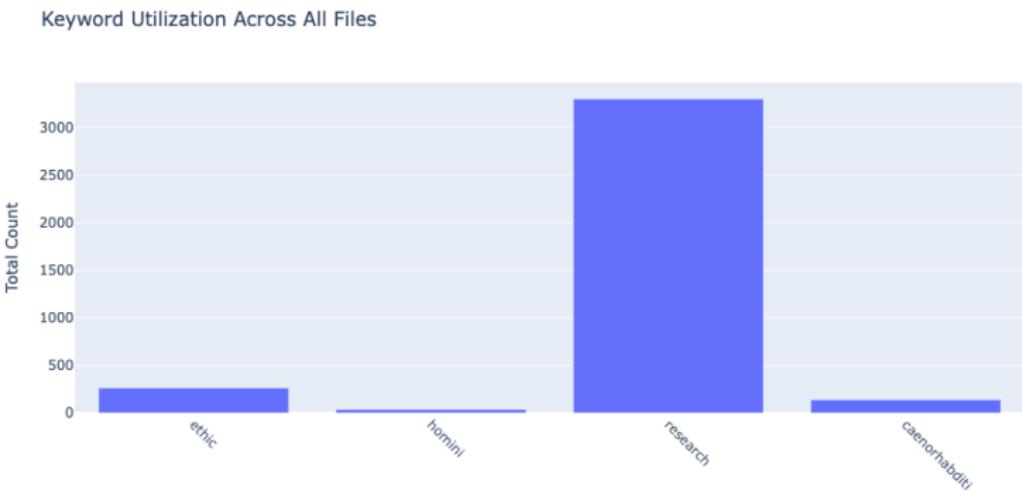
Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!



- Magnitudes represent amount of articles returned by the keywords
- Some keywords can be too general, others not specific enough

K-Means Clustering of Cosine Similarities

Find the Diversity in Results

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

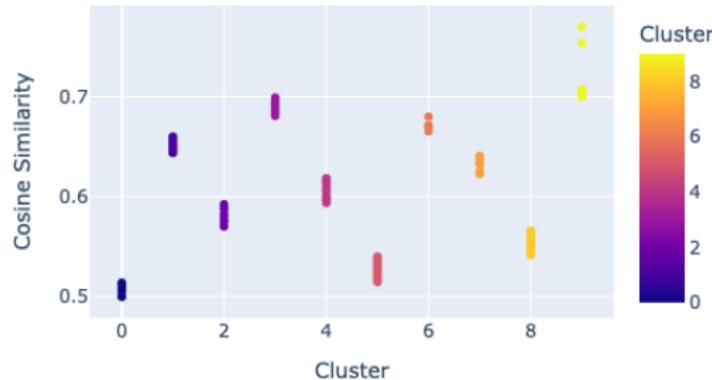
Methods

Results

Conclusions

Reach out!

K-Means Clustering of Cosine Similarities



- K-Means clustering of cosine similarity to show diversity in results



Heatmaps

Determine keyword usage patterns in articles

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-CARTER,
Ph.D

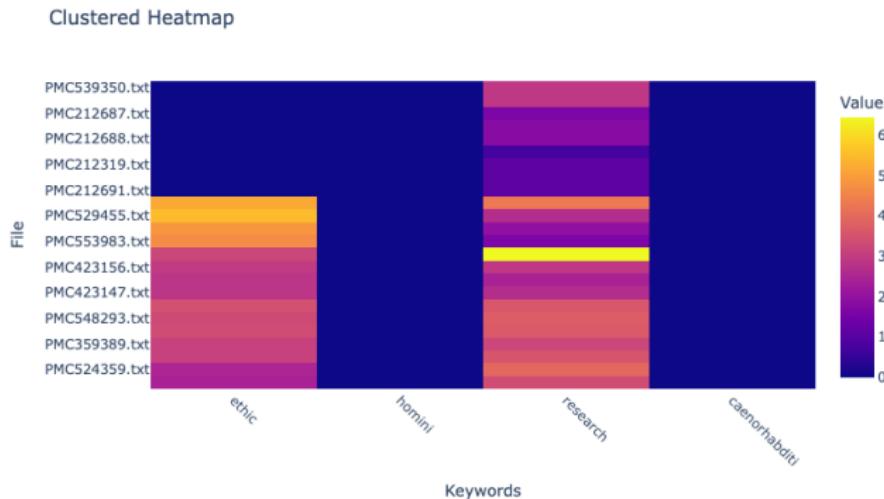
Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!



- Heatmaps show keywords found across the articles
- Articles having more keywords are likely to be better fits for a literature review

Network I

Determine how articles are connected by keyword content

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-CARTER,
Ph.D

Introduction

Troubles In Research

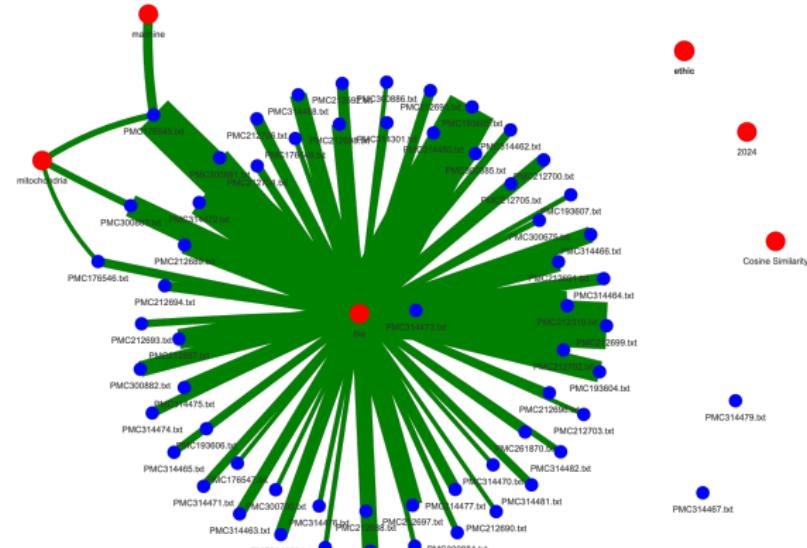
Visualizations

Methods

Results

Conclusions

Reach out!



- Blue nodes are articles, Red nodes are keywords, edges are connections by common keywords
- Node are labeled with document's ID



Network II

Finding the relevant articles of multiple keywords

Enhancing
Student
Research with
Visual Tools

Oliver
**BONHAM-
CARTER,**
Ph.D

Introduction
Troubles In Research

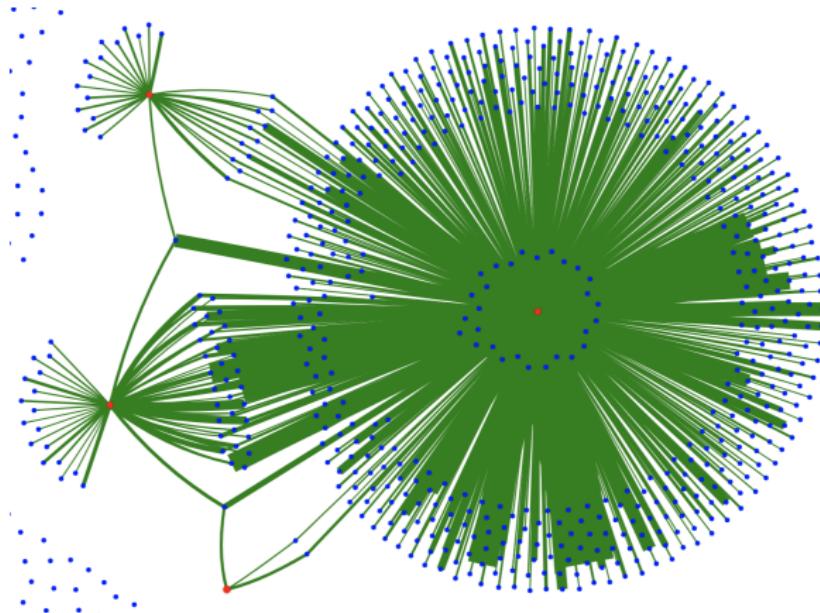
Visualizations

Methods

Results

Conclusions

Reach out!



- Blue nodes connected to multiple red nodes contain more user-selected keywords

Concluding Thoughts

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!

- BeagleTA provides a convenient way to choose foundational articles using interactive and visual outputs
- Reduced frustration when beginning research projects
- keyword usage can (itself) be studied for performance
- Beginning researchers may become energized for own work when creating literature reviews are more *simple* to complete.

Future Works

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

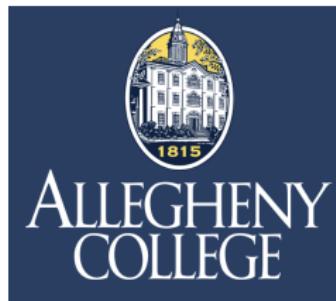
Introduction
Troubles In Research
Visualizations

Methods

Results

Conclusions

Reach out!



- Working on plans to put this tool in the hands of beginners in my research classes
- It is hoped that more students continue into research areas due to improved self esteem.

Enhancing
Student
Research with
Visual Tools

Oliver
BONHAM-
CARTER,
Ph.D

Introduction
Troubles In Research
Visualizations
Methods
Results
Conclusions
Reach out!



Contact and References

- obonhamcarter@allegheny.edu
- <https://www.oliverbonhamcarter.com/>
- <https://github.com/developmentAC/BeagleTA>