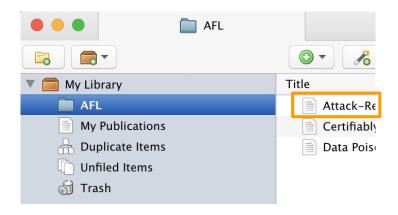
RAGS

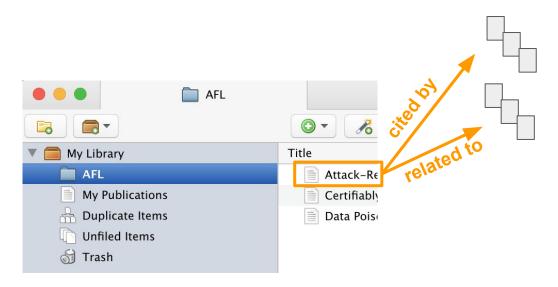
Research Assistant for Graduate Students

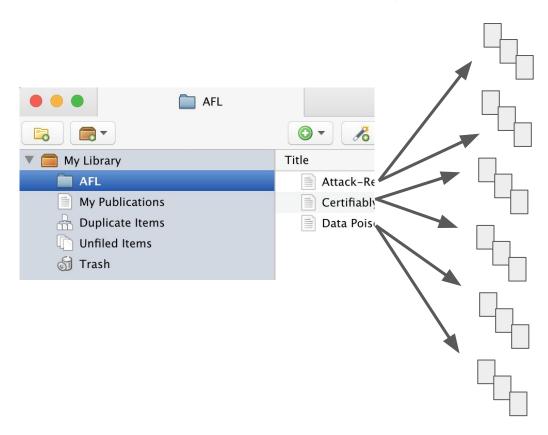
Shaobo Cui, Ryan Sullivan, Kamala Varma

Outline

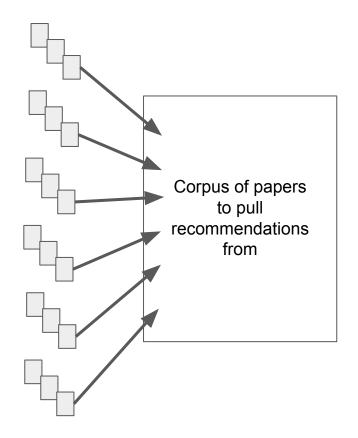
- Paper data collection and storage
- Recommendation algorithm and data used
- User interface
- Demo
- Limitations and future work











Overview of Paper Data Access-Related Functionality

```
def title_from_scholar_search_result(url):
    def find_end_of_title(s, start_ind):
    def paper_data_from_zotero(n=-1):
    def dictionary_from_google_scholar(title):
```

Get paper info from Zotero library given user ID or from Google Scholar given paper title or url of search result Get full paper text or extract paper abstract

```
def text_from_paper_dict(paper, abstract=True):
    def text_from_google_scholar(title, abstract=True):
    def full_text_from_pdf(pdf):
    def abstract_from_pdf(pdf):
    def abstract_from_pub_page(url):
    def extract_abstract(text):
```



Overview of Paper Data Access-Related Functionality

```
def title_from_scholar_search_result(url):
def find_end_of_title(s, start_ind):
def paper_data_from_zotero(n=-1):
      dictionary_from_google_scholar(title):
        Scholar About 101 results (0.04 sec)
   Certifiably-Robust Federated Adversarial
                                                   [PDF] ieee.org
   Learning via Randomized Smoothing
   C Chen, B Kailkhura, R Goldhahn... - 2021 IEEE 18th ..., 2021 -
   ieeexplore.ieee.org Paperpile
   Federated learning is an emerging data-private distributed learning framework,
   which, however, is vulnerable to adversarial attacks. Although several heuristic
   defenses are ...
   ☆ Save 57 Cite Cited by 1 Related articles All 5 versions ১৯
   Federated learning in adversarial settings
                                                   [PDF] arxiv.org
   R Kerkouche, G Ács, C Castelluccia - arXiv preprint
```

Federated Learning enables entities to collaboratively learn a shared prediction

arXiv:2010.07808, 2020 - arxiv.org

Strict assumptions about what defines the start and end of an abstract

```
def text_from_paper_dict(paper, abstract=True):
def text_from_google_scholar(title, abstract=True):
def full_text_from_pdf(pdf):
def abstract_from_pdf(pdf):
   abstract_from_pub_page(url):
   extract_abstract(text):
```

Paper Data Storage-Related Functionality

```
def save_all_data(papers, fname):
def add_title_to_zotero(titles):
                Unfiled Items
      My Library
                              Title
        AFL
        My Publications
        Duplicate Items
      Unfiled Items
     Trash
     Group Libraries
```

Recommendation algorithm

- Content-based filtering:
 - Paper vectorized with SciBERT, sentence by sentence
 - Average of per-sentence paper vectors represent the paper
 - BERT summarization performed on text to reduce memory cost
 - Alternatively: use abstract instead of text
 - User profile formation:
 - For each paper added, add a new vector to the profile
 - Recommendation:
 - For each paper in corpus, compute cosine similarity with each paper in profile
 - Aggregate cosine similarities, with optional weight
 - Recommend papers with top similarity

Data used in recommendation

- Paper identifier
 - Using title as identifier
- Vectorized text/abstract representation
- Corpus inclusion information
 - Using keywords here

User Interface

- Simple text-based interface for setting up projects and requesting recommendations
- Users can create projects associated with their Zotero collections and get recommendations for those papers
- Use Zotero to modify papers used for recommendations
- Saves user settings and projects across uses
 - Create new project
 Get recommendations
 Exit
 Select which action you would like to perform: 1
 Choose a project name: Context-Aware Systems

User Interface

```
    Create new project

Get recommendations
3. Exit
Select which action you would like to perform: 2

    Context-Aware Systems

Recommender Systems
Ontologies
Select project id: 1
Recommended:
Effectful Program Distancing
From Enhanced Coinduction towards Enhanced Induction
Oblivious Algebraic Data Types
Successfully added Effectful Program Distancing to Zotero
Successfully added From Enhanced Coinduction towards Enhanced Induction to Zotero
Successfully added Oblivious Algebraic Data Types to Zotero
```

Documentation

RAGS: Research Assistant for Graduate Students

Instructions

Installing necessary packages

- follow setup instructions for pyzotero
- follow setup instuctions for scholarly

Usage

To use the project, run the run project script:

1 python run project.py

This will take you through setting up your user profile and first project. You will need to get and API key and the name of your personal zotero library using the instructions for <u>pyzotero</u>. These configurations are stored in <u>config.ini</u> and can be edited at any time.

After configuration, the script will provide you with the options of creating a new project, getting recommendations for an existing project, or exiting the application. Any papers searched during the process of producing recommendations will be added to a growing corpus, which is used to provide recommendations more quickly in the future. This corpus is saved in corpus.pt.

Graphical User Interface Design

Lorem ipsum dolor sit amet, consectetur adipiscing elit,

sed do eiusmod tempor incididunt ut labore et dolore

CMSC 818G Project Personal Project Recommended Papers Generate Recommendations Edit Project CMSC 818G Project Title Title Title Abstract Abstract Abstract Deep Learning Lorem ipsum dolor sit amet, consectetur adipiscing elit, Lorem ipsum dolor sit amet, consectetur adipiscing elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore sed do eiusmod tempor incididunt ut labore et dolore sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud magna aliqua. Ut enim ad minim veniam, quis nostrud magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo exercitation ullamco laboris nisi ut aliquip ex ea commodo exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in consequat. Duis aute irure dolor in reprehenderit in consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. voluptate velit esse cillum dolore eu fugiat nulla pariatur. voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in Excepteur sint occaecat cupidatat non proident, sunt in Excepteur sint occaecat cupidatat non proident, sunt in HCI culpa qui officia deserunt mollit anim id est laborum. culpa qui officia deserunt mollit anim id est laborum. culpa qui officia deserunt mollit anim id est laborum. Sed ut perspiciatis unde omnis iste natus error sit Sed ut perspiciatis unde omnis iste natus error sit Sed ut perspiciatis unde omnis iste natus error sit voluntatem accusantium doloremque laudantium, totam voluptatem accusantium doloremque laudantium, totam voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et rem aperiam, eaque ipsa quae ab illo inventore veritatis et rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo quasi architecto beatae vitae dicta sunt explicabo. Nemo quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut enim ipsam voluptatem quia voluptas sit aspernatur aut enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequentur magni dolores eos qui odit aut fugit, sed quia consequentur magni dolores eos qui odit aut fugit, sed quia consequentur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam ratione voluptatem sequi nesciunt. Neque porro quisquam ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, est, qui dolorem ipsum quia dolor sit amet, consectetur, est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora adipisci velit, sed quia non numquam eius modi tempora adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat incidunt ut labore et dolore magnam aliquam quaerat incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum voluptatem. Ut enim ad minima veniam, quis nostrum voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut exercitationem ullam corporis suscipit laboriosam, nisi ut exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum aliquid ex ea commodi consequatur? Ouis autem vel eum aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil iure reprehenderit qui in ea voluptate velit esse quam nihil iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat molestiae consequatur, vel illum qui dolorem eum fugiat molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur? quo voluptas nulla pariatur? quo voluptas nulla pariatur? Title Title Title Abstract Abstract Abstract Settings Lorem ipsum dolor sit amet, consectetur adipiscing elit, Lorem ipsum dolor sit amet, consectetur adipiscing elit,

sed do eiusmod tempor incididunt ut labore et dolore

magna aliqua. Ut enim ad minim veniam, quis nostrud

sed do eiusmod tempor incididunt ut labore et dolore

magna aliqua. Ut enim ad minim veniam, quis nostrud

Demo

 Offline store of paper used as corpus as Scholarly is sometimes detected as bot and prevented from Google Scholar access

Limitations and Future Work

- Better paper-related APIs
 - Current manual abstract extraction is unreliable
 - Scholarly (Google Scholar API) is new and unreliable
- More paper-related APIs
 - Storage (eg. Mendeley, Paperpile)
 - Search/access what if the paper is not available through Google Scholar?
- More context in the recommendation system
 - Activities
- Incorporate other assistance relevant to grad students
 - General planning/scheduling
 - Conference-related planning
 - Note taking/organization