Chapter Class Summ

Miles Jackson

CS 4624 Multimedia, Hypertext and Information Access
Instructor: Edward A. Fox

Virginia Tech
Blacksburg, VA 24061
5/7/2024
Outline

1. Timeline
2. Motivation
3. Design
4. Implementation
5. Testing
6. Problems Faced
7. Future Work
8. Acknowledgements.
**February**

Presentation 1
Dataset Collection and Annotation Setup. Select 100 ETDs for the segmentation dataset. Validate segmentation results. Begin the annotation process. Review the existing codebase and select the Tech stack.

**March**

Prototype testing, Presentation 2, Interim Report
Continue the development of the UI and testing.

**April**

Work on version 2 of the application and address any issues found during testing. Make final adjustments to the UI and models based on feedback. Prepare project documentation, including a guide on how to use the system and a final report.

Continue annotating chapter boundaries for the selected ETDs. Aim to complete the segmentation process. Deliver the first version of the application.

Integrate the trained models with the UI. Ensure the system can process input PDFs and display the results accurately. Conduct thorough testing of the system. Add test cases. Collect feedback from client.

Final presentation of the project to stakeholders. This includes a demonstration of the system, discussion of the project outcomes, and future work.
Most students are required to cite scholarly sources for their reports, essays, and papers.

A valuable type of source is Electronic Theses and Dissertations (ETDs),

Our project aims to reduce the hassle students endure when searching through ETDs to find relevant material for their writings.
If you have never had to search through an ETD, imagine searching through 40 pages that look like this....
To find this

with AM substrates. Highly loaded particle-based inks are typically slightly conductive prior to thermal post treatment, and the conductivity significantly increases when binders and encapsulating materials are removed from the system at temperatures above 100°C [2]. Yet, to achieve ideal conductivity, particle-based inks must be processed in excess of 200°C to sinter the particles [9]. MOD inks typically have a lower required sintering temperature than particle-based inks; e.g., 90–300°C for silver MOD inks [6], [7]. Generally, a higher processing temperature and failure leads to higher conductivity [2], [9]. Micronax C5026 silver particle-based ink has a recommended cure of 160°C for 1 hour, which results in 14–16 times the bulk resistivity of silver [6], [9]. The resistivity reduces to 2.5 times the resistivity of bulk silver when processed at 300°C for one hour [9]. A large challenge in AME is that, regardless of the process, the post-processing temperatures required to achieve low resistivity exceeds the thermal deflection temperature of the polymers that can be printed via AM, as shown in Table 1-1. Even when high temperature AM substrates are used, such as PEI, PEEK or Polyimide, an upper limit of approximately 2–3 times the bulk resistivity for both MOD inks and silver particle-based inks can be expected, with higher thermal post processing steps required to achieve increased conductivity of inks [2], [9]. For example, Joa and coauthors found that a maximum conductivity of Dupont C5026 silver ink was achieved when cured for an hour at temperatures above 300°C, far in excess of the manufacturer specified thermal cure of 160°C for one hour [9]. However, both temperatures exceed the heat deflection temperature for printable dielectric materials (Table 1-1). Thermal performance of common high-performance AM polymers [Table 1-2] indicates that the highest performing materials have thermal limits that exceed the thermal deflection temperatures of the AM substrate, which may contribute to Joule heating of the printed circuit, which can cause failure if the thermal deflection temperatures of the AM substrate or cause failure of thermally sensitive components integrated within the AME structure. In addition to thermal performance, electrical performance of AM substrates must also be considered. An ideal material would have thermal and electrical properties similar to those of common PCB substrates including FR-4 and PMDA-ODA, also called Kapton®. FR-4 is a class of glass-reinforced epoxy which make up the most common dielectric materials in printed circuit board while electrical performance vary by supplier [9]. Kapton® is an all-polymer laminate which exhibits excellent thermal, mechanical, and electromagnetic properties. Kapton® is commercially produced as a film. Two key metrics are the dielectric constant and dielectric strength of a material. Lower dielectric constants in circuit board and small electronic devices improves transmission speed and efficiency in high frequency AC circuit boards and reduces loss in power applications [30]. Dielectric strength corresponds to the strength of an insulator against voltage breakdown. Table 1-2 shows the dielectric performance of common dielectric materials and Kapton®. Ultem™ 1010 and 9085 have lower dielectric constants than Kapton®, but Kapton® has a much higher dielectric strength (>3500V/mil) than Ultem™.

Material Deflection Temperature at 0.45 Mpa (°C) Glass Transition Temperature Tg (°C)
Melt Temperature Tm (°C) Ultem™ 1010 [10] 208 217 >300 Ultem™ 9085 [11] 158 180 >300 PEK [12] 160 162 305 PEEK [13] 140 143 - PPS [14], [15] 80 85 283 Some of the most common high-performance AM polymers which possess thermomechanical properties enabling their use at higher temperatures include PEEK, PEI (Ultem™), PPS, and PEKK. The thermal performance of these materials is described in Table 1-1. None show glass transition temperatures in excess of 200°C required to achieve low resistivity of conductive metal inks discussed in previous sections. Thermal management is a design consideration in all electronics packaging [29]. For AME, the high resistivity of conductive inks contributes to Joule heating of the printed circuit, which can cause failure if more than the thermal deflection temperatures of the AM substrate or cause failure of thermally sensitive components integrated within the AME structure.
Designed Interaction Loop

User upload PDFs on the homepage and view the summarization results on another page. To upload another PDF, the user must restart the cycle.
PDF Uploading
<table>
<thead>
<tr>
<th>Title</th>
<th>Chapter 1: Incidental Ramifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td></td>
</tr>
<tr>
<td>Date Published</td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td></td>
</tr>
<tr>
<td>Advisor</td>
<td></td>
</tr>
</tbody>
</table>

Chapter Summary

Summarization of Uploaded PDFs (ETD Chapter)
Implementation
ETD Summarization

Title: A study of coherence in entity linking

Author: Duncan, Chase

Department: Computer Science

Year Published: 2018

University: University of Illinois at Urbana-Champaign

Discipline: Computer Science

Abstract:

Entity linking (EL) is the task of mapping entities, such as persons, locations, organizations, etc., in text to a corresponding record in a knowledge base (KB) like Wikipedia or Freebase. In this paper we present, for the first time, a controlled study of one aspect of this problem called coherence. Further we show that many state-of-the-art models for EL reduce to the same basic architecture. Based on this general model we suggest that any system can theoretically benefit from using coherence although most do not. Our experimentation suggests that this is because the common approaches to measuring coherence among entities produce only weak signals. Therefore we argue that the way forward for research into coherence in EL is not by seeking new methods for performing inference but rather better methods for representing and comparing entities based off of existing structured data resources such as DBPedia and Wikidata.
PDF is uploaded into FileUploader object
Back End

- The filename is used to determine the ETD ID and determine directories for text storage
- We then check if the file paths for summarization already exist in the database
If the summarization path is not in the database, then that means we need to add that summary to the database and then switch to the summarization page.
The summarization page simply reads data from the tables using the ETD ID passed from the homepage and displays said data.
Testing - Method

- User testing was difficult to conduct without implementing the machine learning models
- Testing was conducted by our client, Bipasha Banerjee
- The client attempted to populate the ETD database while stress testing the code
Testing - Results

- Insufficient input validation before attempting to insert values into database
- Summarization page can be navigated to before a PDF is uploaded causing null errors
- Site currently only works for client’s ETD database
Problems Faced - Accessing a Large CSV File

Problem:
The metadata file used for the project was an Excel spreadsheet that contained 500k entries. Searching through this file was extremely time consuming.

Solution:
We converted the Excel spreadsheet into a MySQL database, which allowed for database searching in a fraction of the time.
Problems Faced - Storing Large Text Files

Problem:
When storing entire ETDs into the database, a chunk of the ETD would be truncated due to a character limit imposed by the database itself.

Solution:
Instead of saving the entire ETD in the database, we saved the ETD into a text file and copied the path of that text file into the database instead.
Problems Faced - Streamlit Formatting

Problem:
Streamlit components are top-aligned by default. This means that elements in the same row have their labels aligned. Unfortunately, this cause most of our buttons to be misaligned due to only some of them requiring labels.

Solution:
We used HTML code to override the native Streamlit formatting and align elements manually.

```
#Override native Streamlit organization with html to align items favorably
st.write(
    "\n    <style>
    [data-testid="stHorizontalBlock"] {
    align-items: flex-end;
    }
    </style>
    
    unsafe_allow_html=True
    
    
)""
```
Future Work

- Integration of machine learning models into the project
- Implementation of user input validation to prevent catastrophic errors
- Expand ETD database to support more ETDs (Currently only supports 500,000)
Acknowledgements

- Yingjie Zhao (Jack)

Bipasha Banerjee

Edward Fox
References

- Lipps, J. (2020, September 2). What is testing? RSS. https://www.headspin.io/blog/what-is-testing
Questions?