

TextMining

Jianchi Wei, Chongyu He & Chenyu Mao



CS 4624 Multimedia, Hypertext & Information Access
Edward Fox
Virginia Tech, Blacksburg VA
24061
May 10, 2022

Outline

- Project Review
- Work Timeline
- Project Design
- Front End
- Lessons Learned
- Future Work
- Acknowledgments
- References

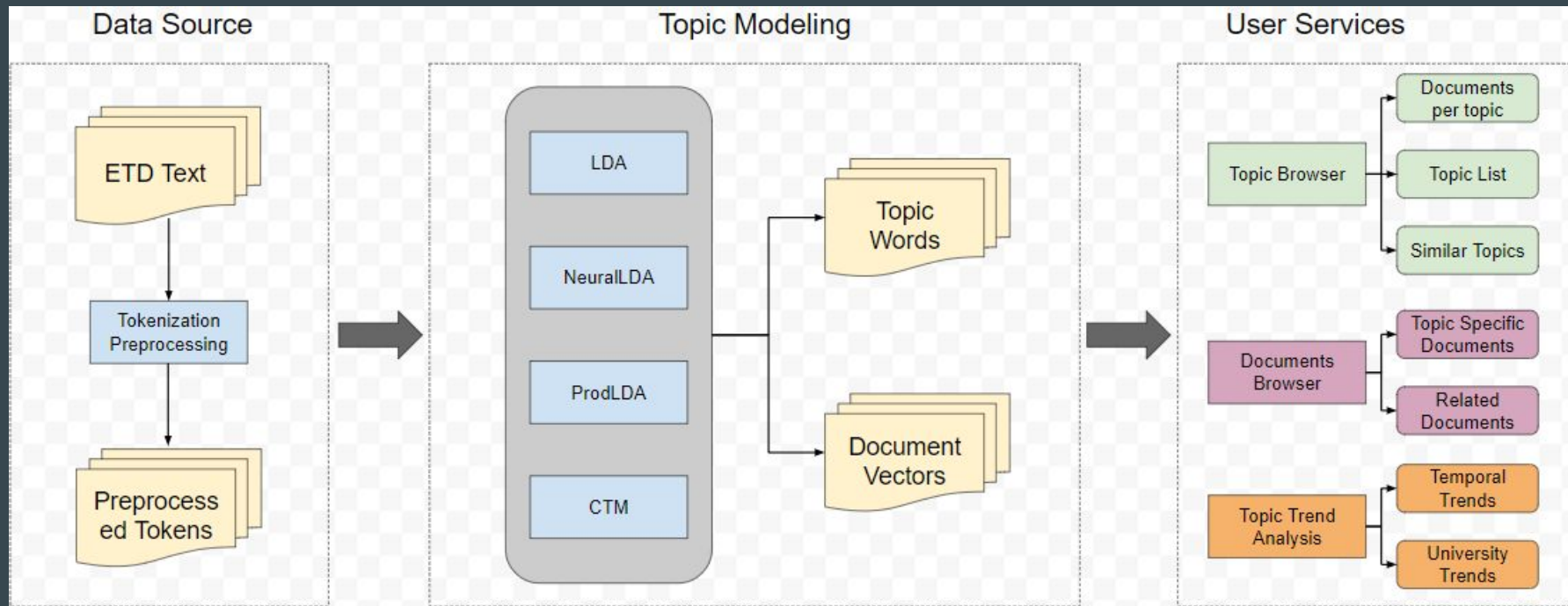
Project Review

- Extract valuable information and keys from Electronic Theses and Dissertations (ETDs).
- Using LDA, ProdLDA, NeuralLDA, and CTM models for summarizing topics.

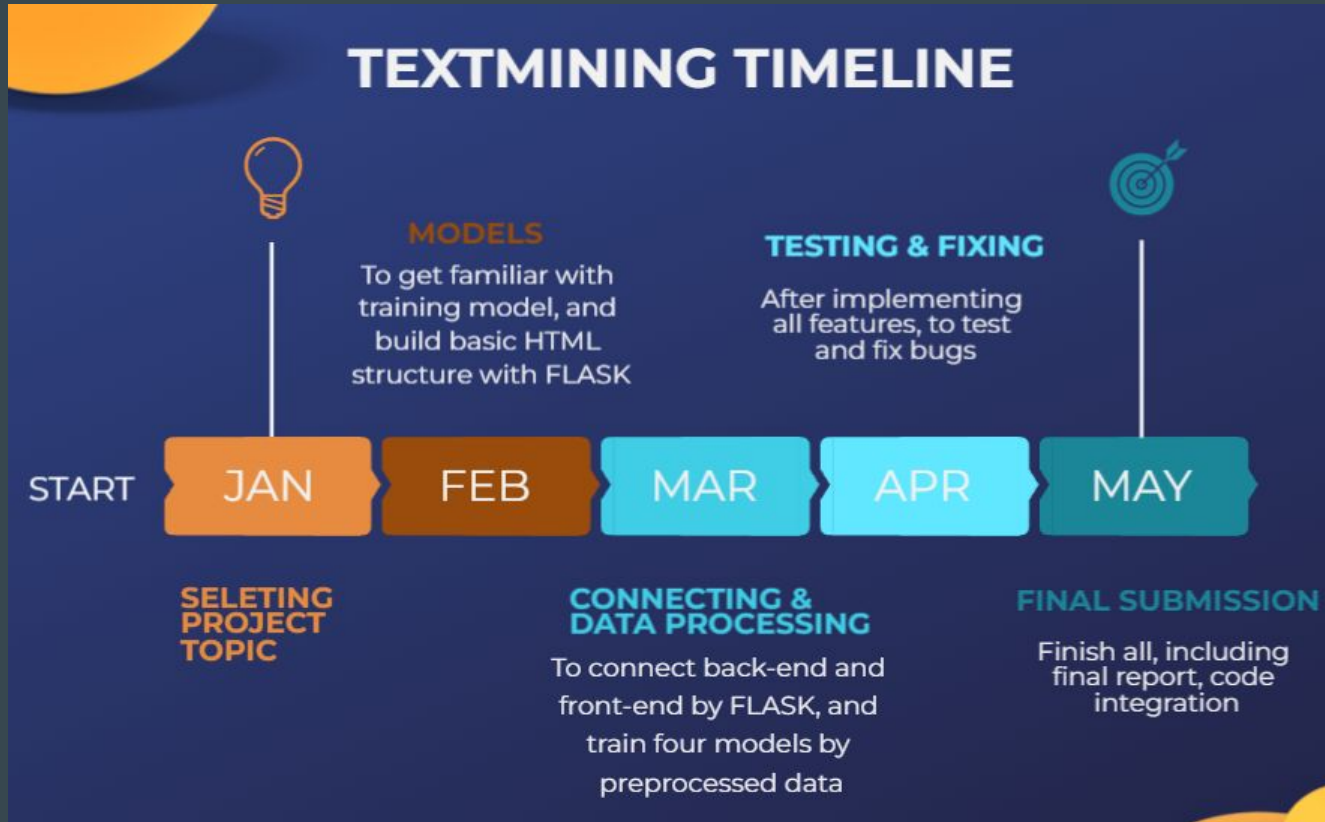
The screenshot displays the MITRE COVID-19 Topic Browser interface. At the top, it shows the MITRE logo, the title "CORD-19 Topic Browser", and navigation options like "20 Topic Model", "Search documents", "Filter", "Download PDF", "Send Feedback", and "Help". The main content area is divided into several sections:

- Sort topics by:** Trending
- Search topics:** A search bar with a close button.
- Topic List:** A vertical list of topic-related terms such as "pregnant, seroprevalence, maternal", "rbd, epitopes, mutations", "trial, randomized, placebo", "assay, assays, qpcr", "dimer, cox, icu", "attitude, attitudes, wearing", "telehealth, telemedicine, visits", "sir, reproduction, seir", "protease, repurposing, docking", "ethical, vaccines, commentary", "pollution, pollutants, meteorological", "embase, meta, medline", "man, woman, manifestations", "depression, anxiety, depressive", "consolidation, ct, glass", "markets, food, holder", "teaching, teachers, learning", "surgery, elective, surgical", "apoptosis, signaling, proliferation", and "convolutional, images, neural".
- Dates:** 2019/12/26 - 2021/5/20
- Bar Chart:** A bar chart showing the prevalence of a topic over time from June 2020 to May 2021. The y-axis represents prevalence, ranging from 0.0000 to 0.0800. The chart shows a relatively stable prevalence around 0.0400-0.0500, with a slight increase in May 2021.
- Document List:** A list of documents related to the topic, including:
 - Cruz-Lemini, Monica; Ferriols Perez, Elena; de la Cruz County, Maria Luisa; Caño Aguilar, Africa; Encinas Pardilla, Maria Begoña; Prats Rodríguez, Pilar; Muner Hernando, Marta; Forcen Acebal, Laura; Pintado Recarte, Pilar; Medina Mallen, Maria Del Carmen; Perez Perez, Noelia; Canet Rodriguez, Judit; Villalba Yarza, Ana; Nieto Velasco, Olga; Del Barrio Fernandez, Pablo Guillermo; Orizales Lago, Carmen Maria; Marcos Puig, Beatriz; Muñoz Abellana, Begoña; Fuentes Ricoy, Laura; Rodriguez Vicente, Agueda; Janeiro Freire, Maria Jesus; Alferez Alvarez-Mallo, Macarena; Casanova Pedraz, Cristina; Alomar Mateu, Onofre; Lesmes Heredia, Cristina; Wizner de Alva, Juan Carlos; Posadas San Juan, Alma; Macia Badia, Montserrat; Alvarez Colomo, Cristina; Sanchez Muñoz, Antonio; Pratorcorona Alicart, Laia; Alonso Saiz, Ruben; Lopez Rodriguez, Monica; Barbancho Lopez, Maria Carmen; Meca Casbas, Marta Ruth; Vaquerizo Ruiz, Oscar; Moran Antolin, Eva; Nuñez Valera, Maria Jose; Fernandez Fernandez, Camino; Tubau Navarra, Albert; Cano Garcia, Alejandra Maria; Soldevilla Perez, Susana; Gattaca Abasolo, Irene; Adanez Garcia, Jose; Puertas Prieto, Alberto; Ostos Serna, Rosa; Guadix Martin, Maria Del Pilar; Catalina Coello, Monica; Espuelas Malon, Silvia; Sainz Bueno, Jose Antonio; *Obstetric Outcomes of SARS-CoV-2 Infection in Asymptomatic Pregnant Women*; *Viruses*; January 01, 2021
 - Cruz-Lemini, Monica; Ferriols Perez, Elena; de la Cruz County, Maria Luisa; Caño Aguilar, Africa; Encinas Pardilla, Maria Begoña; Prats Rodríguez, Pilar; Muner Hernando, Marta; Forcen Acebal, Laura; Pintado Recarte, Pilar; Medina Mallen, Maria Del Carmen; Perez Perez, Noelia; Canet Rodriguez, Judit; Villalba Yarza, Ana; Nieto Velasco, Olga; del Barrio Fernandez, Pablo Guillermo; Orizales Lago, Carmen Maria; Marcos Puig, Beatriz; Muñoz Abellana, Begoña; Fuentes Ricoy, Laura; Rodriguez Vicente, Agueda; Janeiro Freire, Maria Jesus; Alferez Alvarez-Mallo, Macarena; Casanova Pedraz, Cristina; Alomar Mateu, Onofre; Lesmes Heredia, Cristina; Wizner de Alva, Juan Carlos; Posadas San Juan, Alma; Macia Badia, Montserrat; Alvarez Colomo, Cristina; Sanchez Muñoz, Antonio; Pratorcorona Alicart, Laia; Alonso Saiz, Ruben; Lopez Rodriguez, Monica; Barbancho Lopez, Maria Carmen; Meca Casbas, Marta Ruth; Vaquerizo Ruiz, Oscar; Moran Antolin, Eva; Nuñez Valera, Maria Jose; Fernandez Fernandez, Camino; Tubau Navarra, Albert; Cano Garcia, Alejandra Maria; Soldevilla Perez, Susana; Gattaca Abasolo, Irene; Adanez Garcia, Jose; Puertas Prieto, Alberto; Ostos Serna, Rosa; Guadix Martin, Maria del Pilar; Catalina Coello, Monica; Espuelas Malon, Silvia; Sainz Bueno, Jose Antonio; Granell Escobar, Maria Reyes; Cruz Melguizo, Sara; Martinez Perez, Oscar; *Obstetric Outcomes of SARS-CoV-2 Infection in Asymptomatic Pregnant Women*; *Viruses*; January 15, 2021
 - Ahc, Media; *Quantifying the Risk of SARS-CoV-2 Transmission in the United States*; *Infectious Disease Alert*; January 01, 2020
 - Fontanet, A.; Grant, R.; Tondeur, L.; Mader, Y.; Grzelak, L.; Cailleau, I.; Ungeheuer, M.-N.; Renaudat, C.; Fernandes Pellerin, S.; Kuhmel, L.; Staropoli, I.; Anna, F.; Charneau, P.; Demeret, C.; Bruel, T.; Schwartz, O.; Hoen, B.; *SARS-CoV-2 infection in primary schools in northern France: A retrospective cohort study in an area of high transmission*; June 29, 2020
 - Ng, Oon Tek; Marimuthu, Kalisivar; Koh, Vanessa; Pang, Junxiong; Linn, Kyaw Zaw; Sun, Jie; De Wang, Liang; Chia, Wan Ni; Tiu, Charles; Chan, Monica; Ling, Li Min; Vasoo, Shawn; Abdad, Mohammad Yazid; Chia, Po Ying; Lee, Tau Hong; Lin, Ray Junhao; Sadarangani, Sapna P; Chen, Mark I-Cheng; Said, Zubaidah; Kurupatham, Lalitha; Pung, Rachael; Wang, Lin-Fa; Cook, Alex R; Leo, Yee-Sin; Lee, Vernon JM; *SARS-CoV-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study*; *Lancet Infect Dis*; November 02, 2020

Project Detail



Work TimeLine



Project Design

- Index page
- Topic page
- Document page
- Advanced research page
- Search keyword page

Front End Demo

The video for the demonstration is in the file `TextMiningDemo.mp4`.

Lessons Learned

- Front End with FLASK
- Combination of JavaScript, CSS, and HTML
- Testing data



Future Work

Deploy to server



Acknowledgement

Aman Ahuja
aahuja@vt.edu

Professor Edward A. Fox
fox@vt.edu

References

- [1] Aman Ahuja, Chenyu Mao, Chongyu He, Jianchi Wei, Edward A. Fox, 25th International Symposium on Electronic Theses and Dissertations – ETD 2022, Novi Sad, Serbia
September 7 - 9, 2022, submitted for review
- [2] T. Hofmann, Probabilistic latent semantic indexing, in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (1999), pp. 50–57
- [3] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet Allocation, Journal of machine learning research 3, 993 (2003)