CSS4Impact

Computational Social Science for Positive Impact

Master Project in M.Sc. Digital Media - WS22/23 - SS23, University of Bremen

Project Documentation

Participants

1. Ahmed, Mohamad Tofayel
2. Ashraf, Hatim
3. Hesselbarth, Lisa
4. Hoque, Md. Imdadul
5. Hucke, Jule
6. Kakadiya, Ravi
7. Kathiravan, Deepika
8. Keiser, Merle
9. Kuri, Robin
10. Nusrat Ima, Nazia
11. Patel, Apurva
12. Shafiq, Mohammad Saad
13. Siddiqui, Md. Asif

Supervisors

Breiter, Andreas, Prof. Dr.

Molina Leon, Gabriela

Hannah-Marie Büttner

Skitalinska, Gabriella

Introduction

CSS4Impact is a collaboration between Computer Science and Social Science department in the University of Bremen.The project began with a kick-off meeting followed by a retreat organized by the project management team, where the primary objective was to foster better understanding about CSS and collaboration among team members, supervisors, and Prof. Dr. Heiko Pleines and his team (Social Science department). During the retreat, we conducted a poll to determine the

preferences for project tasks. Based on the poll results, As a result, we have decided to break down the project into the following two main components:

1. Webcrawler task in collaboration with CRC 1342.
2. Conducting independent research.

By doing so, we aim to enhance efficiency and effectiveness in our approach towards the project's completion.

As the project progresses, we have established distinct subgroups, while maintaining a unified project management and documentation across all aspects of the project.
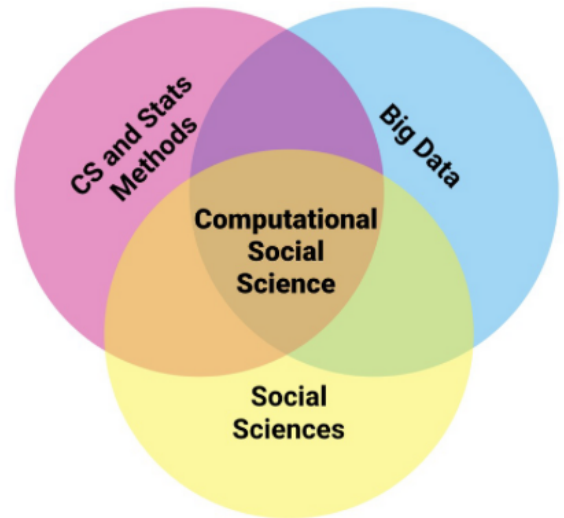
Subgroup Structure:
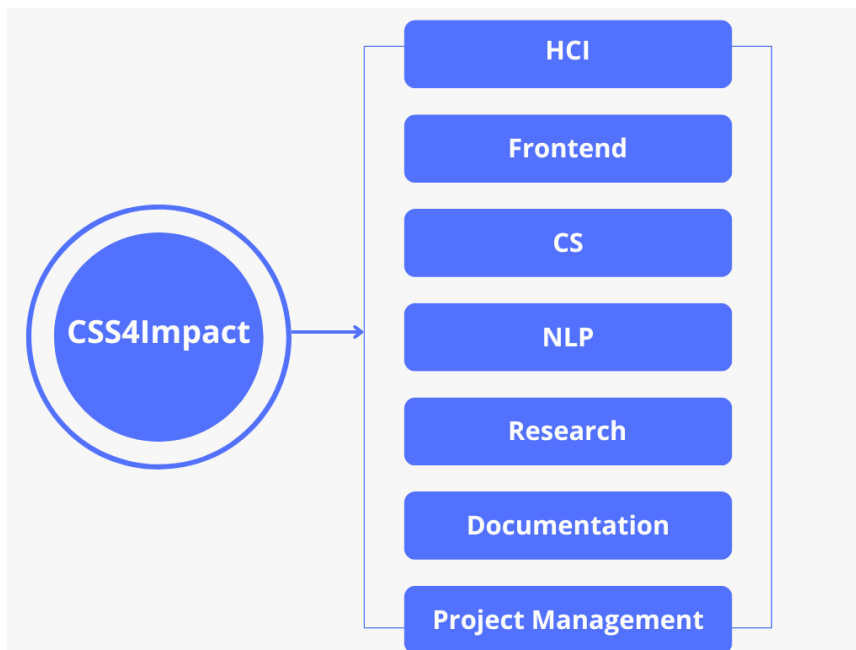
Overview

A detailed description of our project processes, methods, methodologies, goals, challenges and achievements from each subgroups are as follows:

Human-Computer Interaction (HCI)

Team Members worked in HCI: Ravi, Lisa, Apurva, Nazia, Deepika

The HCI team began their work with an insightful presentation by Heiko's team, where they outlined the problem they were facing and the potential solutions they sought. After this presentation, we embarked on compiling a comprehensive list of expectations and requirements from Heiko's team. This step paved the way for in-depth discussions and exploration of various approaches to address the challenges at hand.

The outcome of these discussions was carefully documented in our "Requirement Analysis Document", which outlined the system's purpose, scope, success criteria, and objectives. Throughout the project, we maintained constant communication with both the Computer Science (CS) and Natural Language Processing (NLP) teams to ensure seamless alignment with the set requirements and objectives. Additionally, we regularly scheduled meetings with Heiko to obtain valuable feedback on our progress.

As part of our process, we designed multiple user-interface variations and presented them to the team, seeking their input and preferences. Naturally, the project faced several challenges, especially during the negotiation phase of the initial requirements with both internal teams and Heiko's team. However, our unwavering focus remained on fulfilling the negotiated requirements, aiming to create a system that Heiko's team can effortlessly utilize.

Ultimately, the success of our project hinges on whether we have effectively fulfilled the negotiated requirements and we delivered a system that satisfactorily meets the needs of Heiko's team. Their satisfaction with the final product will be the key measure of our achievement.

Frontend Development

Team Members worked in Frontend: Jule

The front-end team commenced their work after the groundwork from the HCI team and CS team had reached a point where they had a solid data foundation (API) from the CS team and a basic design that they could build upon. The initial focus was on creating a template and establishing connections to the database using REST, enabling them to effectively present data in a table format. Continuous communication with the CS team allowed them to refine the API and implement a filter system from the front-end side.

During the development process, adjustments had to be made repeatedly as formats evolved, requirements were added, and a comprehensive basic concept was developed. To alleviate the workload of the CS team, the front end took on various tasks, such as implementing filtering and pagination functionalities, thus ensuring that these tasks did not burden the back end (server side).

Additionally, the team actively collaborated with the HCI group to address design-related adjustments, accommodating changes in formats like keywords, actors, and paths. Our front-end team promptly implemented these modifications, fostering a smooth and cohesive integration.

As the project progressed, the team underwent changes due to the shifting requirements of other sub-teams, resulting in a reduction in the front-end team's size. Nonetheless, the team members remained committed, drawing upon their expertise to deliver a high-quality outcome.

The ultimate goal is to deliver an intuitive browser application that offers a simple overview of the articles and essential information, coupled with convenient filtering options to access relevant content. To enhance user experience, they created an informative about page to introduce the team, and provide a comprehensive user manual to guide users on utilizing the application effectively.

For implementation, they chose NEXT.js, a React framework, to ensure a robust and efficient development process. To facilitate future development and maintenance, the team provided code documentation. In summary, our front-end team has made significant strides in crafting an easy-to-use and informative browser application. By collaborating with other teams and adapting to evolving requirements, the Frontend team delivered a seamless user experience and contributed to the overall success of the project.

Computer Science (CS)

Team Members worked in CS: Hatim, Robin, Tofayel, Saad

The CS team's task is to develop a comprehensive web scraper that enables users to efficiently extract valuable data from a wide range of websites from Azerbaijan, Ukraine and Georgia . By integrating natural language processing (NLP) code, the CS team have enhanced the capabilities of the scraper, allowing them to process the extracted data more intelligently and set priorities for different elements.

To streamline the execution, they have categorized the tasks into three main areas of focus:

1.   Web Scraper:

The Web Scraper team's dedicated scraping scripts for each site, along with their modular design, allowed concurrent work without compromising quality. Integration of a config file provided them customization options for tailored extraction of essential elements, aligned with search keywords. A shared NLP playground ensured them for a collaborative data harmonization. Database management is optimized through a single call approach, ensuring reliable and high-quality data for our project's success.

2.   Database and APIs:

The websites, front-end applications, and NLP (Natural Language Processing) layer communicate with the DB Layer through APIs. This API-based communication ensured loose coupling between the components, allowing them to evolve independently. The APIs provided them the necessary abstraction and allowed for seamless integration between the different layers of the application.

3.   Architecture / Integration and Deployment:

Seamless API integration, independent machine deployment, automated Docker startup, and user-friendly Bash scripts for application initiation, these established effortless and efficient execution.

Some of their challenges during the course of the development of the product include overcoming obstacles posed by anti-scraping mechanisms on websites, designing website-specific solutions to meet individual requirements effectively, ensuring accurate and reliable data extraction from diverse sources, creating a cohesive architecture that facilitates smooth

collaboration between NLP, Frontend, and Scrapers and ensuring components integrate seamlessly for unified user experience and enhanced functionality.

These architectural improvements significantly enhanced the application's scalability, maintainability, and flexibility. Adopting a microservices approach empowered independent development and deployment of various components, promoting seamless collaboration among team members.

Natural Language Processing (NLP)

Team Members worked in CS: Imdadul, Asif, Apurva, Ravi

The NLP sub-group played a crucial role in our master's project, focusing on assessing document relevancy through NLP techniques. Their main objective was to extract key information, such as actors, locations, and organizations, from crawled files and determine their relevancy.

To kickstart the NLP sub-group, they initiated collaborative brainstorming sessions to explore the potential applications of NLP. Through these sessions, they defined the scope of NLP within our project and identified the necessary technologies and tools. They also sought clarification from our clients to align the NLP efforts with their specific requirements.

The NLP journey began by developing three separate code scripts for three languages. Initially, they focused on topic classification but later shifted their attention to topic modeling, guided by our supervisor's advice. To extract keywords from text files, they explored techniques such as TF-IDF, Count Vectorizer, flair, and word-to-vec. During the process, they realized that tokenization negatively impacted keyword relevancy and decided to exclude it from our workflow.

Recognizing the importance of assessing relevancy, they revised our goal to evaluate the appearance of keywords in the files. Assigning weights to keyword lists (IDP and other search terms) enabled them to prioritize highly relevant documents. They continuously upgraded theNLP code section, fine-tuning the approach and methodology.

They faced challenges when identifying actors, locations, and organizations in Azerbaijani and Georgian documents. After extensive research, they found a language model that offered partial accuracy of around 70%. Ukrainian documents exhibited better results. The NLP team relentlessly strived to overcome obstacles and make improvements.

In the end, they achieved their goal and developed a functional system that accurately determines document relevancy. The integrated database and Docker deployment streamlined their workflow, ensuring seamless execution of the NLP code. The front end displays a file relevancy table, with the most relevant files ranked at the top, and allows users to view associated actors, locations, and organizations. Although accuracy remains a concern, their system provides valuable insights into document relevance.

Our NLP sub-group's journey was characterized by teamwork, adaptability, and continuous improvement. They successfully addressed the challenge of document relevancy, integrating advanced NLP techniques into our project workflow. Their accomplishments and contributions have been vital to the overall success of the project, and they take pride in the outcomes of their efforts.

Research

Team Members worked in Research: Merle, Nazia, Deepika, Ravi, Lisa, Jule, Imdadul, Asif

In the project's initial phase, we explored various areas, and Social Media Analysis emerged as the most intriguing option, aligning with our interest in data science and offering a fresh topic to apply our skills. The idea of connecting Social Media Analysis with the current Russian invasion intrigued us, as we recognized its relevance and potential to contribute to understanding the ongoing events. The subject offered numerous opportunities for investigation.

During the process, we had valuable meetings with Oksana and Artem, seeking guidance for a compelling research topic. Fascinating ideas emerged, including exploring social media's information sources, analyzing reliable sources, propaganda, wording, narratives, and investigating bots and troll factories.

Inspired by CRC1342's research, we initially considered analyzing Telegram and Facebook. However, Social Network Analysis (SNA) on these platforms was impractical due to limited access. Recognizing Twitter as a viable option, we decided to focus on wording and narratives, with valuable input from Artem, who provided relevant keywords to strengthen our decision.

With a substantial collection of research materials, we were prepared to pursue our chosen direction. The prospect of studying Twitter narratives excited us, providing confidence in its potential for an engaging and fruitful

project. The initial exploration and decision-making phase laid a strong foundation, fueling our enthusiasm to delve deeper into the captivating world of Social Media Analysis and its real-world implications.

In the second half, we narrowed our research topic and delved into coding, utilizing a substantial year-long Twitter dataset. Summarizing relevant research papers enriched our investigation. We established metrics to identify influential actors based on engagement, focusing on retweets and likes in the context of the invasion.

We focused on Bot detection, exploring different approaches to identify automated accounts. Additionally, keyword extension and extraction proved highly applicable to our research. Meticulously gathering important keywords enriched our dataset, and we completed the first significant draft of our paper, reflecting substantial progress in understanding the invasion through Twitter data analysis. The devised scoring system allowed us to quantify narratives and keywords, presenting our findings effectively through visualization.

As the project concluded, we consolidated our findings into a comprehensive paper, shedding light on narratives and trends surrounding the Russian invasion of Ukraine on Twitter. The second half saw us deepening our exploration, honing coding skills, and conducting an in-depth analysis, providing valuable insights for a better understanding of this critical event. Our paper reflects our dedication and commitment to making a substantial research contribution.

Documentation

Team Members worked in Documentation: Deepika

As part of our project, one of the crucial responsibilities was to deliver comprehensive project documentation. This documentation serves as a report on various aspects of the project, including the process, participants, methods, and more. Additionally, we facilitated monthly milestone coordination and communication with team members to ensure smooth progress.

The task involved collecting and organizing important project documents, presentations, and Gantt charts ensuring that essential information was readily accessible. We diligently recorded the protocols from Friday meetings to keep track of discussions, decisions, and action items.

However, we faced some challenges along the way with follow-ups and delays. Despite facing challenges, we diligently tackled follow-ups with the team, managed

delays in obtaining documents, and maintained effective coordination with sub-groups for timely project completion. Actively engaging with sub-groups allowed us to gather relevant data and updates, leading to accurate and comprehensive reports.

As contributors to the final project documentation, we took on the task of writing and organizing content to present a coherent and cohesive narrative of the project's progress and outcomes. Gitlab played a central role in our documentation management. We utilized the "Issues" and "Milestones" features to track tasks and project milestones, allowing for efficient collaboration and progress monitoring. Regular communication with the team and supervisors helped ensure that everyone remained on the same page and had the necessary information to contribute effectively.

In close collaboration with the Computer Science (CS) team, we worked on crafting essential documents such as the project framework and status update files. These documents were instrumental in providing clarity on the project's direction and progress. Our dedication to maintaining an organized and accessible project document folder helped streamline information retrieval and ensured that all team members had access to relevant project materials.

Our role in project documentation encompassed a range of tasks, from coordination and communication to writing and organization. By maintaining a well-structured and comprehensive documentation system, we contributed to the success of the project and facilitated effective collaboration among team members.

Project Management

Team Members worked in Project Management: Winter semester - Ravi, Lisa, and Jule. Summer semester - Jule, Merle and sub-team leads.

As project managers, the responsibilities encompassed various tasks to ensure smooth project execution. They diligently organized and prepared both Monday and Friday meetings, facilitating effective communication among team members. Additionally, they maintained open channels of communication with supervisors and the entire team, offering support wherever it was needed.

Planning retreats was an integral part of our project management strategy, providing valuable opportunities for team building and reflection. Implementing Kanban, an agile project management technique, helped us visualize tasks, track progress, and prioritize efficiently.

Throughout the project, they encountered several challenges, which we addressed with dedication and proactive measures like Self-Organization, Keeping overview, Finding time slots for meetings and communication.

But as project managers, we tackled these challenges in a possible way and played a pivotal role in guiding the team, facilitating communication, and promoting self-organization. By addressing challenges proactively and implementing project management techniques, they contributed to the project's success and promoted a cohesive and efficient team dynamic.

Meet our Team

Conclusion

CSS4Impact, which started on October 1, 2022, and concluded on July 31, 2023, was a dynamic journey characterized by multiple processes, challenges and continuous improvement of our approaches. Despite encountering various challenges and

adopting changing methodologies, our collaborative teamwork played a pivotal role in the project's success. The valuable feedback from our supervisors played a crucial role in our achievements. Delivering the final product to Heiko and his team on July 17, 2023, and handing over our research paper to our supervisors on July 14, 2023, marked the successful culmination of our efforts.