

Meeting Minutes — 2025-08-15

****Project:** Claim-Date Extraction (FACTors)**

****Attendees:** Sanjay Bhattacharjee, Adnan Rahman Eshan (student)**

Agenda

- Review progress on claim-date extraction code
- Discuss data handling for `FACTors.csv` [Altuncu et al., 2025](<https://arxiv.org/pdf/2505.09414>)
- Identify and troubleshoot failed URL extractions
- Plan next steps, including batch processing and output validation

Discussion & Notes

- Initially, the dissertation topic was explainable AI using Game Theoretic Solutions, but due to time constraints and complexity, it was dropped. Afterwards, the Supervisor suggested analysing the fact-checking ecosystem dataset `FACTors.csv` ([Altuncu et al., 2025](<https://arxiv.org/pdf/2505.09414>)) and extracting claim dates from the published facts. (03.06.2025)
- The scraper showed claim dates and matched the dates manually with the metadata of the websites and made sure that the date extraction logic was fine.(05.06.2025)
- The scraper was not working for some URLs, such as factcheck and verafiles. (11.05.2025)
- The scraper is mostly functional; asynchronous fetching improves runtime for large datasets.(17.06.2025)
- Some URLs fail due to network timeouts or missing HTML elements; these failures are logged for later review. Also, the runtime was too high. The supervisor recommended using University Raptor for faster and smoother processing of the code.(11.07.2025)
- Was able to extract most of the dates after adding some modifications to the code. Decided to include sample outputs (`sample_results.csv`) for review. (30.07.2025)
- Reviewed requirements for testing plan, meeting logs, and design documentation.

Action Items

1. Update date extraction functions for additional fact-checking sites (Owner: Eshan, Due:2025-09-02)
2. Prepare a representative input CSV sample for submission (Owner: Eshan, Due: 2025-09-02)
3. Complete user guide and developer guide documentation (Owner: Eshan, Due: 2025-09-02)
4. Validate `index.html` links before packaging corpus (Owner: Eshan, Due: 2025-09-02)

Decisions Made

- Use an anonymised input CSV sample if the full dataset is too large.
- Include screenshots of successful runs and logs in the artefacts folder.

References

- Altuncu, E., Baskent, C., Bhattacharjee, S., Li, S., & Roy, D. (2025, July). *FACTors: A New Dataset for Studying the Fact-checking Ecosystem*. In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 3530–3539). [PDF](<https://arxiv.org/pdf/2505.09414>)