

Report on the Fifteenth Temporal Web Analytics Workshop (TempWeb 2025) at The Web Conference 2025 (WWW 2025)

Marc Spaniol
University of Caen Normandy
Caen, France
marc.spaniol@unicaen.fr

Omar Alonso
Amazon
Palo Alto, CA, USA
omralon@amazon.com

Ricardo Baeza-Yates
Universitat Pompeu Fabra
Barcelona, Spain
rbaeza@acm.org

Abstract

For 15 years TempWeb has been a well-established workshop (series) co-located at The Web Conference. Its goals and ambitions are bringing together researchers and practitioners across various domains, taking the constantly evolving Web as a primary object of research. Areas of interest include but are not limited to the investigation of infrastructures, scalable methods, and innovative software for aggregating, querying, and analyzing heterogeneous data at Web scale. For that reason, submissions commonly address core fields of studies in computer science and aspects such as temporal IE/IR, Web mining, Web archiving and large-scale data analysis to just name a few. However, TempWeb does not only cover contributions in core computer science related, but also not purely technical submissions from application domains such as the social sciences, marketing, economics, etc. Therefore, TempWeb serves as a forum for a community from academia and practitioners across various disciplines. TempWeb 2025 was very representative in that regard and covered analytics of Telegram communities, time series analysis, but also urban planning and suicide prevention approaches.

Date: April 29, 2025.

Website: <http://temporalweb.net/>.

1 Introduction

TempWeb 2025 was the fifteenth event in its workshop series and took place co-located on April 29, 2025 in conjunction with The Web Conference WWW 2025¹ in Sydney, Australia. The workshop was organized as an on-site event with the opportunity of a hybrid participation under well-motivated circumstances (e.g., illness). Submissions to TempWeb are available through Open Access via the proceedings Website.²

The focus of TempWeb and the topics perfectly fit within The Web Conference. With digitally born contents existing since about 25 years ago, the need for a more systematic exploitation of our

¹<https://www2025.thewebconf.org/>

²<https://dl.acm.org/doi/proceedings/10.1145/3701716#heading15>

digital cultural heritage as well as new analysis techniques, becomes evident. Hence, TempWeb is an ideal venue to exchange knowledge about temporal analytics at a Web scale with experts from science and industry. Further workshop details can be found on the workshop Website.³ Due to space limitations, the program of TempWeb 2025 was organized as a half-full day workshop, only. Thus, the program had to be restricted to the most essential and was organized in two sessions:

Session 1: Keynote & Short Papers

Session 2: Full Papers & Wrap-Up

Further information about the program of the workshop is available at the workshop Website³.

2 Objectives and Topics

TempWeb aims at nurturing a venue for researchers and practitioners across various domains as a platform to discuss the various possibilities and challenges of studies of (Web) contents having time as a core dimension. As such, TempWeb is a truly interdisciplinary venue of research on the temporal Web. To this end, TempWeb’s ambition is to keep shaping a community of interest on research issues resulting from the introduction of the time dimension in Web analytics and comprehension. Large-scale repositories of Web contents and the maturity of the ever-growing Web makes this a likewise timely and challenging project. We believe, that providing a forum for research on the temporal Web will provide deeper insights beyond the scope of “simple” contents ranging toward - ultimately - a better understanding of societal implications [Alonso et al., 2025b,a].

The focus of TempWeb targets the investigation of infrastructures, scalable methods, and innovative software for aggregating, querying, and analyzing heterogeneous data at Web scale. There is an emphasis to Web data analysis along the time dimension that has been collected over extended time periods. Along these lines, a major challenge arises from the sheer size of the data it exposes and the ability to make sense of it in a useful and meaningful manner for its users. Making “sense” of the continuously evolving Web has become an integral part of our daily life. From an end-user’s perspective leisure related Web consumption (such as social media or online communities) via facilitating our daily life (e.g. online shopping or trip planning) up to serious usage (like online meeting summarization or technology enhanced learning) has become de facto the norm. Therefore, content analysis and their longitudinal dimension has become more than ever relevant for analysts from various domains, including, but not limited to sociology, marketing, environmental studies, politics, etc.

The wide range of all before mentioned aspects are (to some extent) equally important and relevant to be investigated. Research investigated and presented at TempWeb tackles exactly those aspects: Web structure, contents, and/or communities (along the temporal dimension). Therefore, topics relevant to TempWeb include, but are not limited to the following:

- Web scale data analytics
- Temporal Web analytics
- Distributed data analytics
- Web science
- Web dynamics
- Data quality metrics
- Web spam evolution
- Content evolution on the Web

³<https://temporalweb.net/>

-
- Systematic exploitation of Web archives
 - Large scale data storage
 - Large scale data processing
 - Time aware Web archiving
 - Data aggregation
 - Web trends
 - Topic mining
 - Terminology evolution
 - Community detection and evolution

3 Contributions

For its fifteenth edition, TempWeb was organized due to space limitations for the first time ever as a half-day workshop, only. Seven very positively reviewed submissions (three as short papers and four as full papers) were accepted for oral presentation [Spaniol et al., 2025b] covering a wide spectrum of (spatio-)temporal contents. Four submissions plus the keynote were delivered onsite, while three contributors participated virtually. The highlight was the keynote by Oshani W. Seneviratne from Rensselaer Polytechnic Institute (Troy, NY, USA), who is the (so far) one and only PhD student from Sir Tim Berners-Lee. Finally, the workshop was concluded with a wrap-up and feedback session among all participants in order to discuss the lessons learned from the day and exchange ideas for upcoming editions.

TempWeb 2025 was introduced by Marc Spaniol on behalf of the Workshop organizers, giving a quick intro of the workshop objectives and the program of the day [Spaniol et al., 2025a]. The first session then started with the first keynote by Oshani W. Seneviratne on “Verifiable Provenance for the Web” [Seneviratne, 2025]. In her talk, she addressed the fact that the Web originally considered as a decentralized information sharing medium with free and open access. However, during the last three decades the Web has undergone a highly dynamic evolution which has also lead to negative side effects of data misuse and intended data manipulation. In order to overcome these problems, in her talk she presented the aspect of decentralized provenance for the rapidly growing Web consisting of four key components. First, tracking content evolution over time. Second, ensuring the provenance of shared data. Third, the enforcement of semantic agreements between data consumers and providers. Fourth, methods of authorship proof, e.g., through a blockchain-enabled crowdsourcing mechanism. The presentation was accompanied by real life experiences obtained from previous studies that highlighted the importance verifiable provenance for the Web. She concluded her talk by stating that provenance is no longer a problem of tracing the past, but has become an emerging phenomenon to create the future of the Web a more transparent, accountable, and decentralized. The session was then continued with short paper presentations.

The first contribution was by Nicholas Largey et al. (University of Southern Maine, Portland, ME, USA) on “Temporal Aspects of CLIP Embeddings” [Largey et al., 2025]. To this end, the authors explain that OpenAI’s Contrastive Language–Image Pretraining (CLIP) captures a joint semantic representation of image and text. However, the temporal information inherent in the embeddings remains disregarded. In order to pinpoint this observation, they conducted a study on a ten centuries spanning WikiArt data set. Here, they figured out that CLIP disregards the importance of the temporal context and that certain periods were often misclassified due to similar stylistic elements. Hence they argued, that explicit fine-tuning or hybrid approaches are required in order to achieve temporal awareness. In the following, Mandy Ho (University of Southern Maine, Portland, ME, USA) presented her work on “Analyzing the Influence of the Pandemic

on Seasonal Queries” [Ho et al., 2025]. To this end, she investigated seasonal search patterns in Google Trends. To this end, the data set consisted of almost 90 seasonal queries spanning two decades ranging from 2004-2024. Here, she identified that the underlying search patterns are by and large stable. However, certain query topics (e.g., related to education) showed anomalies during the pandemic. Hence, long-term studies might be affected if those aspects might not be sufficiently taken into consideration. The session was concluded by the presentation of Dirk Ahlers (Norwegian University of Science and Technology, Trondheim, Norway) on “Temporal Aspects in Process Support for Urban Digital Twins” [Ahlers, 2025]. In his paper, he investigated the - till now - still underrepresented temporal aspects in urban planning. To this end, he introduced the concept of *timescapes* in order to explore temporal features and data. Thus, he highlighted various dimensions of time in smart cities and how they can be made productive for the planning within Digital Twins.

The second session covered four full papers and the wrap-up discussion. It started with the presentation of Nikhil Jha on “Topic-wise Exploration of the Telegram Group-verse” [Perlo et al., 2025]. In his talk, he presented a broad study covering Telegram groups including, but not limited to, crypto, education, up to erotic. For that purpose, they collected more than 50 million messages through an open-source tool released by them. To this end, they first characterized the collected data by topic and provided statistics about them. Next, they delved into the topics and provided insights about the usage of bots, languages, message length and utilization of non-textual contents. Here, their study revealed a significant heterogeneity in platform usage across the various communities. This showed, that there exists a great diversity of behavioral patterns depending of the group’s topic. As a result, they aim at investigating the key factors driving user engagement as well as the processes of influence and information dissemination. Subsequently, the paper on “BERTDetect: A Neural Topic Modelling Approach for Android Malware Detection” [Ranaweera et al. [2025]] was presented by Nishavi Ranaweera (University of New South Wales, Sydney, Australia). In her talk she addressed the problems of malware detection for Android devices. Here, she argued that existing approaches (e.g. source code analysis) are effective, but often fail to discover the malware due to the dynamics in code adaptation. Complementary techniques, for instance via clustering through their descriptions, are therefore needed to improve malware detection. To this end, they presented an approach to allow a very nuanced analysis of semantic structures present in app descriptions through the BERTopic neural topic modeling. Comprehensive experiments showed the viability of the proposed approach. Next, Adam Jatowt (University of Innsbruck, Austria) presented an approach toward “Predicting Company ESG Ratings from News Articles Using Multivariate Timeseries Analysis” [Aue et al., 2025]. He first motivated that environmental, social, and governance (ESG) reporting has gained public attention and is getting increasingly important for sustainable investment decisions. Thus, rating predictions are getting increasingly relevant. However, standard approaches commonly neglect the temporal dimension(s) between topics, sentiments, and events. To this end, they propose a novel method that combines multivariate time series with techniques of deep learning. In their experiments, they showed the viability of their approach based on an individually created data set containing 3.7 million news articles about 3,000 companies over a period of three years. In particular, they showed that the proposed method can improve the ESG prediction performance over baseline approaches. The final presentation of this session was then given by Jun Li (Hong Kong Polytechnic University, Hong Kong, SAR, China) on “DynaProtect: A Dynamic Factor Influence Learning Framework for

Protective Factor-aware Suicide Risk Prediction” [Li et al., 2025]. In his talk he addressed a societal highly relevant topic of suicide risk detection based on social media. To this end, they proposed a method that overcomes the shortcomings of existing approaches in identifying risk factors, only, while neglecting potentially existing protective factors at the same time. They propose a model to capture the temporal evolution of user posts to reveal critical patterns of mental state evolution. For that purpose, they capture sequential patterns of user postings first and then employ temporal attention to leverage the importance of the individual postings. In their experiments they show the effectiveness of their framework and an improvement over the state-of-the-art baselines.

The workshop was then concluded by a short wrap-up discussion. Here, all participants and audience gave their feedback on the organization and alignment of the workshop. It became clear that the temporal dimension is an aspect which requires to be investigated and needs a forum like TempWeb to exchange ideas and establish opportunities for co-operations with like-minded researchers and practitioners. Further, the participants regretted that TempWeb had to be squeezed into a half-day.

4 Conclusion and Future Directions

As in previous years, TempWeb 2025 was driven by the interactive presentations and the discussions they triggered. Not surprisingly, discussions were more interactive and productive for those papers that were presented in person. This showed again that hybrid events help to overcome distance limitations, but still face problems due to temporal constraints imposed by time zones. Hence, a 100% in presence appears to be ideal to discuss (half baked) ideas not only during the workshop, but also, afterwards.

Experiencing the dynamics of contents along the temporal dimension and many returning contributors and observing, TempWeb was and still is an ongoing endeavor that investigates the important aspect of the temporal Web. Based on the very positive feedback, a novel proposal for TempWeb 2026 has been submitted and accepted for the upcoming edition of The Web Conference. Thus, TempWeb 2026 is supposed to take place as an in presence event to be held on April 13 or April 14, 2026 in Dubai, United Arab Emirates.

Acknowledgements

TempWeb thanks the authors of submitted papers for their efforts, and all workshop attendees. The workshop was jointly organized by University of Caen Normandy (Caen, France), Amazon, Palo Alto, CA, USA, and Universitat Pompeu Fabra (Barcelona, Spain) and. We also wish to express our particular thanks to our program committee members:

- Srikanta Bedathur (IIT Delhi, India)
- Andras A. Benczur (Hungarian Academy of Science)
- Klaus Berberich (University of Applied Sciences Saarbrücken, Germany)
- Behrooz Mansouri (Rochester Institute of Technology, USA)
- Ricardo Campos (Polytechnic Institute of Tomar)
- Govind (Amazon, India)

-
- Adam Jatowt (University of Innsbruck, Austria)
 - Nattiya Kanhabua (Upwork, Bangkok, Thailand)
 - Scott Kirkpatrick (Hebrew University Jerusalem, Israel)
 - Amit Kumar (Tarento Technologies Pvt Limited, India)
 - Frank McCown (Harding University, USA)
 - Michael Nelson (Old Dominion University, USA)
 - Nikos Ntarmos (Huawei Technologies R&D, UK)
 - Kjetil Nørkvåg (Norwegian University of Science and Technology, Norway)
 - Thomas Risse (University Library Johann Christian Senckenberg, Germany)
 - Andreas Spitz (University of Konstanz, Germany)
 - Jannik Strötgen (Bosch Center for Artificial Intelligence, Germany)
 - Torsten Suel (NYU Polytechnic, USA)
 - Masashi Toyoda (Tokyo University, Japan)
 - Gerhard Weikum (Max-Planck-Institut für Informatik, Germany)

Last but not least, we would like to thank the organizers of WWW2025 for hosting the workshops and providing the online meeting facilities.

References

- Dirk Ahlers. Temporal aspects in process support for urban digital twins. In Guodong Long, Michale Blumestein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1772–1773. ACM, 2025. doi: 10.1145/3701716.3717502. URL <https://doi.org/10.1145/3701716.3717502>.
- Omar Alonso, Marc Spaniol, and Ricardo Baeza-Yates. Introduction to the special issue on temporal web: Studying time and the temporal dimension. *ACM Trans. Web*, 19(3), August 2025a. ISSN 1559-1131. doi: 10.1145/3743140. URL <https://doi.org/10.1145/3743140>.
- Omar Alonso, Marc Spaniol, and Ricardo Baeza-Yates, editors. *Transactions on the Web Special Issue on Temporal Web: Studying Time and the Temporal Dimension*, volume 19. Association for Computing Machinery, 2025b.
- Tanja Aue, Adam Jatowt, and Michael Färber. Predicting company ESG ratings from news articles using multivariate timeseries analysis. In Guodong Long, Michale Blumestein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1774–1780. ACM, 2025. doi: 10.1145/3701716.3717509. URL <https://doi.org/10.1145/3701716.3717509>.
- Mandy Ho, Nicholas Largey, and Behrooz Mansouri. Analyzing the influence of the pandemic on seasonal queries. In Guodong Long, Michale Blumestein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference*

-
- 2025, *WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1781–1782. ACM, 2025. doi: 10.1145/3701716.3717504. URL <https://doi.org/10.1145/3701716.3717504>.
- Nicholas Largey, Mandy Ho, Janna Ahrndt, and Behrooz Mansouri. Temporal aspects of CLIP embeddings. In Guodong Long, Michale Blumstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1783–1784. ACM, 2025. doi: 10.1145/3701716.3717503. URL <https://doi.org/10.1145/3701716.3717503>.
- Jun Li, Xiangmeng Wang, Yifei Yan, Haoyang Li, Hong Va Leong, Nancy Xiaonan Yu, and Qing Li. Dynaprotect: A dynamic factor influence learning framework for protective factor-aware suicide risk prediction. In Guodong Long, Michale Blumstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1785–1791. ACM, 2025. doi: 10.1145/3701716.3717510. URL <https://doi.org/10.1145/3701716.3717510>.
- Alessandro Perlo, Giordano Paoletti, Nikhil Jha, Luca Vassio, Jussara M. Almeida, and Marco Melia. Topic-wise exploration of the telegram group-verse. In Guodong Long, Michale Blumstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1792–1801. ACM, 2025. doi: 10.1145/3701716.3717506. URL <https://doi.org/10.1145/3701716.3717506>.
- Nishavi Ranaweera, Jiarui Xu, Suranga Seneviratne, and Aruna Seneviratne. Bertdetect: A neural topic modelling approach for android malware detection. In Guodong Long, Michale Blumstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1802–1810. ACM, 2025. doi: 10.1145/3701716.3717501. URL <https://doi.org/10.1145/3701716.3717501>.
- Oshani Seneviratne. Verifiable provenance for the web. In Guodong Long, Michale Blumstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov, editors, *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1770–1771. ACM, 2025. doi: 10.1145/3701716.3718486. URL <https://doi.org/10.1145/3701716.3718486>.
- Marc Spaniol, Omar Alonso, and Ricardo Baeza-Yates. 15th Temporal Web Analytics Workshop (TempWeb 2025) Overview. In *Companion Proceedings of the ACM on Web Conference 2025, WWW '25*, page 1768–1769, New York, NY, USA, 2025a. Association for Computing Machinery. ISBN 9798400713316. doi: 10.1145/3701716.3717499. URL <https://doi.org/10.1145/3701716.3717499>.
- Marc Spaniol, Ricardo Baeza-Yates, and Omar Alonso, editors. 15th Temporal Web Analytics Workshop (TempWeb 2025), Sydney, Australia, April 29, 2025, 2025b. ACM. ISBN 9798400713316.