Towards Authority-Dependent Risk Identification and Analysis in Online Networks

Frederik S. Bäumer and Sergej Denisov
Bielefeld University of Applied Sciences, Working Group "Applied AI"
GERMANY

frederik.baeumer@fh-bielefeld.de sergej.denisov@fh-bielefeld.de

Michaela Geierhos and Yeong Su Lee
Universität der Bundeswehr München, Research Institute CODE
GERMANY

michaela.geierhos@unibw.de yeongsu.lee@unibw.de

ABSTRACT

Interaction, discussion, and the exchange of diverse information make the Web the place it is today. Texts, images, videos, and even information such as geospatial and health data are shared at an unprecedented scale. This exchange of information on the Web generates an extensive, freely accessible data source for a variety of data-driven applications – with multiple opportunities, but also risks. In this paper, we present the overall idea of the research project ADRIAN – "Authority-Dependent Risk Identification and Analysis in online Networks" which is dedicated to the research and development of AI-based methods for detecting potential threats to individuals and institutions based on heterogeneous, online data sets. We will first monitor selected social sports apps and analyze the collected geospatial data. In a second step, the user profiles of sports apps and social media platforms will be correlated to be able to form a cluster of individuals and enable the identification of potential threats. Since a so-called "digital twin" can be reconstructed in this way, sensitive data is generated. If this data can also be correlated with other confidential data, it is possible to estimate the plausibility of the threat to individuals, groups or locations.

1.0 INTRODUCTION

The modern Web is based on interaction, discussion, and the exchange of information. However, the Web also creates a vast, freely accessible source of information for data-driven applications. Since user-generated data on the Web is effectively linked to existing resources in an automated way, even unintentionally revealed personal information can have damaging consequences. Thus, even trivial, and sometimes unintentionally disclosed information can have potentially harmful effects on individuals, groups, or an entire organization [1,2,3]. Although service providers now have a duty and an interest in ensuring the security and privacy of user data on the Web, there is an increasing number of cases where this data is misused, compromised, or publicly available information is used against the original creator [4] or government institutions [5]. Law enforcement and other groups of people have faced increased potential threats on social media platforms not just since the 2020 riots in the United States. In particular, the collection and linkage of social media accounts and posts (e.g., Twitter or Instagram) with tracking and location data from popular sports apps allows users and their loved ones to be identified, making them traceable and a potential target for cyberattacks (e.g., cyberstalking, doxing, identity theft) [5,6]. Another security-relevant aspect in this context is the fact that military bases can be located using the collected geospatial data of running routes [7]. Since not all information that poses a threat by itself or in combination is obvious, pure data minimization, restricted data access, data avoidance and prevention work are not sufficient [8]. In the research project ADRIAN – "Authority-Dependent Risk Identification and Analysis in online Networks", we take the approach of actively searching, modeling, predicting, and highlighting threats on the Web, and study this particularly with respect to governmental institutions. The goal of our approach is to automatically monitor selected (sports) apps and analyze their collected data, correlate them with social media profiles and form clusters of individuals to identify potential targets and assess their risk potential. This is based on processing texts (e.g., tweets), images (e.g., selfies in front of buildings, maps), and geospatial information (e.g., running routes). This means that we are dealing with a heterogeneous data set. Due to its composition, very different requirements are placed on the processing methods. Since a so-called „digital twin" can be reconstructed in this way during data analysis and knowledge extraction, extremely sensitive (meta-)data is generated [6]. By correlating this information with other classified data, it is possible to determine the threat plausibility for the respective (groups of) individuals or locations. To achieve these goals, the technical implementation must combine, among other things, methods of information retrieval with approaches from forensic linguistics. In addition, methods of network analysis and clustering are to be used to develop novel evaluation functions for the identification of targets (persons, locations, etc.) based on the disclosed information.

In this paper, we present our understanding of the topic, but also our approach and our prototype, which we are continuously developing. The paper is organized as follows. In Section 2, we review the current state of the art in research, focusing on existing methods and definitions, as there is often a lack of a unified terminology. In Section 3, we present our own approach that we are taking in ADRIAN, starting from targeted data collection, data aggregation and enrichment, and interactive visualization. In Section 4, we present our work on a prototype and discuss our approach in Section 5 before concluding and providing an outlook in Section 6.

2.0 THEORETICAL BACKGROUND

In this section, we describe, first, the terminology (s. Sec.2.1) that shapes the research subject, and second, the existing methods and research in this area (s. Sec. 2.2).

2.1 Basic Concepts

In our work, we investigate the threats that individuals expose themselves and others to when they knowingly or unknowingly, actively, or passively, share heterogeneous information on social networks over a short or long period of time. Unlike, for example, the risk of being hit by a car, this threat is abstract and more difficult to name. To understand an abstract threat, it is important to identify what is being threatened and to what extent. Basically, we define threat as a situation that can lead to a negative impact on the individual or others.

We investigate user-generated content on social networks, some of which is shared anonymously and some of which is not, and between which there is a visible or no visible connection. We define a social network as a (commercial) web-based service that provides the ability to create a public or semi-public profile, share information, and build relationships [9]. User-generated content comes from ordinary people contributing data, information, or media that then appears before others on the Web – for example, restaurant reviews, wikis, and videos [10]. This content is often created with a specific intention, for example, to convey a social status. In doing so, every activity on these services gives the impression of having control over the disclosure of data. However, almost every activity on these platforms also conveys more information than it appears. For this reason, there is the term „digital footprint" that users leave behind when using such services. The active digital footprint consists of the information that one willingly shares with the world. This can be, for example, completed online forms or online profiles. The passive digital footprint consists of the information that one unconsciously shares with the world. This can be screen size, installed browser plug-ins, time zone or information in image backgrounds [11]. In addition, there is a cross-service footprint, which results from the combination of many individual pieces of information from different platforms and can arise, for example, from the interlinking of social network profiles. Information on one platform can fill information gaps on another platform so that the overall profile is quite complete. This is not a new insight, but the basis for various commercial services that collect and aggregate information about individuals. However, it is also the basis for de-anonymization, blackmail and bullying on the Web. This systematic seeking for multiple sources of information about a person or institution is called doxing. „Doxing is the intentional public release onto the Internet of personal information about an individual by a third party, often with the intent to humiliate, threaten, intimidate, or punish the identified individual" [6]. Systematically collecting, combining, and analyzing personal information may „remove some degree of anonymity" and violate privacy [6]. Anonymity and privacy are threatened. For example, a person's name, which should be unknown, is instantly known through clever combination of public data. Personal information, locations, or employers, although not deliberately made public, are available. Briefly, anonymity means that a person or group cannot be identified. However, it can also be understood as a form of protection [6], for example, when opinions are expressed that threaten to cause offence. Although anonymity and privacy may be equally affected by a threat and are often mentioned together, they should be kept separate. Privacy can be defined as "someone's right to keep their personal matters and relationships secret"[1]. Moreover, privacy violation refers to the invasion of an Internet user's privacy by collecting, combining, and enriching personal information.

In ADRIAN, we examine not only the acute threat, but also how individual information contributes to the threat and how it evolves over time. Potential threat over time describes the increasing likelihood of facing an individual threat to anonymity and privacy as the number of data points increases. We believe that a certain amount of information, or a certain combination of information, is required to pose an acute threat. Preliminary research suggests that there is such a potential de-anonymization threshold that we can use to develop models that can warn users [12,13]. The potential de-anonymization threshold defines a point in time when there are enough data points to weaken an individual's anonymity to a threatening degree. Methodologically, we are not at the beginning, as there is already prior work we can draw on (s. Sec. 2.2).

2.2 Methods

The subject of this research project can be approached methodically in different ways. In the following, existing methods in the field of identification (s. Sec. 2.2.1), measurement (s. Sec. 2.2.2), prevention of potential privacy risks (s. Sec. 2.2.3), and feature selection in texts (s. Sec. 2.2.4) are discussed.

2.2.1 Risk Identification

As discussed in Sec. 2.1, anonymity and privacy are threatened when a user's identity is publicly revealed in social networking platforms (SNs) against the user's intent. Because users, and often their friends, intentionally or unknowingly reveal information in one SN or another, the combination of this information could provide a clue to a user's identity. Therefore, researchers have tried to combine the information obtained from different SNs and calculate the similarity to determine the identity. The main methods were categorized as either graph-based or public information-based or a combination of both [14,15]. While graph-based de-anonymization has attempted to find user pairs with the largest number of neighbor pairs in community-level or social networks, public information-based models have used public information about SNs to correlate actual user identity.

Based on a graphical model, Halimi and Ayday [15] proposed a belief-based algorithm to match the real user profile from different SNs. They formulated the profile matching as an inference problem that infers the paired profile pairs. They tried their model on three different corpora pairs collected from SNs and showed its feasibility. To identify security risks for users of social networks at an early stage, analysis methods are used which, in addition to the user's personal information, also consider the user's social behavior based on various features (statistical features, social graph features, semantic features). With the help of a security risk

[1] See: https://dictionary.cambridge.org/de/worterbuch/englisch/privacy. Accessed: 2021-08-23.

score calculated based on the security assessment, vulnerable individuals can be identified at an early stage and sensitized to the danger of an attack even before it takes place [14].

A correlation between a person's social media usage behavior and that person's risk of becoming a victim of a cyberattack (victimization risk) has so far only been demonstrated for individual factors. A person's general willingness to take risks when interacting with other people also seems to increase the risk of an attack [16].

2.2.2 Risk Measurement

One way of measuring risk is to determine a privacy score, which can be used to quantify the extent of information disclosure (profile exposure). This considers both the user's handling of his own privacy (privacy preferences and attitude) and the network in which the user moves, and which surrounds him [17]. Pensa et al. [17] have applied the well-known PageRank theory [18] to the concept of privacy risk measurement and hypothesized that the more a person is surrounded by friends who neglect privacy, the more likely that person's privacy is exposed to privacy risk. They proposed a centrality-driven privacy rating system that considers the level of privacy awareness and experimented with a group of real Facebook users to assess the risk derived from privacy attitudes. Their main contribution to measuring objective privacy risk is that each privacy metric should be contextualized within the social graph by considering its influence on the other metrics. Attackers are already using powerful algorithms for profile matching attacks to efficiently link user profiles across platforms with high accuracy. Similarity between two profiles on different platforms is determined based on various attributes (e.g., name, location, gender, profile picture, profile description text, activity patterns, interests, mood, networking). Existing approaches for modeling profile matching in social networks are insufficient and inefficient for real-time privacy risk measurement [15].

2.2.3 Risk Prevention

The most effective strategy for reducing (system) risk is to create an increased awareness among users of the need to protect their own/personal privacy. This strategy also appears to be advantageous from an economic point of view, e.g., compared to investments to identify and eliminate existing security vulnerabilities [19]. Halimi and Ayday [15] also mentions the creation of an awareness of risk on the part of users as an essential contribution to risk minimization [15]. Those who have already been exposed to cybercrime in the past (whether as a victim or acquaintance of a victim) adapt their online behavior [16].

2.2.4 Feature Selection

How the (unconscious) disclosure of information manifests itself in linguistic expressions has been insufficiently studied so far. In our own preliminary work, we were able to show that linguistic expressions often contain more information than it first appears [20] and that they gain significantly in meaningfulness when combined with meta-information (e.g., date of assessment, age, health insurance, location of practice) [21]. In previous work, most predefined patterns were used for recognition, which did only limited justice to the creative freedom of natural language and could only recognize obvious (explicit) information revelations [21]. In this context, there is also other previous work, such as that of Sweeney [22] and Mamede et al. [23]. Also worth mentioning is the NETANOS (Named Entity-based Text ANonymization for Open Science) tool by Kleinberg & Mozes [24], which can detect, and highlight named entities such as person names in texts. However, to date, these have always been named entities whose literal mention may pose a threat to the privacy of the individuals concerned, but whose identification is trivial compared to coping with the expressive complexity of privacy threats in text. This is because there is still a lack of knowledge about precise linguistic expression and linguistic methods that can access it. However, this is indispensable to identify corresponding privacy-threatening text components and provide them with an explanation of potential risks. An early proof-of-concept is the TextBroom tool, which addresses the challenge of detecting information disclosure through a multi-layered, computational linguistic processing pipeline [25]. It has been shown that automatic identification of isolated privacy-prone statements is possible through a step-by-step analysis of user-generated texts and domain-specific knowledge resources. However, this method does not yet fully meet the challenge because it does not consider the interaction of individual pieces of information.

3.0 APPROACH

In this section we present our approach. It includes data acquisition and preprocessing (s. Sec. 3.1), data enrichment and correlation (s. Sec. 3.2), and interactive visualization (s. Sec. 3.3).

3.1 Data Acquisition & Preprocessing

The „digital twin" is created based on data from various popular social networks and publicly available sports apps. Social sports apps do not only record the user's activities, but also provide meta-information on location data, challenge events to achieve the goal, statistics on personal activities including other analytical tools such as personal and global heat maps. They are also connected to the most popular social networking services such as Twitter, Facebook, Instagram, etc. We will collect data from sports apps and create basic user profiles consisting of the usual habits and specific activities, including temporal and geospatial data, as well as basic information such as username and location.

Since the user profiles are often correlated with popular social networks via an API, the correlated user profiles are found and integrated in the corresponding social networks. If no correlated information is available, sports groups are searched in social networks with specific search queries such as 'running in Munich' or 'horse riding in Munich'. Most social networking platforms such as Twitter and Facebook list a variety of groups related to these search queries. The members of a group are then searched in the sports apps. So, we will develop a focused crawler to filter athletes in sports apps on the one hand and search correlates in social networks on the other hand. Furthermore, we will find groups in social networks with specific queries and search the members in sports apps. Although user profiles are often correlated from different social networking platforms, users often do not want to reveal their basic information such as name and location on different social networks. In addition, each social networking platform has its own characteristics. While Twitter is more of a platform that offers micro-blogging with a 280-character limit for non-CJK languages, Facebook is more of a place where friends communicate and share personal information and activities, and Instagram is more of a platform for uploading photos and videos. Due to this aspect, data from different social networks is very heterogeneous and therefore should be prepared for further processing. For this reason, our preprocessing pipeline includes the following four steps: (1) Data cleansing removes noise and marks missing features. (2) Data normalization is used especially for named entities such as name, location, date, geospatial data, etc. (3) Feature selection is the process of selecting the most relevant features for profile matching. (4) Feature conversion transforms the selected features into the appropriate format.

3.2 Data Enrichment & Correlation

Data collected from different sources is correlated and integrated into a graph network together with their labels, attributes, and properties. Figure 3-1 shows an example of correlated data from different sources. Data from two social network services and a sports app are connected. As mentioned in Sec. 3.1, the user provided different names between the social networking platforms (sky and dark green circle) and the sports app (green circle). Nevertheless, it often can be successfully correlated. The orange and purple circles indicate timestamps and activity IDs, respectively.

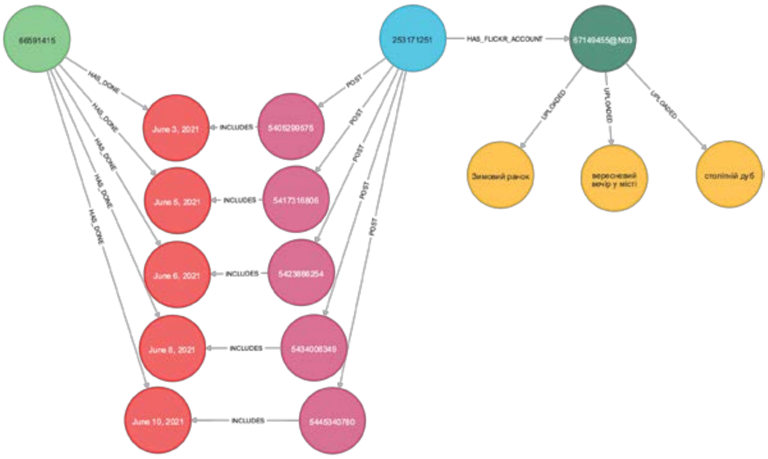


Figure 3-1: Example of data enrichment

The correlates in different social network platforms are determined based on different methods. One of the crucial methods is to compare profile photos in different SNs. Social sports apps often allow the athlete to log in through another SN account and the profile photo is taken from that SN account. So, if the user does not change the profile photo in the SN account, the profile photos remain the same. The comparison can often be easily done using template matching and its histogram, as shown in Figure 3-2.

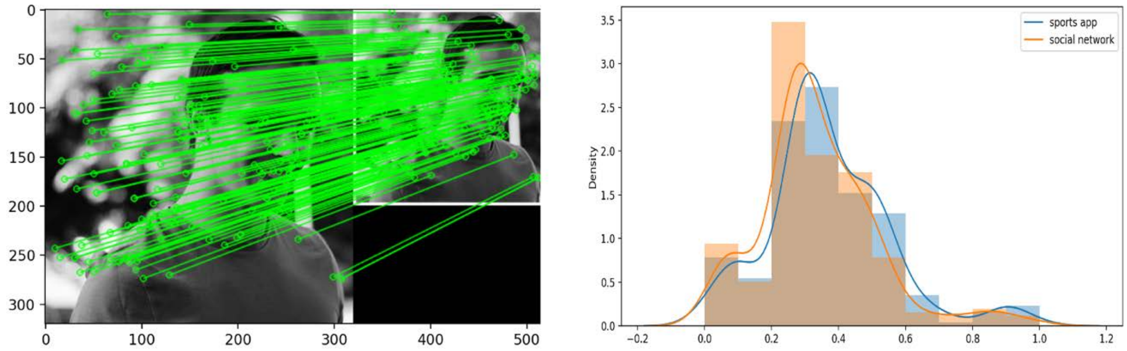
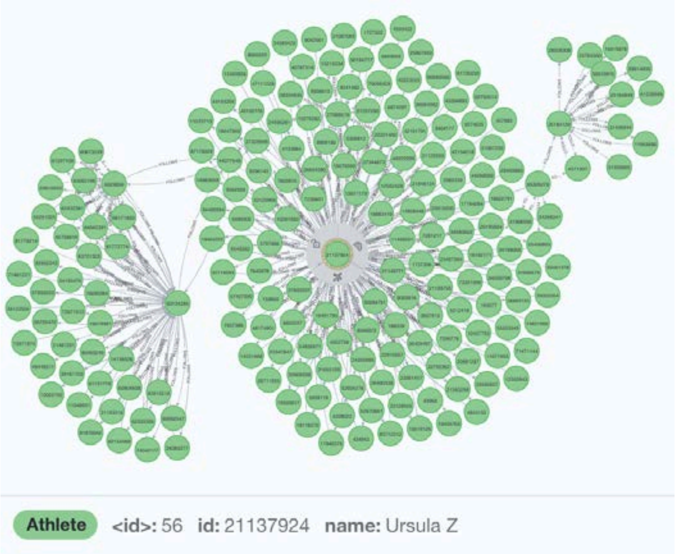


Figure 3-2: Example of template matching and its histogram

Otherwise, we will use Convolutional Neural Networks (CNN) and the openFace tool for deep comparison of profile photos. In addition to profile photo similarity, other similarities are also considered to ensure correlates. For example, Halimi and Ayday [15] define the similarity metrics in the following data points: username, location, gender, profile photo, plain text, activity pattern, interest, mood, and graph connectivity. For the similarity calculation, we will adopt these features. However, since the features are not always available in all social networks and the user may have different names, the features are weighted differently.

3.3 User-Centric Visualization

The correlated data is stored in a graphical database system. The graphical connectivity between users is visualized and the data can be easily retrieved using the queries. For example, various centralities such as degree, closeness and betweenness are calculated and retrieved. Figure 3-3 shows the detailed follower relationship of the athlete Ursula Z. In general, users that have the highest in-degree centrality can influence the information flow more than others in this group.



Athlete <id>: 56 id: 21137924 name: Ursula Z

Figure 3-3: Example of follower relationship

4.0 PROOF-OF-CONCEPT

In this section, we present our proof-of-concept. It includes the framework (s. Sec. 4.1) and the project-accompanying prototype development (s. Sec. 4.2).

4.1 Framework

A core objective of ADRIAN is to develop a framework that documents the entire approach we have developed from data collection to analysis in a traceable way, allows for extensions, and makes the key results reusable even after the end of the project. To facilitate understanding and reusability, we based our framework development on OSEMN (Obtain, Scrub, Explore, Model, and iNterpret) [26], a standardized and widely accepted model for data science research (s. Fig. 4-1). Since there are several challenges in data science [26], the OSEMN process addresses them by providing a precise sequence of operations, although the instantiation of each operation is not described in detail at the framework level. In the following, we describe our approach to each OSEMN step below.

As mentioned earlier, the OSEMN process provides a clear sequence of activities [27]: Obtain, Scrub, Explore, Model and iNterpret. „By following these steps, the entire process can be well planned and organized – starting with data acquisition to the analyzed data results visualization” [27]. The process begins with a step called Obtain, which focuses on identifying relevant data sources to obtain a sufficiently large data corpus. In our case, the data acquisition targets various data sources, such as text, images, and videos from social media networks and social sports platforms. Since the quality of the data can vary widely and user-generated data can usually be assumed to be of poor quality, initial data cleansing is already included in this step. However, this cleansing is far from sufficient and does not include data normalization steps, for example. These take place in the subsequent step Scrub. Since the collected data comes from different sources, data preprocessing is essential. Two necessary preprocessing methods are data standardization and normalization. For example, our data sources contain different date standards and location information. Furthermore, the attribute types and ranges must be normalized. This is important not only when designing custom databases and graphs, but also when linking the resulting resources to other resources using Linked Data standards. By adhering to standards and normalization principles, it becomes possible to perform data exploration across data set boundaries using existing tools and procedures.

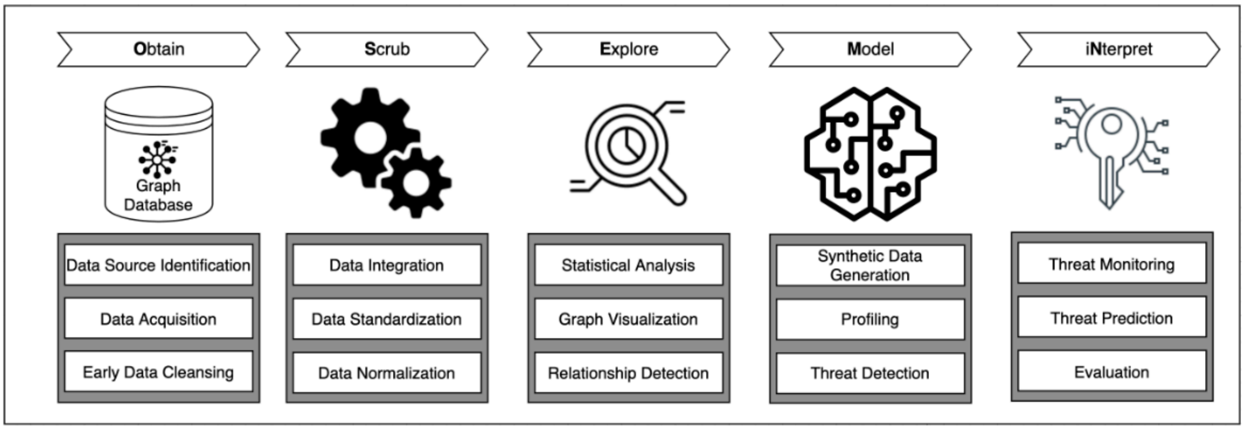


Figure 4-1: ADRIAN framework

In our scenario, the most important task is to identify relevant data points to explore relationships in the data. This task is part of the Explore step in the OSEM process. We use a graph database to store and analyze data from various sources. Thanks to standard compliance, missing information, for example details of locations, can be integrated from other sources (e.g., Linked Open Data). Thereby, revealing correlations between different data points is a crucial detail. Furthermore, the created graphs provide a visualization of the networks and enable graphical exploration of the data sets. We are also developing a web application for statistical analysis and visualization. However, in our research project we must assume that there may also be bottlenecks in data acquisition. Therefore, it is already planned at the beginning that synthetic data will also be generated. Therefore, enriching data with synthetic data is essential for data modeling and AI model performance and is part of the Model step. Data synthesis is a method of creating an „artificial" representation of an original data set. To accomplish this, a model is developed that explains the original data as well as possible. From this model, new data are generated that retain important statistical properties of the original data set. For our approach, it is necessary to identify and compare AI methods that are suitable for working with heterogeneous data. Using state-of-the-art AI models, we generate additional knowledge and analyze the relationships between user profiles in more detail. This knowledge and analysis can be used, for example, to identify groups of people in specific locations to detect potential threats. The use of the results falls into step iNterpret. Since our approach is a human-in-the-loop approach, especially in the current phase of the project, interpreting the results, correcting them, and feeding the corrected data back into the training data is an essential step. In addition, threats are to be made visible. For this purpose, we are developing a dashboard that can show detected threats and operate in live mode. We are investigating whether potential risks can be detected as they arise based on the data we generate and collect. For example, on certain topics that are trending on the networks or based on a growing database of an individual user.

4.2 Prototype

Identifying and visualizing threats on the web based on user data is a topic that thrives on real data. Therefore, we aim to work with mostly real data from the beginning and to develop methods and tools that interact with this data. Therefore, the project will develop a prototype that includes all the features of the presented framework. We expect to encounter constraints of various kinds (data acquisition, model training, interpretation) already during the development and thus we will be able to identify and address the challenges. In terms of agile software development, we try to reduce the design phase to a minimum through an iterative and incremental approach and to arrive at executable software as early as possible in the development process. Now, we are still at the beginning of the research project, so we are focusing on the basic functions during prototype development. Nevertheless, we can already see how helpful the development accompanying the project is, as new questions and challenges emerge in live operation.

In the following, we provide information on the current state of development. As for data acquisition, we are collecting raw data from different social media and social sports platforms. The data collected varies from platform to platform. Therefore, it is important to quickly select the desired data from many data points. In our current approach, we stream tweets by keywords in real time to our Neo4j database via the Twitter API. The prototype then visualizes the amount of data acquired by language and keywords. We can also track the location of tweets in an interactive map. In addition, extensive heatmaps can be created once enough data points are collected. Figure 4-2 shows a screenshot of the current development for Twitter.

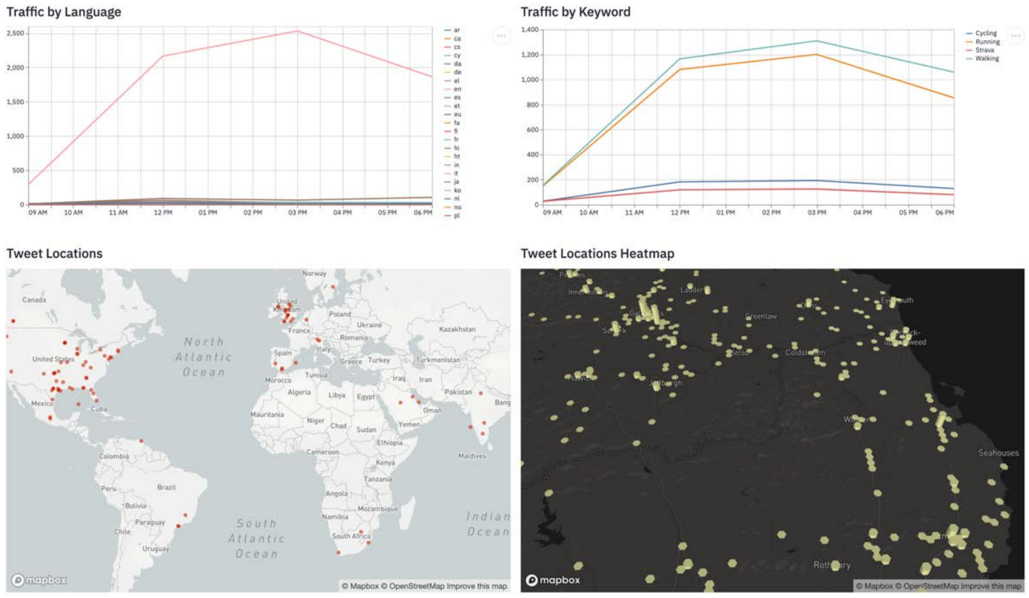


Figure 4-2: Information about the acquired tweets

In the area of sports platforms, we focus on visualizing the public activities of athletes. In Figure 4-3, a screenshot of the current development is shown. The geospatial streams contain information such as altitude, latitude, and longitude. By visualizing this geospatial data, we gain insights into the activities at specific locations. Currently, it is possible to select the athlete and one or more activities and display them on the map. It is also possible to create a heat map to see exactly where the athlete has spent most time. As development continues, places of interest will be marked to allow for targeted analysis of locations or buildings of interest.

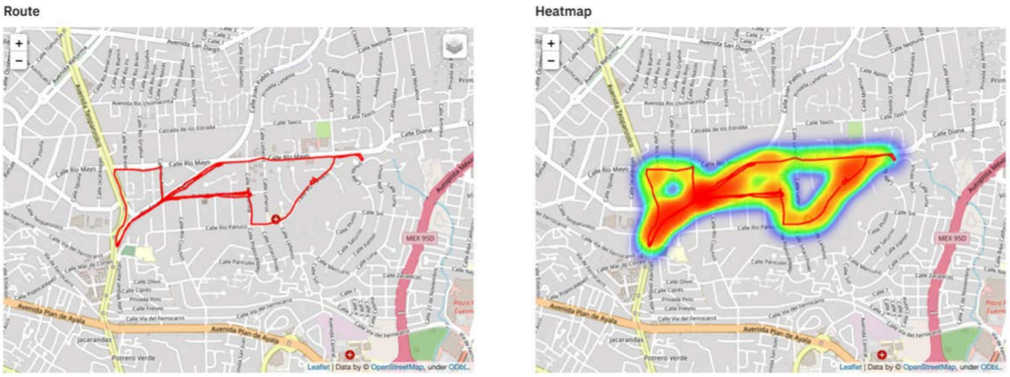


Figure 4-3: Geospatial data for a specific athlete

In addition, the prototype is used to create machine learning inferences for the trained models. In this case, different settings for the models can be quickly tested, visualized, and interpreted.

5.0 DISCUSSION

The project plan is ambitious and foresees several challenges. First, we anticipate a whole range of challenges in data acquisition, management, and enhancement. By its very nature, not all the data we use to train models and alert users will be publicly available. We must limit our data collection to what is publicly viewable. This is consistent with the project goal but needs to be communicated to users. At the same time, it is not realistic to monitor the web live. Therefore, only a retrospective view will ever be possible in the ADRIAN project. Furthermore, it is also questionable how well the synthetically produced data resembles real-world data and how suitable it is for training the models. This is a clear limitation because we rely on synthetic data to address the cold-start problem.

Second, it is also always questionable how models trained on a particular data set will behave on new data. The use case of the project envisions the integration of social networks and especially social sports apps, but of course aims to be applicable beyond this (e.g., for physician rating websites). For this reason, we are already trying to make the data in our project context diverse by including different countries, topics, portals, etc. Nevertheless, it is impossible to predict today what data will be available tomorrow and in what form it will exist. Therefore, there is always a residual risk that the models we develop in ADRIAN will not be usable for future challenges. We think that by developing a framework that is designed to be applied, extended, and interpreted, we are providing a solid foundation for this challenge. When we talk about models that do some form of classification, we always have to talk about the issue of misclassification. Since there is no automatic deletion or blocking of content, and we only try to warn users about data collection, we see no risk of accidental censorship or other restrictions due to misclassification. Nevertheless, it can be problematic if too many misclassifications lead to false warnings that shake trust in the system. On the one hand, we want to counter this with good and sophisticated models in combination with rule sets; on the other hand, we want to educate people about how our methods work and their limitations.

Third, this leads to the fundamental question of when too much information is available and when users should be warned. This is an individual decision, as users define and evaluate the degree of privacy and anonymity quite differently for themselves. Nevertheless, in our project we must work with such a value and the assumption that all users want a certain degree of privacy and anonymity, even if this is not conclusively proven. Therefore, after data collection, we must determine what level of trade-off in terms of the amount of partial information seems appropriate based on the available data.

6.0 CONCLUSION & FUTURE WORK

In this paper, we reported on ADRIAN, a research project that addresses threats on the Web by combining a wide range of information. The threats that individuals and institutions face from careless information sharing can be diverse and therefore warrant a multi-layered identification approach. In our research project, we take a prototype-driven research approach using real-world data to identify and solve the challenges and problems. As shown, we are guided by the OSEMN framework. This framework allows us to work in an agile way and to develop individual components quickly, but also to adapt them to changing conditions. As also pointed out in the discussion, we are aware of the challenges that need to be overcome. We believe that, based on the framework orientation, we can contribute to making the Web more secure and, ideally, also to building an understanding of the threats through the individual components that are being developed in this research project. Not least for this reason, public relations work is also planned in the project.

In our further work in the research project, we are focusing primarily on data acquisition, normalization, and modeling. One goal is to be able to generate and use synthetic data in a timely manner to be able to train initial models for the classification of questionable web content and data collections, which can then already take heterogeneous data sets into account.

ACKNOWLEDGEMENTS

This research is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr within the project MuQuaNet.

REFERENCES

[1] R. Jones, R. Kumar, B. Pang, and A. Tomkins, "'I know what you did last summer': query logs and user privacy," in Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07. Lisbon, Portugal: ACM Press, 2007, p. 909.

[2] A. T. McKenna, A. C. Gaudion, and J. L. Evans, "The Role of Satellites and Smart Devices: Data Surprises and Security, Privacy, and Regulatory Challenges," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3418420, Jul. 2019.

[3] J. Pinchot and D. Cellante, "Privacy Concerns and Data Sharing Habits of Personal Fitness Information Collected via Activity Trackers," JISAR, vol. 14, no. 2, p. 4, Jun. 2021.

[4] J. MacAllister, "The Doxing Dilemma: Seeking a Remedy for the Malicious Publication of Personal Information," Fordham Law Review, vol. 85, no. 5, p. 2451, Apr. 2017.

[5] A. Cheung, "Doxing and the Challenge to Legal Regulation: When Personal Data Become a Weapon," in The Emerald International Handbook of Technology Facilitated Violence and Abuse, J. Bailey, A. Flynn, and N. Henry, Eds. Available: https://www.emerald.com/insight/content/doi/10.1108/978-1-83982-848-520211041/full/html

[6] D. M. Douglas, "Doxing: a conceptual analysis," Ethics and Information Technology, vol. 18, no. 3, pp. 199–210, Sep. 2016. [Online]. Available: https://doi.org/10.1007/s10676-016-9406-0

[7] R. Perez-Peña and M. Rosenberg, "Strava Fitness App Can Reveal Military Sites, Analysts Say," The New York Times, Jan. 2018. [Online]. Available: shorturl.at/gtDJX

[8] D. Lindsay, The 'right to be forgotten' by search engines under data privacy law: a legal and policy analysis of the Costeja decision, ser. Cambridge Intellectual Property and Information Law. Cambridge University Press, 2016, p. 199–223.

[9] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," Journal of computer-mediated Communication, vol. 13, no. 1, pp. 210–230, 2007.

[10] J. Krumm, N. Davies, and C. Narayanaswami, "User-Generated Content," IEEE Pervasive Computing, vol. 7, no. 4, pp. 10–11, Oct. 2008.

[11] I. S. Germany, Munich, "Der digitale Fußabdruck – diese Spuren hinterlassen wir als Online-User," Apr. 2020, accessed: 2021-07-06. [Online]. Available: https://www.infopoint-security.de/der-digitale-fussabdruck-diese-spuren-hinterlassen-wir-als-online-user/a23386/

[12] F. S. Bäumer, N. Grote, J. Kersting, and M. Geierhos, "Privacy Matters: Detecting Nocuous Patient Data Exposure in Online Physician Reviews," in Information and Software Technologies, R. Damaševičius and V. Mikašytė, Eds. Cham: Springer International, 2017, vol. 756, pp. 77–89.

[13] F. S. Bäumer, B. Buff, and M. Geierhos, “Potentielle Privatsphäreverletzungen aufdecken und automatisiert sichtbarmachen,” in DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts, P. Sahle, Ed. Zenodo, 2019, p. 192–193.

[14] B. Feng, Q. Li, Y. Ji, D. Guo, and X. Meng, “Stopping the Cyberattack in the Early Stage: Assessing the Security Risks of Social Network Users,” Security and Communication Networks, vol. 2019, 2019.

[15] A. Halimi and E. Ayday, “Efficient quantification of profile matching risk in social networks using belief propagation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12308 LNCS, 2020.

[16] J. Jalil and G. Sinnamon, “Risks of online victimisation among college students’ on mobile social networks,” International Journal of Cyber Criminology, vol. 13, no. 2, 2019.

[17] R. G. Pensa, G. Di Blasi, and L. Bioglio, “Network-aware privacy risk estimation in online social networks,” Social Network Analysis and Mining, vol. 9, no. 1, 2019.

[18] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” Comput. Netw. ISDN Syst., vol. 30, no. 1-7, pp. 107–117, Apr. 1998.

[19] S. Du, X. Li, J. Zhong, L. Zhou, M. Xue, H. Zhu, and L. Sun, “Modeling Privacy Leakage Risks in Large-Scale Social Networks,” IEEE Access, vol. 6, 2018.

[20] M. Geierhos and F. S. Baeumer, “Erfahrungsberichte aus zweiter Hand: Erkenntnisse über die Autorschaft von Arztbewertungen in Online-Portalen,” in DHd-Tagung 2015, Graz, pp. 69–72.

[21] F. S. Bäumer, N. Grote, J. Kersting, and M. Geierhos, “Privacy Matters: Detecting Nocuous Patient Data Exposure in Online Physician Reviews,” in Information and Software Technologies, R. Damaševičius and V. Mikašytė, Eds. Cham: Springer International Publishing, 2017, pp. 77–89.

[22] L. Sweeney, “Replacing personally-identifying information in medical records, the scrub system.” In Proceedings of the AMIA annual fall symposium. AMIA, 1996, p. 333.

[23] N. Mamede, J. Baptista, and F. Dias, “Automated anonymization of text documents,” in2016 IEEE Congress on Evolutionary Computation (CEC), 2016, pp. 1287–1294.

[24] B. Kleinberg and M. Mozes, “Web-based text anonymization with node.js: Introducing netanos (named entity-based text anonymization for open science),” Journal of OSS, vol. 2, no. 14, p. 293, 2017.

[25] F. S. Bäumer, J. Kersting, M. Orlikowski, and M. Geierhos, “Towards a Multi-Stage Approach to Detect Privacy Breaches in Physician Reviews,” in Proceedings of the Posters and Demos Track of the 14th Int. Conference on Semantic Systems, CEUR Workshop Proceedings, A. Khalili and M. Koutraki, Eds., vol. 2198, no. 97. CEUR-WS.org, 2018.

[26] H. Mason and C. H. Wiggins, “A Taxonomy of Data Science,” 2010, accessed: 2021-07-06. [Online]. Available: http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

[27] K. Dineva, T. Atanasovaet al., “OSEMN process for working over data acquired by IoT devices mounted in beehives,” Current Trends in Natural Sciences Vol, vol. 7, no. 13, pp. 47–53, 2018.