Kraken.me – Multi-Device User Tracking Suite

Immanuel Schweizer
TU Darmstadt - Telecooperation
Hochschulstrasse 10
64289 Darmstadt, Germany
schweizer@tk.informatik.tu-darmstadt.de

Benedikt Schmidt
TU Darmstadt - Telecooperation
Hochschulstrasse 10
64289 Darmstadt, Germany
benedikt.schmidt@tk.informatik.tu-darmstadt.de

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2638728.2641307

Abstract
An in-depths understanding of human activity is a relevant contribution to the design of interactive systems to support human activity. This is of explicit relevance for assistance systems building on prediction and recommendation.

However, the understanding of human activity is limited. Albeit the omnispresence of smart phones and computers, the actual execution of complex activities with those devices in relation to context factors is not completely understood. One possible reason is the limited amount of activity related data to perform actual research.

In this paper, we present the Kraken.me framework to address this lack of information. Kraken.me is the first tracking suite to offer integrated tools for mobile, social, and desktop tracking. It is also, to our knowledge, the first tool to emphasize the collection of data from both physical and soft sensors. In this paper, we will introduce the overall architecture, system components, and future research ideas for Kraken.me.

Author Keywords
User tracking suite, Crowd Sensing, User-centric Sensing

ACM Classification Keywords
H.1.2 [Models and Principles]: User/Machine Systems

Introduction

More and more applications and services support users by eliciting preferences, habits and plans. Consider recommender systems or systems for personal assistance, e.g, Google Now, Microsoft's Cortana or Apple's Siri. To address the resulting requirements, knowledge about human activity is crucial. Therefore an in-depths understanding of human activity, short term as well as long term is an ongoing research topic for science and industry. For this research activity data is required. Data which emerges in the daily business of many companies such as Google and Facebook whose business includes the collection of information about individuals and their world. Activity research in the industry is beneficial as activity knowledge is a business value resulting in better services. However, due to the business value the gained knowledge is only shared partially.

For research this poses a challenge. How to provide researchers with access to data that is as rich or even richer than what companies collect from millions and billions of users? For example, one of our current research directions is intention prediction and personal assistance systems. But, whatever type of prediction one wants to research, the initial requirement is always the same. We require a suitable data set for the topic at hand.

In the past people have collected very focused data sets for individual research questions [6]. These data sets are great to answer these specific questions. However, research domains like intention prediction will benefit from data sets which capture many different characteristics and attributes to enable exploratory studies and enable the comparison of approaches which build on different attributes of the data. Even the data sets available for companies will focus characteristics and attributes from the core services offered by the respective company. We envision a data set with a broader scope. This is where research is in a unique position as it can be vendor agnostic. And this is why we build Kraken.me[1]. Kraken is the first suite to integrate tracking tools for mobile, social, and desktop. But we cannot only track users in a multi-device environment. We also emphasize the integration of both physical (e.g., location, acceleration, etc.) and soft (e.g., software interaction, contacts, calender, etc.) sensors.

To give you an overview about the depth collected by Kraken.me, we will shortly specify the challenge tackled by Kraken.me, scope and design principles. Then we will introduce the overall system architecture of Kraken.me. Afterwards, we will discuss every Kraken.me tool and what information is collected. Last but not least, we want to share and discuss what we see as the overall challenge and possible contributions of the Kraken.me system down the road.

Challenge

The purpose of Kraken.me is the collection of data about activities in terms of interaction histories and user related data.

We know that interaction histories are an invaluable resource for research in the context of human activity and user support. An interaction history provides a detailed log about the interaction of a user with a device or system. The relevance of this data for the analysis of human activity has been recognized early [1]. However, interaction histories have always been created for individual research projects and generally focused on a very specific kind of interaction information.

[1] https://kraken.me

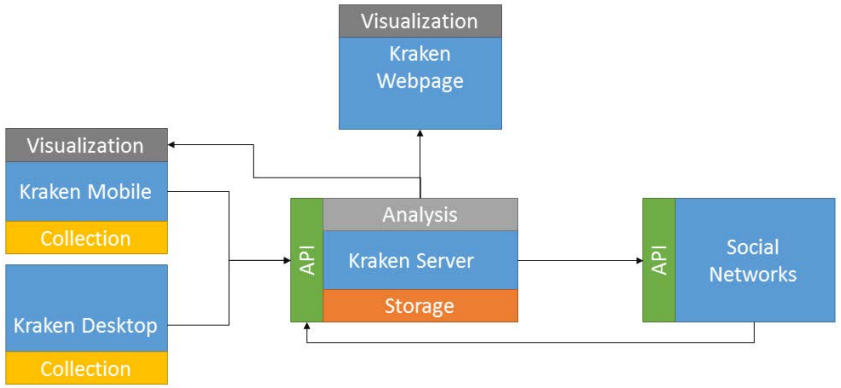

Figure 1: Architecture of Kraken.me

To actually understand interaction and respective activities a more holistic approach is required. Consider the way people interact with devices. The windows user who collects information from a PDF file and a browser, pastes the data into a spreadsheet, asks a colleague in a messaging application for details and finally sends the complete spreadsheet to a mailing list. Consider the mobile user who quickly switches between multiple apps to plan a subway trip while coordinating a meeting with a friend via phone while accessing the named spreadsheet. These examples illustrate that interaction goes beyond application and even device boundaries. Individuals develop complex interaction processes to use information processing and developing tools in an optimized way to realize complex activities and pursue respective goals. A connection between interaction, activity and goal which has not yet completely been understood.

For research on user activity, we see a demand for cross application and cross device data about users: a tool suite to create extensive and rich interaction histories.

Furthermore, interaction is closely related to the interacting subject, a social individual. Therefore, we argue for an integration of information from social networks with interaction histories.

Scope and Design Principles

Primarily, Kraken.me is a platform for the creation of extensive and rich interaction histories across multiple devices combined with information about users. The interaction histories can be connected to data from social networks for more in-depth information.

Furthermore, Kraken.me is a platform for the rapid prototyping of research results in terms of user support functions like recommender and prediction algorithms. Based on a modular architecture and online update mechanisms the development and delivery of such functions is simple.

The design of Kraken.me was guided by the following basic - partially contradictory - design principles:

- Greediness: Kraken.me collects all available data. Exception to this rule is information required to access services and transactions (e.g. passwords, user names, etc.)
- Openness: Users are informed about the collected data, users can access their data and they can trigger the deletion of their data.
- Self-determination: Users can decide which data is actually collected by the system. While the greediness results in the integration of as many sensors as possible the self-determination regulates the actual collection. Kraken.me can be set into a deaf mode without collection of data and considers indication of privacy like the inPrivate mode provided by most web browsers.
- Secureness: The data collected by Kraken.me needs to be stored and transmitted securely.
- Unobtrusiveness: The data collection process of Kraken.me is intended to be as unobtrusive as possible. This means that no actual user interaction is required during the data collection process, access of the system is simple and battery life and transfer data volume are conserved.
- Usefulness: Kraken.me design considers the use of the data to actually help the user by providing information and services based on the collected data.

Kraken.me Architecture

Kraken.me aims at tracking users across devices. The goal is quite literally to use each and every sensor available on each of those devices. The overall architecture of Kraken.me is depicted in Figure 1.

Currently, Kraken.me offers access to three main sources of information. A mobile application available for Android that offers both data collection from soft and physical sensors and a first visualization. A desktop monitor for Windows that runs in the background and collects data from soft sensors (e.g., application interaction, accessed documents, mails, browser history, etc.). Social networks are the third source of information. Users can connect their social logins (e.g., Facebook, Google+, Twitter, etc.), which, given the user's permission, allows us to crawl the respective API.

We decided to go for a server architecture hosted on Amazon Web Services for scalability as the amount of data collected per user is significant. The central server has a number of tasks. The most crucial being data storage. All data is stored using flat tables to retain the raw data in almost original structure. These flat tables are later enriched into more valuable higher level structures, e.g., graph databases. The raw data is not anonymized as it is not feasible given the amount of personal information stored on the system.

The next sections will introduce all Kraken.me tools and their current data collection profile.

Kraken.me Mobile Monitor

Mobile phones are unique tools in that they provide access to a host of physical sensors (e.g., accelerometer, location, temperature, heart rate etc.) and a host of soft sensors (e.g., mails, calendar, contacts, etc.). Hence, mobile phones are at the core of most comprehensive personal tracking suites [2]. Kraken.me is not different, offering an Android application to track both physical and soft sensors. In designing Kraken.me, we drawn heavily from our experience developing other

Figure 2: Kraken.me Mobile

applications [11, 10]. Table 1 illustrates the data collected if the users opts for the full tracking profile.

Most sensors are easily available using the API or some standard procedure. However, since we want to dig deeper into the user's interaction with the device we used a trick that makes tracking on Android even more powerful. For Kraken.me to work the user is asked to enable access to the accessibility framework. The accessibility framework is for example meant to enable applications for the visually impaired, to navigate the operating system. Once an application is registered with the framework every interaction with the device is sent to the application. In theory, this allows for unfiltered tracking of every user interaction. This is also the main reason Kraken.me is available for Android only at the moment. This depth of data cannot yet be collected on iOS.

With this information, Kraken.me is able to identify and visualize (cmp. Fig. 2) means of interaction with the device, i.e. foreground app and interaction with the respective app, and relate this information to the location and physical surrounding of the user. In the future we want to extend both physical sensors and logging of user interactions. Especially health related data, e.g., step count, or environmental data, e.g. temperature and humidity, are not collected currently.

Data	Description
Sound pressure	dBSPL value captured with the microphone
Ring mode	Is the phone on silent, vibration, etc.
Acceleration	Readings from the accelerometer
Activity	Activity Recognition provided by the Android API
Illuminance	Lux as provided by the Android API
Network Connectivity	The active and all available network connections
Browsing History	All visited web sites excluding inPrivate mode
Contacts	All saved contacts
Calendar	All saved calendar events and reminders
Location	Location as provided by the Android API
Call Log	The call history
Foreground apps	Time span apps are in foreground
Keyboard	Statistics about key strokes

Table 1: Data collected by Kraken.me Mobile

What Kraken.me mobile brings to the table is the combination of physical sensor with an interesting take on soft sensors and interaction data.

Kraken.me Desktop Monitor

Kraken.me mobile is already an interesting stand alone tool for personal data collection. However, with the addition of desktop data it is even more valuable. WIMP style interfaces have a high relevance for the execution of

work with computers. The Kraken.me desktop monitor focuses the interaction of the user with the Windows operating system. The monitor captures the foreground processes and the interaction with the respective processes. For some applications additional information like the accessed information objects (e.g., text documents, websites) is captured. Table 2 provides an overview of the information monitored by the desktop monitor if the broadest collection profile is chosen. The desktop monitor is based upon our prior work [9] and relies heavily on functionalities provided by the UI Automation framework and available process diagnostics.

Figure 3: Kraken.me Desktop Monitor

Kraken.me Desktop has a lot of overlap with the mobile monitor when it comes to soft sensors, e.g., emails, contacts. It augments some sensors, e.g., the browsing and location history. As people might only be using the desktop monitor this is already crucial.

The desktop monitor goes above and beyond the mobile monitor by tracking the information objects, i.e. the documents people work on. This tracking of information objects is enabled for PDF, Word, Excel, and Powerpoint files. This means we can extract tracking data for the most common file types (cmp. Fig. 3).

Data	Description
Foreground Process	Program in interaction focus
Interaction with User Interface	Clicks and focus of user interface elements, e.g. buttons, menus, etc. with labels
Office Applications	The information objects accessed with the office suite (Microsoft Outlook, Word, Excel, Powerpoint) and Adobe Acrobat
Web Browsers	URL, title and fulltext of website accessed with Chrome, Internet Explorer, Firefox
Clipboard	Content in the clipboard
Files locked by process	Indicator for accessed information objects

Table 2: Data collected by the Kraken.me Desktop Monitor

Kraken.me Social Connector

The Social Connector is used to connect Kraken.me to existing social networks. To use the social connector, users login through the Kraken.me website to explicitly

connect their Kraken.me account with social networks. Based on collection profiles users can influence the amount of data actually collected from the connected networks. Once the connection is established, the respective API of social networks is accessed. As long as the connection exists, the connected social networks for each account are crawled once a week and new or changed information is persisted. The initial version of Kraken.me comprises connectors for Facebook, LinkedIn, Google+, Xing, Foursquare, and Microsoft Live.

Since each network offers different data, Table 3 provides an overview of the information which is collected for Facebook. Facebook offers the richest data set, hence, it is given as a reference point.

Data	Description
Profile	Profile information
Activity Stream	Recent network activities
Location	GPS position
Photos	Photos of the user and photos accessible by user
Friends	List of friends
Internal Messages	Metadata of exchanged messages without message text

Table 3: Data collected through the Facebook API

Most people are active in at least one social network. Thus, a lot of social data is readily available.

Data storage

Data collected by any of the aforementioned tools of Kraken.me is stored following two different strategies. First, a strategy of unfiltered, complete data sets. The raw data provided by the Kraken.me applications, the social connector, the Desktop monitor and the Android App is persisted in a flat, unfiltered form. Such a direct interface is useful for the creation of basic data sets to be processed with machine learning techniques. Second, a high level perspective is created using graphs. Researchers can create models which provide different perspectives on the data based on so called aggregators. An aggregator creates a Neo4J Database instance for a specific subset of the data. The high level perspective is of specific use for the human interpretation of the data and for application development.

In the future we plan to investigate offline and online analysis of the raw data layer as well as finding meaningful representations on the graph layer.

Kraken.me webpage

Currently, Kraken.me (cmp. Fig. 4) offers links to download the desktop and mobile monitor. A user can log into his profile and connect different social accounts and see some basic visualizations. Over time we hope that Kraken.me can morph into a showcase for our analysis, can offer personal assistance, and a rich visualization of interaction data.

For the time being, however, the most important sites are probably *Download* and *Privacy*.

Initial Directions

A variety of research directions benefit from the data collected by Kraken.me. In the following, we discuss our current research directions with first topics.

Data collection and processing We can already ask interesting questions before any meaningful analysis. How can data collection be as unobtrusive and energy efficient as possible? Which data can be

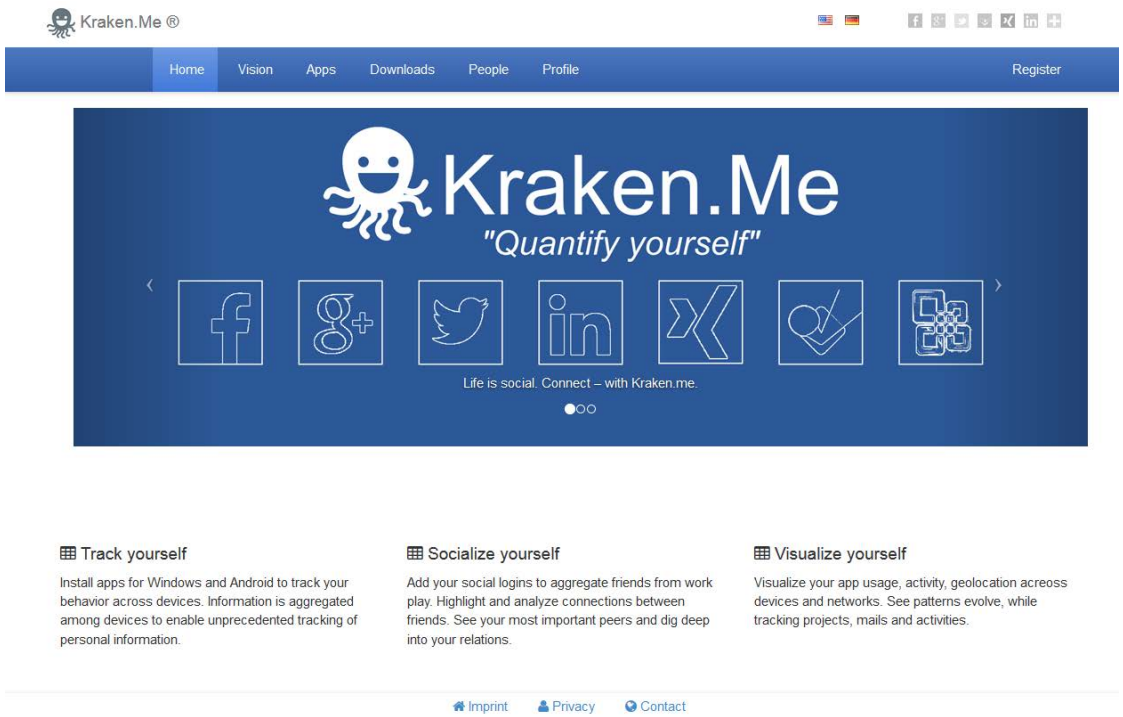

Figure 4: Frontpage of Kraken.me

processed locally? For example, can we reliably implement Named Entity Recognition on mobile phones and not transfer full text? These are only some of the questions we want to answer.

Sociology and interaction research The nature of human activity and relationships between interaction, activity and goal is an important topic. It is the foundation for the development of future recommender and predictive systems. Our first activity is the combination of a diary study with the Kraken.me data set to analyze the relationship between interaction, activity and goal. More specific we will ask people about their plans for the future and use their diary and our tracking to understand the difference between planning and execution.

Mnemonic support, Awareness and Quantified self Interaction data helps people to recall what they have done in the past, thus, getting an awareness of

their past. The data collected with Kraken.me will be the foundation for the development of search services on historic data (answering questions like "what did I do yesterday morning?") and to perform analytics on the individual history (relevant topics, trending topics, time spent with tasks, etc.). These applications are of specific relevance for the quantified self community which is also in the focus of our participant groups.

Recommendation, Predication and Assistance The historic data contains clues for predicting and recommending future actions. We want to understand and predict the intention of users. This will allow for much more powerful personal assistance. Going beyond hard coded use cases into a general intention theory.

Related Work

The Kraken.me tools realize a crowd sensing application which combines interaction monitors and crawlers. In the following we give a brief overview of related work on those three aspects.

Crowd sensing leverages crowds for the collection of data. From historic examples of spotting the route of monarch butterflies following newspaper requests the idea and the term especially stands for the use of mobile devices for the collection of data within a community [5]. Existing crowd sensing applications generally focus a very specific problem, like the collection of data about air pollution, creation of traffic information etc. [6]. Kraken.me applies the idea of data collection with a specific focus on the individual, considering data from physical sensors and from soft sensors for interaction tracking and crawling.

Soft sensors for interaction tracking are software programs which collect information about the interaction with applications. Soft sensors have been used early for the analysis of user activities with shells [1]. Other applications of soft sensors comprise the clustering of related user activities [8], use of the data for activity recognition based on machine learning [7] or for learning analytics in the domain of technology enhanced learning [3]. Different examples show the success of crawlers for the collection of data from social networks [4].

The overview shows that the main contribution of Kraken.me in comparison to the existing work is the integration of several existing data collection techniques. The combination of soft and physical sensors with the crawling of social networks to create interaction data enriched by social information is unique.

Conclusion

In this paper we introduced Kraken.me. Kraken.me provides tracking tools for mobile, desktop, and social. The additional integration of soft and physical sensors enables fine-grained user profiles.

The groundwork has been laid and we hope to collect a comprehensive data set over the next months. We also plan on sharing some anonymized subsets with the research community. Another obvious addition for the future are mobile and desktop monitors for other platforms to include even more possible users. Also, as more data is collected, we will increase the quantity and quality of visualizations, and offer more active support to users.

All of this will help answering some of the questions outlined above, some questions we might not even think of today, and maybe even advance our understand of human behavior at large.

References

[1] Liam Bannon, Allen Cypher, Steven Greenspan, and M.L. Monty. Evaluation and analysis of users' activity organization. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 54–57. ACM Press, 1983.

[2] Ning Ding, Daniel Wagner, Xiaomeng Chen, Y Charlie Hu, and Andrew Rice. Characterizing and modeling the impact of wireless signal strength on smartphone battery drain. In *Proceedings of the ACM SIGMETRICS/International conference on Measurement and modeling of computer systems*, 2013.

[3] Erik Duval. Attention Please ! Learning Analytics for Visualization and Recommendation. In *Proceedings of LAK11: 1st International Conference on Learning Analytics and Knowledge*. ACM Press, 2011.

[4] M Faloutsos, T Karagiannis, and S Moon. Practical Recommendations on Crawling Online Social Networks. *IEEE Journal on Selected Areas in Communications*, 29(9):1872–1892, 2010.

[5] RK Ganti, Fan Ye, and Hui Lei. Mobile crowdsensing: Current state and future challenges. *IEEE Communications Magazine*, (Section 2), 2011.

[6] ND Lane, Emiliano Miluzzo, and Hong Lu. A survey of mobile phone sensing. *IEEE Communications Magazine*, (September):140–150, 2010.

[7] Andreas S Rath and Stefanie N Lindstaedt. UICO: An ontology-based user interaction context model for Automatic Task Detection on the Computer Desktop. In *Proceedings of the 1st Workshop on Context, Information and Ontologies*, number Ciao. ACM Press, 2009.

[8] Tye Rattenbury and John Canny. CAAD: an automatic task support system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 696–706. ACM Press, 2007.

[9] Benedikt Schmidt and Eicke Godehardt. Interaction Data Management. In *Knowlege-Based and Intelligent Information and Engineering Systems*. Springer, 2011.

[10] Immanuel Schweizer. MoSeS: Mobile Sensing System. In *Informing Future Design via Large-Scale Research Methods and Big Data*, 2013.

[11] Immanuel Schweizer, Roman Bärtl, Axel Schulz, Florian Probst, and Max Mühlhäuser. NoiseMap - Real-time participatory noise maps. In *Second International Workshop on Sensing Applications on Mobile Phones*, 2011.