A Tool to Analyze the Reading Behavior of the Users in a Mobile Digital Publishing Platform*

Ludovico Boratto[1], Mattia Cadeddu[1], Salvatore Carta[1], Gianni Deplano[2], and Fabio Mereu[2]

[1] Dipartimento di Matematica e Informatica, Università di Cagliari, Italy
ludovico.boratto@acm.org, mattia.cadeddu@gmail.com, salvatore@unica.it
[2] Xorovo srl — Applix srl, Italy
{gianni.deplano, fabio.mereu}@applixgroup.com

Abstract. In their daily activities, users interact multiple times with mobile applications. This generates huge amounts of data related to these interactions that, when filtered and analyzed, would give insights on the behavior of the users while using an application. In this paper, we consider a real-world mobile digital publishing platform, named *Viewerplus*, which enables a digital, augmented fruition of content from traditional magazines. The objective is to develop a tool that allows the human editors to analyze the reading behavior of the users, by providing analytics that show how the users read magazine issues (i.e., how they browse an issue and move inside the app, which portions of an issue are most frequently read and which frequency, and which topics are of interest for the users during a reading session). The tool has been developed by employing a dataset extracted from the reading sessions of a magazine of an important international publisher. In this work we also employ the dataset to present a preliminary study of the user reading behavior.

Keywords: Reading Behavior, Mobile Application, Data Analysis.

1 Introduction

In order to access information, we interact with different types of devices, from personal computers, to mobile phones, to tablets. These interactions take various forms and the usage of mobile applications is certainly the most diffused nowadays. The vast amounts of data implicitly generated by the users during the interactions might lead to useful information on the behavior of the users while using the applications. In [1], the authors highlight that user behavior on mobile applications is analyzed from three main perspectives, i.e., (*i*) data usage, (*ii*) mobility patterns, and (*iii*) application usage. In this paper we will focus on the first and third types of behavior, by analyzing both the usage and the content

* This work is partially funded by Regione Sardegna under project NOMAD (Next generation Open Mobile Apps Development), through PIA - Pacchetti Integrati di Agevolazione "Industria Artigianato e Servizi" (annualità 2013), and by MIUR PRIN 2010-11 under project "Security Horizons".

browsed by the users of a real-world mobile application, named *Viewerplus*[3], which serves as a magazine reader and provides the users with a digital and augmented fruition of content. More specifically, we will analyze how the users browse the issues of a magazine while reading it, and which topics characterize their interest.

The analysis of the reading behavior of the users is an aspect that is gaining more and more interest nowadays. In their survey, Okoli et al. [2] highlighted that less than 1% of the studies focused on the readers of Wikipedia. However, in [3] it was highlighted that reading can be considered as a form of participation, and in their recent study Lehmann et al. [4] stated that the reading activity of the users can provide insights to the human editors. Indeed, by understanding the reading behavior of the users, human editors can tailor the structure of a product such as a magazine, and improve aspects like the content organization or the placing of the ads.

In the recently mentioned study, the authors focus on the user preferences and reading behavior on Wikipedia [4]. According our knowledge, *no study on the reading behavior in mobile content fruition applications has been performed*, and the two application domains present substantial differences. When analyzing the reading behavior in a mobile content fruition application, it should be noted that the browsing of a magazine in form of a pdf file inside an application presents many differences with respect to web browsing. Indeed, the type of browsing we are considering is usually sequential (users usually move from one page to the next), while web pages usually contain links and this is not the case of a magazine issue, which reports the printed version in a digital file. Moreover, the web pages usually form a hierarchy, and this is not the case in our scenario (users employ the mobile application to read pdf files, which is structured as a sequence of pages).

In order to allow the human editors of a magazine to analyze the reading behavior of the users, in this paper we present *Reader Behavior*, a Java tool that analyzes the interactions of the users with the Viewerplus mobile application. The tool presents analytics on how the users browsed a specific issue, which portions have been read more and with which frequency, and which topics are more interesting for the users.

The scientific contributions of this paper are the following:

- we study for the first time in the literature the reading behavior of the users in mobile applications;
- we present a tool that gives the human editors the possibility to dynamically explore the reading of different magazine issues, by selecting them and seeing how users read their content;
- we perform a preliminary study of the users' reading behavior, based on a real-world dataset extracted from the reading sessions of a magazine published by a famous international publisher.

The rest of the paper is organized as follows: we first present related work (Section 2), followed by a description of the Viewerplus mobile application (Sec-

[3] http://www.viewerplus.com/

tion 3); next, we present the tool developed to analyze the reading behavior of the users along with a preliminary analysis of the results obtained considering the reading sessions of a magazine (Section 4); the paper ends with some concluding remarks and by presenting future work (Section 5).

2 Related Work

The reading behavior of the users in Web environments has been studied from several perspectives. In [4], the readers of Wikipedia are analyzed and the authors found out that the most read article are not the most edited ones, and they identified four patterns that describe how the articles are read. Castillo et al. [5] analyzed the life cycle of online news stories and discovered that the number of visits of a news article and the activity on Twitter and Facebook decay after a short time; moreover, the reactions on the social networks can be employed to predict the future visits an article will receive. Zhang and Ma [6] analyzed the correlations between users' educational level and their reading behavior, and found out that higher educated people pay for academic papers, while the other users prefer online literature. In [7, 8], systems to analyze the web reading behavior of the users by employing eye tracking systems were presented.

Regarding the analysis of the user behavior in mobile applications, some studies analyze the motivation behind their use. In [9], Church and Oliveira compare the use of Whatsapp with respect to traditional SMS, and the results show that WhatsApp is usually employed because of the reduced cost, the social interactions it can offer, and its immediacy, while SMS is considered more reliable and privacy preserving. In [10, 11] the factors that lead to user engagement are studied, and those that emerged as the most important are the perceived enjoyment and usefulness of an application.

The patterns in the usage of mobile applications were also studied. Xu et al. [12] found out that 20% of the applications are local (e.g., radio stations), that some applications co-occur in a smartphone (i.e., they can be treated as a bundle), and that diurnal patterns of different genres of applications can be significantly different. In [13], Tossell et al. identified behavioral patterns associated with browsing, native internet applications' use, and physical locations.

The search behavior of the users has been investigated in [14], which discovered that mobile information access is characterized in 94% of the sessions by browsing content, that 8% of the users are involved in search activities, and that these users have a much richer online behavior than the browsing-only counterpart. In [15] it was highlighted that 70% of mobile information access happens in a stationary place (e.g., at home or at work).

The geospatial dynamics of mobile application usage were mostly analyzed with clustering algorithms. In [16], the authors clustered cell locations and performed an analysis of the cells belonging to different clusters, finding that the byte, packet, flow, and user distributions across different geographical regions are significantly different. Keralapura et al. [17] performed a co-clustering of users and websites, discovering that the browsing behavior of most users can be

classified as either homogeneous in terms of interests and characterized by short sessions, or heterogeneous with very long sessions.

As this analysis showed, no study in the literature is devoted at analyzing the reading behavior of the users in mobile applications, and the problem we are studying is novel.

3 Viewerplus: a Mobile Digital Publishing Platform

In this section we set out the mobile application employed in our study and developed by Applix, called *Viewerplus*[4], by providing an overview of its core features, specifically designed to address the needs of users during their reading activities. For the purpose of this work, it is important to note that Viewerplus is not a prototype, but a full-featured application used by thousands people everyday, freely available for Android- and iOS-powered devices, and available in the main digital distribution platforms, such as Apple's App Store, Google Play, Amazon Marketplace, and Samsung Galaxy Apps. Vierwerplus is the leading application for the visualization and digital fruition of magazine periodicals, and it is employed by the main Italian editorial groups.

The application allows users to browse magazine issues in a mobile device, by interacting with a pdf file through several types of interactions and features (e.g., zoom, underline, page saving, bookmark). The interaction is made possible both offline (users can read a magazine issue without being connected to the Internet) and online, thanks to push notifications and the possibility to access to multimedia content. Indeed, the application supports integrations to include photos, audios, videos, links to external pages, and ads. Moreover, users can share excerpts of what they are viewing or reading on the main social media platforms. Thanks to these online features, Viewerplus is also largely employed by companies who want to provide their customers with their latest catalogue.

The monitoring of the users activity inside the application is made possible by a suite developed by Xorovo, named *APP-BI*[5], which tracks the interactions of the users with the application, and extracts analytics that can be employed for different purposes, such as business intelligence.

With this work, we aim to extend the functionalities offered by APP-BI, by introducing the concept of *reading session* and by analyzing in detail the behavior of the users while reading magazine issues (e.g., which portions are read and with which frequency, which pages are read together based on the reading sessions). In conclusion, we would also like to point out that we will focus on the functionalities offered by Viewerplus as a reader. This means that we will consider a scenario where a user can browse a pdf with a magazine issue and no link, multimedia, or online content is available. Indeed, a user can move through the pages of a pdf by reading a magazine and by exploring its content with classic gestures that allow her to interact with the device and the application (i.e., scroll, tap, zoom, swipe, etc.).

[4] http://www.viewerplus.com/
[5] http://www.app-bi.com/

4 Reading Behavior Analysis

Here, we will present *Reader Behavior*, a Java tool developed to analyze and automatically describe how users behave while reading magazine issues. This section starts with a description of the collected dataset (Section 4.1), the data processing performed to extract the reading sessions employed in our analysis (Section 4.2), ending with a presentation of the tool developed to support the human editors at analyzing the reading behavior and with a preliminary presentation of our findings (Section 4.3).

4.1 Data Collection

In order to build the tool and analyze the user behavior, we analyzed the interactions of the users with the application, considering a magazine of a widely-renowned publisher. APP-BI keeps track of different types of events, but not all of them are related to the reading behavior of the users (e.g., the purchase of an issue).

For this study, we collected the data related to the visualization of a page. Such events are tracked if a user visits a page for at least two seconds (this value was studied internally by the APP-BI development team and set as the optimal one). Each record contains the following attributes: <*deviceID, issueID, time, duration, pageID, pageNumber*>, where $deviceID$ is employed to monitor the behavior of a user that employs the same device, $issueID$ identifies the magazine issue, *time* indicates the timestamp in which the event started, *duration* is the number of seconds that the user spent on the page, $pageID$ is an absolute identifier of the page number, and $pageNumber$ indicates the number of the page indicated in the pdf of the magazine issue. Note that having an anonymous $deviceID$, in order to monitor the activities performed inside a device, helps us analyzing the behavior by respecting the privacy of the users: indeed, the identity of a user is not tracked by APP-BI and no personal information is disclosed.

We monitored the interactions of the users with the application between 01/04/2014 and 04/06/2015, recording 10994 events of this type, which involve 110 different magazine issues.

We would like to point out that no metadata was made available, so we had no table of contents that linked the articles to the pages, and no separation between the title of an article and its text. All this information had to be automatically extracted by us in order to analyze the user behavior, and we will describe this process in the following subsection.

4.2 Data Preparation and Processing

In order to have a more structured data representation and to link the collected events to the content of a magazine issue, we performed three steps to divide the events into reading sessions, get the text of each page in a magazine issue, and automatically extract the topics of the magazine.

Reading sessions definition In [18], the authors define a browsing session as all the activities that occur in less than 30 minutes between an activity and the following. This definition was also employed in [4], to define the reading sessions of the users in Wikipedia. In order to characterize the reading behavior of the users, we also adopted this definition, and considered as *reading session* all the events that involve the same user and for which less than 30 minutes passed between the end of an event ($time + duration$) and the beginning of the following.

Text extraction Given the pdf file of a magazine issue, we used Apache's PDFBox[6] to parse it and get as output the text.

Page topics extraction Given the text of each issue, we automatically extracted the topics that characterize the magazine. This was made thanks to Latent Dirichlet Allocation (LDA), which is usually employed for this purpose (i.e., extract the topics from a set of documents), by employing a Java implementation of the algorithm made available in the MALLET framework [19].

The framework received as input a text corpus with the content of all the 110 issues in the dataset and the number of topics to extract, and produced a set of topics. After a set of experiments (not reported to facilitate the reading of the paper), we extracted seven topics. This choice was made since having a lower number of topics led to having keywords that belong to different domains in the same topic, while having a number of topics higher than seven meant that keywords that belong to the same domain were split into two topics.

Out of the seven detected topics, two were characterized by keywords that occur in all the issues (i.e., the details of the publisher, and common keywords that appear in an issue such as "number", "price", and "data"). These two topics were removed, and we manually assigned the following labels to the remaining five, according to the keywords extracted through LDA:

1. family life;
2. tv;
3. lifestyle;
4. health and self-care;
5. cinema.

The choice to extract the topics for the whole magazine was made since a magazine's articles are usually about the same topics (a magazine is usually directed toward a specific user target), to facilitate the manual labelling of the topics given the keywords extracted by LDA, and to be able to compare the reading behavior on different issues (e.g., the interest generated by the "cinema" articles published in an issue with respect to those published in another).

Given these five topics, we processed each page of each issue through MALLET, and extracted a vector whose elements indicate the relevance of each topic for the page.

[6] https://pdfbox.apache.org/

4.3 A Tool to Analyze the Reading Behavior of the Users

Reader Behavior offers three main types of features to human editors:

1. **Co-readings graph.** By selecting a magazine issue, the tool shows a graph that contains a node for each page, and an undirected weighted edge that connects two pages that have been read one after the other. A human editor has the possibility to interact with the graph and visualize only the edges whose weight is above a certain threshold, in order to isolate the most read subgraphs.
2. **Visualization of the interest toward the pages.** Given a magazine issue, the tool shows each page as a box, whose color is based on the frequency with which the page was read. This allows the human editors to analyze how the readings are distributed and which pages caused more interest on the users.
3. **Clustering of the pages that have been read together.** For each magazine issue, we perform a clustering to put together the pages that have been read in the same sessions. Thanks to this feature, the human editors are allowed to re-organize the content of future issue, by having an automatic description of what users have read together.

In the following, we will describe the details of each feature.

Co-readings graph The first feature offers the possibility to human editors to select a magazine issue and visualize a graph that contains a node for each page in the issue, and an undirected weighted edge that connects two pages that have been read one after the other in the same reading session of a user; the weight represents the number of times the two pages have been read one after the other. A screenshot of the feature is shown in Fig. 1.

As it can be seen, a human editor is provided with the possibility to interact with the graph, in order to see only the edges whose weight is above a certain threshold (in the figure, the threshold is 75). This type of dynamic interaction with the graph offers the possibility to analyze in real time how the graph is split into subgraphs and which components are strongly connected (each subgraph represents a subset of pages for which users have shown the same interest).

We are currently working on an automatic description of each subgraph in terms of the topics that characterize its pages. A preliminary analysis shows that pages that have been read together and with a similar frequency are also characterized by the same topics. This is visually indicated by the colors in the nodes, which are homogeneous in each subgraph (i.e, users tend to read together pages related to the same topics).

Visualization of the interest toward the pages With this feature, the tool visualizes all the pages of a magazine issue. For each page we consider the number of times it has been read and split these values based on the quartiles.

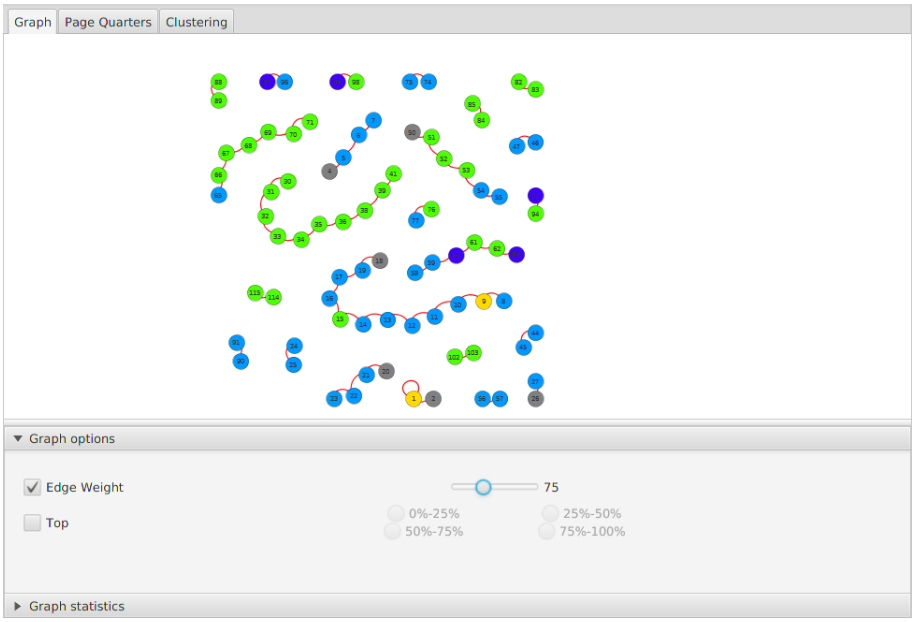

Fig. 1. Co-readings graph feature, showing only the edges whose weight is above 75.

This allows us to obtain four data quarters, which indicate how the interest toward the pages of that issue is distributed.

The tool visualizes the issue in a unique representation, and each page is represented as a box whose color is given by the data quarter associated with the number of times the page has been read. To give a clear differentiation of the data quarters, we chose four vivid colors; the 25% of less read pages (first quarter) is represented with a green color, the 25% of pages under the median (second quarter) has a cyan color, the 25% above the median (third quarter) is given a violet color, and 25% of most read pages (fourth quarter) has a red color.

Fig. 2 shows a representative example of an issue. With representative, we mean that it depicts the usual distribution of the page readings if considering different issues. Indeed, the first part of the issue is usually the most read (red boxes), alternated and followed by the violet boxes that represent the third quarter. The cyan boxes that can be occasionally met even in the first half of the issue represent pages with ads, which have been automatically detected since they are the ones with no text. The less read pages can be found at the end of the issue, represented in green. The fact that the advertising pages do not represent the less considered quarter shows the effectiveness of placing ads in between pages that are of interest for the users.

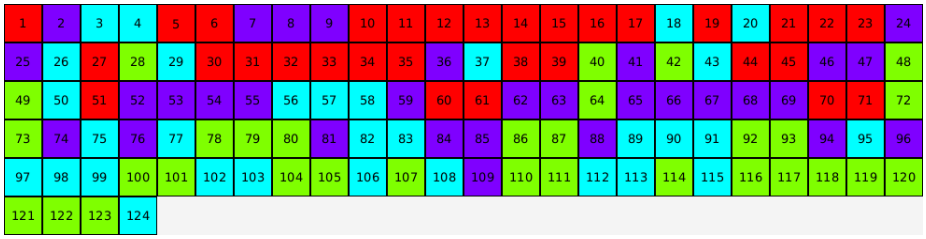

Fig. 2. An example of issue, whose color indicate the frequency with which they have been read.

Clustering of the pages that have been read together The last feature we present is a clustering of the pages in a magazine issue, based on the sessions in which they have been read. Each page is represented by a binary vector, whose elements represent the session IDs and contain 1 if the page was read in the corresponding session, 0 otherwise.

To estimate the number of clusters a priori, the tool employs a technique called *canopy*. This is a fast approximate clustering technique, used to divide the input set of points into overlapping clusters, known as canopies. Although this algorithm may not give accurate and precise clusters, it can detect the optimal number of clusters extremely quickly (i.e., with a single pass over the data). For this reason, the tool runs the algorithm as a pre-processing step to automatically find the optimal number of clusters k, which is given as input to the k-means clustering algorithm, along with the vector representation of the pages, to generate the clusters.

Once the clusters have been detected, the output is given as a set of pages that are in a cluster, plus an automatically-generated description of the cluster in terms of topics, where the relevance of the topic for the cluster is indicated. Let $relevance_{t,p}$ indicate the relevance of a topic t for a page p, and $distance_{p,c}$ indicate the distance of a page p from the centroid of the cluster c in which the page is. The relevance of a topic t for a cluster c is built as follows:

$$relevance_{t,c} = \sum_{p \in c} \frac{relevance_{t,p}}{distance_{p,c}}$$

Thanks to this formula, the highest is the difference in the reading behavior of a page p with respect to the the others in the cluster c, the lower is the weight assigned to the topic t for that cluster (the value $distance_{p,c}$ is seen as an indication of "cohesion" between the page and the rest of the cluster).

In order to give a relative value to the relevance of a topic for a cluster and give the human editor the perception of the reading behavior in a cluster, we normalize the relevance of each topic with a value between 0 and 1, as follows:

$$relevance_{t,c} = \frac{relevance_{t,c} - min(relevance_c)}{max(relevance_c) - min(relevance_c)}$$

where $min(relevance_c)$ and $max(relevance_c)$ respectively indicate the minimum and maximum relevance values obtained by a topic in a cluster c. Trivially, 1 is the score assigned to the most relevant topic, and 0 is the score assigned to the least relevant topic.

Due to space constraints, we will not show a screenshot of this feature, but we will provide an example of the description of an issue, whose pages can be split into three clusters based on the readings sessions. The automatic descriptions generated by the tool are the following:

1. *lifestyle* (1.00), *family life* (0.99), *health and self-care* (0.31), *cinema* (0.00), *tv* (0.00)
2. *family life* (1.00), *lifestyle* (0.56), *health and self-care* (0.24), *cinema* (0.05), *tv* (0.00)
3. *lifestyle* (1.00), *family life* (0.84), *health and self-care* (0.32), *cinema* (0.03), *tv* (0.00)

Apart from the content in terms of pages of these three clusters (which would not be interesting for this purpose of this paper), we can see that based on the reading sessions, the interest of the users varies significantly from cluster to cluster. Indeed, in the first cluster, the pages that are characterized by *lifestyle* and *family life* have been read with similar interest in the same sessions, while *health and self-care* generated a much lower interest, and the users showed the lowest interest for cinema- or tv-related topics. The second cluster of pages is instead much more centered toward pages related to *family life*, while *lifestyle* has been half as relevant in those reading sessions; *health and self-care* is a topic that generates little interest, and *cinema* and *tv* still represent the topics that generate less interest in the users. Finally, the third cluster shows the same ranking as the first one, but with different weights.

This feature gives insights to the human editor on both the ways in which the users read a magazine issue (pages are grouped based on the reading sessions) and on the topics that characterize these sessions with their associated relevance.

It should be noted that the combined use of the features provided by our tool can be very helpful to the human editors. For example, if given these clusters a human editor wanted to increase the relevance of *cinema* and *tv* pages, knowing that they are usually placed at the end of an issue (i.e., the less read portion), the articles related to these topics could be moved into a section of the magazine that appears earlier in future issues.

5 Conclusions and Future Work

In this paper we presented *Reader Behavior*, a Java tool developed to analyze the reading behavior of the users, based on their interactions with a mobile digital publishing platform, named *Viewerplus*.

Our proposal takes the data collected during the browsing of a magazine's issues, extracts the reading sessions of the users, and provides visual and descriptive features of how the users read a given magazine issue. The objective is

to provide the human editors with tools that allow them to get to know their customers more and improve the service they provide to them.

At the moment, the tool provides three features that describe the reading behavior of the users from different perspectives. Future work will extend the tool with additional features, like the automatic description of the subgraphs in the "co-readings graph", or the possibility to click on a box with the interest toward a page, in order to show which topics characterize that page. Moreover, we will employ more real-world datasets and try to develop metrics to describe the reading behavior not only in terms of single magazine issues, but by giving a global view on how a magazine is read.

Acknowledgments

The authors would like to thank Gianluca Zuddas, Andrea Aresu, Giacomo Piseddu, Paolo Tanzi, Davide Melis, Corrado Alvau, and Francesco Argiolas, for their contribution in this research work.

References

1. Yang, J., Qiao, Y., Zhang, X., He, H., Liu, F., Cheng, G.: Characterizing user behavior in mobile internet. IEEE Trans. Emerging Topics Comput. **3** (2015) 95–106
2. Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F.A., Lanamäki, A.: The people's encyclopedia under the gaze of the sages: A systematic review of scholarly research on Wikipedia (2012)
3. Antin, J., Cheshire, C.: Readers are not free-riders: Reading as a form of participation on wikipedia. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. CSCW '10, New York, NY, USA, ACM (2010) 127–130
4. Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., Kaltenbrunner, A.: Reader preferences and behavior on wikipedia. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media. HT '14, New York, NY, USA, ACM (2014) 88–97
5. Castillo, C., El-Haddad, M., Pfeffer, J., Stempeck, M.: Characterizing the life cycle of online news stories using social media reactions. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. CSCW '14, New York, NY, USA, ACM (2014) 211–223
6. Zhang, L., Ma, W.: Correlation analysis between users' educational level and mobile reading behavior. Library Hi Tech **29** (2011) 424–435
7. Beymer, D., Russell, D.M.: Webgazeanalyzer: A system for capturing and analyzing web reading behavior using eye gaze. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. CHI EA '05, New York, NY, USA, ACM (2005) 1913–1916
8. Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in www search. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '04, New York, NY, USA, ACM (2004) 478–479

9. Church, K., de Oliveira, R.: What's up with whatsapp?: Comparing mobile instant messaging behaviors with traditional sms. In: Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services. MobileHCI '13, New York, NY, USA, ACM (2013) 352–361
10. Kim, Y.H., Kim, D.J., Wachter, K.: A study of mobile user engagement (moen): Engagement motivations, perceived value, satisfaction, and continued engagement intention. Decis. Support Syst. **56** (2013) 361–370
11. Verkasalo, H., López-Nicolás, C., Molina-Castillo, F.J., Bouwman, H.: Analysis of users and non-users of smartphone applications. Telemat. Inf. **27** (2010) 242–255
12. Xu, Q., Erman, J., Gerber, A., Mao, Z., Pang, J., Venkataraman, S.: Identifying diverse usage behaviors of smartphone apps. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference. IMC '11, New York, NY, USA, ACM (2011) 329–344
13. Tossell, C., Kortum, P., Rahmati, A., Shepard, C., Zhong, L.: Characterizing web use on smartphones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '12, New York, NY, USA, ACM (2012) 2769–2778
14. Church, K., Smyth, B., Cotter, P., Bradley, K.: Mobile information access: A study of emerging search behavior on the mobile internet. ACM Trans. Web **1** (2007)
15. Church, K., Oliver, N.: Understanding mobile web and mobile search use in today's dynamic mobile landscape. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. MobileHCI '11, New York, NY, USA, ACM (2011) 67–76
16. Shafiq, M.Z., Ji, L., Liu, A.X., Pang, J., Wang, J.: Characterizing geospatial dynamics of application usage in a 3g cellular data network. In Greenberg, A.G., Sohraby, K., eds.: Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012, IEEE (2012) 1341–1349
17. Keralapura, R., Nucci, A., Zhang, Z.L., Gao, L.: Profiling users in a 3g network using hourglass co-clustering. In: Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking. MobiCom '10, New York, NY, USA, ACM (2010) 341–352
18. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the world-wide web. In: Proceedings of the Third International World-Wide Web Conference on Technology, Tools and Applications, New York, NY, USA, Elsevier North-Holland, Inc. (1995) 1065–1073
19. McCallum, A.K.: Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu (2002)