meCurate: Personalized Curation Service using a Tiny Text Intelligence

Woo-Jong Ryu, Jung-Hyun Lee, Kang-Min Kim, SangKeun Lee
Department of Computer Science and Engineering
Korea University, Seoul, Republic of Korea
{skdirwj, jhbslpd, kangmin89, yalphy}@korea.ac.kr

ABSTRACT

The necessity and importance of intelligent content curation have been demanding, as exemplified by Google Photos and Inbox. In this demo, we demonstrate a personalized curation service on a smartphone, called meCurate. By understanding implicit user interests from user smartphone usage data, meCurate intelligently organizes both in-device user content (e.g., SMSs and bookmarks) and external content (e.g., news articles crawled from The New York Times website), which are likely to match the inferred user's interests. In addition, meCurate retrieves semantically relevant in-device user content and external content to an explicit text or voice query. To this end, we utilize a tiny text intelligence which is suitable for smartphones with limited resources. Notably, meCurate works in a stand-alone, privacy-protecting manner without sending out any in-device personal data or content, resulting in a unique user experience.

Keywords

Personalized Curation; User Understanding; Semantic Search; Tiny Text Intelligence

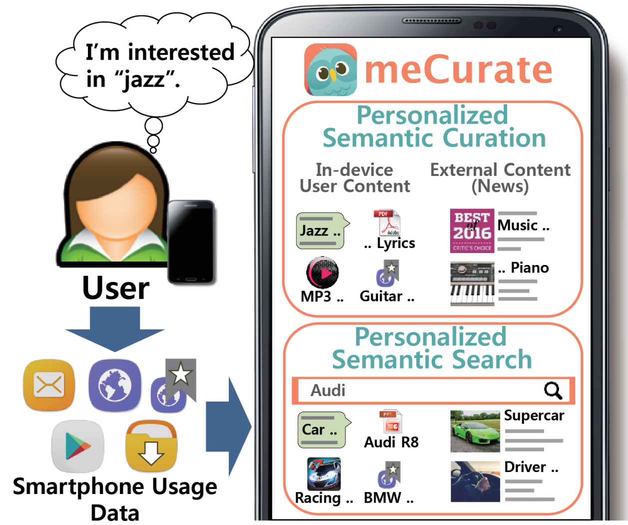

Figure 1: Overview of meCurate

1. INTRODUCTION

As smartphone usage has been increasing, the number of content on a smartphone is rapidly growing, including photos, emails, etc. In recent years, different intelligent curation services on smartphones have been on the market such as Google Photos [4] and Inbox [3]. Google Photos automatically organizes users' photos and helps users to search them with text queries. Google Inbox organizes users' emails by grouping their similar emails together e.g., promotions, purchases, trips, etc. In such services, however, photos or emails should be analyzed on the Google server. Thus, it may lead to privacy breaches although some techniques may be used to protect privacy.

In parallel, the necessity of searching user content stored in a smartphone (i.e., *in-device user content*) has also been demanding. To make the in-device search task easier, popular in-device search services on a smartphone have been introduced, e.g., Google In Apps [2] and Apple Spotlight [1]. These in-device search services enable a user to search in-device user content or content inside apps. Currently, however, they are limited to keyword-based matching. This leads to quite limited search results since they do not consider the semantics of terms.

In this demo, we showcase a novel service, called meCurate[1], a personalized content curation service on a smartphone. Figure 1 shows an overview of meCurate. meCurate first infers *implicit* user interests by semantically classifying smartphone usage data into the corresponding categories in the Open Directory Project (ODP) [5]. It then organizes and presents both in-device user content and external content (e.g., news articles and mobile apps), based on the inferred user interests. The inference of implicit user interests and personalized semantic curation tasks are performed in the form of a background process without any explicit user requests. In addition, a user can perform the personalized, semantic in-device search by making explicit text or voice queries.

Providing intelligent services without privacy breaches is technically challenging. To this end, we utilize a tiny text

[1]The screencast is available at `https://youtu.be/4bGc4jHXmQA`.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054723

intelligence which is suitable for smartphones with limited resources. The tiny text intelligence plays a central role in user understanding, semantic content analysis, and semantic matching. Since the tiny text intelligence resides on a smartphone, it enables meCurate to work in a stand-alone manner without privacy concerns. To the best of our knowledge, this is one of few work that intelligently provides personalized curation service without sending out in-device personal data or content, thereby making a one-of-a-kind user experience.

The contributions of our demo are summarized as follows:

- We design and implement meCurate, a personalized, semantic curation service that is embedded into a smartphone. meCurate proactively organizes and presents in-device content without explicit requests. In addition, meCurate supports a personalized, semantic in-device search with explicit text or voice queries.

- We demonstrate the feasibility of intelligence service without privacy breaches, thereby seeking to break down the social convention that we should pay privacy for intelligence and convenience.

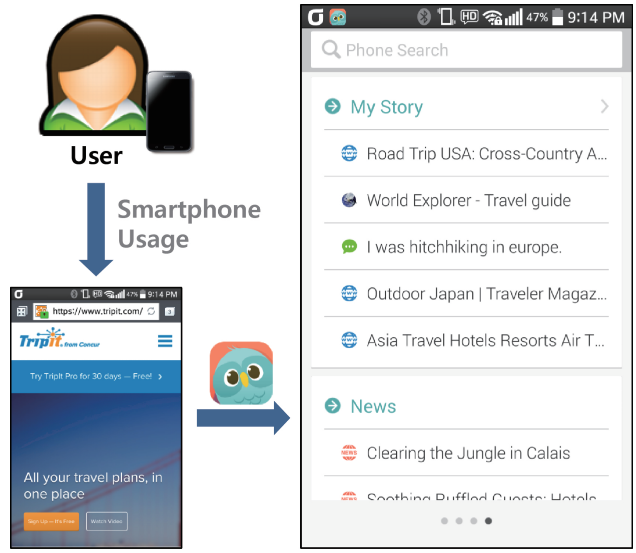

Figure 2: Illustration of personalized semantic curation

2. meCurate

In this section, we propose two main functionalities that meCurate provides, including the personalized semantic curation and the personalized semantic search. With the two functionalities, meCurate serves four types of in-device user content, i.e., SMSs, bookmarks, apps, files, and two types of external content, i.e., news articles and mobile apps that are periodically crawled from The New York Times[2] and Google Play[3] site, respectively. We aggregate both in-device user content and external content alike to form the *in-device content* in the context of meCurate.

2.1 Personalized Semantic Curation

User understanding is fundamental to personalized curation services. We notably observe that in-device usage data is extremely important with respect to user understanding on a smartphone. In daily lives, users use their smartphones in many different ways (e.g., sending or receiving SMSs, visiting webpages, adding bookmarks, installing apps, downloading files, etc.). Smartphone usage data generated by such usage implicitly represents user interests. Thus, we analyze the recent in-device user content to infer the short-term user interests. In particular, we infer each user interest from different usage data since user interests might be different per usage data. In this demo, we infer user interests from five types of in-device user content, including SMSs, Web history, bookmarks, apps, and files.

Using the ODP-based semantic classifier, the personalized semantic curation classifies each type of smartphone usage data into relevant ODP categories. We model such categories as implicit user interests. Then, the personalized semantic curation lists semantically relevant in-device content to the inferred implicit user interests. In the current implementation, a well-organized result is presented in the form of a single card per inferred interest. In each organized result, in-device user content is grouped together as **My Story**, while external content is grouped as **News** and **Apps**. All of the above tasks are performed in background without any explicit requests.

Figure 2 illustrates the personalized semantic curation. For example, suppose that a user recently visited a webpage 'Tripit' (www.tripit.com). From the Web history, the personalized semantic curation infers an implicit user interest as *Travel*. It then organizes and presents *Travel* related in-device content, including three bookmarks, one app, and one SMS, in addition to *Travel* related news articles and mobile apps. In the current implementation, one card corresponds to one user interest like *Travel*.

By utilizing smartphone usage data, meCurate provides the personalized, semantically organized in-device content without explicit requests. In contrast with existing intelligent curation services such as Google Photos and Inbox, meCurate works in a stand-alone, privacy-protecting manner.

2.2 Personalized Semantic Search

Personalized semantic search in meCurate departs from existing in-device search services, such as Google In Apps and Apple Spotlight, in that it provides semantic search while others are limited to keyword matching. In general, queries explicitly represent user interests. We therefore use a given query to infer an explicit user interest. Given a query, the personalized semantic search classifies the query into relevant categories by using an ODP-based classifier. The classified result is modeled as an explicit user interest. The query may be issued in text on a smartphone or in voice via a connected smartwatch.

When a user makes a voice query, a simple meCurate client on the smartwatch converts the voice to the corresponding text and then delivers it to the user's smartphone. The personalized semantic search on the smartphone subsequently obtains a list of ranked in-device content, and presents them in a similar fashion to the personalized semantic curation.

Figure 3 illustrates the personalized semantic search. For example, when a user makes a voice query 'audi' to a connected smartwatch (or as a text query on a smartphone), the personalized semantic search infers an explicit user in-

[2] http://www.nytimes.com
[3] https://play.google.com/store

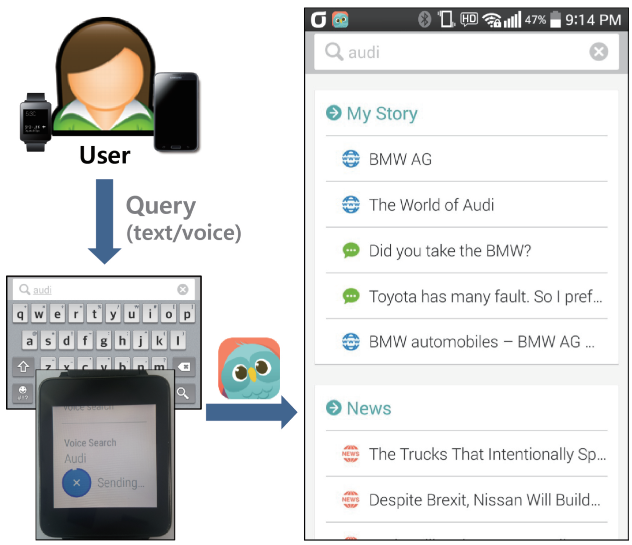

Figure 3: Illustration of personalized semantic search

terest as *Vehicles*. It then searches and presents *Vehicles* related in-device content, including three bookmarks, and two SMSs, in addition to related news articles and mobile apps. This functionality enables a user to perform semantic search for in-device content on a smartphone.

3. ARCHITECTURE

meCurate is based on the ODP-based semantic approach [8]. The semantic approach involves two tasks. First, it classifies a given text into categories (or topics) on the ODP taxonomy. Second, it measures the semantic similarity between two texts based on the topical relevance. However, it is not trivial to use the semantic approach on smartphones with limited resources. We therefore utilize the tiny text intelligence developed in our previous work [6][7]. We design and implement an architecture of meCurate using the tiny text intelligence, which is shown in Figure 4.

3.1 Tiny Text Intelligence

The tiny text intelligence plays an essential role in user understanding, semantic content analysis, and semantic matching on a smartphone. It is built on the well-organized ODP categories. The knowledge base of meCurate consists of 892 categories. On top of the embedded ODP knowledge base, meCurate contains the semantic classifier and the semantic ranker which support semantic classification and similarity model, respectively. Since the tiny text intelligence resides on a smartphone, meCurate does not send out any in-device user content. Instead, it periodically crawls news articles and mobile apps from external websites and builds semantic index on them inside smartphones. Thus, meCurate works in a stand-alone manner without resort to external servers. The tiny text intelligence has been utilized for different types of intelligent services, including personalized recommendation [6] and intelligent album [12].

3.1.1 *Semantic Classifier*

We borrow the *merge-centroid classifier* [8] to infer user interests and analyze in-device content. The classified results of in-device content are utilized to build the semantic index. Similarly, news articles and mobile apps are also semantically indexed whenever they are periodically crawled from external sites and classified.

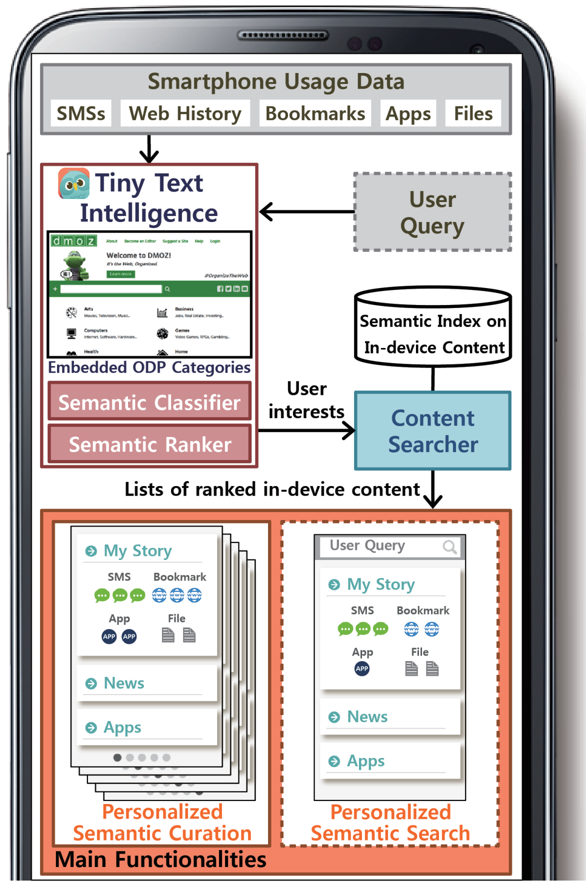

Figure 4: Architecture of meCurate

3.1.2 *Semantic Ranker*

The semantic ranker measures the semantic similarity between inferred user interests and in-device content. We borrow the *GraphScore* metric [8] for the semantic ranking task. It is based on a link analysis technique on the similarity graph derived from the ODP-based topical taxonomy.

3.2 Content Searcher

The content searcher retrieves only semantically relevant in-device content to the inferred user interests by using the semantic index on content.

4. IMPLEMENTATION

We implemented the current meCurate on a LG G2 smartphone with Android OS version 4.4.2 Kitkat. Also, we connected it to a smartwatch, LG G Watch with Android wear OS version 1.5. Using SQLite, we embedded the tiny text intelligence, of which size is 38.9 Mbytes, into the smartphone.

We also developed a simple meCurate client software on top of the LG G Watch to support voice queries. The simple meCurate client converts a given voice query to the corresponding text by utilizing Android Speech API on the smartwatch. It subsequently delivers the text to meCurate app on the connected smartphone.

We analyzed in-device user content within the smartphone from media storage in Android OS. Considering the limited resources, we analyzed SMS messages, merely the title of Web history and bookmarks, and only the names of apps and files to infer user interests, which are usually short texts. We measured the energy consumption for the inference using a Monsoon power meter. We observed that it takes 22.14 μAh, which is equivalent to 0.26 seconds of music play (82.99 μAh/sec) on the same smartphone, when a total of fifty in-device user content (8.08 terms per user content, on average) is accessed. We also observed that the average memory usage for the inference is 16.98 Mbytes.

We carefully designed meCurate such that it infers implicit user interests only when the screen is off. The news articles and mobile apps are periodically crawled from external sites while the smartphone is getting recharged. In building the semantic index on news articles and mobile apps, we analyzed the headline, the lead paragraph, and the section name for each news article, while the app name, the description, and the category name for each mobile app, all of which are usually short texts.

5. DEMONSTRATION

We will showcase the following scenarios. Conference participants will interact with a smartphone with meCurate, which is connected to a smartwatch with a simple meCurate client. A user uses the smartphone as usual, e.g., sending or receiving SMSs, visiting webpages, adding bookmarks, installing apps, or downloading files. Once a user runs the meCurate app, the well-organized results based on the inferred implicit user interests will be presented. In particular, the personalized semantic curation shows the well-organized results in the form of a single card per interest. Alternatively, a user searches in-device content by making text or voice queries. The personalized semantic search will return semantically relevant in-device content to the query. We believe that conference participants will have unique user experiences, since meCurate works in a stand-alone manner without any privacy concerns.

6. DISCUSSION

The concept of portable personality was introduced in [11], which handles the acquisition and management of personalized profiles via mobile devices in ubiquitous computing. An overview of representing, acquiring, managing, and merging personalized profiles was presented in [13]. The portable personality was applied to ambient media [9] and social network services [10]. Note that the concept of portable personality and its related techniques are complementary to our work. Thus, they would help meCurate in improving especially the query interface, as well as the user interests feedback interface between semantic classifier/ranker and content searcher.

7. CONCLUSION

In this demo, we showcase a novel service, called meCurate, an embedded, personalized, semantic curation service. meCurate semantically infers *implicit* user interests from smartphone usage data. It then intelligently organizes in-device user content, and external news articles and mobile apps according to the inferred user interests. In addition, the personalized semantic search retrieves the in-device content semantically relevant to *explicit* user text or voice queries. Notably, meCurate works in a stand-alone, privacy-protecting manner by utilizing a tiny text intelligence. To the best of our knowledge, this is one of few work to provide intelligent services without privacy breaches, thereby breaking down the social convention that we should pay privacy for intelligence and convenience.

8. ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (number 2015R1A2A1A10052665).

9. REFERENCES

[1] Apple Spotlight. `https://en.wikipedia.org/wiki/Spotlight_(software)`.

[2] Google In Apps. `https://www.google.com/search/about/learn-more/in-apps`.

[3] Google Inbox. `https://www.google.com/inbox`.

[4] Google Photos. `http://www.google.com/photos`.

[5] Open Directory Project. `http://www.dmoz.org`.

[6] J. Ha, J.-H. Lee, and S. Lee. EPE: An embedded personalization engine for mobile users. *IEEE Internet Computing*, 18(1):30–37, Jan./Feb. 2014.

[7] J. Ha, J.-H. Lee, K.-S. Shim, and S. Lee. EUI: An embedded engine for understanding user intents from mobile devices. In *Proccedings of ACM International Conference on Information and Knowledge Management*, pages 1935–1936, 2010.

[8] J.-H. Lee, J. Ha, J.-Y. Jung, and S. Lee. Semantic contextual advertising based on the open directory project. *ACM Transactions on the Web*, 7(4):24:1–24:22, Oct. 2013.

[9] A. Lugmayr, S. Reymann, S. Kemper, T. Dorsch, and P. Roman. Bits of personality everywhere: Implicit user-generated content in the age of ambient media. In *Proceedings of International Symposium on Parallel and Distributed Processing with Applications*, pages 516–521, 2008.

[10] S. Reymann, D. S. Alves, and A. Lugmayr. Personalized social networking: an applied scenario in a portable personality environment. In *Proceedings of International Conference on Entertainment and Media in the Ubiquitous Era*, pages 172–176, 2008.

[11] S. Reymann, V. Bruns, and A. Lugmayr. P2-portable personality a middleware solution for smart user profile management and distribution. *Interactive TV: A Shared Experience, TISCP Adjunct Proceedings of EuroITV*, 35, 2007.

[12] H. Shin, H. Oh, W.-J. Ryu, and S. Lee. sigAlbum: An embedded photo service using a tiny text intelligence. In *Proccedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 373–376, 2016.

[13] S. Uhlmann and A. Lugmayr. Personalization algorithms for portable personality. In *Proceedings of International Conference on Entertainment and Media in the Ubiquitous Era*, pages 117–121, 2008.