December 2019

EXTRACTING AND DISPLAYING INFORMATION FROM SCREENSHOTS

Matthew Sharifi

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

EXTRACTING AND DISPLAYING INFORMATION FROM SCREENSHOTS

ABSTRACT

Users often capture and share app screenshots over communication channels such as instant messaging and email. Although a convenient mechanism for the sender, sharing screenshots can be inconvenient for recipients. This is because screenshot information is typically not readily visible in its low-resolution or small version, so that the recipient needs to tap the screen to expand it. Even when expanded, the information in the screenshot is not actionable, e.g., it is not in text format, nor can it be easily copy-pasted. This disclosure describes techniques that automatically extract and display information from screenshots such that relevant information is available to and actionable by a user that receives screenshots.

KEYWORDS

- Instant messaging
- Screenshot
- Optical character recognition (OCR)
- Entity annotation model
- Topicality scoring

BACKGROUND

Users often capture and share app screenshots over communication channels such as instant messaging and email. This is a convenient way to share information, e.g., flight details, restaurant recommendations, etc. with other persons or groups of people.

Although a convenient mechanism for the sender, sharing screenshots can be inconvenient for recipients. This is because screenshot information is typically not readily visible in its low-resolution or small version, so that the recipient needs to tap the screen to expand it.

Even when expanded, the information in the screenshot is not actionable, e.g., it is not in text format, nor can it be easily copy-pasted, e.g., into a search box. The recipient still needs to scan the content to determine information of relevance.
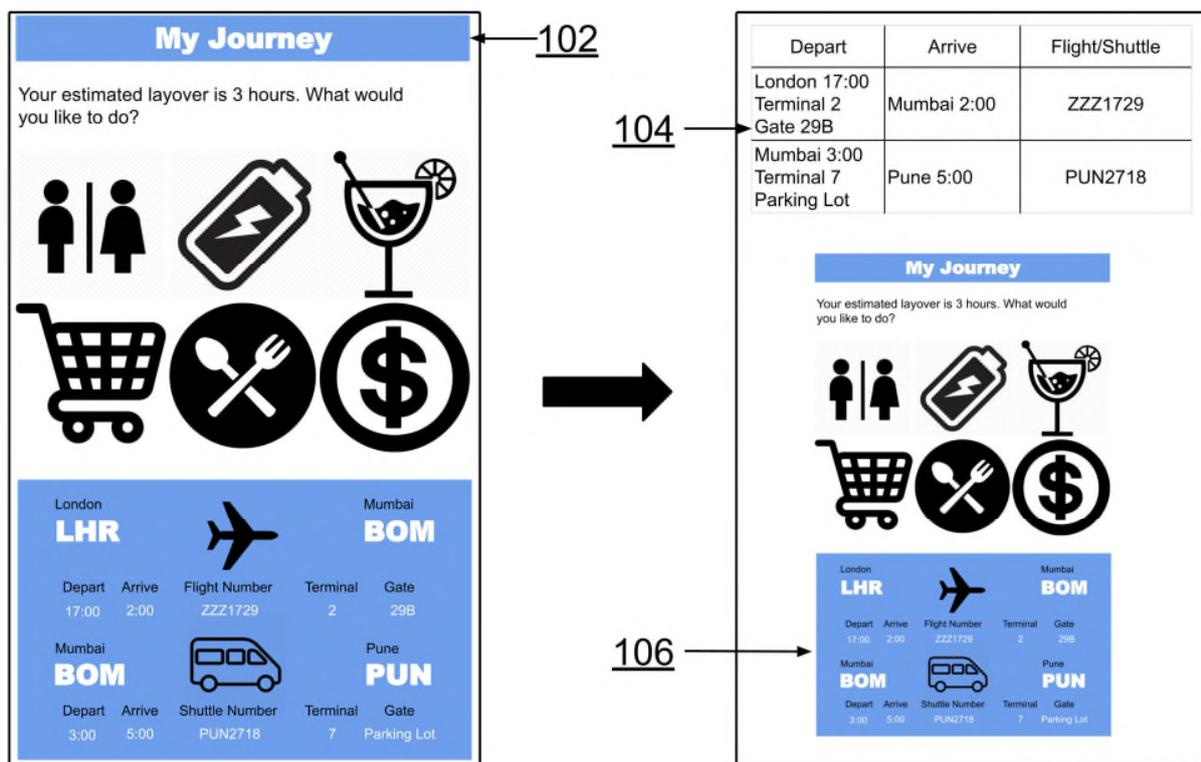
DESCRIPTION

Fig. 1: Extracting and displaying information from screenshots

Fig. 1 illustrates extracting and displaying information from screenshots, per techniques of this disclosure. A user obtains a screenshot (102) to share via instant messaging or email. Textual information (104) is extracted from the screenshot. The user is given the option to send the extracted textual information (when extraction is performed on the sender side), the screenshot image (106), or both. The presence of the extracted textual information in the received message enables the recipient to quickly interpret and use the information within the

screenshot. The presence of text along with the image in the sent message provides the additional benefit of improved accessibility.
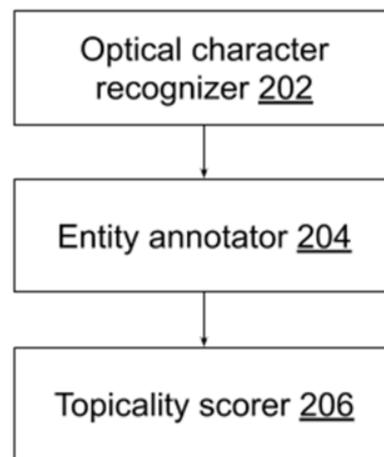

Fig. 2: Analyses performed to extract information from screenshots

As illustrated in Fig. 2, one or more of the following analyses are performed to extract information from screenshot images.

- An optical character recognition engine (OCR, 202) is used to detect text regions and convert the text regions to characters. The OCR engine can also extract style information, e.g., font sizes, types, etc. from the text. Other recognition engines can be executed to infer additional factors, e.g. logos, layouts, objects in images, etc.

- Text derived from OCR is run through an entity annotation model (204), which takes unstructured text as input and extracts entities such as addresses, flight codes, place names, etc. Output from the entity annotation model is used to extract composite entities such as flight reservation details, e.g., flight codes, dates, times, etc.

- Each extracted entity is scored to determine a topicality score (206), e.g., a score that indicates the centrality of the entity to the screenshot. This topicality score can use co-occurring entities, visual factors, style information, e.g., whether the entity is part of a

header or not, etc. Further, in some cases, the sending user may have added visual markup to the screenshot. In such cases, the markup contributes to the topicality score, e.g., the topicality of entities in the screenshot is biased towards marked-up regions.

In addition to the above, if the users participating in the messaging or email conversation permit, the preceding conversation in these applications can be analyzed to contextually guide information extraction from the screenshot. For example, if user A asks user B "when does your flight land?" and user B responds with a screenshot, the extracted subtitle includes the flight landing time that is extracted from the screenshot.

The analysis of screenshots can be performed on the recipient's device, on the sender's device, on a central server (if the message is sent via a central server), or any combination of device and server. Upon extraction of information from the screenshot, relevant details can be shown alongside the shared screenshot. If no information can be derived, only the screenshot itself is displayed, which is equivalent to current behavior.

In this manner, the techniques of this disclosure enable recipients of a message to determine and interpret relevant information being shared via screenshot images, and makes such information actionable, e.g., via a search box, copy-paste, or other actions.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can

be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that automatically extract and display information from screenshots such that relevant information is available to and actionable by a user that receives screenshots.

REFERENCES

[1] Google LLC, "Vision AI," https://cloud.google.com/vision/ accessed Nov. 24, 2019.

[2] Ryan Whitwam, "Google has added limited OCR features to Google Now on tap,"

https://www.androidpolice.com/2016/03/07/google-has-added-limited-ocr-features-to-google-now-on-tap/ accessed Nov. 24, 2019.