Implicit Visual Attention Feedback System for Wikipedia Users

Neeru Dubey
neerudubey@iitrpr.ac.in
Indian Institute of Technology Ropar
Rupnagar, Punjab, India

Amit Arjun Verma
2016csz0003@iitrpr.ac.in
Indian Institute of Technology Ropar
Rupnagar, Punjab, India

S.R.S. Iyengar
sudarshan@iitrpr.ac.in
Indian Institute of Technology Ropar
Rupnagar, Punjab, India

Simran Setia
2017csz0001@iitrpr.ac.in
Indian Institute of Technology Ropar
Rupnagar, Punjab, India

ABSTRACT

The complex collaborative structure of Wikipedia has attracted researchers from various domains, such as social networks, human-computer interaction, and collective intelligence. Yet, a few focus on the readers' perception of Wikipedia. Readers make up the majority of Wikipedia users (editors/readers), and being on the consumption side, readers play a crucial role in its sustenance. The attention patterns of users while reading an article can reveal users' interest distribution as well as content quality of the article. In this paper, we present an Attention Feedback (AF) approach for Wikipedia readers. The fundamental idea of the proposed approach comprises the implicit capture of gaze-based feedback of Wikipedia readers using a commodity gaze tracker. The developed AF mechanism aims at overcoming the main limitation of the currently used "pageview" and "survey" based feedback approaches, i.e., data inaccuracy. Moreover, the incorporation of a single-camera image processing-based gaze tracker makes the overall system cost-efficient and portable. The proposed approach can be extended to enable the research community to analyze various online portals as well as offline documents from the readers' perspective.

CCS CONCEPTS

• **Human-centered computing → Wikis**; **Collaborative and social computing**; **Collaborative and social computing systems and tools**;

KEYWORDS

Wikipedia, gaze tracking, Collaborative analysis tool, gaze dataset, readers

ACM Reference Format:
Neeru Dubey, Amit Arjun Verma, S.R.S. Iyengar, and Simran Setia. 2021. Implicit Visual Attention Feedback System for Wikipedia Users . In *17th International Symposium on Open Collaboration (OpenSym 2021), September 15–17, 2021, Online, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3479986.3479993

OpenSym 2021, September 15–17, 2021, Online, Spain

ACM ISBN 978-1-4503-8500-8/21/09…$15.00
https://doi.org/10.1145/3479986.3479993

1 INTRODUCTION

Wikipedia is a collaboratively developed online encyclopedia. It has been consistently ranked in the top fifteen most-visited websites as per Alexa ranking[1]. Based on the actions performed, Wikipedia users can be broadly divided into producers (editors, moderators) and consumers (readers). Over the last years, research has extensively investigated the group of producers while another group of Wikipedia users, the readers, their preferences, and their behavior have not been much studied.

Online participation research has often characterized Internet readers as non-contributors, who benefit from others' contributions but who contribute little themselves [31]. In general, non-contributing readers constitute 90% or more of any given discussion forum, online community, or social website [25]; a 2011 Readership Survey showed that the number for Wikipedia was 94%. Although often called a "participation inequality", such statistics indicate that reading constitutes the norm, whereas contribution is the anomaly (albeit a necessary one). A few pieces of research highlight the importance of readers in the Wikipedia community. The study by Antin et al. [6] claims that reading can be seen as a form of participation and is, therefore, valuable. Reading a Wikipedia article can be considered legitimate peripheral involvement through which individuals gain knowledge and move towards more active participation. The fact that a user reads an article and not edit, is a good indication of its quality, such as its reliability [3]. Lehmann et al. [23] also emphasize the importance of reading behavior analyses to characterize users' reading preferences for portal content development.

The primary reason for the lack of research focused on readers is the limited availability of suitable techniques to capture readers' attention feedback. The current popular techniques are pageviews and Surveys. Pageviews depict the total number of visits to an article during a given period, but it has several disadvantages. Firstly, the statistics do not quantify the number of unique readers, and secondly, it does not consider the time duration a reader has spent reading the article. Another widely used method is surveying the readers. Nevertheless, it also suffers from drawbacks in terms of survey process scalability and accuracy of information provided by participants. The Wikimedia Foundation has also emphasized that pageviews and surveys are not efficient measures to identify readers' perception of an article [49]. One more technique was presented by Barifah et al. [8]. They used log files as a means to

[1] https://www.alexa.com/topsites

uncover information about users and their behavior when searching for information. This study excluded fine-grained analysis due to lack of details of users' navigation patterns.

This paper proposes a novel framework to capture the Attention Feedback (AF) of Wikipedia readers. The proposed method records the implicit gaze pattern of a reader and extracts the visual feedback features. In past research, it has been well established that our eye movement is closely related to human cognition [4]. There is evidence in reading psychology literature that eye movement patterns during reading are indeed related to the textual features of the document [41]. Ajanki et al. [5] claim that when a system does not have any prior information regarding what the user wants to search, the eye movements help significantly in the search. It is the case in a proactive search, for instance, where the system monitors the reading behavior of the reader in a new document. Xu et al. [55] talk about the relevance between the human text reading pattern and their current cognitive process. The relation between eye gaze pattern and human interest has been vastly utilized in various fields [20].

The previous studies performed to capture users' gaze for online portals utilize dedicated eye trackers like Tobii [33]. The dedicated eye trackers are pretty expensive (hundreds of dollars) and mostly invasive in nature[2]. Additionally, these trackers are rarely available at users' sites. We propose to use a lightweight desk-mounted/laptop camera along with a publicly available image processing based eye tracker to capture readers' attention patterns. We embed the eye tracker and the analysis methods in a web application called "WikiRead". Usage of commodity eye-tracking solution and web application makes it user-friendly and portable. It also enables the collection of a large dataset of Wikipedia readers' AF. This dataset can open doors for the analysis of Wikipedia from a novel perspective.

The rest of the paper is structured as follows: Section 2 describes the necessary background and related works to the proposed method. Section 3 provides an overview of the proposed AF framework. In Section 4, we describe the procedure to perform attention pattern analysis. In Section 5, we discuss the system setup and the experiments performed to evaluate the proposed technique from various perspectives. In Section 6, we provide a discussion on the limitations and possible future directions. Section 7 concludes the work.

2 BACKGROUND & RELATED WORK

Eye gaze tracking is the process of estimating the direction of the sight line and the point of regard. There have been several advances in eye-tracking technology. Based on physical contact requirement with the eye tracker, the research in this field can be classified into two types; invasive eye trackers and non-invasive eye trackers. Invasive eye trackers require physical contact with the users. The most prevalent invasive trackers are Tobii, EyeLink, SMI2, Mirametrix4, and EyeTech3. They provide higher accuracy as compared to non-invasive eye trackers, but they are pretty expensive. The non-invasive eye trackers mainly work using computer vision techniques and some lightweight camera(s). Therefore, they do not require any physical contact with the user. Some recent researches include Searchgazer, xLabs Gaze Tracking, GazePointer, Ogama, OpenEyes, PyGaze, OpenGazer, TurkerGaze. The details of these non-invasive eye trackers can be referred to in [56]. We use GazeFlowAPI[3] as a gaze tracking solution due to it's efficiency and suitability for the current task.

Our eyes show a characteristic behavior composed of a series of fixations and saccades [13]. Fixation is a duration for which the eye is steadily gazing at one point or a collection of proximal points. On average, it is of a time interval of 200-250 ms. A saccade is a rapid eye movement from one fixation to the next one. The time-lined distribution of saccades and fixations reflects significant characteristics of the user's reading behavior and the object being gazed. Henny et al. [4] mentioned that eye gaze is a prominent nonverbal signal compared to pointing, body posture, and other behaviors.

We draw the motivation to use eye gaze as the medium to capture users' attention feedback, from the massive backing of literature on the relationship between eye gaze pattern and human interest [20]. Employment of eye movement data in an educational context has offered insight into how to model gaze to extract human attention patterns. The most noteworthy are the eye movement modeling examples (EMMEs), in which Jarodzka et al. used visual feedback to specifically influence gaze actions in order to enhance subjects' perception performance of medical records [17] and a biological classification task [18, 19]. Eye movement data of experts was visualized in [17] by blurring areas they did not look at, i.e., non-relevant details. Experts' gaze was visualized as yellow circles on a stimulus picture for [18]. The model example in both studies included gaze data post hoc. Orlov and Bednarik's open-source program ScreenMasker [34] created a flexible framework that visualizes gaze actions. Their gaze-contingent system overlays the on-screen stimulus with a pattern mask. It subtracts the pattern or unmasks, where the subject is looking in real-time, using gaze coordinates from the eye tracker. An NVIDIA graphics card with the CUDA architecture was used for this system [34]. Later, [35] proposed a platform integrated with a publicly available eye gaze analysis tool. The multiple plugins integrated offered an experimental center and a real-time gaze feedback option. Their system was tested and capable of running on a standard computer, but the visual search tasks were performed on images.

In this work, we design a real-time gaze-based attention feedback framework that identifies salience areas in mixed media Wikipedia articles (text, images). Wikipedia is a crowd-sourced and openly investigable source of information. As a result, ever-growing research interest has been shown to investigate various aspects of Wikipedia, and different metrics have been proposed for investigation purposes. The metrics include straightforward strategies like word count [10] to the models trained using deep learning [12] to evaluate various aspects of Wikipedia articles. Some other commonly used techniques to analyze Wikipedia articles are reference evaluation [24], editors' contribution [54], talk page evaluation [21], etc. In this work, we propose to use implicit AF from readers of

[2]Some dedicated gaze trackers like EyeLink 1000 provide non-invasive solutions but they also require high end camera and GPU processors. These requirements are rarely available at users' end

[3]https://github.com/szydej/GazeFlowAPI

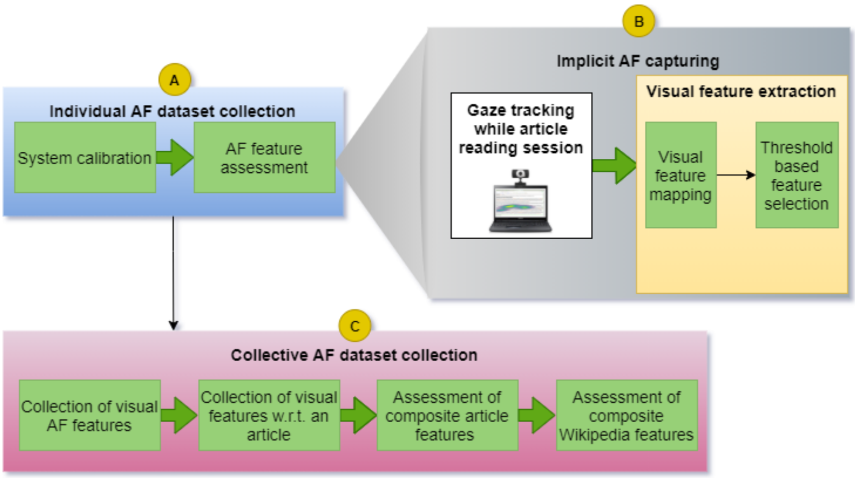


Figure 1: Overall system workflow to capture readers' AF. In block A, dataset collection happens corresponding to each reading session. Block B showcases the procedure to extract visual features. In block C, readers' attention feedbacks are collected to derive composite article/Wikipedia level features.

an article to analyze that article, using the concepts of collective intelligence.

There have been few attempts in the past where readers' input has been used to analyze crowd-sourced web portals. We focus on the crowd-sourced portals because we aim to utilize the knowledge and participation of crowd (users/editors) for portal analysis. In a Wikipedia-based research [40], readers' navigation behavior was analyzed to quantify the importance of references in Wikipedia. Stackoverflow is a successful crowd-sourced QnA portal. There have been some studies on Stackoverflow, which require readers' involvement. In one such study, Peterson et al. [39] proposed a gaze-based exploratory study on readers' information-seeking behavior, esp. developers on Stackoverflow. Gaze-based input from readers has also been utilized to identify the importance of citations in Stackoverflow posts [40]. Linden et al. [50] performed a detailed study to investigate readers' behavior in selecting answers in a thread (QnA post). GitHub is another web portal that supports the collaborative effort. It is a widely-used software development platform that provides features of version control and project hosting. There have been attempts on Github also, which involves input from users. In [16], authors predicted student learning outcomes from a survey of students and teachers while using GitHub in the classroom. In [51], a survey was conducted on GitHub to get a sense of how participants assess and value team composition and diversity.

The Article Feedback Tool (AFT)[4] was a Wikimedia survey for article feedback to engage readers in the assessment of article quality. It suffered from a lack of participation and noisy data. Similarly, the other commonly used approach of pageviews [11] also suffers from data insufficiency since pageviews only provide a count of visits for an article. To the best of our knowledge, the current work presents a novel approach to capture readers' implicit feedback for Wikipedia articles using a low-cost and effective gaze-based solution.

3 SYSTEM OVERVIEW

The generic overview of the proposed system is presented in Fig 1. The first step in the working of the AF system is the calibration of the eye tracker. Subsequently, a user performs reading action while the application captures visual attention pattern-related features. The key part of the overall architecture constitutes the implicit user feedback capturing process. For that purpose, a low-cost and lightweight gaze-tracking framework has been designed. It involves the usage of a single-camera image processing-based gaze-tracker and a dynamic interface for efficiently capturing gaze-related information at sentence level. We use GazeFlowAPI[5] as the eye-tracking solution.

We embed the eye tracker in a web application, WikiRead (https://wikianalysis.herokuapp.com/). We develop the application using NodeJS framework in Javascript and HTML. The application website is hosted using Heroku. Some screenshots of the web application are presented in Fig.2. This web application's primary purpose is to simplify the process of capturing Wikipedia readers' visual activity and analyze the AF-related features. On the main screen of our website, we present the welcome page of Wikipedia along with options to enter the title of the article user wishes to read and buttons to start and stop the reading session. Once a user clicks on the "Start" button, the system requests permission to share the screen content and the camera access. Camera sharing is safe for users because we do not store any images or videos of the user's face. It is required to perform eye-tracking. Once the sharing permission is granted, the control goes to the calibration[6] screen. Users are asked to follow the on-screen instructions to successfully

[4] https://en.wikipedia.org/wiki/Wikipedia:Article_Feedback_Tool

[5] https://github.com/szydej/GazeFlowAPI

[6] Calibration details are mentioned in the experimentation section.

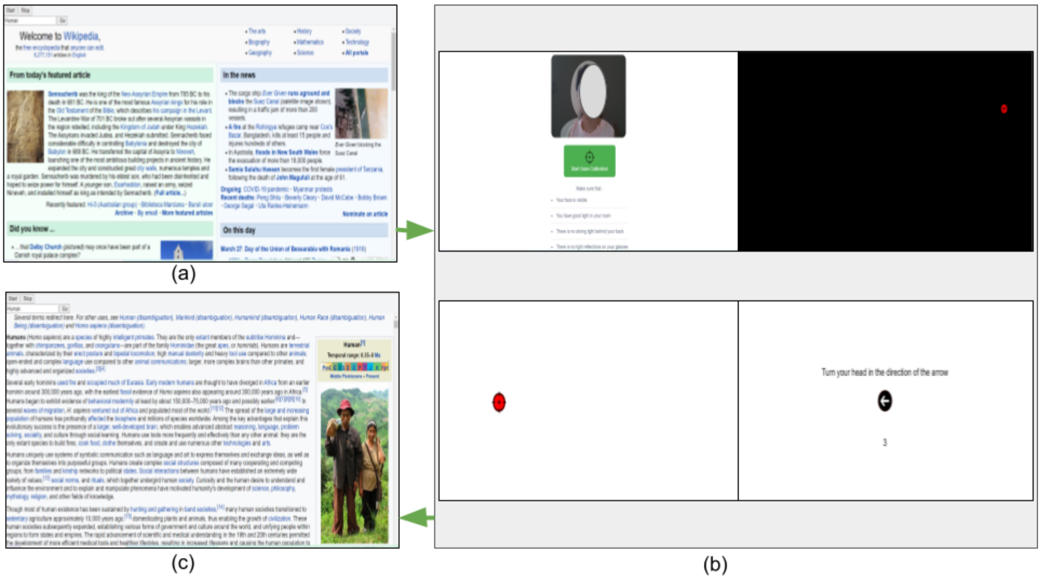


Figure 2: Various screens of the web application, WikiRead. (a) The main screen with options to enter article name and start the reading session. After which the control goes to calibration process. (b) Shows various calibration stages of the eye tracker. (c) Shows the article which user wants to read while system performs gaze tracking in background.

complete system calibration. Post calibration, the interface opens the article which the user wishes to read, and gaze tracking starts. While reading, the user can freely click on any Wikilink present in the article. Gaze data for the redirected articles are also stored accordingly. Users can finish the reading session anytime by clicking on the "Stop" button.

Following the gaze point trajectory estimation, a set of visual features are extracted only for those regions of an article that the user has observed. Then, a threshold-based feature selection procedure is applied for maintaining the most discriminative features. Table 1 shows the set of features that we capture as part of AF.

During dataset collection, we also ask the participants to manually annotate the regions of an article as per their preference level. Using the proposed framework, participants can manually select a region and rate it on a scale of 0 to 10 (10 being the most preferred region). This results in the generation of a set of automatic and manual gaze features for the sentences that have been seen by the user(s). We use MongoDB to store the dataset.

Finally, the visual AF features collected in various reading sessions are grouped as per the article title. It enables the possibility to evaluate article features using the concepts of collective intelligence. The observed benefits of social heuristics trace back to 1907, at least when Francis Galton experimented predicting the weight of an ox [15]. He showed that average estimates of median and mean were only 9 and 1 pounds, respectively, of the actual weight of 1198 pounds. This experiment proves that aggregating the crowd perspective can be helpful in cases where the size and diversity of the crowd are large. Wikipedia also receives participation from a vast and diverse crowd/users. The SOI AF (Sentences Of Interest based Attention Feedback) dataset represents crowd perspective. This knowledge can be efficiently used to analyze Wikipedia from a novel perspective.

4 ATTENTION PATTERN ANALYSIS

To capture readers' attention pattern in an article, we track their visual gaze features. These are valuable pieces of information and can be used to infer the reader's current cognitive processes [9]. Based on the gaze pattern, we extract text from the article. The below-mentioned steps are followed in sequence to track eye gaze and identify areas of interest in an article:

- Gaze tracking framework
- Attention feature selection

4.1 Gaze Tracking Framework

For each new user visiting the website (WikiRead), a randomly generated unique 20 bit code is assigned for identification. The sysId, i.e., the Identification code, is stored in the user's browser storage, and it can be deleted whenever the user deletes the storage for the site. For each unique user, a sessionID is created whenever the user starts the reading session. The sessionId is used to differentiate between various reading sessions for a single user. The website includes the English Wikipedia home page[7], in a sub-frame. The top header contains two buttons: Start and Stop and a search bar for the article title. Using Wikipedia in a sub-frame resulted in a famous web development problem called CORS[8]. This problem prohibits accessing the action events information within the sub-frame. To

[7] https://en.wikipedia.org/wiki/Main_Page
[8] https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS

overcome this problem, we implemented a proxy server and routed every GET request from the Wikipedia website inside the sub-frame through the proxy server. The proxy server makes a GET request to Wikipedia, through MediaWiki API[9], and gets Wikipedia web page with headers. The proxy changes the headers of Wikipedia.org related to CORS and adds our website in the header. Same-Origin-Policy, the process displays the Wikipedia page in our website domain's sub-frame and avoids the CORS problem. However, the GET request returns a plain web page without any styles, so we apply styles explicitly to the web page every time the sub-frame reloads.

As we mentioned earlier, the sophisticated gaze-tracking systems with increased accuracy are commercially available, e.g., Tobii1, SMI2, EyeTech3, and Mirametrix4, to name the most representative ones. However, a prohibitive factor for the extensive use of such specialized equipment constitutes their significantly high cost. To satisfy the requirements of low-cost and availability, the proposed AF system follows an image processing-based approach that uses a single camera for performing gaze-tracking.

We use a non-invasive eye-tracking setup containing Logitech Webcam C922 Pro Stream and open source eye-tracker applications, named GazeFlowAPI[10]. We select this eye-tracking solution because, along with access to sufficiently accurate real-time gaze information, it also provides eye blink, and head position data. These additional pieces of information are very crucial for the consolidated evaluation of attention patterns.

In the case of non-invasive trackers (no physical contact), gaze mapping deals with inherent noise and drift problems. While mapping a single fixation to a text location may be ambiguous, the matching of groups of fixations to a chunk of objects (words/sentences) can resolve this ambiguity [32]. In the proposed method, we perform gaze point mapping at the sentence level.

For further optimization purposes, we do not store the entire reading session's screen recording. Instead, we store only the required frames for further processing. A frame is the display screen snapshot at any given time-stamp. We identify the frames by assessing generic human reading behavior. The mean minimum time to acquire the whole meaning of a word is 151 ms [42]. It has also been shown that the majority of the sentences contain 11-15 words [46]. Since the gaze feedback is performed at the sentence level, we find the average time a reader takes to perceive the meaning of a sentence. We calculate this by multiplying the data mentioned above, i.e., $151 * 15$ ms. It results in approximately 3 seconds. Therefore, we store a frame only if it is on-screen for more than 3 seconds. We store only one instance of each frame.

4.2 Attention Feature Selection

In this work, the gaze features are defined considering only the fixations since they contain more valuable information and less noise than saccades. We adopt the Dispersion-Threshold Identification (I-DT) method [53] for defining fixations. According to this definition, a fixation is considered to occur if the gaze point remains in a circular area of radius R pixels for a minimum of T ms. For the employed gaze-tracker, the following (commonly used) values were selected based on experimentation: R = 25 pixels and T = 180 ms. A fixation is denoted as $F_i(x_i; y_i; t_s; t_e), i \in [1, N]$, where N is the total number of fixations. Point $(x_i; y_i)$ corresponds to the center of the aforementioned circular area. The values of t_s, t_e are the start and end time of the fixation, respectively, where $t_e - t_s > T$. It is to note that before the fixation identification, the gaze trajectory is low-passed for noise removal (estranged points). This is performed by applying a simple mean filter separately to the horizontal and vertical gaze coordinate signals. The low pass filtering is performed according to the technique mentioned in [36].

To identify a reader's regions of interest on a given frame, we need to map their fixations on corresponding frames. On a frame, we map all the fixations with time-stamps (t_s and t_e) within the frame's screen time. After appropriately sprinkling the fixations on each frame, we plot heatmaps based on gaze-points density. The heatmap depicts the time-series gaze density distribution on a frame. Fig. 3 shows a sample frame with gaze points mapping and the corresponding heatmap of gaze points density, post noise removal process.

For each reading session, along with the gaze density heat map, we also provide a set of sentences where a user-focused while reading along with the time for which each sentence was focused. By setting an appropriate threshold on the heatmap, we extract text from each frame using the OCR tool [38]. The extracted sentences might contain noise due to the characters' misidentification by the OCR tool. After processing the sentences, we arrange them in the order they are read along with their gaze quotient. By gaze quotient, we mean the time duration (in seconds) for which a sentence is being gazed at or read. It is calculated by finding the total time duration for which gaze points are mapped within the on-screen span of the sentence's spatial location.

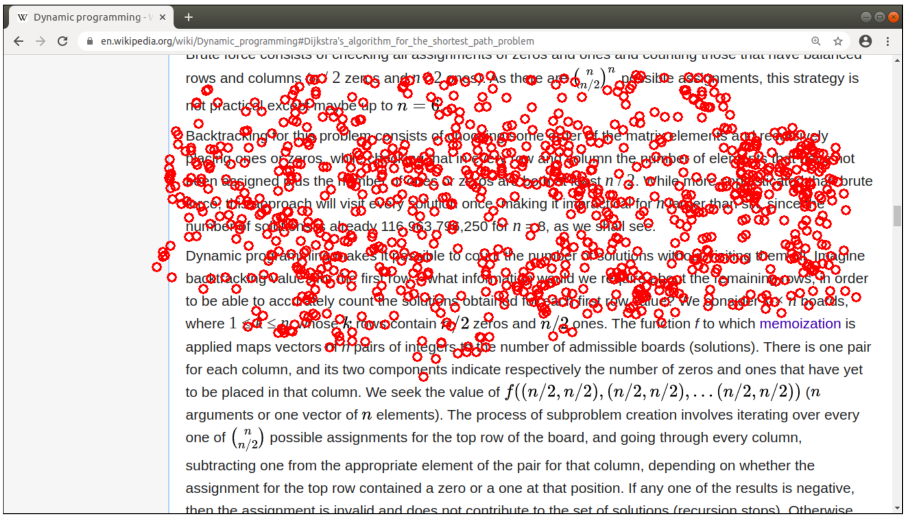
$$GQ(s) = t_s(j) - t_e(i) \tag{1}$$

In equation 1, $GQ(s)$ is the gaze quotient of sentence s. $t_s(j)$ is the start time of the first fixation on the screen span of s and $t_e(i)$ the end time of the last fixation on the screen span of s. We call the set of sentences along with their GQs as Sentences Of Interest (SOI). A user is free to click on any Wikilink. Therefore an SOI can contain sentences belonging to different articles. Later, we arrange the sentences in each SOI according to the article title.

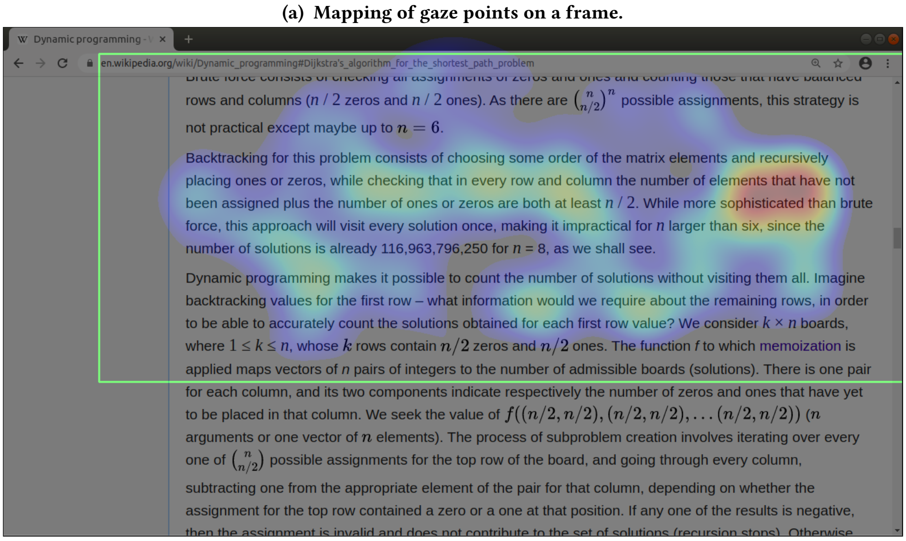
Along with the features mentioned above, the proposed AF framework also captures some additional information listed below:

(1) Wikilink clicks: We store the time series value of Wikilink [48] click data. It can be helpful to evaluate users' navigation path or information search patterns within Wikipedia.
(2) Eye blinks: We store the time series value of readers' eye blinks. The eye blink information has been proven to be critical to assess the reader's attention level [44].
(3) Scroll events: We store the time series value of the scroll events being performed while reading. The scroll information is a good indicator of browsing strategy [26]. The GazeFlowAPI does not provide built-in support to track mouse activity. Therefore, we externally added the scroll tracking feature in the proposed AF system using PyQt5.QtWidgets package of Python.

[9] https://www.mediawiki.org/wiki/API:Main_page
[10] https://github.com/szydej/GazeFlowAPI



(a) Mapping of gaze points on a frame.



(b) Heatmap based on gaze points density.

Figure 3: A sample gaze point distribution and gaze density heatmap on a frame after noise removal process.

Table 1 lists all the features stored in the dataset corresponding to each reading session. All the mentioned features are crucial to provide consolidated feedback from readers.

5 EXPERIMENT AND RESULTS

5.1 Experimental Setup

5.1.1 System Components. Subject's eye movements were recorded using GazeFlowAPI and a Logitech Webcam C922 Pro Stream on a VivoBook ASUSLaptop X430FN_S430FN with Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz, 1992 Mhz, 4 Cores, 8 Logical Processors, and 8GB RAM.

5.1.2 Eye Tracking Setup. The gaze tracker sampled at 30 fps, and the gaze angle was determined by the relative positions of corneal and pupil centers. The video resolution of the laptop during the test recording was 1280×720. Participants were seated 27 inches from a 14-inch flat-panel monitor that displayed the stimuli. Eye data were recorded using the proposed web application on the Google chrome browser. The embedded eye gaze tracker supports head movement. Therefore, we have not used any head/chin rest device during dataset collection.

5.1.3 Calibration Validation. Before starting the reading session, participants were instructed to watch a red dot move around the screen. The computer screen is interpreted as a coordinate system,

Table 1: AF features captured for each Wikipedia reading session performed on the proposed AF framework.

AF Feature	Description	Data Type
SysID	A unique system D for each user	20 bits
SessionID	A unique session ID for each reading session	20 bits
Article title	Title of the article being read	String
Gaze density heatmap	Heatmap based on gaze points density for each frame	Base64 encoded
SOI	Sentences of interest along with gaze quotient	List
Wikilink clicks	Time-series value of clicks on Wikilinks	List
Eye blinks	Time-series value of eye blink	List
Scroll events	Time-series value of scrolls with direction (up/down)	List

where the top left of the screen is (0, 0) and the bottom right of the screen is (1, 1). The ball's calibration movement was fixed at 16 positions in both bright and dark backgrounds sequentially with the frontal head position. Along with this, three positions each to positive and negative 45° yaw and pitch head positions. The head pose variation during dataset collection makes the gaze tracker robust to head pose variation during actual reading.

At each of the above points, the red dot paused and pulsed for ~2s. Post calibration, if the RMSE (Root Mean Square Error) value of the actual and predicted gaze points is higher than the set threshold (0.28), we ask the participant to repeat the calibration task before starting the reading session.

5.1.4 Subjects. Seventy-two clinically normal volunteers participated in the dataset collection. The participants belong to different professional backgrounds, including engineering students, industry professionals, doctors, and professors. The ethics committee approved the study of our university. Subjects underwent informed consent procedures approved by the university, and all subjects provided written and informed consent for participation in the study. Prospective participant-observers were required to have normal or corrected visual acuity. Participants were free to wear spectacles during dataset collection[11]. The mean age of the participants was 29.7 ± 7.6 (range: 22-53). The sample was 38.5% female, and the rest male.

5.2 Evaluation of Eye Tracking Solution

The performance of the eye tracker has a significant impact on the efficiency of the overall AF approach. As a consequence, the accuracy of the employed tracker is examined in this section. Table 2 briefly outlines some of the approaches that have been evaluated for measuring image processing-based gaze-trackers' performance. We select the gaze trackers based on their suitability for the aim of this work i.e. a commodity gaze solution for mass readers. Lu et al. [29] presented an adaptive linear regression based solution for appearance-based gaze estimation. In another work, Lu et al. [28] estimated 3D gaze directions using unlabeled eye images via synthetic iris appearance fitting. Papoutsaki et al. [37] discussed a web based gaze solution using regression analysis informed by user interactions. Computer Vision Center (CVC) released an eye tracking solution [1] which used usage of simple camera and image processing technique but this solution does not support head movements. Wang et al. [52] proposed a real-time 3D eye gaze capture with DCNN-based iris and pupil segmentation. Finally, Liu et al. [27] proposed a 3D model-based gaze tracking via iris features with a single camera and a single light Source. The restriction on light source makes this tracker less applicable in wild. The adopted GazeFlow eye tracker [2] provides high accuracy and satisfies all the requirements for this work. The comparison of GazeFlow with SMI RED 250 can be found here[12].

Table 2: Comparison of the adopted eye gaze tracker with other state-of-the-art image processing based eye trackers

Eye Tracker	Head Movement	Blink Data	Error
Lu et al. [29]	restricted	no	2.59±0.38
Lu et al. [28]	restricted	no	6.91±4.46
Papoutsaki et al. [37]	restricted	no	4.17
CVC [1]	restricted	yes	1.35±1.0
Wang et al. [52]	free	no	2.06
Liu et al. [27]	free	no	1.87±0.4
Adopted [2]	**free**	**yes**	**1.23±1.2**

For producing directly comparable evaluation results, the performance of the developed gaze-tracker was evaluated using the experimental protocols described in the works of [22]. Most of the works reported in Table 2 rely on the usage of static markers in fixed positions for evaluating the gaze tracking performance. In this way, the behavior of the gaze tracker is not adequately examined. For that purpose, a significantly more challenging and thoroughly defined experiment is proposed with the following two key characteristics: a) it can be easily reproduced b) it takes into account both the spatial accuracy (fixations and sccades) and the temporal coherence (timestamp data for Wikilink clicks, eye blinks and mouse scroll) of the tracker.

More specifically, a red dot was depicted on the screen performing the same trajectory as mentioned in the calibration process (Section 5.1.3). The participants were asked to follow the center of this red dot with their gaze. The tracker's accuracy was defined as the RMSE between the fixed points and their corresponding estimations.

Twenty individuals participated in this experiment, each performing the aforementioned task for all seven eye trackers. Regarding the specifications of the defined experiment, the monitor plane was vertically aligned, while the perpendicular vector originating from the monitor's center was maintained to approximately target the nose of the user and to also be perpendicular to the user's face plane. Additionally, the camera was placed on top of the monitor and at the center of the respective monitor's side, with the user's nose to be set to correspond approximately to the center pixel of the captured video sequence.

[11] We took precautions regarding possible glares in the spectacles of the participants during dataset collection

[12] https://www.slideshare.net/szymondeja3/raport-gaze-flow-vs-smi26092013en-1

Table 3: ROUGE-N evaluation of the proposed sentence selection technique with the manual sentence selection

Feature	Average value
P	0.717
R	0.514
F-measure	0.656

Table 4: Statistics of the collected AF dataset

Feature	value
Number of SOI_a	218
Average compression ratio of SOIs	15%
Average reading session duration	3.7 minutes

The comparative evaluation results given in Table 2 show that the developed gaze-tracker outperforms most of the state-of-art approaches and carries the additional advantage of head-movement and eye blink data. It makes this eye tracker a suitable choice for the proposed task of capturing implicit AF of readers.

At this point, it must be highlighted that the focus of this work does not include the proposal of a new gaze-tracker, whose performance needs to be accurately measured and to be superior compared to other state-of-art methods. On the contrary, the aim of this work is the proposal of a novel framework for interpreting the gaze signal and subsequently utilizing this information for realizing RF in the context of image retrieval, irrespective of the particular gaze-tracker that is used.

5.3 Evaluation of AF system

Using the above-mentioned experimental setup with seventy-two participants, we could conduct a total of 153 reading sessions. In these reading sessions, 84 unique articles were visited (including the article redirects using Wikilinks). We evaluate the performance of the proposed method against the extractive summarization task for the articles.

Precise identification of the reading pattern is required for capturing a reader's attention feedback of an article. The proposed approach identifies the sentences of interest (SOI) for each reading session and assembles them according to the article titles. To evaluate the SOI selection technique's quality, we partition all the SOIs with respect to the article titles. We call the partitioned SOIs as SOI_a. They contain sentences of a single article.

We verify the accuracy of the proposed sentence selection approach. We do so by comparing the ROUGE-N value of the sentences selected by the proposed approach with the true value i.e. manually selected sentences. To compare the manual and automated summaries, we use a modified version of ROUGE-N approach to compare the accuracy of sentence selection. ROUGE-N is an n-gram-based metric. It calculates the recall score (R), the precision score (P), and the F-measure score (F) for the text. Let S_{auto} be the set of sentences automatically selected by the proposed approach. Let S_{man} be the set of sentences in the manual summary.

$$P \triangleq \frac{Count(common\ sentences\ in\ S_{auto}\ and\ S_{man})}{|S_{man}|} \tag{2}$$

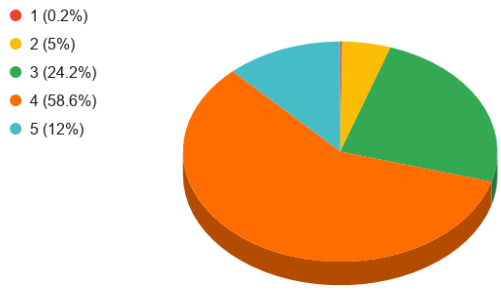


Figure 4: Distribution of users' rating for SOI selection on the scale of 0 to 5. The percentage value for 0 rating is 0% therefore it's not shown in the pie chart.

$$R \triangleq \frac{Count(common\ sentences\ in\ S_{auto}\ and\ S_{man})}{|S_{auto}|} \tag{3}$$

$$F-measure \triangleq \frac{2PR}{P+R} \tag{4}$$

The ROUGE comparison results are shown in table 3. The high value of f-measure, shows that there is high overlap between the sentences selected automatically and the sentences readers manually selected while reading the article.

We collect SOI dataset for further research purposes. Table 4 shows the values of various parameters of the collected dataset of SOIs. The count of SOIs per article indicates the number of SOIs we are able to gather to evaluate each article. Higher the number better will be precision for article analysis. The compression rate is calculated by finding the ratio between the number of bytes of an SOI and the number of bytes of the corresponding article. The average value of the compression rate for all the SOIs across articles is 15%. It implies that on average, users read only 15% of the article content. The average reading session duration is the total time a user takes to read an article, followed by the time to analyze their reading pattern. The timer starts with the calibration phase and ends with the click on the "Stop" button. The average value of completing all these tasks is only 3.7 minutes, which supports the hypothesis that readers use Wikipedia as a quick reference tool [47].

We analyze the efficiency of the SOI identification process using two evaluation techniques. The first evaluation consists of surveying all the participants and requesting them to rate their satisfaction levels with SOIs. The second evaluation involves using a formal metric (ROUGE-N), vastly used to assess text extraction procedures.

Satisfaction Survey. We gather users' perspectives regarding text extraction by conducting a survey. At the end of every reading session during dataset collection, we display on the screen the extracted SOI. We assess the user satisfaction level for SOI selection by requesting them to rate the SOI based on its similarity with the content they read in the article. The rating scale varies from 0 (low satisfaction) to 5 (high satisfaction). The source code along

Table 5: Comparison with other extractive summarization models

Model	R-L	R-1	R-2
LEAD-5	19.24	9.67	3.43
LexRank	26.23	11.89	3.12
SumBasic	22.34	14.17	5.34
SemSenSum	30.11	15.16	4.11
C SKIP	30.16	32.38	4.25
Ours	**30.92**	**37.61**	**6.26**

with the survey form is available at our Github repository[13]. The distribution of user ratings is plotted in Fig. 4. In this pie chart, the legend shows the rating with the corresponding percentage of rating. The 0 ratings are not shown in the figure because it was nil. We observe that 70% of the participants gave four or more ratings for SOI selection. The high user satisfaction ratings show the efficiency of the proposed reading pattern analysis approach.

Comparison with other extractive deep models. We compare the performance of the proposed approach with other extractive models. We include LEAD-k [45], which selects the first k sentences in the document as a summary. In this work, we have used LEAD-5 for comparison. LexRank [14] is a graph-based method, and it uses nodes as text units, and edges define the similarity measure. SumBasic [30] is a frequency-based sentence selection method that uses a component to re-weigh the word probabilities to minimize redundancy. The last extractive baselines are the near state-of-the-art models C SKIP [43] and SemSenSum [7]. The former exploits word embeddings' capability to leverage semantics, whereas the latter aggregates sentence semantic relation graph and graph convolution sentence embedding.

We select trained models and fine-tune them for the Wikipedia data dump[14]. This dump contains the current revision of all the English Wikipedia articles as of February 20, 2021. It does not contain talk or user pages. During fine-tune, we use a learning rate of 0.0001 with ten epochs. All models are fine-tuned on Nvidia GeForce GTX 1080 Ti GPU (60GB RAM, 12 GB dedicated graphic card, and 200 GB Hard drive space). We run all models with their best-reported parameters. Post-fine-tuning the pre-trained models, we compare these models' performance with our approach to the collected dataset of automatic and manual summaries (along with source articles for each summary). We consider three strong reference-based evaluation metrics: ROUGE 1, 2 & L. We use the recalls of ROUGE-1 (unigram), ROUGE-2 (bigram), and ROUGE-L (LCS). We use the official ROUGE script[15] (version 1.5.5) to evaluate the summarization output. In Table 5 we can see the comparison result. Our approach performs better than these models for ROUGE 1 & 2 as well as ROUGE-L metrics. The reason for the better performance of our approach for is that we are directly extracting sentences based on the user's reading pattern. Therefore, when we compare the proposed system generated summary with the manual summary, it results in high similarity.

[13]This is an anonymous repository with lesser details. Link to the principal repository will be provided post-acceptance.

[14]https://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-meta-current.xml.bz2

[15]http://www.berouge.com/

6 LIMITATIONS & FUTURE DIRECTIONS

The goal of this work is to devise a technique to capture AF of Wikipedia readers. However, during dataset collection, we observe that some readers did not perform their natural reading. They claim that the information of their gaze data being recorded impacts their reading behavior. We counter this problem by keeping the reading interface the same as the original portal and not disturbing reading sessions with unnecessary popups of real-time feedback. As a result, most participants reported that none of the feedback conditions were distracting in any way. Therefore, the proposed gaze feedback system appears to be unobtrusive yet effective.

Due to the novelty of the proposed approach, the doors of several new future directions open. For example, we can investigate article readability based on the readers' cumulative reading pattern over an article. We can also predict the popularity of an article based on how readers refer to the article. We can potentially expand our approach into a business analytics framework/artifact that shows the analysis steps. It can also help Design Science researchers to identify the user reference behavior on the portal and thus aid them in designing the user interface. We are in talks with Wikimedia Foundation to include our application as a Wikipedia extension so that users can easily use it and researchers can obtain a vast AF dataset.

It is to note that there still exists a strong potential for further improving the user's AF prediction based on gaze data. Towards this goal, future work includes the investigation and modeling of the factors that affect the way that users see (e.g., facial expression, mood). Furthermore, their integration to the developed framework.

In the future, we also plan to extend the idea for other crowd-sourced portals, such as Stack Exchange[16], Reddit[17], and Quora[18]. We believe the proposed method can help understand users' engagement on these portals by unraveling their eye gaze information. It will help editors to collect low-cost, implicit attention feedback of portal users and may a way to a novel research direction.

7 CONCLUSION

In this paper, a novel gaze-based attention feedback approach was presented for Wikipedia users. The overall approach aimed to estimate the area of interest and reading pattern of the users and subsequently use it to derive results related to their reading behavior on the portal. A novel set of gaze features representing visual characteristics was presented for performing users' attention assessment prediction. Extensive experiments demonstrated their efficiency compared to other feedback approaches used by the Wikipedia research community. The experimental evaluation proved that the proposed approach outperforms representative reading pattern analysis approaches of the literature. Moreover, incorporating a single-camera image processing-based gaze tracker into a web application framework makes the overall system cost-efficient and portable. This study's outcomes are currently being discussed in the Wikimedia Foundation for developing specialized tools to capture readers' implicit feedback.

[16]https://stackexchange.com/

[17]https://www.reddit.com/

[18]https://www.quora.com/

ACKNOWLEDGMENTS

This work was funded by the assistance received from CSRI, Department of Science and Technology India via grant no. SR/CSRI/344/2016.

REFERENCES

[1] [n.d.]. CVC Eye Tracker. https://github.com/tiendan/OpenGazer Accessed: 2016.

[2] [n.d.]. GazeFlow eye tracker. https://github.com/szydej/GazeFlowAPI Accessed: 2021.

[3] B Thomas Adler, Krishnendu Chatterjee, Luca De Alfaro, Marco Faella, Ian Pye, and Vishwanath Raman. 2008. Assigning trust to Wikipedia content. In *Proceedings of the 4th International Symposium on Wikis*. ACM, 26.

[4] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[5] Antti Ajanki, David R. Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. 2009. Can Eyes Reveal Interest? Implicit Queries from Gaze Patterns. *User Modeling and User-Adapted Interaction* 19, 4 (2009), 307–339. https://doi.org/10.1007/s11257-009-9066-4

[6] Judd Antin and Coye Cheshire. 2010. 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*. ACM, 127–130.

[7] Diego Antognini and Boi Faltings. 2019. Learning to Create Sentence Semantic Relation Graphs for Multi-Document Summarization. *arXiv preprint arXiv:1909.12231* (2019).

[8] Maram Barifah and Monica Landoni. 2019. Exploring usage patterns of a large-scale digital library. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 67–76.

[9] David Beymer and Daniel M Russell. 2005. WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze. In *CHI*. 1913–1916.

[10] Joshua E Blumenstock. 2008. Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 1095–1096.

[11] Yuru Cao, Hely Mehta, Ann E Norcross, Masahiko Taniguchi, and Jonathan S Lindsey. 2020. Analysis of Wikipedia pageviews to identify popular chemicals. In *Reporters, Markers, Dyes, Nanoparticles, and Molecular Probes for Biomedical Applications XII*, Vol. 11256. International Society for Optics and Photonics, 112560I.

[12] Quang Vinh Dang and Claudia-Lavinia Ignat. 2016. Quality assessment of wikipedia articles without feature engineering. In *JCDL'16*. ACM, 27–30.

[13] Nathan J Emery. 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *NBR* 24, 6 (2000), 581–604.

[14] Günes Erkan and Dragomir R. Radev. 2004. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *ArXiv* abs/1109.2128 (2004).

[15] Francis Galton. 1907. Vox populi (the wisdom of crowds). *Nature* 75, 7 (1907), 450–451.

[16] Courtney Hsing and Vanessa Gennarelli. 2019. Using GitHub in the classroom predicts student learning outcomes and classroom experiences: Findings from a survey of students and teachers. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 672–678.

[17] Halszka Jarodzka, Thomas Balslev, Kenneth Holmqvist, Marcus Nyström, Katharina Scheiter, Peter Gerjets, and Berit Eika. 2012. Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science* 40, 5 (2012), 813–827.

[18] Halszka Jarodzka, Katharina Scheiter, Peter Gerjets, Tamara van Gog, and Michael Dorr. 2010. How to convey perceptual skills by displaying experts' gaze data. In *Proceedings of the 31st annual conference of the cognitive science society*. Cognitive Science Society, 2920–2925.

[19] Halszka Jarodzka, Tamara Van Gog, Michael Dorr, Katharina Scheiter, and Peter Gerjets. 2013. Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction* 25 (2013), 62–70.

[20] Heather Knight and Reid Simmons. 2013. Estimating human interest and attention via gaze analysis. In *ICRA*. IEEE, 4350–4355.

[21] David Laniado, Riccardo Tasso, Yana Volkovich, and Andreas Kaltenbrunner. 2011. When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. In *AAAI*.

[22] Eui Chul Lee and Kang Ryoung Park. 2009. A Robust Eye Gaze Tracking Method Based on a Virtual Eyeball Model. *Mach. Vision Appl.* 20, 5 (July 2009), 319–337.

[23] Janette Lehmann, Claudia Müller-Birn, David Laniado, Mounia Lalmas, and Andreas Kaltenbrunner. 2014. Reader preferences and behavior on Wikipedia. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 88–97.

[24] Włodzimierz Lewoniewski, Krzysztof Węcel, and Witold Abramowicz. 2017. Analysis of references across Wikipedia languages. In *ICIST*. Springer, 561–573.

[25] C Li and J Bernoff. 2011. Groundswell: winning in a world transformed by social technologies, vol.

[26] Chang Liu, Jiqun Liu, and Yiming Wei. 2017. Scroll up or down? Using Wheel Activity as an Indicator of Browsing Strategy across Different Contextual Factors. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 333–336.

[27] Jiahui Liu, Jiannan Chi, Wenxue Hu, and Zhiliang Wang. 2020. 3D Model-Based Gaze Tracking Via Iris Features With a Single Camera and a Single Light Source. *IEEE Transactions on Human-Machine Systems* (2020).

[28] Feng Lu, Yue Gao, and Xiaowu Chen. 2016. Estimating 3D gaze directions using unlabeled eye images via synthetic iris appearance fitting. *IEEE Transactions on Multimedia* 18, 9 (2016), 1772–1782.

[29] Feng Lu, Yusuke Sugano, Takahiro Okabe, and Yoichi Sato. 2014. Adaptive linear regression for appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence* 36, 10 (2014), 2033–2046.

[30] Ani Nenkova and Lucy Vanderwende. [n.d.]. The impact of frequency on summarization. ([n. d.]).

[31] Blair Nonnecke, Dorine Andrews, and Jenny Preece. 2006. Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research* 6, 1 (2006), 7–20.

[32] Ayano Okoso, Kai Steven Kunze, and Koichi Kise. 2014. Implicit gaze based annotations to support second language learning. In *UbiComp*. Association for Computing Machinery, Inc, 143–146. https://doi.org/10.1145/2638728.2638783

[33] Anneli Olsen. 2012. The Tobii I-VT fixation filter. *Tobii Technology* (2012).

[34] Pavel A Orlov and Roman Bednarik. 2016. ScreenMasker: An open-source gaze-contingent screen masking environment. *Behavior research methods* 48, 3 (2016), 1145–1153.

[35] Kai Otto, Nora Castner, David Geisler, and Enkelejda Kasneci. 2018. Development and evaluation of a gaze feedback system integrated into eyetrace. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–5.

[36] Georgios Th Papadopoulos, Konstantinos C Apostolakis, and Petros Daras. 2013. Gaze-based relevance feedback for realizing region-based image retrieval. *IEEE Transactions on Multimedia* 16, 2 (2013), 440–454.

[37] Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: Scalable Webcam Eye Tracking Using User Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 3839–3845.

[38] Chirag Patel, Atul Patel, and Dharmendra Patel. 2012. Optical character recognition by open source OCR tool tesseract: A case study. *International Journal of Computer Applications* 55, 10 (2012), 50–56.

[39] Cole S Peterson, Jonathan A Saddler, Natalie M Halavick, and Bonita Sharif. 2019. A gaze-based exploratory study on the information seeking behavior of developers on stack overflow. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[40] Tiziano Piccardi, Miriam Redi, Giovanni Colavizza, and Robert West. 2020. Quantifying Engagement with Citations on Wikipedia. In *Proceedings of The Web Conference 2020*. 2365–2376.

[41] Keith Rayner. 1978. Eye movements in reading and information processing. *Psychological bulletin* 85, 3 (1978), 618.

[42] Eyal M Reingold and Keith Rayner. 2006. Examining the word identification stages hypothesized by the EZ Reader model. *Psychological Science* 17, 9 (2006), 742–746.

[43] Gaetano Rossiello, Pierpaolo Basile, and Giovanni Semeraro. 2017. Centroid-based text summarization through compositionality of word embeddings. In *Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres*. 12–21.

[44] Tsugunosuke Sakai, Haruya Tamaki, Yosuke Ota, Ryohei Egusa, Shigenori Inagaki, Fusako Kusunoki, Masanori Sugimoto, Hiroshi Mizoguchi, et al. 2017. EDA-based estimation of visual attention by observation of eye blink frequency. *International Journal on Smart Sensing and Intelligent Systems* 10, 2 (2017), 296–307.

[45] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368* (2017).

[46] HS Sichel. 1974. On a distribution representing sentence-length in written prose. *RSS* 137, 1 (1974), 25–34.

[47] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. 2017. Why we read wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*. 1591–1600.

[48] Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. 2012. Wikilinks: A large-scale cross-document coreference corpus labeled via links to Wikipedia. *University of Massachusetts, Amherst, Tech. Rep. UM-CS-2012* 15 (2012).

[49] Nathan TeBlunthuis, Tilman Bayer, and Olga Vasileva. 2019. Dwelling on Wikipedia: investigating time spent by global encyclopedia readers. In *Proceedings of the 15th International Symposium on Open Collaboration*. 1–14.

[50] Dirk van der Linden, Emma Williams, Joseph Hallett, and Awais Rashid. 2020. The impact of surface features on choice of (in) secure answers by Stackoverflow

readers. *IEEE Transactions on Software Engineering* (2020).

[51] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3789–3798.

[52] Zhiyong Wang, Jinxiang Chai, and Shihong Xia. 2019. Realtime and accurate 3D eye gaze capture with DCNN-based iris and pupil segmentation. *IEEE transactions on visualization and computer graphics* 27, 1 (2019), 190–203.

[53] Heino Widdel. 1984. Operational problems in analysing eye movements. In *Advances in psychology*. Vol. 22. Elsevier, 21–29.

[54] Dennis M Wilkinson and Bernardo A Huberman. 2007. Cooperation and quality in wikipedia. In *EC*. ACM, 157–164.

[55] Songhua Xu, Hao Jiang, and Francis Lau. 2009. User-oriented document summarization through vision-based eye-tracking. In *Proceedings of the 14th international conference on Intelligent user interfaces*. ACM, 7–16.

[56] Petro Zdebskyi, Victoria Vysotska, Roman Peleshchak, Ivan Peleshchak, Andriy Demchuk, and Maksym Krylyshyn. 2019. An Application Development for Recognizing of View in Order to Control the Mouse Pointer.. In *MoMLeT*. 55–74.