# Predicting Next Actions and Latent Intents during Text Formatting

Guanhua Zhang
University of Stuttgart
Stuttgart, Germany
guanhua.zhang@vis.uni-stuttgart.de

Susanne Hindennach
University of Stuttgart
Stuttgart, Germany
susanne.hindennach@vis.uni-stuttgart.de

Jan Leusmann
University of Stuttgart
Stuttgart, Germany
jan.leusmann29@gmail.com

Felix Bühler
University of Stuttgart
Stuttgart, Germany
st117123@stud.uni-stuttgart.de

Benedict Steuerlein
University of Stuttgart
Stuttgart, Germany
st111340@stud.uni-stuttgart.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Mihai Bâce
University of Stuttgart
Stuttgart, Germany
mihai.bace@vis.uni-stuttgart.de

Andreas Bulling
University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

## ABSTRACT

In this work we investigate the challenging task of predicting user intents from mouse and keyboard input as well as gaze behaviour. Different from prior work, we study intent prediction at two different resolutions on the behavioural timeline: predicting future input actions as well as latent intents to achieve a high-level interaction goal. Results from a user study (N=15) on a sample text formatting task show that the sequence of prior actions is more informative for intent prediction than gaze. Using only the action sequence, we can predict the next action and the high-level intent with an accuracy of 66% and 96%, respectively. In contrast, accuracy when using features extracted from gaze behaviour was significantly lower, at 41% and 46%. This finding is important for the development of future anticipatory user interfaces that aim to proactively adapt to user intents and interaction goals.

## KEYWORDS

intent recognition, action prediction, text formatting, machine learning

**ACM Reference Format:**
Guanhua Zhang, Susanne Hindennach, Jan Leusmann, Felix Bühler, Benedict Steuerlein, Sven Mayer, Mihai Bâce, and Andreas Bulling. 2022. Predicting Next Actions and Latent Intents during Text Formatting. In *New Orleans '22: CHI Workshop on Computational Approaches for Understanding, Generating, and Adapting User Interfaces, May 01, 2022, New Orleans, LA.* ACM, New York, NY, USA, 6 pages.



*Computational UI@CHI '22, May 01, 2022, New Orleans, LA*


## 1 INTRODUCTION AND RELATED WORK

Inspired by research in human-human interaction [3, 14], predicting users' intents and anticipating their future actions has the potential to enable a new generation of adaptive user interfaces (UIs) that support users in their everyday tasks. However, most current UIs are *reactive*, i.e. they only react upon receiving explicit user input. There is a growing interest in human-computer interaction (HCI) to develop computational methods to automatically predict users' intents based on their interactive behaviour that can be integrated into *anticipatory* IUIs, i.e. interfaces that are *proactive*. Intent recognition is challenging given that intents are not directly observable and manifest only once the user commits to and starts performing the sequence of actions towards a certain task-related goal. A number of prior works have instead focused on predicting users' next action, such as the next mouse location [8, 12, 17], source code edits [24], button presses [4], or selection of items in virtual reality (VR) [7, 9]. Other works predicted high-level intents while playing games [15, 16, 21] or when purchasing items after browsing an online shopping catalogue [1]. However, only few works studied different behavioural resolutions, i.e. individual actions vs. action sequences. For example, Loyola et al. [13] predicted the next selected item as well as the latent purchasing intent as a consequence of browsing multiple items in online shopping.

We fill this gap by proposing an intent prediction approach using action sequences generated from mouse and keyboard input as well as gaze dynamics. Mouse and keyboard are particularly attractive for intent prediction given that they are readily available on a large number of devices and are typically used frequently and over longer periods of time. From mouse input, Fu et al. [8] and Kwok et al. [12] predicted users' specific next action like annotating and clicking a button with an accuracy between 65% and 76%. Ottley et al. [17] used a hidden Markov model (HMM) to predict mouse interactions in a visual analysis task with an accuracy ranging from 92% to 97%. Agrawal et al. [1] predicted user intent to purchase an item from both mouse and keyboard input using a long short-term memory

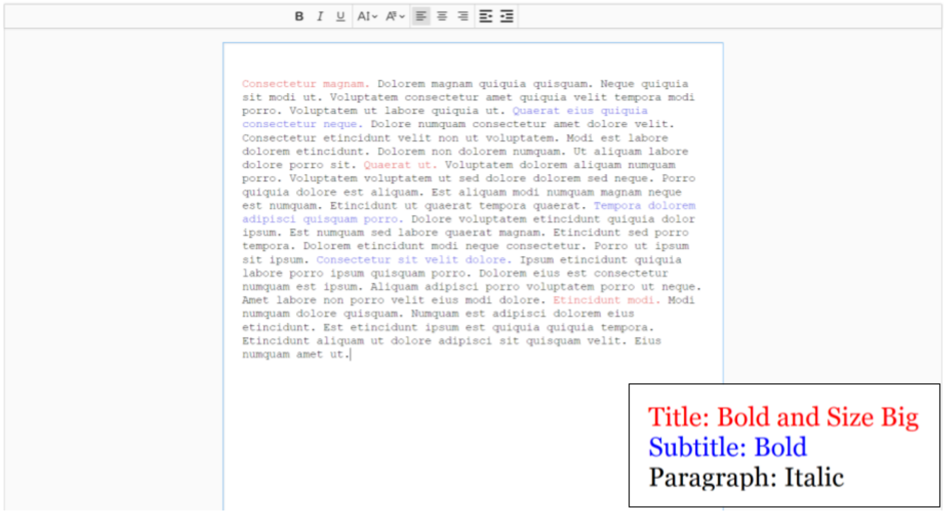


**Figure 1: Sample screenshot from the text formatting task including a rule set that participants had to implement in one trial. The text marked in red had to be formatted according to the formatting rules for titles (e.g. bold and size big), the text in blue according to the rules for subtitles and, finally, the remaining text according to the rules paragraphs.**

(LSTM) neural network and achieved an accuracy of 89%. Gaze was also shown to be indicative of user intent: Bednarik et al. [4] predicted whether users pressed a button in a puzzle. Gomez et al. [9] predicted users' selection from three items in VR using LSTM and achieved an accuracy of 80%. Similarly, David-John et al. [7] used logistic regression to predict the intent to select an item in VR. Despite this link, multimodal prediction using gaze remains under-explored. Singh et al. [21] predicted player intent in a board game using models based on Bayes' rule and showed that including gaze was more beneficial over using only actions. Soleymani et al. [23] predicted three visual search intents with random forest classifier from multimodal input, such as gaze, implicit user interactions (mouse and keyboard actions included), facial expressions and galvanic skin response. They found that implicit user interactions were the most informative from the individual modalities, outperformed only by the late fusion of all modalities.

To analyse if and how mouse and keyboard input as well as gaze dynamics contribute to intent prediction, we picked a text formatting task which, along with text entry and editing, is pervasive and has very high practical relevance in UIs. A large body of research has focused on predictive text entry [2, 5]. However, there are currently no works which predict user formatting intents, which is complementary to text prediction and more independent from the text's semantics. In this work, we are first to study intent prediction at two different resolutions on the behavioural timeline in a text formatting task. Besides predicting the specific action users will perform next, our method can also predict the higher-level goal achieved by a sequence of actions, noted as the overall latent intent.

## 2 USER STUDY

We conducted a data collection study in which participants applied different formatting rules to a given piece of randomly generated text. The text consisted of titles, subtitles and paragraphs, which were marked in three different colours (see Figure 1). In our study, we refer to applying a formatting action (i.e., one out of seven: bold, italic, underline, text size, font family, alignment or indentation) to the selected text as a user *action*. Similarly, a formatting rule is the user's underlying *latent intent*, made of three subrules to three parts of the text. For the full list of seven formatting rules see Appendix A.

The study was designed to have two parts, one covering a predefined, fixed set of formatting rules, and one that allowed participants to create their own custom rules. When applying a specific formatting action, the user was free to either click on the toolbar in the UI (e.g. click on **B** to mark in bold) or use keyboard shortcuts (e.g. CTRL+B for bold). In the first part, participants were given five repetitions of seven predefined formatting rules, resulting in 35 trials. By including each action in at least three rules, we ensured that the similarity between rules was comparable. In the second part, each participant created one rule, then applied the custom rule to the text freely, and hence created a custom number of titles, subtitles and paragraphs. This was repeated five times for a total of five trials. Before each part, participants performed two test trials to familiarise themselves with the task.

To recreate a realistically looking UI, we built a text editor front-end based on CKEditor5[1] (see Figure 1). We recorded mouse and keyboard events, as well as clicks on the toolbar and text selections. Gaze data was collected with the Tobii TX300 eye tracker via its Python SDK[2]. Participants were seated 76 cm away from a Dell U3014 display with a resolution of 2560×1600 pixels. Since the study setup did not involve a chin rest, we asked participants to restrict their head movements. 16 participants (13 male, 3 female), aged between 23 and 30 years old (M=25.25; SD=1.73), were recruited from an internal university self-volunteer pool. Gaze data was

[1] https://ckeditor.com/ckeditor-5/
[2] https://github.com/tobiipro/prosdk-addons-python

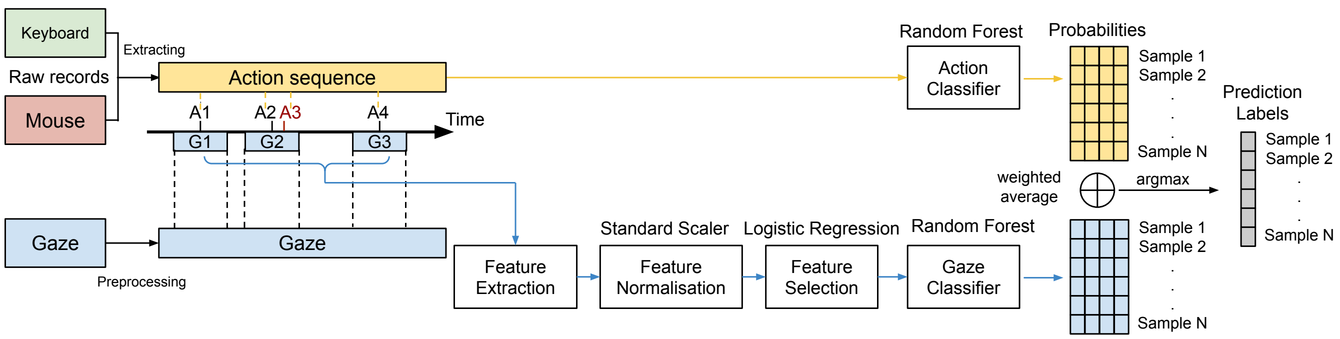


**Figure 2: Method overview.**

missing for one of the participants, hence, we used 15 participants in the following analyses.

## 3 METHOD

Our method has two main components (Figure 2). The first component predicts intent based on prior action sequences, i.e. the formatting history, extracted from mouse and keyboard logs. The second component uses gaze features extracted from a time window centred around actions to predict the same user intent. The two classifiers are trained independently and output a probability distribution over the possible user intents. A late fusion (decision-level fusion) module produces the final classification results by calculating a weighted average between the prediction probabilities of the two classifiers.

We first manually inspected and preprocessed the gaze data. Our method uses 1 s windows of gaze data centred around formatting actions. To denoise and smooth the gaze data, we applied a median filter over a window of seven gaze samples, similar to prior work [7, 19]. Then, we extracted a total of 124 gaze features from each 1 s window, covering the raw gaze data, fixations, and saccades. We used the Dispersion Threshold Identification (I-DT) algorithm [20] to detect fixations. Dispersion is the maximum distance between two consecutive gaze samples. We set the dispersion threshold to 50 pixels ($\approx 1^{\circ}$) and the duration threshold to 100 ms. Saccades were detected using the Velocity Threshold Identification (I-VT) algorithm [20] when the inter-sample velocity exceeded 500 pixels/s ($\approx 10^{\circ}/s$) or the acceleration reached 500 pixels/s$^2$ ($\approx 10^{\circ}/s^2$), and the duration was between 20 and 200 ms. For both I-DT and I-VT we used the implementation from PyGazeAnalyser [6]. Features were normalised to zero mean and unit variance. We trained a linear logistic regression model on the training set to assign weights to features and then selected the top 20% features using recursive feature elimination.

Following prior work on intent recognition [23], we opted for random forest classifiers. Two classifiers were trained separately on the two input modalities and generated the final prediction using late fusion. The label with the maximum weighted average output probability of both classifiers was used as the final prediction. We optimised the weight for maximum performance on the training set using a parameter search in the range $[0.1, 0.9]$ with a step size of 0.1. In addition, for both models we optimised their key parameters, i.e. the number of trees per forest $[10, 50, 100]$ and the maximum depth of the trees $[10, 20, 30]$. The standard scaler, logistic regression, recursive feature elimination, parameter grid search, and random forest classifier were implemented using Scikit-learn [18].

We compared the performance of our method with two baselines inspired by prior research. The **Classifier Baseline** was inspired by Singh et al. [22], in which the authors proposed an inference model based on Bayes' rule to estimate the probability of each intent according to the user's prior action sequences in a board game. However, their model was built on each user's own action history and was hence personalised to each specific user. Additionally, multiple optimal action sequences could lead to the same latent intent in their task but, in ours, given one formatting intent, the optimal action sequence is exclusive regardless of the order of actions. Therefore, we implemented this baseline using two separate Naive Bayes classifiers to predict user intent, Gaussian Naive Bayes from gaze (continuous values) and Multinomial Naive Bayes from past actions (discrete values). The final predictions were also obtained via weighted late fusion. **Gaze Feature Baseline** used the same downsampling method and gaze feature set as [7] which predicted user intents in a fundamentally different task using only gaze. Their task was to predict if a user intends to select an item by clicking the VR hand controller. Note that this baseline used a different feature set for gaze data. For classification, we used a random forest classifier. Therefore, this baseline only influences the component for gaze and shares the same component and results for action sequences with our method.

## 4 EXPERIMENTS

A leave-one-participant-out cross-validation was applied to better illustrate generalisability and real-world performance. N models were trained on the data from N-1 participants and tested on the remaining one. The evaluation metric for the multi-class classification problem is the averaged accuracy across iterations. We compare the performance with ablated versions of our method, i.e. using either action sequences or gaze only, on two different resolutions of the behavioural timeline. All evaluations were carried out on the data collected in the first part of our study in which the formatting rules were predefined. We generated each train/test sample following a windowing approach: a 125-dimensional vector (124 gaze features + 1 action) was produced from 7 prior actions and 7 corresponding 1 s

**Table 1: Top-1, top-2 and top-3 accuracy for predicting the next formatting action using the sequence of actions (A), gaze (G) and their fusion (AG) on given formatting rule sets. Best results in each modality are underlined. Among them, the best results across modalities are marked in bold.**

| Approach | Top-1 | | | Top-2 | | | Top-3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | G | AG | A | G | AG | A | G | AG |
| Classifier Baseline | 0.35 | 0.26 | 0.36 | 0.49 | 0.44 | 0.50 | 0.63 | 0.59 | 0.63 |
| Gaze Feature Baseline | – | 0.29 | 0.65 | – | 0.46 | 0.82 | – | 0.60 | 0.91 |
| Ours | **0.66** | 0.41 | 0.65 | **0.86** | 0.60 | 0.83 | **0.95** | 0.71 | 0.92 |

gaze windows centred around each action. Due to the gaze window, our method needs to wait for an extra 500 ms after each action to collect gaze behaviours, so it might miss actions that appear during this time. In Figure 2, our method uses an example window of size two with actions A1 and A2, meaning it will miss A3 and can only predict A4. However, since A3 is performed nearly immediately after A2 (less than 500 ms), we argue that this is a minor limitation and it is even unclear whether in this case, a prediction would help users in their task. We marked such cases in as incorrect predictions of our method, which negatively impacted the evaluation metrics.

## 4.1 Predicting User Intent Towards a Specific Next Action

This task is a multi-class classification problem with seven classes which correspond to the seven formatting actions, such as making a word italic or left-aligning a paragraph. One practical application could show users the future formatting action they may apply. While recommending multiple actions may improve the UI's friendliness, recommending too many actions results in a longer decision time according to the Hick-Hyman Law [10, 11]. Therefore, besides evaluating the performance of predicting the next action (*top-1*), we also evaluated the performance of predicting the *top-2* and *top-3* most likely actions. Picking an action at random, i.e. a random baseline, the top-1, top-2 and top-3 accuracies are 14%, 29% and 43%, respectively. Table 1 shows the results of our method and the other two baselines using the action sequence only (A), gaze only (G), and the fusion of both modalities (AG). Note that the Gaze Feature Baseline shares the same results for actions as explained in Section 3. We conducted a $3 \times 3$ (approach $\times$ modalities) ANOVA to examine the effect on prediction accuracy. Both factors have a significant main effect for top-1, top-2 and top-3 accuracies[3]. Our approach outperforms the other baselines on all inputs, achieving a top-1 accuracy of 66%, a top-2 accuracy of 86% and a top-3 accuracy of 95%. This difference was significant between Classifier Baseline and our method, and between Gaze Feature Baseline and ours for gaze features only (G) in a post-hoc Tukey HSD test. These top accuracies were obtained on actions only (A). The post-hoc Tukey HSD test also revealed that the ablated version of our method when using gaze (G) achieved significantly lower accuracy: 41% for top-1, 60% for top-2, and 71% for top-3, respectively. The fusion (AG) decreases the top performance to 65% for top-1, 83% for top-2, and 92% for top-3. This difference between AG and A was significant in a post-hoc Tukey HSD test. Therefore, our evaluations show that the action sequence alone performed best for this prediction task.

**Table 2: Accuracy of intent prediction achieved on actions only (A), gaze only (G) and their fusion (AG) on given formatting rule sets. Best results in each modality are underlined. Among them, the best results across modalities are marked in bold.**

| Approach | A | G | AG |
|---|---|---|---|
| Classifier Baseline | 0.22 | 0.15 | 0.20 |
| Gaze Feature Baseline | – | 0.23 | **0.82** |
| Ours | **0.96** | 0.46 | 0.80 |

## 4.2 Predicting Overall Latent User Intent

We predicted the overall latent intent, i.e. one of seven predefined formatting rules (see Section 2). The chance level accuracy for this 7-class classification problem is 14%. Table 2 shows that the best accuracy is achieved using our method and the action sequences only (A) 96% vs. gaze only (G) 46% vs. the fusion of both modalities (AG) 80%. A two-way ANOVA revealed a main effect for both factors[4]. Post-hoc Tukey HSD test showed that our approach significantly outperformed the Classifier Baseline in all modalities, and the Gaze Feature Baseline on gaze dynamics only. The best performance of our approach was obtained when using action sequences (A) (significant in post-hoc Tukey HSD test compared to gaze and the fusion of both). Similarly to our previous evaluation, the action sequences carried more information about the overall latent intent.

# 5 DISCUSSION

## 5.1 Two Resolutions of User Intent

In contrast to prior work, we investigated intent prediction at two behavioural resolutions: user intent towards a specific next action (applying a single formatting action) and the latent intent towards a broader, higher-level goal (implementing the formatting rule set). Our approach achieved similar accuracy for top-3 next action and latent intent prediction: 95% and 96%. The latent intent in our task is a formatting rule set, which consisted of three formats for the title, the subtitle, and paragraphs. Additionally, top-X accuracy measured whether the performed action was within the most probable

[3]Specific next action: Top 1 approach $F(2, 126) = 72.87, p < 0.01$, Top 1 modality $F(2, 126) = 73.10, p < 0.01$; Top 2 approach $F(2, 126) = 115.03, p < 0.01$, Top 2 modality $F(2, 126) = 83.71, p < 0.01$; Top 3 approach $F(2, 126) = 105.83, p < 0.01$, Top 3 modality $F(2, 126) = 78.23, p < 0.01$

[4]Latent intent: approach $F(2, 126) = 546.08, p < 0.01$, modality $F(2, 126) = 313.50, p < 0.01$

X predictions. Hence, knowing the intent constrains the set of possible actions (aside from errors) to three formatting actions. Our approach captured this constraint similarly well at both resolutions.

### 5.2 Contributions of Two Modalities in Intent Prediction

We found that both intents can be inferred from user actions shown in ablation studies ("A" in Table 1 and Table 2), which is in line with findings from prior work in other scenarios [1, 8, 12, 21]. The most likely reason why input actions contribute more is that they are predominantly *goal-oriented* and *sparse*. That is, they are only performed to implement an immediate or long-term intent, e.g. moving the mouse cursor to button **B** and clicking it to make the selected text bold. Additionally, they may not happen frequently, e.g. if the hand is taken off the mouse or during typing breaks. In stark contrast, gaze behaviour occurs all the time – whether to support mouse and keyboard input in the form of "look-aheads" or to explore, navigate, or search the complex visual interface. Larger parts of gaze behaviour will therefore not be directly relevant to the intent at hand.

### 5.3 Limitations and Future Work

Based on our insights, we see four fruitful areas for future work. First, while our approach outperformed the baselines, integrating gaze dynamics did not improve performance. This calls for further research on how to integrate the rich information content available in gaze for this task. Second, so far, we only analysed data collected in the first part of the study in which participants followed predefined formatting rules. We plan to investigate if and how behaviour changed in the second part in which participants defined their own, custom rules. Third, the user study was conducted in a controlled laboratory setting. While this was necessary to collect accurate gaze data, we plan to also study interactive behaviour during everyday tasks, such as in the office. Finally, given the promising results, we plan to build an adaptive UI using this approach that proactively assists users based on the predicted latent intents and actions.

## ACKNOWLEDGMENTS

G. Zhang, S. Hindennach, J. Leusmann, and A. Bulling were funded by the European Research Council (ERC; grant agreement 801708). M. Bâce was funded by a Swiss National Science Foundation (SNSF) Early Postdoc. Mobility Fellowship.

## REFERENCES

[1] Rakshit Agrawal, Anwar Habeeb, and Chih Hsin Hsueh. 2018. Learning user intent from action sequences on interactive systems. *The Workshops of the Thirty-Second AAAI Conference on Artificial Intelligence Learning* (2018), 59–64. arXiv:1712.01328 https://aaai.org/ocs/index.php/WS/AAAIW18/paper/view/17434

[2] Kenneth C Arnold, Krysta Chauncey, and Krzysztof Z Gajos. 2018. Sentiment Bias in Predictive Text Recommendations Results in Biased Writing. In *Graphics Interface*. 42–49.

[3] Chris L. Baker, Julian Jara-Ettinger, Rebecca Saxe, and Joshua B. Tenenbaum. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour* 1, 4 (2017), 1–10. https://doi.org/10.1038/s41562-017-0064

[4] Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What do you want to do next: A novel approach for intent prediction in gaze-based interaction. *Eye Tracking Research and Applications Symposium (ETRA)* 1, 212 (2012), 83–90. https://doi.org/10.1145/2168556.2168569

[5] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. 2019. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2287–2295.

[6] Edwin S Dalmaijer, Sebastiaan Mathôt, and Stefan Van der Stigchel. 2014. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eye-tracking experiments. *Behavior research methods* 46, 4 (2014), 913–921.

[7] Brendan David-John, Candace Peacock, Ting Zhang, T Scott Murdison, Hrvoje Benko, and Tanya R Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–7. https://doi.org/10.1145/3448018.3458008

[8] Eugene Yujun Fu, Tiffany C.K. Kwok, Erin You Wu, Hong Va Leong, Grace Ngai, and Stephen C.F. Chan. 2017. Your Mouse Reveals Your Next Activity: Towards Predicting User Intention from Mouse Interaction. *Proceedings - International Computer Software and Applications Conference* 1 (2017), 869–874. https://doi.org/10.1109/COMPSAC.2017.270

[9] Carlos Gomez Cubero and Matthias Rehm. 2021. Intention Recognition in Human Robot Interaction Based on Eye Tracking. In *IFIP Conference on Human-Computer Interaction*. Springer, 428–437.

[10] William E Hick. 1952. On the rate of gain of information. *Quarterly Journal of experimental psychology* 4, 1 (1952), 11–26.

[11] Ray Hyman. 1953. Stimulus information as a determinant of reaction time. *Journal of experimental psychology* 45, 3 (1953), 188.

[12] Tiffany C.K. Kwok, Eugene Yujun Fu, Erin You Wu, Michael Xuelin Huang, Grace Ngai, and Hong-Va Leong. 2018. Every Little Movement Has a Meaning of Its Own: Using Past Mouse Movements to Predict the Next Interaction. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 397–401. https://doi.org/10.1145/3172944.3173002

[13] Pablo Loyola, Chen Liu, and Yu Hirate. 2017. Modeling User Session and Intent with an Attention-based Encoder-Decoder Architecture. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 147–151. https://doi.org/10.1145/3109859.3109917

[14] Andrew N. Meltzoff. 1995. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology* 31, 5 (1995), 838–850. https://doi.org/10.1037/0012-1649.31.5.838 arXiv:NIHMS150003

[15] Wookhee Min, Bradford Mott, Jonathan Rowe, Barry Liu, and James Lester. 2016. Player Goal Recognition in Open-World Digital Games with Long Short-Term Memory Networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 2590–2596.

[16] Truong-Huy Dinh Nguyen, David Hsu, Wee-Sun Lee, Tze-Yun Leong, Leslie Pack Kaelbling, Tomas Lozano-Perez, and Andrew Haydn Grant. 2011. CAPIR: Collaborative Action Planning with Intention Recognition. In *Proceedings of the Seventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Stanford, California, USA) *(AIIDE'11)*. AAAI Press, 61–66.

[17] Alvitta Ottley, Roman Garnett, and Ran Wan. 2019. Follow The Clicks: Learning and Anticipating Mouse Interactions During Exploratory Data Analysis. *Computer Graphics Forum* 38, 3 (2019), 41–52. https://doi.org/10.1111/cgf.13670 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13670

[18] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[19] Jami Pekkanen and Otto Lappi. 2017. A new and general approach to signal denoising and eye movement classification based on segmented linear regression. *Scientific reports* 7, 1 (2017), 1–13.

[20] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78.

[21] Ronal Singh, Tim Miller, Joshua Newn, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. 2018. Combining planning with gaze for online human intention recognition: Socially interactive agents track. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* 1 (2018), 488–496.

[22] Ronal Singh, Tim Miller, Joshua Newn, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. 2020. Combining gaze and AI planning for online human intention recognition. *Artificial Intelligence* 284 (2020), 103275.

[23] Mohammad Soleymani, Michael Riegler, and Pål Halvorsen. 2017. Multimodal Analysis of Image Search Intent: Intent Recognition in Image Search from User Behavior and Visual Content. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval* (Bucharest, Romania) *(ICMR '17)*. Association for Computing Machinery, New York, NY, USA, 251–259. https://doi.org/10.1145/3078971.3078995

[24] Rui Zhao, David Bieber, Kevin Swersky, and Daniel Tarlow. 2019. Neural Networks for Modeling Source Code Edits. (2019). arXiv:1904.02818 http://arxiv.org/abs/1904.02818

# A FORMATTING RULES

All the seven formatting rules are listed below, including the sub-rules for title, subtitle and paragraph.

(1) Title: size big and bold. Subtitle: bold. Paragraph: italic
(2) Title: font family consolas and alignment center. Subtitle: font family consolas. Paragraph: italic
(3) Title: size big. Subtitle: bold. Paragraph: font family consolas
(4) Title: 1 indent and italic. Subtitle: 1 indent. Paragraph: font family consolas
(5) Title: size big and underline. Subtitle: underline. Paragraph: alignment right
(6) Title: alignment center and underline. Subtitle: underline. Paragraph: 1 indent
(7) Title: bold and underline. Subtitle: bold. Paragraph: 1 indent