Modeling Hierarchical Attentional Control in Web Reading Using ACT-R

Nethmee Rasanga Gunasekara, Joy Karmoker, Umito Sorita, Nilupul Heshan Randika Kodikara, Junya Morita

Shizuoka University
gunasekara.nethmee.24@shizuoka.ac.jp

Abstract

Web-based reading in online learning requires readers to regulate attention while navigating heterogeneous content under cognitive constraints. Rather than reading linearly, learners dynamically alternate between skimming, deep reading, and disengagement. This paper presents a transparent cognitive model of web reading implemented in ACT-R (Anderson et al., 2004) within a unified attentional control framework. Web content is represented hierarchically (viewport, paragraph, sentence) and semantic information is incorporated into symbolic processing through clustered embeddings, enabling realistic simulation of attention allocation, comprehension, and quitting behavior. Simulation results, conditioned on users' screen-captured content, show that reading persistence is primarily driven by sentence-level attentional stability, modulated by higher-level exploration. The model bridges information foraging and text comprehension by providing a mechanistic account of skimming, deep reading, and disengagement in web-based learning.

Keywords: Web reading; Attentional control; ACT-R; Cognitive modeling; Information foraging; Reading engagement

1. Introduction

Web-based reading is central to modern online learning. Students navigate visually heterogeneous content, dynamically alternating between skimming, selective sampling, and deeper reading rather than reading linearly (Liu, 2005). Readers adaptively regulate attention by balancing exploration and exploitation, deciding when to skim, engage deeply, and when to move on (Pirolli & Card, 1999). From this perspective, variability in reading persistence reflects strategic information-seeking rather than failure. Research further shows that frequent local switching disrupts coherence building, while sustained attention supports deeper information integration (DeStefano & LeFevre, 2007; Kintsch, 1998). However, most empirical studies rely on surface-level measures (dwell time, navigation patterns) without modeling the internal mechanisms governing attentional allocation, memory retrieval, and disengagement decisions.

We present a transparent ACT-R model of web reading that captures how readers dynamically shift between skimming and deep reading while pursuing task goals. The model explicitly represents perceptual, attentional, memory, and decision-making processes over a hierarchical representation of web content, enabling simulation of realistic reading behavior within a single webpage. By grounding simulations in empirically observed reading behavior and conditioning on each user's screen-captured content, this work provides a mechanistic account of skimming, deep reading, and disengagement that links observable behavior to underlying attentional control mechanisms.

2. Related Work

Research consistently shows that web reading is rarely linear. Readers dynamically alternate between skimming and deep reading, selectively attending to headings, paragraphs, and salient sentences to assess relevance before committing to detailed processing (Liu, 2005). Eye-tracking and interaction studies demonstrate rapid attentional shifts across webpage regions reflecting strategic resource management (Brumby et al., 2013; Delgado et al., 2018). Studies of web navigation characterize reading as attentional refocusing: readers frequently interrupt reading to reorient attention based on perceived relevance and task demands (Brumby et al., 2013). From this perspective, quitting reflects rational attentional control rather than lack of motivation.

Information foraging theory (Pirolli & Card, 1999) conceptualizes information seeking as an adaptive process balancing exploration costs against expected informational value. The SNIF-ACT model (Scent-based Navigation and Information Foraging in ACT-R; (Fu & Pirolli, 2007)) extends this to describe how attention is allocated across information patches under uncertainty. While effective at explaining high-level navigation, these models abstract away from fine-grained reading processes and do not represent attentional regulation across multiple levels of textual structure or the role of memory retrieval in moment-to-moment reading decisions.

Kintsch's Construction-Integration (CI) model describes comprehension as incremental integration of textual information with prior knowledge to form a situation model (Kintsch, 1988, 1998). Although developed for linear texts, it provides the theoretical basis for the imaginal-memory interaction in the present model. Prior ACT-R-based models have addressed information forag-

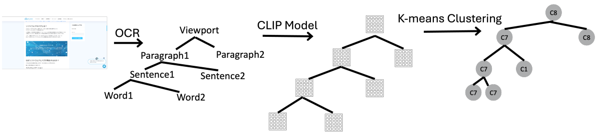

Figure 1　Process of extracting web content and its hierarchical representation.

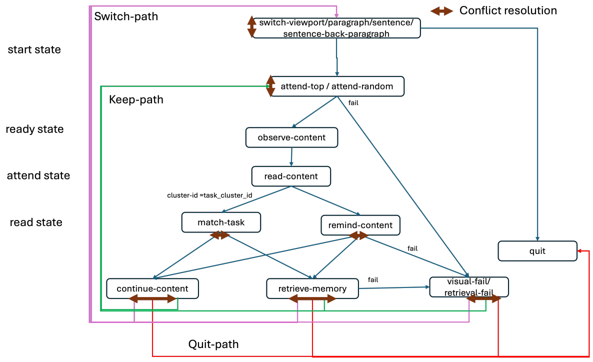

Figure 2　Flow of production rules in the procedural knowledge. The boxes indicate production rules. The color of one-way arrows represent the distinction of paths. The bidirectional arrows indicate conflicts resolved by sub-symbolic computation

ing and text comprehension but typically abstract away from the hierarchical, visually heterogeneous structure of web pages. The present work integrates hierarchical web content representation with ACT-R's transparent control mechanisms, unifying information seeking, attentional regulation, and comprehension within a single framework.

3. Approach

3.1 Hierarchical Representation of Web Content

Reading attention does not operate at a single level; readers dynamically shift between coarse and fine levels depending on task demands and perceived relevance. Web content is organized into three levels: (1) the entire viewport; (2) paragraphs; (3) sentences. For each viewport encountered during reading, textual content was extracted. Then, this content is segmented into paragraphs and sentences forming a nested hierarchy as illustrated in Figure 1. Each extracted text segment is converted into a semantic embedding using a large language model. These embeddings capture the semantics of web content from abstract to detailed levels. The embeddings are clustered to reduce dimensionality and mapped onto a limited set of symbolic representations that serve as input to the ACT-R model. This mapping enables the integration of rich semantic information with cognitively interpretable processing mechanisms.

3.2 ACT-R Module Structure

The model integrates four core ACT-R modules, which are utilized as follows:

Visual module. Represents perceptual access to web content within the current viewport. Locates textual elements as symbolic representations and shifts the attentional spotlight to a selected region. Visual encoding failures (no remaining content at the current level) are explicitly represented and trigger attentional reorganization or disengagement.

Imaginal module. Maintains an internal representation of currently attended content. Consistent with Kintsch's CI model (Kintsch, 1988, 1998), comprehension is modeled as incremental integration of newly read information with prior knowledge to form a situation model. Text semantic cluster identifiers are encoded into the imaginal buffer and combined with task goals and retrieved memory, enabling relevance evaluation, coherence maintenance, and attentional control.

Declarative memory. Stores prior knowledge and previously encountered content indexed by semantic cluster identifiers. Retrieval success or failure critically determines whether reading continues, attention shifts, or disengagement occurs.

Procedural module. Production rules govern attentional shifts, reading strategies (skimming vs. deep reading), memory evaluation, and quitting decisions. Competition is resolved via ACT-R's utility-based conflict resolution. As illustrated in Figure 2 reading behavior of the model cycles among four major paths: *Switch-path* (exploratory reorientation by shifting hierarchical level), *Keep-path* (maintaining current focus), *Reading-path* (perceptual-cognitive cycle encoding content and evaluating via *match-task* and *remind-content*), and *Quit-path* (competing via utility and triggered by a time-based production). The relative utilities of these productions determine the balance between skimming, deep reading, and disengagement.

3.3 Sub-symbolic Parameters

Separate utilities govern viewport-, paragraph-, and sentence-level switching, keep-reading, and quitting. Declarative memory retrieval is controlled by activation and noise parameters: successful retrieval stabilizes attention; failure promotes disruption. The model accumulates time and repeated failures, gradually increasing quitting utility, yielding graded, probabilistic behavior mirroring variability in human web reading.

4. Simulation and Results

4.1 Human Experiment

Twelve undergraduate informatics majors (6 male, 6 female) completed two 30-minute web search and report writing tasks in Japanese under high- and low-interest conditions (July-November 2024)(Sorita et al., 2025). Task difficulty was manipulated via conceptual complexity.Participants freely browsed and composed short reports with no quality evaluation. Behavioral logs, screen recordings, screenshots, and pre/post interest rat-

Smartphone apps, games, and other programs that run on devices can sometimes fail to work properly or need to be fixed immediately. Why do these "failures" occur? Also, please research and summarize your thoughts on how to reduce failures.

Figure 3 High interest task description for User 1.

With the spread of voice dialogue systems and smart devices, designing user-system interaction is becoming increasingly important. These interfaces need to not only be functional, but also be designed to be "ease of use" and "noticeable,"taking into account user perception and behavior. Your assignment is to search for examples and innovations in interface design that take human cognitive characteristics (attention, perception, reasoning, etc.) into account, and explain in your own words how they affect the user experience.

Figure 4 High interest task description for User 2.

ings were collected. Two cases were selected for ACT-R simulation, focusing on high-interest sessions without self-reported mind-wandering.

4.2 Model Simulation

For each screen-captured webpage, text was extracted via Optical Character Recognition (OCR)[*1] and segmented into paragraphs and sentences. Screenshots were sampled whenever a participant remained on a viewport for more than 3 seconds, ensuring simulated reading material closely matched experienced content. Semantic representations were obtained using the Contrastive Language–Image Pre-training model (CLIP) (Radford et al., 2021). These embeddings were computed for the full viewport, each paragraph, and each sentence, then clustered using K-means ($K = 10$) to yield discrete cluster identifiers. Task goals were represented in the same format: embeddings were computed from the assignment texts as shown in Figure 3 and Figure 4 and clustered using the identical procedure, and the resulting cluster identifiers were stored in the ACT-R goal buffer.

The model simulates web reading as an iterative cycle of attentional allocation, perceptual encoding, memory evaluation, and quitting decisions, mirroring the structure of the human task. Simulations systematically varied the initial utility values of switching productions across three factors: (A) Viewport-Switch utility, (B) Paragraph-Switch utility, and (C) Sentence-Switch utility, each set to Low (5.0), Middle (8.0), or High (10.0). All other model parameters were held constant[*2], allowing differences in simulated behavior to be attributed to controlled manipulations of attentional switching preferences rather than ad hoc rule changes.

[*1] pytesseract

[*2] :esc t, :lf .05, :trace-detail medium , :rt -2, :ul t, :egs 0.1, :v t, show-focus t, :visual-num-finsts 10, :visual-finst-span 10, :reward 5 was given to *reminde-content*, *continue-content*, and *match-task* productions, :reward 0 was given to *retrieval-fail* and *visual-fail* productions

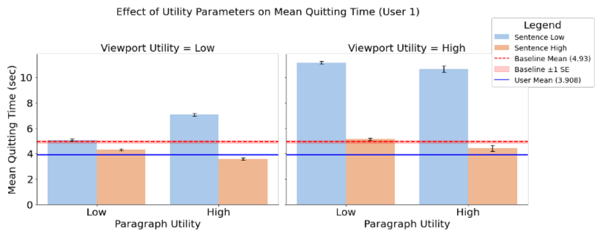

Figure 5 Effects of utility values on mean quitting times of the model (User 1 data)

To systematically explore these effects, we defined nine simulation groups based on combinations of initial utility values for the three switching factors: eight forming a 2×2×2 factorial design (Low/High per factor) plus a baseline with all utilities at Middle (N = screenshots × 5 runs per group). We manipulate hierarchical switching utilities to model skimming and deep reading as competing attentional strategies, and we use quitting time as a global behavioral measure because it emerges naturally from ACT-R's utility-based control mechanisms and reflects the participant's evolving engagement state during web reading.

4.3 Results

User 1 ($N = 200$ screenshots): The baseline condition produced a mean quitting time of 4.93 s, closely approximating the empirical mean of 3.91 s. Figure 5 shows that conditions with low sentence-switch utility produced the longest quitting times, particularly when combined with high viewport ($M = 11.15$, $SE = 0.244$) or paragraph ($M = 10.66$, $SE = 0.241$) switching utilities, biasing the model toward maintaining fine-grained attention and prolonged engagement. High sentence-switch utility markedly reduced quitting times, reflecting frequent attentional reorganization and shallower processing. When viewport switching was low under high sentence-switch conditions, quitting times were shortest ($M = 3.57$, $SE = 0.080$), indicating that restricted high-level exploration combined with frequent fine-grained switching increases disengagement likelihood.

User 2 ($N = 202$ screenshots): As shown in Figure 6, a similar pattern was observed, with baseline quitting time of 5.42 s approximating the empirical mean of 4.53 s. Low sentence-switch utility again produced the longest engagement, especially combined with high viewport ($M = 12.41$, $SE = 0.299$) or paragraph ($M = 11.81$, $SE = 0.266$) switching. A factorial ANOVA on User 2's data confirmed a significant three-way interaction among viewport-, paragraph-, and sentence-level utilities, $F(1, 1608) = 26.08$, $p < .001$, confirming that quitting behavior emerges from coordinated control across hierarchical levels rather than any single mechanism. Although absolute mean quitting times differed between users, reflecting differences in content and reading dynamics, the qualitative effects

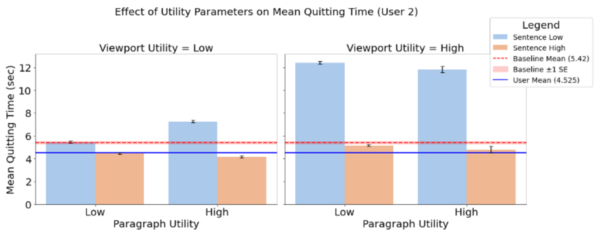

Figure 6　Effects of utility values on mean quitting times of the model (User 2 data)

of hierarchical switching utilities were consistent across both simulated datasets. Prolonged engagement consistently arose when attentional switching was encouraged at coarser levels while constrained at the sentence level, whereas increased sentence-level switching promoted shallow processing and earlier disengagement. Because all simulations for a given user operated on identical content, these effects reflect internal attentional control dynamics rather than content variation.

5. Discussion

This work demonstrates how a transparent cognitive architecture explains variability in web reading persistence by modeling attentional control across hierarchical content levels. Across simulated users, sentence-level attentional control emerged as the primary determinant of persistence. Low sentence-switch utility produced the longest quitting times, particularly paired with high viewport- or paragraph-level switching, suggesting that sustained engagement arises from broad exploration combined with fine-grained stability at the sentence level. High sentence-switch utility led to frequent local reorganization, shallower processing, and earlier disengagement, consistent with behavioral findings on the costs of excessive local scanning (DeStefano & LeFevre, 2007).

The model bridges information foraging and text comprehension by balancing exploration and exploitation across multiple levels of textual granularity while incorporating comprehension-related memory processes. Interactions between *match-task* and *remind-content* mechanisms show how semantic relevance and memory retrieval jointly regulate attentional stability, explaining transitions between skimming, deep reading, and disengagement. Integrating clustered semantic embeddings into ACT-R preserves sensitivity to content meaning while maintaining symbolic interpretability.

From an applied perspective, the model enables inference about when adaptive intervention is beneficial—detecting excessive sentence-level switching preceding disengagement or recognizing productive coarse-level exploration. Explicit representations of attentional and motivational states provide a principled basis for adaptive learning technologies that support sustained attention and effective information processing. Limitations include the focus on two individual cases under high-interest conditions, which constrains generalizability, and the use of quitting time as the sole behavioral outcome.

6. Conclusion

This paper presents a cognitively grounded ACT-R model of web reading that accounts for skimming, deep reading, and disengagement within a unified framework. Simulation results show that reading persistence emerges from interactions among attentional control mechanisms across hierarchical levels, particularly the balance between coarse-grained exploration and fine-grained stability. Conditioning the model on each user's actual reading material demonstrates that these effects reflect internal cognitive control processes rather than content differences. Future work will extend the model to larger samples, incorporate mind-wandering mechanisms, and explore how real-time estimates of attentional state can inform adaptive interventions.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111* (4), 1036–1060.

Brumby, D. P., Cox, A. L., & Back, J. (2013). Recovering from an Interruption: Investigating Speed–Accuracy Trade-offs in Task Resumption Behavior. *Journal of Experimental Psychology: Applied*, *19* (2), 95–107.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't Throw Away Your Printed Books: A Meta-Analysis on the Effects of Reading Media on Reading Comprehension. *Educational Research Review*, *25*, 23–38.

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive Load in Hypertext Reading: A Review. *Computers in Human Behavior*, *23* (3), 1616–1641.

Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, *20* (5), 325–344.

Fu, W.-T., & Pirolli, P. (2007). SNIF-ACT: A Cognitive Model of User Navigation on the World Wide Web. *Human-Computer Interaction*, *22* (4), 355–412.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction–Integration Model. *Psychological Review*, *95* (2), 163–182.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Liu, Z. (2005). Reading Behavior in the Digital Environment: Changes in Reading Behavior over the Past Ten Years. *Journal of Documentation*, *61* (6), 700–712.

Pirolli, P., & Card, S. K. (1999). Information Foraging. *Psychological Review*, *106* (4), 643–675.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. *arXiv preprint arXiv:2103.00020*.

Sorita, U., Kodikara, N. H. R., Karmoker, J., Gunasekara, N. R., & Morita, J. (2025). Experimental Study of Interest Change and Search Behavior in Web-based Report Writing Tasks. *Japanese Society for Artificial Intelligence Research Meeting Materials Advanced Learning Science and Engineering Research Meeting*, *105*, 07–12.