Word Tracker: An Interactive Book System

J. Vinathi

*Dept. of Information Technology*

*Chaitanya Bharathi Institute of Technology*

Hyderabad, India

Ugs22007\_it.vinathi@cbit.org.in

M. Sai Varshitha

*Dept. of Information Technology*

*Chaitanya Bharathi Institute of Technology*

Hyderabad, India

Ugs22015\_it.varshitha@cbit.org.in

N. Manvika

*Dept. of Information Technology*

*Chaitanya Bharathi Institute of Technology*

Hyderabad, India

ugs22017\_it.manvika@cbit.org.in

**Abstract— This project aims to provide an innovative online platform for students enrolled at CBIT (Chaitanya Bharathi Institute of Technology). The platform offers access to a comprehensive collection of academic books spanning across various departments. One of the distinguishing features of this platform is its integrated exploration tool designed to enhance the reading experience. When encountering unfamiliar terms or concepts while perusing the texts, students can simply highlight the word or phrase. In doing so, a contextual tab dynamically appears, furnishing a concise yet informative explanation of the selected term or concept. This feature serves to alleviate comprehension barriers, facilitating a smoother learning process for students. Furthermore, the integration of the exploration tool elevates the platform's utility by providing immediate assistance to students encountering unfamiliar terminology or complex concepts within the texts. For instance, if a student encounters a technical jargon or a specialized term while studying, they can simply ctrl+ right click on the term to summon a succinct yet comprehensive explanation. This feature not only aids in comprehension but also fosters a deeper understanding of the subject matter by providing contextual clarity.**

***Keywords - Online Book Store, Interactive Book, meaning, words, genres***

# **I. Introduction**

In the digital age, reading has evolved with the advent of online platforms that offer portability and interactive features. However, encountering unfamiliar words can disrupt the reading flow. This project addresses this issue by developing a web-based application that allows users to read books and get instant definitions for words they click on.

The platform is designed using HTML, CSS, and JavaScript, for a responsive interface. The backend, built with Node.js, manages data and API interactions, while a dictionary API provides real-time word definitions. The user interface is intuitive, featuring a clean reading environment and unobtrusive pop-ups for definitions.

This project enhances the digital reading experience by integrating vocabulary support, promoting continuous learning, and making literature more accessible and enjoyable. This paper details the design, development, and implementation of the application, highlighting technical solutions and its impact on readers.

# **II. BACKGROUND AND MOTIVATION**

The current system for vocabulary acquisition typically involves traditional methods such as memorization exercises, flashcards, and dictionary definitions. While these methods may be effective to some extent, they often lack engagement and fail to provide meaningful context for learners. Educational initiatives may incorporate vocabulary instruction, but they rarely integrate it with leisure reading or explore the nuanced meanings of words within narrative contexts. As a result, learners may struggle to retain newly acquired vocabulary or apply it effectively in real-world scenarios. The existing system overlooks the potential benefits of integrating contextual learning into leisure reading, especially within genre fiction, horror, and educational novels, thus limiting opportunities for holistic language development and comprehension.

Embarking on the project of highlighting words for their meanings within genre fiction, horror, and educational novels offers a rich tapestry of benefits. Not only does this endeavor expand your vocabulary, but it also deepens your comprehension of the texts you engage with. By actively seeking out and defining words within their narrative contexts, you're not merely learning isolated definitions; you're gaining a nuanced understanding of how language shapes storytelling. This process stimulates cognitive functions, reinforcing memory, attention, and reasoning skills. Moreover, integrating educational goals with leisure reading underscores the practical application of learning, making it both enjoyable and sustainable. Through this project, you explore complex themes within literature while fostering personal growth and self-improvement. Sharing your insights and discoveries through a project report contributes to the collective knowledge base and may inspire others to engage in similar pursuits. Ultimately, completing such a project is a celebration of your dedication, curiosity, and passion for learning, reflecting your commitment to intellectual exploration and academic excellence.

# **III. RELATED WORK**

# The evolution of digital reading platforms and tools has significantly influenced our project. Notable contributions include:

Platforms like Amazon Kindle and Google Books offer vast collections of digital texts but often lack immediate word comprehension tools, which our project aims to provide. Apps such as Instapaper and Pocket allow for article saving and annotation but do not offer instant word definitions, a gap our project addresses with double-click functionality. E-readers like Kindle and Apple Books feature built-in dictionaries that require pressing on words. Our project enhances this by offering real-time definitions via an open-source dictionary API with double-click ease. Apps like Duolingo and Memorize use interactive methods to teach vocabulary. Although focused on language learning, their success in vocabulary enhancement informs our approach within a reading context. APIs such as Dictionary API provide extensive word data, enabling our project to deliver accurate and immediate definitions, enhancing the reading experience without leaving the platform. Research supports that learning words in context is more effective. Our project incorporates instant definitions within the text, aiding better comprehension and retention.

**IV. PROPOSED METHODOLOGY**

A. SYSTEM SPECIFICATIONS

The Chrome extension should be compatible with various operating systems, including Windows, macOS, and Linux, as long as they support the Chrome browser.The extension should be designed to work with the latest stable version of the Chrome browser, as well as previous versions if possible, to ensure broad compatibility.Since the extension primarily operates within the Chrome browser, it should have minimal hardware requirements. It should be able to run smoothly on most modern computers, including laptops and desktops, with standard configurations.

The extension should be optimized to consume minimal memory to ensure smooth browsing performance. It should not significantly impact the overall browsing experience or slow down the browser.A stable internet connection is required for accessing Wikipedia and fetching additional information related to words and concepts. Ensure that the extension gracefully handles scenarios with limited or intermittent internet connectivity.

The PDF viewer integrated into the extension should support common PDF formats and features, ensuring that users can easily navigate through the provided books and documents. The extension's user interface should be intuitive and easy to use, with clear navigation options for accessing different genres, searching for words, and viewing additional information from Wikipedia. Ensure that the extension follows best practices for security, especially when fetching data from external sources like Wikipedia. Users' privacy and data security should be prioritized. Plan for regular updates and maintenance to address any bugs, security vulnerabilities, or compatibility issues that may arise over time. Keeping the extension up-to-date will ensure a smooth and reliable user experience. Provide comprehensive documentation and support resources for users to understand how to use the extension effectively and troubleshoot any issues they encounter.

B. SYSTEM DESIGN

The UI will consist of the Chrome extension interface accessible from the browser toolbar.Thehome page of the extension will display options for searching, selecting genres (e.g., help, fiction, education, horror, novel), and accessing settings.When a genre is selected, the UI will display a list of available PDF books within that genre.Clicking on a book will open it in the integrated PDF viewer.

The backend of the extension will handle interactions with external services such as Wikipedia for fetching word meanings, synonyms, and related information. It will also manage the storage and retrieval of PDF books and metadata related to genres and books. The extension will include a built-in PDF viewer or utilize an existing one to display PDF books seamlessly within the Chrome browser. The viewer should support basic functionalities like scrolling, zooming, and searching within the document. When a user selects a word within the PDF document and performs a specific action (e.g., Ctrl+ right click), the extension will capture the selected word. The selected word will then be sent to the backend for processing. The backend will fetch the meaning, synonyms, and related information for the word from an external source (e.g., Wikipedia API). The retrieved information will be displayed to the user in a pop-up or sidebar within the PDF viewer. Theextension will integrate with the Wikipedia API to fetch additional information about selected words. Upon receiving a word query, the extension will send a request to the Wikipedia API to retrieve relevant articles or definitions. The retrieved information will be presented to the user alongside the word tracking results. Metadata about genres and books will be stored locally within the extension. PDF books themselves may be stored locally or fetched from an external source, depending on the design preferences and copyright considerations. Ensure that all communication with external services (e.g., Wikipedia API) is done over secure HTTPS connections. Implement appropriate security measures to protect user data and prevent unauthorized access to sensitive information. Provide options for users to customize their experience, such as adjusting the font size in the PDF viewer or enabling/disabling certain features. Settings should be accessible from the extension interface and persist across sessions. Implement robust error handling mechanisms to gracefully handle unexpected scenarios, such as network errors or API failures. Log relevant information for debugging purposes and to track user interactions. Conduct thorough testing of the extension to ensure functionality across different Chrome browser versions and operating systems. Perform unit tests, integration tests, and user acceptance tests to verify the reliability and usability of the extension.

By following this system design, you can create a robust and user-friendly "Word Tracker" extension that seamlessly integrates PDF book browsing with word tracking and Wikipedia integration features.

**V. EXPERIMENTAL SETUP**

A. DATA COLLECTION

i. Word meanings and synonyms

We have used World web Dictionary Lookup extension which is available in chrome web store.

The following features of the extensions are:

* One-click lookup in almost any Windows program
* Hundreds of thousands of definitions and synonyms
* The latest international English words
* Works offline, or reference to Wikipedia and web references

Link to download the extension:

<https://chromewebstore.google.com/detail/wordweb-dictionary-lookup/ilikenhndcpmliapkmmhoimckaokmihm.html>

ii. PDF Books

Books that I have used which are downloaded from chrome are:

* Java
* Operating System
* Python
* Electronic devices and circuit theory
* Electronic devices
* Introduction to mechanical engineering
* Thermodynamics

B. FLOW CHART

The “Word Tracker" project Algorithm is:

Step 1: Word Embeddings

Utilize pre-trained word embeddings such as Word2Vec, GloVe, or Fast Text to capture semantic similarities between words. These embeddings can be used to find synonyms or related words.

Step 2: Named Entity Recognition (NER)

Use NER models to identify named entities in text, such as person names, locations, organizations, etc. This can be helpful for extracting relevant entities from Wikipedia articles or identifying important terms within the PDF books.

Step 3: Text Summarization

Implement text summarization algorithms to generate concise summaries of Wikipedia articles or PDF book sections. This can help provide users with quick insights into the content without having to read the entire text.

Step 4: Word Sense Disambiguation

Implement algorithms to disambiguate word senses, especially when a word has multiple meanings. This can improve the accuracy of word tracking and provide more relevant information to the user. Information Retrieval: Use techniques from information retrieval to retrieve relevant Wikipedia articles or definitions based on user queries. Algorithms such as TF-IDF, BM25, or neural network-based models can be used for this purpose.

Step 5: PDF Parsing

Implement algorithms to parse PDF documents and extract text content from them. This may involve using libraries such as PyPDF2 or pdf miner to handle PDF files and extract text for word tracking.

Step 6: Information Retrieval

Use techniques from information retrieval to retrieve relevant Wikipedia articles or definitions based on user queries. Algorithms such as TF-IDF, BM25, or neural network-based models can be used for this purpose.

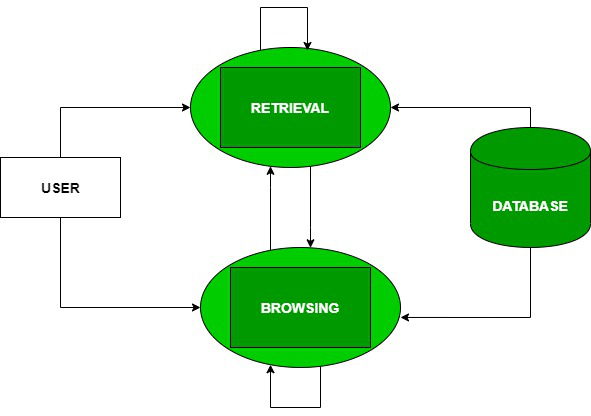


Figure 1: Flowchart

C. RESULTS



Figure 2: Home page

The home page of an online bookstore serves as the primary gateway for users.



Figure 3: Fiction Books page

The fiction page of an online bookstore is dedicated to showcasing a diverse and extensive collection of fiction books.

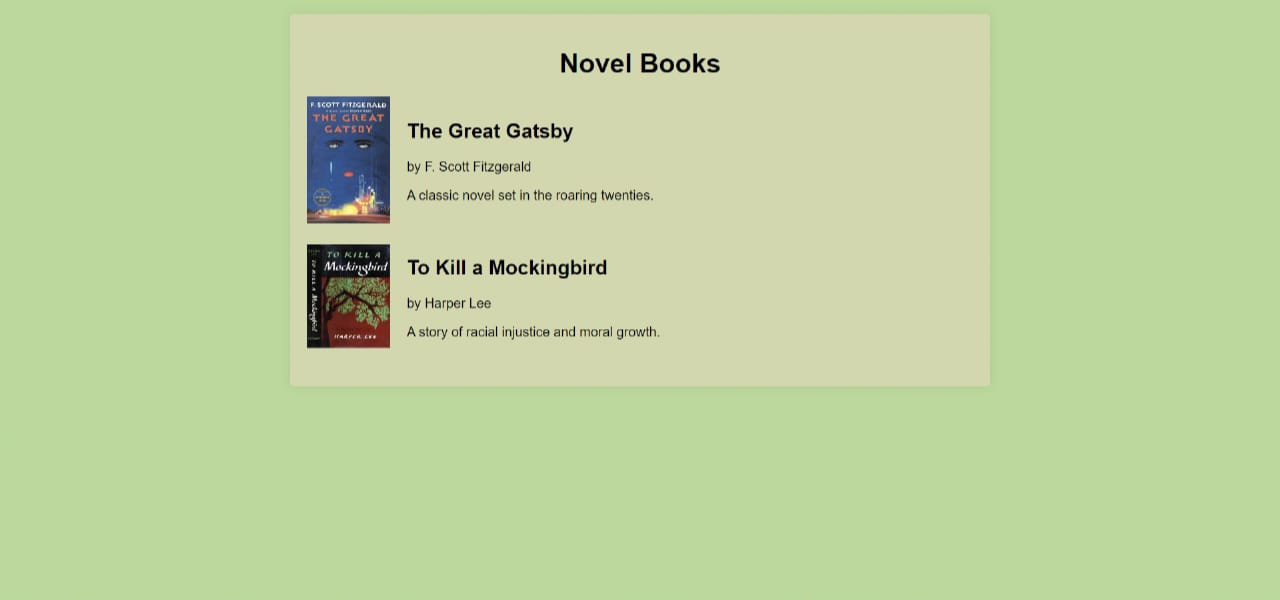


Figure 4: Novel Books page

The novel page of an online bookstore is dedicated to presenting a curated collection of fiction novels.

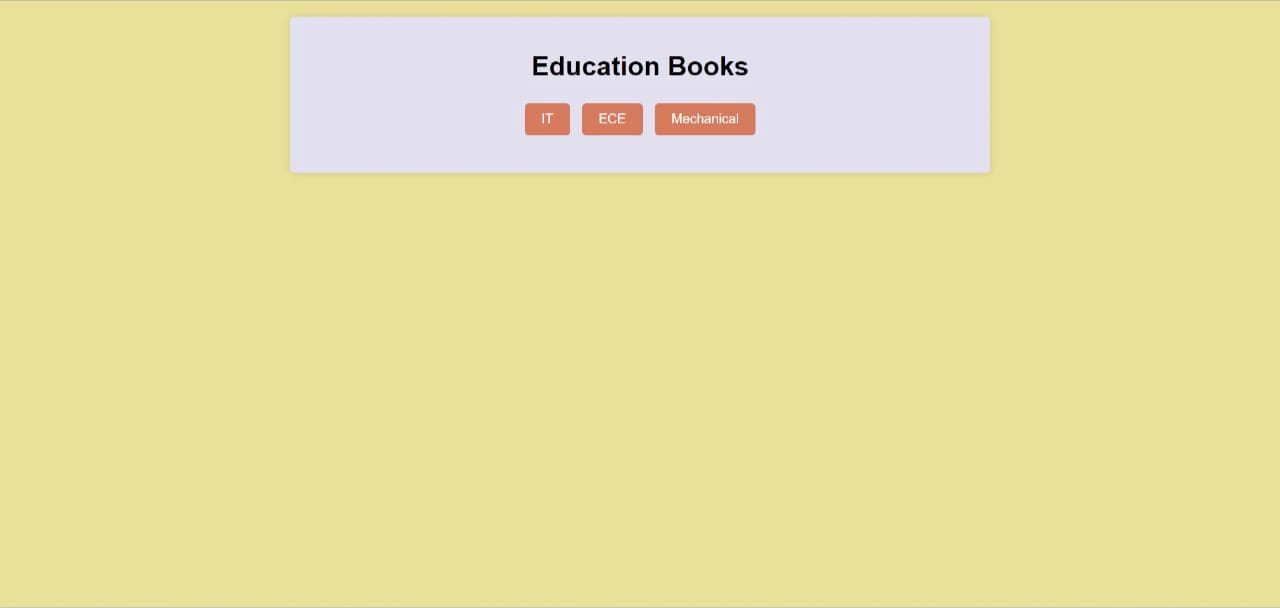


Figure 5: Education books page

The education page of an online bookstore dedicated to educational books is designed to

serve educators, students, and lifelong learners.

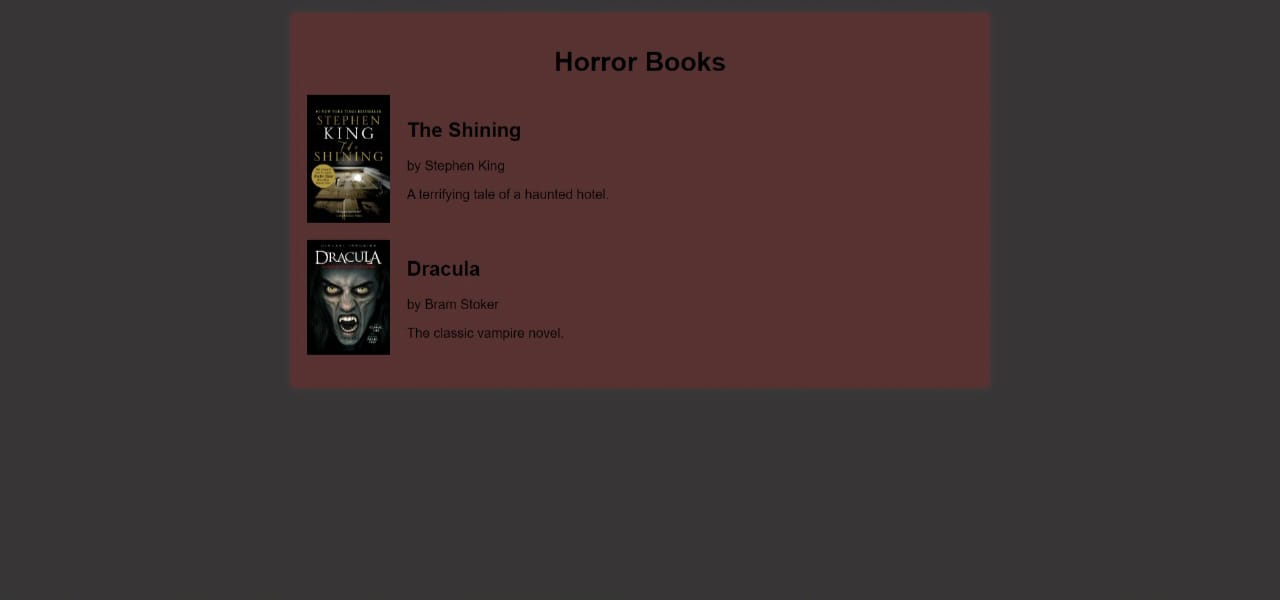


Figure 6: Horror books Page

The horror page of an online bookstore is designed to immerse readers in the thrilling and chilling world of horror literature.

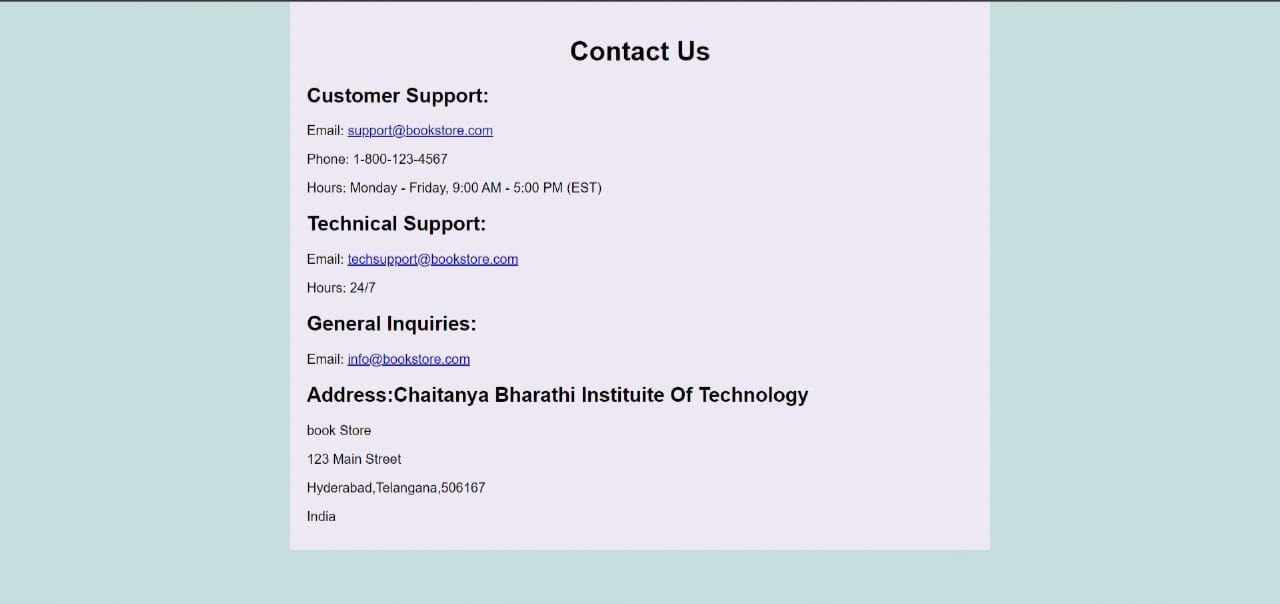


Figure 7: Contact page

The contact page of an online bookstore serves as a central hub for customer support, inquiries, and feedback.

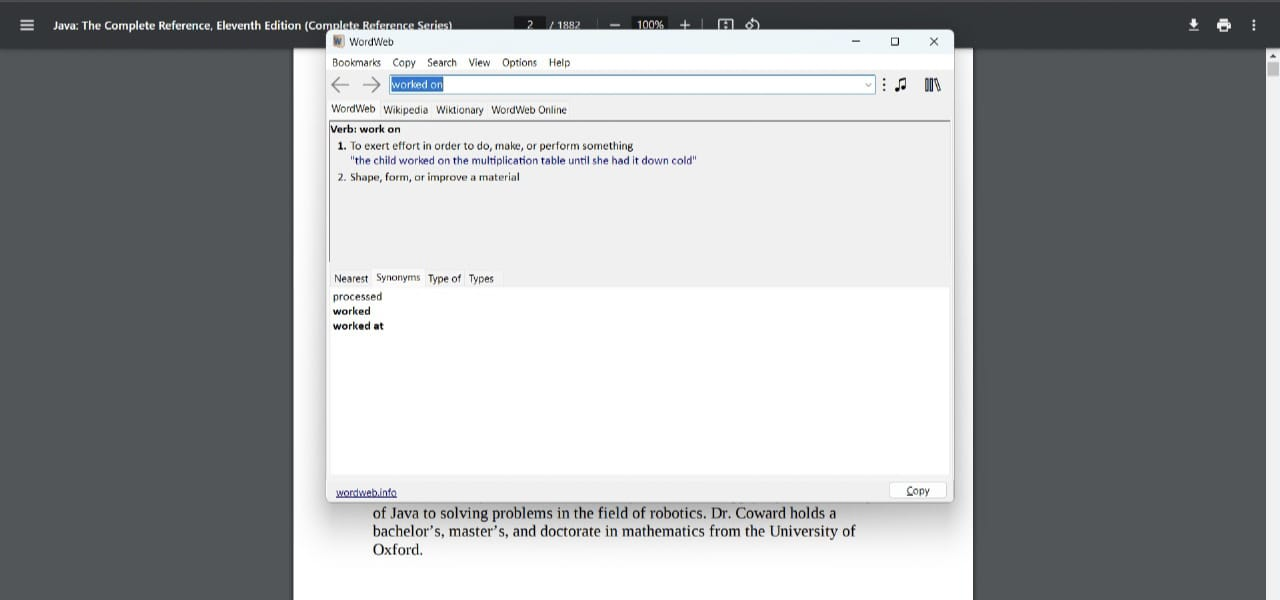


Figure 9: Info of highlighted word

When you Ctrl + right-click on a word, it typically triggers a context menu or pop-up window that provides additional options or information about the selected word.

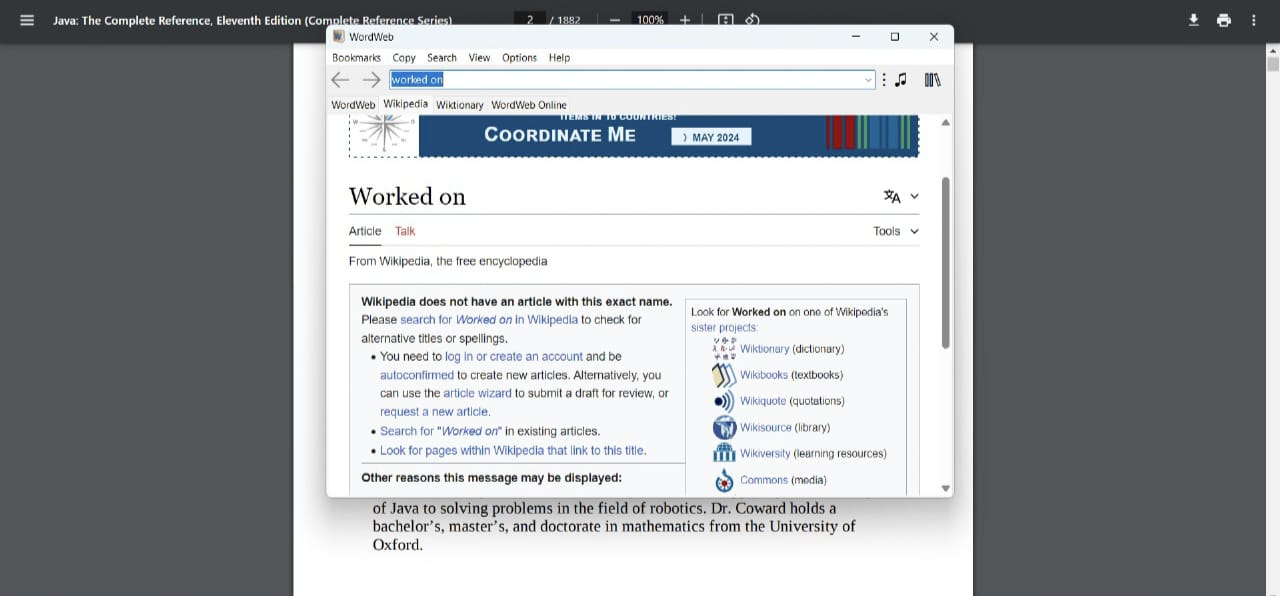


Figure 10: linking to Wikipedia

When the "Define" or "Look up" feature is integrated with Wikipedia, it expands the functionality beyond a traditional dictionary lookup

D. OBSERVATIONS

Integrating the "Look up" feature with Wikipedia offers a dynamic enhancement to the traditional dictionary lookup function. By enabling users to access a wealth of information beyond simple definitions, this integration transforms the reading experience into a journey of exploration and discovery. Through a quick Ctrl + right-click, users can seamlessly delve deeper into the meaning, usage, and cultural significance of the selected word, gaining comprehensive insights from Wikipedia articles. This interactive feature not only enriches users' understanding of the text but also fosters curiosity and engagement by providing access to a vast repository of knowledge. However, ensuring the accuracy and reliability of information retrieved from Wikipedia poses a challenge, requiring careful consideration of verification methods and potential technical limitations. Nevertheless, by empowering users with customizable options and a broader context, this integration enhances the reading experience and promotes lifelong learning.

**VI. FUTURE DIRECTIONS**

Follow-up studies could explore the long-term effects of contextual vocabulary learning on participants' language proficiency and reading habits. Tracking participants over an extended period would provide insights into the sustainability of the intervention's benefits and its impact on lifelong learning. Further research could investigate the applicability of the proposed approach to other literary genres beyond genre fiction, horror, and educational novels. Examining how contextual vocabulary learning can be tailored to different genres would broaden the scope of its potential applications and appeal to a wider range of readers. Explore the integration of technology, such as interactive e-books or mobile applications, to facilitate contextual vocabulary learning in a digital format. Leveraging multimedia features and adaptive learning algorithms could enhance the effectiveness and accessibility of the intervention for diverse learners. Conduct cross-cultural studies to examine the universality of the proposed approach across different linguistic and cultural contexts. Investigating how contextual vocabulary learning operates in diverse cultural settings would provide valuable insights into its generalizability and adaptability. Develop teacher training programs that incorporate the principles of contextual vocabulary learning into language instruction curricula. Equipping educators with the knowledge and skills to implement effective vocabulary teaching strategies can have a multiplier effect on student learning outcomes. Continuously refine and optimize the intervention based on feedback from participants and emerging research findings. Iterative improvement cycles will ensure that the intervention remains responsive to the evolving needs and preferences of learners.

**VII. CONCLUSION**

The proposed project aims to investigate the effectiveness of integrating contextual vocabulary learning into genre fiction, horror, and educational novels. By highlighting and defining words within their narrative contexts, the project seeks to enhance language proficiency, reading comprehension, and overall engagement with the text. Through a rigorous experimental design and comprehensive data analysis, the project has provided valuable insights into the potential benefits of this innovative approach to vocabulary acquisition. This suggests that integrating vocabulary instruction with leisure reading can be an effective strategy for enhancing vocabulary acquisition and promoting deeper understanding of literary texts.

This project represents a significant step towards advancing our understanding of effective vocabulary acquisition strategies within the context of leisure reading. By building on the findings and exploring future research directions, educators and researchers can continue to innovate and enhance literacy instruction for the benefit of learners worldwide.

##### **REFERENCES**

[1] Word Web Dictionary Official Site-The official website for Word Web, a powerful and comprehensive English dictionary and thesaurus software that can be used both online and offline.

[2] Word Web for Windows Features-Overview of the features offered by the Word Web dictionary software, including its extensive word definitions, synonyms, antonyms, and related words.

[3] MDN Web Docs on HTML-Comprehensive resource for learning and referencing HTML, the standard markup language used to create web pages.

[4] MDN Web Docs on CSS-A detailed guide to CSS, the style sheet language used for describing the presentation of web pages.

[5] MDN Web Docs on JavaScript-Extensive documentation and tutorials for JavaScript, a key scripting language for creating dynamic web content.

[6] National Reading Panel. (2000). Report of the National Reading Panel: Retrived from <https://www1.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf>

[7] Reading Rockets. (n.d.). Vocabulary. Retrieved from

<https://www.readingrockets.org/teaching/reading101-course/modules/vocabulary/>

[8] Vocabulary.com.(n.d.).Learn Words-English Dictionary. Retrieved from <https://www.vocabulary.com/>

[9] Scholastic (n.d.). Vocabulary and Spelling. Retrieved from

<https://www.scholastic.com/teachers/teaching-tools/expert-articles/vocabulary-spelling.html>

[10] PDF.js GitHub Repository-The official GitHub repository for PDF.js, a JavaScript library that renders PDF files in web browsers.

[11] Py Mu PDF Documentation-Documentation for PyMuPDF, a Python binding for MuPDF, a lightweight PDF and XPS viewer.

[12] MDN Web Docs on Fetch API-Guide to the Fetch API, which provides an interface for fetching resources, including across the network.

[13] Wikipedia API Documentation-Official documentation for the Wikipedia API, which allows developers to access Wikipedia content programmatically.

[14] Thesaurus.com-An online thesaurus that provides synonyms and antonyms for English words, useful for expanding vocabulary and finding related terms.

[15] NLTK Documentation-Documentation for the Natural Language Toolkit (NLTK), a leading platform for building Python programs to work with human language data.

[16] MDN Web Docs on Event Handling-Comprehensive resource on handling events in JavaScript, including user interactions such as clicks and keypresses.

[17] RESTful API Design-A guide to designing RESTful APIs, which are a common way to structure and access web services.

[18] MDN Web Docs on Browser Extensions-Documentation on creating browser extensions using the Web Extensions API, which is compatible with multiple browsers.

[19] OWASP Web Security Testing Guide-A comprehensive guide to web application security testing, provided by the Open Web Application Security Project (OWASP).

[20] MySQL Documentation- Detailed documentation for MySQL, an open-source relational database management system based on SQL (Structured Query Language).