A
MINI PROJECT REPORT
on
Word Tracker
BE(IT)-IV Sem
By
J. Vinathi Reddy (160122737007)
M. Sai Varshitha (160122737015)
N. Manvika (160122737017)

Under guidance of
Mr. U. Sairam
Assistant professor
IT Department

DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO Certified 9001:2015)

KOKAPET(V), GANDIPET(M),RR District HYDERABAD - 75

Website: www.cbit.ac.in

2023-2024

CERTIFICATE

This is to certify that the project work entitled "**Word Tracker**" submitted to CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY, in partial fulfillment of the requirements for the completion of Mini Project-I of IV Semester B.E. in Information Technology, during the Academic Year 2023-2024, is a record of original work done by J. Vinathi Reddy(160122737007), M. Sai Varshitha (160122737015) and N. Manvika(160122737017) during the period of study in the Department of IT, CBIT, HYDERABAD, under our guidance.

Project Guide   **Head of the Department**

Mr. U. Sairam   **Dr. Rajnikanth Aluvalu**

Assistant Professor, Dept. of IT,   Professor, Dept. of IT,

CBIT, Hyderabad.   CBIT, Hyderabad.

ACKNOWLEDGEMENTS

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.

The online book store caters specifically to the academic needs of CBIT students, housing an extensive repository of textbooks covering diverse subjects ranging from engineering to humanities. With the convenience of digital access, students can easily navigate through their required course materials at any time and from anywhere. This digital platform not only mitigates the logistical challenges associated with physical textbooks but also offers an interactive reading experience.

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 double-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.

By amalgamating accessibility, convenience, and enhanced learning tools, this online book store seeks to empower CBIT students in their academic pursuits. Through seamless access to educational resources and innovative features like the exploration tool, this platform endeavors to cultivate a conducive learning environment conducive to academic excellence.

CONTENTS

LIST OF FIGURES

INTRODUCTION

i. Motivation

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.

ii. Problem statement

WORD TRACKER

This project addresses the challenge of vocabulary acquisition by integrating contextual learning into genre fiction, horror, and educational novels. It aims to enhance comprehension and retention by highlighting and defining words within their narrative contexts. The effectiveness of this approach across different genres and reader demographics is underexplored. By investigating its impact on language proficiency, cognitive skills, and reading experience, the project seeks to provide insights into effective pedagogical strategies for promoting lifelong learning and literacy development.

EXISTING SYSTEM

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.

PROPOSED METHODOLOGY

SYSTEM SPECIFICATIONS

i. Operating System

The Chrome extension should be compatible with various operating systems, including Windows, macOS, and Linux, as long as they support the Chrome browser.

ii. Chrome Browser Version

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.

iii. Hardware Requirements

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.

iv. Memory Usage

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.

v. Internet Connection

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.

vi. PDF Viewer Compatibility

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.

vii. User Interface

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.

viii. Security

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.

ix. Updates and Maintenance

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.

x. Documentation and Support

Provide comprehensive documentation and support resources for users to understand how to use the extension effectively and troubleshoot any issues they encounter.

SYSTEM DESIGN

i. User Interface

- The UI will consist of the Chrome extension interface accessible from the browser toolbar.
- The home 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.

ii. Backend
- 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.

iii. PDF Viewer
- 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.

iv. Word Tracking Feature
- 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.

v. Integration with Wikipedia
- The extension 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.

vi. Data Storage
- 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.

vii. Security
- 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.

viii. User preferences and Settings
- 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.

ix. Error Handling and Logging
- 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.

x. Testing and Quality Assurance
- 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.

LITERATURE SURVEY

Research on online bookstores extensively explores user behavior, preferences, and satisfaction, focusing on factors influencing purchase decisions such as price, selection, convenience, and overall user experience. [1]Studies highlight the significant impact of features like personalized recommendations, user reviews, and ease of navigation on user engagement and satisfaction. In the realm of natural language processing (NLP) and word lookup, the literature delves into techniques for word sense disambiguation and dictionary lookup, discussing various algorithms and methods for identifying word meanings based on context. Challenges such as polysemy and the effectiveness of different NLP models, including rule-based systems, statistical methods, and machine learning approaches, are also examined. [1]The field of user interaction and experience design offers insights into best practices for creating intuitive and user-friendly interfaces for online bookstores. Research covers topics like information architecture, navigation design, visual hierarchy, and feedback mechanisms, all aimed at enhancing user engagement and satisfaction. Additionally, studies investigate the impact of interactive features, such as word meaning lookup, on the perceived usability and utility of these platforms. [2] In e-commerce and information retrieval, research focuses on improving search and recommendation systems in online marketplaces. Techniques for enhancing search relevance, personalization, and recommendation accuracy are explored to help users discover relevant books more effectively. [3]The role of contextual information, user feedback, and social factors in influencing information-seeking and decision-making in e-commerce environments is also a key area of study. Finally, with the increasing digitization of personal data and online transactions, privacy and security in e-commerce platforms are crucial areas of research.[4] Studies address privacy concerns related to user data collection, tracking, and targeted advertising in online bookstores, along with security issues such as data breaches, fraud prevention, and secure payment mechanisms to protect users' sensitive information

DATASETS and ALGORITHMS/ MODELS

DATASETS

i. Word meanings and synonyms

I 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 for Downloading word web 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 systems
- Python
- Electronic devices and circuit theory
- Electronic devices
- Introduction to mechanical engineering
- Thermodynamics

ALGORITHMS OR MODELS
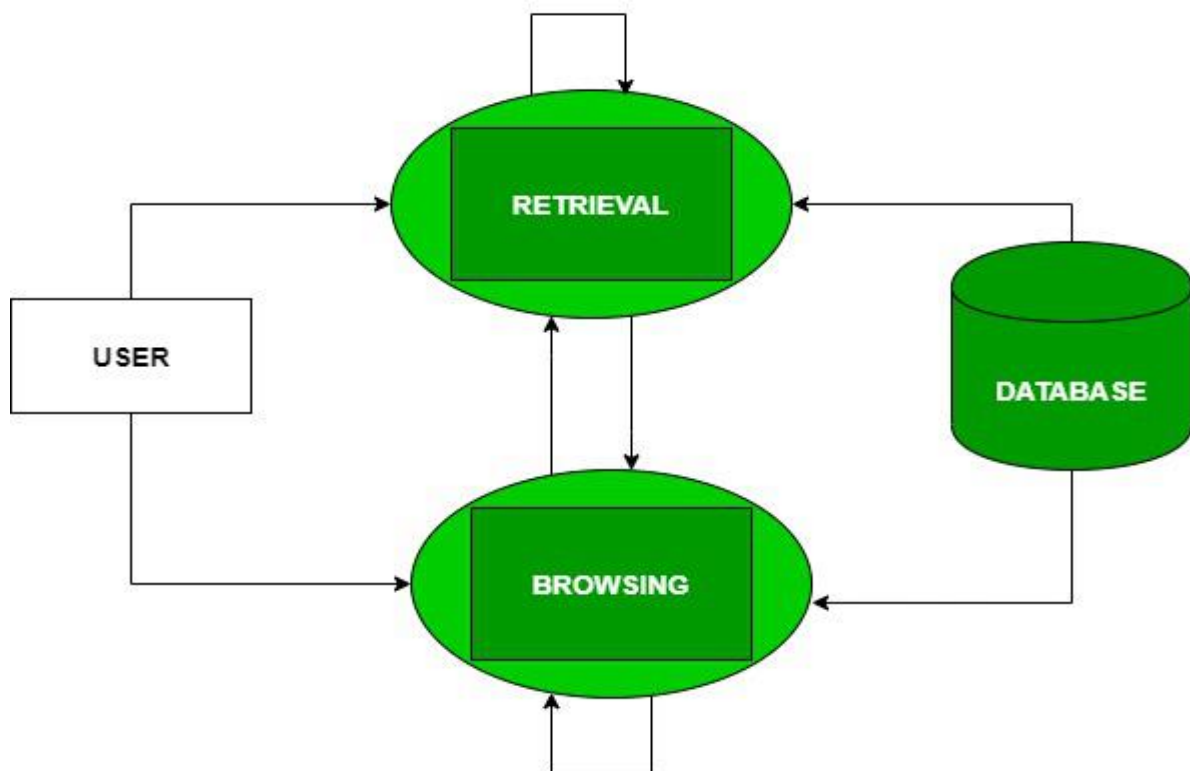
Figure 1: Flow chart

ALGORITHM

The "Word Tracker" project Algorithm is:

Step1: 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.

Step2: 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.

Step3: 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.

Step4: 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.

Step5: PDF Parsing

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

Step6: 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.

IMPLEMENTATION AND RESULTS

IMPLEMENTATION

Implementing a project like "Word Tracker" involves several steps. Here's a high-level outline of how you can proceed with the implementation:

i. Set Up Development Environment
- Install necessary tools and libraries for developing a Chrome extension.
- Set up a development environment for backend services, if applicable.

ii. Design User Interface
- Design the UI for the Chrome extension, including the home page, genre selection, PDF viewer, and word tracking features.
- Consider usability and accessibility principles while designing the UI.

iii. Implement Chrome Extension
- Create the basic structure of the Chrome extension, including manifest.json, background scripts, content scripts, and popup pages.
- Implement functionality for genre selection, fetching PDF books, and displaying them in the PDF viewer.
- Integrate word tracking feature using content scripts to capture selected words and communicate with backend services.

iv. Integrate PDF Viewer
- Choose a PDF viewer library or implement a custom PDF viewer using HTML, CSS, and JavaScript.
- Ensure the PDF viewer supports features like scrolling, zooming, and text selection.

v. Implement Word Tracking Feature
- Set up communication between the content script and the background script to handle word tracking events.
- Implement logic to extract selected words from the PDF document and send them to the backend for processing.
- Retrieve word meanings, synonyms, and related information from external sources (e.g., WordNet, Wikipedia) and display them to the user.

vi. Implement Backend Services (if necessary)
- Set up a backend server to handle requests for fetching word meanings, synonyms, and additional information from external sources.
- Implement APIs to communicate with external services and process word tracking requests from the Chrome extension.

vii. Data Storage and Management
- Implement logic for storing metadata about genres, books, and user preferences within the Chrome extension.
- If necessary, set up a database or file storage system to manage book data and user interactions.

viii. Testing and Debugging
- Test the Chrome extension thoroughly to ensure all features work as expected across different browsers and platforms.
- Debug any issues that arise during testing and make necessary adjustments to improve stability and performance.

ix. User Documentation and Feedback
- Create user documentation explaining how to use the Chrome extension and its features. Gather feedback from users during beta testing and make improvements based on their suggestions and bug reports.

x. Deployment
- Once the Chrome extension is stable and ready for release, publish it to the Chrome Web Store for public distribution.
- Monitor usage metrics and user feedback after deployment to identify areas for further improvement.

Throughout the implementation process, remember to adhere to best practices for code quality, security, and user privacy. Regularly update the extension to address any issues or add new features based on user feedback and emerging requirements.

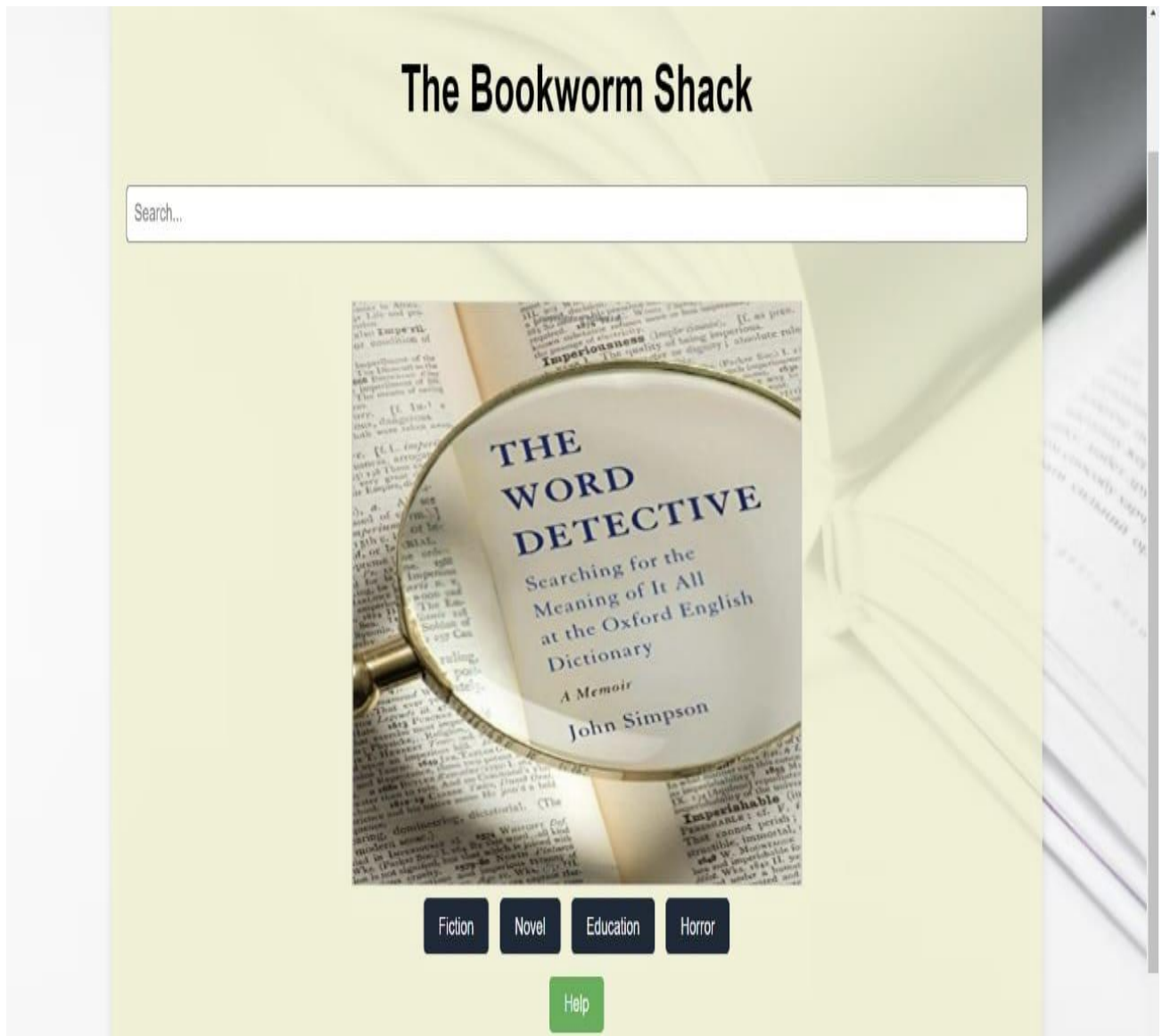# RESULTS

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.
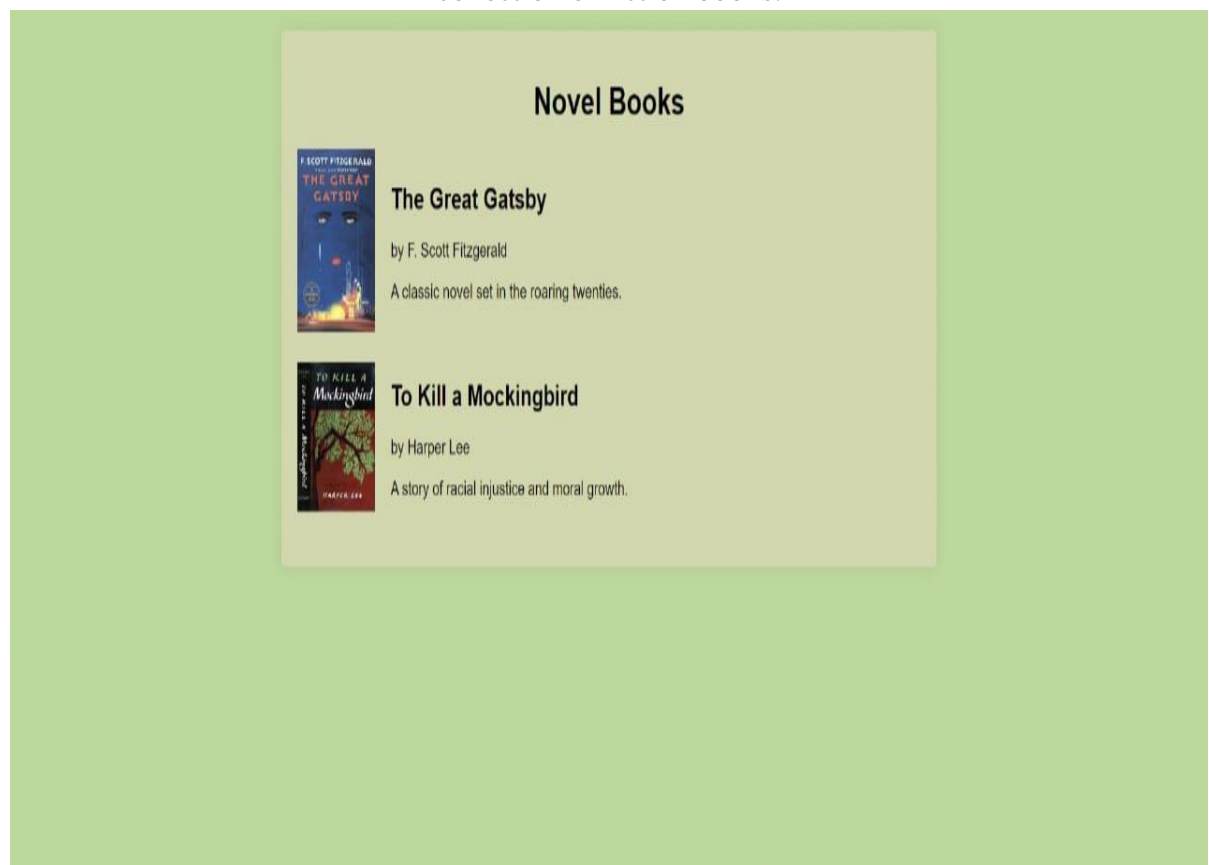
Figure 4: Novel Books page

The novel page of an online bookstore is dedicated to presenting a curated collection of fiction novels.
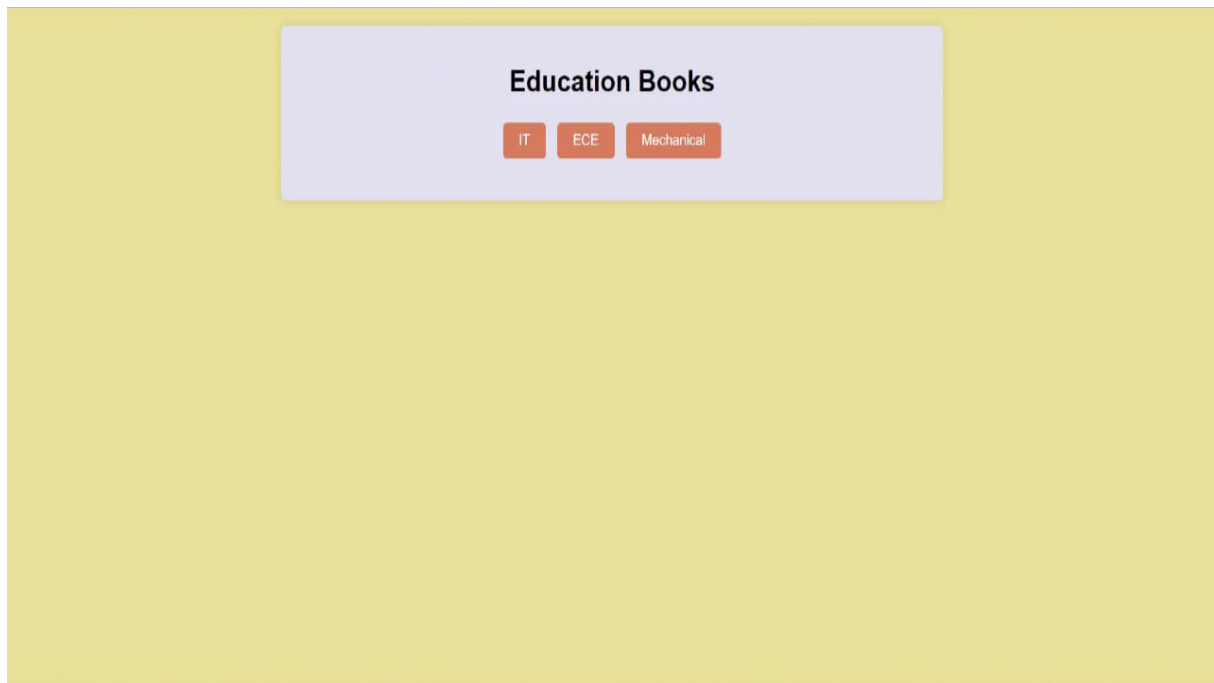
Figure 5: Education books page
The education page of an online bookstore dedicated to educational books is designed to serve educators, students, and lifelong learners.
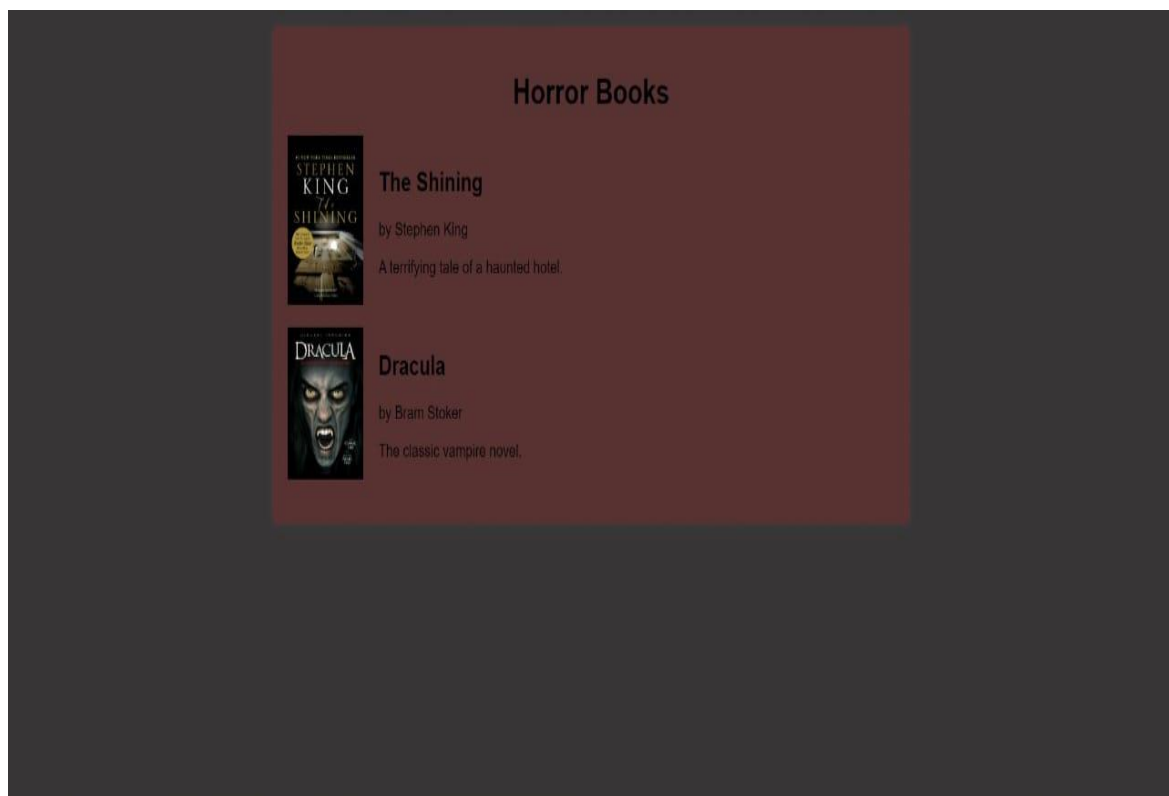
Figure 6: Horror books Page
The horror page of an online bookstore is designed to immerse readers in the thrilling and chilling world of horror literature.

Figure 7: Contact page
The contact page of an online bookstore serves as a central hub for customer support,
inquiries, and feedback.

About the Author

Best-selling author **Herbert Schildt** has written extensively about programming for over three decades and is a leading authority on the Java language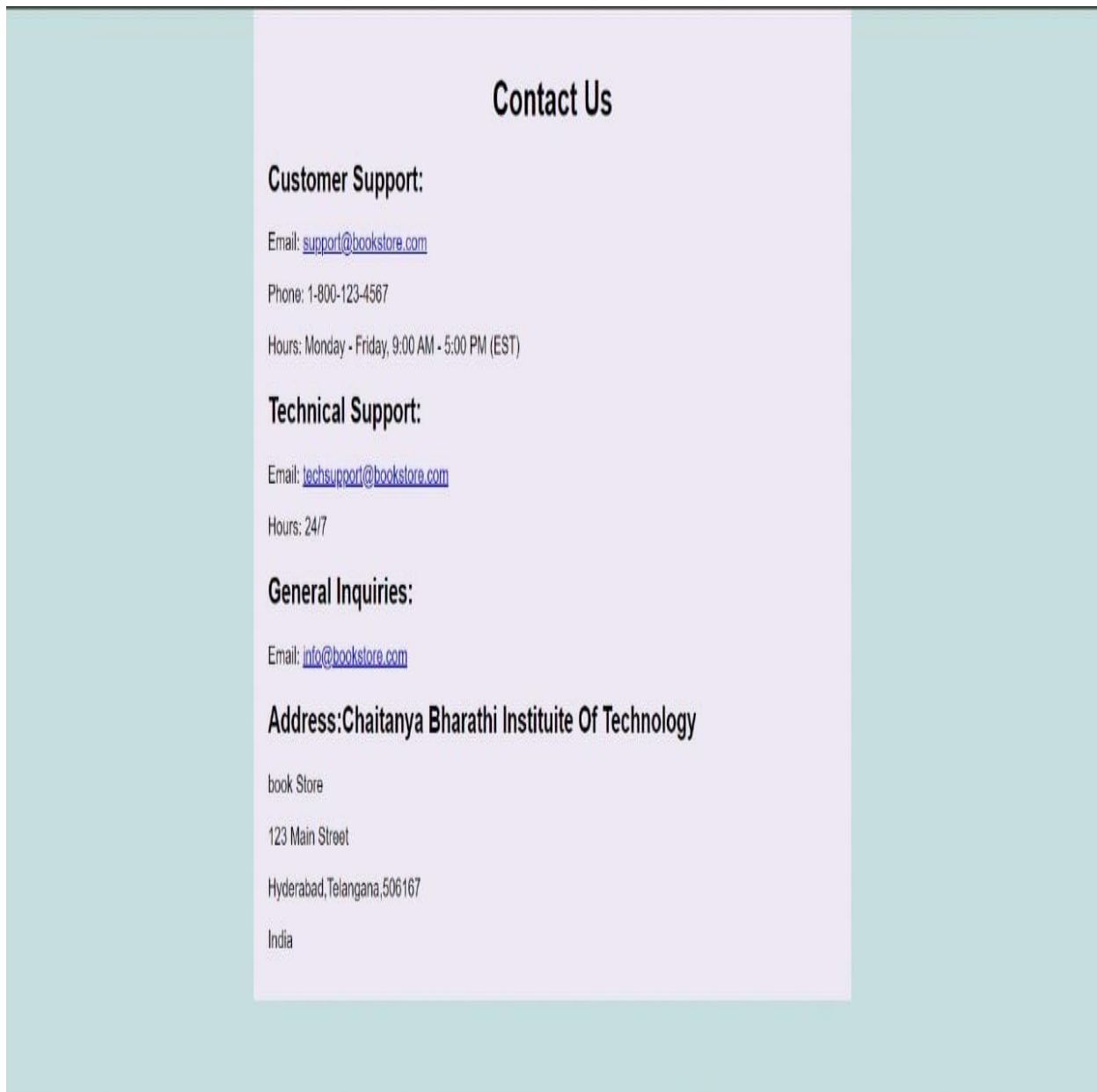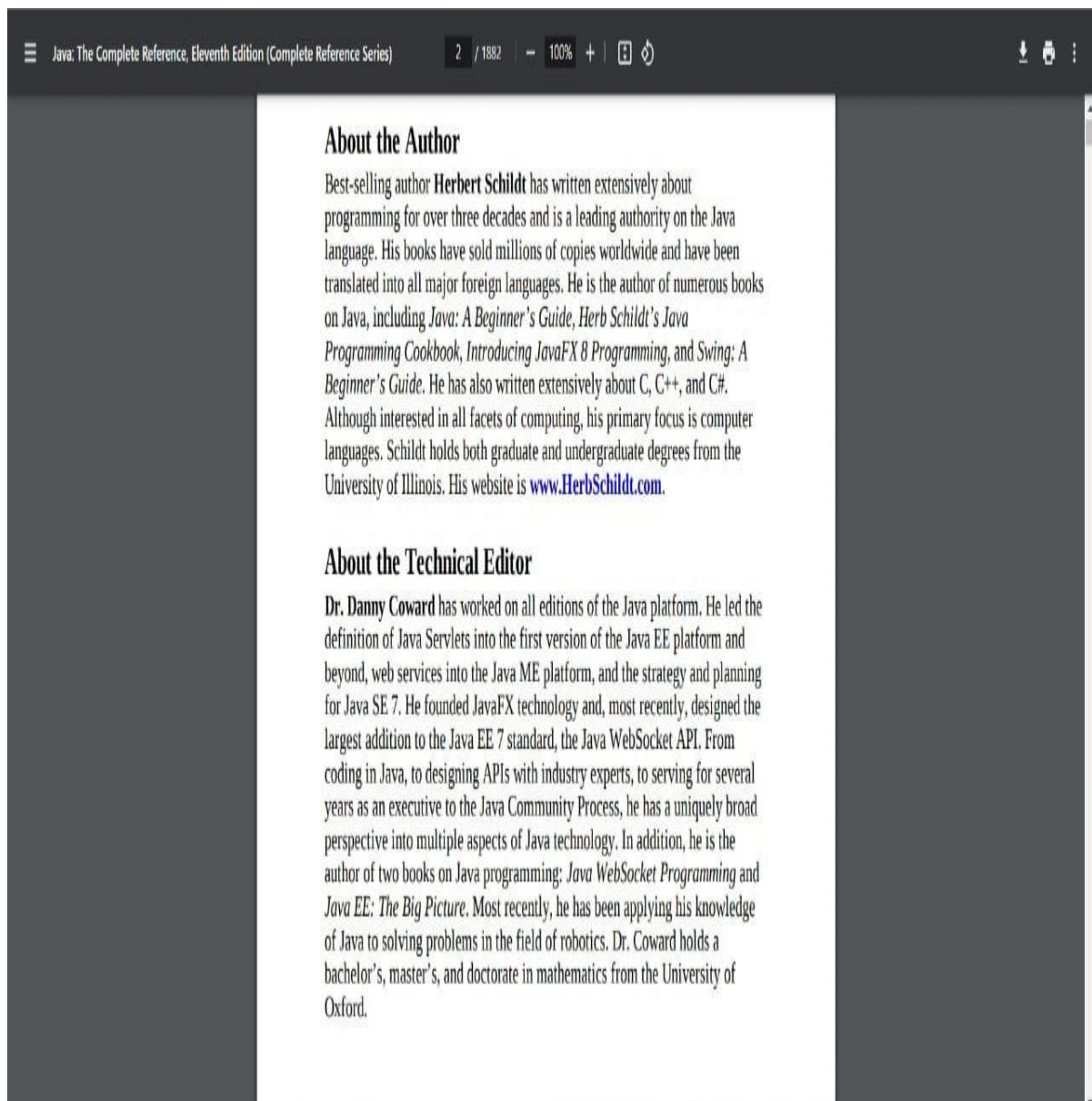. His books have sold millions of copies worldwide and have been translated into all major foreign languages. He is the author of numerous books on Java, including *Java: A Beginner's Guide, Herb Schildt's Java Programming Cookbook, Introducing JavaFX 8 Programming*, and *Swing: A Beginner's Guide*. He has also written extensively about C, C++, and C#. Although interested in all facets of computing, his primary focus is computer languages. Schildt holds both graduate and undergraduate degrees from the University of Illinois. His website is www.HerbSchildt.com.

About the Technical Editor

Dr. Danny Coward has worked on all editions of the Java platform. He led the definition of Java Servlets into the first version of the Java EE platform and beyond, web services into the Java ME platform, and the strategy and planning for Java SE 7. He founded JavaFX technology and, most recently, designed the largest addition to the Java EE 7 standard, the Java WebSocket API. From coding in Java, to designing APIs with industry experts, to serving for several years as an executive to the Java Community Process, he has a uniquely broad perspective into multiple aspects of Java technology. In addition, he is the author of two books on Java programming: *Java WebSocket Programming* and *Java EE: The Big Picture*. Most recently, he has been applying his knowledge of Java to solving problems in the field of robotics. Dr. Coward holds a bachelor's, master's, and doctorate in mathematics from the University of Oxford.

Figure 8: JAVA Book PDF
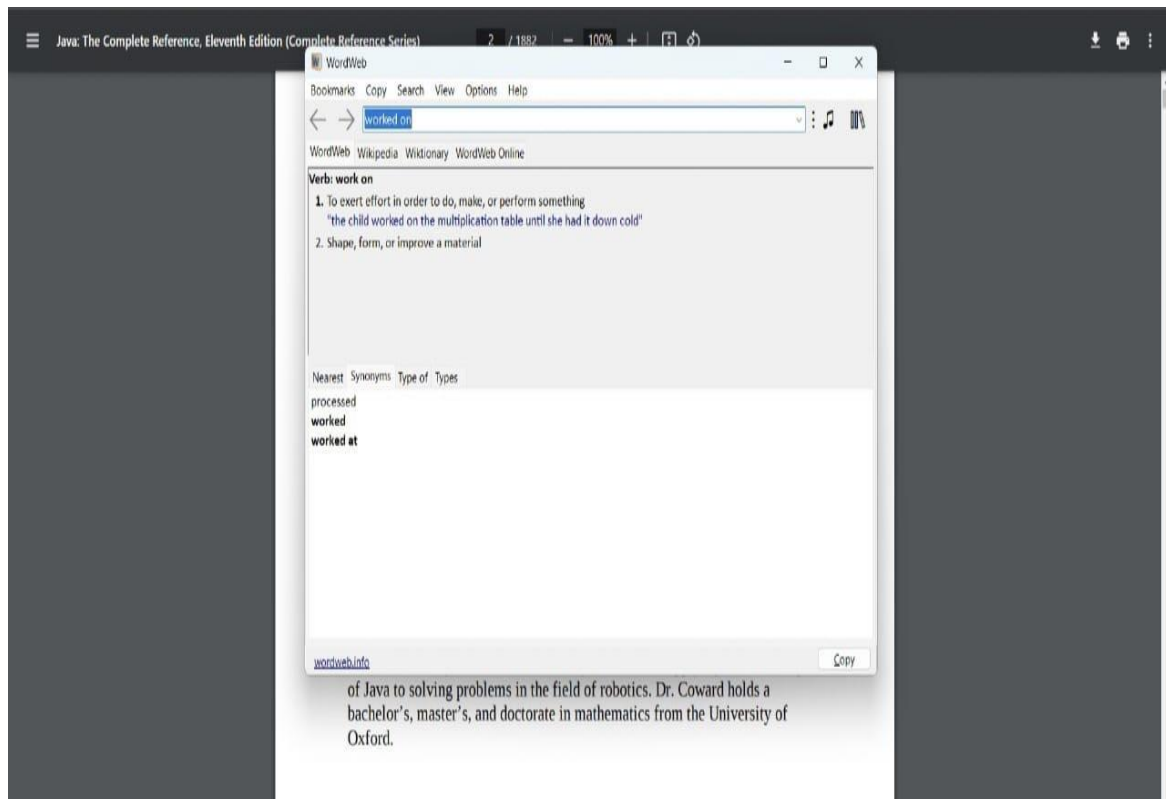Opening a book to search for a unknown word.

16

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.
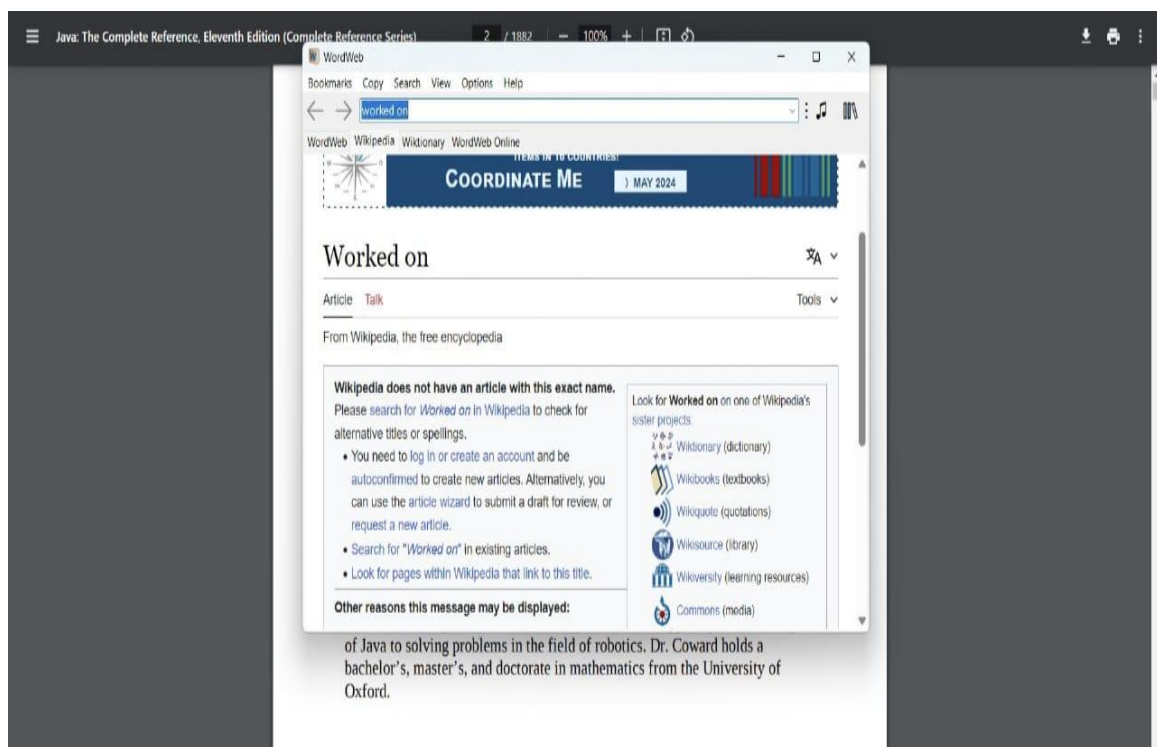

Figure 10: linking to Wikipedia

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

CONCLUSION and FUTURE SCOPE

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.

The findings of the study indicate that participants in the experimental group, who received highlighted novels and engaged in contextual vocabulary learning activities, demonstrated significant improvements in their language proficiency and reading comprehension compared to the control group. 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.

In conclusion, the proposed 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.

Future scope

i. Long-term Effects

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.

ii. Adaptation for different genres

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.

iii. Technology Integration

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.

iv. Cross Cultural Studies

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.

v. Teacher Training Programs

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.

vi. Intervention Optimization

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.

REFERENCES

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

[2] Reading Rockets. (n.d.). Vocabulary. Retrieved from
https://www.readingrockets.org/teaching/reading101-course/modules/vocabulary/

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

[4] Scholastic.(n.d.). Vocabulary and Spelling. Retrieved from
https://www.scholastic.com/teachers/teaching-tools/expert-articles/vocabulary-spelling.html