Analego Explorer: A Chrome Extension for Analego

Yang Zhou*
Ao Shen*
Zepei Li*
Jialuo He*
yz96@illinois.edu
aos5@illinois.edu
zepeili2@illinois.edu
jialuoh2@illinois.edu
University of Illinois Urbana-Champaign
Urbana, Illinois, USA

Abstract

Analego and its Chrome extension, Analego-Explorer, are innovative tools that leverage Large Language Models (LLMs) to enhance educational experiences through analogies. Analego allows users to generate and modify analogies using OpenAI's ChatGPT, storing them in a searchable repository. The Analego-Explorer extension extends this functionality directly into web browsing, providing users with instant access to relevant analogies based on the text they select on web pages. This integration facilitates a seamless educational interaction by employing content-aware searches that adapt to the context of the user's current reading, significantly enhancing usability and learning engagement. Our evaluations demonstrate the effectiveness of Analego-Explorer in delivering contextually relevant and accurate analogies, highlighting its potential to transform educational practices through advanced technology and user-centric design.

ACM Reference Format:
Yang Zhou, Ao Shen, Zepei Li, and Jialuo He. 2024. Analego Explorer: A Chrome Extension for Analego. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Analogies, which connect less familiar concepts (referred to as the target) to more familiar ones (known as the source), play a crucial role in education. They assist in understanding new ideas, solving problems, and boosting learners' interest and motivation, as noted by researchers like Thagard (1992), Novick and Holyoak (1991), and Glynn et al. (1989).

Analego (Bhavya et al. 2024) is a system where users (learners and teachers) can collaboratively generate useful analogies by leveraging Large Language Models (LLMs). The system features two collaborative innovative functions: (1) Users can create analogies for a given target concept by using appropriate prompts with the OpenAI ChatGPT model, as noted by Bhavya, Xiong, and Zhai (2022). Additionally, users have the option to select a source domain, which is a topic they are interested in, to shape the context of the analogy. Once generated, these analogies can be modified and added to our backend database. (2) Users can explore this growing database to find analogies that suit their needs, thanks to the implementation with Elasticsearch. The system also allows users to influence the quality and relevance of analogies by upvoting or downvoting them, thus helping improve the search engine's ability to prioritize analogies based on community feedback over time.

In order to further enhance the usability and user experience of Analego, we have developed an extension based on the Chrome browser: Analog-Explorer. Analego-Explorer inherits the complete database and powerful search capabilities of the Analego system, and adds advanced features such as content-aware, allowing users to select text content on the webpage through keyword selection while browsing, thereby directly obtaining relevant analogies of the selected text on the webpage without the need to jump to the official website of Analego for search, greatly reducing the process and time of searching for analogies and improving the user experience.

2 Related Work

LLMs have recently been used to assist with creating educational content such as questions, explanations, and grading(Moore et al. 2023; MacNeil et al. 2022), as well as generating analogies (Bhavya, Xiong, and Zhai 2022, 2023; Kim et al. 2023). Analego (Bhavya et al. 2024) is a system that

*Both authors contributed equally to this research.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/nnnnnnn.nnnnnnn

enables students and teachers to benefit from the new technology of using LLMs to generate analogies. Our Analego-Explorer has proposed a new feature that is more suitable for web browsing based on Analego. As a Chrome extension, Analego-Explorer is a supplement and extension to the web version of Analego.

3 Motivation

The motivation behind developing this Chrome extension stemmed from a desire to enhance users' browsing experience by providing them with instant access to relevant information while they navigate the web. Recognizing the inconvenience users face when having to switch between multiple tabs or search engines to find additional context on a topic, this extension aims to streamline the research process. By seamlessly integrating Wikipedia links and offering suggested terms within the browser interface, users can efficiently explore related concepts without disrupting their workflow. This project not only addresses a common user pain point but also serves as an exploration of innovative solutions to enhance browsing efficiency and information accessibility.

4 Development

4.1 Main Functionalities

4.1.1 Extension Popup Interface (Control Panel). The Extension Popup Interface of the Analego Explorer Chrome Extension serves as a crucial user interaction hub, allowing users to manage the operational state of the extension through two main toggle switches. This interface is crafted to enhance user autonomy by providing immediate access to the extension's features without necessitating navigation away from the user's current webpage. The functionality of each toggle is as follows:

General Toggle Switch: "Turn On/Off". This primary toggle is strategically positioned to facilitate easy access, enabling users to quickly activate or deactivate the extension's functionalities:

- **"Turn On"**: Engaging this toggle activates the extension, making it responsive to text selections on the webpage. This state allows the extension to prepare to fetch and display contextually relevant data based on the user's interactions with text, ensuring readiness to provide information instantaneously.
- **"Turn Off"**: This option deactivates the extension, rendering its features inactive and ensuring that it does not interfere with other browser activities or consume system resources unnecessarily. This toggle is particularly useful when a clean browsing experience is required or to prevent potential conflicts with other extensions or tasks.

Advanced Toggle Switch: "Content Aware". Located beneath the general toggle, the "Content Aware" switch allows users to engage a more sophisticated analysis of selected text, enhancing the extension's utility by offering more precise information retrieval:

- **"Turn On"**: Activation of this feature extends the capabilities of the extension beyond mere text matching. It involves analyzing the context surrounding the selected text using natural language processing algorithms to extract relevant topics and themes. This enriched understanding aids in refining the backend queries, thereby enhancing the specificity and relevance of the information returned.
- **"Turn Off"**: Deactivating the content-aware feature restricts the extension to consider only the text explicitly selected by the user. This mode is less demanding on system resources and suitable for straightforward queries where contextual depth is unnecessary.

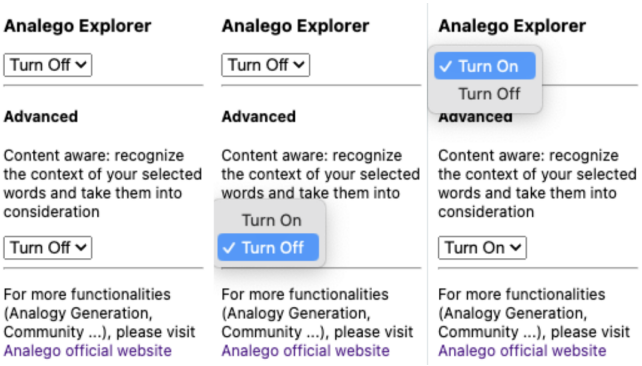

Figure 1. Extension Popup Interface (Control Panel)

4.1.2 Search and Display Panel. The Analego Explorer Chrome Extension is equipped with a powerful "Search and Display" panel, designed to enhance educational experiences directly from a user's browser. This panel actively interacts with a backend API to fetch and display analogies pertinent to terms selected by the user on any webpage.

Operational Workflow. When a user selects a keyword or phrase within the text of a webpage, the extension recognizes this action and triggers a query to the backend API. This API is specifically tailored to search a comprehensive database for analogies that relate directly to the selected word or phrase. The response from the API includes not only the analogies themselves but also contextual prompts that facilitate deeper understanding and learning.

Display of Analogies. The returned analogies are displayed in a user-friendly popup panel adjacent to the text selection. This panel is organized to present each analogy with a corresponding prompt that frames the analogy in a question or explanatory format. For example:

- If the selected word is "nucleus," the panel might display: "Nucleus is to atom as ____ is to cell. Organelles are to cells as atoms are to molecules."
- This format encourages the user to think critically about the analogy, enhancing comprehension and retention of the concept.

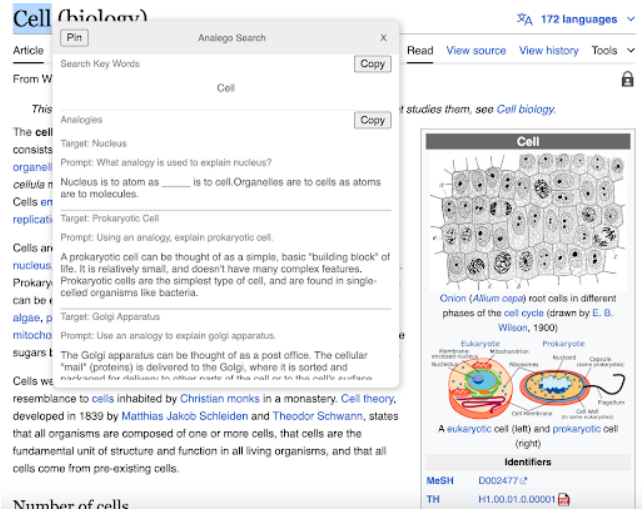

Figure 2. Display of Analogies

Enhanced Learning through Prompts. Each analogy is paired with a prompt designed to engage the user further and deepen their understanding. These prompts vary in format, offering explanations, comparisons, or direct questions that require the user to reflect on the analogy and its application to the selected term. This method serves not only to provide immediate information but also to stimulate cognitive processes that are essential for learning.

Content Aware Functionality. The "Content Aware" feature of the Analego Explorer Chrome Extension represents a sophisticated enhancement to the basic search functionality. When activated, this feature leverages advanced natural language processing (NLP) techniques to conduct a comprehensive analysis of the entire context surrounding the selected keyword. This not only includes the keyword itself but extends to the surrounding text within the paragraph or section, thus capturing a broader spectrum of relevant content for analysis.

- **Context Extraction**: Rather than focusing solely on the selected keyword, this feature dynamically extracts the full textual context in which the keyword is situated. This extraction encompasses not only the immediate text surrounding the keyword but also ancillary metadata, significantly enriching the dataset available for subsequent analysis. This comprehensive approach ensures that the resulting insights are deeply rooted in the actual textual environment of the keyword, providing a robust basis for the generation of relevant analogies and prompts.
- **Topic Extraction via NLP**: Utilizing state-of-the-art NLP techniques, the extension analyzes the extracted text to identify and distill the main topics and themes present. This process involves parsing the text to discern linguistic patterns and thematic elements that characterize the content. The extracted topics are then used to refine and tailor the queries sent to the backend, ensuring that the analogies and learning prompts generated are not only relevant to the selected keyword but also closely aligned with the broader context and thematic nuances of the source material. This targeted approach enhances the relevance and applicability of the educational content delivered to the user, fostering a more insightful and meaningful learning experience. The specifics of the topic extraction process, particularly the use of Latent Dirichlet Allocation (LDA), are discussed in detail in Section 4.2.

This enhanced functionality embodies the extension's commitment to delivering highly personalized and contextually appropriate educational content. By intelligently integrating detailed contextual analysis with user-interaction data, the "Content Aware" feature ensures that each user encounter is both informative and uniquely suited to the textual nuances of their specific reading material.

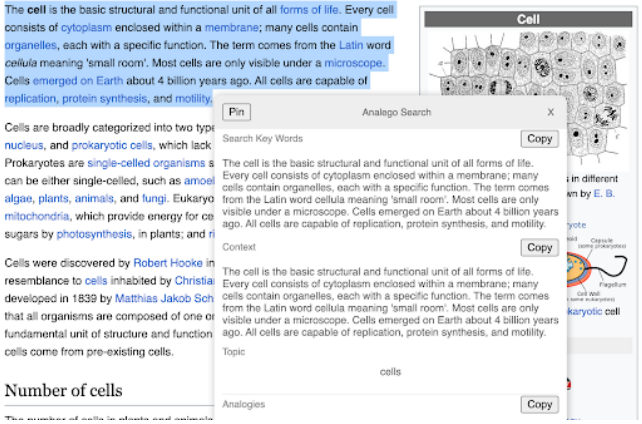

Figure 3. Content Aware Functionality

Interactive Features. The Search and Display panel is designed with several interactive features to enhance usability and learning:

- **Scrollable Display**: Allows users to scroll through multiple analogies if more than one is available, facilitating extensive exploration without leaving the current webpage context.
- **Draggable and Pinnable Interface**: Features drag-and-drop functionality, enabling users to move the

panel around the screen for optimal placement. Users can reposition the display panel across their screen to suit their viewing preferences and the panel can be pinned to the screen, allowing users to keep the information visible while they continue to browse.
- **Copy Functionality**: Incorporates a copy button that enables users to easily transfer the displayed analogies to their clipboard, supporting further educational or research activities outside the extension environment.

Implementation Challenges and Solutions. Implementing the Main Functionality Panel involved overcoming several technical challenges:

- **Real-Time Data Handling**: Ensuring that data fetched from the backend API is rendered almost instantaneously required optimizing both the network calls and the data rendering processes to minimize latency.
- **Accuracy in Context Analysis**: Developing robust NLP algorithms capable of accurately determining the context and extracting relevant topics from unstructured text demanded extensive testing and refinement to align with diverse linguistic structures and idioms encountered across different web pages.

4.2 Topic Extraction with Latent Dirichlet Allocation (LDA)

The "Content Aware" feature of the Analego Explorer Chrome Extension employs Latent Dirichlet Allocation (LDA) to perform sophisticated topic extraction from the text surrounding selected keywords. LDA is a generative probabilistic model that is particularly suited for identifying latent topics within a corpus of text, making it ideal for the dynamic and varied content encountered on the web.

1. **Text Preprocessing**: Prior to applying LDA, the text is preprocessed to optimize it for topic modeling. This involves:
   - **Tokenization**: Segmenting the text into individual words or terms.
   - **Normalization**: Standardizing text by converting all characters to lowercase and removing punctuation and numbers.
   - **Removing Stop Words**: Filtering out common words (such as "and", "the", etc.) that do not contribute to the thematic content of the text.
   - **Stemming**: Reducing words to their root form to ensure that variations of a word are processed as a single term.
2. **Vectorization**: The preprocessed text is converted into a document-term matrix, which LDA uses to discover the distribution of topics. This step typically involves counting the frequency of each term in the document, resulting in a sparse matrix where rows represent documents and columns represent terms.
3. **Applying LDA**: LDA is applied to the document-term matrix with a predefined number of topics set based on the complexity and diversity of the text. LDA assumes each document (a block of surrounding text in our case) is a mixture of topics, where each topic is characterized by a distribution over all known words. The algorithm iterates through each word of each document to assign it to one of the topics, based on the probability distribution of topics in the document and the probability distribution of terms in the topics.
   - **Hyperparameters**: Key hyperparameters include:
     - **Alpha**: Represents document-topic density — with a higher alpha, documents are composed of more topics, and with lower alpha, documents contain fewer topics.
     - **Beta**: Represents topic-word density — with a higher beta, topics are made up of most of the words in the corpus, and with a lower beta, they are made up of fewer words.
4. **Topic Inference and Selection**: After multiple iterations, LDA finalizes the topic probabilities for each word and document. Topics are then inferred by examining the distribution of words across topics. The dominant topics related to the selected keyword and its surrounding context are identified based on their relevance and prevalence.
5. **Refinement and Output**: Finally, the topics most closely aligned with the central themes of the text are selected to enhance the query sent to the backend. This results in more targeted and relevant analogies and prompts that are contextually tied to the user's current text selection.

4.3 Backend Architecture

The backend of Analego-Explorer is generally based on Analego micro-service architecture. We have further optimized and improved the backend based on the Analego backend, combined with the characteristics of Analego-Explorer (such as content-aware advanced feature).

There are three main micro-services in the backend system. The architecture of the micro-service system is shown in Figure 4. Different micro-services communicate and interact with each other through message queues such as RabbitMQ or Redis. The architecture of micro-services ensures the independence and low coupling between modules, ensuring the scalability and availability of the system.

4.3.1 Auth System. The Auth System is mainly used to store user information and logs, providing functions such as user identity authentication and querying user historical behavior.

4.3.2 Search System. The Search system is mainly connected to the Elasticsearch database, providing users with

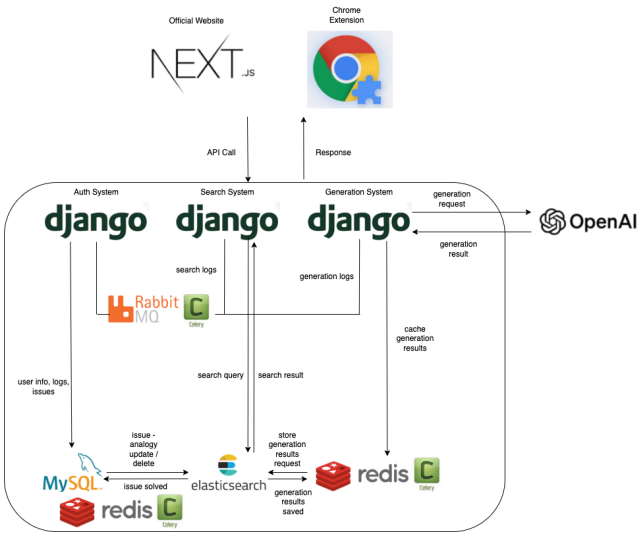

Figure 4. Backend Architecture

an analysis search service, and is also the main system that interacts with the Analog-Explorer Chrome extension.

The analogies were scraped over 10 educational websites to gather science concepts taught from grades one to twelve. We collected approximately 300 concepts. This comprehensive collection of concepts is then used to interact with ChatGPT. The task of the AI is to generate analogies that explain these scientific concepts. Each analogy created by ChatGPT includes a source concept (something familiar) and a target concept (the new science concept being explained), enhancing the understanding by drawing comparisons between the two. This method utilizes both data extraction and AI-powered explanations to simplify and elucidate complex scientific ideas for educational purposes. These concepts were stored in Elasticsearch for faster and more efficient retrieval.

The words selected by users in the webpage through extension will be processed through a series of word segmentation (such as removing all stop words), and then handed over to ElasticSearch to search for relevant words in the three main fields: "target concept", "prompt", and "analogy". ElasticSearch uses a data structure called inverted index to achieve efficient text search. In this index, ElasticSearch associates each independent word with a list of all documents where that word appears. Whenever a search query is executed, ElasticSearch can directly access the list of documents containing the query word, quickly returning relevant information. This method greatly improves search efficiency and allows users to receive search feedback on the extension window in a very short amount of time, greatly enhancing the user experience.

We have designed and implemented a search API that is different from the web version based on the unique features of AnaLego-Explorer and ElasticSearch. We have introduced weight search for the Analego-Explorer content-aware advanced feature to ensure that users can obtain an analogy that is more relevant to the current context through our extension. ElasticSearch provides powerful weight search API support, so in addition to the text selected by the user, the context topic obtained through the content aware feature will also occupy a certain proportion in ElasticSearch's search, making the final returned analogy strongly correlated with the context of the selected text, allowing the user to obtain the most accurate analogy in the current context.

4.3.3 Generation System.

The Generation system mainly provides interfaces for users to interact with Large Language Model LLM (mainly ChatGPT 3.5 model). The user generates an analogy based on LLM by specifying different parameters. The main parameters are as follows:

target: The concept that you would like to generate an analogy for (e.g. cell);

src: The topic area that you would like the analogy to be about;

prompt: Select the prompt or instruction given to the model for generating the analogy;

temp: Loweing results in less ramdom completions. As randomness approaches zero, the model will become deterministic and repetitive;

max_length: The maximum number of tokens to generate shared between the prompt and the completion;

top_p: Controls diversity vai nucleus sampling: 0.5 means half of all likelihood-weighted options are considered;

freq_penalty: How much to penalize new tokens based on their frequency in the text so far. Decreases the likelihood to repeat the same line verbatim;

pres_penalty: How much to penalize new tokens based on whether they appear in the text so far. Increases the likelihood to talk about new topics;

best_of: Returns the 'best' (the one with the highest log probability per token);

5 Evaluation

The purpose of this evaluation is to assess the quality of the Analego by utilizing a combination of language model assessment and manual review to evaluate the accuracy and relevance of the provided analogies.

5.1 Language Model Evaluation

We employed GPT-4, a state-of-the-art language model, to evaluate the content-awareness of analogies generated by Analego. In each evaluation run, we provided GPT-4 with selected text and tasked it with generating an analogy. Subsequently, we assessed whether the generated analogy demonstrated content awareness. The results of ten evaluation runs are presented in Table 1. 1.

Table 1. Language Model Evaluation Results

Evaluation Run	Validity
1	Yes
2	Yes
3	Yes
4	Yes
5	Yes
6	Yes
7	No
8	Yes
9	Yes
10	Yes

5.2 Manual Review

In addition to language model evaluation, we also conducted manual assessments of the analogies provided by the extension. The reviewers (comprising our classmates) meticulously analyzed the accuracy, clarity, and relevance of the analogies. The manual review confirmed the findings of the language model evaluation, highlighting the extension's ability to deliver precise and contextually relevant analogies.

Table 2. Manual Review Ratings

Reviewer	Accuracy	Clarity	Relevance
1	5/5	4/5	5/5
2	4/5	5/5	4/5
3	5/5	5/5	5/5

6 Conclusion

The Analego Explorer Chrome Extension integrates text retrieval technologies and HCI principles to create a tool that significantly enhances the browsing experience. By applying advanced algorithms and user-centric design, the extension offers a model for future developments in web interaction tools, providing users with a powerful means to access and interact with information efficiently and effectively.

References

Novick, L. R.; and Holyoak, K. J. 1991. Mathematical problem solving by analogy. Journal of experimental psychology: Learning, memory, and cognition, 17(3): 398.

Thagard, P. 1992. Analogy, explanation, and education. Journal of Research in science Teaching, 29(6): 537–544.

Glynn, S. M.; Britton, B. K.; Semrud-Clikeman, M.; and Muth, K. D. 1989. Analogical reasoning and problem solving in science textbooks. Handbook of creativity, 383–398.

Bhavya, Yang Zhou, Shradha Sehgal, Suma Bhat, Chengxiang Zhai. Analego: Let's build analogies together! Demo presentation at the AAAI 2024 Workshop on AI for Education (AI4Ed). 2024.

Bhavya, B.; Xiong, J.; and Zhai, C. 2022. Analogy generation by prompting large language models: A case study of instructgpt. arXiv preprint arXiv:2210.04186.

Bhavya, B.; Xiong, J.; and Zhai, C. 2023. Cam: A large language model-based creative analogy mining framework. In Proceedings of the ACM Web Conference 2023, 3903–3914.