

Towards Effective Writing Partnerships Between Humans and AI

David Hyland-Wood
University of Queensland
d.hylandwood@uq.edu.au

Anna Harrison
Ephox
anna.harrison@ephox.com

Ben Kolera
Ephox
Ben.kolera@ephox.com

ABSTRACT

In western industrialised nations, we take for granted that knowledge is acquired and shared through a written medium - and yet, the actual task of writing remains a hard skill to master. In this paper, we present a blueprint for the integration of AI into web-based text editors, and show that this contributes towards simplifying the writing task for authors. Our approach utilises the textual content of a document as input into existing artificial intelligence processes, extended with an RDF knowledge graph. The supplementary material from the AI partner is presented to an author for possible inclusion in their document. Our results from unmoderated observational user tests validate the premise that cognitive technology can simplify the writing task, and show that design is critical to the adoption of cognitive technologies. We conclude with lessons learned from deploying on-line observational tests, and share opportunities for future work.

Keywords: Design, Technology, WYSIWYG, AI, Cognitive, Knowledge Graph, Linked Data, IBM Watson

1. INTRODUCTION

Many types of writing, such as blog posts and business documents, could benefit from supplemental explanatory material to provide context, background, and additional information for exploration. The World Wide Web provides huge quantities of such supplementary material, but it is not generally suitable for inclusion in new writing due to issues of licensing, copyright, formatting, or syntax. Exceptions include clearly identifiable content meant to be embedded in hypertext documents, such as YouTube videos, and Twitter feeds.

We wanted to determine whether the textual content in a general-purpose hypertext editor may be used to automatically identify relevant supplementary material via artificial intelligence (AI) processes, and whether that material may be presented to a document author for possible inclusion in a document in a useful manner.

2. LITERATURE REVIEW

In a broader context, it is important to recognize that the very concept of literacy in the 21st century is undergoing a renaissance. Speaking only of digital literacy, the understanding of computers and their usage, is no longer sufficient. We must now come to understand how we will interact and interoperate with increasingly capable forms of artificial intelligence.

Zheng et al put the situation clearly: “What it means to be literate in the 21st century is being reshaped to include not only traditional literacies, that is, reading and writing in print-based environments, but also the knowledge, skills, and strategies needed for comprehending and communicating via new technologies” (Zheng et al, 2018).

Familiarization with AI presence may help ease children into new forms of literacy. Researchers of primary school education have noted that prior computer use correlates with writing scores when students later wrote with computer writing tools (Tate et al, 2016). Adults may similarly need slow introductions to new technologies.

The techniques necessary for exploring the integration of AI

into general-purpose computer writing tools are now generally available. Linguistic parsing of text content may be conducted relatively easily using generally available tools, at least in many European languages including English. Concept identification from the results of linguistic parsing has been similarly successful.

Concepts may be matched to relevant additional information. We chose to use a form of knowledge representation for the supplemental information where the information is held in an edge-labeled directed graph in the Resource Description Framework (RDF) data model (Cyganiak et al, 2014). RDF graphs contain concepts and information about them (such as “a dog has a tail” or “the author of Moby Dick is Herman Melville”). Each node in an RDF graph is either a Web address (URI) or a simple data type such as a textual string, a number or a date. Such information form “knowledge graphs” of related information. Knowledge graphs in RDF organize information in ways that are as simple to query as relational databases, but have the additional benefits of being more descriptive. Critically for practical scalability, RDF knowledge graphs may be distributed across many physical or virtual computers.

The first general purpose knowledge graph of scale was Freebase (Bollacker, 2008), which became the basis for the Google Knowledge Graph (Singhal, 2012). Web search engines such as Google, Bing, and Yandex now use knowledge graphs to provide summaries of relevant information related to search terms. It is exactly this functionality that we wanted to add to our general-purpose hypertext editors.

During the writing of this paper, Google announced a feature similar to the one described in this paper for their vertically-integrated Google Docs product (Ranjan, 2016). The Google Docs “Explore” button opens a sidebar that exposes a search option, and presents concepts, images, and related information. Images may be dragged into a document; other presented content opens Web pages for further reading.

There are currently many linguistic concept extraction engines we could have used. We chose to use a concept-extraction service called AlchemyLanguage¹ which uses IBM’s Watson

This quality assured paper appeared at the 8th annual conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2017) and the 30th Annual Conference of the National Advisory Committee on Computing Qualifications, Napier, New Zealand, October, 2-4, 2017. Executive Editor: Emre Erturk. Associate Editors: Kathryn MacCallum and David Skelton.

¹ <https://www.alchemyapi.com/products/alchemylanguage/>

question-answering AI system (Ferrucci et al, 2010). Watson has been used to augment human specific human writing tasks such as providing editing suggestions (e.g. Allan, 2016) but not, as far as we could find, as a provider of supplemental material in a general-purpose text editor.

3. ARCHITECTURE

Our company, Ephox, manufactures rich-text hypertext editors deployable on the World Wide Web as applications within Web pages. The editors are written in the JavaScript programming language. We wrapped one of our editors, Text box.io, in a Web page that included a sidebar component for the presentation of supplemental material. The sidebar is an HTML component that accepts HTML fragments for display and possible drag-and-drop of components into the editor's content pane. JavaScript is used to query an online service to search for relevant information based upon the current contents of the editor.

Textual content is written into the editor by a document author. That content is uploaded to an Ephox service for analysis (Figure 1). The service was deployed in an Internet-accessible cloud computing environment as shown in the figure. The Ephox service sends the content to another service parsing, and to extract concepts. We used IBM's online AlchemyLanguage service for linguistic parsing and concept identification, although any number of similar services could have been used for those purposes. AlchemyLanguage is based upon IBM's Watson question-answering AI system (Ferrucci et al, 2010).² One advantage of the AlchemyLanguage API is that its concept tagging service² returns references into RDF knowledge graphs including DBpedia (Auer et al, 2007), Yago (Suchanek et al, 2008) and Freebase URIs for each identified concept. We used DBpedia URIs as indexes into our RDF knowledge graph. We then used those references as the basis for our queries into our own copy of the DBpedia knowledge graph.

To implement our knowledge graph service, we used an Apache Tomcat3 Web application server with a RDF4J version 2.0 RDF database engine. Relevant portions of DBpedia were sharded into six RDF4J databases to speed both data indexing and assist query performance. DBpedia topical concepts, SKOS (Miles and Bechhofer, 2009) categories, labels, image URLs, short abstracts, categories and their labels, transitive instance types, mapping-based objects and literals, and person data were used for our initial experiments. We limited our DBpedia extracts to the English language due to the language limitations of the AlchemyLanguage concept tagging service.

Two SPARQL 1.1 (Harris et al, 2013) queries against the knowledge graph were used to construct supplemental material representations:

- The first SPARQL query is used to identify the type of a concept (i.e. each concept will be either a person, place, organization, movie, species, product, or a general concept).^[1]
- The second SPARQL query is used to identify available data about a specific concept, as anticipated by its type.^[2]

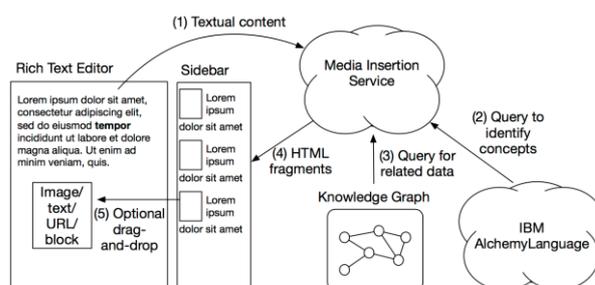


Figure 1: Data flow from editor to sidebar via enhancement services.

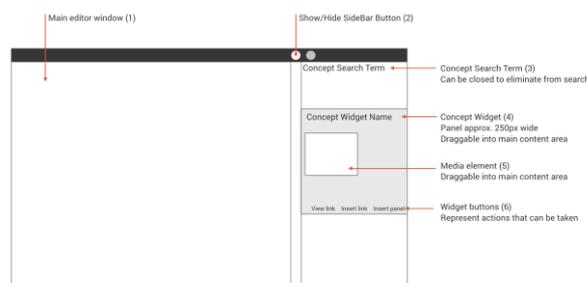


Figure 2: Editor and sidebar user interface elements.

Thus, a concept of type “person” might have a birth date and a birth place. A place might have area and a population estimate, etc. Most data elements for any given type are considered to be optional.^[1] The results of the second query are used to fill a type-specific HTML template. An HTML fragment is produced for each concept, and any empty divs are removed. The completed fragment is returned to the editor's sidebar for display.

SPARQL 1.1 SERVICE clauses were used to execute federated queries (Prud'hommeaux and Buil-Aranda, 2013) across the various databases making up the knowledge graph.

We chose to construct our own knowledge graph instead of using one provided by the AlchemyLanguage service. We wish to eventually refine and contextualize the contents of the knowledge graph over time to provide industry- or even customer-specific data. We also wanted to be able to swap linguistic parsing and concept tagging services if desired at a later time.

A sidebar was added to our editors. Relevant components are illustrated in Figure 2. The show/hide button (2) opens and closes the sidebar panel, shrinking the main editor window content area (1) as it does so. The text typed into the main content area (1) is sent to the server for processing, and the top five concepts returned displayed as keywords (3) in the sidebar area. Individual concepts related to the search terms (3) are displayed as concept widgets (4). Each concept widget has a title and a combination of media and text items. Actions such as view and insert are displayed as buttons at the bottom of each concept widget (6). The concept widgets (4) and media items (5) are able to be drag and dropped into the main content area (1).

² <https://www.alchemyapi.com/products/alchemylanguage/concept-tagging>

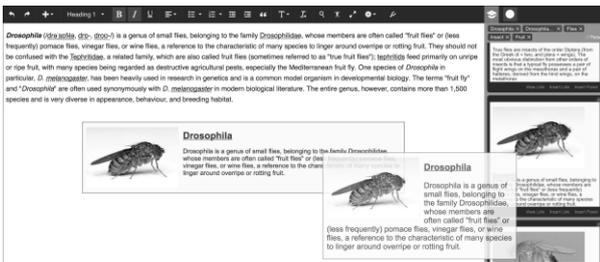


Figure 3: One possible interaction mode; drag-and-drop of suggested content. Content may be reformatted to match editor settings.



Figure 4: Concept filtering interaction sequence.

4. USAGE

We developed five interaction modalities to facilitate the adding of supplemental material to textual content in the editor:

1. **Link insertion.** Inserts a link to the knowledge graph concept. ^[1]_[SEP]
2. **Panel insertion.** Inserts the linked knowledge graph concept widget as an in-line formatted blockquote. ^[1]_[SEP]
3. **Media insertion.** Inserts the media (image or video) from the knowledge graph concept widget into the document. ^[1]_[SEP]
4. **Concept filtering.** Removes concept search terms from the search query. ^[1]_[SEP]
5. **Link viewing.** Opens the link to the knowledge graph concept in a separate browser window. ^[1]_[SEP]

Each interaction type is described below.

4.1 Link insertion

The first interaction type we implemented was the ability to simply add a link to the knowledge graph concept from the sidebar directly into the document. This functionality was delivered through an Insert link button on each concept widget. Pressing the Insert link button would insert a link at the current cursor position, or, if a selection was made in the editor the link would be applied to the selected text.

4.2 Panel insertion

The panel insertion interaction was implemented using two interaction styles. Firstly, like link insertion, a panel could be inserted into the document by clicking the Insert panel button on each concept widget. In addition, a drag-and-drop interaction was implemented on the panel, allowing the concept widget to be dragged directly into the document. In both cases, the resulting concept was displayed as a formatted, in-line blockquote. The insert panel interaction sequence is shown in Figure 3.

4.3 Media insertion

The media insertion interaction was provided as a drag-and-drop interaction only, with no corresponding button on the concept widget. In this case, selecting a specific image (or video) from the sidebar and dragging it into the document embedded the media item only in-line in the document, the formatted blockquote was not included.

4.4 Concept filtering

Identified concepts were shown in a widget near the top of the sidebar (the “search terms” shown in Figure 3). The top five concepts identified by the AI service were listed as keywords within that widget. Each concept could be removed from consideration by deselecting its visual tag. Removal of a concept from the user interface removes it from the knowledge graph search. In the example shown in Figure 4, closing the “John Oxley” concept keyword results in that term being removed from the search, and all concept widgets related to that term being removed from the sidebar. The next most highly weighted concept is added to the widget if one is available.

4.5 Link viewing

The final interaction modality implemented allowed the viewing of the original knowledge graph concept in a new browser window. This was implemented as a View link button on each concept widget, and also as a hyperlink from the concept title. In both cases, clicking the button or the link opened the Wikipedia page related to the knowledge graph concept in a new browser tab.

5. EVALUATION

We ran an unmoderated observational research study in order to test the usage model described in the previous section. Participants were given a writing task and asked to enrich their writing using the Cognitive Assistant. The task description given to users was:

Your Task: Imagine you are writing a blog post to tell your friends about a holiday you recently took (or would like to take) to New York City. Write your story in the editor window below.

Using the Cognitive Assistant, enhance your story with:

- Information about the city itself ^[1]_[SEP]
- Information about any attractions, landmarks, shows, movies or restaurants that you visited (or would like to visit). ^[1]_[SEP]
- Information about famous sports teams based in New York City ^[1]_[SEP]

Please include at least one quote and one video.

The test task intentionally did not prompt the participants on how to achieve the task as we were interested in exploring the natural way in which participants would interact with the sidebar element. Participants were instructed to use a “talk aloud” protocol while completing the task and their interactions recorded using the online testing service Validately³.

The test was deployed to 20 participants aged over 18 years old and having some university education. Of the 20 participants, 11 were female and 9 were male. Thirteen tests were completed either in whole or in part.

³ <https://validately.com>

6. RESULTS

The videos results were coded against the five-part usage model described in section 3, namely: (1) link insertion; (2) panel insertion; (3) media insertion; (4) concept filtering and (5) link viewing. For each of these categories, two of the authors independently evaluated whether the participant:

1. Found the interaction modality (DISCOVERABLE)
2. Was able to interact in an intuitive way (INTUITIVE)
3. Perceived the interaction as having value (USEFUL)

The results were triangulated and discrepancies resolved by reviewing the video footage⁴. Results are shown in Table 1.

Table 1: Results from user tests

	DISCOVERABLE	INTUITIVE	USEFUL
All interactions (1, 2, 3, 4 5), n=35	54%	39%	32%
Any insert interaction (1 or 2 or 3), n=13	100%	92%	85%
Concept filtering interaction (4), n=11	85%	31%	8%
Link viewing interaction (5), n=2	16%	16%	16%

6.1 All interactions

Most participants intuited what the sidebar represented. Some representative comments included:

Looking at this cognitive assistant, that is interesting. The internet tells me a lot, about what New York is about, and I can insert it right here. It gives me a lot of extra information that I need, and it makes writing very simple because it's integrated.

...this is a great concept! I love the idea of a search engine running in the background, finding information to strengthen my writing.

This readily helps me to create a blog as it makes it very easy and quick to write. . . and it guides me with information that inspires me to write more, so I don't have writers block. It's very intuitive, and helps me to write a story with an abundance of information for the viewers of my blog.

[opens cognitive assistant] Oh, how interesting! So it looks like, maybe it searched the text [in the document] and used a search engine to look up the information from my text. That's very interesting.

12 of the 13 participants who completed the task were able to interact with the sidebar to varying degrees of success: the majority of interaction modalities were engaged with (54%),

over a third (39%) of the interactions appeared intuitive and around a third (32%) appeared useful.

A few participants commented on the sidebar being a source of inspiration for ideas that they had not thought of themselves.

The New York Giants...actually I don't follow sports so I will use the cognitive enhancer [to get some info on this topic].

[on opening cognitive assistant] Oh that's cool. It shows me facts about the Statue of Liberty. . . I like this system, how you can get facts on what you are typing about.

6.2 Any insert interactions

The three interaction modalities (link, panel and media insertion) were implemented as alternate modes of interaction. Looking at the results, all of the participants (100%) figured out some way to interact with the sidebar in an intuitive way (92%), and most found these interactions useful (85%).

Link insertion was the least engaged with mode of interaction: only 3 of the 13 participants directly inserted a link using the Insert link button on the concept widget. 9 out of the 13 participants intuitively inserted a panel from the sidebar into their document (using either drag and drop or pressing the Insert panel link). Media insertion was the most intuitive interaction mode, with 10 out of 13 participants dragging media from the sidebar into the document. This result suggests that incorporating images and videos into the concept widgets in the sidebar will promote interaction (as opposed to text-only concept widgets).

For some, while analysis of their interactions revealed that they interacted with the sidebar as expected, their words provided a different narrative:

[uses the cognitive assistant to successfully insert a video and a concept panel] I'm going to say that because it was a prototype, I was unable to use the cognitive assistant.

The above result is not uncommon in this type of testing, and underscores the benefits of observational research over survey-style methods: i.e. the truth lies in what participants do, not what they say that they do.

6.3 Concept filtering

The concept filtering interaction modality was discovered by 85% of participants, however, it was also found to be very confusing: less than a third of participants (31%) found the feature intuitive, and only 8% appeared to find it useful.

Analysis of the participant interactions shows that the main cause for the drop in intuitiveness and utility related to the inability to perform a custom search and to the relevance of the results, i.e. what was returned from the AI service itself. Almost all participants expressed a desire to augment the concepts presented with a custom search feature:

When it came to putting the links in, they just did not have what I was looking for...they showed things that were not as [relevant] to me.

The results [in the sidebar] are not that relevant. . . I wish I could control these keywords, like if there was a search bar up here and I could search for New York Giants. I don't see a way to do that. The search results don't really support what I'm writing about.

Is there a way to search this? Because different things pop up here. . . if I were planning a trip then I could say Oh Ellis Island, Oh yeah, I forgot about that and then I could go there [on concept filters] I don't

⁴ An 8-minute summary video of the most useful results is available at <https://w3id.org/people/prototipo/papers/WWW2017bvideo>

understand what these things are for, I keep xing these thing out and different ones keep popping up...is it figuring out different things based on what I am writing? . . . I don't see a place to search for exactly what I want.

I like the idea of having the cognitive assistant right in the page where the typing is. It didn't seem to be easy to figure out what the search terms were.

The cognitive assistant does have a lot more potential, the only thing is they need to work on technical issues as it's hard to manage [the concept widgets] and search and sometimes you want it to find something, but it does not find it.

6.4 Link viewing

Only 2 of the participants (16%) viewed, or expressed a desire to view, the original sources that corresponded to each concept widget in the sidebar. This result suggests that information presented in the sidebar innately took on a perceived level of authority that one would associate with an encyclopedic-type reference source – the links we tested were all pointing to Wikipedia.com, but need not have been as the architecture of the sidebar allows for the integration of any data source. Other links in our products point to content from additional reachable sources (e.g. content available within an IBM Connections deployment). More research is warranted, but based on our results, care should be taken to ensure that a sidebar type integration into an editor presents reliable information to the author.

6.5 Observations on the on-line testing method

This research used traditional qualitative observational style research methods (Denzin and Lincoln, 2011), carried out in a nontraditional way using the Validately service. Although many of the principles of observational research apply to the design and analysis of the user tests, we discovered interesting nuances in the data collection phase.

Firstly, although the task was relatively simple, our pre-tests indicated that a general audience was unable to complete the writing task. We adjusted the demographics to filter the sample population to include only those participants who had some university-level education. Even then, only 65% of participants were able to complete the test in part or in full.

Secondly, although the participants had some university level education, the task itself was perceived as very hard. 35% of the participants abandoned the test as they were unable to complete the writing component of the task:

I could not complete this task as I have never been to New York and so did not know what to write about.

At the completion of the user test, participants were asked to rate how useful they found the cognitive assistant on a scale of 0 (not at all useful) to 10 (extremely useful). We observed very little correlation between the rating participants assigned and how they interacted with the tool in the test. For example, the average rating for participants who did not even open the sidebar was 2.4, with two of the respondents rating a 6! This result may reflect an underlying pressure that on-line participants feel to “do well” on the test as they are being compensated for it (Ariely, 2008), pp. 278-9. As the field of remote user testing grows, this aspect could warrant its own track of investigation.

7. CONCLUSIONS

Despite the major advances in technology that we have witnessed over the last several decades, the task of writing remains as difficult as it was one hundred years ago. In this

paper, we presented an implementation of widely available AI technology inside a widely deployed commercial web based text editor. Our user testing indicated that the concept created an effective writing partnership, making the writing task easier for authors by reducing context switching while writing.

The component we created utilized existing linguistic concept/entity extraction service (IBM Watson) extended with information from a semantic knowledge graph, and was integrated into two commercial editors, TinyMCE and Textbox.io. The components we tested using unmoderated observational techniques.

Our results showed a high perceived value in the utility of the component, but highlighted that this was directly related to the execution of the user interface elements. In our research, participants found the basic interactions (link insertion, panel insertion and media insertion) intuitive and useful. The concept filtering user interface element, on the other hand, was found to be unintuitive and not useful. Our findings suggest that there is a strong link between the design of the UI elements and the utility perceived by an end user.

The user tests for this research were deployed on-line using Validately. We found that this was an efficient platform for the deployment of unmoderated user tests. Our experience also highlighted the intolerance of this testing medium to mistakes or ambiguities in the design of the test task. We observed interesting nuances in the behaviour of participants, and propose further investigation into this as a topic for future work.

8. FURTHER WORK

The concepts presented in this paper describe one possible way of creating a partnership between AI and human that could make the writing task easier. We are actively exploring this area further, including looking at techniques to simplify writing for multi-lingual authors.

We also expect to grow the size and complexity of the Ephox knowledge graph beyond general encyclopedic content in order to present supplementary material in progressively more contextual ways. We would like to include data that is specific to particular industries, such as manufacturing or pharmaceutical content, and even allow for extension by customers to allow querying of company-specific data.

9. ACKNOWLEDGMENTS

The authors would like to thank James Johnson for prototyping IBM Watson integration into Ephox editors, George Wilson for his review, and Ephox Corporation for supporting this work. We would also like to thank IBM for providing us with access to the Watson API's.

10. REFERENCES

- Allan, D. (2016, June 21). IBM's Watson tries its hand at editing a magazine. Blog post. Retrieved from <http://www.techradar.com/news/world-of-tech/ibm-s-watson-tries-its-hand-at-editing-a-magazine-1323716>
- Ariely, D. (2008). Predictably irrational (p. 20).
- Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. The semantic web, 722-735.
- Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008, June). Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international

- conference on Management of data (pp. 1247-1250). AcM.
- Cyganiak, R., Wood, D., & Lanthaler, M. (2014). RDF 1.1 concepts and abstract syntax. W3c Recommendation, 25(02).
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. Sage.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J., Schlaefel, N., & Welty, C. (2010). Building Watson: An overview of the DeepQA project. *AI magazine*, 31(3), 59-79.
- Harris, S., Seaborne, A., & Prud'hommeaux, E. (2013). SPARQL 1.1 query language. W3C recommendation, 21(10).
- Miles, A., & Bechhofer, S. (2009). SKOS simple knowledge organization system reference.
- Prud'hommeaux, E., & Buil-Aranda, C. (2013). SPARQL 1.1 federated query. W3C Recommendation, 21.
- Ranjan, R. (2016, September) Explore in docs, sheets and slides makes work a breeze — and makes you look good, too. Blog post. Retrieved from <https://www.blog.google/products/docs/explore-docs-sheets-and-slides/>.
- Singhal, A. (2012, May) Introducing the knowledge graph: things, not strings. Blog post. Retrieved from <https://googleblog.blogspot.com.au/2012/05/introducing-knowledge-graph-things-not.html>.
- Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3), 203-217.
- Tate, T. P., Warschauer, M., & Abedi, J. (2016). The effects of prior computer use on computer-based writing: the 2011 NAEP writing assessment. *Computers & Education*, 101, 115-131.
- Zheng, B., Yim, S., & Warschauer, M. (2018). Social Media in the Writing Classroom and Beyond. *The TESOL Encyclopedia of English Language Teaching*.