Exploring the Graphical Interface of Knowledge Structure for Science Texts

Keith Millis, Kyung Kim, Mollie Partee, Madison Milburn, Anne Britt and Christian Steciuch

July 23, 2020

Exploring the Graphical Interface of Knowledge Structure for Science Texts

Keith Millis[1]

Kyung Kim[2]

Molly Partee[1]

Madison Milburn[1]

Anne Britt[1]

Christian Steciuch[1]

[1]Department of Psychology, Northern Illinois University

[2]Department of Educational Technology, Research and Assessment, Northern Illinois University

Author Note

The authors declare that there no conflicts of interest with respect to this preprint.

Correspondence should be addressed to Keith Millis. Email: Kmillis@niu.edu

Abstract

The current study explored the utility of a computerized program called the Graphical Interface of Knowledge Structure (GIKS) that generates and compares a network from a student essay and a master text. We compared different structures of master texts on the same content and also compared GIKS-identified nodes and links to those scored by human experts. We found that GIKS was able to improve node identification from essays by using regular expressions.

Keywords: automatic scoring, knowledge structure, science text, pathfinder, GIKS

Exploring the Graphical Interface of Knowledge Structure for Science Texts

The purpose of the present study was to explore the utility of an approach to represent knowledge proposed by Kim and colleagues (e.g., Kim & Clariana, 2015; Kim, 2018) called the *Graphical Interface of Knowledge Structure* (GIKS). GIKS is a web-based application which creates a visual depiction of knowledge from verbal input using existing computerized visualizing tools. One tool, ALA-Reader computes a co-occurrence matrix of pre-selected terms from the text input (Koul, Clariana, & Salehi, 2005). Another tool, Pathfinder, takes the output of ALA-Reader and converts it to a Pathfinder network (PFnet). Specifically, the PFnet computes all paths between the nodes and searches for the strongest or most direct paths between the nodes. In this way, less salient/weak paths are pruned or removed, and the resulting PFnet has only the most salient connections between the co-occurring terms, reflecting the relations of the well-connected terms/concepts, or visual display of knowledge structure.

In GIKS, students are given a passage to read, and then are asked to write an essay. The passage might be a single document or a series of documents and the essay might be a summary essay or an essay to another prompt. GIKS creates a PFnet of the instructor assigned text (called the master) and computes a PFnet from the student written essay, and computes how they are similar and dissimilar to one another. Figure 1 shows an example output. From the output, the student can quickly recognize which nodes they correctly included or incorrectly omitted, along with correct and missing links. GIKS has been used in diverse domains in physics, biology, business, children and adults, and in different language contexts.

Figure 1:

Screen shot of GIKS

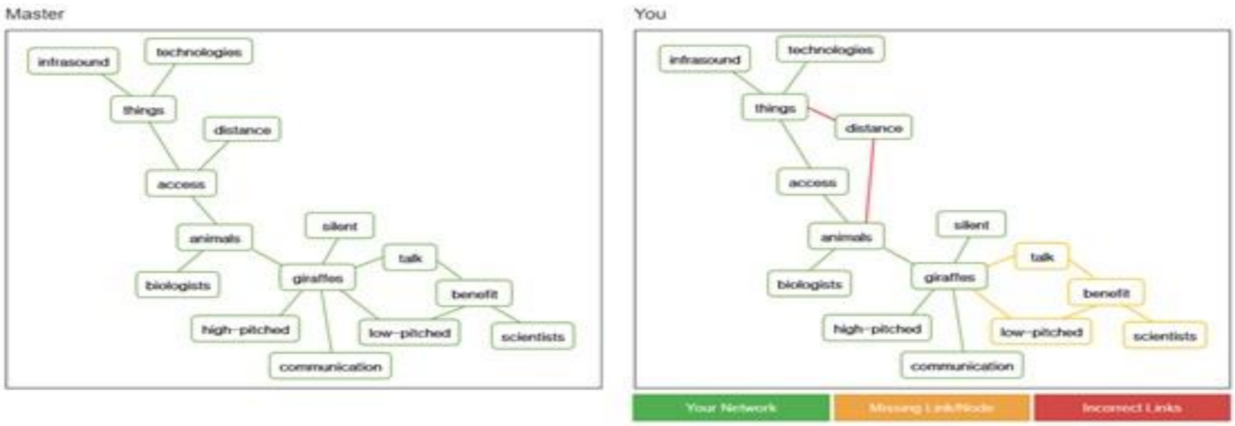


Figure 1: **A GIKS-Text screenshot of knowledge structure network graphs derived from a lesson text (*left*), *Communications through Infrasound*, and a student's summary essay (*right*). Student's knowledge structure feedback consists of a highlighted network graph showing the similarity and difference compared to the referent knowledge structure : *yellow* indicates 'missing' links/nodes and *red* indicates 'incorrect' links/nodes.**

Overview and Research questions. This research is a work in progress. We are ultimately interested in whether GIKS can capture structure of different types of essays at various levels of grain size. For example, how well does the links in the PFnets correspond to human scoring of the same essays and more objective measures (e.g., a multiple choice exam)? At a larger grain size, how well does GIKS do in recognizing parts of an essay, such as claims vs evidence in an argument or the structure of an explanation? Another interest is whether the text(s) which the master PFnet is based upon, matter. Therefore, one research question was how well does GIKS represent the nodes and links as compared to human-scoring of the same essays? Another research question was whether the entire document set performs as well as a summary text when they are used to create a Master document within GIKS.

Methods

Materials

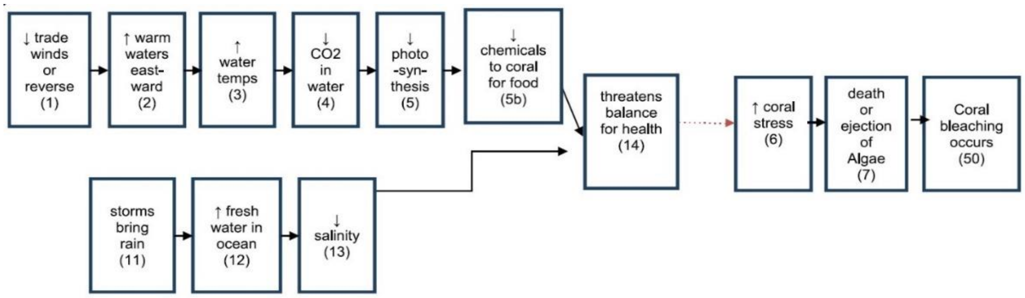
Essays. We sampled 50 essays written by 9th grade Biology students on the topic of coral bleaching (Goldman, et al., 2019). The students were given a set of five documents that together explained the rates of occurrence of coral bleaching. Prior to reading the students were told to read in order to answer the question "What leads to differences in the rates of coral bleaching?" After reading, the students were asked to write an essay that would explain coral bleaching. They had the documents available to them during writing, but it is unknown the extent that they consulted them. The students also took a multiple choice test on the explanatory model. Although not analyzed here, the students also evaluated fictitious peer essays and graphical representations of the causal models.

Each sentence of each essay were scored on correctly expressed concepts and links by two independent raters (Kappa = .81).

Master Documents. The master documents are those which when submitted to GIKS, results in a master knowledge structure that a student-submitted essay can be assessed against. One master was created by submitted the 5 document set that the students in the original study read before writing their essays. We refer to this master as the **Document Master**. The order of the documents was thought to present the most coherent version of the documents. This master had 1,269 words was rated at the 9th grade level according to Flesch Kinkaid. Another master was written as a summary of the causal model itself (see Figure 2). This master had 241 words and was rated at the 8th grade level. We refer to this as the **Causal Model Master**.

Figure 2:

Causal Model of Coral Bleaching (Goldman, et al. 2019)



Essay Selection and Procedure

We submitted each chosen essay to GIKS twice. The first and second submissions used the Document and Causal Model Masters respectively.

Results

At this time, we have computed the correlations between the GIKS-generated measures and human scores of two variables: number of correct nodes and the number of correct links. (We do not have human scoring of the number of missing nodes and links.) The correlations are presented in Table 1.

In regard to the first research question, the correlations were all statistically significant. In regard to the second research question, the correlations of the GIKS master derived from whole document set were slightly higher than the ones derived from a summary of the causal structure.

Table 1: Summary of Correlations between GIKS and Human Scores

	# Nodes		# Links	
GIKS	Summary	Full Document	Summary	Full Document
	Master	Master	Master	Master
# correct nodes	.75*	.86*		
# correct links			.72*	.83*

*p < .05

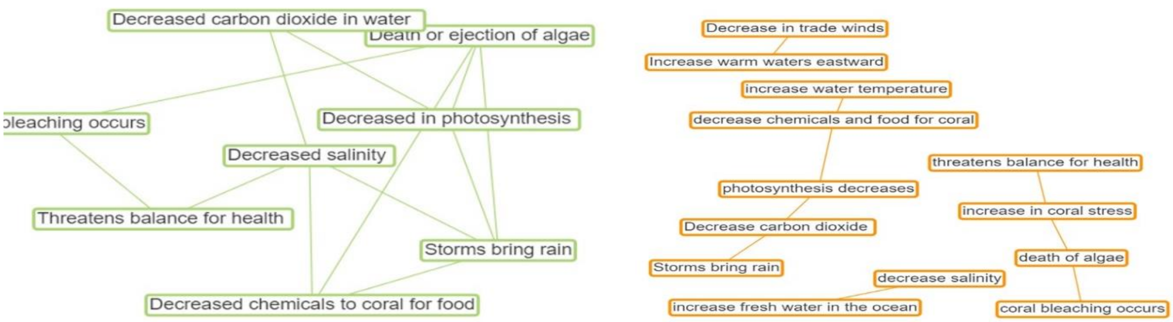
The above findings suggest that GIKS can reliably identify nodes and links from essays. However, the findings are limited in a very important way. Namely, they are correlations based on the **number** of nodes or links rather than other metrics which inform the accuracy of GIKS identifying any given node or link.

We have recently begun to identify ways in which GIKS can increase its accuracy. With GIKS primary developer, Dr. Kyung Kim, we are currently working on how GIKS represents key content in the texts. GIKS in its current form only uses key words together with synonyms rather than key words that express propositional content. For example, "salinity in the water decreases" would be represented by three key words, "salinity" along with synonyms (e.g., salt, salt water), "water" with synonyms (e.g., ocean, sea), and "decreases" with synonyms (e.g., low, small). However, this single keyword approach tends to under- or over-specify term links. For example, the word 'salinity' would be also linked to other words co-occurred with the 'salinity' across a text, thus adding some errors to the data. We have changed the key word approach to a propositional approach in which GIKS uses regular expressions to represent different ways in which a text represents a complete idea. A regular expression for “salinity in the water decreases” would be “[salinity/salt] [water/ocean/sea] [decreas*/low*/small*]” (The “/” represent “or” and * represents wildcards for part of the word). In this case, the regular expression would need three individual words from a single sentence that matches one of the words within each bracket. For each keyword (proposition), there can be several regular expressions which can convey different word orders and options. This newer vision of GIKS allows the users to code propositional content with regular expressions that (1) largely ignore word order and (2) ignore intervening words. This would allow the “salinity” node to fire from the sentence “The salinity in the water is lowered.”

Once we made the jump to regular expressions, we informally saw an increase in the accuracy of the graphs. For example, on the graph below (see left side of Figure 3), one can see a simplified GIKS master graph created from individual key words) and on the right, the resulting master graph from using regular expressions. The largest difference between the two is that the original (using individual concepts) results in a denser graph (more links) than the one based on regular expressions which is more linear. This result is likely based on the fact that many of the content words (and their synonyms) appear more than once in the master text. The occurrence of many instances may cause GIKS (more specifically, ala-reader) to over-link any one given concept. The regular expression approach reduces this over-linking because the added words provide constraints such that only cases in the original text that conveys the intended propositional meaning are identified. This lowers the number of times a node is identified and also the number of adjacent nodes that could be linked to it. Although the regular expression-based graph appears closer to the human-created causal network (see Figure 2), it is not perfect.

Figure 3:

Master graph from concepts (left) and propositions/regular expressions (right)



To access the accuracy of node identification in student essays, we submitted them to the revised GIKS that uses regular expressions. Then we compared GIKS identification to the human scoring. We evaluated accuracy within a signal detection framework. Hit rate was defined by the proportion of correctly identified nodes. False alarms were the proportion of nodes incorrectly identified by GIKS to be located in an essay. Misses was defined by 1 minus the hit rate. Precision was defined by the hit rate divided by the sum of hits and false alarms. Recall was defined by the by the hit rate divided by the sum of hits and misses. F1 was defined by 2 times (recall X precision divided by the sum of recall and precision.

The results are in Table 2 and appear to be encouraging. The hit rate (.70) seems to be fairly high, and the false alarm rate is very low (.03). However, the misses (.30) are higher than what we would like. We think the regular expressions led to the lower false alarms (and the higher misses) because they provided more constraints than single words. However, the regular expressions likely need refinement because almost a third of nodes were missed by the system.

Table 2: *Summary Statistics Assessing the Accuracy of Identifying Nodes using Regular Expressions*

Statistic	Mean	SD
Hits	.70	.25
False Alarms	.03	.12
Misses	.30	.25
Recall	.70	.25
Precision	.96	.11
F1	.79	.19

Discussion

We think GIKS is a promising approach for automatically representing students' knowledge based on essays. It provides a student with instantaneous graphical feedback based on their essay. Although there is some question regarding the meaning and validity of the links because they are unlabeled, GIKS has been empirically shown to increase learning and to represent changes in knowledge structure from reading (e.g., http://giks.herokuapp.com/our_publications). However, we believe that there is room for improvement in regard to the accuracy in which nodes from student essays are identified. We had informally shown that moving from a single word approach to a regular expression approach most likely had lowered false alarms. The next step is to decrease misses, and to increase the accuracy of identifying structures. However, identifying structures from essays is very challenging from NLP and machine learning approaches (Wiley et al., 2017). We are hopeful that by using regular expressions, and by modifications to its subprograms (ALA Reader, Patherfinder algorithms) GIKS will be better at accurately representing students' knowledge.

References

Goldman, S. R., Greenleaf, C., Yukhymenko-Lescroart, Brown, W., Mon-Lin, M., Emirg, J. M., George, M., Wallace, P., Blaum, D., Britt, A. M., Project READI. (2019). Explanatory modeling in science through text-based investigation: Testing the efficacy of the project READI intervention approach. American Educational Research Journal., 56, 1148-1216.

Kim, K. (2018). An automatic measure of cross-language text structures. Technology Knowledge and Learning, 23, 301–314. https://doi.org/10.1007/s10758-017-9320-5.

Kim, K., & Clariana, R. B. (2015). Knowledge structure measures of reader's situation models across languages: Translation engenders richer structure. Technology, Knowledge and Learning, 20(2), 249–268. https://doi.org/10.1007/s10758-015-9246-8.

Koul, R., Clariana, R. B., & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. Journal of Educational Computing Research, 32(3), 261–273.

Wiley, J., Hastings, P., Blaum, D. Jaeger, A. J., Hughes, S., Wallace, P., Griffin, T., & Britt, A. (2017). Different Approaches to Assessing the Quality of Explanations Following a Multiple-Document Inquiry Activity in Science. *International Journal of Artificial Intelligence in Education*, 27, 758–790 (2017). https://doi.org/10.1007/s40593-017-0138-z.