Breaking Down Barriers: Empowering Diabetes Patients with User-Friendly Medical Explanations

Rasha Hendawi
Department of Computer Science
North Dakota State University
Fargo, USA
rasha.hendawi@ndsu.edu

Shadi Alian
Department of Computer Science
North Dakota State University
Fargo, USA
shadi.alian@ndsu.edu

Juan Li
Department of Computer Science
North Dakota State University
Fargo, USA
j.li@ndsu.edu

2024 15th International Conference on Information and Communication Systems (ICICS) | 979-8-3315-4076-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICS63486.2024.10638283

***Abstract*— Effective management of diabetes is contingent upon patients' understanding of their medical conditions and treatments. However, medical documents often contain complex jargon and technical details that can be challenging for patients, especially those with limited health literacy. This paper presents DiaKnow, an innovative tool that simplifies medical documents and customizes explanations to suit individual health literacy levels. Employing a robust self-attention transformer model and a comprehensive diabetes-focused knowledge graph, DiaKnow enhances patient comprehension by providing contextually relevant, simplified medical information. This study assesses DiaKnow's efficacy in real-world clinical settings through a structured use case evaluation method. We tested the tool's ability to accurately identify, link, and simplify crucial medical terms using a diverse set of medical documents. Our findings confirm that DiaKnow not only improves the readability of medical documents but also ensures that explanations are medically accurate, clear, and comprehensive.**

Keywords— health literacy, knowledge graph, ontology, self-attention transformers, medical entity recognition, entity linking.

I. Introduction

Managing diabetes poses a substantial challenge for many individuals, specifically in comprehending medical records. The utilization of specialist medical terminology and complex information can be daunting, and this challenge is further intensified by varying levels of health literacy. People without specialized medical knowledge may find terminology like "glycemic control," "hyperglycemia," "A1C," and pharmacological names difficult to understand. This can lead to ambiguity and misinterpretations about a patient's condition and the necessary actions for appropriate treatment.

Furthermore, effectively managing diabetes necessitates consistent involvement with quantitative data, such as blood glucose readings, insulin dosages, and laboratory test results. Patients rely on this information for their daily decisions regarding their well-being. The significance and implications of these decisions can be overwhelming, especially without clear explanations.

Moreover, patients often find it difficult to implement some lifestyle guidelines provided in medical documents due to their presentation style. Ensuring successful application of diet and exercise guidelines in real-life situations is crucial, and this requires providing clear and practical explanations. It is necessary to tailor these explanations to align with the individual's specific requirements and their health literacy level.

The prevalence of inadequate health literacy is not necessarily dependent on a patient's level of formal education. A significant number of individuals face difficulties in understanding health information, impeding their ability to effectively manage their condition. Studies suggest that almost 60% of patients with diabetes struggle to appropriately understand their blood glucose readings, a critical component of effectively treating the condition [1]. Further research has substantiated a clear link between low health literacy and unfavorable results in the management of diabetes, such as suboptimal medication adherence and glycemic control [2] .

To address these challenges, it is imperative to use innovative technical solutions. The foundation of our approach is rooted in our previous work in health literacy evaluation [3]. We introduce DiaKnow, a customized software that simplifies medical documents and provides tailored explanations according to the user's degree of health literacy. DiaKnow employs medical knowledge bases and assesses the user's health literacy to identify complex diabetes-related terms and provide contextually suitable, clearly understandable explanations. DiaKnow employs a robust self-attention transformer model and an extensive diabetes-specific knowledge graph to deliver concise and pertinent medical information. DiaKnow improves the capacities of individuals with diabetes by overcoming communication barriers and facilitating a deeper understanding of their condition, leading to improved self-management.

The rest of the paper is organized as follows: Section II surveys the related work. Section III explains the details of the proposed methodology. Section IV explains the evaluation methodology and the use case evaluation. Section V concludes the paper.

II. Related work

The simplification of biomedical texts has been the subject of substantial effort. In this discipline, simplification typically employs neural sequence-to-sequence models or dictionary-based methods.

Dictionary-based approaches leverage expert-curated medical dictionaries, such as the Unified Medical Language System UMLS [4], to simplify professional medical sentences [5][6][7][8][9][10][11][12][13] [14][15], or link medical terms with lay definitions [16] and definitions in controlled vocabularies [17]. Although generally effective, these methods can encounter difficulties with terms that have multiple meanings, leading to potential ambiguity. For instance, in the sentence "In patients with DMD, immune cell infiltration of skeletal muscle aggravates disease," the term "DMD" could be associated with multiple meanings, including Duchenne Muscular Dystrophy or the DMD gene, whereas only the first is correct in this context.

To tackle this issue, Sakakini et al. [18] proposed a context-aware medical text simplification tool by improving the lexical replacement methodology by using a language model to elucidate potential replacements with the aid of

This work was supported by the National Science Foundation (NSF) with award number: 2218046.

surrounding context. Similarly, Tran et al. [19] employed a neural Masked Language Model to create context-aware substitution candidates, enhancing the traditional dictionary-based approaches. Nevertheless, these direct substitutions can sometimes diminish the text's readability, continuing to make it difficult for users or patients to comprehend.

Neural methods represent another approach to biomedical text simplification. Supervised methods, such as PharmMT [20], utilize Seq2seq models with a numerical checker to simplify the text. In a similar vein, Devaraj et al. [21] enhanced the BART model [22] with an unlikelihood constraint to reduce the use of technical terms. These approaches generally improve readability over dictionary-based methods, though they can sometimes compromise on the accuracy of term-level simplification.

In the area of unsupervised techniques, Cao et al. [23] came up with a style transfer method for simplifying. They trained different models on nonparallel datasets of professional and everyday language to show the source and target styles, respectively. Meanwhile, Pattisapu et al. [24] explored an unsupervised denoising autoencoder approach. The authors manipulated medical text by replacing simpler terms with their more complex equivalents from knowledge bases to generate training data. The model then learns to revert these texts to their original, simplified forms while ensuring fluency. This denoising method generally outperforms style transfer in human assessments, but both are susceptible to generating unpredictable errors such as repeated words or phrases.

Existing simplification tools typically simplify recognized medical-related terms indiscriminately, regardless of the user's familiarity with them. However, our tool, DiaKnow, provides a context-aware personalized text simplification scheme that tailors simplification based on an individual reader's health literacy level. More terminology is simplified for users with very low literacy levels, whereas only difficult medical terms are simplified for users with higher literacy levels. This personalized approach aims to enhance comprehension and accessibility for users with varying levels of health literacy.

III. Methodology

In this research, we introduce an advanced medical document annotation system to provide diabetic patients with easily understandable explanations. Our method utilizes a self-attention transformer model and a comprehensive knowledge graph to accurately extract entities and provide context-aware annotations. Fig. 1 illustrates the architecture of the system.

A. Knowledge Graph Construction

Foundation for Explanation: We build a comprehensive diabetes-centered knowledge graph. The knowledge graph functions as a basis for generating explanations and enhancing the context of extracted entities. We started by developing the Diabetes Ontology, which serves as a highly organized, formal, and consistent representation of entities and relationships between them in the diabetes field. It also serves as a framework for the knowledge graph. We built our diabetes ontology based on the Diabetes Mellitus Diagnosis Ontology (DDO) [25], which guarantees consistency and interoperability with existing systems. Fig. 2 shows some of the major concepts and relationships of the high-level of the diabetes ontology.

Enhanced Structure: The Diabetes Ontology provides the high-level logical structure of diabetes, including its symptoms, clinical presentation, diagnosis, treatment, and complications. To expand its scope, we have enhanced the ontology with additional classes and properties tailored to encompass lifestyle interventions and physical activities. This expansion enhances our knowledge graph by allowing it to handle various types of entities related to diabetes, including diseases (such as diabetic neuropathy and retinopathy), medications (such as insulin and metformin), procedures (such as blood glucose monitoring and HbA1c test), physiological processes (such as insulin secretion and glucose metabolism), and lifestyle factors (such as diet and exercise).

Relationships and Attributes: The knowledge graph represents each of these concepts as a node. Various properties representing cause-and-effect, treatment options, and other relevant relationships connect these nodes with other related concepts. Furthermore, it is possible to enhance each node by adding features that offer more specific information on the entity, such as medicine dose ranges and symptoms of complications. By utilizing external knowledge sources, including UMLS, SNOMED CT [26], and Wikidata [27], we extend the reach and depth of diabetes-related information within our knowledge graph. We strengthen the ontology's semantic framework by instantiating and linking ontology entities with these repositories, fostering a comprehensive understanding of diabetes-related concepts. This approach not only improves information coverage and granularity, but also enables the seamless integration of emerging knowledge and concepts into the knowledge graph.

B. Self-Attention Transformer for Entity Extraction

To recognize diabetes-related medical terminologies, we employed BioBERT [28]. BioBERT is a robust transformer model that has been specifically trained on a large collection of biomedical documents. This pre-training provides the model with a fundamental comprehension of medical terminology and the relationships between concepts, including those related to diabetes. Employing BioBERT in our system enables us to decrease the time required to fine-tune the model on our diabetes-focused dataset.

BioBERT utilizes the fundamental transformer architecture from the original BERT model [29]. Its pre-training process involved two key tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). During MLM, words within biomedical text were randomly masked, and BioBERT learned to predict the missing words based on the surrounding context. This strengthens the model's ability to understand the nuanced relationships and meanings of medical terms. NSP, on the other hand, trained the model to identify logical connections between sentences. This helps BioBERT grasp how information flows within medical documents — a crucial skill for effective entity recognition.

We enhance the performance of BioBERT specifically for the purpose of finding items related to diabetes. This entails refining the pre-trained model by using a meticulously selected dataset of documents that have been annotated specifically for diabetes. BioBERT's weights are modified during the fine-tuning phase to further improve its sensitivity to diabetes-related phrases. BioBERT utilizes a

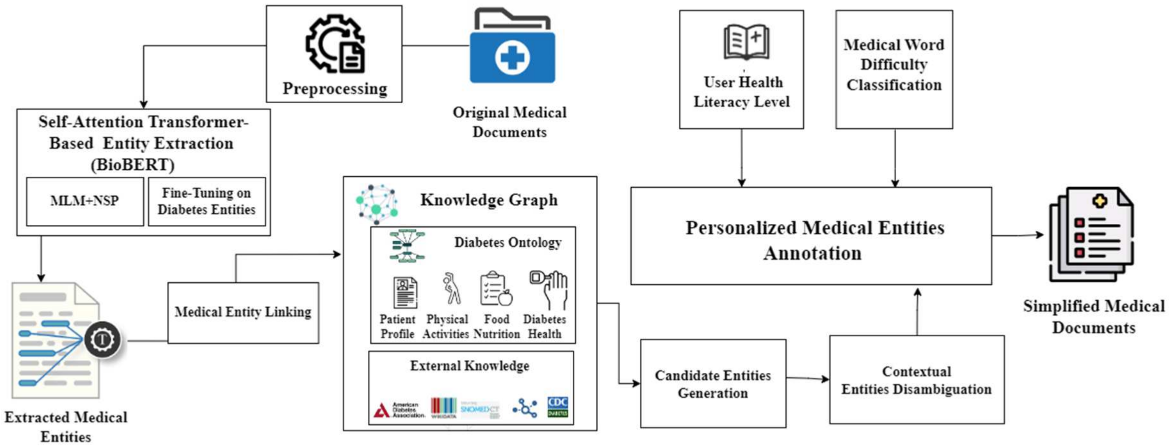

Fig. 1. System Architecture.

special tokenizer specifically created to manage complicated medical terminologies and their variations. The embedding matrix of BioBERT is used to turn each token (or word component) into a numerical representation. This matrix represents the relationships and semantic significance of words in the domain of medicine.

The self-attention mechanism, which is a key component of the transformer design, allows the model to dynamically focus on the most significant phrases to detect medical entities. Within BioBERT, this mechanism works similarly to the principle described previously. The inclusion of many encoder layers in the model allows it to learn complex representations of the input text, thereby enabling precise recognition of diabetes-related entities. Finally, the output layer utilizes this information to predict the most suitable entity type (such as "Disease" or "Medication") for each token in a document.

C. *Knowledge Graph-Based Annotation*

DiaKnow utilizes an extensive knowledge graph to enhance extracted entities with relevant information and offer user-friendly explanations for patients. An essential element of this procedure is entity linking, in which the tool links a medical term recognized by the transformer model with the most suitable corresponding node in the knowledge graph.

Using embeddings, which are vector representations that capture the context and semantic meaning of words, is essential to accurate entity linking. During the process of entity linking, the medical term extracted by the transformer and the descriptive text connected to each node in the knowledge graph are both represented as embeddings. By calculating the cosine similarity between these embeddings, DiaKnow identifies potential candidates for the correct node. Nodes with higher similarity scores are considered a better match.

Even though embeddings provide a potent signal for entity linking, the contextual awareness provided by the transformer model's pre-training is still essential. DiaKnow relies on this context to resolve ambiguities, particularly when multiple knowledge graph nodes have similar embeddings but are used in different contexts. This is achieved by comparing the similarity scores of candidate nodes and the embeddings of surrounding words in the original text. Consider the extracted entity, "metformin." If the knowledge graph contains nodes for both the medication "metformin" and a less common lab test, also called "metformin," the embeddings for these nodes might be relatively similar. However, by examining the contextual embeddings of the surrounding words, the model can often determine whether the extracted "metformin" refers to the medication (surrounded by words like "dose," "side effects") or the lab test (surrounded by terms like "blood levels," "kidney function").

After an entity is effectively connected to a knowledge graph node, DiaKnow utilizes the graph's structure and content to produce a customized explanation. DiaKnow utilizes our well-established health literacy measurement method [3] to evaluate the user's comprehension of medical concepts before producing these explanations. DiaKnow considers the user's assessed health literacy level when annotating the medical terms, determining the level of detail, and prioritizing external resources. In addition, DiaKnow includes the medical term's difficulty classification model that we built in our previous research [3] to determine which mentions should be annotated based on the user's health literacy level, providing personalized annotation. Users with low health literacy levels will receive more annotations, while those with high health literacy levels will only receive annotations for complex medical terms.

By traversing the graph's connections (representing properties like "treats," "is caused by," and "is a type of"), DiaKnow identifies relevant related concepts. The explanation follows pre-defined templates based on the type of entity and the relationships uncovered. Transformer-generated representations continue to enhance the explanation by tailoring it to the specific use of the medical term within the original document. Additionally, the knowledge graph itself can contain pre-populated links to external resources, such as trusted patient education websites, allowing DiaKnow to seamlessly incorporate these into its explanations.

DiaKnow carefully selects external resources (e.g., the American Diabetes Association (ADA) [30], the International Diabetes Federation (IDF) [31], and the Center for Disease Control (CDC)- Diabetes [32] to match the user's health literacy level. Sources with clear explanations, visuals, or simplified language might be prioritized for users with lower assessed health literacy.

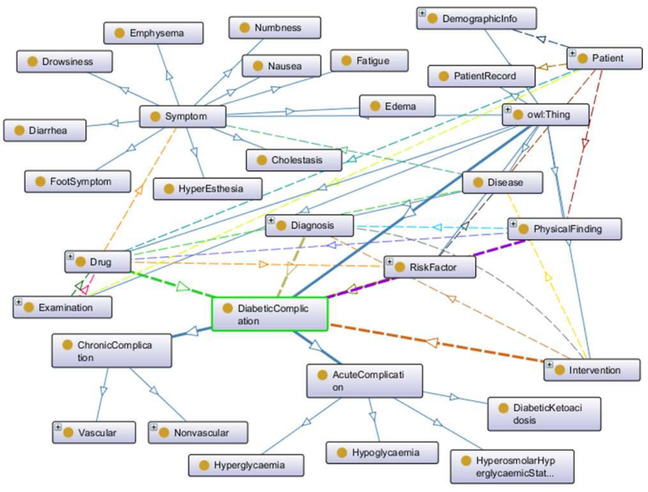

Fig. 2. Part of the Diabetes Ontology (Produced by Protégé 5.5.0 [35]).

IV. EVALUATION

A. Evaluation Methodology

Due to constraints in conducting direct user testing, the initial evaluation of DiaKnow was conducted by our research team, who utilized medical textbooks and relevant literature to assess the tool's effectiveness. This crucial preliminary evaluation aimed to confirm the functionality and accuracy of DiaKnow before its broader deployment and more extensive user-based testing.

The team focused on several fundamental aspects essential to ensuring the operational success of the tool:

1) Correctness of Annotations

Objective: Verify that DiaKnow accurately identifies medical terms from the documents and links them to the appropriate nodes within the knowledge graph.

Methodology: We manually checked each term DiaKnow identified to make sure it matched the correct node in the knowledge graph.

Metrics: We measured success by calculating the percentage of terms correctly linked to their knowledge graph nodes, aiming for a high accuracy rate to ensure reliability.

2) Appropriateness of Explanations

Objective: Verify that the explanations generated by DiaKnow are medically accurate.

Methodology: We reviewed these explanations against current medical texts and guidelines to ensure their accuracy and appropriateness.

Metrics: The percentage of explanations confirmed as medically accurate served as a measure of the annotation's appropriateness.

3) Text Clarity

Objective: Evaluate the simplicity and readability of the explanations provided by DiaKnow, focusing on their ability to be understood by patients with varying levels of health literacy.

Methodology: In evaluating the clarity of language in DiaKnow's explanations, we utilized the Text Ease and Readability Assessor (T.E.R.A.) [33], selectively focusing on components critical to our tool's objectives. Given the unique nature of medical documents, which typically do not require a narrative style or high levels of cohesion due to their instructional nature, we concentrated on the following areas:

- Syntactic Simplicity: Assesses the structure of sentences by examining the number of words and clauses per sentence and the placement of the main verb. Simpler syntax is crucial for ensuring that medical instructions are clear and comprehensible, which is particularly important in healthcare settings where complex syntax can hinder understanding.

- Word Concreteness: Evaluates the use of concrete versus abstract words. Concrete words create clear mental images, making the text easier to understand compared to abstract words that might lack visual specificity.

Metrics: Clarity was scored based on T.E.R.A. outcomes for syntactic simplicity and word concreteness, emphasizing DiaKnow's effectiveness in simplifying complex medical information and using precise language that enhances patient comprehension.

4) Completeness of Terminology Coverage

Objective: Assess DiaKnow's effectiveness in identifying and processing a comprehensive range of diabetes-related medical terms.

Methodology: We compiled an extensive list of essential diabetes-related terms and checked DiaKnow's output to ensure all were recognized and appropriately explained.

Metrics: The completeness metric was evaluated based on the proportion of the predefined list of terms that DiaKnow successfully identified and adequately processed within the medical documents.

By integrating these comprehensive methods and metrics, this evaluation seeks to ensure that DiaKnow's explanations are not only accurate from a medical standpoint but also optimized for patient education, balancing the complexity of medical information with the need for accessibility and clarity.

B. Use Case Evaluation

To validate the practical application of DiaKnow and assess its performance in real-world settings, we implemented a use-case evaluation approach. This method allowed us to observe how DiaKnow functions in scenarios typical of everyday clinical practice and patient education. Fig. 3 depicts a written summary during a doctor's visit for a patient with diabetes, characterized by low health literacy. The summary includes terms that are crucial for understanding their condition and management but are often confusing.

DiaKnow processes the document and identifies key medical terms such as "hyperglycemia," "A1C," "NPH insulin," and "glipizide." It then annotates these terms with simple definitions that are easily understandable by someone with low health literacy and links to additional trusted resources such as the American Diabetes Association (ADA) for further reading and understanding.

We verified that DiaKnow accurately identifies and correctly links all relevant terminology within the document. The focus of the evaluation then shifts to assessing the clarity and completeness of the explanations provided, ensuring they are medically accurate and suitable for a patient with low health literacy.

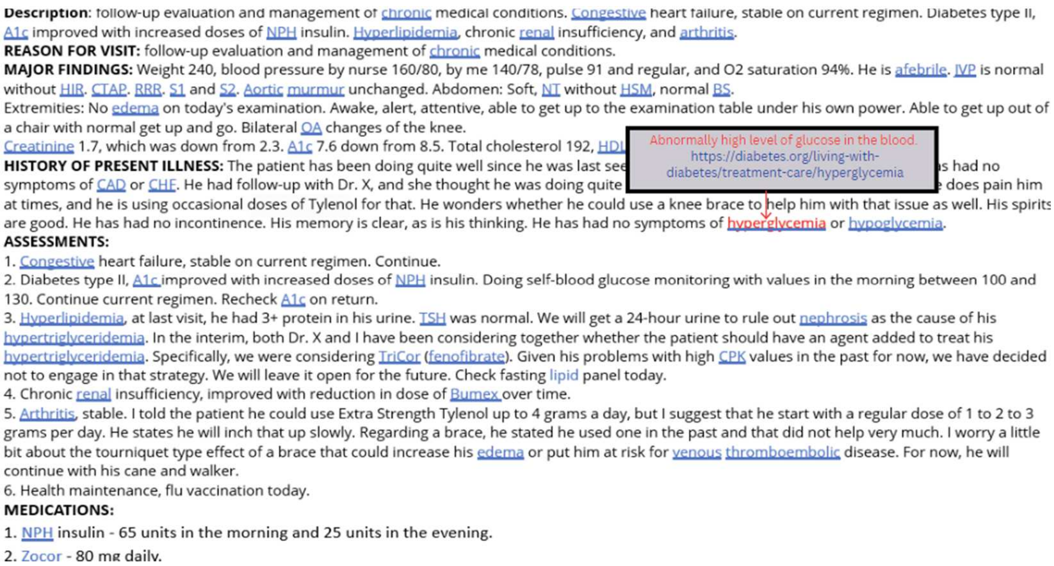

Fig. 3. Example Use Case: Doctor's Visit Summary

TABLE 1. PERFORMANCE METRICS OF DIAKNOW ACROSS DIFFERENT USE CASE SCENARIOS

Use Case Scenario	Number of Cases	Correctness of Annotations (Identification and Linking)	Appropriateness (Explanations)	Text Clarity (T.E.R.A score)		Completeness (Terminology Coverage)
				Syntactic Simplicity	Word Concreteness	
Doctor's Visit Summary	15	95%	98%	92%	85%	93%
Hospital Discharge Instructions	20	96%	98%	94%	87%	94%
Educational Diabetes Materials	25	97%	99%	94%	90%	96%

To comprehensively assess DiaKnow, we evaluated multiple use cases that span various facets of diabetes care and patient education. These scenarios were derived from the Medical Transcription Samples website [34], which hosts an extensive collection of transcribed medical reports across numerous specialties, including diabetes. Table 1 outlines the scenarios, metrics used, and average results for the use cases.

The use-case evaluation demonstrates DiaKnow's ability to effectively enhance patient understanding by simplifying complex medical information. The results from these scenarios indicate that DiaKnow can be a valuable tool in various healthcare settings, aiding in the education of patients with varying levels of health literacy. By providing clear, accurate, and easy-to-understand explanations, DiaKnow has the potential to significantly improve patient comprehension and engagement in their own care.

V. CONCLUSION

The development and evaluation of DiaKnow represent a significant step forward in addressing the challenges of health literacy in diabetes management. By simplifying complex medical documents and tailoring explanations to individual literacy levels, DiaKnow bridges the gap between medical professionals and patients, enhancing understanding and engagement in personal health management.

Our evaluation demonstrates that DiaKnow effectively transforms complex medical jargon into language that is accessible to patients with varying degrees of health literacy, particularly benefiting those with low literacy levels. By accurately identifying and annotating critical medical terms like "hyperglycemia" and "A1C," DiaKnow ensures that patients not only understand their condition but are also better equipped to manage it effectively.

While the use case evaluation provides crucial insights and demonstrates DiaKnow's effectiveness, direct user studies with patients with diabetes are essential. User studies can assess how well DiaKnow addresses real-world challenges, improves patient understanding of medical documents, and ultimately empowers patients to better manage their diabetes care.

REFERENCES

[1] S. C. Bailey et al., "Update on Health Literacy and Diabetes," Diabetes Educ, vol. 40, no. 5, 2014, doi: 10.1177/0145721714540220.

[2] D. Schillinger et al., "Association of health literacy with diabetes outcomes," JAMA, vol. 288, no. 4, 2002, doi: 10.1001/jama.288.4.475.

[3] R. Hendawi, S. Alian, and J. Li, "A Smart Mobile App to Simplify Medical Documents and Improve Health Literacy: System Design and Feasibility Validation," JMIR Form Res 2022;6(4):e35069 https://formative.jmir.org/2022/4/e35069, vol. 6, no. 4, p. e35069, Apr. 2022, doi: 10.2196/35069.

[4] "UMLS - Glossary." Accessed: Nov. 17, 2022. [Online]. Available: https://www.nlm.nih.gov/research/umls/new_users/glossary.html

[5] S. Kandula, D. W. Curtis, and Q. Zeng-Treitler, "A semantic and syntactic text simplification tool for health content.," AMIA Annu Symp Proc, 2010.

[6] L. Ramadier and M. Lafourcade, "Radiological Text Simplification Using a General Knowledge Base," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10762 LNCS, pp. 617–627, 2018, doi: 10.1007/978-3-319-77116-8_46.

[7] L. Zilio et al., “A Lexical Simplification Tool for Promoting Health Literacy.” pp. 70–76, 2020. Accessed: May 04, 2024. [Online]. Available: https://aclanthology.org/2020.readi-1.11

[8] E. Abrahamsson, T. Forni, M. Skeppstedt, and M. Kvist, “Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language,” Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations, PITR 2014 at the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, pp. 57–65, 2014, doi: 10.3115/V1/W14-1207.

[9] M. Alfano, B. Lenzitti, G. Lo Bosco, C. Muriana, T. Piazza, and G. Vizzini, “Design, development and validation of a system for automatic help to medical text understanding,” Int J Med Inform, vol. 138, Jun. 2020, doi: 10.1016/J.IJMEDINF.2020.104109.

[10] B. Qenam, T. Y. Kim, M. J. Carroll, and M. Hogarth, “Text Simplification Using Consumer Health Vocabulary to Generate Patient-Centered Radiology Reporting: Translation and Evaluation,” J Med Internet Res, vol. 19, no. 12, Dec. 2017, doi: 10.2196/JMIR.8536.

[11] “Improving perceived and actual text difficulty for health information consumers using semi-automated methods - PubMed.” Accessed: May 04, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23304324/

[12] “Making texts in electronic health records comprehensible to consumers: a prototype translator - PubMed.” Accessed: May 04, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18693956/

[13] Q. T. Zeng and T. Tse, “Exploring and Developing Consumer Health Vocabularies,” J Am Med Inform Assoc, vol. 13, no. 1, p. 24, 2006, doi: 10.1197/JAMIA.M1761.

[14] J. Chen, A. N. Jagannatha, S. J. Fodeh, and H. Yu, “Ranking Medical Terms to Support Expansion of Lay Language Resources for Patient Comprehension of Electronic Health Record Notes: Adapted Distant Supervision Approach,” JMIR Med Inform, vol. 5, no. 4, Oct. 2017, doi: 10.2196/MEDINFORM.8531.

[15] J. P. Lalor, B. Woolf, and H. Yu, “Improving electronic health record note comprehension with NOTEAID: Randomized trial of electronic health record note comprehension interventions with crowdsourced workers,” J Med Internet Res, vol. 21, no. 1, p. e10793, Jan. 2019, doi: 10.2196/10793.

[16] J. Chen et al., “A Natural Language Processing System That Links Medical Terms in Electronic Health Record Notes to Lay Definitions: System Development Using Physician Reviews,” J Med Internet Res, vol. 20, no. 1, Jan. 2018, doi: 10.2196/JMIR.8669.

[17] “Improving patients’ electronic health record comprehension with NoteAid - PubMed.” Accessed: May 04, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23920650/

[18] T. Sakakini et al., “Context-Aware Automatic Text Simplification of Health Materials in Low-Resource Domains,” EMNLP 2020 - 11th International Workshop on Health Text Mining and Information Analysis, LOUHI 2020, Proceedings of the Workshop, pp. 115–126, 2020, doi: 10.18653/V1/2020.LOUHI-1.13.

[19] L. Tran, E. Velazquez, R. J. Sips, and V. De Boer, “Evaluating Medical Lexical Simplification: Rule-Based vs. BERT,” Stud Health Technol Inform, vol. 281, pp. 1023–1024, Jul. 2021, doi: 10.3233/SHTI210337.

[20] J. Li, C. Lester, X. Zhao, Y. Ding, Y. Jiang, and V. G. Vinod Vydiswaran, “PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions,” Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020, pp. 2785–2796, 2020, doi: 10.18653/V1/2020.FINDINGS-EMNLP.251.

[21] A. Devaraj, I. J. Marshall, B. C. Wallace, and J. J. Li, “Paragraph-level Simplification of Medical Texts,” NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, pp. 4972–4984, Apr. 2021, doi: 10.18653/v1/2021.naacl-main.395.

[22] M. Lewis et al., “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension,” Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880, 2020, doi: 10.18653/V1/2020.ACL-MAIN.703.

[23] Y. Cao, R. Shui, L. Pan, M. Y. Kan, Z. Liu, and T. S. Chua, “Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen,” Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 1061–1071, May 2020, doi: 10.18653/v1/2020.acl-main.100.

[24] N. Pattisapu, N. Prabhu, S. Bhati, and V. Varma, “Leveraging Social Media for Medical Text Simplification,” SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 851–860, Jul. 2020, doi: 10.1145/3397271.3401105.

[25] S. El-Sappagh and F. Ali, “DDO: a diabetes mellitus diagnosis ontology,” Applied Informatics 2016 3:1, vol. 3, no. 1, pp. 1–28, Aug. 2016, doi: 10.1186/S40535-016-0021-2.

[26] “SNOMED CT.” Accessed: Jan. 23, 2022. [Online]. Available: https://www.nlm.nih.gov/healthit/snomedct/index.html

[27] D. Vrandečić and M. Krötzsch, “Wikidata,” Commun ACM, vol. 57, no. 10, pp. 78–85, Sep. 2014, doi: 10.1145/2629489.

[28] J. Lee et al., “BioBERT: a pre-trained biomedical language representation model for biomedical text mining,” Bioinformatics, vol. 36, no. 4, pp. 1234–1240, Jan. 2019, doi: 10.1093/bioinformatics/btz682.

[29] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 4171–4186, Oct. 2018, Accessed: May 04, 2024. [Online]. Available: https://arxiv.org/abs/1810.04805v2

[30] “American Diabetes Association Standards of Medical Care in Diabetes-2018.” Accessed: Sep. 19, 2020. [Online]. Available: https://diabetesed.net/wp-content/uploads/2017/12/2018-ADA-Standards-of-Care.pdf

[31] “International Diabetes Federation.” Accessed: May 04, 2024. [Online]. Available: https://idf.org/

[32] “Division of Diabetes Translation | CDC.” Accessed: May 04, 2024. [Online]. Available: https://www.cdc.gov/diabetes/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdiabetes%2Findex.htm

[33] “T.E.R.A. - Science of Learning and Educational Technology.” Accessed: Jun. 22, 2024. [Online]. Available: https://soletlab.asu.edu/t-e-r-a/

[34] “Medical Transcription Samples.” Accessed: Nov. 06, 2021. [Online]. Available: https://www.medicaltranscriptionsamples.com/

[35] “protégé.” Accessed: Nov. 01, 2023. [Online]. Available: https://protege.stanford.edu/about.php