ENHANCING CHILDREN'S WEB SEARCHES THROUGH AGE-SPECIFIC VOCABULARY REFORMULATION

Improving Readability and Educational Relevance

AUTHOR

Author: Rembrandt Hazeleger
Supervisors: Sole Pera, Hrishita Chakrabarti

AFFILIATIONS

EEMCS, Delft University of Technology, The Netherlands

INTRODUCTION

We are seeing an increasing number of children using search engines for educational purposes [1]. Children often struggle to find and understand educational content online due to the complexity and irrelevance of web search results. This study addresses this issue by testing whether modifying search queries to include simpler, age-appropriate vocabulary improves both the readability and educational relevance of the retrieved content.

RESEARCH QUESTION

To what extent does incorporating age-specific vocabulary into search queries improve the retrieval effectiveness of age-appropriate educational web content?
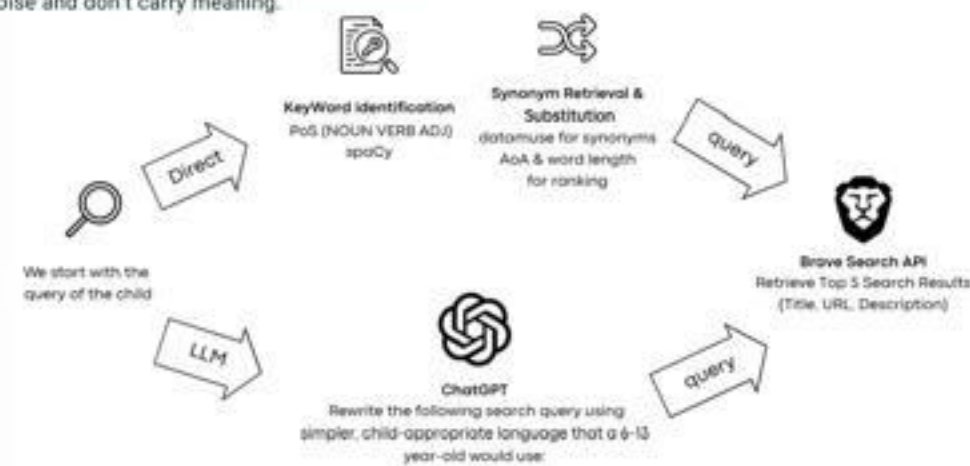
RELATED LITERATURE

Prior studies show that over 90% of content retrieved from children's queries exceeds recommended readability levels [2]. While systems like ReQuIK [3] attempt to generate child-friendly queries externally, few have focused on directly reformulating children's search inputs. Other research highlights how small changes in phrasing can greatly impact retrieved results [4], suggesting a promising intervention point in the query itself.

METHODOLOGY

To extract meaningful keywords from children's queries, we rely on part-of-speech tagging, keeping only nouns, verbs, and adjectives—as these best reflect the user's intent Wang et al [5] We exclude stopwords like "the" or "how" because they add noise and don't carry meaning.

Oshika et al. [6] shows that replacing late-acquired words with earlier-acquired synonyms significantly improved the readability of translated texts without compromising meaning.

We start with the query of the child

KeyWord Identification
PoS (NOUN VERB ADJ)
spaCy

Synonym Retrieval & Substitution
datamuse for synonyms
AoA & word length for ranking

Brave Search API
Retrieve Top 5 Search Results
(Title, URL, Description)

ChatGPT
Rewrite the following search query using simpler, child-appropriate language that a 6-13 year-old would use:

EXPERIMENTAL DESIGN

Educational Relevance:
- BiGBERT [7]
- based on URL and Description

Readability:
- Flesch-Kincaid Grade Level (general)
- Spache Readability Formula (children)
- based on the description

FINDINGS INDICATE A TRADE-OFF: NO SINGLE METHOD OPTIMALLY IMPROVES BOTH DIMENSIONS SIMULTANEOUSLY.

ANALYSIS

We observed the following:
- Rule-based:
  - No improvements on educational relevance
  - Better Spache scores.
- LLM-based:
  - More educationally relevant results.
  - No readability improvements, but showed more consistent readability levels.

EDUCATIONAL RELAVANCE

READABILITY

Flesch-Kincaid Grade

Spache Readability Score

DISCUSSION

Rule-Based Method

Improves Spache readability, but not FKGL, likely due to its design for technical adult readers [8]. Context loss from synonym replacement limits educational relevance [9].

LLM-Based Method

Better contextual understanding improves relevance and aligns content more closely with educational goals [10].

CONCLUSION

Key Insights:
Simplifying query vocabulary helps in retrieving appropriate content for children.
Rule-based improves Readability; LLM-based improves Educational value
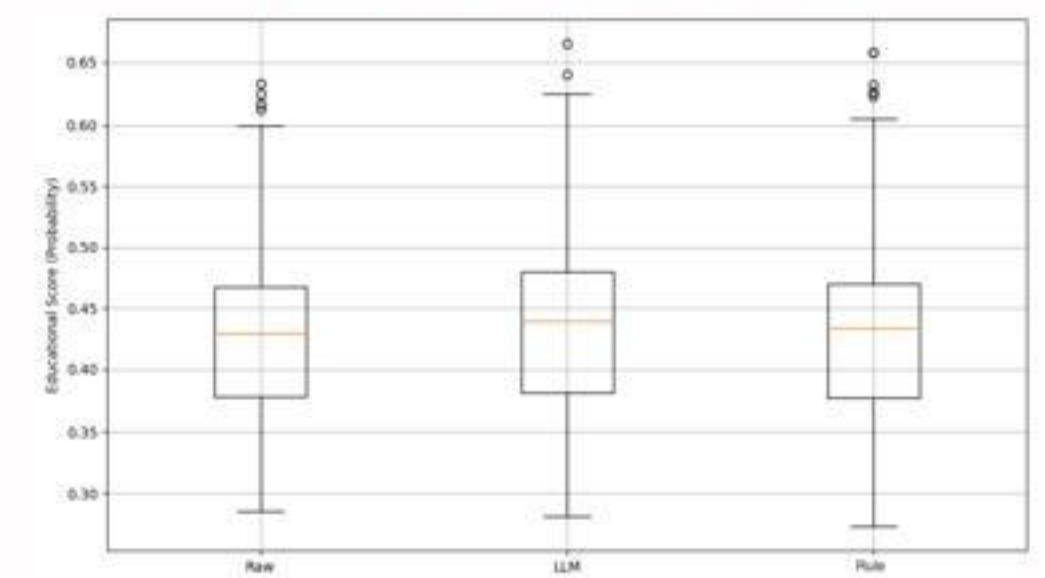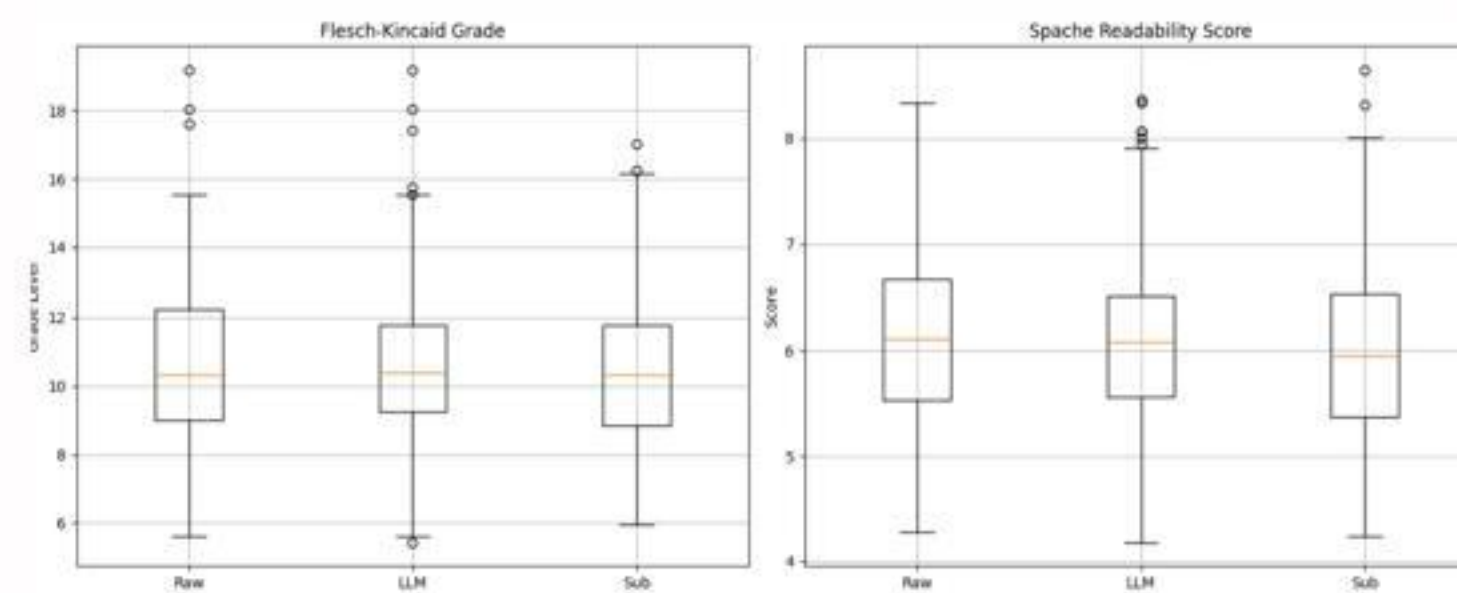
Implications:
Hybrid models may offer a promising future direction.
This research supports creating smarter, child-centred search technologies.

REFERENCES
[1] CARSTEN EICKHOFF, PIETER DEKKER, AND ARJEN P. DE VRIES. SUPPORTING CHILDREN'S WEB SEARCH IN SCHOOL ENVIRONMENTS. IN PROCEEDINGS OF THE 4TH INFORMATION INTERACTION IN CONTEXT SYMPOSIUM, IIIX '12, PAGE 129â€"137, NEW YORK, NY, USA, 2012. ASSOCIATION FOR COMPUTING MACHINERY.
[2] ISHITA BHUA. COMPARING GOOGLE'S READABILITY OF SEARCH RESULTS TO THE FLESCH READABILITY FORMULA: A PRELIMINARY ANALYSIS ON CHILDREN'S SEARCH QUERIES. PROCEEDINGS OF THE 78TH AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY (ASIS&T) ANNUAL MEETING 2015
[3] ION MADRAZO AZPIAZU, NEVENA DRAGOVIC, OGHENEMARO ANUYAH, AND MARIA SOLEDAD PERA. LOOKING FOR THE WORD: SEVEN OR TEN? FOR THE MORE FRIENDLY A MULTI-PERSPECTIVE STRATEGY FOR RECOMMENDING RANCHED FOR CHILDREN. INQWIR 2018 - PROCEEDINGS OF THE 2018 CONFERENCE ON HUMAN INFORMATION INTERACTION AND RETRIEVAL, PAGES 05 - 60, UNITED STATES, FEBRUARY 2018. ACM. 3RD ACM 2018 CONFERENCE ON HUMAN INFORMATION INTERACTION AND RETRIEVAL, CHIIR 2018 CONFERENCE DATE: 11-03-2018 THROUGH 15-03-2018.
[4] MARIA SOLEDAD PERA, EMILIANA MURGIA, MONICA LANDONI, THEO HUIBERS, AND MOHAMMAD ALIANNEJADI. WHERE A LITTLE CHANGE MAKES A BIG DIFFERENCE: A PRELIMINARY EXPLORATION OF CHILDREN'S QUERIES. IN JAAP KAMPS, LORRAINE GOEUROT, FLANE CRESTANI, MARIA MAISTRO, HIDEO JOHO, BRIAN DAVIS, CATHAL GURRIN, ANNALINA CAPUTO, AND UDO KRUSCHWITZ, EDITORS, ADVANCES IN INFORMATION RETRIEVAL, 45TH EUROPEAN CONFERENCE ON INFORMATION RETRIEVAL, ECIR 2023, PROCEEDINGS, LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS), PAGES 601- 620, SPRINGER, 2023.
[5] GREEN OPEN ACCESS ADDED
[6] HALIR J. WANG, ZHAIPING GAO, AND PETER ANDRÉAS. AUTOMATIC KEYWORD EXTRACTION FROM SHORT TEXTS APPLIED TO SPEECH-BASED NATURAL LANGUAGE QUERIES. IN PATRICIA ANTHONY, MITSURU ISHIZUKA, AND DICKSON LUKOSE, EDITORS, PRICAI 2012: TRENDS IN ARTIFICIAL INTELLIGENCE, PAGES 662-668, BERLIN, HEIDELBERG, 2012. SPRINGER BERLIN HEIDELBERG.
[7] MASASHI OSHIKA, MAKOTO MORISHITA, TSUTOMU HIRAO, RYOHEI SASANO, AND KOICHI TAKEDA. SIMPLIFYING TRANSLATIONS FOR CHILDREN: ITERATIVE SIMPLIFICATION CONSIDERING AGE OF ACQUISITION WITH LLMS. IN LUN-WEI KU, ANDRE MARTINS, AND VIVEK SRIKUMAR, EDITORS, FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: ACL 2024, PAGES 1542â€"1554, BANGKOK, THAILAND, AUGUST 2024. ASSOCIATION FOR COMPUTATIONAL LINGUISTICS.
[8] GARRETT ALLEN, BRODY DOWNS, APHILUTA DHANGUL, CASEY KENNINGTON, JERRY ALAN FAILS, KATHERINE LANDAU WRIGHT, AND MARIA SOLEDAD PERA. BIGBERT: CLASSIFYING EDUCATIONAL WEB RESOURCES FOR KINDERGARTEN-9TH GRADES. IN GLENDA HIENTINE, MARIO FRANCINE MORO, JOSIANE MOTHE, AND GABRIELA SEBASTIANI, EDITORS, ADVANCES IN INFORMATION RETRIEVAL, PAGES 176 - 184, CHAM, 2022. SPRINGER INTERNATIONAL PUBLISHING.
[9] GLENDA M. MCCLAAE, READABILITY FORMULAS: USEFUL OR USELESS? IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, PC-24(1):58, 1981.
[10] OLJAH NASSAR, MICHELLE ANNABA-NAJAR, AND GHOLAMRELDA HAFFARI. NEURAL REORDER MODEL FOR SIMPLIFICATION: A CASE STUDY IN HELLO1. MOSTELA, MASSIMO PICCARDI, AND ANDREW MACKINLAY, EDITORS, PROCEEDINGS OF THE 17TH ANNUAL WORKSHOP OF THE AUSTRALASIAN LANGUAGE TECHNOLOGY ASSOCIATION, PAGES 127 -127, SYDNEY, AUSTRALIA, 4-6 DECEMBER 2019. AUSTRALASIAN LANGUAGE TECHNOLOGY ASSOCIATION.
[11] VICTORIA BAKMAN, YISAI GOLDBERG, AND VELT SHAWLEY, LLMO READING COMPREHENSION IS AFFECTED BY PARAMETRIC KNOWLEDGE AND STRUGGLES WITH HYPOTHETICAL STATEMENTS, 2024.