

BÁNH XE ĐẠP CHO TÌM KIẾM TRÊN WEB:

HỌC TẬP ĐA GÓC NHÌN ĐỂ XẾP HẠNG

HỖ TRỢ TRẺ EM TÌM KIẾM THÔNG TIN

LỚP HỌC

qua

Garrett Allen



Một luận án

nộp một phần để hoàn thành

của các yêu cầu về mức độ

Thạc sĩ Khoa học Máy tính

Đại học Boise State

Tháng 12 năm 2021

khoảng năm 2021

Garrett Allen

MỌI QUYỀN ĐƯỢC BẢO LƯU

TRƯỜNG CAO ĐẲNG ĐẠI HỌC BANG BOISE

ỦY BAN QUỐC PHÒNG VÀ PHÊ DUYỆT ĐỌC CUỐI CÙNG

của luận án được nộp bởi

Garrett Allen

Tiêu đề luận án: Bánh xe tập luyện cho tìm kiếm trên web: Học xếp hạng đa góc nhìn để hỗ trợ trẻ em tìm kiếm thông tin trong lớp học

Ngày thi vấn đáp cuối kỳ:

Ngày 18 tháng 12 năm 2021

Những cá nhân sau đây đã đọc và thảo luận luận án do sinh viên Garrett nộp Allen, và họ đã đánh giá bài thuyết trình và phản hồi các câu hỏi trong phần cuối kỳ thi vấn đáp. Họ thấy rằng sinh viên đã vượt qua kỳ thi vấn đáp cuối cùng.

Tiến sĩ Maria Soledad Pera

Chủ tịch, Ủy ban giám sát

Tiến sĩ Jerry Alan Fails

Thành viên, Ủy ban giám sát

Casey Kennington, Tiến sĩ

Thành viên, Ủy ban giám sát

Katherine Landau Wright, Tiến sĩ

Thành viên, Ủy ban giám sát

Sự chấp thuận đọc cuối cùng của luận án đã được chấp thuận bởi Maria Soledad Pera, Tiến sĩ, Chủ tịch Ủy ban giám sát. Luận án đã được Hội đồng sau đại học chấp thuận

Trường cao đẳng.

LỜI CẢM ƠN

Trong suốt quá trình viết luận án này, tôi đã nhận được rất nhiều sự hỗ trợ và hỗ trợ.

Đầu tiên, tôi muốn cảm ơn cố vấn của tôi, Tiến sĩ Sole Pera, người có sự kiên nhẫn đáng kinh ngạc, kiến thức và niềm tin vào tôi luôn thúc đẩy tôi phát huy hết khả năng của mình. tác động của bạn đến công việc này là không thể phủ nhận, cảm ơn bạn.

Tôi cũng muốn cảm ơn Ashlee Milton, người đã sẵn lòng hợp tác, thảo luận và vượt qua những rào cản khó khăn đã đảm bảo rằng luận án này đã thành công kết thúc. Tôi rất trân trọng sự hy sinh thời gian của bạn cho mục tiêu của tôi.

Cuối cùng, tôi muốn cảm ơn cha mẹ tôi, những người đã luôn tin tưởng tôi và cho tôi sức mạnh để không dao động trên con đường của tôi. Tôi yêu, và mãi mãi biết ơn, cả hai Bạn.

TÓM TẮT

Thiết kế xe đạp đã không thay đổi trong một thời gian dài, vì chúng được chế tạo tốt cho những người có kỹ năng cưỡi ngựa, tức là người lớn. Tuy nhiên, những người học cưỡi ngựa thường cần hỗ trợ thêm dưới dạng bánh xe tập đi. Tìm kiếm thông tin trên Web cũng giống như đi xe đạp, nơi các công cụ tìm kiếm hiện đại (xe đạp) được tối ưu hóa cho mục đích sử dụng chung và người dùng trưởng thành, nhưng thiếu chức năng hỗ trợ khán giả và môi trường không theo truyền thống. Trong luận án này, chúng tôi giới thiệu một bộ bánh xe tập đi dưới dạng mô hình học xếp hạng như một sự tăng cường cho các tiêu chuẩn công cụ tìm kiếm dard để hỗ trợ các hoạt động tìm kiếm trong lớp học dành cho trẻ em (từ 6 đến 11 tuổi). Mô hình mới này mở rộng danh sách học tập đã biết để xếp hạng khuôn khổ thông qua sự cân bằng giữa rủi ro và phần thưởng. Làm như vậy cho phép mô hình ưu tiên Web nguồn tài nguyên có sự liên kết giáo dục cao, phù hợp và dễ đọc bằng cách phân tích các URL, đoạn trích và tiêu đề trang của các tài nguyên Web được truy xuất bởi một công cụ tìm kiếm chính thống. Các thí nghiệm bao gồm nghiên cứu cắt bỏ và so sánh với các đường cơ sở hiện có cho thấy tính đúng đắn của mô hình được đề xuất. Kết quả của công trình này chứng minh giá trị của việc xem xét nhiều quan điểm vốn có của bối cảnh lớp học, ví dụ, sự liên kết về giáo dục, khả năng đọc hiểu và khả năng phản đối, khi được áp dụng vào việc thiết kế các thuật toán có thể hỗ trợ thông tin cho trẻ em tốt hơn khám phá.

MỤC LỤC

TÓM TẮT v

DANH SÁCH BẢNG viii

DANH SÁCH HÌNH ẢNH ix

DANH SÁCH CÁC CHỮ VIẾT TẮT x

1 Giới thiệu 1

2 Công trình liên quan 7

 2.1 Học cách xếp hạng 7

 2.2 Xếp hạng các nguồn tài nguyên web dành cho trẻ em. 10

3 Phương pháp luận 13

 3.1 Quan điểm: Từ lý thuyết đến thực hành. 14

 3.1.1 Khả năng đọc 14

 3.1.2 Tài nguyên giáo dục 16

 3.1.3 Tài nguyên phản cảm 20

 3.2 REdORank: Từ lý thuyết đến thực hành 26

 3.2.1 Xếp hạng Ada. 26

 3.2.2 Tối ưu hóa đa góc nhìn với độ nhạy về chi phí. 29

4 Kết quả thực nghiệm	32
4.1 Tìm công thức khả năng đọc phù hợp với tài nguyên web.	32
4.2 Phát hiện tài nguyên web giáo dục.	39
4.2.1 Thiết lập thử nghiệm.	39
4.2.2 Kết quả và thảo luận	41
4.3 Xác định các nguồn tài nguyên đáng ngờ.	44
4.3.1 Thiết lập thử nghiệm.	44
4.3.2 Kết quả và thảo luận	46
4.4 Xếp hạng lại các tài nguyên web bằng REdORank.	48
4.4.1 Thiết lập thử nghiệm.	48
4.4.2 Kết quả và thảo luận	50
5 Kết luận	55
TÀI LIỆU THAM KHẢO	59
Dữ liệu cho các thí nghiệm	80

DANH SÁCH CÁC BẢNG

4.1 Phân tích hiệu suất của BiGBERT (nghiên cứu cắt bỏ cùng với kinh nghiệm các ý liên quan đến việc so sánh với các đối tác). Các hậu tố -U và -S chỉ ra mô hình chỉ áp dụng cho URL và đoạn trích, tương ứng; -E chỉ ra mô hình được tăng cường với dữ liệu giáo dục. * và † đáng kể liên quan đến BiGBERT và đối tác không giáo dục, tương ứng. Ý nghĩa được xác định bằng kiểm định McNemar, $p < 0,05$	41
4.2 Đánh giá các mô hình phân loại có thể phản đối bằng cách sử dụng ObjSet. * chỉ ra ý nghĩa liên quan đến Judgebad được xác định bởi thử nghiệm của McNemar với hiệu chỉnh Bonferroni, $p < 0,05$	47
4.3 Hiệu suất của REORank và các biến thể cắt bỏ bằng RankSet. Các hậu tố -R, -E, -O chỉ ra Chỉ có tính dễ đọc, Chỉ có tính giáo dục, và Chỉ có thể phản đối, tương ứng. -M chỉ ra việc sử dụng máy trộn để căn chỉnh giáo dục và khả năng đọc, và -MER chỉ ra việc sử dụng của bộ trộn với -E và -R. * biểu thị ý nghĩa liên quan đến REORank và chữ in đậm biểu thị hiệu suất tốt nhất cho từng số liệu.	52
4.4 Hiệu suất của REORank và đường cơ sở sử dụng RankSet. * biểu thị ý nghĩa wrtREORank và in đậm cho biết hiệu suất tốt nhất cho mỗi mét.	53

DANH SÁCH CÁC HÌNH ẢNH

2.1 Khung xác định vấn đề học xếp hạng [52]. 8

3.1 Khung RE₀Rank. RE₀Rank xếp hạng lại các tài nguyên Web được truy xuất
từ một SE chính thống để trả lời câu hỏi của một đứa trẻ được đưa ra trong một
thiết lập lớp học bằng cách cân bằng phần thưởng với rủi ro. 13

3.2 Kiến trúc BiGBERT (RU và RS biểu thị URL và đoạn trích của
cho trước nguồn tài nguyên R, tương ứng). 18

4.1 Phân phối ER trên các lớp khác nhau cho các công thức khả năng đọc truyền thống.36

4.2 Phân bố ER cho các công thức lấy cảm hứng từ Spache ở các cấp độ. 38

4.3 NDCG@10 cho các mô hình xếp hạng lại khác nhau sử dụng RankSet. 54

DANH SÁCH CÁC TỪ VIẾT TẮT

AoA - Thời đại của sự tiếp thu

BiGRU - Đơn vị hồi quy có cổng song hướng

CCSS - Tiêu chuẩn chung của tiểu bang

CNN - Mạng nơ-ron tích chập

CS-DCG - Lợi nhuận tích lũy chiết khấu nhạy cảm với chi phí

DCG - Lợi nhuận tích lũy được chiết khấu

ER - Tỷ lệ lỗi

FNR - Tỷ lệ âm tính giả

FPR - Tỷ lệ dương tính giả

GCS - Tìm kiếm tùy chỉnh của Google

ICS - Tiêu chuẩn nội dung của Idaho

IDLA - Liên minh học tập kỹ thuật số Idaho

IR - Truy xuất thông tin

LTR - Học cách xếp hạng

MLM - Mô hình ngôn ngữ che giấu

MP-LTR - Học xếp hạng đa góc nhìn

MRR - Xếp hạng tương hỗ trung bình

NDCG - Chuẩn hóa lợi nhuận tích lũy đã chiết khấu

NGSS - Tiêu chuẩn khoa học thế hệ tiếp theo

ODP - Dự án thư mục mở

RAZ - Đọc AZ

SE - Công cụ tìm kiếm

SERP - Trang kết quả của công cụ tìm kiếm

URL - Bộ định vị tài nguyên thống nhất

CHƯƠNG 1

GIỚI THIỆU

Trẻ em ở các lớp tiểu học (Mẫu giáo- lớp 4) thường sử dụng công cụ tìm kiếm

(SE) để tìm các tài nguyên Web cần thiết để hoàn thành bài tập ở trường của họ [13, 105].

Trong số các SE được xây dựng dành riêng cho trẻ em sử dụng trong môi trường lớp học, chúng tôi thấy

các giải pháp hiện tại yêu cầu bảo trì thường xuyên, chẳng hạn như EdSearch1 và Kidtopia2 .

EdSearch dựa vào việc quản lý thủ công các nguồn tài nguyên (ví dụ: văn bản hoặc phương tiện) để xác định

giáo dục. Kidtopia thay vào đó cung cấp các nguồn tài nguyên từ một số nguồn được liệt kê trong danh sách trắng

các trang web sử dụng nền tảng Tìm kiếm tùy chỉnh (GCS) của Google, sử dụng Tìm kiếm an toàn

tính năng lọc ra các tài nguyên khiêu dâm. Danh sách trắng thông qua việc quản lý thủ công

hạn chế các trang web phải phù hợp với lứa tuổi và mang tính giáo dục, nhưng khi Web phát triển

với tốc độ nhanh chóng, việc duy trì danh sách trắng được cập nhật trở nên nặng nề. Hơn nữa,

SE của trẻ em dựa trên GCS được biết là trả về kết quả ít liên quan hơn gần 30%

thời gian, trao đổi sự liên quan để có kết quả an toàn hơn [42]. Ngoài những sự kém hiệu quả này,

SE chuyên biệt cũng phải vượt qua rào cản của việc áp dụng: trẻ em thích sử dụng

các lựa chọn phổ biến cho SE, chẳng hạn như Google hoặc Bing [32].

SE chính thống được thiết kế và tối ưu hóa cho người lớn và do đó có thể bỏ qua

các yếu tố độc đáo tác động đến trẻ em sử dụng chúng. Ví dụ, trẻ em phải đối mặt với nhiều

những rào cản liên quan đến việc xây dựng truy vấn, một số trong đó các nhà nghiên cứu đã cố gắng

1<https://www.lumoslearning.com/llwp/edsearch.html>

2<https://www.kidtopia.info/>

để cung cấp hỗ trợ cho [16, 82, 123]. Ngoài các rào cản được nghiên cứu kỹ lưỡng liên quan đến truy vấn xây dựng, trẻ em cũng phải vật lộn để nhận ra thông tin gì và bao nhiêu có sẵn trực tuyến, hiếm khi xem qua sáu nguồn đầu tiên được trình bày trên một tìm kiếm trang kết quả của công cụ tìm kiếm (SERP) [43]. Trẻ em cũng gặp khó khăn trong việc hiểu nội dung của các nguồn tài nguyên được thu thập do tính phức tạp của văn bản của chúng, dẫn đến sự không chắc chắn với lựa chọn tài nguyên có liên quan [7]. Khi chuyển sang SE chính thống, trẻ em cũng có thể vô tình tiếp xúc với những nguồn tài nguyên không phù hợp để tiêu thụ.

Đây là một tác dụng phụ đáng tiếc của chức năng, như Tìm kiếm an toàn của Google, được cung cấp bởi SE chính thống chủ yếu lọc nội dung khiêu dâm [131] và không tính đến các nội dung khác nội dung có khả năng gây hại, ví dụ như bạo lực. Chức năng tìm kiếm an toàn cũng bị ảnh hưởng lọc quá mức bằng cách ngăn chặn các tài nguyên được trả về nếu chúng chứa các thuật ngữ có thể bị nhầm là không phù hợp [12].

Chúng tôi hướng tới mục tiêu nâng cao kiến thức trong lĩnh vực Truy xuất thông tin (IR) cho trẻ em, và cụ thể hơn là giúp trẻ em tiếp cận thông tin trực tuyến tốt hơn qua SE. Khi lớn lên, trẻ em cần các mức hỗ trợ khác nhau từ SE mà chúng tương tác với. Là điểm khởi đầu trong quá trình khám phá của chúng tôi, chúng tôi tập trung vào việc điều chỉnh SERP cho đối tượng và bối cảnh cụ thể. Để xác định phạm vi công việc của chúng tôi, chúng tôi chuyển sang khuôn khổ từ [70] cho phép thiết kế và đánh giá toàn diện hệ thống tìm kiếm cho trẻ em thông qua bốn trụ cột. Được định nghĩa cho công việc của chúng tôi, các trụ cột này là: trẻ em từ 6 đến 11 tuổi ở lớp Mẫu giáo đến lớp 4 (K-4) là nhóm người dùng, lớp học là môi trường, khám phá thông tin như một nhiệm vụ và xếp hạng lại các nguồn lực để phù hợp đối tượng và bối cảnh như một chiến lược. Được hướng dẫn bởi các trụ cột, chúng tôi giới thiệu REORank, một khuôn khổ xếp hạng lại mới dựa trên học tập đa góc nhìn để xếp hạng (LTR) có nghĩa là hỗ trợ trẻ em sử dụng SE ưa thích của mình để hoàn thành các nhiệm vụ liên quan đến lớp học.

REORank xây dựng dựa trên hiệu quả truy xuất đã được chứng minh của SE chính thống, cũng như

vì khả năng phản hồi của họ đối với bất kỳ truy vấn nào do nguồn cung cấp được lập chỉ mục lớn của họ trang [75]. Với truy vấn tìm kiếm của trẻ em, REdORank sẽ kiểm tra và xếp hạng lại các tài nguyên được SE chính thống thu thập theo cách mà những thứ phù hợp với giáo dục tiêu chuẩn và trẻ em có thể đọc được được xếp hạng cao hơn, trong khi những tiêu chuẩn có chứa tài liệu không phù hợp để trẻ em xem trong lớp học được đẩy xuống thấp hơn trong danh sách kết quả.

Để REdORank tìm hiểu cách ưu tiên các nguồn lực sao cho phù hợp nhất với nhóm người dùng của chúng tôi và môi trường, nó kiểm tra ba góc nhìn riêng biệt: khả năng đọc, tức là “tổng thể tác động của việc sử dụng ngôn ngữ và sáng tác lên khả năng dễ dàng và nhanh chóng của người đọc hiểu tài liệu” [85], sự liên kết về giáo dục và khả năng phản đối. Các nguồn lực phù hợp với giáo dục được định nghĩa là các nguồn lực phù hợp với các hướng dẫn được trình bày trong Tiêu chuẩn chung của tiểu bang (CCSS). Các hướng dẫn này cung cấp một bộ kết quả học tập cho mỗi lớp K-12 mà học sinh được kỳ vọng đạt được. Ví dụ, kết quả học tập lớp 1 từ CCSS nêu rõ “Xác định mục tiêu chính chủ đề và kể lại các chi tiết chính của một văn bản” [63]. Mặt khác, các nguồn tài nguyên phản đối là những tài nguyên có chứa nội dung vượt ra ngoài nội dung khiêu dâm không phù hợp với trẻ em trong lớp học. Cả sự liên kết giáo dục và khả năng đọc đều đóng vai trò là “phần thưởng” quan điểm, tức là các tín hiệu cần được tối ưu hóa để tăng thứ hạng của một nguồn lực, trong khi khả năng phản đối đóng vai trò là quan điểm “rủi ro”. Một chiến lược xếp hạng được gọi là LTR đa góc nhìn được sử dụng khi mô hình xếp hạng học một hàm xếp hạng ưu tiên các nguồn lực cho nhiều hơn một góc nhìn. Bằng cách ưu tiên các nguồn lực phù hợp với mức độ dễ đọc của nhóm người dùng của chúng tôi, REdORank mang lại lợi ích cho mục tiêu của chúng tôi đối tượng là trẻ em đọc lại trình độ đọc của mình sẽ trải nghiệm đọc kém hơn hiểu biết [8].

Việc ước tính mức độ của các nguồn tài nguyên trực tuyến không phải là vấn đề đơn giản, xét đến phạm vi rộng

phạm vi các công thức có sẵn để ước tính khả năng đọc hoặc trình độ lớp. Ngoài ra, không có sự đồng thuận về việc công thức nào trong số các công thức có sẵn nên được sử dụng cho trực tuyến nguồn lực. Do đó, chúng tôi kiểm tra các công thức và đôn bẫy trong REORank một hiệu quả nhất để dự đoán mức độ dễ đọc của các nguồn tài nguyên nhắm mục tiêu đọc khả năng điển hình của trẻ em từ 6 đến 11 tuổi.

Đáp ứng với môi trường của chúng tôi, REORank xem xét sự liên kết giáo dục của nguồn lực và mục đích thúc đẩy những nguồn lực có giá trị giáo dục như nghiên cứu trước đây đã làm cho thấy rằng việc xếp hạng các nguồn tài nguyên giáo dục cao hơn trong kết quả tìm kiếm có khả năng tăng hiệu quả học tập [117]. REORank xác định sự liên kết giáo dục thông qua phân tích URL và đoạn trích của tài nguyên bằng cách sử dụng các chiến lược biểu diễn văn bản như như nhúng miền cụ thể và BERT [11, 33].

Như đã nêu trước đó, không phải mọi thứ trên Web đều phù hợp với trẻ em, điều đó đưa chúng ta đến với quan điểm phản đối. Ngăn chặn việc hiển thị không phù hợp kết quả trong khi cũng tránh lọc quá mức các kết quả có thể xuất hiện như phản cảm nhưng không phải là, ví dụ, một bài viết về ung thư vú [42], đòi hỏi một giải pháp vượt ra ngoài tìm kiếm an toàn. Do đó, chúng tôi vượt ra ngoài nội dung khiêu dâm và xem xét các nguồn khác sự phản đối, chẳng hạn như bạo lực, ma túy hoặc súng. REORank sử dụng một cách tiếp cận áp dụng chi phí cho thứ hạng của một nguồn lực dựa trên khả năng xác định rằng một nguồn lực là đáng phản đối. Bằng cách học cách đồng thời tối đa hóa các nguồn lực với giáo dục căn chỉnh và khả năng đọc, đồng thời giảm thiểu những thứ có thể phản đối, REORank là rất phù hợp để hỗ trợ các hoạt động tìm kiếm của trẻ em trong lớp học.

Chúng tôi đưa ra giả thuyết rằng chiến lược LTR có thể được tăng cường để đồng thời xem xét nhiều đặc điểm của các nguồn tài nguyên trực tuyến để tạo ra một SERP ưu tiên giáo dục các nguồn tài nguyên có giá trị và dễ hiểu đồng thời giảm thiểu những nguồn tài nguyên không phù hợp. Để hướng dẫn công việc liên quan đến các chủ đề này, luận án này giải quyết những vấn đề sau:

Câu hỏi nghiên cứu:

1. Công thức khả năng đọc nào phù hợp đồng thời với loại tài nguyên, nhu cầu ngữ cảnh,

và nhóm người dùng được phác thảo cho nhiệm vụ của chúng tôi?

2. Các đoạn trích cùng với URL có giúp xác định các nguồn tài nguyên giáo dục không?

Kiến thức cụ thể ảnh hưởng đến việc xác định nguồn lực giáo dục?

3. Từ điển theo chủ đề cụ thể có thể trao quyền xác định Web phản cảm không?

tài nguyên?

4. Việc áp dụng mô hình LTR để tính đến nhiều quan điểm có dẫn đến

để ưu tiên các nguồn lực có liên quan đến cả trẻ em và

bối cảnh lớp học?

Đóng góp chính của công trình này là một khuôn khổ LTR được tối ưu hóa trên nhiều quan điểm đồng thời cho việc tìm kiếm của trẻ em, theo hiểu biết tốt nhất của chúng tôi là khuôn khổ LTR đa góc nhìn đầu tiên như vậy. Những đóng góp tiếp theo bao gồm một mô hình để xác định sự liên kết giáo dục của các nguồn tài nguyên trực tuyến, việc xác định của một công thức khả năng đọc hiệu quả để tính toán mức độ đọc trực tuyến tài nguyên và một mô hình xác định các tài nguyên đáng ngờ vượt quá phạm vi hạn chế của tìm kiếm an toàn. Công việc của chúng tôi có thể giúp tạo điều kiện thuận lợi cho trẻ em tiếp cận nội dung giáo dục trực tuyến và do đó hỗ trợ hướng dẫn trong lớp học. Trên thực tế, REORank có thể được sử dụng trong kết hợp với bất kỳ SE nào và khi kết hợp như vậy có thể cung cấp hỗ trợ cho tìm kiếm như học tập giữa học sinh lớp K-4 [59, 120, 129]. Mô hình liên kết giáo dục có khả năng hỗ trợ giáo viên xác định các nguồn tài nguyên trực tuyến để tận dụng trong lớp học [39]. Cuối cùng, xác định một công thức ước tính chính xác việc đọc

mức độ tài nguyên trực tuyến có thể cung cấp thông tin cho việc thiết kế hệ thống đề xuất hoặc các hệ thống khác trong tương lai hệ thống trực tuyến được thiết kế riêng cho độc giả trẻ [74].

Phần còn lại của bản thảo này được tổ chức như sau. Trong Chương 2, chúng tôi cung cấp lại thông tin cơ bản liên quan đến LTR cho tìm kiếm trên Web và thảo luận về các chiến lược xếp hạng nhằm mục đích hỗ trợ trẻ em sử dụng SE. Sau đó, trong Chương 3, chúng tôi trình bày chi tiết thiết kế của REdORank; tiếp theo là phân tích thực nghiệm chuyên sâu được trình bày trong Chương 4, chúng tôi tiến hành để xác minh hiệu suất của khuôn khổ xếp hạng lại của chúng tôi và đánh giá nhu cầu đưa tất cả các quan điểm vào bảng xếp hạng. Cuối cùng, trong Chương 5, chúng tôi trình bày một số nhận xét kết luận, hạn chế và hướng nghiên cứu trong tương lai lấy cảm hứng từ công trình trình bày trong luận án này.

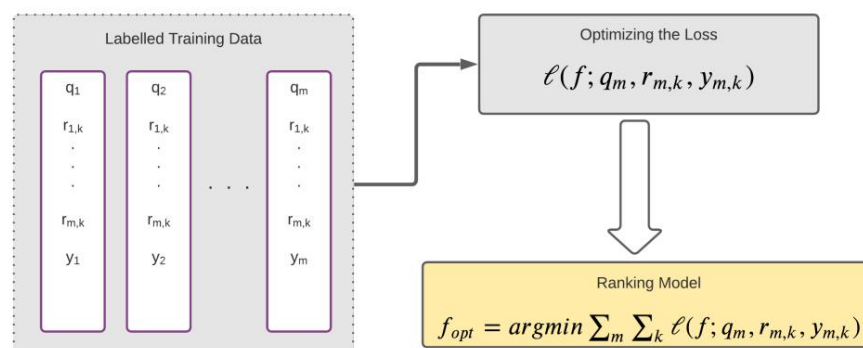
CHƯƠNG 2

CÔNG VIỆC LIÊN QUAN

Trong chương này, chúng tôi cung cấp thông tin cơ bản về LTR và thảo luận về các chiến lược xếp hạng giúp điều chỉnh việc (lấy và) xếp hạng các nguồn tài nguyên cho trẻ em.

2.1 Học cách xếp hạng

LTR là một chiến lược học máy, khi được áp dụng vào việc Truy xuất thông tin, tạo ra một nhiệm vụ trong đó mục tiêu là tự động xác định một mô hình xếp hạng bằng cách sử dụng dữ liệu đào tạo, sao cho mô hình được xây dựng có thể sắp xếp các tài nguyên mới bằng cách sử dụng một chức năng xếp hạng theo mức độ liên quan, sở thích hoặc tầm quan trọng của tài nguyên tance [79]. LTR có thể được thể hiện dưới dạng các truy vấn, nhãn và tài nguyên. Cho một tập hợp m truy vấn $Q = \{q_1, \dots, q_m\}$, tồn tại một tập hợp k tài nguyên $R_m = \{r_{m,1}, \dots, r_{m,k}\}$ cho truy vấn trong q_m . Tương tự, tồn tại một tập hợp các nhãn $y_m = \{y_{m,1}, \dots, y_{m,k}\}$ cho mỗi tài nguyên R_m . Hãy để $f(q_m, r_{m,k})$ là một hàm xếp hạng tính toán điểm xếp hạng cho một cặp truy vấn-tài liệu và để $(f; q_m, r_{m,k}, y_{m,k})$ là hàm mất mát cho dự đoán của hàm f trên cặp truy vấn-tài liệu $(q_m, r_{m,k})$. Vấn đề LTR có thể sau đó được định nghĩa là tìm kiếm hàm xếp hạng tối ưu f_{opt} (trong Phương trình 2.1) thông qua việc giảm thiểu hàm mất mát trên một tập huấn luyện được gán nhãn [76, 79, 52]. Chúng tôi mô tả khuôn khổ cho định nghĩa vấn đề này trong Hình 2.1.



Hình 2.1: Khung xác định vấn đề học xếp hạng [52].

$$f_{opt} = \underset{\text{đánh dấu}}{\operatorname{argmin}} (f; q_m, r_{m,k}, y_{m,k}) \quad (2.1)$$

Theo thời gian, những tiến bộ trong các mô hình LTR đã mở rộng chức năng mất mát thành chấp nhận nhiều hơn một nguồn lực làm đầu vào. Kết quả là, các phân loại sau đây đối với các mô hình LTR đã phát sinh: từng điểm, từng cặp hoặc từng danh sách [76], dựa trên việc một nguồn tài nguyên duy nhất, một cặp tài nguyên hoặc một danh sách các nguồn tài nguyên, tương ứng, được vận hành trong quá trình tối ưu hóa hàm mất mát.

Bất kể chúng thuộc loại nào, các mô hình LTR đều đã thành công được áp dụng cho nhiều lĩnh vực khác nhau của IR, chẳng hạn như trả lời câu hỏi [34], truy xuất tài liệu [79], khuyến nghị [77, 96, 128] và nổi bật nhất là tìm kiếm trên Web [81, 95]. Khi được sử dụng cho tìm kiếm trên Web, các mô hình sử dụng hàm mất mát theo danh sách đã được chứng minh là hiệu quả hơn có hiệu quả về mặt độ chính xác xếp hạng và mức độ chắc chắn của độ chính xác xếp hạng trong mối quan hệ với các điểm tương ứng và các cặp tương ứng [20, 118]. Tồn tại một số của các mô hình dựa trên danh sách nổi tiếng, bao gồm AdaRank [127], ListNet [20], ListMLE [125], online-listMLE [80], SetRank [97] và U-Rank [30]. Mỗi mô hình này đưa ra một bước tiến tới việc nâng cao kiến thức liên quan đến LTR, nhưng tất cả đều tối ưu hóa các chức năng xếp hạng tương ứng trên một thước đo mức độ liên quan duy nhất.

Trong thực tế, mức độ liên quan của kết quả tìm kiếm không phải lúc nào cũng được xác định dựa trên một đặc điểm duy nhất. Ví dụ, người dùng tìm kiếm một nhà hàng hải sản cho bữa tối sẽ xem xét vị trí, giá cả và danh tiếng như những yếu tố quyết định sự phù hợp. Sinh viên tìm kiếm thông tin về John Adams cho bài tập trên lớp sẽ thay vào đó xác định sự liên quan của tài nguyên bằng cách xem xét các yếu tố như liệu một tài nguyên sử dụng ngôn ngữ mà họ có thể hiểu được, cho dù John Adams đang được thảo luận là cá nhân chính xác, và liệu tài nguyên có thảo luận về khía cạnh của John Adams hay không mà họ đang tìm kiếm thông tin, tức là thông tin về nhiệm kỳ của ông với tư cách là Tổng thống so với thông tin về vai trò của ông trong Cách mạng Hoa Kỳ. Để phù hợp hơn với các kịch bản thực tế như vậy, các chiến lược LTR đa mục tiêu tối ưu hóa các hàm mất mát vì nhiều biện pháp liên quan đã được đưa ra [18, 116, 122]. Carmel et al. [21] sử dụng tổng hợp nhân để giảm một vấn đề đa mục tiêu thành một mục tiêu một, tiếp theo là áp dụng LambdaMART [19] để tối ưu hóa cho tổng hợp nhân. Momma et al. [90] cũng sử dụng LambdaMART, kết hợp Augmented Lagrangian, một quá trình đưa một hệ số nhân Lagrange rõ ràng vào tổn thất chức năng được tối ưu hóa [93], để tạo ra một mô hình xử lý tối ưu hóa bị hạn chế bằng cách "giải quyết lặp đi lặp lại các vấn đề không bị ràng buộc". Trong khi các chiến lược này mở rộng LTR từ mục tiêu đơn đến mục tiêu đa, cả hai vẫn lựa chọn cách tiếp cận từng cặp. Khi tính đến nhiều mục tiêu, các phương pháp tiếp cận theo danh sách như AdaRank hiếm khi được xem xét. Với khả năng áp dụng của AdaRank cho các tác vụ Web [20, 118], đây là LTR biến thể mà chúng tôi kết hợp như một phần trong thiết kế của REDORank.

2.2 Xếp hạng các nguồn tài nguyên web dành cho trẻ em

Khi tìm kiếm thông tin bằng SE chính thống, trẻ em có xu hướng (i) khám phá

SERP được tạo ra để đáp ứng các truy vấn của họ bằng cách sử dụng một quy trình tuần tự từ trên xuống dưới

dưới cùng và (ii) nhấp vào các kết quả được xếp hạng cao hơn [35, 43, 50, 53]. Do đó, điều bắt buộc là

để các SE chính thống này ưu tiên các nguồn tài nguyên đã thu thập có liên quan đến thông tin

nhu cầu của trẻ em. Những nỗ lực hiện tại để giải quyết yêu cầu này bao gồm công việc của

Miltsakaki [88], người sắp xếp các nguồn tài nguyên theo mức độ đọc do người dùng xác định (đối với

học sinh trung học cơ sở và trung học phổ thông) và khả năng đọc của tài nguyên được tính toán bằng cách sử dụng

Chỉ số Coleman-Liau [25] cùng với các công thức LIX và RIX [9]. Tương tự như vậy,

Collins-Thompson et al. [26] xếp hạng lại kết quả phù hợp với mức độ đọc của người dùng được suy ra từ

lịch sử tìm kiếm của họ.

Ngoài khả năng đọc, Gyllstrom và Moens [54] giới thiệu AgeRank, một sửa đổi

phiên bản PageRank tận dụng các trang web dành cho đối tượng khán giả trẻ tuổi, theo sau

tiền đề rằng các trang web được thiết kế cho trẻ em có nhiều khả năng liên kết đến các trang web thân thiện với trẻ em khác

trang web. Syed và Collins-Thompson [117] trình bày một thuật toán tìm kiếm xếp hạng lại

kết quả cho tiện ích học tập thông qua phân tích mật độ từ khóa, giả sử rằng

người dùng tiếp xúc với nhiều từ khóa hơn trong ít tài nguyên hơn sẽ tìm hiểu thông tin về một

chủ đề thành công hơn. Các chiến lược đã đề cập ở trên ưu tiên các nguồn lực bằng cách sử dụng

chỉ là một góc nhìn duy nhất, nhưng khi phục vụ một nhóm người dùng cụ thể trong một bối cảnh cụ thể,

chỉ xem xét một góc nhìn sẽ hạn chế kết quả có thể thu được. Chúng tôi

đưa ra giả thuyết rằng bằng cách kết hợp nhiều quan điểm hơn, chẳng hạn như sự liên kết về giáo dục

và khả năng đọc, một tập hợp kết quả đa dạng hơn có thể được cung cấp để phục vụ tốt hơn

nhóm người dùng và bối cảnh.

Nghiên cứu liên quan đến xếp hạng dựa trên giáo dục rất phong phú, đưa ra các chiến lược

dựa trên mô hình chủ đề, nhóm thuật ngữ, chỉ số chất lượng và hồ sơ cộng tác tering [101, 109, 111, 100]. Các ví dụ đáng chú ý bao gồm công trình của Marani [83], tức là, WebEduRank, người định nghĩa bối cảnh giảng dạy (một biểu diễn của các yêu cầu và kinh nghiệm của người hướng dẫn), được sử dụng để xếp hạng các đối tượng học tập nhằm hỗ trợ giảng viên. Estivill-Castro và Marani [41] cũng xếp hạng các nguồn lực cho giảng viên theo phân tích tính phù hợp của một nguồn tài nguyên để giảng dạy một khái niệm. Acuña-Soto et al. [2] coi học sinh là một phần khán giả trong công việc của họ để xếp hạng các video toán học bằng cách sử dụng một khuôn khổ ra quyết định đa tiêu chí. Thật không may, cũng như khả năng đọc và thân thiện với trẻ em, một số tác phẩm này không nhắm đến trẻ em là đối tượng người dùng dự định nhóm và phần lớn tập trung vào một quan điểm duy nhất.

Tập trung vào trẻ em trong bối cảnh giáo dục, Usta et al. [121] đào tạo một LTR mô hình cho một chiến lược xếp hạng phụ thuộc vào truy vấn nhằm mục đích ưu tiên giáo dục nguồn cho học sinh lớp 4-8. Thông qua kỹ thuật tính năng, các tác giả trích xuất các tập hợp tính năng rời rạc từ nhật ký truy vấn của một nền tảng giáo dục Thổ Nhĩ Kỳ được gọi là Vitamin [120]: (i) sự tương đồng văn bản truy vấn-tài liệu, (ii) truy vấn cụ thể, (iii) tài liệu cụ thể, (iv) dựa trên phiên và (v) truy vấn tài liệu dựa trên nhấp chuột. Độc đáo với cách tiếp cận này là trong các nhóm cụ thể của truy vấn và tài liệu cụ thể là các tính năng cụ thể của miền như khóa học, điểm số và loại tài liệu, ví dụ: bài giảng, video hoặc văn bản. Cách tiếp cận này khác với cách tiếp cận của chúng tôi ở chỗ các tính năng được sử dụng trong đào tạo người xếp hạng sử dụng dữ liệu có nguồn gốc từ một SE cụ thể theo miền bao gồm khóa học và thông tin về điểm số của các nguồn lực trong khi chúng tôi thiết kế một trình xếp hạng lại không phụ thuộc vào SE, cho phép xếp hạng lại của chúng tôi được ghép nối với bất kỳ SE chung nào. Ngoài ra, các tính năng được sử dụng bởi Usta et al. [121] bao gồm dữ liệu nhấp chuột có nguồn gốc từ trẻ em, không phải là có sẵn hoặc công khai cho nhóm người dùng của chúng tôi.

Chiến lược có liên quan chặt chẽ nhất đến REDORank là Korsce [87]. Quan điểm đa chiều này

chiến lược kiểm tra tính phù hợp, sự liên kết của chương trình giảng dạy, tính khách quan và khả năng đọc khả năng hiểu được các nguồn tài nguyên để xác định những nguồn phù hợp nhất với học sinh lớp 3 - 5 trẻ em tìm kiếm trong lớp học. Korsce coi các nguồn tài nguyên là không phù hợp nếu họ đề cập đến khiêu dâm và lời nói kích động thù địch, nhưng không đề cập đến những khả năng khác các chủ đề phản cảm như rượu hoặc ma túy. Để phù hợp với chương trình giảng dạy, các tác giả áp dụng phương pháp tiếp cận mô hình hóa chủ đề với Phân bố Dirichlet tiềm ẩn (LDA) như một cách để ước tính mức độ mà một nguồn lực liên quan đến chương trình giảng dạy. Cách tiếp cận này theo một không gian ngữ nghĩa và cấp độ từ ngữ khám phá các nguồn tài nguyên, nhưng không tính đến tính đến thông tin theo ngữ cảnh có thể thu thập được từ việc xem xét tài nguyên văn bản đầy đủ.

Khi xem xét khả năng hiểu đọc, Milton et al. [87] giới thiệu một công thức ước tính điểm dựa trên công thức khả năng đọc Flesch-Kincaid và cosin đường cong phạt các tài nguyên có mức độ dễ đọc vượt quá điểm mong đợi mức độ của người dùng. Có hai khoảng cách lớn trong cách tiếp cận này: (i) việc lựa chọn Công thức Flesch-Kincaid dựa trên việc sử dụng phổ biến hơn là khám phá thực nghiệm, mà chúng tôi tiến hành trong công việc của mình và (ii) điểm hiểu đọc yêu cầu kiến thức về điểm số mong đợi của người dùng, mà chúng ta không thể cho là biết nhóm người dùng của chúng tôi vì chúng tôi đang tập trung vào nhiều cấp độ khác nhau.

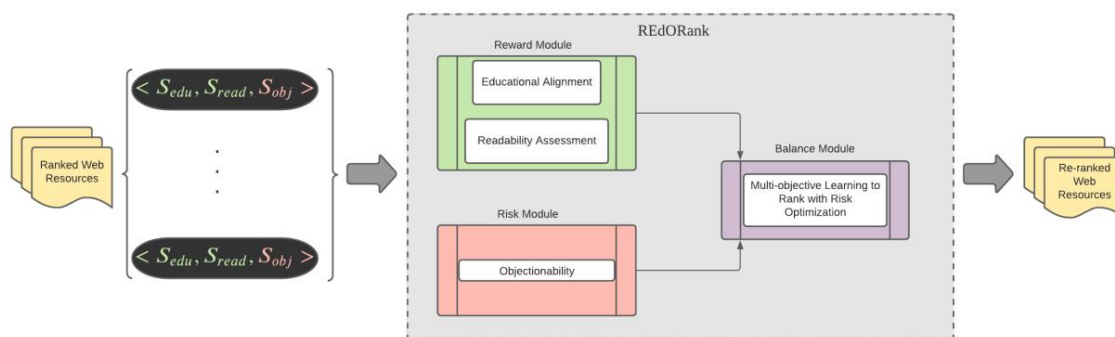
Hơn nữa, Korsce xếp hạng các nguồn lực theo một tập hợp tính các trọng số tối ưu [122]. Các trọng số này được chọn thủ công như là kết quả của một cuộc khám phá thực nghiệm của những người xếp hạng gần tối ưu, trong đó người xếp hạng tối ưu được xác định một cách định tính. người xếp hạng được chọn tạo ra điểm số trên cơ sở từng nguồn lực (tương tự như LTR từng điểm) phương pháp), dẫn đến thứ hạng tương đối dựa trên điểm số được tính toán. Ngoài ra, chúng tôi sử dụng phương pháp tiếp cận theo danh sách, cho phép so sánh mức độ liên quan tuyệt đối giữa các nguồn lực vì tất cả các nguồn lực đều được xem xét cùng một lúc thay vì độc lập.

CHƯƠNG 3

PHƯƠNG PHÁP

Trong chương này, chúng tôi mô tả REORank, một phương pháp học xếp hạng đa góc nhìn (MP-Khung LTR) xếp hạng lại các nguồn tài nguyên thông qua việc kiểm tra song song Khả năng đọc, Sự liên kết về mặt giáo dục và tính phản đối của mỗi nguồn tài nguyên R được thu thập bởi một chính luồng SE để trả lời câu hỏi của trẻ về các khái niệm liên quan đến lớp học.

Tận dụng sức mạnh truy xuất của SE chính thống và thông tin trực tiếp theo các quan điểm đã đề cập ở trên, REORank xác định và ưu tiên các nguồn lực dành cho các lớp học và học sinh K-4. Như thể hiện trong Hình 3.1, REORank bao gồm gồm ba mô-đun: mô-đun phần thưởng, mô-đun rủi ro và mô-đun cân bằng. Mỗi mô-đun phục vụ một mục đích cụ thể trong khuôn khổ chung.



Hình 3.1: Khung REORank. REORank xếp hạng lại các tài nguyên Web được lấy từ một SE chính thống để trả lời truy vấn của trẻ được xây dựng trong bối cảnh lớp học bằng cách cân bằng phần thưởng với rủi ro.

Mô-đun phần thưởng xác định sự tương tác giữa các quan điểm “tích cực”

để phân tích tài nguyên: khả năng đọc và sự phù hợp với chương trình giảng dạy trên lớp. Rủi ro mô-đun xem xét sự tương tác của các quan điểm “tiêu cực” xác định các nguồn lực như không phù hợp với nhóm người dùng. Mô-đun cân bằng đánh đổi các đầu ra của rủi ro mô-đun (một giá trị đóng vai trò là chi phí và do đó làm giảm mức độ ưu tiên tài nguyên) và mô-đun phần thưởng (một giá trị nhằm tăng mức độ ưu tiên tài nguyên trong bảng xếp hạng) đưa đến điểm xếp hạng cuối cùng mà theo đó các nguồn lực được sắp xếp lại.

3.1 Quan điểm: Từ lý thuyết đến thực hành

REdORank xếp hạng lại các tài nguyên Web trực tuyến bằng cách đánh giá lại chúng theo ba quan điểm liên quan đến việc tìm kiếm thông tin của trẻ em trong lớp học. Để đúng cách xếp hạng lại các nguồn lực, chúng ta phải định lượng từng quan điểm. Trong phần còn lại của phần này, chúng tôi mô tả cách chúng tôi thể hiện sự liên kết giáo dục, khả năng đọc và quan điểm phản đối.

3.1.1 Khả năng đọc

Khả năng đọc là một yếu tố quan trọng trong việc hỗ trợ tìm kiếm trên Web của trẻ em dưới dạng văn bản sự linh hoạt và do đó sự hiểu biết ảnh hưởng đến mức độ mà một nguồn tài nguyên là có liên quan đến người dùng [8, 107]. Điều này làm cho việc tính đến văn bản trở nên cấp thiết sự phức tạp của các nguồn tài nguyên được thu thập khi xác định vị trí của chúng trong một SERP. Đây là một vấn đề không tầm thường. Có rất nhiều công thức cho văn bản ước tính độ phức tạp, từ những ước tính truyền thống dựa trên các đặc điểm nông, ví dụ, trung bình từ mỗi câu, đến những câu phức tạp hơn dựa trên học sâu [14, 36, 45, 89]. Mặc dù vậy, vẫn còn thiếu sự đồng thuận về việc nên sử dụng công thức nào cho tự động ước tính độ phức tạp của văn bản trên Web.

Vì hiệu suất tổng thể của RE_DORank bị ảnh hưởng trực tiếp bởi sự lựa chọn công thức khả năng đọc, chúng tôi đánh giá một số công thức trong số đó và so sánh khả năng của chúng đối với đánh giá khả năng đọc hiểu thông qua việc phân công các cấp lớp dựa trên khả năng áp dụng của chúng với nhóm người dùng của chúng tôi, với loại tài nguyên của chúng tôi và tính dễ dàng của việc tính toán. Kết quả là, công thức phù hợp nhất trong phân tích của chúng tôi là Spache-Allen, công thức này cải thiện công thức Spache [113] ban đầu bằng cách mở rộng vốn từ vựng được sử dụng để xác định "dễ dàng từ". Công thức Spache dựa trên vốn từ vựng cho các từ dễ gồm 1.064 từ. Tuy nhiên, vốn từ vựng này chỉ bao gồm các thuật ngữ được coi là dễ hiểu đối với trẻ em hiểu, được thu thập vào những năm 1970 từ các bài báo trên tạp chí và tin tức dành cho người lớn [36]. Thay vào đó, Spache-Allen giải thích thuật ngữ mà trẻ em tiếp xúc trực tuyến bao gồm 47.712 thuật ngữ được trích xuất từ các trang web dành cho trẻ em; danh sách các thuật ngữ ban đầu được biên soạn trong [82]. Hơn nữa, Spache-Allen xem xét thuật ngữ mà trẻ em học thông qua hướng dẫn. Đối với điều này, chúng ta chuyển sang 30.000 từ trong Bộ dữ liệu Tuổi mua lại (AoA) [68]. Cuối cùng, Spache-Allen, được hiển thị chính thức như Phương trình 3.1, dựa vào vốn từ vựng để xác định các từ dễ bao gồm 65.669 thuật ngữ đọc đáo.

$$\text{Spache-Allen}(T) = (0,141 \times w_T / s_T) + (0,086 \times \text{dif}(T)) + 0,839 \quad (3.1)$$

trong đó T là một văn bản, và w_T và s_T là số từ và câu trong T , tương ứng. Hàm $\text{dif}(T)$ xác định tỷ lệ phần trăm các từ khó trong T , nơi một từ được coi là khó nếu nó không xuất hiện trong phiên bản mở rộng của từ vựng dễ dàng.²

¹Chúng tôi cung cấp thông tin chi tiết về quá trình khám phá thực nghiệm và mở rộng vốn từ vựng liên quan đến công thức Spache ban đầu [113] trong Phần 4.1.

²<https://github.com/Neelikh/spache-allen-vocabulary>

REdORank xác định điểm khả năng đọc Sread của R (Phương trình 3.2) dựa trên ước tính khả năng đọc được xác định bằng cách sử dụng Spache-Allen, khi áp dụng cho đoạn mã tương ứng RS.

$$\text{Sread}(R) = \text{Spache-Allen}(RS) \quad (3.2)$$

3.1.2 Tài nguyên giáo dục

Việc ưu tiên tài nguyên theo khả năng đọc có thể giúp hỗ trợ tìm kiếm trên Web của trẻ em. Tuy nhiên, riêng nó, quan điểm này có thể bỏ qua môi trường là trọng tâm của chúng ta nghiên cứu, nghĩa là không phải tất cả các nguồn tài nguyên phù hợp với khả năng đọc của trẻ em đều phù hợp cho lớp học. Vì vậy, trong thiết kế của REdORank, chúng tôi cũng xem xét giáo dục giá trị của tài nguyên, được chúng tôi xác định thông qua quy trình phân loại Web.

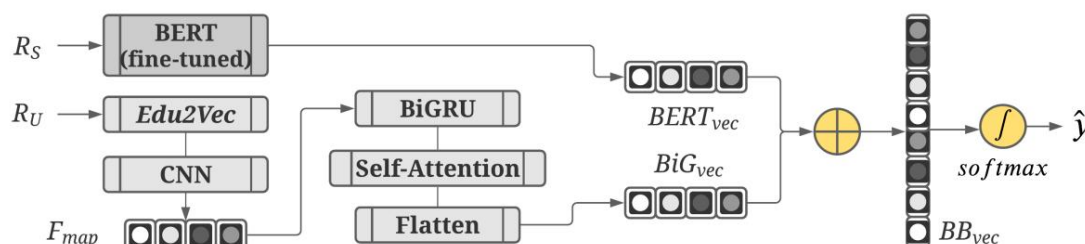
Phân loại tài nguyên web là một lĩnh vực được khám phá kỹ lưỡng trong Truy xuất thông tin [56]. Gần đây, lĩnh vực này đã chứng kiến sự gia tăng mạnh mẽ của các nghiên cứu liên quan đến phân loại theo miền cụ thể. sự phát triển, đặc biệt là trong các lĩnh vực pháp lý, tài chính và y tế [40, 60, 132]. Tuy nhiên, phân loại trong lĩnh vực giáo dục vẫn chưa được khám phá nhiều. Theo thuật ngữ rộng, giáo dục áp dụng cho nhiều nhiệm vụ phân loại khác nhau. Công việc trước đây bao gồm việc phân loại các nguồn lực giáo dục dựa trên “sức mạnh của giáo dục tài nguyên [như] một thuộc tính được đánh giá tích lũy bởi đối tượng mục tiêu của tài nguyên (ví dụ, sinh viên hoặc chuyên gia giáo dục)” bằng cách sử dụng Máy vectơ hỗ trợ (SVM) [57]. Tuy nhiên, mô hình này phụ thuộc rất nhiều vào dữ liệu được chú thích thủ công và chỉ có thể áp dụng đến giáo dục khoa học máy tính. Xia [126] cũng sử dụng SVM để phân loại tài nguyên được hỗ trợ hướng dẫn chuyển hướng, trong khi EduBERT [24] phát hiện các bài đăng trên diễn đàn cấp đại học được viết bởi những học sinh gặp khó khăn. Nhìn chung, những nỗ lực trong lĩnh vực này phân loại các nguồn lực cho những mục đích không xác định nhóm tuổi, học viên người lớn, lĩnh vực môn học hạn chế, giảng viên hoặc cấp độ tổ chức

hiểu biết sâu sắc. Có một khoảng cách trong tài liệu liên quan đến việc nhận ra Web giáo dục tài nguyên dành cho trẻ em từ 6 đến 11 tuổi ở lớp K-4.

Bất kể phạm vi nào, các bộ phân loại có xu hướng dựa vào các tính năng suy ra từ HTML nội dung trang [38, 114]. Xử lý toàn bộ các trang Web đòi hỏi sức mạnh tính toán cao, lưu trữ dữ liệu lớn và thời gian để truy xuất [104] vì các trang Web thường là động và chứa hình ảnh, video hoặc tập lệnh ngoài văn bản [105]. Để giải quyết một số trong những hạn chế, các phương pháp tiếp cận hiện đại chỉ kiểm tra URL [49, 105]. Thật không may, URL không phải lúc nào cũng bao gồm các mã thông báo có ý nghĩa (tức là các thuật ngữ hợp lệ), có thể gây ra phân loại sai. Hãy xem xét URL <https://www.youtube.com/watch?v=pX3V9hoX1eM> cho một video trên YouTube của National Geographic For Kids liên quan đến động vật. Trong trường hợp này, các mã thông báo có ý nghĩa bao gồm "youtube" và "xem", không có điều này cho biết mức độ thân thiện với trẻ em của nguồn tài nguyên tương ứng.

Nhận thức được những hạn chế đã đề cập ở trên, chúng tôi dựa vào kiến thức chuyên môn thu được từ các tiêu chuẩn giáo dục cùng với URL và văn bản mô tả để thông báo công nhận các nguồn tài nguyên giáo dục trên Web của trẻ em. Các tiêu chuẩn giáo dục, chẳng hạn như như CCSS của Hoa Kỳ và Tiêu chuẩn Khoa học Thế hệ Tiếp theo (NGCS), cung cấp kết quả học tập cho học sinh lớp K-4. Đặc biệt, chúng tôi tập trung vào các nguồn tài nguyên giáo dục cung cấp thông tin về các môn học cho lớp K-4, chẳng hạn như ngôn ngữ nghệ thuật, khoa học và nghiên cứu xã hội, được mô tả trong CCSS, NGCS và Nội dung Idaho Tiêu chuẩn (ICS).

Như minh họa trong Hình 3.2, BiGBERT, Đơn vị hồi quy có cổng song hướng (BiGRU) với mô hình BERT chúng tôi giới thiệu để nhận dạng các nguồn tài nguyên Web giáo dục đối với trẻ em, có hai thành phần chính: URL và vectorizer đoạn trích. Cho R, BiGBERT đầu tiên vector hóa URL của nó, kết hợp các nhúng cụ thể theo miền từ Edu2Vec [11] với BiGRU và lớp tự chú ý. Shen et al. [112] hiển thị



Hình 3.2: Kiến trúc BiGBERT (RU và RS biểu thị URL và đoạn trích của một tài nguyên R nhất định).

rằng việc sử dụng tóm tắt thay vì nội dung toàn trang sẽ dẫn đến phân loại tương đương

hiệu suất, do đó chúng tôi sử dụng đoạn trích thay cho nội dung đầy đủ. Để vector hóa đoạn trích của R,

BiGBERT tinh chỉnh mô hình máy biến áp BERT [33] bằng cách sử dụng các tiêu chuẩn giáo dục.

Cuối cùng, BiGBERT nối các vector đoạn mã và URL và áp dụng softmax

chức năng xác định lớp R.

URL Vectorizer. BiGBERT mã hóa URL của R (R_U) thành một chuỗi các thuật ngữ T bằng cách chia thành các ký hiệu không phải chữ và số (ví dụ: dấu chấm, dấu gạch ngang và dấu gạch chéo về phía trước) và sử dụng SymSpell [47] để thực hiện phân đoạn từ vì URL có xu hướng hợp thành các từ với nhau (ví dụ, thay đổi `stackoverflow` thành `stack overflow`). Mỗi mã thông báo $t_i \in T$ được ánh xạ tới những từ tương ứng của nó. Nếu t_i không phải là một phần của những từ điển, chúng tôi quy điều này cho một lỗi chính tả hoặc biến thể chính tả có thể xảy ra, và do đó thử sửa lỗi bằng cách sử dụng một thao tác chỉnh sửa khoảng cách duy nhất (tức là thay thế, thêm hoặc xóa một ký tự). Nếu t_i vẫn không có trong từ điển, chúng tôi loại bỏ nó để đảm bảo chỉ vẫn còn những dấu hiệu có ý nghĩa.

Để tìm hiểu cách biểu diễn của R_U , BiGBERT sử dụng những từ Edu2Vec từ điển [11] vì nó kết hợp kiến thức chuyên ngành từ NGCS, CCSS và ICS. Các tiêu chuẩn này đóng vai trò là nguồn kiến thức có cấu trúc để xác định các thuật ngữ, chủ đề và các môn học dành cho lớp K-4, cho phép BiGBERT nhấn mạnh các khái niệm chương trình giảng dạy K-4 trong

RU có thể bị bỏ qua bởi các nhúng được đào tạo trước mục đích chung. Thay vì phân tích các nhúng độc lập, chúng tôi thiết kế BiGBERT để kiểm tra ngữ cảnh nhạy cảm các chỉ số từ T. Lấy cảm hứng từ Rajalakshmi và Aravindan [103] và để đáp lại URL không tuân theo cú pháp ngôn ngữ truyền thống, chúng tôi kiểm tra các nhóm nhúng (tức là, trigram) sử dụng Mạng nơ-ron tích chập (CNN) – một phương pháp nhanh, hiệu quả và phương pháp nhỏ gọn [65] để tạo ra các vectơ đặc trưng từ các trigram. Tích chập kết quả trong bản đồ đặc điểm $F_{map} = \langle F_1, F_2, \dots, F_x \rangle$, $f = 1..x$ $F_f = \text{relu}(w.x_{i:i+m-1} + b_u)$, trong đó hàm tuyến tính chỉnh lưu relu được áp dụng cho tích vô hướng của hạt nhân w với cửa sổ nhúng $x_{i:i+m-1}$ trong T có kích thước $m=3$; b_u là một thuật ngữ thiên vị. Để khám phá dài khoảng cách phụ thuộc của các tính năng có thể xuất hiện cách xa nhau BiGBERT sử dụng BiGRU mạng, vì nó nắm bắt thông tin ngữ cảnh theo hướng tiến và lùi. A lớp tự chú ý sau đó xác định tầm quan trọng của các tính năng được CNN xác định và BiGRU. Tiếp theo là một lớp phẳng và dày đặc tạo ra một đặc điểm duy nhất Biểu diễn vectơ của RU có kích thước 128, được ký hiệu là BiGvec.

Mã hóa đoạn trích. Vì đoạn trích dài vài câu, không giống như URL nhiều nhất là một vài từ, chúng tôi yêu cầu một mô hình có thể xem xét kỹ lưỡng từng đoạn trích (RS) như một toàn bộ. Do đó, chúng tôi kết hợp mô hình máy biến áp hiện đại BERT [33] vào Thiết kế của BiGBERT. Khả năng xử lý chuỗi của BERT lên đến kích thước tối đa là 512 mã thông báo cho phép BiGBERT khai thác thông tin theo ngữ cảnh, tuần tự trong RS trong toàn bộ. Ngoài ra, kiến trúc của BERT bao gồm 12 khối biến áp và sự tự chú ý của người đứng đầu đảm bảo việc học thông tin ngữ cảnh phong phú từ mỗi đoạn trích. Như vậy, chúng tôi mã hóa RS thành một chuỗi các câu, mã hóa nó thành BERT thông số kỹ thuật và sử dụng BERT để đạt được biểu diễn vectơ đặc điểm tổng hợp của kích thước 768, được ký hiệu là BERTvec.

Đối với các nhiệm vụ phụ thuộc vào miền như nhiệm vụ chúng tôi giải quyết ở đây, BERT được hưởng lợi từ việc tinh chỉnh

điều chỉnh [115]. Do đó, chúng tôi điều chỉnh BERT truyền thống theo định nghĩa giáo dục của chúng tôi bằng khai thác các tiêu chuẩn giáo dục đã được thiết lập. Chúng tôi thực hiện tinh chỉnh như mô tả trong [115], đào tạo 3 nhúng BERT như một trình phân loại văn bản giáo dục bằng cách thêm một tuyến tính lớp phân loại sử dụng entropy chéo nhị phân làm mất mát và trình tối ưu hóa Adam với tốc độ học tập = $1e^{-5}$.

Dự đoán sự liên kết giáo dục. Để tận dụng bằng chứng về sự liên kết giáo dục suy ra từ RU và RS, chúng tôi nối BiGvec với BERTvec thành BBvec.

Sử dụng một lớp được kết nối đầy đủ trên BBvec với hàm kích hoạt softmax, BiGBERT₄ tạo ra một phân phối xác suất \hat{y} trên mỗi lớp, có giáo dục và không, sao cho $\hat{y} \in [0, 1]$.

Hàm softmax đảm bảo rằng tổng xác suất của mỗi lớp là một.

Sử dụng BiGBERT, chúng tôi xác định điểm số liên kết giáo dục $Sedu$ cho R (Phương trình 3.3).

Là một trong hai quan điểm “phần thưởng”, REORank sử dụng điểm số này để xác định của mức độ liên quan của R.

$$Sedu(R) = BiGBERT(RS, RU) \quad (3.3)$$

3.1.3 Tài nguyên phản cảm

Web chứa đựng một bộ sưu tập tài nguyên ngày càng phong phú dành cho người dùng ở nhiều lứa tuổi, ví dụ: kinh nghiệm và trình độ hiểu biết. Do đó, người ta dự đoán một số nguồn tài nguyên này để hòa hợp hơn với một số nhóm người dùng hơn những nhóm khác. Với nhóm người dùng và môi trường là trọng tâm của công việc này, nó trở nên cấp thiết đối với REORank

³Để tinh chỉnh, chúng tôi sử dụng 2.655 đoạn văn bản từ NGCS, CCSS và ICS cùng với 2.725 đoạn từ ngữ liệu Brown [17, 44].

⁴BiGBERT được đào tạo bằng cách sử dụng kích thước lô là 128, hàm mất entropy chéo nhị phân và RMSProp bộ tối ưu hóa [119] với động lượng = 0,2 và tốc độ học = 0,001.

để giảm thiểu rủi ro khi đưa các nguồn lực lên đầu SERP có thể được coi là không phù hợp. Đây là lý do tại sao REdORank kết hợp một sự phản đối quan điểm về thiết kế của nó.

Xác định những gì khiến một nguồn tài nguyên Web trở nên phản cảm đối với trẻ em trong lớp học thiết lập không phải là tầm thường. SE chuyên biệt được xây dựng trên nền tảng GCS (ví dụ: Kiddle và KidRex) sử dụng tính năng tìm kiếm an toàn để xóa bỏ các tài nguyên không mong muốn, nhưng vẫn hiển thị các nguồn tài nguyên được quảng cáo ở đầu SERP dưới dạng quảng cáo. Những quảng cáo có thể chuyển hướng trẻ em đến SERP khác mà không có sự bảo vệ tìm kiếm an toàn, vô tình để lộ chúng với tài liệu không phù hợp [42]. Chức năng tìm kiếm an toàn không phải là không có lỗi vì các nguồn tài nguyên không phù hợp có thể vượt qua bộ lọc của nó [37, 12]. Ví dụ, để trả lời truy vấn “sự thật về chó”, Google đã bật Tìm kiếm an toàn lấy lại một trang web nhà máy bia5 . Đồng thời, có những trang web có thể xuất hiện đáng phản đối nhưng không phải là, ví dụ, một bài viết về ung thư vú [42]. Ngăn chặn việc hiển thị của những kết quả như vậy trong khi cũng tránh lọc quá mức là một vấn đề khó khăn đòi hỏi giải pháp vượt xa tìm kiếm an toàn.

Patel và Singh [98] tăng cường khả năng lọc GCS bằng cách cũng xem xét các nguồn lực chứa lời nói thù địch hoặc bạo lực. Lee et al. [73] đi xa hơn nữa và lọc nội dung đề cập đến phá thai, rượu, thuốc lá, quan hệ bất hợp pháp, ma túy, cờ bạc, cần sa, khiêu dâm, bạo lực, phân biệt chủng tộc và vũ khí. Thật không may, bộ lọc được đề xuất của họ- chiến lược tìm kiếm dựa trên dữ liệu nhấp chuột, điều này hiếm khi tồn tại đối với nhóm người dùng của chúng tôi. Milton et al. [87] giới thiệu một chiến lược không phụ thuộc vào cú nhấp chuột để xác định các tài nguyên không phù hợp với lớp học. Chiến lược này tận dụng các từ điển để giải thích cho sự hiện diện của các thuật ngữ khiêu dâm và ngôn từ kích động thù địch, ngoài các thuật ngữ viết sai chính tả, trong nội dung, thẻ meta và thẻ neo của các nguồn tài nguyên. Tuy nhiên, chiến lược này

coi là tài nguyên đáng phản đối liên quan đến khiêu dâm và ngôn từ kích động thù địch, chỉ có hai của một số chủ đề có thể được coi là không phù hợp với nhóm người dùng và môi trường đang được nghiên cứu. Ngoài ra, toàn bộ nội dung của một nguồn tài nguyên được xem xét, có thể tốn nhiều thời gian để xử lý khi triển khai trong một kịch bản thực tế.

Quan sát những điểm mạnh và hạn chế của các chiến lược lọc hiện có, chúng tôi áp dụng một kỹ thuật đơn giản nhưng hiệu quả để xác định các nguồn tài nguyên bị coi là không phù hợp: Judgebad. Với R, trước tiên chúng ta tạo một biểu diễn nắm bắt thuật ngữ của nó từ nhiều loại khác nhau. Xử lý việc xác định R là đáng phản đối như một nhị phân nhiệm vụ phân loại, sau đó chúng tôi sử dụng mô hình Rừng ngẫu nhiên, đã được chứng minh để duy trì hiệu quả ngay cả khi so sánh với các giải pháp thần kinh gần đây [28, 29].

Các loại phản đối. Để giải thích cho sự đa dạng lớn của các loại phản đối tài liệu hiện có trực tuyến và được lấy cảm hứng từ các chiến lược trước đây để phát hiện ra những phản hồi đáng ngờ nguồn [87, 73], chúng tôi coi là phản cảm đối với trẻ em trong các nguồn tài nguyên lớp học liên quan đến bất kỳ danh mục nào trong ObjCat: Phá thai, Ma túy, Lời nói căm thù, Phi pháp hội chợ, cờ bạc, khiêu dâm và bạo lực. Lưu ý rằng danh mục Ma túy đề cập đến đến các nguồn tài nguyên bao trùm ma túy, mà còn cả rượu, thuốc lá và cần sa. Hơn nữa, Bạo lực tập trung vào nội dung bạo lực, cũng như vũ khí; Lời nói căm thù chiếm phân biệt chủng tộc và nội dung thù hận/phản cảm.

Như đã nêu trước đó, để xác định khả năng các nguồn lực bị phản đối- có thể, chúng tôi áp dụng một kỹ thuật xem xét kỹ lưỡng thuật ngữ của họ và do đó yêu cầu sự tồn tại của các danh sách được xác định trước về các thuật ngữ 'đáng phản đối'. Trong trường hợp của Nội dung khiêu dâm và các danh mục Ngôn từ kích động thù địch, chúng tôi sử dụng các danh sách được xác định trước được sử dụng trong [87], đó là có nguồn từ kho lưu trữ của Google⁶ và trang web của Hate Speech Movement⁷, tôn trọng-

⁶<https://code.google.com/archive/p/badwordslist>
⁷[HateSpeechMovement.org](https://hatespeechmovement.org)

một cách tích cực. Thật không may, không có danh sách thuật ngữ được quản lý nào liên quan đến phần còn lại danh mục trong ObjCat. Do đó, chúng tôi tạo ra chúng thông qua một quy trình mới gọi là danh mục hiểu biết thông qua việc thay thế tên nhãn [86].

Chúng tôi sử dụng các trang web từ Alexa Top Sites [6] được biết là thuộc về các danh mục xuất hiện trong ObjCat như là ngữ liệu của chúng tôi để tạo ra các danh sách thuật ngữ. Đối với mỗi danh mục, không bao gồm Nội dung khiêu dâm và lời nói căm thù, sự xuất hiện của tên danh mục (cũng như các danh mục phụ) tên danh mục, nếu có) trong một trang web từ ngữ liệu được che dấu và một bản trước Bộ mã hóa BERT được đào tạo được sử dụng để tạo ra biểu diễn vectơ theo ngữ cảnh h với tên danh mục được che dấu. Mô hình ngôn ngữ được che dấu của BERT (MLM) tạo ra một phân phối xác suất rằng một thuật ngữ w từ trong vốn từ vựng của BERT sẽ xảy ra tại vị trí của tên danh mục bị che.

Các thuật ngữ có thể xuất hiện trong các bối cảnh khác nhau trong cùng một ngữ liệu. Do đó, các thuật ngữ trong từ vựng được trích xuất được xếp hạng theo xác suất xuất hiện của chúng (Phương trình 3.4), và mỗi thuật ngữ có thể thay thế tên danh mục trong ngữ liệu bao nhiêu lần duy trì bối cảnh.

$$p(w | h) = \text{Softmax}(W_2 \sigma(W_1 h + b)) \quad (3.4)$$

trong đó $\sigma(\cdot)$ là hàm kích hoạt; W_1 , W_2 và b là các tham số đã học cho nhiệm vụ dự đoán ngôn ngữ ẩn, được đào tạo trước trong BERT.

Như trong [86], chúng tôi chọn 100 thuật ngữ hàng đầu cho mỗi danh mục (hoặc toàn bộ danh sách nếu ít hơn 100 được trích xuất) như danh sách thuật ngữ đại diện cuối cùng nằm bất theo ngữ cảnh các thuật ngữ tương tự và đồng nghĩa liên quan đến các danh mục tương ứng.

Biểu diễn đoạn trích. Do sự phức tạp của việc thu thập, máy tính nhu cầu về tài nguyên và lưu trữ để xử lý toàn bộ nội dung của các trang Web (Phần

3.1.2), chúng tôi sử dụng đoạn trích làm proxy cho toàn bộ nội dung trang.

Chúng tôi biểu diễn R bằng một bộ sưu tập gồm 16 tính năng dựa trên văn bản được trích xuất từ các đoạn trích của nó thú cưng. Bây giờ trong số các đặc điểm này giải thích cho sự phổ biến (tức là tần suất xuất hiện) của các thuật ngữ phản đối trong RS . Đối với mỗi danh mục oc trong $ObjCat$, chúng tôi tính toán thuật ngữ tỷ lệ lưu hành, tức là $TP(RS, oc)$, như trong Phương trình 3.5.

$$TP(RS, oc) = \frac{\sum_{t \in T \text{ vị trí}} tf(t, RS)}{|RS|} \quad (3.5)$$

trong đó $T \text{ Loc}$ là danh sách thuật ngữ cho oc , t là một thuật ngữ trong $T \text{ Loc}$ và $tf(\cdot)$ là một hàm tính toán số lần t xuất hiện trong RS . Đóng vai trò là hệ số chuẩn hóa, $|RS|$ là độ dài của RS sau khi phân tách, chấm câu và xóa từ dừng, và phân loại từ vựng (sử dụng thư viện Python NLTK8).

Chúng tôi cũng xem xét phạm vi bao phủ của thuật ngữ phản cảm trong RS , sử dụng bảy các tính năng giải thích cho các tình huống mà một thuật ngữ có thể bị hiểu sai là đối tượng có thể tùy thuộc vào ngữ cảnh. Ví dụ, “ngực” có thể xuất hiện thường xuyên trong nguồn tài nguyên sinh học tự nó không đáng bị phản đối; nó cũng có thể xuất hiện trong một nội dung khiêu dâm tài nguyên. Đối với mỗi loại oc , chúng tôi tính toán phạm vi bảo hiểm thuật ngữ phản đối trong RS , tức là, $T \text{ Cov}(RS, oc)$, sử dụng Phương trình 3.6.

$$T \text{ Cov}(RS, oc) = \frac{\sum_{t \in T \text{ vị trí}} \delta(t, RS)}{|T \text{ Loc}|} \quad (3.6)$$

trong đó $T \text{ Loc}$ và t được định nghĩa trong Phương trình 3.5, $\delta(t, RS)$ là 1 nếu t xảy ra ít nhất một lần trong RS và 0 nếu không, và $|T \text{ Loc}|$, là tổng số các điều khoản trong $T \text{ Loc}$, hoạt động như một hệ số chuẩn hóa.

Chúng tôi giải thích rõ ràng các thuật ngữ viết sai chính tả, với tư cách là nhà sản xuất các nội dung trực tuyến phản cảm

8<https://www.nltk.org/>

nội dung được biết là đưa ra lỗi chính tả có ý như một nỗ lực để vượt qua an toàn

bộ lọc tìm kiếm [87]. Chúng tôi xem xét mức độ phổ biến của các thuật ngữ viết sai chính tả trong Rs-tần suất

lỗi chính tả xảy ra trong Rs-sử dụng Phương trình 3.7.

$$MP(Rs) = \frac{t_{RS} \beta(t, R)}{|RS|} \quad (3.7)$$

trong đó t là một thuật ngữ trong RS , $\beta(t, RS)$ là 1 nếu t là lỗi chính tả và 0 nếu không, và $|RS|$

là một hệ số chuẩn hóa biểu diễn độ dài của RS . Chúng tôi sử dụng thư viện Enchant⁹

để xác định các thuật ngữ viết sai chính tả khi nó bao gồm nhiều thư viện kiểm tra chính tả hiện có, chẳng hạn như

Ispell, Aspell và MySpell.

Cuối cùng, chúng ta xem xét phạm vi lỗi chính tả bằng Công thức 3.8.

$$MC(R) = \frac{t_{RSu} \gamma(t, T_{Lall})}{t_{RSu} \beta(t, Rs)} \quad (3.8)$$

trong đó $\beta(.)$ được định nghĩa như trong Phương trình 3.7, t là một thuật ngữ trong RSu , là tập hợp các

các điều khoản trong RS , T_{Lall} là tập hợp các điều khoản thu được từ việc hợp nhất danh sách các điều khoản cho mỗi

thể loại trong $ObjCat$ và $\gamma(.)$ được đánh giá là 1 nếu t được xác định là lỗi chính tả và nó

xảy ra ở T_{Lall} và 0 ở những trường hợp khác.

Phát hiện khả năng phản đối. Dựa trên hiệu quả của nó trong phân loại tương tự

nhiệm vụ [87], chúng tôi sử dụng mô hình Rừng ngẫu nhiên để xác định các tài nguyên đáng ngờ. Sử dụng

biểu diễn tính năng của R làm đầu vào, một mô hình Rừng ngẫu nhiên được đào tạo¹⁰ tạo ra như

đưa ra một phân phối xác suất nhị phân \hat{y} trên mỗi lớp-có thể phản đối và không-chẳng hạn

rằng $\hat{y} \in [0, 1]$ đối với R . Để phục vụ như là điểm nhạy cảm được khai thác bởi mô-đun rủi ro,

chúng tôi định nghĩa S_{bad} là giá trị xác suất của R được liên kết với điều đáng phản đối

lớp (Phương trình 3.9).

⁹<https://abiword.github.io/enchant/> ¹⁰Số lượng nút

lá tối đa, số lượng mẫu lá tối thiểu và số lượng mẫu chia tối thiểu được đặt thành 32. Độ sâu tối đa được đặt thành 8.

Sbad(R) = Thảm phánbad(RS)

(3.9)

3.2 REdORank: Từ lý thuyết đến thực hành

REdORank được hỗ trợ bởi thuật toán LTR, AdaRank [127], vì nó là một trong những các thuật toán phổ biến hơn trong nghiên cứu LTR [48, 69, 78, 84]. AdaRank sử dụng danh sách cách tiếp cận (được định nghĩa trong Phần 2.1), là cách hiệu quả nhất về mặt xếp hạng độ chính xác khi sử dụng cho tìm kiếm trên Web [20, 118].

3.2.1 Xếp hạng Ada

AdaRank là một thuật toán tăng cường trong đó một tập hợp các thứ hạng được xác định yếu được kết hợp tuyến tính để tạo ra một bảng xếp hạng tổng thể chính xác hơn bất kỳ những người xếp hạng yếu cá nhân. Một người xếp hạng yếu được định nghĩa là $h_t = P_t(i)E(\pi(q_i, d_i, x_k))$, trong đó q_i là một tập hợp các truy vấn, d_i là danh sách được xếp hạng của các tài liệu trên mỗi truy vấn, x_k là sự thật cơ bản cho tài liệu d_i x_k là các biểu diễn đặc trưng cho mỗi x_1, \dots ,

1 tài liệu, E là một biện pháp đánh giá và $P_t(i) = \frac{1}{|H_t(i)|}$ là trọng số ban đầu.

Với một tập dữ liệu đào tạo $\{q_i, d_i, y_i\}$, AdaRank áp dụng phương pháp lặp lại như sau rằng tại mỗi lần lặp t , một tập hợp các ranker yếu được khởi tạo với giá trị hiện tại trọng số, một hoán vị xếp hạng $(\pi(\cdot))$ được dự đoán và đánh giá, và trọng số là đã cập nhật. Mã giả cho quy trình này được trình bày trong Thuật toán 1.

Giống như tất cả các thuật toán LTR, AdaRank học một hàm xếp hạng thông qua tối ưu hóa sự tối ưu hóa của một biện pháp đánh giá. Số liệu được sử dụng phổ biến nhất để tối ưu hóa là Lợi nhuận tích lũy được chiết khấu chuẩn hóa (NDCG) [64, 79]. Mục tiêu của NDCG là để đo lường sự thống nhất giữa danh sách được xếp hạng dự đoán và sự thật cơ bản

Thuật toán 1 Thuật toán AdaRank, được sao chép từ [127].

Đầu vào: $S = \{(q_i, d_i, y_i)\}_{i=1}^m$ là tập dữ liệu đào tạo, $t_{\text{tối}} = 1$, và các tham số E và T .

$P_1(i) = \text{Đối với } t = 1, \dots, T, \frac{1}{t_{\text{tối}}}.$

- Tạo ranker yếu h_t với phân phối có trọng số P_t trên dữ liệu đào tạo S .

- Chọn a_t

tại
$$= \frac{1}{\text{trong } 2} \frac{\sum_{i=1}^{t_{\text{tối}}} P_t(i) \{1 + E(\pi(q_i, d_i, h_t), y_i)\} d_i}{\sum_{i=1}^{t_{\text{tối}}} P_t(i) \{1 - E(\pi(q_i, d_i, h_t), y_i)\} d_i}$$

- Đóng f_t

$$f_t(x) = \sum_{k=1}^t a_k h_k(x)$$

- Cập nhật P_{t+1}

$$\text{Điểm } +1 = \frac{\exp\{ - \sum_{i=1}^{t_{\text{tối}}} E(\pi(q_i, d_i, f_t), y_i)\}}{\sum_{j=1}^{t_{\text{tối}}} \text{biểu thức}\{ - \sum_{j=1}^{t_{\text{tối}}} E(\pi(q_j, d_j, f_t), y_j)\}}$$

Kết thúc cho

Mô hình xếp hạng đầu ra $f(x) = f_T(x)$.

cho một truy vấn q . “Lợi ích” trong NDCG là lợi ích liên quan, hoặc lợi ích của việc hiển thị các nguồn tài nguyên có liên quan cao hơn trong bảng xếp hạng. Mức tăng liên quan của mỗi nguồn tài nguyên được truy xuất để trả lời một truy vấn được xác định dựa trên vị trí của nó trong danh sách được xếp hạng, như trong Phương trình 3.10.

$$g_i = 2 \text{rel}_i - 1 \tag{3.10}$$

trong đó g_i là mức tăng liên quan của i tài nguyên trong danh sách được xếp hạng và rel_i là sự thật cơ bản có liên quan tương ứng.

Phần “Giảm giá” của NDCG là hình phạt áp dụng cho các nguồn lực có liên quan, nhưng chúng xuất hiện thấp hơn trong danh sách được xếp hạng. Cần phải đảm bảo rằng điều này giảm giá không quá cao, để tính đến những người dùng kiên trì có nhiều khả năng khám phá

sâu hơn vào danh sách được xếp hạng [64]. Do đó, NDCG sử dụng chiết khấu logarit dựa trên về vị trí của một nguồn lực trong danh sách được xếp hạng, như thể hiện trong Phương trình 3.11.

$$d_i = \log(\text{rank}_i + 1) \quad (3.11)$$

trong đó d_i là chiết khấu cho i^{th} tài nguyên trong danh sách được xếp hạng và rank_i là vị trí thứ của i nguồn lực trong bảng xếp hạng nói trên.

Khi xem xét danh sách được xếp hạng để trả lời truy vấn q , khía cạnh “Tích lũy” có hiệu lực khi tích lũy các khoản lợi nhuận được chiết khấu, bắt đầu từ đỉnh danh sách được xếp hạng cho đến một vị trí cụ thể. Về mặt chính thức, điều này được gọi là Giảm giá Lợi nhuận tích lũy (DCG) được định nghĩa như trong Phương trình 3.12.

$$\text{DCG}@k(q) = \sum_{i=1}^k \frac{g_i}{d_i} \quad (3.12)$$

trong đó k là giá trị ngưỡng, tức là số lượng tài nguyên được kiểm tra trong danh sách, i là vị trí trong bảng xếp hạng, và g_i và d_i được định nghĩa tương ứng trong Phương trình 3.10 và 3.11.

Tính đến nhu cầu đo lường sự đồng thuận trên toàn bộ danh sách khi số lượng tài liệu có liên quan có thể thay đổi, DCG phải được “Chuẩn hóa”, dẫn đến NDCG, được tính như trong Phương trình 3.13.

$$\text{NDCG}@k(q) = \frac{\sum_{i=1}^k \text{DCG}@k(q)}{\sum_{i=1}^k \text{IDCG}@k(q)} \quad (3.13)$$

trong đó k , i và $\text{DCG}@k(q)$ được định nghĩa trong Phương trình 3.12 và $\text{IDCG}@k(q)$ là DCG được tính toán cho danh sách được xếp hạng hoàn hảo, tức là “lý tưởng”, cho đến vị trí k .

Lợi ích của NDCG là khả năng tính đến các mức độ liên quan khác nhau, do theo cách thức xác định mức độ liên quan. Không giống như các phương pháp đối phó truyền thống hơn

các phần để xếp hạng kiểm tra một giá trị liên quan duy nhất, REdORank xem xét nhiều

tín hiệu cho thấy sự liên quan của một nguồn tài nguyên, cụ thể là sự liên kết về mặt giáo dục và khả năng đọc.

Ngoài ra, NDCG không tính đến các tín hiệu rõ ràng không liên quan, chẳng hạn như

sự phản đối của một nguồn tài nguyên. Do đó, chúng tôi tìm cách mở rộng NDCG để tính đến

để có thêm các tín hiệu liên quan và không liên quan.

3.2.2 Tối ưu hóa đa góc nhìn với độ nhạy chi phí

Mục tiêu của hệ thống tìm kiếm là lấy các tài nguyên từ một bộ sưu tập có

có liên quan cao nhất đối với truy vấn của người dùng. Trong một số trường hợp, các bộ sưu tập này

chứa các tài nguyên không phải để tất cả người dùng nhìn thấy, chẳng hạn như y tế tư nhân

tài liệu hoặc, trong trường hợp của một hệ thống chính phủ, các công văn tuyệt mật. Những loại này

của các nguồn tài nguyên được gọi là các nguồn tài nguyên nhạy cảm. Như một cách để tránh trình bày các nguồn tài nguyên nhạy cảm

vật liệu để đáp ứng các yêu cầu trực tuyến, Sayed và Oard [110] đã giới thiệu một bản mở rộng

phiên bản của số liệu DCG, được gọi là Lợi nhuận tích lũy chiết khấu nhạy cảm với chi phí (CS-

DCG). Chỉ số mới này (Phương trình 3.14) đưa ra mức phạt chi phí hoặc yếu tố rủi ro,

để hiển thị một tài liệu nhạy cảm trong bảng xếp hạng các tài nguyên đã truy xuất.

$$CS \quad DCG_k = \sum_{i=1}^k \frac{g_i}{d_i} \quad c_i \quad (3.14)$$

trong đó k , g_i và d_i được định nghĩa như trong Phương trình 3.12, và c_i là chi phí độ nhạy của

Hiển thị tài liệu nhạy cảm ở vị trí xếp hạng i .

Việc kết hợp CS-DCG vào mô hình LTR như AdaRank giúp tăng cường sức mạnh cho mô hình

để học cách xếp hạng các tài liệu nhạy cảm thấp hơn các tài liệu không nhạy cảm. Điều này

phù hợp với những gì chúng tôi muốn thực hiện với quan điểm phản đối của REdORank:

xóa bỏ khỏi các vị trí xếp hạng cao nhất những nguồn lực có thể được coi là nhạy cảm

cho nhóm người dùng và môi trường là trọng tâm công việc của chúng tôi. Vì vậy, thay vì
 tùy thuộc vào NDCG truyền thống khi đào tạo trình xếp hạng lại LTR của nó, RE₀Rank
 sử dụng CS-DCG cho mục đích tối ưu hóa. Trong trường hợp này, chúng tôi sử dụng chi phí độ nhạy
 ci Sbad (Phương trình 3.9).

CS-DCG tính đến các nguồn tài nguyên không đáng tin cậy, nhưng vẫn chỉ xem xét một chữ ký duy nhất
 na1 để đạt được sự liên quan. Tuy nhiên, trong bối cảnh công việc của chúng tôi, điều bắt buộc là phải tận dụng
 ảnh hưởng của cả sự liên kết giáo dục và khả năng đọc hiểu trong việc xác định
 sự liên quan của một nguồn tài nguyên nhất định. Không đủ để chỉ kết hợp tuyến tính
 điểm số trình độ lớp và trình độ giáo dục tương ứng, Sedu và Sread, được tính toán
 trong Phần 3.1.2 và 3.1.1, tương ứng. Thay vào đó, điều quan trọng là phải hiểu
 sự phụ thuộc lẫn nhau giữa hai điểm số này về mặt quyết định mức độ liên quan.

Để mô hình hóa mối liên hệ giữa sự liên kết giáo dục và khả năng đọc, chúng tôi thực hiện
 lấy cảm hứng từ một chương trình trọng số cốt lõi cho Truy xuất thông tin: TF-IDF. TF (hoặc
 tần suất thuật ngữ) nắm bắt được sự nổi bật của một thuật ngữ trong một tài nguyên, trong khi IDF
 (hoặc tần suất tài liệu nghịch đảo) đặc trưng cho “lượng thông tin được mang theo
 một thuật ngữ, như được định nghĩa trong lý thuyết thông tin” [27] và được tính theo tỷ lệ của
 kích thước của một bộ sưu tập trên số lượng tài nguyên trong bộ sưu tập mà thuật ngữ
 xuất hiện. Trong trường hợp của chúng tôi, lược đồ trọng số này hoạt động như một loại “bộ trộn” cho các đặc điểm
 thông báo sự liên quan. Theo trực giác, chúng tôi coi Sedu là đại diện cho nội dung của
 R (về mặt phù hợp với bối cảnh lớp học) và khả năng đọc hiểu như là yếu tố phân biệt
 yếu tố liên quan đến các nguồn lực được xem xét cho mục đích xếp hạng. Với thường xuyên
 mức độ dễ đọc cao của các nguồn tài nguyên trực tuyến [10, 12], chúng tôi sử dụng 13 làm mức độ dễ đọc
 đại diện của bộ sưu tập, và do đó sử dụng nó như khả năng đọc tối đa trong
 tử số cho IDF. Với điều này trong tâm trí, điểm số của bộ trộn cho R được thông báo bởi hai
 các tín hiệu liên quan đã đề cập ở trên được tính như trong Phương trình 3.15.

$$\text{bộ trộn}(R) = \text{Sread}(R) \times \log_2 \left(\frac{13}{\text{Sedu}(R)} \right) \quad (3.15)$$

Bằng cách kết hợp nhiều tín hiệu có liên quan vào việc xác định mức độ liên quan lợi nhuận và việc mở rộng DCG với hệ số nhạy cảm về chi phí, chúng tôi đã xác định một số liệu được cập nhật dùng để đảm bảo REDORank học cách phản hồi rõ ràng với nhóm người dùng, nhiệm vụ và yêu cầu về môi trường, bằng cách ưu tiên các nguồn lực phù hợp với nhóm người dùng và môi trường của chúng tôi, đồng thời ngăn chặn việc trình bày ở mức cao xếp hạng các nguồn tài nguyên được khai thác có hại cho môi trường của chúng ta.

CHƯƠNG 4

KẾT QUẢ THÍ NGHIỆM

Trong chương này, chúng tôi mô tả các thí nghiệm chúng tôi đã tiến hành để trả lời câu hỏi nghiên cứu. Chúng tôi bắt đầu bằng cách đánh giá tính đúng đắn của khả năng đọc công thức là một phần của thiết kế REdORank, cũng như hiệu suất của đề xuất các chiến lược để phát hiện các nguồn tài nguyên giáo dục và phản cảm. Điều này cho phép chúng tôi chứng minh rằng các phương pháp luận được xem xét để giải thích cho từng phương pháp của chúng tôi quan điểm là hợp lý. Sau đó, chúng tôi đánh giá thiết kế tổng thể của REdORank, thông qua cả một nghiên cứu cắt bỏ và so sánh với các đối tác ban đầu. Trên đường đi, chúng tôi cung cấp phân tích chuyên sâu về kết quả của từng thí nghiệm.

4.1 Tìm công thức khả năng đọc phù hợp với tài nguyên web

Không có công thức khả năng đọc nào là mặc định khi ước tính độ phức tạp của các văn bản. Vì vậy, điều cần thiết là chúng ta phải kiểm tra các công thức một cách thực nghiệm trong nỗ lực xác định phương pháp phù hợp nhất để xác định mức độ phức tạp của tài nguyên Web. Trong bài kiểm tra của mình, chúng tôi xem xét hiệu quả của các công thức khả năng đọc ban đầu của chúng mục đích dự kiến: ước tính mức độ đọc của các văn bản đã xuất bản, tức là sách và bài báo tin tức. Sau đó, chúng tôi điều tra cách các công thức thực hiện khi áp dụng vào đoạn văn bản của các nguồn tài nguyên Web. Thông qua việc so sánh các kết quả trong mỗi phương tiện, sách và Web, chúng tôi lựa chọn công thức phù hợp nhất với đối tượng và bối cảnh của mình.

Chúng tôi bắt đầu khám phá của mình bằng các công thức khả năng đọc truyền thống vì chúng đơn giản để tính toán và được áp dụng rộng rãi [51, 82]. Các công thức truyền thống cũng yêu cầu ít dữ liệu hơn so với các giải pháp máy học và học sâu và có thể truy cập miễn phí, hỗ trợ tính khả dụng mở dành cho REdORank. Ban đầu, chúng ta xem xét (i) Chỉ số Coleman-Liau (Phương trình 4.1), vì nó được thiết kế cho các văn bản kỹ thuật số và được dễ dàng tính toán tự động [25], (ii) Flesch-Kincaid (Phương trình 4.2) [66], vì nó là một công thức nổi tiếng đã được sử dụng để ước tính độ phức tạp của các tài nguyên Web tập trung chủ yếu vào các lớp từ tiểu học đến trung học [15], và (iii) Spache Công thức khả năng đọc (viết tắt là Spache, Phương trình 4.3) [113], dành cho các văn bản mục tiêu cho độc giả ở lớp 1-3. Phần sau dựa trên vốn từ vựng tính gồm 1.064 từ được coi là dễ hiểu đối với trẻ em. Mỗi một trong ba công thức này là được thiết kế để ước tính mức độ đọc của các tài liệu đã xuất bản, ví dụ, sách hoặc tin tức và bài viết tạp chí.

$$\text{Coleman-Liau}(R) = (0,058 \times |L_R|) - (0,296 \times |S_R|) - 15,8 \quad (4.1)$$

trong đó R là một nguồn tài nguyên nhất định, $|L_R|$ là số chữ cái trong R và $|S_R|$ biểu diễn số câu trong R .

$$\text{Flesch-Kincaid}(R) = (0,39 \times \text{s}L_R) + (11,8 \times \text{sp}w_R) - 15,59 \quad (4.2)$$

trong đó $\text{s}L_R$ biểu thị độ dài câu trung bình của R và $\text{sp}w_R$ biểu thị số của số âm tiết trên một từ trong R .

$$\text{Spache}(R) = (0,141 \times w_R/s_R) + (0,086 \times \text{dif}(R)) + 0,839 \quad (4.3)$$

trong đó wR và sR lần lượt là số từ và câu trong R .

hàm $dif(R)$ xác định tỷ lệ phần trăm các từ khó trong R , trong đó một từ là được coi là khó nếu nó không xuất hiện trong từ vựng "từ dễ"¹.

Mặc dù có nhiều tập dữ liệu có thể được sử dụng để đánh giá hiệu suất của các công thức khả năng đọc, theo hiểu biết của chúng tôi, không có công thức nào được thiết kế cho Web nguồn lực, đối tượng mục tiêu của chúng tôi hoặc được dán nhãn cụ thể theo trình độ lớp. Với điều đó trong lưu ý, chúng tôi đã tạo ra TextComp của riêng mình, được biểu thị bao gồm 4.860 trường hợp biểu mẫu <mẫu văn bản, nhãn điểm, nguồn>. Chúng tôi đã bao gồm rõ ràng trong TextComp mẫu tài nguyên từ các phương tiện in và kỹ thuật số cho phép chúng tôi thăm dò khả năng áp dụng các công thức khác nhau cho đối tượng mục tiêu và bối cảnh của chúng tôi. Các mẫu trong TextComp được phân phối như sau:

- 235 trích đoạn sách được trích từ các phụ lục của CCSS [62], mỗi phần liên kết liên quan đến một loạt các cấp lớp. Chúng tôi lựa chọn sử dụng cấp lớp tối thiểu từ những phạm vi này như nhãn, vì trẻ em đọc dưới trình độ đọc của chúng ít gặp khó khăn hơn với sự hiểu biết so với khi đọc ở trên mức độ đọc [7].
- 2.084 cuốn sách từ Reading AZ (RAZ) được dán nhãn với nội dung đọc tương ứng cấp độ 2.
- 2.541 tài nguyên Web từ Liên minh học tập kỹ thuật số Idaho (IDLA), một cộng đồng việc đọc các tài liệu khóa học trực tuyến phục vụ cho học sinh K-12 [5], mỗi tài liệu liên quan đến một điểm số được xác định trước bởi các nhà giáo dục chuyên gia.

¹https://github.com/cdimascio/py-readability-metrics/blob/master/readability/dữ_liệu/spache_easy.txt

2RAZ sử dụng thang điểm 26 chữ cái do các chuyên gia chỉ định để dễ đọc [72]. Để có thể so sánh công bằng giữa các công thức, các nhãn chữ cái này được ánh xạ nhãn lớp từ mẫu giáo đến lớp 6, sử dụng bảng chuyển đổi do RAZ cung cấp [71].

Để định lượng hiệu suất của từng công thức F được xem xét trong quá trình khám phá của chúng tôi, chúng tôi dựa vào Tỷ lệ lỗi (ER), được tính như trong Phương trình 4.4. Chúng tôi xác định ý nghĩa của kết quả của chúng tôi sử dụng kiểm định Kruskal-Wallis H [67] với $p < 0,05$. Trừ khi có quy định khác đã nêu, tất cả các kết quả được báo cáo trong phần còn lại của phần này đều có ý nghĩa.

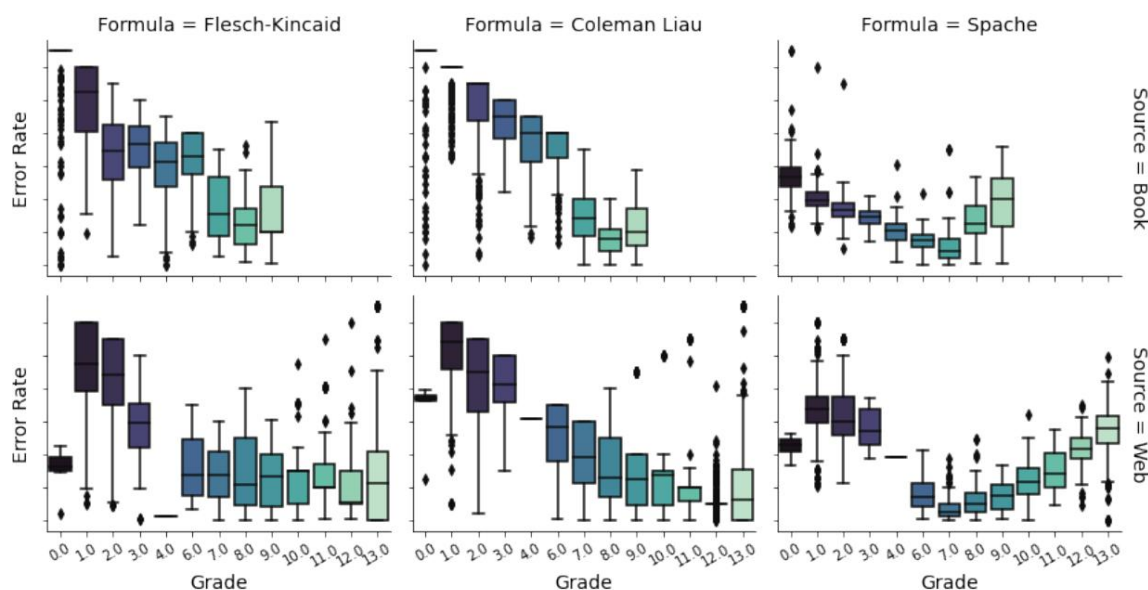
$$ER(F) = \frac{1}{|V_{\text{bảnComp}}|} \sum_{d \in \text{TextComp}} |T C^d, F - T C^d| \quad (4.4)$$

trong đó $|T C^d, F|$ là kích thước của TextComp , d là một thể hiện trong TextComp , $T C^d$ là cấp độ đã biết cho d và $T C^d, F$ là mức độ của d ước tính bằng cách sử dụng F .

Để đạt được sự hiểu biết cơ bản về cách thức mỗi công thức thực hiện khi áp dụng vào nguồn tài nguyên mục tiêu ban đầu, chúng tôi tính toán ER bằng cách sử dụng 2.319 cuốn sách trong TextComp . Như thể hiện trong Hình 4.1, Chỉ số Coleman-Liau thể hiện tỷ lệ lỗi thấp hơn Flesch-Kincaid và Spache ở trình độ lớp 9 trở lên. Thật thú vị, thậm chí mặc dù đây là công thức thường được sử dụng [15], Flesch-Kincaid tạo ra ER lớn nhất trên toàn bộ trình độ lớp khi ước tính độ phức tạp của sách. Spache là lỗi ít nhất dành cho học sinh từ lớp K đến lớp 6, do đó phù hợp nhất với đối tượng học sinh của chúng tôi (lớp K-4).

Để xác thực xem hiệu suất này có chuyển thành tài nguyên Web hay không, chúng tôi lặp lại tương tự thử nghiệm sử dụng các tài nguyên Web trong TextComp . Giống như đối với sách, Spache là công thức ít có khả năng mắc lỗi nhất đối với lớp 1-4 và 6-8 (xem Hình 4.1). Ngược lại đối với sách, Flesch-Kincaid có giá tốt hơn cho các nguồn tài nguyên Web lớp 8 và lớp 10, với Coleman-Liau thực hiện tốt nhất cho các lớp còn lại. Kết quả từ trước phân tích được gửi đi phục vụ như một dấu hiệu cho thấy Spache là công thức phù hợp nhất cho Nhiệm vụ trong tầm tay: ước tính trình độ của các nguồn tài nguyên Web hướng đến người tìm kiếm trẻ tuổi.

Bất kể hiệu quả của nó đối với khán giả và bối cảnh của chúng tôi, chúng tôi lưu ý rằng Spache vốn từ vựng bị hạn chế và được cập nhật lần cuối vào những năm 1970. Khi ngôn ngữ thay đổi



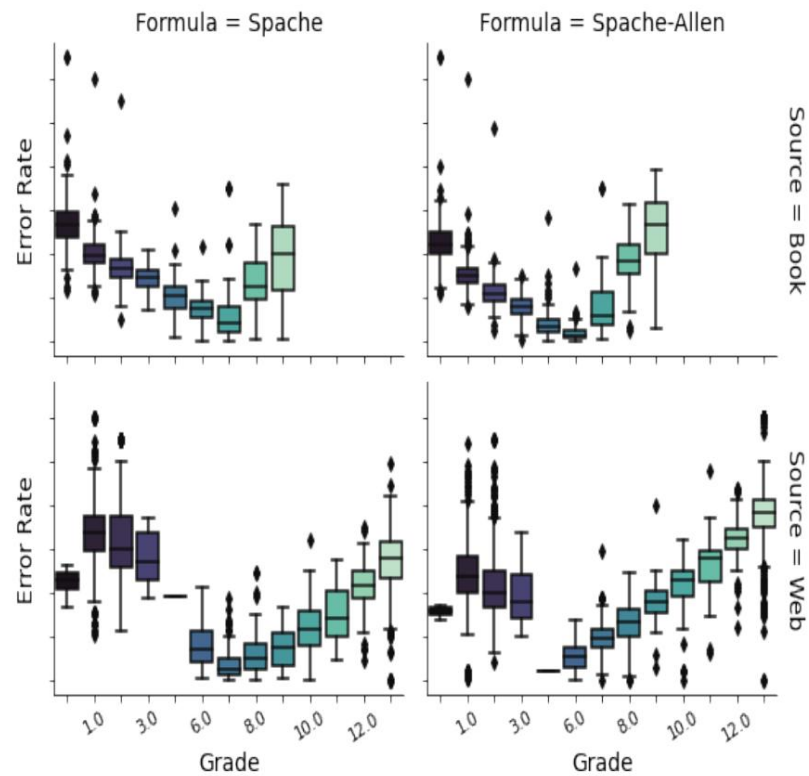
Hình 4.1: Phân phối ER giữa các cấp độ khác nhau cho các công thức khả năng đọc hiểu truyền thống.

theo thời gian [106], một vốn từ vựng lỗi thời có thể không nắm bắt được các thuật ngữ dễ hiểu đối với trẻ em thế giới ngày nay, có khả năng dẫn đến công thức ước tính sai lệch về độ phức tạp của một văn bản. Công thức Dale-Chall mới [22] đã tăng vốn từ vựng được xem xét theo công thức Dale-Chall ban đầu [31] từ 763 đến 3.000 vào những năm 1990 nhằm tìm cách cập nhật công thức để đáp ứng một tập hợp các đoạn văn mới với các cấp lớp được chỉ định để so sánh nhằm xác định độ khó của văn bản, được gọi là “đoạn văn tiêu chuẩn”, đối với sự phát triển của các công thức khả năng đọc [36]. Những hiểu biết sâu sắc này đã truyền cảm hứng cho chúng tôi theo đuổi một phiên bản mở rộng của từ vựng của Spache. Madrazo Azpiazu et al. [82] đã xem xét việc tăng cường danh sách từ vựng gốc của Spache, bằng cách đưa vào một từ điển trong số 48.000 thuật ngữ được lemmat hóa không ngừng mà các tác giả đã trích xuất từ các tài liệu liên quan đến trẻ em trang web (Sven) như một phần của từ vựng được xem xét bởi công thức. công thức đã được sử dụng thành công để xác định xem một truy vấn có giống trẻ em không. Tuy nhiên, điều này sự cải tiến dựa trên phân tích tần suất từ của các trang web liên quan đến trẻ em và giả định rằng các thuật ngữ được thêm vào vốn từ vựng sẽ được trẻ em hiểu, điều này có thể

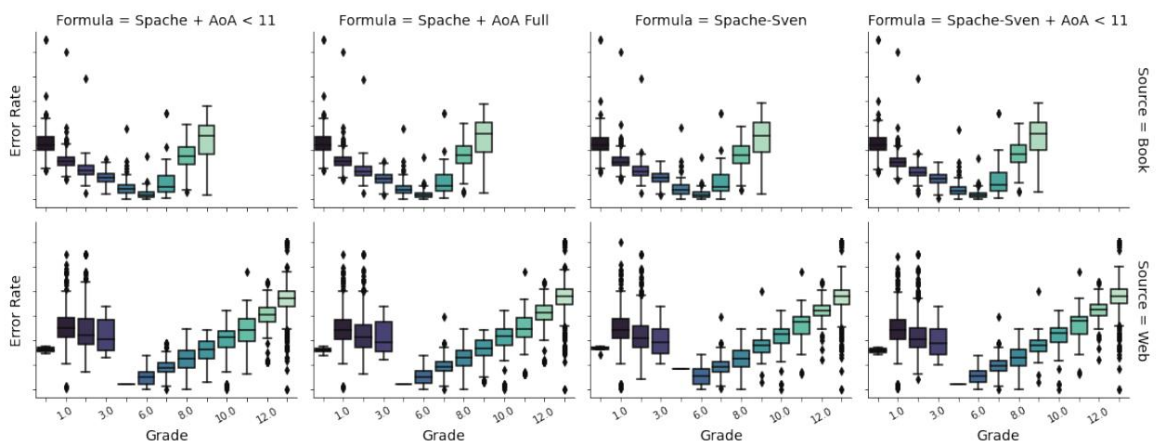
không phải lúc nào cũng như vậy. Với mục đích bao gồm vốn từ vựng mà trẻ em học được thông qua hướng dẫn, chúng tôi tận dụng bộ dữ liệu Tuổi tiếp thu (AoA). Bộ dữ liệu này chứa xếp hạng thu thập dữ liệu theo độ tuổi, từ 1-17 năm, cho khoảng 30.000 từ tiếng Anh [68], mà chúng tôi sử dụng để bổ sung cho Spache gốc từ vựng, cụ thể bao gồm các thuật ngữ từ AoA với độ tuổi tiếp thu trung bình bằng hoặc dưới 11 tuổi vì điều này phù hợp với đối tượng mục tiêu của chúng tôi. Chúng tôi tiếp tục kiểm tra sự gia tăng của Spache mở rộng vốn từ vựng ban đầu với tất cả của các từ trong AoA, cũng như việc các từ AoA đóng vai trò là từ hỗ trợ duy nhất từ vựng. Chúng tôi cho rằng có lợi ích khi đồng thời xem xét thuật ngữ mà trẻ em đã được tiếp xúc thông qua các trang web cũng như thuật ngữ đã được dạy, do đó chúng tôi khám phá sự kết hợp của từ vựng Spache gốc với các thuật ngữ kết hợp từ AoA và từ điển từ [82].

Thực hiện theo cùng một quy trình thử nghiệm được sử dụng để so sánh Flesch-Công thức Kincaid, Coleman-Liau và Spache, chúng tôi so sánh ER của công thức đã sửa đổi Công thức Spache. Đầu tiên, chúng tôi kiểm tra cách thực hiện của từng công thức bằng cách sử dụng các nguồn sách trong TextComp, kết quả có thể được nhìn thấy trong Hình 4.2. Công thức Spache sử dụng từ vựng gốc dễ dàng dễ mắc lỗi hơn so với khi sử dụng từ vựng mở rộng. Trong khi sự khác biệt trong ER ít rõ rệt hơn đối với các công thức sử dụng từ vựng mở rộng, khi từ vựng kết hợp Sven, tỷ lệ lỗi là thấp nhất đối với lớp K-5. Chúng tôi gán lợi ích này cho Sven được thiết kế riêng cho những từ mà trẻ em sử dụng, trong khi bản gốc của Spache từ vựng là danh sách các thuật ngữ chung được coi là dễ hiểu. Không có ý nghĩa quan trọng sự khác biệt trong ER khi áp dụng cho các nguồn lực Lớp 6. Nhìn lại Hình 4.2, chúng ta xem xu hướng tương tự cho các tài nguyên Web, với công thức mang lại tỷ lệ lỗi thấp nhất đối với các tài nguyên dưới lớp 6 là phiên bản Spache sử dụng Sven + AoA

Đầy đủ (tức là Spache-Allen). Từ thí nghiệm này, chúng ta thấy rằng việc mở rộng bản gốc từ vựng Spache có tác động tích cực đến ER.



(a) ER cho Spache và Spache-Allen.



(b) ER cho các biến thể của Spache sử dụng AoA và Sven.

Hình 4.2: Phân bố ER cho các công thức lấy cảm hứng từ Spache theo từng cấp độ.

Công thức khả năng đọc nào phù hợp đồng thời với loại tài nguyên, ngữ cảnh nhu cầu và nhóm người dùng được phác thảo cho nhiệm vụ của chúng tôi? (RQ1) Với trọng tâm của chúng tôi là trẻ em từ lớp K đến lớp 4 và ER giảm đáng kể của công thức Spache với từ vựng đầy đủ của Sven + AoA, chúng tôi coi phiên bản mở rộng này, được định nghĩa là Spache-Allen trong Phương trình 3.1, là phù hợp nhất để đưa vào thiết kế xếp hạng REORank.

4.2 Phát hiện tài nguyên web giáo dục

Việc xác định các nguồn tài nguyên trực tuyến phù hợp với kết quả học tập chung cho Học sinh K-4 trong số những người được xem xét để xếp hạng lại là một phần quan trọng của REORank. Để làm như vậy, chúng tôi đã giới thiệu BiGBERT (Phần 3.1.2), một mô hình phân loại học sâu. Để đảm bảo rằng kết quả của BiGBERT tác động trực tiếp đến hiệu suất của REORank, có thể đóng vai trò là một tín hiệu hiệu quả về sự liên quan, chúng tôi thực hiện một đánh giá BiGBERT, chúng tôi sẽ thảo luận bên dưới.

4.2.1 Thiết lập thử nghiệm

Để đánh giá hiệu suất của BiGBERT, chúng tôi tiến hành một nghiên cứu cắt bỏ để chứng minh rằng mỗi thành phần-vector URL, vector đoạn mã và truyền tải kiến thức miền- là cần thiết để BiGBERT xác định các nguồn lực phù hợp với giáo dục. Ngoài ra, chúng tôi tiến hành so sánh với các mô hình phân loại tương tự để ngữ cảnh hóa hiệu suất của BiGBERT.

Một cuộc khám phá sâu sắc về tài liệu trong lĩnh vực này cho thấy rằng không có tập dữ liệu chúng ta có thể sử dụng để đánh giá hiệu suất của các mô hình xác định giáo dục sự liên kết quốc tế của các nguồn tài nguyên Web. Vì vậy, chúng tôi xây dựng một cái mà chúng tôi gọi là EduSites,

sử dụng URL (có văn bản bằng tiếng Anh) từ Alexa Top Sites [6]-dựa trên Dự án thư mục mở (ODP) [23, 92]. Chúng tôi coi 1.273 URL trong là giáo dục tiểu thể loại Mẫu giáo và Thời gian học từ Trẻ em & Thanh thiếu niên. Chúng tôi cũng ngẫu nhiên chọn 3.998 URL không liên quan đến giáo dục được phân bổ đồng đều giữa Người lớn, Doanh nghiệp, Giải trí và Trò chơi. Để xác thực rằng các nhãn trong EduSites phù hợp (hoặc không) với định nghĩa của chúng tôi về giáo dục, một chuyên gia giáo dục đã chú thích một mẫu đại diện ($n = 527$). Như trong [94], chúng tôi tính toán độ chính xác giữa hai chú thích (Alexa so với chuyên gia) cho mỗi mẫu, đạt được sự nhất trí giữa các chú thích viên là 94,7%.

Để đánh giá hiệu suất, chúng tôi sử dụng Độ chính xác, một số liệu phân loại phổ biến, cùng với tỷ lệ Dương tính giả (FPR) và Âm tính giả (FNR), để cung cấp thông tin chi tiết về loại tài nguyên được phân loại sai. Một kết quả dương tính giả là một tài nguyên được đánh dấu là giáo dục, tức là không. Một kết quả âm tính giả là ngược lại, một nguồn tài nguyên giáo dục được đánh dấu là không mang tính giáo dục. Ý nghĩa của kết quả được xác định bằng McNemar kiểm tra, $p < 0,05$.

Theo hiểu biết của chúng tôi, không có bộ phân loại miền cụ thể nào mà chúng tôi có thể sử dụng để ngữ cảnh hóa hiệu suất của BiGBERT. Do đó, chúng tôi tối ưu hóa và điều chỉnh một số bộ phân loại để phát hiện tài nguyên Web K-4:

1. BoW [49], một mô hình túi từ tính toán độ tương đồng cosin giữa

URL tài nguyên được vector hóa và mô tả danh mục ODP để xác định lại

danh mục tương ứng của nguồn (lưu ý rằng chúng tôi sử dụng văn bản về kết quả học tập từ các tiêu chuẩn giáo dục thay cho mô tả danh mục).

2. BGCNN [105], một mô hình dựa trên BiGRU với CNN xác định trẻ em-

URL thân thiện.

3. BERT4TC [130], một bộ phân loại văn bản sử dụng bộ mã hóa BERT để thực hiện chủ đề và phân loại tình cảm.

4. Hybrid-NB [1], một mô hình lai kiểm tra cả URL và nội dung của các trang web để xác định đối tượng mục tiêu của họ (tức là người dùng Algeria). Đã báo cáo kết quả cho BGCNN và BERT4TC là giá trị trung bình của xác thực chéo 5 lần.

Bên cạnh các bộ phân loại đã đề cập ở trên, chúng tôi khám phá các biến thể của BiGBERT, trong đó U, S và E chỉ ra khi BiGBERT chỉ kiểm tra các URL, đoạn trích và truyền tải giáo dục thông tin quốc tế, tương ứng. Thông qua một nghiên cứu cắt bỏ, chúng tôi trình bày những đóng góp của URL và trình vector hóa đoạn mã hướng tới kiến trúc tổng thể của BiGBERT.

4.2.2 Kết quả và thảo luận

Chúng tôi cung cấp dưới đây một phân tích về kết quả của các thí nghiệm được tiến hành để đánh giá thiết kế của BiGBERT. Chúng tôi tóm tắt những kết quả này trong Bảng 4.1.

Bảng 4.1: Phân tích hiệu suất của BiGBERT (nghiên cứu cắt bỏ cùng với các thí nghiệm liên quan đến việc so sánh với các đối tác). Các hậu tố -U và -S chỉ ra mô hình chỉ áp dụng cho URL và đoạn trích, tương ứng; -E biểu thị mô hình được tăng cường bằng dữ liệu giáo dục. * và † có ý nghĩa đối với BiGBERT và đối tác không phải giáo dục, tương ứng. Ý nghĩa được xác định bằng kiểm định McNemar, $p < 0,05$.

Kiểu hàng		Độ chính xác FPR FNR			
1	Đường cơ sở BoW	.7205	.115	.796	
2	Tình trạng của nghệ thuật	* BGCNN	* .8399	.073 .432	
		BERT4TC .9353	* Hybrid-	.041 .140	
3 4		NB .8600	* BiGBERT-U	.145 .123	
5	Phá hủy Học	.8276	* BiGBERT-UE	.073 .484	
6		.8287	* † .9374	* BiGBERT-	.072 .483
7		BiGBERT-SE	S	.027 .175	
8		.9334	* BiGBERT-US	.038 .155	
9		.9381	*	.035 .146	
10		BiGBERT	.9533 †	.027 .106	

Các báo cáo trong [105] cho thấy hiệu quả của việc chỉ kiểm tra URL để xác định các trang web thân thiện với trẻ em. Điều này thúc đẩy chúng tôi nghiên cứu khả năng áp dụng của phương pháp tiếp cận để phát hiện các nguồn tài nguyên Web giáo dục nhằm mục tiêu vào nhóm dân số K-4. Độ chính xác của BoW không vượt quá mốc 75% đạt được thông qua đường cơ sở ngây thơ (luôn luôn dự đoán không mang tính giáo dục do bản chất mất cân bằng của tập dữ liệu của chúng tôi). BGCNN, BiGBERT-U và BiGBERT-UE vượt trội hơn các mô hình truyền thống về độ chính xác tỷ lệ ở mức phần trăm thấp 80. Chúng tôi cho rằng sự gia tăng hiệu suất là do thực tế các mô hình tiên tiến không cho rằng mã thông báo URL là độc lập, không giống như BoW. Kết quả từ phân tích của chúng tôi chỉ ra rằng khi thông tin giàu ngữ nghĩa và ngữ cảnh có sẵn, URL là nguồn có giá trị để thông báo phân loại. Số lượng Tuy nhiên, các nguồn tài nguyên giáo dục được phân loại sai trong trường hợp này là cao. Trên thực tế, gần một nửa của các mẫu giáo dục, chiếm 25% dữ liệu của chúng tôi, được dán nhãn là không giáo dục (xem FNR tương ứng). Điều này dẫn chúng ta đến việc điều tra các nguồn thông tin bổ sung có thể đóng góp vào quá trình phân loại.

Vì phân tích nội dung là một yếu tố chính của phân loại, nên việc xem xét kiến thức là hợp lý. cạnh được suy ra từ các đoạn trích để hỗ trợ tốt hơn cho việc phân loại giáo dục K-4 Tài nguyên web. Điều này được chứng minh bằng những cải tiến hiệu suất đáng kể của Hybrid-NB, BiGBERT-US và BiGBERT so với các đối tác chỉ xem xét URL (BoW và BGCNN). BiGBERT vượt trội hơn đáng kể so với các mô hình lai về độ chính xác và FPR. Ít kết quả dương tính giả hơn có nghĩa là khả năng xảy ra kết quả không phù hợp tiềm ẩn thấp hơn các trang web được gắn nhãn giáo dục, điều này có tầm quan trọng đặc biệt đối với tên miền và đối tượng của công việc của chúng tôi. Kết quả cho thấy rằng các đoạn trích, kết hợp với URL, có giúp ích xác định các nguồn lực giáo dục. Tuy nhiên, FNR cao hơn của BiGBERT-US so với đến Hybrid-NB, một lần nữa chỉ ra sự phân loại sai các nguồn lực giáo dục. Điều này có thể được nhìn thấy trên các mẫu như www.sesstreet.org, được công nhận là có tính giáo dục bởi

Hybrid-NB nhưng bị BiGBERT-US bỏ qua. Điều này cho thấy rằng việc thiếu

kiến thức chuyên môn rõ ràng gây bất lợi cho BiGBERT-US.

Độ chính xác của BiGBERT tăng lên khi sử dụng Edu2Vec và BERT được tinh chỉnh

nhúng (hàng 9 so với 10 trong Bảng 4.1). Để xác định xem sự cải thiện có phải là

kết quả của việc truyền tải rõ ràng kiến thức giáo dục vào quá trình phân loại, chúng tôi

so sánh BiGBERT-U và BiGBERT-S với các đối tác được tăng cường giáo dục.

các thí nghiệm cho thấy sự giảm đáng kể trong FPR và FNR giữa BiGBERT-U và

BiGBERT-UE; không đáng kể giữa BiGBERT-S và BiGBERT-SE. Không giống như URL

biến thể, hiệu suất của BiGBERT-S-E chỉ được cải thiện trong FNR sau khi tăng cường.

Chúng tôi cho rằng điều này là do bộ đào tạo tương đối nhỏ được sử dụng để tinh chỉnh khi so sánh

vào bộ tiền đào tạo ban đầu cho BERT, dẫn đến ít thông tin ngữ cảnh mới hơn

được học bởi mô hình máy biến áp tiêu chuẩn. Tuy nhiên, sự gia tăng đáng kể trong

độ chính xác và giảm FPR và FNR đối với BiGBERT khi so sánh với BiGBERT-US

đề xuất rằng kiến thức chuyên ngành có thể có tác động tích cực đến việc phân loại

của các nguồn tài nguyên giáo dục. Điều này được minh họa bằng URL www.xpmath.com, một trang web

để hỗ trợ giáo dục toán học ở lớp 2-9, được dán nhãn là không mang tính giáo dục bởi

BiGBERT-US, tuy nhiên nó được BiGBERT công nhận là có mục đích giáo dục.

Các đoạn trích cùng với URL có giúp xác định các nguồn tài nguyên giáo dục không?

kiến thức chuyên ngành ảnh hưởng đến việc xác định các nguồn lực giáo dục

(RQ2) Từ những kết quả được trình bày cho đến nay, có thể thấy rằng cả URL và

cần có các đoạn trích để mô tả đầy đủ sự liên kết giáo dục của các nguồn lực.

Hơn nữa, bằng cách truyền tải rõ ràng kiến thức chuyên môn vào thiết kế của BiGBERT,

phạm vi các nguồn lực giáo dục được xác định mở rộng so với việc chỉ sử dụng truyền thống

BERT quốc tế. Nhìn chung, xét đến thành công mà BiGBERT đạt được trong việc xác định giáo dục K-4

nguồn lực quốc gia trong khi giảm thiểu các kết quả âm tính giả và dương tính giả, chúng tôi coi BiGBERT

như phù hợp để đưa vào thiết kế của REDORank. Bằng cách tối ưu hóa trên đầu ra của BiGBERT như một tín hiệu liên quan, REDORank có thể giải quyết môi trường của chúng ta và khán giả.

4.3 Xác định các nguồn tài nguyên phản cảm

REDORank được thiết kế để hạ cấp các tài nguyên Web có nội dung không phù hợp với trẻ em bối cảnh lớp học. Với việc REDORank dựa vào Judgebad (được giới thiệu trong Phần 3.1.3) để xác định các loại tài nguyên này, điều bắt buộc là phải xác minh độ tin cậy của nó đối với tránh sự lan truyền lỗi. Do đó, chúng tôi tiến hành phân tích sâu về hiệu suất, mà chúng ta sẽ thảo luận dưới đây.

4.3.1 Thiết lập thử nghiệm

Theo hiểu biết của chúng tôi, không tồn tại một tập dữ liệu được gắn nhãn có phạm vi phủ sóng cho tất cả các danh mục trong ObjCat, do đó chúng tôi xây dựng một danh mục: ObjSet. Bộ dữ liệu này, được trích xuất từ thư mục Alexa Top Sites, bao gồm 10.006 mẫu của biểu mẫu <đoạn trích, URL, nhãn>, trong đó nhãn là 1 đối với các mẫu phản cảm và 0 nếu không. Chúng tôi coi 2.096 tài nguyên có liên quan tương ứng là đáng phản đối. Tên danh mục Alexa chứa một trong các danh mục ObjCat và danh mục con dưới dạng chuỗi con tên danh mục. 7.910 tài nguyên bổ sung còn lại từ Alexa đóng vai trò là các đối tác đáng phản đối. Bằng cách chọn các nguồn không đáng phản đối theo tỷ lệ khoảng 4:1 tỷ lệ với những thứ đáng phản đối, chúng tôi mô phỏng một bối cảnh thế giới thực nơi mà những nguồn tài nguyên đáng phản đối sẽ chiếm một phần nhỏ hơn trong SERP.

Để đo hiệu suất, chúng tôi sử dụng Độ chính xác, FPR và FNR. Trong trường hợp này, một dương tính giả là một nguồn tài nguyên được đánh dấu là phản đối nhưng không phải vậy. Một âm tính giả là

ngược lại, một nguồn tài nguyên đáng ngờ được đánh dấu là không đáng ngờ. Hơn nữa, chúng tôi so sánh và đối chiếu Judgebad với một số mô hình đối ứng, mỗi mô hình áp dụng một chiến lược khác để xác định các nguồn tài nguyên đáng ngờ. Thông qua điều này so sánh chúng ta có được hiểu biết sâu sắc và ngữ cảnh hóa cách Judgebad thực hiện một cách tôn trọng với các giải pháp hiện có.

- MNB. Một mô hình Multinomial Naive-Bayes dạng túi từ tính toán TF-IDF cho các mô tả tài nguyên do ODP cung cấp để xác định tài nguyên lớp tương ứng.
- BERT4TC [130]. Một bộ phân loại văn bản sử dụng công nghệ BERT tiên tiến nhất bộ mã hóa kết hợp với một perceptron nhiều lớp để thực hiện chủ đề và tình cảm phân loại.
- TUYỆT VỜI [1]. Một khuôn khổ kết hợp từ vựng tình cảm VADER [61] với BERT để dự đoán cường độ tình cảm của câu.
- Tính phù hợp của KSA [87]. Tập trung vào cùng một nhóm người dùng và môi trường, mô hình này tận dụng một từ điển được quản lý để phân tích tần suất thuật ngữ và thuật ngữ Tỷ lệ trong nội dung của một nguồn tài nguyên Web quyết định tính phù hợp.

Để đảm bảo sự so sánh công bằng giữa các mô hình, ObjSet được chia thành một chương trình đào tạo và bộ kiểm tra sử dụng phân chia 80/20 và tất cả các đoạn mã được xử lý trước theo cùng một cách cách thức: được mã hóa, loại bỏ dấu câu và được lemmat hóa. Ý nghĩa của kết quả là được xác định bằng kiểm định McNemar, $p < 0,05$.

4.3.2 Kết quả và thảo luận

Như thể hiện trong Bảng 4.2, Judgebad đạt được điểm chính xác là 84,9% - một tỷ lệ vượt quá hiệu suất mong đợi của bộ phân loại đa số (79%) cho dữ liệu phân phối đóng góp trong ObjSet. Khi xem xét sâu hơn về hiệu suất của Judgebad, chúng tôi nhận thấy rằng, thú vị là, Judgebad có FPR cao hơn FNR. Chúng tôi quy kết các kết quả âm tính giả, tức là những thứ được đánh dấu là không đáng phản đối khi chúng đáng phản đối, đối với độ dài ngắn hơn của một số từ điển ObjCat, vì từ điển như vậy cung cấp ít khả năng hơn đại diện rõ ràng cho một danh mục nhất định. Điều này có thể nhìn thấy khi kiểm tra thủ công bởi các nguồn tài nguyên liên quan đến rượu và thuốc lá, một tiểu thể loại của Ma túy, bị dán nhãn sai như không thể phản đối. Từ điển về thuốc có tổng cộng 100 thuật ngữ, trong đó chỉ có 21 liên quan đến rượu và thuốc lá. Kiểm tra thủ công các phản đối bị bỏ sót tài nguyên cho thấy rằng trong một số tình huống, Judgebad xác định đúng tài nguyên là đáng phản đối có nhãn không đáng phản đối, nhưng không phù hợp với trẻ em trong lớp học. Ví dụ, hãy xem xét www.kids-in-mind.com, một trang web trực tuyến nền tảng cung cấp các bài đánh giá về nội dung phim (máu me, ngôn ngữ người lớn, khỏa thân, v.v.) để cha mẹ có thể quyết định xem nên cho con xem phim nào, hoặc www.casinocity.com, một trang web thư mục đánh giá liên quan đến sòng bạc và cờ bạc và trò chơi. Cả hai mẫu đều được dán nhãn là không phản cảm theo mô tả của Alexa, tuy nhiên, không có trang web nào phù hợp với trẻ em trong môi trường lớp học. Từ những trường hợp như thế này, chúng tôi cho rằng Judgebad đang đạt được mục tiêu mong muốn.

Để ngữ cảnh hóa màn trình diễn của Judgebad, chúng tôi so sánh nó với màn trình diễn của một số người có liên quan mô hình. MNB thể hiện hiệu suất rất giống với Judgebad, xét về FPR và FNR. Điều thú vị là có một sự khác biệt đáng kể về độ chính xác của hai mô hình. Đi sâu vào những tài nguyên nào đã được phân loại sai, có một số khác biệt

Bảng 4.2: Đánh giá các mô hình phân loại có thể phản đối bằng cách sử dụng ObjSet. * chỉ ra ý nghĩa liên quan đến Judgebad được xác định bởi thử nghiệm của McNemar với Bonferroni Hiệu chỉnh, $p < 0,05$.

Người mẫu	Độ chính xác	FPR	FNR
Judgebad 0,849 0,856* MNB BERT4TC	0,038	0,574	
TUYỆT 0,209*	0,002	0,679	
VỚI 0,209 KSTính phù hợp 0,209	1.0	0.0	
	1.0	0.0	
	1.0	0.0	

được tìm thấy. Kiểm tra thủ công các phân loại sai, chúng tôi thấy rằng MNB có xu hướng để xác định nhầm các trang web được biết là có chứa nội dung khiêu dâm. Chúng tôi kết nối những lỗi này với sự thiếu hụt chiều sâu từ vựng hiện diện trong mô hình túi từ. Thật đáng ngạc nhiên, BERT4TC, AWESOME và KSappropriateness luôn đóng vai trò là bộ phân loại thiếu sót gán nhãn phản cảm cho các tài nguyên. Với phân phối 4:1 của ObjSet, kỳ vọng là BERT4TC sẽ học cách phân loại các vật liệu không gây phản cảm. Chúng tôi gán kết quả không mong đợi cho việc lắp quá mức, theo nghĩa là mô hình đã học được các biểu diễn của các thuật ngữ trong các nguồn tài nguyên phản đối như các định danh rõ ràng bất kể của sự hiện diện hoặc bối cảnh của chúng trong các đối tác không phản đối. Với BERT phục vụ như một thành phần nội bộ cho AWESOME, chúng tôi cũng gán cho sự quá khớp này hành vi đối với AWESOME. Chúng tôi quy kết hiệu suất của KSappropriateness cho sự kết hợp của các tính năng giảm được xem xét (chỉ xem xét nội dung khiêu dâm và thù hận (thuật ngữ diễn đạt) và sự phân bố dữ liệu không đồng đều.

Từ điển theo chủ đề cụ thể có thể trao quyền xác định những điều đáng phản đối không? Tài nguyên web? (RQ3) Việc sử dụng từ điển mở rộng về các chủ đề không phù hợp đối với một lớp học, ngoài nội dung khiêu dâm và lời nói căm thù, hãy cung cấp thêm góc nhìn cho Judgebad để phân biệt những gì là và không phải là đáng phản đối. Nhìn chung, cho thấy

hiệu suất của Judgebad, chúng tôi cho rằng nó phù hợp để đưa vào thiết kế của REdORank.

4.4 Xếp hạng lại các tài nguyên web với REdORank

Cho đến nay, chúng tôi đã chỉ ra tính khả thi của các giải pháp để xác định mục tiêu giáo dục sự liên kết, khả năng đọc và khả năng phản đối của các nguồn tài nguyên. Sử dụng các quan điểm này cùng nhau, REdORank tìm cách xếp hạng lại các nguồn tài nguyên để hỗ trợ trẻ em tìm kiếm trong một môi trường lớp học bằng cách cân bằng bằng chứng về các nguồn lực không phù hợp, và do đó là một rủi ro, với bằng chứng cho thấy các nguồn tài nguyên có tính giáo dục và dễ đọc, và do đó có giá trị hoặc có lợi. Để xác thực rằng thiết kế của REdORank trả lời câu hỏi nghiên cứu của chúng tôi và do đó đạt được mục tiêu này, chúng tôi thực hiện một nghiên cứu toàn diện đánh giá mà chúng ta sẽ thảo luận trong phần còn lại của phần này.

4.4.1 Thiết lập thử nghiệm

Có những tập dữ liệu để đánh giá các mô hình xếp hạng dựa trên LTR, chẳng hạn như Bộ MQ2007 và MQ2008 [102] hoặc bộ OHSUMED [58]. Thật không may, có không có tập dữ liệu LTR bao gồm các truy vấn, tài nguyên và nhãn “lý tưởng” liên quan đến nhóm người dùng và môi trường của chúng tôi. Ngoài ra, không có tập dữ liệu hiện có nào bao gồm các nguồn tài nguyên phản đối đã biết, đây là điều bắt buộc để đánh giá rõ ràng tính hợp lệ của thiết kế REdORank. Dựa trên hai sự kiện này, chúng tôi xây dựng tập dữ liệu RankSet.

Việc xây dựng các tập dữ liệu cho các nhiệm vụ truy xuất thông tin thường tuân theo Mô hình Cranfield [124]. Đối với các nhiệm vụ xếp hạng, quá trình này bao gồm bắt đầu bằng tài nguyên “lý tưởng” đã biết. Tiêu đề của mỗi tài nguyên được sử dụng như một truy vấn để kích hoạt việc truy xuất các nguồn tài nguyên khác để tạo ra một danh sách được xếp hạng. Nguồn tài nguyên lý tưởng là

luôn được xếp ở vị trí đầu bảng xếp hạng vì được coi là sự thật cơ bản.

Các nguồn tài nguyên được xếp hạng N hàng đầu còn lại (trừ nguồn tài nguyên bắt đầu tìm kiếm, nếu có sẵn) được sử dụng để hoàn thành danh sách được xếp hạng. Mô hình Cranfield cho phép xây dựng RankSet để đảm bảo một nguồn tài nguyên lý tưởng ở vị trí hàng đầu cho mọi truy vấn. Tuy nhiên, REDORank cũng nhằm mục đích đẩy các tài nguyên phản đối xuống thấp hơn trong xếp hạng. Để có thể đánh giá khía cạnh này của REDORank, chúng tôi thêm vào cuối trong danh sách một nguồn tài nguyên được biết là “xấu”.

Để hoạt động như nguồn tài nguyên lý tưởng cho RankSet, chúng tôi sử dụng bộ sưu tập gồm 9.540 bài viết với trình độ đọc đã biết và giá trị giáo dục hướng đến trẻ em về nhiều chủ đề khác nhau từ NewsELA [91]. Đối với các tài nguyên xấu, chúng tôi chuyển sang ObjSet (Phần 4.3.1). Sau đây mô hình Cranfield, chúng tôi sử dụng các tiêu đề bài viết lý tưởng làm truy vấn và sử dụng Google API chúng tôi lấy tối đa 20 tài nguyên, tiêu đề, đoạn trích tìm kiếm và vị trí xếp hạng của chúng (chúng tôi loại bỏ các truy vấn không dẫn đến tài nguyên hoặc tài nguyên có nội dung bị thiếu). Chúng tôi gán nhãn liên quan 2 cho các nguồn lực lý tưởng, 0 cho các nguồn lực “xấu” đã biết và 1 cho tất cả các tài nguyên khác được lấy từ Google. Điều này dẫn đến RankSet chứa tổng cộng có 2.617 truy vấn và 46.881 tài nguyên.

Để chứng minh tính đúng đắn của thiết kế REDORank và khả năng áp dụng của nó, chúng tôi tiến hành một nghiên cứu cắt bỏ. REDORank sử dụng AdaRank làm LTR cơ bản thuật toán với số liệu CS-DCG mở rộng để tối ưu hóa. Để xác thực và xem xét cách (i) mở rộng số liệu tối ưu hóa từ phương pháp truyền thống hơn NDCG và (ii) việc kết hợp tính phản đối như một chi phí nhạy cảm, ảnh hưởng đến hiệu suất tổng thể, chúng tôi so sánh REDORank với AdaRank được tối ưu hóa theo tiêu chuẩn Chỉ số NDCG. Mỗi mô hình được cấu hình với các biến thể sử dụng từng góc nhìn như các tính năng độc lập. Để ngữ cảnh hóa hiệu suất của REDORank hơn nữa, chúng tôi thực hiện so sánh với hai mô hình khác: (i) LambdaMART, một danh sách phổ biến

Mô hình LTR sử dụng Cây hồi quy cộng tính đa [46], với tổng thể hàm xếp hạng là tổ hợp tuyến tính của các cây hồi quy và (ii) Korsce [87], một mô hình được thiết kế để xếp hạng các nguồn lực phù hợp với chương trình giáo dục từ lớp 3 đến lớp 5 chương trình giảng dạy, dễ hiểu đối với trẻ em trong cùng phạm vi lớp học, mang tính khách quan trong nội dung (tức là không dựa trên ý kiến), và phù hợp với lớp học, được mô tả chi tiết trong Phần 2.2. Chúng tôi coi LambdaMART là đường cơ sở, trong khi Korsce (phù hợp với nhóm người dùng, môi trường và bối cảnh của chúng tôi) là giải pháp hiện đại.

Để đo lường hiệu suất, chúng tôi sử dụng NDCG@10 và Xếp hạng tương hỗ trung bình (MRR). MRR tìm cách làm nổi bật vị trí xếp hạng trung bình của mục có liên quan đầu tiên. Trong trường hợp của chúng tôi, chúng tôi thấy việc định vị các nguồn tài nguyên phản đối rất quan trọng thấp trong số các kết quả thu được. Do đó, chúng tôi cũng tính toán một phiên bản thay thế của MRR, trong đó thay vì tính đến mục đầu tiên có liên quan (lý tưởng), chúng tôi tính đến cho vị trí của mục đầu tiên đáng phản đối. Chúng tôi gọi đây là MRRBad, trong đó một giá trị cho thấy hiệu suất tốt hơn. Ý nghĩa của kết quả được xác minh bằng cách sử dụng hai đuôi kiểm định t của sinh viên với $p < 0,05$; tất cả các kết quả được báo cáo và thảo luận trong phần còn lại của phần này có ý nghĩa quan trọng trừ khi có quy định khác.

4.4.2 Kết quả và thảo luận

Chúng tôi bắt đầu đánh giá việc áp dụng LTR cho trẻ em tìm kiếm trong lớp học bằng cách xem xét cách một thuật toán LTR theo danh sách đã biết, AdaRank, được tối ưu hóa cho một tiêu chuẩn chỉ số xếp hạng (NDCG), thực hiện khi được đào tạo để xếp hạng theo lựa chọn của chúng tôi quan điểm, sự liên kết giáo dục, khả năng đọc hiểu và khả năng phản đối. Chúng tôi đào tạo nhiều các ations của AdaRank với từng góc nhìn, sự liên kết về giáo dục, khả năng đọc và sự phản đối mỗi cái hoạt động như một tính năng duy nhất. Chúng tôi gọi những biến thể này bằng hậu tố -E, -R và -O tương ứng. Chúng tôi đào tạo cùng một tập hợp các biến thể cho REdORank

với việc bổ sung những cái sử dụng máy trộn (được mô tả trong Phần 3.2.2) để kết hợp sự liên kết giáo dục và quan điểm để đọc thành một tính năng duy nhất. Chúng tôi tham khảo với những hậu tố -M, trong đó các giá trị hỗn hợp là tính năng duy nhất và -MER nơi các giá trị hỗn hợp được sử dụng cùng với các quan điểm cá nhân. Kết quả của các thí nghiệm được trình bày trong Bảng 4.3 và 4.4.

Đầu tiên chúng ta xem xét từng góc nhìn riêng lẻ như một tính năng cho AdaRank. Như một dự đoán, AdaRank-O hoạt động kém nhất, như được thấy bởi NDCG và MRR thấp hơn điểm cũng như MRRBad cao hơn. Chúng tôi cho rằng điều này là do AdaRank-O đang tối ưu hóa cho góc nhìn "rủi ro" và do đó học cách ưu tiên tiềm năng tài nguyên xấu đã biết ở trên lý tưởng đã biết. Khi tối ưu hóa trên "phần thưởng" quan điểm, AdaRank-E và AdaRank-R hoạt động tốt hơn AdaRank-O. Những các mô hình đặt các nguồn tài nguyên phản đối ở vị trí thứ 10 theo MRRBad, trong khi xếp hạng những thứ lý tưởng ở vị trí thứ 5, theo MRR (Hàng 1-3 trong Bảng 4.3). Điều này chỉ ra rằng các mô hình này đang học cách tập trung vào các loại tài nguyên phù hợp với nhóm người dùng và môi trường của chúng tôi. Khi đưa tất cả các tính năng kết hợp lại với nhau, AdaRank vượt trội hơn từng biến thể riêng lẻ, cho thấy giá trị của mỗi quan điểm trong việc xác định mức độ liên quan.

Cho đến nay, chúng tôi đã trình bày rằng các lựa chọn thiết kế để xem xét rủi ro và phần thưởng quan điểm trong nhiệm vụ xếp hạng lại là có cơ sở. Tuy nhiên, chúng tôi cho rằng Các mô hình AdaRank đang học cách xếp hạng các tài nguyên phản cảm thấp hơn thành một tài nguyên có lợi tác dụng phụ của việc tối ưu hóa sự liên kết giáo dục và khả năng đọc. Để tính đến đáng phản đối như một tín hiệu rõ ràng về chi phí, và để cân bằng rủi ro đó với phần thưởng trong số các góc nhìn khác, chúng ta chuyển sang REORank, được tối ưu hóa cho nCS-DCG.

Đối với REORank-E, REORank-R và REORank-O, chúng tôi thấy hiệu suất tương tự như của các đối tác AdaRank của họ (lần lượt là Hàng 5-7 và 2-4 trong Bảng 4.3).

Bảng 4.3: Hiệu suất của REdORank và các biến thể cắt bỏ sử dụng RankSet.
hậu tố -R, -E, -O chỉ ra Chỉ có tính dễ đọc, Chỉ có tính giáo dục và Có thể phản đối
chỉ, tương ứng. -M chỉ ra việc sử dụng máy trộn để căn chỉnh giáo dục và
khả năng đọc và -MER chỉ ra việc sử dụng bộ trộn với -E và -R. * chỉ ra
ý nghĩa liên quan đến REdORank và in đậm cho biết hiệu suất tốt nhất cho mỗi số liệu.

Thuật toán hàng	Tối ưu hóa Hệ mét	NDCG	MRR	MRRBhại
Thứ hạng Ada	NDCG 1	0,778*	0,226*	0,097*
2 AdaRank-E	NDCG	0,765*	0,209	0,110*
3 AdaRank-R	NDCG	0,774*	0,222	0,101*
4 AdaRank-O	NDCG	0,675*	0,148*	0,537*
5 REdORank-E nCS-DCG		0,765*	0,209	0,110*
6 REdORank-R nCS-DCG		0,774*	0,222	0,101*
7 REdORank-O nCS-DCG		0,675*	0,148*	0,537*
8 REdORank-M nCS-DCG		0,765*	0,209	0,110*
9 REdORank-MER nCS-DCG		0,777	0,218	0,089*
10 REdORank nCS-DCG		0,779	0,228	0,097

Điều này càng làm nổi bật thêm rằng các quan điểm là quan trọng. Chúng tôi đưa ra giả thuyết rằng sự giao thoa
sự kết hợp giữa sự liên kết giáo dục và khả năng đọc hiểu sẽ đóng vai trò là một sự kết hợp có lợi
tín hiệu cho sự liên quan của một nguồn tài nguyên. Vì lý do này, chúng tôi sử dụng bộ trộn được mô tả
trong Phần 3.2.2 để kết hợp hai góc nhìn. Thật ngạc nhiên, REdORank-M
có hình thức tệ hơn về mọi mặt khi so sánh với REdORank-R và thực hiện giống nhau
như REdORank-E. Để điều tra đầy đủ liệu quan điểm kết hợp này có thể cung cấp
giá trị cho việc xếp hạng lại, chúng tôi đã tạo ra REdORank-MER. Cho vay sự tin cậy cho ý tưởng
của việc kết hợp một quan điểm kết hợp, REdORank-MER đã vượt trội hơn từng
các biến thể quan điểm cá nhân. Trong khi biến thể này thực hiện tốt hơn đáng kể
so với REdORank về MRRBad, nó lại hoạt động kém hơn ở hai số liệu còn lại.
Điều này nhấn mạnh rằng việc xem xét rõ ràng yếu tố chi phí nhạy cảm, cùng với
nhiều quan điểm có liên quan, có tác động có lợi đến việc xếp hạng lại các nguồn lực cho
trẻ em đang tìm kiếm trong lớp học.

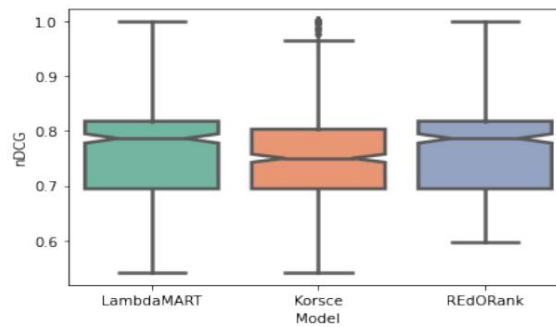
Bảng 4.4: Hiệu suất của REORank và các đường cơ sở sử dụng RankSet. * biểu thị ý nghĩa wrtREORank và in đậm cho biết hiệu suất tốt nhất cho mỗi số liệu.

Thuật toán	Tối ưu hóa Hệ mét	NDCG	MRR	MRRBhại
LambdaMART	NDCG	0,784	0,228	0,081*
Korsce	Không	0,753*	0,209	0,163*
Xếp hạng REORank	có nCS-DCG	0,779	0,228	0,097

Kết quả cho đến nay đã chứng minh rằng thiết kế cho REORank là có cơ sở.

Để hiểu rõ hơn về cách REORank hoạt động, chúng tôi cũng so sánh nó cho cả đối tác hiện đại, Korsce, và thuật toán LTR cơ bản, trong LambdaMART. Kết quả của hai mô hình này xếp hạng các tài nguyên trong RankSet có thể được thấy trong Bảng 4.4. Chúng ta thấy rằng REORank hoạt động tốt hơn đáng kể so với Korsce cho tất cả các số liệu. Điều này được thể hiện trực quan trong Hình 4.3. Chúng tôi gần sự khác biệt về hiệu suất so với thực tế là Korsce xếp hạng theo điểm, có trọng số cách khách quan. Nghĩa là, đối với mỗi nguồn lực, mỗi điểm quan điểm được nhân lên bằng một trọng số được xác định theo kinh nghiệm, sau đó cộng lại với nhau để tạo ra thứ hạng điểm. Ngược lại, REORank học một trọng số động duy nhất tính đến từng quan điểm đồng thời trái ngược với cá nhân. LambdaMART học cách xếp hạng bằng cách tối ưu hóa trên các so sánh từng cặp tài liệu. Đáng ngạc nhiên là LambdaMART thực hiện tốt hơn đáng kể so với REORank cho RankSet. Trong khi điều này là không mong đợi, vì các thuật toán LTR theo danh sách đã được chứng minh là hiệu quả hơn khi được áp dụng cho tìm kiếm trên Web [20, 118], chúng tôi quy sự khác biệt về hiệu suất cho cấu trúc của tập dữ liệu. RankSet chỉ chứa một tài nguyên lý tưởng duy nhất, một thuật toán từng cặp có nhiều khả năng "xác định vị trí" theo bản chất của việc so sánh trực tiếp tài liệu. Mặt khác, REORank có nhiều khả năng bỏ lỡ tài nguyên lý tưởng vì nó không so sánh rõ ràng từng nguồn tài nguyên với mọi nguồn tài nguyên khác, mà thay vào đó xem xét

sự liên quan của chúng theo cách tương đối trong danh sách. Trong các tình huống thực tế, nơi có nhiều hơn một nguồn tài nguyên lý tưởng có thể nằm trong một danh sách duy nhất, một cách tiếp cận theo danh sách là phù hợp hơn với nhiệm vụ xếp hạng lại.



Hình 4.3: NDCG@10 cho các mô hình xếp hạng lại khác nhau sử dụng RankSet.

Việc điều chỉnh mô hình LTR có tính đến nhiều yếu tố không?
 quan điểm dẫn đến việc ưu tiên các nguồn lực có liên quan đến cả hai
 trẻ em và bối cảnh lớp học? (RQ4) Với giới hạn dưới cao hơn rõ ràng của nó
 trên NDCG@10 so với các đối tác của nó (như được thấy trong Hình 4.3), hiệu suất thành công của nó
 về việc xếp hạng các nguồn tài nguyên giáo dục và dễ đọc được biết đến ở vị trí cao trong bảng xếp hạng và
 khả năng khái quát hóa dự kiến của nó đối với các kịch bản xếp hạng lại trong thế giới thực, chúng tôi xem xét thiết kế
 của REDORank là một mô hình thích hợp để cung cấp xếp hạng lại cho các hệ thống tìm kiếm
 hỗ trợ hoạt động tìm kiếm của trẻ em trong lớp học.

CHƯƠNG 5

PHẦN KẾT LUẬN

Công trình trình bày trong luận án này thúc đẩy việc thu thập thông tin cho trẻ em tập trung về thiết kế, phát triển và đánh giá các chiến lược cho phép trẻ em trực tuyến khám phá thông tin. Lĩnh vực nghiên cứu này có tầm quan trọng đặc biệt do sự phổ biến của các công cụ tìm kiếm (SE) và thực tế là trẻ em được coi là người dùng không theo truyền thống, vì phần lớn các nỗ lực học thuật và thương mại trong khu vực này nhắm mục tiêu đến người lớn. Với phạm vi rộng lớn các kỹ năng tìm kiếm của trẻ em khi nó liên quan đến việc tìm kiếm [50], và phạm vi rộng lớn các câu hỏi mà trẻ em tìm đến SE, đặc biệt là với đại dịch COVID-19 đang diễn ra, chúng tôi đã xác định rõ phạm vi công việc của mình tập trung vào trẻ em từ 6 đến 11 tuổi bằng cách sử dụng các công cụ tìm kiếm trên Web trong bối cảnh lớp học. Để hỗ trợ tốt hơn cho nhóm người dùng này, chúng tôi đã trình bày một chiến lược xếp hạng lại mới, REORank, đóng vai trò như bánh xe tập luyện để giúp trẻ em xác định các nguồn lực có liên quan đến nhu cầu thông tin của họ.

Đáp lại những phát hiện được báo cáo trong tài liệu liên quan đến cách thức mà trẻ em tìm kiếm, ví dụ, khuynh hướng tập trung vào các kết quả được xếp hạng cao nhất của chúng [50, 53], cũng như cách thức mà SE phản hồi các câu hỏi của trẻ em [13, 12], ví dụ, cung cấp tài nguyên trẻ em không thể hiểu được, REORank kiểm tra các nguồn tài nguyên được thu thập bởi thương mại SE-được trẻ em ưa chuộng-và ưu tiên chúng theo cách phù hợp nhất cho bối cảnh và nhóm người dùng hiện tại được xếp hạng cao hơn. Để làm như vậy, REORank

đánh giá các nguồn lực dựa trên ba góc nhìn: sự liên kết về giáo dục, khả năng đọc, và khả năng phản đối. Bằng cách kết hợp tất cả các quan điểm này, REdORank đảm bảo rằng SE có thể phản ứng tốt hơn với hành vi tìm kiếm của trẻ em bằng cách cân bằng rủi ro và giá trị phần thưởng của nội dung tài nguyên.

Với sự liên kết giáo dục, mục tiêu là ưu tiên các nguồn lực hỗ trợ việc học tập. ing. Ngay cả khi các nguồn tài nguyên mang tính giáo dục, để nội dung của chúng có giá trị, thì chúng phải có thể đọc được và do đó dễ hiểu. Để đạt được mục đích đó, REdORank xem xét khả năng đọc của các nguồn tài nguyên. Thật không may, không phải tất cả các nguồn tài nguyên trực tuyến đều dành cho, hoặc thậm chí phù hợp cho trẻ em tham gia vào bối cảnh lớp học. Như vậy, REdORank kết hợp quan điểm phản đối để có thể định vị các nguồn lực đó thấp hơn trong bảng xếp hạng mà trẻ em ít có khả năng nhìn thấy và tương tác với chúng.

Một phân tích sâu sắc về REdORank cho thấy rằng một mô hình LTR đa góc nhìn là một giải pháp hiệu quả để xếp hạng lại các nguồn lực cho trẻ em trong lớp học. Thông qua các thí nghiệm được tiến hành, chúng ta có thể kết luận rằng việc đưa vào các quan điểm một cách có chủ đích được kết nối với một nhóm người dùng và môi trường cụ thể có thể mang lại lợi ích cho hiệu suất của một mô hình trong việc xếp hạng lại các nguồn tài nguyên được lấy từ một SE chính thống. Trên thực tế, xem xét rõ ràng các tín hiệu của quan điểm “tiêu cực” và cân bằng chúng với Các tín hiệu liên quan “tích cực” mang lại sự cải thiện đáng kể về hiệu suất. Một phần không thể thiếu trong khả năng xếp hạng lại tài nguyên của REdORank, chúng tôi đã phát hiện ra rằng URL và các đoạn trích cung cấp một proxy hiệu quả cho việc phân loại một tài nguyên trực tuyến như giáo dục, hoặc không. Dựa trên nghiên cứu cắt bỏ được tiến hành với REdORank, chúng tôi cũng thấy rằng các đoạn trích hữu ích trong việc xác định các tài nguyên có thể gây phản đối. Cuối cùng, xét về mặt khả năng đọc, chúng tôi nhận thấy rằng khi xem xét một vốn từ vựng được thiết kế cẩn thận hơn theo ngôn ngữ đương đại và phù hợp với độ tuổi của trẻ em việc thu thập các thuật ngữ có thể cải thiện ước tính khả năng đọc cho các nguồn tài nguyên trực tuyến được nhắm mục tiêu

hướng tới trẻ em.

Là kết quả của một cuộc khám phá thực nghiệm sâu rộng để xác định tính hợp lệ của từng các quan điểm thông báo thiết kế của REdORank, các đóng góp bổ sung sau đây đã xuất hiện: BiGBERT, một mô hình phân loại để xác định các trang web được liên kết với giáo dục nguồn thông qua phương pháp học sâu tổng hợp; và Judgebad, một phương pháp dựa trên từ điển mô hình phân loại để xác định các nguồn tài nguyên trực tuyến có thể gây phản đối cho trẻ em trong bối cảnh lớp học. Ngoài các mô hình phân loại này, chúng tôi giới thiệu một công thức để đọc có hiệu quả trong việc ước tính mức độ dễ đọc của các tài nguyên trực tuyến.

Trong suốt quá trình làm việc của chúng tôi, một số hạn chế và con đường cho lòng thú nghiên cứu của họ đã được đưa ra ánh sáng. Spache-Allen được thiết kế và đánh giá chỉ dựa trên khả năng áp dụng để ước tính khả năng đọc của các nguồn tài liệu tiếng Anh. Tuy nhiên, Web có phạm vi toàn cầu và bao phủ nhiều lĩnh vực thông tin khác nhau. Do đó, thực hiện một cuộc khám phá thực nghiệm tương tự liên quan đến các lĩnh vực khác nhau, ví dụ, pháp lý hoặc y tế, cũng như các công thức khả năng đọc đa ngôn ngữ có thể cung cấp thêm giá trị hiểu biết sâu sắc cho nhiều lĩnh vực nghiên cứu quan tâm đến việc hiểu văn bản, đặc biệt là Xử lý ngôn ngữ tự nhiên (NLP). REdORank tận dụng khả năng đọc như một tính năng, dựa trên Spache-Allen, chỉ xem xét các tài nguyên văn bản để ước tính khả năng đọc. Trong tương lai, chúng tôi có kế hoạch mở rộng quan điểm này để xem xét các phương pháp ước tính khả năng đọc có tính đến sự hiện diện của phương tiện truyền thông bổ sung các yếu tố, ví dụ, hình ảnh và biểu đồ trên các trang web. Một hạn chế khác là thiếu xem xét kiến thức trước đó của người dùng về một chủ đề. Công việc trong tương lai điều tra kết nối giữa kiến thức chủ đề có sẵn và ước tính khả năng đọc có thể thu hẹp khoảng cách này và liên kết chặt chẽ hơn các công cụ hỗ trợ như REdORank với mục tiêu của họ nhóm người dùng.

Khi khám phá các nguồn tài nguyên đáng ngờ, chúng tôi đã tuân theo các ứng dụng tiên tiến hiện có.

tiếp cận và xử lý tất cả các danh mục trong ObjCat như là không thể chối cãi. Làm thế nào-

bao giờ, trẻ em không nhất thiết phải cần một giải pháp phù hợp với tất cả. Ví dụ,

nội dung có thể phản cảm với học sinh lớp 4 nhưng lại không phản cảm với học sinh lớp 12 .

Chúng tôi nhận ra đây là một hạn chế của Judgebad và đề xuất tăng mức độ chi tiết

để xác định điều gì là phản cảm và điều gì không phản cảm đối với trẻ em ở nhiều độ tuổi khác nhau.

Những hạn chế bổ sung liên quan đến RankSet, tập dữ liệu được sử dụng để đánh giá các sản phẩm

mô hình xếp hạng lại được đưa ra, được xây dựng dựa trên mô hình Cranfield. Như

do đó, tập dữ liệu chỉ giới hạn ở một nguồn tài nguyên lý tưởng duy nhất để đáp ứng mỗi truy vấn.

Thật không may, việc chỉ có một nguồn tài nguyên lý tưởng duy nhất không phải là dấu hiệu của thế giới thực.

SERP, do đó dẫn chúng tôi đến việc tiếp tục nghiên cứu sâu hơn về hiệu suất của REORank

trong một môi trường thực tế. Với điều đó, các bước tiếp theo ngay lập tức bao gồm một nghiên cứu người dùng

bao gồm việc kiểm tra hành vi tìm kiếm của trẻ em khi sử dụng hệ thống tìm kiếm

có và không có REORank.

Kết quả từ luận án này có ý nghĩa đối với các nhà nghiên cứu đang điều tra trẻ em

Tìm kiếm trên web. REORank là một bước tiến tới việc áp dụng SE chính thống vào lớp học

sử dụng, nhưng nó vẫn chỉ tập trung vào các quan điểm cụ thể để thông báo mức tăng liên quan. Nó là

đáng để nghiên cứu những lợi ích của việc kết hợp các tín hiệu liên quan bổ sung, từ các nguồn

ngoài văn bản của một nguồn tài nguyên, chẳng hạn như nguồn tài nguyên đó đến từ đâu hoặc ai đã viết

nó. Những yếu tố như vậy góp phần vào độ tin cậy của một nguồn tài nguyên. Thật không may, trẻ em

được biết là không đánh giá độ tin cậy của các nguồn trực tuyến [55], khiến độ tin cậy trở thành

góc nhìn bổ sung có giá trị để đưa vào nhóm xếp hạng lại. Điều này có thể đạt được

nhanh chóng và hiệu quả khi sử dụng chiến lược trộn được REORank sử dụng (Phần 3.2.2)

cho phép tổng hợp đồng thời nhiều điểm thành một điểm duy nhất.

Trong khi máy trộn này hiện đang được sử dụng cho các quan điểm "phần thưởng", nó có thể được sao chép cho

cũng như quan điểm "rủi ro". Ví dụ, trẻ em gặp khó khăn trong việc xác định xem có điều gì đó

thông tin sai lệch hay không [99]. Do đó, họ có thể xem xét một nguồn thông tin sai lệch như đáng tin cậy. Bằng cách mở rộng chi phí để bao gồm các quan điểm vượt ra ngoài sự phản đối, chẳng hạn như thông tin sai lệch, REdORank có thể ưu tiên các nguồn lực có căn cứ thực tế.

Nghiên cứu đang diễn ra về Tương tác giữa Người và Máy tính đã khám phá cách thức hình ảnh các tác động của SERP ảnh hưởng đến tương tác của trẻ em với kết quả [3, 4]. REdORank có thể cung cấp thêm những hướng khám phá về cách xác định loại tài nguyên và các yếu tố trực quan có thể đóng vai trò là manh mối trực quan. Ví dụ, một biểu tượng cuốn sách nhỏ với một con số bên trong có thể được thêm vào cùng với màn hình hiển thị kết quả trên một phương pháp truyền thống SERP để chỉ ra mức độ đọc của một nguồn tài nguyên. Tương tự như vậy, một biểu tượng đại diện cho một có thể thêm một ngôi trường nhỏ để chỉ ra rằng nguồn tài nguyên này có giá trị giáo dục.

Thông qua việc bổ sung các yếu tố trực quan bắt nguồn từ các đặc điểm được xem xét trong quá trình xếp hạng, tính minh bạch của hệ thống tìm kiếm có thể tăng lên. Người dùng có thể được cung cấp cái nhìn thoáng qua đằng sau bức màn về cách một hệ thống cụ thể hoạt động. Điều này có thể tác động đến dễ sử dụng và dễ hiểu của một hệ thống. Ngoài ra, các yếu tố trực quan như vậy có thể mang lại lợi ích cho người dùng đang học cách tìm kiếm bằng cách cung cấp, theo thời gian, một kết nối hữu hình giữa xếp hạng các nguồn tài nguyên và truy vấn mà người tìm kiếm có thể sử dụng để cải thiện việc xây dựng truy vấn.

REdORank là một bước tiến thực tế hướng tới việc thiết kế một công cụ tìm kiếm thích ứng dành cho trẻ em. Có lẽ tác động lớn nhất là khả năng sử dụng REdORank để hỗ trợ tìm kiếm để học một phần của tìm kiếm như mô hình học tập. Tìm kiếm để học là hành động tìm kiếm thông tin để có được kiến thức mới trong bối cảnh giáo dục [13, 108], điều này rất phù hợp với mục đích của REdORank vì tính hiệu quả của xác định và có xu hướng xếp hạng cao hơn các nguồn tài nguyên web có giá trị giáo dục.

TÀI LIỆU THAM KHẢO

- [1] Ouessai Abdessamed và Elberrichi Zakaria. Phân loại trang web dựa trên url và nội dung: Vụ kiện Algeria so với vụ kiện không phải Algeria. Năm 2015, Quốc tế lần thứ 12 Hội thảo về Lập trình và Hệ thống (ISPS), trang 1-8. IEEE, 2015.
- [2] Claudia Margarita Acu~na-Soto, Vicente Liern và Blanca P'erez-Gladish. Một Phương pháp dựa trên vikor để xếp hạng các video hướng dẫn toán học. Quyết định của Ban quản lý, 2019.
- [3] Mohammad Aliannejadi, Monica Landoni, Theo Huibers, Emiliana Murgia, và Maria Soledad Pera. Quan điểm của trẻ em về cách biểu tượng cảm xúc giúp chúng nhận ra kết quả có liên quan: Hành động có nói to hơn lời nói không? Trong Biên bản của Hội nghị năm 2021 về Tương tác và Truy xuất thông tin của con người, các trang 301-305, 2021.
- [4] Garrett Allen, Benjamin L Peterson, Dhanush Kumar Ratakonda, Mostofa Najmus Sakib, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, và Maria Soledad Pera. Engage!: Đồng thiết kế các trang kết quả của công cụ tìm kiếm để thúc đẩy tương tác. Trong Thiết kế tương tác và trẻ em, trang 583-587, 2021.
- [5] Liên minh học tập kỹ thuật số Idaho. Cổng thông tin học tập kỹ thuật số Idaho. <https://portal.idiglearning.net/K12/CourseCatalog>, 2021. (truy cập ngày 19 tháng 1, 2021).

- [6] Inc Amazon. Alexa các trang web hàng đầu. <https://www.alexa.com/topsites/category>, 2020. (truy cập ngày 17 tháng 9 năm 2020).
- [7] Steven J Amendum, Kristin Conradi và Meghan D Liebfreund. sự đẩy đổi với các văn bản khó hơn: Phân tích tốc độ, độ chính xác và khả năng đọc của người mới bắt đầu hiểu biết. *Tâm lý đọc*, 37(4):570-600, 2016.
- [8] Steven J Amendum, Kristin Conradi, và Elfrieda Hiebert. Văn bản có phức tạp không-vấn đề quan trọng trong các lớp tiểu học? tổng hợp nghiên cứu về độ khó của văn bản và khả năng đọc trôi chảy và hiểu của học sinh tiểu học. *Giáo dục Tạp chí tâm lý học*, 30(1):121-151, 2018.
- [9] Jonathan Anderson. Lix và rix: Các biến thể của chỉ số khả năng đọc ít được biết đến. *Tạp chí Đọc*, 26(6):490-496, 1983.
- [10] H'elder Antunes và Carla Teixeira Lopes. Khả năng đọc nội dung web. Năm 2019 Hội nghị lần thứ 14 của Iberia về Hệ thống thông tin và Công nghệ (CISTI), trang 1-4. IEEE, 2019.
- [11] Oghenemaro Anuyah, Ion Madrazo Azpiazu và Maria Soledad Pera. sử dụng kiến thức có cấu trúc và những từ ngữ truyền thống để tạo ra khái niệm đại diện trong lĩnh vực giáo dục. Trong *Biên bản đồng hành của Hội nghị World Wide Web 2019*, trang 274-282, 2019.
- [12] Oghenemaro Anuyah, Ashlee Milton, Michael Green và Maria Soledad Pera. Một phân tích thực nghiệm về phản ứng của công cụ tìm kiếm đối với các truy vấn tìm kiếm trên web liên quan đến bối cảnh lớp học. *Tạp chí Quản lý thông tin Aslib*, 2019.

- [13] Ion Madrazo Azpiazu, Nevena Dragovic, Maria Soledad Pera và Jerry Alan
Thất bại. Tìm kiếm và học trực tuyến: Yum và các công cụ tìm kiếm khác dành cho trẻ em
và giáo viên. Tạp chí Truy xuất thông tin, 20(5):524-545, 2017.
- [14] Rebekah George Benjamin. Tái thiết khả năng đọc: Những phát triển gần đây
và các khuyến nghị trong việc phân tích độ khó của văn bản. Tâm lý giáo dục
Đánh giá, 24(1):63-88, 2012.
- [15] Dania Bilal. So sánh khả năng đọc kết quả tìm kiếm của google với khả năng đọc flesch
công thức khả năng: phân tích sơ bộ về các truy vấn tìm kiếm của trẻ em. Biên bản
của Hiệp hội Khoa học và Công nghệ Thông tin Hoa Kỳ, 50(1):1-9,
2013.
- [16] Dania Bilal và Meredith Boehm. Hướng tới các phương pháp mới để đánh giá
sự liên quan của việc tìm kiếm thông tin từ các công cụ tìm kiếm trên web đối với các truy vấn của trẻ em.
Phương pháp định lượng và định tính trong thư viện, 2(1):93-100, 2017.
- [17] Steven Bird, Ewan Klein và Edward Loper. Xử lý ngôn ngữ tự nhiên với
Python: phân tích văn bản bằng bộ công cụ ngôn ngữ tự nhiên. O'Reilly Media, Inc.,
2009.
- [18] Sebastian Bruch, Shuguang Han, Michael Bendersky và Marc Najork. A
xử lý ngẫu nhiên việc học để xếp hạng các hàm tính điểm. Trong Biên bản
Hội nghị quốc tế lần thứ 13 về tìm kiếm trên web và khai thác dữ liệu, trang 61-69,
2020.
- [19] Christopher JC Burges. Từ ranknet đến lambdarank đến lambdamart: Một
tổng quan. Học tập, 11(23-581):81, 2010.

- [20] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai và Hang Li. Học cách xếp hạng: từ cách tiếp cận từng cặp đến cách tiếp cận từng danh sách. Trong Biên bản báo cáo của lần thứ 24 hội nghị quốc tế về Học máy, trang 129-136, 2007.
- [21] David Carmel, Elad Haramaty, Arnon Lazerson và Liane Lewin-Eytan. Tối ưu hóa xếp hạng đa mục tiêu cho tìm kiếm sản phẩm bằng cách sử dụng nhãn ngẫu nhiên tổng hợp. Trong Biên bản Hội nghị Web 2020, trang 373-383, 2020.
- [22] Jeanne Sternlicht Chall và Edgar Dale. Khả năng đọc được xem xét lại: Dale- mới Công thức khả năng đọc Chall. Brookline Books, 1995.
- [23] Wanyu Chen, Fei Cai, Honghui Chen, và Maarten De Rijke. Cá nhân hóa truy vấn gợi ý đa dạng hóa trong việc truy xuất thông tin. Biên giới của máy tính Khoa học, 14(3):143602, 2020.
- [24] Benjamin Clavié và Kobi Gal. Edubert: Các mô hình ngôn ngữ sâu được đào tạo trước cho phân tích học tập. Bản in trước arXiv arXiv:1912.00690, 2019.
- [25] Meri Coleman và Ta Lin Liau. Một công thức khả năng đọc máy tính được thiết kế cho máy chấm điểm. Tạp chí Tâm lý học Ứng dụng, 60(2):283, 1975.
- [26] Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian De La Chica và David Sontag. Cá nhân hóa kết quả tìm kiếm trên web bằng cách đọc cấp độ. Trong Biên bản báo cáo của hội nghị quốc tế ACM lần thứ 20 về Thông tin và quản lý kiến thức, trang 403-412, 2011.
- [27] W Bruce Croft, Donald Metzler và Trevor Strohman. Công cụ tìm kiếm: Thông tin thu thập thông tin trong thực tế, tập 520. Addison-Wesley Reading, 2010.

- [28] Washington Cunha, S'ergio Canuto, Felipe Viegas, Thiago Salles, Christian Gomes, Vitor Mangaravite, Elaine Resende, Thiersen Rosa, Marcos Andr e Goncalves và Leonardo Rocha. Đường ống xử lý trước mở rộng cho văn bản phân loại: Về vai trò của biểu diễn siêu tính năng, sự thừa thớt và lấy mẫu có chọn lọc. Xử lý và quản lý thông tin, 57(4):102263, 2020.
- [29] Washington Cunha, V'itor Mangaravite, Christian Gomes, S'ergio Canuto, Elaine Resende, Cecilia Nascimento, Felipe Viegas, Celso Franc,ca, Welling-t n Santos Martins, Jussara M Almeida, và những người khác. Về hiệu quả chi phí của phương pháp tiếp cận và biểu diễn thần kinh và phi thần kinh để phân loại văn bản: A nghiên cứu so sánh toàn diện. Xử lý thông tin & Quản lý, 58 (3):102481, 2021.
- [30] Xinyi Dai, Jiawei Hou, Qing Liu, Yunjia Xi, Ruiming Tang, Weinan Zhang, Xiuqiang He, Jun Wang và Yong Yu. Hạng U: Học tập hướng đến tiện ích để xếp hạng với phản hồi ngầm. Trong Biên bản của Hội nghị quốc tế ACM lần thứ 29 Hội nghị về Quản lý thông tin và kiến thức, trang 2373-2380, 2020.
- [31] Edgar Dale và Jeanne S Chall. Một công thức để dự đoán khả năng đọc: Hướng dẫn Bản tin nghiên cứu giáo dục, trang 37-54, 1948.
- [32] Judith H Danovitch. Lớn lên cùng Google: Sự hiểu biết của trẻ em và việc sử dụng các thiết bị dựa trên internet liên quan đến sự phát triển nhận thức. Con người Hành vi và công nghệ mới nổi, 1(2):81-90, 2019.
- [33] Jacob Devlin, Ming-Wei Chang, Kenton Lee và Kristina Toutanova. Bert: Đào tạo trước các bộ biến đổi song hướng sâu để hiểu ngôn ngữ. Bản in trước của arXiv arXiv:1810.04805, 2018.

[34] Yajuan Duan, Long Jiang, Tao Qin, Ming Chu, và Heung Yeung Shum.

Một nghiên cứu thực nghiệm về việc học cách xếp hạng các tweet. Trong Biên bản báo cáo của lần thứ 23 Hội nghị quốc tế về ngôn ngữ học tính toán (Coling 2010), trang 295-303, 2010.

[35] Sergio Duarte Torres và Ingmar Weber. Trẻ em tìm kiếm cái gì và như thế nào trên web. Trong Biên bản báo cáo của hội nghị quốc tế ACM lần thứ 20 về Thông tin và quản lý kiến thức, trang 393-402, 2011.

[36] William H DuBay. Ngôn ngữ thông minh: Người đọc, khả năng đọc và việc chấm điểm Văn bản. ERIC, 2007.

[37] Benjamin Edelman. Phân tích thực nghiệm về tìm kiếm an toàn của google. Trung tâm Berkman về Internet và Xã hội, Trường Luật Harvard, 14, 2003.

[38] Carsten Eickhoff, Pavel Serdyukov, và Arjen P de Vries. Lớp trang web-sự phù hợp của trẻ em. Trong Biên bản của Hội nghị quốc tế ACM lần thứ 19 Hội nghị về Quản lý thông tin và kiến thức, trang 1425-1428, 2010.

[39] Michael D Ekstrand, Katherine Landau Wright và Maria Soledad Pera. Nâng cao hướng dẫn trong lớp học với tin tức trực tuyến. Tạp chí thông tin Aslib Quản lý, 2020.

[40] Ahmed Elnaggar, Christoph Gebendorfer, Ingo Glaser và Florian Matthes. Học sâu đa nhiệm vụ để dịch, tóm tắt và biên dịch tài liệu pháp lý phân loại đa nhãn. Trong Biên bản báo cáo của Hội nghị Trí tuệ nhân tạo và Hội nghị điện toán đám mây, trang 9-15, 2018.

- [41] Vladimir Estivill-Castro và Alessandro Marani. Hướng tới việc xếp hạng các trang web trang cho mục đích giáo dục. Trong CSEDU (1), trang 47-54, 2019.
- [42] Vanessa Figueiredo và Eric M Meyers. Sự đánh đổi sai lầm giữa tính liên quan và sự an toàn trong hệ thống tìm kiếm của trẻ em. Biên bản của Hiệp hội Thông tin Khoa học và Công nghệ, 56(1):651-653, 2019.
- [43] Elizabeth Foss, Allison Druin, Robin Brewer, Phillip Lo, Luis Sanchez, Evan Golub và Hilary Hutchinson. Vai trò tìm kiếm của trẻ em tại nhà: Ý nghĩa dành cho các nhà thiết kế, nhà nghiên cứu, nhà giáo dục và phụ huynh. Tạp chí của Hoa Kỳ Hội Khoa học và Công nghệ Thông tin, 63(3):558-573, 2012.
- [44] W Nelson Francis và Henry Kucera. Sổ tay ngữ liệu Brown. Thư gửi Biên tập, 5(2):7, 1979.
- [45] Thomas Francoys và Eleni Miltsakaki. NLP và máy học có cải thiện không công thức khả năng đọc truyền thống? Trong Biên bản của Hội thảo đầu tiên về Dự đoán và cải thiện khả năng đọc văn bản cho nhóm đối tượng đọc giả mục tiêu, các trang 49-57, 2012.
- [46] Jerome H Friedman. Xấp xỉ hàm tham lam: một phép tăng cường độ dốc máy móc. Biên niên sử thống kê, trang 1189-1232, 2001.
- [47] Wolfe Garbe. SymSpell. <https://github.com/wolfgarbe/SymSpell>, 2020.
- [48] N Geetha và PT Vanathi. Chuyển giao kiến thức cho miền chéo hiệu quả xếp hạng sử dụng thuật toán adarank. Tạp chí quốc tế về trí tuệ kinh doanh và Khai thác dữ liệu, 14(1-2):89-105, 2019.

- [49] Filippo Geraci và Tiziano Papini. Xấp xỉ phân loại văn bản nhiều lớp thông qua việc tự động tạo ra các ví dụ đào tạo. Trong Hội nghị quốc tế về Ngôn ngữ học tính toán và Xử lý văn bản thông minh, trang 585-601. Nhà xuất bản Springer, 2017.
- [50] Tatiana Gossen và Andreas Nürnberger. Chi tiết về việc truy xuất thông tin cho người dùng trẻ: Một cuộc khảo sát. Xử lý & Quản lý thông tin, 49(4):739-756, 2013.
- [51] Kelsey Leonard Grabeel, Jennifer Russomanno, Sandy Oelschlegel, Emily Tester, và Robert Eric HeideI. Sức khỏe được tính toán bằng máy tính so với sức khỏe được tính bằng tay công cụ học tập: so sánh phép đo đơn giản về tiếng lóng (khói bụi) và flesch-kincaid trong tài liệu giáo dục bệnh nhân được in. Tạp chí Y khoa Hiệp hội Thư viện: JMLA, 106(1):38, 2018.
- [52] Jiafeng Guo, Yixing Fan, Liang Pang, Liu Yang, Qingyao Ai, Hamed Zamani, Chen Wu, W Bruce Croft và Xueqi Cheng. Một cái nhìn sâu sắc vào xếp hạng thần kinh mô hình để truy xuất thông tin. Xử lý & Quản lý thông tin, 57(6):102067, 2020.
- [53] Jacek Gwizdka và Dania Bilal. Phân tích các truy vấn của trẻ em và hành vi nhấp chuột về kết quả được xếp hạng và quá trình suy nghĩ của họ trong tìm kiếm của Google. Trong Biên bản hội nghị năm 2017 về tương tác và truy xuất thông tin của con người, trang 377-380, 2017.
- [54] Karl Gyllstrom và Marie-Francine Moens. Trí tuệ của thời đại: hướng tới cung cấp trang web cho trẻ em với thuật toán agerank dựa trên liên kết. Trong Tiến trình-

của hội nghị quốc tế ACM lần thứ 19 về Thông tin và tri thức
quản lý, trang 159-168, 2010.

[55] Elina KH`am`al`ainen, Carita Kiili, Miika Marttunen, Eija R`aikk`onen, Roberto
Gonz`alez-Ib`a`nez, và Paavo HT Lepp`anen. Thúc đẩy uy tín của học sinh lớp sáu
đánh giá các trang web: một nghiên cứu can thiệp. Máy tính trong hành vi của con người,
110:106372, 2020.

[56] Mahdi Hashemi. Phân loại trang web: khảo sát các quan điểm, khoảng cách và
hướng đi trong tương lai. Công cụ và ứng dụng đa phương tiện, trang 1-25, 2020.

[57] Samer Hassan và Rada Mihalcea. Học cách xác định tài liệu giáo dục.
Giao dịch ACM về Xử lý Ngôn ngữ và Giọng nói (TSLP), 8(2):1-18,
2008.

[58] William Hersh, Chris Buckley, TJ Leone và David Hickam. Ohsumed: Một
đánh giá truy xuất tương tác và bộ sưu tập thử nghiệm lớn mới cho nghiên cứu. Trong
SIGIR'94, trang 192-201. Springer, 1994.

[59] Anett Hoppe, Peter Holtz, Yvonne Kammerer, Ran Yu, Stefan Dietze, và
Ralph Ewerth. Những thách thức hiện tại trong việc nghiên cứu tìm kiếm như một quá trình học tập.
Biên bản báo cáo Học tập và Giáo dục với Dữ liệu Web, Amsterdam, Hà Lan
đất đai, 2018.

[60] Mark Hughes, Irene Li, Spyros Kotoulas và Toyotaro Suzumura. văn bản y tế
phân loại sử dụng mạng nơ-ron tích chập. Nghiên cứu về công nghệ y tế
và Tin học, 235:246-50, 2017.

- [61] Clayton Hutto và Eric Gilbert. Vader: Một mô hình dựa trên quy tắc tiết kiệm cho phân tích tình cảm của văn bản phương tiện truyền thông xã hội. Trong Biên bản báo cáo quốc tế Hội nghị AAAI về Web và Truyền thông xã hội, tập 8, 2014.
- [62] Sáng kiến Tiêu chuẩn Nhà nước cốt lõi chung. Phụ lục b: Ví dụ văn bản và các nhiệm vụ thực hiện mẫu, 2020. URL http://www.corestandards.org/tài_sản/Phụ_lục_B.pdf.
- [63] Sáng kiến Tiêu chuẩn Nhà nước cốt lõi chung. Tiêu chuẩn nhà nước cốt lõi chung cho nghệ thuật ngôn ngữ tiếng Anh và văn học trong lịch sử/nghiên cứu xã hội, khoa học và kỹ thuật môn học, 2020. URL http://www.corestandards.org/wp-content/uploads/Tiêu_chuẩn_ELA_1.pdf.
- [64] Kalervo Järvelin và Jaana Kekkonen. Các phương pháp đánh giá Ir để truy xuất tài liệu có liên quan cao. Trong Biên bản báo cáo của Hội nghị quốc tế thường niên lần thứ 23 Hội nghị ACM SIGIR về Nghiên cứu và Phát triển trong Thông tin traeval., trang 41-48. ACM, 2000.
- [65] Zenun Kastrati, Ali Shariq Imran, và Sule Yildirim Yayilgan. Tác động của học sâu về phân loại tài liệu bằng cách sử dụng biểu diễn giàu ngữ nghĩa. Xử lý và quản lý thông tin, 56(5):1618-1632, 2019.
- [66] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers và Brad S Chissom. Đạo hàm của các công thức khả năng đọc mới (chỉ số khả năng đọc tự động, số lượng sương mù và công thức dễ đọc của Flesch) dành cho quân nhân nhập ngũ của hải quân. Báo cáo kỹ thuật, Bộ Tư lệnh Đào tạo Kỹ thuật Hải quân Millington, Chi nhánh Nghiên cứu TN, 1975.

- [67] William H Kruskal và W Allen Wallis. Sử dụng thứ hạng trong phương sai một tiêu chí phân tích. Tạp chí của Hiệp hội thống kê Hoa Kỳ, 47(260):583-621, 1952.
- [68] Victor Kuperman, Hans Stadthagen-Gonzalez và Marc Brysbaert. Tuổi-của-xếp hạng tiếp thu cho 30.000 từ tiếng Anh. Phương pháp nghiên cứu hành vi, 44(4): 978-990, 2012.
- [69] Saar Kuzi, Sahiti Labhishetty, Shubhra Kanti Karmaker Santu, Prasad Pradip Joshi và ChengXiang Zhai. Phân tích đào tạo thích ứng để học xếp hạng trong việc truy xuất thông tin. Trong Biên bản báo cáo của Hội nghị quốc tế ACM lần thứ 28 Hội nghị về Quản lý thông tin và kiến thức, trang 2325-2328, 2019.
- [70] Monica Landoni, Davide Matteri, Emiliana Murgia, Theo Huibers, và Maria Soledad Pera. Sonny, cerca! đánh giá tác động của việc sử dụng giọng hát trợ lý tìm kiếm tại trường. Trong Hội nghị quốc tế về Ngôn ngữ chéo Diễn đàn đánh giá ngôn ngữ châu Âu, trang 101-113. Springer, 2019.
- [71] Inc. LAZEL. Biểu đồ tương quan mức độ. <https://www.readinga-z.com/learninga-z-levels/level-correlation-chart/>, 2021. (truy cập tháng 1 18, 2021).
- [72] Inc. LAZEL. Reading az: Chương trình đọc trực tuyến có thể tải xuống sách để in và lắp ráp. <https://www.readinga-z.com/>, 2021. (truy cập Ngày 18 tháng 1 năm 2021).
- [73] Lung-Hao Lee, Yen-Cheng Juan, Hsin-Hsi Chen, và Yuen-Hsien Tseng. Đối tượng-lọc nội dung có thể thực hiện được bằng dữ liệu nhấp chuột. Trong Biên bản của ACM lần thứ 22

hội nghị quốc tế về Quản lý thông tin và kiến thức, trang
1581-1584, 2013.

[74] Lukas Lerche. Sử dụng phản hồi ngầm cho hệ thống đề xuất: đặc điểm,
ứng dụng và thách thức. Luận án tiến sĩ, Đại học Technischen at Dortmund,
2016.

[75] Dirk Lewandowski. Đánh giá hiệu quả truy xuất của các công cụ tìm kiếm trên web
sử dụng mẫu truy vấn đại diện. Tạp chí của Hiệp hội Thông tin
Khoa học và Công nghệ, 66(9):1763-1775, 2015.

[76] Hang Li. Giới thiệu ngắn gọn về cách học xếp hạng. IEICE TRANSACTIONS
về Thông tin và Hệ thống, 94(10):1854-1862, 2011.

[77] Junjie Liang, Jinlong Hu, Shoubin Dong và Vasant Honavar. Hạng nhất: A
phương pháp xếp hạng theo danh sách có thể mở rộng cho các hệ thống đề xuất. Năm 2018 IEEE
Hội nghị quốc tế về dữ liệu lớn (Big Data), trang 1052-1058. IEEE,
2018.

[78] Hendi Lie, Darren Lukas, Jonathan Liebig và Richi Nayak. Một phương pháp học tập mới lạ-
phương pháp xếp hạng để điều khiển chuyển động camera tự động khi theo dõi thể thao điện tử.
Trong Hội nghị khai thác dữ liệu Australasian, trang 149-160. Springer, 2018.

[79] Tie-Yan Liu. Học cách xếp hạng để tìm kiếm thông tin. Springer Science &
Truyền thông kinh doanh, 2011.

[80] Fan Ma, Haoyun Yang, Haibing Yin, Xiaofeng Huang, Chenggang Yan, và
Xiang Meng. Học trực tuyến để xếp hạng theo cách tiếp cận danh sách để biết thông tin

truy xuất. Trong Hội nghị quốc tế IEEE năm 2019 về đa phương tiện và triển lãm (ICME), trang 1030-1035. IEEE, 2019.

[81] Craig Macdonald, Rodrygo LT Santos và Iadh Ounis. Khi nào và như thế nào của việc học cách xếp hạng cho tìm kiếm trên web. Truy xuất thông tin, 16(5):584-628, 2013.

[82] Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, và Maria Soledad Pera. Tìm kiếm bộ phim bảy hoặc bảy từ bộ phim đông lạnh? một chiến lược đa góc nhìn để đề xuất các truy vấn cho trẻ em. Trong Biên bản Hội nghị năm 2018 về Tương tác thông tin của con người & Truy xuất, trang 92-101, 2018.

[83] Alessandro Marani. Webdurank: một nguyên tắc xếp hạng giáo dục của web nguồn tài nguyên cho việc giảng dạy. Trong ICWL Doctoral Consortium, trang 25-36. Citeseer, 2016.

[84] Ryan McBride, Ke Wang, Zhouyang Ren và Wenyuan Li. Học tập nhảy cảm với chi phí để xếp hạng. Trong Biên bản Hội nghị AAAI về Trí tuệ nhân tạo, tập 33, trang 4570-4577, 2019.

[85] Changping Meng, Muhao Chen, Jie Mao, và Jennifer Neville. Readnet: Một hi-khung chuyển đổi phân cấp để phân tích khả năng đọc bài viết trên web. Tiến bộ trong Tìm kiếm thông tin, 12035:33, 2020.

[86] Yu Meng, Yunyi Zhang, Jiaxin Huang, Chenyan Xiong, Heng Ji, Chao Zhang, và Jiawei Han. Phân loại văn bản chỉ sử dụng tên nhãn: Một mô hình ngôn ngữ phương pháp tự đào tạo. Trong Biên bản báo cáo Hội nghị năm 2020 về Kinh nghiệm Phương pháp xử lý ngôn ngữ tự nhiên, 2020.

- [87] Ashlee Milton, Oghenemaro Anuyah, Lawrence Spear, Katherine Landau Wright và Maria Soledad Pera. Một chiến lược xếp hạng để thúc đẩy nguồn hỗ trợ môi trường lớp học. Trong Biên bản báo cáo năm 2020 Hội nghị chung quốc tế IEEE/WIC/ACM về Trí tuệ Web và Công nghệ tác nhân thông minh (WI-IAT'20), 2020.
- [88] Eleni Miltsakaki. Phù hợp sở thích và kỹ năng đọc của người đọc với văn bản web riêng. Trong Biên bản của Phiên họp trình diễn tại EACL 2009, trang 49-52, 2009.
- [89] Hamid Mohammadi và Seyed Hossein Khasteh. Văn bản như môi trường: A mô hình đánh giá khả năng đọc văn bản học tăng cường sâu. bản in trước arXiv arXiv:1912.05957, 2019.
- [90] Michinari Momma, Alireza Bagheri Garakani, Nanxun Ma, và Yi Sun. Đa-xếp hạng khách quan thông qua tối ưu hóa bị ràng buộc. Trong Biên bản đồng hành của Hội nghị Web 2020, trang 111-112, 2020.
- [91] Tin tức. Nội dung bài viết của Newsela, 2016. URL <https://newsela.com/data>.
- [92] Shastri L Nimmagadda, Dengya Zhu, và Amit Rudra. Cơ sở tri thức thông minh hơn khớp nối cho dự án thư mục mở trong hệ sinh thái kỹ thuật số bền vững. Trong Biên bản báo cáo đồng hành của Hội nghị quốc tế về World Wide Web, trang 1537-1545, 2017.
- [93] Jorge Nocedal và Stephen Wright. Tối ưu hóa số. Springer Science & Truyền thông doanh nghiệp, 2006.

- [94] Stefanie Nowak và Stefan Rügner. Các chú thích thông qua crowdsourcing có đáng tin cậy không? *ing: một nghiên cứu về sự thống nhất giữa các chú thích viên đối với chú thích hình ảnh đa nhân*. Trong *Biên bản báo cáo của hội nghị quốc tế về thông tin đa phương tiện* truy xuất, trang 557-566, 2010.
- [95] Harrie Oosterhuis và Maarten de Rijke. Học tập không thiên vị về nhận thức chính sách xếp hạng cho bảng xếp hạng top-k. Trong *Biên bản báo cáo của Hội nghị quốc tế ACM SIGIR lần thứ 43 Hội nghị về Nghiên cứu và Phát triển trong Thu thập thông tin*, trang 489-498, 2020.
- [96] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky và Elena Zhel-eva. Sửa lỗi cho sự thiên vị lựa chọn trong các hệ thống học xếp hạng. Trong *Biên bản Hội nghị Web 2020*, trang 1863-1873, 2020.
- [97] Liang Pang, Jun Xu, Qingyao Ai, Yanyan Lan, Xueqi Cheng và Jirong Wen. Setrank: Học một mô hình xếp hạng bất biến hoán vị cho thông tin lấy lại. Trong *Biên bản báo cáo của Hội nghị quốc tế ACM SIGIR lần thứ 43 về Nghiên cứu và Phát triển trong Tìm kiếm thông tin*, trang 499-508, 2020.
- [98] Deepshikha Patel và Prashant Kumar Singh. Phân loại tìm kiếm an toàn cho trẻ em mô hình. Trong *Hội nghị quốc tế về truyền thông và điện tử năm 2016 Hệ thống (ICCES)*, trang 1-7. IEEE, 2016.
- [99] Jodi Pilgrim và Sheri Vasinda. Tin giả và “mạng lưới toàn cầu”: Một nghiên cứu của lý luận về độ tin cậy của học sinh tiểu học. *Xã hội*, 11(4):121, 2021.
- [100] Marina A Hoshiba Pimentel, Israel Barreto Sant'Anna và Marcos Didonet Del Fabro. Tìm kiếm và xếp hạng các nguồn tài nguyên giáo dục dựa trên các cụm từ tering. Trong *ICEIS (1)*, trang 507-516, 2018.

- [101] KR Premalatha và TV Geetha. Xếp hạng lại các tài liệu giáo dục dựa trên chủ đề hồ sơ cho e-learning. Trong Hội nghị quốc tế năm 2012 về xu hướng gần đây trong Công nghệ thông tin, trang 217-221. IEEE, 2012.
- [102] Tao Qin và Tie-Yan Liu. Giới thiệu bộ dữ liệu letor 4.0. Bản in trước arXiv arXiv:1306.2597, 2013.
- [103] R Rajalakshmi và Chandrabose Aravindan. Một cách tiếp cận Bayes ngây thơ cho Phân loại url với khung lựa chọn và loại bỏ tính năng có giám sát. Trí tuệ tính toán, 34(1):363-396, 2018.
- [104] R Rajalakshmi, Hans Tiwari, Jay Patel, Ankit Kumar và R Karthik. Thiết kế của phân loại url dành riêng cho trẻ em sử dụng mạng nơ-ron tích chập hồi quy. Procedia Khoa học máy tính, 167:2124-2131, 2020.
- [105] R Rajalakshmi, Hans Tiwari, Jay Patel, R Rameshkannan và R Karthik. Mô hình chú ý dựa trên gru hai chiều để phân loại url dành riêng cho trẻ em. Trong Kỹ thuật học sâu và chiến lược tối ưu hóa trong phân tích dữ liệu lớn, trang 78-90. IGI Global, 2020.
- [106] Christian Ramiro, Mahesh Srinivasan, Barbara C Malt, và Yang Xu. Thuật toán-nhập điều trong sự xuất hiện lịch sử của các giác quan từ. Biên bản của Quốc gia Viện Hàn lâm Khoa học, 115(10):2323-2328, 2018.
- [107] Deborah K Reed và Sarah Kershaw-Herrera. Một cuộc kiểm tra phức tạp của văn bản tính được đặc trưng bởi khả năng đọc và sự gắn kết. Tạp chí Thực nghiệm Giáo dục, 84(1):75-97, 2016.

- [108] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, và Hye-Jung Lee.
Hướng tới tìm kiếm như một quá trình học tập: Một đánh giá về các quan điểm hiện tại và hướng đi trong tương lai. *Tạp chí Khoa học Thông tin*, 42(1):19-34, 2016.
- [109] Javier Sanz-Rodriguez, Juan Manuel Manuel Doderó, và Salvador Sanchez-Alonso. Xếp hạng các đối tượng học tập thông qua tích hợp các chất lượng khác nhau chỉ số. *Giao dịch IEEE về công nghệ học tập*, 3(4):358-363, 2010.
- [110] Mahmoud F Sayed và Douglas W Oard. Cùng nhau lập mô hình liên quan và độ nhạy để tìm kiếm trong số nội dung nhạy cảm. Trong *Biên bản báo cáo của Hội nghị liên ngành lần thứ 42 Hội nghị quốc gia ACM SIGIR về Nghiên cứu và Phát triển trong Thông tin Truy xuất*, trang 615-624, 2019.
- [111] Avi Segal, Kobi Gal, Guy Shani và Bracha Shapira. Xếp hạng độ khó cách tiếp cận cá nhân hóa trong e-learning. *Tạp chí quốc tế về con người Nghiên cứu máy tính*, 130:261-272, 2019.
- [112] Đậu Thần, Zheng Chen, Qiang Yang, Hua-Jun Zeng, Benyu Zhang, Yuchang Lu, và Wei-Ying Ma. Phân loại trang web thông qua tóm tắt. Trong *Biên bản Hội nghị quốc tế ACM SIGIR lần thứ 27 về nghiên cứu và Phát triển trong việc tìm kiếm thông tin*, trang 242-249, 2004.
- [113] George D Spache. Công thức đọc Spache. *Đọc tốt cho người kém đọc giả*, trang 195-207, 1974.
- [114] T Sreenivasulu, R Jayakarthish và R Shobarani. Phân loại nội dung web kỹ thuật dựa trên ontology mở. Trong *máy tính thông minh và đổi mới về Khoa học Dữ liệu (Biên bản báo cáo của ICTIDS 2019)*, trang 189-197. Springer, 2020.

- [115] Chi Sun, Xipeng Qiu, Yige Xu, và Huyền Cảnh Hoàng. Làm thế nào để tinh chỉnh bert cho phân loại văn bản? Tại Hội nghị quốc gia về tính toán Trung Quốc Ngôn ngữ học, trang 194-206. Springer, 2019.
- [116] Krysta M Svore, Maksims N Volkovs, và Christopher JC Burges. Học cách xếp hạng với nhiều hàm mục tiêu. Trong Biên bản báo cáo quốc tế lần thứ 20 hội nghị về World wide web, trang 367-376, 2011.
- [117] Rohail Syed và Kevyn Collins-Thompson. Tối ưu hóa kết quả tìm kiếm cho con người mục tiêu học tập. Tạp chí Truy xuất thông tin, 20(5):506-523, 2017.
- [118] Niek Tax, Sander Bockting, và Djoerd Hiemstra. Một so sánh chéo điểm chuẩn ison của 87 phương pháp học xếp hạng. Xử lý thông tin & quản lý, 51 (6):757-772, 2015.
- [119] T. Tieleman và G. Hinton. Bài giảng 6.5-RmsProp: Chia độ dốc cho một mức trung bình đang chạy của cường độ gần đây của nó. COURSERA: Mạng nơ-ron cho Học máy, 2012.
- [120] Arif Usta, Ismail Sengor Altingovde, Ibrahim Bahattin Vidinli, Rifat Ozcan, .. và Ozg ur Ulusoy. Học sinh từ lớp 1 đến lớp 12 tìm kiếm việc học như thế nào? phân tích một nhật ký công cụ tìm kiếm giáo dục. Trong Biên bản báo cáo của ACM quốc tế lần thứ 37 Hội nghị SIGIR về Nghiên cứu & phát triển trong thu thập thông tin, các trang 1151-1154, 2014.
- [121] Arif Usta, Ismail Sengor Altingovde, Rifat Ozcan, và Ozgur Ulusoy. Học hỏi-xếp hạng cho các công cụ tìm kiếm giáo dục. Giao dịch IEEE về Học tập Công nghệ, 2021.

- [122] Joost van Doorn, Daan Odijk, Diederik M Roijers, và Maarten de Rijke. Bal-tiêu chí liên quan tài chính thông qua tối ưu hóa đa mục tiêu. Trong Biên bản của hội nghị quốc tế ACM SIGIR lần thứ 39 về Nghiên cứu và Phát triển trong Truy xuất thông tin, trang 769-772, 2016.
- [123] Nicholas Vanderschantz và Annika Hinze. Trẻ em nhìn thấy tìm kiếm như thế nào: Một hình ảnh phân tích các công cụ tìm kiếm trên internet. Trong HCI 2017. BISL, 2017.
- [124] Ellen M Voorhees. Sự tiến hóa của cranfield. Trong Truy xuất thông tin Đánh giá trong một thế giới đang thay đổi, trang 45-69. Springer, 2019.
- [125] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang và Hang Li. theo danh sách cách tiếp cận để học xếp hạng: lý thuyết và thuật toán. Trong Biên bản báo cáo của lần thứ 25 hội nghị quốc tế về Học máy, trang 1192-1199, 2008.
- [126] Tian Xia. Phân loại tài nguyên giáo dục dựa trên máy vectơ hỗ trợ. Tạp chí quốc tế về công nghệ thông tin và giáo dục, 6(11):880, 2016.
- [127] Jun Xu và Hang Li. Adarank: một thuật toán tăng cường để truy xuất thông tin. Trong Biên bản báo cáo của hội nghị quốc tế ACM SIGIR lần thứ 30 về Nghiên cứu và phát triển trong tìm kiếm thông tin, trang 391-398, 2007.
- [128] Lưu Dương, Qingyao Ai, Damiano Spina, Ruey-Cheng Chen, Liang Pang, W Bruce Croft, Jiafeng Guo và Falk Scholer. Ngoài ra factoid: Hiệu quả phương pháp để tìm lại câu trả lời không có thực tế. Trong Hội nghị Châu Âu về Truy xuất thông tin, trang 115-128. Springer, 2016.

- [129] Sevgi Yigit-Sert, Ismail Sengor Altingovde, Craig Macdonald, Iadh Ounis, ..
và Ozgür Ulusoy. Đa dạng hóa rõ ràng các kết quả tìm kiếm trên nhiều
kích thước cho tìm kiếm giáo dục. Tạp chí của Hiệp hội Thông tin
Khoa học và Công nghệ, 2020.
- [130] Shanshan Yu, Jindian Su, và Da Luo. Cải thiện phân loại văn bản dựa trên bert
với câu phụ trợ và kiến thức về miền. IEEE Access, 7:176600-176612,
2019.
- [131] Junta Zeniarja, Ramadhan Rakhmat Sani, Ardytha Luthfiarta, Heru Agus
Susanto, Erwin Yudi Hidayat, Abu Salam và Leonardus Irfan Bayu Mahen-
dra. Công cụ tìm kiếm cho trẻ em với chức năng lọc và xếp hạng tài liệu bằng cách sử dụng
bayes phân loại. Trong Hội thảo quốc tế năm 2018 về ứng dụng công nghệ
Thông tin và Truyền thông, trang 560-564. IEEE, 2018.
- [132] Wenjie Zhao, Gaoyu Zhang, George Yuan, Jun Liu, Hongtao Shan và Shuyi
Zhang. Nghiên cứu về phân loại văn bản cho tin tức tài chính dựa trên một phần
thông tin. IEEE Access, 8:100426-100437, 2020.

PHỤ LỤC A

DỮ LIỆU CHO THÍ NGHIỆM

- AoA - Bộ sưu tập các thuật ngữ từ vựng được gắn nhãn theo độ tuổi của trẻ em có được sự hiểu biết về chúng [68]. Được sử dụng trong thiết kế của Spache-Allen (Phần 3.1.1).
- EduSites - Bộ sưu tập các nguồn tài nguyên trực tuyến được trích xuất từ các trang web hàng đầu của Alexa Danh mục thể loại mang tính giáo dục dành cho trẻ em (Mục 4.2).
- ObjSet - Bộ sưu tập các nguồn tài nguyên trực tuyến được trích xuất từ các trang web hàng đầu của Alexa Danh mục danh mục không phù hợp với trẻ em, tức là các tài nguyên thuộc trong các danh mục được mô tả trong Phần 4.3.
- RankSet - Bộ sưu tập các kết quả tìm kiếm được xếp hạng để phản hồi các truy vấn có nguồn gốc từ tiêu đề bài viết của NewsELA (Mục 4.4).
- Sven + AoA Full - Sự kết hợp của các thuật ngữ trong AoA và Sven. Được sử dụng trong thiết kế của Spache-Allen (Mục 3.1.1).
- Sven - Từ điển thuật ngữ được trích xuất từ các trang web liên quan đến trẻ em dành cho công việc được thực hiện trong [82]. Được sử dụng trong thiết kế Spache-Allen (Phần 3.1.1).

- TextComp - Bộ sưu tập sách và tài nguyên web được gắn nhãn với mục đọc mức độ dưới dạng một lớp mà trẻ có thể đọc được tương ứng tài nguyên (Mục 4.1).