NGHIÊN CỨU XÂY DỰNG CÔNG CỤ TRA CỨU THÔNG MINH HƯỚNG DẪN KIỂM SOÁT HOẠT ĐỘNG LUÂN CHUYỂN HÀNG NGUY HIỂM QUA CẢNG BIỂN VIỆT NAM

RESEARCH AND DEVELOPMENT OF AN INTELLIGENT SEARCH TOOL FOR GUIDING THE CONTROL OF DANGEROUS GOODS MOVEMENT THROUGH VIETNAMESE SEAPORTS

NGUYỄN THANH DIỆU*, PHAN VĂN HƯNG
Khoa Hàng hải, Trường Đại học Hàng hải Việt Nam
**Email liên hệ: nguyenthanhdieu@vimaru.edu.vn*
DOI: https://doi.org/10.65154/jmst.969

Tóm tắt

An toàn trong vận chuyển hàng nguy hiểm là yêu cầu cốt lõi đối với thuyền viên, lực lượng quản lý cảng và các đơn vị luân chuyển hàng hóa. Nghiên cứu này xây dựng công cụ tra cứu thông minh dựa trên kỹ thuật tri thức để hỗ trợ kiểm soát hàng nguy hiểm theo Bộ luật Hàng nguy hiểm Hàng hải Quốc tế mới nhất 2024 (International Maritime Dangerous Goods Code - IMDG Code 2024). Hệ thống ứng dụng kỹ thuật Thế hệ tăng cường truy xuất RAG (Retrieval-Augmented Generation) với vector nhúng 3072 chiều (vector embedding) cùng kỹ thuật Tinh chỉnh tăng cường truy xuất (Retrieval-Augmented Fine-Tuning - RAFT) nhằm tối ưu hóa việc ra quyết định của Thuyền trưởng, thuyền viên, Sỹ quan an toàn cảng và các chủ thể có liên quan đến hoạt động kiểm soát hàng nguy hiểm luân chuyển qua cảng biển. Bằng cách mô hình hóa định lượng nỗ lực tác nghiệp và độ phức tạp quy trình kiểm soát, nghiên cứu chứng minh hệ thống đạt độ chính xác trên 95% với thời gian phản hồi dưới 10 giây, giúp giảm thiểu rủi ro từ yếu tố con người. Giải pháp cung cấp một trợ lý số tin cậy, giúp nội địa hóa quy trình an toàn hàng hải và nâng cao hiệu quả phối hợp giữa các lực lượng khai thác cảng biển Việt Nam.

Từ khóa: *Hàng nguy hiểm, quản lý cảng, IMDG Code 2024, thuyền trưởng, thuyền viên, sỹ quan an toàn, RAG (Retrieval-Augmented Generation), 3072-dim Vector Embedding, RAFT (Retrieval-Augmented Fine-Tuning).*

Abstract

Dangerous goods safety is a fundamental requirement for seafarers, port management forces, and logistics units. This study develops an intelligent search tool based on knowledge engineering to support the control of dangerous goods according to the latest International Maritime Dangerous Goods Code 2024 (IMDG Code 2024). The system utilizes Retrieval-Augmented Generation (RAG) with 3072-dimensional vector embeddings and Retrieval-Augmented Fine-Tuning (RAFT) to optimize decision-making for Masters, seafarers, Port Safety Officers, and stakeholders involved in the control of dangerous goods movement through seaports. By mathematically modeling operational effort and control process complexity, the research demonstrates over 95% accuracy with response times under 10 seconds, mitigating risks associated with human factors. The solution provides a reliable digital assistant, localizing maritime safety processes and enhancing operational coordination among Vietnamese seaport authorities.

Keywords: *Dangerous goods, port management, IMDG Code 2024, Masters, seafarers, Port Safety Officers, RAG (Retrieval-Augmented Generation), 3072-dim Vector Embedding, Retrieval-Augmented Fine-Tuning (RAFT).*

1. Mở đầu

Hoạt động vận chuyển hàng nguy hiểm qua các cảng biển Việt Nam đang ngày càng trở nên phức tạp do nhu cầu về hóa chất và năng lượng phục vụ sản xuất tăng cao. Công tác kiểm soát an toàn hiện nay đối mặt với nhiều thách thức, từ việc đánh giá rủi ro chưa đầy đủ đến sự chậm trễ trong cập nhật các quy định quốc tế. Đặc biệt, sự ra đời của Bộ luật Hàng nguy hiểm Hàng hải Quốc tế phiên bản 2024 bổ sung sửa đổi (IMDG Code Amendment 42-24/2024 Edition) với các quy tắc nghiêm ngặt mới về pin công nghệ mới và vật liệu tự phát nhiệt yêu cầu một cơ chế thực thi nhanh chóng tại các bến cảng.

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

Hiện nay, thuyền viên và nhân viên cảng vẫn còn áp dụng cách thức tra cứu thủ công các tài liệu giấy về hàng nguy hiểm dẫn đến không đảm bảo tính kịp thời và dễ dẫn đến sai sót trong các tình huống khẩn cấp. Dữ liệu thanh tra của IMO cho thấy "có tới 34% các lô hàng nguy hiểm tồn tại sai sót về mặt tuân thủ do quy trình tra cứu thủ công lạc hậu" [4]. Bên cạnh đó thì cách thức tra cứu nhanh cũng đã được thực hiện nhưng thường phụ thuộc vào các dịch vụ tra cứu quốc tế như Hazcheck Online với mức phí bản quyền rất cao (dao động từ 182 đến 565 USD/người dùng/năm), gây khó khăn cho việc triển khai áp dụng trên diện rộng. Do đó, việc nghiên cứu xây dựng một công cụ tra cứu điện tử nội địa tích hợp công nghệ mô hình ngôn ngữ lớn (Large Language Model - LLM) không chỉ giúp làm chủ công nghệ mà còn là yêu cầu cấp thiết để đảm bảo an toàn tuyệt đối cho hoạt động cảng biển.

2. Các quy chuẩn an toàn và cập nhật IMDG Code 2024

2.1. Phân loại hàng nguy hiểm luân chuyển qua cảng

Theo Bộ luật IMDG, hàng nguy hiểm được phân loại dựa trên bản chất lý hóa và mức độ rủi ro đối với con người, tàu biển và môi trường. Hệ thống phân loại bao gồm 9 nhóm (Classes) chính, trong đó một số nhóm được chia nhỏ thành các phân nhóm (Divisions) để cụ thể hóa tính chất nguy hiểm.

Bảng 1. Hệ thống phân loại hàng nguy hiểm cập nhật phiên bản 2024

Nhóm	Bản chất nguy hiểm chính	Ví dụ cập nhật 2024
1	Chất và vật phẩm nổ	UN 0514 (Thiết bị dập lửa)
2	Các loại chất khí (Dễ cháy, Độc, Không cháy)	UN 3553 (Khí Disilane)
4	Chất rắn tự phản ứng/bốc cháy	UN 1361 (Carbon - SP978)
8	Chất ăn mòn mạnh	UN 3554 (Gallium trong vật phẩm)
9	Các loại nguy hiểm khác	Pin Natri-ion UN 3551, 3552

Mức độ nguy hiểm của hàng hóa (trừ Nhóm 1, 2, 5.2, 6.2 và 7) được xác định thông qua cách thức yêu cầu đóng gói (Packing Group - PG): PG I (nguy hiểm cao nhất), PG II (nguy hiểm trung bình) và PG III (nguy hiểm thấp). Việc tra cứu chính xác PG là bắt buộc để xác định tiêu chuẩn bao bì và các điều kiện phân tách khu chứa hàng (segregation). Trong phiên bản 2024, chất Tetramethylammonium hydroxide (UN 3423) đã được nâng lên PG I do độc tính cao, dẫn đến các yêu cầu đóng gói khắt khe hơn và cấm vận chuyển dưới dạng số lượng giới hạn.

2.2. Những cập nhật quan trọng trong IMDG Code 2024 (Amendment 42-24)

Phiên bản 2024 mang đến những thay đổi bắt buộc từ 01/01/2026, đặc biệt là Quy định đặc biệt SP978 yêu cầu giám sát nhiệt độ đóng gói nghiêm ngặt cho Carbon để ngăn ngừa tự phát nhiệt. Sự phức tạp này đòi hỏi lực lượng tác nghiệp phải có công cụ hỗ trợ thông minh để đảm bảo tuân thủ.

2.3. Ý tưởng xây dựng giải pháp công nghệ từ nhu cầu thực tế

Xuất phát từ nhu cầu thực tiễn cần tra cứu nhanh thông tin hàng nguy hiểm, tác giả đã hình thành ý tưởng xây dựng công cụ tra cứu nhanh sử dụng thuật toán thông minh dựa trên quy trình như sau:

Từ các thông tin về hàng nguy hiểm và các quy định về an toàn trong IMDG Code cũng như các tiêu chuẩn quy định ở nước ta, nghiên cứu sẽ thực hiện số hoá nguồn tài liệu thông tin này tích hợp dưới dạng cơ sở dữ liệu (database). Thông qua các thuật toán, công cụ tin học và trí tuệ nhân tạo (AI), dữ liệu này được xây dựng thành một ứng dụng cho phép người dùng có thể tra cứu nhanh trực tuyến (online) thông tin hàng nguy hiểm bằng cách nhập thông số để tra cứu là mã số Liên hợp quốc (UN Number) hoặc người dùng có thể tương tác hỏi đáp dạng Trợ lý ảo (Chatbot). Giải pháp này giúp thuyền viên và lực lượng quản lý cảng cũng như các cá nhân cần tra cứu thông tin có thể dễ dàng tiếp cận linh hoạt các kiến thức an toàn quản lý hàng nguy hiểm một cách chuẩn xác và kịp thời.

3. Mô hình hóa hệ thống và Kỹ thuật tri thức

Để hiện thực hóa ý tưởng trên, tác giả đã tiến hành nghiên cứu ứng dụng các phương pháp toán học định lượng để cấu trúc lại quy trình quản trị nghiệp vụ.

3.1. Định lượng nỗ lực tác nghiệp của lực lượng hàng hải

Nghiên cứu đo lường nỗ lực phối hợp giữa các tác nhân (*Quản trị viên, Nhân viên, Khách hàng*) trong mô hình ca hoạt động được biểu diễn bởi hàm mô hình sau:

$$\mathcal{M}_{uc} = (\mathcal{T}, \mathcal{U}, \mathcal{R}) \quad (1)$$

Trong đó:

$\mathcal{T}$ là tập hợp các tác nhân sử dụng hệ thống, cụ thể gồm có 03 tác nhân chính (*Quản trị viên, nhân viên, khách hàng*). Trong đó *khách hàng* tùy theo tính chất nghề nghiệp có thể phân loại thành *thuyền viên*, *quản lý cảng*, *đơn vị Logistics,…*

$\mathcal{U}$ là tập hợp các ca hoạt động, bao gồm các chức năng chính của hệ thống: Quản lý (*người dùng, tài liệu, khách hàng*), đăng ký tài khoản, tìm kiếm thông tin, hỏi đáp chatbot, cập nhật hồ sơ,...

$\mathcal{R}$ là tập hợp quan hệ giữa tác nhân và ca hoạt động hoặc giữa các ca hoạt động với nhau

Độ phức tạp của mô hình $\mathcal{M}_{uc}$ trong tra cứu hàng nguy hiểm hàng hải được định lượng bằng hàm trọng số xác định độ phức tạp $C(\mathcal{M}_{uc})$ dựa trên quy mô của các tập hợp thành phần:

$$C(\mathcal{M}_{uc}) = w_T|\mathcal{T}| + w_U|\mathcal{U}| + w_R|\mathcal{R}| \qquad (2)$$

Trong đó:

$|\mathcal{T}|$ là tập hợp các tác nhân, thể hiện cho các bên liên quan trực tiếp tham gia và quy trình xử lý hàng nguy hiểm, đặc biệt là các loại *khách hàng* tham gia (*Thuyền viên*, *quản lý cảng*, *đơn vị logistics* …) Số lượng loại tác nhân càng lớn, sự phân mảnh trách nhiệm và nhu cầu xác thức thông tin càng cao, làm tăng độ phức tạp chung.

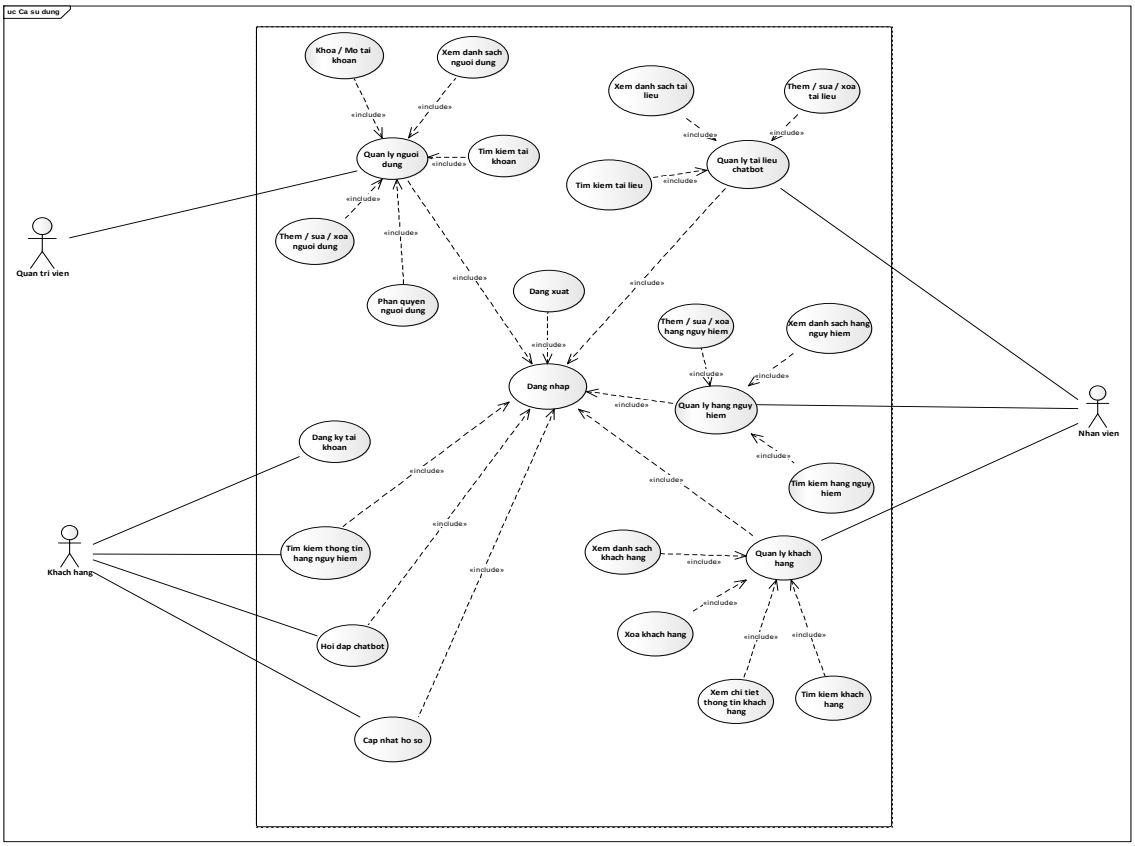
$|\mathcal{U}|$ là tập các tác vụ an toàn cốt lõi cần phải thực hiện theo quy định (Kiểm tra IMDG 2024, lập Safety Checklist …). Đây là thông số quan trọng vì IMDG Code 2024 có càng nhiều quy định thì số lượng $|\mathcal{U}|$ càng tăng, dẫn đến số lượng tác vụ cũng tăng theo tỉ lệ thuận.

$|\mathcal{R}|$ thể hiện số lượng các luồng trao đổi thông tin và tương tác kỹ thuật giữa các các tác nhân và tác vụ.

Bảng 2. Ma trận so sánh cặp

Chỉ số	$\mathcal{T}$	$\mathcal{U}$	$\mathcal{R}$	Trọng số
Tác nhân ($\mathcal{T}$)	1	1/3	1/2	0.16 (w_T)
Tác vụ ($\mathcal{U}$)	3	1	2	0.54 (w_U)
Quan hệ ($\mathcal{R}$)	2	1/2	1	0.30 (w_R)

Các trọng số w_T, w_U và w_R phản ánh mức độ đóng góp của từng thành phần vào độ phức tạp tổng thể. Các trọng số này (với điều kiện $w_T + w_U + w_R = 1$) được tính toán qua ma trận so sánh cặp (Bảng) của phương pháp Phân tích thứ bậc (AHP-Analytic Hierarchy Process). Các giá trị so sánh cặp được xác lập dựa trên đánh giá của các chuyên gia hàng hải về mức độ ảnh hưởng của quy trình kiểm soát và luồng



Hình 1. Sơ đồ ca sử dụng hệ thống tra cứu hàng nguy hiểm

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

thông tin đối với rủi ro sai sót trong thực tế vận hành theo IMDG Code 2024. Bên cạnh đó, việc xác định trọng số w_U có giá trị cao nhất (0,54) nhằm phản ánh thực tế các sai sót chủ yếu hiện nay (34% [4]) xuất phát từ sự phức tạp trong việc thực hiện các tác vụ tra cứu và tuân thủ các quy tắc của IMDG Code 2024. Từ đó, thể hiện chỉ số $C(\mathcal{M}_{uc})$ là căn cứ hợp lý, góp phần hỗ trợ tinh gọn các bước xử lý, giúp giảm thiểu áp lực và rủi ro do yếu tố con người.

Để hợp lý hóa quy trình thiết kế, nghiên cứu thực hiện gộp các ca hoạt động tương đồng ngữ nghĩa thành ca sử dụng chung "Quản lý Tài khoản".

Hình 1 mô tả các tác vụ (*ca hoạt động*) cụ thể mà hệ thống sẽ thực hiện và cung cấp một cách đơn giản, hiệu quả để hiểu và trao đổi yêu cầu với các bên liên quan. Nó thể hiện ứng xử của hệ thống đối với bên ngoài, trong một hoàn cảnh nhất định, xét từ quan điểm của người sử dụng. Nó mô tả các yêu cầu đối với hệ thống, có nghĩa là những gì hệ thống phải thực hiện.

3.2. Kiến trúc hệ thống và mô hình dữ liệu tri thức hàng hải

Hệ thống được thiết kế theo kiến trúc hướng đối tượng, tập trung vào việc nâng cao hiệu quả luồng xử lý thông tin giữa các tác nhân với nhau và gộp các ca hoạt động tương đồng về ngữ nghĩa, giảm thiểu sự phức tạp trong quản lý tài khoản và phân quyền.

Mô hình lớp đối tượng dùng để mô tả cấu trúc tĩnh của hệ thống hướng đối tượng, biểu diễn các lớp (class), thuộc tính, phương thức hoạt động và mối quan hệ giữa chúng, làm cầu nối giữa phân tích và thiết kế hệ thống, hỗ trợ triển khai và bảo trì.

Mỗi lớp C_i trong hệ thống có thể được biểu diễn dưới dạng một bộ ba như sau:

$$C_i=(N_i, A_i, O_i) \quad (3)$$

Trong đó:

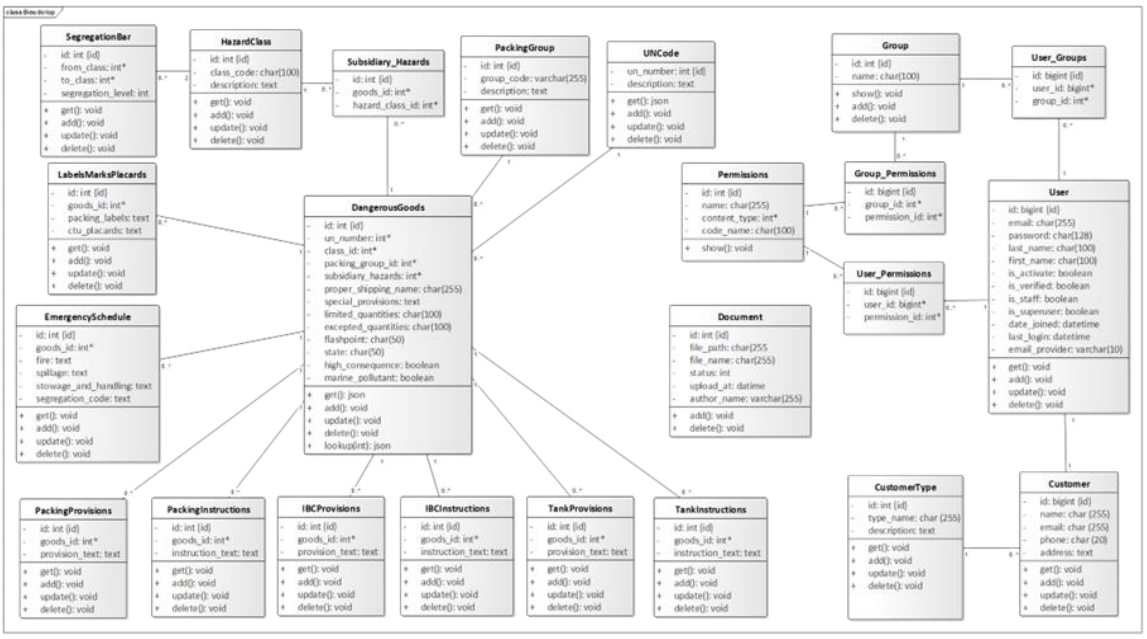
N_i là tên của lớp (Trong sơ đồ Hình 2 thì N_i là tiêu đề ở ô trên cùng của mỗi khối như: DangerousGood, UNCode,…).

A_i là tập hợp các thuộc tính (attributes) của lớp. Mỗi thuộc tính $a \in A_i$ được định nghĩa bởi một tên và một kiểu dữ liệu (Trong sơ đồ Hình 2 thì A_i là danh sách các biến ở ô giữa mỗi khối như: un_number, proper_shipping_name, packing_group).

O_i là tập hợp các phương thức (operations) của lớp. Mỗi phương thức $o \in O_i$ có một định danh, danh sách tham số và kiểu trả về (Trong sơ đồ Hình 2 thì O_i là danh sách các hàm/thao tác ở ô dưới cùng mỗi khối như: get, add, update, delete,…)

Hình 2 thể hiện các lớp (class) xử lý cần thiết phải có trong hệ thống. Với mỗi lớp, đặc tả các trường thông tin và phương thức (hàm chức năng) xử lý thông tin. Các lớp có mối quan hệ 1-1 hoặc 1-n với nhau, thể hiện mối liên kết luồng xử lý dữ liệu, thông tin khi các tác nhân thực hiện các ca sử dụng trong hệ thống.

Cơ sở dữ liệu của các lớp được chuẩn hóa để lưu trữ đầy đủ các thuộc tính của hàng nguy hiểm như: UN Number, Proper Shipping Name (PSN), Class,



Hình 2. Cơ sở dữ liệu hệ thống

Packing Group, và IMDG Segregation. Chỉ số liên kết giữa các lớp (Coupling Between Objects - CBO) được tối thiểu hóa để tăng khả năng bảo trì và nâng cấp hệ thống khi các quy định quốc tế thay đổi, thể hiện qua các chỉ số định lượng sau:

a) Độ liên kết đối tượng (CBO): Phản ánh tính độc lập của các khâu kiểm soát an toàn. Trong quản trị hàng hải, một *CBO* thấp đảm bảo rằng khi quy tắc về Nhóm đóng gói (Packing Group) thay đổi, nó không gây lỗi cho quy trình Phân tách (Segregation) khu vực chứa hàng. *CBO* được xác định theo công thức:

$$CBO_{system} = \sum_{i=1}^{n} |\{C_j | C_i \text{ phụ thuộc vào } C_j\}| \quad (4)$$

Trong đó: C_i là lớp quy trình đang xét; C_j là các lớp khác có liên kết trực tiếp với C_i.

b) Độ phức tạp quy trình nghiệp vụ (M): Đo lường tính tinh gọn trong luồng quyết định của sỹ quan hiện trường thông qua chỉ số Cyclomatic *(M)*:

$$M=|E|-|V|+2P \quad (5)$$

Trong đó: $|E|$ là số cạnh (luồng điều khiển nghiệp vụ); $|V|$ là số đỉnh (các thao tác/quyết định thực tế của sĩ quan quản lý hiện trường); P là số thành phần liên thông ($P=1$).

Chỉ số M chứng minh hệ thống AI đã tối giản hóa "đường đi" trong tư duy ra quyết định của con người, từ đó giảm tỷ lệ sai sót do áp lực công việc.

3.3. Số hóa tri thức kỹ thuật cao: RAG và Vector Embedding 3072 chiều

Hệ thống ứng dụng kỹ thuật Retrieval-Augmented Generation (RAG) với mô hình *text-embedding-3-large* để biến văn bản tài liệu quy định theo IMDG Code và quy định Việt Nam thành các vector 3072 chiều. Độ tương đồng nghiệp vụ được tính toán bằng hàm Cosine Similarity:

$$sim(q,di)=||q||\cdot||di||q\cdot di \quad (6)$$

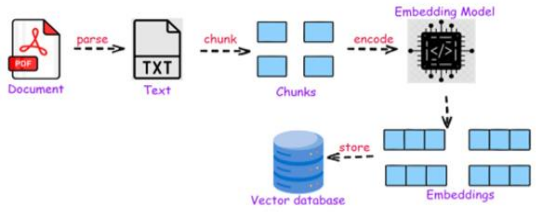
Trong đó: q là vector truy vấn kiến thức của người dùng; di là vector đoạn tri thức theo quy định chuẩn IMDG Code; $q.di$ là tích vô hướng; || . || biểu thị độ dài vector.

Quy trình số hóa tri thức IMDG phục vụ hỗ trợ tra cứu tương tác Chatbot AI ứng dụng kỹ thuật RAG với khả năng truy xuất tri thức động, mô hình này tạo ra một cơ chế sinh văn bản "có dẫn chứng", giúp các hệ thống AI tăng độ tin cậy, dễ kiểm chứng hơn và dễ dàng tùy biến theo ngữ cảnh sử dụng cụ thể.

Một hệ thống RAG điển hình bao gồm bốn thành phần chính: Xây dựng chỉ mục (Indexing), truy xuất (Retrieval), tăng cường ngữ cảnh (Augmentation), và sinh ngôn ngữ (Generation). Mỗi thành phần đóng vai trò cụ thể trong việc đảm bảo hệ thống có thể sinh ra đầu ra chính xác và có căn cứ.

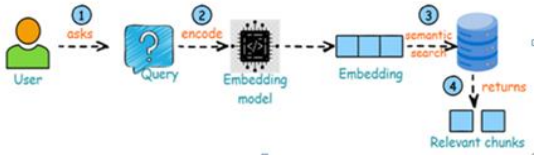
(1) Indexing RAG: Quá trình mã hóa tài liệu IMDG thành vector dữ liệu. Quá trình này nhằm chuẩn bị dữ liệu tri thức để truy xuất sau này. Tài liệu văn bản được chia nhỏ thành các đoạn (chunk), sau đó ánh xạ thành vector ngữ nghĩa thông qua mô hình embedding.

Các vector này kết hợp cùng với metadata sẽ được lưu vào một cơ sở dữ liệu vector (gọi là vector store) như FAISS, Chroma hoặc Qrand. Quá trình này chỉ thực hiện một lần khi dữ liệu nguồn thay đổi (Hình 3).



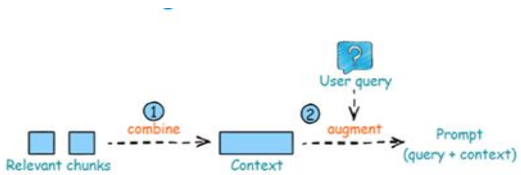
Hình 3. Quá trình Indexing RAG

(2) Retrieval RAG: Quá trình chatbot tìm kiếm đoạn văn bản liên quan đến câu hỏi người dùng. Khi người dùng đặt câu hỏi, hệ thống sẽ ánh xạ truy vấn thành vector và tìm kiếm các đoạn văn bản gần nhất về mặt ngữ nghĩa trong cơ sở dữ liệu vector. Chiến lược phổ biến là top-k retrieval, trong đó k đoạn liên quan nhất sẽ được chọn để đưa vào bước tiếp theo (Hình 4).



Hình 4. Quá trình Retrieval RAG

(3) Augment RAG: Quá trình gộp thông tin truy xuất vào câu hỏi hay còn gọi là tăng cường ngữ cảnh. Các đoạn văn bản được truy xuất sẽ được gộp lại cùng với câu hỏi gốc để tạo thành ngữ cảnh đầu vào cho mô

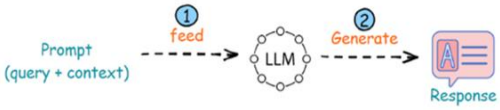


Hình 5. Quá trình Augment RAG

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

hình sinh. Đây là điểm cốt lõi giúp mô hình sinh không chỉ dựa vào tri thức bên trong mà còn tận dụng thông tin từ tài liệu thực tế để tạo câu trả lời (Hình 5).

(4) Generation RAG: Quá trình mô hình ngôn ngữ lớn sinh ra câu trả lời cuối cùng. Mô hình ngôn ngữ lớn (LLM) sẽ sinh ra câu trả lời dựa trên ngữ cảnh được tăng cường. Mô hình có thể là GPT-3.5, GPT-4, LLaMA, Mistral, hay các mô hình mã nguồn mở khác. Đầu ra thường mang tính trôi chảy, tự nhiên và chính xác hơn so với các mô hình không được hỗ trợ bởi tri thức nền (Hình 6).



Hình 6. Quá trình Generation RAG

3.4. Chiến lược chia nhỏ dữ liệu SDPM và kỹ thuật RAFT

Để đảm bảo duy trì sự đồng nhất ngữ cảnh trao đổi trong một khung hội thoại dài giữa người dùng và chatbot nhằm tránh việc chia văn bản làm mất ngữ nghĩa, nghiên cứu ứng dụng chiến lược nghiên cứu áp dụng chiến lược chia nhỏ dữ liệu SDPM (Semantic Double-Pass Merging). Quá trình chia nhỏ (chunking) được thực hiện qua hai bước:

Bước 1: Phân đoạn sơ bộ dựa trên cấu trúc logic.

Bước 2: Sử dụng "cửa sổ trượt" (skip-window) nhìn trước 2 đoạn để gộp các mini-chunks có độ tương đồng ngữ nghĩa cao

Ví dụ SDPM: Thực hiện gộp bảng quy định đặc biệt (Special Provision) bị chia cắt qua nhiều trang PDF thành một khối tri thức thống nhất để AI truy xuất trọn vẹn.

Để xây dựng công cụ tra cứu, mô hình ngôn ngữ được lựa chọn cho để tinh chỉnh là LLaMA3.2 3B - một biến thể nhỏ gọn thuộc thế hệ thứ ba của dòng LLaMA từ Meta AI. Với khoảng 3 tỷ tham số, mô hình cân bằng tối ưu giữa hiệu năng và chi phí triển khai tại cảng. Hệ thống sử dụng mô hình nhúng text-embedding-3-large để chuyển đổi tri thức IMDG thành vector 3072 chiều. Mô hình ngôn ngữ lớn (Large Language Model - LLM) được huấn luyện để phân biệt giữa "tài liệu vàng" (chứa tri thức đúng) và "tài liệu nhiễu". Quá trình huấn luyện bao gồm các bước suy luận Chain-of-Thought (CoT) để đảm bảo Chatbot AI luôn trích dẫn đúng nguồn tri thức từ IMDG Code khi phản hồi thuyền viên.

Hệ thống được tinh chỉnh bằng kỹ thuật RAFT (Retrieval-Augmented Fine-Tuning) hay còn gọi là Tinh chỉnh tăng cường Truy xuất, là một phương pháp huấn luyện mô hình ngôn ngữ tiên tiến, kết hợp giữa sức mạnh truy xuất tri thức động của RAG và khả năng hiểu sâu chuyên ngành của Fine-tuning, ở đây RAFT đóng vai trò là "bộ lọc tư duy" giúp trợ lý AI không chỉ đọc Bộ luật IMDG 2024 mà còn biết cách suy luận như một chuyên gia hàng hải thực thụ.

Việc Fine-tuning mô hình LLaMA3.2 3B được thực hiện theo kỹ thuật RAFT, sử dụng LoRA (Low-Rank Adaptation) để tối ưu hóa tài nguyên tính toán. Mô hình được huấn luyện để phân biệt giữa "tài liệu vàng" và tài liệu nhiễu thông qua chuỗi suy luận Chain-of-Thought (CoT) và ép buộc trích dẫn nguyên văn (Verbatim). Điều này giúp trợ lý số loại bỏ hoàn toàn hiện tượng ảo giác thông tin.

4. Kết quả thực nghiệm và Sự phối hợp đa ngành

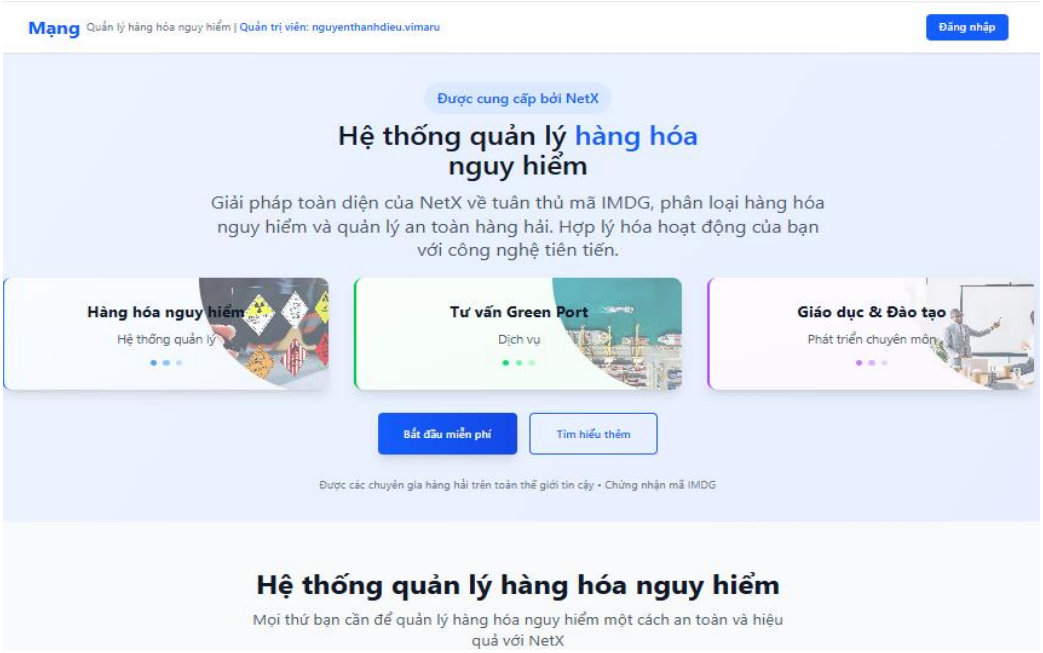
4.1. Hành trình xây dựng trang web tra cứu NetX và sự phối hợp chuyên gia IT

Việc xây dựng thành công trang web tra cứu trực tuyến NetX (https://netx.com.vn/) là kết quả của nỗ lực phối hợp đa ngành. Với vai trò là giảng viên khoa hàng hải kết hợp với kinh nghiệm thực tế thường xuyên làm việc trên tàu hoá chất, tác giả đóng vai trò là "Người định hướng chuyên môn" - thực hiện xác lập các logic an toàn của IMDG Code 2024 và các quy định nhà nước về hàng nguy hiểm, từ đó hoạch định phương pháp và hướng nghiên cứu để hoàn thành mục tiêu "Nghiên cứu xây dựng công cụ tra cứu thông minh hướng dẫn kiểm soát hoạt động luân chuyển hàng nguy hiểm qua cảng biển Việt Nam". Tuy nhiên, để chuyển hóa phần "Nội hàm tri thức hàng hải" này thành ứng dụng ổn định, tác giả đã phối hợp chặt chẽ với các chuyên gia Công nghệ thông tin (IT). Sự đồng hành này giúp tối ưu hóa thuật toán RAG và không gian vector 3072 chiều, đảm bảo hệ thống vận hành trơn tru với dữ liệu thực tế lớn nhằm xây dựng thành công ứng dụng tra cứu điện tử hàng nguy hiểm đáp ứng yêu cầu nghiên cứu đã đề ra.

** Đánh giá kết quả ứng dụng:*

Công cụ tra cứu thông minh tạo ra sự thay đổi mang tính đột phá so với phương pháp tra cứu truyền thống. Giao diện NetX cho phép người dùng truy cập Dashboard thống kê và Chatbot Assistant 24/7.

Việc sử dụng vector embedding 3072 chiều giúp hệ thống xử lý hiệu quả các trường hợp hàng hóa có tên tương tự nhưng tính chất hóa học khác biệt hoàn toàn, từ đó giảm thiểu rủi ro sắp xếp nhầm hàng tại cảng.



Hình 7. Giao diện trang chủ Hệ thống tra cứu trực tuyến hàng nguy hiểm (NetX)

Bảng 3. Đánh giá hiệu quả công cụ tra cứu thông minh

Chỉ số đánh giá	Tra cứu thủ công	Công cụ thông minh	Căn cứ xác định/ Nguồn
Thời gian truy vấn trung bình	5 - 15 phút	< 10 giây	Đo lường thực tế
Độ chính xác thông tin	70 - 80% [4], [21]	> 95% [15, Tr.1], [18]	[4], [21], [15], [18]
Cập nhật quy định mới	Chậm (mua sách mới)	Tức thời (realtime)	Đồng bộ realtime
Khả năng trích dẫn nguồn	Thủ công, tốn thời gian, dễ nhầm lẫn	Tự động, chính xác	Kỹ thuật RAFT
Hỗ trợ ra quyết định	Không có	Có (Chain-of-Thought)	Khả năng suy luận LLM

4.2. Lợi ích thực tiễn và triển vọng

Về mặt kinh tế, công cụ này giúp các doanh nghiệp cảng biển và vận tải biển Việt Nam tiết kiệm được đáng kể phí bản quyền hàng năm so với việc sử dụng các phần mềm bản quyền trên thế giới hiện nay. Về mặt an toàn, khả năng trích dẫn trực tiếp từ "tài liệu chuẩn" của kỹ thuật RAFT giúp người dùng có thể an tâm tin tưởng tuyệt đối vào các chỉ dẫn của Chatbot trong các tình huống thao tác với hàng nguy hiểm hoặc ứng phó sự cố hóa chất bởi vì công nghệ Chatbot của trang Web chỉ cho phép AI tìm kiếm thông tin giới hạn trong tài liệu đã được kiểm duyệt để đưa vào database của ứng dụng.

5. Kết luận

Nghiên cứu đã xây dựng thành công công cụ tra cứu thông minh hướng dẫn kiểm soát hàng nguy hiểm tại cảng biển Việt Nam. Với việc tích hợp các quy định mới nhất từ IMDG Code 2024 và ứng dụng các công nghệ AI tiên tiến như RAG, RAFT và vector embedding 3072 chiều, hệ thống đã giải quyết triệt để bài toán về tốc độ, độ chính xác và chi phí. Công cụ không chỉ là một ứng dụng phần mềm mà còn là một "trợ lý thông minh" giúp chuẩn hóa quy trình an toàn, góp phần đưa ngành hàng hải Việt Nam tiến gần hơn tới tiêu chuẩn cảng biển xanh và thông minh trên trường quốc tế.

TÀI LIỆU THAM KHẢO

[1] Nguyễn Thanh Diệu, Mai Xuân Hương (2023), *Nghiên cứu xây dựng quy trình kiểm soát an toàn quá trình làm hàng nguy hiểm áp dụng cho tàu hóa chất Việt Nam*, Tạp chí Giao thông vận tải. Số 05/2023.

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

[2] Hazcheck Online (2025), *Dangerous Goods compliance tool for maritime industry*, https://hazcheck.com/.

[3] International Maritime Organization (2024), *International Maritime Dangerous Goods Code 2024 (Amendment 42-24)*, London.

[4] International Maritime Organization (2025), *IMO inspection data on dangerous goods shipments - 34% Deficiency Report*. Available online at: https://www.sedge.ai/post/from-48-hours-to-10-minutes-how-ai-is-transforming-maritime-dangerous-goods-operations (Accessed on Feb 12, 2026).

[5] Lewis, P., et al. (2020), *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, Advances in Neural Information Processing Systems (NeurIPS).

[6] OpenAI (2024), *Vector Embeddings technical documentation - text-embedding-3-large model.*

[7] Zhang, T., et al. (2024), *RAFT: Adapting Language Models to Domain-Specific RAG*, arXiv:2403.10131.

[8] Es, S., et al. (2023), *RAGAS: Automated Evaluation of Retrieval Augmented Generation*, arXiv:2309.15217.

[9] Chidamber, S. R., Kemerer, C. F. (1994), *A metrics suite for object oriented design*, IEEE Transactions on Software Engineering.

[10] McCabe, T. J. (1976), *A Complexity Measure*, IEEE Transactions on Software Engineering.

[11] Clarkson, P. J., et al. (2004), *Predicting Change Propagation in Complex Design*, ASME Journal of Mechanical Design.

[12] McKinsey & Company (2023), *The Future of Maritime Automation and AI in Dangerous Goods Management*.

[13] Sedge AI (2024), *From 48 Hours to 10 Minutes: How AI is Transforming Maritime Dangerous Goods Operations*, White Paper.

[14] ICHCA (2025), *IMDG Code Amendment 42-24: Summary of main changes and operational impact*.

[15] Barberi, E., et al. (2025), *Leveraging Artificial Intelligence for Real-Time Risk Detection in Ship Navigation*, Applied Sciences (MDPI), 15(21):11674.

[16] Sadiku, M. N. O., et al. (2025), *Artificial Intelligence in Maritime Industry*, International Journal of Trend in Scientific Research and Development.

[17] Chimakurthi, L. (2024), *Reliable RAG: Implementing Semantic Double-Pass Merging for Technical Content*.

[18] Asia Classification Society (2022), *Maritime RAG Benchmark: 500 expert-validated Q&A pairs for ship classification*, SSMI Conference.

[19] Reach24h (2025), *IMDG Code Amendment 42-24 Takes Effect: What Global Shippers Must Prepare for by 2026*.

[20] Dalaklis, D. (2025), *Artificial Intelligence (AI) Applications and the Shipping Industry*, TransNav Journal.

[21] Griggs, F. J. (2012), *A Human Factors Analysis and Classification System (HFACS) Examination of Commercial Vessel Accidents*, Naval Postgraduate School.

[22] Microsoft Visual Studio (2024), *Code Metrics: Coupling Between Objects (CBO) Thresholds and Maintainability Rules*.

Ngày nhận bài:	26/12/2025
Ngày nhận bản sửa:	13/02/2026
Ngày duyệt đăng:	25/03/2026