

NGHIÊN CỨU XÂY DỰNG CÔNG CỤ KIỂM TRA MỨC ĐỘ PHÙ HỢP CẤU TRÚC KHUNG CƠ SỞ DỮ LIỆU NỀN ĐỊA LÝ QUỐC GIA TỶ LỆ 1:2.000, 1:5.000, 1:10.000

TRẦN ANH TUẤN⁽¹⁾, BÙI THỊ XUÂN HỒNG⁽¹⁾, PHAN QUỐC YÊN⁽²⁾

⁽¹⁾Cục Đo đạc, Bản đồ và Thông tin địa lý Việt Nam

⁽²⁾Trường Đại học Kỹ thuật Lê Quý Đôn, Hà Nội, Việt Nam

Tóm tắt:

Ngày nay, nhu cầu lớn về sử dụng thông tin địa không gian không ngừng tăng lên trong các lĩnh vực khác nhau. Do đó, chất lượng cơ sở dữ liệu nền địa lý quốc gia là vô cùng quan trọng. Nếu chất lượng dữ liệu kém sẽ ảnh hưởng đến công tác quản lý và kết quả cuối cùng của bài toán hỗ trợ ra quyết định. Bài báo giới thiệu kỹ thuật lập trình GIS sử dụng thư viện ArcObject để xây dựng công cụ kiểm tra mức độ phù hợp của dữ liệu với mô hình cấu trúc khung cơ sở dữ liệu nền địa lý quốc gia cho nhóm tỷ lệ 1:2.000, 1:5.000, 1:10.000. Công cụ hoạt động dưới dạng Plug-ins trong phần mềm ArcMap. Công cụ được thử nghiệm kiểm tra trên tập dữ liệu nền địa lý khu vực Nam Định tỷ lệ 1:10.000. Các lỗi phần mềm trả về trên tập dữ liệu thử nghiệm được so sánh với kết quả kiểm tra bằng phương pháp thủ công. Số lượng lỗi về tên lớp, bí danh (Alias), kiểu dữ liệu của trường, miền giá trị hoàn toàn chính xác. Công cụ cho thấy tính hiệu quả rõ rệt về thời gian, độ chính xác và tính minh bạch kết quả kiểm tra.

Từ khóa: Kiểm tra chất lượng dữ liệu GIS; Lập trình GIS; Cấu trúc dữ liệu nền địa lý; ArcObject

1. Giới thiệu

Ngay từ rất sớm, khối lượng lớn dữ liệu số về địa không gian đã được định hình bằng hệ thống thông tin địa lý (GIS). Thiết kế có sự hỗ trợ của máy tính và theo đúng tiêu chuẩn đóng một vai trò quan trọng trong việc chuẩn bị dữ liệu [1]. Theo Luật Đo đạc và Bản đồ [2], dữ liệu nền địa lý là dữ liệu không gian địa lý làm cơ sở để xây dựng dữ liệu không gian địa lý khác. Cơ sở dữ liệu (CSDL) nền địa lý quốc gia là tập hợp dữ liệu nền địa lý được xây dựng trong hệ tọa độ quốc gia, hệ độ cao quốc gia theo tiêu chuẩn, quy chuẩn kỹ

thuật quốc gia, quy định kỹ thuật để sử dụng thống nhất trong cả nước [3].

Để thống nhất một chuẩn chung cho dữ liệu địa lý nhằm thuận tiện trong việc trao đổi, tổ chức tiêu chuẩn thế giới (ISO) đã ban hành bộ tiêu chuẩn ISO-19100. Việt Nam đã ban hành các Quy chuẩn kỹ thuật quốc gia về cơ sở dữ liệu nền địa lý quốc gia, các Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý quốc gia ở các nhóm tỷ lệ khác nhau, như nhóm tỷ lệ 1:2.000-1:10.000, 1:10.000-1:25.000, và nhóm 1:50.000-1:100.000 [4-8]. Trong mỗi quy định kỹ thuật, khung CSDL về chuẩn mô hình cấu trúc dữ

liệu nền địa lý được quy định chặt chẽ thông qua “Lược đồ ứng dụng”, mô tả các lớp đối tượng địa lý trong các gói. Các gói dữ liệu được định nghĩa rõ ràng về tên gói, các lớp trong gói, các trường dữ liệu mô tả lớp đối tượng, kiểu dữ liệu không gian của lớp đối tượng, miền giá trị kiểu dữ liệu thuộc tính, v.v. [4-8]. Việc kiểm tra, kiểm soát chất lượng dữ liệu cũng được quy định thành phụ lục về “Chất lượng cơ sở dữ liệu nền địa lý quốc gia” cho từng quy chuẩn. Ví dụ Phụ lục C của các Quy chuẩn Việt Nam (QCVN) [8], Phụ lục 4 đối với CSDL nền địa lý quốc gia tỉ lệ 1:10.000, 1:25.000 [6]. Các yếu tố về mức độ phù hợp với mô hình cấu trúc được quy định thông qua việc dữ liệu có tuân thủ lược đồ ứng dụng, miền giá trị, định dạng dữ liệu, và quan hệ không gian [4-8]. Xét về mức độ chi tiết lớp đối tượng, đối tượng địa lý, trường dữ liệu và Domain, QCVN 73:2023/BTNMT cho nhóm tỷ lệ 1:2.000, 1:5.000, 1:10.000 là chi tiết nhất. Kiểm tra được dữ liệu ở tỷ lệ này thì có thể bao hàm được các nhóm tỷ lệ khác. Theo Quy chuẩn, giữa 3 tỷ lệ quy định cơ bản giống nhau về lược đồ ứng dụng, chỉ khác nhau tiêu chí thu nhận dữ liệu.

Trong thực tế, công việc kiểm tra có thể được thực hiện theo phương pháp thủ công hoặc tự động bằng phần mềm [5, 9]. Nếu kiểm tra thủ công, người kiểm tra phải rà soát, đối chiếu tập dữ liệu so với quy chuẩn bằng mắt thường dưới sự hỗ trợ các phần mềm bản đồ chuyên dụng. Với hàng trăm lớp dữ liệu và hàng chục tham số chất lượng sẽ có tốn kém thời gian, nhân lực và chi phí [9]. Có một số công cụ hỗ trợ kiểm tra chất lượng dữ liệu phổ biến như ArcGIS Data Reviewer, SQUAD for QGIS, Check for Topological Errors, Check Validity, DSG Tool [10-13]. Tuy nhiên, các công cụ này chủ yếu là hỗ trợ kiểm tra mối quan hệ không gian thông qua các luật đã thiết lập. Dữ liệu thuộc tính, đặc biệt là cấu trúc khung CSDL nền địa lý theo

QCVN thì các phần mềm không thực hiện được vì cần phải xây dựng tập dữ liệu khung tham chiếu theo QCVN.

Hiện nay, CSDL nền địa lý ở nước ta được xây dựng và lưu trữ dưới định dạng GDB của phần mềm ArcGIS. Đây là phần mềm được sử dụng phổ biến trong lĩnh vực địa không gian [14]. ArcGIS hoạt động dựa trên khung công nghệ của ArcObjects [15-17]. Do đó, sử dụng ArcObject để xây dựng các công cụ bổ trợ (Add-ins) cho bộ phần mềm ArcGIS là rất phổ biến [18-20].

Vì các lý do đã tổng quan trên, xây dựng công cụ để hỗ trợ kiểm tra tự động về mức độ phù hợp của các yếu tố này với mô hình cấu trúc theo quy chuẩn Việt Nam là rất cần thiết. Nghiên cứu sử dụng thư viện ArcObject để xây dựng công cụ là phù hợp với công nghệ hiện nay.

2. Phương pháp nghiên cứu xây dựng phần mềm

2.1. Thư viện và ngôn ngữ lập trình

ArcObjects là nền tảng cốt lõi của ArcGIS, cung cấp các chức năng chính của hệ thống. Nhiều chức năng của ArcObjects có thể được truy cập và tự động hóa thông qua các phần Add-ins. Công cụ bổ trợ Add-ins trong ArcMap là một giải pháp phổ biến để tự động hóa công việc kiểm tra mức độ phù hợp của mô hình cấu trúc khung, được phát triển dựa vào ArcObjects SDK for .NET bằng ngôn ngữ C#. Công việc phát triển thường được thực hiện trong môi trường tích hợp Microsoft Visual Studio 2012 hoặc các phiên bản mới hơn [17].

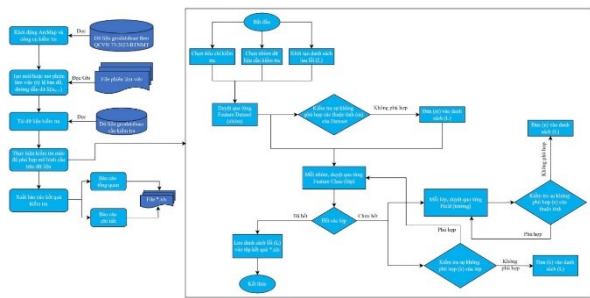
Thư viện ArcObjects SDK có nhiều thành phần. Ứng dụng phát triển công cụ này là sử dụng thư viện để tạo tất cả các thành phần cần thiết của Add-Ins, đọc và ghi dữ liệu từ tệp GDB của ArcGIS. Một số giao diện được sử dụng để thao tác đọc/ ghi dữ liệu như: không

gian làm việc có Iworkspace, IfeatureWorkspace, IworkspaceFactory; Các giao diện để làm việc với nhóm lớp (Dataset) như IenumDataset, Idataset. Giao diện để làm việc với lớp đối tượng (FeatureClass) là IFeatureClass. Ngoài ra, truy vấn ngôn ngữ SQL được sử dụng để thực hiện so sánh, truy vấn dữ liệu trong toàn bộ quá trình kiểm tra mức độ phù hợp của dữ liệu thông qua tập dữ liệu mẫu đã xây dựng trước [17].

2.2. Kiến trúc phần mềm và thuật toán

Phần mềm tổ chức dưới dạng ứng dụng được cài đặt cho máy tính để bàn. Kiến trúc phần mềm được tổ chức làm ba tầng, tầng dữ liệu là nơi lưu trữ các tệp dữ liệu GDB tiêu chuẩn, GDB cần kiểm tra, các file lưu trữ phiên làm việc và báo cáo kết quả kiểm tra. Tầng xử lý và ứng dụng hoạt động trên nền tảng ArcObject dưới dạng Plugin của ArcMap. Tại đây, người dùng thực hiện tạo phiên làm việc, tải dữ liệu cần kiểm tra, lựa chọn các luật cần kiểm tra và thực hiện kiểm tra (Hình 1).

Thuật toán kiểm tra được thực hiện dựa trên luồng logic của công việc kiểm tra. Thực hiện kiểm tra các nội dung bao gồm: tên trường, tên lớp, bí danh (Alias) của lớp, kiểu dữ liệu của trường, giá trị Null của trường, độ dài (kích thước) của trường, kiểu đối tượng không gian, giá trị mặc định của trường, phạm vi miền giá trị thuộc tính (Domain).

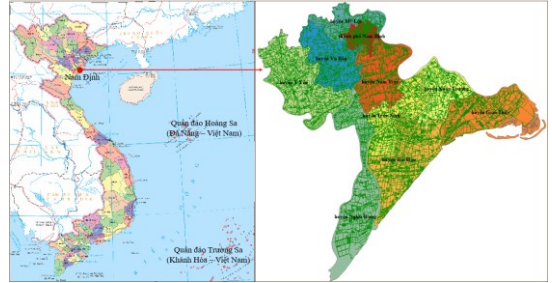


Hình 1: Kiến trúc và thuật toán kiểm tra mức độ phù hợp với mô hình cấu trúc

3. Dữ liệu và kết quả thực nghiệm

3.1. Dữ liệu

Dữ liệu thực nghiệm là cơ sở dữ liệu nền địa lý tỷ lệ 1:10.000 tỉnh Nam Định, gồm các huyện: Hải Hậu, Trực Ninh, Xuân Trường, Ý Yên, Nghĩa Hưng, Mỹ Lộc, Vụ Bản, Nam Trực, Giao Thủy, Thành phố Nam Định (Hình 2).



Hình 2: Khu vực, dữ liệu kiểm thử phần mềm

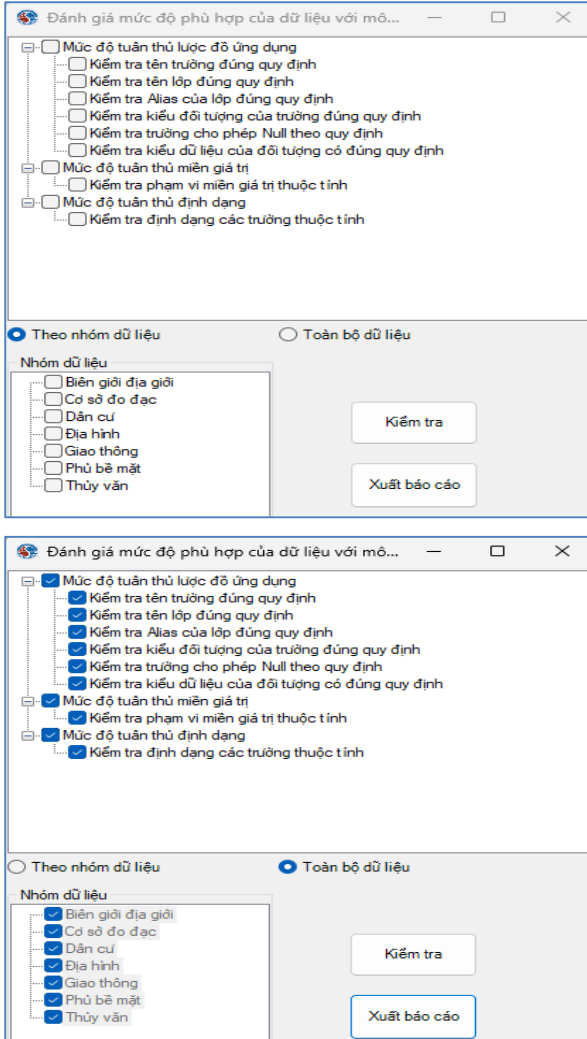
Dữ liệu được xây dựng đảm bảo theo mô hình, cấu trúc nội dung quy chuẩn Thông tư số 73:2023/BTNMT, gồm 7 nhóm dữ liệu, 168 lớp và 119 Domain. Các nhóm gồm Biên giới địa giới có 11 lớp; cơ sở đo đạc có 3 lớp; Dân cư có 45 lớp; Địa hình có 11 lớp; Giao thông có 43 lớp; Phủ bề mặt có 12 lớp; và Thủy văn có 43 lớp.

3.2. Kết quả thực nghiệm

3.2.1. Kết quả phần mềm

Để mở công cụ kiểm tra chất lượng dữ liệu, khởi động phần mềm ArcMap10.8.2, nhấp chuột trái vào nút lệnh có biểu tượng trên thanh công cụ để chạy phần mềm. Tiến hành khởi tạo phiên làm việc, tải dữ liệu cần kiểm tra, và truy cập vào menu “Chức năng/đánh giá mức độ phù hợp của dữ liệu với mô hình cấu trúc”. Giao diện công cụ xuất hiện như tám bên trái (Hình 3). Sau đó, tiến hành lựa chọn các luật về phù hợp với mô hình cấu trúc dữ liệu. Ví dụ tám bên phải sẽ chọn tất cả các luật cho toàn bộ tập dữ liệu Geodatabase. Sau đó nhấp chuột vào nút lệnh “Kiểm tra”, công cụ sẽ thực hiện tự động việc kiểm tra. Nếu kết thúc sẽ hiển thị thông báo “Kiểm tra thành công”. Tiến hành xem báo

cáo các lỗi không phù hợp của dữ liệu, nhấp chuột vào nút lệnh “Xuất báo cáo”. Kết quả và cấu trúc báo cáo tổng thể là báo cáo về từng lớp dữ liệu kiểm tra và có bao nhiêu lỗi hoặc không có lỗi. Báo cáo chi tiết sẽ chỉ ra cụ thể lỗi gì, ở đâu, v.v. đối với từng lỗi.



Hình 3: Giao diện công cụ kiểm tra mức độ phù hợp mô hình cấu trúc của dữ liệu

3.2.2. Kết quả kiểm thử phần mềm

Thực hiện kiểm tra, chạy thử công cụ với tập dữ liệu nền địa lý tỉnh Nam Định với các tham số được cấu hình như tám bên phải của Hình 3. Kết quả báo cáo được xuất ra thành hai tệp *.xls. Trích xuất một phần báo cáo tổng quan và báo cáo chi tiết lỗi như Hình 4 và Hình 5.

Thông kê từ bảng báo cáo lỗi tổng quan, nhóm nghiên cứu tổng hợp thống kê các lỗi theo 10 nhóm lỗi gồm: (1) kiểm tra tên lớp có 03 nhóm chủ đề với tổng số 16 lớp bị lỗi; (2) kiểm tra bí danh của lớp có 36 lớp sai bí danh; (3) kiểm tra kiểu đối tượng không gian có 4 nhóm với tổng 36 lớp đối tượng không đúng kiểu; (4) kiểm tra tên trường có tổng số 34 trường có tên sai so với quy định; (5) kiểm tra kiểu dữ liệu của trường có 36 trường sai kiểu; (6) kiểm tra giá trị Null của trường có 47 trường; (7) kiểm tra độ dài của trường có 39 trường lỗi; (8) kiểm tra giá trị mặc định của trường có 36 trường; (9) kiểm tra phạm vi miền giá trị thuộc tính có 12 Domain lỗi (Bảng 1).

Nhìn chung, các sai số dẫn xuất sẽ được cộng dồn và có liên quan với nhau. Sai tên lớp sẽ dẫn tới sai kiểu đối tượng, những lớp không có tên trong cơ sở dữ liệu tham chiếu công cụ sẽ không kiểm tra được kiểu đối tượng là gì. Ngoài các lớp bị sai tên, các lớp đúng tên mà có tên trường dữ liệu sai cũng sẽ dẫn tới sai các luật khác liên quan đến trường dữ liệu như kiểu dữ liệu, giá trị Null, v.v. (Hình 4, 5).

Với số lỗi được thống kê trong bảng tổng quan, phần mềm cho phép xuất ra.

Bảng thống kê lỗi chi tiết, chỉ đến các nhóm chủ đề, lớp đối tượng, trường dữ liệu của lớp, Domain cụ thể mà không đúng quy định. Ví dụ, Hình 5 trích xuất một số các trường hợp lỗi cụ thể: (1) các lỗi liên quan đến kiểm tra tên trường, ví dụ lớp BeMatCongTrinh và BeMatKhuDanCu đều có 7 lỗi, và liệt kê các trường lỗi kèm theo; (2) kiểm tra tên lớp 16 lớp lỗi được chỉ định cụ thể, ví dụ “Cột điện” trong cơ sở dữ liệu khác với “Cột điện dạng điểm” trong dữ liệu tiêu chuẩn; (3) kiểm tra bí danh của lớp cũng chỉ đến từng trường hợp sai cụ thể, ví dụ như lớp “CotDien” bị lỗi vì trong CSDL tham chiếu là CotDienS, CotDienP, v.v.

STT	Nội dung đánh giá	Dữ liệu đánh giá	Tổng số đối tượng	Số đối tượng đánh giá	Số đối tượng không thỏa mãn	Phần trăm thỏa mãn
161	160 Kiểm tra tên trường đúng quy định	Lớp-Ngầm ô tô qua được dạng vùng	10	10	0	100
162	161 Kiểm tra tên trường đúng quy định	Lớp-Cổng giao thông dạng đường	7	7	0	100
163	162 Kiểm tra tên trường đúng quy định	Lớp-Các đối tượng mặt đường bộ dạng vùng	8	8	0	100
164	163 Kiểm tra tên trường đúng quy định	Lớp-Các đối tượng đường bộ khác dạng vùng	6	6	0	100
165	164 Kiểm tra tên trường đúng quy định	Lớp-Công trình giao thông đường sắt dạng vùng	7	7	0	100
166	165 Kiểm tra tên trường đúng quy định	Lớp-Ga cáp treo dạng vùng	7	7	0	100
167	166 Kiểm tra tên trường đúng quy định	Lớp-Bến thủy nội địa dạng vùng	8	8	0	100
168	167 Kiểm tra tên trường đúng quy định	Lớp-Cầu tàu dạng vùng	8	8	1	87
169	168 Kiểm tra tên trường đúng quy định	Lớp-Nhóm ô tô dạng vùng	6	6	0	100
170	169 Kiểm tra tên lớp đúng quy định	Dataset-BienGioiDiaGioi	11	11	0	100
171	170 Kiểm tra tên lớp đúng quy định	Dataset-CoSoDoDac	3	3	0	100
172	171 Kiểm tra tên lớp đúng quy định	Dataset-DanCu	45	45	6	86
173	172 Kiểm tra tên lớp đúng quy định	Dataset-DiaHinh	11	11	0	100
174	173 Kiểm tra tên lớp đúng quy định	Dataset-GiaoThong	43	43	6	86
175	174 Kiểm tra tên lớp đúng quy định	Dataset-PhuBeMat	12	12	4	66
176	175 Kiểm tra tên lớp đúng quy định	Dataset-ThuyVan	43	43	0	100
177	176 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-BienGioiDiaGioi	11	11	0	100
178	177 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-CoSoDoDac	3	3	0	100
179	178 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-DanCu	45	45	14	68
180	179 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-DiaHinh	11	11	0	100
181	180 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-GiaoThong	43	43	15	65
182	181 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-PhuBeMat	12	12	4	66
183	182 Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-ThuyVan	43	43	3	93
184	183 Kiểm tra kiểu đối tượng trường đúng quy định	Lớp-Đường biên giới quốc gia trên đất liền	9	9	0	100

Hình 4: Trích báo cáo lỗi tổng quan về mô hình cấu trúc dữ liệu

Nội dung đánh giá	Dữ liệu đánh giá	Lỗi chi tiết	Nội dung đánh giá	Dữ liệu đánh giá	Lỗi chi tiết	
86	Kiểm tra tên trường đúng quy định	Lớp-Cây hàng năm	303	TÊN LỚP	Không có lỗi	
87	Kiểm tra tên trường đúng quy định	Lớp-Cây lâu năm	304	Kiểm tra tên lớp đúng quy định	Dataset-BienGioiDiaGioi	Không có lỗi
88	Kiểm tra tên trường đúng quy định	Lớp-Bề mặt công trình	305	Kiểm tra tên lớp đúng quy định	Dataset-CoSoDoDac	Không có lỗi
89		Số lỗi =7 lỗi	306	Kiểm tra tên lớp đúng quy định	Dataset-DanCu	Số lỗi =6 lỗi
90		maNhanDang	307			Cột điện
91		phienBan	308			Cột điện dạng vùng
92		ngayPhienBan	309			Công trình phụ trợ dạng đường
93		giaTriDoChinhXacMatPhang	310			Công trình phụ trợ dạng vùng
94		nguyenNhanThayDoi	311			Cột điện dạng điểm
95		maDoiTuong	312			Nhà dạng đường
96		thucVat	313	Kiểm tra tên lớp đúng quy định	Dataset-DiaHinh	Không có lỗi
97	Kiểm tra tên trường đúng quy định	Lớp-Bề mặt khu dân cư	314	Kiểm tra tên lớp đúng quy định	Dataset-GiaoThong	Số lỗi =6 lỗi
98		Số lỗi =7 lỗi	315			Taluy đường giao thông
99		maNhanDang	316			Đường ra vào của tàu thuyền dạng đường
100		phienBan	317			Đường ra vào của tàu thuyền dạng vùng
101		ngayPhienBan	318			Taluy đường giao thông
102		giaTriDoChinhXacMatPhang	319			Công trình giao thông đường sắt dạng vùng
103		nguyenNhanThayDoi	320			Công trình giao thông đường bộ dạng vùng
104		maDoiTuong				
105		thucVat				
834		TÊN LỚP	2652			TÊN DOMAIN
835	Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-BienGioiDiaGioi	2653	Kiểm tra phạm vi miền giá trị thuộc	Geodatabase	Số lỗi =12 lỗi
836	Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-CoSoDoDac	2654			DoiTuongRanhGioi
837	Kiểm tra bi danh(alias) lớp đúng quy định	Dataset-DanCu	2655			LoaiTramDinhViVeTinh
838		Số lỗi =14 lỗi	2656			LoaiOngDan
839		CotDien	2657			LoaiTrangThaiNuocMat
840		CongTrinhVanHoaP	2658			MauSac
841		CongTrinhVanHoaS	2659			LoaiNguonNuoc
842		CongTrinhThuongMaiDichVuP	2660			LoaiCangHangKhong
843		CongTrinhThuongMaiDichVuS	2661			LoaiRanhGioiNuocMatQuyUoc
844		CongTrinhTonGiaoTinNguongP	2662			HuongBaoHieu
845		CongTrinhXuLyChatThaiP	2663			HinhDang
846		CongTrinhTonGiaoTinNguongS	2664			TrangThaiXuatLo
847		CongTrinhXuLyChatThaiS	2665			LoaiMoc
848		CotDienS				
849		CongTrinhPhuTroC				
850		CongTrinhPhuTroS				
851		CotDienP				
		NhaC				

Hình 5: Trích một số báo cáo lỗi chi tiết về mô hình cấu trúc

Bảng 1: Trích xuất thống kê lỗi kiểm tra

STT	Nội dung kiểm tra	Dataset Dữ liệu kiểm tra	Tổng số đối tượng	Số đối tượng lỗi	Phần trăm đúng
Số đối tượng là lớp					
1	Kiểm tra tên lớp	BienGioiDiaGioi	11	0	100
		CoSoDoDac	3	0	100
		DanCu	45	6	86
		DiaHinh	11	0	100
		GiaoThong	43	6	86
		PhuBeMat	12	4	66

2	Kiểm tra bi danh của lớp (Alias)	ThuyVan	43	0	100
		BienGioiDiaGioi	11	0	100
		CoSoDoDac	3	0	100
		DanCu	45	14	68
		DiaHinh	11	0	100
		GiaoThong	43	15	65
		PhuBeMat	12	4	66
3	Kiểm tra kiểu đối tượng	ThuyVan	43	3	93
		CoSoDoDac	3	0	100

	không gian	DanCu	45	14	68
		DiaHinh	11	0	100
		GiaoThong	43	15	65
		PhuBeMat	12	4	66
		ThuyVan	43	3	93
<i>Số đối tượng là trường</i>					
4	Kiểm tra tên trường	BienGioiDiaGioi	93	0	100
		CoSoDoDac	27	0	100
		DanCu	341	9	97,4
		DiaHinh	83	0	100
		GiaoThong	356	11	96,9
		PhuBeMat	87	14	83,9
		ThuyVan	334	0	100
5	Kiểm tra kiểu dữ liệu của trường	BienGioiDiaGioi	93	0	100
		CoSoDoDac	27	0	100
		DanCu	341	11	96,7
		DiaHinh	83	0	100
		GiaoThong	356	11	96,9
		PhuBeMat	87	14	83,9
		ThuyVan	334	0	100
6	Kiểm tra giá trị Null của trường	BienGioiDiaGioi	93	0	100
		CoSoDoDac	27	2	92,6
		DanCu	341	13	96,2
		DiaHinh	83	1	98,8
		GiaoThong	356	13	96,3
		PhuBeMat	87	15	82,7
		ThuyVan	334	3	99,1
7	Kiểm tra độ dài của trường đúng quy định	BienGioiDiaGioi	93	0	100
		CoSoDoDac	27	0	100
		DanCu	341	11	97,4
		DiaHinh	83	0	100
		GiaoThong	356	11	96,9
		PhuBeMat	87	14	83,9
		ThuyVan	334	0	100
8	Kiểm tra giá trị mặc định của trường đúng quy định	BienGioiDiaGioi	93	0	100
		CoSoDoDac	27	9	97,4
		DanCu	341	0	100
		DiaHinh	83	11	96,9
		GiaoThong	356	14	83,9
		PhuBeMat	87	0	100
		ThuyVan	334	0	100
<i>Số đối tượng là Domain</i>					

9	Kiểm tra phạm vi miền giá trị thuộc tính	Geodatabase	122	12	90
---	--	-------------	-----	----	----

3.2.3. Thảo luận

Các kết quả kiểm tra ở trên được kiểm tra và so sánh thông qua phương pháp kiểm thủ công dưới sự hỗ trợ của phần mềm ArcCatalog trong ArcGIS. Sử dụng phần mềm ArcCatalog mở Geodatabase Nam Định. Sử dụng Phụ lục A về “danh mục đối tượng địa lý cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:2.000, 1:5.000, 1:10.000” của QCVN 73:2023/BTNMT để đối chiếu về tên Feature Dataset; tên, bí danh, kiểu dữ liệu không gian của các Feature Class trong mỗi Feature Dataset; tên, kiểu dữ liệu, domain của các trường trong mỗi Feature Class. Dựa trên kết quả kiểm tra của hai phương pháp cho thấy: (1) về độ chính xác, hai phương pháp cho kết quả giống nhau cả về số lượng và đối tượng lỗi. Số lượng có 16 lớp bị lỗi tên, 36 lớp sai bí danh, 34 trường có tên sai so với quy định, 36 trường sai kiểu, 39 trường lỗi về độ dài, miền giá trị thuộc tính có 12 Domain lỗi trên toàn bộ GDB. (2) về thời gian, phương pháp thủ công tốn 02 ngày để kiểm tra, thống kê. Trong khi đó, phương pháp tự động bằng phần mềm chỉ tốn 2 phút cho tất cả các thao tác. (3) hiệu quả về nhân lực, trong quá trình kiểm tra bằng thủ công, phải sử dụng tối thiểu 2 người thực hiện để kiểm tra chéo, tránh các sai sót do chủ quan, trình độ và kinh nghiệm của người kiểm tra. Ngược lại, phần mềm chỉ cần 1 người thực hiện, biết cách sử dụng phần mềm là có thể thực hiện được thao tác kiểm tra một cách dễ dàng.

4. Kết luận

Bài báo đã tiến hành nghiên cứu đưa ra quy trình, thuật toán kiểm tra mức độ phù hợp của dữ liệu với khung CSDL nền địa lý tỷ lệ 1:2.000, 1:5.000, 1:10.000. Từ quy trình đó,

bài báo đã thực nghiệm phương pháp lập trình GIS bằng thư viện ArcObject để xây dựng phần mềm hỗ trợ kiểm tra theo phương pháp tự động. Kết quả phần mềm được đánh giá thông qua kiểm tra bộ CSDL nền địa lý tỉnh Nam Định tỷ lệ 1:10.000. Qua kết quả cho thấy phần mềm mang lại hiệu quả cao, thời gian xử lý nhanh, kết quả kiểm tra hoàn toàn chính xác, khách quan và minh bạch, không phụ thuộc vào yếu tố chủ quan của con người.

Tài liệu tham khảo

- [1]. Rasheed, M.U., S.R. Ahmad, and R.J.J.o.G.I.S. Saleem, Building a Geodatabase for Parcel and Cadaster Mapping and Add-Ins Development: A Case Study for Defense Housing Authority (DHA), Lahore, Pakistan. 2015. 7(6): p. 588-597.
- [2]. Quốc Hội, Luật đo đạc và bản đồ. Luật số: 27/2018/QH14 ngày 14 tháng 6 năm 2018, 2018.
- [3]. Chính Phủ, Nghị định quy định chi tiết một số điều của Luật đo đạc và bản đồ. số: 27/2019/NĐ-CP, ngày 13/3/2019, 2019.
- [4]. Bộ Tài nguyên và Môi trường, Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:2.000 và 1:5.000. Thông tư số 23/2019/TT-BTNMT ngày 25 tháng 12 năm 2019, 2019.
- [5]. Bộ Tài nguyên và Môi trường, QCVN 42: 2020/BTNMT, Quy chuẩn kỹ thuật quốc gia về chuẩn thông tin địa lý cơ sở. Thông tư số 06/2020/TT-BTNMT ngày 31 tháng 8 năm 2020, 2020.
- [6]. Bộ Tài nguyên và Môi trường, Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:10.000, 1:25.000. Thông tư Số 15/2020/TT-BTNMT ngày 30 tháng 11 năm 2020, 2020.
- [7]. Bộ Tài nguyên và Môi trường, QCVN 71:2022/BTNMT, Quy chuẩn kỹ thuật quốc gia về cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:50.000, 1:100.000. Thông tư số 07/2022/TT-BTNMT ngày 30 tháng 6 năm 2022, 2022.
- [8]. Bộ Tài nguyên và Môi trường, QCVN 73:2023/BTNMT, Quy chuẩn kỹ thuật quốc gia về cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:2.000, 1:5.000, 1:10.000. Thông tư 08/2023/TT-BTNMT ngày 31/7/2023, 2023.
- [9]. Quoc Yen Phan, Anh Tuan Tran, and T.X.H. Bui, Research on spatial consistency checking according to National Technical Regulation on basic Geographic Information Section on Special Construction Engineering, 2023. Vol. 06, No. 02 (Dec. 2023).
- [10]. Kukulska, A., et al., Methodology of evaluation and correction of geometric data topology in QGIS software. Acta Scientiarum Polonorum: Formatio Circumiectus, 2018. 17(1).
- [11]. Sehra, S.S., J. Singh, and H.S. Rai, Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. Future Internet, 2017. 9(2): p. 15.
- [12]. Susetyo, D.B. and A.P. Perdana. Automation of Indonesian Topographic Data Quality Control Using Data Reviewer. in International Conference on Technology, Innovation, and Society (ICTIS). 2016.
- [13]. Hendricks, M., J. Athey, and A. Macpherson, QA/QC of GeMS DataCase Study: The Alaska DGGs Geologic Mapping System. 2022.
- [14]. Booth, B. and A. Mitchell, Getting started with ArcGIS. 2001, Esri Redlands, CA, USA.

- [15]. Eldrandaly, K.J.I.A.J.I.T., Developing a GIS-based MCE site selection tool in ArcGIS using COM technology. 2013. 10(3): p. 276-282.
- [16]. Li, Z., et al., Component GIS, ArcObjects and ArcGIS Server. 2020: p. 103-117.
- [17]. Amirian, P., Beginning ArcGIS for desktop development using .NET. 2013: John Wiley & Sons.
- [18]. Hou, J., H. Yuan, and S.J.A.J.o.G. Sun, Design and application of spatial accessibility plugin based on ArcGIS engine. 2020. 13: p. 1-8.
- [19]. Li, B., et al. Development of an aquaculture suitability assessment system of sea areas based on plug-in technology. in 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). 2020. IEEE.
- [20]. Qin, Y., et al., Automatic optimization model of transmission line based on GIS and genetic algorithm. 2023. 17: p. 100266.

Summary

Research on building a tool to check the suitability of the national fundamental geographic database framework structure at scales of 1:2,000, 1:5,000, 1:10,000

Anh Tuan Tran, Thi Xuan Hong Bui

Department of Survey, Mapping and Geographic Information Viet

Quoc Yen Phan, Le Quy Don Technical University, Ha Noi, Viet Nam

Nowadays, the vast demand for geospatial information is constantly increasing in various fields. Therefore, the quality of the national fundamental geographic database is critical. If the data has many errors and is of inadequate quality, it will affect management and the final result of the decision support problem. The paper introduces GIS programming techniques using the ArcObject library to build a tool to support checking the level of data conformity with the national fundamental geographic database framework at scales of 1:2,000, 1:5,000, 1:10,000. The tool works as a Plug-in in ArcMap software. The tool was tested on a geographic data set of the Nam Dinh area at a scale of 1:10,000. The software errors returned on the test dataset were compared with the results of manual checking. The number of errors in class names, aliases, field types, and value domains was completely accurate. The tool shows apparent efficiency in terms of time, accuracy and transparency of test results.

Keywords: GIS data quality check; GIS programming; Geographic database structure; ArcObject