

**PORÓWNANIE WYBRANYCH ALGORYTMÓW
UPRASZCZANIA LINII NA PRZYKŁADZIE
REPREZENTATYWNEGO OBSZARU TESTOWEGO***

COMPARISON OF SELECTED SIMPLIFICATION
ALGORITHMS ON THE EXAMPLE
OF A REPRESENTATIVE TEST AREA

Krystian Koziol

Katedra Geomatyki, Wydział Geodezji Górniczej i Inżynierii Środowiska
Akademia Górniczo-Hutnicza w Krakowie

Słowa kluczowe: generalizacja, upraszczanie, norma rozpoznawalności rysunku, algorytm Langa, algorytm Douglasa-Peuckera, algorytm Reumanna-Witkama, metoda Chrobaka
Keywords: generalisation, simplification, standard of drawing recognition, Lang algorithm, Douglas-Peucker algorithm, Reumann-Witkam algorithm, Chrobak's method

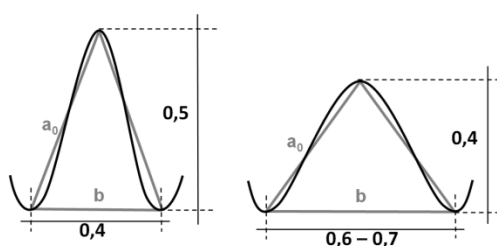
Wstęp

Modelowanie kartograficzne wymaga zastosowania uogólnienia informacji już na poziomie wyboru obiektów, będących przedmiotem tego modelowania. W następstwie abstrahowania informacji, powstały podzbiór obiektów stanowiących wypełnienie modelu podlegać będzie między innymi procesowi upraszczania. Rozwój technologii informatycznych oraz pojawienie się systemów GIS w dużej mierze przyczyniły się do powstania licznych algorytmów rozwiązujących problem upraszczania. Badania w zakresie modelowania kartograficznego wskazują na problem zanikania charakterystycznych cech geograficznych, w wyniku stosowania algorytmów upraszczania. Jednakże pojęcie modelowania jest bardzo szerokim procesem, obejmującym nie tylko problemy generalizacji, ale także samej logicznej konstrukcji zapisu obiektu. Struktura logiczna obiektu, będąca elementem modelu pojęciowego, już na wstępie zadecyduje o możliwości zastosowania lub niezastosowania pewnych funkcji i procedur oraz poprawności ich wyniku. Modelowanie informacji geograficznej wymaga zdefiniowania funkcji przetwarzających informację do wybranego poziomu uogólnienia, określonego przez przeznaczenie i docelową skalę wizualizacji. Decyduje to o braku obiektywizmu ze względu na dowolność treści (dowolne przeznaczenie). Określenie modelowania jako obiektywnego będzie możliwe, gdy obiektami modelowanymi będzie treść georeferencyjna

* Praca wykonana w ramach badań statutowych nr 11.11.150.006 – Katedry Geomatyki, WGGiIŚ, AGH Kraków.

(topograficzna). Istniejące algorytmy upraszczania zakładają wartość pewnej tolerancji liniowej, katowej lub powierzchniowej, a więc powstający uproszczony model musi zakładać jakąś wartość tego parametru. Wprowadzenie normy, jako elementu decydującego o tolerancji, jest rozwiązaniem niejednoznaczności przyjętej tolerancji. Jednocześnie obecne algorytmy nazywane algorytmami upraszczania skupiają się przede wszystkim na odrzuceniu wierzchołków z krzywej pierwotnej. Pokazanie tej samej krzywej w różnych skalach wymusza wykorzystanie innych wierzchołków w reprezentacji cyfrowej.

Wartość rozpoznawalności rysunku jako norma w procesie upraszczania



Rys. 1. Wartości rozpoznawalności rysunku zdefiniowana przez Saliszczewa i wynikający z tej definicji trójkąt elementarny określony przez Chrobaka

$$a_0 = 0,5 \text{ m dla długości podstaw trójkąta } b \in (0,5-0,7) \text{ mm} \quad (1)$$

oraz

$$a_0 \geq 0,5 \text{ m dla długości podstaw trójkąta } b \in (0,4-0,5) \text{ mm} \quad (2)$$

Badania Chrobaka (2009) doprowadziły do uzyskania następujących wartości norm:

- 1) normę rozpoznawalności najkrótszego odcinka linii rysunku o szerokości 0,1 mm definiuje warunek:
 - dla mapy analogowej $a_0 = 0,5$ [mm],
 - dla mapy cyfrowej $a_{01} = 0,5$ [mm] * M,
 - gdzie: a_0, a_{01} – ramiona trójkąta, M – mianownik skali mapy,
- 2) normę rozpoznawalności kształtu linii rysunku o szerokości 0,1 mm definiuje trójkąt elementarny dla map:
 - analogowych, gdy zachowany jest warunek:
 - $a_0 = 0,5$ [mm] dla długości podstaw $b \in (0,5-0,7)$ mm,
 - $a_0 \geq 0,5$ [mm] dla długości podstaw trójkąta, $b \in (0,4-0,5)$ mm,
 - cyfrowych, gdy zachowany jest warunek:
 - $\varepsilon_{01} = 0,5$ [mm]*M i długości podstawy trójkąta, $b \in (0,5 \text{ mm} * M - 0,7 \text{ mm} * M)$,
 - $\varepsilon_{02} = 0,5$ [mm]*M i długości podstawy trójkąta, $b \in (0,4 \text{ mm} * M - 0,5 \text{ mm} * M)$,
 - gdzie: $\varepsilon_{01}, \varepsilon_{02}$ – nazwiemy miarą rozpoznawalności,
- 3) w algorytmach generalizacji trójkąt elementarny eliminuje ustalany przez operatora parametr tolerancji; eliminacja z procesu upraszczania parametru pozwala jednoznacznie określić jego wynik (Chrobak 2010).

Badania Chrobaka (1999, 2007) związane z wartością rozpoznawalności (rys. 1) doprowadziły do zdefiniowania wartości, którą należy uznać za normę w upraszczaniu – cytując za autorem badań.

Ustalona długość ramienia a_0 i wymiary podstaw b (rys. 1) określone przez Saliszczewa (1998), to minimalne wymiary linii rysunku w trójkącie, który nazwano elementarnym (Chrobak, 1999). W trójkącie tym, związek minimalnego wymiaru krótszego ramienia a_0 i podstawy b , mają postać:

Metody badawcze oraz obiekt testowy

Do porównania wybrano trzy algorytmy posiadające swoje implementacje w oprogramowaniu GIS, a więc z punktu użytkownika najczęściej wykorzystywane oraz algorytm Chrobaka. Drugim czynnikiem decydującym o wyborze algorytmów była możliwość wprowadzenia tolerancji liniowej. Wybranymi algorytmami są:

Algorytm Langa (1969) należący do grupy procedur warunkowego rozszerzonego przetwarzania lokalnego. W grupie tej algorytmy wymagają podania oprócz tolerancji liniowych i kątowych, także określenia w badanej linii liczby punktów podlegających grupowaniu.

Algorytm Reumanna-Witkama (1974) należący do grupy procedur bezwarunkowego rozszerzonego przetwarzania lokalnego, badający nie tylko bezpośrednio sąsiadujące współrzędne i oceniający grupy linii.

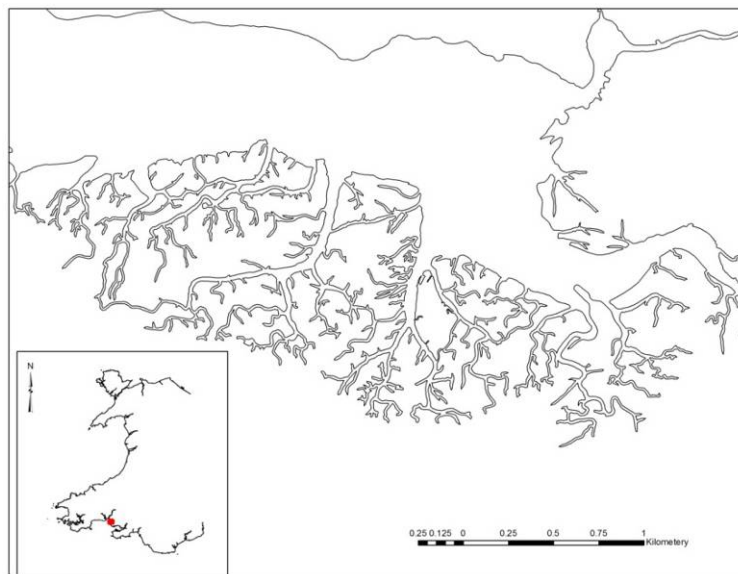
Algorytm Douglasa-Peuckera (1973) należący do grupy procedur globalnych, rozważający całą linię lub jej wyszczególniony fragment, wybierający punkty ekstremalne metodą iteracyjną, uwzględniający odległość strzałki od cięciwy w segmencie. Algorytm Douglasa-Peuckera jest jednym z najczęściej wykorzystywanych algorytmów, a jego implementacja istnieje w większości oprogramowania typu GIS i CAD.

Metoda Chrobaka (1999) jest ostatnim algorytmem, w którym zastosowanie mają zdefiniowane wcześniej normy, a eliminacja punktów odbywa się na podstawie trójkąta elementarnego. Powoduje to, że w algorytmie tym nie jest podawana wartość tolerancji, a jedynie wartość mianownika skali. Punkty ekstremalne wyznaczone są iteracyjnie z uwzględnieniem trójkąta elementarnego.

Jako obszar testowy wybrano fragment wybrzeża Walii (rys. 2) zapisany w postaci wektorowej, udostępniony przez brytyjską służbę kartograficzną. Wybrany fragment wybrzeża składał się z 1604 obiektów liniowych, posiadających łącznie 192 538 wierzchołków. Minimalna odległość pomiędzy sąsiadującymi wierzchołkami wyniosła 0,1 m, natomiast średnia odległość między punktami to ok. 10 m. Najwięcej odcinków (ponad 70%) znajdowało się w przedziale od 0,1 m do 10 m. Wysoki poziom szczegółowości danych (rys. 3) pozwolił na wykonanie upraszczania w szerokim zakresie skal.



Rys. 2. Fragment wybrzeża Walii (oznaczony prostokątem) wybrany do wykonania testu (źródło podkładu: Google Earth)



Rys. 3. a – dane pierwotne, b – poziom szczegółowości danych pierwotnych (zaznaczony kwadrat w powiększeniu)

Metoda badawcza polegała na wykonaniu upraszczania wybranymi algorytmami warstwy wektorowej, z zastosowaniem wartości tolerancji liniowej, określonej przez normę zdefiniowaną przez Chrobaka. Upraszczenie linii wybrzeża wykonano dla skal w zakresie od 1:2000 do 1:500 000, jednakże rozstęp skal był różny dla poszczególnych przedziałów mianownika skali (tab. 1). Dodatkowo wykonano upraszczenie dla skali 1:750 000. W wyniku przyjętej metody uzyskano 34 wyniki upraszczania dla zadanych wybranych algorytmów.

Do wykonania upraszczania algorytmami: Langa, Reumanna-Witkama oraz Douglasa-Peuckera wykorzystano oprogramowanie GRASS GIS v.6.4.0, będące jednym z grupy produktów typu Open Source. Natomiast algorytm Chrobaka został zaimplementowany w oprogramowaniu Microstation V8 XM Edition jako dodatkowa aplikacja o nazwie GenMap v.4. Dane oryginalne zapisane były w formacie *shape file* (ESRI) i zostały zaimportowane do systemu GRASS, a następnie wyeksportowane do formatu dxf. Wykonano kontrolę danych w formatach ostatecznych tj. grass i dxf, która nie wykazała żadnych różnic pomiędzy nimi. Następnie przystąpiono do wykonania upraszczania.

Tabela 1. Zestawienie mianowników skal oraz rozpiętości skal

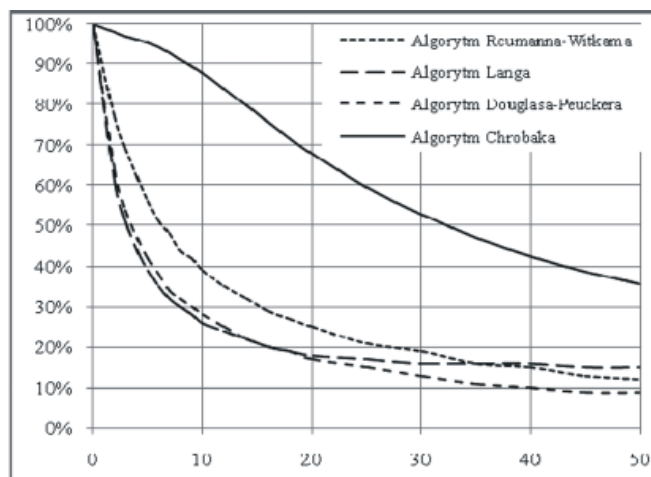
Mianowniki skal		Rozpiętość skal	Mianowniki skal dla których wykonano upraszczenie
od	do		
2000	9000	1000	2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000
10 000	18 000	2000	10 000, 12 000, 14 000, 16 000, 18 000
20 000	45 000	5000	20 000, 25 000, 30 000, 35 000, 40 000, 45 000
50 000	90 000	10 000	50 000, 60 000, 70 000, 80 000, 90 000
100 000	50 0000	50 000	100 000, 150 000, 200 000, 250 000, 300 000, 350 000, 400 000, 450 000, 500 000

Wyniki

W tabeli 2 zestawiono uzyskane wyniki upraszczania w postaci: liczby wierzchołków po procesie oraz procentowego udziału wierzchołków pozostałych względem liczby wierzchołków w krzywej pierwotnej. Tabela posłużyła do konstrukcji wykresów (rys. 4 i 5), na których w postaci graficznej przedstawiono zachowanie się poszczególnych algorytmów względem siebie. Należy zaznaczyć fakt, iż algorytm Langa działał poprawnie do skali 1:50 000, następnie uzyskiwane wyniki były bardzo zbliżone do siebie lub były wręcz identyczne jak poprzednie. Dlatego wyniki algorytmu Langa zostały usunięte z wykresu (rys. 5) przedstawiającego zachowanie się algorytmów od skali 1:50 000 do 1:750 000 i nie były analizowane.

Komentując uzyskane wyniki należy zaznaczyć fakt podobnego zachowania się algorytmu Reumanna-Witkama oraz Douglasa-Peuckera, szczególnie w zakresie skal powyżej 1: 50 000 (rys. 4). Algorytmy Douglasa-Peuckera oraz Langa w skalach do 1:10 000 wykazują największe zmiany w zakresie liczby wierzchołków (rys. 4). Jeżeli zastanowić się nad sensem upraszczania, to im większa liczba wierzchołków usuniętych, tym mniejsza się prawidłowe odwzorowanie informacji geograficznej. Algorytm Chrobaka charakteryzuje bardzo

Rys. 4. Wykres procentowego udziału wierzchołków po upraszczeniu w zakresie skal od 1:2000 do 1:50 000



Rys. 5. Wykres procentowego udziału wierzchołków po upraszczeniu w zakresie skal od 1:50 000 do 1:750 000

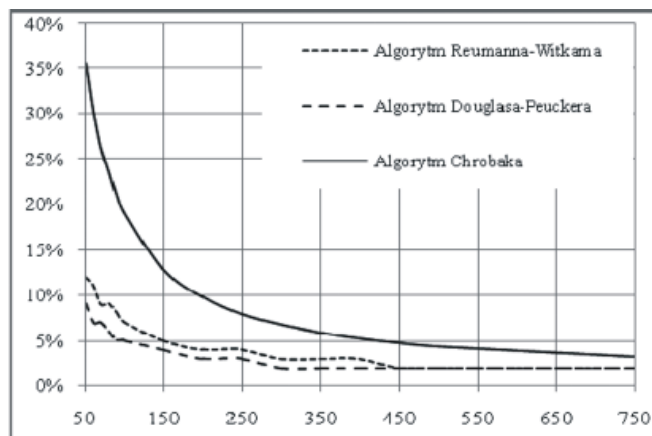
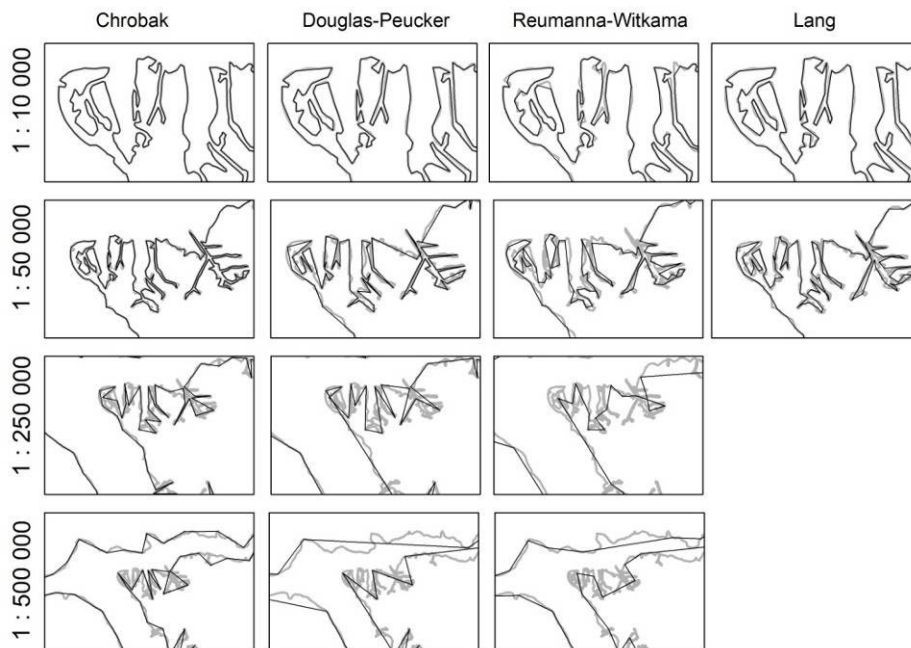


Tabela 2. Zestawianie liczby wierzchołków pozostałych w wyniku upraszczania oraz ich udział procentowy

Mianownik skali	Liczba wierzchołków w oryginale	Algorytmy							
		Reumanna-Witkama		Langa		Douglasa-Peuckera		Chrobaka	
		po uproszczeniu	[%]	po uproszczeniu	[%]	po uproszczeniu	[%]	po uproszczeniu	[%]
2000	192 538	149 598	77	118 783	61	124 515	64	188 423	98
3000		133 380	69	99 269	51	105 015	54	186 533	97
4000		119 507	62	85 946	44	91 784	47	184 890	96
5000		108 548	56	76 117	39	81 780	42	182 981	95
6000		99 993	51	68 513	35	73 843	38	180 850	94
7000		92 767	48	62 663	32	67 370	34	178 207	93
8000		86 441	44	57 870	30	62 239	32	175 236	91
9000		81 132	42	54 074	28	57 773	30	172 085	89
10 000		76 472	39	50 978	26	54 111	28	168 632	88
12 000		68 927	35	46 247	24	48 309	25	161 240	84
14 000		62 642	32	42 797	22	43 569	22	153 329	80
16 000		57 299	29	40 150	20	39 753	20	145 456	76
18 000		52 923	27	38 129	19	36 701	19	137 688	72
20 000		49 243	25	36 570	18	34 063	17	130 402	68
25 000		42 190	21	33 950	17	29 019	15	114 535	59
30 000		36 595	19	32 426	16	25 410	13	101 448	53
35 000		32 695	16	31 470	16	22 748	11	90 635	47
40 000		29 450	15	30 869	16	20 575	10	81 734	42
45 000		26 813	13	30 488	15	18 891	9	74 469	39
50 000		24 788	12	30 230	15	17 440	9	68 285	35
60 000		21 455	11	29 932	15	15 241	7	58 323	30
70 000		19 088	9	29 779	15	13 602	7	50 820	26
80 000		17 339	9	29 706	15	12 308	6	45 099	23
90 000		15 998	8	29 659	15	11 258	5	40 388	21
100 000		14 686	7	29 636	15	10 432	5	36 563	19
150 000		10 968	5	29 583	15	7968	4	24 734	13
200 000		8981	4	29 579	15	6639	3	18 814	10
250 000		7797	4	29 578	15	5886	3	15 240	8
300 000		6997	3	29 577	15	5401	2	12 909	7
350 000		6463	3	29 577	15	5078	2	11 219	6
400 000	6005	3	29 577	15	4785	2	10 042	5	
450 000	5705	2	29 577	15	4560	2	9142	5	
500 000	5372	2	29 577	15	4382	2	8350	4	
750 000	4546	2	29 577	15	3874	2	6227	3	

mała, w porównaniu do innych algorytmów, liczba odrzuconych wierzchołków (rys. 4 i 5). W tym podejściu ciekawe jest zachowanie się algorytmu Chrobaka, który z jednej strony eliminuje wierzchołki, ale z drugiej dodaje inne niezbędne do prawidłowej reprezentacji graficznej. Dodawanie, a raczej zastępowanie wierzchołów innymi, nie jest niczym nowym w kartografii. Przywołując generalizację pierwotną, otrzymamy właśnie podobne rezultaty. Jeżeli do reprezentacji łuku pozyskamy 100 wierzchołów charakteryzujących łuk, to będzie on posiadał poprawną wizualizację w określonych skalach. Zmniejszając skalę strzałka ulega wydłużeniu, a pewna liczba wierzchołków może zostać zastąpiona przez inne. W związku z tym, pozyskane wierzchołki dla skali mniejszej będą innymi wierzchołkami niż dla poprzedniej skali, a ich liczba będzie mniejsza. Nie wpłynie to jednak negatywnie na wizualizację łuku. Czy opisany powyżej przykład generalizacji pierwotnej powoduje zmianę w opisie faktu geograficznego jakim jest łuk? Odpowiedź jest oczywista. Z tego względu należy podkreślić ciekawe rozwiązanie algorytmu Chrobaka, który nie tylko eliminuje wierzchołki krzywej, ale także potrafi wstawić nowe wierzchołki nie pokrywające się z istniejącymi. Jest to zachowanie zgodne z definicją generalizacji pierwotnej.

Na rysunku 6 zostały zaprezentowane wyniki przeprowadzonego upraszczania dla wybranych skal. Analizując je można dojść do wniosku, że algorytmy które w szybki sposób usuwają punkty, jednocześnie powodują dużą liczbę konfliktów wewnętrznych oraz zewnętrznych. Problem konfliktów charakteryzuje jednak wszystkie algorytmy. Najszybciej konflikty pojawiają się dla algorytmów Douglasa-Peuckera i Langa (skala 1:25 000). Problem konfliktów był przedmiotem licznych prac badawczych, wśród których należy podkreślić wyniki Żukowskiej, której badania nad wykorzystaniem normy rozpoznawalności rysunku zaowocowały rozwiązaniem umożliwiającym nie tylko odnajdywanie konfliktów, ale także przez

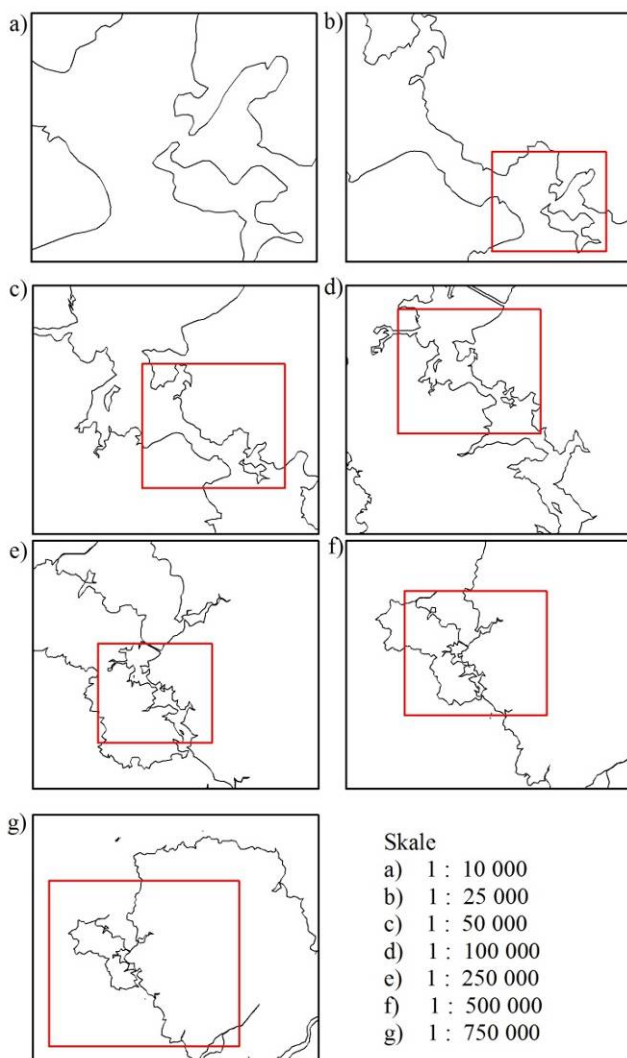


Rys. 6. Przedstawienie zmian w geometrii kształtu dla poszczególnych algorytmów w wybranych skalach – kolejne fazy upraszczania

ich grupowanie i klasyfikację doprowadziły do ich usunięcia (Chrobak i in., 2007). Innym sposobem jest odpowiedni sposób zapisu danych geograficznych, uwzględniający osnowę kartograficzną obiektów. Metoda ta, znajduje się obecnie w trakcie badań (Chrobak, 2009).

Wszystkie algorytmy powodują pojawienie się ostrych krawędzi, jednakże algorytm Chrobaka powoduje pojawienie się kątów ostrych stosunkowo późno, bo dla skali 1:100 000. Dla porównania, w algorytmie Douglasa-Peuckera pierwsze kąty ostre pojawiają się już przy skali 1:50 000 (rys. 6). Pod tym względem ciekawie zachowuje się algorytm Reumanna-Witkama, który już w skali 1:25 000 charakteryzuje się pojawieniem kątów ostrych, które wraz ze zmniejszaniem skali, zanikają (rys. 6). Dla skali 1:750 000 algorytm Chrobaka pozostawił najwięcej wierzchołków spośród porównywanych algorytmów i prawie dwukrotnie więcej niż algorytm Douglasa-Peuckera.

Po analizie wyników można postawić następujące pytania: Czy algorytm Chrobaka jest



tylko algorytmem upraszczającym krzywą? Czy jego zachowanie odzwierciedla w danej skali rzeczywistość geograficzną?

Odpowiedzią na postawione pytania może być zaprezentowane zestawienie map wybrzeża w wybranych skalach, uzyskane za pomocą algorytmu Chrobaka (rys. 7). Na zestawieniu tym widać jak zmienia się liczba wierzchołków, przy jednoczesnym zachowaniu pewnych istotnych cech geograficznych. Możliwe jest to dzięki cesze tego algorytmu jaką jest nie tylko usuwanie punktów krzywej, ale także ich dodawanie w miejscach określanych na podstawie normy rozpoznawalności. Należy podkreślić, iż zaprezentowane fragmenty wybrzeża nie podlegały żadnym innym przekształceniom typu: wygładzanie czy eliminacja. Konflikty, które pojawiły się w trakcie upraszczania, także nie zostały usunięte, aby można było zaobserwować rzeczywisty efekt algorytmu Chrobaka.

Rys. 7. Wyniki działania algorytmu Chrobaka dla wybranych skal wraz ze zwiększeniem obszaru wizualizacji (kwadratem oznaczono zakresy dla skali poprzedzającej)

Wnioski

Norma rozpoznawalności wprowadzona do procesu upraszczania powoduje jednoznaczność otrzymanych wyników procesu upraszczania. Zastosowanie normy jest możliwe dla innych algorytmów, lecz należy je przekształcić w zależności od parametrów wymaganych przez poszczególne algorytmy. Algorytm Chrobaka wykazał się najmniejszymi zmianami w zakresie liczby wierzchołów usuniętych z krzywej, czego efektem jest prawidłowe odzwierciedlenie rzeczywistości geograficznej.

Wybrane do porównania algorytmy (Douglasa-Peuckera, Langa, Reumanna-Witkama) cechuje jedna wspólna wada jaką jest jedynie odrzucanie wierzchołków krzywej pierwotnej. Autor podaje propozycję aby określać te algorytmy mianem algorytmów „odrzućcia”. Pojawienie się konfliktów jest nie do uniknięcia, jednakże liczba tych konfliktów jest jedną z miar jakości przeprowadzonego procesu upraszczania.

Algorytm Chrobaka można stosować bez ograniczeń ilościowych dla danych źródłowych i powinien on stać się jednym z algorytmów stosowanych w narzędziach GIS, jako alternatywny w stosunku do zaprezentowanych algorytmów. Algorytm Chrobaka jest algorytmem globalnym (analiza całej krzywej), jednakże jego wynik zachowuje lokalne ekstrema krzywej, dzięki czemu przebieg krzywej wynikowej jest adekwatny do faktów geograficznych.

Literatura

- Chrobak T., 1999: Badanie przydatności trójkąta elementarnego w komputerowej generalizacji kartograficznej, UWND AGH Kraków.
- Chrobak T., Koziol K., Szostak M., Żukowska M., 2007: Podstawy cyfrowej generalizacji kartograficznej, UWND AGH Kraków.
- Chrobak T., 2009: Przydatność osnowy kartograficznej i metody obiektywnego upraszczania obiektów do aktualizacji danych w TBD, *Geoamtics and Environmental Engineering*, Vol. 3, No. 1/1, 81-90, UWND AGH Kraków.
- Chrobak T., 2010: The role of least image dimensions in generalization of object in spatial databases, *Geodesy and Cartography*, Polish Committee for Scientific Research (w druku).
- Douglas D.H., Peucker T.K., 1973: Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature, *The Canadian Cartographer*, Vol. 10, No. 2, 112-22.
- Lang T., 1969: Rules for the Robot Draughtsmen, *The Geographical Magazine*, 42(1), 50-51.
- Reumann K., Witkam A.K.P., 1974: Optimizing curve segmentation in computer graphics, Proceedings. International Computing Symposium, North-Holland Publishing Company, Amsterdam, 52-78.
- Saliszczyk K.A., 1998: Kartografia ogólna, Wydawnictwo naukowe PWN, Warszawa.

Abstract

The operator of simplification is one of the most widely used in cartography, and now there are many algorithms performing this process automatically. In 1999, Chrobak developed his own algorithm based on the value of drawing recognition with the use of elementary triangle for testing curve. This algorithm does not provide the possibility of introducing any value of tolerances (linear or angular), but only the value of the target scale. In this paper, the author made attempt to compare Chrobak's algorithm and other algorithms functioning in GIS. The results are not so surprising but very interesting from the point of view of rendering geographical facts in different scales. When comparing the number of vertices, remaining after simplification of various algorithms with the result of Chrobak's algorithm, the obtained results differed even four times. The three algorithms compared should be called rejection algorithms, because they do not fulfill the basic condition that is „similarity” in relation to the initial curve.

dr inż. Krystian Koziol
krystian.koziol@agh.edu.pl
home.agh.edu.pl/koziol