



Aleksander Żarnowski

**E-JĘZYK I GENEROWANIE SYMBOLI E-ALFABETU
NA PODSTAWIE TRANSFORMACJI „TEKST-GRAFIKA”**

**E-LANGUAGE AND GENERATING SYMBOLS
OF E-ALPHABET BASED ON „TEXT – GRAPHICS”
TRANSFORMATION**

*Katedra Fotogrametrii i Teledetekcji, Uniwersytet Warmińsko-Mazurski w Olsztynie
e-mail: Aleksander.Zarnowski@uwm.edu.pl*

STRESZCZENIE: Jak pokazuje historia cywilizacji, nowy stopień jakościowy jej rozwoju jest ściśle powiązany ze sposobem zapisywania i przekazywania informacji. Sposób zapisywania informacji w postaci tekstu praktycznie nie zmienił się od czasów odkrycia druku.

W referacie zdefiniowano podstawowe pojęcia (alfabet, litera, słowo, zdanie i tekst), które są wykorzystywane przy opracowaniu E-alfabetu. Pokazano, że dowolne słowo w języku naturalnym może być zapisane jako symbol E-alfabetu przy odpowiednio dobranym alfabecie. Zdefiniowano schemat formowania symboli E-alfabetu i przedstawiono przykłady takich symboli.

Wykorzystane dla generowania symboli E-alfabetu podstawy teoretyczne mogą być zastosowane do opracowania SIP-alfabetu i modernizacji technologii geoinformatycznej tradycyjnie wykorzystywanej w SIP.

SŁOWA KLUCZOWE: Kartografia, SIP, technologia geoinformatyczna, alfabet elektroniczny

1. WPROWADZENIE

Z historii rozwoju cywilizacji wynika, że poziom jej osiągnięć jest ściśle związany, przede wszystkim, ze sposobem zbierania, przechowywania i przekazywania informacji. Jej dzisiejszy poziom, opierający się na zapisywaniu informacji w postaci tekstu, praktycznie nie zmienił się od czasów odkrycia druku. Człowiek, który nauczył się alfabetu 400 lat temu, mógłby czytać współczesne książki korzystając z umiejętności zdobytych wcześniej.

Na Ziemi istnieje około 6000 tradycyjnych języków, przy czym wykorzystują one tylko 15 typów alfabetów. Zbiór danych, który przesunął ludzkość na drobinie biosocjalnej ewolucji, został graficznie zapisany tymi alfabetami. Z punktu widzenia historycznego czas tworzenia alfabetów to aż taka odległa przeszłość, ale przy ich opracowaniu nie mogło być wzięte pod uwagę ich późniejsze wykorzystanie w technologiach komputerowych. Rozwój języków naturalnych i ich obecny stan jest mało związany z dyskretnymi

przetwornikami informacji, jakimi są komputery. Z drugiej strony, rozwój gramatyk formalnych w latach 50-tych XX wieku stworzył nowe możliwości wykorzystania komputerów w trybie opracowania języków algorytmicznych (Algol, Fortran, Pascal, BASIC, C itp.). Jednakże proces rozwoju tych języków przebiega w swojej dziedzinie, słabo powiązanej z dziedziną języków naturalnych.

Rozwiązaniem problemu mogłoby być opracowanie alfabetu podwójnego naznaczenia, który mógłby być wykorzystany przez komputery zarówno do odczytywania i transferu danych, jak i przez człowieka. Nie jest to proste, ale w procesie uczenia się powstanie drugi poziom systemu sygnałowego człowieka – od związku „słowo kluczowe => obraz” do związku „symbol graficzny => obraz”, co może zwiększyć efektywność dialogu między człowiekiem i komputerem.

1. E-ALFABET I ZASADY FORMOWANIA SYMBOLI E-ALFABETU

Aksjomat 1. Dowlone słowo w języku naturalnym może być zapisane w postaci symbolu E-alfabetu, gdy dobierze się odpowiedni alfabet podstawowy (\mathbf{V}).

Definicja 1. Symbol E-alfabetu jest to wygenerowany według matematycznie określonych zasad, symbol graficzny na obrazie rastrowym, którego główną cechą jest odzwierciedlenie słowa napisanego w języku wybranym bez strat informacyjnych.

Wymiary rastra dla transformowanego słowa ustala się na jednej osi (np. – X) równej długości słowa, na drugiej osi, w tym przypadku Y , równej wymiarowi alfabetu. Wszystkim pikselom rastra przyporządkowuje się określony poziom jasności spektralnej.

Definicja 2. Dla dowolnej litery (a) słowa w języku naturalnym (a) \in (\mathbf{V}) jest prawdziwe, że może być ona również rzutowana na obraz cyfrowy \mathbf{C} ($i=0,1,\dots,m$, $j=0,1,\dots,n$) jako piksel o współrzędnych (indeksach):

$$\begin{aligned} i &= r, \\ j &= R, \end{aligned} \quad (1)$$

gdzie:

r – bieżący numer litery w alfabecie (\mathbf{V}),

R – bieżący numer litery w słowie,

K – wymiar alfabetu,

N – liczba liter w słowie.

Definicja 3. Funkcję (regułę) formowania symboli E-alfabetu można określić w następujący sposób: dla każdej litery transformowanego słowa (aa) w języku naturalnym i określonego alfabetu (\mathbf{V}) należy obliczyć indeksy (współrzędne) odpowiedniego piksela rastra zgodnie z definicją 2 i przypisać mu określony poziom jasności spektralnej, na przykład $c_{i,j} = 1$.

2. SCHEMAT FORMOWANIA SYMBOLI E-ALFABETU

Określmy alfabet:

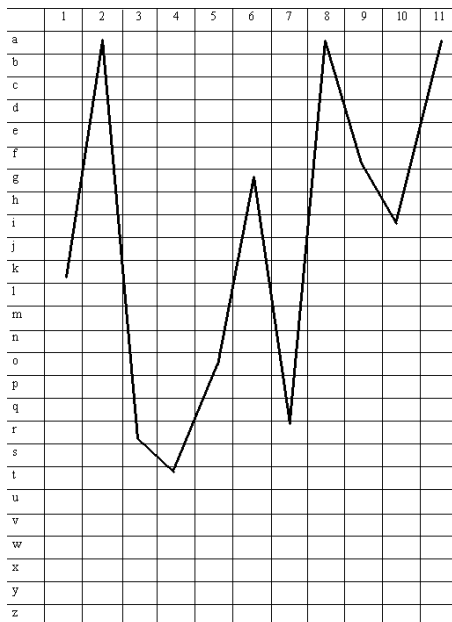
$$(\mathbf{V}) = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\} \quad (2)$$

którego wymiar równa się 26. Zgodnie z definicjami 1–3, słowo **kartografia** (liczba liter N w słowie równa się 11) można wygenerować jako symbol E-języka wg schematu przedstawionego na rys. 1:

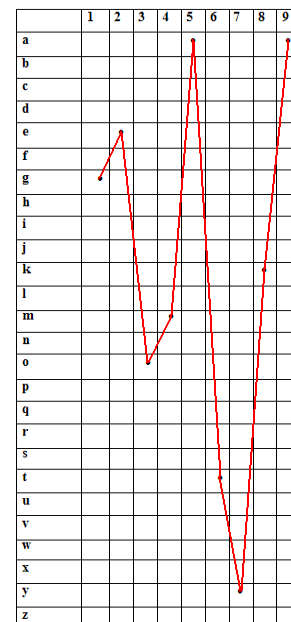
$$\begin{aligned}
 S &\Rightarrow k \Rightarrow C_{11,} = 0 \\
 &\Rightarrow a \Rightarrow C_{1,2} = 0 \\
 &\Rightarrow r \Rightarrow C_{18,3} = 0 \\
 &\Rightarrow t \Rightarrow C_{22,4} = 0 \\
 &\Rightarrow o \Rightarrow C_{15,5} = 0 \\
 &\Rightarrow g \Rightarrow C_{7,6} = 0 \\
 &\Rightarrow r \Rightarrow C_{18,7} = 0 \\
 &\Rightarrow a \Rightarrow C_{1,8} = 0 \\
 &\Rightarrow f \Rightarrow C_{6,9} = 0 \\
 &\Rightarrow i \Rightarrow C_{9,10} = 0 \\
 &\Rightarrow a \Rightarrow C_{1,11} = 0
 \end{aligned}$$

Rys. 1. Schemat formatowania słowa **kartografia**
Fig. 1. Forming the word „**kartografia**” – a diagram

Odpowiednio, symbol E-alfabetu na obrazie cyfrowym dla słowa **kartografia** będzie zapisany w postaci pokazanej na rys. 2 a dla słowa **geomatyka** – na rys. 3.



Rys. 2. Symbol E-alfabetu dla słowa
„**kartografia**”
Fig. 2. E-alphabet symbol for the word
„**kartografia**”



Rys. 3. Symboli E-alfabetu dla słowa
„**geomatyka**”
Fig. 3. E-alphabet symbol for the word
„**geomatyka**”

Definicja 4. Rozmiar symbolu E-alfabetu nie wpływa na jego identyfikację, jeśli skala zmiany w poziomie i pionie jest stała. Zmiana skali daje możliwości optymalizacji transferu danych.

Przy zmianie wymiaru rastra w taki sposób, że po osi **X** ustala się wymiar alfabetu i odpowiednio po osi **Y** – długość słowa, schemat wygenerowania symbolu można przedstawić jak na rys. 4.

Definicja 5. Symbol E-alfabetu, wygenerowany według definicji 3 jest zapisany w macierzy **A**. Aby przekształcić symbol E-alfabetu, wystarczy wyliczyć macierz transponowaną \mathbf{A}^T , która powstaje z **A** przez zamianę wierszy z kolumnami zgodnie ze wzorem:

$$\mathbf{A}^T = (C_{ji}) \quad (3)$$

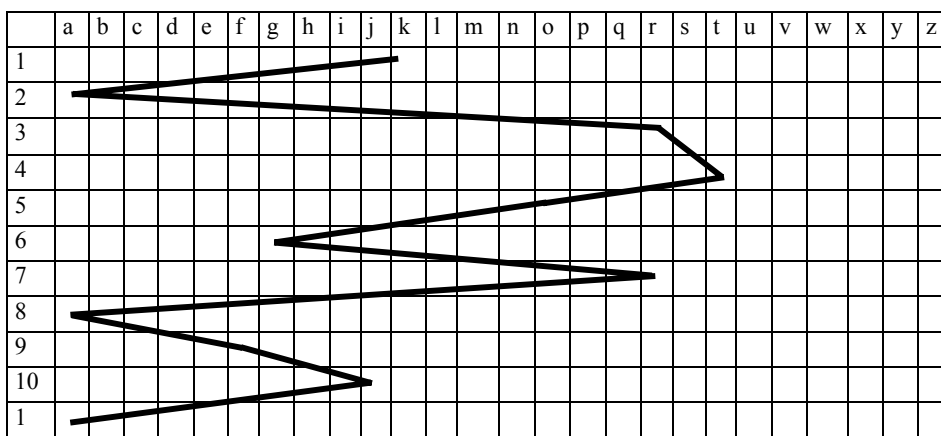
gdzie:

$$i = 1, 2, \dots, N; \quad j = 1, 2, \dots, K.$$

$$\begin{aligned} S &\Rightarrow k \Rightarrow C_{1,11} = 0 \\ &\Rightarrow a \Rightarrow C_{2,1} = 0 \\ &\Rightarrow r \Rightarrow C_{3,18} = 0 \\ &\Rightarrow t \Rightarrow C_{4,22} = 0 \\ &\Rightarrow o \Rightarrow C_{5,15} = 0 \\ &\Rightarrow g \Rightarrow C_{6,7} = 0 \\ &\Rightarrow r \Rightarrow C_{7,18} = 0 \\ &\Rightarrow a \Rightarrow C_{8,1} = 0 \\ &\Rightarrow f \Rightarrow C_{9,6} = 0 \\ &\Rightarrow i \Rightarrow C_{10,9} = 0 \\ &\Rightarrow a \Rightarrow C_{11,1} = 0 \end{aligned}$$

Rys. 4. Schemat słowa „kartografia” przy zmianie osi

Fig. 4. Forming the word „kartografia” – a diagram when the axis is altered



Rys. 5. Symbolu E-alfabetu dla słowa „kartografia” przy zmianie osi

Fig. 5. E-alphabet symbol for the word „kartografia” when the axis is altered

3. SIP-ALFABET I FORMOWANIE JEGO SYMBOLI NA PRZYKŁADZIE BAZY DANYCH TOPOGRAFICZNYCH

Jakość dialogu między użytkownikiem i komputerem stanowi poważny problem przy opracowaniu i wdrażaniu dowolnych systemów informatycznych. Problem ten ma szczególną wagę dla SIP, gdy stosuje się mapy wektorowe i rastrowe do celów wizualizacji, pozyskiwania i odzwierciedlania danych przestrzennych. Opracowany E-alfabetu, może być przydatny do polepszenia jakości takiego dialogu.

Kryteria jakości mapy i jej przydatności z punktu widzenia człowieka są: podstawy matematyczne budowy mapy, treść, system znaków kartograficznych i kolorystyka, czytelność i estetyka obrazu. Do głównych z nich można zaliczyć czytelność mapy, pod którą mamy na myśli proces dostrzegania i zrozumienia jej treści lub pozyskanie informacji o terenie na podstawie jego kartograficznego wizerunku.

Już w bardzo wczesnych pracach naukowych przy określeniu miejsca kartografii w naukach o Ziemi wskazano na powiązanie języka kartograficznego z językiem naturalnym. Stwierdzono na przykład (Boczarow, 1966), że „czytać mapę – to znaczy łączyć symbole alfabetu kartograficznego (znaki kartograficzne i barwy) w słowa i zdania; to znaczy nauczyć się, po połączeniu symboli widzieć teren takim, jakim on jest naprawdę”. Czytanie mapy polega zwykle na czytaniu nazw tekstowych, słowno-liczbowych oraz pozyskiwaniu informacji z modelu kartograficznego i przekazywaniu wiadomości w formie językowej. W związku z tym staje się zrozumiałe, jaką rolę w czytaniu mapy odgrywa wyrazistość obrazu kartograficznego, która w największym stopniu ułatwia identyfikację obiektów terenu na podstawie ich odwzorowania na mapie. Najlepiej jest, gdy znak kartograficzny kształtem przypomina obiekt lub jego najbardziej charakterystyczną cechę (Boczarow, 1966). Można uznać za interesującą ideę, że nawet ten sam model geometryczny użytkowany w różnych warunkach powinien otrzymać inną – adekwatną do tych warunków – symbolikę (Głażewski, 2001).

Słuszne jest pytanie, jaką jest rola komputera w numerycznych technologiach kartograficznych i jak wykorzystuje się komputer dla czytania mapy? Niestety, do czasów obecnych rola komputera w kartografii i SIP jest głównie ograniczona do pełnienia następujących funkcji: instrumentu dla opracowania mapy, elektronicznego magazynu danych, środka obliczeniowego dla analiz ilościowych i wydruku map. Związane to jest ze specyfiką opracowywania obrazów graficznych na komputerze i brakiem odpowiednich teoretycznych uzasadnień realizacji metod sztucznej inteligencji w geoinformatycznych technologiach i aplikacjach.

Z powyższego wynika, że praktycznie wszystkie właściwości mapy, z których korzysta człowiek odczytując z niej informacje o terenie, są słabo sformalizowane lub nawet bezużyteczne przy „czytaniu” mapy przez komputer. Jeżeli spojrzeć na proces czytania mapy jako na łączenie symboli alfabetu kartograficznego w słowa i zdania, to pytanie brzmi: jakie powinny być symbole alfabetu kartograficznego, które w równym stopniu mogłyby być efektywnie zarówno przez komputer, jak i przez człowieka przy pracy z mapą?

Jednym z takich rozwiązań może być opracowanie SIP-alfabetu na podstawie teorii generowania symboli E-alfabetu i zastosowanie go w technologiach geoinformatycznych, tradycyjnie wykorzystywanych w SIP. Bezwarunkowo, rozwiązanie optymalne nie może jednocześnie być:

- maksymalnie wyrazistym przy czytaniu mapy przez człowieka,
- maksymalnie przydatne przy opracowywaniu informacji obrazowej przez komputer.

W artykule przedstawiono propozycje opracowań zmodyfikowanych produktów kartograficznych. Na pierwszym etapie przyjmuje się rozwiązanie kompromisowe: mapy tradycyjne na papierowych nośnikach uwzględniają zapotrzebowanie człowieka, jako użytkownika końcowego, ale już mają wpasowane symbole SIP-alfabetu, co zwiększa możliwości opracowywania map przez komputer, natomiast mapy komputerowe, których wizualizacja odbywają się na monitorze, w większym stopniu uwzględniają specyfikę opracowania informacji kartograficznej przez komputer.

Zgodnie z klasyfikacją obiektów bazy danych topograficznych (Piotrowski, 2001) systematyka przestrzennego zagospodarowania terenu składa się z 351 pozycji, w czym wyróżniamy: 42 – terenowe układy sieciowe, 130 – obiektów topograficznych, 151 – przedmiotów terenowych, 28 – elementów krajobrazowych.

W pracach eksperymentalnych przeanalizowano 8 modyfikacji alfabetów, uwzględniając czytelność wygenerowanych symboli E-alfabetu dla użytkownika. Odczytywanie tych symboli przez komputer jest jednoznaczne, i z tego powodu na wybór alfabetu nie ma on wpływu. Do końcowego opracowania wybrano następujący alfabet:

$$(V) = \{.,aąbcćdeęfghijklmńnoóprśstuwyzż0123456789-\} \quad (4)$$

Tabela 1

Table 1

Terenowe układy sieciowe
Terrain net systems

Klasa Class	Typ Type	Rodzaj Kind	Symbol SIP-alfabetu GIS-alfabet symbol	Poziom hierarchii Hierarchy level
Sieć pierwszorzędna Primary network				1
”	Sieć wodna Water network			2
”	_”_	Rzeki o długości pow. 20 km Rivers oper 20 km long		3
”	_”_	Rzeka żeglowna Navigable river		3
”	_”_	Kanał żeglowny Navigable channel		3
”	_”_	Kanał Channel		3











Wyberzemy funkcje generacji symboli SIP-alfabetu, zgodnie z definicją 3, oraz przyjmujemy 3-poziomowy model hierarchii symboli. W tym przypadku, terenowe układy sieciowe będą opisane symbolami SIP-języka (tab. 1). Do generacji symboli wykorzystano system „GT” opracowany przez autora. Można go również stosować do translacji symboli SIP-alfabetu na język polski.

Przykłady obiektów topograficznych, przedmiotów terenowych i elementów krajobrazowych pokazano w tabeli 2.

Tabela 2

Table 2

Obiekty topograficzne, przedmioty terenowe i elementy krajobrazowe
Topographic objects, terrain objects and landscape elements

Klasa Class	Typ Type	Rodzaj Kind	Symbol SIP-alfabetu GIS-alphabet symbol	Poziom hierarchii Hierarchy level
Grunty Land				1
–”–	Grunty zabudowane Developed land			2
–”–	Grunty zadrzewione Gorest land			2
–”–	Grunty pod wodami Underwater land			2
Subobszary Subareas				1
–”–	Tereny rekreacyjne Recreational areas			2
–”–	–”–	Park Park		3
Elementy liniowe Linear elements				1
–”–	Alejka Alley			2
–”–	Żywopłot Hedges			2

4. PODSUMOWANIE I WNIOSKI

Na podstawie opracowanych symboli E-języka można stwierdzić, że:

1. Każdy symbol SIP-alfabetu ma swój indywidualny wygląd, co umożliwia jego jednoznaczną identyfikację zarówno „ręczną”, jak i komputerową.
2. Ważnym etapem generacji symboli SIP-alfabetu jest wybór słów kluczowych do określenia klasy obiektów. Takie symbole mogą być stosowane również jako znaki kartograficzne na mapach topograficznych i specjalnych.
3. Trzy poziomowy system hierarchii symboli SIP-alfabetu umożliwia określenie dostatecznie elastycznego sposobu ich wykorzystania jako źródła danych semantycznych.

LITERATURA

- Boczarow M. K., 1966, Podstawy teorii projektowania system znaków kartograficznych. Nedra, Moskwa, str. 31–45.
- Głazewski A., 2001, Natura znaku w odniesieniu do oznaczanego przedmiotu w badaniach nad podstawami metodycznymi projektowania systemu znaków dla map topograficznych, Mapa w systemach komputerowych, Materiały Ogólnopolskich Konferencji Kartograficznych tom 23, XXVIII, Szczecin, str. 76–88.
- Piotrowski R., 2001, Topograficzna baza danych. Program działania, GISPOL, Stowarzyszenie Użytkowników Krajowego Systemu Informacji o Terenie, Warszawa.

E-LANGUAGE AND GENERATING SYMBOLS OF E-ALPHABET BASED ON „TEXT – GRAPHICS” TRANSFORMATION

S u m m a r y

As illustrated by the history of civilization, progress has always involved some means of recording and transferring information. The text form of recording information has remained virtually unchanged since Gutenberg developed printing.

Basic terms such as alphabet, letter, word, sentence, and text, which all are used for an E-alphabet, are defined in this paper. This paper demonstrates that any word from a natural language may be represented as a symbol of an E-alphabet if the alphabet is appropriately chosen. Additionally, this paper defines a schema of forming E-alphabet symbols and presents examples of such symbols.

Generating E-alphabet symbols is based on theory that can also be applied to writing an SIP-alphabet as well as modernising the geoinformatics technology traditionally used in SIP.

Recenzent: prof. dr hab. inż. Ewa Krzywicka-Blum, Akademia Rolnicza, Wrocław