

Andrzej Ziabicki

Komunikacja globalna w czasach internetu – dostosowanie (do) języków narodowych

Komunikacja globalna (internet, poczta elektroniczna) wymaga rozwoju narzędzi umożliwiających zapisanie i przesyłanie informacji w rozmaitych językach. Nowy system kodowania pisma Unicode pozwala na przekazywanie informacji zapisanych dowolnym alfabetem bądź systemem ideogramów bez konieczności transkrypcji lub transliteracji. Trudności sprawia natomiast, niezbędne do efektywnego korzystania z baz danych, porządkowanie wyrazów i danych. Zasady porządkowania w różnych językach (nawet w językach posługujących się alfabetem łacińskim) są sprzeczne i porządkowanie zbiorów wielojęzycznych wymaga przyjęcia jednolitych (kompromisowych) zasad. W artykule przedyskutowano konsekwencje wprowadzenia międzynarodowej normy ISO 12199 dla różnych języków europejskich.

Wstęp

Komunikacja multimedialna odgrywa coraz większą rolę. Oprócz szybkiego przekazywania informacji (tekstowych i graficznych) oraz możliwości dyskusji, wielkie znaczenie ma dostęp do encyklopedii, słowników, baz danych, katalogów i indeksów. Pewne ograniczenia w tej dziedzinie wynikają z różnorodności języków narodowych i niedostosowania ich do komunikacji globalnej. Na przestrzeni wieków wykształciły się różne pisma i odmienne zasady porządkowania haseł leksykograficznych (wyrazów, danych). Konieczne jest więc zarówno **dostosowanie programów komputerowych do istniejących języków**, jak i **ujednoczenie narodowych systemów porządkowania**. Niedawno analizowałem techniczne aspekty kodowania pisma i porządkowania wielojęzycznych baz danych (por. Ziabicki 2002). W tym artykule skoncentruję się na cywilizacyjnych i psychologicznych konsekwencjach zmian w językach europejskich zalecanych przez międzynarodową normę ISO (por. *Norma ISO 12199*, 2000).

Kodowanie pisma

Pisma alfabetyczne i sylabiczne posługują się stosunkowo niewielkimi zbiorami znaków (10–100). W pismach ideograficznych, obejmujących tysiące ideogramów (np. w chińskim, japońskim), kodowanie jest trudniejsze. Kodowanie polega na przy-

pisaniu każdemu znakowi określonej wartości liczbowej. W początkach komputerowego zapisu i transmisji tekstów podstawą były 7-bitowe kody ASCII (por. *ASCII...*) różniące 128 ($= 2^7$) znaków, w tym duże i małe litery, cyfry i znaki przestankowe standardowego 26-literowego alfabetu łacińskiego. Następny krok polegał na zastosowaniu kodów 8-bitowych, obejmujących do 256 znaków. Tak zwane strony kodowe, utworzone przez normę ISO 8859, oprócz znaków ASCII objęły niestandardowe litery alfabetu łacińskiego, litery łacińskie ze znakami diakrytycznymi (por. *Norma ISO 8859-x*), a także kombinacje alfabetu łacińskiego z cyrylicą, alfabetem greckim, arabskim i hebrajskim (por. *Norma ISO 8859-x*). Powszechnie stosowane w komputerach 256-elementowe strony kodowe pozwalają kodować i transmitować teksty w poszczególnych językach, ale nie wystarczają do pracy z **tekstami wielojęzycznymi**. Nawet w obrębie alfabetu łacińskiego stosuje się różne zbiory znaków niestandardowych i tekst wysłany może przybrać inną formę u odbiorcy, jeżeli w jego komputerze zainstalowany jest inny zestaw znaków. Rozwiązanie tego problemu osiągnięto przez wprowadzenie nowego, jednolitego systemu kodowego Unicode. Kalifornijskie konsorcjum Unicode we współpracy z Międzynarodową Organizacją Normalizacyjną postawiło sobie ambitny cel opracowania jednolitego kodu do transmisji tekstów **we wszystkich językach świata**. System Unicode i związana z nim norma ISO/IEC 10646 tworzy przestrzeń kodową składającą się z 17 płaszczyzn, z których każda zawiera 65 536 kodów (łącznie 1 114 112 elementów). Każdemu znakowi pisma (litera, cyfra, sylaba, ideogram, znak specjalny) odpowiada liczba pozycyjna o podstawie szesnastkowej, Dotychczas opracowano kody składające się z ok. 100 tys. znaków. Pierwsza płaszczyzna, zwana BMP (*Basic Multilingual Plane*), kodowana jest 4-cyfrowymi liczbami szesnastkowymi (UXXXX) oraz zawiera znaki ważniejszych pism alfabetycznych i sylabicznych, podstawowe ideogramy chińskie, znaki specjalne, znaki matematyczne, muzyczne, techniczne itp. „U” oznacza system Unicode, a „X” jest dowolną cyfrą w układzie szesnastkowym (0, 1, ..., E, F). Zbiór 65 536 kodów na płaszczyźnie BMP pokrywa bieżące potrzeby transmisji tekstów, pozostawiając wiele miejsca na rozszerzenie repertuaru znaków. W odniesieniu do języków opartych na alfabecie łacińskim wykorzystuje się tylko dwa dwucyfrowe kody szesnastkowe (U00XX i U01XX) – łącznie 512 znaków.

Zbiór znaków Unicode został już zastosowany w systemie Windows NT. Powszechne wprowadzenie tego kodu umożliwi transmisję tekstów wielojęzycznych bez utraty jakichkolwiek cech graficznych właściwych poszczególnym językom.

Znaczenie uporządkowania katalogów i baz danych

Alfabetyczne uporządkowanie baz danych i katalogów ma olbrzymie znaczenie. Wyszukiwanie hasła w bazie **nieuporządkowanej** wymaga przeprowadzenia liczby operacji równej (średnio) połowie liczby elementów w przeszukiwanym zbiorze, $x_{un} = 1/2N$. W przypadku bazy **uporządkowanej alfabetycznie** liczba potrzebnych operacji (a więc i czas wyszukiwania) jest równa **logarytmowi liczby elementów**, $x_{ord} = \log_2 N$. Stosunek liczby operacji w zbiorach nieuporządkowanych i uporządkowanych (charakteryzujący efektywność porządkowania) podano w tabeli 1.

Tabela 1

Stosunek liczby operacji (czasów wyszukiwania) w bazach danych: nieuporządkowanej i uporządkowanej alfabetycznie

Liczba rekordów w bazie N	Stosunek czasów wyszukiwania
	$\frac{x_{un}}{x_{ord}} = \frac{t_{un}}{t_{ord}} = \frac{N}{2\log_2 N}$
10	1,5
100	7,5
1 000	50,2
10 000	373,3
100 000	3 010,3
1 000 000	25 085,8

Jak widać z tabeli 1, w przypadku niewielkich zbiorów zysk czasu wynikający z uporządkowania jest niewielki, ale w przypadku dużych katalogów i baz (a z takimi mamy do czynienia w internecie) uporządkowanie haseł (rekordów) jest absolutnie konieczne.

W odróżnieniu od kodowania, opracowanie jednolitych zasad porządkowania, zgodnych z zasadami języków narodowych, jest niemożliwe. Przyczyną tego jest odmiennosc (często sprzeczność) zasad porządkowania w różnych językach. W poszczególnych krajach stosuje się rozmaite systemy porządkowania lokalnych baz danych i katalogów. Sam tworzyłem dla własnego użytku komputerowe systemy porządkowania w językach polskim, rosyjskim i nowogreckim (por. Ziabicki 1990).

Ogólne zasady porządkowania haseł leksykograficznych

We wcześniejszych pracach analizowałem systemy porządkowania haseł leksykograficznych w różnych językach i opracowałem ogólny algorytm porządkowania w dowolnym języku alfabetycznym (por. Ziabicki 1988; 1990; 1992).

Porządkowanie jest dwustopniowe. **Pierwszy stopień** obejmuje właściwe dla każdego języka elementy pisma, nazwane ostatnio **sortemami** (por. Ziabicki 2002). Rolę sortemów mogą odgrywać **litery standardowe** (a–z), **litery niestandardowe** (np. *æ, ø* w języku duńskim), **litery ze znakami diakrytycznymi** (ą, ó w języku polskim, ö w językach szwedzkim, fińskim i węgierskim), a także wieloznaki (np. „ch” w czeskim, „cs”, „dzs” w węgierskim). Uporządkowany zbiór sortemów nazywamy **alfabetem**.

Porządkowaniu **drugiego stopnia** poddaje się tylko wyrazy zbudowane z identycznych sortemów, a różniące się **znakami diakrytycznymi**. Na przykład w języku francuskim litery ze znakami diakrytycznymi nie są sortemami i traktowane są w pierwszym stopniu porządkowania tak jak litery bez znaków diakrytycznych. Porównajmy kilka wyrazów francuskich. W pierwszym stopniu porządkowania akcenty nie wpływają na kolejność haseł i pozycje wyrazów *élève* i *élevé*, a także *élude* i *éludé* są jednakowe. W drugim stopniu porządkowania pojawiają się znaki diakrytyczne, które wprowadzają dodatkowy porządek różnie akcentowanych wyrazów.

I stopień	II stopień
<i>élan</i> (rozpęd)	<i>élan</i>
<i>élève</i> (uczeń) ↔ <i>élevé</i> (wychowany)	<i>élève</i>
<i>élite</i> (elita)	<i>élevé</i>
<i>élude</i> (omijać) ↔ <i>éludé</i> (ominięty)	<i>élite</i>
<i>emballage</i> (opakowanie)	<i>élude</i>
<i>être</i> (być)	<i>éludé</i>
<i>exulter</i> (nie posiadać się z radości)	<i>emballage</i>
	<i>être</i>
	<i>exulter</i>

Znaki diakrytyczne mają określoną hierarchię. W języku francuskim litera nieakcentowana pojawia się najwcześniej, dalej następuje ta sama litera z akcentem *acutus*, potem z akcentem *gravis*, a wreszcie z akcentem *circumflexus*. W języku francuskim litery za znakami diakrytycznymi porządkuje się *od prawej do lewej*.

W języku polskim wszystkie litery alfabetu zmodyfikowane znakami diakrytycznymi (**ą, ć, ę, ł, ń, ó, ś, ź, ż**) są sortemami i porządkowanie ogranicza się do pierwszego stopnia.

Wyższe stopnie porządkowania mogą dotyczyć dowolnych atrybutów pisma (np. krój czcionki, kursywa, wytłuszczenie) lub własności wyrazów, takich jak kategoria gramatyczna, kategoria semantyczna itp.

Porządkowanie zbiorów wielojęzycznych

Problem porządkowania zbiorów **wielojęzycznych** pojawił się wraz z powstaniem wielkich bibliotek i archiwów dysponujących tekstami w różnych językach, a także międzynarodowych baz danych dostępnych w internecie. Najwcześniejsze systemy automatycznego porządkowania baz danych opierały się na kodach ASCII, a więc na literach standardowego alfabetu łacińskiego. Nie uwzględniano liter specjalnych występujących w językach narodowych (np. polskich liter **ą, ć, ..., ź, ż**, skandynawskich **æ, å, ä, ö, ø**), a także wieloznaków podlegających porządkowaniu jak pojedyncze litery alfabetu (np. czeskie i hiszpańskie **ch**, węgierskie **cs, dz, dzs, ..., zs**). W połowie lat dziewięćdziesiątych w Instytucie Podstaw Informatyki PAN w Warszawie odbyła się dyskusja polonistów i informatyków, podczas której rozważano konieczność dostosowania polskiego alfabetu do komunikacji elektronicznej. Wydawało się wówczas, że najprostszym rozwiązaniem będzie zastąpienie liter polskich ich standardowymi odpowiednikami (**ą** → **a**, **ć** → **c**, **ź** → **z** itd.), co zresztą robiło już (i robi) wielu użytkowników poczty elektronicznej i internetu. Ciekawe, że w tym samym czasie (1994) formalną decyzję o zmianie systemu porządkowania w języku hiszpańskim przyjął Kongres Akademii Języka Hiszpańskiego (por. *X Congreso...* 1994). Zrezygnowano z traktowania jako samodzielnych elementów alfabetu (sortemów) dwuznaków **ch** i **ll**, zamieniając je na pary liter (**ch** → **c + h**, **ll** → **l + l**). Nie zdecydowano się natomiast na rezygnację z litery specjalnej **ñ**, położonej w alfabecie pomiędzy **n** i **o**.

Rezygnacja ze specyfiki alfabetu narodowego na pierwszy rzut oka może się wydawać szokująca, ale dostosowanie narodowych zasad porządkowania wyrazów do komunikacji elektronicznej nie pociąga za sobą zmiany ortografii. Ortograficzną wierność transmitowanych tekstów gwarantuje system Unicode, obejmujący wszystkie znaki pisma.

Uniwersalny system porządkowania według normy ISO 12199

Opublikowana ostatnio norma międzynarodowa ISO 12199 (2000) zaleca następujące, jednolite zasady porządkowania haseł we wszystkich językach posługujących się rozszerzeniami alfabetu łacińskiego:

- Przygotowanie tekstu do porządkowania polega na przekształceniu wszystkich nie-standardowych liter i ligatur na standardowe znaki alfabetu łacińskiego. Wieloznaki rozбивa się na litery składowe (**ch** → **c + h**, **dzs** → **d + z + s**).
- Pierwszy stopień porządkowania opiera się na standardowym alfabecie łacińskim, do którego dopisano literę runiczną **þ** (*thorn*). 27 liter, spacja (SPC) i 10 cyfr tworzy pełny zbiór sortemów – jedyną podstawę uporządkowania alfabetycznego.

**(SPC) 0 1 2 3 4 5 6 7 8 9 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 þ**

W pierwszym stopniu porządkowania nie odróżnia się małych i dużych liter oraz ignoruje znaki diakrytyczne. Porządkowanie sortemów w wyrazie odbywa się *od lewej do prawej*.

- Drugi stopień porządkowania opiera się na określonej hierarchii 22 pojedynczych znaków diakrytycznych i ich 21 kombinacji.
- Trzeci stopień porządkowania dotyczy małych i dużych liter. Inaczej niż w systemie ASCII, małe litery poprzedzają duże.
- Czwarty stopień porządkowania dotyczy znaków specjalnych według norm ISO/IEC 14651 i ISO/IEC 10646-1.

Zasady zalecane przez normę ISO 12199 w różnym stopniu zmieniają tradycyjne systemy porządkowania w językach narodowych. Porządkowanie drugiego stopnia ma znaczenie marginalne, gdyż dotyczy nielicznych wyrazów, które bez znaków diakrytycznych są homografami. Przeanalizujemy konsekwencje wprowadzenia tej normy w różnych językach.

Odchylenia narodowych alfabetów i zasad porządkowania od normy ISO 12199

Porównamy zbiory sortemów (alfabety) 17 wybranych języków europejskich posługujących się alfabetem łacińskim. Dla uproszczenia pominiemy spację i cyfry 0–9, które we wszystkich językach poprzedzają w alfabecie literę „a”. Pogrubioną czcionką zaznaczymy *sortemy nietypowe* wyeliminowane przez normę ISO 12199.

angielski, francuski, niemiecki, niderlandzki, włoski, portugalski
(standardowy alfabet łaciński)

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

hiszpański (po zmianach w 1994 r.)

a b c d e f g h i j k l m n ñ o p q r s t u v w x y z

szwedzki

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 å ä ö

norweski

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 æ ø å

duński

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 å æ ø

fiński

a b c d e f g h i j k l m n o p q r s š t u v w x y z ž å ä ö

polski

a ą b c ć d e ę f g h i j k l ł m n ń o ó p q r s ś t u v w x y z ź ż

chorwacki

a b c č ć d đ ž đ e f g h i j k l l j m n n j o p q r s š t u v w x y z ž

rumuński

a â ä b c d e f g h i i j k l m n o p q r s ș ț t u v w x y z

węgierski

a b c c s d d z d z s e f g g y h i j k l l y m n n y o ö p q r s s z t t y u ü v w x y z z s

czeski

a b c č d e f g h c h i j k l m n o p q r ř s š t u v w x y z ž

litewski

a ą b c č d e é ę f g h i y j j k l m n o p q r s š t u ū v w x z ž

Odchylenie narodowych zasad porządkowania od systemu ISO wynika z obecności w alfabetach narodowych nietypowych sortemów, a także nietypowego położenia w alfabecie sortemów standardowych. „Nietypowe” cechy porządkowania zebrano w tabeli 2.

Tabela 2

Cechy porządkowania alfabetycznego (porządkowanie pierwszego stopnia)
w wybranych językach europejskich

Język	Nietypowe zasady porządkowania		
	niestandardowe litery	wieloznaki	nietypowy porządek
Angielski	–	–	–
Francuski	–	–	–
Niemiecki	–	–	–
Włoski	–	–	–
Portugalski	–	–	–
Niderlandzki	–	–	–
Hiszpański (po 1994 r.)	ñ	–	–
Szwedzki	å, ä, ö	–	z < å < ä < ö
Duński	å, æ, ø	–	z < å < æ < ø
Norweski	æ, ø, å	–	z < æ < ø < å
Fiński	š, ž, ä, å, ö	–	z < ž < ä < å < ö
Polski	ą, ć, ę, ł, ń, ó, ś, ź, ż	–	–
Chorwacki	č, ć, đ, š, ž	dž, lj, nj	–
Rumuński	â, ä, î, ș, ț	–	–
Węgierski	ö, ü	cs, dz, dzs, gy, ly, ny, sz, ty, zs	–
Czeski	č, ř, š, ž	ch	h < ch < i
Litewski	ą, č, é, ę, į, š, ū, ū, ž	–	i < y < į < j

Zmiany porządku haseł w różnych językach wynikające z zastosowania normy ISO

Języki zachodnioeuropejskie postępujące się standardowym alfabetem łacińskim (m.in. **angielski, francuski, niemiecki, włoski, portugalski i niderlandzki**) nie zawierają żadnych nietypowych elementów porządkowania i norma ISO 12199 nie wprowadza do nich żadnych zmian. Wieloznaki, skróty i ligatury (np. **æ, œ**) rozwijane są na pojedyncze litery i porządkowane według standardowego, 26-literowego alfabetu łacińskiego.

Inne analizowane w tej pracy języki odbiegają od normy ISO. W uproszczonym języku **hiszpańskim** pozostawiono niestandardowy sortem **ñ**, położony w alfabecie między **n** i **o**. Zalecane przez normę ISO zastąpienie **ñ** standardową literą **n** prowadzi do zmian w porządku wyrazów. Uporządkowanie II stopnia w znikomym stopniu poprawia kolejność haseł (tabela 3).

Tabela 3
Porządek wyrazów w języku hiszpańskim

Według zasad narodowych	Według normy ISO 12199	
	I stopień	II stopień
<i>nada</i> (nic)	<i>nada</i>	<i>nada</i>
<i>nota</i> (notatka)	<i>ñaque</i>	<i>ñaque</i>
<i>nupcial</i> (ślubny)	<i>ñoño</i>	<i>ñoño</i>
<i>nutrir</i> (żywić)	<i>nota</i>	<i>nota</i>
<i>nylon</i> (nylon)	<i>ñu</i>	<i>ñu</i>
<i>ñaque</i> (kupa gratów)	<i>nupcial</i>	<i>nupcial</i>
<i>ñoño</i> (głupi)	<i>nutrir</i>	<i>nutrir</i>
<i>ñu</i> (antylopa gnu)	<i>nylon</i>	<i>nylon</i>
<i>objeto</i> (przedmiot)	<i>objeto</i>	<i>objeto</i>
<i>sonar</i> (dźwięczec)	<i>sonar</i> ↔ <i>soñar</i>	<i>sonar</i>
<i>sonrisa</i> (uśmiech)	<i>sonrisa</i>	<i>soñar</i>
<i>soñar</i> (śnić, marzyć)		<i>sonrisa</i>

Alfabety języków wschodnioeuropejskich zawierają wiele nietypowych sortemów (litery niestandardowe, wieloznaki, skróty). W językach, w których niestandardowe litery i wieloznaki występują w alfabecie obok swoich standardowych odpowiedników (tak jest w językach **polskim, chorwackim, rumuńskim, węgierskim**) przemieszczenia haseł mają charakter lokalny (w obrębie jednej litery). Występowanie w języku **polskim** dziewięciu niestandardowych sortemów (**ą, ć, ... ż**) prowadzi jednak do częstego występowania takich odchyłań. Zamiana (w I stopniu porządkowania) liter polskich na niemodyfikowane litery standardowe (**ą = a, ć = c, ... ż = z**) powoduje oddalenie haseł obecnie sąsiadujących w słownikach, indeksach i katalogach (tabela 4).

System porządkowania jest szczególnie czuły na nietypowe położenie sortemów w alfabecie. O ile obecność niestandardowych liter obok standardowych odpowiedników (w hiszpańskim **ñ** po **n**, w polskim **ą** po **a**, **ź** i **ż** po **z** itd.) powoduje tylko lokalne przemieszczenia, o tyle nietypowa kolejność sortemów całkowicie burzy strukturę uporządkowa-

nia. Sytuacja taka występuje m.in. w językach skandynawskich (**szwedzki, duński, norweski**), a także w **fińskim**, który posługuje się podobnym alfabetem. Rozbieżność uporządkowanych indeksów wynika z obecności niestandardowych liter **å, ä, æ, ö, ø** położonych **na końcu alfabetu**, po **z**. Zastąpienie tych liter standardowymi literami **a**, (**a + e**), i **o** powoduje, że hasła z końca alfabetu przemieszczają się na początek (wyrazy na **å, ä, æ**) lub lokują się w drugiej połowie alfabetu (wyrazy na **ö i ø**). Widać to wyraźnie w przykładach z języka **szwedzkiego** podanych w tabeli 5.

Tabela 4
Porządek wyrazów w języku polskim

Według zasad narodowych	Według normy ISO 12199	
	I stopień	II stopień
macać	mać	mać
mać	macać	macać
martyrologia	mącić	mącić
marynarz	mądrość	mądrość
mat	mąka	mąka
mazać	martyrologia	martyrologia
maż	marynarz	marynarz
mącić	mat	mat
mądrość	maż ↔ mąż	maż
mąka	mazać	mąż
mąż	pak ↔ pąk	mazać
pak	wąchać	pak
pąk	wahać	pąk
wahać	wąż	wąchać
ważyć	ważyć	wahać
wąchać		wąż
wąż		ważyć

Sytuacja podobna do tej, jaką obserwowaliśmy w językach skandynawskich, występuje w językach **czeskim** i **litewskim**. W czeskim nietypowa kolejność sortemów dotyczy dwuznaku **ch**, położonego w alfabecie nie między literami **c** i **d**, lecz między **h** oraz **i**. Zalecane przez normę ISO traktowanie sortemu **ch** jako sumy liter (**c + h**) całkowicie zaburza tradycyjny, narodowy system porządkowania. Przykłady podaje tabela 6.

W **litewskim** nietypowe miejsce w alfabecie zajmuje litera **y**, położona pomiędzy standardowym **i** i nosowym **į**. Traktując sortem **į** jako równoważny **i** oraz lokując standardową literę **y** pomiędzy **x** i **z** otrzymujemy porządek mało przypominający narodowe zasady porządkowania (tabela 7).

Tabela 5
Porządek wyrazów w języku szwedzkim

Według zasad narodowych	Według normy ISO 12199	
	I stopień	II stopień
<i>agg</i> (złość)	å	å
<i>akademisk</i> (akademicki)	<i>agg</i> ↔ <i>äg</i>	<i>agg</i>
<i>djur</i> (zwierzę)	<i>akademisk</i>	<i>äg</i>
<i>flicka</i> (dziewczyna)	<i>åldras</i>	<i>akademisk</i>
<i>onsdag</i> (środa)	<i>djur</i>	<i>åldras</i>
<i>zoológ</i> (zoolog)	<i>flicka</i>	<i>djur</i>
å (rzeczka)	ö	<i>flicka</i>
<i>åldras</i> (starzec się)	<i>öde</i>	ö
<i>äg</i> (jajko)	<i>onsdag</i>	<i>öde</i>
ö (wyspa)	<i>zoológ</i>	<i>onsdag</i>
<i>öde</i> (opustoszały)		<i>zoológ</i>

Tabela 6
Porządek wyrazów w języku czeskim

Według zasad narodowych	Według normy ISO 12199	
	I stopień	II stopień
<i>celek</i> (całość)	<i>čaditi</i>	<i>čaditi</i>
<i>cíl</i> (cel)	<i>čas</i>	<i>čas</i>
<i>čaditi</i> (dymić)	<i>celek</i>	<i>celek</i>
<i>čas</i> (czas)	<i>charakter</i>	<i>charakter</i>
<i>čtvrť</i> (ćwierć)	<i>chlád</i>	<i>chlád</i>
<i>dálka</i> (odległość)	<i>chov</i>	<i>chov</i>
<i>hvězda</i> (gwiazda)	<i>chyba</i>	<i>chyba</i>
<i>hybný</i> (poruszający się)	<i>cíl</i>	<i>cíl</i>
<i>charakter</i> (charakter)	<i>čtvrť</i>	<i>čtvrť</i>
<i>chlád</i> (chtód)	<i>dálka</i>	<i>dálka</i>
<i>chov</i> (hodowla)	<i>hvězda</i>	<i>hvězda</i>
<i>chyba</i> (omyłka, błąd)	<i>hybný</i>	<i>hybný</i>
<i>identický</i> (identyczny)	<i>identický</i>	<i>identický</i>
<i>impuls</i> (impuls)	<i>impuls</i>	<i>impuls</i>

Tabela 7
Porządek wyrazów w języku litewskim

Według zasad narodowych	Według normy ISO 12199	
	I stopień	II stopień
<i>ilsūs</i> (męczący)	<i>įbauginti</i>	<i>įbauginti</i>
<i>istubà</i> (izba)	<i>ilsūs</i>	<i>ilsūs</i>
<i>išdarbis</i> (wyrób)	<i>išdarbis</i>	<i>išdarbis</i>
<i>ižus</i> (kruchy)	<i>įsmągti</i>	<i>įsmągti</i>
<i>yda</i> (brak, feler)	<i>istubà</i>	<i>istubà</i>
<i>yštaka</i> (ujście)	<i>įžambūs</i>	<i>įžambūs</i>
<i>įbauginti</i> (nastraszyć)	<i>ižus</i>	<i>ižus</i>
<i>įsmągti</i> (ściszyć)	<i>jaunas</i>	<i>jaunas</i>
<i>įžambūs</i> (ukośny)	<i>jutrumàs</i>	<i>jutrumàs</i>
<i>jaunas</i> (młody)	<i>yda</i>	<i>yda</i>
<i>jutrumàs</i> (czujność)	<i>yštaka</i>	<i>yštaka</i>

Konsekwencje dostosowania języków narodowych do uniwersalnego systemu porządkowania ISO

Nie ulega wątpliwości, że wprowadzenie w komunikacji globalnej uniwersalnego systemu kodowania oraz porządkowania pisma jest konieczne i nieuniknione. Jednolity system porządkowania przewidziany w normie ISO 12199 jest zgodny z tradycyjnymi zasadami porządkowania w niektórych językach zachodnioeuropejskich i nie stwarza tam żadnego problemu. W Szwecji, Norwegii, Danii, Finlandii, a także w krajach Europy Środkowej i Wschodniej, nowy system porządkowania wymaga jednak radykalnych zmian w językach narodowych.

Perspektywa wprowadzenia normy ISO 12199 nasuwa szereg pytań. Jaki ma być zakres funkcjonowania normy i czas wprowadzania nowego systemu? Można sobie wyobrazić kilka rozwiązań.

- Nowy system porządkowania stosuje się tylko do **międzynarodowych** baz danych, słowników, katalogów i indeksów dostępnych w formie **elektronicznej**. Zbiory **lokalne**, a także **drukowane** katalogi i słowniki porządkowane są nadal według dotychczasowych zasad.
- Nowy system stosuje się tylko **wewnętrznie**. Dla użytkownika internetowej bazy danych **system porządkowania** jest w gruncie rzeczy obojętny. System wyszukuje informacje na temat podanego hasła i nie wymaga od użytkownika znajomości wewnętrznej struktury systemu. Elektroniczne bazy danych, encyklopedie, katalogi biblioteczne istniały już 20 lat temu i każda posługiwała się jakimś systemem porządkowania. Stosowane obecnie wyszukiwarki internetowe (Yahoo, Altavista, Google, Onet) porządkują swoje zbiory według własnych zasad. Wewnętrzna struktura baz danych i programów wyszukiwawczych nie dotyczy użytkowników internetu.
- Nowy system obowiązuje **powszechnie** oraz dotyczy zarówno zbiorów elektronicznych, jak i drukowanych, międzynarodowych i lokalnych. Wszystkie nowo two-

rzony zbiory muszą być porządkowane według nowych zasad. Można tu zapytać o los **istniejących katalogów** w bibliotekach i archiwach, a także drukowanych **słowników i encyklopedii**. Powszechne wprowadzenie nowego systemu porządkowania będzie budzić zrozumiałe opory psychiczne. Konieczne będzie wprowadzenie nowych zasad do programów szkolnych, a także kosztowne dostosowanie, przynajmniej najważniejszych, baz danych i katalogów. Program wprowadzania takich zmian musiałby być zaplanowany i realizowany w ciągu wielu lat.

- Ekstrapolując obecne tendencje, można sobie wyobrazić, że drukowane słowniki, encyklopedie, katalogi i bazy danych stopniowo będą wychodzić z użycia, aby ustąpić miejsca zbiorom **multimedialnym**. Gdy to nastąpi, problem porządkowania stanie się zmartwieniem informatyków budujących bazy danych i systemy wyszukiwawcze, a nie użytkowników.
- Zakres, czas i sposób wprowadzania zmian w narodowych systemach porządkowania powinien być poddany dyskusji z udziałem wszystkich zainteresowanych, a więc językoznawców, informatyków, pedagogów i ekonomistów.

Literatura

Alphabetical Ordering...

Alphabetical Ordering of Multilingual Terminological and Lexicographical Data Represented in the Latin Alphabet.

ASCII...

ASCII, *American Standard Code for Information Interchange*, 7-bit Coded Characters (ASCII).

Bojar B. 2002

Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych, Wyd. SBP, Warszawa.

X Congreso... 1994

X Congreso de la Asociación de Academias de la Lengua Española (www.spanishpronto.com).

Norma ISO 12199 2000

International Standardisation Organisation. Norma ISO 12199.

Norma ISO 8859-x

International Standardisation Organisation. Norma ISO 8859-x: *8-bit, Single-byte Coded Graphic Character Sets*. ISO 8859-1 - 8859-4 *Latin Alphabet*. Includes Characters Appearing in European Languages Using Modified Latin Alphabet.

Norma ISO 8859-6

ISO 8859-5 „*Latin & Cyrillic Alphabet*”; ISO 8859-6 „*Latin & Arabic Alphabet*”; ISO 8859-7 „*Latin & Greek Alphabet*”; ISO 8859-8 „*Latin & Hebrew Alphabet*”.

Norma ISO/IEC 10646-1 (2000)

Information Technology – Universal Multiple Octet Coded Character Set (UCS).

Meyer Myklestad J., Søråas H. 1955

Damms Lommeordbøker, Engelsk-Norsk, Norsk-Engelsk, Damm & Søn, Oslo.

Podlawska D., Ptóciennik I. 2002

Leksykon nauki o języku, Wyd. Park, Bielsko-Biała.

Unicode Consortium (www.unicode.org)

Wierzbicki J., Rachmanović U., Chlabin J. 1966

Mały słownik serbochorwacko-polski, Wiedza Powszechna, Warszawa.

Ziabicki A. 1988

Automatyczne porządkowanie łańcuchów alfanumerycznych według różnych alfabetów i zasad porządkowania, „Prace Instytutu Podstawowych Problemów Techniki PAN”, nr 40.

Ziabicki A. 1990

Dwustopniowe porządkowanie haseł leksykograficznych w różnych językach, „Prace Instytutu Podstawowych Problemów Techniki PAN”, nr 47.

Ziabicki A. 1992

The Theory of Ordering Lexicographic Entries: Principles, Algorithms and Computer Implementation, „Computers and the Humanities”, t. 26, s. 119–137.

Ziabicki A. 2002

Problem automatycznej transmisji i porządkowania haseł leksykograficznych w zbiorach wielojęzycznych, „Przegląd Biblioteczny”, nr 4.