

ANDRZEJ NIENARTOWICZ, ADAM BARCIKOWSKI,
KAZIMIERZ WIECZORKOWSKI

ZASTOSOWANIE METOD KOMPUTEROWYCH W TAKSONOMII, GEOGRAFII I SOCJOLOGII ROŚLIN

THE APPLICATION OF COMPUTER METHODS IN PLANT TAXONOMY, PLANT GEOGRAPHY AND PHYTOSOCIOLOGY

Wstęp

Zmienność populacji i zbiorowisk roślinnych oraz podobieństwo flory różnych regionów geograficznych mogą być wyrażane za pomocą liczb. Do końca lat pięćdziesiątych bieżącego stulecia w badaniach taksonomicznych, fitogeograficznych i syntaksonomicznych obliczenia wskaźników liczbowych stanowiących wyniki porównań obiektów przyrodniczych wykonywane były przy zastosowaniu prostych urządzeń liczących. Czasochłonność obliczeń, wprost proporcjonalna do liczby analizowanych obiektów oraz niewielkie możliwości umysłu ludzkiego ogarniania fenetycznej różnorodności, ograniczały zakres przeprowadzanych analiz oraz zmniejszały precyzję wyróżniania kategorii obiektów w obrębie badanego materiału. Poszukiwania bardziej wydajnych metod, pozbawionych powyższych ograniczeń, inspirowane ogromem zmienności populacji i zbiorowisk roślinnych oraz potrzebą zwiększenia obiektywności badań, przyczyniły się do rozwoju metod taksonomii numerycznej [57]. Pojawienie się nowych sposobów analizowania zmienności obiektów roślinnych w oparciu o sprzęt komputerowy Feoli i Ganis [17] określili jako naturalną ewolucję sztuki liczb w taksonomii i geografii roślin.

Wraz ze wzrostem stosowania metod numerycznych do analizy coraz większych zbiorów danych botanicznych, na przestrzeni ostatnich trzydziestu lat rozszerzył się zakres zastosowań komputerów na takie obszary działań taksonomów, geografów i socjologów roślin, jak katalogowanie obiektów badań oraz gromadzenie ich opisów. Zbiory danych powiększono o informacje dotyczące przynależności taksonomicznej lub syntaksonomicznej badanych obiektów, lokalizacji geograficznej, warunków siedliskowych ich występowania, pochodzenia, użytkowania, stopnia

zagrożenia działalnością ludzką, a także o dane bibliograficzne. Banki danych stały się podstawą do przeprowadzania rewizji klasyfikacji opracowanych metodami klasycznymi, sporządzania techniką komputerową map rozmieszczania gatunków, opracowań flory oraz list gatunków zagrożonych. Możliwości zastosowań komputerów w taksonomii i geografii roślin przedstawili między innymi Furlow i in. [19] oraz Hawks i in. [26].

W bazach danych botanicznych przechowywane są zazwyczaj informacje o gatunkach. Dane dotyczące zbiorowisk roślinnych opisanych metodą zdjęć fitosocjologicznych wg Braun-Blanqueta, są gromadzone rzadziej. Jest to spowodowane niejednorodnym statusem jaki osiągnęła fitosocjologia w krajach najbardziej rozwiniętych ekonomicznie.

Wraz z powiększaniem informacji zawartej w bazach rozbudowywane są biblioteki programów do analizy zgromadzonych zbiorów danych. Zawierają one często algorytmy ze standardowych bibliotek statystycznych. Jednakże wiele programów opracowanych zostało do rozwiązywania specyficznych problemów taksonomicznych, florystycznych i fitosocjologicznych.

W tworzonych ostatnio systemach informatycznych dąży się do zwiększenia powiązań pomiędzy bazą danych taksonomicznych, bazą danych fitosocjologicznych oraz biblioteką programów. Wzrost integracji systemu przyspiesza jego działanie, a tym samym zmniejsza koszty przetwarzania danych.

Korespondencję pomiędzy zbiorami danych analizowanymi wg algorytmów pochodzących z biblioteki programów, a informacjami zawartymi w bazach danych zapewniają kody gatunków.

Niniejsza praca stanowi ogólny przegląd metod numerycznych stosowanych w taksonomii, geografii i socjologii roślin oraz podaje charakterystykę niektórych banków informacji botanicznej i sposoby kodowania gatunków.

W dalszej części artykułu przedstawiono projekt zintegrowanego systemu informatycznego FITOTAX, obejmującego bazy danych i biblioteki programów, tworzonego w ramach podprogramu CPBP 04.10.01 w Instytucie Biologii UMK.

Metody taksonomii numerycznej

Cele stosowania metod komputerowych i tradycyjnej techniki porównywania obiektów w badaniach taksonomicznych i fitogeograficznych są zbieżne. Feoli i Ganis [17] sformułowali je następująco:

- wyróżnienie jednostek podstawowych, gatunków w taksonomii oraz prowincji i regionów florystycznych w geografii roślin,
- sklasyfikowanie hierarchiczne tych jednostek,
- porządkowanie (ordynacja) jednostek podstawowych,
- określenie zależności ewolucyjnych pomiędzy badanymi obiektami, czyli określenie drogi ewolucji każdego z nich.

Powyższą listę uzupełnić można wymieniając w punkcie drugim możliwość przeprowadzenia klasyfikacji niehierarchicznej oraz podając w dodatkowym punkcie

jeszcze jeden cel, identyfikację obiektów. Rozszerzając listę celów na socjologię roślin, w kolejnych punktach należy wymienić wyróżnianie kategorii w zbiorze zdjęć fitosocjologicznych, przeprowadzanie klasyfikacji i ordynacji wyróżnionych grup, określanie zależności syndynamicznych oraz identyfikowanie zbiorowisk.

Osiągnięcie przy użyciu komputera celów wymienionych w punkcie pierwszym i drugim cytowanej listy wymaga zastosowania metod klasyfikacji numerycznej. Rozwiązywanie trzeciego problemu to stosowanie metod porządkowania numerycznego. Czwarty cel osiągnąć można dzięki metodom kladystyki.

Klasyfikacja numeryczna polega na obliczaniu podobieństwa fenetycznego pomiędzy badanymi obiektami, OTU — Operational Taxonomical Units [56], OGU — Operational Geographical Units [5] lub zdjęciami fitosocjologicznymi. Nie jest ona wolna od zarzutów stawianych tradycyjnym metodom klasyfikacji. Zarzuty te dotyczą zależności wyniku klasyfikacji od doboru cech.

Cechy użyte w klasyfikacji numerycznej stanowią jedynie podzbiór (próbę) ze zbioru wszystkich właściwości porównywanych obiektów [55]. Ten sam zbiór obiektów można sklasyfikować w różny sposób, zależnie od doboru cech oraz zastosowanej metody porównywania i grupowania. Klasyfikacja numeryczna nie określa więc jednoznacznie podobieństwa porównywanych obiektów i w związku z tym każdy jej wynik może być kwestionowany [34]. Precyzyjniej problem ten przedstawia Wanntorp [59] stwierdzając, że tylko niektóre wyniki klasyfikacji numerycznej można zakwestionować na naukowej podstawie. Są to wyniki, których niezgodność z faktami przyrodniczymi wykazano na podstawie testów. W innych wypadkach podstawą kwestionowania wyników klasyfikacji numerycznej mogą być jedynie opinie osobiste, czyjs autorytet lub dążenie do utrzymania jednomyślności [59].

Mimo niejednoznaczności wyników metody klasyfikacji numerycznej, stosowane dziś powszechnie, ułatwiają stawianie hipotez dotyczących zależności pomiędzy badanymi obiektami. Feoli i Ganis [17] stwierdzają, że klasyfikacja numeryczna pozwala wykryć zachowanie się grupy cech względem zbioru innych cech, ułatwia interpretację modeli porządkowania obiektów, pozwala określić siłę dyskryminacyjną cech na różnych poziomach klasyfikacji hierarchicznej oraz wykryć zbiory cech o podobnym zachowaniu we wszystkich badanych obiektach (analiza redundancji). Heywood [25] podaje, że szczególne znaczenie ma znajdowanie korelacji i asocjacji pomiędzy cechami. Według niego znajomość tych zależności ułatwia, po usunięciu redundancji, określenie związku pomiędzy zbiorami cech pozostałych a różnymi przejawami adaptacji ekologicznej gatunków, która może przysłonić zależności genetyczne.

Metody ordynacji numerycznej polegają na znajdowaniu współrzędnych określających położenie badanych obiektów w uproszczonej przestrzeni fenetycznej. Punktem wyjścia w obliczeniach jest macierz odległości obiektów, a współrzędne uzyskuje się poprzez znajdowanie wartości własnych i wektorów własnych symetrycznej macierzy korelacji lub kowariancji cech [40].

Ordynację badanych obiektów przeprowadzoną na podstawie danych fenetycznych porównuje się z wynikiem porządkowania ich w przestrzeni ekologicznej. Na podstawie rozmieszczenia populacji lub zdjęć fitosocjologicznych w obu prze-

strzeniach wnioskować można o wpływie środowiska na różnicowanie się gatunku lub zbiorowisk. O przydatności metod ordynacji numerycznej w badaniach zmienności roślin, szczególnie zmienności klinowej, pisali Jardine i Sibson [27].

Ostatnio metody klasyfikacji i ordynacji numerycznej są stosowane do określania stopnia konwergencji gatunków i zbiorowisk roślinnych. Przedmiotem badań są zbiorowiska rozwijające się w podobnych warunkach ekologicznych na różnych kontynentach. Obserwacje prowadzone są głównie w ramach włosko-kanadyjskiego programu badawczego „Ewolucja konwergentna zbiorowisk roślinnych na terenach odległych” [31, 32, 41]. W badaniach dąży się do określenia metodami numerycznymi podobieństwa homoplastycznego [56], tj. wywołanego procesem ewolucji w podobnych warunkach ekologicznych, po wyeliminowaniu, poprzez pomijanie w zbiorze danych niektórych cech, podobieństwa ojcowskiego [56], wywołanego wspólnym pochodzeniem gatunków.

Zależności ewolucyjne pomiędzy jednostkami taksonomicznymi bada kladystyka numeryczna. Pozwala ona tworzyć klasyfikację filogenetyczną na drodze dedukcji konceptualnej związków ewolucyjnych pomiędzy stanami poszczególnych cech i stosowania modeli algebraicznych, umożliwiających budowanie drzew genealogicznych [13, 56]. Drzewa oddają zależność przodek-potomek w kontekście użytych danych i z reguły różnią się od dendrogramów wyrażających podobieństwo fenetyczne [4].

Rekonstrukcjom ewolucyjnym zarzuca się większy subiektywizm niż klasyfikacjom fenetycznym [17]. Występuje on zarówno przy wyborze cech, jak i w momencie przyjmowania założeń dotyczących związku każdej cechy z procesami ewolucyjnymi. Klasycznym przykładem przyjmowania takich założeń jest np. zależność pomiędzy stopniem poliploidalności gatunków a czasem ewolucyjnym ich rozwoju. Jednakże kladystyka numeryczna jako narzędzie symulujące myślenie, poprzez stosowanie modeli matematycznych opartych na decyzjach biologicznych, tworzy drzewa genealogiczne zgodne z logicznymi zasadami rozumowania [17].

Kladystykę numeryczną do opracowania klasyfikacji filogenetycznej gatunków stosowali Duncan [11], Estabrook i Anderson [14], La Duke i Crawford [29], Seaman i Funk [53], Stuessy [58]. Jej metody mogą być również stosowane w geografii roślin. Wymaga to jednak opracowania ogólnych zasad określania rangi każdego gatunku lub innych cech adaptacyjnych flory i roślinności, w poszczególnych etapach ewolucji roślin w aspekcie geograficznym [17].

Identyfikację okazów botanicznych lub zbiorowisk roślinnych przy użyciu komputera przeprowadzić można na dwa sposoby. Pierwszy sposób, numeryczny, polega na porównywaniu stanów cech badanego obiektu w stosunku do stanu cech poszczególnych kategorii obiektów. Wykonuje się to metodami zaczerpniętymi z teorii informacji, np. obliczając tzw. rozbieżność informacyjną [40].

Drugi sposób polega na oznaczaniu okazu według klucza dichotomicznego, wygenerowanego przez komputer na podstawie wiedzy o poszczególnych taksonach, zgromadzonej w jego pamięci masowej. Metody identyfikacji okazów wg kluczy na drodze dialogu z komputerem rozwinęli Dallwitz [9], Goodall [20], Hall [23], Johnston [28], Morse [35, 36], Morse et al. [37], Pankhurst [42, 43, 45, 46, 47], Pankhurst i Aitchison [48], Shelter [54]. Przykłady zastosowań metod

tego typu przedstawili: Ceska i Trempour [1] do prowadzenia identyfikacji w obrębie rodziny skalnicowatych, Forget i in. [18] do oznaczania drzew rosnących w Gujanie Francuskiej, Watson i Milne [61] do identyfikowania gatunków traw. Najbardziej znane programy (systemy ekspertowe) do identyfikowania okazów botanicznych to XPER [18] i GENKEY [49].

Bazy danych

Tworzenie banków informacji botanicznej na szerszą skalę rozpoczęto na początku lat siedemdziesiątych. Jednym z pierwszych przedsięwzięć było opracowanie flory Ameryki Północnej. System „Flora of North America” opisał Shelter [54]. Baza danych zawiera opisy około 10 tys. taksonów.

Nieco później w Meksyku bank danych „Flora of Veracruz” obejmujący opisy około 8 tys. gatunków należących do 250 rodzin [47]. Informacje o ponad 300 rodzajach traw umieszczono w banku danych zorganizowanym w końcu lat siedemdziesiątych w Australii [60]. W Republice Południowej Afryki powstała baza danych taksonomicznych PRECIS, a przy stanowym ogrodzie botanicznym w Missouri, USA, baza danych o roślinach tropikalnych TROPICOS. W kooperacji, głównie Wielkiej Brytanii i Stanów Zjednoczonych Ameryki Północnej, utworzono bazę danych ILDIS, zawierającą informacje o roślinach strączkowych [2]. W latach siedemdziesiątych skomputeryzowano też szereg kolekcji zielnikowych [22, 24, 38].

Jedną z pierwszych organizacji międzynarodowych, które przystąpiły do tworzenia baz danych botanicznych na dużą skalę była Międzynarodowa Unia Ochrony Przyrody i Zasobów Naturalnych (IUCN). Już w 1975 r. baza danych tej organizacji, tworzona w Cambridge i w Królewskim Ogrodzie Botanicznym Kew, w Wielkiej Brytanii, obejmowała informacje o około 5 tys. gatunkach roślin endemicznych i zagrożonych działalnością człowieka.

Również w Polsce w połowie lat siedemdziesiątych przystąpiono do tworzenia baz danych botanicznych. Zakres informacji zgromadzonej w jednej z największych baz danych ATPOL opracowanej w Instytucie Botaniki UJ w Krakowie, przedstawili ostatnio Zajęc i in. [63].

W trakcie organizowania baz opracowano szereg standardowych formatów zapisywania danych taksonomicznych [8, 37, 44].

Do zarządzania bazami danych wykorzystywano uniwersalne systemy informacyjne, jak np. ISIS, oraz specjalnie utworzone dla zbiorów danych taksonomicznych, np. TAXIR [15, 16].

Gwałtowny rozwój baz danych nastąpił na początku lat osiemdziesiątych w Europie Zachodniej. W roku 1983 Rada Europejska EWG zainicjowała program organizowania banków danych dla zwiększenia efektywności działań na rzecz ochrony środowiska. W katalogu opracowanym po drugim kolokwium poświęconym problematyce komputeryzowania informacji o środowisku przyrodniczym, które odbyło się w Strasburgu w roku 1985, wymieniono 70 banków danych tworzonych w ramach tego programu w 14 krajach Europy Zachodniej oraz w Nairobi. W 22

bankach gromadzone są informacje dotyczące flory i roślinności. Największą ilość informacji z tego zakresu zawierają: bank Uniwersytetu Rzymskiego, System Europejskiej Dokumentacji Taksonomicznej, Florystycznej i Biosystematycznej (EDS) w Reading, bank danych Szwedzkiego Narodowego Komitetu Ochrony Środowiska (NSEPB) w Uppsali oraz bank danych CMC Międzynarodowej Unii Ochrony Przyrody i Zasobów Naturalnych.

Baza danych Uniwersytetu Rzymskiego zawiera m. in. dane chorologiczne i ekologiczne dla ponad 6 tys. gatunków roślin. System EDS obejmuje dane dotyczące przynależności systematycznej, rozmieszczenia geograficznego i ekologii gatunków występujących na terenie Europy. W bazie zgromadzono też dane cytologiczne, biochemiczne, bibliograficzne oraz listę synonimów. Szwedzką bazę danych tworzą informacje dotyczące rozmieszczenia gatunków, struktury zbiorowisk roślinnych oraz siedlisk na terenie Skandynawii.

W banku danych CMC (Conservation Monitoring Centre) dane taksonomiczne gromadzone są głównie przez jednostkę ds. roślin zagrożonych TPU (Threatened Plant Unit) i przez jednostkę ds. terenów chronionych (Protected Areas Data Unit). Baza ma charakter globalny i obejmuje informacje o florze regionów geograficznych, dane o gatunkach, jak rozmieszczenie, status ochrony, występowanie w kolekcjach ogrodów botanicznych, znaczenie ekonomiczne gatunku. Do roku 1986 zgromadzono tu opisy ponad 41 tys. taksonów roślin, w tym ponad 17 tys. wymierających i zagrożonych.

Z centrum informatycznym Międzynarodowej Unii Ochrony Przyrody i Zasobów Naturalnych współpracuje ściśle 138 ogrodów botanicznych. W 1986 roku w wydawnictwach IUCN podano, że 29% z nich jest skomputeryzowanych, a dalsze 53% organizuje bazy danych botanicznych. Dla ujednoczenia zapisu informacji o ogrodach botanicznych Cullen i in. [6] opracowali międzynarodowy, standardowy format danych ITF (International Transfer Format). Do roku 1986 ponad 250 ogrodów botanicznych nadesłało do CMC informacje o swoich kolekcjach wg tego systemu.

Ostatnio do CMC przenosi się do pamięci komputerów informacje zawarte w czerwonych księgach roślin wymierających i zagrożonych, opracowanych w poszczególnych krajach [30]. Przystąpiono też do tworzenia systemu PECS (Plant Existence Category System). Jest to propozycja nowego sposobu zapisywania informacji o zasięgu występowania gatunku, jego pochodzeniu, użytkowaniu, stopniu zagrożenia i ewentualnej przynależności do endemitów.

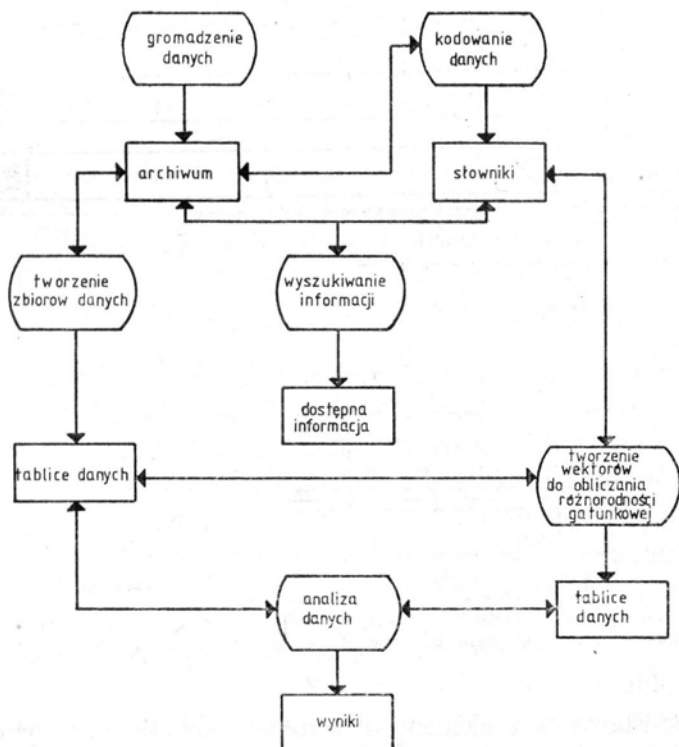
Baza danych IUCN wchodzi w skład Globalnego Systemu Informacji o Środowisku (GEMS), który utworzony został w ramach Programu Środowiskowego Narodów Zjednoczonych (UNEP). W skład tego systemu wchodzi również rozwijana w Nairobi baza GRID (Global Resource Information Data Base), zawierająca szereg danych florystycznych.

Poza programem zainicjowanym przez Radę Europejską EWG rozwijają się największe europejskie banki danych fitosocjologicznych. Są to banki danych Uniwersytetu w Nijmegen oraz Uniwersytetu w Trieście. W obydwu zgromadzono znaczne zbiory

zdjęć fitosocjologicznych i opracowano oprogramowanie pozwalające analizować duże zbiory danych. W banku danych Uniwersytetu w Trieście przechowywane są też informacje dotyczące ponad 9 tys. gatunków roślin [50].

System FITOTAX

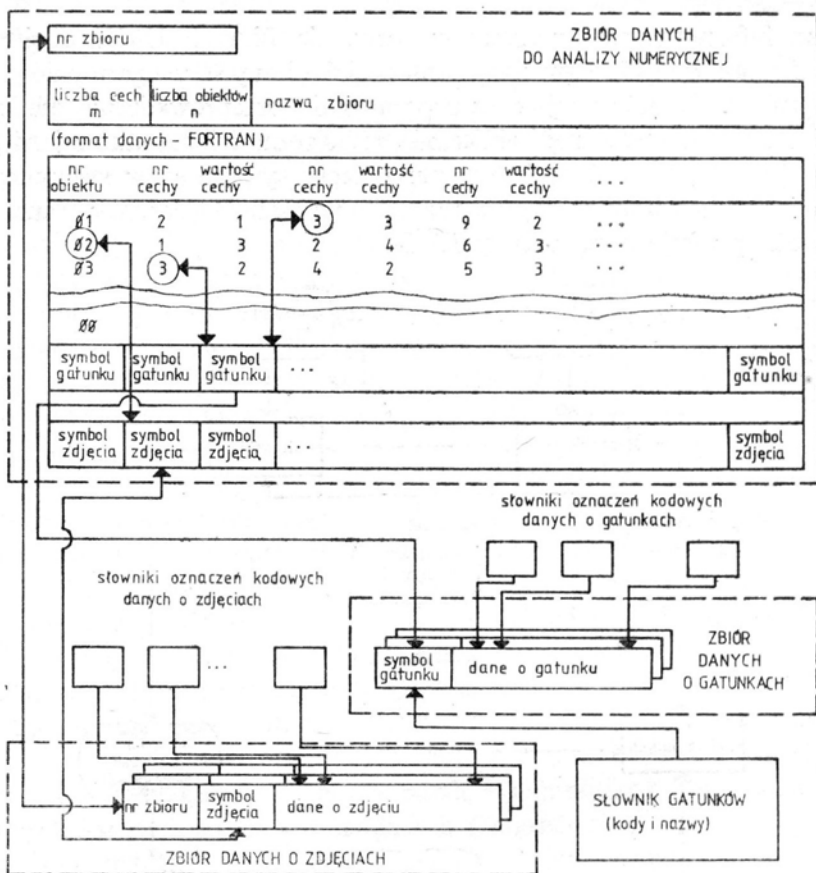
System informatyczny tworzony w Instytucie Biologii UMK obejmuje bazę danych taksonomicznych, bazę danych fitosocjologicznych, programy do gromadzenia i wyszukiwania informacji oraz programy do przeprowadzenia analizy metodami taksonomii numerycznej i obliczania różnorodności gatunkowej zdjęć i flory regionów (ryc. 1). Bazy danych systemu przechowywane są w komputerze Riad 32, natomiast zbiory informacji gromadzonych na bieżąco są archiwizowane w mikrokomputerach profesjonalnych klasy IBM-PC.



Ryc. 1. Sieć funkcjonalna systemu informatycznego do analizy danych taksonomicznych, fitogeograficznych i fitosocjologicznych wg Feoliego i Ganis [17]

Punktem wyjścia przy rozbudowywaniu software systemu jest baza fitosocjologiczna. Sposób zapisu informacji o składzie gatunkowym i ilościowym gatunków w zdjęciach fitosocjologicznych jest akceptowany przez programy numeryczne. Z drugiej strony baza fitosocjologiczna ma łączność z bazą danych taksonomicz-

nych. Integrację obu zapewniają identyfikatory kodowe gatunków. Taka organizacja systemu przyspiesza przesyłanie zbiorów informacji zawartych w bazach do obliczeń. Usprawnia też proces powiększania bazy o nowe informacje. Opisy nowych obiektów wprowadzane do obliczeń numerycznych są równocześnie, w niezmięnionej postaci zapisywane w bazach danych.



Ryc. 2. Rola oznaczeń kodowych gatunków w integrowaniu zbiorów danych systemu FITOTAX

Spójność podstawowych elementów systemu pozwala na szybkie tworzenie podzbiorów zdjęć, okazów lub populacji charakteryzujących się jednakowym stanem określonej cechy lub grupy cech. Podbiory te można przesłać do ponownych obliczeń.

Przy wprowadzaniu nowego zbioru zdjęć do pamięci komputera zastosowano sposób przygotowania danych stosowany w programach opracowanych w Uniwersytecie Cornell w stanie Nowy Jork w USA. Zbiór danych składa się z trzech części (ryc. 2). Część pierwsza zawiera informacje o składzie gatunkowym. Na początku każdego wiersza zapisany jest numer zdjęcia, a na następnych polach podaje

się na przemian numery gatunków nadane wg kolejności ich występowania w tabeli fitosocjologicznej oraz stopień pokrycia gatunku. Gatunki występujące w tabeli, lecz nie występujące w danym zdjęciu, opuszcza się. Zdjęcia zbioru i gatunki w poszczególnych zdjęciach muszą być uszeregowane wg rosnących numerów.

W drugiej części zbioru danych zapisane są ośmiocyfrowe oznaczenia kodowe gatunków, po 10 w wierszu. Kod każdego gatunku zapisany jest w polu, którego numer jest równy numerowi gatunku w części pierwszej zbioru danych.

Część trzecia zawiera ośmiocyfrowe symbole zdjęć fitosocjologicznych. Podobnie jak w wypadku oznaczeń gatunków wymagana jest zgodność numeru pola, na którym zapisany jest identyfikator kodowy z numerem tego zdjęcia w części pierwszej zbioru danych.

Poszczególnym zbiorom (tabelom fitosocjologicznym) nadawane są pięciocyfrowe numery wg kolejności wprowadzania do bazy danych. Baza organizowana jest więc na 99 999 zbiorów.

Kodowanie gatunków można przeprowadzić na trzy sposoby [50]. Może to być kodowanie *ad hoc*, bez jakiegokolwiek zasady logicznej. Polega ono na przyporządkowywaniu gatunkom kolejnych cyfr. Taką metodę stosuje się dla konkretnego zadania i praktycznie nie ma możliwości zastosowania tego samego kodu dla innego zbioru danych.

Dla archiwowania zbiorów w bazach należy przeprowadzić kodowanie oparte na zasadach logicznych. W tym wypadku można utworzyć zamknięty lub otwarty system kodów. System zamknięty oznacza, że zbiór kombinacji cyfr, które można utworzyć tym sposobem kodowania jest skończony, przyjęty *a priori*. W drugim przyjmuje się *a priori* zasady kodowania i istnieje możliwość rozszerzenia kodu na dane nawet nie przewidywane [50].

W bazach zastosowano zamknięty system kodów, przewidziany na zaszyfrowanie nazw gatunków roślin Europy Środkowej. Przyjęty sposób kodowania pozwala utworzyć liczbę możliwych kombinacji cyfr znacznie większą od liczby gatunków występujących na tym obszarze. Listę gatunków roślin naczyniowych dla baz danych przygotował Rutkowski [52]. W wypadku mchów przyjęto listę gatunków, którą opracowali Ochyra i Szmajda [39]. Listę wątrobowców ustalono na podstawie opracowania Duella [10], a przy sporządzaniu listy porostów wykorzystano literaturę, której Cieśliński i in. [3] użyli przy opracowywaniu nomenklatury wymierających i zagrożonych w Polsce gatunków porostów.

Zastosowane w systemie FITOTAX oznaczenia kodowe są zbliżone do kodów gatunków Ehrendorfera [12]. Dwie pierwsze pozycje kodu są zarezerwowane dla oznaczenia klas (od 01 do 50). Na następnych trzech polach zapisany jest symbol rodzaju, a na dalszych dwóch, gatunków. Ósmy znak kodu oznacza niższą od gatunku jednostkę taksonomiczną. Zastosowane identyfikatory kodowe w porównaniu do pięciocyfrowych symboli nazw gatunków wg Ehrendorfera [12] charakteryzują się większą zawartością informacyjną (pozycja pierwsza i druga) oraz oznaczeniem zróżnicowania wewnątrzgatunkowego oddzielną cyfrą (pozycja ósma). Jest on znacznie krótszy od kodów wg *Flora Europaea* (12 cyfr) i kodu Goulda [21], który zawiera 17 cyfr. Pod względem długości jest zbliżony do siedmiocyfrowego kodu sto-

sowanego w bazie danych Uniwersytetu w Trieście. W tym systemie rodzaje, przyjęte wg opracowania Dalla Torre i Harmsa [7], są kodowane czterema cyframi, a gatunki oznaczone są trzema cyframi wg kodu *Flora Europaea*. W kodzie z Triestu brak jest informacji o przynależności gatunku do wyższej jednostki taksonomicznej, co np. Pignatti [50] uznał za wadę kodu Ehrendorfera [12].

W części trzeciej zbioru danych oznaczenia zdjęć fitosocjologicznych mogą być dowolne, jednakże przewidziane jest umieszczenie tu cyfr oznaczających miejsce zdjęcia w systemie „Zbiorowisk Roślinnych Polski” Matuszkiewicza [33]. W tym celu na oznaczenie klasy zespołów należy użyć trzech cyfr, po jednej na oznaczenie rzędu i związku zespołów oraz dwie cyfry na oznaczenie zespołu. W wypadku braku możliwości zaklasyfikowania zbiorowiska do zespołu umieszczane są tylko numery wyższych jednostek syntaksonomicznych.

Zbiory zdjęć fitosocjologicznych są połączone poprzez kody ze zbiorami informacji o gatunkach i zdjęciach fitosocjologicznych. Informacje o zdjęciach są zapisywane po numerze zbioru, w którym znajduje się to zdjęcie i jego oznaczeniu kodowym. Obejmują one wszystkie dane zawarte w nagłówkach tabel fitosocjologicznych. Dodatkowo na odpowiednich polach rekordów zapisane są informacje dotyczące autorstwa, miejsca publikowania zdjęcia, lokalizacji fizycznogeograficznej, fitogeograficznej, wg jednostek podziału administracyjnego i wg jednostek gospodarczych, np. nadleśnictw. Przewiduje się też możliwość określenia miejsca wykonywania zdjęcia w sieci stałych kwadratów stosowanej w „Systemie Informacji o Środowisku Przyrodniczym” (podprogram CPBP 04.10.12).

Zbiory informacji o gatunkach zawierają informacje z zakresu morfologii, anatomii, cytologii, biochemii, ekologii, chorologii, zapisane w postaci ciągów symboli po oznaczeniu kodowym gatunku. W zbiorach danych ekologicznych wykorzystano ekologiczne liczby wskaźnikowe Zarzyckiego [64].

Dzięki korespondencji pomiędzy bazami danych istnieje możliwość uzyskania spektrum form życiowych roślin lub map rozmieszczenia gatunków występujących w analizowanym zdjęciu lub florze określonego terenu. W wypadku wyszukiwania w bazie fitosocjologicznej zdjęć charakteryzujących się obecnością konkretnego gatunku sprawdzana jest najpierw część druga każdego zbioru danych (ryc. 2). Przy stwierdzeniu poprzez identyfikator kodowy obecności szukanego gatunku rozpoznawany jest jego numer w części pierwszej zbioru danych, wg numeru pola. Numer ten jest następnie poszukiwany w kolejnych zdjęciach. Wyszukiwanie informacji o składzie gatunkowym konkretnego zdjęcia rozpoczyna się odnajdywaniem w zbiorze danych o zdjęciach fitosocjologicznych jego kodu i numeru zbioru, w którym znajduje się to zdjęcie. Następnie poszukiwana jest część trzecia tego zbioru danych. Numer pola, na którym zapisany jest poszukiwany numer kodowy, informuje o miejscu tego zdjęcia w części pierwszej zbioru danych.

System FITOTAX obejmuje ponadto listę synonimów i skrótów nazw gatunkowych oraz zbiory słownikowe, tj. listy wszystkich symboli z wyjaśnieniem ich znaczenia.

Zasadniczą część biblioteki programów tworzonego systemu stanowią specjalistyczne programy do badań taksonomicznych i syntaksonomicznych opracowane

w Uniwersytecie Cornella (COMPCLUS, DECORANA, ORDIFLEX, TWINS-PAN), Uniwersytecie w Nijmegen (CLUSLA), Uniwersytecie w Trieście (IAHOPA, THREE-PA), oraz programy, które napisali Podany [51], Wildi i Orłóci [62]. Bibliotekę uzupełniają programy opracowane w Ogólnouczelnianym Ośrodku Obliczeniowym UMK oraz pakiety statystyczne SSP, SPSS i inne.

Wymienione programy pozwalają analizować duże zbiory danych fitosocjologicznych, np. przy pomocy programu DECORANA można przeprowadzić ordynację 1600 zdjęć zawierających 800 gatunków. Program ten zastosowano m. in. do analizy zbiorowisk halofilnych Polski, zbioru obejmującego 579 zdjęć fitosocjologicznych i około 400 taksonów. Wyniki porządkowania zdjęć pozwoliły na bardziej obiektywne ustosunkowanie się do klasyfikacji tej grupy zbiorowisk roślinnych opracowanej klasyczną metodą Braun-Blanqueta.

Przy pomocy programu DECORANA przeprowadzono również ordynację taksonów występujących w zbiorowiskach halofilnych. W oparciu o wyniki obu analiz porównano poszczególne gatunki halofitów pod względem zakresu i optimum występowania w stosunku do natężenia jednego z głównych czynników warunkujących ich rozwój, zasolenia gleby.

Program DECORANA wykorzystywany jest również do analizy zmian struktury runa rezerwatu Las Piwnicki. Przetwarzane są informacje zebrane na 970 stałych powierzchniach badawczych rozmieszczonych regularnie na terenie obiektu badań.

Wyniki analizy numerycznej zbiorowisk halofilnych Polski oraz wyniki badań prowadzonych w rezerwacie Las Piwnicki będą przedmiotem odrębnych publikacji.

Wnioski

Metody numeryczne nie były dotychczas zbyt często stosowane w polskiej taksonomii, geografii i socjologii roślin. Ich zalety, jak możliwość wielokrotnego analizowania dużych zbiorów danych, powoduje, że zainteresowanie nimi ciągle rośnie. Do ich upowszechnienia przyczyni się zapewne wzrost dostępności sprzętu komputerowego notowany w Polsce w ostatnich latach. Stosowanie metod numerycznych obok tradycyjnych spowoduje, że analiza struktury szaty roślinnej i zachodzących w niej procesów będzie wszechstronniejsza.

Metody komputerowe stwarzają też szansę usprawnienia przepływu informacji pomiędzy różnymi ośrodkami botanicznymi w kraju. Wymaga to jednak opracowania zunifikowanego systemu gromadzenia i analizowania danych, w oparciu o który prowadzona byłaby wymiana informacji przenoszonych na dyskietkach, taśmach magnetycznych, a być może, w dalszej przyszłości, również za pośrednictwem Krajowej Akademickiej Sieci Komputerowej. Przedstawiony system FITO-TAX może stanowić jedno z rozwiązań.

Dr Andrzej Nienartowicz,

dr Adam Barcikowski

Zakład Taksonomii, Ekologii Roślin i Ochrony Przyrody
Instytut Biologii Uniwersytetu Mikołaja Kopernika, ul. Gagarina 9
87-100 Toruń

Mgr inż. Kazimierz Wieczorkowski
Ogólnouczelniany Ośrodek Obliczeniowy UMK
ul. Grudziądzka 5, 87-100 Toruń

LITERATURA

- [1] Ceska A., Trumppour A. D., 1979, Computer editing of serial and indented identification keys. *Taxon* 28 : 329—335.
- [2] Charlwood B. V., Morris G. S., Grenham M. J., 1984. A chemical database for the *Leguminosae*. In: Allkin R., Bisby F. A. (eds). *Databases in systematics*. Academic Press, London, 201—208.
- [3] Cieśliński S., Czyżewska K., Fabiszewski J., 1986. Czerwona lista porostów zagrożonych w Polsce. In: Zarzycki K., Wojewoda W. (eds). *Lista roślin wymierających i zagrożonych w Polsce*. PWN, Warszawa, 83—107.
- [4] Crovello T. J., 1976. Numerical approaches to the species problem. *Plant Syst. Evol.* 125 : 179—187.
- [5] Crovello T. J., 1981. Quantitative biogeography: an overview. *Taxon* 30 : 563—575.
- [6] Cullen J., Lear M., Mackinder D. C., Syngé H., 1986. Principles and standards for the computerization of garden record schemes, as applied to conservation, with proposals for an International Transfer Format (A contribution to the botanic gardens Conservation Strategy). In: Bramwell D., Hamann O., Heywood V. H., Syngé H. (eds.) *Botanic Gardens and the World Conservation Strategy*. Academic Press, London.
- [7] Dalla Torre C. G., Harms H. 1900—1907. *Genera Siphono-garnanem*. Lipsiae, 2 vol. Reprint Wiesbaden (1963).
- [8] Dallwitz M. J., 1980. Usher's guide to the DELTA System. A general system for coding taxonomic descriptions. CSIRO Division of Entomology Report No 13, Canberra, Australia, 71 pp.
- [9] Dallwitz M. J., 1984. Automatic typesetting of computer — generated keys and descriptions. In: Allkin R., Bisby F. A. (eds.) *Databases in systematics*. Academic Press, London, 279—290.
- [10] Duell R., 1983. Distribution of the European and Macaronesian liverworts (Hepatophytina). *Bryologische Beiträge* 2 : 1—114.
- [11] Duncan T., 1980. A cladistic analysis of the *Ranunculus hispidus* complex. *Taxon* 29 : 441—454.
- [12] Ehrendorfer F., 1973. *Liste der Gefäßpflanzen Mitteleuropas*. Granz 2. Aufl. Stuttgart, G. Fisher Verlag, 318 pp.
- [13] Estabrook G. F., 1972. Cladistic methodology: a discussion of the theoretical basis for the induction of evolutionary history. *Ann. Rev. Ecol. Syst.* 3 : 427—456.
- [14] Estabrook G. F., Brill R. C., 1969. The theory of the Taxir accessioner. *Mathem. Biosc.* 5 : 327—340.
- [15] Estabrook G. F., Anderson W. R., 1979. An estimate of philogenetic relationships within genus *Crusea* (*Rubiaceae*) using character compatibility analysis. *Syst. Bot.* 3 : 179—196.
- [16] Estabrook G. F., Brill R. C., 1984. Management of almost flat files in systematic Biology using TAXIR. In: Allkin R., Bisby F. A. (eds.) *Databases in systematics*. Academic Press, London, 53—67.
- [17] Feoli E., Ganis P., 1984. On the application of numerical and computer methods in plant taxonomy and plant geography: an integrated information system for data banking and numerical classifications and ordinations. *Webbia* 38 : 165—184.
- [18] Forget P. M., Lebbe J., Puig H., Vignes R., Hideux M., 1986. Microcomputer-aided identification: an application to trees from French Guiana. *Botanical Journal of the Linnean Society*. 93 : 205—223.

- [19] Furlow J. J., Morse L. E., Beaman J. H., 1971. Computers in biological systematics, a new university course. *Taxon* 20 : 283—290.
- [20] Goodall D. W., 1968. Identification by computer. *Bi-Science* 18 : 485—488.
- [21] Gould S. W., 1962. Family names of the plant kingdom. New Haven, New York.
- [22] Greene D. M., 1972. A taxonomical data bank and retrieval system for a small herbarium. *Taxon* 21 : 621—630.
- [23] Hall A. V., 1970. A computer-based system for forming identification keys. *Taxon* 19 : 12—18.
- [24] Hall A. V., 1972. Computer-based data banking for taxonomic collections. *Taxon* 21 : 13—25.
- [25] Heywood V. H., 1973. Ecological data in practical taxonomy. In: Heywood V. H. (ed). *Taxonomy and ecology*. Academic Press, London, New York.
- [26] Hawks J. G., Kershaw B. L., Readett R. C., 1968. Computer mapping of species distributions in a county flora. *Watsonia* 6 : 350—364
- [27] Jardine N., Sibson R., 1971. *Mathematical taxonomy*. Wiley, London, New York.
- [28] Johnston B. C., 1980. Computer programs for constructing polyclave key from data matrices. *Taxon* 29 : 47—51.
- [29] La Duke J. C., Crawford D. J., 1979. Character compatibility and phyletic relationships in several closely related species of *Chenopodium* of Western United States. *Taxon* 28 : 307—314.
- [30] Leon Ch., 1986. From RDBs to databases. *Threatened Plants Newsletter* 16 : 20.
- [31] Lausi D., Nimis P. L., 1985. The study of convergent evolution in plants and plant communities. A quantitative approach. *Abstracta Botanica* 9 : 67—77.
- [32] Lausi D., Nimis P. L., 1986. Leaf and canopy adaptations in a high-elevation desert on Tenerife, Canary Islands. *Vegetatio* 68 : 19—31.
- [33] Matuszkiewicz W., 1982. *Przewodnik do oznaczania zbiorowisk roślinnych Polski*. PWN, Warszawa, 298 pp.
- [34] Meeuse A. D. J., 1981. Again: cladistics in botany. *Taxon* 30 : 642—644.
- [35] Morse L. E., 1968. Construction of identification keys by computer. *Amer. Jour. Bot.* 55 : 737.
- [36] Morse L. E., 1971. Specimen identification and key construction with timesharing computers. *Taxon* 20 : 269—282.
- [37] Morse L. E., Peters J. A., Hamel P. B., 1971. A general data format for summarizing taxonomic information. *Bio-Science* 21 : 174—180.
- [38] Morris J. W., Leistner O. A., 1975. Computerization of the National Herbarium, Pretoria. *Taxon* 24 : 261—270.
- [39] Ochyra R., Szmajda P., 1978. An annotated list of Polish mosses. *Fragm. Flor. et Geobot.* 24 : 93—145.
- [40] Orlóci L., 1978. *Multivariate analysis in vegeration research*. 2nd ed. Junk, Den Hag, 451 pp.
- [41] Orlóci L., Feoli E., Lausi D., Nimis P. L., 1986. Estimation of character structure convergence (divergence) in plant communities: a nested hierarchical model. *Coenoses* 1 : 11—20.
- [42] Pankhurst R. J., 1970. A computer program for generating diagnostic keys. *Computer Journal* 13 : 145—151.
- [43] Pankhurst R. J., 1971. Botanical keys generated by computer. *Watsonia* 8 : 357—368.
- [44] Pankhurst R. J., 1978a. Taxonomic data format, version 2. Botany Dept., British Museum (Nat. Hist.), London.
- [45] Pankhurst R. J., 1978b. The printing of taxonomic descriptions by computer. *Taxon* 27 : 65—68.
- [46] Pankhurst R. J., 1978c. Biological identification. The principle and practice of identification methods in Biology. Edward Arnold, London.
- [47] Pankhurst R. J., 1983. The construction of floristic database. *Taxon* 32 : 193—202.
- [48] Pankhurst R. J., Aitchison R. R., 1975. An on-line identification. In: Pankhurst R. J., (ed). *Biological identification with computers*. Academic Press, London, 181—196.
- [49] Payne R. W., 1975. GENKEY: a program for constructing diagnostic keys. In: Pankhurst R. J. (ed). *Biological identification with computers*, Academic Press, London, 65—72.
- [50] Pignatti S., 1976. A system for coding plant species for dataprocessing in phytosociology. *Vegetatio* 33 : 23—32.

- [51] Podanyi J. 1984. SYN-TAX II. Computer programs for data analysis in ecology and systematics. *Abstracta Botanica* 8 : 73—94.
- [52] Rutkowski L., 1986. Lista roślin naczyniowych Polski. Manuskrypt, Toruń.
- [53] Seaman F. C., Funk V. A., 1983. Cladistic analysis of complex natural products: developing transformation series from sesquiterpene lactone data. *Taxon* 32 : 1—27.
- [54] Shelter S. G., 1975. A generalized descriptive data bank as a basis for computer assisted identification. In: Pankhurst R. J. (ed). *Biological identification with computers: 197—235*. Academic Press, London, New York, 197—235.
- [55] Sneath P. H. A., 1976. Phenetic taxonomy at the species level and above. *Taxon* 24 : 447—450.
- [56] Sneath P. H. A., Sokal R. R., 1973. *Numerical taxonomy*. Freeman W. H., San Francisco.
- [57] Sokal R. R., Sneath P. H. A., 1963. *Principles of numerical taxonomy*. Freeman W. H., San Francisco.
- [58] Stuessy T. F., 1979. Cladistics of *Melampodium* (*Compositae*) *Taxon* 28 : 385—389.
- [59] Wannrop H. E., 1983. Cladistics misunderstood — Again. *Taxon* 32 : 97—98.
- [60] Watson L., Dallwitz M. J., 1980. Australian grass genera, Anatomy, morphology and keys. Research School of Biological Sciences, Australian National University, Canberra, 209 pp.
- [61] Watson L., Milne P., 1972. A flexible system for automatic generation of species-purpose dichotomous keys, and its application to Australian grass genera. *Austral. Jour. Bot.* 20 : 331—352.
- [62] Wildi O., Orlóci L., 1983. *Management and multivariate analysis of vegetation data*. 2nd ed. Swiss Federal Institute of Forestry Research, Birmensdorf. Reports no 215, 139 pp.
- [63] Zając A., Kotońska B., Zając M. 1987. Stan prac nad „Atlasem rozmieszczenia roślin naczyniowych w Polsce” (ATPOL) w końcu 1985 roku. *Wiadomości Botaniczne* 31 : 109—123.
- [64] Zarzycki K., 1984. *Ekologiczne liczby wskaźnikowe roślin naczyniowych Polski*. PAN — Instytut Botaniki. Kraków, 45 pp.