
Katedra Automatyki, Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

mgr Maria Zając

**Zastosowanie algorytmów rozpoznawania obrazów
do indywidualizacji procesu nauczania w multimedialnych
systemach komputerowych**

ROZPRAWA DOKTORSKA

Promotor:
Prof. zw. dr hab. inż. Ryszard Tadeusiewicz

Kraków 1998

Spis treści.

1	<i>Wstęp.</i>	1
2	<i>Idea i elementy składowe inteligentnego systemu komputerowego wspomagającego nauczanie.</i>	5
2.1	Ogólna struktura systemu.	5
2.2	Moduł gromadzenia danych o uczniu.	7
2.3	Moduł rozpoznawania.	8
2.4	Moduł nauczający.	10
2.4.1	Struktura modułu nauczającego.	10
2.4.2	Baza wiedzy.	10
2.4.3	Układ sterujący.	11
2.4.4	Ilość planowanych metod nauczania.	11
2.4.5	Charakterystyka proponowanych metod nauczania.	14
3	<i>Przedmiot badań w ujęciu psychologicznym.</i>	17
3.1	Rola inteligencji w procesach poznawczych.	17
3.2	Pojęcie uczenia się.	19
3.3	Uczenie się wspomagane komputerowo.	21
3.3.1	Tradycyjne formy wykorzystania komputerów w edukacji.	21
3.3.2	Rola aplikacji multimedialnych w nauczaniu.	21
3.3.3	Perspektywy wykorzystania sieci Internet w uczeniu się i nauczaniu.	22
3.4	Pomiar w psychologii.	23
3.4.1	Skale pomiarowe.	23
3.4.2	Stosowane metody pomiaru.	23
3.5	Charakterystyka i przedmiot badań.	25
3.5.1	Opis modelu badawczego.	25
3.5.2	Założenia i sposób doboru próby.	26
3.5.3	Wykorzystane narzędzia i skale pomiarowe.	27
4	<i>Przestrzeń cech.</i>	29
4.1	Charakterystyka badanych obiektów.	29
4.2	Ujęcie matematyczne.	31
4.3	Kodowanie danych.	33
4.3.1	Kodowanie 1-z-N.	33
4.3.2	Kodowanie z wykorzystaniem szeregów kumulacyjnych.	34
4.4	Redukcja wymiaru przestrzeni cech.	36
4.4.1	Uporządkowanie cech według rosnącej ilości informacji.	37
4.4.2	Redukcja wymiaru przestrzeni przez wybór cech reprezentantek.	37
4.4.2.1	Metoda taksonomii struktur.	38
4.4.2.2	Zastosowanie analizy skupień.	40
4.4.3	Redukcja wieloetapowa.	40
4.5	Ilustracja graficzna wybranych zestawów cech.	43
5	<i>Rozpoznawanie.</i>	46

5.1 Ciąg uczący.	46
5.1.1 Definicja ciągu uczącego.	46
5.1.2 Klasyfikacja elementów ciągu uczącego.	47
5.1.2.1 Klasyfikacja na podstawie oceny nauczyciela.	47
5.1.2.2 Indywidualna opinia ucznia.	47
5.1.2.3 Określanie przynależności obiektów poprzez analizę skupień	48
5.1.2.4 Ilość klas obiektów.	48
5.1.2.5 Przyporządkowanie numerów klas grupom obiektów.	49
5.2 Algorytm rozpoznawania.	50
5.2.1 Odwzorowanie obiektów w przestrzeń cech – recepcja.	51
5.2.2 Funkcje przynależności.	51
5.2.2.1 Metryki w przestrzeniach liczbowych.	53
5.2.2.2 Metryki dla danych jakościowych.	55
5.2.2.3 Porównanie uzyskanych rezultatów rozpoznawania w zależności od przyjętej metryki.	55
5.2.3 Funkcje decyzyjne.	56
5.3 Wykorzystane metody rozpoznawania.	56
5.3.1 Metody minimalnoodległościowe.	57
5.3.1.1 Metoda najbliższego sąsiada (NN).	57
5.3.1.2 Metoda α -najbliższych sąsiadów (α NN).	58
5.3.1.3 Metoda j_N -tego najbliższego sąsiada (j_N NN).	59
5.3.2 Metoda najbliższej mody (NM) jako przykład metody wzorców.	61
5.3.3 Możliwość zastosowań innych metod rozpoznawania.	61
5.4 Wyniki rozpoznawania.	62
6 Implementacja modułów systemu.	66
6.1 Implementacja algorytmu rozpoznawania.	66
6.2 Realizacja modułu gromadzenia wiedzy o uczniu.	66
6.2.1 Budowa modułu.	67
6.2.2 Identyfikacja użytkownika.	68
6.2.3 Ankieta.	68
6.2.4 Test umiejętności studiowania.	72
6.2.5 Wymagania sprzętowe.	73
7 Prezentacja metod nauczania na przykładzie twierdzenia Pitagorasa.	75
7.1 Idea i założenia realizacji.	75
7.2 Propozycja realizacji tematu w zależności od metody nauczania.	76
7.2.1 Metoda hipertekstowa.	76
7.2.2 Metoda z przewodnikiem.	78
7.2.3 Metoda prezentacyjna.	80
7.2.4 Metoda interakcyjna.	82
7.3 Podsumowanie przykładu i komentarz.	84
8 Zakończenie.	85
8.1 Podsumowanie.	85
8.2 Wnioski.	87
8.3 Kierunki dalszych badań.	89

<i>Dodatek A – Zestawienie wyników rozpoznawania dla różnych sposobów kodowania i różnych zestawów cech.</i>	<i>91</i>
<i>Dodatek B - Porównanie grupowania obiektów w wyniku analizy skupień z klasyfikacją nauczyciela.....</i>	<i>93</i>
<i>Dodatek C – Zestaw pytań psychologa</i>	<i>94</i>
<i>Dodatek D – Zestaw pytań ankiety.</i>	<i>96</i>
<i>Dodatek E – Macierz korelacji dla wybranych cech (fragmenty).</i>	<i>108</i>
<i>Dodatek F – Jedna z grup elementów ciągu uczącego utworzona na podstawie analizy skupień.</i>	<i>111</i>
<i>Bibliografia</i>	<i>112</i>

1 Wstęp

Przydatność komputerów w edukacji nie budzi dziś chyba żadnych wątpliwości. Coraz więcej szkół posiada pracownie informatyczne wyposażone w dobrej klasy sprzęt komputerowy. Opracowuje się programy autorskie i wprowadza klasy o profilu informatycznym. Z drugiej jednak strony, gdy zapytać studentów pierwszego roku kierunków nie-informatycznych jaką wiedzę na temat komputerów wynieśli ze szkoły, zazwyczaj okazuje się ona dość nikła. Nie sposób w tym miejscu zanalizować wszystkie przyczyny tego zjawiska, gdyż jest ich, jak się wydaje, bardzo wiele. Warto może jednak zasygnalizować niektóre z nich. Jak potwierdzają to badania, prowadzone m.in. przez pracowników Instytutu Informatyki Uniwersytetu Wrocławskiego (prezentowane na XIII Konferencji „**Informatyka w Szkole**” w Lublinie, we wrześniu ub. roku. [27]), komputer wykorzystywany jest w szkole głównie na lekcjach tzw. *Elementów informatyki*. Wynika to najczęściej z koncentracji niemal całego sprzętu komputerowego będącego w posiadaniu szkoły w jednej, dobrze strzeżonej pracowni, do której dostęp ma tylko nauczyciel informatyki. W dodatku odpowiedzialny materialnie za posiadany sprzęt. Do wyjątków należą dyrektorzy jak np. pan Jerzy Cywicki ze Szkoły Podstawowej Nr 3 w Zabierzowie, który ustawił w bibliotece szkolnej ogólnodostępny komputer pełniący rolę elektronicznej tablicy informacyjnej oraz lokalnego serwera pocztowego.

Niezależnie od postawy dyrekcji, czy łatwości dostępu do sprzętu trudno powiedzieć, że nauczyciele innych przedmiotów są zainteresowani wykorzystaniem komputerów na swoich lekcjach (znowu poza nielicznymi wyjątkami). Do niedawna zasadniczą barierą była po prostu nieznajomość obsługi tych urządzeń i lęk przed poznawaniem tego co, jak się niejednokrotnie wydaje, jest zbyt skomplikowane. Dziś, gdy praktycznie każdy absolwent wyższej uczelni zetknął się w sposób praktyczny z techniką komputerową trudno mówić o lęku przed nieznanym. Znacznie jednak istotniejszą barierą jest chaos informacyjny. Nauczyciele często nie wiedzą, które programy są naprawdę wartościowe, a wobec ograniczonych środków finansowych szkoły nie decydują się na ich zakup. Również chyba zbyt niska jest wśród nauczycieli świadomość ogromnego potencjału informacyjnego jaki niesie z sobą Internet.

W ostatnich latach pojawiło się, także na rynku polskim, wiele programów multimedialnych, które pretendują do miana edukacyjnych. Nie rosząc sobie prawa do ich oceny, gdyż wymagałoby to oddzielnych studiów, należy jednak zaznaczyć, iż niezależnie od wartości dydaktycznej, czy merytorycznej tych programów, nie cieszą się one wystarczającym zainteresowaniem. W każdym razie na pewno biorąc pod uwagę częstotliwość ich wykorzystywania, zdecydowanie ustępują one grom, które ciągle stanowią grupę najczęściej używanych programów, zwłaszcza wśród młodych ludzi. Tymczasem komputer wyposażony w odpowiedni program mógłby z powodzeniem pomóc uczniom, i to zarówno w czasie lekcji, jak i podczas nauki własnej. Dlaczego tak nie jest? Częściowa odpowiedź tkwi w fakcie, że komputer gdy trafił do domów i do szkół służył głównie rozrywce, „poważne” programy komputerowe były zbyt

skomplikowane i trudne w obsłudze aby mógł z nich korzystać przeciętny użytkownik komputera.. Co więcej, jeżeli nie ma ukształtowanego nawyku wykorzystania komputera w nauczaniu innych, poza informatyką, przedmiotów szkolnych, to w domu tym bardziej nie będzie się on kojarzył ze źródłem wiedzy, czy narzędziem pracy. Pozostaje jednak pytanie: czy atrakcyjność gier wynika wyłącznie z naturalnej chęci człowieka do rozrywki i zabawy? Odpowiedź na to pytanie prowadzi do stwierdzenia, które będzie kluczowe dla tej pracy. Podstawą sukcesu gier jest to, że oddziałują one na odpowiednie elementy psychiki młodego człowieka. Ciekawa, acz przerażająca była dyskusja drukowana parę miesięcy temu na łamach tygodnika dla młodzieży *Droga*, w której młodociany miłośnik gier komputerowych z przekonaniem dowodził skuteczności gier typu *Mortal Kombat* jako środka na odreagowanie stresów szkolnych. Nie sposób zaprzeczyć, że gry komputerowe mają silny wpływ na ludzką psychikę, często ją niestety deformując. W jakim stopniu jest to świadome zamierzenie ich twórców trudno w tej chwili rozstrzygnąć, natomiast nasuwa się spostrzeżenie, że autorzy programów edukacyjnych zdają się nie dostrzegać związku pomiędzy osobowością ucznia a sposobem przyswajania przez niego oferowanej mu wiedzy. I tu chyba należy upatrywać głównej przyczyny słabego powodzenia programów wspomagających nauczanie.

Wielu nauczycieli narzeka, nie bez racji, na przeładowanie programów nauczania i na dużą liczebność klas podając je jako przyczyny uniemożliwiające dostosowanie poziomu nauczania do potrzeb i możliwości każdego z uczniów. Zwykle w szkole realizuje się wersję programu dostosowaną do uczniów przeciętnych, poświęcając niewiele uwagi uczniom słabym, i jeszcze mniej - uczniom bardzo zdolnym. Zdarzają się oczywiście nauczyciele, którzy próbują dzielić swój czas i uwagę tak, aby każdego potraktować indywidualnie. Przy trzydziesto- a nawet czterdziestoosobowych klasach nie jest to jednak możliwe. Z myślą o tych nauczycielach, którzy dostrzegają różne preferencje uczniów i chcieliby uwzględnić je w swojej pracy powstała idea inteligentnego programu nauczającego (Artificial Intelligence Based Tutoring System), różnicującego sposób i formę przekazu wiedzy w zależności od indywidualnych predyspozycji osoby nauczanej. Rozwinięcie tej idei stanowić będzie główny wątek pracy.

Podstawą do rozpoczęcia badań będących przedmiotem niniejszej rozprawy był, opracowany przez autorkę kilka lat wcześniej, komputerowy program wspomagający nauczanie na temat baz danych [74]. Ideą jego powstania była próba przybliżenia złożonej problematyki baz danych przyszłemu ich użytkownikowi. Kwestie np. wyszukiwania informacji w bazach danych, aczkolwiek towarzyszące nam w praktyce codziennej nie są oczywiste ani proste dla przeciętnego użytkownika. Wykorzystanie wspomnianego programu komputerowego w czasie zajęć dydaktycznych pozwoliło zauważyć zróżnicowane oczekiwania użytkowników względem programów tego typu. Chodzi tutaj nie tylko o różny poziom wiadomości i umiejętności w momencie rozpoczynania pracy, ale przede wszystkim o różny sposób przyswajania wiedzy oferowanej przez postulowany system. Jedni preferowali instrukcje słowne, inni rysunkowe (np. w formie schematów). Jedni chcieli się dowiedzieć więcej szczegółów, podczas gdy inni oczekiwali możliwie szerokiej (acz nie pogłębionej) wiedzy na temat

zastosowań baz danych. Kolejnym elementem różnicującym była bądź duża samodzielność i gotowość do własnych poszukiwań, bądź przeciwnie - oczekiwanie na ciągłe podpowiedzi ze strony systemu i zdanie się na drogę postępowania sugerowaną, krok po kroku, przez program.

Te właśnie obserwacje przyczyniły się do powstania idei opracowania komputerowego systemu nauczającego, który wyposażony w określoną bazę wiedzy potrafiłby dostosowywać sposób przekazu tej wiedzy do indywidualnych predyspozycji ucznia. Pomyślna realizacja takiego systemu pozwoliłaby przynajmniej w części rozwiązać problem zróżnicowanych preferencji uczniów w zakresie przyswajania wiadomości. Powstaje natychmiast pytanie jak badać te różne preferencje związane z uczeniem się, a ponadto na jakiej podstawie system komputerowy miałby decydować o tym, jaki sposób przekazywania wiedzy będzie najbardziej odpowiedni dla danego ucznia. Należy w tym miejscu podkreślić, że jakkolwiek prowadzone były badania związku między nauczaniem wspomaganym komputerowo (Computer Assisted Teaching) a wynikami nauczania (Erik de Corte w [27]), jest to jednak problem jakościowo różny od podejmowanego tu zagadnienia dostosowywania komputerowej metody nauczania do indywidualnych potrzeb ucznia. Nie chodzi tu tylko o wykorzystanie komputera jako narzędzia wspierającego nauczyciela, czy nawet czasami go zastępującego, bowiem w tym przypadku system nauczania jest jednakowy dla wszystkich. Chodzi natomiast o opracowanie systemu umożliwiającego każdemu uczniowi przyswajanie wiedzy metodą najlepiej dostosowaną do jego faktycznych możliwości.

Istnieją wprawdzie próby tworzenia systemów, które pozwalają na pracę bez nauczyciela (np. Super Memo [4]), ale ich efektywność mierzona jest głównie ilością czasu potrzebnego do utrwalenia określonej porcji wiedzy i jedyną kontrolowaną w nich zdolnością jest sprawność zapamiętywania tych porcji. Istotnym jest także fakt, iż ocena tej sprawności jest czysto subiektywna, gdyż dokonuje jej sam uczeń mając do dyspozycji sześciostopniową skalę. W przypadku systemu, który ma powstać na bazie prowadzonych badań, nie przewiduje się sprawdzania zdobytych wiadomości, choć nie jest wykluczone dołączenie takiej opcji w fazie realizacji systemu. Główny nacisk położono natomiast na sposób pomiaru i oceny określonych cech osobowości ucznia decydujących o różnym sposobie przyswajania przez niego wiedzy. W tym zakresie realizowane badania napotykają istotną trudność wynikającą z faktu, że jakkolwiek psychologia dysponuje całą gamą różnego rodzaju testów pomiaru bądź to inteligencji ucznia, bądź poszczególnych zdolności, brak jest jednoznacznych „wzorców” czy „profilu” osobowości umożliwiających automatyczny dobór metody uczenia do konkretnej sylwetki ucznia. Wydaje się, że metody rozpoznawania obrazów [67,68] mogą być pomocne w rozwiązaniu tego problemu. W wielu bowiem dziedzinach, w których są z powodzeniem stosowane, jak chociażby diagnostyka medyczna ([37],[53]) często występuje sytuacja kiedy nie umiemy podać algorytmu przyporządkowania badanych obiektów do poszczególnych klas ale potrafimy jednak nauczyć maszynę jak wykorzystać wiedzę o pewnej grupie obiektów do tworzenia reguł decyzyjnych pozwalających klasyfikować inne, nieznane obiekty.

Perspektywa ta wydała się na tyle obiecująca, że skłoniła ona autorkę do sformułowania następującej tezy:

Wykorzystanie technik sztucznej inteligencji, a w szczególności wybranych metod rozpoznawania obrazów (*pattern recognition*) umożliwia uzyskanie informacji dotyczących cech osobowości ucznia, które zastosowane jako parametry sterujące multimedialnego systemu komputerowego wspomagania nauczania, pozwolą na dostosowanie sposobu przekazu wiedzy do indywidualnych predyspozycji osoby nauczanej.

Próba sprawdzenia czy i na ile jest to możliwe oraz wskazania praktycznego sposobu rozwiązania wielu szczegółowych zagadnień, wchodzących w skład postawionego problemu stanowi przedmiot niniejszej pracy. Należy przy tym wyraźnie zaznaczyć, że pojęcie cechy osobowości przyjęte jest tutaj w znaczeniu szerszym niż czyni się to np. w psychologii uczenia. Nie chodzi tu tylko o cechy tradycyjnie uważane za główne czynniki mające wpływ na sposób przyswajania wiedzy, takie jak: preferencje wzrokowe bądź słuchowe, zakres pamięci, zdolności matematyczne czy umiejętność logicznego myślenia, ale także o pewne nawyki ukształtowane w trakcie dotychczasowej nauki szkolnej. Dlatego bierze się pod uwagę warunki środowiskowe i społeczne ucznia, sposób nauki we wcześniejszych etapach kształcenia, a także chęć rozwijania własnych umiejętności i poszerzania zdobytej wiedzy. Szczegółowe omówienie tych cech nastąpi w dalszej części pracy.

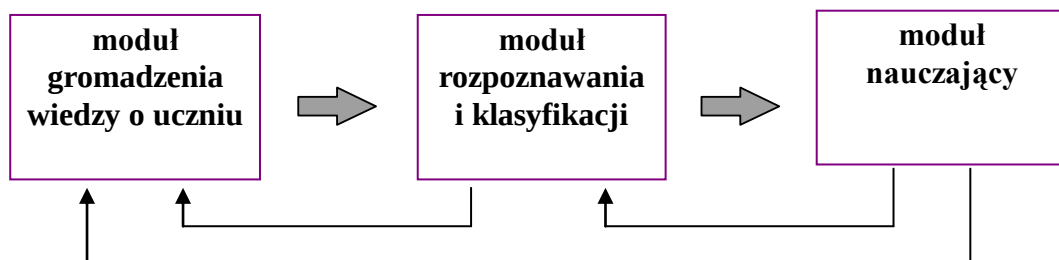
Prezentacja przeprowadzonych badań oraz uzyskanych rezultatów rozpocznie się od omówienia proponowanej struktury komputerowego systemu wspomagającego nauczanie. System ten będzie punktem odniesienia w całych dalszych badaniach. Silne powiązania prowadzonych badań z psychologią wymagają naświetlenia tej tematyki na gruncie psychologii eksperymentalnej. Problematyce tej poświęcono zatem osobny rozdział. Kwestii związanych z określeniem przestrzeni cech osobowości ucznia, sposobu ich pomiaru oraz analizy dotyczy rozdział kolejny. Sposób opisu zgromadzonych informacji oraz zbadanie możliwości wykorzystania technik rozpoznawania obrazów (*pattern recognition*) do określenia najbardziej efektywnej metody uczenia stanowi treść rozdziału piątego. Osobny rozdział przeznaczono na omówienie narzędzi służących gromadzeniu wiedzy o uczniu. Idea działania modułu nauczającego, który sam w sobie nie jest przedmiotem niniejszej pracy, została zilustrowana na przykładzie, który przedstawiono w rozdziale siódmym. Podsumowanie uzyskanych wyników oraz nasuwające się wnioski wraz z sugestiami dotyczącymi kierunków rozwoju zapoczątkowanych badań stanowią treść rozdziału ósmego, zamykającego całą pracę.

Autorka wyraża gorące podziękowanie promotorowi pracy prof. zw. dr hab. inż. Ryszardowi Tadeusiewiczowi bez którego pomocy merytorycznej nie byłoby możliwe powstanie tej pracy.

2 Idea i elementy składowe inteligentnego systemu komputerowego wspomagającego nauczanie

2.1 Ogólna struktura systemu

Próbując wykazać prawdziwość postawionej tezy badawczej należy najpierw określić warunki, w jakich miałyby nastąpić jej realizacja. W związku z tym została zaprojektowana struktura zaproponowanego przez autorkę i postulatywnie przyjmowanego w całej dalszej treści pracy inteligentnego systemu nauczającego. Wobec szybko postępujących zmian w dziedzinie sprzętu komputerowego, od strony technicznej nie precyzuje się żadnych szczegółów implementacji. Można tylko zaznaczyć, że przewiduje się w systemie szerokie wykorzystanie elementów multimedialnych i audiowizualnych (animowana grafika, dźwięk), dostępu do sieci rozległych (głównie Internet), multimedialnych baz wiedzy, technik hipertekstu i innych. Jednak szczegóły dotyczące zakresu i sposobu wykorzystania wymienionych elementów wykraczają poza ramy niniejszej pracy i nie są istotne z punktu widzenia jej tezy. Ponadto, ze względu na wspomnianą dużą zmienność w czasie komponenty te muszą być określone dopiero w fazie realizacji systemu i powinny być tak zaprojektowane aby możliwa była modyfikacja i modernizacja stosownych elementów modułu nauczającego wraz z rozwojem technik i metod multimedialnych. Istotne natomiast, z punktu widzenia celu przeprowadzonych badań, jest określenie struktury logicznej przyszłego systemu komputerowego wspomagania nauczania. Został on przedstawiony na rysunku 2.1.



Rys. 2.1. Schemat struktury systemu

Przewiduje się, iż system składać się będzie z trzech głównych modułów, ściśle ze sobą powiązanych.

Moduł M I - **gromadzenie wiedzy o uczniu** - zawiera zestaw pytań i testów, których zadaniem jest zebranie informacji o osobie

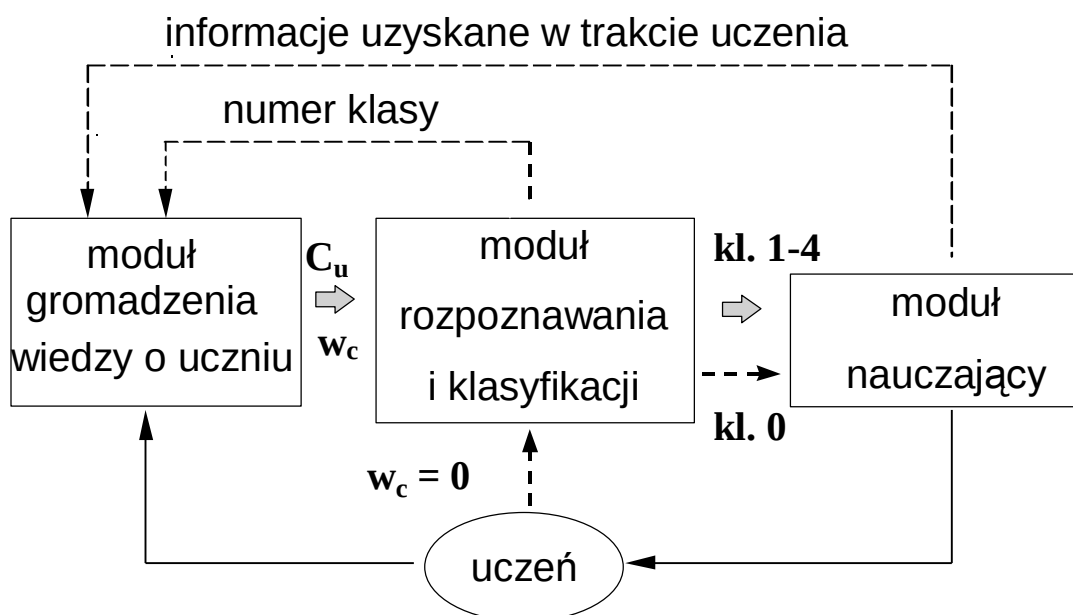
uczającej się, w kontekście jej indywidualnych preferencji, upodobań i przyzwyczajęń, związanych z procesem uczenia.

Moduł M II - **rozpoznawanie i klasyfikacja** - dokonuje analizy danych zebranych w części pierwszej i podejmuje decyzję o tym, jaka metoda nauczania będzie najbardziej odpowiednia dla danego ucznia.

Moduł M III - **nauczanie** - zawiera bogatą bazę wiedzy z danej dziedziny oraz zaprogramowane różne algorytmy przekazu tej wiedzy. W zależności od wprowadzonej bazy wiedzy możliwe jest nauczanie różnych przedmiotów, bądź różnych treści w obrębie tego samego przedmiotu.

System ma podstawową strukturę typu feed forward (z jednokierunkowym przepływem sygnałów), jednak dla badań i eksperymentów warto dopuścić także możliwość występowania zaznaczonych na rysunku 2.1 sprzężeń zwrotnych. Każdy z modułów posiada swoją specyfikę, która zostanie dalej szczegółowo omówiona

Schemat ten można dodatkowo rozbudować i wzbogacić. W szczególności przewiduje się, że w przyszłości moduł M III będzie mógł rejestrować i oceniać te cechy ucznia, które ujawniają się w trakcie procesu uczenia. Tak uzyskane informacje trafią wtedy do modułu M I (gromadzenia danych) jako dodatkowe dane wejściowe. Także (oprócz pierwszej sesji) rezultat poprzedniego rozpoznawania w postaci numeru klasy może być dodatkowym źródłem informacji wejściowych pełniących jednak tylko rolę pomocniczą a nie rozstrzygającą w kolejnych etapach uczenia. Można także przyjąć opcję, w której uczeń rozpoczyna naukę bez poddawania się testowi wstępnemu. Wszystkie składowe wektora cech otrzymają wówczas wartość początkową równą zero. W tej sytuacji system nie podejmie rozpoznawania tylko przyjmie klasę 0, odpowiadającą metodzie "bez określonych preferencji" i ta właśnie informacja zostanie przekazana do modułu nauczającego. Takie ustawienie początkowe nie wyklucza oczywiście zbierania danych o uczniu w trakcie nauki. Ponieważ jednak w pierwotnej wersji projektowanego systemu nie przewiduje się opcji modyfikowania profilu psychologicznego ucznia w trakcie nauki, metoda ta nie będzie uwzględniana w dalszej części pracy. Schemat systemu komputerowego, w którym uwzględniono powyższe modyfikacje przedstawia rysunek 2.2. Linia ciągłą oznaczono na nim operacje projektowane, linią przerywaną - operacje opcjonalnie możliwe, ale nie poddawane w tej pracy szczegółowej analizie.



Rys.2.2. Zmodyfikowany schemat systemu komputerowego wspomagania nauczania

Interesującym przedmiotem dalszych badań byłoby porównanie wyników nauczania właśnie takimi dwiema metodami - tą dobraną do automatycznie wykrytych i ustalonych preferencji ucznia i tą odpowiadającą klasie zero. Będzie to jednak możliwe dopiero po opracowaniu i zaimplementowaniu modułu nauczającego, czyli po skompletowaniu całości systemu.

2.2 Moduł gromadzenia danych o uczniu

Moduł gromadzenia wiedzy o uczniu składa się w wersji badawczej z ankiety i serii testów, których zadaniem jest ustalenie predyspozycji psychologicznych związanych z uczeniem się. Na etapie opisywanych w pracy badań naukowych mających na celu sprawdzenie metodologicznej poprawności samej koncepcji dane o uczniach zbierane były i przetwarzane ręcznie. W rozwiązaniu docelowym, w którym działanie całego systemu będzie w pełni autonomiczne i automatyczne, moduł ten będzie wykonany jako niezależny system generujący stosowne pytania ankietowe, a także porządkujący i przetwarzający dane z prezentowanych użytkownikowi specjalnych testów. W obecnej wersji zasadniczym kryterium jest stopień samodzielności ucznia, jego chęć rozwijania własnych umiejętności i rozszerzania swojej wiedzy, umiejętność gospodarowania własnym czasem, a także preferencje związane ze sposobem odbioru treści wpływające na formę przekazu - przekaz werbalny czy symboliczny, sekwencyjny lub według powiązań tematycznych, czy wreszcie statyczny bądź ruchomy. W praktyce oznacza to, że uczeń podczas pierwszej sesji pracy z systemem musi najpierw odpowiedzieć na pytania zawarte w ankiecie i rozwiązać zadania testowe. Uzyskane wyniki stanowią dane wejściowe dla modułu rozpoznającego. W kategoriach

rozpoznawania obrazów zestaw informacji opisujących jeden obiekt (jednego ucznia) nazywany jest wektorem cech, zaś odwzorowanie przyporządkowujące badanym obiektom ($d^k \in \mathbf{D}$) punkty w n -wymiarowej przestrzeni cech określane jest mianem recepcji. Przyjmuje się następujące oznaczenia, używane w teorii rozpoznawania obrazów [68]:

$$\underline{x}^k = (x^k_1, x^k_2, \dots, x^k_n) - \text{wektor cech opisujących } k\text{-ty obiekt} \quad (2.1)$$

gdzie: n - wymiar przestrzeni cech

$$B: \mathbf{D} \rightarrow \mathbf{X} - \text{funkcja recepcji} \quad (2.2)$$

gdzie: $\mathbf{D}$ - przestrzeń obiektów
 $\mathbf{X}$ - przestrzeń cech

Szczegółowemu omówieniu zagadnień związanych z określeniem wymiaru przestrzeni cech oraz sposobem definiowania jej elementów poświęcono osobny rozdział w dalszej części pracy.

2.3 Moduł rozpoznawania

W module rozpoznawania wykorzystywany jest zbiór danych nazywany ciągiem uczącym, którego elementami są również wektory cech opisujące poszczególnych osobników, ale rozszerzone o numer klasy, do której należy obiekt charakteryzowany przez dany wektor cech. Symbolicznie ciąg uczący można zapisać jako zbiór par:

$$\mathbf{C}_U = \{(\underline{x}^k, i^k), k = 1 \dots m\} \quad (2.3)$$

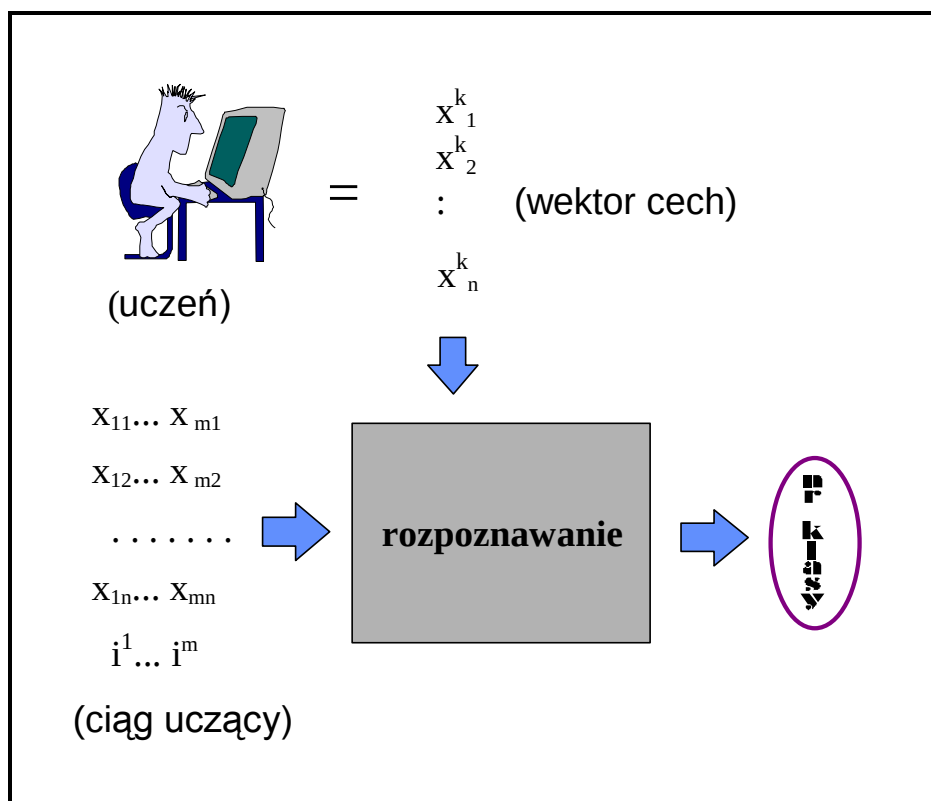
gdzie: $m = \# \mathbf{C}_U$

$$\underline{x}^k = B(d^k) \wedge d^k \in \mathbf{D}$$

oraz: $i^k \in \mathbf{I} \wedge i^k = A(d^k)$

Zbiór $\mathbf{I}$ jest zbiorem indeksów (identyfikatorów) klas, do których zaliczać będziemy rozpoznawane obiekty (czyli uczniów wykazujących określony typ osobowości i tym samym wymagających określonego podejścia w trakcie procesu nauczania). Sprawa struktury zbioru $\mathbf{I}$ będzie dyskutowana dalej. Odwzorowanie A użyte we wzorze (2.3) reprezentuje w sposób symboliczny idealną klasyfikację. Zakłada się (choć jest to

założenie silnie idealizujące rzeczywistość), że dla elementów ciągu uczącego znane są prawidłowe klasyfikacje. Oznacza to, że dla uczniów, których cechy tworzą wektory $\underline{x}^k$ możemy znaleźć idealne (prawdziwe) wartości identyfikatorów klas i^k , do których powinni być zaliczeni. Ponieważ uczniowie stanowiący elementy zbioru uczącego byli poddawani starannej obserwacji także w trakcie procesu uczenia, można przyjąć, że wartości i^k zebrane w bazie danych są dobrze zweryfikowane. Odwzorowanie A , ze względu na jego wyidealizowany charakter nie da się zrealizować, dlatego w pracy dążymy do budowy odwzorowania $\hat{A}$. Nazywamy je algorytmem rozpoznawania. Algorytm $\hat{A}$ aproksymuje (na tyle dokładnie, na ile to jest możliwe) idealne odwzorowanie A , zmierzając do ustalenia dla nieznanych obiektów $d \in D$ ich klas przynależności i . Porównuje on wektor cech opisujących badany obiekt z tymi, które są zapamiętane w ciągu uczącym i podejmuje decyzję o przynależności nowego obiektu do i -tej klasy na podstawie największego „podobieństwa” między porównywanymi wektorami. Słowo podobieństwo zostało ujęte w cudzysłów dla zaznaczenia, iż jakkolwiek odwołujemy się tutaj do intuicyjnego rozumienia tego pojęcia, to jego ścisłe znaczenie wynika z przyjętej metody rozpoznawania i będzie sprecyzowane w rozdziale piątym. Wynikiem pracy tego modułu jest, jak już powiedziano wcześniej, numer klasy, do której należy badany obiekt. Odpowiada on numerowi metody, która na podstawie dokonanej analizy została wybrana jako najbardziej odpowiednia dla ucznia o zbadanych właśnie predyspozycjach. Schemat działania tej części systemu przedstawia rysunek 2.3.



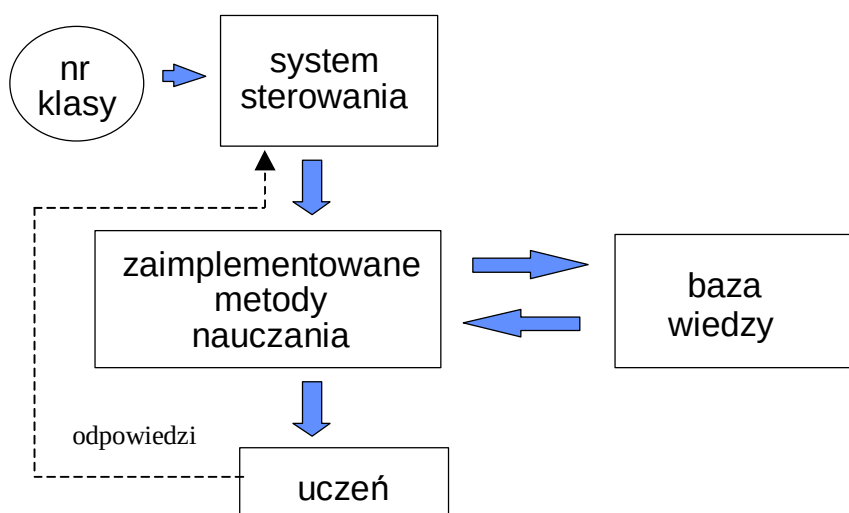
Rys 2.3. Schemat działania modułu rozpoznającego

2.4 Moduł nauczający

Moduł M I (gromadzenia danych) i M II (rozpoznawania) nie były opisywane dokładnie, gdyż ze względu na ich zasadnicze znaczenie dla całości pracy będą im poświęcone oddzielne rozdziały. Niniejszy podrozdział zawiera nieco obszerniejszą charakterystykę modułu M III, który jako wykraczający poza obszar opisywanych badań nie będzie dalej szczegółowo omawiany.

2.4.1 Struktura modułu nauczającego

Metody nauczania są integralnym elementem ostatniej części systemu, czyli właściwego modułu nauczającego. Zawiera on ponadto dwa inne, równie ważne elementy, jakimi są: baza wiedzy i system sterujący przepływem informacji w obu kierunkach. To, jaki materiał wprowadzimy do bazy wiedzy decyduje o nauczanych treściach. Schemat działania modułu nauczającego przedstawia rysunek 2.4.



Rys.2.4. Schemat modułu nauczającego

2.4.2 Baza wiedzy

Jak już wzmiankowano wcześniej idea opracowania multimedialnego systemu nauczającego pojawiła się w związku z próbą przybliżenia studentom problematyki baz danych. Taka też była pierwsza planowana baza wiedzy. Treści nauczania miałyby dotyczyć typów baz danych, ich budowy, aktualizacji ich zawartości oraz szerokiej gamy zagadnień związanych z wykorzystaniem baz danych w życiu codziennym,

ze szczególnym uwzględnieniem problemów wyszukiwania informacji. W trakcie badań przyjęto jednak założenie, że pytania ankiety i zadania testowe mają charakter bardziej ogólny i nie są ukierunkowane na żadną konkretną dziedzinę wiedzy. Pozwala to na postawienie tezy, iż projektowany system może mieć zastosowanie uniwersalne, zależne tylko od zawartości dołączonej bazy wiedzy. Wymieniając bazy wiedzy można nie tylko zmieniać rodzaj nauczanego przedmiotu, ale także wprowadzać różne poziomy szczegółowości.

2.4.3 Układ sterujący

Budowa i optymalizacja funkcjonalna układu sterującego stanowi zupełnie odrębny problem. Dla potrzeb niniejszej pracy zakłada się tylko, że będzie on miał konstrukcję zbliżoną do systemu ekspertowego, który wykorzystując techniki sztucznej inteligencji będzie sterował przepływem informacji stosownym do wybranej metody nauczania. System ten odpowiadać będzie za wielkość porcji wiedzy, za formę jej przekazu oraz za tempo prezentowania nowych wiadomości. Moduł ten zapewni również obsługę połączeń hipertekstowych, korzystanie z zasobów sieci rozproszonej (np. Internet) oraz interakcję z użytkownikiem. Nie są jeszcze opracowane metody komunikowania się z osobą nauczaną, ale przewiduje się, że będzie możliwe wykonywanie określonych operacji zarówno przez wskazanie ich myszką na ekranie, jak i wprowadzanie odpowiednich wartości (np. wyników rozwiązania) z klawiatury. Docelowo warto będzie także rozważyć możliwość użycia systemów wprowadzających dane do komputera za pośrednictwem sygnału mowy.

2.4.4 Ilość planowanych metod nauczania

Wynikiem pracy modułu rozpoznającego jest, jak już powiedziano wcześniej, numer klasy, do której należy badany obiekt. Odpowiada on numerowi metody nauczania, która na podstawie dokonanej analizy została wybrana jako najbardziej odpowiednia dla ucznia o zbadanych właśnie predyspozycjach.

Chcąc rozstrzygnąć jakie powinny być proponowane metody nauczania należy odpowiedzieć przede wszystkim na dwa pytania:

1. Jakie są profile osobowości uczniów ?
2. Jakimi środkami technicznymi dysponujemy ?

Poszukując odpowiedzi na pytanie pierwsze należy sięgnąć do doświadczeń psychologów, którzy potrafią określić pewne prawidłowości związane z procesem edukacji człowieka. I tak na przykład stwierdzono, że dla jakości uczenia się istotny jest wiek ucznia, który rzutuje między innymi na jego sposób przyswajania i utrwalania nabytej wiedzy, na zainteresowania, samodzielność itd. W niniejszych badaniach zdecydowano się przyjąć za próg wiekowy pierwszy rok studiów, jako początek nowego etapu zdobywania wiedzy, w którym z jednej strony wymienione wyżej cechy są już

uksztaltowane, czyli dadzą się badać, a z drugiej strony jest to etap jakościowo różny od poprzednich, tak ze względu na sposób uczenia się, jak i na wybór poznawanych treści. Można się spodziewać, że ktoś kto miał szeroki dostęp do wiedzy i nauczył się z tego faktu korzystać będzie bardziej samodzielny, dociekliwy i zainteresowany zdobywaniem możliwie szerokiej wiedzy. Jeżeli dodatkowo wziąć pod uwagę jego dobrze wyćwiczoną pamięć oraz rozwinięte zdolności logicznego myślenia, otrzymuje się obraz studenta, któremu można zaproponować nauczanie oparte na systemie połączeń hipertekstowych, z możliwością zagłębiania się, w miarę zainteresowań, w obszary wiedzy bardziej szczegółowej. Przeciwnieństwem podanego wyżej schematu będzie osoba, która całą swoją dotychczasową edukację odbywała na zasadzie programu minimum. Nie tylko w sensie ilości zdobywanej wiedzy, ale także sposobu jej pozyskiwania, tzn. praca własna sprowadzająca się tylko do odrabiania zadań domowych, żadnych prób poszukiwań własnych (najlepiej aby na lekcji była podana notatka, ze wskazaniem co należy zapamiętać), czytanie ograniczone tylko do lektur, albo wręcz do ich opracowań, itd. Powstaje wtedy obraz studenta pasywnego, któremu należy wiedzę przekazywać w niewielkich porcjach, w sposób usystematyzowany, bez dygresji i rozszerzeń, bo i tak nie potrafi z nich skorzystać.

Oczywiście nie trudno sobie wyobrazić postawy pośrednie, takie jak np. uczeń zainteresowany, potrafiący myśleć logicznie, któremu nikt do tej pory nie wskazał możliwości rozszerzania swojej wiedzy poza szkolne minimum, czy odwrotnie - uczeń o wielkiej ciekawości świata i chęci uczenia się, ale o kiepskiej pamięci, słabej zdolności koncentracji itd. Konkludując należy stwierdzić, że różnych postaw jest, jeżeli nie tyle, co ludzi uczących się, to na pewno wielokrotnie więcej niż możliwych do zaimplementowania metod nauczania.

Pozostaje ponadto odpowiedź na drugie z postawionych pytań, czyli możliwości techniczne posiadanego sprzętu i wynikające stąd formy przekazu nauczanych treści. Z założenia ma to być multimedialny system komputerowy, a zatem pozwalający na przekaz barwny, ilustrowany, także animowany, w razie potrzeby wzbogacony dźwiękiem i sekwencjami video. Wobec powszechnego dostępu uczelni do sieci Internet, warto także wziąć pod uwagę możliwość korzystania z bogatych zasobów wiedzy zgromadzonych w różnych miejscach świata. I tu znowu, podobnie jak w pytaniu poprzednim widać wielość potencjalnych metod wynikających z bogatych i ciągle rozwijanych możliwości sprzętu komputerowego.

Warto w związku z tym zastanowić się, jak bardzo wymienione wyżej aspekty różnicują przyszłe metody nauczania. Biorąc na przykład pod uwagę możliwości graficzne należy najpierw rozstrzygnąć, czy chodzi o grafikę w sensie ilustracji do tekstu, czy o obraz w formie schematu, wykresu czy grafu zastępującego tekst. Wydaje się, że to pierwsze zastosowanie, które ma oddziaływanie wzbogacające przekaz słowny może mieć tylko skutek pozytywny, to znaczy nawet dla tych uczniów, którzy preferują przyswajanie wiedzy przez czytanie (czy słuchanie) dodanie rozsądnej ilości ilustracji do tekstu nie będzie przeszkodą, dla innych zaś będzie wyraźną pomocą. Czyli ilustracje nie są istotnym wyróżnikiem przy wyborze różnych metod nauczania.

Podobnie wygląda sprawa animacji dołączonego obrazu. Można, bez tworzenia oddzielnej klasy, przyjąć że jeżeli w tekście są animowane ilustracje, to na życzenie można je uczynić statycznymi i odwrotnie. Inaczej przedstawia się kwestia zastępowania tekstu obrazem. To jest już cecha bardzo istotna i może ona, po zbadaniu odpowiednim testem, służyć jako kryterium różnic. Rolę dźwięku w przekazywaniu wiedzy można zdefiniować podobnie. Można przyjąć, że dołączane obrazy będą nie tylko animowane na życzenie (tak jak to zresztą ma miejsce we współczesnych programach multimedialnych), ale i towarzyszący im dźwięk może być w razie potrzeby włączony, bądź wyłączony.

Oddzielnego rozpatrzenia wymaga natomiast kwestia tzw. preferencji wzrokowych, bądź słuchowych. Czynnikiem ten jest silnie różnicujący, ale warto zwrócić uwagę, że uczeń który jest zdecydowanym słuchowcem raczej w ogóle nie zainteresuje się nauczaniem komputerowym, które w sposób naturalny kojarzy się mu z obrazem na ekranie monitora, nawet jeżeli jest to "czysty" tekst. Takim osobom wypadałoby raczej zaproponować słuchanie wykładów z taśmy, czy oglądanie filmów z opisem słownym, gdyż są to narzędzia z natury swojej wykorzystujące narząd słuchu jako środek przekazu. Z przedstawionych powyżej rozważań dotyczących technicznych aspektów komputerowego wspomagania nauczania wynika, że w zasadzie jedynym istotnie różnicującym metody nauczania (a zatem klasy w systemie rozpoznającym) czynnikiem są preferencje do pracy z tekstem, bądź z obrazem. Oczywiście ktoś mógłby w tym momencie postawić zarzut, iż w takim razie nie ma sensu proponować komputerowego nauczania komuś, komu wystarczy sam tekst gdyż znajdzie go w odpowiednich podręcznikach. Zarzut ten jest słuszny do momentu, gdy nie bierze się pod uwagę ogromnego potencjału wiedzy kryjącego się w zasobach Internetu. Jeżeli założyć iż system, którego dotyczą obecne badania, mógłby stanowić przydatne narzędzie w procesie nauczania na odległość (ang. *distance learning*), to zarówno możliwość porozumiewania się z nauczycielem (kierownikiem naukowym) bez konieczności dojeżdżania często z odległych miejscowości, jak i dostęp na przykład do zasobów bibliotecznych uczelni, w której się studiuje stanowią niepodważalne atuty takiej metody w stosunku do metod tradycyjnych.

Z powyższych rozważań wynika, że przynajmniej jeżeli brać pod uwagę względy techniczne, udało się zredukować ilość czynników mogących różnicować proponowane klasy uczniów, a co za tym idzie - postulowane metody nauczania. Inaczej ma się sprawa z cechami osobowości, gdyż tych jest po pierwsze znacznie więcej, a przede wszystkim niełatwo dają się one eliminować. Z przyczyn oczywistych nie wchodzi także w grę dobór metodą 'prób i błędów', z konieczności zatem pozostaje decyzja arbitralna. W obecnej wersji projektu przyjęto – traktując to jako jedną z hipotez badawczych - cztery metody nauczania. Wybór najbardziej odpowiedniej dla danego ucznia (w świetle wiedzy pochodzącej z ciągu uczącego) zapewnia algorytm rozpoznawania diskutowany szczegółowo w dalszej części pracy.

2.4.5 Charakterystyka proponowanych metod nauczania

Jak już powiedziano wcześniej, w obecnej wersji projektu systemu przewidziane są cztery metody nauczania: Ich ilość jest wynikiem kompromisu pomiędzy pożądaną różnorodnością oferty dydaktycznej proponowanej uczniom, a możliwościami implementacyjnymi. Natomiast ich rodzaj jest po prostu arbitralnym wyborem autorki projektu. Nawet jeżeli uwzględnimy fakt, że niektóre cechy są ze sobą skorelowane, to i tak przy ilości około osiemdziesięciu badanych cech trudno jest sobie wyobrazić rozsądny podzbiór metod nauczania uwzględniających wszelkie możliwe kombinacje rozpatrywanych cech. Zdecydowano zatem przyjąć jako podstawowe kryteria różnicujące poszczególne metody nauczania takie cechy ucznia jak: samodzielność, chęć poszukiwań i rozszerzania posiadanej wiedzy, bierny bądź aktywny odbiór przekazywanych treści, a także podatność na różne formy tego przekazu. Wybrane metody charakteryzują się następującymi własnościami:

- A - Metoda hipertekstowa** - oparta na technice hipertekstu, pozwalająca poruszać się po nauczonym materiale w sposób dość swobodny, także z wykorzystaniem łączy internetowych. Wybrane słowa - łączniki prowadzą do miejsc zawierających informacje związane z poznawany tematem, często o charakterze podstawowym, a czasami też rozszerzającym. Metoda ta przeznaczona jest dla uczniów samodzielnych, nastawionych na aktywne poszerzanie posiadanej wiedzy, gotowych do poszukiwań, preferujących tekst jako formę przekazu treści;
- B - Metoda prezentacyjna** - tworzona przez sekwencję zaprogramowanych ekranów graficznych traktowanych jako ustalona seria przeźroczy, pozwalająca na zdobycie ogólnego poglądu na dane zagadnienie. Przewidziana dla osób biernych, preferujących raczej obraz niż tekst, chcących nauczyć się szybko, choć może niezbyt szczegółowo;
- C- Metoda z przewodnikiem** - wzorowana na komputerowych asystentach lub samouczkach, polega na systematycznym prezentowaniu uczniowi poszczególnych fragmentów wiedzy i na sugerowaniu kolejnych kroków postępowania. Udział ucznia ogranicza się do odczytania bądź obejrzenia przekazywanych wiadomości i do wykonania sugerowanych działań. Nieco podobna do metody prezentacyjnej, choć pozwala na przekazanie większej liczby szczegółów. Oparta głównie o przekaz słowny, ilustracje pełnią tylko rolę pomocniczą. Przeznaczona raczej dla osób wygodnych i mało ambitnych, lub też dla mało samodzielnych;
- D - Metoda interakcyjna** - nastawiona na dialog z uczniem, zagadnienia prezentowane są głównie w formie zadań do rozwiązania, oferuje jednak możliwość skorzystania w razie potrzeby ze wskazówki. Uczeń wpisuje swoje odpowiedzi, decyzje i w zależności od nich rozwija się dalszy przebieg nauki. Metoda ta wykorzystuje zarówno obraz, jak i tekst, przeznaczona jest

dla ucznia aktywnego, samodzielnego, o rozwiniętych umiejętnościach myślenia abstrakcyjnego.

W powyższej charakterystyce pominięto informacje dotyczące sposobu implementacji poszczególnych metod. Jest to działanie celowe, częściowo podyktowane faktem, że przy obecnym tempie rozwoju techniki komputerowej trudno jest przewidzieć wszystkie możliwości oferowane przez sprzęt i oprogramowanie w chwili realizacji modułu nauczającego. Ponadto, istotne znaczenie ma rodzaj nauczanego przedmiotu. Gdy na przykład nauczanie dotyczy jednego z działów matematyki można sobie wyobrazić, że w metodzie prezentacyjnej zostanie podana ilustracja graficzna twierdzenia, bądź wynikających z niego zastosowań, w metodzie z przewodnikiem uczeń otrzyma praktycznie gotowy dowód twierdzenia z wyjaśnieniem poszczególnych jego kroków, a w metodzie interakcyjnej przeprowadzenie dowodu będzie właśnie głównym zadaniem ucznia (por. rozdział 7 – Przykład dotyczący twierdzenia Pitagorasa).

Inaczej, gdy nauczany przedmiotem jest np. geografia. Rozpatrując przykładowo problem opadów w różnych rejonach świata można korzystać ze szczegółowych map satelitarnych i przy pomocy odpowiednich elementów graficznych (strzałek, cieniowania, kolorowania) ilustrować wpływ poszczególnych czynników środowiska na ilość opadów na danym obszarze. Warto zwrócić uwagę, że dzięki zasobom sieci Internet można łatwo wyświetlić na ekranie mapy różnych regionów i na ich podstawie dokonywać analizy porównawczej poszczególnych obszarów. Metoda z przewodnikiem zawierałaby wówczas omówienie konkretnych czynników decydujących o ilości opadów na danym terenie, a metoda interakcyjna prowadziłaby do uzyskania przez ucznia pewnych wniosków i uogólnień dotyczących zaobserwowanych zjawisk, tak aby nie tylko potrafił on wskazać istniejące prawidłowości ale także podać ewentualne następstwa określonych zjawisk bądź zaproponować działania zapobiegawcze.

Na zakończenie tego rozdziału warto wspomnieć, że była także brana pod uwagę piąta klasa obiektów odpowiadająca metodzie nauczania określanej jako **gra strategiczna** (bądź decyzyjna), która w pewnym stopniu odpowiadałaby metodzie interakcyjnej, ale miałaby przewagę treści obrazowych (schematów, grafów) nad tekstem. Ze względu jednak na jej mniejszą uniwersalność (nadaje się głównie do nauczania wybranych zagadnień ekonomicznych [70]) zdecydowano się w chwili obecnej z niej zrezygnować.

Pozostaje jeszcze kwestia różnicowania sposobu uczenia w zależności od stopnia wstępnej znajomości przez ucznia nauczanego przedmiotu, ale wydaje się, że można to zagadnienie rozwiązać dopuszczając rozpoczynanie nauki (w każdej z wymienionych metod) od kilku wyróżnionych etapów odpowiadających różnym poziomom apriorycznej wiedzy ucznia, przy czym decyzję o zezwoleniu na pominięcie etapów wcześniejszych podejmowałby system sterujący modułu nauczania M III po sprawdzeniu testowym posiadanej przez ucznia wiedzy.

Należy podkreślić, iż jakkolwiek w charakterystyce metod podano pewne wskazania dotyczące cech ucznia predestynujących do korzystania z określonej metody, zostały one użyte tylko w celu lepszego zilustrowania różnic pomiędzy metodami. W systemie nie istnieją modele osobowości, które odpowiadałyby poszczególnym metodom, gdyż zbudowanie ich nie jest możliwe. Podstawowym motywem sięgnięcia po metody rozpoznawania obrazów jako techniki klasyfikacji była właśnie trudność określenia takich idealnych modeli uczniów i niemożność zdefiniowania kryterium przydziału konkretnych uczniów do poszczególnych klas. Jak pokazały przeprowadzone badania, różnice między poszczególnymi osobnikami są często bardzo subtelne i udział poszczególnych cech w ostatecznej decyzji klasyfikacyjnej zmienia się w zależności od metody. Problemy te ze względu na ich szczególną wagę zostaną dalej szczegółowo omówione.

3 Przedmiot badań w ujęciu psychologicznym

Na początku pracy sformułowano tezę, która mówi, że wykorzystując metody sztucznej inteligencji można zaprojektować komputerowy system wspomagający nauczanie, który pozwoli na indywidualizację tego procesu przez dostosowanie sposobu przekazu wiedzy i umiejętności do rzeczywistych predyspozycji ucznia związanych z uczeniem się. Próbując ustalić jakie cechy osobowości, tudzież indywidualne doświadczenia ucznia mają wpływ na sposób i skuteczność przyswajania przez niego wiedzy należy odwołać się do psychologii jako do dziedziny, na gruncie której pojęcia takie jak: styl poznawczy, profil osobowości czy uzdolnienia znajdują swoje wyjaśnienie.

Jak wspomina B.Niemierko w [28] powołując się na pracę J.Kozieleckiego¹ „psychologia przeżyła trzy okresy fascynacji bezpośrednią obserwowalnością i mierzalnością zjawisk: pierwszy, jako „psychofizyka”, w połowie dziewiętnastego wieku, skoncentrowany na pomiarze dokładności odbioru pojedynczych wrażeń, drugi, jako „psychofizjologia”, na przełomie dziewiętnastego i dwudziestego wieku, skupiający się na pomiarze funkcji fizjologicznych podmiotu, takich jak oddech, tętno, ciśnienie krwi, oraz trzeci, jako „nauk a o zachowaniu” (*behawioryzm*), trwający od I wojny światowej do lat osiemdziesiątych naszego wieku, kładący nacisk głównie na pomiar reakcji organizmu na różnego rodzaju bodźce” Zdaniem Niemierki żaden z tych kierunków nie potrafił przekonywająco objaśnić osobowości człowieka. Obecnie próby takie podejmuje psychologia nazywana poznawczą (*cognitive science*). Co zatem rozumie się w psychologii poznawczej przez cechy osobowości związane z uczeniem się?

3.1 Rola inteligencji w procesach poznawczych

Cechą dość często kojarzoną z procesem uczenia się jest inteligencja lub zdolności intelektualne. Jednak różni autorzy proponują nie tylko różny sposób pomiaru inteligencji ([8],[13],[16],[29],[31]), ale przede wszystkim różnorodne definicje tego pojęcia. Otóż, cytując za A.Matczak [42] „...zdolności intelektualne to możliwości warunkujące efektywność wykonywania określonego typu zadań poznawczych. Zdolności te powstają w efekcie angażowania inteligencji płynnej w rzeczywiste działania, wobec czego poziom każdej z nich zależy, z jednej strony od indywidualnych doświadczeń – od tego jak często i jak silnie był w danym zakresie angażowany potencjał inteligencji płynnej, z drugiej zaś – od wielkości tego potencjału.” Takie ujęcie zagadnienia wymaga z kolei wyjaśnienia pojęcia **inteligencji płynnej** i nie wymienionego tutaj ale funkcjonującego równolegle pojęcia **inteligencji skryształizowanej**. Rozróżnienie to, wprowadzone przez Catella² określa jakby dwie oddzielne części zjawiska zwanego inteligencją. Mianowicie przez **inteligencję płynną** Cattell rozumie tę

¹ Kozielecki J. *Koncepcje psychologiczne człowieka*, Żak, Warszawa 1997.

² Cattell R.B. *Abilities: their structure, growth and action*, Houghton Mifflin, Boston 1971.

jej część, która uwarunkowana jest biologicznymi właściwościami mózgu, a zatem w znacznej mierze dziedziczona, natomiast **inteligencja skryształizowana** powstaje pod wpływem indywidualnych doświadczeń, wynikających tak z wpływu środowiska, jak i ze sposobu w jaki jednostka korzysta z posiadanych „możliwości płynnych”. Wprowadzone tutaj rozróżnienie jest istotne ze względu na sposób, w jaki należy dokonywać pomiaru wymienionych własności. Inteligencję płynną (I_p) rozumianą jako pewną zdolność ogólną można mierzyć za pomocą jednego wskaźnika jakim jest np. iloraz inteligencji (zazwyczaj oznaczany w literaturze jako IQ od angielskiej nazwy Intelligence Quotient), podczas gdy inteligencja skryształizowana (I_s) powstała, jak powiedziano, na skutek oddziaływań wielu czynników, musi być określona poprzez pomiar tychże elementów składowych. Niestety w literaturze psychologicznej panuje dość duże zróżnicowanie poglądów nie tylko w kwestii rozumienia samego pojęcia inteligencji, ale i jej elementów składowych. Nie wdając się w analizę różnych poglądów na ten temat warto zaznaczyć, że generalnie istnieją dwie szkoły. Jedna, wywodząca się od Francisca Galtona uznaje inteligencję za zdolność ogólną, pewną globalną dyspozycję poznawczą leżącą u podstaw wszelkich czynności intelektualnych i odpowiedzialną za ich zróżnicowanie [42]. Poglądy te wiążą się z przekonaniem o istnieniu tak zwanego **czynnika ogólnego**, znanego jako czynnik **g**, wprowadzony po raz pierwszy przez Spearmana³. Przejawy działania tego czynnika można znaleźć w każdym niemal „inteligentnym” działaniu.

Szkółą drugą, reprezentowaną między innymi przez Guilforda, twórcę tzw. modelu struktury intelektu a zapoczątkowaną przez Alfreda Bineta utrzymuje, że inteligencja stanowi zbiór w znacznym stopniu od siebie niezależnych zdolności zwanych czynnikami. Sam Guilford wyróżnia 120 takich czynników [23].

Jeszcze dalej posuwają się psychologowie współcześni, (m.in.. Jensen) twierdząc, że inteligencję człowieka należy rozumieć jako posiadany przez niego „repertuar” wiedzy, szczegółowych zdolności czy sprawności oraz przyswojonych strategii rozwiązywania problemów. Z kolei E. Nęcka w swojej książce zatytułowanej *Inteligencja i procesy poznawcze* [45] prezentuje pogląd w myśl którego nie wiadomo co to jest inteligencja a co za tym idzie czy istnieje jakiś realny desygnat tego pojęcia. Jego zdaniem, prawdziwym powodem, dla którego jednych ludzi uważa się za bardziej inteligentnych niż inni jest występowanie różnic indywidualnych w zakresie czynności poznawczych. Wspomniane wcześniej przez Niemierkę [28] powstanie psychologii poznawczej ściśle wiąże się z sygnalizowanym przez Neckę przejściem od tzw. koncepcji czynnikowych do koncepcji operujących pojęciem przetwarzania informacji ([36],[39],[45],[77]). W podejściu tym szczególną uwagę zwraca się na działanie elementarnych operacji umysłowych, związanych z funkcjonowaniem systemów uwagi i pamięci.

³ Spearman C. *The abilities of man: their nature and measurement*, Macmillan, London 1927.

3.2 Pojęcie uczenia się

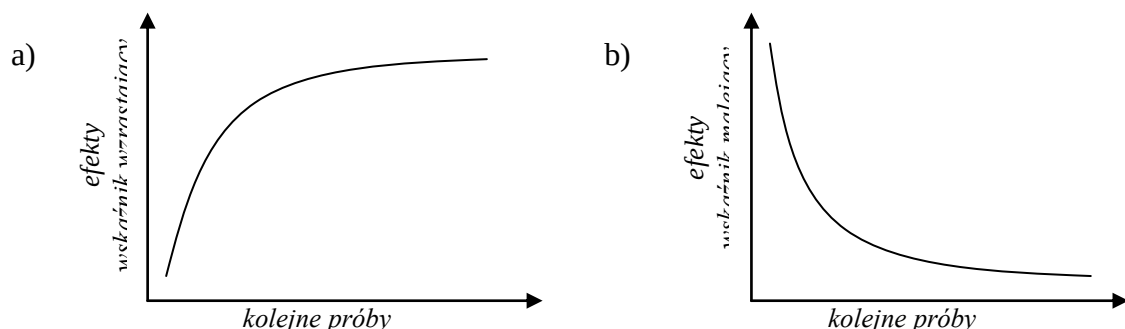
Poznanie nowych zjawisk czy faktów niekoniecznie musi być związane z zapamiętywaniem. Jak już wspomniano w poprzednim podrozdziale pamięć i uwaga są czynnikami warunkującymi sprawny przebieg procesów poznawczych. czy jednak poznanie świata jest równoznaczne z uczeniem się? Odwołując się ponownie do doświadczeń psychologów ([11],[36]) można przytoczyć stosowaną często definicję Colvina i Bagleya, w myśl której **uczenie się to proces prowadzący do modyfikacji zachowania jednostki w wyniku jej uprzednich doświadczeń**. Z.Włodarski w pracy [11] ujmując to w formie następującego schematu (rys. 3.1):



Rys. 3.1. Sytuacja uczenia się.

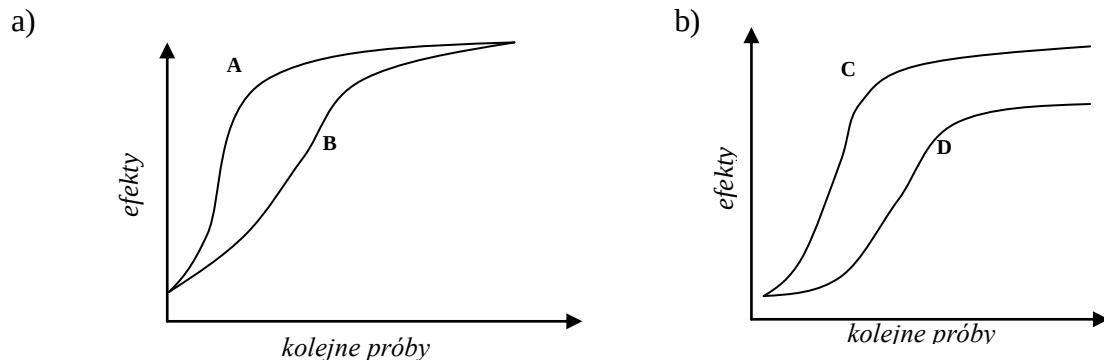
Definicja ta jest pojęciem bardzo szerokim, gdyż przewiduje możliwość uczenia się w każdym obszarze ludzkich zachowań, a nie tylko w zakresie zdobywania wiedzy i umiejętności. Ponadto uściślenia wymaga zwrot *modyfikacja zachowania*. Po odpowiednim doprecyzowaniu pojęć można zatem stwierdzić, że „**za wynik uczenia się będziemy uważać te zmiany, które są względnie trwałe, nie zależą od procesów obwodowych w receptorach i efektorach i nie są wyłącznie skutkiem dojrzewania organizmu.**” (Ida Kurcz [36] str.99)

Tradycyjnie miarami uczenia się są - szybkość przyswajania przez jednostkę nowych doświadczeń lub liczba błędów w tym procesie oraz szybkość zmiany i przekształceń utrwalonych wcześniej form zachowania. Najczęstszym sposobem przedstawiania przebiegu procesu uczenia się są krzywe uczenia wyrażające zależność uzyskiwanych efektów od ilości prób. Krzywe uczenia się mogą mieć różnorodny kształt w zależności nie tylko od przebiegu procesu uczenia się (rys. 3.3) ale także od sposobu przedstawiania otrzymanych wyników (rys 3.2).



Rys.3.2 Idealne krzywe uczenia się a) krzywa wypukła b) krzywa wklęsła

Krzywą wypukłą uzyskuje się wówczas, gdy wielkość wskaźników obrazujących efekty uczenia się wzrasta w miarę uzyskiwanych postępów (takim wskaźnikiem może być liczba prawidłowych wyborów w labiryncie, liczba znaków napisana w ciągu minuty na maszynie itp.), natomiast krzywa ła ilustruje przebieg wskaźników, których wartość maleje wraz z czynionymi postępami (np. liczba popełnionych błędów).



Rys. 3.3. a) krzywe obrazujące szybki i wolny przebieg uczenia się
 b) krzywe obrazujące jednakowo szybki przebieg uczenia się ale różny poziom stabilizacji wprawy

Krzywa **A** na rys 3.3 a) obrazuje szybszy przebieg uczenia się niż krzywa B. Krzywe te wskazują jednak, że w obu przypadkach ostateczny poziom stabilizacji wprawy jest ten sam, to znaczy że pod względem końcowych efektów uczenia się osoby te nie różnią się między sobą. Praktyka życia codziennego wskazuje jednak, że różnice międzypersonne dotyczą nie tylko tempa uczenia się, ale także poziomu stabilizacji wprawy. Rysunek 3.3 b) przedstawia krzywe przebiegu uczenia się mniej więcej jednakowo szybkiego ale prowadzącego do zdecydowanie różnych efektów końcowych. Wszystkie krzywe na rysunku 3.3. zostały wykreślone na podstawie fikcyjnych danych.

Psychologia oferuje rozbudowany aparat pojęciowy dotyczący procesu uczenia się. Wymienia się różne formy zachowań warunkujących różne rodzaje uczenia się. Niemal w każdej pracy poświęconej j uczeniu się występuje podział na dwa elementarne rodzaje uczenia się:

- warunkowanie klasyczne (zapoczątkowane przez Pawłowa)
- warunkowanie instrumentalne (istotny wkład w jego rozwój przypisuje się polskim fizjologom Konorskiemu i Millerowi oraz amerykańskiemu Skinerowi)

Gdy mowa o warunkowaniu wymienia się takie czynniki jak:

- wzmocnienie
- powtarzanie
- wygasanie

Szczegółowa analiza tych pojęć będzie niewątpliwie nieodzowna na etapie tworzenia modułu nauczającego (M III), gdyż chcąc zaprojektować możliwie efektywny system uczący należy umiejętnie wykorzystać wszystkie prawidłowości występujące w zachowaniu się jednostki a prowadzące do utrwalenia się bądź zaniku wprowadzonych zmian.

W niniejszej pracy proces uczenia się rozpatrywany jest w znaczeniu węższym, bliższym raczej podejściu pedagogów. Dotyczy on tych zmian w zachowaniu ucznia, które wiążą się z nabywaniem przez niego określonych wiadomości i umiejętności.

3.3 Uczenie się wspomagane komputerowo

Wokół problemu zdobywania wiadomości i nabywania umiejętności toczy się, nie tylko wśród pedagogów, ożywiona dyskusja. Kiedyś przeważał pogląd, że ważne są przede wszystkim wiadomości (świadczy o tym funkcjonujący do dziś model polskiej szkoły), obecnie coraz więcej wagi przywiązuje się do posiadania określonych umiejętności. Jest to, jak już powiedziano, temat oddzielnych dyskusji, tutaj należy tylko zaznaczyć, że podejście preferujące wiadomości miało zasadniczy wpływ na sposób wykorzystania komputerów w nauczaniu..

3.3.1 Tradycyjne formy wykorzystania komputerów w edukacji

Choć zagadnienia dotyczące komputerowego wspomagania nauczania pojawiały się w literaturze stosunkowo dawno przez długi okres czasu dotyczyły one właśnie badania wpływu wykorzystania komputera na wyniki uczenia się ([2],[36]). Zresztą gdy przyjrzeć się praktycznym zastosowaniom komputerów w szkołach to, jeżeli komputer jest wykorzystywany poza lekcjami informatyki najczęściej służy on właśnie przeprowadzaniu różnego rodzaju testów wiadomości [75]. Ograniczenie to było zresztą w znacznym stopniu uwarunkowane sposobem komunikacji z komputerem. Dziś, w epoce dominacji wśród komputerów osobistych systemu Windows z interfejsem graficznym niewiele osób pamięta, że jeszcze w połowie lat osiemdziesiątych podstawowym trybem pracy komputera był tryb tekstowy. Komputerowe testy wiadomości dawały się stosunkowo łatwo zaprogramować i najczęściej nie wymagały użycia grafiki. Jednak, jak twierdzi Siemieniecki [59] „ten zautomatyzowany system kontroli służy przede wszystkim pomiarowi osiągnięć w zakresie niższych obszarów taksonomii tj. głównie wiedzy biernej.”

3.3.2 Rola aplikacji multimedialnych w nauczaniu

Stopniowy rozwój aplikacji graficznych umożliwił stosowanie programów wspomagających nauczanie matematyki czy fizyki. Pojawiły się liczne programy nie tylko ułatwiające skomplikowane obliczenia matematyczne, ale także pozwalające na graficzną

interpretację analizowanych danych i uzyskiwanych wyników. Wystarczy wymienić programy takie, jak: Cabri, Derive, Math Cad, Math Lab, Mapple i wiele innych. Upowszechnienie się tzw. pakietów biurowych umożliwia stosunkowo łatwe korzystanie z redaktorów tekstu i arkuszy kalkulacyjnych. Wraz z burzliwym rozwojem multimediów na początku lat dziewięćdziesiątych pojawiły się bogate możliwości wykorzystania dźwięku i obrazu, pozwalające nie tylko na wierne odtwarzanie otaczającej nas rzeczywistości (liczne encyklopedie multimedialne), ale także na symulowanie zjawisk, których ze względu na wysoki koszt bądź duże ryzyko uczeń nigdy nie mógłby zobaczyć w rzeczywistości. Formą programów wspomagających nauczanie, które wywodzą się z testów są programy do nauki języków obcych. Powstały one przez dołączenie do testów opcji wprowadzania nowych pojęć i zwrotów. Pomimo wprowadzenia obrazu i dźwięku podstawową zasadą działania tych programów jest w dalszym ciągu sprawdzanie stopnia opanowania określonych porcji wiedzy.

3.3.3 Perspektywy wykorzystania sieci Internet w uczeniu się i nauczaniu

Za kolejną rewolucję można uznać powstanie sieci Internet i coraz ciekawszych programów pozwalających wędrować po zasobach sieci. Ciekawa analizę możliwości wykorzystania sieci rozległych w zastosowaniach edukacyjnych przedstawia B Siemieniecki w [60]. Zwraca on uwagę na ideę powstawania tzw. uniwersytetów sieciowych (np. Globewide Netwide Academy), w których z powodzeniem realizuje się zajęcia z astronomii, literatury czy języków obcych. Na gruncie polskim także pojawiają się już próby tworzenia tzw. uniwersytetów otwartych wykorzystujących sieć Internet jako środek dydaktyczny. Eksperymenty takie podejmowano już na Uniwersytecie Wrocławskim, a w bieżącym roku także na Akademii Górniczo-Hutniczej w Krakowie. W tym kontekście pojawia się konieczność opracowania nowych pomocy naukowych w formie podręczników elektronicznych oraz komputerowych programów nauczających. Gdy chodzi o podręczniki, to w żadnym wypadku nie mogą one być prostym przełożeniem formy drukowanej na elektroniczną lecz powinny wykorzystywać te właśnie możliwości, które daje nowe medium, jakim jest komputer. Oznacza to, że przekaz słowny powinien być wzbogacony obrazem i dźwiękiem, a także, co jest niezmiernie ważne dla skuteczności procesu kształcenia, powinien aktywizować ucznia do podejmowania własnych prób rozwiązywania postawionych problemów bądź przynajmniej do samodzielnych poszukiwań tego, co wymyślili inni. To samo kryterium odnosi się do komputerowych programów nauczających ale od nich wymaga się jeszcze więcej. O ile podręcznik zwykle adresowany jest do szerokiego kręgu słuchaczy (np. studentów wybranego kierunku) i powinien być w miarę uniwersalny, o tyle komputerowy program nauczający można adresować do każdego studenta indywidualnie tak, aby pracując w swoim pokoju, we własnym tempie mógł przyswajać potrzebną wiedzę i umiejętności w sposób najbardziej mu odpowiadający. Zbadanie możliwości opracowania takiego inteligentnego programu nauczającego to główny cel niniejszej pracy.

3.4 Pomiar w psychologii

Problemy związane z pomiarem w psychologii mają znaczenie zasadnicze dla przeprowadzania jakichkolwiek badań dlatego psychologia wypracowała już osobny dział zajmujący się metodami pomiaru, zwany psychometrią. Stosowane narzędzia charakteryzują się ogromną różnorodnością w zależności od przedmiotu badań i wybór właściwego zależy ściśle od tego co i jak dokładnie ma zostać zmierzone. Generalnie można stwierdzić że zakres metod pomiarowych rozciąga się od nieprecyzyjnych form wywiadu, czy obserwacji aż po badania laboratoryjne i standaryzowane testy psychometryczne. Z kolei otrzymane wyniki pomiarów można przedstawiać na różnych skalach w zależności od charakteru mierzonych zmiennych i od zastosowanych narzędzi pomiarowych.

3.4.1 Skale pomiarowe

W teorii pomiaru psychologicznego stosuje się cztery rodzaje skal pomiarowych zaproponowane przez Stevensa [20,32,65]. W kolejności od najsłabszej jest to: skala nominalna, skala porządkowa, skala przedziałowa oraz skala ilorazowa. Każda z nich różni się nie tylko sposobem reprezentacji mierzonych wielkości ale i dopuszczalnymi operacjami na danych uzyskanych w wyniku przeprowadzonego pomiaru. Zdania na temat możliwości stosowania innych operacji na danych niż wynikałoby to ze specyfiki użytej skali są mocno podzielone i to zarówno wśród statystyków, jak i teoretyków pomiaru [32], wszyscy oni wydają się być jednak zgodni co do tego, że aby móc dokonywać pewnych przekształceń na danych pomiarowych a następnie wyciągać wnioski z otrzymanych wyników wszystkie pomiary muszą być dokonywane na tej samej skali. Gdy tak nie jest należy dążyć do ujednolicenia zastosowanej skali pomiarowej, pamiętając, że możliwe jest tylko przejście ze skali silniejszej do słabszej, nigdy zaś na odwrót. Taka operacja oznacza zwykle utratę części informacji, co jednakże w przypadkach szczególnych może nawet okazać się korzystne. Jak podają autorzy w [65] „ umiejętne posłużenie się skalą słabszą pozwala na efektywniejszą eliminację tzw. białych szumów i na tej drodze wyraźniejsze wyodrębnienie poszukiwanej prawidłowości”. Na przykład wprowadzenie kilku kategorii wiekowych (zapisywanych na skali porządkowej) zamiast dokładnej liczby lat badanych (podawanych na silniejszej skali przedziałowej) pozwoli wyraźniej uchwycić badane zależności.

3.4.2 Stosowane metody pomiaru

Jak podaje Brzeziński w [6] w przypadku badania osobowości psycholog zazwyczaj ma do dyspozycji trzy metody:

- a) metodę kliniczną
- b) metodę laboratoryjną
- c) metodę testów psychologicznych

W ramach metody klinicznej wyróżnia się rozmowę (wywiad) z badanym oraz obserwację. U podstaw tej metody leży założenie, że osoba badana umie i chce odpowiadać zgodnie z prawdą na stawiane jej przez badacza pytania. Otrzymywane wyniki mają postać danych jakościowych. Fakt ten uznawany jest zwykle za podstawową wadę metody, gdyż utrudnia (a czasem wręcz uniemożliwia) stosowanie metod statystycznych do analizy danych. Dlatego głównym zarzutem wysuwany przeciwko metodzie klinicznej jest jej subiektywizm, który może prowadzić do wysuwania hipotez opartych wyłącznie na wiedzy i doświadczeniu przeprowadzającego wywiad. Podkreśla się, że ze względu na osobisty charakter odpowiedzi każdego badanego nie uzyskuje się układu odniesienia pozwalającego na dokonywanie porównań interindywidualnych.

Eksperyment laboratoryjny w odniesieniu do badania osobowości stwarza wiele problemów natury metodologicznej i organizacyjnej. Wprawdzie uzyskane wyniki pomiarów dają się przedstawić na skali ilorazowej umożliwiając tym samym ich skuteczną analizę ale pozostaje do rozstrzygnięcia zasadnicza kwestia relacji pomiędzy mierzonymi wskaźnikami fizjologicznymi a określonymi cechami osobowości człowieka. O ile w przypadku badania typów osobowości (np. neurotyczny, nieneurotyczny) możliwe jest powiązanie pewnych reakcji organizmu (tętno, ciśnienie krwi, badanie EEG) z typowymi przejawami zachowań jednostki, to nie opracowano jeszcze podobnych powiązań w odniesieniu do pojedynczych cech osobowości. Pojawiają się wprawdzie w psychologii tendencje do wiązania np. zdolności intelektualnych ze stanem pobudzenia emocjonalnego człowieka [19], ale są to na razie raczej ciekawe idee niż udowodnione hipotezy naukowe. Podobnie, niektórzy naukowcy próbują wiązać na przykład czynnik płci ze sposobem działania ludzkiego mózgu [43], ale nie oznacza to, że potrafią oni wskazać laboratoryjne metody pomiaru stopnia tej zależności.

Najchętniej, i bodaj najczęściej stosowaną metodą pomiaru w psychologii są testy. Należą do nich precyzyjnie opracowane, sprawdzone pod względem rzetelności i trafności testy standaryzowane [41], jak również wszelkiego rodzaju kwestionariusze i ankiety. Te ostatnie, stosunkowo proste do przeprowadzenia w formie pisemnej pozwalają szybko zgromadzić dane dotyczące nawet dużej próby z populacji. Najprostszą formą uzyskanych wyników są dane jakościowe, czy wręcz opisowe. Można jednak próbować przedstawiać je na skali przedziałowej poprzez zabieg rangowania [5]. Zarówno test, jak i ankieta może mieć formę pytań otwartych, bądź zamkniętych. Te ostatnie są zresztą chętnie stosowane ze względu na łatwiejszy sposób skalowania wyników. Pomiary ilościowe w testach otrzymuje się poprzez:

- ◆ zliczanie odpowiedzi poprawnych
- ◆ zliczanie ilości popełnionych błędów
- ◆ mierzenie czasu trwania wykonania zadania
- ◆ zliczanie ilości zapamiętanych słów, liczb itp.

Stosunkowo często występuje praktyka łączenia pojedynczych testów w tzw. baterie testów. Uzyskuje się w ten sposób wielokrotny pomiar tej samej cechy (np. Test matryc progresywnych Ravena, który sprawdza umiejętności myślenia przez analogię,

przez tożsamość, przez syntezę i analizę oraz myślenia kombinatorycznego), bądź pomiar wielu czynników, które w całości składają się na obraz jednego zespołu cech (Skala inteligencji WAIS-R [8], Krótka Skala Inteligencji [13]). Testy mogą mieć charakter werbalny (testy wiadomości, słownikowe, rozumienia tekstu, podobieństwa) bądź niewerbalny (uzupełnianie obrazków, konstrukcje z klocków, układanie historyjek obrazkowych we właściwej kolejności, składanki - łamigłówki itp.). Od czasu upowszechnienia się komputerów osobistych pojawiają się także próby opracowywania testów komputerowych np. stosowane w Instytucie Psychologii UJ testy pamięci i testy uwagi [45].

3.5 Charakterystyka i przedmiot badań

Przedmiotem badań jest zależność sposobu nauczania (**SN**) od indywidualnych predyspozycji (**IP**) ucznia. Przy czym termin uczeń jest używany do określania osoby uczącej się, niezależnie czy jest nią uczeń, student czy ktoś, kto po prostu próbuje poznać wybraną dziedzinę wiedzy.

Przeprowadzone badania mają charakter reprezentacyjny. Po przebadaniu grupy około 100 studentów I roku studiów (różnych kierunków) próbuje się uogólnić otrzymane wyniki na całą populację studentów (najpierw I roku, a po dodatkowych badaniach wszystkich studentów).

3.5.1 Opis modelu badawczego

Modelem **M** jest populacja wszystkich studentów I roku studiów. Przedmiotem pomiarów będą ich indywidualne zdolności i predyspozycje związane z możliwością zdobywania i utrwalania nabytej wiedzy. Pomiarom i obserwacji podlegają zarówno cechy niezależne zawarte w zbiorze $\{Z_N\}$, jak i cechy towarzyszące, będące elementami zbioru $\{Z_W\}$. W wyniku przeprowadzonych badań zostaną zdefiniowane różne style nauczania (**SN**), a opracowany algorytm rozpoznający $\hat{A}$ ma za zadanie na podstawie dokonanej analizy cech badanego ucznia wskazać najbardziej dla niego odpowiedni styl nauczania (**SN**). Charakterystyka stylów nauczania została podana w poprzednim rozdziale (por. rozdz. 2.4.4 i 2.4.5), natomiast do odpowiednich zbiorów zmiennych należą:

Z_N - zmienne niezależne: metoda pracy
 zdolności percepcyjne
 zainteresowania
 zdolności intelektualne

Z_W – zmienne towarzyszące: środowisko
 przebieg dotychczasowej nauki
 doświadczenia komputerowe

poziom inteligencji
płeć

3.5.2 Założenia i sposób doboru próby

Z psychologicznego punktu widzenia nie jest obojętnym wiek osób poddawanych badaniom, jako że zdolności intelektualne i percepcyjne różnią się u tych samych osób w zależności od wieku. Należało zatem albo ustalić próg wieku, albo przeprowadzić analogiczne badania dla różnych grup wiekowych. Ze względu jednak na to, że badania mające na celu określenie związku między indywidualnymi zdolnościami i predyspozycjami ucznia a sposobem przekazywania mu wiedzy z pomocą komputera należą do prototypowych zdecydowano na początek ograniczyć się do jednej grupy wiekowej. Jak wiadomo, rozpoczęcie nauki na studiach wiąże się z całkowicie odmiennym sposobem przyswajania wiedzy. Student musi „przestawić się” z systemu lekcyjnego na system ćwiczeń i wykładów, który charakteryzuje się nie tylko innym rozkładem czasowym ale i większą samodzielnością pracy studiumującego. Pojawia się konieczność samodzielnego dotarcia do potrzebnej wiedzy, która najczęściej rozproszona jest w różnych pozycjach literatury, mniej lub bardziej dostępnych. Zazwyczaj wykładowca lub prowadzący ćwiczenia podają wykaz literatury, ale rzadko jest to jeden konkretny podręcznik. Istotnym jest więc sposób korzystania z różnych źródeł informacji.

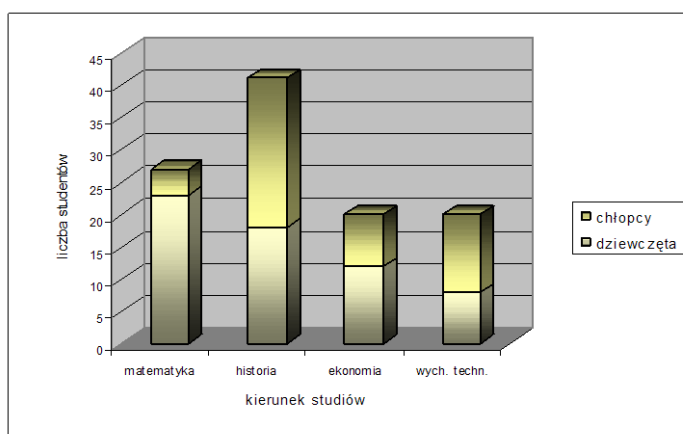
W ramach prowadzonych badań zgromadzono dane dotyczące 108 studentów, w tym 32 z matematyki, 30 z historii, 23 z wychowania technicznego i 23 z ekonomii. Wśród badanych było 68 kobiet i 40 mężczyzn. Charakterystykę liczebności badanych w zależności od płci i kierunku studiów przedstawia rys.3.4. Można łatwo zauważyć przewagę liczebności kobiet na kierunkach takich jak matematyka, czy historia, ale stan ten odzwierciedla rzeczywisty fakt dominacji ilościowej kobiet na tych kierunkach studiów.

O wyborze tych właśnie kierunków studiów zadecydowały następujące fakty:

1. Konieczność zapewnienia różnych profiliów poznawczych badanych osób przy równoczesnym zachowaniu możliwie równomiernego rozłożenia cech (np. nie ma większego sensu wybór kierunku ścisłego jakim jest matematyka i zestawianie go z typowo humanistyczną polonistyką gdyż w tym wypadku różnice wynikające właśnie ze specyfiki studiów będą dominować i być może przesłaniać inne, bardziej istotne właściwości).
2. Istotnym jest również w miarę równomierny udział w badaniach osobników obu płci, co w przypadku WSP jest w sposób istotny związane z kierunkiem studiów (a zatem przewaga dziewcząt na matematyce i ekonomii jest równoważona przewagą chłopców na historii i wychowaniu technicznym).

3. Brak podziału na kierunki stosowany od paru lat na AE (pierwszy rok liczy kilkuset studentów, podział na kierunki następuje na wyższych latach studiów) zapewnia większą losowość i zarazem różnorodność reprezentowanych cech.

Zatem czynnikami określonymi już w momencie doboru próby są takie zmienne towarzyszące, jak: uczelnia, kierunek studiów, płeć, wiek (w kategorii lat jest to około 19-21, ale dla badań istotną informację stanowi I rok, czyli początek studiowania). Pozostałe cechy muszą zostać zmierzone w trakcie badań, czy to poprzez wypełnienie odpowiednio przygotowanego kwestionariusza, czy też na podstawie specjalnie w tym celu skonstruowanych i przystosowanych testów. Szczegółowy opis mierzonych zmiennych znajdzie się w rozdziale dotyczącym przestrzeni cech (rozdz. 4).



Rys. 3.4. Rozkład liczebności badanych studentów wg kierunków studiów z uwzględnieniem płci.

3.5.3 Wykorzystane narzędzia i skale pomiarowe

W obecnej fazie opisywanych badań za podstawowe źródło informacji o uczniu przyjmuje się ankietę i test rozumienia tekstu zwany także testem umiejętności studiowania. Treść ankiety została zaprojektowana specjalnie dla potrzeb prowadzonych badań. Powstała ona przy współpracy z mgr Małgorzatą Mazurkiewicz z Katedry Psychologii Klinicznej UJ oraz w oparciu o doświadczenia psychologów z Poradni Wychowawczo-Zawodowej Nr 1 w Krakowie.

Należy podkreślić, że jakkolwiek nie było możliwe użycie gotowych testów psychometrycznych ani kwestionariuszy osobowości (prowadzone badania są prototypowe i żaden z typowych testów psychologicznych nie przystawał do ich potrzeb), niektóre z tych narzędzi były inspiracją i wzorem w podjętym opracowaniu. W szczególności schemat do testu umiejętności studiowania zaczerpnięto z testów stosowanych podczas egzaminów wstępnych na kierunek psychologii UJ (kopie testów z lat ubiegłych są dostępne w bibliotece instytucyjowej). Pytania ankiety mają charakter pytań zamkniętych, co oznacza, że badany wybiera jedną z podanych odpowiedzi. Nie może podać odpowiedzi własnej ani też nie może pytania pominąć. Należy wyraźnie

podkreślić, że przyjęte formy gromadzenia wiedzy o uczniu są tylko wstępną propozycją, którą zdecydowanie trzeba rozwijać w trakcie dalszej pracy nad podjętym problemem, w szczególności poprzez zmniejszenie liczby pytań w ankiecie na rzecz zadań testowych, a także dołączając opcję samodzielnego „nabywania” przez system wiedzy o uczniu na podstawie stylu jego pracy podczas nauki z pomocą tegoż systemu.

Ze względu na dominującą rolę pytań ankiety większość danych ma charakter jakościowy lub opisowy możliwy do przedstawienia na skali nominalnej bądź porządkowej. Tę ostatnią skalę można zastosować po odpowiednim uszeregowaniu odpowiedzi w myśl kryterium: mało, średnio, dużo gdy pytanie dotyczy np. ilości czasu poświęcanego na naukę własną lub na rozrywkę. Tylko wyniki testów (ilość poprawnych odpowiedzi, ilość powrotów do tekstu, ilość korekt oraz czas wykonania zadań) mają charakter liczbowy. W opisie przestrzeni cech utworzonej przez te dane (por. rozdział 4) próbowano w związku z tym różnego sposobu opisu danych. Kodowanie proste i kodowanie 1-z-n odpowiada sytuacji, gdy wszystkie dane sprowadzono do skali nominalnej, a kodowanie skumulowane koresponduje z próbą potraktowania danych tak, jakby były one mierzone na skali porządkowej.

Pełną treść ankiety zamieszczono w Dodatku D, natomiast informacje dotyczące implementacji komputerowej modułu gromadzenia danych (M I) zawiera rozdział 6.

4 Przestrzeń cech

Kwestia właściwego określenia przestrzeni cech stanowi zawsze w zagadnieniach rozpoznawania problem o podstawowym znaczeniu. Szczególnie ważna jest tu relacja rzeczywistej ilości cech opisujących badane obiekty do liczby tych, które można praktycznie zmierzyć. Często czynnikiem decydującym o użyciu pewnej cechy lub wpływającym na jej odrzucenie staje się bądź wysoki koszt przeprowadzanych badań, bądź też czas trwania eksperymentu. W omawianym przypadku dodatkowym problemem jest określenie, jakie cechy osobowości rzeczywiście opisują predyspozycje indywidualne poszczególnych uczniów związane z uczeniem się. Wynika to po części z faktu, że tak naprawdę ciągle za mało wiadomo o naturze procesów zachodzących w ludzkim mózgu i dlatego wszelkie próby definicji czy usystematyzowania pojęć dotyczących procesów poznawczych są w mniejszym lub większym stopniu oparte na przypuszczeniach. W niniejszej pracy przyjęto własną koncepcję wypracowaną na podstawie poglądów prezentowanych w dostępnej literaturze psychologicznej (por. rozdz. 3) z nadzieją, że nie pozostanie ona w zasadniczej sprzeczności z panującymi w tej dziedzinie przekonaniem.

4.1 Charakterystyka badanych obiektów

Poszukując odpowiedzi na pytanie o sposób i tempo uczenia się postawiono kilka hipotez roboczych, których wynikiem są właśnie proponowane zbiory cech. Przyjęto, że:

1. Człowiek posiada pewne zdolności intelektualne (**ZI**) takie jak: pamięć, zdolności werbalne, zdolności liczbowe, zdolności logicznego myślenia, zdolności abstrahowania i syntezy pojęć oraz znaczeń, zdolności twórcze. Cechy te, aczkolwiek w znacznej mierze uwarunkowane genetycznie, podlegają zmianom w trakcie całego życia człowieka. Rozwijane pozwalają uzyskać większą sprawność, nie wykorzystywane – stopniowo zanikają.
2. Określone biologicznie są również zdolności związane ze sposobem percepcji otaczającego świata (**ZP**). Należą do nich cechy takie, jak: uwaga, zdolność koncentracji, szybkość wykonywania zadań poznawczych, preferencje wzrokowe bądź słuchowe, słuch muzyczny, wyobraźnia przestrzenna. Również i w tym wypadku posiadane preferencje można rozwijać i doskonalić lub zaniedbać i zatracić.
3. Niebagatelny wpływ na rozwój posiadanych zdolności mają warunki w jakich człowiek rodzi się i wychowuje. Istotne w tym momencie jest nie tylko pochodzenie społeczne, i to zarówno w sensie miejsca pochodzenia (duże lub małe miasto, wieś), ale i rodzina, z której się wywodzi (np. rodzina pełna czy niepełna), jej status społeczno-ekonomiczny, wykształcenie rodziców, czy nawet dziadków, liczba

rodzeństwa (jednak, najstarszy czy najmłodszy z dzieci). Wymienione tutaj przykładowe cechy ujęte są w zbiór nazwany czynnikami środowiskowymi (**CS**).

4. Skoro badania indywidualnych stylów uczenia się dotyczą studentów jest oczywistym, że posiadają oni w tym względzie pewne nawyki wynikające z wcześniejszej nauki. Można tutaj wymienić trzy grupy czynników składających się na zbiór nazwany historią edukacji (**HE**). Do pierwszej z nich należą: rodzaj ukończonej szkoły, profil klasy, występujące wcześniej trudności w nauce (np. powtarzanie roku lub zmiana szkoły). Drugą grupę stanowią własności związane ze sposobem zdobywania wiedzy, m.in. zwyczaj samodzielnego poszerzania wiedzy (np. sięganie do literatury popularno-naukowej), udział w konkursach i olimpiadach przedmiotowych, zamiłowanie do czytania książek. Grupa trzecia obejmuje m.in. styl nauki – samodzielnie, czy z kimś, z koleżanką / kolegą teraz a w dzieciństwie z mamą, babcią lub starszym rodzeństwem oraz formy utrwalania zdobytych wiadomości – głośne powtórki, brak powtórek, uczenie się tylko od sprawdzianu do sprawdzianu, preferowanie odpowiedzi ustnych w stosunku do form pisemnych, bądź na odwrót.
5. Ze względu na wykorzystanie w systemie diagnostyczno-nauczającym specyficznego narzędzia nauki jakim jest komputer należy w trakcie gromadzenia danych uwzględnić również indywidualne doświadczenia w tym zakresie (**DK**). I tak tutaj istotne informacje dotyczą: posiadania komputera bądź dostępu do niego na uczelni i poza nią, znajomości jego obsługi, rodzaju stosowanych programów użytkowych oraz ogólnej wiedzy na temat przydatności i zastosowań komputerów w życiu codziennym.
6. I wreszcie grupa ostatnia, nazwana umownie aktualnymi uwarunkowaniami (**AW**), która ma zawierać ogólnie rzecz ujmując dwa rodzaje informacji. Jeden rodzaj to: płeć, wiek, kierunek studiów, uczelnia, miejsce zamieszkania (dom rodzinny, stacja, akademik), ilość godzin zajęć w tygodniu. Do tej grupy należą również pytania o sposób spędzania wolnego czasu, liczba godzin przeznaczanych w tygodniu na naukę własną, motyw wyboru kierunku studiów, dostęp do potrzebnej literatury.

Wymienione wyżej przykładowe elementy wektora cech mają bardzo różnorodny charakter. Część z nich to cechy ilościowe (np. wiek, liczba rodzeństwa, ilość godzin zajęć na uczelni), część to typowe cechy jakościowe (np. miejsce zamieszkania, wykształcenie rodziców, profil szkoły itd.), a także cechy o wartościach typu tak – nie, np. czy lubisz powtarzać na głos wyuczony materiał, czy uczysz się sam? Jednakże w trakcie badań postanowiono tę ostatnią grupę, w celu ujednolicenia procedury gromadzenia danych przekształcić na grupę cech jakościowych dopuszczając odpowiedzi typu: rzadko, wcale, często, zazwyczaj itp. Każda z wyszczególnionych grup wnosi pewien wkład w opis badanego obiektu, którym w tym wypadku jest człowiek. Z oczywistych względów opis ten jest niepełny i nieprecyzyjny. Przyjęto jednak, że zgromadzone dane dają podstawy do rozpoczęcia procesu rozpoznawania

4.2 Ujęcie matematyczne

W rozdziale dotyczącym struktury systemu zdefiniowano odwzorowanie nazwane recepcją (por. wzór 2.2), które obiektom przestrzeni $\mathbf{D}$ przyporządkowuje punkty w n -wymiarowej przestrzeni $\mathbf{X}$. Elementy zbioru $\mathbf{D}$ (zarówno wzorcowe należące do ciągu uczącego, jak i te nieznane podlegające rozpoznawaniu) zostały scharakteryzowane w poprzednim rozdziale (por. rozdz.3.5.), pora zatem przyjrzeć się bliżej strukturze przestrzeni $\mathbf{X}$. Jest ona w znacznym stopniu zdeterminowana przez przyjęte metody pomiaru właściwości rozważanych obiektów.

Najbardziej pożądaną postacią składowych wektora $\underline{x}$ byłyby liczby rzeczywiste, tak aby uzyskana przestrzeń cech była przestrzenią euklidesową ($\mathbf{X} \subseteq \mathbb{R}^n$), gdyż wtedy bez problemu można posługiwać się pojęciem metryki i stosować typowe algorytmy rozpoznawania. Nie zawsze jednak jest to możliwe, czego dowodzi także przykład badań opisywanych w niniejszej pracy.

Zdecydowana większość danych zgromadzonych dla potrzeb opracowywanego algorytmu rozpoznawania pochodzi z ankiety zawierającej 74 pytania, na które można wybrać dokładnie jedną z kilku ponumerowanych odpowiedzi. Zakładając, że każde pytanie ankiety odzwierciedla jedną badaną cechę naturalną formą zdefiniowania funkcji recepcji (B) wydaje się być przyporządkowanie każdej cesze liczby będącej numerem odpowiedzi wskazanej przez ankietowanego na odpowiednie pytanie. Odwzorowanie B_1 przyjmie wówczas postać:

$$B'_p: \mathbf{D} \rightarrow \mathbf{X}' \quad (4.1)$$

gdzie:

$$\mathbf{X}' = \{ \underline{x}_i : \underline{x}_i = [x_i^1, x_i^2, x_i^3, \dots, x_i^n]; x_i^k \in \{1..6\}, k=1..n, i=1..M \} \quad (4.2)$$

n – liczba pytań w ankiecie

M – liczba badanych obiektów

W rozważanym przypadku liczba pytań ankiecie wynosi 74, a liczba badanych obiektów $M = 108$. Zapis $x_i^k \in \{1..6\}$ ma w omawianym kontekście znaczenie czysto umowne, gdyż oznacza on, że poszczególne składowe wektora są zapisane w postaci symboli z sześcioelementowego zbioru cyfr $\{1..6\}$.

Nie jest to jednak jeszcze kompletny opis przestrzeni $\mathbf{X}$, gdyż oprócz pytań ankiety brane są pod uwagę wyniki testu $\mathbf{WT}$ (4 parametry opisane przez: ilość poprawnych odpowiedzi, ilość dokonanych poprawek, ilość powrotów do prezentowanego tekstu, czas wykonania zadań testowych) oraz wspomniane wcześniej zmienne towarzyszące: płeć i kierunek studiów (należące do grupy $\mathbf{AW}$). Razem tworzą one wektor o długości 80 składowych. Zdefiniowaną wyżej przestrzeń $\mathbf{X}'$ należy zatem rozszerzyć

o wymienione wyżej elementy. Choć wyniki testów są rzeczywiście liczbami, dla zachowania jednolitego charakteru danych sprowadzono je także do postaci nominalnej wprowadzając - dla liczby poprawnych odpowiedzi, liczby powrotów do tekstu oraz liczby dokonanych poprawek - wartości:

- ♦ brak
- ♦ mało
- ♦ dużo
- ♦ bardzo dużo

oraz dla czasu wykonania zadań wartości:

- ♦ krótki
- ♦ średni
- ♦ długi

Podobnie zmienna płeć otrzymała oznaczenia:

1. kobieta
2. mężczyzna

a kierunek studiów:

1. matematyka
2. ekonomia
3. historia
4. wychowanie techniczne

Należy podkreślić, że jakkolwiek uzyskana w ten sposób przestrzeń cech pozornie opisana jest przez wartości ze zbioru liczb całkowitych, nie da się praktycznie nic powiedzieć o jej charakterze. Nie są znane wektory bazowe tej przestrzeni, a co za tym idzie, nie można stwierdzić, czy są ortogonalne, co jak okaże się później, upoważniałoby do stosowania metryk definiowanych np. dla przestrzeni euklidesowych. Po wprowadzeniu do definicji odwzorowania B''_p modyfikacji dotyczących przekształcania dołączanych wartości na odpowiadające im cyfry otrzymuje się odwzorowanie B_p , które przekształca przestrzeń obiektów $\mathbf{D}$ w 80-wymiarową przestrzeń cech $\mathbf{X}$.

$$B_p : \mathbf{D} \rightarrow \mathbf{X} \quad (4.3)$$

gdzie

$$\mathbf{X} = \mathbf{ZI} \cup \mathbf{ZP} \cup \mathbf{CS} \cup \mathbf{HE} \cup \mathbf{DK} \cup \mathbf{AW} \cup \mathbf{WT} \quad (4.4)$$

Indeks p przy symbolu odwzorowania zastosowano dla odróżnienia tej postaci (prostej) funkcji recepcji od wprowadzonych poniżej bardziej właściwych sposobów odwzorowania zbioru obiektów w przestrzeń cech.

4.3 Kodowanie danych

Opisane w poprzednim punkcie przyporządkowanie kolejnym cechom obiektu numeru wybranej odpowiedzi jest intuicyjnie najprostszym, ale z punktu widzenia praktycznych zastosowań najgorszym sposobem tworzenia wektora cech. Sprowadza on bowiem wszystkie dane, niezależnie od ich początkowego charakteru, do poziomu skali nominalnej, na której formalnie rzecz biorąc jedyną dopuszczalną operacją jest zliczanie. A zatem próba potraktowania tak otrzymanych wartości w sposób przyjęty dla danych ilościowych jest daleko idącym uproszczeniem. Wprawdzie, w przypadku pomiarów psychologicznych niektórzy teoretycy [32] dopuszczają możliwość wykonywania operacji przewidzianych dla wyższych skal w odniesieniu do skali nominalnej ale czynią to z zastrzeżeniem, że badacz musi zachować daleko idącą ostrożność przy wyciąganiu wniosków na podstawie tak uzyskanych wyników. Choć w opisywanych badaniach sprawdzono użyteczność prostego kodowania cech w zadaniu rozpoznawania, i to z zaskakująco dobrym skutkiem, próbowano także stosować inne formy ujednolicenia charakteru danych.

4.3.1 Kodowanie 1-z-N

W przypadku niejednolitego charakteru cech najczęściej stosowaną formą opisu jest kod nazywany 1-z-N. Algorytm kodowania w tym systemie można przedstawić w następujących krokach:

1. Dla każdej cechy tworzony jest tzw. wektor instrumentalny [9] o długości N_k równej liczbie różnych wartości przyjmowanych przez tę cechę ($N_k \in \langle 2, 6 \rangle$);
2. Łączna długość wektora cech, a zatem wymiar przestrzeni X równy jest sumie długości wektorów instrumentalnych

$$N = \sum_{k=1}^n N_k \quad (4.5)$$

gdzie

n – liczba cech opisujących w przestrzeni X obiekt $d \in D$

3. Wektor instrumentalny opisujący k -tą zmienną ma postać:

$$\underline{x}^k = [x^{k,1}, x^{k,2}, \dots, x^{k,N_k}] \quad (4.6)$$

przy czym:

$$x^{k,j} = \begin{cases} 1 & \text{dla } j = x^{k(p)} \\ 0 & \text{dla } j \neq x^{k(p)} \end{cases}$$

gdzie indeks (p) sygnalizuje wartość cechy w odwzorowaniu B_p , czyli numer odpowiedzi wybranej na danego pytania.

4. Wszystkie wektory instrumentalne zapisane w kolejności wystąpienia cech tworzą wektor w N-wymiarowej przestrzeni $\mathbf{X}$ (N określone wg wzoru (4.5))

Odwzorowanie B , oznaczone dla odróżnienia od poprzedniego symbolem B_{1zN} przyjmie teraz postać daną wzorem (4.7):

$$B_{1zN} : \mathbf{D} \rightarrow \mathbf{X} \quad (4.7)$$

gdzie:

$$\mathbf{X} = \{\underline{x}_i : \underline{x}_i = [\underline{x}_i^1, \underline{x}_i^2, \dots, \underline{x}_i^N]\} \quad (4.8)$$

przy czym: $\underline{x}_i^k$ - wektor instrumentalny k-tej składowej i-tego obiektu określony wg wzoru (4.6)

4.3.2 Kodowanie z wykorzystaniem szeregów kumulacyjnych

Analiza wybieranych odpowiedzi pozwoliła stwierdzić, że część z nich, choć teoretycznie prawdopodobnych, praktycznie nigdy nie występowała. Potwierdziła to próba zastosowania testu niezależności χ^2 podjęta w celu redukcji ilości cech, która wykazała, iż liczebności niektórych grup odpowiedzi były zerowe lub bliskie zeru. Biorąc pod uwagę słabą skuteczność stosowanych wcześniej sposobów kodowania oraz występujące w nich ograniczenia sformułowano kolejną próbę odwzorowania rzeczywistości. Otóż, tym razem bierze się pod uwagę ilość osób, które wybrały dany wariant odpowiedzi na k-te pytanie i oblicza frakcje tych odpowiedzi w stosunku do ogółu badanych. Przy czym dla kolejnych odpowiedzi uwzględnia się sumę osób, które wybrały ten lub jeden z poprzednich możliwych wariantów. W ten sposób liczebności zerowe mogą wystąpić tylko dla pierwszej odpowiedzi na każde pytanie, a jak wynika z uzyskanych danych sytuacja taka nigdy nie zaistniała. Otrzymane kody przyjmują wartości liczb rzeczywistych z przedziału $<0,1>$. Algorytm kodowania zostanie zapisany z wykorzystaniem, stosowanej często w literaturze informatycznej, notacji pascalo podobnej.

Przyjmując oznaczenia:

k – numer cechy

N – liczba cech
 N_k – ilość różnych wariantów k -tej cechy (= ilość możliwych odpowiedzi na k -te pytanie)
 j – numer odpowiedzi
 i – numer obiektu
 x_i^k – wartość k -tej cechy dla i -tego obiektu
 $ilob$ – liczba wszystkich badanych obiektów
 s – liczba obiektów, których odpowiedzi są jednakowe
 $f[1..N_k]$ – nowe wartości k -tej cechy

to sposób obliczania nowych wartości k -tej cechy można przedstawić w następującej postaci:

```

procedure frakcje (k; var f);
begin
  for j := 1 to  $N_k$  do
    begin
      for i := 1 to  $ilob$  do if  $x_i^k \leq j$  then inc(s);
      f [j] := s /  $ilob$ ;
    end;
  end;

```

Z powyższego zapisu wynika również, że dla każdej cechy obliczone frakcje (f) mają następujące własności:

$$\bigvee_{j=1..N_k} s < ilob \Rightarrow f[j] \leq 1 \quad (4.9)$$

oraz:

$$f[j] = 0 \Leftrightarrow s = 0 \quad (4.10)$$

Wyniki rozpoznawania przy pełnym zestawie (80) cech przedstawia tabela A.I (Dodatek A). Niezależnie od przyjętego sposobu kodowania są one dalekie od zadowalających. Prowadzi to do przypuszczenia, że przyjęty sposób konstruowania przestrzeni X nie ma w przypadku opisywanych badań decydującego znaczenia dla procesu rozpoznawania, a w konsekwencji, że należy poszukiwać innych czynników „znieskształcających” analizowany obraz predyspozycji psychologicznych ucznia.

4.4 Redukcja wymiaru przestrzeni cech

Pierwszym z przyjętych kierunków poszukiwań zmierzających do poprawy rezultatów rozpoznawania było zmniejszenie wymiaru przestrzeni X . Duża ilość składowych wektora cech daje, z jednej strony szansę na dokładne odwzorowanie badanych obiektów, równocześnie jednak stwarza ryzyko wprowadzenia dużej liczby zmiennych nie tylko nieistotnych z punktu widzenia wnoszonych przez nie informacji, ale co więcej, być może zaciemniających rzeczywisty obraz badanych obiektów. Przypuszczenia te potwierdziła analiza macierzy korelacji rozpatrywanych cech. Przy poziomie istotności 0.05 współczynnik korelacji przyjmuje wartości od -0.41 (cechy nr 9 i 10) do $+0.75$ (cechy nr 41 i 42).

W trakcie analizy pełnej macierzy korelacji zauważono pewną prawidłowość, która zostanie wykorzystana w dalszych rozważaniach. Z obserwacji tej wynika np., że silnie skorelowane ze sobą są cechy: opisujące początki nauki czytania (***Gdzie poznawałeś pierwsze litery?*** – pytanie nr 9 oraz ***Kiedy zacząłeś czytać książki samodzielnie?*** – pytanie nr 10), oraz cechy dotyczące ocen szkolnych (***Jakie oceny dostawałeś najczęściej z matematyki w szkole średniej?*** – pytanie nr 41 oraz ***Jak sam oceniałbyś swoją znajomość matematyki w szkole średniej?*** – pytanie nr 42). Przy czym ujemna korelacja w przypadku pierwszej pary pytań wynika z faktu, że z odpowiedzią o najwyższym numerze w pytaniu 9 (*D – w domu, uczyłem się sam*) najlepiej koresponduje odpowiedź A w pytaniu 10 (*w pierwszej klasie*), czyli im bardziej samodzielny był uczeń w nauce czytania, tym wcześniej zaczął wykorzystywać tę umiejętność.

Podane przykłady pokazują jedynie pary cech najsilniej ze sobą skorelowanych, dokładniejsza ich analiza zostanie dokonana w jednym z kolejnych podrozdziałów. Dołączenie pełnej macierzy korelacji cech uznano za bezcelowe gdyż ze względu na swoje rozmiary (80×80) macierz ta zajmuje 16 stron formatu A4 co sprawia, że analizowanie jej w całości jest bardzo utrudnione. Natomiast przedstawione w Dodatku E odpowiednie fragmenty tej macierzy zawierają współczynniki korelacji (najwyższe i najniższe) dla wybranych cech.

Spotykane w literaturze ([21],[63]) sposoby podejścia do problemu redukcji cech można najogólniej podzielić na dwie grupy.

Grupa pierwsza – to metody zmierzające do uporządkowania zbioru cech według przyjętej hierarchii ważności i odrzucenia cech mniej istotnych;

Grupa druga – skupia metody polegające na grupowaniu cech i wyborze reprezentantów poszczególnych grup;

4.4.1 Uporządkowanie cech według rosnącej ilości informacji

Najczęściej stosowanym kryterium oceny przydatności cech jest ilość dostarczanych przez nie informacji, zwłaszcza umożliwiających prawidłową klasyfikację obiektów.

Operację taką umożliwia np. przekształcenie Karhunen-Loewego ([17],[63]), znane także w literaturze jako metoda składowych głównych [66]. Polega ona na obliczeniu macierzy rozprożeń wewnątrzgrupowych i międzygrupowych (macierzy kowariancji) wszystkich cech, obliczeniu jej wartości własnych a następnie uszeregowaniu tych wartości w kolejności malejącej. Wektory własne n pierwszych wartości własnych tworzą wektory bazowe nowej, zredukowanej przestrzeni cech. Można udowodnić, że otrzymane wektory są ortogonalne, czyli umożliwiają metryzację przestrzeni X , a w konsekwencji pozwalają na wykorzystanie w algorytmie rozpoznawania metod minimalnoodległościowych.

Inne podejście zmierzające również do selekcji cech na podstawie ilości wnoszonej przez nie informacji odwołuje się do tzw. miary Shannona [63]. W podejściu tym rozpatruje się jednak nie udział poszczególnych cech w określaniu klasyfikacji obiektów, ale ich wybranych podzbiorów, zmierzając do ustalenia, który podzbiór dostarcza największą ilość informacji.

Poważnym mankamentem tej metody, nawet przy znanych prawdopodobieństwach badanych obiektów, jest konieczność obliczania wartości ilości dostarczanej informacji dla wszystkich możliwych n-elementowych podzbiorów N-elementowego zbioru cech, co przy dużych wartościach N (a tak jest w omawianym przypadku) jest zadaniem praktycznie niewykonalnym nawet przy pomocy komputera.

Niestety, obie wymienione wyżej metody redukcji dotyczą danych ilościowych, dla których możliwe jest obliczanie macierzy kowariancji i parametrów rozkładu prawdopodobieństwa. W badaniach opisywanych w niniejszej pracy charakter danych (głównie jakościowych) uniemożliwia wykorzystanie opisywanych metod.

4.4.2 Redukcja wymiaru przestrzeni przez wybór cech reprezentantek

W obu wymienionych wcześniej metodach wszystkie zmienne traktowano jako wartości niezależne, różniące się jedynie ilością wnoszonej przez nie informacji. Tymczasem okazuje się, że często składowe wektorów przestrzeni cech układają się w naturalne skupienia, reprezentujące cechy podobne lub powiązane ze sobą. Można wówczas zredukować liczbę cech, a w konsekwencji zmniejszyć szum informacyjny wnoszony przez cechy mniej istotne dzięki zastosowaniu metody polegającej na wyborze tzw. cech reprezentantek [21]. Algorytm postępowania zawiera się w dwóch krokach:

1. wyłonienie naturalnych skupisk cech (o ile takie istnieją)
2. wskazanie dla każdego skupiska jednej cechy, reprezentującej wszystkie czynniki umieszczone w danym skupieniu.

Możliwe jest też postępowanie odwrotne, tzn.:

1. najpierw wskazuje się cechy wiodące (centra skupień)
2. do podanych cech dołącza się cechy do nich podobne (w sensie zdefiniowanego kryterium)

Podejście takie można zaobserwować w analizie czynnikowej, czy w aglomeracyjnej metodzie analizy skupień ([9],[48]).

Poniżej zamieszczono krótki opis użytych metod statystycznego grupowania cech oraz ocenę ich przydatności z punktu widzenia prowadzonych badań.

4.4.2.1 Metoda taksonomii struktur

Metodą grupowania stosunkowo prostą w obliczeniach i, jak się wydaje, adekwatną dla posiadanych danych była, zaproponowana przez Sokołowskiego i Chomątowskiego, metoda taksonomii struktur. Zdefiniowany przez autorów algorytm nazwany metodą eliminacji wektorów został opisany w [12]. Kolejne jego kroki w zastosowaniu do problemu grupowania cech dla potrzeb opisywanych badań można przedstawić następująco:

1. Punktem wyjścia dla obliczeń omawianą metodą jest utworzenie macierzy P. Powstaje ona z tablicy kontyngencji o wymiarach $N \times N$ tworzonej tak, jak dla testu niezależności χ^2 .
2. Następnie należy odczytać z tablic wartość krytyczną $w_{kryt} = \chi^2$ dla $(t-1)(r-1)$ stopni swobody

oraz

3. utworzyć zero - jedynkową macierz Q według następującej reguły:

$$Q(i, j) = 0 \Leftrightarrow P(i, j) > \alpha \chi^2_{(t-1)(r-1)} \quad (4.11)$$

oraz

$$Q(i, j) = 1 \Leftrightarrow P(i, j) \leq \alpha \chi^2_{(t-1)(r-1)}$$

gdzie:

i, j - numery pytań $i, j = 1 \dots n$

N - wymiar przestrzeni X (przestrzeni cech)

t - ilość odpowiedzi dla pytania i
 r - ilość odpowiedzi dla pytania j
 $\alpha = 0.01$ - przyjęty poziom ufności
 $\chi^2_{(t-1)(r-1)}$ - wartość krytyczna χ^2 przy poziomie ufności α i $(t-1)(r-1)$ stopniach swobody

4. Kolejnym krokiem jest obliczenie składowych wektora $\mathbf{p}_0$ według wzoru:

$$\mathbf{p}_0 = \mathbf{Q} \bullet \mathbf{I}$$

gdzie:

$\mathbf{I}$ - wektor kolumnowy o n składowych równych jednościami. Każda składowa wektora $\mathbf{p}_0$ jest więc sumą odpowiedniego wiersza macierzy $\mathbf{Q}$.

5. Maksymalna wartość w wektorze $\mathbf{p}_0$ wskazuje obiekt, który jest niepodobny (na poziomie w_{kryt}) do największej liczby pozostałych obiektów. Wartość ta wskazuje, którą kolumnę i odpowiadający jej wiersz należy wyeliminować z macierzy $\mathbf{Q}$. Jeżeli w wektorze $\mathbf{p}_0$ więcej niż jedna składowa jest równa wartości maksymalnej to do wyboru wiersza i kolumny eliminowanych w danym kroku można wykorzystać kryterium pomocnicze, choć, jak podają autorzy, w praktyce kolejność wyboru kolumny i wiersza nie ma zasadniczego wpływu na uzyskiwane rezultaty.

6. Czynności podane w krokach 4 i 5 należy powtarzać do momentu, aż wszystkie składowe wektora $\mathbf{p}_0$ będą równe 0 (co odpowiada sytuacji, gdy wszystkie wiersze i kolumny pozostałe w macierzy $\mathbf{Q}$ mają wartość 0).

7. Obiekty odpowiadające wierszom, które pozostały w macierzy $\mathbf{Q}$ tworzą pierwszą podgrupę.

8. Pozostałe po wyeliminowaniu obiektów bieżącej podgrupy wiersze i kolumny macierzy $\mathbf{Q}$ tworzą nowe wartości tej macierzy i dla niej powtarzane są kroki 3-7, aż do momentu gdy wszystkie obiekty zostaną pogrupowane.

Aby możliwe było wykorzystanie powyższego algorytmu dla zbioru ponad 100 obiektów opisanych przez 80 cech konieczna okazała się jego implementacja komputerowa. Algorytm zapisano w języku Pascal, którym posłużono się także przy programowaniu innych elementów projektowanego systemu (ankiety i algorytmu rozpoznawania).

Macierz $\mathbf{Q}$ stanowiła dane wejściowe dla algorytmu, którego zadaniem było wyłonienie grup cech. Niestety rezultatem jego działania okazał się być podział zupełnie nieprzydatny z punktu widzenia redukcji przestrzeni cech. Uzyskano wprawdzie 13 grup, ale 12 z nich było jednoelementowych, podczas gdy wszystkie pozostałe cechy znalazły się w jednej grupie. Porównanie grup pytań z naturalną strukturą wynikającą z charakteru mierzonych danych doprowadziło do odrzucenia hipotezy o wystąpieniu tak nierównomiernego podziału pytań w rzeczywistości i skłoniło do poszukiwania innej metody zmniejszenia długości wektora cech.

4.4.2.2 Zastosowanie analizy skupień

Podobny efekt dała również próba wykorzystania analizy skupień do grupowania cech. O ile w odniesieniu do obiektów grupowanie przyniosło oczekiwane rezultaty w postaci 4 skupień {por. rozdz. 5.1), to dla cech, podobnie jak w metodzie taksonomii struktur otrzymany podział był nie tylko zbyt nierównomierny, ale przede wszystkim całkowicie niezgodny z rzeczywistym układem cech. Próbę podziału zbioru cech wykonywano dla różnych ilości skupień (od 7 do 13). Za każdym razem występowało jedno duże skupienie, liczące od 23 do 26 cech, 3 -5 skupień kilkuelementowych i podobna ilość skupień zawierających tylko po jednym elemencie.

Analiza pogrupowanych numerów pytań ujawniła np., że do tej samej grupy zostały zakwalifikowane pytania badające zupełnie odmienne cechy osobowości (np. *Ile czasu dziennie spędzasz przed telewizorem?* oraz *Co sądzisz o studiowaniu na odległość?*), natomiast wyraźnie ze sobą powiązane pary pytań takich jak: *Czy, kiedy byłeś młodszy potrzebowaleś obecności dorosłych w czasie odrabiania lekcji?* oraz następujące po nim: *Jakiego rodzaju pomocy potrzebowaleś?* trafiły do dwóch różnych skupień.

4.4.3 Redukcja wieloetapowa

Wobec słabej użyteczności wcześniej omawianych metod postanowiono odwołać się do charakteru analizowanych danych i kryteriów ich wyłaniania.

Redukcję przeprowadzono w dwóch etapach:

Etap I – wyeliminowanie cech nadmiarowych, nie wnoszących informacji istotnych z punktu widzenia procesu rozpoznawania;

Etap II – wybór reprezentantów odzwierciedlających rzeczywistą strukturę rozpatrywanych cech.

Etap I przebiegał również w kilku krokach. Jedną z zasad tworzenia ankiety jest uzyskiwanie tych samych informacji przez powtarzanie pytań w zmienionej formie. Chodzi o to, aby wyeliminować działanie tzw. czynnika kłamstwa, wynikającego z udzielania odpowiedzi nieprawdziwych, bądź nieścisłych. Na przykład pytania 20 i 21: *Czy masz w domu inne obowiązki poza nauką?* oraz *Ile czasu dziennie zajmują ci te obowiązki?*, przy wybraniu odpowiedzi A (*tak, ale niewiele*) w pytaniu 20 i odpowiedzi A (*kilka godzin dziennie*) w pytaniu 21 pozostają ze sobą w sprzeczności. Taką parę lub grupę pytań należy wówczas rozpatrywać łącznie i dopiero poprzez koniunkcję wyników uzyskuje się odpowiedź końcową. W zestawie pytań będących źródłem danych w niniejszej pracy było 7 takich par pytań i jedna grupa złożona z 6 pytań. Pomocnicze

tablice powiązań logicznych między pytaniami umożliwiły znalezienie odpowiedzi wynikowych. Zastąpienie wyników surowych odpowiedziami przetworzonymi pozwoliło na redukcję 14 cech.

W trakcie tworzenia tablic wielodzielnych dla testu niezależności χ^2 okazało się, że niektóre odpowiedzi nie były nigdy wybierane, dając zerowe liczebności w niektórych polach tablic. Nasunęło to przypuszczenie, iż niektóre cechy rzadko występują również w całej populacji, skoro nie natrafiono na nie w ponad 100-osobowej próbie i wobec tego ich udział w opisie badanych predyspozycji jest na tyle mało istotny, że można go pominąć. Co więcej, szczegółowa analiza wszystkich odpowiedzi ujawniła jeszcze jedną prawidłowość. Mianowicie w grupie zadawanych pytań były również takie, na które wszyscy, lub prawie wszyscy wybierali tę samą odpowiedź. Z punktu widzenia wartości informacyjnej pytania te również miały wartość zerową (lub bliską zeru). Podjęto zatem decyzję o pominięciu tych pytań w wektorze cech, jako nie wnoszących żadnego czynnika różnicującego.

Dodatkowym źródłem informacji była wspomniana w rozdziale 4.4. macierz korelacji cech i zaobserwowane w niej prawidłowości. Ujawniła ona mianowicie istnienie silnych powiązań między niektórymi grupami pytań (por Dodatek E), co potraktowano również jako wskazanie do wyeliminowania części skorelowanych cech.

Usunięcie pytań wynikających z podanych wyżej obserwacji pozwoliło na istotną redukcję cech, dając w rezultacie wektor o długości 31 składowych.

Wobec wyczerpania możliwości prostej redukcji cech przystąpiono do **etapu II** zmierzającego do wyłonienia cech - reprezentantów. Zbiór cech dzielony jest zwykle na kilka grup, a następnie podejmuje się decyzję, które cechy są najbardziej reprezentatywne dla poszczególnych grup.

W niniejszej pracy przyjęto następujący algorytm postępowania:

1. Podział na grupy

W trakcie tworzenia ankiety (por. rozdz. 4.2) czynniki opisujące predyspozycje psychologiczne ucznia zgrupowane były w następujących sześciu kategoriach :

- zdolności intelektualne - **ZI**
- zdolności poznawcze - **ZP**
- czynniki środowiskowe - **CS**
- historia edukacji - **HE**
- doświadczenia komputerowe - **DK**
- aktualne uwarunkowania - **AW**

do których następnie dołączono kategorię siódmą zawierającą wyniki testu (**WT**).

Suma wymienionych zbiorów: tworzy przestrzeń **X**. (por. wzór 4.4)

$$X = ZI \cup ZP \cup CS \cup HE \cup DK \cup AW \cup WT$$

Początkowo wektor $\underline{x}$ liczył 80 cech. W wyniku redukcji przeprowadzonej w **etapie I** długość wektora zmniejszyła się do 31 składowych. Pozostałe cechy rozkładały się mniej więcej równomiernie pomiędzy grupami dając ich liczebności odpowiednio: 3, 4, 4, 5, 6, 5 i 4 cechy w grupie, przy czym ostatnia wartość oznacza parametry otrzymywane z testu i nie była redukowana.

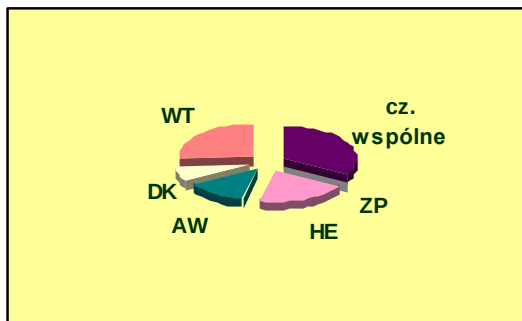
2. Wybór reprezentantów

Dla każdej grupy obliczono (na podstawie wzoru Shannona) ilość informacji dostarczanej przez kolejne cechy, przyjmując za wartość prawdopodobieństwa każdej kategorii cechy częstość jej wystąpień w obrębie badanej populacji. Jako reprezentantów grup wybrano cechy wnoszące największą ilość informacji w poszczególnych grupach. Otrzymano w ten sposób 7 cech podstawowych. Jednak rozpoznawanie przeprowadzone na zbiorze obiektów opisywanych przez tak zredukowany zestaw cech pokazało, że wyłoniłone cechy nie różnicują badanych obiektów w stopniu umożliwiającym ich poprawną klasyfikację.

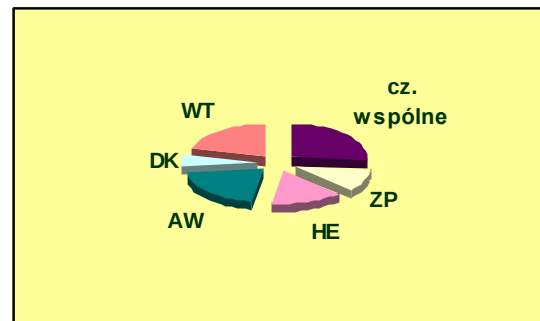
3. Dołączenie cech uzupełniających

W tej sytuacji zdecydowano się dołączyć zbiór cech pomocniczych. Wobec braku jasno określonego kryterium wyboru, chcąc zminimalizować ryzyko błędu wynikającego z niepoprawnych decyzji przyjęto trzy zestawy cech, nazwane odpowiednio Wariant I, Wariant II i Wariant III. Za każdym razem zestaw zawierał 7 cech wyłoniłonych w kroku drugim, oraz od 4 do 8 cech uzupełniających.

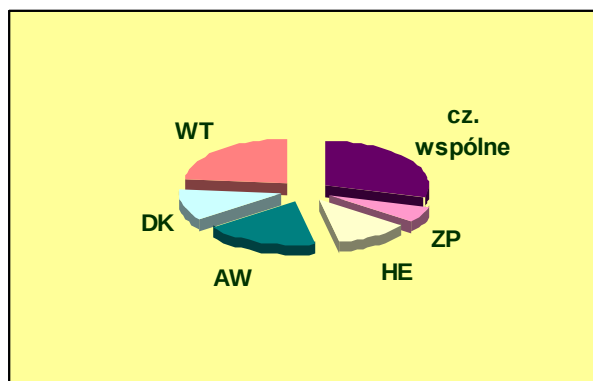
Wybrane cechy utworzyły nowe wektory przestrzeni cech o znacznie zredukowanej liczbie wymiarów (11,15,13) i posłużyły jako podstawa do opisu ciągu uczącego i zbioru badanych obiektów w algorytmie rozpoznawania (por. rozdz. 5). Rys. 4.1 przedstawia wykres udziału poszczególnych grup czynników w różnych zestawach cech.



Rys 4.1a) Wariant I



Rys 4.1b) Wariant II



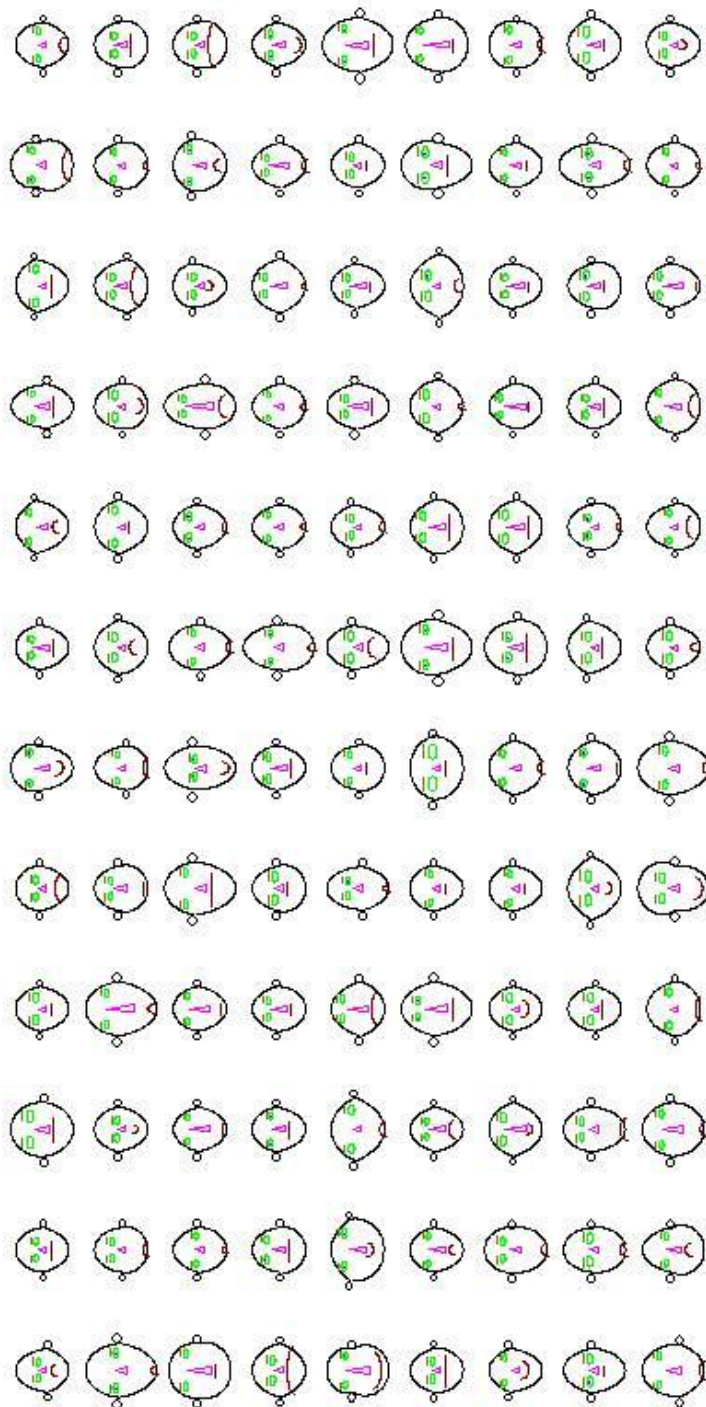
Rys. 4.1c) Wariant III

4.5 Ilustracja graficzna wybranych zestawów cech

Jak już wielokrotnie podkreślano w tej pracy, nie istnieją jednoznaczne kryteria pozwalające określić jakie cechy osobowości i czynniki zewnętrzne mają największy wpływ na kształtowanie się stylu nauki. Ta sytuacja zresztą była głównym powodem sięgnięcia po metody sztucznej inteligencji przy doborze indywidualnego stylu uczenia się. Pokazany na rysunku 4.2 wykres obrazkowy (tzw. twarze Chernoffa) ilustruje rozkład cech dla wariantu I, zapisanych w kodzie kumulacyjnym. Sposób kodowania nie ma jednak w tym przypadku większego znaczenia, gdyż jak sprawdzono, graficzna złożoność układu cech przy różnych systemach kodowania jest porównywalna. Każda twarz na wykresie przedstawia jednego ucznia. Każdy element rysów twarzy odpowiada jednej składowej wektora cech zgodnie z legendą zamieszczoną pod wykresem. Już pobieżna analiza przedstawionych twarzy pokazuje jak trudne jest zadanie klasyfikacji tak opisanych obiektów.

Nieco bardziej czytelny wykres dla tego samego zestawu cech przedstawia rysunek 4.3, na którym poszczególne składowe wektora $\underline{x}$ przyjmują kształt promieni o wspólnym środku. Jednak i tym razem przedstawione obiekty, choć pozornie do siebie podobne trudno byłoby podzielić na cztery grupy (a taka jest przewidywana ilość metod nauczania) tylko na podstawie układu cech.

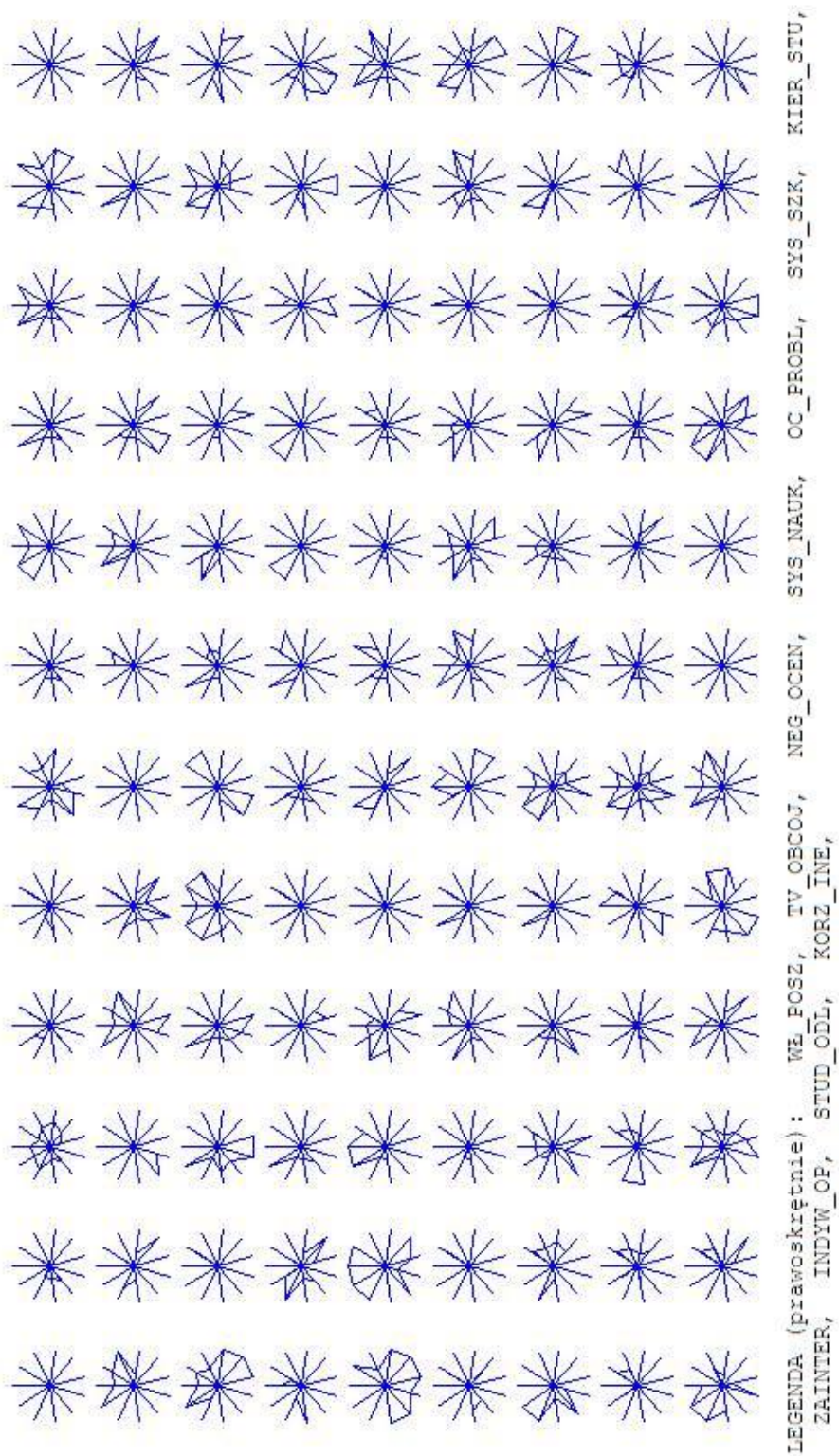
Wykres obrazkowy (szereg w1.STA 11v*108c)



LEGENDA: twarz/szer. = WŁ_POSZ, uszy/poziom = TV_OBCOJ, twarz/po1.wys. = NEG_OCEN,
 górna po1. twarz/eksc. = SYS_NAUK, dolna po1. twarz/eksc. = OC_PROBL, nos/dig. = SYS_SZK,
 usta/środ. = KIER_SIU, usta/krzyw. = ZAINIER, usta/dig. = INDIW_OP, uszy/wys. = STUD_ODL,
 oczy/odst. = KORZ_INE,

Rys. 4.2. Wykres układu cech dla wariantu I (Twarze Chernoffa) 11 cech x 108 przypadków

Wykres obrazkowy (szereg gw1.STA 11v*108c)



Rys. 4.3. Wykres układu cech dla wariantu I (Promienie) 11 cech x 108 przypadków

5 Rozpoznawanie

Omówione w poprzednim rozdziale problemy związane z reprezentacją przestrzeni cech stanowią w praktyce bardzo istotny etap zadania rozpoznawania. Od wyboru cech, opisujących badane obiekty zależy zarówno jakość samego rozpoznawania (wyrażana ilością rozpoznań poprawnych), jak i złożoność algorytmów obliczeniowych. Rodzaj zgromadzonych danych implikuje bowiem klasę metod jakie mogą być użyte do rozpoznawania, jak również czasami rodzaj przyjętej metryki. Im dokładniej potrafimy określić właściwości przestrzeni cech, wyrażające się na przykład w ortogonalnym charakterze wektorów bazowych, tym prościej dobrać odpowiednią metrykę i optymalny algorytm. Tak idealne przypadki występują jednak niezwykle rzadko i praktycznie tylko tam, gdzie, zazwyczaj na podstawie wcześniejszych badań, dopracowano się już pewnych wzorców postępowania.

5.1 Ciąg uczący

Zastosowane w prowadzonych badaniach algorytmy zakładają tzw. rozpoznawanie z uczeniem się. Konieczne jest zatem zdefiniowanie zbioru przykładów o znanej przynależności, na podstawie których algorytm rozpoznający mógłby podejmować decyzje dotyczące innych, nieznanych obiektów.. Zbiór ten nazywany jest ciągiem uczącym.

5.1.1 Definicja ciągu uczącego

Przez ciąg uczący rozumie się zbiór par $(\underline{x}^k, i^k)$, w których $\underline{x}^k$ – to wektor przestrzeni cech opisujący k -ty obiekt, a i^k – to numer klasy, do której ten obiekt należy. Symbolicznie ciąg uczący można zapisać w następującej postaci:

$$C_u = \{ (\underline{x}^k, i^k), k = 1, \dots, M \} \quad (5.1)$$

gdzie:

M – liczba obiektów ciągu

$$\underline{x}^k = B(d^k) \wedge d^k \in \mathbf{D}$$

oraz

$$i^k \in \mathbf{I} \wedge i^k = \hat{A}(d^k)$$

Zakłada się również, że w momencie tworzenia ciągu uczącego znany jest podział zbioru $\mathbf{D}$ na odpowiednie klasy i wybrane przykłady reprezentują wszystkie klasy.

5.1.2 Klasyfikacja elementów ciągu uczącego

Ze względu na rolę, jaką odgrywają elementy ciągu uczącego w zadaniu rozpoznawania prawidłowe określenie ich przynależności może w zasadniczy sposób wpływać na jakość działania całego algorytmu rozpoznającego. Z definicji ciągu uczącego wynika, że przed przystąpieniem do rozpoznawania prawidłowa klasyfikacja elementów ciągu jest znana. Może ona być ustalana tak, jak np. w diagnostyce medycznej na podstawie kartotek szpitalnych pacjentów lub, jak w zadaniach językoznawczych - na podstawie reguł leksykalnych.

5.1.2.1 Klasyfikacja na podstawie oceny nauczyciela

W rozpatrywanym zadaniu numer klasy obiektu wyznacza wybór metody, według której badana osoba miałaby być nauczana w sposób najbardziej odpowiadający jej indywidualnym predyspozycjom. W projekcie modułu nauczającego (rozdz. 2.4) przewidziano cztery różne metody nauczania.

Dla obiektów wchodzących w skład ciągu uczącego oraz grupy kontrolnej do wyboru metody wykorzystano opinię osób prowadzących zajęcia z ankietowanymi studentami. Otrzymali oni opis poszczególnych metod i mieli za zadanie wskazać, na podstawie wyników nauczania oraz własnych obserwacji w czasie lekcji, optymalną (ich zdaniem) metodę uczenia. Ten, jak się wydawało, najbardziej naturalny sposób podejmowania decyzji okazał się być obarczony dużym błędem wynikającym głównie z różnic indywidualnych w interpretacji przedstawionych opisów. Wydaje się, że subiektywne rozumienie form aktywności w czasie lekcji, oraz równie subiektywna ocena zarówno stopnia trudności, jak i atrakcyjności poszczególnych metod w znacznej mierze zaburzyły otrzymany obraz. Porównanie wyników uzyskiwanych przez osoby z różnych grup ćwiczeniowych, a zatem od różnych nauczycieli potwierdziło te przypuszczenia. Nic też dziwnego, że rezultaty rozpoznawania dalekie były od zadowalających (por. tabela 5.6 rozdz. 5.4).

5.1.2.2 Indywidualna opinia ucznia

Interesującym wydaje się być również zestawienie opinii nauczycieli w kwestii doboru najodpowiedniejszej metody nauczania z indywidualną oceną samych zainteresowanych. Kwestia ta przypuszczalnie mogłaby być przedmiotem osobnych badań psychologicznych, ale to wykracza już poza ramy niniejszej pracy. W omawianym kontekście informacje pochodzące od uczniów a dotyczące najbardziej korzystnej dla nich metody uczenia się (w znaczeniu „bycia nauczonym”), miały charakter czysto porównawczy i nie zostały ujęte w formie danych dla modułu rozpoznającego.

5.1.2.3 Określanie przynależności obiektów poprzez analizę skupień

Wobec dużego udziału czynnika subiektywnego w określaniu optymalnej metody nauczania zdecydowano się wykorzystać dostępne narzędzia statystyczne aby sprawdzić, na ile w przestrzeni cech da się zaobserwować naturalne skupiska rozpatrywanych obiektów. Wykorzystano w tym celu metody analizy skupień. Zastosowano dwie metody dostępne w pakiecie **SPSS v.2.0 for Windows**:

❖ **metodę K-means**, w której zbiór obiektów dzielony jest na zadaną przez użytkownika ilość grup;

oraz

❖ **aglomeracyjną metodę hierarchicznego grupowania danych**, która analizuje ilość skupień, w jakie układają się obiekty i podaje numery obiektów należących do poszczególnych grup.

Otrzymany w wyniku analizy skupień podział zbioru obiektów różnił się dość znacznie od zaproponowanego przez uczących (por. Dodatek B), co w sposób wyraźny wpłynęło na poprawę wyników rozpoznawania. Porównanie uzyskiwanych rezultatów na przykładzie wariantu I przedstawia tabela 5.6 w rozdziale dotyczącym wyników rozpoznawania (5.4). W konsekwencji zdecydowano się przyjąć podział zbioru **D** otrzymany na podstawie analizy skupień za podstawę określania przynależności elementów ciągu uczącego, a w dalszej kolejności także dla grupy testowej.

Początkowo przyjęty ciąg uczący, zawierający 55 przykładów, okazał się za mało reprezentatywny, dlatego musiał on zostać rozszerzony. Uzyskano to poprzez przesunięcie części obiektów z grupy kontrolnej. Ostatecznie ciąg uczący liczył 78 obiektów. Pozostałe ze 108 obiektów, dla których zgromadzono kompletne dane, postanowiono wykorzystać właśnie w charakterze grupy kontrolnej, aby móc w ten sposób sprawdzić jakość dokonanej klasyfikacji. Liczebność grupy walidacyjnej wynosiła zatem 30 obiektów

5.1.2.4 Ilość klas obiektów

W metodzie K-means, która wymaga podania przewidywanej liczby skupień badano podział zbioru na cztery i na trzy grupy. Natomiast metoda aglomeracyjna hierarchicznego grupowania danych dała podział na trzy grupy. Porównując wcześniejsze przydziały dokonane przez prowadzących oraz wektory cech opisywanych obiektów zdefiniowano odwzorowanie f_{gm} , które pozwoliło przyporządkować numerom grup odpowiednie metody nauczania przewidziane w module nauczającym. Odwzorowanie f_{gm} można zdefiniować w sposób następujący:

$$f_{gm}: G \rightarrow L \Leftrightarrow f_{gm}(g_i) = m_i \wedge f_{gm}(g_i) = f_{gm}(g_k) \Rightarrow g_i = g_k \quad (5.2)$$

gdzie:

G – zbiór grup wyłonionych w procesie analizy skupień

L – zbiór klas (metod nauczania)

g_i, g_k – numery skupisk obiektów

m_i – i -ta metoda nauczania

W przypadku podziału na cztery grupy wymóg wzajemnej jednoznaczności odwzorowania f_{gm} jest spełniony w sposób automatyczny. Jednej grupie obiektów odpowiada dokładnie jedna metoda. Wystarczy zatem wskazać odpowiadające sobie numery i odwzorowanie jest zdefiniowane.

W aglomeracyjnej metodzie hierarchicznej zbiór obiektów został podzielony nie na cztery, a na trzy skupienia. Jakkolwiek opracowując strukturę projektowanego systemu nauczającego zaplanowano cztery metody nauczania odpowiadające czterem klasom obiektów, to gdy podczas analizy statystycznej wystąpiły przesłanki wskazujące na mniejszą liczbę grup obiektów, a także wobec trudności z uzyskaniem poprawnych rezultatów rozpoznawania postanowiono sprawdzić jak zachowa się algorytm rozpoznający przy mniejszej liczbie klas. Aby dostosować liczbę metod nauczania do ilości klas można wyeliminować jedną z planowanych wcześniej metod lub próbować połączyć dwie z już istniejących. W jednym i w drugim przypadku decyzja taka oznacza ograniczenie możliwości projektowanego systemu (mniejszą indywidualizację stylu nauczania), ale podjęto ją jako jedną z hipotez badawczych, uzależniając ostateczne rozstrzygnięcie od otrzymanych wyników (por. rozdz. 5.4). W niniejszej pracy zdecydowano się na połączenie **metody z przewodnikiem z metodą prezentacyjną**. Na takim właśnie wyborze zaważyło znaczne podobieństwo obu metod, których głównym wyróżnikiem jest forma prezentacji (w metodzie z przewodnikiem przeważa tekst, a w metodzie prezentacyjnej – obraz), ale obie przewidują raczej bierny odbiór treści przez ucznia i prowadzenie go „krok po kroku”. Uzyskano w ten sposób trzy grupy obiektów i trzy metody a zatem, podobnie jak poprzednio, należało tylko ustalić odpowiednie przyporządkowanie.

5.1.2.5 Przyporządkowanie numerów klas grupom obiektów

Uzgodniwszy ilość grup i ilość metod należy dokonać odpowiedniego przyporządkowania pomiędzy nimi. Jako wynik działania programu SPSS można otrzymać wykaz utworzonych skupień wraz z numerami obiektów do nich należących. Dane te po przekształceniu pozwalają przyporządkować każdemu obiektowi numer grupy, do której on należy. Łącząc numery obiektów z numerami ich klas oraz wektorami cech opisujących te obiekty otrzymuje się obraz danej grupy (por. Dodatek F). Analiza porównawcza tak otrzymanych zbiorów pozwala ustalić na ile wartości pojedynczych cech powtarzają się w obrębie grupy, lub innymi słowy, czy obiekty wykazują cechy wspólne. W przeciwnym bowiem razie dokonane uszeregowanie elementów nie miałoby praktycznego zastosowania.

Rzeczywiście dla niektórych cech można zaobserwować wyraźną tendencję powtarzania się tej samej wartości w obrębie całej grupy. I tak, w grupie D badani często wykazywali samodzielność w poszerzaniu swojej wiedzy (cecha nr 1), także z wykorzystaniem programów obcojęzycznych w telewizji (cecha nr 2) i Internetu (cecha nr 11) co najlepiej koresponduje z opisem hipertekstowej metody nauczania. Równie samodzielni uczniowie, którzy w dodatku wypracowali własny system uczenia się (grupa C) mogą próbować nauki metodą interakcyjną. W grupach A i B przeważały osoby bierne, mniej samodzielne, które wybierały studia nie pod wpływem własnych zainteresowań ale opinii innych osób (cecha nr 7), które jednak chętnie podjęłyby studiowanie z indywidualnym opiekunem (cecha nr 9) – dla nich przewidziano metodę z przewodnikiem i prezentacyjną. Głównym kryterium pozwalającym wybrać jedną z tych dwu metod były preferencje do odbioru tekstu, bądź obrazu i na ich podstawie dokonano ostatecznego wyboru. Dla trzech klas obiektów to ostatnie rozróżnienie nie było potrzebne.

Dokonane przyporządkowanie pozwoliło na określenie wartości funkcji przynależności dla obiektów ciągu uczącego i grupy testowej. Podana niżej tabela (5.1) dla trzech zestawów cech w kodzie 1-z-N przedstawia otrzymane wyniki przy podziale na trzy (K-M3) i na cztery klasy (KM-4).

Liczba klas	Kod prosty		Kod 1-z-N		Kod kumulowany	
	m. euklid.	m. uliczna	m. euklid.	m. uliczna	m. euklid.	m. uliczna
K-M 4	60.00%	60.00%	50.00%	50.00%	40.00%	40.00%
K-M 3	60.00%	66.67%	63.33%	63.33%	46.67%	46.67%

Tabela 5.1. Wyniki rozpoznawania przy podziale na trzy i na cztery grupy.

Jak wynika z powyższego zestawienia redukcja liczby klas przyniosła nieznaczną poprawę jakości rozpoznawania. Należy jednak pamiętać, że uzyskano ją za cenę zmniejszenia liczby metod nauczania, co oznacza zubożenie możliwości systemu nauczającego, który mając dobrze spełnić zadanie indywidualizacji sposobu nauczania powinien uwzględniać możliwie dużo różnych stylów uczenia się. Nie jest to więc, jak się wydaje, właściwy sposób usprawniania algorytmu rozpoznającego.

5.2 Algorytm rozpoznawania

Zadanie rozpoznawania utożsamiane jest w literaturze [67] z konstrukcją algorytmu **A**, który umożliwia podział pewnego zbioru obiektów **D** na rozłączne podzbiory (zwane dalej klasami) i określenie, do którego z tych podzbiorów należy niezany obiekt $d^k \in \mathbf{D}$. Ponieważ taki idealny podział zbioru **D** nie jest możliwy do zrealizowania w praktyce zazwyczaj przyjmuje się pewne przybliżenie algorytmu **A** oznaczane, dla odróżnienia $\hat{\mathbf{A}}$. Symbolicznie algorytm $\hat{\mathbf{A}}$ można przedstawić jako złożenie trzech odwzorowań zapisane w sposób następujący:

$$\hat{\mathbf{A}} = \mathbf{F} \bullet \mathbf{C} \bullet \mathbf{B} \quad (5.3)$$

czyli złożenie :

$$\blacklozenge \text{ funkcji recepcji } B: \mathbf{D} \rightarrow \mathbf{X} \quad (5.4)$$

$$\blacklozenge \text{ funkcji przynależności } C: \mathbf{X} \rightarrow \mathbf{R}^L \quad (5.5)$$

oraz

$$\blacklozenge \text{ funkcji decyzyjnych } F: \mathbf{R}^L \rightarrow \mathbf{I} \cup \{\mathbf{i}_0\} \quad (5.6)$$

gdzie:

$\mathbf{D}$ – zbiór obiektów;

$\mathbf{X}$ – przestrzeń cech;

$\mathbf{R}^L$ – wartości miary podobieństwa nieznanego obiektu do pozostałych obiektów zbioru $\mathbf{D}$

L – liczba klas

$\mathbf{I}$ – zbiór indeksów (numerów klas)

5.2.1 Odwzorowanie obiektów w przestrzeń cech – recepcja

Zadaniem funkcji recepcji, oznaczonej w poprzednim rozdziale symbolem B (wzór 5.4), jest przekształcenie badanych obiektów $d \in \mathbf{D}$ w punkty pewnej n -wymiarowej przestrzeni, zwanej przestrzenią cech i oznaczanej symbolem $\mathbf{X}$. Sposób definiowania tej przestrzeni, czyli wyboru odpowiednich cech opisujących obiekty $d \in \mathbf{D}$ został przedstawiony obszernie w rozdziale poprzednim (4). Omówiono tam aspekt psychologiczny konstrukcji przestrzeni cech (jakie cechy osobowości oraz czynniki związane z procesem edukacji mają wpływ na sposób uczenia się), a także związane z tym problemem kryteria obliczeniowe (wymiar przestrzeni $\mathbf{X}$, reprezentatywność zbioru cech, itd.)

5.2.2 Funkcje przynależności

Drugie z odwzorowań składowych algorytmu $\hat{A}$ (wzór 5.5) polega na wprowadzeniu miary podobieństwa badanych obiektów do innych, znanych już obiektów przestrzeni $\mathbf{D}$, dla których została ustalona wcześniej prawidłowa przynależność do odpowiednich klas. Od tego jak dokładnie można ustalić prawidłową przynależność obiektów wzorcowych zależy dokładność całego algorytmu $\hat{A}$. W praktyce brak jest jednoznacznych reguł postępowania i przyjęte rozwiązanie zależy od specyfiki obiektów będących przedmiotem rozpoznawania. Na przykład w diagnostyce medycznej, którą zalicza się do klasycznych obszarów zastosowań metod rozpoznawania obrazów, zazwyczaj podstawą do skonstruowania kryterium podziału przestrzeni $\mathbf{D}$ są informacje pochodzące z kartotek pacjentów poddawanych zabiegom chirurgicznym. W kartotekach zapisywane są objawy choroby, wyniki przeprowadzonych badań, postawiona diagnoza oraz rezultat dokonanego zabiegu. Przykładowo, w diagnostyce chorób trzustki prowadzonej z wykorzystaniem metod

rozpoznawania obrazów, opisywanej przez panią Annę Popielę-Mizerę w jej rozprawie doktorskiej [53], wykorzystano pochodzące z kartotek szpitalnych informacje na temat objawów towarzyszących potwierdzonemu przez zabieg chirurgiczny nowotworowi trzustki oraz - traktowanym alternatywnie innym schorzeniom tego narządu. Dane o tych pacjentach, zawierające zarówno zestaw objawów, jak i potwierdzoną chirurgicznie diagnozę umieszczono w tzw. **ciągu uczącym**, stanowiącym podstawę klasyfikacji nowych obiektów. Rola tego zbioru została omówiona w rozdziale 5.1, tutaj zasygnalizowano jedynie jego udział w tworzeniu algorytmu rozpoznawania. Sprawą zasadniczą dla skonstruowania funkcji przynależności jest, jak już powiedziano wcześniej, określenie miary podobieństwa między obiektami. W minimalnoodległościowych metodach rozpoznawania, a takie głównie stosowano w opisywanych badaniach, za miarę podobieństwa przyjmuje się funkcję odległości, spełniającą następujące aksjomaty:

1. $\forall_{x, y \in X} \rho(x, y) \geq 0 \quad \rho(x, y) = 0 \Leftrightarrow x = y$
2. $\forall_{x, y \in X} \rho(x, y) = \rho(y, x)$
3. $\forall_{x, y, z \in X} \rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Wybór odpowiedniej metryki zależy przede wszystkim od struktury przestrzeni cech. I to nie tylko od rodzaju wektorów bazowych wyznaczających przestrzeń, w szczególności od tego, czy są ortogonalne, czy nie, ale także od charakteru danych opisujących badane obiekty. Inne bowiem metryki stosuje się dla danych wyrażanych na skali liczbowej, a inne dla danych mierzonych na skali nominalnej. W omawianym przypadku dodatkową trudność stanowi mieszany charakter danych, czyli występowanie zarówno cech ilościowych, jak i jakościowych. W wyniku zastosowanej transformacji cech zostają one wyrażone w postaci liczb, które wprawdzie nie są miarą ilościową rozpatrywanych cech, ale niosą w sobie pewien element uporządkowania wynikający z przyjętej metody kodowania. Ten problem rzutuje w sposób istotny na wynik pracy algorytmu rozpoznającego dlatego jego rozwiązaniu poświęcono osobny rozdział (4.3).

Poniższe dwa podrozdziały zawierają krótki opis zastosowanych metryk z rozróżnieniem na dane o charakterze ilościowym oraz jakościowym. Zamieszczona w rozdziale 5.2.2.3. tabela ilustruje różnice w uzyskiwanych wynikach w zależności od przyjętej definicji odległości

5.2.2.1 Metryki w przestrzeniach liczbowych

Zanim jednak osiągnięto praktycznie użyteczny sposób kodowania cech podejmowano próby rozpoznawania w przestrzeni nie transformowanej, w której wartości poszczególnych cech były bezpośrednimi wynikami dokonanych pomiarów (por. rozdział 4.3).

Powszechnie stosowana metryka euklidesowa o postaci:

$$\rho(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5.7)$$

gdzie:

- $\underline{x}$ – wektor cech badanego obiektu
- $\underline{y}$ – wektor cech elementu ciągu uczącego
- x_i, y_i – i-ta składowa odpowiedniego wektora
- n – liczba cech (wymiar przestrzeni X)

lub w ogólniejszej postaci metryka Minkowskiego:

$$\rho(\underline{x}, \underline{y}) = \left[\sum_{i=1}^n |x_i - y_i|^p \right]^{\frac{1}{p}} \quad (5.8)$$

wymagają założenia o ortogonalności wektorów bazowych przestrzeni cech. W przypadku przestrzeni, której dotyczą opisywane badania nie ma praktycznie podstaw do przyjęcia takiego założenia. Tym niemniej, w zastosowaniach metod rozpoznawania obrazów wielokrotnie wykorzystuje się te metryki z dobrym skutkiem, choć często nie da się stwierdzić ortogonalnego charakteru wektorów bazy. Okazuje się bowiem, że przyjęte założenie, choć na pewno zniekształca nieco opisywaną rzeczywistość pozwala uzyskać całkiem dobre przybliżenia zależności pomiędzy badanymi obiektami.

Jeden z wariantów zastosowanego algorytmu najbliższego sąsiada wykorzystywał do obliczania odległości między obiektami następujące metryki:

- metrykę euklidesową daną wzorem (5.7)
- uogólnioną metrykę euklidesową:

$$\rho(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^n [\lambda_i (x_i - y_i)]^2} \quad (5.9)$$

gdzie współczynniki lambda obliczano według wzoru:

$$\lambda_i = \frac{1}{x_{i \max} - x_{i \min}} \quad (5.10)$$

$x_{i \max}$ – maksymalna wartość i-tej współrzędnej

$x_{i \min}$ – minimalna wartość i-tej współrzędnej;

- metrykę uliczną:

$$\rho(\underline{x}, \underline{y}) = \sum_{i=1}^n |x_i - y_i| \quad (5.11)$$

- ♦ uogólnioną metrykę uliczną

$$\rho(\underline{x}, \underline{y}) = \sum_{i=1}^n \lambda_i |x_i - y_i|$$

gdzie:

λ_i – określone jak wyżej;

- metrykę Czebyszewa:

$$\rho(\underline{x}, \underline{y}) = \max_i |x_i - y_i| \quad (5.13)$$

Próba zastosowania podanych metryk w algorytmie rozpoznającym w początkowym etapie badań nie wypadła zbyt pomyślnie dając poprawność nie przekraczającą 50% (tabela 5.2).

5.2.2.2 Metryki dla danych jakościowych

Wobec niskiej przydatności wyżej wymienionych metryk, przynajmniej w odniesieniu do danych surowych, zdecydowano się zastosować metryki zaprojektowane specjalnie dla cech o charakterze jakościowym. Posłużono się dwiema miarami odległości tego typu:

- statystyką Chi-kwadrat, wyrażoną wzorem:

$$\rho(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^n \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum_{i=1}^n \frac{(y_i - E(y_i))^2}{E(y_i)}} \quad (5.14)$$

gdzie: $E(\underline{x})$ – wartość oczekiwana zmiennej losowej $\underline{x}$;

oraz znormalizowaną statystyką Phi-kwadrat o postaci:

$$\rho(\underline{x}, \underline{y}) = \sqrt{\frac{\sum_{i=1}^n \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum_{i=1}^n \frac{(y_i - E(y_i))^2}{E(y_i)}}{n}} \quad (5.15)$$

5.2.2.3 Porównanie uzyskanych rezultatów rozpoznawania w zależności od przyjętej metryki

Poniższa tabela zawiera zestawienie początkowych wyników rozpoznawania dla pełnego zestawu cech przy wykorzystaniu różnych metryk.

Jak wynika z tabeli 5.2 zastosowanie metryk Chi-kwadrat i Phi-kwadrat nie przyniosło znaczącej poprawy sytuacji. Prowadzi to do wniosku, że w algorytmie rozpoznającym prawdopodobnie występuje czynnik zakłócający związany nie z metryką, a z konstrukcją przestrzeni cech i w tym obszarze należy poszukiwać sposobu na poprawę uzyskiwanych rezultatów.

		Wyniki rozpoznawania	
Lp.	Metryka	kod prosty	kod 1-z-N
1	Euklidesowa	36.67%	43.33%

2	euklid. uogólniona	40.19%	45.75%
3	Uliczna	38.30%	43.33%
4	uliczna uogólniona	34.53%	45.75%
5	Czebyszewa	30.19%	30.00%
6	Chi-kwadrat	44.54%	47.53%
7	Phi-kwadrat	44.53%	47.53%

Tabela 5.2 Wyniki rozpoznawania dla różnych metryk przy pełnym zestawie cech (przynależność przykładów ustalana na podstawie opinii uczących).

5.2.3 Funkcje decyzyjne

W opisie algorytmu rozpoznawania funkcje decyzyjne zostały przedstawione za pomocą wzoru:

$$F: \mathbf{R}^L \rightarrow \mathbf{I}$$

Stanowią one trzeci z elementów składowych algorytmu rozpoznawania ($\hat{\mathbf{A}}$). Nazwa ich wynika z roli, jaką pełnią w algorytmie. Służą one bowiem klasyfikacji badanych obiektów, czyli podejmowaniu decyzji o przynależności tych obiektów do odpowiednich klas (podzbiorów) zbioru D . Ogólnie stosowanym kryterium decyzyjnym jest reguła majoryzacyjna miary podobieństwa (opisanej w rozdziale 5.2.2.). Oznacza ona, iż na przykład gdy przyjętą miarą podobieństwa jest funkcja odległości za obiekty należące do tej samej klasy przyjmuje się te, które znajdują się (w sensie odległości) najbliżej siebie.

Niestety w praktycznych zastosowaniach metod rozpoznawania obrazów stosunkowo często występują trudności z jednoznacznym zaklasyfikowaniem badanego obiektu do jednej konkretnej klasy. Stosuje się wówczas na oznaczenie takiej decyzji indeks i_0 oznaczający brak rozpoznania (*nie umiem sklasyfikować*).

Rzadziej, chociaż zdarzają się takie przypadki, uwzględnia się możliwość rozpoznań niejednoznacznych, które oznaczają decyzję: obiekt reprezentowany przez wektor cech $\underline{x}^k$ należy do klasy i_s lub klasy i_t , przy czym $t \neq s$.

5.3 Wykorzystane metody rozpoznawania

Niestety problem doboru metody nauczania do właściwości psychologicznych człowieka nie ma swoich wzorców na gruncie teorii rozpoznawania obrazów i prezentowane tutaj badania można zaliczyć do nielicznych w tej dziedzinie. Dlatego podjęte próby doboru właściwej metody rozpoznawania oraz odpowiedniej metryki należy traktować jedynie jako punkt wyjścia do bardziej szczegółowych i systematycznych badań, które ze względu na znaczenie praktyczne problemu, powinny, jak się wydaje, być rozwijane. Próby te rozpoczęto od zastosowania metod

minimalnoodległościowych. Za ich wykorzystaniem przemawiają następujące argumenty:

- brak empirycznych wskazań co do większej przydatności innych metod;
- względna prostota i intuicyjna zrozumiałość;
- dobre rezultaty zastosowań w dziedzinach o podobnym charakterze danych, zwłaszcza w diagnostyce medycznej (wysoki udział czynnika subiektywnego), oraz w analizie procesów ekonomicznych (dużo danych jakościowych)

5.3.1 Metody minimalnoodległościowe

Nawa tych metod pochodzi od sposobu definiowania funkcji przynależności. Sprowadza się on bowiem do obliczenia odległości nieznanego obiektu, będącego przedmiotem rozpoznawania od wszystkich elementów ciągu uczącego C_n . Indeks klasy, do której należy najbliższy z nich (w sensie przyjętej metryki) wskazuje numer klasy nowego obiektu. Jeżeli ρ oznacza przyjętą metrykę, to wartości funkcji przynależności oblicza się według wzoru:

$$C^i(\underline{x}) = \frac{1}{\rho(\underline{x}, \underline{x}^{i,k}) + \varepsilon} \quad \text{dla } i = 1, \dots, L \quad (5.16)$$

gdzie:

$\underline{x}^{i,k}$ – element ciągu uczącego należący do i-tej klasy
 ε - mała stała dodatnia wprowadzona w celu zapewnienia wartości $C^i(\underline{x}) < \infty$ na wypadek gdyby $\rho(\underline{x}, \underline{x}^{i,k}) = 0$.

Sposób podejmowania decyzji na podstawie tak wyliczonych funkcji przynależności decyduje o różnych wariantach metod minimalnoodległościowych.

5.3.1.1 Metoda najbliższego sąsiada (NN)

Jeżeli decyzję o przynależności badanego obiektu, opisanego przez wektor cech $\underline{x}$, do klasy i^k podejmuje się na podstawie indeksu najbliższego elementu ciągu uczącego to jest to najprostsza z metod minimalnoodległościowych nazywana algorytmem **najbliższego sąsiada** lub, w skrócie, **NN** od angielskiej nazwy *nearest neighbour*. Wybór elementu $\underline{x}^{i,k}$ we wzorze (5.16) odbywa się zgodnie z regułą:

$$\rho(\underline{x}, \underline{x}^{i,k}) = \min_{t \in U^i} \rho(\underline{x}, \underline{x}^t) \quad (5.17)$$

Ze względu na dostępność podstawowych algorytmów minimalnoodległościowych w literaturze (np. [67]) nie będą one tutaj szczegółowo przedstawiane.

Jeżeli nie wprowadzi się żadnych dodatkowych warunków, to w przypadku wystąpienia dwóch elementów ciągu uczącego jednakowo odległych od badanego obiektu indeks klasy pierwszego z nich będzie numerem klasy, do której zostanie zakwalifikowany nowy obiekt.

Tabela 5.3 przedstawia wyniki rozpoznawania z wykorzystaniem algorytmu NN dla pełnego zestawu cech oraz dla poszczególnych wariantów przy różnych rodzajach kodowania i zastosowanych metrykach: euklidesowej oraz ulicznej. Przynależność przykładów ustalono na podstawie analizy skupień.

	Pełny zestaw cech		Wariant I		Wariant II		Wariant III	
Kod	euklid.	uliczna	euklid.	uliczna	euklid.	uliczna	euklid.	uliczna
Prosty	60.00%	66.67%	86.67%	83.33%	66.67%	63.33%	80.00%	76.67%
1-z-N	50.00%	50.00%	80.00%	80.00%	66.67%	60.00%	76.67%	70.00%
Kumulowany	40.00%	40.00%	80.00%	76.67%	56.67%	56.67%	63.33%	63.33%

Tabela 5.3. Wyniki rozpoznawania dla algorytmu NN przy różnych rodzajach kodowania.

Otrzymywane rezultaty zostaną poddane szczegółowej analizie na końcu rozdziału, tutaj warto tylko zauważyć, że wyraźnie zaznacza się poprawa jakości rozpoznawania po zredukowaniu ilości składowych wektora $\underline{x}$ (por. rozdz. 4.4).

5.3.1.2 Metoda α -najbliższych sąsiadów (α NN)

Z praktycznych zastosowań metod minimalnoodległościowych wynika, że pewną poprawę jakości rozpoznawania można uzyskać wprowadzając uporządkowanie elementów ciągu uczącego zgodnie z regułą:

jeżeli t, s są numerami elementów zbioru C_u to

$$\forall_{t,s} \left| \frac{1}{s} - \frac{1}{t} \right| > s \Rightarrow \rho(\underline{x}, \underline{x}^t) > \rho(\underline{x}, \underline{x}^s) \quad (5.18)$$

i rozpatrując tylko podzbiór C_u^α zbioru C_u zawierający α najbliższych sąsiadów. Parametr α zwykle wybiera się arbitralnie, jednak w ten sposób aby:

$$\alpha \leq \min_{i \in I} N^i \quad (5.19)$$

gdzie

I – zbiór indeksów ($I = 1, 2, \dots, L$)

N^i – ilość elementów C_u należących do i -tej klasy

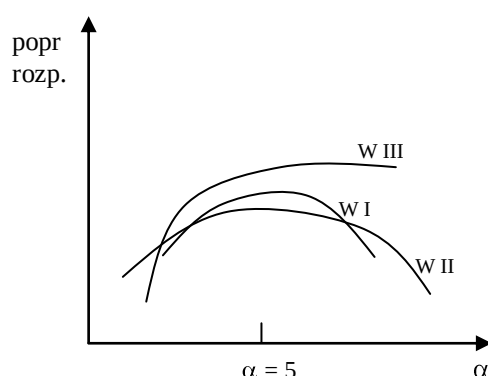
Wybrany podzbiór C_u^∞ dzieli się z kolei na podzbiory odpowiadające poszczególnym klasom. O wartości funkcji przynależności decyduje liczebność odpowiednich podzbiorów, przy czym jako klasę nieznanego obiektu przyjmuje się tę, która jest najliczniej reprezentowana.

Wyniki rozpoznawania przy pomocy algorytmu α NN dla różnych wartości parametru α zawiera tabela 5.4.

α	Kod prosty			Kod 1-z-N		
	Wariant I	Wariant II	Wariant III	Wariant I	Wariant II	Wariant III
2	86.67%	73.33%	70.00%	60.00%	50.00%	63.33%
3	90.33%	73.33%	86.67%	76.67%	66.67%	53.33%
5	96.67%	76.67%	86.67%	80.00%	70.00%	66.67%
7	93.33%	70.00%	90.00%	83.33%	73.33%	76.67%
9	86.67%	66.67%	90.00%	86.67%	70.00%	83.33%

Tabela 5.4. Wyniki rozpoznawania dla różnych wartości α .

Z przedstawionego zestawienia widać, że tylko do pewnego momentu poprawność rozpoznawania zwiększa się ze wzrostem α . Przy czym, dla kodu prostego maksimum poprawnych rozpoznań osiąga się w przybliżeniu dla $\alpha=5$, a w kodzie 1-z-N dla $\alpha=7$ lub 9 w zależności od zestawu cech. Sytuację tę ilustruje wykres na rysunku 5.1.



Rys. 5.1. Wykres zależności poprawności rozpoznawania od wartości parametru α (kod prosty).

5.3.1.3 Metoda j_N -tego najbliższego sąsiada (j_N NN)

Kolejną modyfikacją algorytmu NN, również wykorzystującą uporządkowanie zbioru C_u (zgodnie z regułą 5.18) jest metoda j_N -tego najbliższego sąsiada (j_N NN). Badanemu obiektowi zostaje przyporządkowany numer tej klasy, do której należy j_N -ty w kolejności element uporządkowanego zbioru C_u . Parametr j_N , podobnie jak α jest niewielką liczbą dodatnią.

W tabeli 5.5 zebrano wyniki rozpoznawania za pomocą algorytmu j_N NN. Wynika z niej wyraźnie, że jakość rozpoznawania przy pomocy algorytmu j_N NN jest niższa niż

w przypadku metody NN w wersji podstawowej jak i w postaci α NN. Kolejna obserwacja dotyczy zestawu cech. Otóż, we wszystkich prezentowanych wyżej metodach (tabele 5.3-5.5) najmniejsza ilość poprawnych rozwiązań występuje w przypadku opisu obiektów za pomocą cech z zestawu W II, nieco lepsze efekty uzyskuje się przy wariancie W III, a najlepsze przy zestawie W I.

J_N	Kod prosty			Kod 1-z-N		
	Wariant I	Wariant II	Wariant III	Wariant I	Wariant II	Wariant III
1	83.33%	73.33%	70.00%	60.00%	50.00%	63.33%
3	86.67%	66.67%	76.67%	66.67%	66.67%	43.33%
5	80.00%	53.33%	70.00%	73.33%	63.67%	63.33%
7	73.33%	50.00%	63.33%	63.33%	50.00%	56.67%
9	66.67%	43.33%	56.67%	53.33%	43.33%	50.00%

Tabela 5.5. Wyniki rozpoznawania z wykorzystaniem algorytmu j_N NN.

W trakcie analizy uzyskanych wyników zwrócono uwagę na następujący szczegół. Otóż, algorytm **najbliższego sąsiada** podejmuje decyzję o przynależności nieznanego obiektu, przyjmując za jej wartość numer tej klasy, do której należy najbliższy, w sensie przyjętej metryki, element ciągu uczącego. W sytuacji, gdy badany obiekt jest jednakowo odległy od kilku elementów C_u , należących do różnych klas, jako wartość funkcji przynależności zostanie przyjęty numer tej klasy, do której należy pierwszy (w kolejności występowania w ciągu uczącym) z równoodległych elementów przykładowych. Jeżeli wziąć pod uwagę, iż kolejność ta jest zazwyczaj czysto przypadkowa, nie trudno sobie wyobrazić, do jakich zniekształceń w poprawności rozpoznawania może to doprowadzić. W praktyce rozpoznawania stosuje się często zabieg polegający na zwiększeniu liczebności zbioru klas rozszerzając go o tak zwany element neutralny, oznaczający brak decyzji (*nie umiem rozpoznać*). Zapisuje się go jako jednoelementowy zbiór $\{i_0\}$. Można też przyjąć rozwiązanie dopuszczające wybór więcej niż jednej klasy (np. 1 i 3), który w przypadku opisywanych badań odpowiadałby wskazaniu nie jednej a dwu możliwych metod nauczania. Rozwiązanie to, aczkolwiek możliwe, nie będzie jednak tutaj szerzej dyskutowane gdyż wymagałoby ono wprowadzenia dodatkowych reguł decyzyjnych dopuszczających kombinację jednych, a wykluczających połączenie innych metod nauczania. Nie jest np. możliwe równoczesne dopuszczenie metod: interakcyjnej i z przewodnikiem, gdyż pierwsza z nich zakłada dużą samodzielność i aktywność ucznia, a druga - zgodnie z nazwą – pracę tylko pod kierunkiem „przewodnika”. Zresztą przy obecnej liczbie metod zostały one tak dobrane, aby różnice między nimi były istotne. Alternatywą mogłoby być dopuszczenie metod „mieszanych”, ale pomysł ten, w chwili obecnej, należy raczej potraktować jako sugestię dotyczącą rozwoju systemu w momencie, kiedy będzie możliwa ocena funkcjonowania już zaimplementowanych metod.

5.3.2 Metoda najbliższej mody (NM) jako przykład metody wzorców

Idea tej metody jest intuicyjnie prosta i wydaje się świadczyć o jej większej użyteczności w sytuacji, gdy trudno jest o jednoznaczne dopasowanie nieznanego obiektu do któregoś z przykładów zapamiętanych w ciągu uczenia. W metodzie NM porównuje się badane obiekty nie z kolejnymi elementami zbioru C_u , ale z reprezentantami poszczególnych klas sprawdzając, czy obiekt ten należy do otoczenia któregoś z nich. Metoda ta daje bardzo korzystne rezultaty gdy obiekty w przestrzeni niejako w sposób naturalny układają się w określone skupiska, gdyż wówczas procent poprawnych rozpoznań jest stosunkowo wysoki. Przy dużym przemieszaniu cech w przestrzeni metoda okazała się zupełnie nieprzydatna (por. Dodatek A, tabela. A.V).

Podstawową trudność przy próbach stosowania tej metody stanowi kwestia wskazania obiektów modalnych dla poszczególnych klas. Brak rekomendowanych w literaturze reguł postępowania dla tego typu przypadku problem ten dodatkowo pogłębia. Stosowane często określanie cech nowego obiektu jako średnich arytmetycznych cech obiektów należących do tej samej klasy:

$$m_j^i = \frac{1}{N_i} \sum_{k=1}^{N_i} x_j^{i,k} \quad (5.20)$$

gdzie:

$i = 1, \dots, L$ – numer klasy

$j = 1, \dots, n$ – numer cechy

okazało się być, w zastosowaniu do opisywanych badań, rozwiązaniem zupełnie nieadekwatnym. W sytuacji, gdy o różnicach między obiektami decydują czasem wartości pojedynczych cech, próby uśredniania dają efekty zgoła odwrotne niż oczekiwano.

5.3.3 Możliwość zastosowań innych metod rozpoznawania

Wobec niepowodzeń związanych z zastosowaniem metody najbliższej mody (NM), a wynikających głównie z charakteru przestrzeni cech zrezygnowano także z prób dotyczących innych metod, również wykorzystujących wzorce.

Rozważano możliwość wykorzystania metod probabilistycznych. Te jednak wymagają bądź posiadania pewnych informacji dotyczących rozkładu prawdopodobieństwa analizowanych danych, bądź przynajmniej możliwości oszacowania parametrów takiego rozkładu. W omawianej sytuacji uznano, iż na obecnym etapie badań rodzaj uzyskanych

danych nie daje podstaw do uogólnień pewnych prawidłowości na całą populację uczniów w stopniu uprawniającym do oszacowania choćby przybliżonych parametrów rozkładu. Zwłaszcza zaobserwowane przy zastosowaniu algorytmu NM duże przemieszanie obiektów w przestrzeni nakazywało daleko idącą ostrożność. Wydaje się, że podjęcie próby rozpoznawania przy pomocy metod probabilistycznych ma sens dopiero po przeprowadzeniu badań uzupełniających, które dostarczą bardziej szczegółowych informacji na temat rozkładu badanych predyspozycji w populacji uczniowskiej, zwłaszcza potrzebnych we wzorze Bayesa warunkowych rozkładów gęstości prawdopodobieństwa [37].

5.4 Wyniki rozpoznawania

Jak w każdym zadaniu rozpoznawania obrazów podstawową trudność stanowi właściwe określenie przestrzeni cech. Jest to zarazem ta część zadania, która za każdym razem wymaga indywidualnego rozwiązania, zależnie od specyfiki rozwiązywanego problemu. Ewentualne błędy popełnione podczas definiowania przestrzeni cech są główną przyczyną niepowodzenia całego procesu.

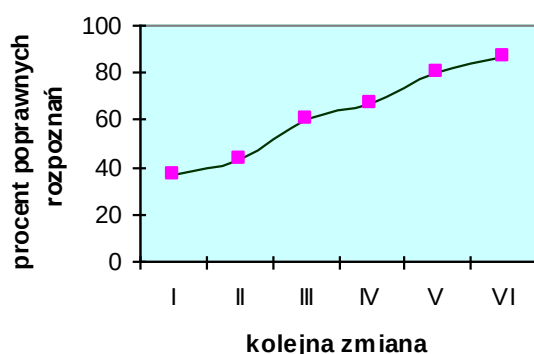
Celem badań opisywanych w niniejszej pracy było wykazanie, że dzięki wykorzystaniu metod rozpoznawania obrazów jest możliwe określenie i pomiar predyspozycji psychologicznych ucznia tak, aby otrzymane dane posłużyły jako parametry sterujące inteligentnym systemem komputerowego wspomagania nauczania, który potrafiłby na ich podstawie wybrać najlepszą dla danego ucznia metodę uczenia.

Jakkolwiek współczesna psychologia eksperymentalna chętnie odwołuje się do metod wypracowanych na gruncie matematyki ([14],[32]), a zwłaszcza statystyki ([2],[6],[20],[24]), sięgając do coraz bardziej wyrafinowanych technik, takich jak: wielozmiennowa analiza wariancji [9], analiza ścieżek [10] czy analiza skupień z elementami teorii podzbiorów rozmytych [48], to próby wykorzystania metod rozpoznawania obrazów do badania cech osobowości ucznia, jak dotąd, w literaturze nie były opisywane. W ostatnich latach prowadzone są na Politechnice Lubelskiej pod kierunkiem dr inż. Marka A. Jakubowskiego (prezentowane podczas VIII Ogólnopolskiego Sympozjum „Techniki komputerowe w przekazie edukacyjnym”, Kraków 1998) prace wykorzystujące sieci neuronowe w połączeniu z teorią zbiorów rozmytych do badania zjawiska „wypalenia zawodowego nauczycieli” oraz do badania operacji myślowych młodzieży. Są to jednak, jak przyznają sami autorzy, dopiero wstępne próby wykorzystania przez nich metod sztucznej inteligencji w zakresie badań pedagogicznych.

Problematyka opisu cech osobowości człowieka jest sama w sobie na tyle złożona (por. rozdział 3), iż trudno oczekiwać jednoznacznych rozwiązań w tej dziedzinie. Nic więc dziwnego, że i w postawionym zadaniu rozpoznawania ta właśnie część badań sprawiła najwięcej trudności. Świadczy o tym nie tylko ilość modyfikacji jakie należało wprowadzić od momentu wstępnego zdefiniowania zbioru cech opisujących badane obiekty, ale i zmieniające się rezultaty rozpoznawania. Poniższy

wykres (rys.5.2) ilustruje wpływ wprowadzanych zmian w zakresie struktury przestrzeni cech na ilość poprawnych rozpoznań. Wyróżniono trzy zasadnicze etapy zmian. Dotyczyły one:

1. Sposobu opisu elementów przestrzeni X (kodowanie);
2. Zmniejszenia wymiaru tej przestrzeni;
3. Metody klasyfikacji przykładów (elementów ciągu uczącego)



zmiana	
nr	typ
I	stan początkowy
II	kod 1-z-N
III	an. skupień
IV	I etap redukcji
V	wariant I
VI	metoda (α NN)

Rys. 5.2. Wykres zależności wyników rozpoznawania od kolejno wprowadzanych zmian.

Każdy z wymienionych czynników miał pewien wpływ na jakość rozpoznawania, jakkolwiek wydaje się, że najistotniejszy był sposób określania przynależności elementów ciągu uczącego (tabela 5.6). Stosowany pierwotnie wybór klas obiektów, subiektywny z natury, dodatkowo zaburzało różne rozumienie proponowanych metod nauczania przez uczących. Jednak wymóg reprezentatywności próby narzucał przeprowadzenie badań wśród studentów różnych kierunków, a co za tym idzie, uniemożliwiał zebranie opinii tylko jednej osoby. Natomiast zastosowanie metod statystycznych pozwoliło na ujednolicenie procedury klasyfikacyjnej, co zaowocowało radykalną poprawą jakości rozpoznawania.

Metryka	Kod prosty			Kod 1-z-N		
	Euklid.	Uliczna	Czebysz.	Euklid.	uliczna	Czebysz.
K-M 4	96.67%	93.33%	83.33%	80.00%	80.00%	40.00%
U 4	46.67%	36.67%	40.00%	40.00%	40.00%	30.00%

Metryka	Kod kumulowany		
	Euklid.	Uliczna	Czebysz.
K-M 4	80.00%	76.67%	76.67%
U 4	73.33%	76.67%	76.67%

Tabela 5.6. Porównanie wyników rozpoznawania dla przynależności określonych metodą K-Means (K-M 4) oraz na podstawie opinii uczących (U4). Symbol 4 oznacza liczbę klas obiektów.

Jeżeli chodzi o sposób reprezentacji danych pewną różnicę, aczkolwiek mniejszą niż oczekiwano, przyniosło wykorzystanie kodu typu 1-z-N, choć jednoznaczna interpretacja wpływu tego czynnika nie jest prosta, gdyż jest on różny w zależności od rozpatrywanego zestawu cech (por. tabela 5.6).

Dość duży wpływ na jakość rozpoznawania miała ilość mierzonych cech. Przede wszystkim już samo wyeliminowanie niektórych z nich (por. rozdział 4.3) wprowadziło wyraźną poprawę (tabela 5.7), a w połączeniu z pozostałymi (omówionymi wcześniej) zmianami umożliwiło pomyślną realizację zadania.

	Pełny zestaw cech	I etap redukcji (31 cech)	II etap redukcji		
			Wariant I	Wariant II	Wariant III
m. euklidesowa	60.00%	66.67%	86.67%	66.67%	80.00%
m. uliczna	66.67%	70.33%	83.33%	63.33%	76.67%
m. Czebyszewa	46.67%	53.33%	66.67%	63.33%	56.67%

Tabela 5.7. Zmiany jakości rozpoznawania wynikające z redukcji wymiaru przestrzeni cech.

Jeżeli chodzi o rodzaj wykorzystanych metryk, należy podkreślić, że jakkolwiek stosowano także metryki przewidziane specjalnie dla danych jakościowych (metryka Chi-kwadrat i Phi-kwadrat – rozdz. 5.2.2.2), to otrzymywane rezultaty były porównywalne z tymi dla metryki euklidesowej i ulicznej. Metryka Czebyszewa w większości przypadków dawała gorsze wyniki (tabela 5.2).

W zakresie stosowanych metod rozpoznawania były to głównie metody minimalnoodległościowe, bazujące na algorytmie najbliższego sąsiada. Różnice pomiędzy wynikami uzyskanymi za pomocą algorytmów NN, α NN oraz j_N NN wynosiły od kilku do kilkunastu procent (tabela 5.8)

metryka $\Rightarrow$	euklid.	uliczna	Czebysz.	euklid.	uliczna	Czebysz.
metoda rozpozn.	Wariant I			Wariant II		
NN	86.67%	83.33%	66.67%	66.67 %	63.33 %	53.33 %
αNN, $\alpha=5$	96.67%	90.00%	76.67%	76.67%	73.33%	60.00%
j_N NN, $j=3$	86.67%	73.33%	66.67%	66.67%	66.67%	43.33%

metryka $\Rightarrow$	euklid.	uliczna	Czebysz.
metoda rozpozn.	Wariant III		
NN	80.00%	76.67 %	56.67 %
αNN, $\alpha=7$	90.00%	86.67%	63.33%
j_N NN, $j=3$	76.67%	73.33%	56.67%

Tabela 5.8. Porównanie uzyskanych wyników dla różnych metod minimalnoodległościowych

Reasumując można stwierdzić, że w wyniku zastosowania metod rozpoznawania obrazów otrzymano podział przestrzeni cech, opisanej przez predyspozycje psychologiczne uczniów związane z uczeniem się, na 4 klasy odpowiadające czterem zaproponowanym metodom nauczania, a jakoś klasyfikacji analizowanych obiektów upoważnia do kontynuacji badań opartych na tak uzyskanych danych. Należy bowiem podkreślić, że o ile celem niniejszej pracy było zbadanie czy możliwa jest klasyfikacja obiektów przestrzeni cech, opisywanej przez predyspozycje psychologiczne uczniów, przy pomocy metod rozpoznawania obrazów, to celem głównym podjętych badań jest budowa komputerowego systemu nauczającego, który wykorzystując otrzymane informacje dotyczące optymalnej metody uczenia się mógłby właściwy proces nauczania prowadzić.

W dalszych badaniach proponuje się przyjąć za podstawę klasyfikacji elementów ciągu uczącego wyniki grupowania obiektów na podstawie analizy skupień, jako pozbawioną wpływu czynników subiektywnych wynikających np. z różnej interpretacji przez uczących podanych kryteriów oceniania. Natomiast w kwestii sposobu pomiaru i rodzaju cech opisujących obiekty przestrzeni **D**, sugeruje się konieczność powołania zespołu ekspertów, traktując przedstawione wyżej rezultaty jako zachętę do podjęcia tej ciekawej tematyki. Dlatego też ostateczna decyzja co do sposobu opisu badanych obiektów, tak w sensie ilości mierzonych cech, jak i systemu ich kodowania powinna być uzależniona od wyników prac tego zespołu, który mógłby poczynione tutaj obserwacje uzupełnić i rozszerzyć.

6 Implementacja modułów systemu

Jak wspomniano we wstępie pracy, docelowo, opracowywany system komputerowy będzie się składał z trzech części: gromadzenia danych o uczniu, rozpoznawania i klasyfikacji zakończonej wyborem najbardziej odpowiedniej metody uczenia oraz właściwego nauczania. Ostatni z wymienionych elementów nie jest przedmiotem badań w niniejszej pracy, został on tylko naszkicowany w celach poglądowych (por. rozdział 2.4) i nie będzie szczegółowo omawiany.

6.1 Implementacja algorytmu rozpoznawania

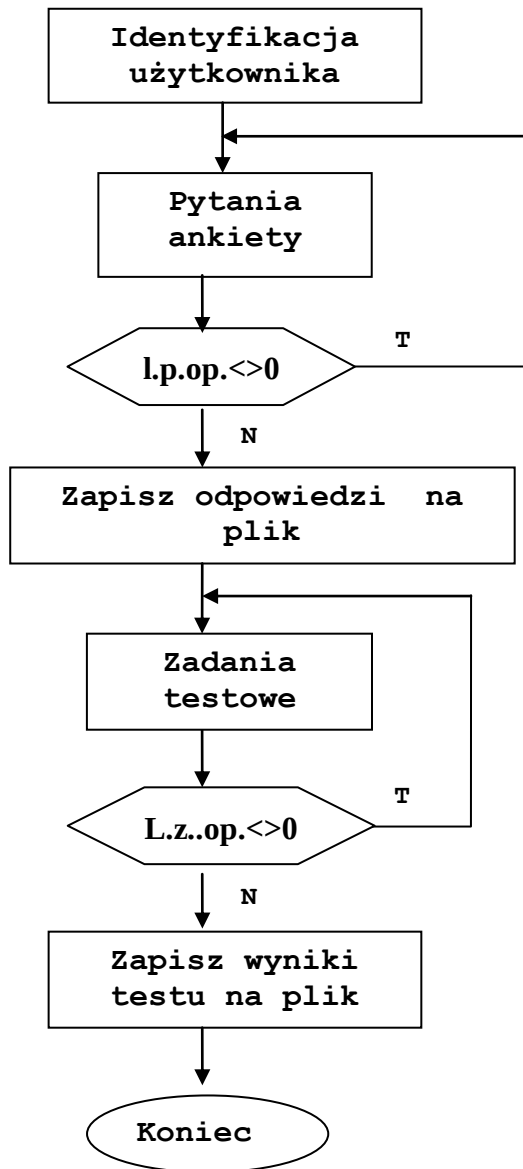
Drugi moduł rozważanego systemu (por. rys. 2.1 i 2.2), będący podstawowym przedmiotem badań, został przedstawiony w rozdziale poprzednim (rozdział 5). Od strony implementacji komputerowej jest to po prostu algorytm najbliższego sąsiada (nearest neighbour) w wersji NN, α NN i j_N NN, zapisany w języku Pascal. Dane wejściowe dla algorytmu stanowi wektor cech badanego obiektu, zapisany jako jednowymiarowa tablica liczb rzeczywistych o długości odpowiednio 11, 15 lub 13 składowych, w zależności od przyjętego wariantu (por. 4.4.3). Sposób pomiaru i zapisu poszczególnych cech został omówiony w rozdziałach 4.1 i 4.2, natomiast kryteria wyboru cech zawiera rozdział 4.4. Informacją pamiętaną w systemie jest ciąg uczący (oznaczany C_u (5.1)) zawierający charakterystyki obiektów, których przynależność do odpowiednich klas jest znana. W Pascalu jest to dwuwymiarowa tablica o wymiarach $l_{cu} \times l_{cech}+1$, gdzie l_{cu} – liczba elementów w ciągu uczącym, l_{cech} – to liczba składowych wektora opisujących każdy element ciągu. Zatem wierszami macierzy C_u są poszczególne obiekty (o znanej przynależności), natomiast kolumny (oprócz ostatniej) tworzą wektory zmiennych losowych odpowiadających wartościom poszczególnych cech na całym zbiorze uczącym. Kolumna ostatnia zawiera numery klas poszczególnych obiektów ciągu uczącego (por. rozdział 5.1.3.). Istota algorytmu najbliższego sąsiada sprowadza się obliczania odległości nieznanego (badanego) obiektu od wszystkich elementów ciągu uczącego i zaklasyfikowania go do tej klasy, do której należy najbliższy mu (w sensie odległości) element ciągu uczącego. Analizę algorytmów rozpoznawania wraz z dyskusją dotyczącą wyboru zastosowanej metody zawiera rozdział 5.

6.2 Realizacja modułu gromadzenia wiedzy o uczniu

Od strony realizacji komputerowej najbardziej pracochłonnym i czasochłonnym było przygotowanie modułu gromadzenia danych o uczniu. W rozdziale 4.1 przedstawiono sposób opracowania zestawu badanych cech, niniejszy punkt dotyczy realizacji procedur w Pascalu służących temu zadaniu.

6.2.1 Budowa modułu

Moduł gromadzenia danych ma budowę sekwencyjną, co oznacza, że istotna jest kolejność poszczególnych etapów. Jej schemat przedstawia rysunek (6.1) Proces rozpoczyna się zidentyfikowaniem użytkownika połączonym ze sprawdzeniem, czy w systemie zawarte są już dane na jego temat Całość kończy się zapisem na plik zgromadzonych danych.



Rys 6.1. Schemat budowy modułu gromadzenia danych.

6.2.2 Identyfikacja użytkownika

W wersji testowej każdy uczeń przystępując do pracy z systemem najpierw wprowadza kod nadany mu przez osobę prowadzącą badania. W wersji użytkowej, w momencie instalacji programu będzie tworzona baza informacji o użytkownikach. Każdorazowe rozpoczęcie pracy będzie się wiązało z wprowadzeniem kodu użytkownika. Odpowiednia procedura w programie sprawdzi, czy w systemie pamiętane są dane na temat tego użytkownika. Pozwoli to z jednej strony uniknąć przesłaniania informacji przy przypadkowym powtórzeniu kodu (dość powszechnie stosowanym sposobem identyfikacji jest np. podanie imienia, a te, jak wiadomo, często się powtarzają), z drugiej zaś zwalnia osobę, która kontynuuje naukę od konieczności powtarzania od początku całej procedury testującej. Choć nie wyklucza się powtórzenia fragmentu testu, badającego np. posiadane aktualnie wiadomości na temat będący treścią nauczania celem wskazania właściwego punktu startowego procedury nauczającej. Postać kodu musi spełniać wymagania takie jakie stawia system DOS wobec nazw plików bez rozszerzenia, czyli zawiera nie więcej niż 8 liter lub cyfr bez odstępów. Informację na temat dopuszczalnej postaci kodu otrzymuje użytkownik wyświetloną na ekranie bezpośrednio po uruchomieniu systemu. Wymóg taki wynika z faktu, iż wszystkie dane o uczniu pamiętane są w pliku tekstowym, którego unikatowa nazwa tworzona jest właśnie w momencie podania kodu. Standardowym rozszerzeniem plików z danymi o uczniach jest **odp** (plik odpowiedzi) dla odróżnienia od innych plików danych.

W przyszłości przewiduje się również uzupełnianie informacji wejściowych „spostrzeżeniami” czynionymi przez system w trakcie nauki danej osoby przy pomocy tego programu. Oznacza to wzbogacenie interfejsu wejściowego o elementy samouczące i samokorygujące (por. rozdział 2.1 rys.2.2).

6.2.3 Ankieta

W wersji testowej każdy badany odpowiada na 74 pytania. Ankieta (Dodatek D) ma charakter zamkniętego testu wyboru, co oznacza, że należy wskazać jedną (i tylko jedną) z podanych odpowiedzi. Ograniczenie to narzucono świadomie ze względu na to, aby nie komplikować i tak już złożonej procedury analizy danych. Nietrudno sobie wyobrazić jak różne mogłyby być odpowiedzi dotyczące osobistych upodobań i preferencji gdyby nie podano jednolitego sposobu ich opisu. Zatem każdy plik dyskowy o nazwie:

Identyfikator_użytkownika.odp

jest zbiorem numerów odpowiedzi wskazanych przez ucznia. Plik ten pamiętany jest w oddzielnym katalogu danych, na dysku wskazanym w trakcie instalacji systemu. Pytania ankiety są tak skonstruowane aby pozwalały uzyskać informacje dotyczące wszystkich podanych w rozdziale czwartym grup czynników. Sposób interpretacji poszczególnych cech został omówiony w tym właśnie rozdziale.

Można odpowiadać na pytania praktycznie w dowolnej kolejności. Rysunek 6.2 przedstawia wygląd ekranu z pytaniem przed udzieleniem odpowiedzi.

Przycisk  pozwala przejść do pytania następnego bez udzielenia odpowiedzi na pytanie bieżące.

Po wybraniu odpowiedzi dostępne są następujące opcje (rys. 6.3) :

-  - umożliwia przejście do listy pytań opuszczonych z możliwością wyboru jednego z nich (rys.6.4)
-  - cofa zaznaczenie wybranej odpowiedzi i umożliwia wybór innej odpowiedzi
-  - wyświetla krótką informację dotyczącą dostępnych opcji
-  - przejście do następnego pytania

Należy podkreślić, że wymienione cztery przyciski są widoczne dopiero po wybraniu odpowiedzi. Nie ma zatem możliwości przejścia do następnego pytania za pomocą przycisku „**Dalej**” bez udzielenia odpowiedzi. Opcja taka jest dostępna jedynie przez przycisk „**Opuść**”, co gwarantuje pełną kontrolę nad pominiętymi pytaniami.

Przed zakończeniem ankiety system sprawdza, czy któreś z pytań pozostało bez odpowiedzi i jeżeli tak, wyświetla kolejno wszystkie opuszczone pytania oczekując uzupełnienia wyboru. Przejście do części testowej następuje dopiero wówczas, gdy uczeń odpowiedział na wszystkie pytania ankiety. W ten sposób zostało zagwarantowane, że wszystkim cechom zostanie nadana niezerowa wartość.

Pytanie 34:

Czy w trakcie nauki szkolnej wypracowałeś własny system uczenia się ?

Odpowiedzi:

A. tak

B. nie

C. nie wiem

D. nie odczuwałem takiej potrzeby

Wskaż myszką wybraną odpowiedź

Jezeli nie mozesz się zdecydować naciśnij klawisz OPUŚĆ ➡ i wróć do tego pytania później

Opuść

Rys. 6.3. Ekran pytania ankiety przed wybraniem odpowiedzi

Pytanie 34:

Czy w trakcie nauki szkolnej wypracowałeś własny system uczenia się ?

Odpowiedzi:

A. tak

B. nie

C. nie wiem

D. nie odczuwałem takiej potrzeby

Idź do...

Anuluj

Pomoc

Dalej

Rys. 6.4. Ekran pytania wraz z opcjami dostępnymi po wskazaniu odpowiedzi

6.2.4 Test umiejętności studiowania

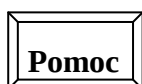
Dodatkowe źródło informacji o uczniu stanowi dołączony **test umiejętności studiowania** znany także w psychologii pod nazwą **test rozumienia tekstu**. Ma on również postać testu zamkniętego, ale pytania dotyczą treści zawartych w tekście prezentowanym na ekranie. Tutaj również, podobnie jak w ankiecie, oprócz treści pytania i wariantów odpowiedzi wyświetlane są cztery przyciski:



- pozwala wrócić do trybu wyświetlania tekstu, aby ponownie go przeczytać



- cofa zaznaczenie wybranej odpowiedzi i umożliwia zmianę decyzji



- wyświetla krótką informację dotyczącą dostępnych opcji



- przejście do następnego pytania

Przejście do następnego pytania bez wskazania odpowiedzi jest również możliwe, ale podobnie jak w ankiecie, do zakończenia testu potrzebne są wszystkie odpowiedzi. Dlatego też przed przejściem do następnego zadania na ekranie pojawia się informacja o pominiętych pytaniach i ich treść wyświetlana jest ponownie. Równocześnie z treścią zadania wyświetlanych jest kilka pasków wskaźników.

Są to:

Procent wykonanych zadań (w obrębie jednego testu)

Ilość powrotów (w celu ponownego odczytania prezentowanego tekstu)

Ilość poprawek (gdy anulowano już wybraną odpowiedź po to by wskazać inną)

Czas wykonania (łączny czas zużyty na zapoznanie się z tekstem i udzielenie odpowiedzi)

Stopień zaciemnienia paska pokazuje aktualny stan wskaźnika. Po zakończeniu testu obliczane są wartości końcowe odpowiadające: ilości zadań wykonanych poprawnie, łącznej liczbie poprawek, ilości powrotów do tekstu oraz łącznemu czasowi trwania wszystkich zadań. Informacje te są następnie dołączane jako kolejne składowe wektora cech opisujących badane obiekty (por. rozdz. 4.1).

Istotą tego testu jest sprawdzenie umiejętności odnajdywania w tekście potrzebnych informacji. Uczeń ma bowiem odpowiedzieć na pytania nie na podstawie wiadomości, które być może posiada już na dany temat, lecz dokładnie na podstawie tekstu wyświetlanego na ekranie. Podstawową trudność przy konstrukcji tego typu testów stanowi konieczność takiego sformułowania pytań, aby odpowiedź poprawna

wymuszała zrozumienie tekstu, a równocześnie nie kolidowała z ewentualną wcześniejszą wiedzą ucznia na dany temat. Zaletą komputerowej wersji testu umiejętności jest możliwość oszacowania, na ile dla danej osoby forma przekazywania treści za pomocą ekranu komputera jest pomocą, a na ile utrudnieniem w przyswajaniu wiedzy. Wynikami tego testu są właśnie wymienione wyżej parametry czyli: ilość powrotów, ilość poprawek, czas wykonania oraz procent odpowiedzi poprawnych. Należy przy tym podkreślić, że ich znaczenie analizowane jest kontekstowo, to znaczy, np.:

większość odpowiedzi poprawnych, mała ilość poprawek (lub ich brak), ale długi czas wykonania i wielokrotny powrót do tekstu;

lub:

krótki czas wykonania, brak powrotów i brak poprawek ale mało odpowiedzi poprawnych,

dużo poprawnych odpowiedzi, test wykonany szybko i bez pomyłek;

Każdemu z tych zestawów można nadać odpowiednią interpretację psychologiczną i z pewnością stanowiłaby ona ciekawy przyczynek do analizy porównawczej badanych cech, ale równocześnie wykraczałaby poza ramy niniejszej pracy. Tutaj cechy te były rozpatrywane w znaczeniu: „staranny, samodzielny, pracuje szybko i dokładnie” lub „szybki, mało dokładny, niezbyt zainteresowany” itd.

Na marginesie prac związanych z przygotowaniem tej właśnie części systemu nasuwa się wniosek o celowości nawiązania szerszej współpracy z fachowcami z innych dziedzin (psychologia, socjologia, dydaktycy poszczególnych przedmiotów) w celu opracowania dalszych narzędzi diagnostycznych, które umożliwiłyby bardziej wnikliwą analizę badanych cech, czy wręcz zmodyfikowanie istniejącego zestawu.

6.2.5 Wymagania sprzętowe

Opracowane dotychczas dwie części systemu nie narzucają żadnych szczególnych wymagań sprzętowych, co oznacza, że mogą być uruchamiane praktycznie na dowolnym komputerze poczynając od 386, pracującym w trybie MS DOS. Dopóki nie ma modułu nauczania nie jest nawet wymagany dysk twardy, gdyż zarówno moduł gromadzenia danych, jak i moduł rozpoznający mieszczą się na jednej dyskietce. Również na tej samej dyskietce zostaną zapisane dane o uczniu w postaci wyników ankiety i testu.

Inaczej wygląda sprawa w przypadku modułu nauczającego, gdyż dla wykorzystania możliwości grafiki animowanej, sekwencji video czy też efektów dźwiękowych konieczna jest obecność karty grafiki wysokiej rozdzielczości (SVGA), karty dźwiękowej oraz kolorowego monitora, co zresztą w kontekście parametrów współczesnych komputerów nie wydaje się być żądaniem wygórowanym. Również,

ze względu na rozmiary plików graficznych i dźwiękowych, niezbędny będzie czytnik CD-ROM gdyż praktycznie niemożliwe jest umieszczenie obszernej bazy wiedzy na jednej, czy nawet kilku dyskietkach. Ze względu na metodę hipertekstową wskazany byłby również dostęp do sieci Internet.

7 Prezentacja metod nauczania na przykładzie twierdzenia Pitagorasa

Moduł nauczający, jak to już wielokrotnie podkreślano, znajduje się na razie w fazie projektu wstępnego. Jego realizacja wymaga nie tylko wykorzystania zaawansowanych technik sztucznej inteligencji (systemy ekspertowe, sieci neuronowe), ale przede wszystkim aktywnej współpracy z nauczycielami dyscyplin, dla których odpowiednie bazy wiedzy mają się znaleźć w systemie.

Pierwsze próby współpracy podjęto w zakresie nauczania matematyki. Przedstawiony poniżej przykład dotyczący twierdzenia Pitagorasa⁴ i jego zastosowań może służyć jako ilustracja idei projektowanych metod nauczania.

W rozdziałach 2.4.4 i 2.4.5. została omówiona ilość postulowanych metod nauczania wraz z krótkim uzasadnieniem dokonanego wyboru. Przewiduje się zastosowanie na początek czterech metod nauczania, różniących się nie tylko sposobem prezentacji wiedzy (tekstowy lub graficzny), ale przede wszystkim preferencjami, jakie powinien wykazywać uczeń, aby wybrana dla niego przez system metoda nauczania przyniosła jak najlepsze efekty. Projektowane metody nazwano: hipertekstową, z przewodnikiem, prezentacyjną oraz interakcyjną. Zostaną one teraz krótko omówione w odniesieniu do konkretnego przykładu, którym jest twierdzenie Pitagorasa

7.1 Idea i założenia realizacji

Podstawowym założeniem jest, aby każda z projektowanych metod nauczania pozwalała uzyskać pewne, ustalone z góry, minimum wiadomości związanych z prezentowanym tematem i aby dawała pewien pogląd na możliwości praktycznych zastosowań omawianych zagadnień. Natomiast różnice decydujące o specyfice poszczególnych metod nauczania pozwolą na poszerzenie poznawanych wiadomości, lub na dokładniejsze ich poznanie.

W organizacji bazy wiedzy postuluje się zachowanie podziału nauczanych treści na jednostki nazywane lekcjami, ale w znaczeniu nieco szerszym niż stosowane w szkole. Otóż, przez lekcję rozumie się podstawowy zestaw wiadomości dotyczących jednego zagadnienia. W prezentowanym przykładzie twierdzenie Pitagorasa stanowi właśnie treść jednej lekcji. Mieści się w niej: prezentacja twierdzenia, dowód oraz podstawowe przykłady zastosowań. Kolejność i stopień szczegółowości lekcji zależy od konkretnej metody nauczania. Podstawowym warunkiem funkcjonalności systemu jest istnienie „zbiorów” lekcji, które można dowolnie ze sobą łączyć za pomocą tzw. hiperlinków. Celowo użyto tutaj słowa zbiór, a nie np. zestaw, aby nie sugerować, że ma to być zamknięta całość przypominająca kilka rozdziałów podręcznika. Wręcz przeciwnie, im

⁴ Przykład ten opracowany we współpracy z mgr Katarzyną Wójcik, był prezentowany na I ogólnopolskiej konferencji IPN, Kraków 97.

więcej ciekawych powiązań między lekcjami uda się utworzyć, im większą swobodę uzyska uczeń w poruszaniu się po bazie wiedzy, tym większa szansa, że zainteresuje się realizowanym tematem i będzie chciał w nim aktywnie uczestniczyć.

7.2 Propozycja realizacji tematu w zależności od metody nauczania

.W przypadku twierdzenia Pitagorasa zakłada się, że każdy uczeń po zakończeniu lekcji posiada następujące minimum wiedzy i umiejętności:

- ◆ zna treść twierdzenia
- ◆ rozumie jego dowód
- ◆ potrafi rozwiązać proste zadanie oparte na wykorzystaniu twierdzenia

Natomiast poszczególne metody oferują różną kolejność realizacji zakładanego minimum oraz różne formy poszerzenia i pogłębienia tematu.

7.2.1 Metoda hipertekstowa

Zakłada ona przede wszystkim dużą niezależność, a zarazem dyscyplinę ucznia. W konsekwencji oczekuje się, że potrafi on samodzielnie rozwiązywać stawiane mu zadania, czy to podejmując próby dowodzenia twierdzeń, bądź przynajmniej uzupełniania naszkicowanych dowodów, czy też znajdując rozwiązania ciekawych problemów obliczeniowych opartych na wykorzystaniu poznanych właśnie wiadomości. Niezależność przejawia się także w chęci własnych poszukiwań, które są możliwe dzięki systemowi połączeń hipertekstowych zarówno lokalnych (w obrębie penetrowanej właśnie bazy wiedzy) jak i rozproszonych, z wykorzystaniem np. podanych adresów internetowych⁵. Krótki scenariusz lekcji w oparciu o tę metodę można przedstawić następująco⁶:

Jednym z największych matematyków i filozofów starożytności był Pitagoras⁷ z Krotony. Wśród wielu problemów jakimi się zajmował sformułował on i udowodnił twierdzenie przedstawiające związki między długościami boków trójkąta prostokątnego.

⁵ Na uwagę zasługuje np. inicjatywa pod hasłem Internet dla Szkół (<http://www.ids.pl>), w której nauczyciele prezentują m.in. swoje propozycje scenariuszy lekcji z wykorzystaniem zasobów internetowych, czy też oferowany przez WSiP serwis edukacyjny **Matematyka 2001**, będący rozwinięciem serii książkowej i uzupełnieniem wersji elektronicznej wydanej na krążkach CD pod tym samym tytułem (<http://www.wsip.com.pl>)

⁶ Uwaga! Treść lekcji dla odróżnienia od opisu poszczególnych metod nauczania będzie podawana czcionką pochyłą.

⁷ Podkreślenie słowa wskazuje, że jest ono połączeniem z innymi obszarami bazy wiedzy, w tym wypadku można pod tym hasłem dowiedzieć się więcej o samym Pitagorasie i jego osiągnięciach.

Zaobserwowaną przez niego prawidłowość można sformułować następująco:

*Jeżeli trójkąt jest prostokątny,
to pole kwadratu zbudowanego na przeciwprostokątnej tego
trójkąta jest równe sumie pól kwadratów zbudowanych na jego
przyprostokątnych.*

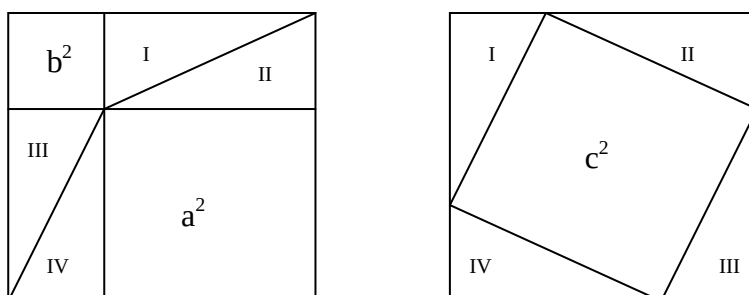
Twierdzenie to można zapisać symbolicznie:

Z: a, b – długości przyprostokątnych w trójkącie prostokątnym
 c – długość przeciwprostokątnej

T: $c^2 = a^2 + b^2$

Twierdzenie Pitagorasa udowodniono na około 200 sposobów. Czy potrafisz sam je udowodnić? W razie potrzeby skorzystaj ze wskazówki.

Wskazówką może być ilustracja geometryczna dowodu przypisywanego samemu Pitagorasowi postaci:



Aby umożliwić uczniowi lepsze zrozumienie przedstawionych zależności należy zlecić mu rozwiązanie kilku zadań dotyczących np. przekątnych kwadratu i prostokąta, wysokości w trójkącie równobocznym itp. W metodzie hipertekstowej, podobnie jak w interakcyjnej dopuszcza się możliwość wyboru, czy uczeń chce sam rozwiązywać zadania, czy z pomocą systemu.

Rozszerzeniem lekcji jest propozycja wyboru innych pokrewnych tematów. Oto klika przykładowych możliwości:

Inne dowody twierdzenia Pitagorasa – przykłady innych dowodów, kompletnych bądź w formie praktycznych wskazówek;

Trójkaty pitagorejskie – po zapoznaniu się z definicją trójkątów pitagorejskich uczeń szuka odpowiednich przykładów oraz próbuje uogólnić swoje spostrzeżenia;

Pitagoras w Internecie - tu uczeń otrzymuje listę adresów w sieci Internet, pod którymi może znaleźć ciekawe informacje związane z prezentowanym twierdzeniem⁸;

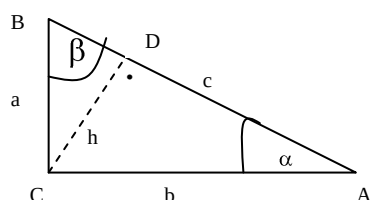
Ciekawe zadania – niebanalne zadania, np. konstrukcje liczb niewymiernych;

7.2.2 Metoda z przewodnikiem

W metodzie z przewodnikiem każdy element lekcji wprowadzany jest wraz z niezbędnymi objaśnieniami. Nie przewiduje się zbytniego rozszerzania podstawowych wiadomości, główny nacisk kładąc na dokładne zrozumienie tak samego twierdzenia, jak i jego praktycznych zastosowań.

Ze względu na preferencje słowne uczniów, dla których ta metoda jest przeznaczona, rysunek pełni w niej tylko rolę pomocniczą, natomiast wszystkie elementy lekcji opisane są słowami. Z tego samego względu przyjęto algebraiczną postać zarówno twierdzenia, jak i jego dowodu, przy czym analiza występujących związków algebraicznych prowadzi do zaobserwowania istniejących prawidłowości, które dopiero na końcu zostają ujęte w formę twierdzenia.

Zarys scenariusza dla metody z przewodnikiem można przedstawić w następujących krokach:



1. Na rysunku przedstawiono trójkąt prostokątny ABC. Wysokość h dzieli ten trójkąt na dwa trójkąty, które również są prostokątne. Zauważ, że można teraz wskazać na rysunku trzy pary trójkątów podobnych.

2. Uzupełnij następujące równości:

$$\begin{array}{l} | \angle BCA | = | \angle CDA | = 90^\circ \\ | \angle CAB | = | \dots\dots\dots | = \alpha \end{array}$$

⁸Np.: <http://router.staszic.waw.pl/~wslaw> – 15 różnych dowodów twierdzenia Pitagorasa opracowanych na podstawie książki S.Jeleńskiego pt. *Śladami Pitagorasa*.

$$|\angle CBA| = |\angle \dots\dots\dots| = \beta$$

3. Zatem pary trójkątów podobnych to:

$$\begin{aligned} \triangle ACB &\sim \triangle \dots\dots\dots && (\text{wspólny kąt } \beta) \\ \triangle ACB &\sim \triangle \dots\dots\dots && (\text{wspólny kąt } \alpha) \\ \triangle \dots\dots\dots &\sim \end{aligned}$$

4. Z podobieństwa trójkątów ACB i BDC otrzymujemy

$$\frac{a}{c} = \frac{|DB|}{a} \quad \text{czyli:} \quad a^2 = \dots\dots\dots \quad (1)$$

a z podobieństwa trójkątów ACB i CDA wynika, że:

$$\frac{b}{c} = \frac{|DA|}{b} \quad \text{czyli} \quad b^2 = \dots\dots\dots \quad (2)$$

Dodając stronami (1) i (2) otrzymamy następującą równość:

$$a^2 + b^2 = \dots\dots\dots$$

Jeżeli teraz zauważysz, że $|DB| + |DA| = c$, to powyższe równanie przyjmie postać:

$$a^2 + b^2 = c^2$$

5. Udowodniliśmy twierdzenie Pitagorasa, które brzmi:

W trójkącie prostokątnym suma kwadratów przyprostokątnych jest równa kwadratowi przeciwprostokątnej.

6. Twierdzenie to można zapisać inaczej:

Z: a, b - długości przyprostokątnych
 c - długość przeciwprostokątnej

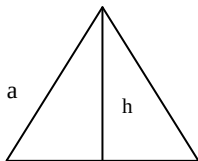
T: $a^2 + b^2 = c^2$

7. Twierdzenie Pitagorasa ma bardzo wiele praktycznych zastosowań. Uważa się, że prawidłowość ta była znana i stosowana już w starożytności (np. przy budowie egipskich piramid), a Pitagoras był pierwszym, który sformułował ją w postaci twierdzenia.

Rozważ następujące zadanie:

Oblicz długość wysokości w trójkącie równobocznym o boku a .

- narysuj podany trójkąt równoboczny



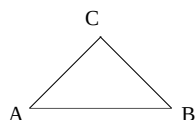
- w narysowanym trójkącie wskaż trójkąty prostokątne.
- zaznacz długości ich boków.
- zapisz odpowiednie związki wynikające z twierdzenia Pitagorasa.
- z podanej równości oblicz h (nie zapomnij o tym, że długość wysokości jest liczbą dodatnią).

Proponowana lekcja może (i powinna) zawierać znacznie więcej przykładów oraz zadań do wykonania. Podany szkic ma tylko odzwierciedlać ideę samej metody, a nie jej szczegółowy scenariusz.

7.2.3 Metoda prezentacyjna

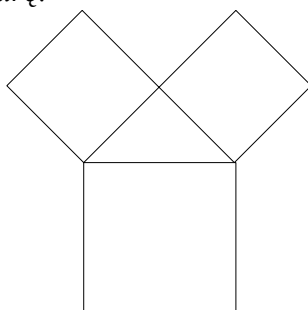
W metodzie tej zasadnicze znaczenie ma obraz, dlatego też lekcja zaczyna się od przedstawienia na ekranie odpowiedniej konstrukcji. Poszczególne elementy rysunku pojawiają się w takiej kolejności aby ilustrowały przebieg dowodu. Może to następować na przykład zgodnie z poniższym schematem:

1.

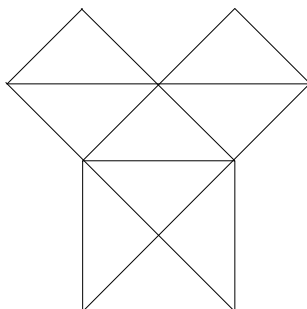


Przedstawiony na rysunku trójkąt ABC jest trójkątem prostokątnym równoramiennym.

2. *Jeżeli na każdym z boków tego trójkąta zbudujemy kwadrat to otrzymamy następującą figurę:*



3. Gdy teraz narysujesz przekątne w tych kwadratach, to co możesz powiedzieć o polach otrzymanych trójkątów ?

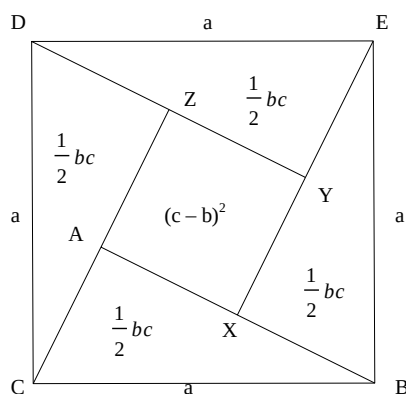


4. Czy potrafisz obliczyć pola kwadratów zbudowanych na poszczególnych bokach trójkąta ABC ?

Jeżeli symbol a oznacza bok kwadratu zbudowanego na przeciwprostokątnej, a symbol b oznacza bok kwadratu zbudowanego na przyprostokątnej trójkąta ABC to jak można wyrazić występujący między nimi związek?

5. Trójkąt ABC był trójkątem równoramiennym, jak myślisz czy równość taka zachodzi także dla innych trójkątów prostokątnych ?

Przyjrzyj się kolejnej ilustracji i spróbuj wyrazić pole kwadratu BCDE jako sumę pól figur, które w nim zaznaczono. Czy potrafisz powiedzieć jaka jest długość boków trójkąta ABC ?



6. Prezentowana własność trójkątów prostokątnych została zauważona już w starożytności. Grecki matematyk Pitagoras sformułował twierdzenie, które mówi, że:

Jeżeli trójkąt jest prostokątny to pole kwadratu zbudowanego na przeciwprostokątnej jest równe sumie pól kwadratów zbudowanych na przyprostokątnych.

W dalszej części lekcji wystąpią przykłady obliczeń i konstrukcji opartych na wykorzystaniu twierdzenia Pitagorasa. Powinny one być tak dobrane, aby analizowane zależności były łatwo widoczne na dołączonych ilustracjach.

7.2.4 Metoda interakcyjna

Metoda ta zakłada, podobnie jak metoda hipertekstowa, dużą aktywność ucznia. Przewidziana jest dla osób, które podejmują próby poszukiwania własnych dowodów bądź rozwiązań, chętnie poszerzają swoją wiedzę i chcą samodzielnie dotrzeć do przyczyn badanych zjawisk oraz wynikających zeń konsekwencji. Szczegóły realizacji zależą oczywiście od nauczanych treści. W omawianym przykładzie proponuje się najpierw rozwiązanie zadania, a następnie uogólnienie zaobserwowanych zależności w postaci twierdzenia. Udowodnienie ich można także w całości pozostawić uczniowi, lub, w sytuacji gdy zachodzi uzasadniona obawa, że nie poradzi sobie z zadaniem, zlecić mu dokończenie rozpoczętego dowodu, czy też uzupełnienie opisu słownego na podstawie ilustracji graficznej. Z kolei uczniom, którzy dla lepszego zrozumienia treści potrzebują opisu słownego, można polecić wykonanie ilustracji graficznej zapisanego dowodu.

Poniżej podano zarys scenariusza lekcji realizowanej metodą interakcyjną.

1. Rozwiąż zadanie:

Dany jest trójkąt prostokątny o przyprostokątnych a i b . Oblicz pole kwadratu zbudowanego na przeciwprostokątnej.

Wpisz
odpowiedź _

Wskazówka*

**Wskazówka może mieć postać tekstową lub graficzną. Powinna zawierać propozycję jednego z dowodów twierdzenia Pitagorasa.*

2. Uzyskana zależność nosi nazwę **twierdzenia Pitagorasa**. Mówi ono, że:

Jeżeli trójkąt jest prostokątny to pole kwadratu zbudowanego na przeciwprostokątnej jest równe sumie pól kwadratów zbudowanych na przyprostokątnych.

3. Zapisz to twierdzenie w sposób symboliczny.

Wpisz
odpowiedź _

Wskazówka*

**Wskazówka: a) zwróć uwagę na niezbędne założenia
b) jaka jest teza twierdzenia?*

4. Czy prawdziwe jest twierdzenie odwrotne ?

Wpisz
odповідź _

Wskazówka*

**Wskazówka zawiera informację co to jest twierdzenie odwrotne.*

5. *Ale jeszcze ciekawsze od samego twierdzenia są jego wielorakie zastosowania. Poniżej podano kilka z nich. Korzystając z twierdzenia Pitagorasa można obliczyć na przykład:*

- ♦ *długość przekątnej kwadratu, prostokąta, sześcianu,*
- ♦ *wysokość trójkąta równobocznego,*
- ♦ *promień koła opisanego na trójkącie i wpisanego w trójkąt,*
- ♦ *długość odcinka mając dane współrzędne jego końców,*

Każdy z elementów tej grupy można rozwinąć na podstawie przykładów i zadań do samodzielnego rozwiązania.

Kolejną formą rozszerzenia prezentowanych treści jest propozycja:

6. *Wybierz interesujące Cię tematy:*

Inne dowody
tego twierdzenia.

Uczeń otrzymuje fragmenty różnych dowodów, często tylko rysunki lub wskazówki. Jego zadaniem jest uzupełnienie dowodu.

Trójkąty
pitagorejskie

Po zapoznaniu się z definicją trójkątów pitagorejskich, uczeń szuka odpowiednich przykładów oraz powinien uogólnić swoje spostrzeżenia.

Ciekawe
zadania

Niebanalne zadania, łamigłówki związane z twierdzeniem Pitagorasa.

Zastosowania

Do czego może przydać się w praktyce znajomość tego twierdzenia.

7.3 Podsumowanie przykładu i komentarz

Przykład twierdzenia Pitagorasa został wybrany ze względu na wielość praktycznych jego zastosowań, różnorodność dowodów oraz możliwość interpretacji geometrycznej, co zapewnia bogaty wybór zarówno treści nauczania jak i form ich przekazu.

Nauczyciel może, znając możliwości ucznia sprawdzić, czy osiągnął on oczekiwany poziom wiedzy i umiejętności w zakresie prezentowanego tematu. Może zlecić mu samodzielnie rozwiązanie zadania lub przeprowadzenie dowodu twierdzenia. Uczeń, który korzystał z metody hipertekstowej może w czasie kolejnej lekcji zaprezentować w klasie wiadomości, które znalazł w Internecie, a ten który pracował z pomocą metody interaktywnej może przedstawić własną propozycję dowodu.

Można oczekiwać, że zastosowanie komputerowego systemu nauczającego uwzględniającego indywidualne predyspozycje ucznia nie tylko pozwoli uatrakcyjnić sam proces uczenia się, ale także zaktywizuje uczniów zdolnych, którzy prezentując wiadomości uzyskane w wyniku własnych poszukiwań będą bardziej zainteresowani przebiegiem lekcji, gdyż sami mogą ją współtworzyć. Równocześnie uczniowie słabsi nie muszą się stresować tym, że nie nadążają za innymi kolegami, gdyż system komputerowy umożliwi im pracę w tempie dostosowanym do ich własnych możliwości.

8 Zakończenie

Rola i zakres wykorzystania komputerów w nauczaniu powoli, ale systematycznie wzrasta. Podstawowe znaczenie mają w tym względzie programy dydaktyczne właściwie dopasowane do potrzeb i zainteresowań uczniów. Również intensywnie rozwijająca się ogólnoswiatowa sieć Internet jest coraz częściej postrzegana jako przydatne narzędzie edukacyjne. Także u źródeł badań opisywanych w niniejszej pracy była chęć wyjścia naprzeciw narastającym tendencjom poprzez opracowanie idei inteligentnego systemu komputerowego wspomagającego nauczanie.

8.1 Podsumowanie

Celem pracy było wykazanie, że

Wykorzystanie technik sztucznej inteligencji, a w szczególności wybranych metod rozpoznawania obrazów (pattern recognition) umożliwia uzyskanie informacji dotyczących cech osobowości ucznia, które zastosowane jako parametry sterujące multimedialnego systemu komputerowego wspomagania nauczania, pozwolą na dostosowanie sposobu przekazu wiedzy (lub: procesu uczenia) do indywidualnych predyspozycji psychologicznych osoby nauczanej.

Próba udowodnienia powyższej tezy wymagała rozwiązania następujących zadań szczegółowych:

1. Określenia ilości i rodzaju badanych cech, z ewentualnym wskazaniem, które z nich mają większe, a które mniejsze znaczenie z punktu widzenia postawionego problemu, jakim był dobór metody nauczania do indywidualnych predyspozycji psychologicznych ucznia.
2. Skonstruowanie odpowiednich narzędzi komputerowego pomiaru badanych cech.
3. Wykorzystanie opracowanych narzędzi do zgromadzenia danych o uczniach, tak aby mogły one być wykorzystane jako informacje wejściowe dla modułu rozpoznawania.
4. Sprawdzenie, czy możliwe jest wykorzystanie metod rozpoznawania obrazów do klasyfikacji uczniów opisywanych przez zdefiniowany zestaw cech, co pozwoliłoby na indywidualizację procesu nauczania tych uczniów w oparciu o multimedialny system komputerowy.

Pierwsze z postawionych zadań było równocześnie najtrudniejsze. Jak zwykle w problemach rozpoznawania, właściwe zdefiniowanie przestrzeni cech jest sprawą kluczową. W omawianym przypadku dodatkowym utrudnieniem był mieszany charakter

danych, z których tylko niewielka część była danymi mierzalnymi, a zdecydowana większość – danymi symbolicznymi czy wręcz opisowymi (por. Dodatek D). Podstawowa trudność polegała jednak na zdefiniowaniu samego zestawu cech. Bowiem, choć psychologowie dostrzegli już użyteczność komputera w badaniu pojedynczych zdolności i umiejętności ([42],[45],[46]) to, jak dotychczas, brak jest opracowań dotyczących wykorzystania komputera do opisu i pomiaru cech osobowości ucznia związanych z uczeniem się. Jeżeli nawet pojawiają się opracowania nawiązujące do tego problemu ([41],[72]) to głównie dotyczą one diagnostyki dokonywanej pod kątem wyboru zawodu, a nie procesu zdobywania wiedzy. Pozornie duża liczba publikacji nawiązujących do zagadnień nauczania wspomaganego komputerowo dotyczy głównie oceny przydatności komputera w sprawdzaniu nabytych umiejętności rzadko natomiast przedstawia komputer jako narzędzie diagnozowania sposobu przyswajania wiedzy i czynników warunkujących efektywność tego procesu.

Szczegółowe omówienie realizacji zadania pierwszego zamieszczono w rozdziale trzecim. Jako rozwiązanie otrzymano zestaw cech połączony z propozycją ich opisu i sposobu kodowania.

Po zdefiniowaniu cech, które należy mierzyć kolejnym zadaniem było opracowanie narzędzi pomiaru wskazanych parametrów. To znalazło swoje odzwierciedlenie w rozdziale szóstym dotyczącym implementacji modułów systemu, w tym modułu gromadzenia danych. Efektem realizacji zadania drugiego jest zestaw 74 pytań **ankiety** oraz **test umiejętności studiowania**. Podstawą do opracowania pytań ankietowych były zarówno źródła literaturowe ([5],[10],[1],[11],[12],[28]), jak i bezpośrednie konsultacje z psychologami pracującymi w Poradni Wychowawczo-Zawodowej nr 1 w Krakowie, którzy na co dzień zajmują się diagnostyką zdolności uczniów. W dodatku C zamieszczono zestaw pytań, które psycholog w poradni zwykle zadaje bądź to bezpośrednio badanemu dziecku, bądź jego rodzicom w celu ustalenia przyczyn ewentualnych trudności związanych z uczeniem się, a także w celu stwierdzenia, czy uczeń rzeczywiście wykazuje wybitne uzdolnienia w zakresie jednego, bądź grupy przedmiotów.

Deklarując wykonanie postawionego zadania należy wyraźnie podkreślić, że opisywane badania stanowią tylko jedną z możliwych prób rozwiązania problemu pomiaru cech osobowości związanych z uczeniem się. Jak to już stwierdzono w trakcie opisu opracowywanych narzędzi pomiaru wskazanych cech osobowości ucznia (por. rozdz.4.1), wydaje się być celowym powołanie specjalnego interdyscyplinarnego zespołu ekspertów (z dużym udziałem psychologów), który zająłby się rozwinięciem propozycji opisanych w niniejszej pracy. Pozwoliłoby to na opracowanie w pełni profesjonalnego zestawu testów komputerowych służących pomiarowi predyspozycji psychologicznych ucznia związanych z uczeniem się. O dostrzeganiu znaczenia problemu indywidualizacji nauczania może świadczyć zainteresowanie, jakie towarzyszyło prezentacji wyników badań będących przedmiotem niniejszej pracy na XIII Konferencji *Informatyka w Szkole* w Lublinie (17-20 września 1997) [28] oraz na III Międzynarodowej Konferencji *Wykorzystanie nowych technologii w nauczaniu matematyki* w Koblencji (27 września – 2 października 1997) [76].

Zgromadzenie potrzebnych danych w momencie, gdy zostało już określone co i w jaki sposób należy mierzyć nie przedstawiało większych trudności. Studenci poddawani badaniom (po wcześniejszym wyrażeniu zgody na ich przeprowadzenie) odpowiadali na pytania ankiety i testu, a wyniki były automatycznie zachowywane w postaci pliku na tej samej dyskiecie, na której znajdował się zestaw testujący (por. rozdział 6.2.5).

Autorka wyraża niniejszym serdeczne podziękowanie pod adresem studentów, którzy zgodzili się uczestniczyć w badaniach (łącznie 198 osób), jak też wobec koleżanek i kolegów z WSP i AE, którzy przeznaczili część swoich zajęć na przeprowadzenie badań oraz podzielili się swoimi obserwacjami dotyczącymi ankietowanych studentów.

Ostatnim z zadań tej pracy, mającym równocześnie podstawowe znaczenie dla udowodnienia sformułowanej tezy badawczej, było stwierdzenie na ile jest możliwe wykorzystanie informacji na temat predyspozycji psychologicznych związanych z uczeniem się do sterowania multimedialnym systemem komputerowym wspomagającym nauczanie, tak aby na podstawie tych informacji inteligentny system potrafił automatycznie dobrać metodę nauczania najbardziej odpowiednią dla danego ucznia.

Początkowe wyniki (por. Dodatek A) zdawały się nie potwierdzać przydatności zastosowanych metod. Jednak świadomość jaką rolę odgrywa właściwie zdefiniowana przestrzeń cech w całym procesie rozpoznawania obrazów skłoniła do przypuszczeń, że w tym właśnie punkcie należy upatrywać źródeł niepowodzeń. Podjęto zatem próby modyfikacji określonej uprzednio przestrzeni polegające na zmianie systemu kodowania mierzonych zmiennych zmierzającym do ujednolicenia charakteru danych. Niezbędna okazała się również redukcja wymiaru przestrzeni oraz bardziej obiektywna ocena przynależności elementów ciągu uczącego, dokonana z pomocą analizy skupień, co zostało potwierdzone wyraźną poprawą wyników rozpoznawania (por. rozdział 5.4).

Wobec uzyskanych 96.67% poprawnych rozpoznań, oznaczających wskazanie przez moduł rozpoznawania tej samej metody nauczania, która wynikała z analizy skupień, można przyjąć, że i to zadanie zostało pomyślnie zrealizowane, co jest równoznaczne z potwierdzeniem prawdziwości sformułowanej tezy pracy.

8.2 Wnioski

Na bazie realizacji zarówno zadań szczegółowych, jak i ogólnego celu pracy można wysnuć następujące wnioski: i spostrzeżenia:

1. Przyjęcie metody komputerowego pomiaru cech narzuciło pewne ograniczenia – badanie objęło tylko tych studentów, którzy mieli zajęcia przy komputerze. W tym samym czasie prowadzący zajęcia dokonywali wyboru właściwej metody nauczania dla każdego ze studentów, w oparciu o własne doświadczenie dydaktyczne oraz

obserwacje poczynione w czasie trwania zajęć. W konsekwencji, do subiektywnej opinii prowadzącego zajęcia dołączył jeszcze dodatkowy czynnik zaburzający wynikający ze specyfiki zajęć „informatycznych”, innych niż przedmioty nauczane metodą tradycyjną.

2. Wspomniane w punkcie pierwszym ograniczenie miało też wydźwięk pozytywny. Dało możliwość obserwacji studenta w czasie pracy przy komputerze, a zatem w warunkach, w jakich ma być nauczany z pomocą postulowanego w niniejszej pracy systemu komputerowego.
3. Zarówno studenci poddawani badaniom, jak i nauczyciele, którym prezentowano ideę systemu na konferencjach w Lublinie, Krakowie i Koblencji ([28],[75],[76]) zwracali uwagę na fakt, że sposób przyswajania wiedzy zależy w znacznym stopniu nie tylko od indywidualnych cech osobowości, ale różni się także, dla tego samego ucznia w zależności od nauczanego przedmiotu. Prowadzi to do konkluzji, iż w projekcie modułu gromadzenia danych należy uwzględnić warianty testów sprawdzających różne zdolności zależnie od przedmiotu będącego treścią nauczania. Szczególnie istotne jest rozróżnienie na tzw. przedmioty ścisłe i humanistyczne.
4. Przeprowadzone badania wykazały jak wąski jest zakres zastosowań komputerów w nauczaniu, i to zarówno w ramach lekcji szkolnych, jak i podczas nauki własnej. Jeżeli zajęcia komputerowe w szkole nie sprowadzają się do nauki programowania, to najczęściej ograniczają się do obsługi redaktora tekstu, czasem arkusza kalkulacyjnego.
5. Wykazały także ubóstwo naszego systemu nauczania. Często badani nie umieli wybrać alternatywnej metody uczenia się, gdyż nigdy nie brali takiej możliwości pod uwagę. Dominujący system: podręcznik - nauczyciel - notatki zdawał się wykluczać formy własnych poszukiwań, czy indywidualnych studiów. Fakt ten zdają się potwierdzać odpowiedzi na jedno z pytań ankiety (por. Dodatek D, pyt. 56 i 57) dotyczące studiowania z indywidualnym opiekunem. Spośród czterech możliwych odpowiedzi: **A. jeszcze nie wiem** / **B. tak chętnie** / **C. zdecydowanie nie** / **D. raczej nie**, najczęściej wybieraną była odpowiedź **A** (jeszcze nie wiem).
6. Choć prezentowana praca nie dotyczy, jak to już niejednokrotnie podkreślano, modułu nauczającego, warto już teraz zwrócić uwagę na pewne spostrzeżenia, które nasunęły się w trakcie opracowywania projektu. Otóż pewną trudność może stanowić fakt, że różne treści, nawet w obrębie tego samego przedmiotu, w różnym stopniu poddają się komputeryzacji. W szczególności może się okazać, że pewne tematy trudno będzie zaprezentować przy pomocy wszystkich czterech metod, a takie jest założenie systemu. Nie oznacza to jednak, iż należy zaniechać podjętych prób, a jedynie potwierdza konieczność współpracy z dydaktykami z poszczególnych dziedzin, aby wykorzystując ich wiedzę i doświadczenie uniknąć ewentualnych błędów, nie tylko merytorycznych ale i dydaktycznych.

7. Na uwagę zasługują także obserwacje wynikające z analizy wyników testów. W przypadku ankiety otrzymywano jedynie zestaw odpowiedzi, które wybrał badany uczeń. W testach uwzględniono także: poprawność odpowiedzi, ilość powrotów do prezentowanego tekstu, ilość dokonanych korekt odpowiedzi oraz czas wykonania zadania. Oprócz uwzględnienia tych parametrów w charakterystyce poszczególnych uczniów, warto zwrócić uwagę na pewne prawidłowości, które wystąpiły w badanej grupie. Mianowicie tylko około 10% badanych dokonywało korekty wybranej odpowiedzi, choć ogólnie liczba popełnionych błędów była stosunkowo wysoka (większość studentów odpowiedziała poprawnie na cztery lub pięć z ośmiu postawionych pytań). Równie niewiele, bo tylko 24% badanych korzystało z możliwości ponownego przeczytania tekstu. Trudno to uzasadnić presją czasową gdyż przed przystąpieniem do rozwiązywania zadań studenci byli informowani, iż czas odpowiedzi nie ma znaczenia dla ich końcowej charakterystyki. Przyczyn takiego podejścia do testów należy raczej upatrywać w przyzwyczajeniach związanych z dotychczasowym stylem nauki. W funkcjonującym w Polsce systemie szkolnym, gdzie liczą się głównie wiadomości a nie umiejętności {[58],[72]} uczniów, często nawet podświadomie, nastawiony jest na wykazanie się znajomością określonych porcji wiedzy. Wahanie, bądź próba korekty już udzielonej odpowiedzi odbierana jest jako dowód zbyt słabego opanowania zadanego materiału. Natomiast w społeczeństwie informacyjnym, a do takiego zmierzamy (niezależnie od naszej woli), umiejętność wyszukiwania potrzebnej informacji wśród coraz obfitszych zasobów wiedzy wydaje się być cenniejsza niż samo jej zapamiętanie. Zdaniem autorki komputerowy system wspomagający nauczanie, którego założenia są przedmiotem badań w niniejszej pracy, może skutecznie pomóc w opanowaniu umiejętności swobodnego przeglądania i wyszukiwania w elektronicznych bazach wiedzy.

Powyższe obserwacje zostały tutaj podane jedynie w celu zasygnalizowania pewnych zjawisk, jak się wydaje dość powszechnych w polskim systemie edukacji. Głębsza ich analiza wymaga jednak osobnych studiów na gruncie psychologii uczenia, dydaktyk szczegółowych bądź technologii kształcenia.

8.3 Kierunki dalszych badań

Jak wynika z przedstawionych wyników podjęcie problemu badawczego, zmierzającego do opracowania podstaw dla inteligentnego systemu komputerowego wspomagania nauczania było celowe i słuszne. Opisywane badania wykazały, że wykorzystanie metod rozpoznawania obrazów umożliwi indywidualizację stylu uczenia się przez wybór metody nauczania adekwatnej do naturalnych predyspozycji i ukształtowanych nawyków ucznia w tym zakresie. Należy jednak wyraźnie podkreślić, iż była to tylko próba. Kolejnym krokiem w badaniach powinno być opracowanie i implementacja komputerowa brakującego modułu nauczającego, który umożliwi rzeczywistą ocenę przydatności projektowanego systemu.

Równolegle powinny być prowadzone dalsze prace nad zdefiniowaną przestrzenią cech i narzędziami pomiaru. Niewielkie różnice pomiędzy wynikami z różnych zestawów cech, oraz wątpliwości dotyczące sposobu opisu tych cech wskazują na konieczność dopracowania tej części algorytmu rozpoznawania, przy czym dotychczasowe rezultaty powinny być potraktowane jako ważne źródło wiedzy na temat istotności i wartości informacyjnej poszczególnych cech.

Reasumując warto zaznaczyć, iż praktycznie każda rozmowa nawiązująca do problematyki multimedialnego systemu wspomagającego nauczanie, postulowanego przez autorkę pracy przynosiła konkretne uwagi na temat tego, co powinno się w nim znaleźć i jakie czynniki różnicujące sposób przyswajania wiedzy należałoby uwzględnić. Nie świadczy to jednak, zdaniem autorki, o słabości projektowanego systemu, a jedynie o dostrzeganiu przez zainteresowanych (nauczycieli i studentów) potrzeby i możliwości pełniejszego wykorzystania komputerów w nauczaniu i uczeniu się. Dla autorki niniejszej pracy stanowi to zachętę do dalszych badań nad podjętym tematem i potwierdzenie słuszności dotychczasowych działań. Autorka składa gorące podziękowanie wszystkim, którzy swoimi uwagami, bądź pytaniami przyczynili się do rozwinięcia i poszerzenia pierwotnej idei systemu, a tym samym do wzbogacenia treści niniejszej pracy

.

Dodatek A – Zestawienie wyników rozpoznawania dla różnych sposobów kodowania i różnych zestawów cech

Tabela A.1. Wyniki rozpoznawania dla pełnego zestawu (80) cech przy różnych rodzajach kodowania

	Kod prosty			Kod 1-z-N		
Metryka	Euklid.	uliczna	Czebysz.	Euklid.	uliczna	Czebysz.
K-M 4	60.00%	66.67%	46.67%	50.00%	50.00%	33.33%
K-M 3	60.00%	60.00%	33.33%	63.33%	63.33%	40.00%
U 4	28.30%	28.30%	30.19%	43.33%	43.33%	30.00%

	Kod kumulowany		
Metryka	Euklid.	uliczna	Czebysz.
K-M 4	40.00%	40.00%	40.00%
K-M 3	40.00%	40.00%	40.00%
U 4	33.33%	34.53%	30.119%

Tabela A.II. Wyniki rozpoznawania dla pierwszego wariantu (11 cech) przy różnych rodzajach kodowania

	Kod prosty			Kod 1-z-N		
Metryka	Euklid.	uliczna	Czebysz.	Euklid.	uliczna	Czebysz.
K-M 4	86.67%	83.33%	73.33%	80.00%	80.00%	40.00%
K-M 3	83.33%	83.33%	70.00%	86.67%	86.67%	40.00%
U 4	46.67%	36.67%	40.00%	40.00%	40.00%	30.00%

	Kod kumulowany		
Metryka	Euklid.	uliczna	Czebysz.
K-M 4	80.00%	76.67%	66.67%
K-M 3	70.00%	70.00%	73.33%
U 4	43.33%	36.67%	36.67%

Oznaczenia: K-M4 – podział obiektów na grupy dokonany metodą K-Means dla 4 klas
 K-M3 – jw. ale dla 3 klas
 U4 – podział zbioru obiektów na podstawie opinii uczących – 4 klasy

Tabela A.III. Wyniki rozpoznawania dla drugiego wariantu (15 cech) przy różnych rodzajach kodowania

metryka	Kod prosty			Kod 1-z-N		
	Euklid.	uliczna	Czebysz.	Euklid.	uliczna	Czebysz.
K-M 4	66.67%	63.33%	53.33%	66.67%	66.67%	50.00%
K-M 3	70.00%	70.00%	63.33%	73.33%	73.33%	50.00%
U 4	30.00%	43.33%	33.33%	53.33%	53.33%	43.33%

Kod kumulowany			
Metryka	Euklid.	uliczna	Czebysz.
K-M 4	56.67%	56.67%	46.67 %
K-M 3	60.00%	60.00%	50.00%
U 4	30.00%	43.33%	36.67%

Tabela A.IV. Wyniki rozpoznawania dla trzeciego wariantu (13 cech) przy różnych rodzajach kodowania

Metryka	Kod prosty			Kod 1-z-N		
	Euklid.	uliczna	Czebysz.	Euklid.	uliczna	Czebysz.
K-M 4	80.00%	76.67%	56.56%	76.67%	70.00%	63.33%
K-M 3	83.33%	73.33%	63.33%	80.00%	73.33%	70.00%
U 4	46.67%	40.00%	46.67%	50.00%	50.00%	43.33%

Kod kumulowany			
Metryka	Euklid.	uliczna	Czebysz.
K-M 4	86.67%	90.00%	76.67%
K-M 3	83.33%	76.67%	73.33%
U 4	66.67%	63.33%	63.33%

Oznaczenia: K-M4 – podział obiektów na grupy dokonany metodą K-Means dla 4 klas

K-M3 – jw. ale dla trzech klas

U4 – podział zbioru obiektów na podstawie opinii uczących – 4 klasy

Tabela A.V. Wyniki rozpoznawania za pomocą algorytmu najbliższej mody

zestaw cech	m. euklidesowa	m. uliczna	m. Czebyszewa
Wszystkie	33.33%	33.33%	33.33%
Wariant I	40.00%	40.00%	40.00%
Wariant II	23.33%	23.33%	23.33%
Wariant III	30.00%	30.00%	30.00%

Dodatek B - Porównanie grupowania obiektów w wyniku analizy skupień z klasyfikacją nauczyciela

nr obiektu	fp-ucz.	fp-an.skup.	nr obiektu	fp-ucz.	fp-an.skup.	nr obiektu	fp-ucz.	fp-an.skup.
1	1	4	37	3	4	73	1	2
2	1	4	38	3	3	74	4	3
3	1	3	39	3	4	75	2	4
4	1	1	40	4	4	76	1	2
5	1	4	41	4	4	77	3	4
6	1	2	42	4	4	78	2	4
7	1	4	43	4	3	79	4	1
8	1	3	44	4	3	80	2	3
9	1	1	45	4	3	81	1	1
10	1	4	46	4	2	82	2	4
11	1	2	47	4	2	83	1	1
12	1	1	48	4	2	84	1	4
13	1	2	49	4	1	85	1	4
14	1	2	50	4	1	86	2	1
15	1	1	51	4	3	87	2	4
16	1	3	52	4	2	88	1	2
17	2	2	53	4	2	89	1	1
18	2	3	54	4	3	90	2	3
19	2	4	55	4	4	91	2	1
20	2	1	56	1	2	92	3	2
21	2	2	57	1	2	93	4	1
22	2	4	58	3	2	94	3	2
23	2	3	59	3	2	95	3	3
24	3	4	60	4	2	96	1	1
25	3	3	61	4	4	97	1	2
26	3	3	62	4	1	98	2	3
27	3	3	63	1	4	99	2	3
28	3	4	64	3	2	100	3	2
29	3	4	65	1	2	101	3	4
30	3	2	66	4	4	102	4	3
31	3	3	67	3	2	103	3	3
32	3	2	68	4	2	104	1	2
33	3	4	69	3	1	105	2	2
34	3	4	70	1	4	106	4	1
35	3	3	71	3	2	107	4	1
36	3	4	72	4	4	108	4	2

Uwaga: Zaznaczone pola wskazują zgodność klasyfikacji uczących z klasyfikacją na podstawie analizy skupień.

Dodatek C – Zestaw pytań psychologa*

Podane niżej pytania ilustrują scenariusz rozmowy psychologa z dzieckiem i jego rodzicami przeprowadzanej w poradni. Typowymi sytuacjami, w których odbywa się rozmowa są albo trudności w nauce, albo wykazywane przez ucznia ponadprzeciętne uzdolnienia. W jednej i drugiej sytuacji próbuje się ustalić faktyczny poziom tych uzdolnień oraz wskazać metody dalszego postępowania. Nie wszystkie z wymienionych pytań są użyteczne dla badań opisywanych w niniejszej pracy, ale zostały tutaj przytoczone w celu zilustrowania potencjalnych czynników branych pod uwagę w badaniach dotyczących cech osobowości ucznia.

Cz. I. Informacje uzyskiwane na podstawie pytań:

- Ciąża, poród (choroby, lekarstwa, samopoczucie matki, powikłania, skala Apgar);
- Rozwój we wczesnym dzieciństwie (ruchowy i mowy);
- Choroby dziecka (wypadki, pobyty w szpitalu, zażywane lekarstwa, operacje, wady wzroku, słuchu);
- Środowisko rodzinne (wykształcenie rodziców, praca, rodzeństwo, czy pełna rodzina);
- Warunki mieszkaniowe (czy jest kąt do pracy);
- Czy dziecko chodziło do „zerówki”, jak było przygotowane do szkoły;
- Kiedy nauczyło się czytać, czy lubi czytać
- Jak wygląda plan dnia dziecka
- Czy odrabia lekcje (jak chętnie, ile godzin, czy samodzielnie, czy z pomocą rodziców, jak ta praca wygląda);
- Czy lubi chodzić do szkoły;
- Czy wie, co ma zadane, jak wyglądają jego notatki ze szkoły;
- Oceny, przebieg nauki szkolnej;
- Zainteresowania pozaszkolne;
- Zajęcia dodatkowe
- Jak rodzice reagują gdy dziecko nie chce się uczyć, sięść do nauki;
- Odporność na frustrację, jak reaguje na niepowodzenia;
- Ocena dziecka przez rodziców, jakie jest dziecko

Cz. II. Informacje uzyskiwane na podstawie sprawdzianów:

- możliwości intelektualne
- pamięć (słuchowa, wzrokowa)
- koncentracja uwagi
- koordynacja wzrokowo-ruchowa

* Opracowano przy współpracy mgr Ewy Znamirowskiej Jagielak z Poradni Wychowawczo-Zawodowej nr 1 w Krakowie.

- motywacja do pracy (sposób pracy, czy uczeń łatwo się zniechęca, czy jest gotów podejmować wysiłek i trudne zadania)
- umiejętność samodzielnej pracy (rozumienie instrukcji złożonej z 2-3 poleceń w jednym zdaniu – ciche czytanie ze zrozumieniem)
- analiza i synteza słuchowa
- lateralizacja
- czynniki emocjonalne: nadpobudliwość, zahamowanie, niepewność, brak wiary w siebie, reakcja na stres

Dodatek D – Zestaw pytań ankiety

Grupa I Środowisko:

Jakie miał warunki rozwoju intelektualnego w domu i w najbliższym środowisku, tzn.:

- dostęp do wiedzy, ośrodków kultury
- wykształcenie rodziców, dziadków
- charakter pracy rodziców
- jak wcześnie zaczął edukację (zerówka, przedszkole)

Pytania:

1. Miejscowość, z której pochodzisz

- A. duże miasto
- B. małe miasto
- C. miasteczko
- D. wieś

2. Wykształcenie ojca

- A. wyższe
- B. średnie
- C. zawodowe
- D. podstawowe

3. Wykształcenie matki

- A. wyższe
- B. średnie
- C. zawodowe
- D. podstawowe

4. Czy ktoś z twoich dziadków miał wyższe wykształcenie ?

- A. tak, ze strony matki
- B. tak, ze strony ojca
- C. nikt

5. Rodzaj wykształcenia ojca

- A. humanistyczne
- B. techniczne
- C. medyczne / biologiczne
- D. matematyczne / ekonomiczne
- E. artystyczne
- F. ukierunkowane zawodowo

6. Rodzaj wykształcenia matki

- A. humanistyczne

- B. techniczne
- C. medyczne / biologiczne
- D. matematyczne / ekonomiczne
- E. artystyczne
- F. ukierunkowane zawodowo

7. Czy chodziłeś do przedszkola ?

- A. tak, kilka lat
- B. tylko do zerówki
- C. nie chodziłem

Grupa II Początki edukacji

- *kiedy zaczął czytać, uczył się sam, czy z pomocą innych, pomagał innym w nauce, miał trudności z nauką*

Pytania:

8. Czy pomagasz / pomagałeś kiedyś rodzeństwu w nauce ?

- A. tak, wiele razy
- B. sporadycznie
- C. nie, nigdy

9. Gdzie poznawałeś pierwsze litery ?

- A. w szkole
- B. w przedszkolu / zerówce
- C. w domu - uczyła mnie mama/ babcia..
- D. w domu - uczyłem się sam

10. Kiedy zacząłeś czytać książki samodzielnie ?

- A. w pierwszej klasie
- B. w drugiej klasie
- C. w trzeciej klasie
- D. gdy zaczęły się lektury

Grupa III Ukształtowane nawyki

- *czy lubi czytać, co czyta, czy woli bierny odbiór przez telewizję, czy sam poszerza wiedzę...*

Pytania:

11. Czy lubisz czytać książki ?

- A. tak, bardzo
- B. niezbyt
- C. czytanie to moje utrapienie

12. Jakie książki najczęściej czytasz ?

- A. lekkie, nie wymagające myślenia
- B. psychologiczne
- C. literaturę faktu / eseje
- D. literaturę popularno - naukową
- E. literaturę piękną / klasykę

13. Czy przed przeczytaniem książki ważna jest dla ciebie ilość stron ?

- A. nie
- B. tak, i to bardzo

14. Czy zdarza ci się sięgać po książki popularno - naukowe ?

- A. tak, często
- B. raczej rzadko
- C. nie, nigdy

15. Czy w czasie nauki szkolnej próbowałeś się samodzielnie dowiedzieć czegoś więcej niż to, co było na lekcji ?

- A. nie odczuwałem takiej potrzeby
- B. tak ,ale rzadko
- C. tak, często

16. Ile czasu dziennie spędzasz przed telewizorem ?

- A. każdą wolną chwilę
- B. ok. 3 godzin
- C. ok. 1 godziny
- D. 15-30 minut

17. Dlaczego zasiadasz przed telewizorem ?

- A. dla wypełnienia wolnego czasu
- B. aby obejrzeć wybrany program
- C. z przyzwyczajenia
- D. nie zasiadam, bo szkoda mi czasu
- E. nie zasiadam, bo nie mam na to czasu

18. Jeżeli dostępny jest film nakręcony na podstawie powieści będącej lekturą obowiązkową, to:

- A. najpierw oglądam film, potem czytam książkę
- B. nie czytam, wystarczy mi film
- C. czytam, a jak jest możliwość oglądam również film

19. Czy oglądasz (oglądałbyś mając taką możliwość) programy w wersji obcojęzycznej (w TV sat. lub kablowej) ?

- A. tak, często
- B. tak, czasami

- C. nie, bo ich nie rozumiem
- D. nie interesują mnie

Grupa IV Gospodarowanie własnym czasem.

- czy umie właściwie organizować sobie czas, dzielić go między obowiązki a przyjemności, czy brak czasu wynika z ilości zajęć, czy z niewłaściwego nim gospodarowania.

Pytania:

20. Czy masz w domu inne obowiązki poza nauką ?

- A. tak, ale niewiele
- B. tak, dużo
- C. nie mam
- D. sporadycznie

21. Ile czasu dziennie zajmują ci te obowiązki ?

- A. kilka godzin
- B. nie więcej niż jedną godzinę
- C. zero

22. Czy masz jakieś regularne zajęcia dodatkowe (także dla rozrywki) ?

- A. kursy językowe, zawodowe
- B. zajęcia sportowe
- C. zajęcia artystyczne (taniec, teatr, muzyka.)
- D. nie mam

23. Ile czasu w tygodniu zajmują ci te zajęcia ?

- A. ok. 2 godzin
- B. ok. 4 godzin
- C. 6 lub więcej godzin
- D. zero

24. Jak najczęściej spędzasz wolny czas ?

- A. z przyjaciółmi
- B. na świeżym powietrzu
- C. w kinie, na dyskotecę, w pubie
- D. czytam
- E. oglądam telewizję
- F. nie mam wolnego czasu

25. Ile czasu dziennie poświęcasz na naukę własną ?

- A. 1 godzinę
- B. 2 lub 3 godziny
- C. 4 godziny lub więcej
- D. każdą wolną chwilę

E. uczę się tylko przed egzaminami

Grupa V Sposób pracy

- *samodzielność, uczenie się przez przekazywanie wiedzy, motywacja (rola oceny), ewent. trudności w nauce, sposób sprawdzania wiedzy - ustny, czy pisemny*

Pytania:

26. Czy lubisz się uczyć z kolegą / z koleżanką ?

- A. tak, często
- B. tak, ale tylko przed egzaminem
- C. nie, zawsze uczę się sam

27. Czy, kiedy byłeś młodszy, potrzebowałeś obecności dorosłych w czasie odrabiania lekcji ?

- A. tak, zazwyczaj towarzyszyła mi mama
- B. często pomagał mi ktoś z rodzeństwa
- C. nie, zawsze uczyłem się sam

28. Jakiego rodzaju pomocy potrzebowałeś ?

- A. wystarczyła tylko sama obecność
- B. często potrzebowałem wyjaśnień
- C. tylko sprawdzenia tego, co zrobiłem
- D. nie potrzebowałem pomocy

29. Czy uważasz, że masz wystarczająco dużo czasu na naukę własną ?

- A. tak
- B. nie
- C. nie wiem

30. Jaką wolisz formę sprawdzianów z przedmiotów ścisłych - pisemną, czy ustną ?

- A. jest mi to obojętne
- B. zdecydowanie pisemną
- C. wolę ustną

31. Jaką wolisz formę sprawdzianów z przedmiotów ogólnych i humanistycznych - pisemną, czy ustną ?

- A. jest mi to obojętne
- B. zdecydowanie pisemną
- C. wolę ustną

32. Czy dostajesz lepsze oceny ze sprawdzianów pisemnych, czy z odpowiedzi ustnych ?

- A. mniej więcej takie same
- B. z pisemnych
- C. z ustnych

33. Negatywna ocena ze strony nauczyciela oznaczała dla ciebie

- A. zaniechanie dalszych wysiłków
- B. mobilizację do pracy
- C. nigdy nie miałem złych ocen

Grupa VI Metoda pracy

- *mobilizacja tylko w obliczu oceny końcowej, systematycznie / korzystanie z bibliotek, notatki, fiszki / sposób przyswajania wiedzy - pamiętanie, zrozumienie*

Pytania:

34. Czy w trakcie nauki szkolnej wypracowałeś własny system uczenia się ?

- A. tak
- B. nie
- C. nie wiem
- D. nie odczuwałem takiej potrzeby

35. Czy zdarzyło ci się być nieprzygotowanym do sprawdzianu / egzaminu z powodu braku potrzebnych podręczników ?

- A. tak
- B. nie
- C. nie pamiętam

36. W jaki sposób starasz się przyswoić potrzebną wiedzę (np. przed egzaminem) ?

- A. uczę się na pamięć
- B. staram się przede wszystkim zrozumieć tekst
- C. robię dużo notatek, potem je przeglądam

37. W jaki sposób pomagasz sobie w zapamiętywaniu potrzebnej wiedzy ?

- A. nie mam konkretnej metody
- B. powtarzam tekst na głos
- C. podkreślam w tekście ważniejsze zdania
- D. robię krótkie notatki

38. Czy w szkole na zakończenie semestru miałeś problemy z ocenami ?

- A. nie, uczyłem się systematycznie
- B. zwykle poprawiałem je w ostatniej chwili
- C. nie pamiętam

39. Z czego korzystasz najczęściej przygotowując się do sprawdzianu (egzaminu)

- A. zazwyczaj wystarczają mi notatki
- B. z różnych podręczników po trochu
- C. notatki + jeden wybrany podręcznik

40. Przekazany ci w szkole system zdobywania wiedzy polegał na:

- A pamięciowym opanowaniu materiału
- B umiejętności wyszukiwania informacji
- C rozwiązywaniu problemów
- D umiejętności kojarzenia faktów

Grupa VII Dotychczasowe osiągnięcia

- uzyskiwane oceny, znajomość języków obcych

Pytania:

41. Jakie oceny dostawałeś najczęściej z matematyki w szkole średniej ?

- A. bardzo dobre
- B. dobre
- C. dostateczne
- D. niedostateczne

42. Jak sam ocenilbyś swoją znajomość matematyki w szkole średniej ?

- A. bardzo dobra
- B. dobra
- C. dostateczna
- D. niedostateczna

43. Jakie oceny dostawałeś najczęściej z historii w szkole średniej?

- A. bardzo dobre
- B. dobre
- C. dostateczne
- D. niedostateczne

44. Jak sam ocenilbyś swoją znajomość historii w szkole średniej ?

- A. bardzo dobra
- B. dobra
- C. dostateczna
- D. niedostateczna

45. Jakie oceny dostawałeś najczęściej z fizyki w szkole średniej ?

- A. bardzo dobre
- B. dobre
- C. dostateczne
- D. niedostateczne

46. Jak sam ocenilbyś swoją znajomość fizyki w szkole średniej ?

- A. bardzo dobra
- B. dobra
- C. dostateczna

D. niedostateczna

47. Ilu języków obcych uczyłeś się dotychczas ?

- A. żadnego
- B. jednego
- C. dwóch
- D. trzech lub więcej

48. W ilu językach obcych jesteś w stanie się porozumieć ?

- A. w żadnym
- B. w jednym
- C. w dwóch lub więcej

49. Jak oceniasz swoją znajomość angielskiego ?

- A. nie znam wcale
- B. słaba
- C. dobra
- D. bardzo dobra

Grupa VIII Kierunek studiów

- jego wybór, akceptacja, stopień zainteresowania

Pytania:

50. Czy studiowałeś już wcześniej ?

- A. nie, to moje pierwsze studia
- B. tak, zaczynałem już inne studia
- C. tak, mam już jeden dyplom

51. Jeżeli przerwałeś studia, to z jakiego powodu ?

- A. nie przerwałem
- B. nie spełniały moich oczekiwań
- C. to była tylko ucieczka przed wojskiem
- D. miałem kłopoty z egzaminami

52. Dlaczego wybrałeś obecny kierunek studiów ?

- A. zawsze mnie interesował
- B. tu były akurat wolne miejsca
- C. jest teraz modny
- D. wybierałem pod wpływem innej osoby
- E. daje szansę na dobrze płatną pracę

53. Czy, gdybyś miał wybierać jeszcze raz, wybrałbyś ten sam kierunek ?

- A. tak, zawsze
- B. nie nigdy
- C. raczej nie

54. Od jak dawna interesujesz się tym, co studiujesz ?

- A. od czasu, gdy zacząłem studia
- B. od szkoły podstawowej
- C. zacząłem pod koniec liceum
- D. chyba mnie to w ogóle nie interesuje

55. Przystosowanie się do systemu nauczania na studiach jest dla Ciebie:

- A. łatwiejsze niż w szkole średniej
- B. trudniejsze niż w szkole średniej
- C. nie widzę różnicy

Grupa IX Gotowość, umiejętność przystosowania się do nowych metod pracy

Pytania:

56. Czy odpowiada ci organizacja zajęć na uczelni (np. o różnych porach dnia) ?

- A. nie, wolę system lekcyjny
- B. nie, cały dzień mam zajęty
- C. tak, łatwiej zorganizować sobie czas
- D. tak, daje większą swobodę wyboru zajęć

57. Czy gdybyś mógł wybrać studiowanie z indywidualnym opiekunem naukowym, skorzystałbyś z tej możliwości ?

- A. jeszcze nie wiem
- B. tak, chętnie
- C. zdecydowanie nie
- D. raczej nie

58. Jaki motyw decydowałby o twoim wyborze ?

- A. w grupie łatwiej ukryć własną niewiedzę
- B. lubię pracę w grupie
- C. indywidualizm daje szerszą wiedzę
- D. mógłbym się uczyć tego, co mnie interesuje

59. Co sądzisz o studiowaniu na odległość - otrzymujesz potrzebne materiały, a z nauczycielem porozumiewasz się np. poprzez Internet ?

- A. to by było bardzo wygodne
- B. chętnie, o ile będę dostawał konkretne zadania
- C. nie, to wymaga zbyt dużej dyscypliny
- D. nic nie zastąpi tradycyjnego wykładu

Grupa X Doświadczenia komputerowe

- znajomość komputera i dostęp do niego, wiedza na temat współczesnych technik komputerowych, zastosowań komputerów w edukacji, znajomość programów edukacyjnych

Pytania:

60. Czy miałeś w szkole przedmiot *Elementy informatyki* lub podobny ?

- A. tak, tylko przez rok
- B. tak, przez dwa lata lub więcej
- C. nie, wcale

61. Czy masz dostęp do komputera poza uczelnią ?

- A. tak, w każdej chwili
- B. tak, ale tylko grzecznościowo
- C. nie mam

62. Czy chciałbyś się nauczyć korzystać z komputera ?

- A. już umiem
- B. tak, chętnie
- C. nie wiem
- D. nie lubię komputerów

63. Przez ile godzin w tygodniu masz na uczelni zajęcia przy komputerze ?

- A. nie mam w ogóle
- B. 1 x 1.5 godziny
- C. 2 x 1.5 godziny
- D. 4 lub więcej godzin

64. Do czego najczęściej używasz komputera ?

- A. przede wszystkim do gier
- B. jako maszynę do pisania
- C. jako narzędzie do obliczeń
- D. nie używam go na co dzień

65. Czy zgadzasz się z twierdzeniem, że komputer może pomóc w zdobywaniu wiedzy ?

- A. chyba tak, choć nie wiem w jaki sposób
- B. sądzę, że niewiele
- C. może o ile mamy odpowiednie programy
- D. nie wiem

66. Które z podanych niżej zastosowań edukacyjnych wskazałbyś jako najbardziej użyteczne ?

- A. testy komputerowe
- B. encyklopedie multimedialne
- C. programy do nauki języków obcych

- D. programy użytkowe (edytor, arkusz kalkulacyjny)
- E. gry edukacyjne
- F. Internet

67. Czy wiesz co to jest Internet ?

- A. tak
- B. nie mam pojęcia
- C. właściwie nie, choć znam to słowo

68. Jakie twoim zdaniem są najistotniejsze korzyści płynące z sieci Internet?

- A. nie miałem okazji ich poznać
- B. to tylko pochłaniacz czasu
- C. to jest doskonałe źródło wiedzy
- D. możliwość szybkiej wymiany informacji

Grupa XI Odbiór treści nauczania

- *forma przekazu treści, rola ilustracji, wykresów i tabel*

Pytania:

69. W tekście, który masz opanować jest dużo tabel, wykresów. Czy czytając tekst analizujesz ich treść ?

- A. tak, zazwyczaj studiuję je dokładnie
- B. przeglądam je, ale bez wnikania w szczegóły
- C. nie, zwykle je pomijam

70. Jakie znaczenie mają dla ciebie ilustracje w tekście ?

- A nie są mi potrzebne, ignoruję je
- B przyjemnie urozmaicają tekst
- C są dodatkowym źródłem informacji

71. Czy podczas studiowania tekstu, w którym zawarto wiele danych liczbowych, zebranie ich w jednej tabeli stanowi dla ciebie istotną pomoc?

- A tak
- B nie, nie korzystam z nich
- C nie wiem

72. Czy ucząc się tekstu, w którym podano wiele faktów i danych liczbowych sam sporządzasz odpowiednie zestawienia, aby ułatwić sobie ich zapamiętanie ?

- A tak
- B nie

73. Jaką formę prezentowania wiadomości lubisz najbardziej ?

- A opis tekstowy
- B rysunki, schematy
- C nie zastanawiałem się

74. Jak często zdarza ci się powtarzać na głos materiał w trakcie nauki ?

A zazwyczaj tak się właśnie uczę

B nigdy tego nie robię

**Dodatek E – Macierz korelacji dla wybranych cech
(fragmenty)**

STAT.	Korelacje (obiektyw1.sta)										
PODST.	Oznaczone wsp. korelacji są istotne z $p < ,05000$										
STATYST.	N=108 (Braki danych usuwano przypadkami)										
Zmienna	WŁ_POSTV_OBC	NEG_OCSYS_NA	OC_PROSYS_SZ	KIER_S	ZAINT	INDYW_STUD_O	KORZ_I				
	Z	OJ	EN	UK	BL	K	TU	R	OP	DL	NE
WŁ_POSZ	1,00	-,04	,08	-,23*	,04	,27*	-,13	-,18	,12	-,05	,18
TV_OBCOJ	-,04	1,00	-,11	,25*	-,10	,03	-,02	,28*	,11	,10	-,05
NEG_OCEN	,08	-,11	1,00	,02	-,10	-,15	-,00	,08	-,01	,08	-,03
SYS_NAUK	-,23*	,25*	,02	1,00	,19*	-,04	-,17	,11	,01	-,04	-,15
OC_PROBL	,04	-,10	-,10	,19*	1,00	,05	,16	-,17	,02	-,25*	,08
SYS_SZK	,27*	,03	-,15	-,04	,05	1,00	-,23*	-,10	,03	-,10	,08
KIER_STU	-,13	-,02	-,00	-,17	,16	-,23*	1,00	,00	,18	-,19	,13
ZAINTER	-,18	,28*	,08	,11	-,17	-,10	,00	1,00	,10	,12	-,16
INDYW_OP	,12	,11	-,01	,01	,02	,03	,18	,10	1,00	-,01	-,06
STUD_ODL	-,05	,10	,08	-,04	-,25*	-,10	-,19	,12	-,01	1,00	-,11
KORZ_INE	,18	-,05	-,03	-,15	,08	,08	,13	-,16	-,06	-,11	1,00

Rys. E.I. Macierz korelacji cech po zredukowaniu (na przykładzie wariantu I).

STAT. Korelacje (obiekty.sta)
 PODST. Oznaczone wsp. korelacji są istotne z $p < ,05000$
 STATYST. N=108 (Braki danych usuwano przypadkami)

Zmienna	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14
VAR1	1,00	,34*	,24*	,13	,06	,03	,09	-,15	,26*	-,21*	-,00	-,01	,03	,18
VAR2	,34*	1,00	,61*	,23*	-,13	-,02	,23*	-,07	,13	-,15	,03	,13	,04	,01
VAR3	,24*	,61*	1,00	,32*	-,10	-,38*	,38*	-,06	,16	-,10	,00	,10	-,02	,03
VAR4	,13	,23*	,32*	1,00	-,12	-,04	,21*	-,05	,14	-,18	-,05	,16	,00	-,12
VAR5	,06	-,13	-,10	-,12	1,00	,26*	-,08	-,02	-,05	,08	,16	,03	-,00	,01
VAR6	,03	-,02	-,38*	-,04	,26*	1,00	-,09	-,08	-,10	-,02	,07	,11	,08	-,10
VAR7	,09	,23*	,38*	,21*	-,08	-,09	1,00	-,27*	,34*	-,17	-,18	,08	-,08	,10
VAR8	-,15	-,07	-,06	-,05	-,02	-,08	-,27*	1,00	-,08	,14	,08	-,08	-,00	-,02
VAR9	,26*	,13	,16	,14	-,05	-,10	,34*	-,08	1,00	-,41*	,04	-,03	-,24*	,05
VAR10	-,21*	-,15	-,10	-,18	,08	-,02	-,17	,14	-,41*	1,00	,12	-,10	,06	,04
VAR11	-,00	,03	,00	-,05	,16	,07	-,18	,08	,04	,12	1,00	,13	,33*	,15
VAR12	-,01	,13	,10	,16	,03	,11	,08	-,08	-,03	-,10	,13	1,00	-,04	-,11
VAR13	,03	,04	-,02	,00	-,00	,08	-,08	-,00	-,24*	,06	,33*	-,04	1,00	,14
VAR14	,18	,01	,03	-,12	,01	-,10	,10	-,02	,05	,04	,15	-,11	,14	1,00
VAR15	-,10	,01	,05	-,02	-,19*	-,12	-,02	-,06	,02	-,05	,03	,08	-,23*	-,05
VAR16	,05	-,12	-,20*	-,13	-,07	,09	-,07	-,26*	,04	-,07	-,01	-,05	-,04	,07
VAR17	-,02	,03	,06	,11	-,12	-,01	,07	-,25*	-,12	-,23*	-,14	,19	,03	,02
VAR18	-,01	,03	,18	,26*	-,16	-,12	,16	,05	,01	-,33*	-,21*	,25*	-,02	-,03
VAR19	,23*	,15	,05	,11	-,12	-,01	,06	,09	-,02	-,19	-,08	,15	-,03	,12
VAR20	-,09	-,05	-,08	-,09	,05	-,06	,00	-,15	-,09	,22*	,10	,11	,04	-,02
VAR21	-,15	-,11	-,22*	-,03	-,00	,06	-,17	,30*	-,01	-,03	-,00	-,13	,02	-,03
VAR22	,40*	,32*	,41*	,08	,01	-,25*	,24*	,00	,33*	-,23*	-,03	-,06	,09	,14
VAR23	,25*	,29*	,23*	,12	-,05	-,19*	,22*	,07	,25*	-,28*	-,06	-,14	,05	,18
VAR24	,12	-,01	-,04	-,10	-,19*	-,10	-,03	-,02	-,08	,03	-,02	-,06	,01	,14
VAR25	-,08	-,18	-,07	-,01	,10	-,06	-,08	,05	,00	-,02	,19*	-,26*	,13	,18
VAR26	-,05	,04	,08	,03	-,06	-,06	,02	,25*	,09	-,08	-,24*	,04	-,07	-,04
VAR27	,10	-,07	,07	,07	,01	,01	,13	,05	,15	-,22*	-,05	,06	-,13	-,01
VAR28	,14	,02	,14	,17	-,04	-,07	,15	-,03	,16	-,16	-,04	-,05	-,13	,11

Rys. EII. Wartości współczynnika Pearsona dla najsilniej skorelowanych cech – cz I
 (macierz korelacji obliczana dla pełnego zestawu cech)

STAT. Korelacje (obiekty.sta)
 PODST. Oznaczone wsp. korelacji są istotne z $p < ,05000$
 STATYST. N=108 (Braki danych usuwano przypadkami)

Zmienna	VAR49	VAR50	VAR51	VAR52	VAR53	VAR54	VAR55	VAR56	VAR57	VAR58	VAR59
VAR16	,05	-,23*	-,20*	,10	,02	-,05	,13	,09	,09	-,11	,05
VAR17	,16	-,10	-,20*	,08	,16	-,13	-,09	-,26*	-,06	-,13	,07
VAR18	-,02	-,16	-,21*	,05	,08	-,00	,01	-,29*	-,03	,09	,04
VAR19	-,35*	,25*	,13	-,02	,02	,28*	,00	-,04	,11	-,13	,10
VAR20	-,07	,10	,03	-,03	,04	-,04	-,06	-,20*	-,16	,06	,17
VAR21	,27*	,29*	,20*	,17	-,06	-,04	,08	,01	,09	-,21*	-,07
VAR22	-,22*	-,04	-,01	-,15	,05	-,05	-,01	-,07	-,07	,02	,06
VAR23	-,14	-,05	-,03	-,14	,05	,06	,00	-,08	-,03	-,12	,04
VAR24	-,15	-,02	,01	,15	,10	,12	,06	-,15	,09	-,04	-,04
VAR25	-,01	,14	,21*	-,03	,18	,02	-,01	,05	,30*	-,00	-,18
VAR26	-,10	,11	,19*	,13	-,13	-,00	,19*	-,19*	,00	,03	,19*
VAR27	,09	,02	-,01	-,03	,09	,07	,08	-,10	-,04	,01	-,01
VAR28	,17	,03	-,01	-,06	,01	,04	,01	-,17	-,01	,04	-,10
VAR29	-,05	-,13	-,04	-,02	,04	-,02	,16	-,27*	-,20*	,05	-,17
VAR30	,04	,14	,20*	,07	-,08	-,14	-,07	-,01	,12	,21*	-,10
VAR31	,02	,02	,07	,06	,01	-,05	,05	-,11	,14	,16	-,00
VAR32	,03	,22*	,22*	,02	,02	-,05	,02	,04	,14	-,01	-,00
VAR33	,00	-,07	-,13	-,00	-,10	,08	-,10	,22*	-,01	,07	,08
VAR34	-,07	,06	,05	-,17	-,04	,11	-,08	,09	,01	-,19*	-,04
VAR35	,08	-,11	-,16	-,05	,02	,01	-,03	,09	-,14	-,11	-,05
VAR36	,04	-,08	,00	,14	,08	-,04	,05	-,19	,09	-,12	-,07
VAR37	-,08	-,04	-,03	,03	,12	,00	-,05	-,23*	-,09	,13	,18
VAR38	,02	-,02	,11	,16	-,00	-,17	,07	,09	,02	,07	-,25*

Rys. EIII. Wartość współczynnika Pearsona dla najsilniej skorelowanych cech – cz. II
 (macierz korelacji obliczana dla pełnego zestawu cech)

Dodatek F – Jedna z grup elementów ciągu uczącego utworzona na podstawie analizy skupień

Grupa I, wariant I, kod prosty

1	2	3	4	5	6	7	8	9	10	11	12	13
Nr obiektu	Wartość cechy											Indeks K lasy
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	
1	2	4	2	3	1	1	1	3	1	3	1	1
2	2	1	2	1	1	1	1	2	2	4	1	1
5	2	2	2	4	1	1	2	3	4	4	1	1
7	2	2	2	1	1	3	1	2	2	4	1	1
15	2	2	2	1	1	1	3	4	1	3	1	1
19	3	2	2	3	3	1	1	3	1	4	1	2
22	3	4	2	3	2	4	1	3	4	4	1	2
24	2	2	2	1	2	4	1	2	2	3	1	3
33	2	4	3	3	3	2	1	3	2	4	1	3
34	2	3	2	2	1	1	2	4	1	4	1	3
36	2	2	2	4	2	1	1	3	3	4	1	3
37	3	1	2	1	1	1	1	3	3	4	1	3
40	2	2	2	1	1	4	1	2	2	4	4	4
66	3	2	2	1	3	4	5	2	2	1	4	4
69	2	3	2	3	1	1	5	3	2	4	1	3
75	2	2	1	3	2	4	1	3	1	3	4	2
76	2	2	1	3	1	3	1	2	2	4	1	1
77	3	1	1	1	2	4	1	2	2	3	1	3

Objaśnienia:

1. kolumna pierwsza - numer elementu w zbiorze wszystkich badanych obiektów.
2. kolumny od 2 do 12 - wartości składowych wektora cech dla poszczególnych obiektów
3. kolumna 13 – numer klasy wskazany przez uczącego

Bibliografia

1. Andrews H..C. *Computer techniques in image processing*, Academic Press New York and London 1970.
2. Atkinson R.L., Atkinson R.C., Hilgard E.R. *Introduction to Psychology*, New York, Harcourt, Brace, Jovanovich, 1983.
3. Azzopardi G. *Oblicz twój IQ*, Livra Editions, 1996.
4. Biedalak K., Woźniak P. „Metoda Super Memo. Optimalizacja procesu uczenia się”, *Informatyka*, Nr 10/1992.
5. Blalock H.M. *Statystyka dla socjologów*, Państwowe Wydawnictwo Naukowe, Warszawa 1975.
6. Brzeziński J. *Elementy metodologii badań psychologicznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1980.
7. Brzeziński J. Red. *Problemy teorii, rzetelności, konstrukcji i analizy wyników testów psychologicznych*, Polskie Towarzystwo Psychologiczne, Warszawa 1988.
8. Brzeziński J., Hornowska E. *Skala inteligencji Wechslera WAIS-R. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej*, Wydawnictwo Naukowe PWN, Warszawa 1993.
9. Brzeziński J. red. *Wielozmiennowe modele statystyczne w badaniach psychologicznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1987.
10. Brzeziński J., Stachowski R. *Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych*, Państwowe Wydawnictwo Naukowe, Warszawa 1984.
11. Budohoska W., Włodarski Z. *Psychologia uczenia się. Przegląd badań eksperymentalnych i teorii*, Państwowe Wydawnictwo Naukowe, Warszawa 1984.
12. Chomątowski S., Sokołowski A. „Taksonomia struktur”, *Przegląd Statystyczny*, R XXV, Zeszyt 2, 1978.
13. Choynowski H. red. *Testy psychologiczne w poradnictwie wychowawczo-zawodowym. Poszerzone podręczniki dla badającego*, Państwowe Wydawnictwo Naukowe, Warszawa 1980.

14. Coombs C.H., Dawes R.M., Tversky A. *Wprowadzenie do psychologii matematycznej*, Państwowe Wydawnictwo Naukowe, Warszawa 1977.
15. *Edukacja Medialna*, Kwartalnik Polskiego Towarzystwa Technologii i Mediów edukacyjnych, nr 2, grudzień 1996.
16. Eysenck H.J. *Sprawdź swoją inteligencję*, Wydawnictwo Dobrucki i Malicki DMS, Otwock 1992.
17. Fu K.S. *Sequential methods in pattern recognition and machine learning*, Academic Press New York and London 1968.
18. Gajek L., Kałuska M. *Wnioskowanie statystyczne. Modele i metody*. Wydawnictwo Naukowo-Techniczne, Warszawa 1994.
19. Goleman D. *Inteligencja emocjonalna*, Wydawnictwo Naukowe PWN, Warszawa 1997.
20. Góralski A. *Metody opisu i wnioskowania statystycznego w psychologii i pedagogice*, Państwowe Wydawnictwo Naukowe, Warszawa 1987.
21. Grabiński T. *Metody taksonometrii*, Wydawnictwo Akademii Ekonomicznej, Kraków 1992.
22. Green J., Lewis D. *Kim możesz być ?*, Wydawnictwo Dobrucki i Malicki DMS, Otwock 1992.
23. Guilford J.P. *Natura inteligencji człowieka*, Państwowe Wydawnictwo naukowe, Warszawa 1978.
24. Guilford J.P. *Podstawowe metody statystyczne w psychologii i pedagogice*, Państwowe Wydawnictwo Naukowe, Warszawa 1960.
25. Gurbiel E., Hardt-Olejniczak G., Kołczyk E., Krupicka H., Sysło M.M. *Technologia informacyjna w kształceniu ogólnym*, WSiP, Warszawa 1997.
26. Hornowski B. *Rozwój inteligencji i uzdolnień specjalnych*, WSiP, Warszawa 1986.
27. *Informatyka w szkole. Materiały z XIII Konferencji*, Lublin, 17-20 września 1997.
28. *Informatyka w szkole. Materiały z XIV Konferencji*, Lublin, 16-19 września 1998.
29. Jaworowska A., Szustrowa T. *Podręcznik do testu matryc Ravena wersja standard. Polska standaryzacja 1989*, Pracownia testów psychologicznych Polskiego Towarzystwa Psychologicznego, Warszawa 1991.
30. Kałuska M. *Rachunek prawdopodobieństwa i statystyka*. WNT, Warszawa 1997.

31. Kostrzewski J. *Wprowadzenie do skali inteligencji D. Wechslera dla dzieci od 5 do 15 lat*, Polskie Towarzystwo Higieny Psychiczej, Warszawa 1973.
32. Kozielecki J. red. *Problemy psychologii matematycznej*, Państwowe Wydawnictwo Naukowe, Warszawa 1971.
33. Krishnaiah P.R., Kanal L.N. *Classification, pattern recognition and reduction of dimensionality*, North-Holland Publishing Company, Amsterdam 1987.
34. Kulikowski J. *Cybernetyczne układy rozpoznające*, PWN, Warszawa 1972.
35. Kuraś B., Mieszkowska A., Moszner P., Szydełko A., Władyka H., Zając M. *Elementy Informatyki, Wybrane zagadnienia dla studentów kierunków humanistycznych*, Wydawnictwo WSP, Kraków 1996 (wyd. II rozszerzone).
36. Kurcz I. *Pamięć, uczenie się, język* z cyklu *Psychologia ogólna* pod red. Tomaszewskiego, Wydawnictwo Naukowe PWN, Warszawa 1995.
37. Kurzyński M. *Rozpoznawanie obiektów - metody statystyczne*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1997.
38. Kwiatkowski S.M., *Komputery w procesie kształcenia i zarządzania szkołą*, Instytut Badań Edukacyjnych, Warszawa 1994.
39. Lindsay P.H., Norman D.A. *Procesy przetwarzania informacji u człowieka. Wprowadzenie do psychologii*, Państwowe Wydawnictwo Naukowe, Warszawa 1984.
40. Lorayne H. *Superpamięć dla uczących się*, Wydawnictwo RAVI, Łódź 1996.
41. Magnusson D. *Wprowadzenie do teorii testów*, Państwowe Wydawnictwo Naukowe, Warszawa 1991.
42. Matczak A. *Diagnoza intelektu*, Wydawnictwo Instytutu Psychologii PAN, Warszawa 1994.
43. Moir A., Jassel D. *Płeć mózgu*, PIW, Warszawa 1995.
44. Morbitzer J. red. *Współczesna technologia kształcenia*, Wydawnictwo Naukowe WSP, Kraków 1997.
45. Nęcka E. *Inteligencja i procesy poznawcze*, Oficyna Wydawnicza „Impuls”, Kraków 1994.
46. Nęcka E. *Proces twórczy i jego ograniczenia*, Oficyna Wydawnicza „Impuls”, Kraków 1995.

47. Nowakowaska M. *Psychologia ilościowa z elementami naukometrii*, PWN Warszawa 1975.
48. Noworol Cz. *Analiza skupień w badaniach empirycznych. Rozmyte modele hierarchiczne*, PWN, Warszawa 1989.
49. Papert S. *Burze mózgów, dzieci i komputery*, PWN, Warszawa 1996.
50. Podolak I.T. *Problemy reprezentacji danych w warstwowym sieciach neuronowych*, Rozprawa doktorska, AGH, Kraków 1998.
51. Podolak I.T., Bielecki A. *Input and output representation problems in supervised learning of feedforward ANN's*, Proceedings of Int. Conf. Neural Networks and their application possibilities, Košice, 1996.
52. *Podstawy programowe obowiązkowych przedmiotów ogólnokształcących*, Załącznik do Zarządzenia Nr 8 Ministra Edukacji narodowej z dnia 15 maja 1997 (dostępne także w *Komputer w Szkole* Nr 2/3 1997).
53. Popiela–Mizera A. *Ocena przydatności metod automatycznego diagnozowania trzustki*, Rozprawa doktorska, AGH, Kraków 1989.
54. Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P. *Numerical recipes in C*, Cambridge University Press, New York 1994.
55. Raven, J.C., *Test matryc, wersja standardowa*, Polskie Towarzystwo Psychologiczne, Warszawa 1986.
56. Raven, J.C., *Test matryc, wersja dla zaawansowanych*, Polskie Towarzystwo Psychologiczne, Warszawa 1986.
57. Seligman D. *O inteligencji prawie wszystko. Kontrowersje wokół ilorazu inteligencji*, Wydawnictwo Naukowe PWN, Warszawa 1995.
58. Siemieniecki B. „Komputer a humanizm – podstawowe dylematy edukacji” w *Komputer w Edukacji* Nr 1 1994.
59. Siemieniecki B. *Komputery i hipermedia w procesie edukacji dorosłych*, Wydawnictwo Adam Marszałek, Toruń 1995.
60. Siemieniecki B. red., *Perspektywa edukacji z komputerem*, Wydawnictwo Adam Marszałek, Toruń - Płock 1995.
61. Siemieniecki B. *Skutki powszechnego stosowania metod informatycznych w edukacji*” w *Komputer w Edukacji*, Nr 1 1995.

62. Skrzydlewski W. *Technologia kształcenia, przetwarzanie informacji, komunikowanie*, Wydawnictwo Naukowe UAM, Poznań 1990.
63. Sobczak W., Malina W. *Metody selekcji informacji*, WNT, Warszawa 1978.
64. Sobczyk M. *Statystyka*, PWN, Warszawa 1996.
65. Steczkowski J., Zeliaś A. *Statystyczne metody analizy cech jakościowych*, PWE, Warszawa 1981.
66. Tadeusiewicz R., Izworski A., Majewski J. *Biometria*, Wydawnictwa AGH, Kraków 1993.
67. Tadeusiewicz R., Flasiński M. *Rozpoznawanie obrazów*, Państwowe Wydawnictwo Naukowe, Warszawa 1991.
68. Tadeusiewicz R. *Rozpoznawanie obrazów w zastosowaniach ekonomicznych*, Wydawnictwo Akademii Ekonomicznej, Kraków 1985.
69. *Techniki komputerowe w przekazie edukacyjnym - 7 Ogólnopolskie Sympozjum naukowe*, Kraków, 26-27 września 1997 (*Streszczenia referatów*).
70. Van Der Heijden K., Eden C. *General and strategic management*, Book 3, University of Strathclyde, Glasgow 1994.
71. Waitley D. *Imperium umysłu*, Wydawnictwo MEDIUM, Warszawa 1997.
72. Włodarski Z. *Odbiór treści w procesie uczenia się*, Państwowe Wydawnictwo Naukowe, Warszawa 1985.
73. Włodarski Z. *Z tajemnic ludzkiej pamięci*, WSiP, Warszawa 1984.
74. Zając M., Maksylewicz B. *Możliwość wykorzystania mikrokomputerów w systemie testowego badania wiadomości uczniów*, Rocznik Naukowo-Dydaktyczny WSP, Zeszyt 157, Kraków 1992.
75. Zając M., Wójcik K. *Indywidualizacja nauczania z uwzględnieniem predyspozycji psychologicznych ucznia*, Materiały z I Ogólnopolskiego Sympozjum „Informatyczne Przygotowanie Nauczycieli”, WSP Kraków 1997 (w druku).
76. Zając M. *A multimedia computer system as a tool of individualization of teaching process*, ICTMT 3 Proceedings, Koblenz 1997 (na krążku CD).
77. Zimbardo P.G., Ruch F.L., *Psychologia i życie*, Wydawnictwo Naukowe PWN, Warszawa 1994.