

## C. INFORMACJA O GŁÓWNYCH KIERUNKACH PRAC JEDNOSTKI

1. Główne kierunki badań naukowych i prac rozwojowych objęte co najmniej trzyletnim programem badań, w tym wiążące się z rządowymi programami gospodarczymi lub społecznymi albo międzynarodowymi programami współpracy naukowej i naukowo-technicznej.

Katedra Informatyki Stosowanej (KIS) jest jednostką naukowo-dydaktyczną Uniwersytetu Mikołaja Kopernika (UMK) podejmującą **interdyscyplinarne** programy badawcze w dziedzinach, które wykorzystują wyrafinowane metody informatyczne. Profil naszej jednostki w klasyfikacji KBN najbliższy jest sekcji T11F (metody komputerowe w nauce) oraz T11C (informatyka stosowana) i mieści się w głównym nurcie prac związanych z tworzeniem społeczeństwa informacyjnego i gospodarki opartej na wiedzy. W terminologii anglojęzycznej nasza działalność odpowiada pojęciu „informatics”, występującym w takich złożeniach jak bioinformatics, neuroinformatics, medical informatics, czy chemical informatics. Stąd w języku angielskim używamy nazwy „Department of Informatics”.

KIS prowadzi badania i organizuje specjalistyczne wykłady i seminaria dotyczące zastosowań informatyki w naukach ścisłych (fizyce i chemii), naukach o życiu (biologia, medycyna), naukach ekonomicznych, technicznych, oraz w kognitywistyce (naukach o poznaniu). Współpracując z kadrą naukową z różnych dyscyplin naukowych – fizyki, chemii, biologii, nauk humanistycznych – dążymy do rozwinięcia tych dziedzin nauk podstawowych i stosowanych, które stosują metody informatyki zarówno w zakresie obliczeń numerycznych (symulacyjnych) jak i zagadnień wymagających obliczeń symbolicznych i metod inteligencji obliczeniowej.

Problematyka badawcza KIS skoncentrowana jest wokół kilku głównych tematów:

1. Zagadnień inteligencji obliczeniowej (computational intelligence), a w szczególności szeroko pojętych zagadnień neuroinformatyki i metod uczenia maszynowego.
2. Metod komputerowych w zastosowaniu do problemów fizyki, chemii i biologii molekularnej (*life sciences*).
3. Kognitywistyki (*cognitive sciences*), włączając w to neuroinformatykę, komputerowe modele procesów poznawczych i filozofię umysłu.
4. Dydaktyki wspomaganiej metodami komputerowymi.

W 2003 roku rozwijaliśmy następujące kierunki badań:

### I. Teoria i zastosowania metod inteligencji obliczeniowej.

Jest to obecnie najważniejszy kierunek naszych badań; prace dotyczą teorii i zastosowań modeli typu sieci neuronowych, logiki rozmytej, drzew decyzji, algorytmów opartych na podobieństwie oraz fundamentalnych mechanizmów działania struktur adaptujących się. Nazwy „Computational Intelligence” używaliśmy od wielu lat; warto wspomnieć, że IEEE Neural Network Society zmieniło w 2003 roku nazwę na IEEE Computational Intelligence Society. Prace w tej dziedzinie wspierane są przez oddział firmy Fujitsu w Polsce, FQS Poland, pomagający w integracji opracowanych przez nas algorytmów, zawartych w naszym systemie programów do data mining o nazwie GhostMiner, w innych narzędziach, np. systemach bankowych. Wystąpiliśmy z wnioskiem o grant KBN na dalszy rozwój tych metod, prace te wspierane były wcześniej grantami KBN zakończonymi w 2002 roku.

**Metody rozumienia danych** (data understanding), czy też wyjaśnienia struktur danych, to coś więcej niż statystyczne modele których celem jest osiągnięcie wysokiej dokładności. Badania te są obecnie

jedną z najważniejszych rozwiniętych przez nas specjalności w ramach tej grupy zagadnień. Należą do nich metody ekstrakcji reguł logicznych z danych treningowych za pomocą drzew decyzji i sieci neuronowych, metody oparte na wizualizacji danych i na regułach wykorzystujących podobieństwo do prototypów. Badanie powiązań pomiędzy metodami opartymi na podobieństwie, a metodami logiki rozmytej, oraz rola kontekstu w definiowaniu zmiennych lingwistycznych używanych w regułach, jest bardzo obiecującą dziedziną. Odkryte reguły można wykorzystać w procesie rozumowania.

**Teoria podobieństwa** w zastosowaniu do integracji metod inteligencji obliczeniowej w oparciu o ocenę podobieństwa przypadków referencyjnych (SBL, Similarity Based Learning), pozwala na nowy sposób rozumienia danych przez analizę ich podobieństwa do wyróżnionych prototypów (jest to temat pracy doktorskiej mgr M. Blachnika). Wymaga to optymalizacji funkcji podobieństwa, selekcji cech ze względu na które oceniane jest podobieństwo, oraz optymalizacji samych prototypów. Praca magisterska M. Grochowskiego zawiera obszerny przegląd i porównanie znanych metod selekcji prototypów. W szczególności metody optymalizacji prototypów mogą być konkurencją dla systemów neurorozmytych, pozwalając na tworzenie funkcji przynależności dla danych dowolnego typu. Nowy algorytm aktywnego uczenia SVNN (Support Vector Neural Networks) opracowany przez W. Duchę pozwolił na znaczną poprawę wyników uzyskiwanych przez sieci neuronowe dla trudnych, dużych baz danych, wyszukując niewielką liczbę wektorów brzegowych, które pozostają w końcowej fazie uczenia bardzo ją przyspieszając. Przeanalizowano też relacje pomiędzy niepewnością pomiarów a typami funkcji transferu w sieciach neuronowych, wprowadza nowe, bardzo obiecujące funkcje transferu. Praca ta pozwala zrozumieć relacje pomiędzy wielowarstwowymi perceptronami a systemami logiki rozmytej, otwiera nowe perspektywy przed modelowaniem rozmytym. Ocena podobieństwa struktur złożonych może być w prostych przypadkach wykonana za pomocą algorytmów programowania dynamicznego. Tego typu metody będą bardzo przydatne w wielu dziedzinach, od bioinformatyki do ekonomii, niestety trudno jest znaleźć dobre dane, na których można je testować – danych nie zbierano, gdyż nikt dotychczas nie potrafił takich problemów rozwiązywać.

**Systemy heterogeniczne** wprowadzają nową jakość do algorytmów inteligencji obliczeniowej. W ostatnich latach opracowaliśmy kilka systemów tego rodzaju. Należą do nich ontogeniczne sieci neuronowe, czyli algorytmy budowy i adaptacji sieci neuronowych zmieniających swoją strukturę. Dr Janowski napisał na ten temat monografię, która ukazała się w 2003 roku. Wszystkie dotychczasowe modele sieci neuronowych powielają elementy przetwarzające (neurony) tego samego typu. Heterogeniczne modele neuronowe tworzą końcowe rozwiązanie z różnych elementów, zależnie od potrzeb wynikających z analizowanych danych; w efekcie końcowy model ma szansę odkryć najprostsze struktury w tych danych. Eksperymenty z wyborem różnych funkcji transferu w czasie dodawania nowych węzłów do sieci były bardzo udane, obecnie poszukujemy (jest to temat pracy magisterskiej) efektywnych algorytmów uczenia tego typu sieci. Kilka artykułów napisaliśmy na temat heterogenicznych drzew decyzji, pracujemy też nad wersją heterogenicznego systemu opartego na podobieństwie. Celem jest lepsze zrozumienie struktur danych za pomocą prostszych systemów uczących się.

**Meta-uczenie**, czyli automatyczne tworzenie optymalnych i jak najprostszyc modeli, oparte jest na procesach szukania w przestrzeni modeli o wzrastającym stopniu złożoności. Modele takie można wygenerować w ramach spójnego schematu tworzenia metod opartych na podobieństwie, lub tworzyć z elementów systemów heterogenicznych. Mając do dyspozycji wiele różnorodnych opracowanych przez nas algorytmów próbujemy zastosować te idee do automatycznego wyboru najlepszego modelu, włączając w to wszystkie etapy jego tworzenia: transformację i selekcję danych, krosvalidację, uśrednianie wyników i wybór samego modelu i jego parametrów. Meta-uczenie jest kolejnym krokiem w stronę rozumienia danych przez automatyzację odkrywania jak najprostszyc struktur.

**Metody uczenia oparte na algorytmach szukania** są znacznie prostsze od analitycznych metod minimalizacji gradientowej. Dotychczas w zastosowaniach do sieci neuronowych rozwijano tylko stochastyczne metody szukania, podczas gdy algorytmy wykorzystujące szukanie deterministyczne (SMLP, Search-based MLP) i gradienty numeryczne są znacznie szybsze i dają doskonale rezultaty (praca doktorska M. Kordosa). Nie nakłada to żadnych ograniczeń na rodzaj funkcji transferu sieci, jest to więc bardzo ogólne podejście do optymalizacji systemów uczących się. Algorytm szukania o zmiennym kroku okazał się szybszy i dokładniejszy od tak wyrafinowanych metod jak SCG (metoda skalowanych gradientów sprzężonych) i Levenberg-Marquard (znana metoda drugiego rzędu). W 2003 roku napisano kilka prac na temat efektywności takich algorytmów używanych do uczenia sieci

MLP i do odkrywania reguł klasyfikacyjnych za pomocą sieci neuronowych. Obecnie prowadzone są obliczenia dla dużych baz danych.

**Komitet systemów klasyfikujących** to popularne zagadnienie w uczeniu maszynowym. Pracujemy tu nad dwoma zagadnieniami: integracją modeli uzyskanych na różnych partycjach w krosvalidacji (optymalizacja parametrów i sposobu ich głosowania), oraz komitetach uwzględniających kompetencję głosujących członków. Takie komitety inspirowane są przez sposób działania mózgow, które włączają tylko te moduły do oceny danych, które mogą się do tego przydać, obniżając aktywność pozostałych modułów. Członkowie komitetu – różne modele klasyfikujące – oddają swoje głosy, które skalowane są przez czynniki kompetencji w danym obszarze przestrzeni cech. Uzyskano bardzo dobre wyniki (w ramach pracy magisterskiej), metoda ta jest nadal rozwijana i zostanie zintegrowana z innymi w pakiecie programów GhostMiner.

**Selekcja cech i agregacja dostępnej informacji** jest niezbędnym krokiem we wszystkich metodach uczących się w złożonych domenach. Zagadnienia bioinformatyczne charakteryzują się często tysiącami a nawet setkami tysięcy cech, z których należy wybrać nieliczne cechy istotne dla opisu danego problemu. Badania metod selekcji cech i redukcji wymiarowości danych realizowane są przy współpracy z dr J. Biesiadą i dr A. Kachelem z Politechniki Śląskiej. Opracowano zestaw programów realizujących liczne metody typu filtrów, oraz metody selekcji wykorzystujące te filtry, oparte na algorytmach szukania. Badano nowe metody redukcji wymiarowości danych, oparte zarówno na projekcjach liniowych w oparciu o kryteria maksymalizacji miar teorio-informatycznych lub miar typu Fishera, oraz metod nieliniowych typu skalowania wielowymiarowego. W tej dziedzinie popełnia się często błędy metodologiczne, stosując metody selekcji dla całego zbioru danych, co nie zapewnia odpowiedniej generalizacji. Szereg nowych metod jest w opracowaniu a bardzo dobre rezultaty osiągnięte w najważniejszym międzynarodowym konkursie „NIPS 2003 Feature Selection Challenge” doprowadziły do zaproszenia do napisania trzech rozdziałów do powstającej po tym konkursie książki.

**Wizualizacja wielowymiarowych danych** jest kolejną metodą rozumienia danych. Opracowano nowe sposoby projekcji danych pozwalające zrozumieć relację w przestrzeniach o stosunkowo niewielkiej liczbie wymiarów (rzędu 8-10) za pomocą projekcji na siatkę wierzchołków hipersześcianu (lattice projection), oraz dwuwymiarowych projekcji wykorzystujących kryteria Fishera (W. Duch). Metody te zastosowano do analizy odwzorowań pobudzeń neuronów w przestrzeni wewnętrznej i wyjściowej typowych sieci neuronowych (MLP i RBF), co pozwala zrozumieć sposób ich działania. Wykorzystano je również do obserwacji trajektorii wielowymiarowych układów dynamicznych opisujących chaotyczne pamięci asocjacyjne (P. Matykwicz), co pozwala zaobserwować ich atraktory oraz ocenić ich baseny atrakcji. Nieliniowe mapowanie barycentryczne opracowane zostało przez współpracujących z nami studentów informatyki (F. Piękniewski, L. Rybicki). Projekcje na dwie składowe główne zastosowane zostały do analizy trajektorii parametrów adaptacyjnych (wag i progów) sieci neuronowych, co pozwoliło lepiej zrozumieć proces uczenia się w tych sieciach (M. Kordos). Skalowanie wielowymiarowe (MDS, multi-dimensional scaling) połączono z procedurami klasteryzacji typu LVQ (Learning Vector Quantization), dendrogramów lub metody k-średnich, pozwalającymi na zmniejszenie liczby centrów (codebook vectors) przeznaczonych do mapowania. Daje to mapę MDS dla położenia niewielkich klastrów, co w połączeniu z procedurą domapowywania pojedynczych wektorów pozwala na szybkie umiejscowienie nowych przypadków w otoczeniu przypadków znanych. Program pozwala na wierniejsze odwzorowanie wielowymiarowych danych niż bardzo popularny model samoorganizującej się sieci neuronowej (SOM) Kohonena.

**System GhostMiner do dogłębnej analizy danych** integruje większość rozwiniętych w naszej Katedrze algorytmów i zawiera nowatorskie rozwiązania informatyczne, zarówno w zakresie metod inteligencji obliczeniowej jak i inżynierii oprogramowania. W 2003 roku dokonano w tym pakiecie szeregu zmian, dodając nowe sposoby transformacji danych i selekcji informacji. Usprawniono zarządzanie tworzeniem i testowaniem modeli, realizowane za pomocą drzewa projektów, na którym umieszczane są modele i procedury testowania. System powstał przy współpracy z firmą FQS Poland z Krakowa (oddział firmy Fujitsu), która wspomaga prace programistyczne. Od połowy 2002 roku firma FQS rozpowszechnia ten system komercyjnie i jest on stosowany w wielu miejscach do analizy danych o różnym charakterze.

**Zastosowania naszych algorytmów** i wyników badań teoretycznych możliwe są w analizie danych dla potrzeb medycyny, nauk ścisłych i bioinformatyki. Ekstrakcja reguł logicznych z danych jest podstawą automatyzacji akwizycji wiedzy i konstrukcji systemów inteligentnego wspomaganie decyzji. Prowadzimy współpracę z Department of Biomedical Informatics, CHMC. Dr J. Meller (adiunkt KIS) pracuje tam obecnie jako asisstant professor, dr Rafał Adamczak (nasz były doktorant) jako research associate, a mgr Łukasz Itert (nasz obecny doktorant) przebywa od listopada 2002 roku na kilkuletnim stażu.

**Metody semantycznego szukania informacji** rozwijane są przy współpracy z Department of Biomedical Informatics (kierowanym przez Johna Pestiana), Children's Hospital Medical Center (CHMC) w Cincinnati, Ohio, USA, głównie pod kątem szukania informacji w zgromadzonych tam medycznych bazach danych. W 2003 roku prowadzono głównie prace nad budową korpusu tekstów, na którym trenowane będą różne systemy uczące się. Z jednej strony do tekstów dodano nazwy kategorii gramatycznych (POS tagger), częściowo automatycznie a częściowo robiła to grupa lingwistów, a z drugiej strony szukano reguł pozwalających na osiągnięcie jednoznaczności skrótów i akronimów (potrzeba około 900 takich reguł). Oprócz metod inteligencji obliczeniowej stosowane będą również metody lingwistyczne. W kolejnym kroku robione będą automatyczne adnotacje tekstów przy wykorzystaniu ontologii Unified Medical Language System (UMLS) i XML (Xtended Markup Language). Tworzeniem korpusu zajmuje się Ł. Itert (doktorant W. Ducha). Opracowano też wstępnie interesującą ideę połączenia metod analizy tekstów z modelami pamięci rozpoznawczej, semantycznej i epizodycznej. W rozszerzonym projekcie pracować będzie P. Matykiewicz (również doktorant W. Ducha). Badania te należą częściowo do grupy badań kognitywnych. Przewidujemy znaczny rozwój tej problematyki: kilka tematów z tej grupy zaproponowaliśmy w projekcie utworzenia Centrum Doskonałości (przy Instytucie Podstaw Informatyki PAN), do którego chcemy się przyłączyć.

**Działalność organizacyjna** w organizacjach międzynarodowych świadczy o pozycji naszej grupy w dziedzinie inteligencji obliczeniowej. W. Duch był w 2003 roku:

- członkiem Executive Board, European Neural Network Society (ENNS),
- członkiem Board of Governors, International Neural Networks Society (wśród 3 osób spoza USA),
- członkiem IEEE Neural Networks Society Technical Comittee,

a więc jedyną osobą, która została wybrana na członka wszystkich trzech najważniejszych w tej dziedzinie organizacji na świecie. W 2003 roku był też członkiem redakcji wielu najważniejszych pism specjalistycznych w tej dziedzinie:

- IEEE Transactions on Neural Networks,
- Neural Information Processing Letters and Reviews
- International Journal of Computational Intelligence
- International Journal of Signal Processing
- Machine Graphics and Vision
- Computer Physics Communications

W. Duch bierze udział w kilku największych odbywających się regularnie konferencjach dotyczących inteligencji obliczeniowej i jest członkiem ich komitetów naukowych, w tym:

- World Congress on Computational Intelligence (WCCI)
- International Conference On Neural Information Processing (ICONIP)
- International Joint Conference on Neural Networks (IJCNN)
- International Conference on Artificial Neural Networks (ICANN)
- European Symposium on Artificial Neural Networks (ESANN)
- Engineering Applications of Neural Networks (EANN)

- International Conference on Artificial Intelligence and Soft Computing (ICAISC)
- Intelligent Information Systems (IIS-IIPWM)
- International Conference on Intelligent Sensing and Information Processing (ICISPI)
- International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA)

i wielu innych.

Tematy należące do tego kierunku rozwijane są obecnie przy współpracy z kilkoma ośrodkami zagranicznymi:

- Department of Biomedical Informatics, Cincinnati, Ohio, prof. John Pestian;
- Department of Computer Science, National University of Singapore, prof. R. Setiono;
- School of Computer Engineering, Nanyang Technological University, prof. G-S. Ng, J. Rajaspake;
- Department of Computer Science, Univ. of Louisville, Kentucky, USA, prof. Jackiem Żuradą;
- Department of Computer Science, Meiji University, Tokyo, prof. T. Takagi i Y. Hayashi;
- Department of Life Sciences, Kyushu Institute of Technology, prof. M. Ishikawa;
- Max Planck Inst. of Astrophysics, Garching (Monachium), prof. G.H.D. Dierksen;

**Słowa kluczowe i dyscypliny pokrewne:** inteligencja obliczeniowa, sztuczna inteligencja, sieci neuronowe, logika rozmyta, reguły logiczne, dogłębna analiza danych, systemy neurorozmyte, rozpoznawanie obiektów, systemy eksperckie, biocybernetyka, bioinformatyka, data mining, inteligentne wspomaganie decyzji, optymalizacja, analiza obrazów, wizualizacja danych.

## II. Zastosowanie metod komputerowych do problemów fizyki, chemii i biologii molekularnej.

- Prof. J. Wasilewski z współpracownikami prowadził badania w ramach tematu: „Struktura energetyczna małych cząsteczek”, zajmując się zastosowaniami teorii funkcjonałów gęstości (DFT – Density Functional Theory) do obliczeń struktury energetycznej i widm elektronowo-oscylicyjnych cząsteczek. Uczestniczył też w badaniach kierowanych przez prof. K. Jankowskiego z Instytutu Fizyki UMK, dotyczących nowych metod opisu efektów korelacji elektronowej w stanach niezamknięto-powłokowych atomów i cząsteczek.
- Dr J. Meller prowadził badania nad zastosowaniem metod obliczeniowych typu dynamiki molekularnej oraz metod bioinformatyki rozpoznawania struktur i funkcji białek, metod dopasowania sekwencji w biologii molekularnej i genomice. Prace te publikowane są w najlepszych pismach bioinformatycznych. W 2003 roku dr Meller pracował jako *assistant professor* w Children’s Hospital Research Foundation, zajmując się przede wszystkim bioinformatyką. Jego praca habilitacyjna broniąca będzie w połowie 2004 roku przez Radą Naukową Instytutu Biocybernetyki i Inżynierii Biomedycznej PAN w Warszawie.
- Prof. W. Duch prowadził badania nad zastosowaniem metod inteligencji obliczeniowej do analizy danych z mikromacierzy DNA oraz danych z proteomicznych widm spektrografów masowych. Dr Jankowski i dr Grąbczewski dokonali analizy bardzo dużych zbiorów danych w ramach konkursu metod selekcji cech na konferencji NIPS 2003.

Nasza Katedra współorganizowała w 2003 roku po raz drugi 3-dniowe warsztaty ogólnopolskie z udziałem J. Meller i kilku innych wykładawców z UMK i innych uczelni. Prace nad zastosowaniami naszych metod w bioinformatyce rozwijają się powoli, pomimo tego, że jedna z najważniejszych firm w tej dziedzinie, Abbott Laboratories (USA), zakupiła stworzony przez nas system GhostMiner. Dwie próby otrzymania grantu KBN na rozwój metod inteligencji obliczeniowej pod kątem ich zastosowań w bioinformatyce skończyły się niepowodzeniem ze względu na ocenę recenzenta-biologa. W naszej

ocenie dziedzina ta została zmonopolizowana przez biologów i biofizyków i szanse na prowadzenie w tej dziedzinie badań informatycznych są niewielkie, z wielką szkodą dla nauki polskiej.

**Słowa kluczowe i dyscypliny pokrewne:** bioinformatyka, proteomika, metody komputerowe w fizyce i chemii, symulacje komputerowe, fizyka teoretyczna, fizyka komputerowa, chemia kwantowa, chemia teoretyczna, chemia komputerowa, dynamika molekularna.

### **III. Kognitywistyka, neuroinformatyka, komputerowe modele procesów poznawczych i filozofia umysłu.**

Kognitywistyka (nauki o poznaniu) powoli rozwija się w Polsce w odrębną dyscyplinę badawczą. Należy podkreślić, że interdyscyplinarne badania w tej dziedzinie należą do priorytetowych w ramach programów ramowych Unii Europejskiej. Niedorozwój tej dziedziny w Polsce związany jest ze złą organizacją nauki, uniemożliwiająca robienie doktoratów wymagających badań interdyscyplinarnych, które nie znajdują się na oficjalnej liście uprawianych u nas dziedzin nauki. W. Duch był w 2003 roku (i w latach poprzednich) recenzentem tego typu projektów w ramach VI Programu Ramowego. Jesteśmy jedną z nielicznych jednostek w kraju, w której prowadzi się badania kognitywne.

W. Duch jest jednym z członków-założycieli i członków zarządu Polskiego Towarzystwa Kognitywistycznego. Przewodniczy też komitetowi naukowemu pisma „Kognitywistyka i Media w Edukacji” (Wyd. A. Marszałek), które wychodzi przynajmniej dwa razy do roku (każdy numer ma ponad 300 stron). Numery specjalne publikowane są w języku angielskim. Jest to oficjalne pismo Polskiego Towarzystwa Kognitywistycznego. Ukazało się już 6 tomów. Otrzymaliśmy bardzo dobre recenzje z pierwszych lat działalności naszego pisma napisane przez wybitnych specjalistów z filozofii, biocybernetyki, pedagogiki, psychologii i neurobiologii. W maju 2003 zorganizowaliśmy na UMK drugą krajową Konferencję i Zjazd Towarzystwa.

W 2003 roku trzy jednostki UMK, Katedra Informatyki Stosowanej, Zakład Technologii Kształcenia (Inst. Pedagogiki) i Zakład Teorii Poznania i Metodologii Nauk utworzył grupę inicjatywną i otrzymał grant JM Rektora UMK w celu utworzenia Centrum Kognitywistyki. Dążymy do powołania interdyscyplinarnego kierunku studiów doktoranckich w zakresie kognitywistyki. Nasze badania prowadzone są przy współpracy z tymi Zakładami. Pracujemy nad empirycznie uzasadnioną teorią umysłu w oparciu o modele funkcji mózgu na poziomie systemowym jak i na poziomie modelowania neuronowego. Teoria ta powinna nie tylko być pomocną w rozwiązaniu podstawowych problemów nauk o poznaniu, lecz również prowadzić do nowych konstrukcji systemów uczących się, i w tym kierunku zmierzają nasze badania.

Zarysy takiej teorii, w postaci hipotez określających prawdopodobne związki pomiędzy zdarzeniami w mózgu i ich korelacji z percepcją, wyższymi czynnościami poznawczymi i treścią świadomości zostały już nakreślone w kilku pracach i wystąpieniach konferencyjnych, zarówno o charakterze filozoficznym jak i bardziej technicznym. W 2003 roku wysłano do druku pracę „Brain-inspired conscious computing architecture” do Journal of Mind and Behavior (W. Duch jest członkiem redakcji tego pisma), związaną z modelowaniem umysłu i świadomością w układach sztucznych. W. Duch został poproszony o napisanie dwóch komentarzy dla Brain and Behavioral Science, jednego z najważniejszych pism w tej dziedzinie. W. Duch jest też członkiem rady redakcyjnej International Journal of Transpersonal Studies.

Kontynuowaliśmy współpracę z filozofami umysłu w Polsce (UAM Poznań, IFiS PAN Warszawa) i za granicą, biorąc udział w konferencjach, seminariach i dyskusjach na temat fundamentalnych problemów dotyczących natury świadomości. Utrzymujemy też kontakty z Instytutem Biologii Eksperymentalnej im. M. Nenckiego w Warszawie. Utrzymywaliśmy też kontakty z prof. Johnem Taylorem z Center for Neural Networks, King's College, London, na temat modelowania mechanizmów uwagi za pomocą metod obliczeniowej neurobiologii kognitywnej, którego w 2003 roku odwiedzał P. Matykiewicz, doktorant W. Ducha. Niestety współpraca z tą grupą okazała się trudna ze względu na plany komer-

cializacji pewnych idei przez prof. Taylora. P. Matykiewicz napisał dwie prace o chaotycznych modelach pamięci asocjacyjnej wprowadzając interesujące modyfikacje w układach neurodynamicznych, dzięki czemu tak skonstruowane sieci neuronowe mogą wędrować pomiędzy dwoma obiektami nałożonymi na siebie, rozpoznając obydwa.

Nowym tematem jest rozwój zhumanizowanych interfejsów (HIT, Humanized InTerfaces) integrujących technologie awatarów (mówiącej głowy), syntezę i rozpoznawanie mowy, umożliwiających naturalną interakcję ze stronami internetowymi. Jednym z zastosowań będą gry słowne, takie jak np. gra w 20 pytań. Gry słowne były niegdyś ważnym elementem zabaw sprzyjającym myśleniu analitycznemu. Rozpowszechnienie się podręcznych komputerów i inteligentnych telefonów będzie sprzyjać popularności tego rodzaju technologii. Prowadziliśmy rozmowy w sprawie utworzenia laboratorium gier słownych w Computer Engineering School, NTU, Singapur, gdzie przy współpracy z KIS prawdopodobnie uda się je powołać w przyszłym roku. Zagadnienie to wiąże się również z badaniami nad strukturą przestrzeni semantycznej i jest tematem pracy dwóch doktorantów zaocznych W. Ducha (J. Szymańskiego i T. Sarnatowicza).

W. Duch podał zarys teorii współpracy złożonych sieci agentów programowych, rozważając coraz bardziej złożone sieci oddziaływujących elementów. Przejście od prostych sieci neuronowych do sieci złożonych oddziaływujących agentów dokonać można zwiększając lokalną wiedzę i złożoność oddziaływań, dlatego warto konstruować systemy o pośredniej złożoności, których na razie brakuje.

**Słowa kluczowe i dyscypliny pokrewne:** kognitywistyka, modele umysłu, filozofia umysłu, neuroinformatyka, układy złożone, układy dynamiczne, dynamika symboliczna, teoria informacji, symulacje komputerowe, matematyka stosowana, biocybernetyka, sieci neuronowe, reprezentacje mentalne.

#### **IV. Dydaktyka wspomagana metodami komputerowymi.**

Badania w tym zakresie prowadzone są przez zatrudnionych w naszej Katedrze wykładowców i profesorów. Bierzymy udział zarówno w restrukturyzacji programów nauczania w zakresie specjalizacji komputerowych i przedmiotów o charakterze informatycznym na Wydziale Fizyki i Astronomii UMK (prof. Wasilewski), w tworzeniu nowych programów na uczelni przy współpracy z Centrum Kształcenia Komputerowego, programów studiów podyplomowych oraz w tworzeniu nowych programów nauczania dla regionu.

W 2003 roku prowadzono eksperymenty z nauczaniem na odległość pomiędzy Ann Arbor (USA), Singapurem i Toruniem. W. Duch będąc w Singapurze wygłosił referat na konferencji Polskiego Towarzystwa Kognitywistycznego w Toruniu korzystając z systemu zdalnego nauczania Centra, biorąc też udział w dyskusji pomiędzy trzema ośrodkami. Otworzył też Poznańskie Seminarium Kognitywistyczne, wygłaszając ponad godzinny referat. Planujemy więcej eksperymentów z nauczaniem na odległość pomiędzy Singapurem i Polską.

Utrzymujemy strony WWW dotyczące wszystkich spraw związanych z komputerami, fizyką komputerową, sieciami neuronowymi, sztuczną inteligencją, kognitywistyką, badaniami nad mózgiem. Ze stron tych korzysta bardzo wiele osób. W ramach eksperymentów z nauczaniem na odległość stworzono dostępne w Internecie wersje notatek do kilku wykładów w formacie HTML i PPT. Niestety tworzenie takich materiałów jest czasochłonne i brakuje środków na wsparcie prac w tym zakresie.

**Słowa kluczowe i dyscypliny pokrewne:** multimedia w nauczaniu, nauczanie na odległość, dydaktyka fizyki, dydaktyka matematyki, dydaktyka chemii teoretycznej, pedagogika, fizyka komputerowa, symulacje środowiska naturalnego.

**C2. Informacja o zawartych i realizowanych przez jednostkę umowach dotyczących wykonywania zadań badawczych wspólnie z podmiotami krajowymi lub zagranicznymi, w tym zadań w ramach programów ramowych Unii Europejskiej.**

Nie prowadzimy w tej chwili badań w oparciu o takie umowy; podejmowaliśmy następujące próby:

- Przyłączyliśmy się do wniosku grupy irlandzkiej (Univ. of Ulster) o finansowanie projektu „Breeding creative information societies” w ramach programu Unii Europejskiej „Future Emerging Technologies” (FET), ale nie został on zakwalifikowany.
- W kwietniu 2003 roku złożyliśmy wspólnie z 31 innymi grupami projekt w ramach „Excellence Network, 6FP” w zakresie informatyki medycznej, ale program ten nie został zakwalifikowany.
- Firma DaimlerChrysler zwróciła się do nas z prośbą o udział w projekcie „One-shot learning”, który zarejestrowany został na stronie projektów FET serwera Unii Europejskiej. Bierze w nim udział tylko 4 partnerów (łącznie z nami), projekt jest nadal w przygotowaniu.
- Firma Bayer Diagnostics (z centrum w Terrytown pod Nowym Jorkiem), zajmująca się między innymi produkcją aparatury do testów hematologicznych, przysłała nam dane do analizy i utrzymuje z nami kontakt. Na razie nie udało się utworzyć sieci w ramach VI programu ramowego UE do gromadzenia i analiz danych hematologicznych, do której zaprosili nas hematolodzy z Uniwersytetu w Rzymie, ale sprawa jest w toku.
- Od 6 lat prowadzimy współpracę z firmą FQS Poland, która jest oddziałem Fujitsu w Polsce, finansującą rozwój narzędzi do data mining, opracowanych w naszej Katedrze.



## **D. INFORMACJA ROCZNA O MERYTORYCZNYCH I FINANSOWYCH WYNIKACH DZIAŁALNOŚCI STATUTOWEJ JEDNOSTKI W ROKU POPRZEDZAJĄCYM ROK ZŁOŻENIA WNIOSKU**

1. Syntetyczny opis merytoryczny zrealizowanych zadań wg planu zadaniowo-finansowego, w tym:
  - 1) wykonanych badań naukowych i prac rozwojowych,
  - 2) najważniejszych osiągnięć poznawczych,
  - 3) najważniejszych zastosowań praktycznych, w tym opracowanych nowych: technologii, materiałów, wyrobów,
  - 4) wdrożeń dokonanych poza jednostką
  - 5) uzyskanych nagród o znaczeniu międzynarodowym i ogólnokrajowym,
  - 6) innych ważniejszych osiągnięć.

### **Opis merytoryczny prac realizowanych w roku 2003**

**Finansowanie:** działalność statutowa, badania własne, współpraca międzynarodowa. W ramach działalności statutowej pracowaliśmy nad następującymi tematami:

#### **I. Teoria i zastosowania metod inteligencji obliczeniowej.**

Kierownik prof. W. Duch, finansowanie – badania statutowe.

W ramach rozwoju teorii inteligencji obliczeniowej, a szczególnie modeli sieci neuronowych, metod opartych na podobieństwie, metod rozpoznawania wzorców i wizualizacji, oraz metod sztucznej inteligencji opartych na logicznym opisie danych pracowaliśmy nad następującymi tematami.

Główne osiągnięcia w tej grupie tematów to:

1. Zakończenie pracy doktorskiej przez Krzysztofa Grąbczewskiego, „Zastosowanie kryterium separowalności do generowania reguł klasyfikacji na podstawie baz danych”, Rozprawa doktorska. Instytut Badań Systemowych PAN, marzec 2003. Praca została obroniona w czerwcu 2003 roku.
2. Ukazanie się monografii: Norbert Jankowski, Ontogeniczne sieci neuronowe. Kontrola złożoności sieci neuronowych. Akademicka Oficyna Wydawnicza, Warszawa 2003, str. 312.
3. W międzynarodowym konkursie „NIPS 2003 Feature Selection Challenge” system programów GhostMiner, którego autorami są K. Grąbczewski i N. Jankowski zajęli 3 miejsce (na 78 uczestników) w konkursie selekcji cech dla zadań klasyfikacji. NIPS, czyli Neural Information Processing Systems, jest najpoważniejszą konferencją w tej dziedzinie.
4. Praca W. Ducha, R. Setiono (Signapur) i J. Zurady (USA) ukazała się na okładce numeru majowego 2004 pisma *Proceedings of IEEE*, najbardziej prestiżowego pisma IEEE.
5. Przy współpracy z oddziałem firmy Fujitsu w Polsce, FQS Poland, N. Jankowski i K. Grąbczewski rozwinęli pakiet programów GhostMiner 2.0, oparty w większości o algorytmy do data mining opracowane w naszej Katedrze.

6. Przy współpracy z J. Biesiadą z Politechniki Śląskiej opracowano pakiet programów realizujący kilkanaście metod przydatnych do dyskretyzacji i selekcji cech.
7. Opracowano bardzo interesujący algorytm aktywnego uczenia, SVNN (Support Vector Neural Networks), który pozwolił na znaczną poprawę wyników uzyskiwanych przez sieci neuronowe dla trudnych, dużych baz danych.
8. Dokonano znacznego postępu w zrozumieniu mechanizmów działania sieci neuronowych dzięki wizualizacji aktywności warstwy ukrytej oraz wyjściowej w wielu wymiarach, oraz wizualizacji wielowymiarowych powierzchni błędów (M. Kordos).
9. Przeanalizowano relacje pomiędzy niepewnością pomiarów a typami funkcji transferu w sieciach neuronowych; praca ta otwiera nowe perspektywy przed modelowaniem rozmytym, pozwala zrozumieć relacje pomiędzy wielowarstwowymi perceptronami a systemami logiki rozmytej, oraz wprowadza nowe, bardzo obiecujące funkcje transferu (W. Duch).
10. Opracowano duży korpus tekstów medycznych, które zostały opisane przez lingwistów, po czym zawarte w nich słowa są automatycznie opisywane (tagowane, czyli określane są części mowy), w celu mapowania zawartości na ontologię medyczną UMLS i automatycznego przekształcenie tekstu na XML, co pozwoli na inteligentne przeszukiwanie tekstów medycznych.
11. Dokonano obszernego przeglądu i porównania znanych metod selekcji prototypów (praca magisterska M. Grochowskiego).
12. W. Duch wygłosił 3-godzinny tutorial „Computational intelligence tools for data understanding”, na połączonych konferencjach International Conference on Artificial Neural Networks (ICANN), and International Conference on Neural Information Processing (ICONIP), Istanbul, Turkey.
13. Zredagowano numer międzynarodowego kwartalnika TASK Quarterly, Vol. 1/2003, „Artificial and Computational Intelligence” (W. Duch i D. Rutkowska).

## I.1. Metody rozumienia struktur danych.

Jest to część bardzo modnej obecnie dziedziny, określanej jako „data mining”, czyli dogłębna analiza danych, lub „knowledge discovery in databases” (KDD), czyli odkrywania wiedzy w bazach danych. Szukanie wiedzy w bazach danych (data mining) nabiera coraz większego znaczenia i na wielu konferencjach organizuje się sesje specjalne poświęcone temu zagadnieniu. Temat ten będzie przez naszą grupę nadal intensywnie rozwijany. W roku 2003 wykonaliśmy następujące prace:

I.1.1 Praca doktorska Krzysztofa Grąbczewskiego, „Zastosowanie kryterium separowalności do generowania reguł klasyfikacji na podstawie baz danych”, napisana pod opieką W. Duchy, obroniona została przed komisją Instytut Badań Systemowych PAN, w czerwcu 2003 roku. Podano w niej kryterium separowalności (SSV, Separability Split Value) danych, które zostało zastosowane do: sprawnego tworzenia drzew decyzyjnych; ekstrakcji reguł decyzji z tych drzew; tworzenia „lasu drzew decyzyjnych”, o podobnej dokładności, ale różniących się między sobą wykorzystywanymi cechami jak i dokładnością (czułością i wrażliwością) dla specyficznych klas; tworzenia reguł decyzji opartych na funkcjach odległości, co daje bardzo proste reguły prototypowe; rankingu i selekcji cech; dyskretyzacji danych; oraz uciągania danych.

1. Krzysztof Grąbczewski, „Zastosowanie kryterium separowalności do generowania reguł klasyfikacji na podstawie baz danych”, Rozprawa doktorska. Instytut Badań Systemowych PAN, marzec 2003.

I.1.2. W. Duch napisał obszerną pracę „Computational intelligence methods for rule-based data understanding” na zamówienie *Proceedings of IEEE*, najbardziej prestiżowego pisma IEEE (współautorami są R. Setiono, Singapur, i J. Żurada, USA). Praca ukazała się na okładce numeru majowego 2004 tego pisma.

2. Duch W, Setiono R, Żurada J, Computational intelligence methods for rule-based data understanding. Proc. of the IEEE (w druku, ukazało się w maju 2004)

I.1.3 Rozwijaliśmy algorytmy uczenia sieci neuronowych oparte na deterministycznych metodach szukania optymalnych dyskretnych parametrów opisujących sieć neuronową (M. Kordos). Obecne testy pokazują, że będą to metody konkurencyjne ze względu na dokładność i szybkość działania w stosunku do najlepszych metod gradientowych, a jednocześnie znacznie prostsze w implementacji. Można za ich pomocą generować również reguły logiczne i udało się w ten sposób otrzymać interesujące, proste reguły. Ponieważ do oceny perturbacji parametrów wykorzystuje się tylko informację o liczbie błędów jest to podejście ogólne, szczególnie przydatne w poszukiwaniu ostrych reguł logicznych, gdy obszar niezerowych gradientów kurczy się do zera i sieci z wsteczną propagacją błędów nie uczą się.

3. Kordos M, Duch, W, Search-based training for logical rule extraction by Multilayer Perceptron. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, pp. 86-89

I.1.4. W przypadku złożonych granic decyzji reguły logiczne mogą nie być wystarczająco dokładne lub wystarczająco proste. Alternatywnym sposobem rozumienia struktur danych jest użycie sieci neuronowej, pod warunkiem, że rozumiemy co ta sieć robi. Jest to możliwe dzięki wizualizacji aktywności jednostek wyjściowych i jednostek ukrytych. Wizualizacja aktywności jednostek wyjściowych pozwala poza tym poznać szczegółowe cechy mapowania, którego nauczyła się sieć, porównywać sieci o tych samych liczbach błędów, oceniać wiarygodność klasyfikacji, zdolność do generalizacji, efekty zbytniego dopasowania się do danych i niedostatecznej liczby neuronów, szanse na osiągnięcie zbieżności i wiele innych rzeczy. Dłuższa praca na ten temat znajduje się w przygotowaniu.

4. Duch, W, Coloring black boxes: visualization of neural network decisions. International Joint Conference on Neural Networks, Portland, Oregon, 2003, Vol. I, pp. 1735-1740

I.1.5. Zbadano relacje pomiędzy niepewnością danych, funkcjami sigmoidalnymi w sieciach neuronowych i funkcjami przynależności systemów rozmytych. Doprowadziło to do zrozumienia przyczyn, dla których funkcje sigmoidalne pełnią tak ważną rolę w sieciach neuronowych, nowego spojrzenia na sieci jako systemy oceniające prawdopodobieństwo spełnienia pewnej reguły, a także wyprowadzenia nowych funkcji transferu.

5. Duch W, Uncertainty of data, fuzzy membership functions, and multi-layer perceptrons. IEEE Transactions on Neural Networks (w druku)

## I.2. Teoria podobieństwa.

W ostatnich latach opracowaliśmy i rozszerzyliśmy ogólny schemat dla metod opartych na podobieństwie (Similarity-Based Methods, SBM), obejmujący jako szczególne przypadki większość znanych metody rozpoznawania obiektów, takich jak różne warianty k-NN czy kwantyzacji wektorowej LVQ, wiele modeli sieci neuronowych, w tym najbardziej popularne modele MLP i RBF, oraz wiele innych metod. Schemat ten oparty jest na parametryzacji prawdopodobieństw *a posteriori* za pomocą funkcji podobieństwa, sposobu wyboru prototypowych wektorów, sposobu ważenia ich wpływu na obliczane prawdopodobieństwo, liczby wykorzystywanych wektorów prototypowych, sposobu ich wyboru i optymalizacji ich położenia, sposobu wyboru cech do klasyfikacji, sposobu mieszania różnych modeli ze sobą i wielu innych procedur. Opisałiśmy zastosowania tego schematu w zagadnieniach klasyfikacji, aproksymacji, asocjacji i szukania brakujących danych.

I.2.1 Pracowaliśmy teraz nad metodami pozwalającymi na zrozumienie danych w oparciu o charakterystyczne prototypy. Pozwala to na wprowadzenie nowego rodzaju reguł opartych na podobieństwie do prototypów. Dokonano przeglądu algorytmów, implementacji i systematycznego porównania tych metod ze względu na ich liczbę i dokładność klasyfikacji.

1. Grochowski M, Wybór wektorów referencyjnych dla wybranych metod klasyfikacji. Praca magisterska, UMK Torun 2003

I.2.2 Opracowano metodę generacji funkcji przynależności logiki rozmytej odpowiadających różnym metrykom (M. Błachnik). Prowadzi to do tworzenia funkcji przynależności na podstawie danych, zamiast powszechnie stosowanej metody przyjmującej *a priori* proste funkcje przynależności, których

parametry są następnie optymalizowane. Przyjęcie miary odległości w formie addytywnych przyczynków zależnych od pojedynczych cech (przykładem jest miara odległości Manhattan) pozwala na bezpośrednią interpretację takich reguł w sensie logiki rozmytej. Reguły klasyczne otrzymać można dla używając metryki Chebyscheva. Można też wykorzystać probabilistycznie określone miary podobieństwa i znaleźć reguły dla danych symbolicznych, dla których metody logiki rozmytej nie działają. Reguły prototypowe można generować można będzie za pomocą wielu metod inteligencji obliczeniowej: sieci neuronowych, drzew decyzyj i metod opartych na podobieństwie. Kolejnym krokiem będzie zbadanie możliwości optymalizacji prototypów przez optymalizację ich własności.

2. Duch W, Blachnik M, Deriving membership functions from similarity measures. International Conference on Neural Information Processing 2004, wysłane.

I.2.3 Pracowaliśmy również nad połączeniem metod najbliższych sąsiadów z regularyzowanymi metodami oceny gęstości rozkładu prawdopodobieństwa, a więc pełną metodą Bayesa. Jest to temat pracy magisterskiej M. Szkulmowskiego, które powinna zostać zakończona w połowie 2004 roku.

### I.3. Integracja metod inteligencji obliczeniowej i nowe modele systemów uczących się.

Znanych jest obecnie wiele metod inteligencji obliczeniowej, wyrastających z różnych dziedzin: uczenia maszynowego, sieci neuronowych, algorytmów ewolucyjnych, logiki klasycznej i romytej, rozpoznawania wzorców, statystyki. Każde środowisko ma inne tradycje stosuje własne metody, stąd stosunkowo niewielkie przenikanie się idei opracowanych w pokrewnych dziedzinach. Naszym celem jest integracja metod uczenia maszynowego (ML, machine learning) i statystycznych metod rozpoznawania obiektów z metodami typu sieci neuronowych, opracowanie nowych metod, wynikających z takiego zintegrowanego ujęcia, oraz automatyzacja tworzenia optymalnych modeli do konkretnych zastosowań.

I.3.1. Sieci ontogeniczne dopasowują swoją złożoność do danych pozwalając na znalezienie nieobciążonych modeli klasyfikatorów o minimalnej wariancji. Systemy heterogeniczne używają różnych elementów w celu odkrycia najprostszyc struktur w danych. Naszym celem jest połączenie możliwości tych dwóch rodzajów sieci w celu stworzenia modeli działających automatycznie, bez konieczności doboru architektury sieci, tworzących modele o minimalnej złożoności.

W ramach prac nad tymi zagadnieniami w 2003 roku N. Jankowski przygotował i opublikował monografię na temat sieci ontogenicznych, opisując w niej również własne modele, sieć IncNet raz sieci realizujące algorytm Optimal Transfer Functions.

1. Norbert Jankowski, Ontogeniczne sieci neuronowe. Kontrola złożoności sieci neuronowych. Akademicka Oficyna Wydawnicza, Warszawa 2003, str. 312.

I.3.2. Systemy heterogeniczne (idea wprowadzona przez nas) to sieci neuronowe wykorzystujących funkcje transferu różnego typu, drzewa decyzyj oparte na różnych typach testów, oraz systemy oparte na podobieństwie wykorzystujące różne funkcje podobieństwa dla różnych prototypów. Wstępne testy dla heterogenicznych drzew decyzyj dały bardzo obiecujące wyniki. Jest tu nadal dużo wyzwań i otwartych problemów. Zagadnienie budowy sieci heterogenicznych jest obecnie tematem pracy magisterskiej (M. Wiśniewski). Wiąże się to też z eksperymentami z użyciem nowych funkcji transferu, które prowadzimy od dłuższego czasu.

I.3.3 Ponieważ nie istnieje jeden systemu, który zawsze najlepiej działa na wszystkich możliwych danych, należy w każdym przypadku szukać najlepiej działającej metody wśród tych, które można wygenerować w schemacie metod opartych na podobieństwie. Korzystając ze schematu metod opartych na podobieństwie rozwinęliśmy **algorytm meta-uczenia**, czyli automatycznego konstruowania optymalnej metody mieszczącej się we wspólnym schemacie, metody najlepiej działającej dla konkretnej bazy danych. Oprócz możliwości wyboru liczby sąsiadów, wyboru i optymalizacji wielu funkcji odległości, można w nim dokonywać selekcji cech i wektorów referencyjnych (prototypów). Optymalizacja miary podobieństwa obejmuje skalowanie odległości i minimalizację funkcji kosztów za pomocą

metody symulowanego wyżarzania i metody wielosympleksowej. W 2003 roku pracowaliśmy nad bardziej ogólnym algorytmem przydatnym do wyboru najprostszyc i jednocześnie możliwie najdokładniejszych transformacji danych, metod selekcji, metod klasyfikacji i ich parametrów.

Krokiem w kierunku automatycznego doboru metod i parametrów jest praca o optymalnych hiperparametrach dla metody SVM.

3. N. Jankowski and K. Grąbczewski, Toward optimal SVM. In The Third IASTED International Conference on Artificial Intelligence and Applications, pp. 451-456, Anaheim, Calgary, Zurich, September 2003. The International Association of Science and Technology for Development, ACTA Press.

I.3.4 Sieci neuronowe uczone są prawie zawsze metodą wstecznej propagacji błędów. W metodzie tej gradienty ocenia się za pomocą analitycznych wzorów, działających dla funkcji o ciągłych pochodnych. Alternatywnym podejściem jest wymyślona przez nas metoda oceniająca gradienty w sposób numeryczny. Jej wielką zaletą jest prostota implementacji a rezultaty wydają się być obiecujące.

4. Kordos M, Duch, W, Multilayer Perceptron Trained with Numerical Gradient. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, pp. 106-109

I.3.5 Niektóre algorytmy, takie jak sieci neuronowe, zorientowane są na dane ciągłe, lub też dane dyskretne o wartościach uporządkowanych. Przekształceniu danych symbolicznych na dane ciągłe poświęcona była praca:

5. Grąbczewski K, Jankowski N, Symbolic data transformations for continuous data oriented models. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, Lecture Notes in Computer Science 2714, pp. 359-366

I.3.6. Metody wektorów wsparcia zdobyły wielką popularność. Podobne metody aktywnego uczenia można zastosować w klasycznych algorytmach sieci typu MLP lub RBF. Udało się dzięki temu zredukować liczbę wektorów uczących do mniej niż 1% wszystkich wektorów, co oczywiście przyspiesza wykonywanie obliczeń, wystrzegając znacznie granice decyzji. Dla dwóch trudnych zbiorów danych osiągnięto w ten sposób najlepsze dotychczas znane wyniki.

6. Duch W, Support Vector Neural Training. IEEE Transactions on Neural Networks (wysłane)

I.3.7 Rozwijaliśmy też metody integracji różnych modeli w oparciu o ideę komitetów modeli. Jest to obecnie duża poddziedzina metod inteligencji obliczeniowej, organizowane są konferencje, numery specjalne a nawet towarzystwa, zajmujące się tylko tym tematem. Celem jest nie tylko poprawa wyników dla złożonych problemów z którymi nie radzi sobie pojedynczy model, ale i zmniejszenie wariancji wyników modelu. Opracowano nową, bardzo obiecującą wersję komitetów „lokalnie kompetentnych modeli”. Testy dały bardzo dobre rezultaty. Inne podejście do tworzenia komitetu modeli (N. Jankowski, K. Grąbczewski, w przygotowaniu) wykorzystuje modele utworzone na różnych partycjach w procesie krosvalidacji.

7. Duch W, Itert L, Committees of Undemocratic Competent Models. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, pp. 33-36

I.3.8. P. Matykiewicz (doktorant W. Ducha) napisał dwie wysłane do druku prace na temat chaotycznych pamięci asocjacyjnych, które zdolne są do rozdzielenia nałożonych na siebie wzorców.

## **I.4. Selekcja cech i agregacja informacji.**

I.4.1 W międzynarodowym konkursie „NIPS 2003 Feature Selection Challenge” system programów GhostMiner, którego autorami są K. Grąbczewski i N. Jankowski zajął 3 miejsce (na 78 uczestników) w konkursie selekcji cech dla zadań klasyfikacji. NIPS, czyli Neural Information Processing Systems, jest najpoważniejszą coroczną konferencją w tej dziedzinie, o wyjątkowo wysokim progu akceptacji (jedynie 20% nadesłanych prac). Bazy danych były bardzo duże, a liczba cech sięgała 100 tysięcy. Osiągnięcie dobrych wyników wymagało opracowania metodologii selekcji zorientowanej na generalizację traktując selekcję łącznie z klasyfikacją. W przygotowaniu są trzy prace.

I.4.2 Prowadziliśmy prace nad porównaniem różnych metodami rankingu cech, selekcji i agregacji dostępnej informacji w celu redukcji wymiarowości. Dokonano przeglądu literatury i opracowano zestaw programów realizujących liczne znane metody. Rozwinięto kilka nowych metod selekcji w oparciu o alternatywne indeksy oceny jakości cech. Przeprowadzono bardzo wiele testów mających na celu określenie wpływu dyskretyzacji na wyniki, ustalenie błędów związanych z dyskretyzacją w oszacowaniu ilości informacji łącznej, warunkowej i wzajemnej, i innych wielkości. Opracowano model sztucznych danych, dla których można jednoznacznie porównać różne metody. Przeprowadzono wiele testów i wykonano obliczenia za pomocą szeregu metod na podzbiorach cech otrzymanych z różnych metod selekcji. Reguły logiczne wymagają wstępnej dyskretyzacji danych. Opracowano specjalną wersję programu do celów dyskretyzacji i pokazano, że informacja wzajemna (mutual information), obliczona pomiędzy klasami i cechami, jest dla realnych zbiorów danych ponad dwukrotnie większa niż w przypadku prostych metod dyskretyzacji, opartych na stałej liczbie wektorów lub wielkości przedziałów. Zamierzamy udostępnić bibliotekę programów do selekcji i dyskretyzacji.

1. Duch W, Winiarski T, Biesiada J, Kachel, A, Feature Ranking, Selection and Discretization. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, pp. 251-254

I.4.3. Selekcja i ważenie cech w metodach najbliższych sąsiadów można znacznie poprawić jakość działania takich modeli. Numeryczna metoda szukania najlepszego wektora wag zaimplementowana została w systemie GhostMiner i opisana została w pracy:

2. N. Jankowski. Discrete feature weighting and selection algorithm. 2003 International Joint Conference on Neural Networks, pages 636-641, Portland, USA, July 2003. The IEEE Neural Networks Society.

## **I.5. Wizualizacyjna eksploracja wielowymiarowych danych.**

Wizualizacja służy dwóm celom: wstępnej eksploracji danych i zrozumieniu struktury wielowymiarowych danych dzięki tworzeniu ich wizualnych reprezentacji, oraz próbie zmniejszenia wymiarowości danych. Podejście oparte na minimalizacji miar zachowania topograficznego sąsiedztwa wielowymiarowych danych, czyli skalowanie wielowymiarowe (MDS) oraz odwzorowanie Sammona, daje ciekawsze rezultaty niż samoorganizująca się mapa Kohonena, ale jest bardzo kosztowne obliczeniowo.

I.5.1 W przypadku dużej liczby danych skalowanie wielowymiarowe (MDS, multi-dimensional scaling) połączono z procedurami klasteryzacji typu LVQ (Learning Vector Quantization), pozwalającymi na zmniejszenie liczby centrów (codebook vectors) przeznaczonych do mapowania. Daje to mapę MDS dla położenia niewielkich klastrów, co w połączeniu z procedurą domapowywania pojedynczych wektorów pozwala na szybkie umiejscowienie nowych przypadków w otoczeniu przypadków znanych. Program pozwala na wierniejsze odwzorowanie wielowymiarowych danych niż bardzo popularny model samoorganizującej się sieci neuronowej (SOM) Kohonena. Praca jest w przygotowaniu.

I.5.2 Opracowano nowe sposoby projekcji danych pozwalające zrozumieć relację w przestrzeniach o stosunkowo niewielkiej liczbie wymiarów (rzędu 8-10) za pomocą projekcji na siatkę wierzchołków hipersześcianu (lattice projection), oraz dwuwymiarowych projekcji wykorzystujących kryteria Fishera (W. Duch). Projekcje te zastosowano do analizy map obrazujących wyjścia sieci neuronowej. Zastosowano też rzut na rogi wielokąta. Takie mapy niosą wiele przydatnych informacji, pozwalając unikać

modeli, które są zbyt „pewne siebie” i wybierać optymalne modele spośród kilku o tej samej dokładności. Wykorzystano je również do obserwacji trajektorii wielowymiarowych układów dynamicznych opisujących chaotyczne pamięci asocjacyjne (P. Matykewicz), co pozwala zaobserwować ich atraktory oraz ocenić ich baseny atrakcji.

1. Duch, W, Coloring black boxes: visualization of neural network decisions. International Joint Conference on Neural Networks, Portland, Oregon, 2003, Vol. I, pp. 1735-1740

I.5.3. Projekcje na dwie składowe główne zastosowane zostały do analizy trajektorii parametrów adaptacyjnych (wag i progów) sieci neuronowych, co pozwoliło lepiej zrozumieć proces uczenia się w tych sieciach. Próbujemy również zastosować tą metodę do skrócenia ostatnich etapów uczenia się sieci neuronowych.

2. Kordos M, Duch W, On Some Factors Influencing MLP Error Surface. The Seventh International Conference on Artificial Intelligence and Soft Computing (ICAISC), Zakopane, 7-11.06.2004 (submitted, 11/03)

I.5.4 Hierarchiczne podejście do skalowania wymiarowego jest testowane przy współpracy z prof. B. Schmidtem z SCE NTU w Singapurze.

## **I.6. Zastosowanie metod inteligencji obliczeniowej do analizy danych.**

I.6.1 Kontynuowano współpracę z firmą FQS Poland (Fujitsu Group) zainteresowaną stworzeniem nowego systemu do dogłębnej analizy danych (data mining), opartego na naszych algorytmach. Wersja 2.0 systemu została ukończona i jest obecnie sprzedawana przez FQS, jest w niej wiele nowych elementów, dokumentacja do wersji 2.0 liczy 255 stron a tutorial rzędu 100 stron. Wersję demonstracyjną pobrało już bardzo wiele osób a program stosowany jest między innymi przez wiodącą firmę bioinformatyczną Abbott Laboratories (USA) oraz stał się częścią systemu do wykrywania podejrzanych transakcji bankowych, budowanego przez firmę FQS. Jest to udane wdrożenie, które pozwoliło również członkom naszej grupy na łatwiejsze porównywanie i integrację nowych algorytmów, ułatwiając korzystanie z poszczególnych, rozwijanych dotychczas niezależnie programów. Obecnie pracujemy nad wersją 3.0 tego systemu.

1. N. Jankowski, K. Grąbczewski, and W. Duch. GhostMiner 2.0 User's Guide. FQS Poland, Fujitsu 2003, 255 stron.
2. N. Jankowski, K. Grąbczewski, and W. Duch. GhostMiner 2.0 Tutorial. FQS Poland, Fujitsu 2003, 99 stron.

I.6.2 Korzystając z publicznie dostępnych baz danych porównywaliśmy wiele systemów klasyfikujących dane jak i systemów dokonujących ekstrakcji danych. Wyniki tych badań są ciągle uzupełniane i udostępniane naszych stronach internetowych.

<http://www.phys.uni.torun.pl/KIS/projects/datasets.htm>

Odkryte przez nas reguły są proste i posługują się niewielką liczbą ważnych diagnostycznie testów; w kilku przypadkach udało się odkryć najprostszą znaną strukturę danych. Porównanie z innymi metodami uczenia maszynowego, w tym z metodami generującymi reguły rozmyte, pokazuje jednoznacznie wyższość proponowanych przez nas metod. Odkryte przez nas reguły te podane są na stronie:

<http://www.phys.uni.torun.pl/KIS/projects/rules.htm>

I.6.3 Otrzymaliśmy nowe rezultaty dla danych medycznych oraz danych dotyczących datowania obiektów archeologicznych (z Austrii) na podstawie analizy spektrograficznej próbek znalezionych szkieł. Odkryto interesujący problem związany z obecnością kilku próbek z tego samego obiektu w bazie danych. Takie sytuacje zdarzają się często również w bazach medycznych. Prowadzi to do zbyt optymistycznych wyników. Na razie nie ma dobrego sposobu na uwzględnienie tego typu przypadków.

3. Grudzinski K, Karwowski M, Duch W, Computational Intelligence Study of the Iron Age Glass Data. International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, June 2003, pp. 17-20

I.6.4 W. Duch wygłosił 3-godzinny tutorial „Computational Intelligence for Data Understanding”, na połączonych konferencjach Joint International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), w Istanbule, pokazujący zastosowania systemu Ghostminer;

4. Duch W, Computational Intelligence Tools for Data Understanding. In: ICANN/ICONIP, Istanbul, 06.2003 (tutorial, 50 str, osobna broszura).

## **I.7. Metody semantycznego szukania informacji.**

I.7.1. Prowadzono badania nad metodami semantycznego szukania informacji w informatyce medycznej, a w szczególności w pediatrycznych anotacjach Children's Hospital Medical Center (CHMC) w Cincinnati, Ohio, USA. W 2003 roku prowadzono głównie prace nad budową korpusu tekstów, na którym trenowane będą różne systemy uczące się. Tworzeniem korpusu zajmuje się mgr Ł. Itert (doktorant W. Ducha). Z jednej strony do tekstów dodano nazwy kategorii gramatycznych (POS tagger), częściowo automatycznie a częściowo robiła to grupa lingwistów, a z drugiej szukano reguł pozwalających na osiągnięcie jednoznaczności skrótów i akronimów (około 900 takich przypadków). Jedna praca jest w druku, druga w przygotowaniu.

1. J. Pestian, L. Itert, W. Duch, Development of a Pediatric Text-Corpus for Part-of-Speech Tagging Trends in Intelligent Information Processing and Web Mining, w druku

I.7.2. Opracowano też wstępnie interesującą ideę połączenia metod analizy tekstów z modelami pamięci rozpoznawczej, semantycznej i epizodycznej. Badania te należą częściowo do grupy badań kognitywnych. Zamierzamy stworzyć słownik cech obiektów, częściowo automatycznie, oraz przeanalizować strukturę przestrzeni semantycznych. Badania te są w fazie wstępnej. Przewidujemy znaczny rozwój tej problematyki: kilka tematów z tej grupy zaproponowaliśmy w projekcie utworzenia Centrum Doskonałości przy Instytucie Podstaw Informatyki PAN, do którego chcemy się przyłączyć.

## **II. Metody komputerowe w fizyce, chemii i naukach o życiu.**

Najważniejsze wyniki uzyskane w tej grupie tematów to:

1. Zakończenie badań nad efektywnością stosowania różnych wariantów teorii funkcjonałów gęstości (DFT) w opisie efektów anharmonicznych w widmach oscylacyjnych cząsteczek trójatomowych, na przykładzie stanu podstawowego cząsteczki  $SO_2$ . Przedmiotem badań było także porównanie wyników otrzymywanych metodami numerycznymi i zaburzeniowymi.
2. Prowadzono badania nad metodami uwzględniania korelacji elektronowej, w oparciu o konsyistentne wymiarowo metody wieloreferencyjnego oddziaływania konfiguracyjnego (dr Meller) oraz relację pomiędzy różnymi wariantami teorii funkcjonałów gęstości (DFT) a orbitalami Brucknera dla modelowych potencjałów korelacyjno-wymiennych.
3. Przeprowadzono szereg zaawansowanych symulacji procesów biologicznych, ściśle przy tym współpracując z doświadczalnikami. Zaproponowano nowe mechanizmy kontroli transkrypcji poprzez post-translacyjną degradację polimerazy RNA w wyniku wiązania z białkiem VHL. Dwa artykuły związane z tą tematyką zostały opublikowane w renomowanych czasopismach.
4. W oparciu o techniki modelowania komputerowego oraz data mining zaproponowano szereg hipotez dotyczących oddziaływań pomiędzy białkiem otoczki wirusa Norwalk i receptorami na



powierzchni komórek nabłonka jelitowego. Komputerowo zidentyfikowany binding pocket został następnie potwierdzony doświadczalnie, otwierając drogę do racjonalnego projektowania inhibitorów hamujących infekcję powodowaną przez wirusy typu Norwalk. Praca opublikowana w Journal of Virology, następna praca w przygotowaniu. Jest to kluczowe zagadnienie dla zrozumienia sposobu wnikania wirusów przez błony komórkowe, potencjalnie bardzo ważny mechanizm biologiczny dla opracowania leków antywirusowych.

5. Pracowano nad problemem diagnostyki artretyzmu dziecięcego za pomocą analizy profili ekspresji genów z tkanki maziowej otrzymanych z użyciem mikromacierzy oraz innych metod pomiaru ekspresji genów, praca została wysłana do druku.
6. Opracowano nowe metody przewidywania dwurzędowych struktur białkowych oraz dokładne metody przewidywania (w oparciu o metody regresji liniowej, np. Support Vector Regression oraz regresji nieliniowej) stopnia uwodnienia (solvent exposure) aminokwasów w natywnych strukturach trzeciorzędowych. Trzy prace związane z tą tematyką zostały wydrukowane bądź są w druku. Nowe metody są dostępne w postaci serwera: <http://sable.cchmc.org>
7. Zastosowano różne metody selekcji cech dla wielu problemów bioinformatycznych, publikacje na ten temat są w druku.
8. Opracowano nowy algorytm klasyfikacyjny dla optymalizacji parametrów potencjałów oddziaływań białek, w oparciu o komitety słabych klasyfikatorów zoptymalizowanych w użyciem technik programowania liniowego i heurystyki Maximum Feasibility. Wykorzystano techniki threading, dopasowania sekwencji (sequence alignment), techniki optymalizacji liniowej i analizy statystycznej struktur białkowych oraz tworzenia bibliotek unikalnych fragmentów białek za pomocą porównywania struktura-struktura przy zastosowaniu metod programowania dynamicznego. Potencjały zawierają parametry adaptacyjne a procedura uczenia się uwzględnia istniejące struktury natywne białek i wygenerowane struktury sztuczne, nie istniejące w przyrodzie.

## **II.1 Teoretyczne badania struktury energetycznej małych cząsteczek i uwzględniania efektów korelacji elektronowej.**

Badania te wykonywane były częściowo w ramach grantu KBN (kierowanego przez prof. K. Jankowskiego z Instytutu Fizyki UMK), w realizacji którego uczestniczy prof. Wasilewski,

**II.1.1.** Prof. J. Wasilewski z współpracownikami prowadził badania w ramach tematu: „Struktura energetyczna małych cząsteczek”, zajmując się zastosowaniami teorii funkcjonałów gęstości (DFT – Density Functional Theory) do obliczeń struktury energetycznej i widm elektronowo-oscyłacyjnych cząsteczek. Uczestniczył też w badaniach kierowanych przez prof. K. Jankowskiego z Instytutu Fizyki UMK, dotyczących nowych metod opisu efektów korelacji elektronowej w stanach niezamkniętopowłokowych atomów i cząsteczek.

1. S. Zelek, L. Cyrnek, J. Wasilewski: Lower vibrational states of  $SO_2$ . Density Functional and Discrete Variable Representation Approach. – J. Mol. Struct. (THEOCHEM), 629:61-70 (2003)

**II.1.2.** Kontynuowano badania teoretyczne niższych stanów elektronowych pirazyny i pirymidyny przy zastosowaniu różnych wariantów metody funkcjonałów gęstości (DFT). Do obliczeń energii przejścia elektronowego  $S_0 - T_1$  zastosowano, z dobrym wynikiem, modyfikowany funkcjonał hybrydowy typu B3. Do wyznaczenia powierzchni otwartopowłokowego stanu singletowego zastosowano wariant czasowo zależny teorii (TD DFT); tego typu zastosowania są w literaturze nieliczne. Obliczenia teoretyczne mają na celu interpretację wyników doświadczalnych, otrzymanych przez prof. Heldta w ramach współpracy międzynarodowej (z grupą prof. C. Ottingera w Getyndze) najnowszą techniką laserowo indukowanej fluorescencji” (LIF – Laser Induced Fluorescence). Wyniki te wchodzą w zakres prowadzonych aktualnie prac magisterskich.

**II.1.3.** Prowadzone były badania układów modelowych w ramach nowych wersji wieloreferencyjnych metod sprzężonych klasterów (MRCC - Multi-Reference Coupled Clusters) w sformułowaniach wa-

lencyjnie-universalnym, stanowo-universalnym i stanowo-specyficznym. Badania dotyczyły wpływu bazy jednoczątkowej (formy orbitali molekularnych) na zbieżność rozwinięcia klasterowego dla stanów podstawowych i wzbudzonych, ze szczególnym uwzględnieniem występowania tzw. stanów obcych („intruderów,„). Prof. Wasilewski jest aktywnym uczestnikiem seminarium naukowego dotyczącego teorii korelacji i podstaw teorii funkcjonałów gęstości.

**II.1.4.** Ukazała się dłuższa praca na temat konsystentnej wymiarowo wersji metody oddziaływania konfiguracyjnego, nad którą pracowano w ostatnich kilku latach.

2. J. Meller, J. P. Malrieu and J. L. Heully; Size-consistent multireference CI through the dressing of the norm of determinants, *Molecular Physics*, vol. 101 (13), 2029-2041 (2003)

## **II.2. Symulacje komputerowe w fizyce i chemii**

II.2.1. Prowadzono prace rozpoznawcze i studialne w następujących kierunkach: oddziaływanie tlenu molekularnego z atomami i małymi cząsteczkami; zastosowanie różnego typu funkcjonałów mieszanych w metodzie DFT; zastosowanie metod DFT do opisu stanów otwartopowłokowych; zastosowanie reprezentacji zmiennej dyskretnej (DVR - Discrete Variable Representation) do numerycznego wyznaczania poziomów oscylacyjnych małych cząsteczek; komputerowe modelowanie procesów dekompozycji ozonu – w ramach pracy magisterskiej kierowanej przez dr S. Zelka.

II.2.2. Kontynuowano prace nad interpretacją widm elektronowo-oscylacyjnych pirazyny i pirymidyny, uzyskanych metodą laserowo indukowanej fluorescencji (LIF) w międzynarodowym zespole prof. J. Heldta (Uniwersytet Gdański) i prof. C. Ottingera (Max-Planck-Institut fuer Stroemungsforschung w Getyndze). W badaniach zastosowano szereg wariantów teorii funkcjonałów gęstości i rachunku zaburzeń Moellera-Plesseta.

II.2.3. Ukazała się praca na temat wpływu efektów izotopowych na oddziaływania hydrofobowe; jest to bardzo istotne zagadnienie dla biomolekuł.

1. M. Turowski, N. Yamakawa, J. Meller, K. Kimata, T. Ikegami, K. Hosoya, N. Tanaka and E.R. Thornton; Deuterium Isotope Effects on Hydrophobic Interactions. The Importance of Dispersion Interactions in the Hydrophobic Phase, *Journal of American Chemical Society*, 125: 13836-13849 (2003)
2. M. Zduniak Teoretyczne badania niższych stanów elektronowo-oscylacyjnych cząsteczki pirymidyny. Praca magisterska, obroniona 11.2003.

## **II.3. Bioinformatyka i nauki o życiu.**

W ramach współpracy z jedną z wiodących instytucji amerykańskich w dziedzinie nauk o życiu i ich klinicznych zastosowań, Children's Hospital Research Foundation w Cincinnati, gdzie dr J. Meller pracuje w Department of Biomedical Informatics jako assistant professor, kontynuowano badania w zakresie bioinformatyki i nauk o życiu. Bierze w nich udział były doktorant KIS, Rafał Adamczak.

W 2003 roku pracowano nad następującymi tematami:

II.3.1 Opracowano nowe metody przewidywania dwurzędowych struktur białkowych oraz dokładne metody przewidywania (w oparciu o metody regresji liniowej, np. Support Vector Regression oraz regresji nieliniowej) stopnia uwodnienia (solvent exposure) aminokwasów w natywnych strukturach trzeciorzędowych. Trzy prace związane z tą tematyką zostały wydrukowane bądź są w druku. Nowe metody są dostępne w postaci serwera: <http://sable.cchmc.org>

1. M.-D. Filippi, C. E. Harris, J. Meller, Y. Zheng and D. A. Williams; Localization of Rac2 Specifies Superoxide Generation, Actin Polarity and Chemotaxis in Neutrophils, *Nature Immunology*, w druku
2. R. Adamczak, A. Porollo and J. Meller, Accurate Prediction of Solvent Accessibility Using Neural Networks Based Regression. *Proteins: Structure, Function and Bioinformatics*, w druku
3. M. Wagner, R. Adamczak, A. Porollo and J. Meller; Linear Regression Models for Solvent Accessibility Prediction in Proteins, wysłana do *Journal of Computational Biology*

II.3.2 Pracowano nad problemem diagnostyki artretyzmu dziecięcego za pomocą analizy profili ekspresji genów z tkanki maziowej otrzymanych z użyciem mikromacierzy oraz innych metod pomiaru ekspresji genów:

4. J. Meller, R. Adamczak, M. Scola, M. Barnes, S. Thompson, M. Passo, H. Brunner, D. N. Glass and A. A. Grom, Pattern Recognition Analysis of Synovial Tissue Cytokine Expression Profiles Helps Identify Patients with Systemic Onset Juvenile Rheumatoid Arthritis, wysłana do *Arthritis Research*

II.3.3 Przy współpracy z grupą eksperymentalną nowe mechanizmy kontroli transkrypcji poprzez post-translacyjną degradację polimerazy RNA w wyniku wiązania z białkiem VHL. Dwa artykuły związane z tą tematyką zostały opublikowane w renomowanych czasopismach.

5. V. Kuznetsova, J. Meller, P. O. Schnell, J. A. Nash, Y. Sanchez, J. W. Conaway, R. C. Conaway and M. F. Czyżyk-Krzeska; VHL binds hyperphosphorylated large subunit of RNA Polymerase II through a praline hydroxylation motif and targets it for ubiquitination, *PNAS* vol. 100 (5), 2706-2711 (2003)
6. M. F. Czyżyk-Krzeska and J. Meller; Von Hippel-Lindau Tumor Suppressor: Not Only HIF's Executioner, *Trends in Molecular Medicine*, w druku

II.3.4 W oparciu o techniki modelowania komputerowego oraz data mining zaproponowano szereg hipotez dotyczących oddziaływań pomiędzy białkiem otoczki wirusa Norwalk i receptorami na powierzchni komórek nabłonka jelitowego. Komputerowo zidentyfikowany binding pocket został następnie potwierdzony doświadczalnie, otwierając drogę do racjonalnego projektowania inhibitorów hamujących infekcję powodowaną przez wirusy typu Norwalk. Praca opublikowana w *Journal of Virology*, następną pracą w przygotowaniu.

7. M. Tan, P. Huang, J. Meller, W. Zhong, T. Farkas and X. Jiang; Mutations within P2 Domain of Norovirus Capsid Affect Binding to Human Histo-Blood Group Antigens: Evidence for a Binding Pocket, *Journal of Virology*, 77 (23): 12562-71 (2003)

II.3.5 Opracowano nowy algorytm klasyfikacyjny w oparciu o komitety słabych klasyfikatorów zoptymalizowanych z użyciem technik programowania liniowego i heurystyki Maximum Feasibility; artykuł konferencyjny opublikowany, kolejny w przygotowaniu.

8. A. Porollo, R. Adamczak, M. Wagner and J. Meller; Maximum Feasibility Approach for Consensus Classifiers: Applications to Protein Structure Prediction, *Proceedings of The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems, Computational Intelligence, Robotics and Adaptive Systems, Singapore, Dec. 2003*
9. R. Adamczak and J. Meller; On the Transferability of Folding and Threading Potentials and Sequence-Independent Filters for Protein Folding Simulations, *Molecular Physics*, w druku
10. M. Wagner, J. Meller and R. Elber; Large-Scale Linear Programming Techniques for the Design of Protein Folding Potentials, *Mathematical Programming*, w druku

II.3.6 Opracowano graficzny program do adnotacji strukturalnej i funkcjonalnej białek:

11. A. Porollo, R. Adamczak and J. Meller; Polyview: A Flexible Visualization Tool for Structural and Functional Annotations of Proteins, *Bioinformatics*, w druku

II.3.7 Ukazała się dłuższa praca na temat dynamiki molekularnej w wydanej przez Nature Publishing encyklopedii genomu ludzkiego.

12. J. Meller, Molecular Dynamics, w: *Encyclopedia of the Human Genome*. Nature Publishing Group, Macmillan Publishers Ltd 2003.

### **III. Kognitywistyka, neuroinformatyka, komputerowe modele procesów poznawczych i filozofia umysłu**

Nasze prace w tej dziedzinie zmierzają w kilku kierunkach. W 2003 roku pracowano nad następującymi zagadnieniami:

#### **III.1. Modele procesów poznawczych i zhumanizowane interfejsy.**

**III.1.1** Podejmujemy próby stworzenia empirycznie uzasadnionej teorii umysłu w oparciu o modele sieci neuronowych. Teoria powinna nie tylko być pomocną w rozwiązaniu podstawowych problemów nauk o poznaniu, lecz również prowadzić do nowych konstrukcji systemów uczących się. Zarysy takiej teorii, w postaci hipotez określających prawdopodobne związki pomiędzy zdarzeniami w mózgu i ich korelacji z percepcją, wyższymi czynnościami poznawczymi i treścią świadomości, zostały już nakreślone w kilku pracach i wystąpieniach konferencyjnych. Matematyczne metody opisu zdarzeń mentalnych wiążą się z teorią układów złożonych w zastosowaniu do sieci neuronowych z rekurencją. Takie podejście napotyka jednak trudności praktyczne i może służyć jedynie uzasadnieniu możliwości przeprowadzenia pewnych redukcji do uproszczonych modeli.

1. Duch W, Neurokognitywna teoria świadomości. *Studia z kognitywistyki i filozofii umysłu* (red. W. Dziarnowska i A. Klawiter). Tom. I, Subiektywność a świadomość. Zysk i S-ka, Poznań 2003, str. 133-154.
2. Duch W, Geometryczny model umysłu. *Kognitywistyka i Media w Edukacji*, Vol. 6 (2002) 199-230 (numer pojawił się w 2003 r).

**III.1.2** Opracowano zarys teorii złożonych agentów, rozważając coraz bardziej złożone sieci oddziaływujących elementów. Przejście od prostych sieci neuronowych do sieci złożonych oddziaływujących agentów dokonać można zwiększając lokalną wiedzę i złożoność oddziaływań, dlatego warto konstruować systemy o pośredniej złożoności, których na razie brakuje. Zaproponowano szkic takiej teorii, od najprostszycy sieci z węzłami o pojedynczym parametrze i liniowymi oddziaływaniami, przez sieci o złożonych elementach wewnętrznych z coraz większą lokalną wiedzą, sieci złożone z podsieci, sieci wyspecjalizowanych modułów, do sieci agentów wymieniających informację w symboliczny sposób.

3. Duch W, Mandziuk J, Quo vadis, computational intelligence? In: *Quo Vadis Machine Intelligence? The Progressive Trends in Intelligent Technologies*. Eds: P. Sinčák, J. Vaščák, K. Hirota, *Studies in Fuzziness and Soft Computing*, Springer Verlag, w druku
4. Duch W, Complex Agents for Socio-Cognitive Engineering. *First International Workshop on Socio-Cognitive Engineering Foundations (SCEF-2003)*, Rome, Italy, Sept 30 - Oct 1, 2003

**III.1.3** Opracowano założenia minimalnego systemu, który będzie wyrażał przekonanie, że jest świadomy – jest to nowy punkt widzenia, pozwalający uniknąć pułapek dyskusji filozoficznych związanych z definiowaniem świadomości. Artykuł prezentuje realistyczne podejście do problemu zrozumienia natury świadomości i pokazuje, jak można stworzyć świadome komputery, chociaż jest to technicznie bardzo trudne, gdyż nie będą to urządzenia przypominające komputery.

5. Duch W, Brain-inspired conscious computing architecture. *Journal of Mind and Behavior* (submitted, Oct. 2003)

III.1.4 Tworzenie uproszczonych modeli procesów poznawczych wiąże się z próbą stworzenia naturalnych, zhumanizowanych interfejsów programowych, z którymi można będzie porozumieć się przy pomocy głosu. Coraz więcej osób używa komputerów podręcznych typu PDA, coraz większe moce obliczeniowe są też w telefonach lub palmofonach komórkowych. Projekty zhumanizowanych interfejsów (HIT, Humanized InTerfaces) są dobrym polem do testowania różnych teorii kognitywistycznych oraz wymuszają integrację wielu technologii dotyczących rozpoznawania i syntezy mowy, prozodii wymowy, korelacji z emocjonalną treścią tekstu, dopracowania graficznej strony mówiącej głowy, synchronizacji ruchu ust z dźwiękiem i wielu innych zagadnień. Planujemy rozwiązać je w ramach współpracy z School of Computer Engineering, NTU, Singapur. Przygotowano podanie o grant na rozwój interfejsów (HIT, Humanized InTerfaces) integrujących technologie awatarów (mówiącej głowy), syntezę i rozpoznawanie mowy, oraz interakcję poprzez takie interfejsy ze stronami internetowymi. Jednym z zastosowań będą gry słowne, takie jak np. gra w 20 pytań.

### III.2. Gry słowne i modele umysłu.

III.2.1 Dialog w języku naturalnym z takimi interfejsami wymaga prac dotyczących analizy języka naturalnego. Sformułowaliśmy ogólne założenia dotyczące takich projektów, wykonaliśmy (mgr J. Szymański, T. Sarnatowicz) wstępne prace w oparciu o awatary Hapteka, demonstrując możliwość gry w 20 pytań z programem dołączonym do strony internetowej. Stworzenie dobrego programu tego rodzaju wymaga szeregu badań o naturze podstawowej. Do sprawnej gry w 20 pytań i do stworzenia dobrego systemu odpowiadającego na pytania potrzebny jest słownik cech obiektów, uproszczona reprezentacja pojęć w oparciu o wektory kontekstowe. Sformułowaliśmy różne projekty, łącząc zagadnienia lingwistyczne z teorią przestrzeni semantycznych, modelowaniem pamięci rozpoznawczej, semantycznej i epizodycznej. Badania te są w fazie wstępnej, ale wydają się bardzo obiecujące.

III.2.2 Jednym z problemów jest rozumienie roli podobieństwa i reguł logicznych w procesach poznawczych. Psycholodzy patrzą na to zagadnienie coraz bardziej jako dwa bieguny tego samego zjawiska, jednak podejście psychologiczne prowadzi do wielu nieporozumień. Reguły progowe realizowane przez sieci neuronowe nie zawsze przekładają się w prosty sposób na reguły propozycjonalne. Do druku w *Behavioral and Brain Sciences*, najlepszym piśmie nauk behawioralnych i nauk o mózgu, przyjęto komentarz W. Duchy do pracy na ten temat:

6. Duch W, Rules, Similarity, and Threshold Logic. Commentary on Emmanuel M. Pothos, The Rules versus Similarity distinction. *Behavioral and Brain Sciences* (w druku)

III.2.3 Artykuł o synchroniczności omawia realistyczne możliwości wyjaśnienia anomalii obserwowanych w eksperymentach prowadzonych od wielu lat w Princeton (zakładając, że nie ma w nich jakiegoś systematycznego błędu).

7. Duch W, Synchronicity, Mind and Matter. *Neuroquantology*, Vol. 1, January 2003, pp. 2-17

### III.3. Filozofia umysłu.

III.3.1 Podejmujemy też filozoficzne próby zrozumienia koncepcji umysłu i zagadnień dotyczących świadomości. Do druku w najlepszym piśmie nauk behawioralnych i nauk o mózgu (*Behavioral and Brain Sciences*, BBS) przyjęto komentarz do pracy o modelu „gestalt” subiektywnej świadomości; W. Duch został wpisany na listę BBS associates.

8. Duch W, Just bubbles? Commentary on Steven Lehar, Gestalt Isomorphism and the Primacy of Subjective Conscious Experience: A Gestalt Bubble Model. *Behavioral and Brain Sciences* (w druku).

Udostępniono również wiele tekstów referatów wygłoszonych na konferencjach i na zaproszenie ośrodków polskich i zagranicznych. Można je znaleźć pod adresem:

<http://www.phys.uni.torun.pl/~duch/ref.html>

## IV. Badania w zakresie dydaktyki wspomagananej metodami komputerowymi

W 2003 roku wykonano następujące prace:

IV.1 Utrzymujemy i aktualizujemy strony WWW dotyczące wszystkich spraw związanych z inteligencją obliczeniową, sztuczną inteligencją, inteligentnym szukaniem i ekstrakcją informacji, kognitywistyką, badaniami nad mózgiem, neurobiologią teoretyczną, naukami informacyjnymi, fizyką komputerową. Ze stron tych korzysta bardzo wiele osób.

- o Inteligentne metody szukania informacji,  
<http://www.phys.uni.torun.pl/~duch/IR.html>
- o Inteligencja obliczeniowa, statystyka, układy dynamiczne  
<http://www.phys.uni.torun.pl/~duch/CI.html>
- o Sztuczna inteligencja i uczenie maszynowe,  
<http://www.phys.uni.torun.pl/~duch/ai-ml.html>
- o Modelowanie neuronowe, badania nad mózgiem i neurobiologia:  
<http://www.phys.uni.torun.pl/~duch/neural.html>
- o Kognitywistyka  
<http://www.phys.uni.torun.pl/~duch/cognitive.html>
- o Oprogramowanie i bazy danych do testów metod inteligencji obliczeniowej  
<http://www.phys.uni.torun.pl/~duch/software.html>
- o Fizyka komputerowa,  
<http://www.phys.uni.torun.pl/~duch/fiz-komp/index.html>  
AI w szukaniu informacji i przetwarzaniu tekstów
- o Nauka i jej peryferia para, trans, futuro i religio-logiczne  
<http://www.phys.uni.torun.pl/~duch/science.html>
- o Rezultaty - porównanie metod klasyfikacji i ekstrakcji reguł logicznych.  
<http://www.phys.uni.torun.pl/kmk/projects/datasets.html>

Przykładowo, w internetowej encyklopedii Wikipedia przytoczono reguły logiczne, które otrzymaliśmy analizując zbiór danych dla grzybów.

IV.2 W ramach eksperymentów z nauczaniem na odległość udostępniłmy i zaktualizowaliśmy wiele stron Internetowych z notatkami do wykładów, pod adresem:

<http://www.phys.uni.torun.pl/~duch/Wyklady/index.htm>

Duch W, Wykłady wstępne o komputerach i oprogramowaniu, Studium Podyplomowe.

Duch W, *Wstęp do kognitywistyki* (wykład monograficzny).

Duch W, *Mózg, umysł i zachowanie* (wykład monograficzny).

Duch W, *Sztuczna inteligencja* (wykład monograficzny).

Duch W, *Inteligencja obliczeniowa* (wykład monograficzny).

IV.3 Utrzymujemy strony WWW dotyczące zagadnień związanych z komputerami i oprogramowaniem. Dziedzina ta, zwana w krajach anglojęzycznych „*information science*“.

Adres: <http://www.phys.uni.torun.pl/~duch/book-fsk.html>

Ze stron tych korzysta bardzo wiele osób (około 1000 nowych załadowań strony miesięcznie). Są to pierwsze książki nadające się jako podręcznik do dziedziny określanej w krajach anglosaskich mianem „*information science*”, dotyczącej zastosowań komputerów do przetwarzania informacji, i w niektórych ośrodkach akademickich już znalazły takie zastosowanie.

IV.4 Prof. J. Wasilewski koordynował prace redakcyjne nad czwartym wydaniem Wydziałowego Informatora ECTS - Studia na Wydziale Fizyki, Astronomii i Informatyki Stosowanej UMK w roku akademickim 2003/04, ISBN 83-231-1576-1, str. 412, Wydawnictwo UMK, Toruń, 2003 (skład i łamanie: mgr H. Małek z KIS).

IV.5 Prowadziliśmy eksperymenty z nauczaniem na odległość za pomocą narzędzi Blackboard, używanych w School of Computer Engineering, NTU, Singapur, serwerów Centra pomiędzy USA (Ann Arbor), Singapurem i Toruniem, oraz wykorzystując bezpośrednią łączność internetową pomiędzy Singapurem i Poznaniem, gdzie W. Duch wygłosił godzinny referat i brał udział w dyskusji.

### **D 1.5) Uzyskane nagrody o znaczeniu międzynarodowym i ogólnokrajowym.**

1. W międzynarodowym konkursie „NIPS 2003 Feature Selection Challenge” system programów GhostMiner, którego autorami są K. Grąbczewski i N. Jankowski zajął 3 miejsce (na 78 uczestników) w konkursie selekcji cech dla zadań klasyfikacji. NIPS, czyli Neural Information Processing Systems, jest najpoważniejszą konferencją w tej dziedzinie.
2. Praca na temat ekstrakcji reguł logicznych z danych empirycznych została umieszczona na okładce *Proceedings of IEEE*, najbardziej prestiżowego pisma IEEE (współautorami są R. Setiono, Singapore i J. Żurada, Louisville), numer z maja 2004.

### **D1.6) Inne ważniejsze osiągnięcia.**

Opisane z załączniku 2b

### **D2. Informacja o upowszechnianiu i popularyzacji wyników działalności jednostki.**

W. Duch udzielił wywiadu dla “Computer Times” w Singapurze, który pojawił się w dwóch artykułach “Artificial Intelligence: Still a man's world”, oraz “When AI runs your life”, w numerze z 26 listopada 2003.