

SPRAWOZDANIE Z DZIAŁALNOŚCI NAUKOWO-BADAWCZEJ KATEDRY INFORMATYKI STOSOWANEJ UMK W ROKU 2005

I. TEMATY BADAWCZE

1. Liczba realizowanych tematów.

- 14, w tym 9 w ramach działalności statutowej, 5 w ramach badań własnych.

2. Liczba tematów zakończonych: 1 - w planowanym terminie 1, - z opóźnieniem 0.

- Prowadzone badania mają w większości charakter długoterminowy.

3. Z liczby tematów zakończonych przekazano:

- do wykorzystania, - do wdrożenia 1.

4. Z liczby tematów przekazanych do wdrożenia, w 2005 r. wdrożono:

- ogółem 1, - w tym w resorcie edukacji narodowej 0.

- Wdrożenie dotyczy metod inteligencji obliczeniowej zawartych w pakiecie programów GhostMiner 3 do data mining, opracowanym przy współpracy z firmą FQS Poland (oddział Fujitsu); obecnie powstaje całkiem nowa wersja systemu do analizy danych w oparciu o prowadzone badania.

5. Liczba tematów, których realizację zaniechano (w roku sprawozdawczym).

- nie zaniechano żadnego tematu.

Liczba realizowanych tematów ogółem		14
Działalność statutowa		9
Badania własne		5
Granty UMK		0
Liczba tematów zakończonych		
w tym	W planowanym terminie	1
	z opóźnieniem	0
z liczby tematów zakończonych przekazano	Do wykorzystania	0
	Do wdrożenia	1
z liczby tematów do wdrożenia, w 2005 wdrożono	Ogółem	1
	W tym w resorcie edukacji narodowej	0
Liczba tematów, których realizację zaniechano		0

II. CZĘŚĆ OPISOWA

II.1. Przyczyny zaniechania tematów prowadzonych i rozpoczętych w 2005 roku.

W 2005 roku nie zaniechaliśmy realizacji żadnego tematu.

II.2. Krótka charakterystyka i podsumowanie wyników uzyskanych w 2005 r.

A) Inteligencja obliczeniowa (computational intelligence)

Główne osiągnięcia w tej grupie tematów to:

1. 11-15 września odbyła się w Warszawie 15th International Conference on Artificial Neural Networks (ICANN 2005), oficjalna doroczna konferencja European Neural Network Society, a na jej zakończenie 3 warsztaty na UMK w Toruniu (bioinformatyczny, neuroinformatyczny oraz budowy sztucznych mózgow). Konferencję zorganizowała nasza grupa razem z Instytutem Badań Systemowych PAN, jej przewodniczącymi byli W. Duch i J. Kacprzyk. Materiały w objętości ponad 1100 stron wydano w dwóch tomach Springer Lecture Notes in Computer Science, Vol 3696 i 3697.
2. Dla Future Emerging Technologies (FET) 6 Programu Ramowego UE opracowano duży projekt zintegrowany „Artificial Brain Architecture and Cognitive Control Understanding System” (ABACCUS), przy udziale 9 partnerów; w 2005 roku projekt ten został poprawiony i powtórnie złożony, okazał się jednak zbyt kosztowny, dlatego wspólnie z 4 partnerami (King’s College London, NCSR “Demokritos” w Atenach, FRT – Hellas na Krecie, oraz SCE NTU w Singapurze) złożyliśmy we wrześniu 2005 podanie o grant na projekt „Brain as Complex System (BRACS)” z mniejszym budżetem. Jest to najbardziej ambitny z dotychczas sformułowanych na świecie projektów w dziedzinie neuroinformatyki, napisany w znacznej części przez W. Duchą. Zaproszenie do udziału w tym projekcie uważamy za wielkie wyróżnienie.
3. W lipcu 2005 r. rozpoczęliśmy realizację grantu KBN na rozwój metod selekcji informacji. Dr J. Biesiada z Politechniki Śląskiej otrzymał stypendium FNP by spędzić 3 miesiące na UMK pracując nad tym tematem. Rozwijano pakiet programów realizujący znane i nowe metody przydatne do dyskretyzacji i selekcji cech. Wybrane metody dodano do pakietu GhostMiner 3.0. W złożonej do druku książce „Feature extraction, Foundations and Applications”, Springer 2006, pod redakcją I. Guyon, S. Gunn, M. Nikravesh, i Lofti Zadeha, znajdują się 3 rozdziały napisane przez naszą grupę. Jest to jak dotychczas najlepsza książka w tej dziedzinie, napisana przez grono wybitnych specjalistów z USA i Europy. Na temat selekcji informacji napisano szereg prac prezentowanych na wielu międzynarodowych konferencjach.
4. W lipcu 2005 r. rozpoczęliśmy też realizację grantu KBN na rozwój metod meta-uczenia, polegających na automatycznym wyborze najlepszych alternatywnych modeli dla danego punktu złożoności/dokładności. Temat ten będzie podstawą nowego pakietu do data mining, nad którego założeniami pracowaliśmy znaczną część 2005 roku.
5. W związku z tymi pracami zakończono współpracę z firmą FQS Poland (oddziałem Fujitsu w Polsce) nad rozbudową pakietu programów GhostMiner, opartego w większości na algorytmach do data mining opracowanych w naszej Katedrze. FQS zamierza nadal rozwijać ten pakiet samodzielnie, głównie w zastosowaniach bankowych. Licencję na Ghostminera wykupiła między innymi jedna z najbardziej znanych firm zajmujących się bioinformatyką, Abbot Laboratories.
6. Rozwijano nowe, konkurencyjne podejście do systemów neurorozmytych, oparte na szukaniu prototypów i optymalizacji funkcji odległości, w tym probabilistycznych miarach podobieństwa. Pokazano, że generowane w ten sposób reguły można przedstawić w postaci reguł rozmytych i w niektórych przypadkach reguły rozmyte można zamienić na prototypy i funkcje odległości. Coraz lepiej widać związki takiego podejścia z innymi metodami, prowadzące do ogólnej teorii metod inteligencji obliczeniowej. W 2006 roku mgr Blachnik spędzi 6 miesięcy w laboratorium prof. Erkki Oja w Helsinkach, rozwijając ten temat.
7. Rozwijano nowe metody uczenia sieci neuronowych. M. Kordos obronił w czerwcu 2005 pracę doktorską na temat algorytmów szukania, całkiem nowej klasy algorytmów uczących. Prace na ten temat zostały już częściowo opublikowane, dłuższa praca (Variable Step Search Training Algorithm for Feedforward Networks) wysłana została do IEEE Transactions on Neural Networks. Wizualizacja wielowymiarowych powierzchni błędów pozwoliła na wyciągnięcie szeregu konkluzji usprawniających uczenie. Wizualizacja aktywności warstwy ukrytej oraz wyjściowej w wielu wymiarach (opisana w pracy Internal representations of multi-layered perceptrons i pracach wcześniejszych) doprowadziła do nowej koncepcji teoretycznej, k-separowalności, która będzie podstawą nowych algorytmów jak i pozwoli na głębsze analizy w teorii inteligencji obliczeniowej. Opublikowano algorytm aktywnego uczenia SVNN (Support Vector Neural Networks), który po-

zwolił na znaczną poprawę wyników uzyskiwanych przez sieci neuronowe dla trudnych, dużych baz danych (W. Duch, Lecture Notes in Computer Science, Vol 3697, 67-72, 2005)

- Rozwinęliśmy badania związane z analizą tekstów medycznych i ogólnymi metodami analizy tekstów; opracowano szkic neurokognitywnego podejścia do rozumienia tekstów, integrujące pamięć rozpoznawczą, semantyczną i epizodyczną, zaczęto tworzenie sieci semantycznych z konkurencją zawierających wiedzę a priori pozwalająca na interpretację tekstu, grafów spójnych koncepcji, oraz opracowano nowy algorytm stosujący programowanie liniowe do klasyfikacji tekstów. Kilka prac już opublikowano, kolejne są w przygotowaniu. Mamy nadzieję, że podejście neurokognitywne spowoduje w analizie języka naturalnego znaczny postęp.
8. O pozycji naszego zespołu w świecie świadczy fakt, że W. Duch został wybrany na prezydenta European Neural Network Society (kadencja 2005-2008), oraz jest obecnie członkiem redakcji 10 międzynarodowych pism specjalistycznych, w tym IEEE Transaction on Neural Networks.

B) Metody komputerowe w fizyce, chemii i naukach o życiu.

Prof. J. Wasilewski i dr S. Zelek prowadzili 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 i uczestnicząc w badaniach kierowanych przez prof. K. Jankowskiego z Instytutu Fizyki UMK, dotyczących metod opisu efektów korelacji elektronowej w stanach niezamknięto-powłokowych atomów i cząsteczek. Dr hab. J. Meller i dr R. Adamczak prowadzili badania nad zastosowaniem metod bioinformatyki i biologii obliczeniowej do problemu rozpoznawania struktur i funkcji białek, metod dopasowania sekwencji w biologii molekularnej, analizy podobieństwa genomów (ang. *synteny*) a także identyfikacji markerów stanów chorobowych i fenotypicznych w danych klinicznych i genomicznych dotyczących polimorfizmów i profili ekspresji genów. Najważniejsze wyniki uzyskane w tej grupie tematów to:

1. Prowadzono badania nad metodami uwzględniania korelacji elektronowej, oraz badano relacje pomiędzy przestrzeniami orbitalnymi i rozkładami gęstości elektronowej otrzymywanymi w różnych wariantach teorii funkcjonałów gęstości (DFT).
 - Jankowski K., I. Grabowski, K. Nowakowski, **J. Wasilewski**, *Ab initio* correlation effects in density functional theory: an electron-distribution-based study for neon, Collect. Czech. Chem. Commun., 70:1157-1176 (2005)
2. Ulepszono potencjały do przewidywania struktury białek i zaimplementowano algorytmy w kolejnych wersjach serwera obliczeniowego LOOPP (Learning, Observing and Outputting Protein Patterns). Rozpoznaje on struktury białek (fold recognition). 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. Program jest dostępny w postaci serwera sieciowego umożliwiającego automatyczną anotację sekwencji genomicznych pod adresem:

<http://www.tc.cornell.edu/reports/NIH/resource/CompBiologyTools/looppp/>
3. Opracowano nową metodę oceny solwatacji reszt aminokwasowych w natywnych strukturach białkowych (ang. *solvent accessibility*) z użyciem nowatorskich podejść opartych na modelach regresji nieliniowej oraz liniowej (Support Vector Regression). Rozwinięto serwer SABLE (<http://sable.cchmc.org>), który osiągnął dokładność najlepszych dostępnych metod również w przewidywaniu struktur drugorzędowych. Zaproponowano jak takie metody mogą być następnie użyte do przewidywania oddziaływań między-białkowych i białek membranowych.
 - Adamczak, A. Porollo and J. Meller; Combining Prediction of Secondary Structures and Solvent Accessibility in Proteins, Proteins: Structure, Function and Bioinformatics, 59(3): 467-75 (2005)
 - Wagner M, Adamczak R, Porollo A, Meller J, Linear regression models for solvent accessibility prediction in proteins, Journal of Computational Biology 12(3):355-69 (2005)
4. Przeprowadzono szereg zaawansowanych symulacji procesów biologicznych, ściśle przy tym współpracując z doświadczalnikami. Wyniki zawarte są w dwóch opublikowanych pracach:

- Ivanenko V.V., Meller J., and T. L. Kirley; Characterization of disulfide bonds in human nucleoside triphosphate diphosphohydrolase 3 (NTPDase3): implications for NTPDase structural modeling, *Biochemistry*, 44(25):8998-9012 (2005)
- Petre-Draviam C.E., Williams E.B., Burd C.J., A. Gladden, H. Moghadam, J. Meller, J. A. Diehl, and K. E. Knudsen; A Central Domain of Cyclin D1 Mediates Nuclear Receptor Co-repressor Activity, *Oncogene*, 24(3): 431-44 (2005)

C) Modelowanie kognitywne (cognitive modeling).

W 2005 roku W. Duch przy współpracy z Center for Computational Intelligence, NTU Singapur napisał szereg wniosków o granty w oparciu o międzynarodową współpracę z Singapurem, USA, Koreą i Wielką Brytanią, na temat: 1) rozpoznawania i modelowania emocji (Emovere); 2) integracji percepcji i architektur kognitywnych w zhumanizowanych interfejsach (HIT, Humanized InTerfaces), oraz do sterowania głową androida w czasie rzeczywistym (DREAM, Developmental Robot-Embedded Artificial Mind), modelowania procesu nabywania umiejętności manualnych, odśwaidomego działania do pełnej automatyzacji („Brain-Inspired Model of Skill Learning: from Conscious Cognition to Subconscious Actions”), monitorowania niemowląt za pomocą smoczków telemetrycznych i innych czujników (projekt IDoCare); neurokognitywne podejście do procesów twórczych w lingwistyce; sformułowano też założenia nowych projektów dotyczących analizy EEG i zastosowania neurofeedback w leczeniu zaburzeń traumatycznych i depresji. Wnioski te są obecnie oceniane lub w fazie wstępnej realizacji.

W styczniu 2005 W. Duch był współorganizatorem workshopu na temat nauk kognitywnych w Nanyang Technological University w Singapurze, w maju IV Krajowej Konferencji Kognitywistycznej, "Znak, znaczenie, kontekst w badaniach kognitywistycznych", w Toruniu, oraz konferencji „Sztuka – Mózg – Emocje” w Toruniu; we wrześniu brał udział jako jeden z 5 wykładowców plenarnych w pierwszym Cognitive Science Symposium w Singapurze. Główne osiągnięcia w 2005 roku to:

1. Opracowano ideę wektorów opisu koncepcji (Concept Description Vectors) do reprezentacji semantyki i zrobienia pamięci semantycznej, która wystarcza do formułowania pytań precyzujących zagadnienie. Opracowano szereg algorytmów zbierania informacji do pamięci semantycznych z encyklopedii, słowników i tekstów bez struktury. Zastosowano tak stworzoną pamięć semantyczną do gry w 20 pytań oraz tworzenia zagadek słownych. Program do gry w 20 pytań zintegrowano z mówiącą głową i analizą mowy, tworząc prototyp awatara do gier słownych.
- Duch W, Szymanski J, Sarnatowicz T, Concept description vectors and the 20 question game. *Intelligent Information Processing and Web Mining, Advances in Soft Computing*, Springer Verlag, ISBN 3-540-25056-5 (Eds. Kłopotek, M.A., Wierzhon, S.T, Trojanowski, K., pp. 41-50, 2005.
- Szymanski J, Sarnatowicz T, Duch W, Semantic memory for avatars in cyberspace. *International Conference on Cyberworlds, Singapore 23-25 Nov. 2005*, T.L. Kunii, S.H. Soon and A. Sourin (eds), IEEE Computer Society, pp. 165-171
2. Ukazał się artykuł prezentujący realistyczne podejście do problemu zrozumienia natury świadomości (Duch W, Brain-inspired conscious computing architecture. *Journal of Mind and Behavior*, Vol. 26(1-2), 1-22, 2005). W 2006 roku na kongresie World Congress of Computational Intelligence (Vancouver) odbędzie się sesja specjalna i dyskusja panelowa na temat strategii dojścia do systemów o ludzkiej kompetencji, zorganizowana przez W. Ducha i N. Kasabova.
3. W najlepszym piśmie nauk behawioralnych i nauk o mózgu (*Behavioral and Brain Sciences, BBS*) ukazał się komentarz W. Ducha na temat podobieństwa i reguł logicznych w psychologii: W. Duch, Rules, Similarity, and Threshold Logic. *Commentary on Emmanuel M. Pothos, The Rules versus Similarity distinction. Behavioral and Brain Sciences* 28 (1): 23-23, 2005.

III. PUBLIKACJE

III.1. W 2005 roku ukazało się ogółem 57 prac, w tym:

Liczba prac opublikowanych w 2005		57
w tym:	Publikacje z listy filadelfijskiego IIN	11
	Publikacje w innych recenzowanych pismach zagranicznych	0
	Publikacje popularno-naukowe i dydaktyczne	23
	Redakcja książek naukowych, tłumaczenia	2
	Artykuły, recenzje	7

25 oryginalnych prac badawczych: w tym

11 z listy filadelfijskiego Instytutu Informacji Naukowej,

13 prac na konferencjach międzynarodowych

1 w recenzowanych wydawnictwach krajowych;

32 prace inne:

2 redakcje książek naukowych (Springer)

6 recenzji

12 = 1 praca doktorska: 2 magisterskie, 9 licencjackich/inż.

12 prac dydaktycznych dostępnych przez WWW

III.2. Liczba prac przyjętych do druku: 10

III.3. Przykłady publikacji związanych z działalnością naukową KIS:

Najważniejsze prace (max 5), finansowane z dz. statutowej (po stronie UMK).

1. Adamczak, A. Porollo, J. Meller; Combining Prediction of Secondary Structures and Solvent Accessibility in Proteins. *Proteins: Structure, Function and Bioinformatics*, 59(3): 2005, 467-75.
2. Duch W, Uncertainty of data, fuzzy membership functions, and multi-layer perceptrons. *IEEE Transactions on Neural Networks* 16(1), 2005, 10-23.
3. Duch W, Brain-inspired conscious computing architecture. *Journal of Mind and Behavior*, Vol. 26(1-2), 2005, 1-22.
4. Itert L, Duch W, Pestian J, Medical document categorization using *a priori* knowledge, *Lecture Notes in Computer Science*, Vol 3696, 641-646, 2005
5. Wagner M, Adamczak R, Porollo A, Meller J, Linear regression models for solvent accessibility prediction in proteins, *Journal of Computational Biology* 12(3), 2005, 355-369.

Pełny spis znajduje się w Załączniku 2b.