

# SPRAWOZDANIE Z DZIAŁALNOŚCI NAUKOWO-BADAWCZEJ KATEDRY INFORMATYKI STOSOWANEJ UMK W ROKU 2006

## I. TEMATY BADAWCZE

1. Liczba realizowanych tematów.

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

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

- Prowadzone badania mają w większości charakter długoterminowy.

3. Z liczby tematów zakończonych przekazano:

- do wykorzystania, - do wdrożenia 0.

4. Z liczby tematów przekazanych do wdrożenia, w 2006 r. wdrożono:

- ogółem 6, - w tym w resorcie edukacji narodowej 0.

- Nie wdrożyliśmy wyników nowych tematów ale dokonano pewnych ulepszeń w pakiecie programów GhostMiner 3 do data mining, opracowanym przy współpracy z firmą FQS Poland (oddział Fujitsu) i wdrożonym w poprzednich latach; 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.

6. Liczba realizowanych grantów KBN

2 granty własne:

- Grant KBN „Selekcja informacji i odkrywanie struktur w danych”, lata 2005-2007, kierownik W. Duch.

- Grant KBN „Meta-uczenie w inteligencji obliczeniowej”, lata 2005-2007, kier. N. Jankowski.

## II. CZĘŚĆ OPISOWA

### II.1. Przyczyny zaniechania tematów prowadzonych i rozpoczętych w 2006 roku.

W 2006 roku nie zaniechaliśmy realizacji żadnego tematu.

### II.2. Krótka charakterystyka i podsumowanie wyników uzyskanych w 2006 r.

#### A) Inteligencja obliczeniowa (computational intelligence)

Główne osiągnięcia w tej grupie tematów to:

1. Przeanalizowano zachodzące w mózgu procesy w czasie rozumienia słów, w szczególności eksperymenty psychologiczne w wykorzystaniu torowania, w których widoczne są efekty rezonansu stochastycznego. Opracowano nowe neurokognitywne podejście do symulacji procesów tworzenia nowych słów. Jest to pierwszy krok w stronę autentycznych możliwości twórczych komputerów. Rezultaty przedstawiono na konferencjach, w rozdziałach książkowych i wysłanej do druku publikacji. Program tworzący nowe słowa znalazł około 2/3 nazw pokrywających się z wymyślonymi przez ludzi.
2. Ukończono prace nad redakcją książki: W. Duch i J. Mandziuk, Challenges for Computational Intelligence. Cały materiał – w tym dwa rozdziały napisane przez W. Duchę – został wysłany do druku Springer "Studies in Computational Intelligence" Series, 2007. Książka składa się z kilkunastu rozdziałów na temat wyzwań stojących przed inteligencją obliczeniową, napisanych przez najlepszych specjalistów w tej dziedzinie.
3. W 2006 r. realizowaliśmy grant KBN na rozwój metod selekcji informacji. Dr J. Biesiada z Politechniki Śląskiej korzystając ze stypendium FNP spędził 3 miesiące w KIS pracując nad tym tematem. Rozwijano pakiet programów realizujący znane i nowe metody przydatne do dyskretyzacji i selekcji cech. Ukazała się książka „Feature extraction, Foundations and Applications”, Springer 2006, pod redakcją I. Guyon, S. Gunn, M. Nikraves, i Lofti Zadeha, a w niej 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 2006 r. realizowaliśmy też grant KBN na rozwój metod meta-uczenia, polegających na automatycznym wyborze najlepszych alternatywnych modeli dla danego punktu złożoności/dokładności. Prowadzono prace nad nowym pakietem do data mining, pierwsze prace na temat jego konstrukcji przyjęto do druku. Napisano pracę na temat k-separowalności, po raz pierwszy rozbijając zagadnienia problemów nieseparowalnych liniowo na klasy o różnej złożoności. Wprowadzono nowe cele uczenia i nowe algorytmy, które powinny pozwolić na uczenie się w przypadkach trudnych, z którymi nie radzi sobie obecnie żaden istniejący algorytm.
5. Napisano szereg prac na temat nowego, konkurencyjnego podejścia do systemów neurorozmytych, oparte na szukaniu prototypów i optymalizacji funkcji odległości, w tym probabilistycznych miarach podobieństwa. Mgr Blachnik otrzymał stypendium na 6 miesięcy w laboratorium prof. Erkki Oja w Helsinkach, gdzie będzie pracował nad tym tematem (z powodów osobistych wyjazd został przesunięty na 2007 rok).
6. W. Duch został członkiem EU COST Action B27 Group, Electric neuronal oscillations and cognition (ENOC), na koszt tego programu biorąc udział w spotkaniu w maju 2006 w Skopie, gdzie przedstawił wyniki badań zarówno swoich jak i innych grup z Polski. Przy współpracy z neurologami z Collegium Medicum UMK rozpoczęto prace nad modelowaniem pnia mózgu i analizą wpływu jego uszkodzeń na rytmy oddechowe (dane spirograficzne).
7. 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 pozwalającą na interpretację tekstu, grafów spójnych koncepcji, oraz opracowano nowy algorytm stosujący programowanie liniowe do klasyfikacji tekstów. Kilka prac już opublikowano, kolejne są w przygotowaniu. Przygotowano dane do międzynarodowego konkursu w tej dziedzinie, ogłoszonego w styczniu 2007.

8. 10-14 września odbyła się w Atenach 16th International Conference on Artificial Neural Networks (ICANN 2006), oficjalna doroczna konferencja European Neural Network Society. Prezydentem tej organizacji jest W. Duch, który pełnił również rolę Program Co-chair i był współredaktorem materiałów w objętości ponad 2000 stron wydanych w dwóch tomach w ramach Springer Lecture Notes in Computer Science, Vol. 4131 i 4132.
9. Napisano prace na temat redukcji wymiarowości danych za pomocą algorytmu skalowania wielowymiarowego (MDS) pod kątem zastosowania algorytmu do wizualizacji większych zbiorów wielowymiarowych danych. Testy na kilku zbiorach pokazały skuteczność tej metody oraz pewne skrócenie czasu dokonania wizualizacji.

## **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 wybranymi aspektami teorii funkcjonałów gęstości (DFT – Density Functional Theory) i jej zastosowaniami do obliczeń struktury energetycznej i widm elektronowo-oscylacyjnych cząsteczek, uczestnicząc też w badaniach kierowanych przez prof. K. Jankowskiego z Instytutu Fizyki UMK, dotyczących tej tematyki. Ponadto dr S. Zelek współpracował z Zespołem Teoretycznej Biofizyki Molekularnej kierowanym przez prof. W. Nowaka z Instytutu Fizyki UMK, w ramach tematu komputerowej analizy mechanizmu reakcji hydrolizy nityli. W szczególności:

1. Kontynuowano badania relacji między przestrzeniami orbitalnymi otrzymywanymi w metodach: Hartree-Focka, Bruecknera, CASSCF i w różnych wariantach teorii funkcjonałów gęstości (DFT). Przedmiotem szczególnego zainteresowania była ewolucja indeksów podobieństwa przestrzeni orbitalnych wzdłuż krzywych potencjalnych cząsteczek dwuatomowych, na przykładzie izoelektronowej serii CO, NO+, CN-. Pokazano, że w dość szerokim otoczeniu geometrii równowagowych przestrzenie orbitali Kohna-Shama zachowują bliskie podobieństwo do przestrzeni orbitali Bruecknera i mają istotny charakter wielokofiguracyjny. Dzięki temu w ramach różnych wariantów DFT uzyskuje się znacznie bardziej realistyczny niż w metodzie Hartree-Focka opis energetyki oscylacji molekularnych. Podobieństwo przestrzeni orbitalnych zachowuje się w przypadku anionów, natomiast jest zmniejszone dla kationów, w wyniku kontrakcji orbitali spowodowanej nadmiarowym ładunkiem dodatnim jąder.

- J. Wasilewski, S. Zelek, Evolution of orbital spaces along potential curves for diatomic molecules. A comparative study of Hartree-Fock, Kohn-Sham, Brueckner and multiconfigurational orbital spaces, komunikat na Central European Symposium for Theoretical Chemistry (CESTC 2006), Zakopane, 24-27.09.2006; publikacja w przygotowaniu

2. W badaniach dotyczących komputerowej analizy mechanizmu reakcji hydrolizy nityli podjęto próbę rozstrzygnięcia, przy pomocy obliczeń kwantowomechanicznych, który z zaproponowanych przez Hanga i innych (Structure 5, 691 (1997)) mechanizmów reakcji hydrolizy jest najbardziej prawdopodobny.

- L. Peplowski, K. Kubiak, S. Zelek, W. Nowak, A comparative DFT study of substrates and products of industrial enzyme nitrile hydratase, International Journal of Quantum Chemistry - wysłana do druku

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. W szczególności:

1. Dokonano oceny przydatności metod przewidywania obszarów transmembranowych do przewidywania białek GCPR
- R. Adamczak, Ł. Peplowski and W. Nowak, Performance of Neural Networks Based Transmembrane Helix Prediction Methods Applied to Mosquito Anopheles Gambiae Gprotein Coupled Odorant Receptors, in Applications of Statistical and Machine Learning in Bioinformatics; Advances in Computational and Systems Biology, eds. J. Meller and W. Nowak, Peter Lang GmbH w druku.
  - Cao B, Porollo A, Adamczak R, Jarrell M, Meller J. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. Bioinformatics. 2006 Feb 1;22(3):303-9. Epub 2005 Nov 17.

2. Przeprowadzono badania profili ekspresji genów w chorobach reumatologicznych dzieci. Wyniki zostały opisane w następującej pracy:
  - J. Meller, R. Adamczak, M. P. Scola, M. Barnes, S. D. Thompson, M. H. Passo, H. I. Brunner, D. N. Glass, and A. A. Grom; Machine Learning Analysis of Expression Profiles of Synovial Tissue Cytokines Helps Identify Patients with Systemic Onset Juvenile Rheumatoid Arthritis, in Applications of Statistical and Machine Learning in Bioinformatics; Advances in Computational and Systems Biology, eds. J. Meller and W. Nowak, Peter Lang, w druku.
3. Opracowano nową metodę badania oddziaływań międzybiałkowych:
  - Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. Proteins. 2006 Dec 6; [Epub, na okładce wydania pierwszego numeru w 2007 roku]
4. Przeprowadzono szereg zaawansowanych symulacji procesów biologicznych, ściśle przy tym współpracując z doświadczalnikami. Wyniki zawarte są w pracach:
  - Slaven BE, Meller J, Porollo A, Sesterhenn T, Smulian AG, Cushion MT. Draft Assembly and Annotation of the *Pneumocystis carinii* Genome. J Eukaryot Microbiol. 53,S1:S89-91, 2006.
  - Slaven BE, Porollo A, Sesterhenn T, Smulian AG, Cushion MT, Meller J, Large-Scale Characterization of Introns in the *Pneumocystis carinii* Genome. J Eukaryot Microbiol. 2006 Nov;53 Suppl 1:S151-3.
  - Tan M, Meller J, Jiang X. C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein. J Virol. 2006 Aug;80(15):7322-31.
5. Praca przeglądowa: Duch W, Swaminathan K, Meller J, Computational Intelligence Approaches to Rational Drug Design and Discovery, przyjęta została do druku w specjalnym wydaniu Current Pharmaceutical Design.

### **C) Modelowanie kognitywne (cognitive modeling).**

W lipcu 2006 roku W. Duch zorganizował sesję specjalną i panel dyskusyjny „Roadmap to human level intelligence” w ramach IEEE World Congress on Computational Intelligence w Vancouver, Kanada. Wspólnie z prof. J.G. Taylorem pracujemy nad wydaniem książki, opartej na dyskusji i referatach w czasie tego kongresu. Pracowaliśmy również nad kolejnym projektem na ten temat w ramach 7 Programu Ramowego UE. Szereg projektów w tym kierunku prowadzono przy współpracy z School of Computer Engineering, Nanyang Technological University w Singapurze. W ramach projektu z robotyki kognitywnej „Developmental Robot-Embedded Artificial Mind” (DREAM), napisanego przez W. Ducha, stworzono szereg programów sterujących głową androida.

W listopadzie 2006 W. Duch był współorganizatorem workshopu na temat nauk uczenia się niemowląt w Nanyang Technological University (NTU) w Singapurze. Przy współpracy z National Institute of Education w oraz NTU Division of Psychology rozpoczęto prace nad projektem stworzenia laboratorium do prowadzenia takich badań w oparciu o ideę interakcyjnej stymulacji niemowląt zaproponowaną przez W. Ducha.

Rozwijano ideę wektorów opisu koncepcji (Concept Description Vectors) w zastosowaniu do reprezentacji semantyki i tworzenia pamięci semantycznej. Rozwijano algorytmy zbierania informacji do pamięci semantycznych z encyklopedii, słowników i tekstów bez struktury, dodając mechanizmy aktywnego uczenia oraz inicjowania aktywnych dialogów w celu zdobycia brakującej informacji. Program do gry w 20 pytań zintegrowany z mówiącą głową i rozpoznawaniem oraz syntezą mowy pracuje obecnie w wersji zarówno lokalnego programu jaki i interfejsu do aplikacji sieciowych. Wyniki zostały opisane w pracach konferencyjnych i w pracy będącej w druku:

- Szymanski J, Sarnatowicz T, Duch W, Towards Avatars with Artificial Minds: Role of Semantic Memory. Journal of Ubiquitous Computing and Intelligence, American Scientific Publishers.

### III. PUBLIKACJE

#### III.1. W 2006 roku ukazało się ogółem 57 prac, w tym:

Liczba prac opublikowanych w 2006	57
w tym:	
a) Publikacje z listy filadelfijskiego IIN	11
b) nie dotyczy	
c) Publikacje w innych recenzowanych pismach	2
e) Redakcja książek naukowych, tłumaczenia	4
Artykuły, recenzje, popularno-naukowe i dydaktyczne	40

#### III.2. Liczba prac przyjętych do druku: 19

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

Najważniejsze prace (max 5).

1. Cao B, Porollo A, Adamczak R, Jarrell M, Meller J. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. *Bioinformatics*. 2006 Feb 1;22(3):303-9. Epub 2005 Nov 17.
2. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins*. 2006 Dec 6; [Epub, na okładce wydania pierwszego numeru w 2007 roku]
3. Duch, W, K-separability. *Lecture Notes in Computer Science* 4131 (2006) 188-197.
4. Matykiewicz P, Duch W, Pestian J, Nonambiguous Concept Mapping in Medical Domain, *Lecture Notes in Artificial Intelligence*, Vol. 4029 (2006) 941-950
5. Blachnik M, Duch W Prototype-based threshold rules. *Lecture Notes in Computer Science*, Vol. 4234 (2006) 1028-1037.

Pełny spis znajduje się w Załączniku 2b.