

Lidia OGIELA

AGH, AKADEMIA GÓRNICZO-HUTNICZA W KRAKOWIE
Al. Mickiewicza 30, 30-059 Kraków

Modelowanie biologiczne w systemach E-UBIAS

Dr Lidia OGIELA

Informatyk, matematyk, ekonomista. W roku 2005 uzyskała tytuł doktora nauk technicznych w dyscyplinie informatyka. Jest pracownikiem Akademii Górniczo-Hutniczej od roku 2000, gdzie zajmuje się zagadnieniami informatyki kognitywnej. Jest autorką ponad stu prac naukowych o zasięgu międzynarodowym w tym trzech książek z zakresu informatyki kognitywnej. Członek międzynarodowych towarzystw SIAM, SPIE oraz CSS-Cognitive Science Society.



e-mail: logiela@agh.edu.pl

Streszczenie

W pracy przedstawiono propozycję wykorzystania metod modelowania biologicznego opartych na kryptografii DNA do zadań znaczeniowej analizy danych. Proponowane rozwiązania zostaną omówione na przykładzie systemów E-UBIAS prowadzących analizę danych obrazowych w połączeniu z analizą identyfikacyjną (personalizacja osobowa). Wykorzystanie w procesach analizy danych kryptografii DNA pozwoli w sposób jednoznaczny na etapie identyfikacji osobowej przypisać analizowane dane do jednostki osobowej. Jednocześnie wzbogacenie systemu o procesy analizy znaczeniowej prowadzone na podstawie interpretacji semantycznej pozwoli wskazać występujące ewentualne zmiany (patologie), które dana osoba posiada.

Słowa kluczowe: kryptografia DNA, systemy E-UBIAS (Extended Understanding Based Image Analysis Systems), analiza znaczeniowa, procesy personalizacji i identyfikacji osobowej.

Biological modeling in E-UBIAS systems

Abstract

In this paper the author proposes using biological modeling methods based on the DNA cryptography for semantic data analysis. The solutions proposed are illustrated with an example of E-UBIAS (*Extended Understanding Based Image Analysis Systems*) systems which analyse the image data in combination with the identity analysis – personal verification and identification processes (personalisation of individuals). The use of DNA cryptography to analyse data makes it possible to unanimously assign the analysed data to an individual at the personal identification stage. At the same time, supplementing the system with semantic analysis processes conducted based on semantic interpretation allows the possible lesions that the person suffers from to be identified. In this publication the author proposes a new class of biometric data analysis systems supplemented with semantic analysis executed in E-UBIAS systems. Extended UBIAIS (*Understanding Based Image Analysis Systems*) systems are dedicated for semantic analysis of biometrics personal features (data).

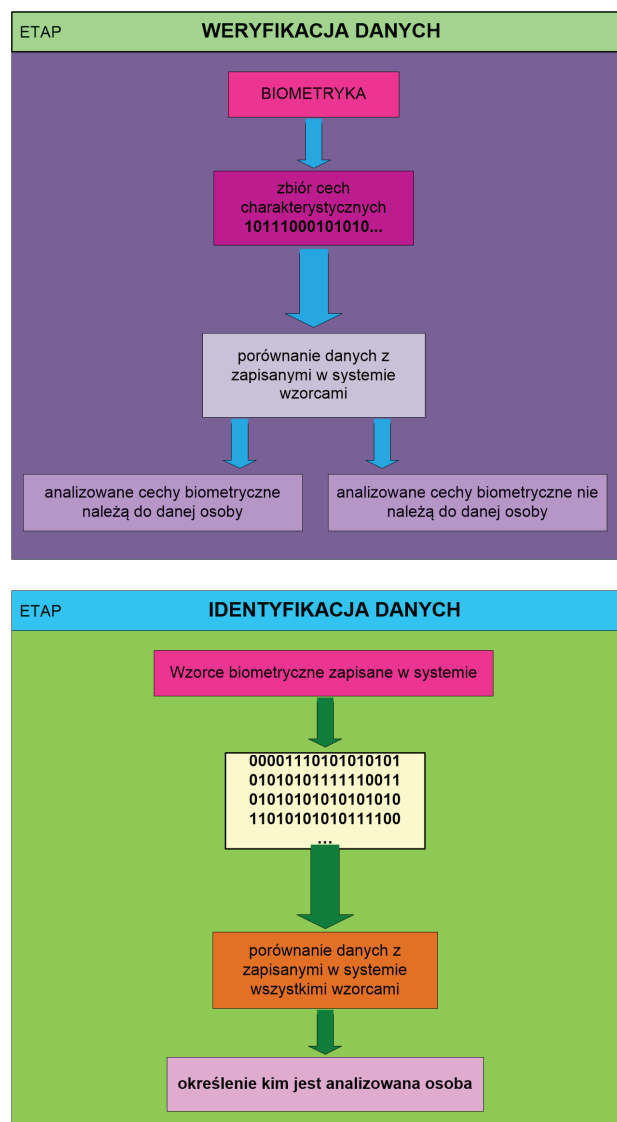
Keywords: DNA cryptography, E-UBIAS systems (Extended Understanding Based Image Analysis Systems), semantic analysis, personalisation and personal identification processes.

1. Wstęp

Coraz szybszy rozwój inteligentnych systemów analizy danych 1-3, 20 sprawia, iż dedykowane są one coraz to innym dziedzinom nauki a w wyniku prac o charakterze aplikacyjnym systemy te znajdują zastosowanie w różnorodnych dziedzinach życia. Obecnie coraz częściej prowadzone są prace dotyczące analizy znaczeniowej danych obrazowych. Tego rodzaju opracowania dotyczą budowy systemów klasy UBIAS (*Understanding Based Image Analysis Systems*), systemów analizy obrazowej. Istotą prezentowanej klasy systemów jest analiza znaczeniowa interpretowanych obrazów, w szczególności obrazów medycznych przedstawiających zmiany patologiczne różnych organów [4-19]. Prace dotyczące analizy znaczeniowej obrazów medycznych (rozumianych jako wybrana klasa danych obrazowych) prowadzone są w odniesieniu

do różnorodnych danych. Ta różnorodność danych dotyczy różnych rodzajów zmian patologicznych występujących w różnorodnych częściach ludzkiego ciała. Badania takie były prowadzone w zakresie analizy obrazów medycznych następujących organów ludzkich – centralny układ nerwowy – rdzeń kręgowy [3], złamania i deformacje kości długich kończyn [5], zmiany patologiczne kości nadgarstka, dłoni oraz kości stopy [11].

Szczególną klasą systemów UBIAS są systemy prowadzące analizę obrazów kości dłoni z uwagi na fakt, iż to właśnie ta klasa systemów może zostać wykorzystana nie tylko do analizy znaczeniowej patologii kości dłoni, ale także stała się podstawą do budowy nowej klasy systemów E-UBIAS (*Extended Understanding Based Image Analysis Systems*). Systemy E-UBIAS prowadzą analizę personalizacyjną oraz identyfikacyjną (rys. 1)



Rys. 1. Procesy weryfikacji i identyfikacji osobowej
Fig. 1. Personal verification and identification processes

W systemach E-UBIAS prowadzone są procesy weryfikacji i identyfikacji osobowej wzbogacone o analizę obrazową kształtu dłoni oraz wykrywania patologii dłoni. Systemy te powstały w wyniku prowadzonych prac dotyczących analizy obrazowej kości dłoni w systemach UBIAS [4-19].

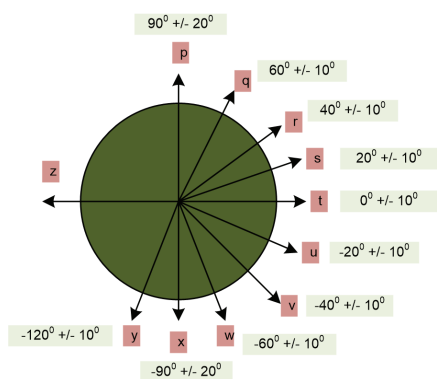
Celem niniejszej pracy jest zatem charakterystyka nowej klasy systemów znaczeniowej analizy danych E-UBIAS dedykowanych analizie wybranych danych obrazowych (np. obrazy RTG kości dłoni) w połączeniu z analizą identyfikacyjną (tj. personalizacja osobowa). Ponadto dokonując charakterystyki przykładu wykorzystania metod modelowania biologicznego opartego na kryptografii DNA do zadań znaczeniowej analizy danych, możliwe będzie wskazanie kierunków dalszego rozwoju systemów analizy znaczeniowej.

2. Analiza kości dłoni fundamentem systemów klasy E-UBIAS

Systemy prowadzące analizę znaczeniową obrazów przedstawiających zmiany kości dłoni były opisywane w wielu pracach naukowych, a szczególnie w [14, 19].

Definicja gramatyki formalnej zawiera nazwy kości występujące w obrębie dłoni, do których zalicza się następujące z nich [11]:

- kość łokciowa – ulna (ul),
- kość łódeczkowata – os scaphoideum (sc),
- kość księżycowata – os lunatum (lt),
- kość trójgraniasta – os triquetrum (tq),
- kość grochowata – os pisiforme (pf),
- kość czworoboczna większa – os trapezium (tz),
- kość czworoboczna mniejsza – os trapezoideum (tm),
- kość główkowata – os capitatum (c),
- kość haczykowata – os hamatum (h),
- kość śródręcza – os metacarpale (m),
- trzeszczka ręki – ossa sesamoidea (ses),
- paliczek bliższy – phalanx proximalis (pip),
- paliczek środkowy – phalanx media (pm),
- paliczek dalszy – phalanx distalis (pd).



Rys. 2. Graf Γ_{hand} określający relacje pomiędzy kośćmi dłoni
Fig. 2. The Γ_{hand} graph defining relations between hand bones

Definicja formalnej gramatyki grafowej w oparciu o formalizmy lingwistyczne przedstawia się następująco:

$$G_{hand} = (N_{hand}, T_{hand}, \Gamma_{hand}, S_{hand}, P_{hand})$$

gdzie:

N_{hand} – oznacza zbiór symboli nieterminalnych, zdefiniowany następująco:

$N_{hand} = \{ST, ULNA, OS\ SCAPHOIDEUM, OS\ LUNATUM, OS\ TRIQUETRUM, OS\ PISIFORME, OS\ TRAPEZIUM, OS\ TRAPEZOIDEUM, OS\ CAPITATUM, OS\ HAMATUM, m1, m2, m3, m4, m5, SES1, SES2, PIP1, PIP2, PIP3, PIP4, PIP5, PM2, PM3, PM4, PM5, PD1, PD2, PD3, PD4, PD5\}$

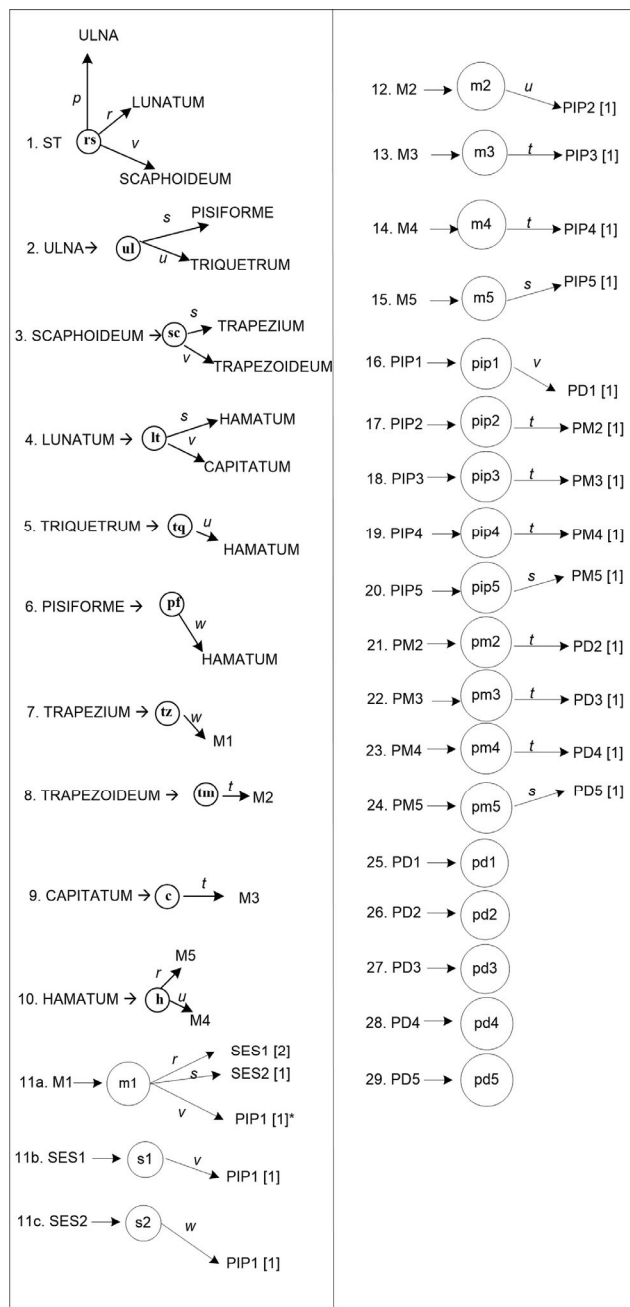
T_{hand} – oznacza zbiór symboli terminalnych, zdefiniowany następująco:

$T_{hand} = \{rs, ul, sc, lt, tq, pf, tz, tm, c, h, m1, m2, m3, m4, m5, s1, s2, pip1, pip2, pip3, pip4, pip5, pm2, pm3, pm4, pm5, pd1, pd2, pd3, pd4, pd5\}$

$\Gamma_{hand} = \{p, q, r, s, t, u, v, w, x, y, z\}$ – graf określający relacje pomiędzy poszczególnymi kośćmi dłoni (rys. 2)

$S_{hand} = ST$ – oznacza symbol startowy gramatyki G_{hand}

P_{hand} – oznacza zbiór produkcji (rys. 3).



Rys. 3. Zbiór produkcji P_{hand} .
Fig. 3. The set of productions P_{hand}

Zaproponowany formalizm lingwistyczny do znaczeniowej analizy danych obrazowych posłużył także do zaproponowania opisu formalnego analizy cech biometrycznych dłoni. Zbiór cech biometrycznych wzbogacony o elementy znaczeniowej analizy danych pozwalają na poszerzoną analizę identyfikacyjną. Do analizy cech biometrycznych kości zostało zaproponowane rozwiązanie formalne oparte na definicji zbioru określającego zbiór biometrycznych cech charakterystycznych dłoni:

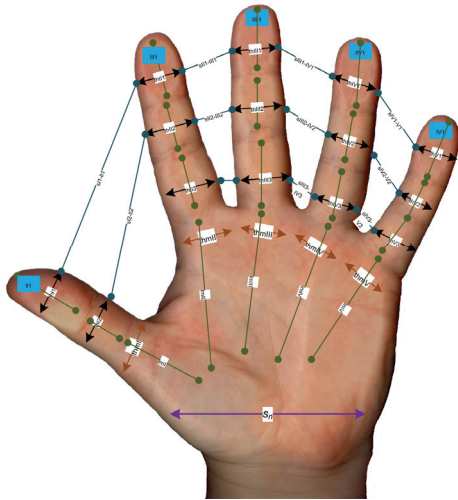
$$B_{hand} = \{th_{ij}, l_{ij}, s_{ij}, thm_i, lm_i, s_n\}$$

gdzie:

th_{ij} – oznacza grubość kości i -tego palca i j -tego paliczka dla $i = \{I, II, III, IV, V\}, j = \{1, 2, 3\}$

l_{ij} – długość kości i -tego palca i j -tego paliczka
 s_{ij-ij} – wielkość obszarów pomiędzy poszczególnymi kośćmi dłoni
 thm_i – grubość i -tej kości śródrezcza
 lm_i – długość i -tej kości śródrezcza
 s_n – wielkość kości nadgarstka.

Elementy zbioru B_{hand} przedstawiono na rysunku 4.



Rys. 4. Składowe zbioru B_{hand}
Fig. 4. Elements of the set B_{hand}

Zaproponowany zbiór B_{hand} określający cechy biometryczne dłoni został poszerzony o następujące elementy określające kształty odcisków linii papilarnych dłoni:

$$L_{hand} = \{p_i, o_j\}, i = \{1, \dots, 5\}, j = \{1, \dots, 3\}$$

gdzie:

p_i – oznacza kształt odcisku palca i -tej ręki (od pierwszego do piątego)
 o_j – kształt odcisków biometrycznych śródrezcza (od pierwszego do trzeciego).

Składowe zbioru L_{hand} przedstawiono na rysunku 5.



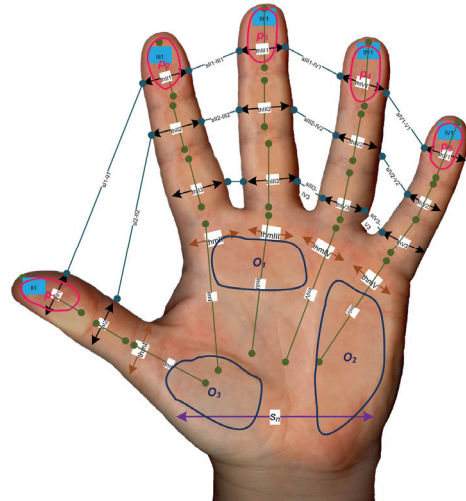
Rys. 5. Składowe zbioru L_{hand}
Fig. 5. Elements of the set L_{hand}

Połączenie zbiorów B_{hand} i L_{hand} daje zbiór BL_{hand} (rys. 6) następującej postaci:

$$BL_{hand} = \{th_{ij}, l_{ij}, s_{ij-ij}, thm_i, lm_i, s_n, p_i, o_j\}$$

gdzie poszczególne elementy zbioru zdefiniowane są jak powyżej.

Zbiór BL_{hand} określa cechy biometryczne oraz kształty odcisków dłoni, a także kształt linii papilarnych palców, na podstawie których możliwa jest analiza biometryczna. Zapisane w systemie E-UBIAS dane dotyczące cech biometrycznych pozwalają na identyfikację personalną i weryfikację osobową.

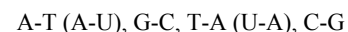


Rys. 6. Składowe zbioru BL_{hand}
Fig. 6. Elements of the set BL_{hand}

3. Kryptografia DNA w systemach E-UBIAS

Systemy klasy E-UBIAS wzbogacone o elementy kryptografii DNA wykorzystywane do celów generowania kluczy w oparciu o kody DNA przynależne każdej osobie prowadzą do określenia wektora cech biometrycznych danej osoby. Każda osoba posiada niepowtarzalny kod DNA, przy pomocy którego można ją opisać.

Ponadto można wykorzystując indywidualny nośnik informacji genetycznej DNA połączyć go z cechami biometrycznymi opisanymi przy pomocy zbioru BL_{hand} . Polimer nukleotydów (DNA) złożony jest z zasad purynowych, w skład których wchodzi adenina (A) i guanina (G) oraz zasad pirymidynowych, do których zaliczane są cytozyna (C) i tymina (T) oraz reszty deoksyrybozowe i reszty kwasu fosforowego. W niektórych polimerach nukleotydów zamiast tyminy może występować uracyl (U). Zasady połączone są w pary według następującego schematu:



Połączenie zasad purynowych i pirymidynowych pozwala zapisać kod DNA w postaci jednoznacznej, a sam kod DNA może posłużyć do analizy biometrycznej. Zakodowany kod DNA stanowi podstawę analizy kryptograficznej, identyfikacyjnej i weryfikacyjnej.

Takie połączenie przyjęte zostało do zdefiniowania zbioru, na podstawie którego możliwe jest wykonanie analizy biometrycznej, identyfikacyjnej oraz weryfikacyjnej:

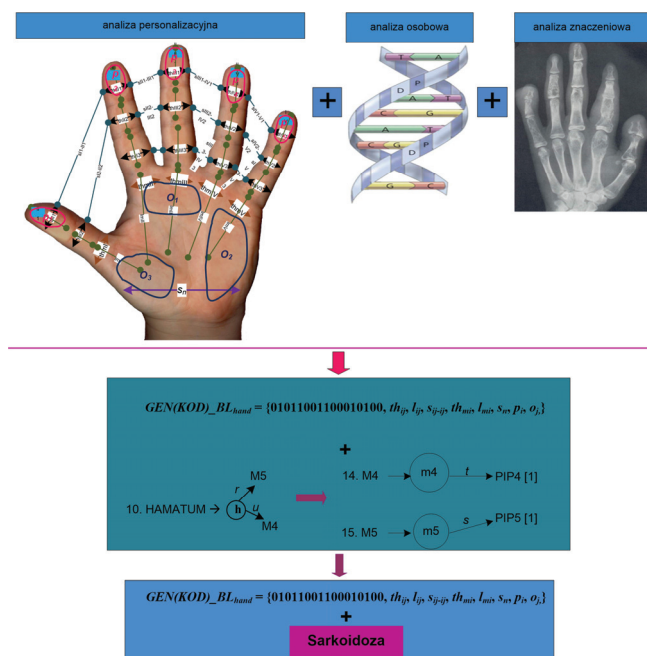
$$GEN_{BL_{hand}} = \{\text{kod DNA (A-T G-C T-A C-G)}, th_{ij}, l_{ij}, s_{ij-ij}, thm_i, lm_i, s_n, p_i, o_j\}$$

Kodowanie informacji genetycznej ma na celu utajnianie zawartych w niej informacji. Zbiór $GEN_{BL_{hand}}$ z zakodowaną postacią informacji genetycznej przyjmuje następującą postać:

$$GEN(KOD)_{BL_{hand}} = \{01011001100010100, th_{ij}, l_{ij}, s_{ij-ij}, thm_i, lm_i, s_n, p_i, o_j\}$$

W wyniku kodowania kodu DNA, identyfikacji osobowej oraz analizy kognitywnej prowadzonej w systemach E-UBIAS, w wektorach cech przypisanych danej osobie w systemach biometrycznej analizy danych mogą znajdować się informacje dotyczące (rys. 7):

- kodu DNA,
- cech biometrycznych,
- cech fizycznych,
- deformacji i patologii kości dłoni.



Rys. 7. System biometrycznej analizy danych wzbogacony o analizę znaczeniową prowadzoną w systemach E-UBIAS

Fig. 7. A biometric data analysis system supplemented with semantic analysis executed in E-UBIAS systems

Na rysunku 7. przedstawiony został opis działania systemów biometrycznej analizy danych wzbogaconych o elementy znaczeniowej analizy danych obrazowych. Taki rodzaj analizy pozwala na przypisanie danej osobie jej informacji genetycznej w postaci zakodowanej, danych biometrycznych oraz cech fizycznych i ewentualnych zmian patologicznych. Wieloetapowy proces identyfikacji personalnej znacznie ogranicza możliwości popełnienia błędu wynikającego z nieprawidłowego procesu wnioskowania oraz weryfikacji personalnej.

Przedstawiony na rysunku 7. etap analizy personalnej dotyczy procesów wyłonienia indywidualnych cech osobowych, które w połączeniu z informacjami zawartymi w kodzie DNA oraz z informacjami semantycznymi dotyczącymi występujących u danej osoby jednostek chorobowych (zmian chorobowych), stają się podstawą działania systemów E-UBIAS, systemów znaczeniowej analizy danych. Informacje te zapisywane w rekordach personalnych pozwalając prowadzić rozbudowaną (wnikliwą) analizę personalną danej osoby, grup osobowych oraz całych zbiorowości osób, u których będzie np. występować identyczna jednostka chorobowa. W takich przypadkach system rozpoznaje daną osobę na podstawie zbioru cech $GEN(KOD)_{BL_hand}$, który zawiera informacje odnośnie cech personalnych i osobowych każdej osoby z osobna (rys. 7).

W wektorach tych można zamieścić także informacje dotyczące innych cech patologicznych przynależnych danej osobie np. zmian w obszarze tętnic wieńcowych, deformacji kości długich kończyn, zmian w narządach wewnętrznych. Tego rodzaju informacje mogą służyć do wnikliwej analizy danych, która nie jest prowadzona w klasycznych systemach weryfikacyjnych, ale która może w szczególnych przypadkach posłużyć złożonym zadaniom identyfikacji osobowej.

4. Wnioski

Przedstawiona w niniejszej pracy idea biometrycznej analizy danych w systemach identyfikacji i weryfikacji personalnej prowadzona przy wykorzystaniu kognitywnych systemów klasy

E-UBIAS pozwala na analizę i interpretację złożonych (niezwykle rozbudowanych) zbiorów danych. Jednocześnie systemy tego rodzaju mogą wykonywać nie tylko prostą analizę identyfikacyjną na podstawie zgromadzonych w systemie cech biometrycznych (takich jak chociażby kształt linii papilarnych, czy analiza tęczy oka), ale mogą wykonać analizę różnorodnych zbiorów danych. Ta różnorodność danych pozwala także na zastosowanie różnorodnych sposobów analizy. Możliwość zgromadzenia w bazach wiedzy kognitywnych systemów informacyjnych, informacji dotyczących cech biometrycznych, cech fizycznych, cech budowy poszczególnych organów ludzkich, występowania zmian patologicznych, a także kodu DNA (w postaci zakodowanej i niedostępnej dla przeciętnego użytkownika systemu) stwarza możliwości poszerzenia systemów E-UBIAS o dowolne dane i informacje. Procesy modelowania danych zawartych w systemach E-UBIAS pozwalają na całkowite utajnienie zawartych w systemach kognitywnych E-UBIAS danych personalnych.

Niniejsza praca została sfinansowana ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2012/05/B/HS4/03625.

5. Literatura

- [1] Cohen H., Lefebvre C. (Eds.): Handbook of Categorization in Cognitive Science. Elsevier, The Netherlands, 2005.
- [2] Meystel A.M., Albus J.S.: Intelligent Systems – Architecture, Design, and Control. A Wiley-Interscience Publication John Wiley & Sons, Inc., Canada, 2002.
- [3] Ogiela L.: Cognitive Systems for Medical Pattern Understanding and Diagnosis. in: I. Lovrek, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems. Springer-Verlag Berlin Heidelberg, pp. 394-400, 2008.
- [4] Ogiela L.: Syntactic Approach to Cognitive Interpretation of Medical Patterns. in: Caihua Xiong, Honghai Liu, Yongan Huang, Youlun Xiong (Eds.), Intelligent Robotics and Applications, Springer-Verlag Berlin Heidelberg, pp. 456-462.
- [5] Ogiela L.: UBIAS Systems for Cognitive Interpretation and Analysis of Medical Images. Opto-Electronics Review 17(2), Springer Verlag Heidelberg, pp. 166-179, 2009.
- [6] Ogiela L.: Computational Intelligence in Cognitive Healthcare Information Systems, in: Bichindaritz I., Vaidya S., Jain A., Jain L. (Eds.), Computational Intelligence in Healthcare 4, Advanced Methodologies. Studies in Computational Intelligence, Springer-Verlag Berlin Heidelberg, vol. 309, pp. 347-369, 2010.
- [7] Ogiela L.: Cognitive Informatics in Automatic Pattern Understanding and Cognitive Information Systems. in: Yingxu Wang, Du Zhang, Witold Kinsner (Eds.), Advances in Cognitive Informatics and Cognitive Computing. Studies in Computational Intelligence (SCI), Springer-Verlag Berlin Heidelberg, vol. 323, pp. 209-226, 2010.
- [8] Ogiela L., Ogiela M.R.: Cognitive Techniques in Visual Data Interpretation. Studies in Computational Intelligence 228, Springer-Verlag Berlin Heidelberg, 2009.
- [9] Ogiela L., Ogiela M.R.: Semantic Analysis Processes in UBIAS Systems for Cognitive Data Analysis. in: Ilun You, Leonard Barolli, Feilong Tang, Fatos Xhafa (Eds.), IMIS 2011 - 2011 Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Seoul, Korea, 30 June-2 July 2011, pp. 182-185, 2011.
- [10] Ogiela L., Ogiela M.R.: Semantic Analysis Processes in Advanced Pattern Understanding Systems. in: Tai-hoon Kim, Hojjat Adeli, Rosslin John Robles, Maricel Balitanas (Eds.), Advanced Computer Science and Information Technology. Communications in Computer and Information Science 195, Springer-Verlag, Berlin Heidelberg, pp. 26-30, 2011.
- [11] Ogiela L., Ogiela M.R.: Advances in Cognitive Information Systems. COSMOS 17, Springer-Verlag, Berlin-Heidelberg, 2012.
- [12] Zhong N., Raś Z.W., Tsumoto S., Suzuki E. (Eds.): Foundations of Intelligent Systems. 14th International Symposium, ISMIS, Maebashi City, Japan 2003.