

Od inżynierii dokumentu do inżynierii wiedzy

Bogdan Wiszniewski
Katedra Inżynierii Wiedzy

1. Wstęp

Dzięki możliwości włączenia w procesy obliczeniowe realizowane przez współczesne systemy rozproszone ludzkiej inteligencji, staje się możliwe rozwiązywanie problemów dla których rozwiązania algorytmiczne nie istnieją lub przynajmniej nie są jeszcze znane. Kluczowe w tej integracji są *dokumenty elektroniczne*, czytelne równocześnie dla człowieka i komputera dzięki zastosowaniu języka XML, oraz *serwisy WWW*, pozwalające budować aplikacje na zasadzie dynamicznego wyszukiwania i konfigurowania dostępnych usług. Ten model przetwarzania rozproszonego stwarza nieograniczone możliwości budowania systemów informatycznych wspierających realizację ludzkich przedsięwzięć wymagających dostępu do informacji, szczególnie w dziedzinach takich jak gospodarka, administracja rządowa, ochrona zdrowia, kultura, bezpieczeństwo narodowe czy ochrona środowiska.

2. Historia katedry

Za punkt początkowy ewolucji zasugerowanej w tytule należy przyjąć rok 1989, gdy z Zakładu Systemów Liczących kierowanego przez legendarnego docenta Tadeusza Bartkowskiego wyłonił się Zakład Inżynierii Oprogramowania, którego kierownictwo objął (podówczas) docent Witold Malina. Profil badań i dydaktyki nowego zakładu został ukierunkowany na zagadnienia związane z metodami wytwarzania oprogramowania, strukturami języków programowania, sztuczną inteligencją i zaawansowanymi aplikacjami z zakresu grafiki komputerowej, przetwarzania i rozpoznawania obrazów. W tym okresie aktywność pracowników Katedry pozostawała jednak w głównym nurcie inżynierii oprogramowania. Po radykalnej reformie struktury Wydziału w roku 1992 i powołaniu kilkunastu katedr, specjalizujących się w dziedzinach pokrywających szczerlnie całe spektrum technologii informacyjnych, zagadnienia badawcze i przedmioty Katedry oddzieliły się ostatecznie od głównego nurtu inżynierii oprogramowania i skoncentrowały się na aspektach inteligentnej interakcji człowieka z komputerem. Tak powstała Katedra Techniki Programowania wprowadziła do programu studiów na Wydziale nową specjalność magisterską *Techniki Multimedialne*. Z upływem czasu, w miarę realizacji w Katedrze kolejnych projektów badawczych finansowanych z funduszy krajowych i europejskich, zagadnienia inteligentnej interakcji będące przedmiotem zainteresowania pracowników Katedry były stopniowo rozszerzane z relacji człowiek-komputer na relacje człowiek-człowiek za pośrednictwem komputera, a także komputer-komputer w trybie zdalnym – prowadząc do takich zagadnień jak współpraca ludzi i komputerów w cyberprzestrzeni, rzeczywistość wirtualna czy systemy uczące się. Ostatecznie w roku 2004 w związku z kolejną znaczną przebudową programu studiów na Wydziale, Katedra Techniki Programowania przekształciła się w Katedrę Inżynierii Wiedzy, prowadzącą specjalność *Przetwarzanie Dokumentów Cyfrowych*, przemianowaną nieco później na *Inżynierię Dokumentu*. Obecna nazwa specjalności określa jedną z najnowszych, dopiero kształtujących się dziedzin informatyki, zapoczątkowaną powołaniem w 2004 roku przez Wydział Informatyki Uniwersytetu Berkeley w Kalifornii centrum badawczo-dydaktycznego pod nazwą *Center for Document Engineering*. Katedra Inżynierii Wiedzy jest drugim ośrodkiem badawczym na świecie i pierwszym w Polsce, w całości zorientowanym na badania i dydaktykę w tej nowej dziedzinie.

3. Badania i dydaktyka

Ze względu na dynamiczny rozwój bardzo młodej dziedziny jaką jest inżynieria dokumentu, aktualizacja treści przedmiotów wykładanych w Katedrze musi odbywać się praktycznie co 1-2 lata. Z jednej strony wprowadza to dość znaczne obciążenia dla kadry, zmuszanej do nieustannej modyfikacji realizowanych przedmiotów i rozszerzania ich o nowe treści, ale z drugiej strony pozytywnie stymuluje wykładowców i licznych doktorantów do intensywnej pracy badawczej. Wśród zagadnień będących przedmiotem zainteresowania Katedry można wyróżnić trzy główne osie tematyczne:

1. Prezentacja informacji w postaci graficznej, obejmująca takie zagadnienia jak wizualizacja informacji, grafika trójwymiarowa, systemy graficzne i rzeczywistość wirtualną.
2. Pozyskiwanie i gromadzenie informacji, obejmujące takie zagadnienia jak przetwarzanie obrazów, systemy widzenia komputerowego, dokumenty i biblioteki cyfrowe.
3. Inteligentna interakcja systemów i ludzi w cyberprzestrzeni, obejmująca takie zagadnienia jak przetwarzanie języka naturalnego, systemy uczące się, wirtualne zespoły robocze i multimedialne systemy interaktywne.

Działania prowadzone w tych trzech strumieniach są realizowane przy założeniu komplementarności zadań badawczych, tworzących rdzeń wszystkich projektów realizowanych w Katedrze. Wspomniana komplementarność zadań pozwala na harmonijną współpracę tworzonych dynamicznie zespołów zadaniowych.

Na przykład, aktualnie grupa pracowników i doktorantów skupionych wokół prof. Witolda Maliny podejmuje zagadnienia z zakresu przetwarzania obrazów i widzenia komputerowego, podczas gdy dr Mariusz Szwoch koordynuje prace w zakresie bibliotek cyfrowych dokumentów muzycznych, kierując zespołem wykonawców własnego projektu badawczego finansowanego przez KBN. Dr Jacek Lebieź koordynuje prace związane z interaktywną wizualizacją sceny dynamicznej, zainspirowane zakończonym niedawno projektem celowym KBN, w którym kierował zespołem realizującym zadanie interaktywnej wizualizacji terenu na klastrze obliczeniowym, a prof. Bogdan Wiszniewski, oprócz koordynacji wszystkich zespołów zadaniowych pracujących w Katedrze (mającej na celu integrację ich działań), jeszcze dodatkowo kieruje bezpośrednio zespołem wykonawców opracowujących nowe architektury dokumentów elektronicznych umożliwiających innowacyjne modele przetwarzania w otwartych systemach rozproszonych.

4. Osiągnięcia

Pracownicy Katedry skupieni we wspomnianych wcześniej zespołach zadaniowych mogą poszczycić się wieloma wynikami, osiągniętymi zarówno indywidualnie jak i zespołowo. Do najważniejszych indywidualnych osiągnięć badawczych pracowników Katedry należą:

1. Opracowanie przez prof. Witolda Malinę metody optymalizacji liniowej transformacji przestrzeni obrazów na bazie analizy dyskryminacyjnej i rozwinięcia K-L. W tym celu sformułował rozszerzone kryterium Fishera i dwu-parametrowe kryterium Fishera, które opublikował odpowiednio w *IEEE Trans. PAMI*, w 1981 oraz *IEEE Trans. SMC-B* w 2001. Prace te były bezpośrednią inspiracją dla dwu rozpraw doktorskich: dr Agaty Kołakowskiej i dr Macieja Smiatacza obronionych w Katedrze w latach 2002 i 2003. Wyniki opublikowane w tych pracach dorobiły się już kilkudziesięciu cytowań w literaturze światowej
2. Rozwiązanie przez prof. Bogdana Wiszniewskiego problemu testowania ścieżek programów sekwencyjnych w obecności pętli, co pozwoliło skutecznie zredukować liczbę testów jakie należy wykonać dla pełnego przetestowania programu. Zbudowane na podstawie tego rozwiązania narzędzie CAST zostało wdrożone w środowisku przemysłowym w USA, a wyniki opisane w piśmie *Software Practice & Experience* w 1991 były wielokrotnie cytowane w literaturze światowej, w tym w *Encyklopedia of Software Engineering* wydanej przez John Wiley & Sons w 1994 roku.
3. Metoda testowania programów równoległych i rozproszonych opracowana przez prof. Bogdana Wiszniewskiego, która wykorzystuje ideę syntezy maszyn protokołu do tworzenia reguł deterministycznego wykonywania eksperymentów z kodem programu, pozwalających wyeliminować skutki eksplozji kombinatorycznej akcji komunikacyjnych będących wadą wcześniej stosowanego modelu przeplotowego. Narzędzie CAST przeznaczone do powtarzalnego wykonywania programów równoległych i rozproszonych napisanych w języku C z wykorzystaniem biblioteki PVM zostało wdrożone w serii europejskich projektów badawczych COPERNICUS i opisane szczegółowo wraz z metodą w międzynarodowej monografii przygotowanej wspólnie z prof. Henrykiem Krawczykiem pt. *Analysis and Testing of Distributed Software Applications* i wydanej w W. Brytanii przez Wiley w 1998 roku. Monografia ta dorobiła się kilkunastu cytowań.
4. Opracowanie przez dr Jacka Lebieźa kilku nowych algorytmów rysowania linii, m. in. bardzo efektywnego całkowitoliczbowego algorytmu punktu środkowego (ang. *midpoint*) dla krzywych Béziera.
5. Opracowanie przez dr Mariusza Szwocha lingwistycznego modelu opisu i rozpoznawania notacji muzycznej oraz algorytmów segmentacji i rozpoznawania obrazów, wdrożonych w kompletnym systemie rozpoznawania partytur drukowanych.
6. Opracowanie przez dr inż. Jana Daciuka nowych algorytmów tworzenia słowników w postaci automatów skończonych. Algorytmy zostały opublikowane m. in. w 2 artykułach w *Computational Linguistics*, są wielokrotnie cytowane i zostały zaimplementowane w postaci programów i bibliotek używanych w wielu projektach realizowanych przez niego w różnych ośrodkach na świecie.

Wymienione wyżej wyniki naukowe przełożyły się na szereg konkretnych systemów i narzędzi informatycznych, zrealizowanych siłami pracowników Katedry i zespołów przez nich kierowanych w ramach międzynarodowych projektów europejskich, bilateralnych programów międzynarodowych, projektów badawczych i celowych finansowanych na szczeblu krajowym (KBN, MNiSzW), a także umów współpracy z przemysłem krajowym. Wśród projektów europejskich wymienić należy wspomnianą wcześniej serię projektów COPERNICUS, finansowanych z 4 Programu Ramowego UE i realizowanych na Wydziale wspólnie z pracownikami i doktorantami z Katedry Architektury Systemów Komputerowych, oraz projekt MEMORIAL finansowany z 5 Programu Ramowego UE, w wyniku którego powstał zestaw narzędziowy DDW do automatycznego wydobywania danych osobowych z archiwalnych dokumentów pisanych na maszynie. Wdrożenia zrealizowane w projektach finansowanych z funduszy krajowych obejmują m. in. gdański system informacyjny dla osób niepełnosprawnych GDASKON, system Vis3D do interaktywnej wizualizacji 3D terenu z wykorzystaniem klastra obliczeniowego, oraz narzędzia ScoreExplorer i Guido do automatycznego

rozpoznawania treści dokumentów (partytur) muzycznych. W ramach lokalnych projektów wydziałowych w Katedrze opracowano i wdrożono m. in. systemy AATE do rozpoznawania i przetwarzania treści ankiet egzaminacyjnych oraz PollReader do rozpoznawania i przetwarzania treści studenckich ankiet oceny nauczycieli akademickich.

5. Wizja przyszłości katedry

Ze względu na dynamikę młodej dziedziny, którą Katedra od kilku lat intensywnie rozwija, jej perspektywy na przyszłość rysują się obiecująco. Zbiegają się one z faktem niedawnej akcesji Polski do Unii Europejskiej, w rozwoju której jednym z głównych priorytetów jest budowa społeczeństwa wiedzy, w którym dokumenty elektroniczne i rozmaite „e-systemy” je przetwarzające będą odgrywały kluczową rolę. Za projektem takiego społeczeństwa idą olbrzymie środki, które Polska zaczęła z wolna absorbować. Środki te są udostępniane w ramach rozmaitych programów operacyjnych i funduszy strukturalnych z przeznaczeniem na projekty badawcze, rozwój infrastruktury i zasobów ludzkich, wspieranie dydaktyki w zakresie eksploracji wiedzy, itp. Katedra bierze aktywny udział w inicjowaniu i przygotowywaniu wniosków badawczych i inwestycyjnych do wielu programów finansowanych z tych funduszy, przy ścisłej współpracy z firmami Wybrzeża z sektora małych i średnich przedsiębiorstw innowacyjnych, jednostek samorządu terytorialnego na wszystkich szczeblach od gminnego po wojewódzki, które wiążą się z osiami tematycznymi wspomnianymi wcześniej w p.2. Aktualnie przygotowywane jest uruchomienie czterech projektów, przewidzianych do realizacji przez różne konsorcja krajowe i międzynarodowe: wirtualnego muzeum dziedzictwa kultury materialnej Pomorza, inteligentnych dokumentów rozproszonych dla potrzeb administracji państwowej i sądownictwa, negocjacyjne protokoły uzgadniania współpracy jednostek eBiznesu z wykorzystaniem rejestru ebXML oraz wdrażania metod widzenia komputerowego w procesach technologicznych wybranych firm Wybrzeża w ramach europejskiej sieci transferu technologii TTN. Jednym z podstawowych założeń tych inicjatyw jest osiągnięcie sukcesu nie tylko na płaszczyźnie naukowej (w postaci nowych opracowanych technologii i narzędzi) ale także wdrożeniowych, przy ścisłej współpracy z przemysłem krajowym.



Bogdan Wiszniewski, ur. 1954 r., mgr inż. (1977), dr inż. (1984), dr hab. (1998), profesor (2006). Dyscyp. nauk. informatyka, spec. inżynieria oprogramowania i inżynieria dokumentu. Ponad 150 publikacji, w tym 3 monografie w John Wiley & Sons, MIKOM i PWN, kilkanaście rozdziałów w książkach Springer-Verlag, Nova Science, Elsevier, North Holland, Chapman&Hall i MIKOM, artykuły w czasopismach Softw. Pract.&Exp., Lecture Notes in Comp. Sci, Microproc.& Microprogr. J., oraz referatów w materiałach konferencji m-nar. i kraj. Kierował lub był głównym wykonawcą wielu projektów finansowanych z funduszy europejskich i krajowych, w tym 3 projektów Copernicus, 2 projektów z 4 i 5 Programu Ramowego UE i 6 grantów badawczych KBN.



Witold Malina, ur. 1937 r., mgr inż. (1960), dr inż. (1974), dr hab. (1983), profesor (2003). Dyscyp. nauk. informatyka, spec. metody przetwarzania i rozpoznawania obrazów oraz inżynieria dokumentu. Ponad 120 publikacji i 9 książek (monografii i podręczników akademickich). Był głównym wykonawcą projektów europejskich CPBR i TEMPUS JEP, 4 grantów badawczych KBN oraz autorem systemów w Katedrze takich, jak system analizy EKG, system rozpoznawania obrazów, system automatycznej analizy testów egzaminacyjnych. Po przejściu w 2007 r. na emeryturę nadal aktywnie uczestniczy w pracy naukowej i dydaktycznej.



Jacek Lebieź, ur. 1963 r., mgr inż. informatyk (1986), mgr matematyki (1989), dr inż. (2001). Dyscyp. nauk. informatyka, spec. grafika komputerowa. Autor lub współautor kilkudziesięciu publikacji krajowych i międzynarodowych, w tym artykułów w Machine Graphics & Vision, Lecture Notes in Comp. Sci. i Lecture Notes in Artif. Intelligence, oraz referatów konferencyjnych. Główny wykonawca w kilku projektach finansowanych z funduszy europejskich i krajowych, a także kilku projektów realizowanych w ramach umów dwustronnych z przemysłem regionalnym.



Mariusz Szwoch, ur. 1967 r., mgr inż. (1991), dr inż. (2002). Dyscyp. nauk. informatyka, spec. przetwarzanie i rozpoznawanie obrazów oraz biblioteki cyfrowe. Ponad 40 publikacji, w tym 2 rozdziały w książkach, 1 podręcznik akademicki, kilkanaście artykułów w czasopismach i wydawnictwach, m. in. Lecture Notes in Comp. Sci. Kierownik indywidualnego grantu badawczego KBN oraz autor 2 wdrożeń przemysłowych.



Jan Daciuk, ur. 1962 r., mgr inż. (1986), dr inż. (1999). Dyscyp. nauk. informatyka, spec. automaty skończone i przetwarzanie języka naturalnego. Autor lub współautor kilkudziesięciu publikacji międzynarodowych, w tym artykułów w Computational Linguistics, Theoretical Computer Science, Journal of Natural Language Engineering i Lecture Notes in Computer Science. Autor kilku pakietów oprogramowania do przetwarzania języka naturalnego z użyciem automatów skończonych, które są wykorzystywane w różnych projektach zarówno krajowych, jak i zagranicznych (np. Alpino).