

**Jacek KITOWSKI, Krzysztof BORYCZKO, Marian BUBAK,
Witold DZWINEL, Włodzimierz FUNIKA, Renata SŁOTA***

AKADEMIA GÓRNICZO-HUTNICZA, WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI I ELEKTRONIKI,
KATEDRA INFORMATYKI

Rozwój środków i narzędzi informatyki dla potrzeb symulacji i budowy aplikacji gridowych

Prof. dr hab. inż. Jacek KITOWSKI

Jest kierownikiem Zespołu Systemów Komputerowych w Katedrze Informatyki AGH. Pracuje również w ACK Cyfronet-AGH, w zakresie rozwoju systemów obliczeniowych. Jest autorem lub współautorem ok. 200 publikacji. Jego zainteresowania dotyczą architektur komputerowych i ich zastosowań, gridowych środowisk obliczeniowych, przechowywania i udostępniania danych oraz semantyki, ontologii i zarządzania wiedzą dla organizacji obliczeń gridowych. Członek PAU.



e-mail: kito@agh.edu.pl

Dr hab. inż. Witold DZWINEL

Z-ca kierownika Katedry Informatyki AGH. Autor ok. 150 publikacji w dziedzinach: symulacji rzeczywistości, fizyki komputerowej, obliczeń równoległych, obliczeń ewolucyjnych, metod rozpoznawania obrazów. Członek PAU.



e-mail: dzwinel@agh.edu.pl

Dr hab. inż. Krzysztof BORYCZKO

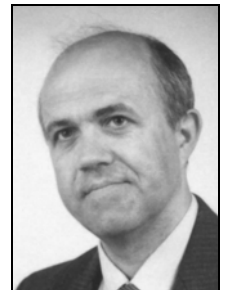
Jest absolwentem Wydziału Mechanicznego Politechniki Krakowskiej oraz Wydziału Elektrotechniki Automatyki Informatyki i Elektroniki AGH w Krakowie. Stopień doktora nauk technicznych z dyscypliny Informatyka uzyskał w roku 1995, a doktora habilitowanego w roku 2004. Zajmuje się obliczeniami równoległymi dużej skali, rozpoznawaniem obrazów oraz symulacjami zjawisk zachodzących w różnych skalach przestrzenno-czasowych.



e-mail: boryczko@agh.edu.pl

Dr Włodzimierz FUNIKA

Ukończył Uniwersytet w Kiszyniowie na kierunku "Matematyka teoretyczna". Od 1978 do 1993 roku pracował w Akademii Nauk Mołdawii. Od 1993 roku związany jest z AGH w Krakowie, gdzie uzyskał stopień doktora nauk technicznych w zakresie informatyki. Wśród zainteresowań naukowych są systemy rozproszone, monitorowanie, systemy agentowe, wizualizacja. Uczestnik projektów UE: CrossGrid, K-Wf Grid, CoreGRID, int.eu.grid, Virolab.



e-mail: funika@agh.edu.pl

Dr inż. Marian BUBAK

Ukończył studia i uzyskał stopień doktora z informatyki na Wydziale EAIiE AGH; adiunkt w Katedrze Informatyki i w ACK CYFRONET AGH oraz Professor of Distributed System Engineering w Universiteit van Amsterdam. Prowadzi badania w zakresie obliczeń równoległych i rozproszonych oraz systemów gridowych; współautor ok. 200 publikacji; uczestnik projektów IST EU: CrossGrid, K-WfGrid, CoreGRID, ViroLab i GREDIA; Associate Editor czasopisma Journal of Future Generation Computer Systems.



e-mail: bubak@agh.edu.pl

Dr inż. Renata SŁOTA

Jest obecnie adiunktem Katedry Informatyki AGH w Krakowie. Stopień doktora nauk technicznych w dyscyplinie Informatyka uzyskała na AGH w 1989 roku. W zakresie jej zainteresowań znajdują się systemy rozproszone, ze szczególnym uwzględnieniem systemów typu grid. Do tej pory opublikowała około 60 artykułów, a także uczestniczyła w wielu projektach naukowych zarówno krajowych (ostatnio: SGIGrid, Clusterix), jak i zagranicznych (Pellucid, Cross-Grid, K-Wf Grid, int.edu.grid, Gredia)



e-mail: rena@agh.edu.pl

Streszczenie

W artykule przedstawiono zasadnicze wyniki prac uzyskane w okresie ostatnich kilku lat w Grupie Systemów Komputerowych Katedry Informatyki AGH oraz kierunki dalszych badań. Poruszane zagadnienia dotyczą opracowania warstwy pośredniej infrastruktury rozproszonej i gridowej dla potrzeb obliczeń o dużym nakładzie z zakresu eScience i aplikacji biznesowych, rozwoju i efektywnej implementacji rozproszonych algorytmów z zakresu biotechnologii i zastosowań medycznych, jak również metod i implementacji ontologicznej reprezentacji wiedzy i algorytmów rozwoju bazy wiedzy dla potrzeb tworzenia organizacji wirtualnych w zakresie wspomaganie obliczeń i zwiększenia ich elastyczności w wymienionych obszarach problemowych.

Słowa kluczowe: środowiska i interaktywne obliczenia gridowe oraz ich monitorowanie, algorytmy klasteryzacji i rozpoznawania obrazów, symulacje w mezoskali, generacja siatek obliczeniowych, symulacja sedymentacji, ontologiczna reprezentacja wiedzy, organizacje wirtualne, optymalizacja dostępu do rozproszonych danych.

Development of Tools and Environments for Computer Simulations and Grid Applications

Abstract

Some main results of the research recently obtained by Computer Systems Group of the Department of Computer Science are presented in the paper. They are supported by descriptions of future plans and further development. The achievements include architecture of grid middleware for high performance computing and business applications, tools for grid computing, parallel algorithms for mesoscopic simulations and clustering, ontological representation and management of knowledge as well as virtual organizations.

Keywords: grid environments and tools, interactive grid computing, monitoring grid applications, cluster and pattern recognition algorithms, mesoscopic simulations, 3D mesh generation, simulation of sedimentation, ontological representation of knowledge, virtual organizations, data access optimizations.

1. Wprowadzenie

W ostatnich latach w prowadzeniu badań naukowych i działalności biznesowej pojawiają się nowe tendencje, które powodują konieczność wykorzystania zaawansowanych technologii informatycznych. Te nowe tendencje to zespołowość prac naukowych

* Autorzy artykułu składają podziękowanie pozostałym Współautorom, których nazwiska ze względów formalnych nie mogły się znaleźć na liście. Są to: Witold ALDA, Monika DEKSTER, Barbara GLUT, Darin NIKOLOW, Paweł TOPA, Katarzyna RYCERZ, Bartosz BALIŚ, Tomasz JURCZYK, Marta MAJEWSKA i Maciej MALAWSKI

i współpraca w zakresie realizacji celów komercyjnych. Wspólną cechą tych odrębnych nieco segmentów działalności jest wykorzystanie Internetu jako infrastruktury komunikacyjnej.

Niezwykle ważną cechą obecnie realizowanych prac zespołowych jest ogólnościowy ich charakter, rosnąca złożoność i interdyscyplinarność, wymagająca dostępu do zaawansowanych usług obliczeniowych, archiwizacyjnych i wizualizacyjnych, formułując nowy paradygmat badań, nazywany coraz częściej e-Science. Stale rośnie liczba zespołów naukowych, do istniejących już zespołów dołączają nowi partnerzy, zespoły naukowe intensywnie ze sobą współpracują, tworząc mniej lub bardziej formalne sieci instytucji, które wymieniają uzyskane wyniki, a to oznacza konieczność posiadania narzędzi informatycznych umożliwiających współpracę w skali globalnej, gromadzenie i wymianę uzyskanej wiedzy.

Obecnie wyniki eksperymentalne to bardzo często olbrzymie, rozproszone zbiory danych o różnorodnej strukturze, które z kolei wymagają środowisk umożliwiających integrację danych, selektywny dostęp do nich i ich przetwarzanie we współpracy z licznymi partnerami w sposób gwarantujący spójność uzyskiwanych wyników pośrednich i finalnych.

Symulacja komputerowa jest obecnie w pełni akceptowaną metodą badawczą, dzięki której jest możliwe przejście od eksperymentów „in vivo” czy też „in vitro” do eksperymentów „in silico”. Co więcej, coraz częściej łączone są ze sobą wyniki uzyskane obiema metodami. Jednym z kierunków rozwoju w obrębie e-Science jest tzw. *system level science* (nauka na poziomie systemowym), w której badania nie są zogniskowane jedynie na indywidualnych zjawiskach, ale mają na celu zrozumienie złożonych systemów wielo-zjawiskowych i wieloskalowych, obejmujących zjawiska należące do wielu dyscyplin.

Niezwykle ważnym etapem w interpretacji wyników badań jest zaawansowana wizualizacja, często realizowana w tzw. wirtualnej rzeczywistości. Takie nowatorskie podejście jest najbardziej widoczne w naukach biologicznych i medycznych, fizyce wysokich energii, w chemii, w astrofizyce, w naukach o Ziemi.

W zakresie zastosowań komercyjnych konieczność rozwoju środków i narzędzi informatycznych wynika z potrzeb globalnej współpracy między organizacjami, opartej na wiedzy, skutkującej w obniżeniu kosztów działalności przedsiębiorstwa, poprawie poziomu usług i podniesieniu elastyczności (*business networking*). Jego celem jest zapewnienie łatwej organizacji współpracy oraz rozwój środowisk dla wymiany doświadczeń i podejmowania inicjatyw gospodarczych przez jednostki, które posiadając swoje własne doświadczenie biznesowe i wiedzę na temat realizowanych przez siebie procesów, skłonne są do nawiązania szerszej współpracy z innymi podmiotami gospodarczymi dla poszerzenia rynków zbytu i globalizacji swojej własnej działalności. Wirtualizacja zasobów, wirtualne organizacje i opis semantyczny są metodologiami dla realizacji biznesowej przestrzeni współpracy. Warto podkreślić, że wymienione podejście może być wykorzystane także poza sferą ustabilizowanej działalności gospodarczej, w szczególności do zarządzania kryzysowego w trakcie powodzi, pożarów i innych klęsk żywiołowych.

Nakreślona tematyka naukowa jest w centrum zainteresowania Grupy Systemów Komputerowych Katedry Informatyki AGH od wielu lat [1]. W ostatnim okresie szczególnie nacisk położono na badania w zakresie metodologii wykorzystania infrastruktury rozproszonej i gridowej dla potrzeb obliczeń z zakresu e-Science i aplikacji biznesowych, rozwój algorytmów równoległych i rozproszonych, m.in., z zakresu biotechnologii i symulacji systemów medycznych oraz metod i implementacji ontologicznej reprezentacji wiedzy i algorytmów rozwoju bazy wiedzy dla potrzeb tworzenia organizacji wirtualnych w odniesieniu do realizacji obliczeń z zakresu e-Science i aplikacji biznesowych, a także grafiki i wizualizacji naukowej.

W poniższych rozdziałach przedstawiamy najważniejsze osiągnięcia w poszczególnych grupach problemowych. Większość z nich powstała w wyniku współpracy międzynarodowej, bilateralnej, w ramach grantów Unii Europejskiej lub grantów narodowych.

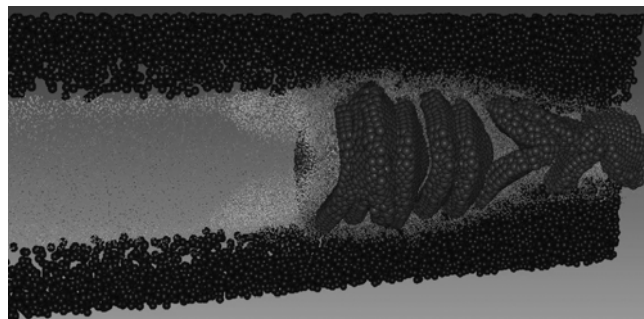
2. Problematyka obliczeniowa

2.1. Symulacje przepływu krwi w naczyniach włosowatych

Do początku lat 90-tych, symulacja zjawisk zachodzących w przestrzeniach o rozmiarach od kilku mikrometrów do dziesiątych części milimetra w czasie do jednej sekundy, były praktycznie niemożliwe ze względu na brak odpowiedniej metody symulacji. Przełom przyniosło opracowanie modelu dyssypatywnej dynamiki cząstek [2]. Metoda ta postrzega materię jako złożoną nie z atomów, czy molekuł a z obszarów o rozmiarach liniowych odpowiadających mezoskali. Podejście to umożliwiło symulację zjawisk zachodzących w zawiesinach koloidalnych, płynach biologicznych czy wprost w organizmach żywych. Stanowi dobry przykład badań w zakresie e-Science, przy uwzględnieniu własności algorytmów (takich, jak złożoność i efektywność), metod obliczeń równoległych i metodologii odwzorowania algorytmów na wieloprocesorową architekturę komputerową.

Układ naczyń włosowatych stanowi integralną część układu krwionośnego. Szacuje się, że u dorosłego człowieka powierzchnia ścian naczyń włosowatych w samych tylko mięśniach wynosi ok. 600 m², zaś ich liczba w przekroju o powierzchni 1 mm² ok. 200. Przez tak skomplikowany układ naczyń przepływa krew, zbudowana w ok. 50% z ciał upostaciowanych (głównie erythrocyty oraz trombocyty), której wartość współczynnika lepkości kinematycznej jest 2.5-3.5 razy większa niż wody. Badania dowodzą, iż tajemnica sprawności układu naczyń włosowatych tkwi we własnościach fizycznych komórek endothelium budujących ściany naczyń oraz budowie i kształcie upostaciowanych składników krwi.

Złożoność zjawisk zachodzących podczas przepływu krwi w naczyniach włosowatych oraz ich skala przestrzenno-czasowa stanowiły inspirację do zbudowania unikatowego modelu wykorzystującego model dyssypatywnej dynamiki cząstek. Model składa się z trzech podstawowych składników, a mianowicie ścianek naczyń, osocza oraz zawieszonych w nim erythrocytów. Współczynniki oddziaływania cząstek modelu zostały tak dobrane, aby wartości parametrów fizycznych modelu, jak np. sztywność erythrocytów, lepkość osocza, podatność ścianek były zgodne z danymi laboratoryjnymi. Na rys. 1 przedstawiono symulację przepływu erythrocytów o kształtach prawidłowych przez zwiężające się naczynie krwionośne. Wyniki uzyskano na maszynie SGI Altix (8 procesorów), z użyciem modelu wymiany komunikatów.



Rys. 1. Symulacja przepływu erythrocytów o kształtach prawidłowych w zwiężającym się naczyniu włosowatym.

Fig. 1. Simulation of the flow of properly shaped erythrocytes in a narrowing capillary vessel

Zakres zastosowań zbudowanego modelu jest bardzo szeroki. Pozwala on między innymi na badanie wpływu sztywności erythrocytów na jakość przepływu, a więc symulację działania leków o działaniu przeciwważalowym. Badanie mechanicznego oddziaływania erythrocytów na ścianki naczyń umożliwia śledzenie ich

patologicznego rozwoju. Jest to szczególnie istotne w pracach mających doprowadzić do zaprojektowania leków hamujących rozwój naczyń krwionośnych w tkankach nowotworowych. Weryfikacja modelu oraz liczne wyniki symulacji przepływu w naczyniach o różnych kształtach, różnych rodzajów erytrocytów zostały przedstawione w [3, 4]. Model został uznany w świecie i jest wykorzystywany [5]. Badania prowadzone są we współpracy z University of Minnesota.

2.2. Algorytmy klasteryzacji

Celem analizy skupień, realizowanej przez algorytmy klasteryzacji, jest podział zbioru danych na grupy (klastry) obiektów o podobnych, lub przeciwnych, cechach. Podejście to posiada szereg zastosowań w wielu zakresach problematyki e-Science, w szczególności dotyczących dziedzin nauki, takich jak: bioinformatyka, medycyna, statystyka, ekonomia, symulacje komputerowe i inne. Pomimo ciągłego rozwoju, metody klasteryzacji wciąż napotykać na szereg problemów i ograniczeń.

Literatura dotycząca metod rozpoznawania wzorców, prezentuje szereg algorytmów klasteryzacji, dzielonych na dwie główne klasy [6] tj.: metody hierarchiczne i metody podziałowe. Wśród metod hierarchicznych wyróżniamy klasę algorytmów opartych na idei łączenia małych klastrów w większe grupy (tj. algorytmy aglomeratywne) oraz klasę algorytmów budujących strukturę hierarchiczną poprzez kolejne podziały analizowanego zbioru danych.

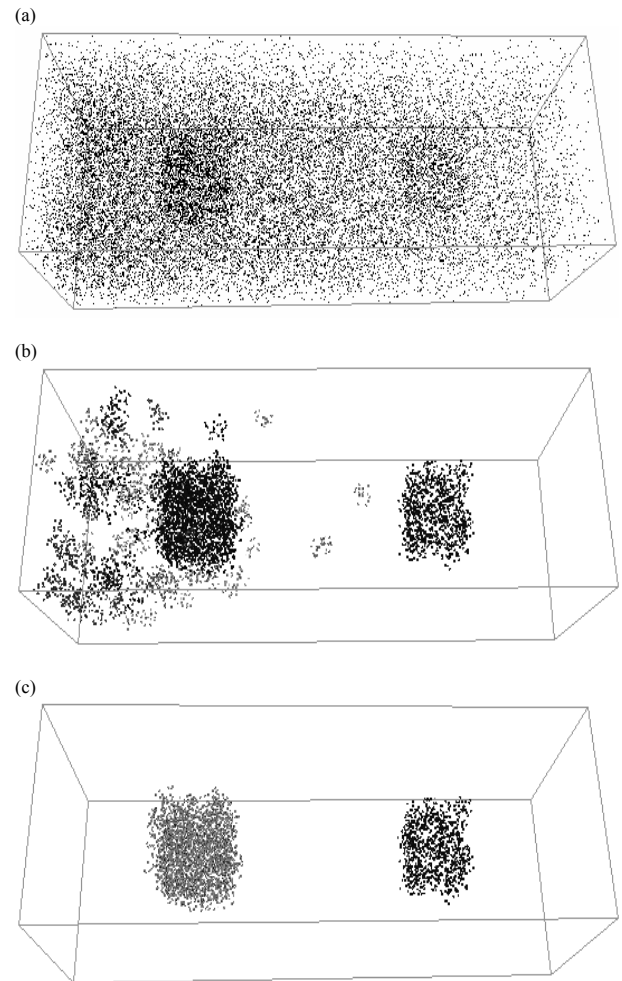
Klasycznymi metodami aglomeratywnymi są algorytmy pojedynczego, całkowitego oraz średniego połączenia (ang. *single link*, *complete link*, *average link algorithms*). Algorytmy te budują strukturę hierarchiczną poprzez iteracyjne łączenie klastrów o najmniejszej wzajemnej odległości, przy czym każdy z nich definiuje odległość klastrów w odmienny sposób.

Metody podziałowe budują niehierarchiczną strukturę klastrów. Główną klasę tego typu algorytmów stanowią metody bazujące na minimalizacji odpowiednio zdefiniowanej, kwadratowej funkcji błędów (ang. *squared error criterion function*). Wymienić tu należy w szczególności algorytm k-means, będący jedną z najczęściej stosowanych metod klasteryzacji. Główną wadą tego algorytmu jest fakt, iż nadaje się do detekcji klastrów jedynie o kształtach globularnych. Co więcej, w pewnych przypadkach algorytm k-means może utknąć w lokalnym minimum funkcji błędów, co prowadzi do błędnych rezultatów klasteryzacji. Innym, ważnym typem metod podziałowych są algorytmy bazujące na teorii grafów. Głównym przedstawicielem tej klasy algorytmów jest metoda oparta o ideę minimalnego drzewa rozpinającego (ang. *minimum spanning tree*). Udowodniono jednak, iż znajomość kształtów poszukiwanych klastrów może być konieczna dla jej użycia [7].

Jak wynika z powyższego przedstawienia podstawowych algorytmów klasteryzacji, zdecydowana większość z nich jest nieskuteczna w obecności szumu. Co więcej, w wielu przypadkach metody te nie radzą sobie z obserwacjami nietypowymi. Warto też wspomnieć, iż część z zaprezentowanych algorytmów wymaga od użytkownika wyspecyfikowania liczby poszukiwanych klastrów. Przy tym liczba ta najczęściej nie jest znana apriori. Z tych względów, w ostatnich latach podjęto wysiłki zmierzające do opracowania nowych algorytmów klasteryzacji, dobrze radzących sobie z szumem, automatycznie określających liczbę klastrów oraz zdolnych do wykrywania klastrów o nieregularnych kształtach. Badania te zaowocowały opracowaniem metod takich jak: Chameleon [8], DBSCAN [9] oraz CURE [10]. Niestety, w większości przypadków skuteczność tych algorytmów została zaprezentowana jedynie dla danych o niewielkiej liczbie wymiarów.

Wymienione powyżej algorytmy klasteryzacji, ze względu na swoje ograniczenia, są nieskuteczne w analizie danych biomedycznych, które są przedmiotem naszych zainteresowań. W związku z powyższym, istotne jest stworzenie nowych algorytmów, zdolnych do detekcji złożonych struktur klastrowych

w danych wielowymiarowych, w obecności szumu i przy zastosowaniu miar podobieństwa w zastępstwie poprawnie zdefiniowanych metryk odległości. W ramach tego zagadnienia, przeprowadzono badania nad metodami klasteryzacji wykorzystującymi relację najbliższego sąsiedztwa do oceny odległości obiektów oraz estymacji lokalnej gęstości danych. Algorytmy te bazują na grafie najbliższych sąsiadów (ang. *Shared Nearest Neighbours Graph*) [11] oraz miarze odległości wspólnego najbliższego sąsiedztwa (ang. *Mutual Nearest Neighbours Distance Measure*) [12]. Przykładowy wynik, otrzymany dzięki zastosowaniu omawianego podejścia, przedstawiono na rys 2. Szczegółowy opis algorytmu zaprezentowano w [13].



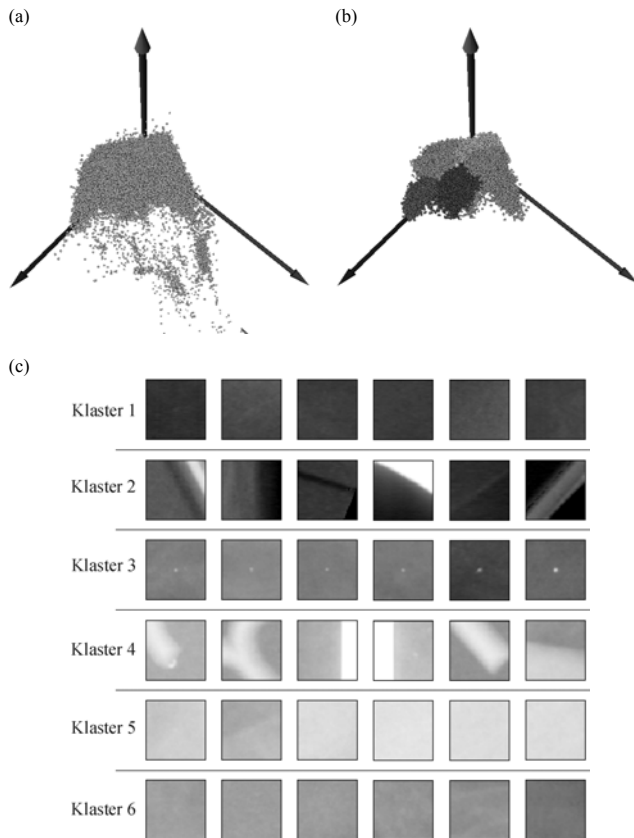
Rys. 2. Rezultat klasteryzacji trójwymiarowego zbioru danych o dużej zawartości szumu, z wykorzystaniem zaproponowanego algorytmu, łączącego ideę grafu wspólnych najbliższych sąsiadów z miarą odległości wzajemnego najbliższego sąsiedztwa. (a) Wejściowy zbiór danych, zawierający dwa skupiska punktów zanurzone w gradiencie szumu. (b) Rezultat klasteryzacji. (c) Dwa największe skupiska, zbudowane przez algorytm. Algorytm klasteryzacji usunął ze zbioru danych większość szumu i poprawnie zidentyfikował obydwa skupiska

Fig. 2. Results of cluster analysis of three-dimensional data set with large amount of noise using the proposed algorithm which combines the idea of a nearest-neighbor graph and a distance measure of mutual nearest neighbourhood. (a) The input data set containing two agglomerations of points immersed in a gradient of noise. (b) The result of clustering. (c) Two greatest agglomerations built by the algorithm. The algorithm removed most of the noise from data set and properly identified both agglomerations

Podstawowym obszarem zastosowań opracowywanych algorytmów jest analiza danych biomedycznych. W ramach tego zagadnienia, zajmowano się między innymi problemem detekcji zmian nowotworowych na cyfrowych zdjęciach mammograficznych [14-17]. W jednym z podejść, opracowany algorytm klasteryzacji został z powodzeniem wykorzystany do detekcji skupisk

w 27-wymiarowym zbiorze wektorów cech, reprezentujących podejrzane obszary zdjęć (rys. 3). Większość obszarów zawierających mikrozwapnienia, które są ważnym markerem zmian patologicznych, skupiła się w jednym z wyznaczonych klastrów.

Przedstawione wyniki uzyskano na maszynie SGI Altix (16 procesorów) z wykorzystaniem modelu pamięci współdzielonej, we współpracy z University of Minnesota.



Rys. 3. Rezultat klasteryzacji 27-wymiarowego zbioru wektorów cech, reprezentujących podejrzane obszary zdjęć mammograficznych. (a) Oryginalny zbiór danych zawierający około 93.000 wektorów. (b) Zbiór danych po klasteryzacji. Odkryte skupiska wektorów znaczone różnymi kolorami. Wektory uznane przez algorytm za szum zostały usunięte ze zbioru danych. (c) Przykładowe obszary zdjęć mammograficznych pogrupowane zgodnie z przynależnością odpowiadających im wektorów cech do klastrów. W klastrze nr 3 występują obszary zawierające mikrozwapnienia, będące częstym objawem zmian patologicznych. Wektory cech zobrazowano za pomocą metody PCA

Fig. 3. Results of cluster analysis of 27-dimensional set of feature vectors representing the regions of mammographic images. (a) Data set after analysis. The recovered agglomerations of vectors are marked with different colours. The vectors classified by the algorithm as the noise were removed from data set. (c) Sample regions of mammographic images grouped according to affiliations of corresponding feature vectors. There are regions of microcalcifications in the cluster #3 which often are symptoms of pathological changes. Feature vectors were represented using the PCA method

2.2.1. Ekstrakcja cech w algorytmach rozpoznawania obrazów

Ekstrakcja cech jest problemem obszernym, w którym stosowane metody zależą silnie od dziedziny badań i własności analizowanych sygnałów. Prowadzimy badania (we współpracy z University of Minnesota i CM UJ) nad deskryptorami dwójkięgo typu. Pierwszy rodzaj dotyczy obiektów rejestrowanych na zdjęciach, w szczególności twarzy oraz zmian nowotworowych w mammogramach. W ramach rozpoznawania twarzy, oryginalnym rezultatem badań jest zdefiniowanie nowej rodziny przekształceń inwariantnych względem translacji, rotacji i skalowania obiektu na zdjęciu [18] Przekształcenia te wykorzystują m.in.

transformatę Radona oraz transformatę Fouriera. Dzięki zastosowaniu zaproponowanych transformacji, możliwe jest stworzenie zestawu cech pozwalających na identyfikację obiektów, np. twarzy, bez konieczności estymacji i kompensacji zmian w ich ułożeniu na zdjęciu. W dziedzinie analizy mammogramów, badania skupione były na stworzeniu metody detekcji mikrozwapnień, tj. silnie zlokalizowanych znamion nowotworowych widocznych na zdjęciach rentgenowskich [19].

Najnowsze z badanych przez nas zagadnień dotyczących ekstrakcji cech wiąże się z analizą grafową. Jest to nowatorskie podejście do badań układów złożonych, dających się zamodelować w postaci grafu czy też dynamicznie zmieniającej się sieci. Na bazie modelu grafowego stosuje się szereg deskryptorów, wyupukających poszczególne aspekty struktury i dynamiki systemu. O ile modele grafowe były badane już w latach pięćdziesiątych ubiegłego wieku, o tyle w ostatnich latach dokonano istotnego przeorientowania tej dziedziny w kierunku grafów o niejednorodnym, silnie jednostronnym rozkładzie prawdopodobieństwa połączeń. Wiąże się to z odkryciem powszechnego występowania prawa potęgowego, zarówno w systemach będących wytworem człowieka (np. sieci połączeń lotniczych) jak i naturalnych (np. sieci metaboliczne). W ramach modelu sieci złożonych, globalna struktura obiektu reprezentowanego przez graf wyrażana jest poprzez analizę lokalnych cech strukturalnych w ramach grafu. Przykładem tego typu deskryptorów, mierzących złożoność grafu za pomocą informacji o sąsiedztwie każdego z wierzchołków, są m.in. średni stopień wierzchołków grafu, miara połączenia wierzchołka z innymi wierzchołkami czy współczynnik skupienia. Inne deskryptory wykorzystują informację mniej zlokalizowaną, dotyczącą np. długości dróg pomiędzy wierzchołkami. Do tego typu deskryptorów należą m.in. średnia odległość wierzchołka od pozostałych wierzchołków czy promień grafu.

Nasze badania w tym zakresie idą w kierunku zastosowania metodologii sieci złożonych jako nowatorskiego sposobu tworzenia cech na potrzeby dalszego przetwarzania metodami uczenia maszynowego. W szczególności, badane jest wykorzystanie deskryptorów grafowych do opisu sieci naczyniowych powstających w procesie angiogenezy związanej ze wzrostem guzów nowotworowych. Celem jest zbadanie, czy możliwe jest opracowanie zestawu deskryptorów silnie skorelowanych z obserwacjami medycznymi, takimi jak np. wzrost guza czy stopień reakcji na terapię. Ponadto, wartości deskryptorów obliczone dla sieci naczyniowych zamodelowanych przez graf automatów komórkowych mogą służyć jako metoda porównywania komputerowego modelu angiogenezy z rzeczywistością. Dalej, prowadzimy badania porównawcze pomiędzy opisanym wyżej podejściem opartym o sieci złożone a bardziej tradycyjnym podejściem bazującym na spektralnej teorii grafów, w celu ewentualnego wykorzystania, połączenia i rozszerzenia najlepszych cech obu metodologii. Wreszcie, we wstępnych badaniach nad zależnościami pomiędzy deskryptorami sieci złożonych, wykazano, iż prawo potęgowe jest możliwe do zaobserwowania również w oparciu o deskryptory bazujące na krawędziach grafu, nie tylko na wierzchołkach [20].

Drugim ogólnym kierunkiem badań są metody klasyfikacji, rozważane w oderwaniu od konkretnego typu cech opisujących wzorce. Szczególny nacisk położony został na metody wykorzystujące dynamiczną zmianę położony został na metody wykorzystujące dynamiczną zmianę funkcji celu w trakcie uczenia. W efekcie, podczas procesu treningu uwaga algorytmu skupiana jest na coraz to innych elementach dostępnych danych. Metody tego typu powiązane są z ideą uczenia zespołowego. W najczęściej spotykanym przypadku, tj. problemach klasyfikacyjnych, decyzja podejmowana jest kolektywnie przez grupę wyuczonych wcześniej instancji pewnego klasyfikatora. Aby budowanie zespołu dawało spodziewany rezultat w postaci większej skuteczności, konieczne jest zapewnienie różnorodności poszczególnych instancji danej reguły decyzyjnej. Dlatego też mechanizmy przenoszenia uwagi podczas uczenia kolejnych klasyfikatorów mają duże znaczenie dla tego podejścia.

Punktem wyjściowym naszych badań były obserwacje wskazujące na to, iż dla klasyfikatorów liniowych zastosowanie schema-

tów zespołowych nie przynosi zwiększenia skuteczności. Wynika to z niedostatecznej różnorodności pomiędzy składowymi tego typu zespołami. W celu przezwyciężenia tej wady, zaproponowaliśmy destabilizowanie liniowej metody decyzyjnej poprzez wykorzystanie zmiennych podzbiorów cech [21]. Przykładem tego typu podejścia jest algorytm, w którym, przy zachowaniu mechanizmów wymuszających zwracanie uwagi na przykłady trudne, każdemu klasyfikatorowi liniowemu prezentowany jest inny, losowo wybrany podzbiór cech opisujących przykłady [22]. Innym zaproponowanym rozwiązaniem jest dobór podzbiorów atrybutów dla każdego klasyfikatora bazowego w sposób deterministyczny, poprzez przypisanie poszczególnym cechom wag, dynamicznie zmienianych w trakcie uczenia. Tak uzyskane klasyfikatory wykazują dużą skuteczność nawet dla zespołów o niewielkiej liczbie składowych.

Wykorzystanie dyskryminatorów liniowych do tworzenia zespołu ma ponadto zasadniczą zaletę - możliwe jest określenie dla nich wielkości, których zdefiniowanie dla bardziej skomplikowanych klasyfikatorów jest problematyczne. Wykorzystując jedną z takich wielkości, tj. margines błędu klasyfikacji, zaproponowaliśmy nową miarę różnorodności zespołu klasyfikatorów, która lepiej obrazuje związek pomiędzy różnorodnością zespołu a jego skutecznością [23].

Istotnym zagadnieniem związanym z problemami klasyfikacji jest automatyczny wybór cech niosących najwięcej informacji o przykładach. W ramach schematu zespołowego, zaproponowaliśmy metodę, która łączy w jeden algorytm uczenie klasyfikatora i selekcję cech [24]. Równoległe do trenowania kolejnych klasyfikatorów bazowych rozszerzany jest, początkowo pusty, zbiór cech uznanych za istotne. Nowa cecha jest jednak wybierana tylko wtedy, gdy dla dotychczas wybranych cech dołączanie kolejnych klasyfikatorów bazowych przestaje zmniejszać błąd klasyfikacji.

Pod względem struktury, klasyfikator zespołowy składający się z liniowych reguł decyzyjnych jest tożsamy z jednokierunkową siecią neuronową o jednej warstwie ukrytej. Pozwala zatem na klasyfikację nieliniową, jednak istnieją pewne ograniczenia co do granic decyzyjnych, które jest on w stanie konstruować. Aby przełamać tę barierę, zaproponowaliśmy metodę rozwijającą schemat zespołowy w kierunku klasyfikatora o wielu warstwach. Klasyfikator taki potrafi rozwiązywać problemy, które pozostają poza zasięgiem możliwości prostego zespołu klasyfikatorów liniowych.

2.2.2. Generowanie i adaptacja siatek obliczeniowych

Generowanie siatki polega na przybliżonym pokryciu obszaru skończoną liczbą prostych, połączonych ze sobą elementów, tworzących w ten sposób siatkę. Siatki znajdują szerokie zastosowanie w komputerowej symulacji procesów fizycznych, ale także w innych dziedzinach takich jak: grafika komputerowa, geodezja, geometria obliczeniowa itp.. W numerycznej symulacji procesów z zastosowaniem np. metody elementów skończonych adaptacja siatki elementów jest istotnym czynnikiem wpływającym na jakość numerycznego rozwiązania. Zadaniem generatora jest w tym przypadku zapewnienie możliwości w pełni automatycznego konstruowania siatki o ściśle określonej strukturze i wysokiej jakości elementów. W ramach prac nad tym tematem rozwijany jest generator dla siatek dwuwymiarowych, powierzchniowych oraz objętościowych trójwymiarowych [25, 26].

Generacja siatek

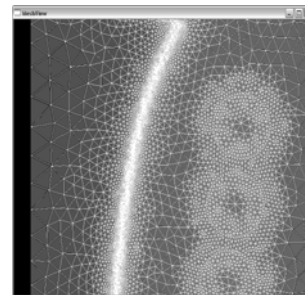
Opracowany generator siatek niestrukturalnych działa w oparciu o triangulację Delaunay'a. Tworzone siatki dwuwymiarowe oraz powierzchniowe mogą być trójkątne lub czworokątne, a siatki objętościowe – czworosienne oraz sześciosienne. Przykładowe siatki pokazano na rysunku 4.

Siatki czworokątne tworzone są na podstawie wcześniej wygenerowanych siatek trójkątnych [27, 28]. Konwersja ta przeprowa-

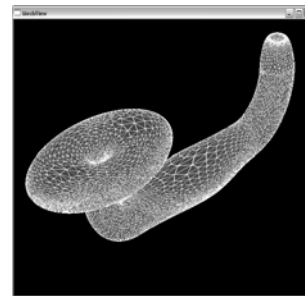
dzana jest w taki sposób, aby zachować strukturę siatki (rozmiary i kształty elementów) oraz uniknąć pozostawiania pojedynczych trójkątów. Zaimplementowano trzy metody: jedną opartą na podziale trójkątów, dwie – na łączeniu trójkątów.

W przypadku siatek sześciosiennych zajmujemy się dwoma metodami ich tworzenia – pośrednimi: siatka sześciosienna generowana jest na podstawie wcześniej utworzonej siatki czworosiennej, oraz bezpośrednimi: siatka konstruowana jest w oparciu o strukturę zwaną Medial Axis Transform [29, 30]. Dla tej ostatniej techniki opracowano osobny generator siatek strukturalnych.

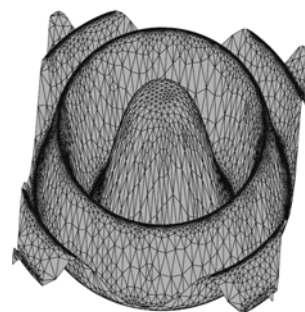
(a) fragment trójkątnej, dwuwymiarowej siatki o zróżnicowanym rozmiarze elementów



(c) siatka powierzchniowa 3D

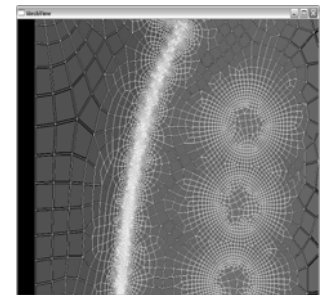


(e) dostosowana do krzywizny

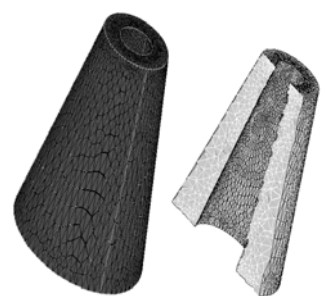


Rys. 4. Przykładowe siatki
Fig. 4. Sample meshes

(b) fragment czworokątnej siatki 2D wygenerowanej na podstawie siatki z rys. (a)



(d) siatka objętościowa 3D



Adaptacja siatek

Tematyka badawcza obejmuje również problem adaptacji siatek do rozpatrywanego modelu – tj. ich optymalizację z jednoczesnym zachowaniem możliwie najmniejszej liczby elementów. Dotyczy to zarówno problemów adaptacji do geometrii obszaru [31, 32] jak również do symulowanego procesu [32-35]. W trakcie generacji siatki stosowana jest definicja metryki w postaci macierzowej reprezentującej odpowiednią transformację współrzędnych [26].

W ramach prac nad tym zagadnieniem przetestowano różne sposoby tworzenia przestrzeni kontrolnej, metody jej wykorzystania w trakcie konstrukcji siatek oraz wpływ wyboru węzłów siatki

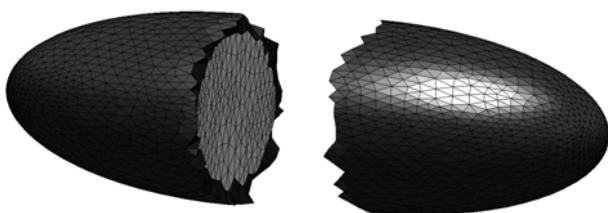
kontrolnej na jakość generowanych siatek oraz efektywność generatora [31, 36]. Opracowano i przeanalizowano też metodę automatycznego wyboru odpowiedniej metody interpolacji metryki [36].

Wymagania wobec kształtu i ewentualnego wydłużenia elementów mogą pochodzić z różnych źródeł. Tym samym każde z nich wprowadza inną metrykę. Celem pogodzenia tych wymagań wprowadzono metodę uzgadniania wspólnej metryki [31, 32, 35]. Opracowane metody pozwalają zatem również na wielokryterialną adaptację siatek [35]. Opracowany generator spełnia te wymagania i był wykorzystywany w symulacji zarówno procesów stacjonarnych jak i niestacjonarnych, dodatkowo także ze zmieniającym się w czasie obszarem obliczeń [33, 34].

Dekompozycja siatek oraz równoległa generacja siatek

Tematem związanym z generacją siatek jest dekompozycja siatek na potrzeby obliczeń rozproszonych. Opracowano metodę łączącą generowanie siatki z jej dekompozycją [37], rys. 5. Dzięki przeprowadzeniu dekompozycji na wczesnym etapie konstrukcji siatki, zarówno czas działania jak i jakość podziału mogą zostać zoptymalizowane. Temat ten jest realizowany we współpracy z Université de Technologie de Compiègne z Francji.

Opracowano strategię dekompozycji wyjściowej siatki powierzchni 3D dla późniejszego równoległego generowania siatki objętościowej [38]. Siatka powierzchni dekomponowana jest na podobzary poprzez rozcięcie jej separatorami (z możliwością rekurencji). Generowanie siatki objętościowej dokonywane może być następnie dla każdego obszaru zamkniętego na maszynie równoległej.



Rys. 5. Ilustracja dekompozycji geometrycznej dla równoległej generacji siatki objętościowej

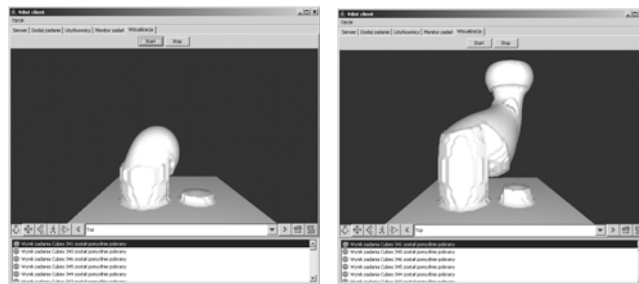
Fig. 5. The geometrical decomposition for parallel generation of a volumetric mesh

We wszystkich omówionych powyżej zagadnieniach szczególny nacisk jest położony na efektywność algorytmów oraz związany z tym odpowiedni dobór struktur danych. Opracowano także metodę przewidywania końcowej liczby elementów w tworzonej siatce, co umożliwi lepszy dobór rozmiaru struktur wykorzystywanych do ich przechowywania [39].

2.2.3. Środowisko do rozproszonej symulacji i wizualizacji wyników

W ramach współpracy z ośrodkiem TASK (Gdańsk) opracowano i zaimplementowano rozproszony system realizujący sekwencyjną i równoległą wizualizację danych symulacyjnych pochodzących z programów symulacyjnych wymagających dużego nakładu obliczeń, uruchamianych również poziomu systemu w trybie sekwencyjnym lub równoległym [40, 41]. Jako bazę dla systemu przyjęto istniejące, sprawdzone w innych zastosowaniach rozproszonych technologie, obejmujące serwery aplikacji J2EE, relacyjne bazy danych, klienta w języku Java, interfejs użytkownika z wykorzystaniem API Swing. Do przesyłania danych pomiędzy serwerem a klientem wykorzystano format X3D. Efektem pracy jest działający, rozszerzalny system oraz zbiór zaimplementowanych algorytmów służących do wizualizacji danych, pochodzących z symulacji różnych rodzajów.

W chwili obecnej system jest zintegrowany z programami obejmującymi symulację metodą dynamiki molekularnej, symulację granulatu i symulację metodą gazu siatkowego Boltzmann (LBG) uruchamianych i nadzorowanych z poziomu systemu. Spośród wielu technik wizualizacyjnych obecnie wykorzystywane jest przedstawianie izopowierzchni, mapy wysokości i molekuł (rys. 6).



Rys. 6. Przykładowe obrazy z części wizualizacyjnej. Wypychanie cieczy zwilżającej przez niezwilżającą w symulacji metodą gazu siatkowego Boltzmann

Fig. 6. Sample screenshots of the visualization part. Pushing up of moistening liquid by non-moistening one in the cellular Boltzmann's gas simulation

Oprócz wymienionych powyżej badań prowadzone są również zaawansowane prace nad infrastrukturą sprzętową i oprogramowaniem powstającego w Katedrze laboratorium 3-wymiarowej wizualizacji i stereoskopowej grafiki komputerowej (typu *Power Wall*). Laboratorium to oferuje unikalne w ramach Polski możliwości w tym zakresie.

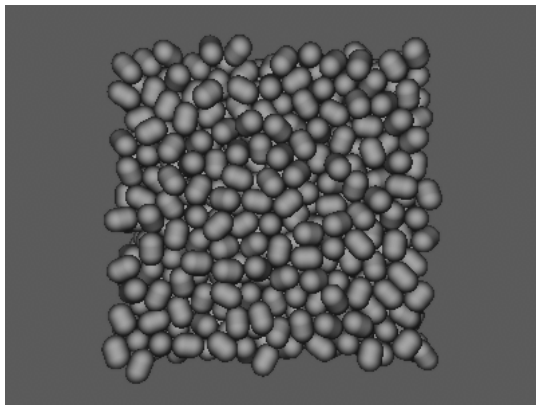
2.2.4. Numeryczne zagadnienia sedymentacji

Eksperymentalnie wykryto trzy zasadnicze cechy charakterystyczne dla zawiesin wieloskładnikowych: segregacja poszczególnych składników różniących się gęstością lub wielkością, propagacja gradientu koncentracji oraz inwersja gęstości spowodowana globalnymi ruchami zespołów cząstek. Istnieją zaawansowane metody numeryczne, służące do wykrywania nieciągłości koncentracji, które teoretycznie występują w zawiesinach. Stosowanym podejściem jest symulacja zachowania poszczególnych cząstek w celu uzyskania globalnych własności zawiesiny. Nowatorskie podejście polega to na założeniu, że prędkość cząstki zależy od koncentracji w obszarze bezpośrednio pod lub nad tą cząstką. Tego typu symulacja pozwala na odtworzenie cech rzeczywistych zawiesin, włączając propagację gradientu gęstości i wygładzenie oscylacji.

Proste modele zawiesin jednoskładnikowych, np. [42, 43], mogą w prosty sposób zostać rozszerzone do zawiesin wieloskładnikowych [44-46] i problemu fluidyzacji. Jak zaznaczono we wnioskach tych prac są one zgodne z eksperymentami opisanymi w literaturze. Ponieważ istotną własnością zawiesin jest ich stabilność, tak więc do opisu układów niestabilnych konieczne jest zastosowanie modelu trójwymiarowego. Komplikuje to zasadniczo algorytm obliczeniowy dostarczając w zamian wniosków dotyczących zarówno własności termodynamicznych jak również wariantów implementacji [47].

Problemem istotnym w zakresie wielu zastosowań jest zagadnienie gęstego losowego upakowania sztywnych kul. Tradycyjnie, problem ten postrzegano jako dobrze zdefiniowany stan o gęstości upakowania około 0.6366. Jednak późniejsze prace, np. [48] pokazały, że można otrzymać wyższe gęstości upakowania, jednak kosztem zwiększenia stopnia uporządkowania systemu. Ponieważ nie istnieje dokładna definicja uporządkowania bądź losowości systemu, precyzyjny podział na układy uporządkowane i losowe nie jest możliwy. W pracach [49, 50] opisano miarę lokalnego uporządkowania jako różnicę między każdym podukładem wybranej liczby kul a odpowiadającym jej fragmentem regularnej siatki.

Obecnie istnieje wiele algorytmów generujących upakowania sfer o różnych gęstościach i własnościach geometrycznych. Jednak w wielu zastosowaniach praktycznych mamy do czynienia z cząstkami o różnych rozmiarach i / lub kształtach. Niesferyczne cząstki muszą być traktowane inaczej niż kule ze względu na dodatkowe stopnie swobody. Zmodyfikowano więc istniejące algorytmy tak, by mogły generować upakowania kul o różnych promieniach lub cząstek niesferycznych [51, 52]. Jako modelu niesferycznych cząstek użyto sferocylindrów, które w znacznie lepszym stopniu przybliżają rzeczywiste układy spotykane na przykład w chemii czy mineralogii. Jednym z wniosków z przeprowadzonych symulacji jest fakt, iż sferocylindry o współczynniku wydłużenia około 0.4 mogą być upakowane znacznie gęściej niż kule (gęstość upakowania przekracza 0.71). Przy dalszym wydłużaniu sferocylindrów gęstość spada w przybliżeniu zgodnie z funkcją hiperboliczną, co potwierdzają również rozważania teoretyczne. Rysunek 7 przedstawia przykładowe upakowanie sferocylindrów o współczynniku wydłużenia 0.4. Prace prowadzone są wspólnie z Mount Allison University w Kanadzie.



Rys. 7. Przykładowe upakowanie sferocylindrów o współczynniku wydłużenia 0.4
Fig. 7. Sample packing of spherocylinders with aspect ratio 0.4

3. Problematyka Środowisk i Narzędzi

3.1. Badania naukowe w zakresie nowych technologii gridowych

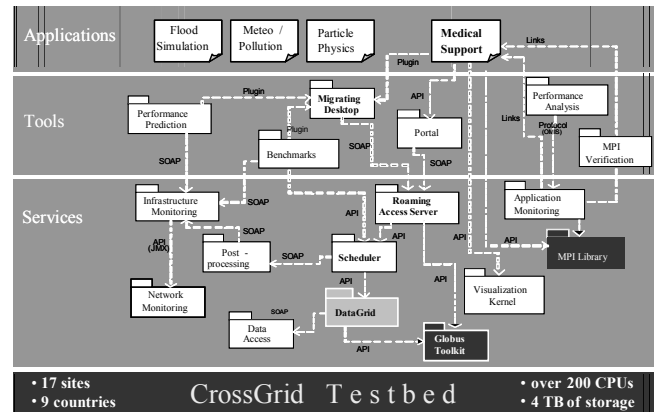
Środowisko gridowe dla aplikacji interaktywnych

Zainteresowanie rozwojem technologii gridowych rozpoczęło się wraz z podjęciem realizacji projektu EU IST CrossGrid, który został z sukcesem zakończony w lutym 2005 [53]. W pierwszym etapie prac badawczych koncentrowano się na rozwoju metod, narzędzi i serwisów gridowych zorientowanych na aplikacje, które wymagają wykonywania ich w trybie interaktywnym, mają dużą złożoność obliczeniową i konieczna jest ich realizacja w trybie high performance computing, wykorzystują rozproszone dane, dużych rozmiarów, mogą mieć postać tzw. *workflow* [54]. Naszym istotnym osiągnięciem było opracowanie architektury gridowej dla aplikacji interaktywnych (rys. 8), która w ramach projektu CrossGrid została zaimplementowana i zweryfikowana.

Analiza jakości aplikacji gridowych i monitorowanie rozproszonych aplikacji w Javie

Realizacja aplikacji interaktywnych wymaga monitorowania ich przebiegu w trybie on-line [61] oraz analizy jakości działania tych aplikacji. W pracach [56, 57] przedstawiono oryginalne podejście do konstruowania metryk pozwalających oceniać jakość działania tego typu aplikacji. Metryki te mogą mieć bardzo złożoną budowę

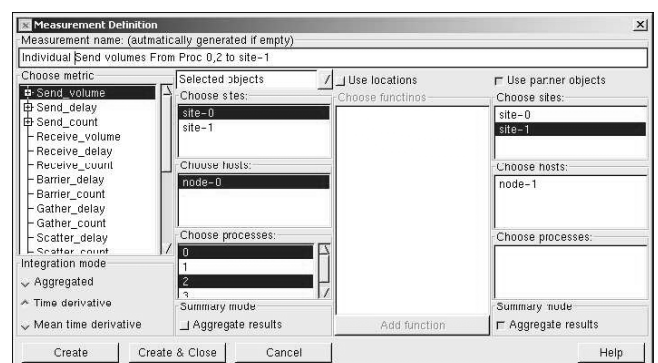
i mogą być definiowane przez użytkowników interaktywnych aplikacji gridowych dzięki specjalnemu językowi [57].



Rys. 8. Architektura środowiska gridowego projektu CrossGrid
Fig. 8. The grid architecture of the CrossGrid Project

Interaktywność aplikacji zakłada, że od specyfikacji aplikacji w dużym stopniu zależy określenie pomiarów, które będą wykonywane w trakcie działania programu. Jest to przesłanka ku temu, żeby użytkownika wyposażać w środki, umożliwiające dopasowanie analizy wydajności do konkretnych warunków, w których znajduje się użytkownik i aplikacja. Takimi środkami są mechanizm sond (ang. probes) oraz język Performance Measurement Specification Language (PMSL), umożliwiający na podstawie metryk standardowych, jak np. czas, ilość danych, ilość wywołań oraz danych monitorowania uzyskanych z sond, obliczenie takich metryk, które są semantycznie związane z monitorowaną aplikacją, np. „tempo zbiegania się rozwiązania równania różniczkowego”.

Narzędzie G-PM, realizujące pomiary na podstawie metryk standardowych, sond i metryk użytkownika, wyposażono w interfejs graficzny przeznaczony do sterowania pomiarami poprzez wybór metryk i ograniczeń pomiarów w czasie i przestrzeni (rys. 9) oraz definiowania sposobu wizualizacji, który najbardziej odpowiada specyfice wybranego pomiaru, np. danym zmieniającym się w czasie w większym stopniu odpowiada wykres w postaci krzywej, a zbiorowi skalarnych danych bardziej pasuje histogram.

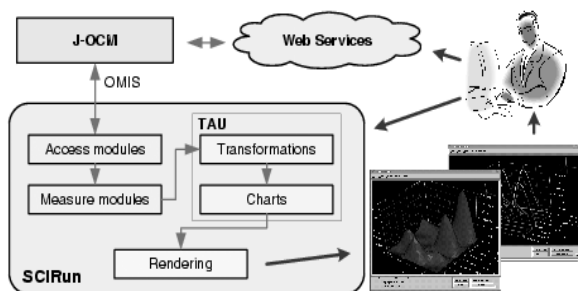


Rys. 9. Interfejs graficzny do definiowania pomiarów wydajności
Fig. 9. The graphical user interface used to define efficiency tests

Współpracę między narzędziem pomiarowym G-PM a systemem monitorującym OCM-G oparto o specyfikację interfejsu OMIS, rozszerzoną o monitorowanie aplikacji gridowych. Interfejs ten umożliwia oddzielenie warstwy narzędzi od warstwy monitorowania, pozwalając na podniesienie przenośności narzędzi, tak, że szczegóły platformy programistyczno-sprzętowej niskiego poziomu są ukryte przed warstwą narzędziową.

Programowanie rozproszone w języku Java wymaga silnego wsparcia narzędziowego, przede wszystkim monitorowania, ze

względu na problemy związane z wykorzystaniem tej platformy, głównie natury wydajnościowej. Prace nad monitorowaniem wydajności aplikacji w Javie oparte są o opracowaną modyfikację specyfikacji OMIS, przeznaczoną dla Javy – J-OMIS [58]. System monitorujący J-OCM [59] zrealizowany na podstawie J-OMIS umożliwia śledzenie zarówno obiektów aplikacji (np. klas, metod), jak i obiektów wykonawczych (wątki, Maszyna Wirtualna Javy (JVM)). Możliwości monitorowania aplikacji rozproszonych rozszerzono o monitorowanie usług sieciowych (ang. *Web Services*), charakteryzujących się problemami wydajnościowymi oraz wsparcie narzędziowe oparte o integrację ze środowiskiem SCIRun [60].



Rys. 10. Monitorowanie aplikacji rozproszonych opartych o usługi sieciowe
Fig. 10. Monitoring of distributed, Web service based applications

Monitorowanie aplikacji gridowych

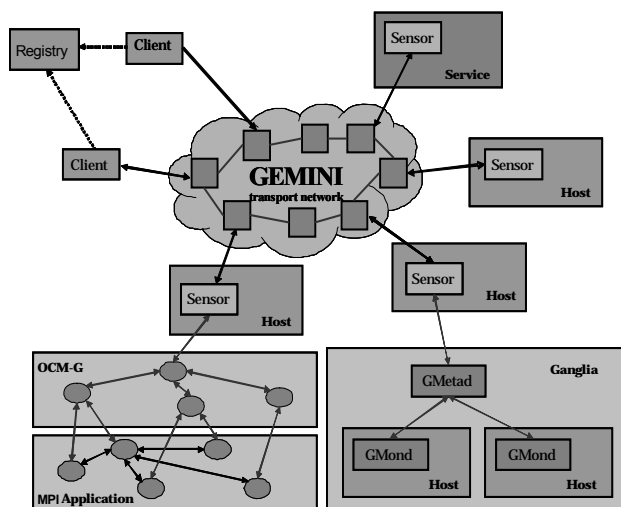
Uzyskiwanie na bieżąco informacji o sposobie realizacji rozproszonej aplikacji wykonywanej w środowisku gridowym wymaga odpowiedniego systemu monitorowania w trybie on-line. Prace [66] i [74] przedstawiają sposób monitorowania zarówno infrastruktury gridowej jak i aplikacji typu workflow.

Aplikacje typu workflow w środowisku gridowym charakteryzują się m.in. (1) rozproszeniem na skalę wielu niezależnych lokalizacji (*multi-site*) i decentralizacją, (2) luźnym powiązaniem w czasie i przestrzeni, (3) hierarchiczną, wielowarstwową budową, (4) heterogenicznością ze względu na język implementacji i środowisko uruchomieniowe. Każdy z tych aspektów stanowi wyzwanie z punktu widzenia monitorowania. Ważnym aspektem jest wsparcie dla kodu zastanego (*legacy code*), który często stanowi zasadniczą część, w której odbywają się obliczenia w workflow.

Aby rozwiązać te problemy, zaprojektowano i stworzono system monitorowania GEMINI oraz powiązane z nim technologie i narzędzia do instrumentacji. GEMINI jest uniwersalną infrastrukturą, która oddziela system zbierania danych monitorowania – *sensory*, od systemu subskrypcji i transportu tych danych (sieć monitorów GEMINI). GEMINI wykorzystuje standardową reprezentację danych w oparciu o XML, a także standardowy język zapytań o dane monitorowania, PDQS (*Performance Data Query and Subscribe*), również oparty o XML [74]. Te rozwiązania w połączeniu z ustandaryzowaną procedurą dynamicznej instalacji sensorów powodują, że GEMINI nadaje się do monitorowania wszelkiego typu zasobów – sprzętowych i programowych – takich jak węzły obliczeniowe, łącza sieciowe, usługi systemowe i aplikacje. Ogólna architektura systemu GEMINI przedstawiona jest na rys. 11. Na rysunku pokazano jak sensory rozproszone w Gridzie podłączone są do globalnej sieci monitorów GEMINI, która również jest rozproszona. Sensory mogą być różnego typu – zarówno autonomiczne jak i osadzone wewnątrz wykonującego się kodu; sensory mogą same wykonywać pomiary lub adoptować istniejące systemy monitorowania, takie jak Ganglia lub OCM-G, włączając je do globalnej infrastruktury.

W przypadku aplikacji workflow ważne jest to, iż rozproszone w przestrzeni i czasie zasoby programowe mogą stanowić logicznie jedną aplikację. System monitorowania musi umożliwić identyfikację tego logicznego powiązania, aby przedstawić dane moni-

torowania dotyczące jednego workflowu w sposób spójny i kompletny; użytkownik może zażądać np. „wszystkich danych dotyczących workflowu *wf-xyz* w okresie 2 godzin od chwili obecnej”. W tym kontekście ważne jest generowanie specjalnych *identyfikatorów korelacji* (*correlation identifier*) i przekazywanie ich do wszystkich części danego workflow, jak również wykrywanie zasobów, tzn. identyfikowanie tych monitorów, które dostarczają jakichkolwiek danych dotyczących określonego workflow. Do rejestrowania i wykrywania zasobów obecnie wykorzystywany jest scentralizowany rejestr, co jednak nie może być rozwiązaniem docelowym. Trwają prace nad zintegrowaniem systemu GEMINI z sieciami *peer-to-peer* typu DHT (*Distributed Hash-Table*), aby umożliwić efektywne i skalowalne wykrywanie zasobów w rozproszonym środowisku zdecentralizowanym [75].



Rys. 11. System monitorowania GEMINI
Fig. 11. The GEMINI monitoring system

Niezwykle istotnym aspektem monitorowania aplikacji jest instrumentacja kodu (statyczna bądź dynamiczna). W GEMINI zastosowana instrumentacja *dynamicznie aktywowana*, aby uniknąć generowania dużej ilości danych monitorowania (pełnych śladów wykonania) i umożliwić użytkownikowi elastyczne definiowanie pomiarów w trakcie działania aplikacji. Aktualnie możliwa jest dynamiczna instrumentacja Javy na poziomie *byte-code*'u; dodatkowo, aby umożliwić monitorowania zastanych aplikacji MPI napisanych w języku C (planowane jest również wsparcie dla Fortranu) zaadoptowano system monitorowania OCM-G [73], który zapewnia instrumentację i monitorowanie aplikacji MPI. GEMINI udostępnia standardowy interfejs służący do instrumentacji, oparty o język WIRL (*Workflow Instrumentation Request Language*). Używając żądań WIRL, użytkownik określa, które miejsca aplikacji mają być zainstrumentowane i jakie pomiary mają być dla nich wykonane.

Aby umożliwić instrumentację selektywną i drobnoziarnistą (tj. na poziomie poszczególnych regionów kodu), GEMINI wspiera koncepcję *standardowej reprezentacji* kodu. Jest to wysokopoziomowa, niezależna od języka reprezentacja aplikacji jako zbioru jednostek. Reprezentacja ta jest prezentowana użytkownikowi i dzięki niej może on konstruować żądania instrumentacji na wielu poziomach abstrakcji – od workflow po poszczególne regiony kodu. GEMINI implementuje wsparcie dla istniejącej specyfikacji standardowej reprezentacji, zwanej SIRWF (*Standard Intermediate Representation for Workflows*), która jest rozszerzeniem wcześniejszej specyfikacji – SIR [74].

Instrumentacja i monitorowanie możliwe są na różnych poziomach abstrakcji – od zdarzeń dotyczących workflowu jako całości, pochodzących z usługi uruchamiającej workflow (*workflow enactment engine*), poprzez aktywności workflowu oparte o usługi sieciowe (*web services*) napisane w Javie, aż po zastane aplikacje

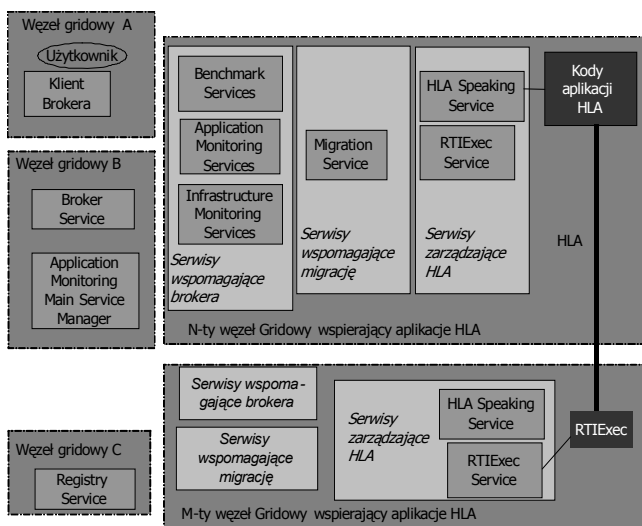
MPI oparte napisane w C. Dzięki ustandaryzowanym interfejsom, reprezentacji danych i koncepcji SIRWF, dostęp do wszystkich tych poziomów odbywa się w jednolity sposób.

Realizacja aplikacji HLA w środowisku gridowym

Bardzo wiele aplikacji z zakresu ochrony środowiska to złożone, rozbudowane symulacje, w tym symulacje typu *discrete event*. Dla konstruowania tego typu symulacji bardzo często stosowane jest podejście i system HLA (*High Level Architecture*). Nasze badania dotyczyły realizacji symulacji wykorzystujących efektywnie HLA w środowisku gridowym: opracowano sposób monitorowania takich aplikacji oraz przenoszenia ich między węzłami systemu gridowego w celu uzyskania lepszego zrównoważenia obciążenia; wyniki tych badań zostały przedstawione w pracach [62, 63].

Zaproponowano system zarządzania symulacjami HLA realizowanymi na gridzie – Grid HLA Management System (G-HLAM) [62]. Architektura systemu oparta jest na serwisach – *Service Oriented Architecture* (SOA). Przedstawione rozwiązanie jest możliwe do wykorzystania nie tylko dla tworzenia nowych aplikacji HLA, lecz pozwala ono również na łatwe dostosowanie aplikacji już istniejących do efektywnego działania na gridzie.

Architektura systemu G-HLAM, przedstawiona na rys. 12, zawiera serwis typu broker (Broker Service), który używa serwisu rejestru (*Registry Service*) w celu wyboru serwisów obsługujących pojedyncze elementy aplikacji zwanych wg terminologii HLA federatami (*federate*). Wybrane poprzez brokera serwisy obsługują zarówno proces koordynujący aplikacje HLA (o nazwie RTIExec) - do tego służy *RTIExec Service*, jak również poszczególnych federatów aplikacji - do tego służą *HLA-Speaking Services*. Te ostatnie stanowią interfejs pomiędzy faktycznymi procesami federatów a systemem G-HLAM. Ponadto G-HLAM zawiera serwisy do monitorowania aplikacji HLA (*Monitoring Services*), oparte na systemie monitorującym OCM-G, serwisy służące do testowania wydajności danej konfiguracji środowiska gridowego (*Benchmark Services*) oraz serwisy odpowiedzialne za bezpośrednie przeprowadzenie migracji (*Migration Services*).



Rys. 12. System zarządzania symulacjami HLA realizowanymi na gridzie
Fig. 12. The Grid HLA Management System (G-HLAM)

Stworzono mechanizm migracji umożliwiający federatom, których wydajność spada, zmianę miejsca wykonania. Zaproponowano mechanizm zapisu stanu aplikacji na poziomie użytkownika. W tym celu zaprojektowano bibliotekę o nazwie *GridHLA Controller library* (GHCL), która ukrywa szczegóły zapisu wewnętrznego stanu infrastruktury komunikacyjnej HLA (*Runtime Infrastructure* – RTI) i udostępnia przyjazny dla użytkownika interfejs do zapisu i odczytu danych.

W celu pokazania, że dzięki G-HLAM aplikacje HLA mogą efektywnie wykorzystać technologie gridowe, użyto dwóch przykładów ilustrujących wspomaganie aplikacji z różnymi typami rozproszenia oraz interaktywności przez system G-HLAM.

Jedną z aplikacji jest symulacja metodą wielu ciał [72]. Jest to przykład użycia systemu G-HLAM na gridzie dla aplikacji o różnych typach interaktywności, takich jak pasywne oglądanie wyników symulacji napływających w trakcie jej działania, aktywna eksploracja tych wyników, jak również cofanie stanu symulacji do poprzedniego kroku na żądanie użytkownika. Wyniki pokazują, w jaki sposób system G-HLAM może poprawić efektywność symulacji interaktywnej.

W pracy [63] pokazano zalety użycia systemu G-HLAM dla aplikacji dużej skali opartej o standard HLA – jest to aplikacja medyczna mająca na celu wspomaganie planowania operacji chirurgicznych układu krążenia. Opisano, w jaki sposób G-HLAM radzi sobie z aplikacją o różnych typach rozproszenia (jedna symulacja dla jednego użytkownika, wiele symulacji dla jednego użytkownika, środowisko dla wielu użytkowników). Skoncentrowano się w szczególności na badaniu środowiska dla wielu użytkowników. Przedstawione eksperymenty pokazują, że G-HLAM może być użyty do znaczącej poprawy wydajności również takiej aplikacji.

Workflowy, programowanie aplikacji gridowych

Obecnie aplikacje mają często postać tzw. workflowów, dotyczy to zwłaszcza tych sytuacji, gdy aplikacje są konstruowane poprzez odpowiedni przepływ danych i sterowania między serwisami (Web Services, Grid Services). Praca [64] opisuje opracowaną w ACK CYFRONET AGH metodę konstruowania tzw. *abstract workflows* w oparciu o informacje o charakterze semantycznym (zapisaną w postaci ontologii) o aplikacji, którą należy utworzyć, oraz o właściwościach składowych serwisów, wykorzystującą język opisu workflowów oparty o sieci Petriego. Istotną rolę w tworzeniu takich aplikacji odgrywają rejestry, w których przechowywane są informacje o serwisach; w pracy [65] przedstawiono sposób budowy rozproszonego, semantycznego rejestru tego typu. Na podstawie tych doświadczeń opracowywany jest nowy model programowania aplikacji gridowych [71].

Bardzo ważnym jest też przenoszenie tzw. oprogramowania zastanego (*legacy code*) do środowiska typu SOA oraz komponentowych; w pracy [70] przedstawiono koncepcję i prototyp systemu, który to umożliwia.

Podejście komponentowe

Przedmiotem naszego zainteresowania były również komponentowe środowiska gridowe [55], będące konkurencyjnym rozwiązaniem w stosunku do środowisk typu SOA. W pracy [67] opisano eksperymenty ze środowiskiem zbudowanym w oparciu o system H2O i JXTA, a w pracy [68] przedstawiono propozycję tzw. lekkiego środowiska komponentowego, zaś praca [69] opisuje koncepcję monitorowania tego typu systemu komputerowego.

Jako wkład w rozwój podejścia komponentowego powstała koncepcja oraz prototypowa implementacja rozproszonego środowiska do programowania opartego o model komponentowy CCA (*Common Component Architecture*). Środowisko to, nazwane MOCCA [78], jest dostosowane do systemów rozproszonych z współdzielonymi zasobami, takich jak środowiska gridowe oraz bardziej drobnziarniste systemy *peer-to-peer*, w których pojedynczy uczestnicy mogą współdzielić swoje zasoby. MOCCA wykorzystuje H2O, lekką platformę do współdzielenia zasobów.

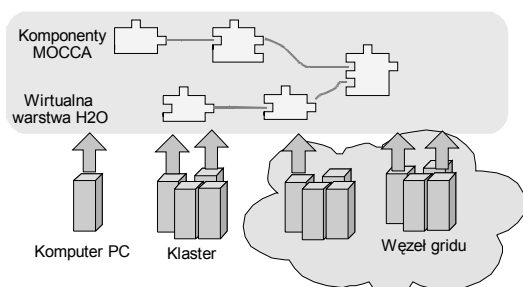
Główne założenia środowiska komponentowego MOCCA to:

- Dynamiczne tworzenie i uruchamianie komponentów na zdalnych współdzielonych zasobach
- Mechanizmy komunikacji dostosowane do systemów o różnym stopniu rozproszenia
- Współpraca z komponentami napisanymi w różnych językach programowania

- Współpraca ze standardami middleware gridowego oraz innymi modelami komponentowymi

Rys. 13 pokazuje schemat aplikacji komponentowej uruchamianej przy użyciu środowiska MOCCA. Pierwszym etapem jest udostępnienie zasobów poprzez uruchomienie platformy H2O. Wirtualna warstwa H2O stanowi pulę kontenerów, w których mogą być uruchamiane komponenty wchodzące w skład aplikacji. Porty komponentów są następnie łączone i aplikacja jest gotowa do działania.

Interesującą aplikacją było zastosowanie podejścia komponentowego do aplikacji obliczeniowych z dziedziny fizyki w środowisku gridowym oraz peer-to-peer. Przeprowadzonym obliczeniem było modelowanie klastrow złota metodą *simulated annealing*. W metodzie tej uzyskanie optymalnego wyniku wymaga przeprowadzenia dużej liczby symulacji, dlatego użycie środowiska rozproszonego ma wpływ na przyspieszenie tego typu obliczeń oraz poprawia jakość uzyskiwanych wyników, natomiast wybór różnych metod symulacji jako komponentów ułatwia budowanie aplikacji i badanie różnych jej wariantów.



Rys. 13. Środowisko MOCCA umożliwia budowanie aplikacji na heterogenicznych zasobach: pojedynczych komputerach PC, klastrach oraz infrastrukturach gridowych

Fig. 13. The MOCCA environment allows for building of applications on heterogenous resources: single personal computers, clusters and grid infrastructures

MOCCA stanowi wartościowy wkład do rozwiązania problemów związanych z programowaniem gridu, dzięki prostemu i eleganckiemu modelowi CCA oraz wykorzystaniu dynamicznych możliwości tworzenia aplikacji rozproszonych przy użyciu H2O. MOCCA oferuje mechanizmy uruchamiania komponentów na zdalnych współdzielonych zasobach oraz dynamicznego łączenia ich w aplikacje w sposób przyjazny dla użytkownika.

Wirtualne laboratorium – infrastruktura dla e-Science

Dla realizacji tego nowego paradygmatu uprawiania badań naukowych, zwanego eScience, konieczne jest posiadanie środowiska informatycznego – wirtualnego laboratorium, które powinno obejmować:

- infrastrukturę komputerową o stale rozbudowywanej mocy obliczeniowej i systemach przechowywania danych umożliwiających realizację zaawansowanych symulacji komputerowych,
- laboratorium wirtualnej rzeczywistości,
- infrastrukturę gridową, obejmującą oprogramowanie umożliwiające współdzielenie zasobów komputerowych oraz współpracę z innymi systemami gridowymi (międzynarodowymi – jak EGEE, DEISA oraz narodowymi – jak D-GRID, Austrian Grid),
- Problem Solving Environment – środowisko dla uruchamiania złożonych aplikacji, umożliwiające interaktywną współpracę między członkami różnych zespołów badawczych.

Warto podkreślić, że taka infrastruktura informatyczna może być wykorzystywana również poza obszarem badań naukowych; w szczególności do zarządzania kryzysowego w trakcie powodzi, pożarów i innych klęsk żywiołowych.

Od marca 2006 uczestniczymy w pracach badawczych projektu ViroLab [77], którego podstawowym zadaniem jest opracowanie

i zaimplementowanie wirtualnego laboratorium wspierającego zarówno prace naukowe z dziedziny wirusologii, immunologii i epidemiologii oraz praktyczne leczenie chorób wirusowych. Laboratorium to będzie miało zasięg europejski. Już obecnie szczególne znaczenie w leczeniu chorób wirusowych, takich jak np. HIV, ma wykorzystanie informacji o charakterze genetycznym; informacja tego typu staje się coraz powszechniej dostępna. Równocześnie coraz obszerniejsze i dokładniejsze stają się bazy danych zawierające informacje o poszczególnych pacjentach, o przebiegu ich leczenia, o aplikowanych środkach medycznych, o stosowanych procedurach leczenia i ich wynikach. Intensywnie rozwijane są metody symulacyjnego badania przyczyn i wyników leczenia chorób wirusowych oparte m.in. na metodach dynamiki molekularnej i automatów komórkowych; wymagają one zorganizowania dostępu do odpowiednich mocy obliczeniowych oraz narzędzi wizualizacyjnych. Istotną rolę w decydowaniu o stosowaniu konkretnych medykamentów w walce z chorobami wirusowymi odgrywa dostęp do wyników badań zamieszczane w bardzo licznych publikacjach naukowych. Dzięki zastosowaniu technologii gridowych ViroLab umożliwił integrację tych trzech wyżej wymienionych grup użytkowników. Cechą charakterystyczną wirtualnych laboratoriów jest ich ciągły rozwój i udoskonalanie, stąd też konieczne będzie prowadzenie badań w zakresie:

- nowych technologii systemów gridowych umożliwiających efektywną realizację obliczeń dużej skali,
- metod programowania aplikacji gridowych z wykorzystaniem opisu semantycznego i ontologii,
- wykorzystania informacji o poprzednich eksperymentach w wirtualnym laboratorium do ich kolejnych optymalnych realizacji,
- narzędzi umożliwiających współdziałanie partnerów.

3.2. Rozwój metod wspierających działalność biznesową przedsiębiorstw

Tematyka naukowa dotycząca zagadnień wspierania przedsiębiorstw w codziennej działalności realizowana była w ramach projektu finansowanego przez Unię Europejską, pn. Pellucid [79]. Podstawowym celem było opracowanie elastycznej platformy programowej dla wspomaganie działalności mobilnych pracowników średniego i wysokiego poziomu organizacji sektora publicznego. Przez mobilnych pracowników rozumie się tutaj tych spośród pracowników organizacji, którzy zmieniają stanowisko pracy przenosząc się do innych departamentów lub jednostek w ramach rozwoju swojej kariery zawodowej. Ich doświadczenie na dotychczasowym stanowisku jest zazwyczaj istotne i wartościowe, a także godne wykorzystania przez następcę. Z kolei, przenosząc się na nowe stanowisko pracy istotne jest skrócenie czasu adaptacji i wykorzystanie doświadczeń swojego poprzednika, jeśli tylko istniałaby metodyka ich gromadzenia i udostępniania.

Rozwój platformy przebiegał w dwóch aspektach [80-84]:

- Poprawa skuteczności i wydajności organizacji przez wprowadzenie formalnych działań prowadzących do wychwytywania, zapisu i ponownego udostępniania doświadczeń i wiedzy pracownika. Przy braku tego typu działań opuszczenie stanowiska pracy przez doświadczonego pracownika wiąże się ze stratami organizacji, wynikłymi z utraty jego doświadczeń. Realizowane to było przez indeksację dokumentów oraz możliwość wprowadzania notatek przez użytkownika.
- Wspomaganie pracowników w czasie podejmowania pracy na nowym stanowisku, poprzez udostępnienie doświadczeń poprzednika, zgromadzonych w czasie jego działalności. Redukuje się wtedy konieczność klasycznych szkoleń, frustrację oraz zaangażowania innych pracowników dla wyjaśniania wątpliwości nowoprzyjętemu. Etap ten polegał na pozyskiwaniu dokumentów z bazy wiedzy, które wykorzystywane były w poprzednich działaniach podobnego typu oraz wprowadzonych poprzednio notatek dotyczących zblizonych zadań.

Zyskiem dla organizacji jest możliwość ponownego wykorzystania wiedzy pracowników, jej współdzielenie, ewolucja „najle-

szej praktyki” oraz możliwość wielokrotnego wykorzystania. System tworzy możliwość wspomagania potrzeb i możliwości pracowników organizacji bez zasadniczych zmian istniejących praktyk i systemów.

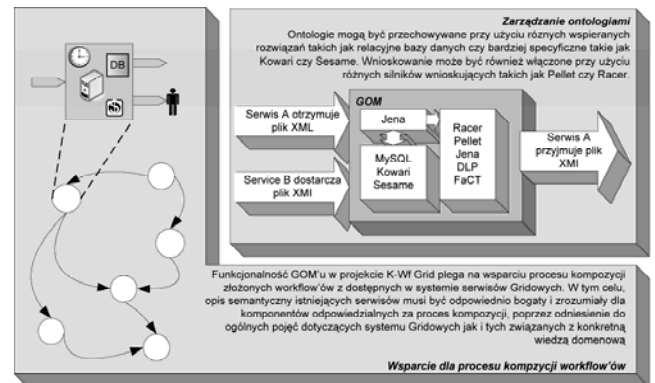
W ramach prac badawczych opracowano model organizacyjnej bazy wiedzy, wykorzystujący ontologiczną reprezentację wiedzy domenowo-specyficznej, możliwość jej rozwijania poprzez wprowadzenia oceny podpowiedzi ze strony użytkownika oraz zaprojektowano i zaimplementowano warstwę dostępu do repozytoriów dokumentów wraz z ich semantyczną indeksacją. Praktyczna weryfikacja w trzech przykładowych instytucjach potwierdziła przydatność zaproponowanych rozwiązań [85]. Badania prowadzone we współpracy z Rutherford Appleton Lab. (UK) i Instytutem Informatyki Słowackiej Akademii Nauk.

3.3. Ontologiczna reprezentacja wiedzy dla potrzeb zastosowań gridowych i wsparcia biznesu

W przypadku wzrostu skomplikowania zarówno zasobów gridowych, jak i ich zastosowań niezbędnym jest wspieranie użytkowników poprzez udostępnienie wiedzy w różnych typowych zadaniach takich jak rozwiązywanie problemów w danej dziedzinie czy wyszukiwanie danych. Może to być osiągnięte poprzez Wirtualne Organizacje. W takim przypadku znaczącą rolę zaczynają odgrywać metadane, które umożliwiają indeksowanie rozległych zasobów, jak i zarządzanie wymaganiami użytkownika. Do niedawna, większość systemów umożliwiających taką funkcjonalność opierała się o standardowe technologie takie jak np. relacyjne bazy danych. Jednak rozwój i osiągnięcia w dziedzinie technologii semantycznych w ostatnich latach stworzyły nową możliwość w postaci definiowania zarówno metadanych jak i wiedzy zgromadzonej w danej Wirtualnej Organizacji w sposób zunifikowany.

Unifikacja metadanych dotyczących Gridu jak i konkretnej domeny [86], w sensie jednorodnego formalizmu umożliwiającego wykorzystanie opisu na różnych poziomach w celu anotowania danych i opisu problemów, umożliwia tworzenie nowej jakości systemów automatyzujących większość aktywności użytkownika Wirtualnej Organizacji. Wirtualne Organizacje mogą tutaj dotyczyć zarówno aplikacji naukowych (e-Science) jak również aplikacji biznesowych. Aktywności użytkownika w systemie Gridowym w ramach pewnej Wirtualnej Organizacji często są definiowane w postaci workflow, czyli specyfikacji kolejnych wywołań serwisów gridowych oraz wymienianych między nimi danych. W tym zakresie, krok dalej idzie projekt K-Wf Grid [54, 87], który dzięki zunifikowanemu, semantycznemu opisowi wszystkich aspektów środowiska Gridowego oraz domeny danej Wirtualnej Organizacji, umożliwia tworzenie takich struktur typu *workflow* w sposób automatyczny, na podstawie jedynie wstępnej specyfikacji problemu przez użytkownika. Dzięki pracom nad analizą tekstu naturalnego w kontekście dopasowania go do istniejących ontologii, możliwe jest tworzenie takich opisów w języku naturalnym. W celu stworzenia skalownego modelu metadanych w projekcie K-Wf Grid, stworzono odpowiedni model polegający na podziale przestrzeni pojęć na pod-ontologie. Jednym z typów ontologii są tutaj ontologie ogólne, opisujące podstawowe pojęcia, wspólne dla wszystkich Wirtualnych Organizacji, umożliwiające stworzenie zbioru serwisów Gridowych pracujących na odpowiednio abstrakcyjnym poziomie. Każda Wirtualna Organizacja definiuje swoje ontologie poprzez rozszerzenie i specyfikowanie ontologii ogólnych [86]. Poza zdefiniowaniem samych ontologii, konieczne jest również istnienie odpowiedniej bazy wiedzy, zarządzającej i umożliwiającej innym serwisom oraz użytkownikom dostęp do wiedzy w nich zgromadzonej. Taką właśnie rolę spełnia rozproszona, Gridowa baza wiedzy zwana Grid Organizational Memory (GOM) (rys. 14). Umożliwia ona przechowywanie, wnioskowanie i wydobywanie informacji, zdefiniowanej w postaci ontologii [88]. Główną cechą tej bazy wiedzy jest jej modularna struktura umożliwiająca łatwe poszerzanie funkcjonalności na poziomie nawet pojedynczych pod-ontologii. Funkcjonalność ta

może wynikać np. z wymagań konkretnej aplikacji, czyli dostarczać specyficznych języków zapytań czy akceptowania przycho-dzących informacji w istniejących już standardach w danym środowisku. Domyślnie GOM wspiera OWL jako język definiowania ontologii, RDQL oraz SPARQL jako języki zapytań. W celu umożliwienia modyfikowania zawartości przechowywanych ontologii na poziomie pojedynczych klas czy instancji ontologicznych, stworzona została specjalna ontologia umożliwiająca definiowanie dowolnych, możliwych w ramach języka OWL, zmian jednostkowych. Zaimplementowana została również biblioteka, która umożliwia akceptowanie takich opisów. Na tej podstawie stworzony został mechanizm ewolucji ontologii w ramach danej Wirtualnej Organizacji, umożliwiający rozszerzanie i aktualizowanie wiedzy gromadzonej w postaci ontologii w bazie wiedzy. Aktualizacja wiedzy jest przeprowadzana zarówno poprzez inne komponenty systemów jak i użytkowników Wirtualnej Organizacji [89]. W celu umożliwienia integracji z innymi systemami GOM udostępnia odpowiednie interfejsy takie jak Web Service czy Gridowy WSRF. Dla użytkowników została natomiast stworzona specjalna wtyczka do środowiska Protege umożliwiająca bezpośrednią pracę w zaawansowanym środowisku wizualnym nad ontologiami zarządzanymi przez GOM.



Rys. 14. Schemat współpracy GOM ze środowiskiem K-WfGrid

Fig. 14. The schema of collaboration between GOM and the K-WfGrid environment

Ze względu na skomplikowany proces aktualizacji wiedzy, niezbędne są narzędzia wspomagające ten proces, nie tylko w postaci graficznego przedstawiania zawartości ontologicznej bazy wiedzy, ale wręcz automatycznie rozszerzające dotychczasową wiedzę. Jednym z takich narzędzi opracowanych w Katedrze Informatyki, jest narzędzie wspierające proces rozszerzania ontologicznej wiedzy domenowej podczas procesu rejestracji nowych serwisów gridowych opisanych za pomocą OWL-S. Przyjmuje się, że takie serwisy mogą być opisane semantycznie za pomocą zewnętrznych w stosunku do GOM, już istniejących, ontologii. Rejestracja takiego serwisu wymagałaby dostosowania ontologii GOM do pojęć stosowanych w ontologii zewnętrznej. Proces ten nie jest prosty i wymaga współpracy inżynierów i administratorów wiedzy. Poprzez wykorzystanie metod uzgadniania ontologii wspieranych przez podobieństwo ontologii, otrzymano narzędzie automatyzujące taki proces [90]. W trakcie rejestracji nowego serwisu, pojęcia z zewnętrznej ontologii automatycznie mapowane są na pojęcia z GOM i/lub dołączane są nowe, niewystępujące w GOM'ie, pojęcia. Dzięki temu narzędziu wiedza domenowa może być rozszerzana, a co za tym idzie możliwe jest zwiększenie różnorodności i efektywności proponowanych *workflow*.

Dzięki uniwersalnej architekturze i elastycznej implementacji GOM, jako baza wiedzy, może być oczywiście stosowana w innych zastosowaniach. Rozważane jest tutaj m.in. wsparcie automatycznego procesu udostępniania aplikacji obliczeniowych dla Grid'u [91] czy jako podstawa platformy integrującej różne heterogeniczne środowiska Gridowe w zastosowaniach biznesowych poprzez mapowanie odpowiednich pojęć i tłumaczenie żądań pomiędzy ich odpowiednimi komponentami [92].

Dalsze prace nad środowiskiem GOM idą w kierunku automatyzacji procesów związanych z tworzeniem i zarządzaniem Wirtualnymi Organizacjami na poziomie semantycznym, poprzez stworzenie zarówno odpowiednich modeli kolaboracji jak i ich ontologicznych specyfikacji. Współpraca prowadzona jest z wieloma partnerami europejskimi, w szczególności z Fraunhofer Institute FIRST, Instytutem Informatyki Słowackiej Akademii Nauk i Dresden University.

3.4. Optymalizacja dostępu do danych w środowiskach rozproszonych typu Grid

Aplikacje gridowe [93, 94] często produkują lub przetwarzają wielkie ilości danych, które ze względów ekonomicznych są na ogół przechowywane w hierarchicznych systemach składowania danych typu HSM za pomocą urządzeń o dostępie pośrednim (ang. *near-line storage*). Pozwalają one znacznie obniżyć koszt przechowywania danych szczególnie w przypadku, gdy ilość danych jest olbrzymia. Ze względu na duży rozmiar danych oraz różnorodności systemów przechowywania danych i nośników fizycznych pojawia się problem optymalizacji dostępu do danych. Prace prowadzone w Katedrze Informatyki odnośnie optymalizacji dostępu do danych skupiają się wokół następujących zagadnień: przyspieszanie dostępu do dużych plików na taśmach poprzez fragmentację, estymacja czasu dostępu do danych i replikacja danych.

Przyspieszanie dostępu do dużych plików na taśmach

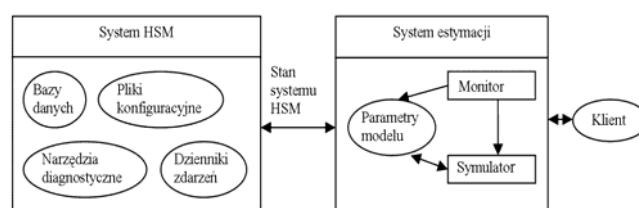
W systemach HSM dane często są zapisywane na nośnikach taśmowych ze względu na mniejszy koszt przechowywania. Gdy plik znajdujący się na taśmie jest odczytywany w większości systemach musi być najpierw skopiowany do dyskowej pamięci podręcznej – w przypadku dużych plików operacja ta może być czasochłonna [95]. Na dodatek bloki, które już się znajdują w dyskowej pamięci podręcznej nie są dostępne dopóki cały plik nie znajdzie się na dysku. W opracowanym podsystemie [96] duże pliki przy zapisie są transparentnie fragmentowane poprzez oprogramowanie warstwy pośredniej. Taki plik widziany jest przez system HSM jako zbiór mniejszych plików, natomiast przez użytkownika - jako jeden duży plik. W przypadku sekwencyjnego odczytu takiego pliku dane są dostępne jak tylko odpowiedni fragment znajdzie się na dysku. Technika ta pozwala zmniejszyć czas odczytu całego pliku o czas potrzebny na jego odczyt z dysku. Natomiast w przypadku dostępu do początkowych bloków pliku (np. w przypadkach odczytu nagłówku pliku) czas ten skrącający jest do minimum, czyli do czasu potrzebnego na ściągnięcie jednego fragmentu. Umożliwia to również udostępnianie usługi dostarczania dowolnego fragmentu pliku.

Estymacja czasu dostępu do danych przechowywanych w systemach HSM

Systemy HSM składają się ze sprzętu do przechowywania danych (najczęściej o dużej pojemności), serwera (lub serwerów) obsługujący te urządzenia oraz specyficznego oprogramowania do zarządzania systemem składowania danych HSM. System HSM organizuje dostępne urządzenia w hierarchie i umożliwia automatyczne przeniesienie (migracje) danych pomiędzy różnymi poziomami hierarchii, tak, aby zapewnić minimalny średni czas dostępu do danych oraz najlepszy stosunek pojemności do kosztu przechowywania danych. Ze względu na dużą złożoność systemów HSM oraz dużą różnorodność urządzeń i nośników przechowywania danych, estymacja czasu dostępu do danych w HSM nie jest łatwym zadaniem. Czas dostępu do danych zależy od wielu czynników takich jak: charakterystyka wydajnościowa urządzenia i/lub nośnika, lokalizacja danych w obrębie nośnika, chwilowe obciążenie systemu komputerowego, na którym jest zainstalowane oprogramowanie systemu HSM, wydajność podsystemu wejścia/wyjścia, rozmiar danych, przepustowość magistral znajdujących się na ścieżce przepływu danych, długość aktual-

nej kolejki żądań oraz ilość aktualnie obsługiwanych żądań, w niektórych przypadkach nawet przyszłe żądania mogą powodować zmiany w czasie dostępu dla żądań już znajdujących się w kolejce.

W systemie estymacji czasu dostępu do danych lokowanych w systemach HSM, opracowanego w Katedrze Informatyki AGH, wykorzystano symulacje działania takiego systemu celem otrzymania czasu dostępu [97]. W prezentowanym systemie wykorzystano podejścia szarej skrzynki. Informacje na temat wewnętrznego stanu systemu otrzymywane są w sposób pośredni poprzez analizę konfiguracji i dostępnych dzienników zdarzeń, uruchamianie narzędzi diagnostycznych dostarczonych przez producenta oprogramowania HSM (rys. 15). Opracowaną metodologię wykorzystano do budowy estymatorów czasu dostępu do takich systemów jak DiskXtender [97] i Castor [98], które to następnie wykorzystano w projekcie EU CrossGrid [53].



Rys 15. Estymacja czasu dostępu
Fig. 15. Access time estimation

Replikacja danych

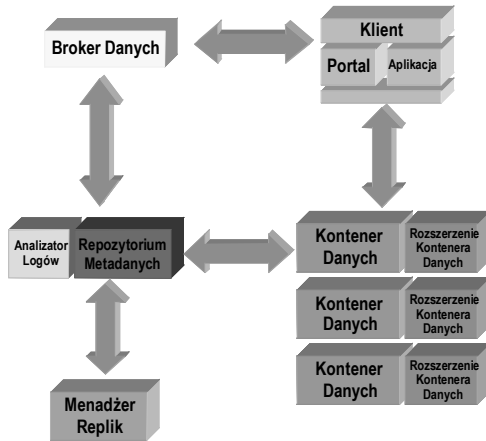
Replikacja danych jest wykorzystywana celem zapewnienia większego bezpieczeństwa i/lub efektywniejszego dostępu do danych. Większa dostępność jest uzyskiwana poprzez tworzenia pasywnych kopii danych, które używane są wtedy jeśli dane z głównej lokalizacji nie są dostępne z jakiegoś powodu jak np., brak dostępu do sieci, awaria systemu, prace konserwacyjne. Szybszy dostęp uzyskiwany jest poprzez wybór najlepszej pod względem przewidywanego czasu dostępu repliki danych. Szybszy czas dostępu można również uzyskać inicjując większą liczbę jednoczesnych transferów różnych fragmentów danych z różnych lokalizacji.

Replikacja może być statyczna lub dynamiczna (automatyczna). W przypadku dynamicznej replikacji system sam decyduje, kiedy i gdzie stworzyć nową replikę lub, którą replikę należy już usunąć. Celem dynamicznej replikacji jest uzyskanie minimalnego ogólnego czasu dostępu do danych dla zmieniających się warunków obciążenia i rozkładu żądań. Aby móc poprawnie funkcjonować system replikacji danych musi posiadać informacje o obciążeniu sieci i o estymowanym czasie dostępu do danych.

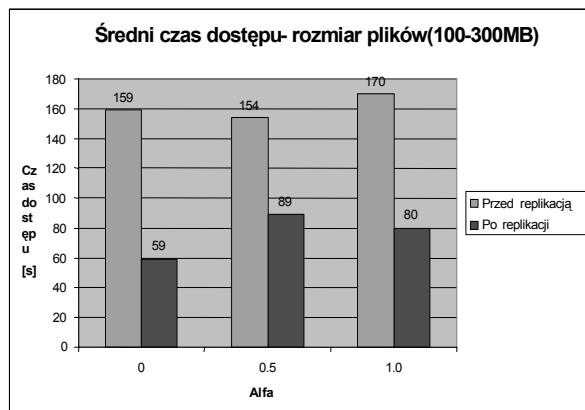
Taki system automatycznej replikacji został zaimplementowany w ramach projektu SGGrid [93], w którym to jednym z zadań było stworzenia wirtualnego systemu archiwizacji danych (WSA). Celem systemu WSA [99] jest integracja zasobów przechowywania danych w różnych ośrodkach obliczeniowych w Polsce w jeden logiczny system plików. WSA zapewnia również dostęp do fragmentów plików oraz zamawianie plików. Architektura systemu jest pokazana na rys. 16, natomiast wyniki przeprowadzonych testów efektywności replikacji na rys. 17. Testy były przeprowadzone dla trzech różnych schematów obciążenia określone różnym parametrem w rozkładzie Zipfa dla generowanych żądań. Te same żądania generowane były najpierw dla losowo umieszczonych plików i następnie powtarzane po zakończeniu replikacji.

W projekcie Clusterix [94] automatyczną replikację zastosowano dla systemu zarządzania danymi [100]. Zastosowano trzy rodzaje replikacji: replikacja początkowa, replikacja na żądanie i replikacja w oparciu o analizę statystyczną. Replikacja początkowa służy do automatycznego tworzenia replikę potrzebnych do obliczeń danych przy uruchamianiu zadania na gridzie. W replika-

cja na żądanie tworzone są określonej ilości replik na żądanie użytkownika. Replikacja statystyczna służy do tworzenia replik w oparciu o analizę statystyczną wcześniejszych żądań dostępu do danych. W tym przypadku system sam podejmuje decyzje o stworzeniu dodatkowej repliki oraz określa najbardziej odpowiedniej lokalizacji dla niej.



Rys. 16. Architektura WSA
Fig. 16. WSA architecture



Rys. 17. Wyniki testów
Fig. 17. Results of tests

4. Podsumowanie

Symulacja komputerowa pozwala coraz częściej na rezygnację z eksperymentów typu „in vivo” lub „in vitro” na rzecz eksperymentów „in silico”, dostarczających wniosków na poziomie systemowym, wieloskalowym, w oderwaniu od indywidualnych zjawisk. Odpowiada to badaniom złożonych zjawisk otaczającego nas świata. Z kolei w zakresie współczesnych tendencji w rozwoju biznesu obserwuje się potrzebę globalnej współpracy opartej na wiedzy.

W celu realizacji takiego systemowego podejścia, konieczne są prace w zakresie algorytmów i ich efektywnej implementacji w różnych środowiskach obliczeniowych. Szczególnie interesująca wydaje się ostatnio tendencja obliczeń gridowych i rozwoju środowisk gridowych z uwagi na ich elastyczność oraz potencjalnie duże możliwości, wsparta dodatkowo efektywnym wykorzystaniem zasobów na podstawie zgromadzonej wiedzy i doświadczenia.

Podziękowania. Prace w obszarze problematyki obliczeniowej zrealizowano we współpracy z Prof. D.A. Yuenem i A.Z. Dudkiem (University of Minnesota) oraz Prof. E.M. Tory (Mount Allison University). Prace z zakresu technologii gridowych zosta-

ły zrealizowane we współpracy z Kolegami z ACK CYFRONET AGH: Tomaszem Gubałą, Piotrem Nowakowskim, Marcinem Radeckim, Tomaszem Szepeńcem, dr. Łukaszem Dutką, Bartoszem Kryżą i Łukaszem Skitałem, a także we współpracy z następującymi zespołami zagranicznym: Prof. Peter M.A. Slood (Universiteit van Amsterdam), Prof. Roland Wismueller (Siegen University), Prof. Vaidy Sunderam (Emory University), Prof. Thomas Fahringer (Innsbruck University), Prof. Peter Brezany (Vienna University). Zagadnienia wsparcia biznesowego, wiedzy i jej ontologicznej reprezentacji rozwijano wspólnie z dr. Ladislavem Hluchy (Institute of Informatics, Bratislava), dr. Simonem Lambertem (Rutherford Appleton Laboratory, UK) i dr. Steffenem Ungerem (Fraunhofer Institute FIRST, Berlin).

5. Literatura

- [1] Kitowski J., Alda W., Boryczko K., Bubak M., Dzwinel W., Funika W., Moscinski J., Nikolow D., Pogoda M., Slotka R., Wcislo R., Large Scale Computing as a Vehicle for Studies in Computational and Computer Science, Zeszyty Naukowe Uniwersytetu Jagiellońskiego, Prace informatyczne Z. 10, Dept. of Computer Science, Jagiellonian University, 2000, pp. 51-68.
- [2] Hoogerbrugge P.J. i Koelman J.M., „Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics”, Europhys. Lett., 19/155, 1992.
- [3] Boryczko K., Dzwinel W., Yuen D., „A”, „Dynamical Clustering of Red Blood Cells in Capillary Vessels”, J Mol. Modeling, 9, 16-33, 2003.
- [4] Dzwinel W., Boryczko K., Yuen D., „A”, „A Discrete-Particle Model of Blood Dynamics in Capillary Vessels”, J Colloid Int Sci, 258/1, 163-173, 2003.
- [5] Tsubota K., Wada S., Kamada H., Kitagawa Y., Lima R., Yamaguchi T., A Particle Method for Blood Flow Simulation, Application to Flowing Red Blood Cells and Platelets, J.Earth Simulator, 5, 2-7, March 2006.
- [6] Jain A.K., Murty M.N. i Flynn P.J., „Data clustering: a review”, ACM Comput. Surv., 31/3, 264-323, 1999.
- [7] Zahn C.T., „Graph-theoretical methods for detecting and describing gestalt clusters”, IEEE Trans. Comput., C-20, 68-86, 1972.
- [8] Karypis G., Han E.H. i Kumar V., „Chameleon: A hierarchical clustering algorithm using dynamic modeling”, IEEE Computer: Special Issue on Data Analysis and Mining, 32/8, 68-75, 1999.
- [9] Sander J., Ester M., Kriegel H.P. i Xu X., „Density-based clustering in spatial databases: The algorithm gdbscan and its applications”, Data Mining and Knowledge Discovery, 2/2, 169-194, 1998.
- [10] Sudipto K.S. i Rastogi R., „Cure: An efficient clustering algorithm for large databases”, w SIGMOD 1998, Proc. ACM SIGMOD International Conference of Management of Data, 73-84, Waszyngton, 1998.
- [11] Ertöz L., Steinbach M. i Kumar V., „Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data”, Proc. Third SIAM Intl. Conf. on Data Mining, San Francisco, 2003.
- [12] Gowda C.K. i Krishna G., „Agglomerative clustering using the concept of nearest neighborhood”, Pattern Recognition, 10/2, 105-112, 1978.
- [13] Arodz T., Kurdziel M., Sevre E. i Yuen D.A., „Pattern Recognition Techniques for Automatic Detection of Suspicious-looking Anomalies in Mammograms”, Computer Methods and Programs in Biomedicine, Elsevier Science, (w druku).
- [14] Boryczko K. i Kurdziel M., „Approximate clustering of noisy multidimensional data”, wysłane do Pattern Recognition.
- [15] Arodz T., Kurdziel M., Popiela T.J., Sevre E. i Yuen D.A., „A 3D Visualization System for Computer-Aided Mammogram Analysis”, University of Minnesota, Research Report UMSI 2004/181, 2004.
- [16] Boryczko K. i Kurdziel M., „Recognition of Subtle Microcalcifications in High-Resolution Mammograms”, w Proceedings of 4th International Conference on Computer Recognition Systems CORES'05, Advances in Soft Computing, 485-492, Wrocław, 2005.
- [17] Popiela T., Urbanik A., Arodz T. i Kurdziel M., „Computer-aid system for the detection of clustered microcalcifications in digital mammography”, Polish J.Radiology, 69(S1), 67-68, 2004, abstract.
- [18] T. Arodz. Invariant Object Recognition Using Radon-based Transform. Computing and Informatics, 24(2):183-199, Slovak Academy of Sciences, 2005.

- [19] T. Arodz, M. Kurdziel, T.J. Popiela, E.O.D. Sevre, D.A. Yuen. Detection of Clustered Microcalcifications in Small Field Digital Mammography. *Computer Methods and Programs in Biomedicine*, 81(1):56-65, Elsevier, 2006.
- [20] S. Goryczka, T. Arodz. Complex-network-based Methodology for Analysis of Biomedical Data. *Bio-Algorithms and Med-Systems*, CM UJ, 2006, (w druku).
- [21] T. Arodz. Boosting the Fisher Linear Discriminant with Random Feature Subsets. w: Kurzyński, M., Woźniak, M., Puchała, E., Żolnierek, A., (red.) "Computer Recognition Systems - CORES'05", *Advances in Soft Computing*, Springer, 2005.
- [22] T. Arodz, D. A. Yuen, A. Z. Dudek. Ensemble of Linear Models for Predicting Drug Properties. *Journal of Chemical Information and Modeling* 46(1):416-423, American Chemical Society, 2006.
- [23] T. Arodz. Margin-based Diversity Measures for Ensemble Classifiers. "Computer Recognition Systems - CORES'05", *Advances in Soft Computing*, Springer, 2005.
- [24] T. Arodz. Training Set Size in Ensemble Feature Selection for Clinical Proteomics. *Bio-Algorithms and Med-Systems*, 1(1/2): 107-110, CM UJ, 2005.
- [25] T. Jurczyk and B. Głut. Generation of triangular meshes for complex domains. *Computer Science, Annual of University of Mining and Metallurgy, Kraków*, 3:71-93, 2001.
- [26] T. Jurczyk and B. Głut. Metric 3d surface mesh generation using Delaunay criteria. *LNCS*, 3992:302-309, Springer, 2006.
- [27] T. Jurczyk and B. Głut. Triangular and quadrilateral meshes on 3D surfaces. *Proc. Fifth World Congress on Computational Mechanics (WCCM V)*, Vienna, Austria, July 7-12 2002. Vienna University of technology, Austria. ISBN 3-9501554-0-6, <http://wccm.tuwien.ac.at>.
- [28] T. Jurczyk and B. Głut. Generation of Good Quality Quadrilateral Elements for Anisotropic Surface Meshes, *Proc. of the ADMOS2003 Conference on Adaptive Modeling and Simulation*, Göteborg, Szwecja, 29 Sept. - 1 Oct. 2003. Chalmers University of Technology, Szwecja.
- [29] A. Adamek and B. Głut. Symboliczna dekompozycja blokowa w generatorze siatek sześciociennych. *Computer Science, Annual of University of Science and Technology, Kraków*, 7:7-31, 2005.
- [30] A. Adamek and J. Kitowski. Vertex location for symbolic block decomposition method of linear polyhedron. In *Proc. CMS'05 Conference on Computer Methods and Systems*, vol. 1, pages 407-418, Kraków, November 2005. ONT Kraków.
- [31] T. Jurczyk and B. Głut. Efficiency Aspects of Control Space Definition for the Generation of Surface Meshes. *Proc. of the SIAM Conf. on Computational Science and Engineering*, Orlando, Florida, USA, 10 - 15 February 2005.
- [32] B. Głut and T. Jurczyk. Mesh Adaptation Based on Discrete Data. *LNCS* 391:559-566, Springer, 2006.
- [33] B. Głut, T. Jurczyk, and M. Pietrzyk. Adaptive mesh generation for non-steady state heat transport problems, *Proc. Fifth World Congress on Computational Mechanics (WCCM V)*, <http://wccm.tuwien.ac.at>, Vienna, Austria, July 7-12 2002. Vienna University of technology, Austria. ISBN 3-9501554-0-6, <http://wccm.tuwien.ac.at>.
- [34] B. Głut, T. Jurczyk, and M. Pietrzyk. Adaptive remeshing in the FE modeling of moving boundary problems, *Proc. of ADMOS2003 Conf. on Adaptive Modeling and Simulation*, Göteborg, Szwecja, 29 Sept - 1 Oct. 2003. Chalmers University of Technology, Szwecja.
- [35] B. Głut and T. Jurczyk. Quality and efficiency of FEM mesh adaptation with respect to control space. *Proc. SIAM Conf. on Computational Science and Engineering*, Orlando, Florida, USA, Feb.10-15, 2005.
- [36] B. Głut and T. Jurczyk. Definition and interpolation of discrete metric for mesh generation on 3D surfaces. *Computer Science, Annual of University of Science and Technology, Kraków*, 7: 89-103, 2005.
- [37] T. Jurczyk and B. Głut. Domain decomposition coupled with Delaunay mesh generation. *LNCS*, 2329:353-360, Springer, 2002.
- [38] B. Głut, T. Jurczyk, P. Breitkopf, A. Rassineux and P. Villon. Geometry Decomposition Strategies for Parallel 3D Mesh Generation, *Proc. CMS'05 Conference on Computer Methods and Systems*, 1:443-450, Kraków, 14-16 Nov. 2005. ONT Krakow.
- [39] T. Jurczyk and B. Głut. Organization of the mesh structure. *LNCS* 3037:646-649, Springer, 2004.
- [40] Alda, W., Białoskorski, M., Gorecki, R., Rybicki, J., Grid Approach to Heat Transfer Simulation in Atomic-continuum Model, *Proc. Cracow Grid Workshop - CGW'04*, December 13-15 2004, ACC-Cyfronet UST, 2005, Kraków, p. 462.
- [41] Alda, W., Balcerek, B., Dyczkowski, M., Kowaliczek, G., Polak, S., Wcislo, R., Zachara, A., Distributed Visualization System Based on WebServices, *Proc. Cracow Grid Workshop - CGW'04*, Dec. 13-15 2004, ACC-Cyfronet UST, 2005, Kraków, p. 479.
- [42] Tory EM, Bargiel M, Honeycutt RL. A three-parameter Markov model for sedimentation III. A stochastic Runge-Kutta method for computing first-passage times. *Powder Technol.* 1994;80:133-146.
- [43] M. Bargiel, E.M. Tory, „A five-parameter Markov model for simulating the paths of sedimenting particles”, *Applied Mathematical Modelling*, submitted.
- [44] M. Bargiel, E.M. Tory, „Simulation of sedimentation and fluidization of polydisperse suspensions via a Markov model”, *Chemical Engineering Science* 61(17) (2006) 5575-5589.
- [45] M. Bargiel, E.M. Tory, “An extension of the particle-based approach to simulating the sedimentation of polydisperse suspensions”, *Int. J. Min. Pro.* 79(4) (2006) 235-252.
- [46] M. Bargiel, R.A. Ford, E.M. Tory, Simulation of sedimentation of polydisperse suspensions: a particle-based approach, *AIChE Journal* 51 (2005) 2457-2468.
- [47] Bargiel, M., Tory, E.M., 2005. Stability of tridisperse suspensions. *Comput. Visual. Sci.*, in press.
- [48] Mościński, J., Bargiel, M., Rycerz, Z.A., and Jacobs, P.W.M., „The force-biased algorithm for the irregular close packing of equal hard spheres”, *Molecular Simulation*, 3 (1989) 201.
- [49] Bargiel, M. and Tory, E.M., „Packing fraction and measures of disorder of ultradense irregular packings of equal spheres. I. Nearly ordered packing.”, *Advanced Powder Technology* 4 (1993) 79-101.
- [50] Bargiel, M. and Tory, E.M., „Packing fraction and measures of disorder of ultradense irregular packings of equal spheres. II. Transition from dense random packing.” *Advanced Powder Technology* 12 (4)(2001) 533-557.
- [51] Bezrukov, A., Bargiel, M., and Stoyan, D., „Statistical analysis of Simulated Random Packings of Spheres.” *Part. Part. Syst. Charact.* 19 (2002) 111-118.
- [52] Bezrukov, A., Stoyan, D. and Bargiel, M., „Spatial statistics for simulated packings of spheres.” *Image Anal. Stereol.* 20 (2001) 203-206.
- [53] Project EU IST CrossGrid – www.crossgrid.org
- [54] Bubak, M., Fahringer, T., Hluchy, R., Hoheisel, A., Kitowski, J., Unger, S., Viano, G., Votis, K., K-WfGrid Consortium, K-WfGrid – Knowledge Based Workflow System for Grid Applications, *Proc. of Cracow Grid Workshop - CGW'04*, Dec. 13-15 2004, ACC-Cyfronet UST, 2005, Kraków, pp. 39.
- [55] Bubak, M., Funika, W., Balis, B., Gubala, T., Malawski, M., Radecki, M., Rycerz, K., Smetek, M., Szepieniec, T., CYFRONET Contribution to CoreGRID: Problem Solving Environments, Tools and GRID Systems, *Proc. of Cracow Grid Workshop – CGW'04*, Dec. 13-15 2004, ACC-Cyfronet UST, 2005, Kraków, pp. 45-57.
- [56] R. Wismueller, M. Bubak, W. Funika, B. Balis, A Performance Analysis Tool for Interactive Applications on the Grid, *Intl. Journal of High Performance Computing Applications*, 18(3)(2004) 305-316.
- [57] R. Wismueller, M. Bubak, W. Funika, High-Level Application Specific Performance Analysis Using the G-PM Tool, *Proc. PVM/MPI 2005*, LNCS 3666, 317-324, Springer, 2005.
- [58] Bubak, M., Funika, W., Wismueller R. Mełtel P., Orłowski R., Monitoring of Distributed Java Applications, *Future Generation Computer Systems*, 19 (2003) 651-663.
- [59] Funika, W., Bubak, M., Smetek, M., Monitoring System for Distributed Java Applications, *Computational Science – ICCS 2004*. 4th International Conference, Kraków, Poland, June 2004, LNCS, 3038, 472-479, Springer, 2004.
- [60] W. Funika, M. Koch, D. Dziok, M. Smetek, R. Wismueller, Performance Visualization of Web Services Using J-OCM and SCIRun/TAU, in: *Proc. HPCC 2005*, LNCS 3726, 666-671, Springer, 2005.
- [61] B. Balis, M. Bubak, M. Radecki, T. Szepieniec, and R. Wismueller. Application Monitoring in CrossGrid and Other Grid Projects. In *Grid Computing. Proc. 2nd European Across Grids Conference*, pp. 212-219, Nicosia, Cyprus, Jan 2004. Springer.
- [62] K. Rycerz, M. Bubak, M. Malawski, P. M. A. Slood, A Framework for HLA-Based Interactive Simulations on the Grid, *SIMULATION*, 81(1)(2005) 67-76.
- [63] Rycerz, K., Bubak, M., Malawski, M., Slood, P., HLA Grid Based Support for Simulation of Vascular Reconstruction, *Proc. CoreGRID Workshop "Integrated Research in Grid Computing"*, Nov. 28-30, 2005, Technical Report TR-05-22, 2005, pp. 165-174.

- [64] T. Gubala, M. Bubak, M. Malawski, K. Rycerz, Semantic-Based GridWorkflow Composition, Parallel Processing and Applied Mathematics: 6th International Conference, PPAM 2005, Poznan, Poland, September 11-14, 2005, LNCS, 3911, 651-658, Springer, 2006
- [65] Marian Bubak, Tomasz Gubala, Michal Kapalka, Maciej Malawski, and Katarzyna Rycerz. Workflow composer and service registry for grid applications. *Future Generation Computer Systems*, 21(1):79-86, 2005.
- [66] B. Balis, M. Bubak, J. Dziwisz, H.-L. Truong, T. Fahringer, Integrated Monitoring Framework for Grid Infrastructure and Application, in *Innovation and the Knowledge Economy. Issues, Applications, Case Studies*, IOS Press, 2005, pp. 269-276.
- [67] Jurczyk, P. Golenia, M., Malawski, M., Kurzyniec, D., Bubak, M., Sunderam, V. S., A System for Distributed Computing Based on H2O and JXTA, Proc. Cracow Grid Workshop - CGW'04, Dec. 13-15 2004, ACC-Cyfronet UST, 2005, Kraków, pp. 257-268.
- [68] R. M. Badia, O. Beckmann, M. Bubak, D. Caromel, V. Getov, S. Isaiadis, V. Lazarov, M. Malawski, S. Panagiotidi, J. Thiyagalangam, Lightweight Grid Platform: Design Methodology, in *CoreGRID Integration Workshop 2005*, 2005.
- [69] Badia R.M., Bubak, M., Funika, W., Smetek, M., Performance Monitoring of GRID superscalar applications with OCM-G, Proc. of the CoreGRID Workshop "Integrated Research in Grid Computing", Nov. 28-30, 2005, Technical Report TR-05-22, 2005, pp. 229-236.
- [70] B. Balis, M. Bubak, and M. Wegiel. A solution for adapting legacy code as web services. In *Component Models and Systems for Grid Applications*, 57-75. Springer, 2005. ISBN 0-387-23351-2.
- [71] T. Gubala, M. Bubak, GridSpace – Semantic Programming Environment for the Grid, in *Parallel Processing and Applied Mathematics: 6th International Conference, PPAM 2005*, Poznan, Poland, September 11-14, 2005, LNCS, 3911, 172-179, Springer, 2006.
- [72] K. Rycerz, M. Bubak, M. Malawski, P. M. A. Sloot, A Grid Service for Management of Multiple HLA Federate Processes, in *Parallel Processing and Applied Mathematics: 6th International Conference, PPAM 2005*, Poznan, Poland, September 11-14, 2005, LNCS 3911, 699-706, Springer, 2006.
- [73] B. Balis, M. Bubak, K. Guzy, Fine-Grained Instrumentation and Monitoring of Legacy Application in a Service-Oriented Environment, in *Proc. Computational Science – ICCS 2006*, 6th Intl Conference, Reading, UK, May 2006, LNCS, 3992, 542-548, Springer, 2006.
- [74] H.-L. Truong, B. Balis, M. Bubak, J. Dziwisz, T. Fahringer, A. Hoheisel, Towards Distributed Monitoring and Performance Analysis Services in the K-WfGrid Project, 6th Intl. Conf. PPAM 2005, Poznan, Poland, Sept. 11-14, 2005, LNCS, 3911, 156-163, Springer, 2006.
- [75] B. Balis, M. Bubak, J. Dziwisz, K. Rozkwitalski, Towards a Generic P2P Monitoring Infrastructure for Grid Resources and Applications. Proc. 2nd CoreGRID Workshop on GRID and Peer to Peer Systems Architecture, Paris, 2006.
- [76] P. Jurczyk, M. Golenia, M. Malawski, D. Kurzyniec, M. Bubak, V. S. Sunderam, Enabling Remote Method Invocations in Peer-to-Peer Environments: RMIX over JXTA, 6th Intl. Conf., PPAM 2005, Poznan, Poland, Sept. 11-14, 2005, LNCS 3911, 667-674, Springer, 2006.
- [77] P. Sloot, C. Boucher, M. Bubak, A. Hoekstra, P. Plaszcak, A. Posthumus, D. van de Vijver, VIROLAB - a Virtual Laboratory for Decision Support in Viral Diseases Treatment, Proc. of Cracow Grid Workshop - CGW'05, Nov. 20-23 2005, ACC-CYFRONET AGH, 2006, Kraków, pp. 33.
- [78] M. Malawski, D. Kurzyniec, V. Sunderam. MOCCA - towards a distributed CCA framework for metacomputing. Proc. 19th IEEE Intl Parallel and Distributed Processing Symposium (IPDPS'05) – Joint Workshop on High-Performance Grid Computing and High-Level Parallel Programming Models – HIPS-HPGC, April 4-8, 2005, Denver, Colorado, USA, p. 174a. IEEE Computer Society Press, 2005.
- [79] Pellucid homepage (IST-2001-34519): <http://www.sadiel.es/Eropa/pellucid/>.
- [80] M. Majewska, K. Krawczyk, R. Slota, J. Kitowski, L. Hluchy, S. Lambert, Information Searching for an Experience Management Platform of the UE Pellucid Project, TASK QUARTERLY. Sci. Bull. of Academic Centre in Gdansk, 8(4)(2004) 513-523, TASK Publ..
- [81] Krawczyk, K., Majewska, M., Dziewierz, M., Slota, R., Balogh, Z., Kitowski, J., Lambert, S., Reuse of Organisational Experience Harnessing Software Agents, Proc. Intl. Conf. Comput. Science – ICCS 2004, LNCS, 3038, 583-590, Springer, 2004.
- [82] Kitowski, J., Krawczyk, K., Majewska, M., Dziewierz, M., Slota, R., Model of Experience for Public Organizations with Staff Mobility, Proc. 5th IFIP Int. Conf. Knowledge Management in Electronic Government – KMGov 2004, LNAI, 3035, 75-84, Springer, 2004.
- [83] Slota, R., Majewska, M., Dziewierz, M., Krawczyk, K., Laclavik, M., Balogh, Z., Hluchy, L., Kitowski, J., Lambert, S., Ontology Assisted Access to Document Repositories in Public Sector Organizations, Proc. Intl. Conf. Parallel Processing and Applied Mathematics – PPAM 2003, LNCS, 3019, 700-705, Springer, 2004.
- [84] Laclavik, M., Zoltan, B., Hluchy, L., Slota, R., Krawczyk, K., Dziewierz, M., Distributed Knowledge Management Based on Software Agents and Ontology, Proc. Intl. Conf. Parallel Processing and Applied Mathematics – PPAM 2003, LNCS, 3019, 694-699, Springer, 2004.
- [85] Slota R., Majewska M., Kitowski J., Lambert S., Laclavik M., Hluchy L., Viano G. Evaluation of experience-based support for organizational employees, Chapter in: J. Lu, et al (Eds.), *E-service Intelligence – Methodologies, Technologies and Applications*, vol.37, Springer, 2006, in press.
- [86] B. Kryza, M. Majewska, R. Slota and J. Kitowski. Unifying Grid Metadata Representations Through Ontologies, Proc. 6-th Intl. Conf. on Parallel Processing and Applied Mathematics PPAM'2005, LNCS 3911, 683-690, Springer, 2006.
- [87] K-Wf Grid Project website: <http://www.kwfgrid.net>
- [88] B. Kryza, J. Pieczykolan, M. Babik, M. Majewska, R. Slota, L. Hluchy and J. Kitowski: Managing Semantic Metadata in K-Wf Grid with Grid Organizational Memory, Proc. Cracow Grid Workshop '05, Cracow, Nov. 2005, ACK Cyfronet AGH, 2006, pp. 66-73.
- [89] B. Kryza, R. Slota, M. Majewska, J. Pieczykolan, J. Kitowski. Grid Organizational Memory - Provision of a High-Level Grid Abstraction Layer Supported by Ontology Alignment. *Future Generation Computer Systems*, 2006, Elsevier, in print.
- [90] R. Slota, J. Zieba, M. Majewska, B. Kryza, J. Kitowski, Ontology Alignment and Ontology Similarity for Extension of Service Ontology in Grid Environment, Proc. Cracow Grid Workshop '05, Nov. 2005, ACK-Cyfronet AGH, 2006, Kraków, pp. 41-48.
- [91] J. Pieczykolan, B. Kryza, J. Kitowski. Semi-automatic Creation of Adapters for Legacy Application Migration, Proc. 6th Intl. Conf. Comput. Science, ICCS2006, LNCS 3994, 252-259, Springer, 2006.
- [92] B. Kryza, Ł. Skitał, J. Kitowski, Maohzen Li, Takebumi Itagaki. Analysis of Interoperability Issues Between EGEE and VEGA Grid Infrastructures, In Proc. of Int. Conf. On High Performance Computing and Communications, Munich, 2006, LNCS, Springer, in print.
- [93] SGGrid: Large-scale computing and visualization for virtual laboratory using SGI cluster (in Polish), KBN Project, <http://www.wcss.wroc.pl/pb/sggrid/>
- [94] Clusterix - <http://www.clusterix.pcz.pl>
- [95] Nikolow, D., Slota, R., Kitowski, J., Nyczyk, P., Otfinowski, J., Tertiary Storage System for Index-based Retrieving of Video Sequences, LNCS, 2110, 435-444, Springer, 2001.
- [96] D. Nikolow, R. Slota, J. Kitowski, Performance Aspects of Data Delivery Using Hierarchical Storage Management Systems in the Grid, TASK QUARTERLY. Scientific Bulletin of Academic Centre in Gdansk, 8(4) 537-548 (2004) TASK Publishing, 2004.
- [97] Nikolow, D., Slota, R., Kitowski, J., Gray Box Based Data Access Time Estimation for Tertiary Storage in Grid Environment, Proc. Int. Conf. Parallel Proc. and Mathematics (PPAM 2003), LNCS, 3019, 182-188, Springer, 2004.
- [98] M. Kuta, D. Nikolow, R. Slota, J. Kitowski, Data Access Time Estimation for the CASTOR HSM System, Proc. Parallel Processing and Applied Mathematics: 6th Intl. Conf., PPAM 2005, LNCS 3911, 148-155, Springer, 2006.
- [99] Nikolow, D., Slota, R., Kitowski, J., Skitał, Ł., Virtual Storage System for the Grid Environment, Proc. Intl. Conf. Comput. Science - ICCS 2004, LNCS, 3036, 458-461, Springer, 2004.
- [100] R. Slota, L. Skitał, D. Nikolow, J. Kitowski, Algorithms for Automatic Data Replication in Grid Environment, Proc. Parallel Processing and Applied Mathematics: 6th Intl. Conf., PPAM 2005, LNCS, 3911, 707-714, Springer, 2006.

Artykuł recenzowany