

Piotr Tarłowski

Sieć Laboratoriów Wirtualnej Organizacji Działań:

Łódź-Olsztyn-Polkowice-Poznań

Wyższa Szkoła Informatyki w Łodzi

MODELOWANIE POJĘCIOWE SYSTEMÓW PARTNERSKICH W WIRTUALNEJ ORGANIZACJI DZIAŁAŃ

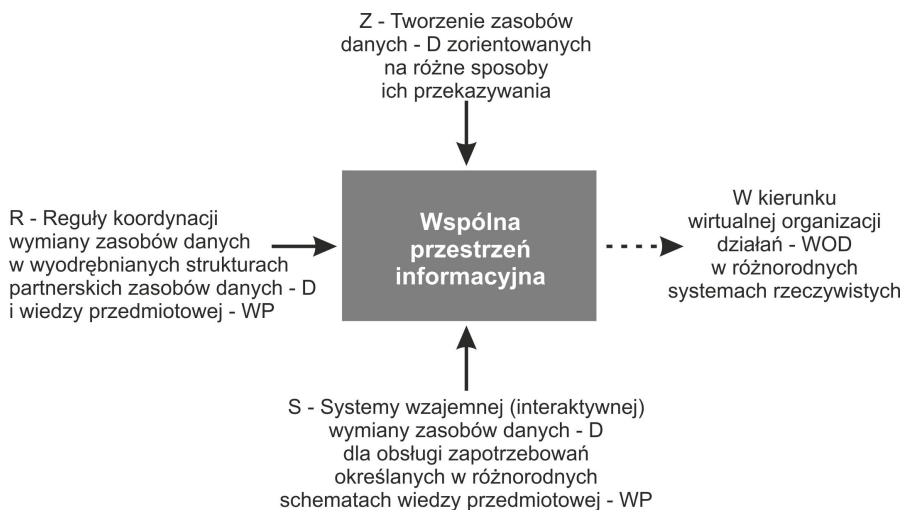
Streszczenie – W pracy omówiono wybrane problemy tworzenia wspólnych przestrzeni informacyjnej, środowiska wirtualnej współpracy ludzi i wirtualnego współdziałania organizacji, korzystając z modeli systemów partnerskich – P2P. Wśród problemów szczegółowych, zwraca się uwagę na: (1) sposoby koordynacji wymiany zasobów cyfrowych, (2) utrzymywanie systemów partnerskich, które bazuje na procesach dostępu do węzłów sieci partnerskiej. Rozważane problemy są ważne dla modelowania konceptualnego struktur gridowych, co jest potrzebne dla budowy różnorodnych systemów wirtualnej organizacji działań.

1 Wprowadzenie

Kreowanie systemów wirtualnej organizacji działań (WOD) [5, 8] wymaga tworzenia wspólnej przestrzeni informacyjnej, za pośrednictwem której możliwa jest realizacja różnorodnych funkcji i zadań w strukturach współdziałania organizacji i współpracy ludzi. Chodzi o kreowanie środowiska systemów WOD, polegającego na rozwiązywaniu problemów organizacji i użytkowania wspólnej przestrzeni informacyjnej (rys. 1).

Proces kształtowania wspólnej przestrzeni informacyjnej opiera się na trzech wymiarach [4]: Z, S, R. Pierwszy wymiar – Z wynika z zasad digitalizacji zasobów danych – D. Polegają one na stosowaniu określonych metod tworzenia dokumentów źródłowych w postaci plików komputerowych uwzględniających różne tryby dostępu do tych zasobów (komputerowe odczytywanie, odczytywanie z możliwościami wprowadzania zmian). Drugi wymiar – S określa procedury komunikacyjne przekazywania, czy wymiany zasobów. Polega to na implementowaniu aplikacji systemów komunikacyjno-wykonawczych przeznaczonych do obsługi zapotrzebowań informacyjnych, odwzorowywanych w schematach wiedzy przedmiotowej – WP, ustalanych przez uczestników systemów WOD. Trzeci wymiar – R, warunkujący zasady współdziałania instytucji i współpracy ludzi w systemach WOD, wymaga ustalania reguł

koordynacji wymiany danych. Polega to na wyodrębnianiu określonych struktur partnerskich zasobów D&WP we wspólnej przestrzeni informacyjnej, przeznaczonych do realizacji wspomaganie informacyjnego, rozwiązywania definiowanych funkcji czy zadań w działaniach systemów wirtualnych.



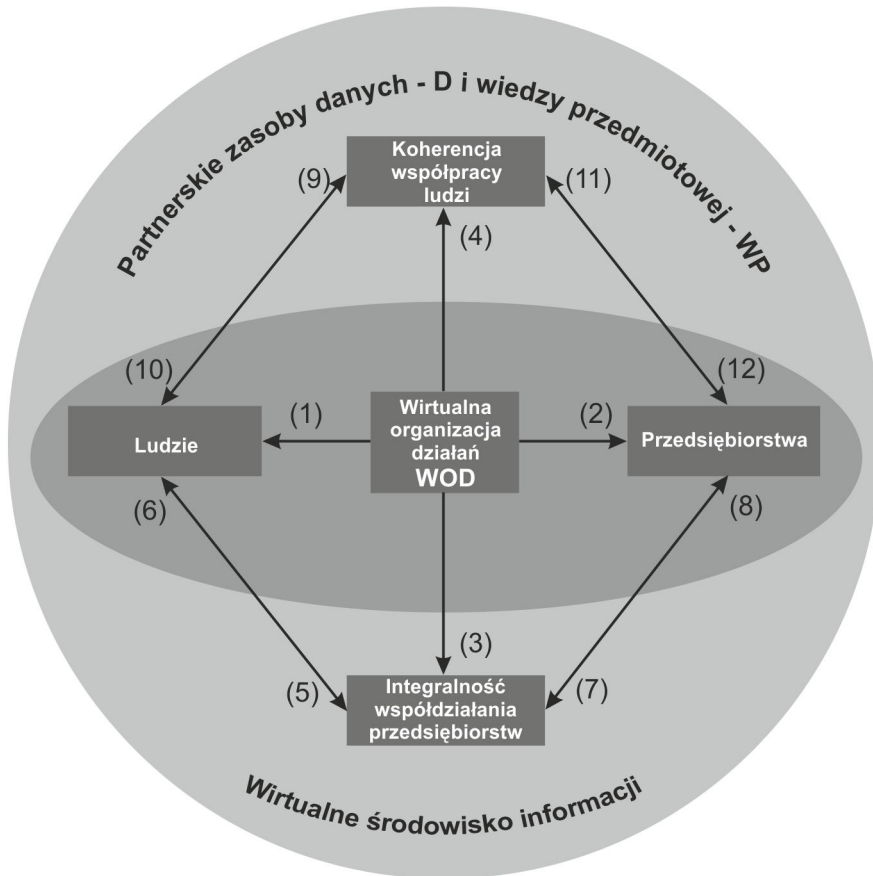
Rys. 1. Główne czynniki kształtowania środowiska informacyjnego dla kreowania systemów wirtualnej organizacji działań

2 Środowisko informacyjne i procesowa organizacja działań

Problemy organizacji i użytkowania wspólnej przestrzeni informacyjnej możemy zaliczyć do podstawowych czynników kreowania systemów WOD. We wspólnej przestrzeni informacyjnej możliwa jest procesowa integracja współdziałania różnorodnych organizacji [6, 9]. Modelem procesowej integracji współdziałania w systemie partnerskich zasobów jest logistyka różnorodnych procesów zarządzania informacją (rys. 2).

Logistyka wynika z zestawiania modeli tematycznych procesów przetwarzania danych przeznaczonych do obsługi wspólnie (kooperacyjnie) rozwiązywanych wielozadaniowych problemów.

Znaczenie łuków 1-12 grafu z rys. 2 wynika z obsługi różnorodnych tematycznych problemów przetwarzania danych. Zasoby danych i procedury obsługi zadań (1-9) pozwalają modelować złożoną sieć danych – D i wiedzy przedmiotowej – WP, przeznaczone do wspomaganie informacyjnego funkcji, zadań i rozwiązywanych problemów w systemach WOD, a więc w złożonym systemie informacyjnym.



Rys. 2. Ogólna struktura logistyki współdziałania systemów WOD

3 Modelowanie zasobów danych w systemach partnerskich

Wspólną przestrzeń informacyjną systemów WOD i konkretne sieci danych D&WP kształtują metody wymiany plików inspirowane zastosowaniami architektury gridowej integracji zasobów – określanymi mianem partnerskich systemów cyfrowych: P2P. Chodzi o organizację wymiany danych konsolidującą procedury tworzenia i wymiany dokumentów źródłowych w schematach WP poszczególnych partnerów WOD. Architektura P2P udostępnia rozproszone mechanizmy wymiany danych z znaczeniem odwzorowań między wybranymi zasobami innych partnerów, wynikających ze znaczenia partnerstwa, m.in. tzw. sąsiedztwa, czy znajomości. W architekturze P2P nie istnieje schemat globalnego opisu zasobów. Grupy partnerów mogą tworzyć sieci dla rozwiązywania różno-

rodnych zadań, np. grupa pracująca nad rejestracją pacjentów, grupa zajmująca się obsługą procesu diagnozowania, leczenia itp.

W architekturze P2P partnerzy wymieniają między sobą polecenia, realizują te polecenia, a serwery pełnią rolę centralnych katalogów i wspomaganie wyszukiwania zapotrzebowań informacyjnych partnerów. Zasada działania sieci polega na tym, że wszyscy partnerzy włączani są w proces wymiany plików. Zwykle plik dzielony jest na fragmenty stanowiące jednostki wymiany. Niezbędna jest kontrola autentyczności fragmentów.

W architekturze P2P odwzorowywane są procesy wymiany danych. P2P składa się [2] z otwartej sieci rozproszonych aplikacji sieciowych *partnerów* (ang. *peer*), gdzie każda aplikacja może wymieniać dane i usługi ze zbiorem innych aplikacji, tworzących zbiór jego *znajomych* (ang. *acquaintances*). Zakłada się, że aplikacje są w pełni autonomiczne w wyborze swoich znajomych. Ponadto zakłada się, że nie ma kontroli w formie globalnych rejestrów, usług, ani globalnego zarządzania zasobami.

Dane pamiętane w różnych bazach danych mogą mieć współzależność (semantyczną). Związki te mogą być wyrażane za pomocą formuł koordynacyjnych określających sposób powiązań aplikacji z jego znajomymi.

Formuły koordynacyjne wymuszają spójność danych i propagację uaktualnień. Dodatkowo aplikacje wymagają znajomości protokołu wymiany i ustalenia poziomu koordynacji między nimi. Poziom koordynacji powinien być dynamiczny, co oznacza, że zestaw formuł koordynacyjnych może się zmieniać w zależności od tego, czy powiązania między parami są luźne, zacieśniają się, czy są w ogóle likwidowane (z powodu zmiany zainteresowań i zadań poszczególnych aplikacji).

Przy tak dynamicznych zmianach nie można przyjąć istnienia globalnego schematu dla wszystkich zasobów danych w sieci P2P, a nawet dla wszystkich *znajomych* zasobów danych. Co więcej, zmiany te oznaczają, że partnerzy powinni mieć możliwość wzajemnej integracji przy minimalnej interwencji człowieka.

Mamy więc do czynienia ze środowiskiem dynamicznym i rozszerzalnym. Partnerzy mogą wchodzić i wychodzić z systemu. Partner wchodząc do systemu wnosi swoje dane, swoje schematy oraz odwzorowania między schematami swoimi i swoich partnerów. W ten sposób dynamicznie kształtuje się sieć na różne sposoby powiązanych ze sobą systemów współpracujących w procesach integracji i wymiany danych. Wymiana danych między partnerami wiąże się z problemami *koordynacji* tej wymiany oraz *integracji* danych w procesie wymiany.

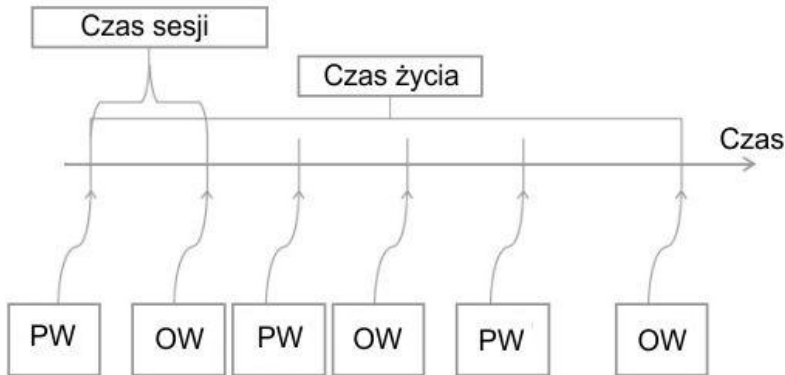
4 Wybrane problemy utrzymywania systemów partnerskich

Problem utrzymywania systemów partnerskich – bardzo dynamicznych i zmiennych – jest jednym z ważniejszych zagadnień wymagających analizy podczas tworzenia systemów WOD.

Ponieważ w systemach P2P węzły dowolnie przyłączają się i opuszczają sieć (w przewidywalnym czasie), pojawia się zjawisko „kotłowania” (ang. *churn*) węzłów sieci [7]. Dla zapewnienia niezawodności sieci, węzły sąsiadujące wymagają uaktualnień z węzłami dołączającymi lub opuszczającymi sieć. Jeżeli tablica tras (ang. *routing table*) węzła straci dokładność, opóźnienie systemu wzrośnie z powodu wysyłania wiadomości do nieistniejącego węzła. Jeżeli zjawisko *churn* będzie dalej wzrastać, sieć zacznie się w końcu dzielić, powodując wyszukiwania zapytań do powracających sprzecznych wyników i znaczne obniżenie wydajności i jakości sieci.

Mechanizmy utrzymania sieci wymagają skutecznych metod znajdowania nieaktualnych wpisów w tablicach tras i zamiany ich na aktualne. Ponieważ wyszukiwanie zapytań w sieci i utrzymanie ruchu sieci zależy od przepustowości łącza, to utrzymanie ruchu powinno zostać ograniczone tylko do tego, co jest ściśle potrzebne do posiadania precyzyjnej tablicy tras. Tak więc utrzymanie ruchu przez węzeł powinno być proporcjonalne do tempa zmian jego sąsiadów. Algorytmy utrzymania sieci powinny być skonstruowane w taki sposób, by zawierać skuteczne mechanizmy radzące sobie z tak dynamicznymi zmianami – muszą być przygotowane na ciągły rozwój sieci P2P oraz na to, że operacje dołączenia i odłączania węzłów sieci zdarzają się jednocześnie. Muszą brać pod uwagę zmieniające się tempo zmian, decydować czy prędkość obsługi wzrośnie aby utrzymać sieć stabilną, lub czy system zaakceptuje wzrost opóźnienia obsługi sieci wprowadzając jakieś limity. Protokół utrzymania sieci musi zapewnić, że ilość łącza wykorzystana przez węzeł do wymiany wiadomości „utrzymujących” nie wzrośnie nadmiernie w przypadku rozrastania się sieci. Inaczej przepustowość dostępu węzłów może zostać zminimalizowana i sieć staje się bezużyteczna.

Wydajność protokołu utrzymania jest tempem w którym każdy węzeł zużywa zasoby sieci dla utrzymania. Konsumpcja zasobów potrzebna do utrzymania sieci powinna być utrzymana na możliwie niskim poziomie, ponieważ zasoby te są niedostępne dla aplikacji partnerskich wykonywujących „zwykłą” pracę (czyli np. wymianę plików). Minimalne tempo uaktualnień, które musi osiągnąć każdy węzeł w sieci aby utrzymać sieć w dobrym stanie oraz jak wiele pracy potrzeba aby utrzymać właściwy stan trasy potrzebny do szybkiego i poprawnego wyszukiwania zapytań pokazuje w dużym uproszczeniu rys. 3.



Rys. 3. Dostępność węzła w sieci partnerskiej

Czas sesji węzła jest czasem pomiędzy przyłączeniem (*PW*) i opuszczeniem (*OW*) sieci (jednorazowo). Czas życia węzła jest czasem między przyłączeniem się do sieci P2P pierwszy raz i czasem jej trwałego opuszczenia. Suma czasów sesji węzła podzielona przez jego czas życia jest jego *dostępnością*. Czas sesji węzła jest ważną miarą w odniesieniu do wyszukiwania w rozproszonej tablicy mieszającej (ang. *DHT - Distributed Hash Table*). Tymczasowa strata sąsiada osłabia poprawność i osiągi DHT. Niedostępność sąsiadów redukuje spójność węzła, powodując wybieranie przez węzła ścieżek nieoptymalnych, co prowadzi do opóźnień wyszukiwania i wzrostu nieefektywności utrzymania sieci (zwłaszcza w przypadku dalszych dynamicznych zmian). Kiedy węzły są niestabilne, z krótkimi i nieprzewidywalnymi czasami sesji, to zapamiętywanie takich sąsiadów jest mało wartościowe (w odniesieniu do wykonywania zapytań).

Stopień obniżenia wydajności sieci spowodowany *churn*, zależy od jego zdolności wykrywania zdarzeń przyłączeń i opuszczenia sieci przez węzły i jego zdolności wyboru nowych zgłoszeń do sieci. Wyróżniamy dwa podejścia do utrzymania sieci: aktywne i bierne.

W podejściu aktywnym węzeł przetwarza awarię lub odłączenie sąsiedniego węzła (lub dołączenie nowego) poprzez rozesłanie kopii informacji o nowym sąsiedzie do innych węzłów w sieci. Dla zachowania przepustowości łącza, węzeł może wysłać tylko różnice od ostatniego stanu sieci, ale całkowita liczba stanów ma złożoność kwadratową $O(k^2)$ dla sąsiedniego zbioru k -węzłów.

Podejście aktywne niesie ze sobą ryzyko tworzenia się dodatkiego sprzężenia zwrotnego: Rozważmy przypadek, gdy połączenie dostępne węzła do sieci jest tak zablokowane, że powoduje

przekroczenia czasu dla węzła, tak że wierzy iż jeden z jego sąsiadów zepsuł się (lub opuścił system). Jeżeli węzeł będzie pracował w trybie aktywnym, to rozpocznie operacje konserwacji. Ta operacja doda więcej pakietów do już zablokowanego fragmentu sieci, co zwiększy prawdopodobieństwo, że także inne węzły mylnie wydedukują, że inni sąsiedzi zawiedli w systemie. Ponieważ proces ten będzie postępował, może nastąpić „załamanie się” sieci w tym miejscu.

W przypadku niskiej dynamiki zmian, aktywne utrzymanie jest skuteczne i skalowalne, ponieważ wiadomości są wysyłane tylko w odpowiedzi na faktyczne zmiany. Jednak gdy tempo zmian wzrasta, proces ten staje się bardziej kosztowny - węzeł widzi więcej zmian, gdy jego zbiór sąsiadów się rozrasta.

W podejściu biernym, węzeł okresowo dzieli się swoim zbiorem sąsiadów z każdym członkiem tego zbioru. Ten proces ma miejsce niezależnie czy węzeł wykryje zmiany w zbiorze sąsiadów czy nie. Węzeł wybiera jednego losowego członka ze zbioru sąsiadów, by dzielić się swoim stanem w każdym okresie optymalizacji. To podejście chroni przepustowość łącza i oznacza złożoność logarytmiczną $O(\log k)$, gdzie k oznacza rozmiar zbioru sąsiadów.

Poprzez rozdzielenie tempa konserwacji sieci od wykrywania zmian węzłów, unikamy sprzężenia zwrotnego. Ponadto, przez wydłużenie okresu utrzymania systemu z obserwacją czasów stanu systemu, metoda ta wprowadza ujemny sprzężenie zwrotne, które zwiększa elastyczność sieci. Taki proces może być bardziej zachowawczy w wykrywaniu węzłów, co prowadzi do zmniejszenia niestabilności związanej z aktywnym utrzymaniem sieci. Jedynym minusem tej techniki jest to, że sąsiedzi, którzy rzeczywiście odeszli, pozostają przez jakiś czas w tabeli tras węzła. Wyszukiwanie zapytań, które zostałyby wysłane przez tych odłączonych sąsiadów, jest w ten sposób opóźnione, doprowadzając do długiego oczekiwania na odpowiedź. Aby rozwiązać ten problem, węzeł przestaje wysyłać przez sąsiada, jeżeli zauważy jego kolejne opóźnienia. Ponadto, niektóre sieci używają równoległego wyszukiwania aby zrekompensować posiadanie niedokładnych wpisów w tabeli tras.

Podczas niskiej dynamiki zmian, metoda bierna staje się nieekonomiczna. Jednakże, kiedy zbiór sąsiadów węzła się rozrasta (i rośnie wielkość sieci), metoda staje się bardziej wydajna i przepustowość łącza zużyta w czasie wysokiej dynamiki zmian jest kontrolowana.

5 Podsumowanie

W rozwoju modeli pojęciowych systemów komputerowego wspomaganie działań pojawiają się obecnie nowe składowe, wynikające

ze stosowania usług międzysieciowych informacyjnych i komunikacyjnych. Wiele nowych kwestii wynika ze współczesnych trendów rozwoju systemów komputerowo integrowanej organizacji działań, których zwieńczeniem jest wirtualna organizacja działań.

Budowa systemów wirtualnej organizacji działań bazuje na koordynacji gromadzenia zasobów i wymiany informacji. Problemy koordynacji wynikają z logistyki procesów współużytkowania sieciowego partnerskich zasobów cyfrowych, odwzorowujących konsolidację funkcji i zadań partnerów uczestniczących w organizacji działań.

Punktem wyjścia budowy systemów wirtualnej organizacji działań jest strukturalizacja sieciowego gromadzenia danych cyfrowych, traktowanych jako partnerskie zasoby cyfrowe, a także synteza usług koordynacji ich wymiany.

W środowisku informacyjnym wirtualnych usług modeluje się procesy koordynacji wymiany zasobów. Rezultatem jest ustanawianie partnerstwa we współdziałaniu, które wynika z ustalenia reguł koordynacji współdzielenia partnerskich zasobów danych cyfrowych (peer digital data resources).

Strukturyzacja gromadzenia zasobów cyfrowych oraz procesy koordynacji ich wymiany są podstawą budowy systemów wirtualnej organizacji działań o różnym zasięgu. Systemy te, oparte na sieciach partnerskich, muszą zmierzyć się z dynamicznymi zmianami i potrzebą utrzymania sieci jako całości.

Literatura

- [1] Górecki J., Kierzkowski Z., Kleban H., Kołopieńczyk A., Kwiatkowska L., Lubiatowski M., N'sir M., Tarłowski P., *Services oriented data storing and access to information* (eds.: Skołod B. and Krenczyk D), WNT, Warszawa 2003; pp. 186-191.
- [2] Tarłowski P., Pankowski T., *Koordynacja wymiany danych w środowisku P2P*, [Rozdział w książce]: Bazy danych – Struktury, Algorytmy, Metody, Wydawnictwo Komunikacji i Łączności, Warszawa 2006; s. 37-45.
- [3] Sikorski A., Tarłowski P., *Zarządzanie współbieżnością w systemie integrowania informacji z uwzględnieniem uzgadniania i tolerowania sprzeczności*, *Bazy Danych Nowe Technologie*, Rozdz. 8, Kozielski S., Małysiak B., Kasprowski P., Mrozek D. (red.), Wydawnictwa Komunikacji i Łączności, Warszawa 2007, str. 91-103.
- [4] Tarłowski P., *Koordynacja partnerstwa w środowisku informacyjnych współdziałania*, [w]: Komputerowe wspomaganie dydaktyki (red. E. Kaćki i J. Stemoczyńska), Polskie Tow.

- Informatyki Med. i Wyższa Szkoła Informatyki w Łodzi, Łódź, 25-26 czerwca 2008; s.115-117.
- [5] Kierzkowski Z., *Virtual Activities Organization within Information Society Structures*, 21st International CODATA Conference: Scientific Information for Society – from Today to the Future, [In]: Abstracts, Kyiv, Ukraine, 5-8 October 2008; pp.81-82.
- [6] Kierzkowski Z., N'sir M., Polkowski Z., Tarlowski P., *Data and Problem Knowledge Environment within Co-operation Structures*, 21st International CODATA Conference: Scientific Information for Society – from Today to the Future, [In]: Abstracts, Kyiv, Ukraine, 5-8 October 2008; pp.54-55.
- [7] Buford J.F., Heather Y., Lua E.K., *P2P networking and Applications*, Elsevier 2009.
- [8] Kierzkowski Z., *Wirtualna organizacja działań w rozwoju technologii społeczeństwa informacyjnego*, Przegląd Elektrotechniczny, Nr 9/2009, s. 172-178.
- [9] Kierzkowski Z., Tarlowski P., *Wirtualne systemy informacji przedmiotowej*, Przegląd Elektrotechniczny, Nr 9/2009, s. 179-181.

CONCEPTUAL MODELING OF THE PARTNER SYSTEMS IN VIRTUAL ACTIVITIES ORGANIZATION

Summary - The paper discusses some problems of creating a common information space, which is the environment of the virtual collaboration and of people and the virtual cooperation of organizations, using the partner systems – P2P models. Among the detailed problems, attention is paid mainly to: (1) how to coordinate the exchange of digital resources, (2) maintenance of the partner systems based on the processes of access to the partner network nodes. Considered problems are important for conceptual modeling of grid structures, what is needed for the building of various virtual activities organization systems.