

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Systemy rozproszone dużej skali		Kod 1010512321010519247
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Systemy rozproszone	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca:		
dr inż. Anna Kobusińska email: Anna.Kobusinska@cs.put.poznan.pl tel. 61 6652964 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów operacyjnych, rozproszonych systemów operacyjnych oraz sieci komputerowych.
2	Umiejętności:	Powinien posiadać umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	Kompetencje społeczne	Powinien również rozumieć konieczność poszerzania swoich kompetencji i mieć gotowość do podjęcia współpracy w ramach zespołu. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
Cel przedmiotu:		
1. Przekazanie studentom podstawowej wiedzy związanej z wyzwaniem konstrukcji systemów rozproszonych dużej skali, w zakresie prezentacji teoretycznych i praktycznych aspektów konstrukcji rozproszonych systemów dużej skali. 2. Prezentacja problematyki przetwarzania dużych ilości danych (big data) w systemach rozproszonych dużej skali. 3. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z przetwarzaniem w systemach rozproszonych dużej skali.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie algorytmów i złożoności, architektury systemów komputerowych, systemów operacyjnych, technologii sieciowych, języków i paradygmatów programowania - [K_W4] 2. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: architektura i klasyfikacja systemów rozproszonych dużej skali i ich środowiska komunikacyjnego, protokołów epidemicznych i plotkujących, systemów przetwarzania dużej ilości danych - [K_W5] 3. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w informatyce i w wybranych pokrewnych dyscyplinach naukowych - [K_W6] 4. ma podstawową wiedzę o cyklu życia systemów informatycznych sprzętowych lub programowych - [K_W7] 5. zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z wybranego obszaru dotyczącego przetwarzania w rozproszonych systemach operacyjnych - [K_W8]		
Umiejętności:		

<ol style="list-style-type: none">1. potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K_U1]2. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - [K_U5]3. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne oraz eksperymentalne - [K_U9]4. potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych) oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne - [K_U10]5. potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi - [K_U12]6. potrafi ocenić przydatność i możliwości wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych - [K_U13]7. potrafi zaproponować ulepszenia (usprawnienia) istniejących rozwiązań technicznych - [K_U21]8. potrafi ocenić przydatność metod i narzędzi służących do rozwiązania zadania inżynierskiego, polegającego na budowie lub ocenie systemu informatycznego lub jego składowych, w tym dostrzec ograniczenia tych metod i narzędzi - [K_U24]9. potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne, w tym zadania nietypowe oraz zadania zawierające komponent badawczy - [K_U27]
Kompetencje społeczne:
<ol style="list-style-type: none">1. zna przykłady i rozumie przyczyny wadliwie działających systemów informatycznych, które doprowadziły do poważnych strat finansowych, społecznych lub też do poważnej utraty zdrowia, a nawet życia - [K_K4]2. potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania - [K_K6]3. potrafi myśleć i działać w sposób przedsiębiorczy - [K_K8]

Sposoby sprawdzenia efektów kształcenia
<p>Ocena formująca:</p> <ol style="list-style-type: none">a) w zakresie wykładów:<ul style="list-style-type: none">- na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach,b) w zakresie laboratoriów:<ul style="list-style-type: none">- na podstawie oceny bieżącego postępu realizacji zadań, <p>Ocena podsumowująca:</p> <ol style="list-style-type: none">a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:<ul style="list-style-type: none">- ocenę wiedzy i umiejętności wykazanych na pisemnym egzaminie o charakterze problemowym składającym się z 4 otwartych pytań. Za każde pytanie można uzyskać 10 punktów, stąd maksymalna liczba punktów do zdobycia na kolokwium wynosi 40. Aby uzyskać z kolokwium zaliczeniowego ocenę pozytywną należy zdobyć minimum 20 punktów- omówienie wyników egzaminu,b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:<ul style="list-style-type: none">- ocenę przygotowania studenta do poszczególnych sesji zajęć laboratoryjnych (sprawdzian wejściowy) oraz ocenę umiejętności związanych z realizacją ćwiczeń laboratoryjnych,- ocenę sprawozdania przygotowywanego częściowo w trakcie zajęć, a częściowo po ich zakończeniu; ocena ta obejmuje także umiejętność pracy w zespole,- ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych / laboratoryjnych poprzez 1 kolokwium w semestrze,- ocenę i obronę przez studenta sprawozdania z realizacji projektu, <p>Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:</p> <ul style="list-style-type: none">- omówienia dodatkowych aspektów zagadnienia,- efektywność zastosowania zdobytej wiedzy podczas rozwiązywania zadane go problemu,- umiejętność współpracy w ramach zespołu praktycznie realizującego zadanie szczegółowe w laboratorium,- uwagi związane z udoskonaleniem materiałów dydaktycznych,- wskazywanie trudności percepcyjnych studentów umożliwiające bieżące doskonalenia procesu dydaktycznego.
Treści programowe

Program wykładu obejmuje następujące zagadnienia:

1. Przedstawienie wyzwań związanych z budową i przetwarzaniem w systemach rozproszonych dużej skali: klasyfikacje systemów, idea działania, zastosowania, przykłady implementacji
2. Architektury systemów P2P (ustrukturalizowane, nieustrukturalizowane, hybrydowe); organizacja węzłów, topologie, skalowalność, równoważenie obciążeń; gossip-based small word networks, Kleinberg's peer sampling
3. Nieustrukturalizowane systemy P2P: dołączanie/odłączanie węzłów, peer-sampling, strategie wyszukiwania zasobów (flooding-based multicast, tree-based multicast, random walk, expanding ring, rendezvous point, bubblecast) na przykładzie systemów: FreeNet, FastTrack, eDonkey, Gnutella
4. Ustrukturalizowane systemy P2P: dołączanie/odłączanie węzłów; przestrzeń identyfikatorów, Distributed Hash Table (DHT), utrzymywanie informacji o stanie przetwarzania, redukcja opóźnień, odporność na awarie na przykładzie systemów Pastry, Chord, Tapestry, CAN (ang. Content Addressable Network), Kademia, Kad.
5. Monitorowanie systemów P2P (aggregation, polling, slicing, renaming), anonimowość w systemach P2P: miary anonimowości, mechanizmy zapewnienia anonimowości
6. Protokoły Bittorrent i Bitcoin ? zasada działania, przeznaczenie
7. Wprowadzenie do tematyki dużych danych: definicje i charakterystyka ?Big data?, źródła dużych danych, aspekty przetwarzania dużych danych, wyzwania związane z ?Big data?
8. Architektura systemów Big Data: pozyskiwanie dużych danych, składowanie dużych danych, przetwarzanie wsadowe/przetwarzanie strumieni danych, analiza dużych danych; algorytmy rozproszone w przetwarzaniu dużych danych; przykłady wykorzystania narzędzi dużych danych w istniejących systemach (m.in. na przykładzie Google, Facebook, Yahoo, LinkedIn, Cloudera, Microsoft)
9. Wprowadzenie do tematyki baz danych NoSQL: klasyfikacja ze względu na modele danych (key value, column-oriented, document-oriented, column-oriented, graph-oriented); budowa i działanie systemów NoSQL (data partitioning, load balancing, replication, data versioning, membership management, failure handling) na przykładzie Amazon Dynamo; Google BigTable, HBase, Cassandra.
10. Przetwarzanie dużych danych z wykorzystaniem platformy Hadoop (MapReduce);
11. Przetwarzanie dużych danych z wykorzystaniem Resilient Distributed Datasets (RDD) i platformy Apache Spark
12. Systemy interaktywnych zapytań w systemach dużych danych (systemy Hive i Shark)
13. Przetwarzanie danych reprezentujących grafy dużej skali ? system Pregel (Bulk Synchronous Parallel model, Vertex-centric, Superstep - sequence of iterations, Master-worker model)
14. Systemy zarządzania zasobami Mesos i YARN (architektura, algorytmy lokowania zasobów)

Program laboratorium obejmuje zagadnienia związane z problematyką przetwarzania dużych ilości danych w systemach Redis, Casandra, Spark, MongoDB. Dla każdego z systemów omawiane są następujące zagadnienia:

1. Instalacja, konfiguracja, interfejs programistyczny, typy danych, podstawowe operacje dostępne w danym systemie.
2. Praktyczne wykorzystanie systemu: modelowanie danych; sortowanie, filtrowanie, opcje tabel, indeksy, kolekcje, liczniki, lightweight transactions i inne zaawansowane funkcje, stronicowanie wyników. Wykorzystanie znaczników czasowych. Poziomy spójności. Awarie węzłów, partycje sieciowe i rozszynchronizowanie zegarów. Konflikty znaczników czasowych.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie zadań
2. ćwiczenia laboratoryjne: demonstracja, dyskusja, warsztaty, ćwiczenia praktyczne, praca w zespole

Literatura podstawowa:

1. Peer-to-Peer Computing ? principles and applications, Q. Vu, M. Lupu, B.C Ooi, Springer, 2010
2. Peer-to-peer systems and applications, R. Steinmetz, K. Wehrle, Springer, 2005
3. Large Scale Network-Centric Distributed Systems, H.Sarbazi-Azad, A.Y.Zomaya, Wiley-IEEE Computer Society Press, 2013
4. NoSQL distilled, P. Sadalage, M. Flower, Addison-Wesley, 2013
5. Spark in Action, Bonać M., Zečević P., Manning, 2015

Literatura uzupełniająca:

1. M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, M. Van Steen: Gossip-based peer sampling. ACM Trans. Comput. Syst 25(3) 2007.
2. Andrew Oram, redaktor. Peer-to-peer: Harnessing the bene?ts of a disruptive technology. O'Reilly, Sebastopol, CA, 2001.
3. Shen X, Yu H, Buford J, Akon M, Handbook of Peer-to-peer networking, 2010
4. Nitin Sawant, Himanshu Shah, Big data application architecture Q&A, Springer, 2013
5. J.Berman, Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information, Morgan-Kaufman, 2013

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
----------	--------------

1. udział w zajęciach laboratoryjnych / ćwiczeniach	30
2. przygotowanie do ćwiczeń laboratoryjnych	7
3. dokończenie (w ramach pracy własnej) sprawozdań z ćwiczeń laboratoryjnych: 3 x 1 godz.	3
4. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych / projektu (część konsultacji może być realizowana drogą elektroniczną)	6 30
5. udział w wykładach	8
6. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 80 stron	14
7. przygotowanie do egzaminu i obecność na egzaminie: 12 godz. + 2 godz.	
Obciążenie pracą studenta	
forma aktywności	godzin
ECTS	
Łączny nakład pracy	98
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	68
Zajęcia o charakterze praktycznym	37