

<b>KARTA OPISU MODUŁU KSZTAŁCENIA</b>		
Nazwa modułu/przedmiotu <b>Projekt eksploracji danych</b>		Kod <b>1010512321010510011</b>
Kierunek studiów <b>Informatyka</b>	Profil kształcenia (ogólnoakademicki, praktyczny) <b>ogólnoakademicki</b>	Rok / Semestr <b>1 / 2</b>
Ścieżka obieralności/specjalność <b>Technologie wytwarzania oprogramowania</b>	Przedmiot oferowany w języku: <b>polski</b>	Kurs (obligatoryjny/obieralny) <b>obligatoryjny</b>
Stopień studiów: <b>II stopień</b>	Forma studiów (stacjonarna/niestacjonarna) <b>stacjonarna</b>	
Godziny Wykłady: <b>30</b> Ćwiczenia: - Laboratoria: - Projekty/seminaria: <b>30</b>		Liczba punktów <b>4</b>
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) <b>kierunkowy z danego kierunku</b>		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki <b>nauki techniczne</b>		Podział ECTS (liczba i %) <b>4 100%</b>
<b>Odpowiedzialny za przedmiot / wykładowca:</b>  dr inż. Krzysztof Dembczyński email: krzysztof.dembczynski@put.poznan.pl tel. 61 6652936 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
<b>Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:</b>		
1	<b>Wiedza:</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_W1-2, K_W4, K_W6-15, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl  Przedmiot wymaga w szczególności dobrego przygotowania w ramach programowania, podstawowej wiedzy z zakresu systemów baz danych oraz statystyki, analizy danych i uczenia maszynowego.
2	<b>Umiejętności:</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
3	<b>Kompetencje społeczne</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_K1-9, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl  Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
<b>Cel przedmiotu:</b> Przekazanie studentom praktycznej wiedzy w zakresie eksploracji masywnych danych (bardzo dużych zbiorów danych) poprzez rozwiązywanie zadań konkursowych dotyczących tej dziedziny. Rozwijanie u studentów umiejętności rozwiązywania problemów dotyczących eksploracji masywnych danych.		
<b>Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia</b>		
<b>Wiedza:</b>		
1. Ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w informatyce związanych z eksplozją danych i technologiami eksploracji masywnych zbiorów danych. - [K_W6]		
2. Ma podstawową wiedzę o cyklu życia systemów eksploracji masywnych danych. - [K_W7]		
3. Zna podstawowe metody, techniki i narzędzia stosowane do złożonych zadań inżynierskich w dziedzinie eksploracji masywnych danych. - [K_W8]		
<b>Umiejętności:</b>		

<ol style="list-style-type: none"><li>1. Potrafi pozyskiwać informacje z literatury (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie. - [K_U1]</li><li>2. Potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia. - [K_U5]</li><li>3. Potrafi planować i przeprowadzać eksperymenty, w tym pomiary i symulacje komputerowe, interpretować uzyskane wyniki i wyciągać wnioski. - [K_U8]</li><li>4. Potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne oraz eksperymentalne. - [K_U9]</li><li>5. Potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne. - [K_U10]</li><li>6. Potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi w zakresie eksploracji masywnych danych. - [K_U12]</li><li>7. Potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych w zakresie eksploracji masywnych danych. - [K_U13]</li><li>8. Potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne, w tym zadania nietypowe oraz zadania zawierające komponent badawczy dotyczące eksploracji masywnych danych - [K_U25]</li></ol>
<b>Kompetencje społeczne:</b>
<ol style="list-style-type: none"><li>1. Rozumie, że w informatyce, a zwłaszcza w eksploracji masywnych danych, wiedza, technologie i umiejętności bardzo szybko stają się przestarzałe. - [K_K1]</li><li>2. Zna możliwości dalszego dokształcania się w zakresie eksploracji masywnych danych. - [K_K3]</li><li>3. 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]</li><li>4. Potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania. - [K_K6]</li></ol>

### Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
  - na podstawie odpowiedzi na pytania dotyczące materiału omówionego na wykładach,
- b) w zakresie projektu:
  - na podstawie oceny bieżącego postępu realizacji projektu.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:
  - ocenę dwóch sprawozdań dotyczących postępów w rozwiązywaniu zadania konkursowego oraz przygotowanego seminarium dotyczącego wybranego tematu z zakresu eksploracji masywnych danych.
- b) w zakresie projektu weryfikowanie założonych efektów kształcenia realizowane jest przez:
  - ocenę ostatecznego wyniku w konkursie eksploracji danych.

### Treści programowe

Program wykładu obejmuje następujące zagadnienia:

- Przedstawienie zadania konkursowego, jego głównych aspektów, trudności i wyzwań. W każdym roku zadanie konkursowe będzie dotyczyło innego problemu eksploracji danych. Zadania konkursowe będą pochodzić z takich serwisów jak [www.kaggle.com](http://www.kaggle.com), konkursów organizowanych przy konferencjach eksploracji danych lub od firm, które w ramach swojej działalności spotykają się z problemami eksploracji danych (w 2014 r. dane do konkursu zostały dostarczone przez firmę OLX).
  - Część wykładów będzie prowadzonych na zasadzie seminarium, na którym studenci, osoby zaproszone lub prowadzący przedmiot będą przedstawiać zagadnienia związane bezpośrednio z konkursem. Mogą to być, na przykład, nowe rozwiązania technologiczne dotyczące przetwarzania i eksploracji masywnych danych, omówienie zaawansowanych algorytmów eksploracji danych lub podstaw teoretycznych metod eksploracji danych.
- Zajęcia projektowe prowadzone są w formie piętnastu dwugodzinnych spotkań dyskusyjnych. Projekty realizowane są w zespołach 2-osobowych. Program projektu obejmuje udział w konkursie eksploracji masywnych danych:
- W pierwszej części projektu główny nacisk jest kładziony na organizację danych i ich wstępne przetwarzanie. Na poszczególnych spotkaniach omawiane są zadania konkursowe i postępy w realizacji poszczególnych zadań związanych z organizacją i wstępnym przetwarzaniem danych.
  - W drugiej części projektu główny nacisk kładziony jest na dobór odpowiednich metod analitycznych i ich efektywną implementację. Na poszczególnych spotkaniach omawiane są algorytmy eksploracji danych, które mogą posłużyć do rozwiązania zadania konkursowego, jak także ich efektywna implementacja.

Metody dydaktyczne:

1. Wykład: prezentacja multimedialna ilustrowana przykładami podawanymi na tablicy.
2. Projekt: dyskusja dotycząca możliwych rozwiązań zadań konkursowych.

<b>Literatura podstawowa:</b>		
1. Mining of Massive Datasets, A. Rajaraman, J. D. Ullman, Cambridge University Press, 2012 ( <a href="http://infolab.stanford.edu/~ullman/mmds.html">http://infolab.stanford.edu/~ullman/mmds.html</a> )		
<b>Literatura uzupełniająca:</b>		
1. Hadoop in Action, Ch. Lam, Manning Publications Co., 2011.		
2. Data-Intensive Text Processing with MapReduce, J.Lin, Ch. Dyer, Morgan and Claypool Publishers, 2010 ( <a href="http://lntool.github.com/MapReduceAlgorithms/">http://lntool.github.com/MapReduceAlgorithms/</a> )		
3. Elements of Statistical Learning: Second Edition, T. Hastie, R. Tibshirani, J. Friedman, Springer, 2009. ( <a href="http://www-stat.stanford.edu/~tibs/ElemStatLearn/">http://www-stat.stanford.edu/~tibs/ElemStatLearn/</a> )		
<b>Bilans nakładu pracy przeciętnego studenta</b>		
<b>Czynność</b>	<b>Czas (godz.)</b>	
1. Udział w zajęciach projektowych	30	
2. Praca nad projektem:	20	
3. Udział w konsultacjach związanych z realizacją procesu kształcenia (częściowo mogą być realizowane drogą elektroniczną)	8	
4. Przygotowanie do zajęć.	4	
5. Praca nad zadaniem projektowym	10	
6. Udział w wykładach/seminariach	30	
7. Zapoznanie się ze wskazaną literaturą i materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 60 stron	6	
<b>Obciążenie pracą studenta</b>		
<b>forma aktywności</b>	<b>godzin</b>	<b>ECTS</b>
Łączny nakład pracy	108	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	68	3
Zajęcia o charakterze praktycznym	60	2