



Syllabus przedmiotu

w ramach projektu „@kademia na rzecz rozwoju ICT – studia podyplomowe”

Rok akademicki 2010/2011

Nazwa przedmiotu	Web Mining
Kod przedmiotu	WEB

1. Opis

Nazwa kierunku	Metody informatyczne w badaniach społecznych	Kod kierunku	MIBS
Jednostka prowadząca	Wydział Informatyki, Warszawa		
Prowadzący przedmiot	dr Marcin Sydow		
Semestr studiów	1		
Język wykładowy	Polski		
Plan godzinowy	Wykłady	12	
	Laboratoria	22	
Warunki zaliczenia	zaliczenie (ćwiczenia) + egzamin		



2. Wymogi

<p>Założenia i cele</p>	<p>Celem kursu jest praktyczne wprowadzenie do współczesnych zagadnień eksploracji danych ze szczególnym uwzględnieniem sieci WWW, wyszukiwania w niej informacji oraz zjawisk ekonomicznych i społecznych obserwowanych w sieci WWW. Wykłady mają dwa główne wątki: uczenie maszynowe i jego praktyczne zastosowania w analizie danych oraz budowa i zagadnienia związane z wyszukiwarkami internetowymi. W trakcie ćwiczeń studenci m.in. Rozwiązują praktyczne zadania dotyczące eksploracji danych rzeczywistych przy użyciu odpowiedniego oprogramowania specjalistycznego.</p>
<p>Treści programowe</p>	<p>Wprowadzenie do Eksploracji Danych</p> <ul style="list-style-type: none"> - Podstawy i historia sztucznej inteligencji - Wprowadzenie do eksploracji danych i uczenia maszynowego, uczenie z nadzorem i bez, tabele decyzyjne, atrybuty, klasyfikacja i regresja - Reprezentacja wiedzy, tabele decyzyjne, drzewa decyzyjne - Selekcja modelu, problem przetrenowania i ewaluacja modeli - Podstawy Sieci Neuronowych, Perceptron <p>Zagadnienia Wyszukiwarek WWW</p> <ul style="list-style-type: none"> - podstawy technologiczne sieci WWW, rola i historia wyszukiwarek - budowa wyszukiwarek WWW - automatyczne zbieranie kolekcji dokumentów WWW - podstawy przetwarzanie tekstu w kontekście wyszukiwania informacji i eksploracji danych - graf WWW, jego własności i rola w eksploracji danych WWW - algorytmy analizy struktury grafu WWW i ich zastosowania w eksploracji danych WWW, PageRank, Hits i ich warianty - aspekty ekonomiczne sieci WWW, reklamy wyszukiwarkowe - aspekty społeczne sieci WWW, spam wyszukiwarkowy i jego konsekwencje, wykrywanie i zwalczanie <p>Praktyczne aspekty eksploracji danych</p> <ul style="list-style-type: none"> - własności i posługiwanie się pakietem Weka - przygotowanie i przetwarzanie danych rzeczywistych - rozwiązywanie praktycznych zadań eksploracji danych przy użyciu pakietu Weka <p>Ponadto, w ramach ćwiczeń studenci wykonują krótkie prezentacje na wybrane tematy związane z eksploracją danych WWW uprzednio przygotowane pod kierunkiem prowadzącego w oparciu o publikacje naukowe, inne ogólnie dostępne źródła i własne doświadczenia.</p>
<p>Wykaz literatury</p>	<p>Literatura podstawowa:</p>



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



	<ol style="list-style-type: none">1. Ian H. Witten et al. „Data Mining”, 2nd edition, Morgan Kaufmann, 20052. Christopher D. Manning et al. „Introduction to Information Retrieval”, Cambridge University Press, 2008 <p>Literatura uzupełniająca</p> <ol style="list-style-type: none">1. Bing Liu „Web Data Mining”, Springer, 20082. Ian H. Witten et al. „Web Dragons”, Morgan Kaufmann, 2007
Kontakt do prowadzącego przedmiot	e-mail: msyd@poljap.edu.pl
Kontakt do Koordynatora merytorycznego kierunku	Mgr inż. Radosław Nielek, -mail: nielek@pjawst.edu.pl