Politechnika Krakowska im. Tadeusza Kościuszki

Karta przedmiotu

obowiązuje studentów rozpoczynających studia w roku akademickim 2014/2015

Wydział Fizyki, Matematyki i Informatyki

Kierunek studiów: Informatyka | Profil: Ogólnoakademicki

Forma sudiów: stacjonarne | Kod kierunku: I

Stopień studiów: I

Specjalności: Brak specjalności

1 Informacje o przedmiocie

Nazwa przedmiotu	Przetwarzanie języka naturalnego
Nazwa przedmiotu w języku angielskim	
Kod przedmiotu	WFMiI I oIS D2 14/15
Kategoria przedmiotu	Przedmioty specjalnościowe
Liczba punktów ECTS	5.00
Semestry	7

2 Rodzaj zajęć, liczba godzin w planie studiów

Semestr	Wykład	Ćwiczenia	Laboratorium	Laboratorium komputerowe	Seminarium	Projekt
7	30	0	30	0	0	0

3 Cele przedmiotu

Cel 1 Poznanie zagadnień związanych z przetwarzaniem dokumentów tekstowych.

Cel 2 Nauka projektowania algorytmów przetwarzania dokumentów tekstowych.

Cel 3 Praktyka implementacji programów do przetwarzania dokumentów tekstowych.

4 Wymagania wstępne w zakresie wiedzy, umiejętności i innych kompetencji

1 Algorytmy i struktury danych, projektowanie

2 Podstawy programowania języków skryptowych (Perl)

5 Efekty kształcenia

EK1 Kompetencje społeczne Potrafi pracować indywidualnie i w zespole informatyków, w tym także potrafi zarządzać swoim czasem oraz podejmować zobowiązania i dotrzymywać terminów.

EK2 Wiedza Zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu prostych zadań informatycznych z zakresu budowy systemów komputerowych, sieci komputerowych i technologii sieciowych oraz systemów wbudowanych.

EK3 Umiejętności Umie sformułować specyfikację prostych systemów informatycznych w odniesieniu do sprzętu, oprogramowania systemowego i cech funkcjonalnych aplikacji.

EK4 Umiejętności Umie - zgodnie z zadaną specyfikacją - zaprojektować oraz zrealizować prosty system informatyczny, używając właściwych metod, technik i narzędzi.

6 Treści programowe

Laboratorium		
Lp	Tematyka zajęć Opis szczegółowy bloków tematycznych	Liczba godzin
L1	Wprowadzenie do języka Perl, wyrażenia regularne.	2
L2	Zaawansowane typy danych i ich persystencja.	2
L3	Parsowanie dokumentów tekstowych i analiza leksykalna.	2
L4	Algorytmy pozyskiwania tekstów i budowy korpusu tekstowego.	2
L5	Analiza statystyczna korpusu tekstowego.	2
L6	Indekser dokumentów tekstowych, wyszukiwarka dokumentów.	2
L7	Automaty słownikowe.	2
L8	Własności stop-listy, algorytm budowy.	2
L9	Analiza morfologiczna tekstu, lematyzacja bezkontekstowa.	2
L10	Binaryzacja i wektoryzacja tekstu, wskaźniki dopasowania.	2
L11	Programowanie dynamiczne i wskaźniki podobieństwa.	2
L12	Badanie kontekstu informacji, konstrukcja ngramów.	2
L13	Reprezentacja wiedzy, generowanie tekstu, odpowiedzi.	2
L14	Projekt i implementacja systemu przetwarzania tekstu.	2

Laboratorium		
Lp	Tematyka zajęć Opis szczegółowy bloków tematycznych	Liczba godzin
L15	Testowanie systemu przetwarzania tekstu.	2

Wykład		
Lp	Tematyka zajęć Opis szczegółowy bloków tematycznych	Liczba godzin
W1	Wprowadzenie do tematyki, podstawowe wiadomości wstępne	2
W2	Narzędzia i języki programowania w przetwarzaniu dokumentów	2
W3	Podstawy gramatyki języka polskiego, formalizacja języka	2
W4	Korpusy tekstowe, Web-crawling, ekstrakcja informacji z tekstu	2
W5	Analiza statystyczna tekstu w języku polskim, słowa stop i kluczowe	2
W6	Indeksowanie dokumentów tekstowych, wyszukiwarki internetowe	2
W7	Analiza leksykalna i algorytmy wyszukiwania wzorca	2
W8	Słowniki, automaty słownikowe, słowniki wielojęzyczne	2
W9	Bezkontekstowa analiza morfologiczna tekstu	2
W10	Kontekst tekstu, technika n-gramów	2
W11	Algorytmy badania podobieństwa dokumentów tekstowych	2
W12	Reprezentacja i przedstawianie wiedzy, generowanie tekstu	2
W13	Podstawy translacji maszynowej tekstu	2
W14	Systemy i aplikacje do przetwarzania języka naturalnego	2
W15	Projektowanie aplikacji do przetwarzania i zarządzania wiedzą	2

7 Narzędzia dydaktyczne

N1 Wykłady

N2 Ćwiczenia laboratoryjne

N3 Praca w grupach

8 Obciążenie pracą studenta

Forma aktywności	Średnia liczba godzin na zrealizowanie aktywności
Godziny kontaktowe z nauczycielem akademickim, w tym:	
Godziny wynikające z planu studiów	60
Konsultacje przedmiotowe	0
Egzaminy i zaliczenia w sesji	0
Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	10
Opracowanie wyników	10
Przygotowanie raportu, projektu, prezentacji, dyskusji	10
Realizacja zadań programistycznych	70
Sumaryczna liczba godzin dla przedmiotu wynikająca z całego nakładu pracy studenta	**160**
Sumaryczna liczba punktów ECTS dla przedmiotu	5.00

9 Sposoby oceny

Ocena formująca

F1 Ćwiczenie praktyczne

F2 Sprawozdanie z ćwiczenia laboratoryjnego

F3 Odpowiedź ustna

Ocena podsumowująca

P1 Średnia ważona ocen formujących

Warunki zaliczenia przedmiotu

W1 Obecność na zajęciach.

W2 Aktywność na zajęciach.

Kryteria oceny

Efekt kształcenia 1	
Na ocenę 2.0	Opanowanie nie więcej niż 50% materiału.

Na ocenę 3.0	Opanowanie więcej niż 50% materiału.
Na ocenę 3.5	Opanowanie więcej niż 60% materiału.
Na ocenę 4.0	Opanowanie więcej niż 70% materiału.
Na ocenę 4.5	Opanowanie więcej niż 80% materiału.
Na ocenę 5.0	Opanowanie więcej niż 90% materiału.
Efekt kształcenia 2	
Na ocenę 2.0	Opanowanie nie więcej niż 50% materiału.
Na ocenę 3.0	Opanowanie więcej niż 50% materiału.
Na ocenę 3.5	Opanowanie więcej niż 60% materiału.
Na ocenę 4.0	Opanowanie więcej niż 70% materiału.
Na ocenę 4.5	Opanowanie więcej niż 80% materiału.
Na ocenę 5.0	Opanowanie więcej niż 90% materiału.
Efekt kształcenia 3	
Na ocenę 2.0	Opanowanie nie więcej niż 50% materiału.
Na ocenę 3.0	Opanowanie więcej niż 50% materiału.
Na ocenę 3.5	Opanowanie więcej niż 60% materiału.
Na ocenę 4.0	Opanowanie więcej niż 70% materiału.
Na ocenę 4.5	Opanowanie więcej niż 80% materiału.
Na ocenę 5.0	Opanowanie więcej niż 90% materiału.
Efekt kształcenia 4	
Na ocenę 2.0	Opanowanie nie więcej niż 50% materiału.
Na ocenę 3.0	Opanowanie więcej niż 50% materiału.
Na ocenę 3.5	Opanowanie więcej niż 60% materiału.
Na ocenę 4.0	Opanowanie więcej niż 70% materiału.
Na ocenę 4.5	Opanowanie więcej niż 80% materiału.
Na ocenę 5.0	Opanowanie więcej niż 90% materiału.

10 Macierz realizacji przedmiotu

Efekt kształcenia	Odniesienie danego efektu do szczegółowych efektów zdefiniowanych dla programu	Cele przedmiotu	Treści programowe	Narzędzia dydaktyczne	Sposoby oceny
EK1	I1_K04	Cel 3	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3
EK2	I1_W11	Cel 1	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1	P1
EK3	I1_U21	Cel 2	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3
EK4	I1_U23	Cel 3	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3

11 Wykaz literatury

Literatura podstawowa

[1] **Lubaszewski W.** — *Słowniki komputerowe I automatyczna ekstrakcja informacji z tekstu*, Kraków, 2009, AGH

[2] **Kłopotek M.A.** — *Inteligentne wyszukiwarki internetowe*, Warszawa, 2001, Akademicka Oficyna Wydawnicza Exit

[3] **Saloni Z., Świdziński M.** — *Składnia współczesnego języka polskiego*, Warszawa, 2011, PWN

[4] **Daciuk J.** — *Informacje: http://www.eti.pg.gda.pl/jandac/*, Internet, 2011, Politechnika Gdańska

[5] **Gusfield D.** — *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, Cambridge, 1997, Cambridge University Press

[6] **Jurafsky D., Martin J.H.** — *Speech and language processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, New Jersey, 2000, Prentice-Hall

12 Informacje o nauczycielach akademickich

Osoba odpowiedzialna za kartę

dr inż. Krzysztof Rzecki (kontakt: krz@pk.edu.pl)

Osoby prowadzące przedmiot

1 dr inż. Krzysztof Rzecki (kontakt: krz@iti.pk.edu.pl)

13 Zatwierdzenie karty przedmiotu do realizacji

(miejscowość, data) (odpowiedzialny za przedmiot) (dziekan)

Przyjmuję do realizacji (data i podpisy osób prowadzących przedmiot)

. . . . . .