



## KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Teoria informacji i metody kompresji danych

### Przedmiot

Kierunek studiów

Informatyka

Studia w zakresie (specjalność)

ITI

Poziom studiów

drugiego stopnia

Forma studiów

stacjonarne

Rok/semestr

1/2

Profil studiów

ogólnoakademicki

Język oferowanego przedmiotu

polski

Wymagalność

obieralny

### Liczba godzin

Wykład

30

Laboratoria

30

Inne (np. online)

Ćwiczenia

Projekty/seminaria

### Liczba punktów ECTS

5

### Wykładowcy

Odpowiedzialny za przedmiot/wykładowca:

Robert.Susmaga

email: Robert.Susmaga@cs.put.poznan.pl

tel: 61 6652934

Instytut Informatyki, Wydział Informatyki

i Telekomunikacji

adres: ul. Piotrowo 2, 60-965 Poznań

Odpowiedzialny za przedmiot/wykładowca:

### Wymagania wstępne

Podstawowa wiedza z zakresu: a) struktur danych (tablice jedno i dwuwymiarowe, listy, drzewa), b) rachunku prawdopodobieństwa i statystyki (prawdopodobieństwo, w tym prawdopodobieństwo warunkowe, zmienna losowa, rozkład zmiennej losowej, wartość średnia, wartość oczekiwana), c) analizy matematycznej (funkcja logarymiczna, wykładnicza, pochodne funkcji), d) algebry liniowej (wektory).

Umiejętność projektowania, implementowania i testowania programów komputerowych (w wybranym języku programowania) realizujących proste operacje na danych statycznych (wektory) oraz dynamicznych (listy, drzewa).

(Pożądane) ciekawość poznawcza i wytrwałość w dążeniu do poszerzania swojej wiedzy.



## Cel przedmiotu

Przedstawienie wybranych elementów teorii informacji, jednej z najbardziej podstawowych teorii leżących u podstaw współczesnej informatyki i zajmującej się reprezentowaniem, zapisywaniem i przesyłaniem informacji z użyciem symboli. Ponieważ ważne zastosowania tej teorii wykraczają poza ramy samej informatyki, prezentowany przedmiot prezentuje je jedynie w podstawowym zakresie, a skupia się na zastosowaniach teorii informacji w dziedzinie szeroko rozumianej kompresji danych. Owoce rozwoju tej dziedziny, stanowiące rozwiązania problemów, początkowo nieco akademickich i pozornie nierozwiązywalnych, okazały się niezwykle przydatne i praktyczne. Rozwiązania te są dziś spotykane w niemal każdym systemie komputerowym, nie wyłączając domowych i osobistych, które trudno sobie wyobrazić bez wszechobecnych treści multimedialnych, zawdzięczających wiele ze swej popularności dynamicznemu rozwojowi metod kompresji.

Szczegółowe cele przedmiotu obejmują przekazanie wiedzy niezbędnej do:

- a) rozumienia fundamentalnych pojęć leżących u podstaw teorii informacji i metod kompresji danych,
- b) identyfikowania, formułowania i rozwiązywania problemów związanych z teorią informacji i metodami kompresji, z wykorzystaniem leżących u ich podstaw elementów i wywodów teoretycznych,
- c) projektowania i tworzenia programów implementujących prezentowane metody.

## Przedmiotowe efekty uczenia się

Wiedza

Student/-ka

- ma zaawansowaną i pogłębioną wiedzę z zakresu szeroko rozumianych systemów informatycznych, podstaw teoretycznych ich budowania oraz metod, narzędzi i środowisk programistycznych wykorzystywanych do ich implementacji (K2st\_W1)
- ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z kluczowymi zagadnieniami z zakresu informatyki (K2st\_W2)
- ma zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu informatyki (K2st\_W3)
- zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych (K2st\_W6)

Umiejętności

Student/-ka

- potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł w języku polskim (ewentualnie: angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie (K2st\_U1)
- potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne, symulacyjne oraz eksperymentalne. (K2st\_U4)
- potrafi — przy formułowaniu i rozwiązywaniu zadań inżynierskich — integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych) oraz stosować podejście systemowe, uwzględniające także aspekty pozatechniczne (K2st\_U5)
- potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych w zakresie kompresji danych (K2st\_U6)
- potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych oraz zaproponować ich



ulepszenia (usprawnienia) (K2st\_U8)

-- 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 (K2\_stU9)

-- potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia (K2st\_U16)

Kompetencje społeczne

Student/-ka

-- rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe (K1st\_K1)

-- ma świadomość znaczenia wiedzy w rozwiązywaniu problemów inżynierskich oraz świadomość tego, że nieprawidłowe rozwiązania tych problemów prowadzą do powstawania wadliwie działające systemów informatycznych (K1st\_K2)

-- rozumie znaczenie działalności popularyzatorskiej dotyczącej najnowszych osiągnięć z zakresu informatyki (K2st\_K3)

### Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

(laboratoria):

-- na podstawie oceny bieżącego postępu realizacji przydzielanych zadań.

(wykłady):

-- ocena wiedzy i umiejętności wykazanych na pisemnym sprawdzianie zawierającym kilka (4-6) zadań (analogicznych do zadań prezentowanych na zajęciach); czas przewidziany na zaliczenie to 60--90 (wykłady); aby uzyskać ocenę pozytywną trzeba zdobyć przynajmniej  $1 + [m/2]$  (zaokrąglenie w dół) punktów, gdzie  $m$  jest punktacją maksymalną (np. aby uzyskać ocenę pozytywną przy  $m = 30$  należy zdobyć przynajmniej 16 punktów).

### Treści programowe

Program przedmiotu obejmuje następujące zagadnienia:

Podstawowe wiadomości z:

-- algebry liniowej (kombinacje liniowe, kombinacje wypukłe);

-- analizy matematycznej (właściwości i przebiegi funkcji  $\log(x)$  i  $x \cdot \log(x)$ );

-- rachunku prawdopodobieństwa (rozkłady zmiennej losowej dyskretnej, wartości oczekiwane, prawdopodobieństwo, prawdopodobieństwo warunkowe).

Informacja i miara informacji. Podstawowe właściwości miary informacji. Miara Hartley'a. Entropia Shannona: podstawowe właściwości (minima, maksima, wykresy przebiegu /rozkłady dwuelementowe i trzelementowe/). Entropia jako miara informacji. Zastosowania miary Hartley'a i miary Shannona w obliczaniu zawartości informacyjnej przekazów tekstowych, obrazowych i dźwiękowych.

Wielowymiarowe aspekty entropii (dane dyskretne): entropia łączna, entropia warunkowa, informacja wzajemna; zależność między entropią a informacją wzajemną; interpretacje w zastosowaniach związanych z szeroko rozumianą analizą danych.

Idea różnych rodzajów przekształceń danych: szyfrowanie, kompresowanie i skracanie. Kompresja (dane



dyskretne): idea i cele. Kompresowanie danych jako forma kodowania. Rodzaje kompresji (bezstratna i stratna). Techniki kompresji (kodowa i słownikowa).

Kodowanie (dane dyskretne). Idea i cele (w tym: niezawodność transmisji). Definicja kodu, przykłady kodów, drzewa kodowe, kody przedrostkowe, nierówność Krafta. Kodowanie jako forma kompresji danych.

Kompresja bezstratna (dane dyskretne). Metody kompresji kodowej. Kodowanie Shannona-Fano i kodowanie Huffmana: idea, kody optymalne, drzewa kodów, przykłady, algorytmy, właściwości.

Zastosowanie entropii w analizie kodowych metod kompresji.

Metody kompresji słownikowej Lempela-Ziva: idea, słowniki, przykłady, algorytmy (LZW), właściwości.

Dzień dzisiejszy systemów kompresji. Przyszłość kompresji.

(Opcjonalnie) Poszerzone zastosowania entropii w analizie danych (dywergencja Kullbacka-Leiblera /entropia względna/ i jej zastosowania); wersja dynamiczna algorytmu kodowania Huffmana; kodowanie arytmetyczne: idea, przedziały prawdopodobieństw, przykłady, algorytmy, właściwości; buforowe algorytmy słownikowe (LZ77, LZ78); transformata Burrowsa-Wheelera: idea, kodowanie.

Zajęcia laboratoryjne:

Wprowadzenie do języka Python i wybranych bibliotek tego języka: NumPy i Matplotlib.

Tworzenie prostych programów działających na danych skalarnych i wektorowych, generowanie i wizualizowanie rozkładów prawdopodobieństwa, generowanie probabilistycznych przybliżeń języka.

Zastosowania wybranych wielowymiarowych aspektów entropii w analizie danych (selekcja atrybutów, przeszukiwanie dokumentów).

Zastosowania wybranych metod kodowych w kompresji danych: metoda Shannona-Fano i metoda Huffmana.

Zastosowania wybranych metod słownikowych w kompresji danych: metoda LZW.

(Opcjonalnie) Adaptacyjna metoda Huffmana. Kodowanie arytmetyczne. Metoda LZ77/LZ78.

## Metody dydaktyczne

Wykłady: prezentacja multimedialna uzupełniona przykładami obliczeniowymi, demonstracja wybranych aspektów obliczeniowych.

Laboratoria: modelowanie przykładowych problemów dotyczących przetwarzania i wizualizacji danych wielowymiarowych i rozwiązywanie tych problemów, wykonywanie eksperymentów symulacyjnych, dyskusja, praca w zespole, pokaz multimedialny i demonstracja.

## Literatura

### Podstawowa

1. A. Drozdek: "Wprowadzenie do kompresji danych", WNT, Warszawa, 1999.
2. A. Przelaskowski: "Kompresja danych. Podstawy, metody bezstratne, kodery obrazów", BTC, Legionowo, 2005.

### Uzupełniająca

1. Materiały wykładowe.
2. T.M. Cover, J.A. Thomas, "Elements of Information Theory", 2nd Edition, Wiley & Sons, Hoboken, New Jersey, 1991.



3. D.J.C. MacKay: "Information Theory, Inference, and Learning Algorithms", Cambridge University Press, Cambridge, UK, 2003.
4. K. Sayood (red.): "Lossless Compression Handbook", Academic Press, Elsevier Science, San Diego, California, 2003.
5. K. Sayood: "Introduction to Data Compression", 3rd Ed., Morgan Kaufmann Publishers, San Francisco, California, 2006.

### Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,0
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	60	2,0
udział w wykładach <sup>1</sup>	30	1,2
udział w zajęciach laboratoryjnych	30	1,2
teoretyczne przygotowanie do ćwiczeń laboratoryjnych (literatura przedmiotu, języki i systemy programowania)	20	0,8
praktyczna praca nad zebraniem danych, tworzeniem i testowaniem oprogramowania	20	0,8
przeprowadzenie badań / eksperymentów obliczeniowych	10	0,4
tworzenie sprawozdań	5	0,2
przygotowanie do sprawdzianu (wykład)	10	0,4

<sup>1</sup> niepotrzebne skreślić lub dopisać inne czynności