



## KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Metody kompresji danych [S1S1E>MKD]

### Przedmiot

Kierunek studiów

Sztuczna inteligencja/Artificial Intelligence

Rok/Semestr

2/3

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

pierwszego stopnia

Język oferowanego przedmiotu

angielski

Forma studiów

stacjonarne

Wymagalność

obieralny

### Liczba godzin

Wykład

15

Laboratorium

0

Inne

0

Ćwiczenia

15

Projekty/seminaria

0

### Liczba punktów ECTS

3,00

### Koordynatorzy

dr hab. inż. Robert Susmaga

robert.susmaga@put.poznan.pl

### Wykładowcy

### Wymagania wstępne

Podstawowa wiedza dotycząca: a) rachunek różniczkowy (funkcja logarytmiczna, funkcja wykładnicza, pochodna funkcji), b) algebra liniowa (wektory, macierze, operacje na wektorach/macierzach), c) teoria prawdopodobieństwa i statystyka (prawdopodobieństwo, w tym prawdopodobieństwo warunkowe, zmienne losowe, rozkłady prawdopodobieństwa i funkcje rozkładu prawdopodobieństwa, wartości średnie, wartości oczekiwane, wariancja), d) struktury danych (tablice jedno- i dwuwymiarowe, listy, drzewa). Podstawowe umiejętności w zakresie projektowania, tworzenia i testowania programów komputerowych (w języku programowania do wyboru), które realizują proste przetwarzanie statyczne (wektory i macierze) oraz dynamiczne (listy, drzewa) struktury danych. (zalecane) Spora dawka ciekawości poznawczej i nie mniejsza wytrwałość w dążeniu do celów rozwój osobisty.

## Cel przedmiotu

Celem kursu jest przedstawienie aspektów Kompresji Danych w kontekście Informatyki Teoria, jedna z najbardziej podstawowych teorii leżących u podstaw teoretycznej informatyki współczesnej. Teoria informacji zajmuje się reprezentowaniem, przechowywaniem i przekazywaniem informacji wyrażonej w formie symboli. Z uwagi na fakt, że wiele ważnych zastosowań tej teorii sięga daleko poza nią stanowiącym rdzeń informatyki, przedstawiony wybór aspektów ograniczy się do większości fundamentalnych, głównie związanych z takimi dziedzinami informatyki jak eksploracja danych oraz Kompresja Danych, w szczególności: Bezstratna Kompresja Danych. Owoce szybkiego rozwoju Pojęcia Teorii Informacji, początkowo dyskusyjne i rzekomo nierozwiązywalne, wkrótce okazały się być niezwykle przydatne i mieć dużą wartość praktyczną. We współczesnych rozwiązaniach pojawiają się praktycznie we wszystkich możliwych do wyobrażenia istniejących systemach komputerowych, które z trudem przetrwają obecnie bez wszechobecnych treści multimedialnych, których popularność i wszechstronność była dotychczasowa stale i niezmiennie pod wpływem dostępności metod kompresji danych. Szczegółowe cele kursu obejmują dzielenie się umiejętnościami i wiedzą wystarczającą do: a) zrozumienie podstawowych idei leżących u podstaw teorii informacji, zwłaszcza tych z nią związanych do bezstratnej kompresji danych, b) identyfikowanie, formułowanie i rozwiązywanie podstawowych problemów Teorii Informacji i Danych Bezstratnych Kompresja, c) projektowanie i tworzenie programów komputerowych skutecznie realizujących przedstawione metody i algorytmy.

## Przedmiotowe efekty uczenia się

Wiedza:

Uczniowie:

1. posiadać podstawową, uporządkowaną i ugruntowaną wiedzę niezbędną w ważnych obszarach informatyki nauki ścisłe takie jak algorytmika i języki programowania -- [K1st\_W2]
2. znać i rozumieć podstawowe techniki, metody, algorytmy i narzędzia stosowane do rozwiązywania problemów komputerowych problemy związane z teorią informacji i jej zastosowaniami w kompresji danych - [K1st\_W4]
3. znać i rozumieć podstawowe techniki, metody, algorytmy i narzędzia stosowane do rozwiązywania problemów komputerowych problemów i problemów związanych ze sztuczną inteligencją, w tym zautomatyzowanego rozpoznawania wzorców w dane różnych typów i ich przetwarzanie -- [K1st\_W5]
4. posiadać podstawową wiedzę na temat kluczowych kierunków i najważniejszych sukcesów sztucznej inteligencji rozumiana jako istotna subdziedzina informatyki, korzystająca z osiągnięć innych dyscypliny naukowe, takie jak kompresja danych, w celu zapewnienia rozwiązań o dużym wpływie praktycznym -- [K1st\_W5]

Umiejętności:

Uczniowie:

1. potrafi zbierać informacje z odpowiednich źródeł o różnym charakterze, dokonywać ich krytycznej analizy, interpretacji i syntezy oraz kompleksowo wyciągać i uzasadniać sformułowane wnioski dotyczące informacji, szczególnie w kontekście teorii informacji i jej zastosowań w danych Kompresja -- [K1st\_U1]
2. Potrafi sprawnie planować i przeprowadzać eksperymenty, w tym pomiary komputerowe i symulacje związanych z różnymi aspektami kompresji danych, interpretować uzyskane wyniki i wyciągać wnioski na podstawie wyników eksperymentów -- [K1st\_U4]
3. Potrafi wyszukiwać, analizować i przekształcać różnego rodzaju dane oraz dokonywać syntezy danych do wiedzy i wnioski przydatne przy rozwiązywaniu różnorodnych problemów pojawiających się w informatyce, zwłaszcza w AI -- [K1st\_U10]
4. potrafi – kierując się z góry określoną specyfikacją – zaprojektować i stworzyć system informatyczny dokonując najpierw wyboru i następnie wykorzystując dostępne metody, techniki i narzędzia komputerowe (w tym języki programowania) -- [K1st\_U8]

5. Potrafi adaptować istniejące algorytmy oraz formułować i wdrażać nowe, w tym algorytmy typowe dla różnych strumieni AI, wykorzystujące przynajmniej jedno znane narzędzie -- [K1st\_U9]

Kompetencje społeczne:

Uczniowie:

1. rozumie, że wiedza i umiejętności szybko się dezaktualizują i dostrzega potrzebę ich ciągłego doskonalenia

dodatkowe szkolenia i podnoszenie kwalifikacji -- [K1st\_K1]

2 mają świadomość znaczenia wiedzy naukowej i badań związanych z informatyką i sztuczną inteligencją w rozwiązywaniu problemów istotnych dla funkcjonowania jednostek, firm, organizacji

całe społeczeństwo -- [K1st\_K2]

## Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Przedstawione powyżej efekty uczenia się weryfikowane są w następujący sposób:

Ocena kształtująca (zajęcia laboratoryjne): ocena rozwiązań zadanych programowania problemu (w miarę ich pojawiania się).

Ostateczna ocena:

-- (zajęcia laboratoryjne): ocena rozwiązań zadanych problemów programistycznych (końcowa),

-- (wykłady): ocena wyników testu pisemnego (45-60 min) zarówno wielokrotnego wyboru, krótkiego odpowiedzi i (małe) pytania obliczeniowe (głównie: mikroproblemy do rozwiązania w formie pisemnej).

## Treści programowe

Kurs obejmuje między innymi następujące elementy.

Podstawy:

-- rachunek różniczkowy (podstawowe właściwości i wykresy  $\exp(x)$ ,  $\log(x)$  i  $x \cdot \log(x)$ );

-- algebra liniowa (kombinacje liniowe/wypukłe wektorów, kadłuby wypukłe);

-- teoria prawdopodobieństwa (dyskretne zmienne losowe): prawdopodobieństwo, funkcje rozkładu prawdopodobieństwa.

Idea informacji Shannon (w skrócie: informacja). Miara treści informacyjnej:

konstrukcja i właściwości. Miara Hartleya treści informacyjnej. Pomysł Shannon

entropia (w skrócie: entropia). Entropia jako miara zawartości informacji: budowa i właściwości,

w szczególności: nieruchomość podziałowa.

Przesyłanie danych i potrzeba procedur kodowania/dekodowania.

Trzy rodzaje przetwarzania danych: kodowanie/dekodowanie, szyfrowanie/desyfrowanie i mieszanie (ze szczególnym uwzględnieniem

pierwszy).

Kodowanie/dekodowanie (dane dyskretne): idea i cele (w tym: niezawodność przesyłania danych). Kody:

definicje i przykłady, drzewiaste reprezentacje kodów, kody przedrostkowe. Nierówność Krafta. Kodowanie danych

mający na celu kompresję.

Kompresja (dane dyskretne i ciągłe, z naciskiem na te pierwsze): pomysł i cele (m.in.:

prędkość przesyłu danych). Dwa podstawowe rodzaje kompresji danych (bezstratna i stratna, ze szczególnym uwzględnieniem pierwszego).

Miary kompresji danych (dane dyskretne). Entropia jako granica kompresji. Uniwersalny metoda kompresji.

Kompresja bezstratna (dane dyskretne): kompresja oparta na kodowaniu. Shannon, Shannon-Fano i Huffman

kodowanie: idea, drzewa kodowe, kody optymalne, algorytmy, przykłady, właściwości, podejścia adaptacyjne.

Kompresja bezstratna (dane dyskretne): kompresja słownikowa. Metoda Lempela-Ziv-Welcha:

idea, reprezentacje, słowniki, algorytmy, przykłady, właściwości, podejścia adaptacyjne.

Różne metody kompresji i techniki wspierające kompresję: Kodowanie długości przebiegu

(pomysł, algorytmy); Kodowanie Move-to-Front (idea, algorytmy), transformacja Burrowsa-Wheelera (pomysł, algorytmy).

Systemy i platformy kompresji.

Materiał opcjonalny (ze względu na czas i postęp):

-- (kodowanie) adaptacyjne kodowanie Huffmana (idea, algorytmy),

--(słownikowe) metody Lempla-Ziv-77 i Lempel-Ziv-78 (idea, algorytmy),  
-- (inne) kodowanie arytmetyczne (idea, kodowanie ciągów symboli, przedziały prawdopodobieństwa, algorytmy).

## Tematyka zajęć

Kurs obejmuje między innymi następujące elementy.

Podstawy:

-- rachunek różniczkowy (podstawowe właściwości i wykresy  $\exp(x)$ ,  $\log(x)$  i  $x \cdot \log(x)$ );  
-- algebra liniowa (kombinacje liniowe/wypukłe wektorów, kadłuby wypukłe);  
-- teoria prawdopodobieństwa (dyskretne zmienne losowe): prawdopodobieństwo, funkcje rozkładu prawdopodobieństwa.

Idea informacji Shannon (w skrócie: informacja). Miara treści informacyjnej:

konstrukcja i właściwości. Miara Hartleya treści informacyjnej. Pomysł Shannon

entropia (w skrócie: entropia). Entropia jako miara zawartości informacji: budowa i właściwości, w szczególności: nieruchomość podziałowa.

Przesyłanie danych i potrzeba procedur kodowania/dekodowania.

Trzy rodzaje przetwarzania danych: kodowanie/dekodowanie, szyfrowanie/desyfrowanie i mieszanie (ze szczególnym uwzględnieniem pierwszego).

Kodowanie/dekodowanie (dane dyskretne): idea i cele (w tym: niezawodność przesyłania danych). Kody: definicje i przykłady, drzewiaste reprezentacje kodów, kody przedrostkowe. Nierówność Krafta. Kodowanie danych

mający na celu kompresję.

Kompresja (dane dyskretne i ciągłe, z naciskiem na te pierwsze): pomysł i cele (m.in.: prędkość przesyłu danych). Dwa podstawowe rodzaje kompresji danych (bezstratna i stratna, ze szczególnym uwzględnieniem pierwszego).

Miary kompresji danych (dane dyskretne). Entropia jako granica kompresji. Uniwersalny metoda kompresji.

Kompresja bezstratna (dane dyskretne): kompresja oparta na kodowaniu. Shannon, Shannon-Fano i Huffman

kodowanie: idea, drzewa kodowe, kody optymalne, algorytmy, przykłady, właściwości, podejścia adaptacyjne.

Kompresja bezstratna (dane dyskretne): kompresja słownikowa. Metoda Lempela-Ziv-Welcha: idea, reprezentacje, słowniki, algorytmy, przykłady, właściwości, podejścia adaptacyjne.

Różne metody kompresji i techniki wspierające kompresję: Kodowanie długości przebiegu (pomysł, algorytmy); Kodowanie Move-to-Front (idea, algorytmy), transformacja Burrowsa-Wheelera (pomysł, algorytmy).

Systemy i platformy kompresji.

Materiał opcjonalny (ze względu na czas i postęp):

-- (kodowanie) adaptacyjne kodowanie Huffmana (idea, algorytmy),

--(słownikowe) metody Lempla-Ziv-77 i Lempel-Ziv-78 (idea, algorytmy),

-- (inne) kodowanie arytmetyczne (idea, kodowanie ciągów symboli, przedziały prawdopodobieństwa, algorytmy).

## Metody dydaktyczne

Wykłady: prezentacje w postaci pokazów slajdów (elementy teoretyczne, objaśnienia, przykłady, ćwiczenia).

Zajęcia laboratoryjne: projektowanie i tworzenie (w wybranym przez siebie języku programowania) programów obsługujących

rozwiązywać zadane problemy (ilustrujące idee i pojęcia prezentowane na wykładach).

## Literatura

Literatura podstawowa

1. D.J.C. MacKay: Information Theory, Inference, and Learning Algorithms, Cambridge University Press, Cambridge, UK, 2003.
2. T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd Edition, Wiley and Sons, Hoboken, New Jersey, 1991.
3. K. Sayood (red.): Lossless Compression Handbook, Academic Press, Elsevier Science, San Diego,

California, 2003.

Literatura uzupełniająca

1. Lecture notes (slide show presentations)

2. K. Sayood: Introduction to Data Compression, 3rd Ed., Morgan Kaufmann Publishers, San Francisco, California, 2006.

3. A. Drozdek: Wprowadzenie do kompresji danych WNT, Warszawa, 1999

4. A. Przelaskowski: Kompresja danych. Podstawy, metody bezstratne, kodery obrazów, BTC, Legionowo, 2005

### Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	75	3,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	1,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwίων/egzaminu, wykonanie projektu)	45	1,50