KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu
Podstawy analizay informacji [S1Cybez1>PAI]

Przedmiot

Kierunek studiów
Cyberbezpieczeństwo

Rok/Semestr
3/6

Studia w zakresie (specjalność)
–

Profil studiów
ogólnoakademicki

Poziom studiów
pierwszego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
stacjonarne

Wymagalność
obieralny

Liczba godzin

Wykład
16

Laboratorium
16

Inne
0

Ćwiczenia
0

Projekty/seminaria
16

Liczba punktów ECTS

3,00

Koordynatorzy

dr hab. inż. Mariusz Żal
mariusz.zal@put.poznan.pl

dr hab. inż. Sławomir Hanczewski
slawomir.hanczewski@put.poznan.pl

Wykładowcy

Wymagania wstępne

• Podstawowa znajomość języka Python 2 • Znajomość podstaw matematyki stosowanej, w tym statystyki i algebry liniowej. • Umiejętność korzystania z narzędzi do analizy danych (np. Pandas, NumPy, scikit-learn). • Podstawowe zrozumienie procesów przetwarzania języka naturalnego (NLP).

Cel przedmiotu

Celem przedmiotu jest wprowadzenie studentów w zagadnienia analizy informacji z naciskiem na przetwarzanie języka naturalnego (NLP) w kontekście cyberbezpieczeństwa. Kurs koncentruje się na przygotowaniu i przetwarzaniu danych tekstowych, zrozumieniu mechanizmów stojących za popularnymi modelami i narzędziami NLP oraz wykorzystaniu ich potencjału w analizie zagrożeń, wykrywaniu ataków socjotechnicznych i innych aspektach związanych z bezpieczeństwem systemów i sieci teleinformatycznych.

Przedmiotowe efekty uczenia się

Wiedza:
• Student zna rolę analizy informacji w wykrywaniu ataków, moniotorowaniu zagrożeń i ocenie ryzyka [K1_W05]
• Rozumie kluczowe koncepcje (tokenizacja, stemming, lematyzacja, embeddingi) i ich znaczenie w przetwarzaniu tekstu. [K1_W16]
• Zna najbardziej rozpowszechnione biblioteki oraz wie, do jakich zastosowań się nadają. [K1_W09]
• Rozumie, czym są modele takie jak BERT, GPT, word2vec i jakie mają możliwości oraz ograniczenia.[K1_W06]
• Zna techniki wstępnego przetwarzania danych [K1_W05]

Umiejętności:
• Potrafi oczyścić i sformatować dane tekstowe. [K1_U02]
• Potrafi wybrać właściwe narzędzie do realizacji konkretnych zadań (klasyfikacja tekstu, analiza sentymentu, rozpoznawanie nazwanych encji). [K1_U02]
• Umie zaimplementować modele językowe (np. word2vec, BERT) i zinterpretować uzyskane wyniki. [K1_U04]
• Umie przetwarzać i analizować treści dotyczące ataków, a także wyszukiwać wzorce [K1_U04]

Kompetencje społeczne:
• Jest świadomy regulacji dotyczących danych osobowych, np. RODO, oraz implikacji etycznych wynikających z analizy tekstów. [K1_K05]
• Potrafi współpracować w grupie. [K1_K05]
• Wykazuje gotowość do ciągłego uczenia się i otwartość na opinie innych osób.[K1_K01][K1_K02]

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:
Wiedza zdobyta w ramach wykładu weryfikowana jest przez egzamin w formie pisemnej lub ustnej. W formie pisemnej studenci muszą udzielić odpowiedzi na 7 - 10 pytań (testowych i otwartych) różnie punktowanych. Są trzy lub cztery grupy punktowe. Natomiast w przypadku egzaminu ustnego student losuje po jednym pytaniu z każdej grupy punktowej. W formie ustnej, do każdego wylosowanego pytania, student może otrzymać dodatkowe pytanie (związane z wylosowanym pytaniem). Ocena pytania (obejmuje odpowiedź zarówno na pytanie wylosowane jak i pytanie dodatkowe) obejmuje zakres odpowiedzi oraz głębię zrozumienia zagadnienia. Do każdego egzaminu przygotowywanych jest 50 - 60 pytań. Warunkiem pozytywnego zaliczenia egzaminu jest otrzymanie minimum 50% punktów możliwych do zdobycia.
Umiejętności nabyte w ramach projektów oceniane są na podstawie prezentowanych projektów. Ocenie poddawane jest zaangażownie w przygotwanie projektu, wykorzystane narzędzia oraz zakres dodatkowej wiedzy, jaką studneci musieli posiąść. Projekty są jedno lub dwuosobowe. Skala ocen 2,0 - 5,0.
Umiejętności nabyte w ramach zajęć laboratoryjnych weryfikowane są na bieżąco. Na każdych zajęciach laboratoryjnych oceniana jest poprawność wykonania ćwiczeń w skali od 0 do 10 punktów. Warunkiem pozytywnego zaliczenia ćwiczeń laboratoryjnych jest otrzymanie minimum 50% punktów możliwych do zdobycia.
liczba punktów ocena
<=50% 2,0
51% - 60% 3,0
61% - 70% 3,5
71% - 80% 4,0
81% - 90% 4,5
91% - 100% 5,0

Treści programowe

Program obejmuje praktyczne przygotowanie danych tekstowych do analiz, począwszy od ich oczyszczania i standaryzacji, aż po wybrane techniki lematyzacji i reprezentacji (np. word embeddingi). Uczestnicy zapoznają się z popularnymi bibliotekami NLP, dzięki czemu nauczą się implementować rozwiązania do klasyfikacji tekstu, rozpoznawania nazwanych encji czy wykrywania phishingu. Ważnym elementem jest także kontekst cyberbezpieczeństwa - studenci poznają scenariusze wykorzystania NLP

do analizy logów, wiadomości e-mail oraz wpisów w mediach społecznościowych pod kątem zagrożeń. Całość dopełniają aspekty etyczne (prywatność, anonimizacja danych) oraz umiejętność prezentacji wyników w formie przyjaznych raportów i wizualizacji.

Tematyka zajęć

1. Wprowadzenie do analizy informacji, ze szczególnym uwzględnieniem zastosowań w cyberbezpieczeństwie
2. Przygotowanie danych tekstowych
3. Podstawowe narzędzia i biblioteki NLP
4. Zastosowanie popularnych modeli NLP
5. Analiza tekstu a bezpieczeństwo IT
6. Wizualizacja i raportowanie
6. Najnowsze trendy i kierunki badawcze

Metody dydaktyczne

• Wykłady problemowe, połączone z analizą studiów przypadków.
• Ćwiczenia laboratoryjne, polegające na implementacji i testach narzędzi NLP, pracy z rzeczywistymi (lub bliskimi realnym) danymi.
• Projekty zespołowe, podczas których studenci opracowują kompleksowe rozwiązania do detekcji zagrożeń.

Literatura

Podstawowa:
1. Jurafsky, D., Martin, J. H. Speech and Language Processing, Pearson, 2020.
3. Goldberg, Y. Neural Network Methods for Natural Language Processing, Morgan & Claypool Publishers, 2017.
Uzupełniająca
1. Clark, K., Luong, M.-T., Le, Q. V., Manning, C. D. ELECTRA: Pre-training Text Encoders as Discriminators
Rather Than Generators (2020).
2. Blogi, dokumentacja i tutoriale oraz dokumentacja narzędzi i bibliotek.
3. Raporty o cyberzagrożeniach (np. CERT Polska, ENISA, Cisco Talos) - dla obserwacji rzeczywistych zastosowań NLP w bezpieczeństwie.

Uzupełniająca:
Uzupełniająca
1. Clark, K., Luong, M.-T., Le, Q. V., Manning, C. D. ELECTRA: Pre-training Text Encoders as Discriminators
Rather Than Generators (2020).
2. Blogi, dokumentacja i tutoriale oraz dokumentacja narzędzi i bibliotek.
3. Raporty o cyberzagrożeniach (np. CERT Polska, ENISA, Cisco Talos) - dla obserwacji rzeczywistych zastosowań NLP w bezpieczeństwie.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	88	3,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	48	1,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwiów/egzaminu, wykonanie projektu)	40	1,50