

Popis předmětu

Zkratka předmětu:	KIV/IR-E	Strana:	1 / 3
Název předmětu:	Vyhledávání informací		
Akademický rok:	2023/2024	Tisknuto:	27.05.2024 19:05

Pracoviště / Zkratka	KIV / IR-E			Akademický rok	2023/2024
Název	Vyhledávání informací			Způsob zakončení	Zkouška
Akreditováno/Kredity	Ano, 6 Kred.			Forma zakončení	Kombinovaná
Rozsah hodin	Přednáška 3 [HOD/TYD] Cvičení 2 [HOD/TYD]			Zápočet před zkouškou	ANO
Obs/max	Statut A	Statut B	Statut C	Počítán do průměru	ANO
Letní semestr	0 / -	0 / -	6 / -	Min. (B+C) studentů	1
Zimní semestr	0 / -	0 / -	0 / -	Opakovaný zápis	NE
Rozvrh	Ano			Vyučovaný semestr	Letní semestr
Vyučovací jazyk	Angličtina			Počet dnů praxe	0
Volně zapisovatelný předmět	Ano			Hodn. stup. zp. před zk.	S N
Hodnotící stupnice	1 2 3 4				
Počet hodin kontaktní	Ne				
Automat. uzn. záp. před zk.					
Periodicita	K				
Nahrazovaný předmět	Žádný				
Vyloučené předměty	KIV/IR				
Podmiňující předměty	Nejsou definovány				
Předměty informativně doporučené	KIV/IDT nebo KIV/PT				
Předměty, které předmět podmiňuje	KIV/NLP				

Cíle předmětu (anotace):

Dát studentům důkladné znalosti jak vytvářet komplexní softwary na zpracování textů přirozeného jazyka.

Požadavky na studenta

Vypracování a obhájení semestrálního softwarového projektu (získat minimálně 50% z možných bodů), získat 50% z bodů za aktivní účast na cvičeních, absolvovat písemný kontrolní test (získat 50% bodů), písemná zkouška s ústním dozkoušením (minimum je opět 50% možných bodů). Konečný termín pro získání zápočtu je konec května.

Z důvodu průběžné aktualizace předmětu je pro získání zápočtu při opakovaném zapsání předmětu (viz SZŘ čl. 24 odst. 3) nutné souhlasné vyjádření garanta předmětu.

Upozornění:

Termíny a forma ověřování splnění požadavků mohou být upraveny s ohledem na opatření vyhlášená v souvislosti s vývojem epidemiologické situace v ČR.

Obsah

1. Taxonomie úloh strojového zpracování přirozeného jazyka. Typické problémy a aplikace.
2. Tokenizace, stemming, Porterův algoritmus, lematizace, POS značkování, parsing. Slovníky, editační vzdálenost.
3. Vyhledávání informace, booleovský model, indexování.
4. Podobnost dotazu s dokumentem, vektorový model dokumentu, výběr top hitů.
5. Hodnocení vyhledávacího systému, standardní kolekce.
6. XML vyhledávání, vektorový model v XML prostředí, vyhodnocování relevantnosti.
7. Pravděpodobnostní modely pro IR. Maticové dekompozice, latentní sémantické indexování.
8. Klasifikace textů, výběr vlastností, vyhodnocování klasifikace, klasifikace ve vektorovém prostoru. Detekce plagiátů, spamů.
9. Shlukování textů, volba počtu shluků. Systémy shlukování zpravidajských textů.
10. Extrakce informací, extrakce událostí, extrakce relací.
11. Sumarizační metody, generování textu.

12. Vyhledávání názorů. Aplikace na texty sociálních médií.

13. Získávání informací z Webu, analýza obsahu, prohledávače Webu, distribuované indexy, Web jako graf, analýza struktury Webu, PageRank, HITS.

Studijní opory

Studentům jsou k dispozici ke stažení jednotlivé přednášky ve formátu PDF a také podstatná část přednášek v ucelené podobě jako elektronická kniha (PDF).

Garanti a vyučující

- **Garanti:** Doc. Ing. Dalibor Fiala, Ph.D. (100%)
- **Přednášející:** Doc. Ing. Dalibor Fiala, Ph.D. (100%)
- **Cvičící:** Ing. Ondřej Pražák (100%)

Literatura

- **Základní:** Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hinrich. *Introduction to information retrieval*. 1st pub. New York : Cambridge University Press, 2008. ISBN 978-0-521-86571-5.
- **Doporučená:** Baeza-Yates, R.; Ribeiro-Neto, Berthier. *Modern information retrieval*. Harlow : Addison-Wesley, 1999. ISBN 0-201-39829-X.
- **Doporučená:** Jurafsky, Daniel; Martin, James H. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. 2nd ed. Upper Saddle River : Pearson/Prentice Hall, 2009. ISBN 978-0-13-504196-3.

Časová náročnost

Všechny formy studia

Aktivita	Časová náročnost aktivity [h]
Projekt individuální [40]	40
Příprava na zkoušku [10-60]	55
Kontaktní výuka	65
Celkem:	160

Hodnotící metody

Odborné znalosti - odborné znalosti dosažené studiem předmětu jsou ověřovány hodnotícími metodami:

Individuální prezentace,
Průběžné hodnocení,
Test,
Kombinovaná zkouška,

Odborné dovednosti - odborné dovednosti dosažené studiem předmětu jsou ověřovány hodnotícími metodami:

Výstupní projekt,
Demonstrace dovedností (praktická činnost),
Kombinovaná zkouška,

Obecné způsobilosti - obecné způsobilosti dosažené studiem předmětu jsou ověřovány hodnotícími metodami:

Kombinovaná zkouška,
Individuální prezentace,

Předpoklady**Odborné znalosti - pro úspěšné zvládnutí předmětu se předpokládá, že je student před zahájením výuky schopen:**

orientovat se v možnostech aplikačního programového vybavení s cílem dosáhnout lepší orientaci v narůstajícím množství informací
 popsat principy programování v imperativních a objektových jazycích včetně základních řídicích struktur a způsobů reprezentace dat, vysvětlit základní datové struktury a algoritmy pro práci s nimi
 vysvětlit principy relačních databází, datové integrity a základních SQL příkazů, popsat postupy datového modelování

Odborné dovednosti - pro úspěšné zvládnutí předmětu se předpokládá, že student před zahájením výuky dokáže:

navrhnout databázový systém či informační systém menšího až středního rozsahu, navrhnout a realizovat jednodušší samostatnou a webovou aplikaci
 ovládat zásady vytváření dobře dokumentovaných a robustních programových kódů, prakticky využít teoretické a praktické poznatky o práci s algoritmy, datovými strukturami a konkrétními vývojovými prostředky
 utřídit, zpracovat a prezentovat získané informace písemnou i ústní formou v anglickém jazyce; vytvořit dokumentaci k realizovanému dílu nebo jeho součásti
 získávat a zpracovávat informace ze zdrojů v anglickém jazyce

Obecné způsobilosti - před zahájením studia předmětu je student schopen:

mgr. studium: používají své odborné znalosti, odborné dovednosti a obecné způsobilosti alespoň v jednom cizím jazyce,

Vyučovací metody**Odborné znalosti - pro dosažení odborných znalostí jsou užívány vyučovací metody:**

Cvičení (praktické činnosti),
 Přednáška s diskusí,
 Řešení problémů,
 Samostatná práce studentů,
 Samostudium,
 Výuka podporovaná multimédií,

Odborné dovednosti - pro dosažení odborných dovedností jsou užívány vyučovací metody:

Demonstrace dovedností,

Obecné způsobilosti - pro dosažení obecných způsobilostí jsou užívány vyučovací metody:

Přednáška s diskusí,

Výsledky učení**Odborné znalosti - po absolvování předmětu prokazuje student znalosti:**

popsat principy zpracování přirozeného jazyka a vyhledávání v textových datech
 vysvětlit a ilustrovat metody a modely pro reprezentaci a zpracování rozsáhlých nestrukturovaných dat

Odborné dovednosti - po absolvování předmětu prokazuje student dovednosti:

efektivně použít metody a technologie pro vyhledávání v rozsáhlých nestrukturovaných datech
 realizovat různé metody webového vyhledávání a základní metody zpracování přirozeného jazyka

Obecné způsobilosti - po absolvování předmětu je student schopen:

mgr. studium: dle vyvíjejících se souvislostí a dostupných zdrojů vymezí zadání pro odborné činnosti, koordinují je a nesou konečnou odpovědnost za jejich výsledky,
 Po absolvování je student nejen schopen realizovat různé metody zpracování přirozeného jazyka, ale také získá profesionální znalosti o možnostech jejich využití v oblasti softwarového inženýrství, Business Intelligence, Social Media monitoring, odhalování závadných a společensky nebezpečných textů a názorů, analýzy názorů apod. Získá i schopnost používat formální metody pro konstruování takového softwaru.

Předmět je zařazen do studijních programů: