

☑ Магистърска програма: Извличане на информация и откриване на знания

Срок на обучение: 3 семестъра

Форма на обучение: редовна

Минимален брой ECTS кредити: 90

Ръководител: доц. д-р Иван Койчев

ФМИ, бул. Дж. Баучър 5, София 1164

тел.: 02 971 04 00

e-mail: koychev@fmi.uni-sofia.bg

В програмата са привлечени водещи учени от различни области на извличането на информация (технологии за разработване на машини за търсене), методите за обработка на естествен език и анализа на големи масиви от данни. Магистърската програма се реализира в сътрудничество с европейския проект по програмата COST: Action: IC1002: Multilingual and Multifaceted Interactive Information Access.

Програмата цели подготовка на специалисти в следните модерни и перспективни направления: извличане на информация (Information Retrieval or Search Engine Technologies) и откриване на знания (Knowledge Discovery in Data, a.k.a. Data Mining, incl. Business Intelligence and Web Mining). Това са бързо развиващи се области през последното десетилетие и прогнозата е тази тенденция да се запази и през следващото.

В момента представените в електронен вид неструктурирани данни (текст) значително надхвърлят по обем структурираните данни. Нараства и необходимостта от автоматичен анализ на тези големи корпуси от текст за търсене и откриване на полезна информация в тях.

Паралелно с това събирането на структурирани данни чрез различни сензори и системи също е във възход. Натрупани са огромни масиви от данни, които са потенциален източник на полезна информация и знания. Но за да бъдат открити тези полезни знания в морето от данни, е необходимо да се използват съвременни алгоритми за интелигентен анализ на данни.

Тенденцията е двете направления за откриване на полезна информация в текст и в данни взаимно да се допълват, защото за много от съвременните приложения данните са и от двата типа. Използването на всички възможни източници води до постигане на значително по-добри резултати.

В момента това са едни от най-бързо растящите пазари в ИТ сектора, което предопределя и голямото търсенето на специалисти, които задълбочено да познават алгоритмите и методите в областта. Прогнозата е тази тенденция да се запази и през следващото десетилетие.

Програмата е ориентирана към кандидати, завършили бакалавърски програми в областите на информатиката, компютърните науки и математиката.

Магистърската програма си поставя за цел да запознае студентите с основните методи и алгоритми на извличането на информация и откриване на знания, които добавят и степен на интелигентност в разработваните системи. Студентите ще се запознаят с основните подходи за автоматичен анализ на данни и обработка на естествен език. Ще развият умения за решаване на конкретни задачи, като комбинират успешно необходимите последователности от стъпки и преценяват ролята на необходимите ресурси и технологии за успешна реализация.

Обучението по дисциплините в програмата се осъществява чрез комбиниране/допълване на лекции + семинари + дискусии + индивидуални задачи. Основният материал се поднася на студентите по време на лекции, като се използва също и формата на дискусия. Студентите ще получават и задачи за самостоятелна работа.

Семинарните занятия включват решаване на допълнителни задачи по темата, разглеждане на варианти на задачата за самостоятелна работа с цел да се илюстрират специфични алгоритми и методи.

Необходимите ресурси за обучение са следните: компютър и мултимедия за презентация на

слайдове.

Програмата ще даде на студентите:

- знания за основните алгоритми и методите зад уебтърсачките, сайтовете за е-търговия, препоръчващите системи, филтрирането на имейл спам, корекцията на правописа, машинния превод, автоматичната класификация и групиране и др. Освен това студентите ще имат възможността да се запознаят с методи за проектиране и разработване на софтуерни системи и услуги;

- умения, придобити от опита при разработване курсови проекти на реални приложения, като например: обработка на текст; създаване на машини за търсене; интелигентни инструменти за подобряване на точността на търсене; системи за филтриране на спам, категоризация и групиране на документи; подходи за автоматичен анализ на данни и текст с цел откриване в тях на неявно представени зависимости и образци.

- отворени врати към възможности за работа в компании за търсене на информация и анализ на големи масиви от данни.

Завършилите обучението по учебна програма ще придобият следните основни компетенции:

- умения за прилагане на алгоритми и методи, необходими за изграждане на системи за търсене и извличане на информация от колекции от текстови документи и уеб;

- умения за прилагане на подходи за обработка на естествен език и реализиране на системи, подпомагащи потребителя в анализа и обработката на текст;

- умения за използване на методи, необходими за анализ на големи масиви от данни и текст с цел откриване на неизвестни знания в тях, включително и при разработване на „горещи“ приложения като автоматичен анализ на мнения по дадени теми и бизнес разузнаване;

- умения за използване на софтуерните технологии за проектиране и разработване на разпределени софтуерни системи и услуги.

Магистрите ще могат да намерят реализация като:

- разработчици на системи за анализ на текстове и извличане на информация;

- разработчици на платформи за обработка на големи масиви от данни;

- експерти, използващи инструменти за анализ на данни за биз-неса, държавни институции, издателства, е-търговия и др.

- изследователи и преподаватели в академичните институции;

Завършилите тази магистърска програма също ще бъдат доста-тъчно добре квалифицирани, за да продължат своето образование в докторски програми.

УЧЕБЕН ПЛАН

Дисциплина	ECTS-кредити		Хорариум седмичен
	семестриален	седмичен	
I семестър			
1. Извличане на информация	7	75	3+0+2
2. Подходи за обработка на естествен език	6	60	3+0+1
3. Бази от знания	5	60	3+0+1
4. Обектно-ориентиран анализ и проектиране на софтуерни системи	5	60	2+2+0
5. Размити множества и приложения	5	60	3+0+1
6. Изкуствен интелект	7	75	3+0+2
7. Професионална етика	4	60	2+0+2
8. Технологично предприемачество	5	60	2+0+2
9. Архитектури на софтуерни системи	5	45	2+1+0
II семестър			
10. Откриване на знания в данни	7	75	3+0+2
11. Откриване на знания в текст	6	60	3+0+1
12. Софтуерни системи, базирани на услуги	5	60	2+0+2
13. Семантичен Уеб	5	60	2+0+2
14. Интернет технологии и уеб програмиране	5	60	2+1+1
15. Изграждане на приложения в облачна среда	6	60	2+2+0
16. Уеб технологии и цифрови библиотеки	5	60	2+0+2

17. Управление на проекти	6	60	2+2+0
18. Методология на изследователската работа	4	60	2+2+0

III семестър

19. Преддипломен курсов проект	15	150	
20. Стаж	15	150	
21. Разработване и защита на дипломна работа	15	150	

Курсовете изписани с тъмен шрифт са задължителни