

УДК 004

Составители: Е.А. Кулешова, А.В. Мандрика

Рецензент

Кандидат технических наук, доцент Ю.А. Халин

Информационно-поисковые системы: методические указания по выполнению самостоятельной работы / Юго-Зап. гос. ун-т; сост.: Е.А. Кулешова, А.В. Мандрика – Курск, 2019. – 16 с.: табл. 5. – Библиогр.: с. 16.

Содержатся сведения о темах для самостоятельного изучения по дисциплине «Информационно-поисковые системы», необходимые для успешного освоения дисциплины. Указывается порядок выполнения самостоятельной работы всех предусмотренных учебным планом видов, приводятся рекомендации по оформлению результатов работы.

Методические указания соответствуют требованиям программы, утвержденной учебно-методическим объединением по специальности.

Предназначены для студентов направления подготовки бакалавриата 02.03.03 Математическое обеспечение и администрирование информационных систем.

Текст печатается в авторской редакции

Подписано в печать 16.12.2019 . Формат 60x84 1/16.
Усл.печ. л. 1,34. Уч.-изд. л. 1,21. Тираж 100 экз. Заказ 991. Бесплатно.

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Содержание

1 Введение.....	4
2 Учебно-методическое обеспечение для самостоятельной работы.....	6
3 Запланированные виды самостоятельной работы по дисциплине.....	7
4 Оценивание знаний, умений, навыков.....	13
5 Рекомендации по выполнению самостоятельной работы.....	14
6 Контрольные вопросы для самоконтроля.....	16
7 Библиографический список.....	16

1 Введение

Самостоятельная работа - это индивидуальная или коллективная учебная деятельность, осуществляемая без непосредственного руководства преподавателя, но по его заданиям и под его контролем.

Самостоятельная работа студентов включает:

- изучение лекционного материала по конспекту с использованием рекомендованной литературы;
- отработку изучаемого материала по печатным и электронным источникам, конспектам лекций;
- подготовку к выполнению лабораторных или практических работ;
- выполнение отчетов по лабораторным или практическим работам и подготовку к их защите;
- индивидуальные задания (решение задач, подготовка сообщений, докладов, исследовательские работы и т.п.);
- работу над творческими заданиями;
- подготовку кратких сообщений, докладов, рефератов, самостоятельное составление задач по изучаемой теме (по указанию преподавателя).

Назначение самостоятельной работы студентов.

- **Овладение знаниями**, что достигается:

чтением текста (учебника, первоисточника, дополнительной литературы), составлением плана текста, графическим структурированием текста, конспектированием текста, выписками из текста, работой со словарями и справочниками, поиском информации в сети Интернет и т.п.;

- **закрепление знаний**, что достигается:

работой с конспектом лекций, обработкой текста, повторной работой над учебным материалом (учебником, первоисточником, дополнительной литературой), составлением плана, составлением таблиц для систематизации учебного материала, ответами на контрольные вопросы, заполнением рабочей тетради,

аналитической обработкой текста (аннотирование, рецензирование, реферирование, конспект-анализ и др), составлением библиографии и т.п.;

- **формирование навыков и умений**, что достигается:

решением задач и упражнений по образцу, решением вариативных задач, выполнением схем, выполнением расчетов, решением ситуационных задач, подготовкой к дискуссиям, проектированием и моделированием разных видов и компонентов профессиональной деятельности, математическим описанием опытно экспериментальной работой и т.п.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов.

Контроль результатов внеаудиторной самостоятельной работы студентов может осуществляться в пределах времени, отведенного на обязательные учебные занятия по дисциплине и внеаудиторную самостоятельную работу студентов по дисциплине, может проходить в письменной, устной или смешанной форме.

Текущий контроль качества выполнения самостоятельной работы может осуществляться с помощью:

- контрольного опроса;
- собеседования;
- автоматизированного программированного контроля (машинного контроля, тестирования с применением ЭВМ).

Контроль выполнения индивидуальных заданий осуществляется поэтапно в соответствии с разработанным преподавателем графиком.

2 Учебно-методическое обеспечение для самостоятельной работы

Студенты могут при самостоятельном изучении отдельных тем и вопросов дисциплины пользоваться учебно-методическими пособиями и методическими разработками кафедры в рабочее время, установленное Правилами внутреннего распорядка работников.

Учебно-методическое обеспечение для самостоятельной работы обучающихся по данной дисциплине организуется:

библиотекой университета:

- библиотечный фонд укомплектован учебной, методической, периодической, справочной литературой в соответствии с Учебным планом (УП) и Рабочей программой данной дисциплины (РПД);

- имеется доступ к основным информационным образовательным ресурсам, информационной базе данных, в том числе библиографической, возможность выхода в Интернет;

кафедрой:

- путем обеспечения доступности всего необходимого учебно-методического и справочного материала;

- путем предоставления сведений о наличии учебно-методической литературы, современных программных средств;

- путем разработки:

- заданий для самостоятельной работы;

- вопросов к экзамену;

- тестов;

- методических указаний к выполнению практических работ и

т.д.;

типографией университета:

- помощь авторам в подготовке и издании научной, учебной и методической литературы;

- удовлетворение потребности в тиражировании научной, учебной и методической литературы.

3 Запланированные виды самостоятельной работы по дисциплине

В соответствии с учебным планом, на самостоятельную работу студентов в рамках дисциплины «Информационно-поисковые системы» отводится 73,85 часов, включая подготовку к экзамену. Распределение часов самостоятельной работы по темам (видам деятельности) приведено в рабочей программе дисциплины (Таблица 1).

В таблице ниже приведены соответствующие сведения, взятые из Рабочей программы дисциплины.

Таблица 1 – Самостоятельная работа студентов в соответствии с рабочей программой дисциплины

№	Наименование раздела (темы) дисциплины	Срок выполнения	Время, затрачиваемое на выполнение СРС, час.
1.	Проблема информационного поиска.	2 неделя	6
2.	Информационный поиск	3 неделя	6
3.	Выборка данных	4 неделя	6
4.	Анализ задачи информационного поиска.	5 неделя	6
5.	Становление ИПС Web: краткий исторический обзор	6 неделя	6
6.	Математические модели документального поиска.	7 неделя	6
7.	Алгоритмы документального поиска	8 неделя	6
8.	Классификация и кластеризация документов	9 неделя	6
9.	Информационный поиск в Веб.	10 -11 неделя	6
10.	Анализ гиперссылок при информационном поиске в Веб	12-13 неделя	6
11.	Архитектура инфомационнопоисковой системы Google	14-15 неделя	4
12.	Контекстно- сфокусированный поиск	16-17 неделя	4
13.	Самоорганизация в сети Веб	18 неделя	5,85

№	Наименование раздела (темы) дисциплины	Срок выполнения	Время, затрачиваемое на выполнение СРС, час.
Итого:			73.85

Текущий контроль знаний, основанный на выяснении качества самостоятельной работы студентов при работе с конспектом лекций и учебной основной и дополнительной литературой, производится в соответствии с Рабочей программой дисциплины (Таблица 2) и предусматривает контрольный опрос (КО) и собеседование (С).

В таблице ниже приведены соответствующие сведения, взятые из Рабочей программы дисциплины.

Таблица 2 – Формы текущего контроля в соответствии с рабочей программой дисциплины

№ п/п	Раздел (тема) дисциплины	Виды деятельности			Учебно-методические материалы	Формы текущего контроля успеваемости и (по неделям семестра)	Компетенции
		лек., час	№ лаб.	№ пр.			
1	Проблема информационного поиска.	1	1		У-1, У-2, У-3, У-5 МУ-1, МУ-2	УО, ЛР (1 неделя)	ПК-3 ПК-6
2	Информационный поиск vs. выборка данных	1	1		У-1, У-3, У-6, МУ-1, МУ-2	УО, ЛР (2 неделя)	ПК-3 ПК-6
3	Анализ задачи информационного поиска.	1	2		У-1, У-2, У-4, МУ-1, МУ-2	УО, ЛР (3 неделя)	ПК-3 ПК-6
4	Становление ИПС Web: краткий исторический обзор	1	2		У-3, У-4, У-5, МУ-1, МУ-2	УО, ЛР (4 неделя)	ПК-3 ПК-6
5	Математические модели документального поиска.	1	3		У-1, У-2, У-6,	УО, ЛР (5-6 недели)	ПК-3 ПК-6

№ п/п	Раздел (тема) дисциплины	Виды деятельности			Учебно-методические материалы	Формы текущего контроля успеваемости и (по неделям семестра)	Компетенции
		лек., час	№ лаб.	№ пр.			
					МУ-1, МУ-2		
6	Алгоритмы документального поиска	1	4		У-3, У-2, У-4, У-6, МУ-1, МУ-2	УО, ЛР (7 неделя)	ПК-3 ПК-6
7	Классификация и кластеризация документов	1	5		У-1, У-2, У-5, МУ-1, МУ-2	УО, ЛР (8-9 недели)	ПК-3 ПК-6
8	Информационный поиск в Веб.	1	6		У-3, У-4, У-5, МУ-1, МУ-2	УО, ЛР, Р (10 неделя)	ПК-3 ПК-6
9	Анализ гиперссылок при информационном поиске в Веб	1	7		У-3, У-4, У-6, МУ-1, МУ-2	УО, ЛР (11-12 недели)	ПК-3 ПК-6
10	Архитектура инфомационно-поисковой системы Google	1	8		У-1, У-2, У-4, МУ-1, МУ-2	УО, ЛР (13-14 недели)	ПК-3 ПК-6
11	Контекстно- сфокусированный поиск	2	9		У-1, У-2, У-5, МУ-1, МУ-2	УО, ЛР (15-16 недели)	ПК-3 ПК-6
12	Самоорганизация в сети Веб	2	10		У-1, У-3, У-4, У-6 МУ-1, МУ-2	УО, ЛР (17-18 недели)	ПК-3 ПК-6

Лекционные занятия проводятся в соответствии с Рабочей программой дисциплины (Таблица 3) и включают следующие темы.

Таблица 3 – Краткое содержание лекционного курса

№ п/п	Раздел (тема) дисциплины	Содержание
1.	Проблема	Проект “How Much Information”. Объемы производимой

	информационного поиска.	информации. Суммарное годовое потребление информации жителями США. Проблема экспоненциального разрыва между производством и потреблением. Подходы к решению проблемы. Жизненный цикл информационного ресурса. Средства навигации для текстовой информации: этапы становления. Библиотечные поисковые системы. Информационный поиск: этапы становления. Исторические исследовательские документально-поисковые системы.
2.	Информационный поиск vs. выборка данных	Классификация. Определение информационного поиска. Структура процесса информационного поиска. Задача тематического поиска. Особенности Web, затрудняющие классический информационный поиск
3.	Анализ задачи информационного поиска.	Информационный поиск vs. выборка данных. Определение информационного поиска (Information Retrieval). Структура процесса информационного поиска. Особенности Web, затрудняющие классический информационный поиск.
4.	Становление ИПС Web: краткий исторический обзор	Становление ИПС Web: краткий исторический обзор. Становление ИПС Web в России. Обобщенное описание модели документального поиска. Типы математических моделей документального поиска. Теоретико-множественная модель. Метрики подобия. Простая булевская модель. Расширенная булевская модель. Достоинства и недостатки булевских моделей.
5.	Математические модели документального поиска.	Векторная модель. Закон Ципфа (Zipf). Принцип Луна (Luhn). Расчет весов терминов. Достоинства и недостатки векторной модели. Правило принятия решения. Оценка вероятности на основе обратной связи по релевантности (Robertson&Jones). Достоинства и недостатки вероятностных моделей.
6.	Алгоритмы документального поиска	Полнотекстовое сканирование. Файлы сигнатур. Сигнатуры слов. Сигнатуры документов. Ложные совпадения. Инверсия. Использование семантической информации. Обработка естественного языка (NLP). Латентносемантическое индексирование (LSI).
7.	Классификация и кластеризация документов	Классификация, автоматическая классификация, кластеризация документов. Использование автоматической классификации при информационном поиске. Кластеризация. Методы кластеризации. Метод К-средних. Иерархическая кластеризация. Поиск кластера. Кластеризация в распределенных системах. Подходы к кластеризации в распределенных системах. Выполнение запроса в распределенной системе. Классификация документов на основе гиперссылок. Модификация запросов. Метод Rocchio.
8.	Информационный поиск в Веб.	Особенности Web, затрудняющие классический информационный поиск. Характеристики статического (поверхностного) Веб. Размер базы поисковых систем (2005). Глубинный Веб. Действия пользователя. Таксономия поиска в Веб. Эволюция поисковых систем Веб. Веб-

		каталоги vs. поисковые системы. Усредненный “портрет” пользователя ИПС Веб.
9.	Анализ гиперссылок при информационном поиске в Веб	Базовые допущения при анализе гиперссылок. Алгоритмы анализа гиперссылок. Алгоритм PageRank. Расчет коэффициентов PageRank. Алгоритм HITS. Авторитеты и концентраторы. Процедура расчета весов авторитетности и концентрации. Алгоритм HITS: проблемы. Алгоритм HITS: расширения. Различие между PageRank и HITS.
10.	Архитектура информационнопоисковой системы Google	Общая структура ИПС Google (1998 г.). Основные структуры данных. Хранилище. Документальный индекс. Словарь. Списки вхождений (hits). Прямой индекс. Обратный индекс. Архитектура Google сегодня.
11.	Контекстно-сфокусированный поиск	Регулярность в распределении гиперссылок. Степенной закон распределения гиперссылок в вебграфе. Модель предпочтительного прикрепления. Структура “бабочка”. Сфокусированный обход вебграфа. Радиус-1 гипотеза. Радиус-2 гипотеза. Контекстный граф. Сфокусированный сетевой робот. Контекстно-сфокусированный сетевой робот.
12.	Самоорганизация в сети Веб	Веб-сообщества. Модель веб-сообщества. Идентификация веб-сообществ на основе алгоритма поиска максимального потока минимальной стоимости.

Основная литература для освоения дисциплины включает источники [1-2], дополнительная литература включает источники [3-32].

Лабораторные работы, предусмотренные Рабочей программой дисциплины, описаны в таблице 4.

Таблица 4.2.1 – Лабораторные работы

№	Наименование лабораторной работы	Объем, час.
1	Русскоязычные вербальные ИПС: поиск	4
2	Глобальные вербальные ИПС: сравнительный анализ	6
3	Глобальные вербальные ИПС: изучение	6
4	Глобальные вербальные ИПС: поиск	6
5	ИПС AltaVista и Google	6
Итого:		28

Рекомендации по выполнению практических работ приведены в соответствующих методических указаниях. Методические указания содержат полные требования к видам и объему самостоятельной работы при подготовке, выполнении, оформлении отчетов и защите практических работ.

4 Оценивание знаний, умений, навыков

Процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций, регулируются следующими нормативными актами университета:

- положение П 02.016–2018 Обально-рейтинговой системе оценивания результатов обучения по дисциплинам (модулям) и практикам при освоении обучающимися образовательных программ;
- методические указания, используемые в образовательном процессе, указанные в списке литературы.

Для *текущего контроля успеваемости* по дисциплине в рамках действующей в университете балльно-рейтинговой системы применяется следующий порядок начисления баллов:

Для промежуточной аттестации обучающихся, проводимой в виде тестирования, используется следующая методика оценивания знаний, умений, навыков и (или) опыта деятельности. В каждом варианте КИМ –16 заданий (15 вопросов и одна задача).

Каждый верный ответ оценивается следующим образом:

- задание в закрытой форме –2балла,
- задание в открытой форме – 2 балла,
- задание на установление правильной последовательности – 2 балла,
- задание на установление соответствия – 2 балла,
- решение компетентностно-ориентированной задачи – 6 баллов.

Максимальное количество баллов за тестирование –36 баллов.

5 Рекомендации по выполнению самостоятельной работы

Изучение теоретических основ дисциплин

Изучение теоретической части дисциплин способствует углублению и закреплению знаний, полученных на аудиторных

занятиях, а также развивает у студентов творческие навыки, инициативу и умение рационально организовать свое время.

Самостоятельная работа при изучении дисциплины включает:

- работу над конспектом лекций;
- изучение рекомендованной литературы;
- поиск и ознакомление с информацией в сети Интернет;
- подготовку к различным формам контроля (контрольный опрос, собеседование, тесты);
- подготовку ответов на вопросы по различным темам дисциплины, в том числе заданным преподавателем по результатам контроля знаний.

Материал, законспектированный в течение лекций, необходимо регулярно прорабатывать и дополнять сведениями из других источников и литературы.

При освоении дисциплины сначала необходимо по каждой теме изучить рекомендованную литературу и составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в этой теме для освоения последующих тем курса. После этого следует разобраться с обоснованием утверждений. Для расширения знания по дисциплине рекомендуется использовать Интернет-ресурсы; проводить поиски в различных системах и использовать материалы сайтов, рекомендованных преподавателем.

По требованию преподавателя конспект лекций предоставляется ему для проверки. Замеченные недостатки и внесенные замечания и предложения следует отработать в приемлемые сроки.

6 Контрольные вопросы для самоконтроля

Задание в закрытой форме:

1. Лингвистическая переменная может принимать значения:
 - а) либо слова, либо числа
 - б) слова
 - в) числа
2. Гипертекстовые системы предназначены для ...?

- а) Реализации контекстной помощи
- б) Обеспечения голосового ввода команд в системах управления
- в) Реализации поиска по ключевым словам в базах текстовой информации
- г) Доступа к интеллектуальным базам данных

Задание в открытой форме:

Какой знак нужно использовать в запросе, чтобы поиск прошел без учета морфологии?

Задание на установление правильной последовательности:

Установите последовательность этапов развития информационной технологии:

- "электрическая" технология
- "механическая" технология
- "электронная" технология
- "компьютерная" технология
- "ручная" технология

Задание на установление соответствия:

Укажите соответствие для всех 5 вариантов ответа:

- 1) информационно-поисковая система
- 2) управляющая информационная система
- 3) интеллектуальная информационная система

- Информационная библиотечная система
- Медицинские информационные системы
- Компьютеризированная продажа железнодорожных билетов
- Система бухгалтерского учета
- Система оперативного планирования выпуска продукции

Компетентностно-ориентированная задача:

Составить и сохранить поисковый запрос для Google и Yandex используя поисковые операторы и команды логического исключения и объединения - найти резюме жителей Москвы со знанием французского языка, имеющих опыт работы с Windows, Linux. При этом уровень знания французского языка имеет первостепенное значение.

Библиографический список

Основная учебная литература

1. Информационные ресурсы и технологии в экономике [Текст] : учебное пособие для студентов высших учебных заведений / под ред.: Б. Е. Одинцова и А. Н. Романова. - Москва : Вузовский учебник : Инфра-М, 2019. - 462 с.
2. Информационные системы и технологии управления [Электронный ресурс] : учебное пособие / ред. Г. А. Титоренко. - 3-е изд., перераб. и доп. - Москва : Юнити-Дана, 2015. - 591 с.
3. Информационные технологии в науке и образовании [Текст] : учебное пособие / Е. Л. Федотова, А. А. Федотов. - Москва : ФОРУМ : ИНФРА-М, 2015. - 334 с.

Дополнительная учебная литература

4. Шмидт, Эрик. Как работает Google [Текст] : пер. с англ. / Э. Шмидт, Д. Розенберг, при участии А. Игла. - Москва : Эксмо, 2015. - 320 с.
5. Уткин, В. Б. Информационные системы и технологии в экономике [Электронный ресурс] : учебник / В. Б. Уткин, К. В. Балдин. - Москва : Юнити-Дана, 2015. - 336 с.
6. Щипицина, Лариса Юрьевна. Информационные технологии в лингвистике [Текст] : учебное пособие / Л. Ю. Щипицина. - Москва : Флинта : Наука, 2015. - 128 с.

8.3 Перечень методических указаний

1. Информационно-поисковые системы [Электронный ресурс] : методические рекомендации по выполнению лабораторных работ для студентов направления подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем» / Юго-Зап. гос. ун-т ; сост.: С. Ю. Сазонов, Е. А. Кулешова. - Курск : ЮЗГУ, 2019. - 26 с.