

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Факультет компьютерных систем и информационных технологий
Кафедра информатики и программной инженерии

УТВЕРЖДАЮ

Декан факультета компьютерных
систем и информационных технологий

Кочевский А. А.

« 19 » 04 2023 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по учебной дисциплине

«Обработка естественного языка»

09.03.03 Прикладная информатика

«Прикладная информатика в экономике»

Разработчик:

доцент _____ Петрущенко Т. В.

ФОС рассмотрен и одобрен на заседании кафедры информатики и
программной инженерии от 18 апреля 2023 г., протокол № 17

Заведующий кафедрой _____ Кочевский А.А.

Луганск 2023 г.

**Паспорт
фонда оценочных средств по учебной дисциплине
«Обработка естественного языка»**

**Перечень компетенций (элементов компетенций),
формируемых в результате освоения учебной дисциплины**

№ п/п	Код контролируемой компетенции	Формулировка контролируемой компетенции	Контролируемые темы учебной дисциплины	Этапы формирования (семестр изучения)
1	ПК-2	Способен осуществлять проектирование программного обеспечения ИС и разрабатывать техническую документацию на его компоненты	Тема 1. История и концепция управления проектами. Тема 2. Разработка проекта. Тема 3. Подсистемы управления проектами.	8

**Показатели и критерии оценивания компетенций,
описание шкал оценивания**

№ п/п	Код контролируемой компетенции	Показатель оценивания (знания, умения, навыки)	Контролируемые темы учебной дисциплины	Наименование оценочного средства
1	ПК-2	Знать: как использовать современные языки и системы программирования, технологии проектирования программного обеспечения. Уметь: сформулировать требования к разрабатываемому программному обеспечению, выполнить его реализацию и оформить техническую документацию на его компоненты. Владеть: навыками осуществления проектирования программного обеспечения конкретной ИС и разработку технической	Тема 1. Введение в обработку естественных языковых текстов Тема 2. Методы обработки естественных языков Тема 3. Вопросно-ответные системы Тема 4. Программирование и проектирование систем обработки естественных языков	Лабораторные работы; контрольные работы; промежуточная аттестация (зачет)

		документации на ее компоненты.		
--	--	--------------------------------	--	--

Фонды оценочных средств по дисциплине «Обработка естественного языка»

Лабораторные работы:

Типовые задания к лабораторным работам

Лабораторная работа 1. Построение транскриптора слов русского языка *Задание.*

Из связного текста размером не менее 200 слов сформировать словарь. Упорядочить полученный словарь по алфавиту в порядке возрастания. В словах, состоящих более, чем из одного слога, обозначить ударение символом '/'. Сформировать базу продукционных правил, используя правила транскрипции русского языка. Разработать программное обеспечение, которое на базе созданных правил строит транскрипцию слов русского языка

Лабораторная работа 2. Обработка естественно-языковых текстов на морфологическом уровне. *Задание.*

Для предметной области, заданной текстом из Л. р. № 1 создать грамматический словарь, на основе которого разработать программу морфологического синтеза слов.

Лабораторная работа 3. Морфологический анализ слов русского языка. *Задание.*

Для словаря и правил словоизменения, полученных в Л. р. № 2, разработать программу морфологического анализа слов русского языка.

Лабораторная работа 4. Элементы синтаксического анализа. *Задание.*

Используя библиотеку, разработать программу, определяющую наличие синтаксической связи между парами слов.

Критерии и шкала оценивания по оценочному средству лабораторная работа

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Лабораторная работа выполнена на высоком уровне (правильность выполнения 90-100%)
4	Лабораторная работа выполнена на среднем уровне (правильность выполнения 75-89%)
3	Лабораторная работа выполнена на низком уровне (правильность выполнения 50-74%)
2	Лабораторная работа выполнена на неудовлетворительном уровне (правильность выполнения менее чем на 50%)

Контрольные работы:

Типовые вопросы к контрольным работам

1. Лингвистика как наука о языке.
2. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика.
3. Лингвистика и прагматика.
4. Лингвистическое моделирование.
5. Действующие модели языка.
6. Теория «Смысл – Текст» как фундамент для построения систем автоматической обработки текста.
7. Анализ и синтез текста.
8. Морфологический и синтаксический анализ.
9. Парсинг.
10. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх».
11. Языковая неоднозначность как принципиальное свойство языка и методы ее
12. разрешения при автоматической обработке текста.
13. Интерактивное разрешение лексической и синтаксической неоднозначности.
14. Правильные и статистические подходы к автоматической обработке текста.
15. Алгоритм синтаксического анализа.
16. Синтаксические отношения.
17. Синтагмы.
18. Синтаксическая структура предложения.
19. Вопросно-ответные системы: основы вопросно-ответной системы, архитектуры вопросно-ответной системы, установление смысла вопроса и порождение ответов.
20. Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация.
21. Понятия лексемы, словоформы, леммы, морфемы, псевдо-основы и псевдоокончания.
22. Грамматические категории.
23. Словоизменительная парадигма.
24. Морфотактика.
25. Структура данных морфологического словаря, лексикона.
26. Грамматические модели русского языка в контексте автоматической обработки. Минимальное расстояние редактирования.
27. Алгоритм подсчета расстояния Левенштейна.
28. Практика по подсчету минимального расстояния Левенштейна.
29. Понятие статистической языковой модели.

Критерии и шкала оценивания по оценочному средству «контрольная работа»

Шкала оценивания	Критерий оценивания
5	Контрольная работа выполнена на высоком уровне (правильные ответы даны на 90-100% вопросов/задач)
4	Контрольная работа выполнена на среднем уровне (правильные ответы даны на 75-89% вопросов/задач)
3	Контрольная работа выполнена на низком уровне (правильные ответы даны на 50-74% вопросов/задач)
2	Контрольная работа выполнена на неудовлетворительном уровне (правильные ответы даны менее чем на 50%)

Оценочные средства для промежуточной аттестации (зачет)

Типовые вопросы к зачету

1. Снятие неоднозначности морфологического разбора на основе скрытых марковских моделей.
2. Методы обработки неизвестных слов в морфологическом анализе.
3. Использование сглаживания Гуда-Тьюринга в языковом моделировании.
4. Основные метрики оценки качества информационного поиска.
5. Источники данных для задачи автоматического реферирования.
6. Методы повышения производительности формирования веб-сниппетов.
7. Подходы к повышению качества статистических парсеров на основе составляющих.
8. Оценки вычислительной сложности вероятностного синтаксического анализа на основе составляющих и деревьев зависимостей.
9. Формирование словарей тонально окрашенных слов и фраз с помощью запросов к машинам поиска интернета.
10. Группы признаков для выделения именованных сущностей из текста.
11. Подходы к извлечению отношений между именованными сущностями из больших корпусов текстов.
12. Методы оценки вопросно-ответного поиска.
13. Сингулярное разложение матриц в задаче векторного представления семантики слов.
14. Недостатки метрики BLEU.
15. Задачи обработки естественного языка, этапы обработки текста.
16. Наивный байесовый классификатор и другие традиционные методы классификации.
17. Классификация с помощью нейронных сетей прямого распространения.
18. Лингвистические приложения методов классификации.
19. Текст как последовательность символов или слов.

20. N-грамм модели, модели основанные на рекуррентных нейронных сетях.
 21. Сглаживание, оценка модели.
 22. Классификация элементов последовательности символов и слов.
 23. Контекстно-свободные грамматики в описании естественных языков, дерево разбора текста.
 24. Контекстно-свободные грамматики в описании естественных языков, граф зависимостей.
 25. Разные подходы к определению понятия смысла языковых единиц: исчисление предикатов.
 26. Разные подходы к определению понятия смысла языковых единиц: дистрибутивная семантика.
 27. Кластеризация.
 28. Нейросетевые подходы к векторному представлению слов (word embedding).
 29. Разрешение кореферентности.
 30. Понятие трансформера, энкодера, декодера, механизм внимания, современные нейросетевые модели.
 31. Машинный перевод, генерация текста.
- Критерии и шкала оценивания по оценочному средству промежуточный контроль (зачет)

Шкала оценивания	Критерий оценивания
зачтено	Студент глубоко и в полном объеме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
	Студент знает только основной программный материал, допускает неточности, недостаточно четкие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
не зачтено	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)

Экспертное заключение

Представленный фонд оценочных средств (далее - ФОС) *по дисциплине «Обработка естественного языка»* соответствует требованиям ГОС ВО.

Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной образовательной программы по направлению подготовки 09.03.03 Прикладная информатика.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы обучающегося представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки бакалавров, по указанному направлению.

Председатель учебно-методической
комиссии факультета компьютерных
систем и информационных
технологий



Ветрова Н. Н.