**ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ**

**ФИЗИКО-ТЕХНИЧЕСКИЙ ФАКУЛЬТЕТ**

**КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Инструкции и методические указания   
к выполнению лабораторных работ**

**Современные проблемы информатики**

|  |  |
| --- | --- |
| **Направление подготовки:** | *09.04.01 - Информатика и вычислительная техника*. |
| **Программа подготовки:** | *Интеллектуальные информационные системы* |
| **Квалификация (степень) выпускника:** | *Магистр* |
| **Форма обучения:** | *Дневная* |

Донецк, 2020

лабораторныЕ РАБОТЫ

1. **Знакомство с интеллектуальными агентами ИПС в Internet**

***Задание***

Выявить в сети Internet динамику изменения количества ***всех*** публикаций, касающихся ***предмета*** исследований вашей магистерской диссертации за последние годы. Построить в отчёте MS Office Excel к лабораторной работе соответствующую таблицу и график, которые в дальнейшем вы сможете использовать в магистерской диссертации для обоснования её актуальности.

***Указания и рекомендации***

1. Для выполнения работы вам необходимо выбрать в Internet какие-либо три ИПС. Для них найти и кратко описать основные правила[[1]](#footnote-1) словарного поиска (AND, OR, +, –, ~, «…», рассто­яние между словами…). Для выбранных ИПС составьте требуемые выражения-запросы по [булевой модели](http://www.tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/Модели%20информационного%20поиска.ppt). Запросы составляйте по образцу слайда «[Примеры профессиональных запросов к ИПС](http://www.tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/Характеристики%20ИПС.ppt)». То есть, выберите 2-3 ключевых понятия по вашему ***предмету[[2]](#footnote-2)*** исследований и все возможные синонимы, фразы для каждого из понятий. Запишите всё это в виде соответствующего булевского выражения для выбранных ИПС. Корректность выражений согласуйте с преподавателем. Результаты ответов ИПС должны быть релевантными, не содержать ссылок на посторонние документы.
2. Для получения количества публикаций (количества документов, найденных ИПС) за каждый год и построения необходимой таблицы и графика, предусмотрите отсев документов, в которые лишь упоминаются публикации других лет и которые не являются публикациями заданного года. Например, при поиске публикаций за 2019 год необходимо в запросе убрать их упоминания во всех последующих годах(2020, 2021…).
3. Не забудьте, в дополнение к полученным таблицам и графикам, включить в отчёт тему магистерской диссертации, предмет исследований и выражение-запрос для каждой ИПС.

При сдаче лабораторной работы надо знать ответы на контрольные вопросы.

***Контрольные вопросы***

* 1. Как работает булева модель поиска?
  2. Как работает векторная модель поиска?
  3. Что такое релевантность и как она может вычисляться?
  4. Что такое SEO и что такое поисковый ранг?
  5. Как повысить поисковый ранг сайта?

1. **Интеллектуальный RSS-агент. Семантика и базы знаний**

***Задание***

Разработать Internet-агент (мультиагентную систему) в виде робота по сбору, доставке и показу в браузере или Win-приложении новостей из RSS-лент по ключевым словам, заданным пользователем.

***Указания и рекомендации***

Такая система может выглядеть следующим образом. Пользователь вводит фразу, для которой он хочет получить новости. Агент выбирает значимые слова из фразы запроса (например, исключая незначимые слова, используя подходящие базы таких слов), просматривает RSS**-**ленты из заданного Internet-набора лент и выбирает новости, подходящие по значимым словам. Показывает новости пользователю в порядке частоты встречаемости значимых слов. Полученная мультиагентная система оценивается в баллах в два раза выше, чем обычно.

В простейшем варианте (оценивается стандартным количеством баллов) агент должен показывать в браузере новости RSS**-**ленты по указанному URL-адресу. Для тех, кому трудно программировать, достаточно написать xml-пример своей RSS-ленты (оценивается до 0.75 от стандартной оценки).

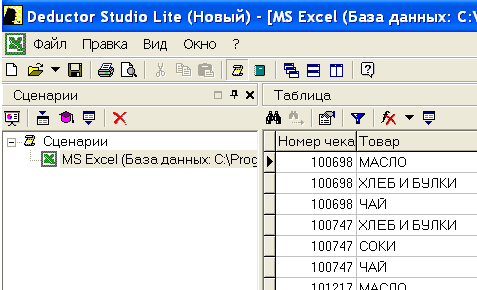
При сдаче лабораторной работы надо знать ответы на вопросы.

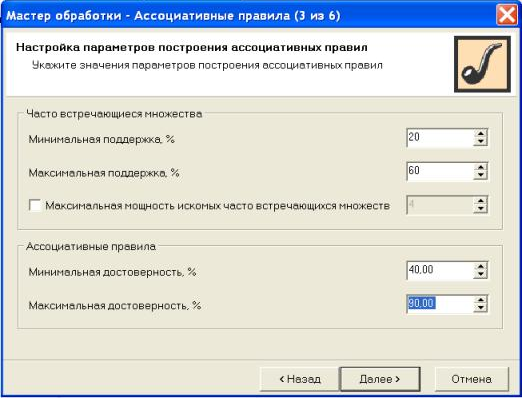
***Контрольные вопросы***

* 1. Понятие интеллектуального агента, мультиагентной системы, их типичные свойства.
  2. Какие свойства интеллектуальных агентов были реализованы в вашей лабораторной работе?
  3. Что дают онтологии для семантического Web?
  4. Что собой представляют метаданные в RSS-лентах?
  5. Нарисуйте семантическую сеть RSS-ленты. Покажите в ней классы, свойства. Классифицируйте сеть по количеству отношений и по типу отношений.
  6. Что такое база знаний? Как она формируется и заполняется?
  7. Как добываются новые знания из базы знаний?
  8. Где в RSS-ленте данные, а где знания? Чем отличаются языки запросов к базе знаний от языка SQL запроса к данным?
  9. Как микроданные помогают Google правильно распознавать семантику документа?

1. **Поиск ассоциативных правил в ИАД**

***Задание***

Создать в среде MS Excel транзакционную базу данных для двух-трёх десятков товаров. База должна состоять из двух столбцов: идентификатор транзакции (ID), наименование товара. Обязательно предусмотрите возможность покупки товаров одним чеком, т.е. с одной и той же транзакцией. Именно в таких покупках могут присутствовать ассоциативные правила. Используя методический материал <http://www.intuit.ru/studies/courses/6/6/lecture/186?page=5>, на основе платформы для анализа данных – Deductor (или Loginom ), найти наибо­лее часто встречающиеся наборы товаров (поддержка) и ассоциативные правила с достоверностью. Обоснуйте выбранные границы поддер­жки и достоверности.



При сдаче лабораторной работы надо знать ответы на вопросы:

***Контрольные вопросы***

* 1. Что такое скрытые знания?
  2. Расскажите алгоритм обнаружения знаний в ИАД.
  3. Что такое шаблоны и зачем они нужны в ИАД?
  4. Чем отличается классификация от кластеризации?
  5. Что такое прогнозирование?
  6. Что делает шаблон последовательностей?
  7. Что такое ассоциативные правила?
  8. На что влияют достоверность, поддержка в шаблонах ассоциативных правил?
  9. Как в ассоциативных правилах избавиться от очевидных и неинтересных закономерностей?

1. **Социальные сети**

***Задание***

Построить кластер ваших друзей в виде графа в социальной сети ВКонтакте. Найти коэффициент кластеризации *C* полученного кластера сети относительно вас. Рассчитать среднюю длину пути *L* в кластере.

***Указания и рекомендации***

1. Построить кластер друзей можно, например, с помощью приложения ВКонтакте <https://vk.com/socialgraph3d> (V/L - кол-во вершин и связей).
2. Локальный (относительно вас) коэффициент кластеризации *C* можно найти используя материалы ([слайды](http://tolstykh.com/docs/Слайды/Intelligent%20Web-technologies/%D0%92%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%B2%20%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5%20Web-%D1%82%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D0%B8.pptx)) лекций. По результатам найденного значения коэффициента кластеризации сделайте вывод: удовлетворяет ли рассматриваемый кластер сети модели малого мира?
3. Среднюю длину пути *L* можно найти используя материалы ([слайды](http://tolstykh.com/docs/Слайды/Intelligent%20Web-technologies/%D0%92%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%B2%20%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5%20Web-%D1%82%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D0%B8.pptx)) лекций. По результатам найденного значения длины пути *L* сделайте вывод: на сколько полученное значение *L* отличается от средней длины пути в модели случайного графа? Каково значения длины пути *L* было бы если ваш кластер был полносвяз­ным?
4. Предположить, что скорость диффузии информации в вашем кластере равна 0.5 дня за один средней шаг пути, то через какое время весь ваш социальный кластер будет информирован новостью, созданной любым членом кластера?

При сдаче лабораторной работы надо знать ответы на вопросы.

***Контрольные вопросы***

* 1. Какие типичные процессы моделируют и исследуют в комплексных сетях?
  2. Что такое модель случайного графа?
  3. Что такое средняя длина пути в сети и как она рассчитывается в модели случайного графа?
  4. Что такое кластеризация сети и как рассчитывается коэффициент кластеризации сети?
  5. Что такое полносвязная сеть и как определяется количество связей в ней?
  6. Особенности модели сети малого мира?
  7. Как зависит кластеризация и средняя длина пути в сети от случайных связей?
  8. Особенности модели сети предпочтительного соединения?
  9. О чём свидетельствует схожесть структуры Вселенной, социальных сетей и мозга человека?
  10. Что такое диффузионная модель сети?

Организация текущего, модульного и итогового контроля

**Критерии оценок:**

|  |  |  |
| --- | --- | --- |
| **Зачётные модули** | **Форма контроля, баллы** | **Итого баллы** |
|  | Четыре лабораторные работы, по 20 баллов каждая | 80 |
| 1 | Модульная контрольная работа | 20 |
|  |
| Общий итог |  | 100 |

**Вопросы для модульного контроля**

1. Что такое «Интеллектуальные агенты» и их типичные свойства?
2. Задачи агентов информационно-поисковых систем. Каков смысл их задач?
3. Булева модель поиска?
4. Векторная модель поиска?
5. Основные идеи технологии PageRank. Как разработчик Web-ресурсов может влиять на SEO?
6. Что такое Semantic Web?
7. Semantic Web: Что такое семантическая сеть? Классификация сетей.
8. Semantic Web: Приведите пример построения семантической сети.
9. Semantic Web: Что такое база знаний и для чего нужен язык SPARQL?
10. Что такое адаптация (персонализация) Web-ресурсов и её основные функции?
11. Основные приёмы адаптации представления Web-ресурсов?
12. Web Mining: Алгоритм построения модели пользователя для персонализации?
13. Web Mining: Алгоритм интеллектуального анализа данных для обнаружения знаний в Web?
14. Web Mining: Типичные виды закономерностей, выявляемые при интеллектуальном анализе данных?
15. Web Mining: Ассоциативные правила интеллектуального анализа данных? Приведите пример.
16. Комплексные сети: Что такое средняя длина пути в сети, что она характеризует?
17. Комплексные сети: Что такое кластеризация сети, что она характеризует?
18. Комплексные сети: Как зависит кластеризация и средняя длина пути в сети от случайных связей?
19. Комплексные сети: Особенности модели случайного графа?
20. Комплексные сети: Особенности модели сети малого мира?
21. Комплексные сети: Особенности модели сети предпочтительного соединения?
22. Комплексные сети: Особенности диффузионной модель сети?

**Методическое обеспечение**

Сайт В. К. Толстых <http://www.tolstykh.com>, слайды лекций:

[Введение в интеллектуальные Web-технологии](http://tolstykh.com/docs/Слайды/Intelligent%20Web-technologies/%D0%92%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%B2%20%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5%20Web-%D1%82%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D0%B8.pptx);

[Поиск в динамической новостной среде](http://tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/%D0%9F%D0%BE%D0%B8%D1%81%D0%BA%20%D0%B2%20%D0%B4%D0%B8%D0%BD%D0%B0%D0%BC%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B9%20%D0%BD%D0%BE%D0%B2%D0%BE%D1%81%D1%82%D0%BD%D0%BE%D0%B9%20%D1%81%D1%80%D0%B5%D0%B4%D0%B5.ppt);

[Характеристики ИПС](http://tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/%D0%A5%D0%B0%D1%80%D0%B0%D0%BA%D1%82%D0%B5%D1%80%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B8%20%D0%98%D0%9F%D0%A1.ppt);

[Модели информационного поиска](http://tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D0%B8%20%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B3%D0%BE%20%D0%BF%D0%BE%D0%B8%D1%81%D0%BA%D0%B0.ppt);

[Text-Mining](http://tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/New%20Media.ppt);

[Семантический Web](http://tolstykh.com/docs/Слайды/Advanced%20Internet-technologies/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9%20Web.pptx);

**Рекомендуемая литература**

1. **Боженюк A.B.** Интеллектуальные интернет-технологии : учебник / A.B. Боженюк, Э.М. Котов., A.A. Целых. — Ростов н/Д: Феникс, 2009С — 381.
2. **Гаврилова Т. А.** Базы знаний интеллектуальных систем : учебник / Т. А. Гаврилова, Ф. В. Хорошевский. — СПб.: Питер, 2001. — 384 с.
3. **Дари Кристиан, Сирович Джейми**. Поисковая оптимизация на ASP.NET для профессионалов. Руководство разработчика по SEO.- М.: ООО «И. Д. Вильямс», 2008.- 400 с.: ил.
4. **Жуков Л.** Теория социальных сетей. — <http://leonidzhukov.ru/hse/2012/socialnetworks/>
5. **Ландэ Д. В.** Поиск знаний в Internet.- М: Диалектика, 2005.- 272с.
6. Web Mining: интеллектуальный анализ данных в сети Internet // Управление знаниями. — <https://sites.google.com/site/upravlenieznaniami/>

1. Например, продвинутый поиск в Яндекс - <https://hackware.ru/?p=6045>, в Гугл - <https://hackware.ru/?p=5996> [↑](#footnote-ref-1)
2. Предмет исследований вашей диссертации согласуйте с руководителем. [↑](#footnote-ref-2)