**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего профессионального образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет бизнес-информатики

Отделение прикладной математики и информатики

**Программа дисциплины**

**«Социальные сети»**

Для направления 035800.68 «Фундаментальная и прикладная лингвистика»

Для магистерской программы "Компьютерная лингвистика"

Автор программы:

Жуков Леонид Евгеньевич, PhD, профессор, lzhukov@hse.ru

Одобрена на заседании кафедры [Введите название кафедры] «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 20 г

Зав. кафедрой [Введите И.О. Фамилия]

Рекомендована секцией УМС [Введите название секции УМС] «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 20 г

Председатель [Введите И.О. Фамилия]

Утверждена УС факультета [Введите название факультета] «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_20 г.

Ученый секретарь [Введите И.О. Фамилия] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ [подпись]]

Москва, 201\_

*Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.*

# Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов направления специальности 035800.68 Фундаментальная и прикладная лингвистика , обучающихся по магистерской программе "Компьютерная лингвистика" изучающих дисциплину «Социальные сети».

Программа разработана в соответствии с:

* Образовательным стандартом федерального государственного автономного образовательного учреждения высшего профессионального образования национального исследовательского университета «Высшая школа экономики», в отношении которого установлена категория «национальный исследовательский университет»
* Учебным планом университета по направлению подготовки 035800.68 «Фундаментальная и прикладная лингвистика» для подготовки магистра для магистерской программы «Компьютерная лингвистика» утвержденным в 2012г

# Цели освоения дисциплины

Целями освоения дисциплины «**Социальные сети**» являются овладение студентами моделями и методами анализа сетевых данных и их применения к лингвистическому анализу, а также приобретение навыков исследователя и разработчика математических моделей, методов, алгоритмов и совершенствование навыков работы по анализу реальных данных.

# Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

* Знать основные модели и методы анализа социальных сетей
* Уметь адекватно применять указанные модели и методы, а также программные средства, в которых они реализованы
* Иметь навыки (приобрести опыт) анализа реальных сетевых данных с помощью изученных методов

В результате освоения дисциплины студент осваивает следующие компетенции:

| Компетенция | Код по ФГОС/ НИУ | Дескрипторы – основные признаки освоения (показатели достижения результата) | Формы и методы обучения, способствующие формированию и развитию компетенции |
| --- | --- | --- | --- |
| Способен рефлексировать (оценивать и перерабатывать) освоенные научные методы и способы деятельности. | СК-1 | Понимание места и ценности методов машинного обучения и разработки данных в современной науке и практической деятельности. | Лекции и практикумы. Работа по изучению научных статей и подготовке доклада. |
| Способен анализировать, верифицировать, оценивать полноту информации в ходе профессиональной деятельности, при необходимости восполнять и синтезировать недостающую информацию. | СК-6 | Умеет самостоятельно находить научно-техническую литеру по тематике и критически подходить к оценке моделей и методов машинного обучения и разработки в ней рассмотренных. | Работа по изучению научных статей. Групповая работа над анализом реальных данных. |
| Способен управлять проектами (подпроектами), анализировать риски | ПК-7 | Корректно применяет полученные ранее знания в таких дисциплинах как математический анализ, дифференциальные уравнение, дискретная математика и стохастическое моделирование при изучении материалов курса. | Работа по изучению научных статей и подготовке доклада. Практикумы. Работа над проектом по анализу реальных данных. |
| Способен обеспечивать аналитическую поддержку выработки и принятия решений в сферах, касающихся языковой политики, мониторинга текстовых данных, сохранения языкового наследия | ПК-11 | Адекватно определяет тип задачи, строит модель и подбирает параметры методов. | Работа по изучению научных статей и подготовке доклада. Практикумы. Работа над проектом по анализу реальных данных. |
| Способен осуществлять лингвистическуюобработку текстов в производственно-практических целях | ПК-17 | Корректно применяет математические модели и методы прикладной математики в анализе данных. | Работа по изучению научных статей и подготовке доклада. Работа над проектом по анализу реальных данных. |
| Способен разрабатывать методики и алгоритмы дигитализации гуманитарного наследия, в том числе специальные языковые разметки корпусов  | ПК-20 | Умеет представлять результаты проведенного исследования в виде отчета или устного рассказа с показом мультимедийной презентации. | Работа по изучению научных статей и подготовке доклада. Подготовка и представление результатов отчета по проекту. |

# Место дисциплины в структуре образовательной программы

Для магистерской программы «Компьютерная лингвистика» настоящая дисциплина является дисциплиной по выбору.

Изучение данной дисциплины использует элементы следующих дисциплин:

* Математика(адаптационная дисциплина)
* Математические основания компьютерной лингвистики
* Программирование для лингвистов

Для освоения учебной дисциплины, студентам желательно владеть следующими знаниями и компетенциями:

* Знание основных понятий линейной алгебры
* Знание основных понятий теории вероятности и математической статистики.
* Основные определения теории графов и способы представления графов в виде структур данных.

# Тематический план учебной дисциплины

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название раздела | Всего часов  | Аудиторные часы | Самостоя­тельная работа |
| Лекции | Семинары | Практические занятия |
| 1 | Введение в анализ сложных сетей | 20 | 2 | 2 |  | 16 |
| 2 | Метрики центральности узлов и анализ связей | 26 | 4 | 2 |  | 20 |
| 3 | Методы нахождения сетевых сообществ | 32 | 4 | 4 |  | 24 |
| 4 | Структурные эквивалентности узлов | 20 | 2 | 2 |  | 16 |
| 5 | Модели формирования сетей | 20 | 2 | 2 |  | 16 |
| 6 | Визуализация сетей | 26 | 2 | 4 |  | 20 |
|  | Итого | 144 | 16 | 16 |  | 112 |

# Формы контроля знаний студентов

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 2 год | Параметры  |
| 1 | 2 | 3 | 4 |
| Текущий(неделя) |  |
| Эссе по прочитанной научной статье | Текущий(неделя) |  |  |  | Список тем и сроки обговариваются со студентами. |
| Проект по анализу реальных данных |  |  |  |  | Предполагается промежуточный контроль в рамках зачета первого модуля и итоговая защита в рамках экзамена во втором модуле. |
| Домашнее задание |  |  |  |  | срок выполнения и требования указываются в тексте задания |
| Промежу­точный | Коллоквиум | 1 |  |  |  | в устной форме, а также собеседование по результатам домашнего задания и проекта |
| Итоговый | Зачет |  | 2 |  |  | в письменной форме, а также собеседование по результатам домашнего задания и проекта |

## Критерии оценки знаний, навыков

Оценки по всем формам контроля выставляются по 10-ти балльной шкале.

Проект по анализу реальных данных подразумевает индивидуальное выполнение задания, связанного с применением программных средств и методов анализа социальных сетей к текстовым данным. Защита проекта происходит в два этапа, на первом модуле к зачету происходит выбор данных и обоснование методов для решения задачи, в рамках коллоквиума проходит собеседование с группой или индивидуально. К зачету каждый студент высылает отчет по проекту, а рамках устной презентации происходит его защита.

Осуществляется следующая дистанционная поддержка при проведении контроля: задания и материалы лекций, рассылаются по электронной почте, учебная литература и материалы лекций доступны на вебстранице класса.

## Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на практических занятиях: учитывается активность студентов и правильность решения задач. Оценки за работу на практических занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на практических занятиях определяется перед промежуточным или итоговым контролем – *Оаудиторная*.

Преподаватель оценивает самостоятельную работу студентов: оценивается правильность выполнения домашних работ (*имеются ввиду домашние работы, которые не включаются в РУП, это не форма текущего контроля "Домашнее задание"*), полнота освещения темы, которую студент готовит для выступления с докладом. Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за самостоятельную работу определяется перед промежуточным или итоговым контролем – *Осам. работа*.

Накопленная оценка за текущий контроль учитывает результаты студента следующим образом:

О*накопленная*= 0,8∙ *О текущий* + 0,1∙ О ауд + 0,1∙ О сам.работа

где О текущий рассчитывается как взвешенная сумма всех форм текущего контроля

О*текущая* = 0,4·О*эссе* +0,2 ·О*д/з* +0,4 ·О *коллоквиум* ;

Способ округления накопленной оценки текущего контроля: арифметический.

Результирующая оценка за дисциплину рассчитывается следующим образом:

О *результат* = 0,2 О *накопл* + 0,8∙О *зачет*

Способ округления накопленной оценки промежуточного (итогового) контроля в форме экзамена: арифметический.

В случае особых обстоятельств студент может получить возможность пересдать низкие результаты за текущий контроль или работу на занятиях, самостоятельную работу.

На зачете студент может получить дополнительный вопрос (дополнительную практическую задачу, решить к пересдаче домашнее задание), ответ на который оценивается в 1 балл.

На экзамене студент может получить дополнительный вопрос (дополнительную практическую задачу, решить к пересдаче домашнее задание), ответ на который оценивается в 1 балл.

В диплом выставляет результирующая оценка по учебной дисциплине.

# Содержание дисциплины

**Тема 1. Введение в анализ сложных сетей**

**Вопросы лекционной части курса**

Введение в теорию комплексных систем. Основные понятия в теории сетей. Свойства и метрики анализа сетей. Степенное распределение. Масштабно-инвариантные сети (scale-free networks). Случайные графы. Распределение Парето, нормализация, моменты,Закон Ципфа.Граф ранк-частота Диаметр и кластерный коэффициент

**Основная литература**

## Albert-Laszlo Barabasi and Eric Bonabeau. Scale-Free Networks, 2003.

## L.A.N. Amarala and J.M. Ottino. Complex networks: Augmenting the framework for the study of complex systems, 2003.

1. M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46, pp 323-351 (2005).

**Дополнительная литература**

## R. Albert and A.-L. Barabasi, Statistical mechanics of complex networks, Rev. Modern Phys., 74 (2002), pp. 47–97.

## M. E. J. Newman. The Structure and Function of Complex Networks, 2003

## S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, Adv. in Phys., 51 (2002), pp. 1079–1187.

**Тема 2. Метрики центральности узлов и анализ связей**

**Вопросы лекционной части курса**

Анализ узлов и связей / Nodes metrics and link analysis . Метрики центральности узлов / Centrality metrics. Понятия центральности и престижа. Модельные графы. Degree centrality, closeness centrality, betweenness centrality, статус/rank prestige (eigenvector centrality). Центральность сети (сentralization). Анализ связей. Алгоритм PageRank. Стохастические матрицы. Hubs и Authorities. Алгоритм HITS.

**Вопросы практической части курса**

Работа в среде программирования R

**Основная литература**

1. L.C. Freeman, Centrality in Social Networks Conceptual Clarification, Social Networks 1, pp 215-239 (1978/79)
2. P. Bonacich, Power and Centrality: A Family of Measures. American Journal of Sociology, vol 92, N 5, 1987

**Тема 3. Методы нахождения сетевых сообществ**

**Вопросы лекционной части курса**

Понятие сетевых сообществ (network communities). Плотность связей. Метрики. Разделение графа на части (graph partitioning). Разрезы (cuts) в графе. Min-cut, quotent and normalized cuts метрики. Divisive and agglomerative algorithms. Repeated bisection. Корреляционная матрица. Clustering. Классификация алгоритмов нахождения сообществ. Алгоритмы нахождения сетевых сообществ Edge Betweenness. Newman-Girvin algorithm. Spectral methods. Modularity maximization algorithm

**Вопросы практической части курса**

 Библиотека для работы с сетями iGraph, нахождение сообществ

**Основная литература**

1. M.E.J Newman and M. Grivan. Finding and evaluating community structure in networks. Physical Review E 69, 026113, 2004
2. S. Fortunato. Community detection in graphs. Physics Reports, V. 486, Issues 3–5, pp75-174, 2010

**Дополнительная литература**

1. M. E. J. Newman. The Structure and Function of Complex Networks, SIAM Review, V 45, N2, pp 167-256, 2003

**Тема 4. Структурные эквивалентности узлов**

**Вопросы лекционной части курса**

Метрики структурной эквивалентности узлов. Эвклидово расстояние. Расстояни Хэмминга. (Eucleadean and Hamming distance). Корреляционный коэффициент. Сходство по косинусу (cosine similarity). Ассортативнoe смешивание (homophily). Модулярность (modularity). Ассортативный коэффициент (Assortativity coefficient). Смешивание по степеням узлов (Mixing by degree).

**Вопросы практической части курса**

Библиотека для работы с сетями iGraph, нахождение структурно эквивалентых узлов

**Основная литература**

1. Social Network Analysis: Methods and Applications. Stanley Wasserman, Katherine Faust. Cambridge University Press, 1995.

**Дополнительная литература**

## M. E. J. Newman. The Structure and Function of Complex Networks, 2003

**Тема 5. Модели формирования сетей**

**Вопросы лекционной части курса**

Случайные графы (Random graphs). Модель "малого мира". Модель предпочтительного присоединения. Стратегические модели формирования сетей.

**Основная литература**

1. D. J. Watts and S. H. Strogatz, Collective dynamics of “small-world” networks, Nature, 393 (1998), pp. 440–442.
2. P. Erdos, A . Renyi, On Random Graphs, in Publ. Math. Debrecen 6, p. 290–297, 1959
3. A.L. Barabasi, R. Albert, Emergence of Scaling in Random Networks, Science, vol 286, p. 510, 1999

**Дополнительная литература**

1. Matthew O. Jackson, ``The Economics of Social Networks,'' Chapter 1 in Volume I of Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press, 2006.

**Тема 6. Визуализация сетей**

**Вопросы лекционной части курса**

Методы и инструменты для визуализации сетей.

**Вопросы практической части курса**

Визуализация сетей в библиотеке iGraph, NetworkX, yEd

# Образовательные технологии

Необходимое для выполнения работ программное обеспечение, как правило, находится в свободном доступе и его можно загрузить в сети Интернет или скопировать с диска, прилагающегося к курсу.

Дополнительно к каждой из тем прилагаются слайды лекций, изложение которых адаптировано с учетом используемых программных пакетов.

В рамках курса используется также проектная групповая форма работы.

## Методические рекомендации преподавателю

Преподавателю рекомендуется использовать демонстрацию работы методов в программных продуктах во время лекционных занятий.

# Оценочные средства для текущего контроля и аттестации студента

## Тематика заданий текущего контроля

## Выполнение студентами самостоятельной практической курсовой работы по созданию лингвистической сети на основе анализа текста, определении основных свойств сети, нахождение метрик центральности узлов, сетевых сообществ и визуализации сети. Формулировка выводов и заключений полученных в работе

## Вопросы для оценки качества освоения дисциплины

1. Привести примеры комплексных сетей
2. Дать определение расстояние между узлами на графе
3. В чем отличие между понятиями связанные и сильно связанные компоненты
4. Перечислить методы измерения центральности узлов в социальной сети
5. Объяснить понятие «малого мира”
6. Привести примеры распределения Парето
7. Закон Ципфа.
8. Рассказать о методах построения случайных графов
9. Рассказать о модели предпочтительного присоединения
10. Рассказать о методах визуализации сетей

## 9 Учебно-методическое и информационное обеспечение дисциплины

## 9.1 Базовый учебник

* Jurafsky, Daniel, and James H. Martin. (2009). [Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics](http://www.cs.colorado.edu/~martin/slp.html) . 2nd edition. Prentice-Hall

## 9.1 Основная литература

* [Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце](http://www.ozon.ru/context/detail/id/5497130/#tab_person) Введение в информационный поиск, М. Вильямс, 2011

## 9.2 Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

* Среда программирования R, Python
* Библиотека  iGraph (<http://igraph.sourceforge.net>)
* Библиотека  NetworkX (http://networkx.github.io)
* Системы визуализации yEd (http://www.yworks.com/), Gephi (http://gephi.org)

## 9.3 Дистанционная поддержка дисциплины

Дистанционная поддержка осуществляется с помощью вебсайта курса.

# 10 Материально-техническое обеспечение дисциплины

Используется проектор (для лекций или семинаров), слайды мультимедийных презентаций и компьютеры с предустановленным программным обеспечением и доступ в Интернет.