



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

Факультет гуманитарных наук  
Школа лингвистики

**Рабочая программа дисциплины Введение в Digital Humanities 1**

для цикла свободного минора факультета гуманитарных наук  
уровень: бакалавриат

Разработчик(и) программы

Франк Фишер, PhD, доцент школы лингвистики, [ffischer@hse.ru](mailto:ffischer@hse.ru)

Бонч-Осмоловская А.А., к.ф.н., доцент школы лингвистики, [abonch@hse.ru](mailto:abonch@hse.ru)

Орехов Б.В., к.ф.н., доцент школы лингвистики, [borekhov@hse.ru](mailto:borekhov@hse.ru)

Скоринкин Д.А., преподаватель школы лингвистики, [dskorinkin@hse.ru](mailto:dskorinkin@hse.ru)

Одобрена на заседании школы лингвистики «5» июня 2018 г.

Руководитель школы Е.В. Рахилина \_\_\_\_\_

Рекомендована Академическим советом образовательной программы

«05» июня 2018 г., № протокола 15

Утверждена «05» июня 2018 г.

Академический руководитель образовательной программы

Ю.А. Ландер \_\_\_\_\_

Москва, 2018

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



## 1. Область применения и нормативные ссылки

Настоящая программа учебной дисциплины майнора устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину Введение в Digital Humanities 1, учебных ассистентов и студентов, выбравших данный майнор.

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ;
- Программой свободного минора

## 1 Цели освоения дисциплины

Целью освоения дисциплины Введение в Digital Humanities 1 является знакомство студентов с новыми исследовательскими подходами и задачами в истории, литературоведении, лингвистике, культурологии, опирающимся на применение методов компьютерной обработки текста. Эти методы включают в себя создание и анализ корпусов текстов, построение статистических моделей, работу с большими данными, формализацию параметров текста, картографирование, сетевой анализ. Упомянутые подходы широко используются в современной гуманитарной науке, задача курса состоит в том, чтобы показать студентам исследовательский потенциал этих методов и вооружить их практическими навыками работы с существующими инструментами компьютерных исследований в гуманитарной сфере.

## 2 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать современные методы получения, обработки и анализа гуманитарных данных
- Уметь использовать в исследовательской работе корпуса текстов, свободные инструменты текстовой аналитики, сетевого анализа
- Иметь навыки (приобрести опыт) корпусных исследований, количественного анализа художественных текстов, анализа географических данных и социальных сетей в гуманитарных областях.

Уровни формирования компетенций:

**РБ** — ресурсная база, в основном теоретические и предметные основы (знания, умения);

**СД** — способы деятельности, составляющие практическое ядро данной компетенции;

**МЦ** — мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать.

В результате освоения дисциплины студент осваивает компетенции:

| Компетенция   | Код по ОС ВШЭ | Дескрипторы – основные признаки освоения (показатели достижения результата)  | Формы и методы обучения, способствующие формированию и развитию компетенции | Форма контроля уровня сформированности компетенции                  |
|---|---------------|--|---|---|
| Способен учиться, приобретать новые знания, умения, в том числе в области, отличной от профессиональной | УК-1(СК-Б1)   | Дает определение понятию цифровых гуманитарных исследований (Digital Humanities), может ограничить круг таких, воспроизводит актуальные методы | Лекции, семинары, итоговые презентации проектов                             | Задания, выполняемые на семинарах, подготовка и презентация проекта |



| Компетенция   | Код по ОС ВШЭ | Дескрипторы – основные признаки освоения (показатели достижения результата)   | Формы и методы обучения, способствующие формированию и развитию компетенции | Форма контроля уровня сформированности компетенции                  |
|---|---------------|---|---|---|
|   |               | цифровых исследований в гуманитарных науках на предложенном или самостоятельно найденном материале, владеет инструментами цифровых гуманитарных исследований. |   |   |
| Способен вести исследовательскую деятельность, включая анализ проблем, постановку целей и задач, выделение объекта и предмета исследования, выбор способа и методов исследования, а также оценку его качества | УК-6 (СК-Б7)  | Представляет связи между измеримыми количественными параметрами текста и его значимыми содержательными, в т.ч. художественными, композиционными свойствами    | Лекции, семинары, итоговые презентации проектов                             | Задания, выполняемые на семинарах, подготовка и презентация проекта |
| Способен работать в команде   | УК-7 (СК-Б8)  | Владеет навыками работы в команде, умеет распределять задачи, брать ответственность за работу, выделять и выполнять собственную функцию.                      | Семинары, итоговые презентации проектов                                     | Подготовка и презентация проекта                                    |
| Способен ориентироваться в системе общечеловеческих ценностей и ценностей мировой и российской культуры, понимает значение гуманистических ценностей для сохранения и развития современной цивилизации        | ПК-24 (СК-Б9) | Понимает ценность культурного наследия, необходимость его сохранения в цифровой форме.  | Лекции  | Работа с материалом на семинарских занятиях                         |

### 3 Место дисциплины в структуре образовательной программы

Для освоения учебной дисциплины, студенты должны владеть следующими знаниями и компетенциями:<sup>1</sup>

- Базовые навыки работы с персональным компьютером на уровне начального пользователя

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

<sup>1</sup> Если дисциплина не предполагает пререквизитов-постреквизитов, соответствующие абзацы можно удалить



- Введение в Digital Humanities 2

#### 4 Тематический план учебной дисциплины

| № | Название раздела   | Всего часов | Аудиторные часы |          |                      |                                 | Самостоятельная работа |
|---|--|-------------|-----------------|----------|----------------------|---------------------------------|------------------------|
|   |  |             | Лекции          | Семинары | Практические занятия | Другие виды работы <sup>2</sup> |                        |
| 1 | Обзор новых методов в гуманитарных науках. Компьютерные методы в филологии: «дальнее чтение» Ф. Моретти, макроанализ М. Джокерса |             | 4               | 4        |                      |                                 | 2                      |
| 2 | Корпусные исследования в лингвистике, филологии, истории, культурологии, искусствоведении  |             | 6               | 4        |                      |                                 | 30                     |
| 3 | Стилеметрия и количественные методы атрибуции авторства  |             | 4               | 2        |                      |                                 | 30                     |
| 4 | Теория графов, сетевой анализ, их применение в гуманитарных науках   |             | 4               | 4        |                      |                                 | 20                     |
|   | <b>ИТОГО</b>   | 114         | 18              | 14       |                      |                                 | 82                     |

#### 5 Формы контроля знаний студентов

| Тип контроля  | Форма контроля | 2 или 3 курс |   |   |   | Параметры **                                       |
|---------------|----------------|--------------|---|---|---|--|
|               |                | 1            | 2 | 3 | 4 |  |
| Текущий       | Эссе           |              |   |   |   | 1-2 тыс. слов                                      |
|               | Проект         | 1            |   |   |   | Презентация из 10-20 слайдов, групповой доклад     |
| Промежуточный | Экзамен        |              |   |   |   |  |
| Завершающий   | Экзамен        | 1            |   |   |   | Итоговое представление исследовательского проекта. |

<sup>2</sup> Указать другие виды аудиторной работы студентов, если они применяются при изучении данной дисциплины.



## 6 Критерии оценки знаний, навыков

Итоговый экзамен проводится в форме защиты реализованного проекта цифрового гуманитарного исследования. Также студенты делают доклады на семинарах, в ходе которых демонстрируют освоение и проработку выбранного для освещения материала.

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

Задания доступны на вики-странице курса.

## 7 Содержание дисциплины

**Тема 1. Обзор новых методов в гуманитарных науках.** Компьютерные методы в филологии: «дальнее чтение» Ф.Моретти, макроанализ М. Джокерса.

Гуманитарные науки и вызов цифровой эпохи. Анализ данных в литературоведении, лингвистике, исторической науке, культурологии, истории искусства и креативных индустрий. Цифровые издания, картографические проекты, количественное отображение культурных трендов, визуализация, сети и графы в гуманитарных науках. Проблема репрезентативности литературного канона. «Дальнее чтение» (Distant Reading) Франко Моретти как способ объективизации филологического исследования. Макроанализ М. Джокерса. Идея «масштабируемого чтения». Использование инструмента Google Ngram Viewer; ограничения этого инструмента.

8 часов аудиторной работы.

Литература (полные обозначения см. в конце программы):

Liu, 2011; Hoover, 2007; Moretti, 2005; Schreibman et al, 2004; Moretti, 2013/Моретти, 2016; Jockers, 2013; Michel et al, 2011; Ярхо, 2006

### Тема 2. Корпусные исследования.

Понятие корпуса. Корпус текстов как исследовательский объект. Виды корпусов. Лингвистические корпуса, национальные корпуса. Параллельные, устные, мультимедийные корпуса. Разметка и метаданные в корпусе. Запросы к корпусу. Типология корпусных исследований.

Корпусные исследования прозы: на какую помощь со стороны цифрового знания рассчитывают литературоведы? Что можно посчитать в романе? Корпусные исследования поэзии. Семантический ореол метра.

10 часов аудиторной работы.

Литература (полные обозначения см. в конце программы):

Hoover et al, 2014; Orekhov, 2015; Ляшевская, 2016

**Тема 3. Компьютерные методы в филологии:** стилеметрия (компьютерная стилистика), тематическое моделирование.

История формальных исследований авторского стиля. Проблема спорного авторства и количественные подходы к её решению. Спорное авторство произведений Шекспира, «Записок федералиста», Музиля, Шолохова и др. Пример раскрытия авторства Дж. Роулинг (Juola, 2006). Понятие идиостиля автора. Программный пакет stylo для определения авторства (Eder et al, 2013).

6 часов аудиторной работы.

Литература (полные обозначения см. в конце программы):

Mendenhall, 1887; Mosteller, Wallace, 1963; Burrows, 2002; Juola, 2006; Eder et al, 2013; Мухин,



#### **Тема 4. Теория сетей (графов) и её применения в гуманитарных науках.**

Математический граф и его основные элементы (ребра, вершины). Применение теории графов в анализе социальных сетей. История становления сетевого анализа в гуманитарных областях. Гарвардский прорыв 1960-х. Социологические, исторические, культурологические исследования с использованием теории графов. Применение сетевого анализа в литературоведении.

Значимые количественные (математические) параметры графа. Плотность и диаметр графа. Основные свойства ребер и вершин. Степень вершины. Метрики центральности (betweenness centrality, closeness centrality).

Программа для анализа и визуализации графов Gephi. Основные возможности Gephi: импорт графа, алгоритмы укладки на плоскости, подсчет метрик и статистики, кластеризация графа с использованием Лувенского алгоритма.

8 часов аудиторной работы.

Литература (полные обозначения см. в конце программы):  
Newman, 2005; Bastian et al, 2009; Moretti, 2011;

### **8 Образовательные технологии**

На лекциях используются презентации, аудиозаписи и фрагменты видео. В ходе практических занятий применяются современные средства компьютерного анализа данных (соответствующие библиотеки языков программирования R, Python) и их визуализации (Gephi).

### **9 Оценочные средства для текущего контроля и аттестации студента**

#### **9.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля**

Итоговый экзамен по дисциплине проводится в форме защиты проекта (proof of concept) цифрового гуманитарного исследования или ресурса. От студентов ожидается подготовка такого проекта в качестве текущей работы по курсу; в процессе защиты студент должен показать глубокое знакомство с теми из основных современных методов гуманитарных исследований, которые применяются в её/его проекте.

### **10 Порядок формирования оценок по дисциплине**

Преподаватель оценивает работу студентов на семинарских занятиях: активность студентов в дискуссиях, работу с рекомендуемой литературой, доклады на выбранные темы. Оценки за работу на семинарских и практических занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем -  $O_{аудиторная}$ .

Результирующая оценка за итоговый контроль в форме экзамена выставляется по следующей формуле, где  $O_{экзамен}$  – оценка за работу непосредственно на экзамене:

$$O_{итоговый} = 0,6 \cdot O_{экзамен} + 0,4 \cdot O_{аудиторная}$$

Способ округления накопленной оценки итогового контроля в форме экзамена: в пользу студента.

### **11 Учебно-методическое и информационное обеспечение дисциплины**

#### **11.1 Базовый учебник**

Цифровые гуманитарные науки: хрестоматия / под ред. М. Террас, Д. Найхан, Э. Ванхутта, И. Кижнер. — Красноярск: Сиб. федер. ун-т, 2017. — 352 с.



Schreibman S., Siemens R., Unsworth, J. (eds.) (2004) *Companion to Digital Humanities*. Blackwell Companions to Literature and Culture. Oxford: Blackwell.

## 11.2 Основная литература

Hoover, D. Culpeper, J., O'Halloran, K. (2014) *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Routledge Advances in Corpus Linguistics.

Moretti, F. (2013) *Distant Reading*. London: Verso.

Моретти Ф. Дальнее чтение. М., Издательство Института Гайдара, 2016.

Сборники конференций Альянса организаций цифровых гуманитарных исследований (ADHO):

- a. <https://dh2018.adho.org/abstracts/>
- b. <https://dh2017.adho.org/program/abstracts/>
- c. <http://dh2016.adho.org/abstracts/>
- d. <http://dh2015.org/abstracts/>

## 11.3 Дополнительная литература

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.

Blei D. (2012) Introduction to Probabilistic Topic Models // *Communications of the ACM*. — С. 77–84.

Burrows, J. (2002) 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17:267-87

Eder, M. Kestemont, M. and Rybicki, J. (2013). Stylometry with R: a suite of tools. In: "Digital Humanities 2013: Conference Abstracts". University of Nebraska-Lincoln, Lincoln, NE, pp. 487-89.

Eder M. (2011) Style-markers in authorship attribution: A cross-language study of the authorial fingerprint. *Studies in Polish Linguistics*, 6:99–114.

Foote K., Lynch M., (2015) *Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions*". The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder.

Hoover, D. (2007) The End of the Irrelevant Text: Electronic Texts, Linguistics, and Literary Theory. *Digital Humanities Quarterly* 1.2.

Jockers M. (2013) *Macroanalysis: Digital Methods and Literary History*. Champaign, IL: University of Illinois Press.

Juola, P. (2006). *Authorship Attribution*. Foundations and Trends in Information Retrieval

Juola, P., Baayen, H. (2005) A Controlled-corpus Experiment in Authorship Identification by Cross-Entropy, *Literary and Linguistic Computing* 20 (Suppl 1), pp. 59-67

Kestemont, M. (2014) Function words in authorship attribution. from black magic to theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 59–66, Gothenburg, Sweden.

Lieberman E., Michel, J.B., Jackson J., Tang T., Nowak M. (2007) Quantifying the Evolutionary Dynamics of Language. *Nature*. p. 449.

Liu, A. (2011) The State of the Digital Humanities: A Report and a Critique. *Arts and Humanities in Higher Education* 2.1-2. pp. 8-41.

Mendenhall T. (1887) The Characteristic Curves Of Composition. *Science* : Vol. 9, Issue 214S, pp. 237-246

Michel, J.B., (2011). Quantitative analysis of culture using millions of digitized books . *Science*, 331(6014): pp. 176–82.

Moretti F. (2005) *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.

Moretti F. (2011) Network Theory, Plot Analysis. *New Left Review* 68, March-April

Mosteller F., Wallace D., (1963) Inference in an Authorship Problem. *Journal of the American Statistical Association*, Volume 58, Issue 302, pp. 275 - 309.

Newman M. (2010) *Networks: An Introduction*. Oxford: Oxford University Press.



- Orekhov B. (2015) Bashkir poetic corpus as a DH-resource / NRU HSE. Series WP BRP "Linguistics".
- Robinson P. (2013) Towards a Theory of Digital Editions. *Variants: The Journal of the European Society for Textual Schola*. Vol. 10, p. 105
- Schreibman S. (2013) Digital Scholarly Editing. *Literary Studies in the Digital Age: An Evolving Anthology* // <https://dlsanthology.mla.hcommons.org/digital-scholarly-editing/>
- Schreibman S., Siemens R. (eds.) (2008) *A Companion to Digital Literary Studies*. Oxford: Blackwell.
- Воронцов К.В. Вероятностное тематическое моделирование // <http://www.machinelearning.ru/wiki/images/2/22/Voron-2013-ptm.pdf>, 2013.
- Журкин И. Г., Шайтура С. В. *Геоинформационные системы*. — Москва: Кудиц-пресс, 2009.
- Ляшевская О. Н. *Корпусные инструменты в грамматических исследованиях русского языка*. М.: Языки славянской культуры, 2016.
- Мухин М. Ю. *Лексическая статистика и идиостиль автора: корпусное идеографическое исследование : на материале произведений М. Булгакова, В. Набокова, А. Платонова и М. Шолохова: диссертация на соискание ученой степени доктора филологических наук. [Место защиты: ГОУВПО "Уральский государственный университет"]*. - Екатеринбург, 2011. - 383 с.
- Ярхо Б. И., *Методология точного литературоведения: Избранные труды по теории литературы*, Ред. М. В. Акимова, И. А. Пильщиков и М. И. Шапир, Под общей редакцией М. И. Шапира, Москва: Языки славянских культур, 2006, xxxii, 927 с.

#### 11.4 Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

- AntConc ([laurenceanthony.net/software/antconc](http://laurenceanthony.net/software/antconc))
- Stylo ([sites.google.com/site/computationalstylistics/stylo](http://sites.google.com/site/computationalstylistics/stylo))
- Gephi ([gephi.org](http://gephi.org))

## 12 Материально-техническое обеспечение дисциплины

Для практических занятий используются персональные компьютеры; для лекций и семинаров необходим проектор.