

Цифровые гуманитарные технологии: ресурсы, инструменты, кейсы. Лекция 1. Вводная

Борис Орехов

НИУ Высшая школа экономики

nevmenandr@gmail.com

8 февраля 2018

- 1 Структура курса, организационная информация
- 2 Гуманитарные науки и цифра
- 3 Гуманитарные науки и естественные науки
- 4 Формализация в литературоведении

- 1 Структура курса, организационная информация
- 2 Гуманитарные науки и цифра
- 3 Гуманитарные науки и естественные науки
- 4 Формализация в литературоведении

Гарри Поттер и цифровые гуманитарные технологии: ресурсы, инструменты, кейсы

4 лекции, 10 семинаров.

Таблица: Структура курса

8.02	Лекция 1	Лекция 2
22.02	Лекция 3	Лекция 4
22.03	Семинар 1	Семинар 2
??.04	Семинар 3	Семинар 4
??.04	Семинар 5	Семинар 6
??.05	Семинар 7	Семинар 8
??.05	Семинар 9	Семинар 10

Лекции:

- 1 Вводная лекция. О цифровой гуманитарности.
- 2 Оцифровка текстов.
- 3 Компьютерная лингвистика и цифровые гуманитарные науки.
- 4 Сетевой анализ художественных текстов

Семинары:

- 1 Оцифровка в «Текстографе» (4 часа).
- 2 Стилеметрический анализ (4 часа).
- 3 Гуманитарные карты (2 часа).
- 4 Reading club (10 часов)

Из чего складывается оценка?

- Оценка складывается из 2 обязательных практических заданий, 1 прочитанной статьи по теме курса и 1 дополнительного задания.
- За каждое из обязательных заданий начисляется 3 балла при достойном уровне исполнения.
- За халтурное выполнение — 1 или 2 балла в зависимости от степени халтурности.
- За несделанное задание — 0 баллов.
- Задания сдаются через интернет-сервис github.com.
- Дополнительное задание «стоит» 1 балл.

Обязательные задания:

- 1 Оцифровка черновика А. С. Пушкина в системе «Текстограф» (я предоставлю сканы и транскрипции).
- 2 Силеметрическое исследование (можно делать в группе).

Дополнительное задание выдается индивидуально (географическая карта, сеть, компьютерно-лингвистическое исследование корпуса художественных текстов).

- TEI XML транскрипции стихотворения в вашем репозитории на гитхабе;
- материалы стилеметрического исследования на гитхабе в MD-формате;
- результат прочтения статьи из списка.

Результат прочтения статьи должен представлять собой её устный рассказ о ней на одном из пяти последних семинаров, или краткий аналитический пересказ по-русски, положенный в репозиторий на гитхабе.

К рассказу должен прилагаться (и лежать на гитхабе) доступный всем иллюстративный материал (презентация или хендаут).

- 1 Структура курса, организационная информация
- 2 Гуманитарные науки и цифра**
- 3 Гуманитарные науки и естественные науки
- 4 Формализация в литературоведении

Гуманитарные науки — это науки о культуре, то есть об искусственном, созданном человеком.

Продукты человеческой деятельности — объекты культуры в той мере, в какой они и управляются людьми (проблема языка).

Что такое «цифровой»?

Цифровой — так или иначе связанный с двумя сущностями:

- Компьютер как средство обработки информации.
- Количественные методы работы с данными.

Что привнёс компьютер?

- Сохранность информации благодаря удешевлению копирования (от древнерусских рукописей дошло одно промилле написанного).
- Облегчение доступа к информации (но люди по-прежнему, в основном, смотрят на котиков на Youtube).
- Упрощение подсчётов (количественные методы).

Можно определить через противопоставление качественным методам. Качественные методы дают возможность **установить связи** между объектами (взаимовлияния в культуре, цитаты в текстах, в т.ч. не дословные: ср. Вильгельм Баскервильский)

- Математическая статистика.
- Также позволяют сопоставлять объекты.
- Но только если их формализовать.
- В какой степени тексты, их форма, содержание и материал могут быть изучены с помощью количественных методов?

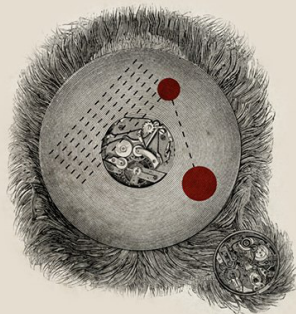
- Цифровые гуманитарные науки.
- Цифровая гуманитаристика.

<http://whatisdigitalhumanities.com/>: 817 определений.

- 1 Структура курса, организационная информация
- 2 Гуманитарные науки и цифра
- 3 Гуманитарные науки и естественные науки**
- 4 Формализация в литературоведении

Опыт применения количественных методов в естественных науках

- Большой опыт описания явлений природы при помощи числовых характеристик.
- Можно измерить скорость, массу, светимость, сопротивление, силу. . .
- Можно вывести формулы, которые увязывали бы эти явления в одну систему: $E = mc^2$.



ПРОТИВОЕСТЕСТВЕННЫЕ НАУКИ СМОГУТ ЛИ ДОГОВОРИТЬСЯ ФИЗИКИ И ЛИРИКИ?

Отличия гуманитарной науки от естественной

- Естественные науки и протиеоеетественные.
- В естественных науках модель создаётся из количественных характеристик того, что повторяется.
- В гуманитарных науках важным оказывается не повторяющееся, а единичное, не общее, а уникальное (не может быть класса пушкиных).
- «Каждый объект, будучи повторимым в самых разных своих деталях, уникален как единство смысла во всей полноте и в любых тонкостях его материального воплощения в чувственно воспринимаемой форме» (М. И. Шапир).
- Значение, выраженное в других словах — уже другое значение.

Отличия гуманитарной науки от естественной

- Для естественных наук важной является прогностическая функция.
- Гуманитарная наука изучает прошлое не для того, чтобы управлять будущим или (как частный случай) прогнозировать его.
- Гуманитарной науке достаточно знания самого по себе.
- Изучение семантики искусства не может базироваться на точных (т.е. формализованных до степени простого знака) методах, потому что здесь принципиально не работает закон логического тождества ($A=A$).
- Для разных людей один и тот же текст будет значить разное.

Отличия гуманитарной науки от естественной

- В естественных науках уже давно ясно, что именно следует формализовать и обратить в число.
- В гуманитарной науке не очевидно, что является сущностным, а что второстепенным.
- Можно посчитать длину всех романов в буквах. Ну и что?
- В естественных науках законы мира неизменны.
- В гуманитарной науке мир постоянно меняется.
- Для писателя древности традиционность выражения была добродетелью, для писателя эпохи модерна главное — новаторство.

- Некоторые важные для гуманитариев понятия совсем не поддаются формализации и переводу в число.
- Стилль?
- Понимание?
- Намеренность приёма?

М. И. Шапир: «С успехом применяя точные методы для теоретического описания физического мира, мы бессильны построить универсальную математическую модель даже самого «простого» и «поверхностного» явления духовной культуры. Причина банальна: нельзя сконструировать математизированную картину универсума, которым управляет множество относительно равноправных демиургов; кто-нибудь из них во что бы то ни стало отменит закон, установленный другим».

- Нельзя механически перенести на гуманитарные науки уже имеющиеся инструменты, не учитывая специфики знания.
- Компьютер хорош тем, что может считать и таким образом обрабатывать большие массивы данных.
- В гуманитарных науках не просто данные не структурированы (с этим компьютеры научились справляться), но и плохо формализуемо то знание, которое из этих данных хотят извлечь.

- Нет, кое-что всё-таки можно формализовать;
- Формализовав, встроить в систему накопленного знания;
- Хитрость именно в том, чтобы не опровергнуть, а соединить количественные методы с традиционными;
- Б. И. Ярхо: «Ни один математический акт не должен совершаться, пока в него не будет вложен конкретный литературоведческий смысл».

- 1 Структура курса, организационная информация
- 2 Гуманитарные науки и цифра
- 3 Гуманитарные науки и естественные науки
- 4 Формализация в литературоведении**

Чем занимается литературоведение?

C50 $\frac{\Gamma-4}{4175}$

Кабаков, Иван Флегонтович.

Сорта и агротехника кормовых культур Свердловской области. Свердловск, Кн. изд., 1954.

71 с. 20 см. 1.500 экз. 1 р.

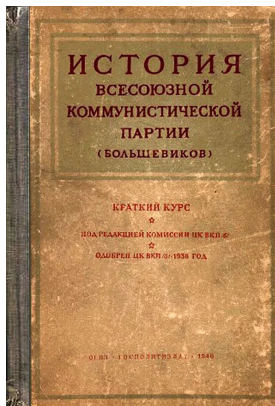
— — 1. Кормовые культуры — Возделывание — Свердл. обл.

№ 27825 [54-54723] и 17-4
XVI № 4820; № 12479
Всесоюз. кн. палата 16 XII 54

633.2
633.2/4(47.811)
K12

7.1.55

Не только художественные тексты



Гловинский М. «Краткий курс истории ВКП(б)» как мифическое сказание // Новое литературное обозрение. 1996. № 22. С. 142–160.

34-8

1011

Новиков, А.
Алешкин, В.

соавт.

... Захватные приспособления. Сопроводит. текст
к серии диапозитивов 300. М., Цутехпроп. Всес.
трест „Оргвыставка НКПС“, тип. „Гудок“. 1933.

24 с. (15×11) 400 [Б. ц.]

Перед загл.: В. Алешкин, А. Новиков, Вс. Остольский. Под ред.
В. Альтмача, В. Бубнова и А. Новикова. 621.87 : 778.28

И. Новиков, А. соавт. II. Остольский, Вс. О. соавт. III. „Оргвы-
ставка НКПС“, трест. Москва. — 1. Захватные приспособления—
Пояснительный текст к диапозитивам, кинофильмам и т. п.

5606/34 г. VIIA

Кн. л. № 14

5442

Что можно формализовать?

- Конвенционально принятая область: стиховедение.
- Язык художественной литературы.
- Всё то, что имеет дело не с высокоуровневыми. понятиями (стиль, литература), а с атомарными фактами: слоги, метры, слова.

- Понятно, что сущностно, а что второстепенно: метры, ритм, рифма, строфика.
- Можно посчитать количественное распределение метров в определённый период.
- Количественно описать динамику и связать её с имеющимся запасом знаний об истории литературы (четырёхстопный ямб).

- Давно возникла математическая лингвистика.
- Тогда мы научились считать слова и делать выводы об их распределениях.
- Частотные слова важны для текста при определённых оговорках.

Традиционный филолог = комментатор

Н. В. Брагинская: «Угроза репутации комментатора со стороны электронных средств поиска информации существует. А именно той репутации, которая строится на эрудиции, на больших мнемонических способностях, огромном многолетнем труде по накоплению знаний. Например, авторитетный филолог считает, что в таком-то тексте есть влияние аттицизма: дескать, вот это, и это, и это слово характерны для аттикизирующей стилизации. И приводит примеры. Ему верят как эксперту, на этом основании датируют памятник. Но по базе данных, где помещается практически вся греческая литература, выясняется, что названные им слова равномерно распределены по доаттикизирующему, аттикизирующему и уже не аттикизирующему периоду».

Для чего нужны компьютеры в филологии?

Н. В. Брагинская: «Экспертные оценки часто терпят крах, мнемонический ресурс падает в цене. Однако наличие баз данных дает возможность ставить такие научные и комментаторские задачи, о которых до их появления нечего было и думать. Непосредственно базы данных, ресурсы Интернета дают доступ только к готовому знанию-справке и освобождают от рутины, а не от профессии».



- Susan Schreibman, Ray Siemens, John Unsworth (eds.): Companion to Digital Humanities. Blackwell Companions to Literature and Culture. Oxford: Blackwell. 2004.
<http://www.digitalhumanities.org/companion/>
- Companion to Digital Humanities. 2016.
- Debates in the Digital Humanities. 2012 Edition.
<http://dhdebates.gc.cuny.edu/debates/1>
- Debates in the Digital Humanities. 2016 Edition.
<http://dhdebates.gc.cuny.edu/debates/2>
- Doing Digital Humanities. A DARIAH Bibliography.
https://www.zotero.org/groups/doing_digital_humanities_-_a_dariah_bibliography/items