

Каким должно быть электронное представление черновых редакций и вариантов?

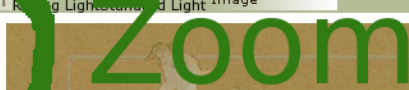
Борис Орехов (НИУ ВШЭ)

Рукописи vs. черновики

Мы уже умеем хорошо оцифровывать рукописи (прежде всего древние манускрипты).

Достижения электронного представления:

- Zoom
- Линковка прориси, транскрипта и изображения, подсветка



ЛИНКОВК

fppt.com

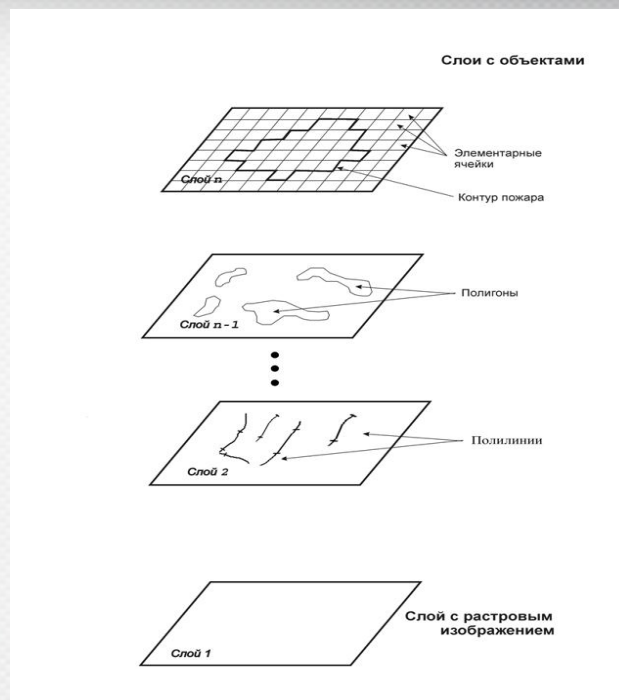
Рукописи vs. черновики

Но (просто) рукописи и авторские черновые редакции структурно не одно и то же.

Рукописи vs. черновики

Дело в иерархии:

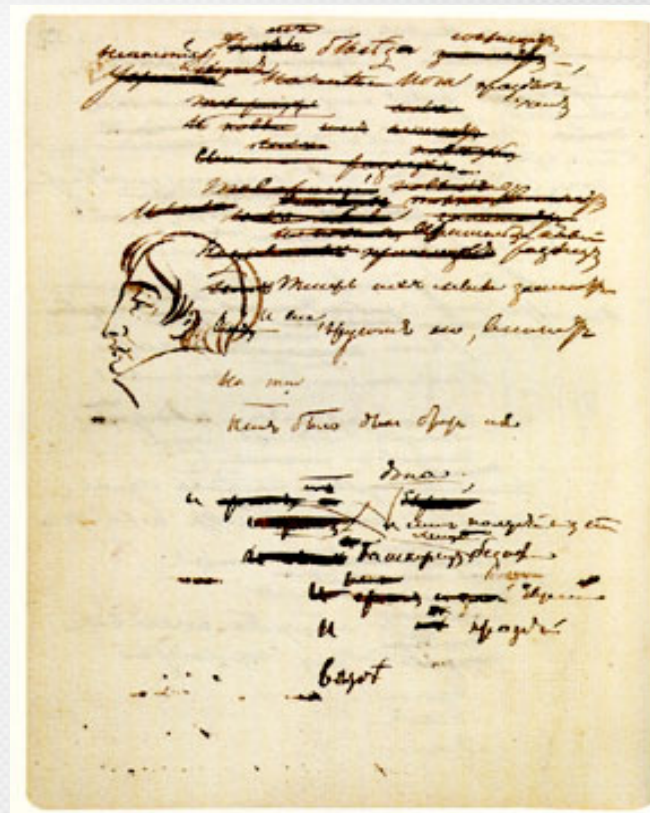
- Текст манускрипта устроен монолитно
- Черновики представляют текст, состоящий из множества слоёв.



Рукописи vs. черновики

Слои

взаимодействуют друг с другом, связи между ними многообразны, сложны, нелинейны и не всегда могут быть чётко описаны.



Задачи

Чего бы мы хотели от электронного представления такого типа системы?

Задачи

- Удобство сопоставления слоёв для пользователя
- Динамическая архитектура с возможностью перекombинации слоёв
- Помощь электронных инструментов в оценке связей между слоями

Что мы имеем сейчас?

Как существующие ресурсы справляются с этими задачами?

- Электронные библиотеки
- Корпуса

Электронные библиотеки

Библиотеки идут вслед за бумажными изданиями, рассматривают редакции и варианты как *отдельные тексты* и располагают их в разных документах. Связи не учитываются.

Неудобно сравнивать

Корпуса

Для лингвистов филологической проблемы вариантов не существует. В корпуса всегда включается один вариант, другими пренебрегают (схожим образом поступают и авторы словарей языка писателя).

Проблема игнорируется

Корпуса

Для лингвиста (создателя корпуса) текст внеиерархичен, плоскостен и оценивается двоично: есть в корпусе (1) или нет в корпусе (0). Корпуса плохо приспособлены к тому, чтобы отображать слои. Единственный тип корпусов, который к этому приближается: параллельные.

Параллельные корпуса

Тоже, впрочем, не идеально

Найдено 23 документа, 44 вхождения.

Поискать в других корпусах: [основном](#), [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [поэтическом](#), [синтаксическом](#), [устном](#).

Страницы: 1 [2](#) [3](#) [следующая страница](#)

1. [Московский государственный университет. Отчет № 1 о влиянии игр в рамках программы "Исследование влияния Олимпийских игр" \[ABBYY LingvoPRO\] \(2009-2010\) \[омонимия не снята\] Все примеры \(1\)](#)

ru	В этом смысле предлагаемый МОК вариант объединения индикаторов в блоки (экологический, экономический и социо-культурный) недостаточен для поставленных задач. [Московский государственный университет. Отчет № 1 о влиянии игр в рамках программы "Исследование влияния Олимпийских игр" [ABBYY LingvoPRO] (2009-2010) [омонимия не снята] ←...→
en	Consequently, the division of indicators suggested by the IOC (into environmental, economic and socio-cultural) is not sufficient to achieve the research objective. [Report № 1 on Games Impact in the framework of the programme "Olympic games impact – OGI"] [омонимия не снята] ←...→

2. [Operating instructions for a BOSCH gas hob \[ABBYY LingvoPRO\] \(2010\) \[омонимия не снята\] Все примеры \(1\)](#)

ru	А если рядом с вашим домом такого пункта нет, положите емкость с маслом в мусорное ведро (оттуда оно попадет на контролируемую свалку; возможен вариант , но таким образом мы можем избежать загрязнения воды). [Operating instructions for a BOSCH gas hob [ABBYY LingvoPRO] (2010)] [омонимия не снята] ←...→
en	Dispose of it in a closed container at your nearest collection point or, failing this, dispose of it in your rubbish bin. Although this last solution is far from perfect, you can also dispose of it to a controlled refuse dump where it will not be allowed to pollute water. [Руководство по эксплуатации варочной панели BOSCH] [омонимия не снята] ←...→

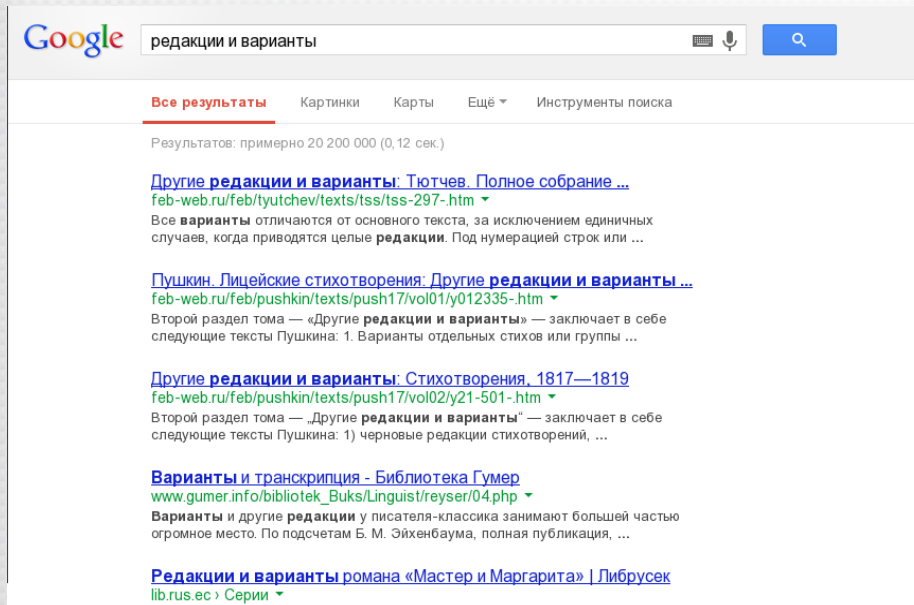
Что за пределами филологии?

Как решаются задачи визуализации
многослойной информации вне
филологических наук?

Элементы вертикального поиска

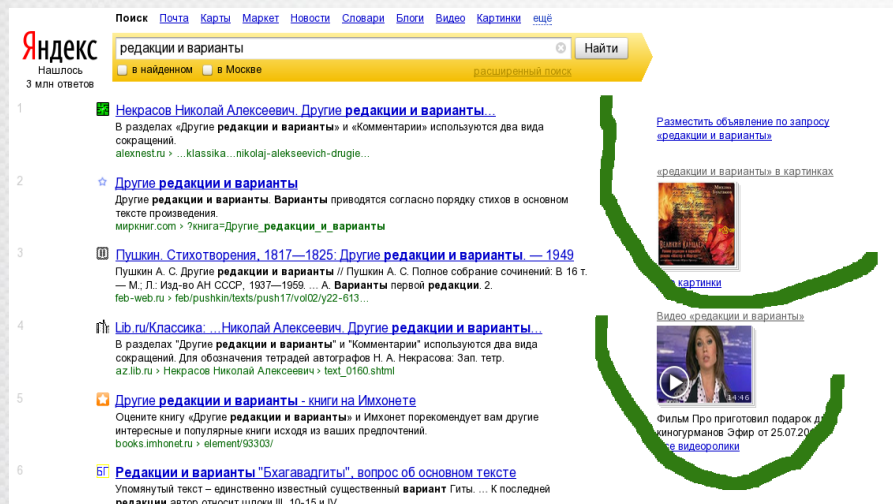
Самая близкая аналогия — элементы «вертикальности» в выдаче поисковых машин, когда поисковик показывает нам что-то дополнительное, такое, о чём мы его, может быть, и не спрашивали напрямую.

Элементы вертикального поиска



Обычный
«горизонтальный»
поиск

Элементы вертикального поиска



Элементы
вертикального поиска
в выдаче: поиск по
картинкам, поиск по
видео

Итог

Ничего особенного: разноплановая информация располагается рядом в отдельных блоках. Связи не визуализируются и не подчёркиваются.

Что дальше?

Раз так, нужно изобретать самостоятельно.

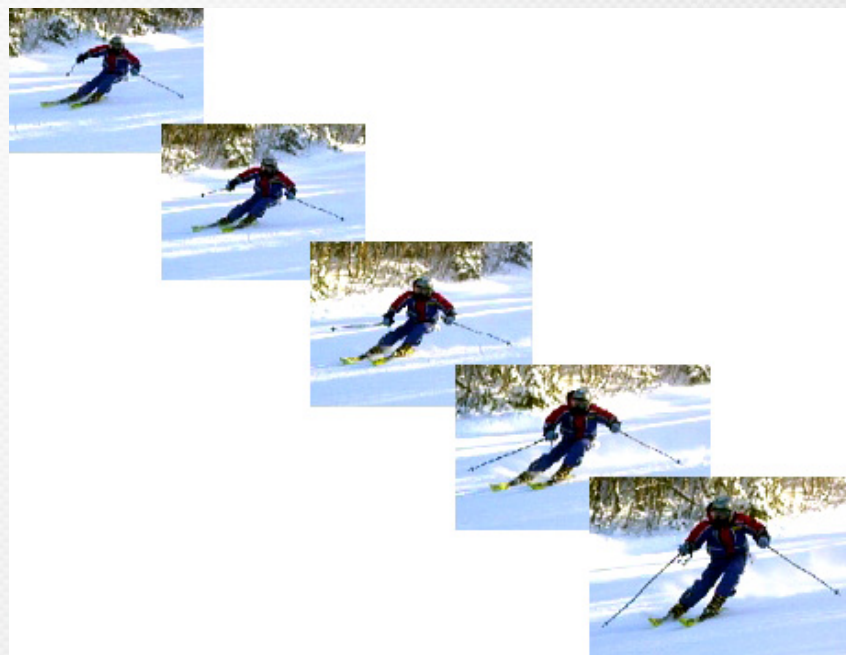
Но не с нуля.

Принципы кодирования видео

Кино — это ряд статических картинок, иллюзия движения в которых появляется благодаря тому, что они сменяют друг друга у нас перед глазами.

Принципы кодирования видео

Как и наши
черновики, эти
картинки (кадры)
связаны друг с
другом: во многом
похожи и в чём-то
различаются



Принципы кодирования видео

Любая картинка (в том числе и кадр) для компьютера — это набор точек



Принципы кодирования видео

Когда происходит оцифровка видео, нужно ли запоминать все точки во всех кадрах?

Мы имеем 24 кадра. Каждый размером 800x600 точек. Верно ли, что в памяти компьютера такое видео будет $800 \times 600 \times 24 = 11520000$ точек?

Принципы кодирования видео

Нет.

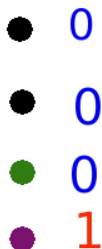
На самом деле, достаточно запомнить точки только в первом кадре, а для кадра 2, 3, ... n вычислять разницу с кадром 1.

Принципы кодирования видео

Если схематично. «0» означает отсутствие разницы, «1» означает разницу.



Кадр 1



Кадр 2



Кадр 3

Как работает эта аналогия?

В системе вариантов такие «точки» — это слова, части слов, пунктуация, положение относительно других частей текста. И разницу между ними тоже можно рассчитывать.

Как сделать, чтобы по-настоящему было хорошо?

Транскрипты
черновиков должны
храниться в базе
данных

Обращаясь к базе
программа должна
видеть варианты так,
чтобы уметь
рассчитывать разницу
между ними

На основе этой разницы
(«диффов») машина
смогла бы строить
гипотетическую модель
творческой истории
текста

Эта модель не
обязательно правильна
и не абсолютна, но это
помощь и подсказка
исследователю

Как сделать, чтобы по-настоящему было хорошо?

Разница между вариантами может быть визуализирована («диффы» выделены цветом). Это тоже компьютерная помощь

Конечно, не нужно терять из виду и графического воспроизведения рукописей. Zoom и линковку никто не отменял

Польза от «диффов»

Автоматика поможет избежать «простых» ошибок, когда текстолог не видит разницы между различающимися документами.

Каких ошибок не будет

«В комментариях в «Полному собранию сочинений и писем» Жуковского о публикации в «Собрании стихотворений, относящихся к незабвенному 1812 году» сообщено, что она «отличается как от первого, так и от второго изд<ания>, являясь промежуточным звеном между ними» (Жуковский В.А. Полное собрание сочинений и писем. Т. 1. С. 598), а тексты первого отдельного издания «Певца» и публикации в «Вестнике Европы» названы идентичными (Там же. С. 595). <...> названные идентичными тексты первого отдельного издания и публикации в «Вестнике Европы» различаются между собой не только отдельными строками, но и целыми строфами».

Бодрова А. Кто же был составителем «Собрания стихотворений, относящихся к незабвенному 1812 году»? // Новое литературное обозрение. 2012. № 6 (118). — С. 158—167.

Прототип

Вот пример того, как автоматическое сравнение вариантов может работать на черновиках пушкинского «Анчара»:

<http://nevmenandr.net/cgi-bin/anchar-sample.pl>

Прототип

Первая версия

СПЗ

В пустыне чахлой и скупой,
На почве, зноем раскаленной,
Анчар, как грозный часовой,
Стоит — один во всей вселенной.

Природа жаждущих степей
Его в день гнева породила,
И зелень мертвую ветвей
И корни ядом напоила.

Яд каплет сквозь его кору,
К полудню растопясь от зною,
И застывает ввечеру
Густой, прозрачною смолою.

К нему и птица не летит
И тигр нейдет — лишь вихорь черный
На древо смерти набежит
И мчится прочь уже тлетворный.

Вторая версия

Г. Второй черновой автограф (ПД 102).

В пустыне чахлой и скупой
На почве зноем раскаленной
Анчар как грозный часовой
Стоит один во всей вселенной

Природа пламенных степей
Его в день гнева породила
И жилы мощные ветвей
И корни ядом напоила —

Яд каплет сквозь кору его
К полудню растопясь от зною
И застывает ввечеру
Густой, прозрачною смолою, —

Кругом нет жизни — всё молчит
Недвижно всё — лишь вихорь черный
На древо смерти налетит
И мчится прочь уже тлетворный,

Варианты для сравнения

☐ Собрание сочинений в 10 томах. Т. 2. Москва, 1959

☐ Собрание в 10 томах. Москва,

[Показать
правки](#)

☐ В. Отдельный набросок (ЛБ 71, л. 22 об.)

☐ В. Отдельный набросок об.)

☐ Г. Второй черновой автограф (ПД 102).; слой правки: 1

☐ Г. Второй черновой автограф (ПД 102).; слой правки: 1

☐ СЦ
☐ СПЗ

☐ СЦ
☐ СПЗ

Отправить

<http://nevmenandr.net/cgi-bin/anchar-sample.pl>

Как сделать, чтобы по-настоящему было хорошо?

При этом инструмент должен представлять как можно меньше готовых решений, принятых человеком. В архитектуре должна быть предусмотрена эталонная творческая история от авторитетных текстологов.

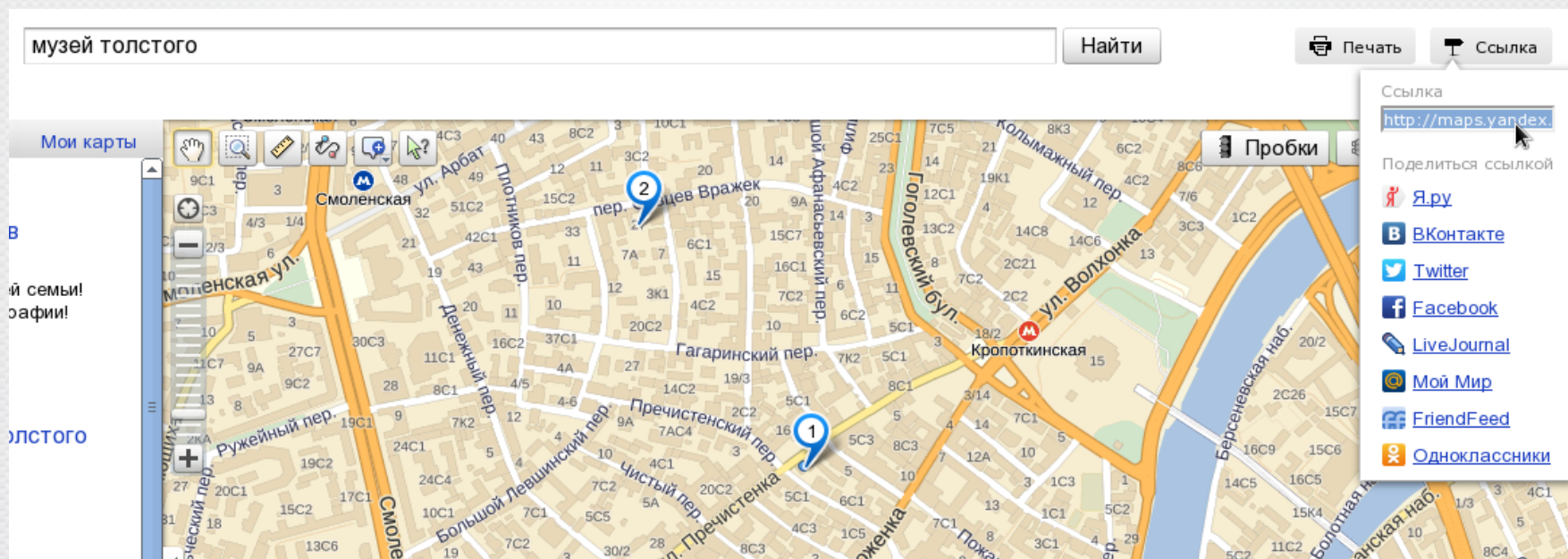
Но также и возможность для каждого пользователя самостоятельно реконструировать ход работы над рукописью.

Элементы интерфейса

Сервер должен запоминать
пользовательскую версию творческой
истории текста и затем восстанавливать
её в любой момент по уникальной
ссылке вроде тех, которые работают в
Яндекс.Картах

Элементы интерфейса

<http://maps.yandex.ru/-/CVV9BF5W>



Элементы интерфейса

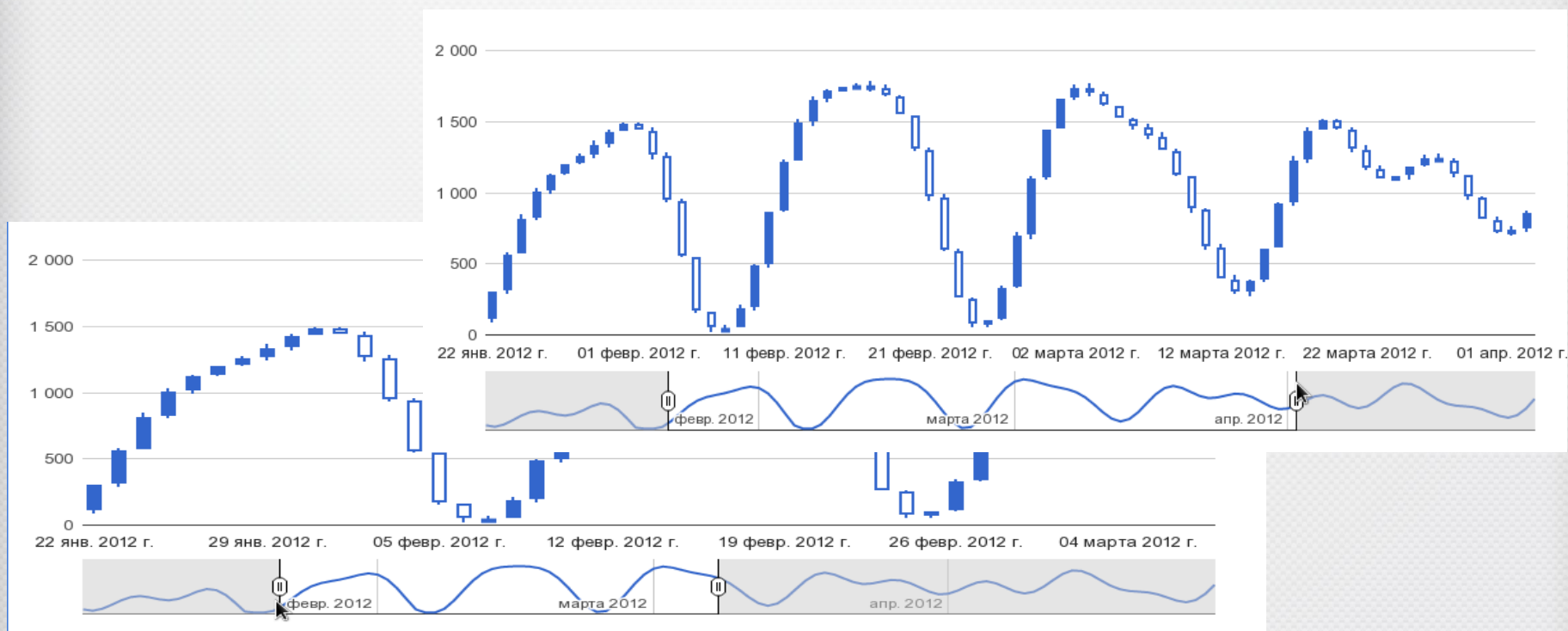
Творческая история должна быть
положена на timeline и связана с
управляющими элементами типа
«слайдер»



Элементы интерфейса

Хороший пример того, как это бывает:

https://code.google.com/apis/ajax/playground/?type=visualization#chartrangefilter_control



Принципиальные вопросы

Важно учесть нелинейный
характер развития текста,
древовидную структуру его
ЭВОЛЮЦИИ

Принципиальные вопросы

Важно учесть, что
текстологические гипотезы,
которые будет предлагать
машина, всего лишь гипотезы
в помощь исследователю, не
более