



Гуманитарные проблемы  
актуальных наук: цифровая дисциплина и  
проект

Михаил Шаповалов  
магистрант 2 курса ВШЭ-Пермь

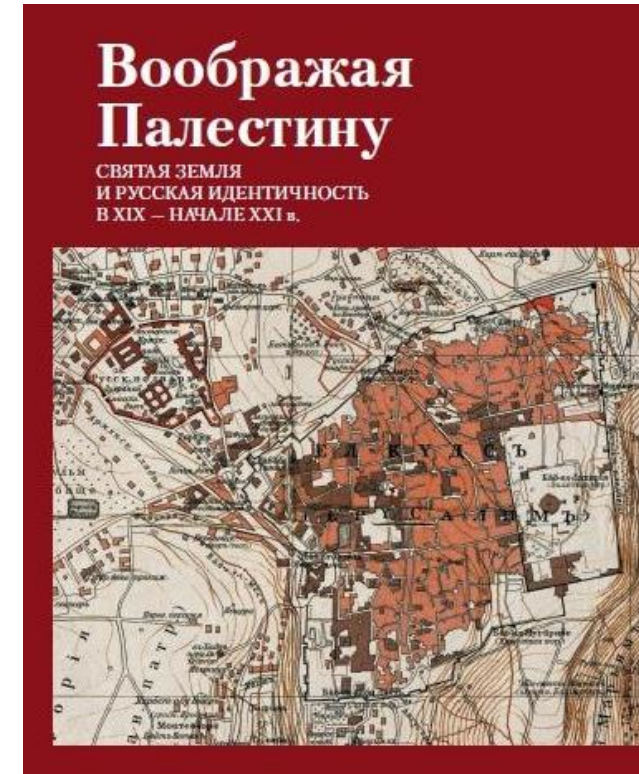
15 апреля 2024 г.  
ИТМО

# Паломнические записки в журналах епархиальных ведомостей XIX — начала XX вв.: от формирования полного корпуса к сетевому анализу

Доклад подготовлен в ходе проведения исследования (№24-00-034 «Паломничество на страницах епархиальных ведомостей во второй половине XIX – начале XX вв.: база данных, источниковедческий и сетевой анализ») в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)»

## Почему вы стали работать над этой темой?

- 1) На сегодняшний момент не собран полный корпус паломнических записок в базе Епархиальных ведомостей.
- 2) Исследователи используют лишь незначительную, избранную часть текстов, но на основе которой делают выводы о всей паломнической практике.
- 3) Вопрос на сколько на сегодняшний день мы располагаем релевантными данными о паломничестве на Святую землю, остается открытым.
- 4) Потребность в цифровом повороте в исследовании паломничества и палестиноведения (палестиноведение 3.0, термин М.В. Каиль).



Воображая Палестину : Святая земля и русская идентичность в XIX - начале XXI в / М. С. Шаповалов, А. Ю. Бокатов, А. А. Валитов [и др.]. Санкт-Петербург : Нестор-История, 2021. 776 с.



## Что я делаю?

(1) на основе выявленного полного комплекса паломнических текстов из Епархиальных ведомостей (2) провести комплексное источниковедческое исследование (3) с применением сетевого анализа и цифровых методов обработки информации.

## Гипотеза и методы

полный комплекс паломнических текстов из Епархиальных ведомостей позволит (1) проводить сетевой и стилометрический анализ текстов, существенно (2) расширит возможности источниковедческого анализа паломнических текстов, (3) изучения социокультурного портрета авторов записок. Созданный **корпус паломнических текстов** станет одновременно **и результатом, и новым цифровым инструментом исследования.**



## Как формировался корпус ?

1) Выпуски Епархиальных ведомостей (63 епархии, с 1860-х гг. по 1919 г., **более 60 тыс. выпусков**).

2) Технический этап. Для работы с базой данных оцифрованных журналов ЕВ. Был использован пакет Python для анализа документов HTML и XML - BeautifulSoup. Поиск в текстах проходил по ключевым словам.

3) «Ручной этап» работы. Алгоритм выявляет от нескольких **сотен до четырех тысяч** записей, которые необходимо просматривать в ручную (около 189 тыс. записей).

📁 Проект: "Паломнические записки в региональных Епархиальных ведомостях".ipynb"" ☆  
Файл Изменить Вид Вставка Среда выполнения Инструменты Справка [Последнее изменение: 15 июня](#)

+ Код + Текст

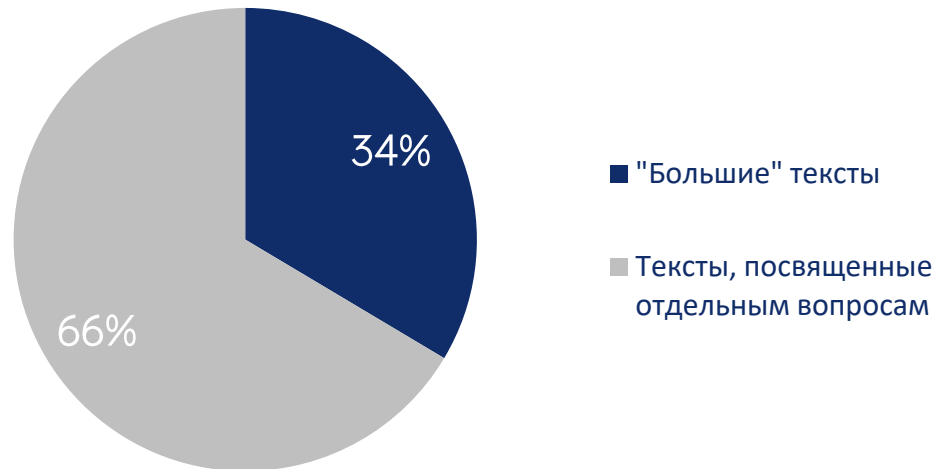
```
[ ] keywords = ['палом', 'русал', 'палес']  
article = []  
  
for i in os.listdir():  
    if i != '.ipynb_checkpoints':  
        text = []  
        reader = PdfReader(i)  
        number_of_pages = len(reader.pages)  
        for p in range(len(reader.pages)):  
            page = reader.pages[p]  
            t = page.extract_text()  
            text.append(t)  
        for j in text:  
            for w in keywords:  
                if w in j.lower():  
                    t = j  
                    t = t.replace('\n', '')  
                    l = []  
                    l.append(i)  
                    l.append(p)  
                    l.append(t)  
                    article.append(l)
```

**Важно:** огромная благодарность Анне Васильевне Сениной, без которой эта работа была бы невозможна!!!



## Какой получился корпус паломнических текстов?

### Корпус записок



### Уникальность корпуса





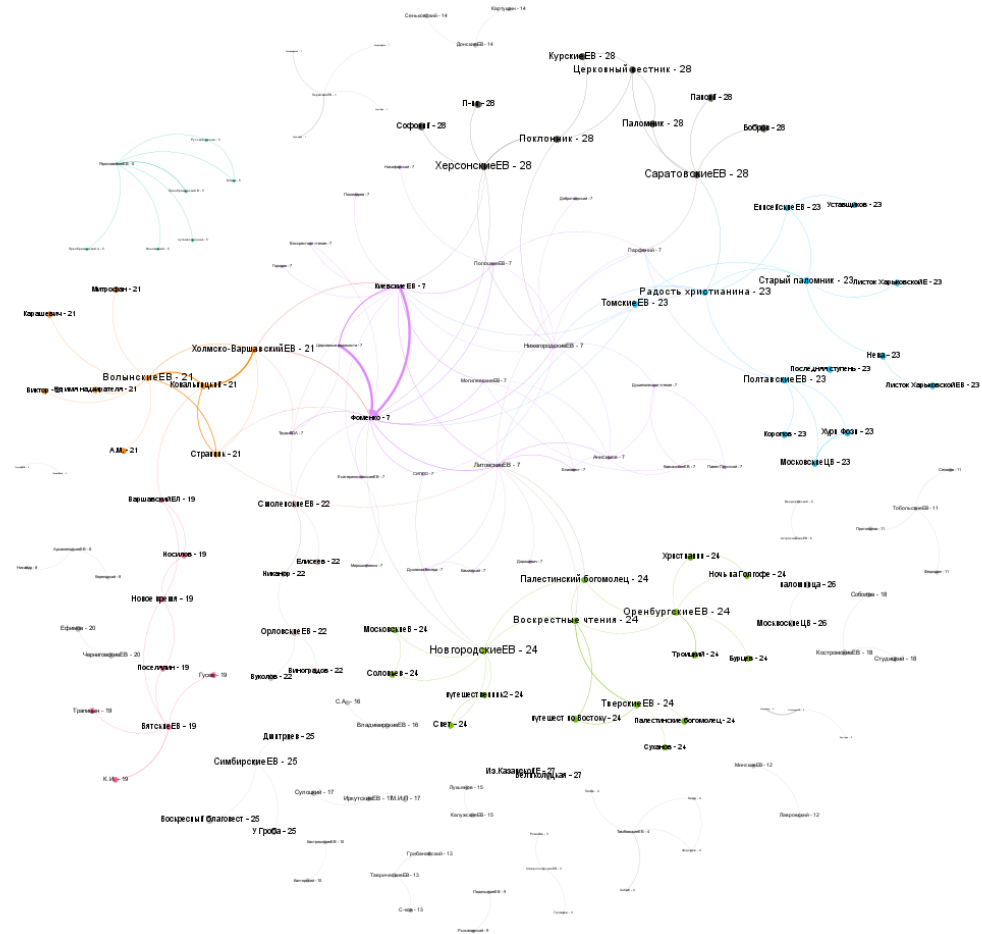


## Сетевой анализ

1) Подтверждается атомарность изданий паломнических текстов в региональных ЕВ.

2) Сеть позволяет увидеть:

- выделить центры формирования регионального паломнического контента (взвешенная степень);
- связи между разными редакциями (модулярность);
- посредников через которых распространялись тексты, ЕВ перепечатывали материалы из других журналов: СИППО, Новое время, Нива, Душеполезное чтение и др.) (центральности).



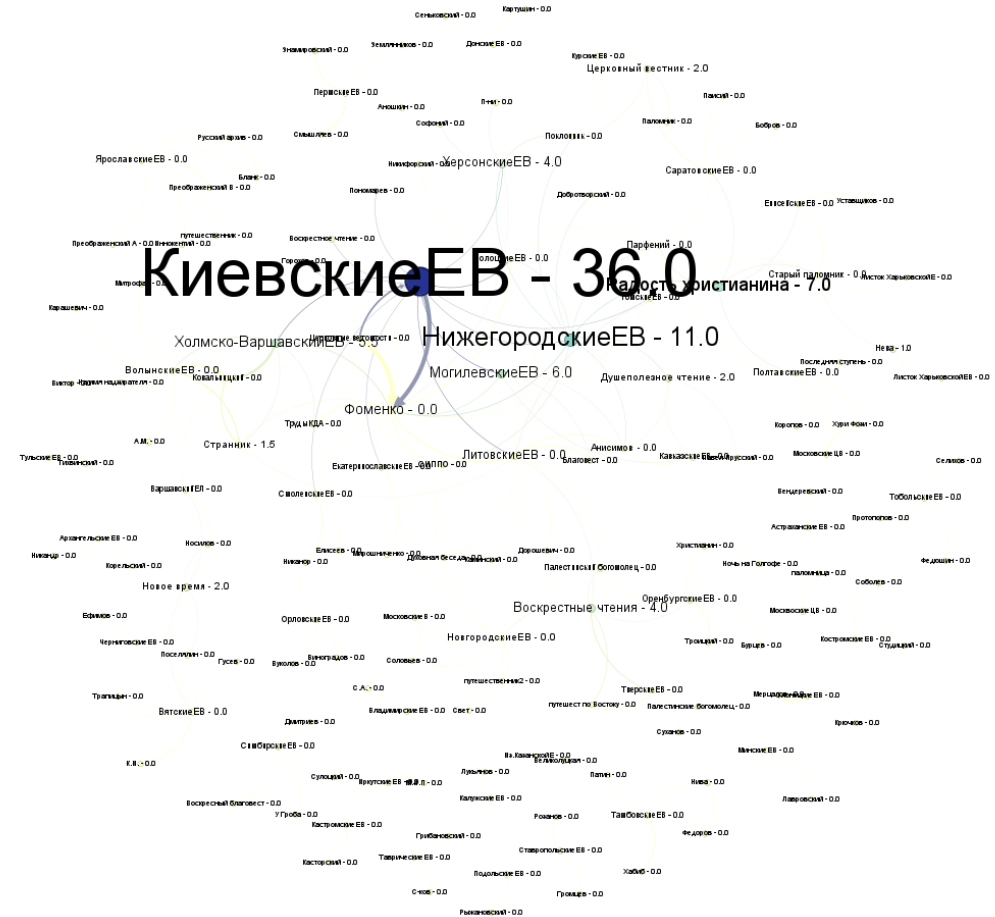


## Сетевой анализ

1) Подтверждается атомарность изданий паломнических текстов в региональных ЕВ.

2) Сеть позволяет увидеть:

- выделить центры формирования регионального паломнического контента (взвешенная степень);
- связи между разными редакциями (модулярность);
- посредников через которых распространялись тексты, ЕВ перепечатывали материалы из других журналов: СИППО, Новое время, Нива, Душеполезное чтение и др.) (центральности).



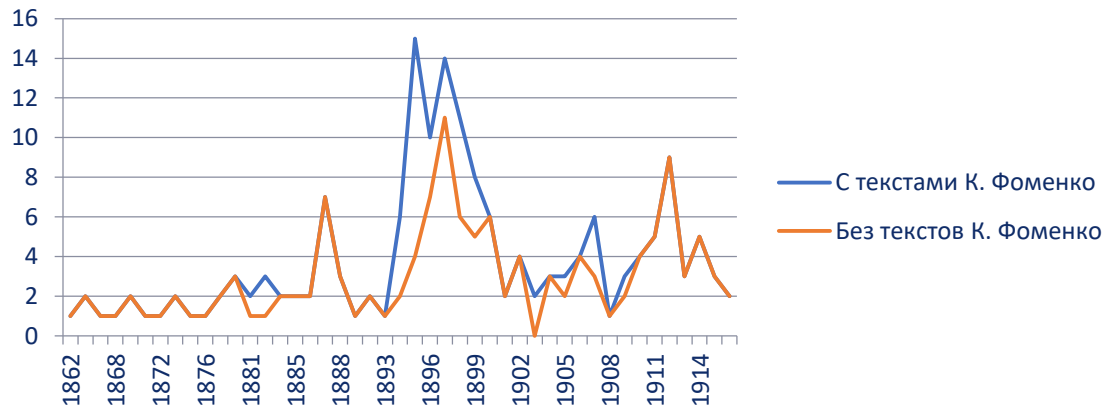


## Источниковедческий анализ паломнических текстов

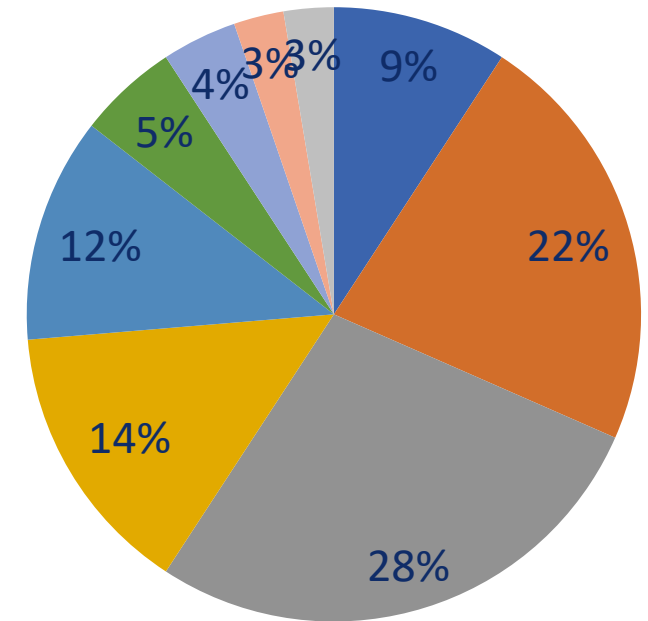
**Ковальницкий:** 1 основной текст в 16 выпусках + 11 отдельных текстов по мотивам основного текста).

**Никанор:** 4 текста = 1 воспоминание о путешествии на Иордан, 2 беседы о путешествии в Вифлеем (Смоленские ЕВ) + 1 воспоминания о путешествии в Египет (Орловские ЕВ).

Феномен протоирея К. Фоменко – 23 текста и 22 копии = **45 текстов**



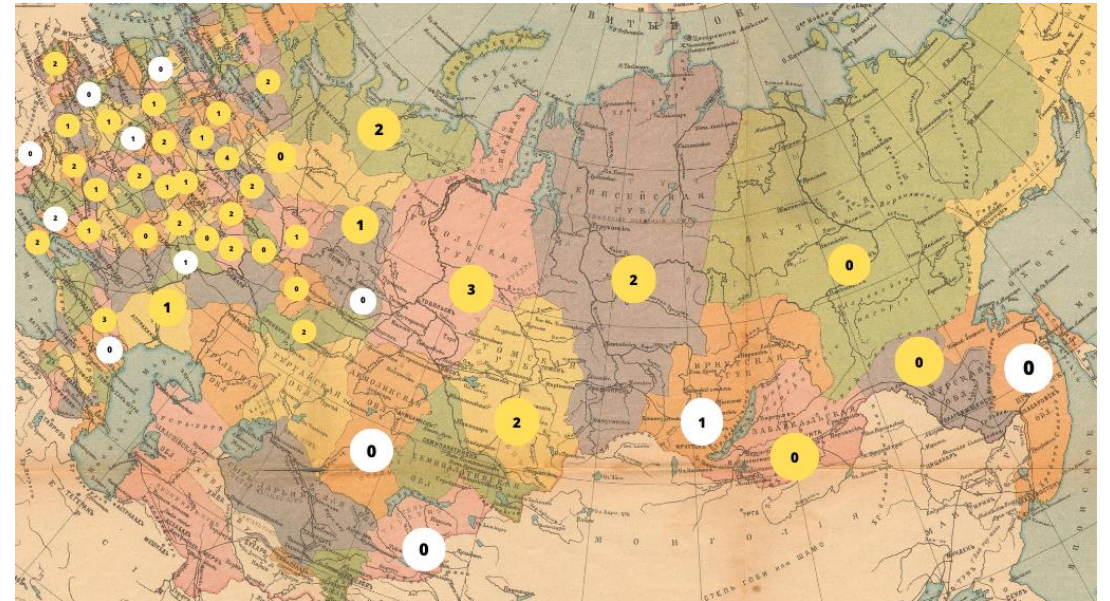
### Жанры



- Письма
- Записки
- Воспоминания
- Дневники
- Заметки
- Рассказ
- Очерк
- Стихотворение
- Доклад

## С какой проблемой/проблемами вы столкнулись?

- 1) Алгоритм на Python нашёл не все паломнические тексты (потребовалась перепроверка).
- 2) Сети оказались достаточно атомарными («некрасивые» сети).
- 3) Необходима дополнительная визуализация сети с привязкой к карте (R).
- 4) Как посчитать тексты (сложные тексты, копии это отдельный текст или нет? ).
- 5) Что мы относим к паломническим запискам. Источниковедческая проблема жанрового определения.



## Что было самое неожиданное для вас при работе?

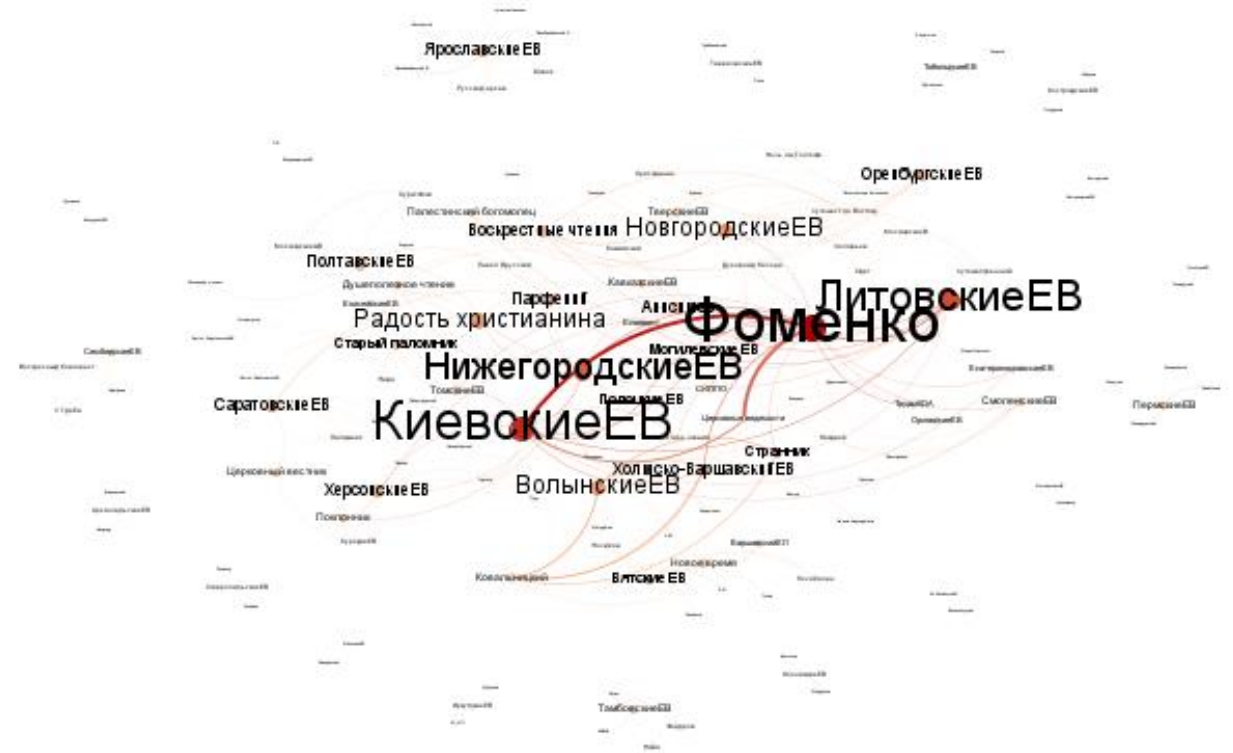
На формировании корпуса (машинный и «ручной» отбор) ушло около года рабочего времени.

### Результаты

Небольшие открытия: священник К-н и протоирей Фоменко один и тот же человек; красноярский паломник Уставщиков и воронежский паломник Розанов путешествовали в составе одной группы.

### Предварительные выводы

С точки зрения сетевого анализа нас интересуют взаимосвязанные тексты, те что собственно и создают сеть. А с точки зрения исследования паломничества, наиболее интересны уникальные тексты, которые в сети максимально удалены от других.

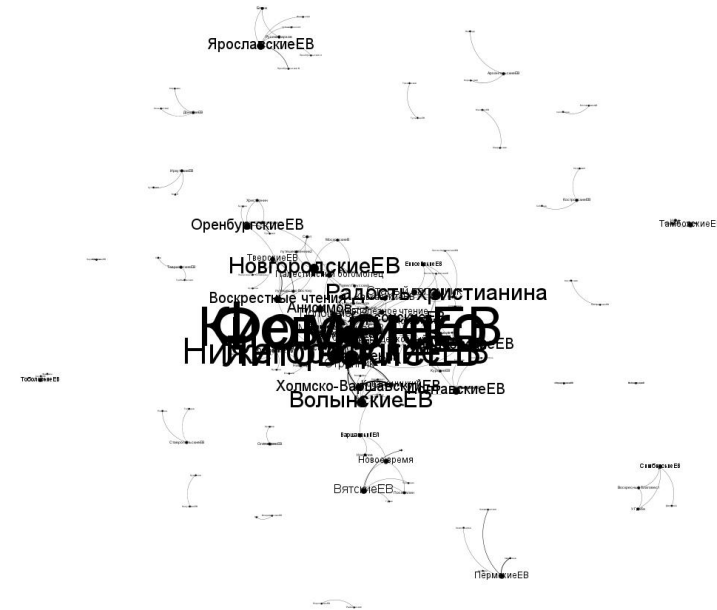


Почему тот способ проведения работы  
который выбрали вы считается самым  
приемлемым?

- 1) Это пошаговый путь наполненный ошибками.
- 2) Позволяет перейти от «пристального» к  
«дальному» чтению (Франко Моретти).

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-6-3da3a9329350> in <cell line: 2>()
      1 # page0 = requests.get(urls[0-19])
----> 2 page0.status_code

NameError: name 'page0' is not defined
```



Что хотелось бы обсудить с аудиторией?

- 1) Что делать, если цифровые методы не дают  
дополненные данные?
- 2) Как можно использовать сети внутри текста и  
какие дополнительные инструменты можно  
применить для работы внутри паломнических  
записок (токен анализ).