

**Документация, содержащая описание функциональных характеристик программного обеспечения «PROMT Analyzer SDK» и информацию, необходимую для установки и эксплуатации программного обеспечения**

## Содержание

---

Назначение и состав программного обеспечения «PROMT Analyzer SDK» .....	4
Общие сведения и понятия .....	4
Сущности.....	4
Пользовательские сущности .....	5
Актант .....	5
Факты.....	5
Системные требования программного обеспечения .....	5
Серверная часть.....	5
Клиентская часть (веб-интерфейс).....	6
Процедура установки.....	6
Функциональные возможности .....	11
Пользовательский интерфейс .....	12
Выполнение анализа текста .....	13
Тип текста .....	14
Поле "Результат разбора".....	14
Поле "Образ документа".....	15
Взаимосвязь полей "Результат разбора" и "Образ документа" .....	17
Сохранение образа документа.....	18
Программный интерфейс .....	18
Общие сведения о сущности .....	18
Свойства сущности .....	19
Свойства факта .....	25
Свойства актанта.....	26
Пример разбора (JSON) .....	27
Аутентификация .....	29
Метод Login (специальный сервис аутентификации) .....	29
Описание .....	29
URI.....	29
Входные данные .....	30
Методы веб-сервиса .....	30
Метод Types .....	30
Описание .....	30
URI.....	30
Входные данные .....	30
Выходные данные .....	30
Метод Subtypes .....	31
Описание .....	31
URI.....	31
Входные данные .....	31
Выходные данные .....	31
Метод Profiles .....	31
Описание .....	31
URI.....	31
Входные данные .....	31
Выходные данные .....	31
Метод Languages .....	32
Описание .....	32
URI.....	32
Входные данные .....	32
Выходные данные .....	32
Метод Localizations .....	32
Описание .....	32
URI.....	32
Входные данные .....	32
Выходные данные .....	32

Метод ExtractEntities .....	32
Описание .....	32
URI.....	32
Входные данные .....	33
Выходные данные .....	33
Метод Analyze.....	33
Описание .....	33
URI.....	33
Входные данные .....	34
Выходные данные .....	34

## **Назначение и состав программного обеспечения «PROMT Analyzer SDK»**

Программное обеспечение «PROMT Analyzer SDK» (далее – «Анализатор», «PROMT Analyzer») представляет собой масштабируемое серверное программное обеспечение для анализа текстов на естественных языках с целью поиска, извлечения и обобщения информации о сущностях, фактах, событиях и их связях, путем лингвистического анализа соответствующих текстов с учетом синтаксиса и семантики.

Программное обеспечение «Анализатор» разработано для использования в различных информационных и аналитических системах, предназначенных для сбора и систематизации разнородной информации, построения аналитических отчетов, в тех случаях, когда одним из объектов анализа являются текстовые документы любых видов и форматов.

Использование программного обеспечения «Анализатор» предполагает обращение к его функциям через кросс-платформенный программный интерфейс, реализованный как web-сервис.

Преимущества программного обеспечения «Анализатор» определяются методом, лежащим в его основе - это глубокий синтактико-семантический разбор предложения, как единицы языка, текста, как единого комплекса составляющих его предложений, а также наличием словарной базы, содержащей несколько миллионов слов с приписанными им синтаксическими и семантическими признаками.

Программное обеспечение «Анализатор» обеспечивает высокую точность извлечения нужной информации и позволяет разработчикам информационных и аналитических систем создавать современные программные решения для бизнес-аналитики и интеллектуального поиска, а также автоматизации и поддержки бизнес-процессов.

Программное обеспечение «Анализатор» содержит следующие компоненты:

- сервер приложения, распределяющий нагрузку между серверами перевода и предоставляющий как визуальный, так и программный интерфейсы для обращения;
- один или несколько серверов, осуществляющих непосредственно анализ текста.

Все эти компоненты представляют собой ASP.NET веб-приложения, которые работают под управлением веб-сервера Microsoft Internet Information Services (IIS).

Для хранения данных используется также система управления базами данных DBS, являющаяся собственной разработкой правообладателя (ООО «ПРОМТ»).

## **Общие сведения и понятия**

### **Сущности**

«Анализатор» получает на вход фрагмент текста, анализирует его и выделяет из текста следующие именованные сущности:

- персоны (имена собственные), т.е. упоминания конкретных людей с фамилиями, именами и отчествами, если они находятся в тексте. Кроме этого, для найденных персон «Анализатор» выделяет ряд атрибутов, если информация о них присутствует в тексте (должность, профессия, родственные связи, партийная принадлежность, национальность и т.п.)

- названия организаций – имена собственные. Для найденных организаций выделяется тип (ОАО, ООО и т.п.) и подтип организаций (производственная организация, банк и т.п.), если информация о них присутствует в тексте или в онтологиях программного обеспечения «Анализатор»

- географические названия с указанием типа (страна, город и т.п.) и подтипа (часть света, водоем и т.п.), если информация о типе и подтипе имеется в онтологиях программного обеспечения «Анализатор»

- геополитические сущности – географические названия в организационном контексте (Москва заявила, требование Анкары) с указанием типа и подтипа, если информация о типе и подтипе имеется в онтологиях программного обеспечения «Анализатор»

- другие именованные сущности разных типов, найденные при разборе текста (события, устройства, документы и т.д.) с указанием типа и подтипа, если информация о типе и подтипе имеется в онтологиях программного обеспечения «Анализатор».

«Анализатор» выделяет сущности указанных выше типов не только в случае, когда они известны данному программному обеспечению (т.е. имеют соответствующие коды в ее внутренней базе данных), но и когда они представляют собой ранее неизвестные ей объекты.

### **Пользовательские сущности**

Кроме персон, организаций и географических названий, в «Анализатор» предусмотрена работа с так называемыми пользовательскими сущностями. Для выделения пользовательских сущностей в «Анализатор» можно загрузить пользовательскую онтологию, описывающую трехуровневую классификацию, например, буровых установок (мобильные, стационарные, принадлежность, расположение и т.п.). После этого «Анализатор» начнет их выделять аналогично сущностям, указанным выше.

### **Актант**

Актант— активный, значимый участник ситуации, конструкция в тексте, заполняющая семантическую или синтаксическую структуру предиката.

### **Факты**

Для каждой сущности «Анализатор» анализирует и выделяет факты - действия и события, в которых зафиксировано участие сущности. Для каждого выделенного факта строится фрейм, описывающий действие (кто, что, где, когда и т.п.) или событие с выводом других найденных сущностей, являющихся актантами в этом действии.

## **Системные требования программного обеспечения**

### **Серверная часть**

Системные требования к конфигурации компьютера для установки серверной части программного обеспечения «Анализатор»:

- процессор уровня Intel Core i5 или выше;
- оперативная память – не менее 4 Гб;
- свободное место на диске – не менее 10 Гб.

Требования к программной среде на серверной стороне:

- операционная система:
  - Microsoft Windows Server 2016;
  - Microsoft Windows Server 2012;
  - Microsoft Windows Server 2008.
- Microsoft Internet Information Services 6.0 или выше (в IIS должны быть разрешены Web Server Extensions: Active Server Pages и ASP.NET);
- .NET Framework 4.0 или выше.

## Клиентская часть (веб-интерфейс)

Системные требования к конфигурации компьютера, выступающего клиентом «Анализатор» и использующего только веб-интерфейс программного обеспечения «Анализатор»:

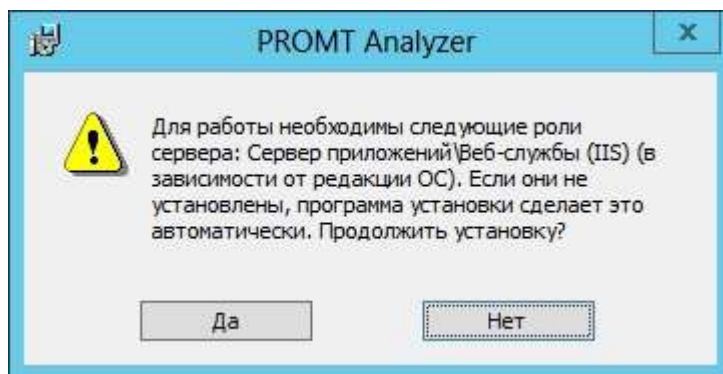
- Microsoft Internet Explorer версии 8 и выше;
- Microsoft Edge;
- Mozilla Firefox версии 24 и выше;
- Google Chrome версии 25 и выше.

## Процедура установки

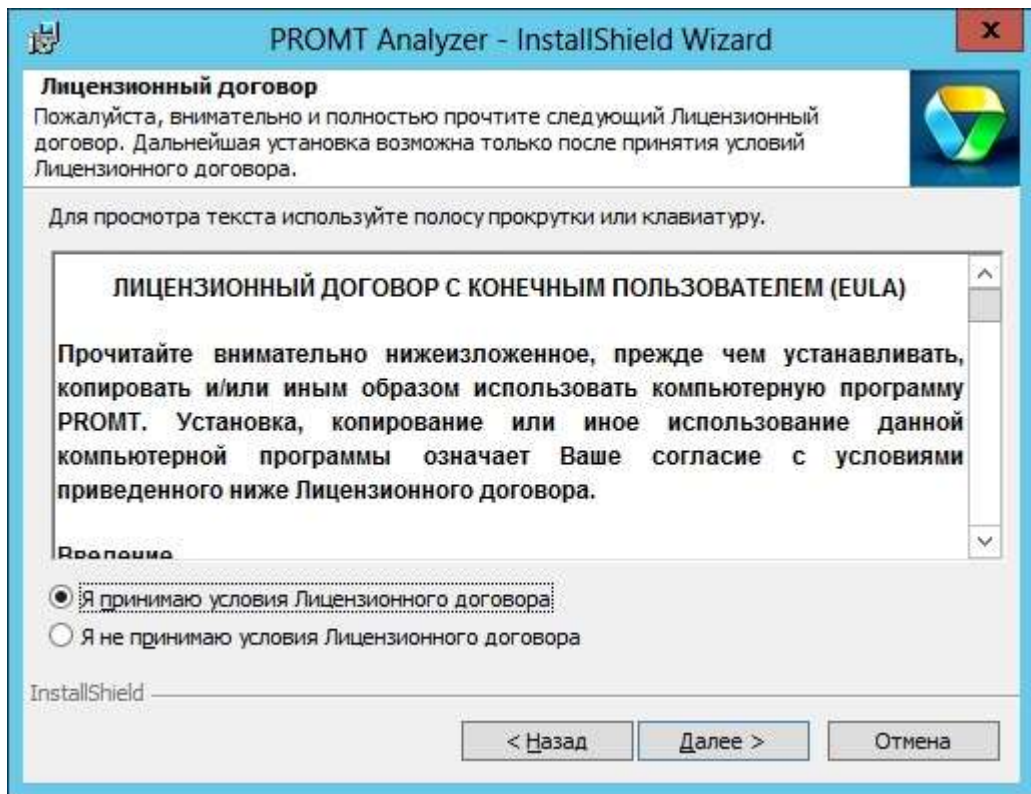
Для запуска процесса инсталляции программного обеспечения «Анализатор» необходимо запустить файл setup.exe из состава соответствующего инсталляционного набора указанного программного обеспечения.

Для корректной работы «Анализатор» на сервере, где производится инсталляция данного программного обеспечения, требуется наличие определенного набора пререквизитов. Мастер установки самостоятельно определит их наличие в используемой операционной системе и в случае необходимости предложит доустановить недостающие компоненты – необходимо согласиться и установить указанные пререквизиты. После этого мастер установки перейдет к основному процессу инсталляции «Анализатор»:

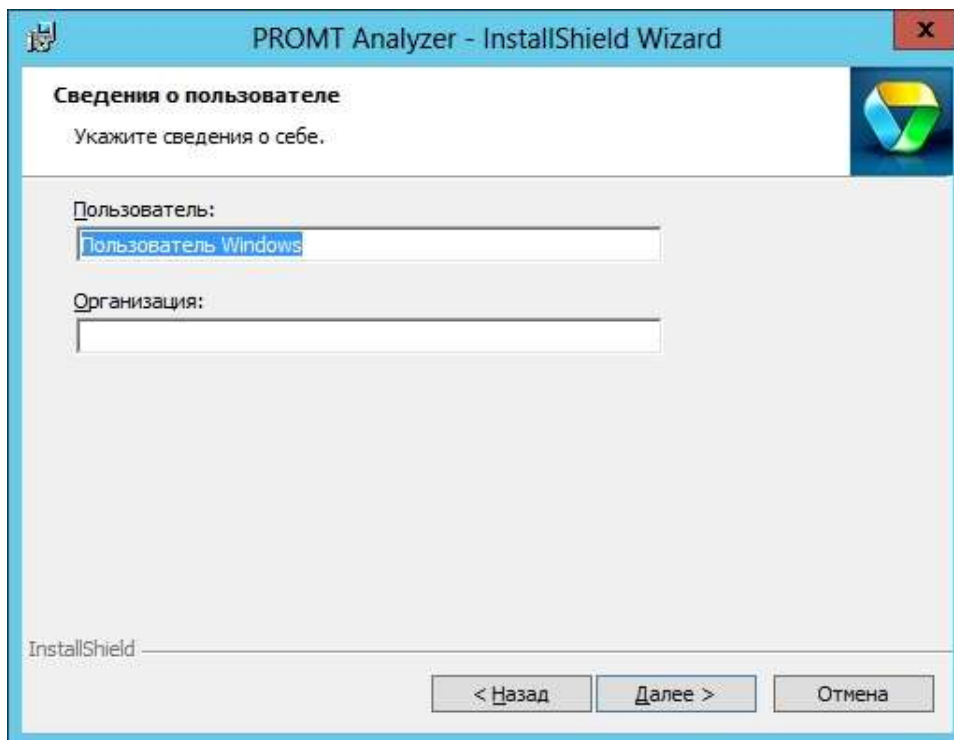
1. Будет предложено включить необходимые роли сервера – нужно нажать кнопку «Да»:



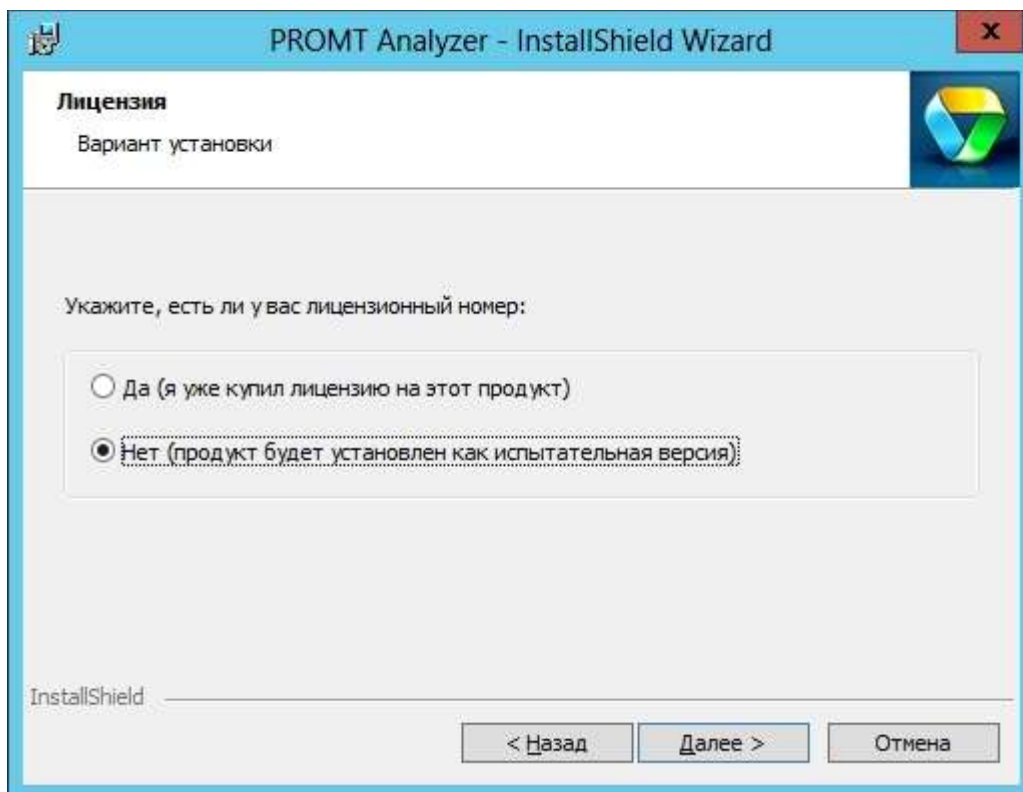
2. Будет предложено принять условия Лицензионного договора – нужно выбрать «Я принимаю» и нажать кнопку «Далее»:



3. Далее необходимо указать Ваши пользовательские данные (опционально):

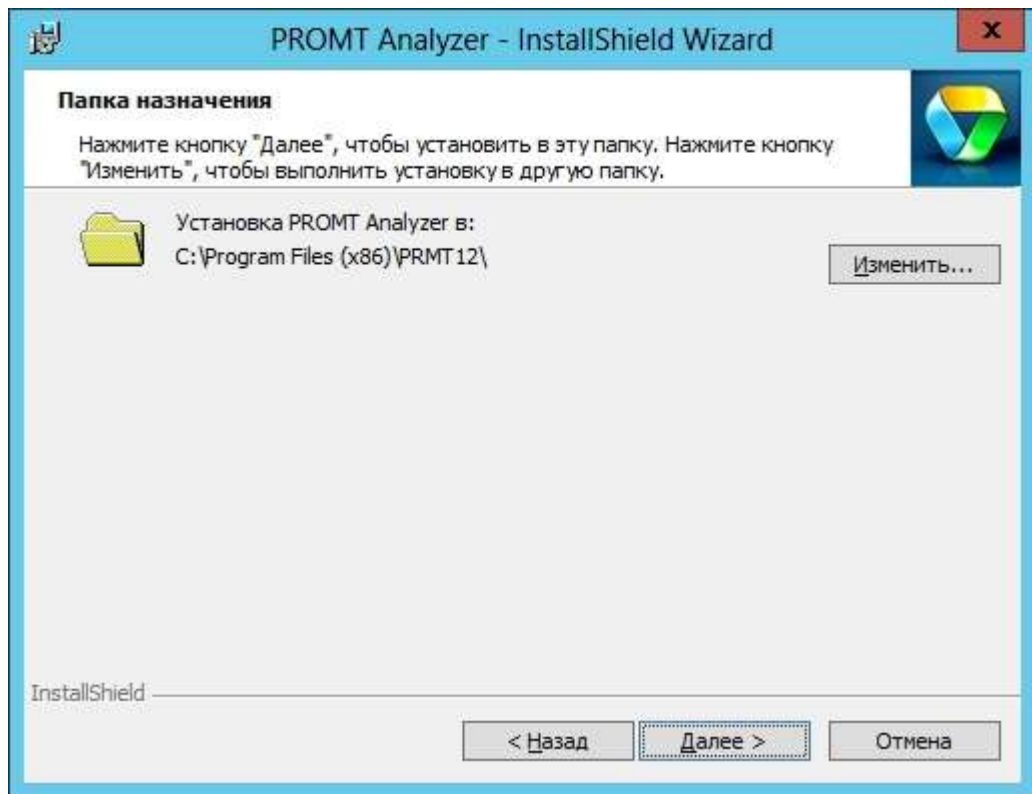


4. На этапе выбора варианта установки в случае, если у Вас еще нет лицензионного номера для продукта, выберите вариант установки испытательной версии:

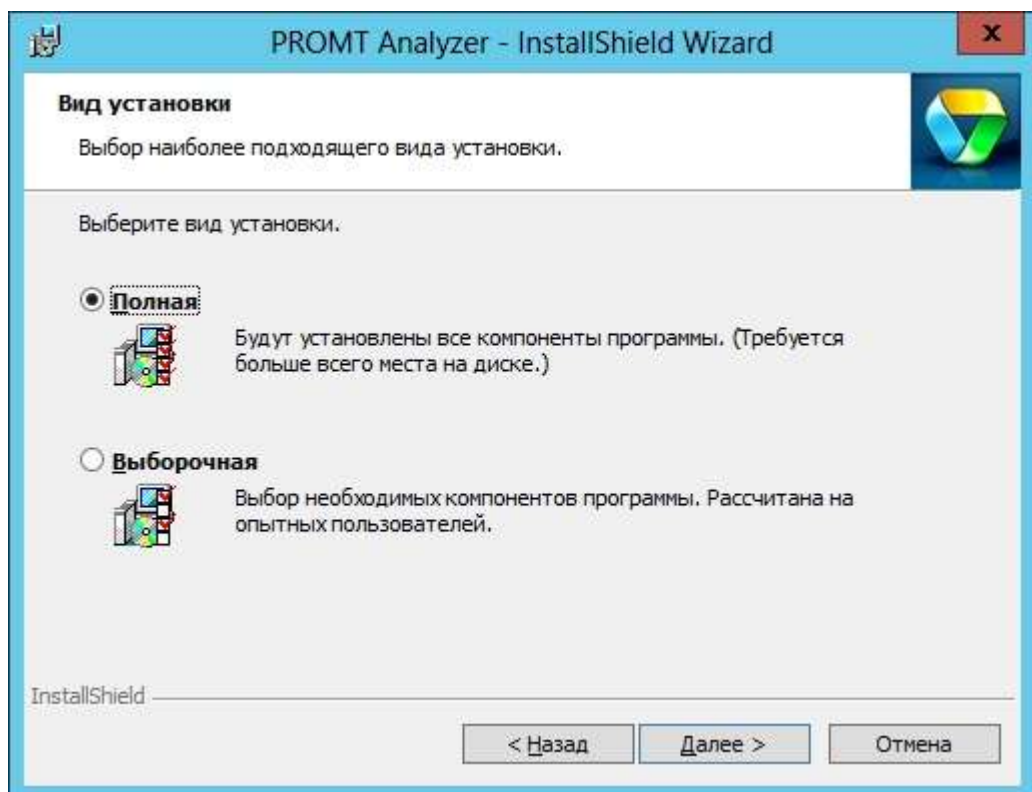


5. На этапе выбора каталога установки можно воспользоваться умолчательным вариантом и нажать кнопку «Далее»:

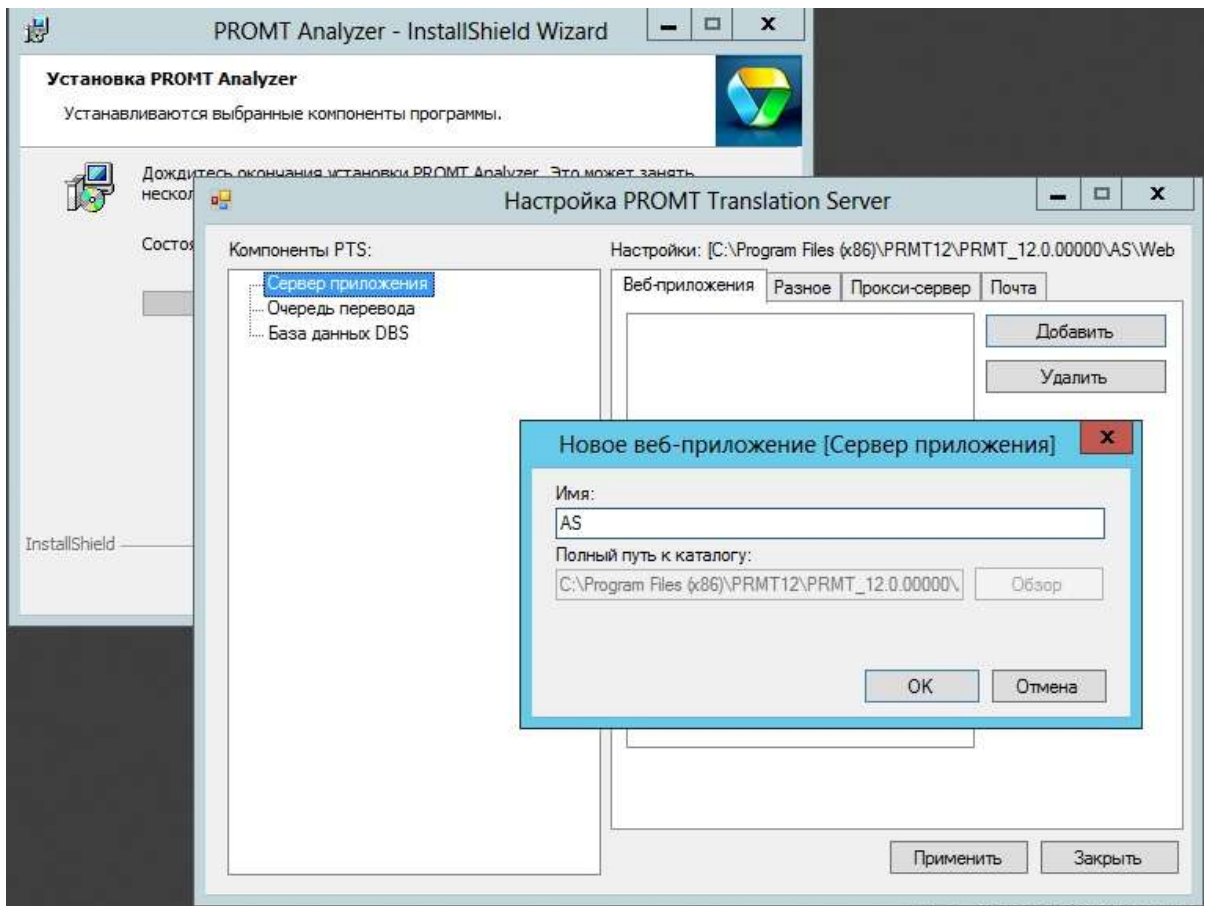




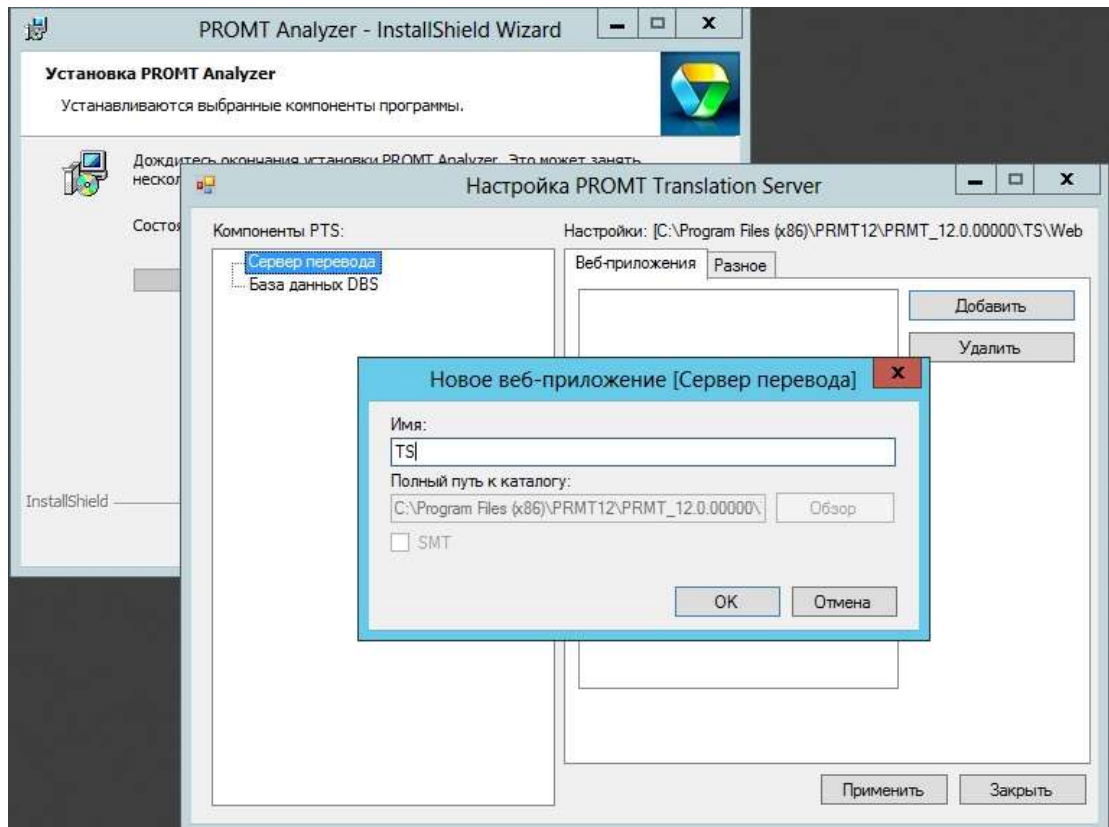
6. Далее выберите вид установки «Полная», нажмите кнопки «Далее» и «Установить»:



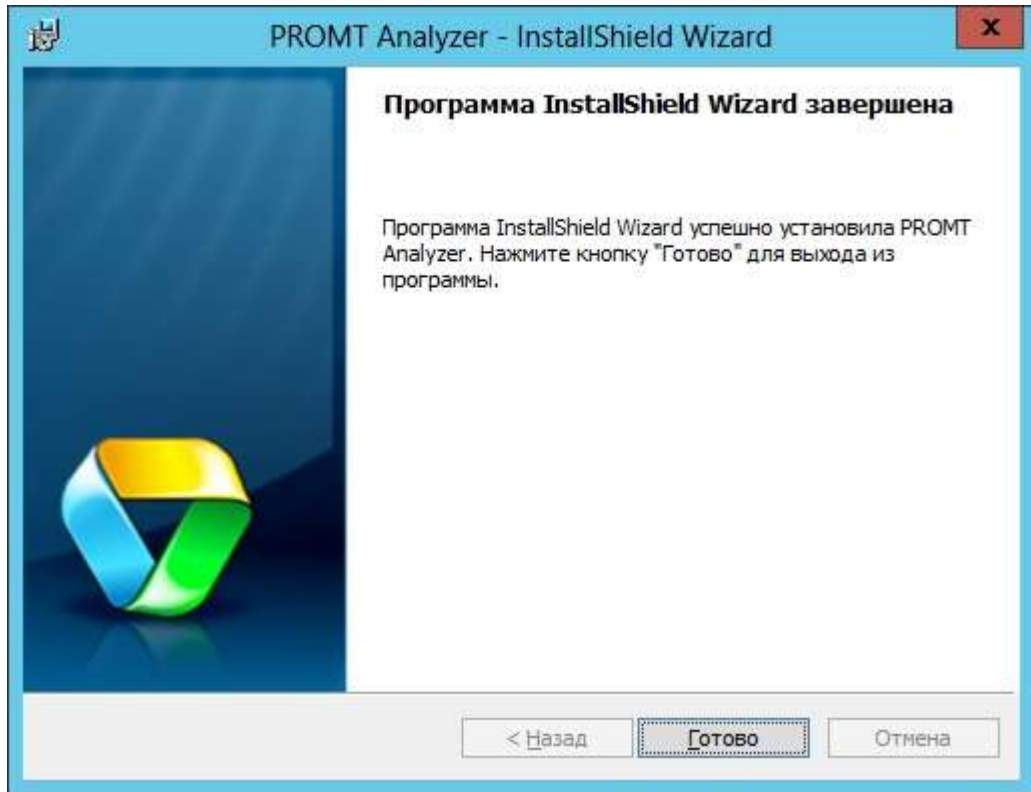
7. В конце установки запустится мастер настройки приложений и на первом этапе будет необходимо создать\добавить веб-приложение для Сервера приложения (по умолчанию предлагается имя AS) – последовательно нажать кнопки «Добавить», «Ок» в окне «Новое веб-приложение» и «Закреть» для перехода к следующему этапу. Прочие настройки можно оставить без изменений.



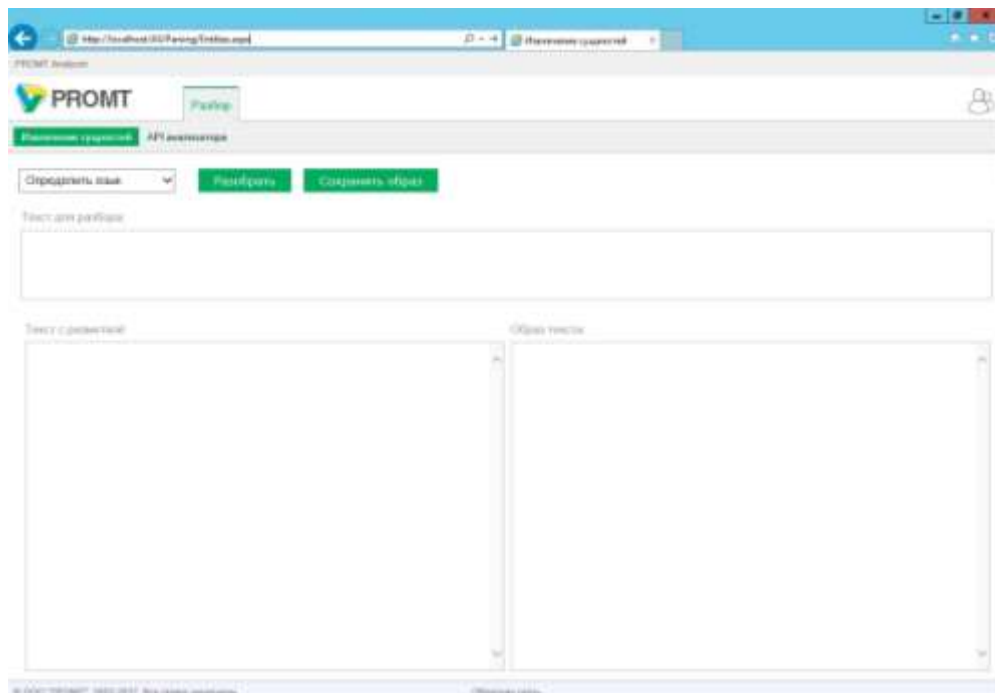
8. На втором этапе будет предложено создать/добавить веб-приложение для Сервера перевода (по умолчанию предлагается имя TS) – последовательно нажать кнопки «Добавить», «Ок» в окне «Новое веб-приложение» и «Закреть» для перехода к завершению общей установки. Прочие настройки можно оставить без изменений.



9. На последнем этапе вы получите сообщение об успешном окончании установки «Анализатор»:



10. Проверьте в браузере доступность созданного веб-приложения Сервера приложения «Анализатор»:



## Функциональные возможности

«Анализатор» обеспечивает следующие функции:

- извлечение из текстовых документов на различных языках сущностей, а именно: персон, географических названий, организаций и других именованных сущностей; нахождение новых (ранее не известных и не зафиксированных в программном обеспечении «Анализатор») сущностей;
- возможность раздельного выделения географических названий, используемых для указания физических локаций, и географических названий, используемых в организационном контексте (*находиться в России / Россия приняла решение*);
- объединение упоминаний одних и тех же сущностей в рамках одного документа с учетом синонимов, кореферентных связей и анафоры;
- выделение атрибутов сущностей: ФИО, должность, национальность и т.д. для персон; геотипа для географических названий; типа организаций и т.п.;
- выделение фактов (действий и событий), связанных с выделенными сущностью;
- определение актантной структуры факта с выделением субъекта, объекта, адресата, места, времени, причины и других косвенных связей;
- поиск среди найденных актантов других именованных сущностей и установка связей факта с ними;
- создание пользователем сущностей новых по отношению к предопределенным в «Анализатор» сущностям;
- возможность учета тематики и типа документа (например, информационные сообщения / резюме и т.п.), а также источника текста (например, распознанная речь / письменный документ) при анализе документа;
- возможность тонкой настройки на предметную область по требованию пользователя;
- возможность анализа текста на русском, английском, немецком, арабском, китайском и других языках (всего 20 языков);
- выделение названия сущности на русском языке и языке оригинала при обработке документов на иностранных языках
- автоматическое определение языка анализируемого документа следующих форматов: plain text в различных кодировках, RTF, HTML, XML, PDF, DOC(X), XLS(X), PPT(X);
- автоматическое определение формата анализируемого документа;
- пользовательский интерфейс для интерактивного анализа документов;
- API для интеграции с информационно-аналитическими или поисковыми системами;
- возможность масштабирования вычислительных ресурсов при изменении нагрузки «на лету»;
- возможность использования как в виде облачного сервиса (Cloud), так и развертывания на площадке пользователя.

## **Пользовательский интерфейс**

Пользовательский интерфейс «Анализатор» изображен на Рисунке 1.

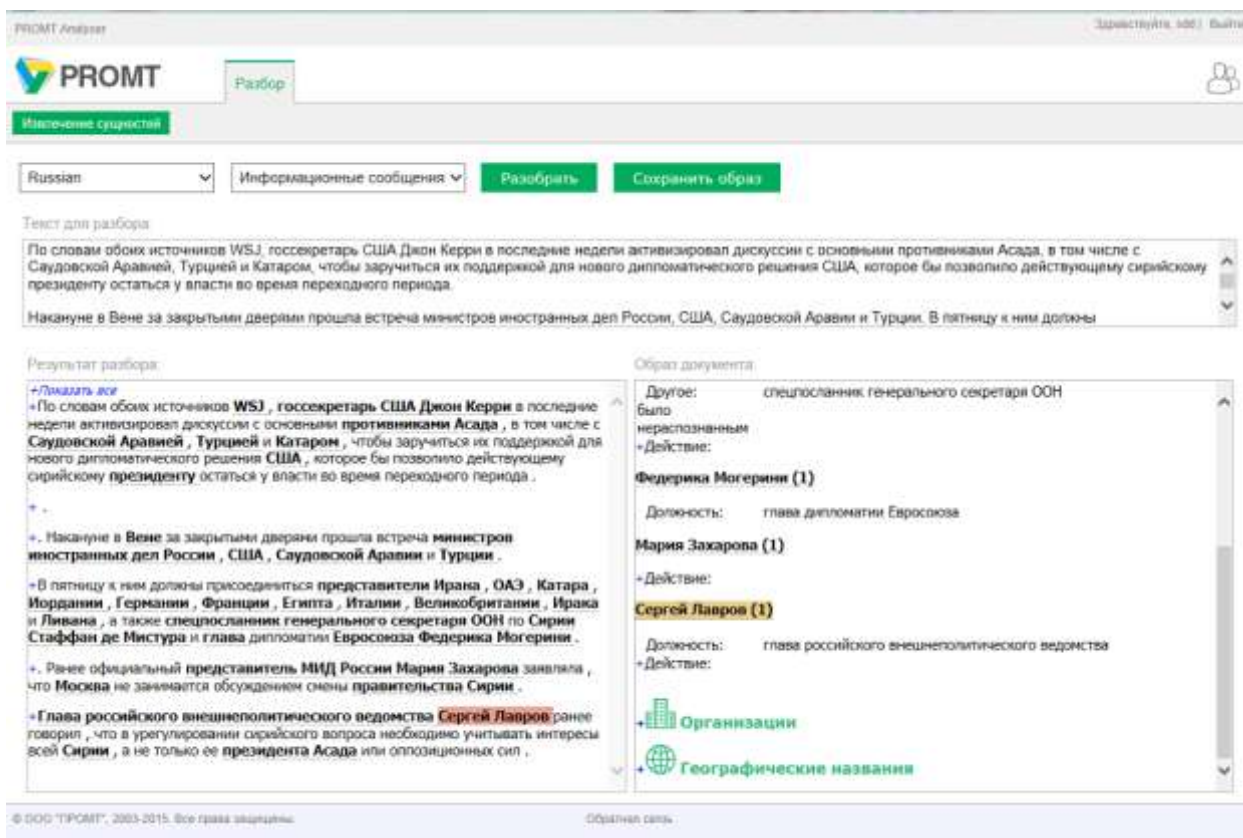


Рисунок 1

Интерфейс «Анализатор» состоит из следующих элементов:

1. Выпадающий список для выбора языка анализируемого текста (поддерживаются русский и английский языки)
2. Выпадающий список для типа анализируемого текста
3. Кнопка «Разобрать» для запуска процесса анализа
4. Кнопка «Сохранить образ» для сохранения результата анализа
5. Текстовое поле «Текст для разбора» для ввода анализируемого текста
6. Поле «Результат разбора» для отображения проанализированного текста с выделением найденных сущностей
7. Поле «Образ документа» для отображения найденных в проанализированном тексте сущностей с информацией про них, собранной по всему тексту.

### Выполнение анализа текста

Для выполнения анализа текста нужно выполнить следующие действия:

1. Ввести текст в окно «Текст для разбора» или вставить текст из буфера обмена.
2. Установить язык анализируемого текста, выбрав его из соответствующего списка
3. Установить тип анализируемого текста, выбрав его из соответствующего списка
4. Нажать кнопку «Разобрать»

Проанализированный текст с выделенными сущностями будет отображен в окне «Результат разбора», в окне «Образ документа» отобразятся названия групп найденных сущностей, как показано на Рисунке 2.

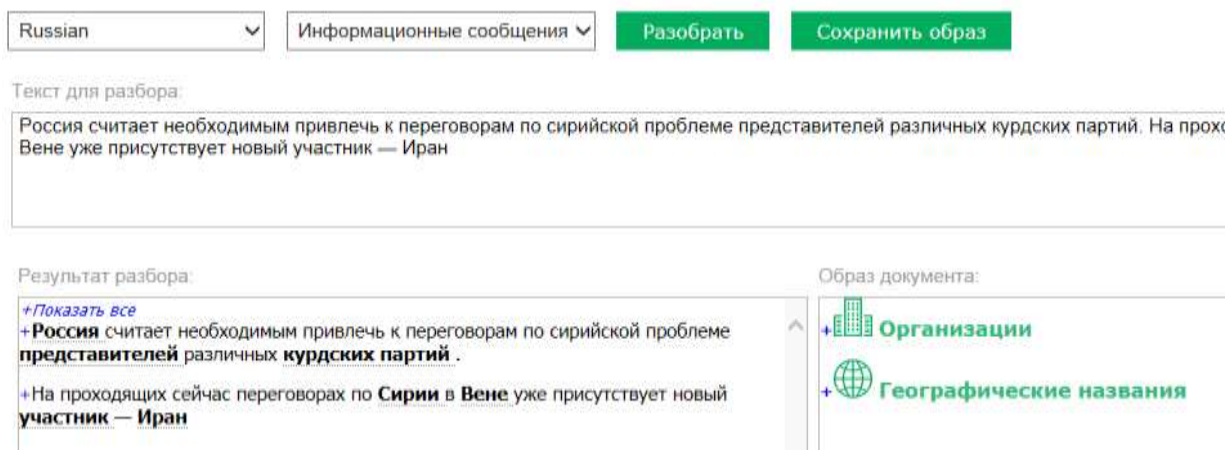


Рисунок 2

### Тип текста

Тип текста – это совокупность настроек (словарей, правил анализа, набора синонимических рядов), управляющих алгоритмами извлечения и визуализации сущностей для текстов определенной композиционной структуры и стиля.

В продукте доступны типы текстов:

1. Информационные сообщения – предназначен для анализа новостных сообщений политического характера
2. CV – предназначен для анализа автобиографий
3. Универсальный - для других текстов

### Поле “Результат разбора”

В окне “Результат разбора” отображается разделенный на предложения проанализированный текст, в котором выделены ключевые слова найденных в нем сущностей следующих типов:

- Персоны:  
имена собственные (имена, фамилии, отчества конкретных людей), нарицательные (имена существительные, имеющие семантический признак одушевленности, например, представитель или спортсмен) и личные местоимения
- Организации:  
имена собственные (названия конкретных организаций) и нарицательные (имена существительные, имеющие семантический признак организаций, например, школа, завод, спортивное общество и т.п.)
- Географические названия:  
имена собственные (названия географических мест) и нарицательные (имена существительные, имеющие семантический признак локации, например, река, местность и т.п.)
- Геополитические сущности (географические названия в организационном контексте) – имена собственные и нарицательные.
- Прочие сущности – имена собственные и нарицательные.

При клике на значок “+” в начале каждого предложения раскрывается подробная информация по найденным в предложении сущностям (см. Рисунок 3).



Рисунок 3

Указанная информация включает:

- Полное название выделенной сущности
- Тип сущности, с уточнением, если это имя собственное
- Подтип сущности
- Другую информацию, которая отнесена «Анализатор» к выделенной сущности

Для персон-имен собственных дополнительно может быть выведена информация о должности, профессии, национальности, родственных связях, если она была извлечена «Анализатор» из текста (см. Рисунок 4).



Рисунок 4

При клике на значок “+ Показать все” раскрывается информация по всем предложениям.

### Поле “Образ документа”

Окно “Образ документа” содержит интегральную информацию по сущностям, найденным в документе, сгруппированную по типу:

- Персоны
- Организации
- Географические названия
- Геополитика
- Событие, документ, вещество и другие именованные сущности

Группа "Персоны" содержит информацию о найденных персонах-именах собственных, персоны – имена нарицательные не включаются, даже если выделены в поле "Результат разбора"

Группа "Организации" содержит информацию об организациях-именах собственных. Организации – имена нарицательные не включаются, даже если выделены в поле "Результат разбора"

Группа "Географические названия" содержит информацию о найденных локациях-именах собственных, локации – имена нарицательные не включаются, даже если выделены в поле "Результат разбора"

Группа «Геополитика» содержит информацию о найденных географических названиях – именах собственных в организационном контексте (Москва заявила протест.)

Прочие группы содержат информацию о найденных именованных сущностях определенного типа.

При клике на название группы раскрывается информация о составе группы:

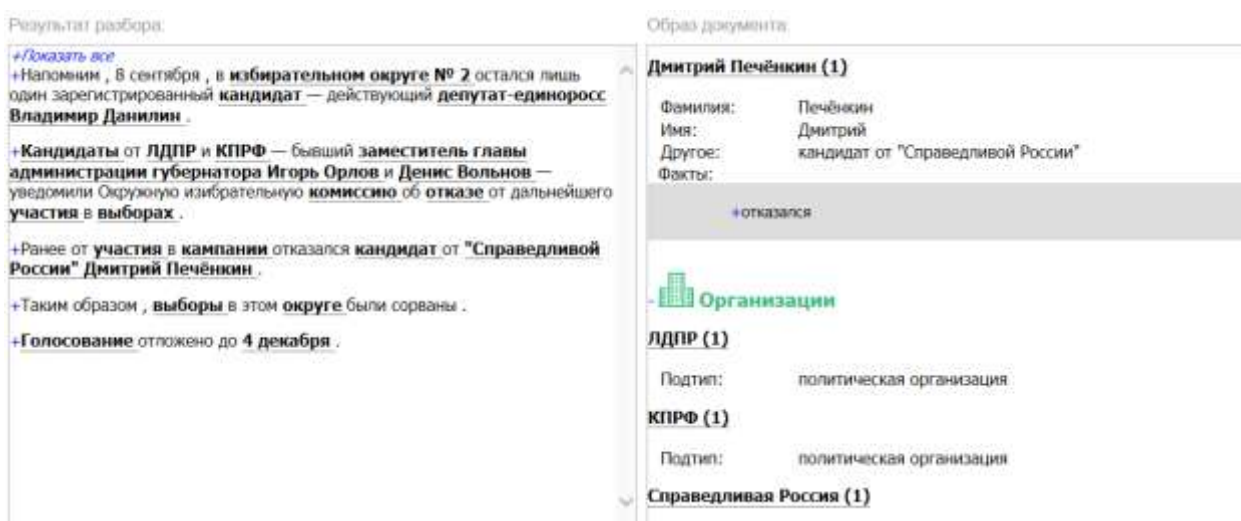


Рисунок 5

Для каждой найденной сущности отображается:

- количество найденных упоминаний в тексте (цифра в скобках)
- семантический подтип для организаций, географических названий, геополитических сущностей и событий
- другая информации, которая отнесена «Анализатор» к выделенной сущности
- для персон (если выделены из текста):
  - o должность
  - o профессия
  - o национальность
  - o родственная связь
- факты, в которых упоминалась сущность

Список найденных действий раскрывается при клике на значок "+" рядом с надписью "Факты" (см. Рисунок 6).



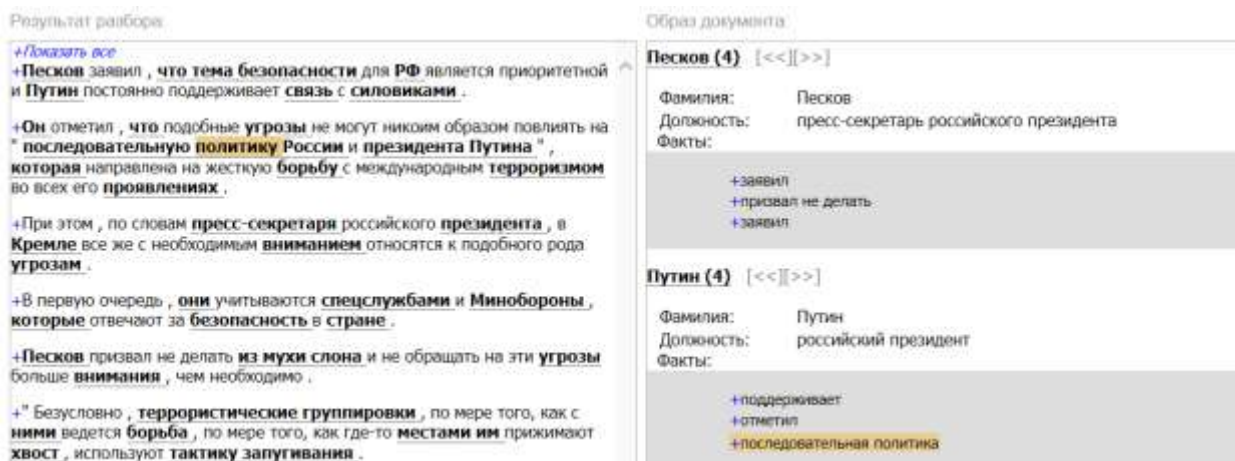


Рисунок 6

Факты сгруппированы по типам актантов, которыми является выделенная сущность по отношению к найденным действиям.

При клике на значок "+" рядом с действием раскрывается глагольный фрейм, описывающий данный факт (см. Рисунок 7).

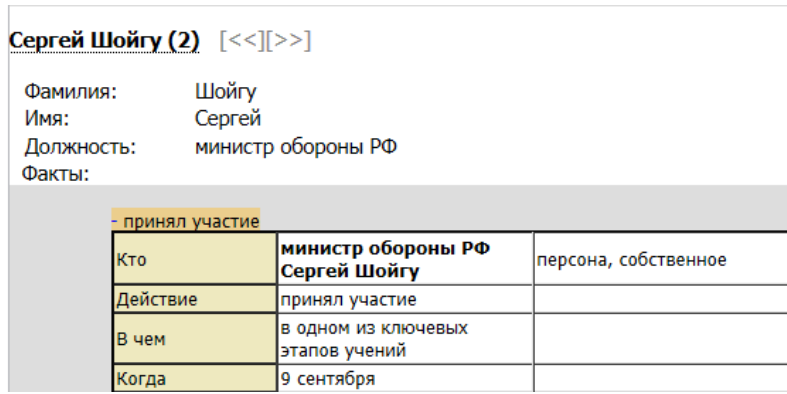


Рисунок 7

### Взаимосвязь полей "Результат разбора" и "Образ документа"

При клике на выделенную сущность в поле "Образ документа" в поле "Результат разбора" выделяются все найденные упоминания данной сущности, включая косвенные. При этом первое упоминание выделяется розовым цветом (см. Рисунок 8).

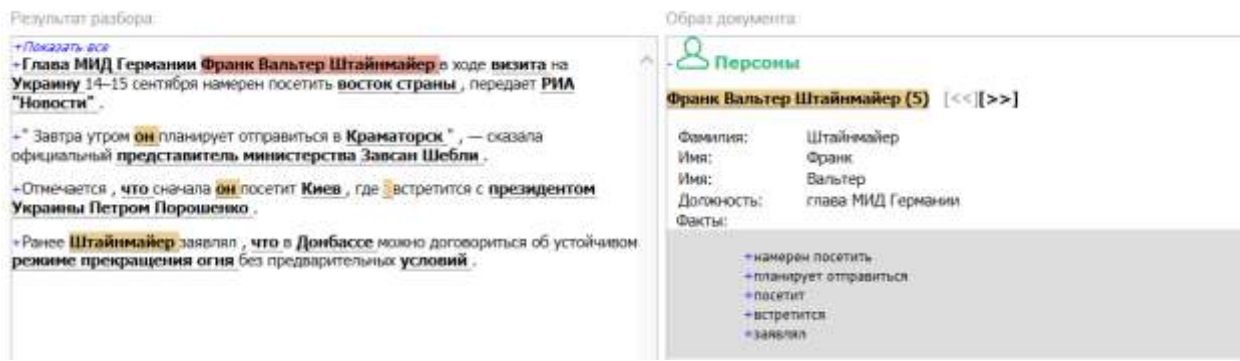


Рисунок 8

При клике на значок ">>" в поле "Образ документа" в поле "Результат разбора" розовым выделяется следующее упоминание сущности, а на значок "<<" – предыдущее.

При клике на действие или событие в списке "Факты" в поле "Образ документа" в поле "Результат разбора" действие выделяется коричневым, а сама сущность зеленым цветом (см. Рисунок 9).

+." Следует провести инвентаризацию индивидуальных средств защиты **граждан** , определить , какие из них морально и технически устарели , и выработать меры по восполнению запасов таких средств за счет современных образцов " , — **сказал президент** .

Кто/что	президент
Действие	сказал

Рисунок 9

## Сохранение образа документа

Образ документа может быть сохранен в виде HTML файла, для этого необходимо нажать на кнопку "Сохранить образ", откроется стандартный диалог сохранения в файл, в котором следует указать имя файла с сохраняемым образом.

## Программный интерфейс

В этой секции описаны методы веб-сервиса для извлечения сущностей из текста.

Входные и выходные данные запросов к веб-сервису могут быть либо в формате JSON, либо в формате XML. Формат можно указать в поле HTTP-заголовка запроса. Например, чтобы выполнить запрос с данными в формате JSON, необходимо задать следующие поля:

```
Content-Type: application/json
```

```
Accept: application/json
```

а для выполнения запроса в формате XML:

```
Content-Type: application/xml
```

```
Accept: application/xml
```

Ниже по тексту, примеры запросов и ответов даны в формате JSON (для метода ExtractEntities приводятся оба варианта). Если в процессе обработки запроса на сервере произошла ошибка, то код ответа от сервера будет равен 500, а тело ответа будет содержать описание ошибки.

Для извлечения сущностей в примерах используется следующий входной текст:

***Вчера пресс-секретарь Дмитрий Песков сказал в Москве господину Франсуа Олланду, что подобные угрозы не могут повлиять на последовательную политику России и президента Путина.***

### Общие сведения о сущности

«Анализатор» получает на вход фрагмент текста, анализирует его и выделяет из текста следующие именованные сущности:

- персоны (имена собственные), т.е. упоминания конкретных людей с фамилиями, именами и отчествами, если они находятся в тексте. Кроме этого, для найденных персон «Анализатор» выделяет ряд атрибутов, если информация о них присутствует в тексте (должность, профессия, родственные связи, партийная принадлежность, национальность и т.п.)
- названия организаций – имена собственные. Для найденных организаций выделяется тип (ОАО, ООО и т.п.) и подтип организаций (производственная организация, банк и т.п.), если информация о них присутствует в тексте или в онтологиях «Анализатор»
- географические названия с указанием типа (страна, город и т.п.) и подтипа (часть света, водоем и т.п.), если информация о типе и подтипе имеется в онтологиях «Анализатор»
- геополитические сущности – географические названия в организационном контексте (Москва заявила, требование Анкары) с указанием типа и подтипа, если информация о типе и подтипе имеется в онтологиях «Анализатор».

- другие именованные сущности разных типов, найденные при разборе текста (события, устройства, документы и т.д.) с указанием типа и подтипа, если информация о типе и подтипе имеется в онтологиях «Анализатор».

Для каждой сущности «Анализатор» анализирует и выделяет факты - действия и события, в которых зафиксировано участие сущности. Для каждого выделенного факта строится фрейм, описывающий действие (кто, что, где, когда и т.п.) или событие с выводом других найденных сущностей, являющихся актантами в этом действии. Пример:

Кто	пресс-секретарь Дмитрий Песков	персона, собственное
Действие	сказал	
Цитата	подобные угрозы не могут повлиять на последовательную политику России и президента Путина	высказывание с отрицательной тональностью
Кому	господину Франсуа Олланду	персона, собственное
Где	в Москве	география, собственное
Когда	Вчера	

### Свойства сущности

Сущность (entity) описывается следующим набором свойств (в скобках даны идентификаторы свойств для JSON и XML):

- **Идентификатор сущности** (id)

Используется для ссылки на сущность. Идентификаторы сущностей уникальны для сущностей одного разбора

```
{
  "id": 0
}
```

- **Имя сущности** (name)

Имя сущности в тексте

```
{
  "name": "Дмитрий Песков"
}
```

- **Тип сущности** (type)

```
{
  "type": "персона"
}
```

Возможны следующие типы сущностей (ниже в таблице содержатся 63 уникальных типа):

Тип	Подтип 1-го уровня	Подтип 2-го уровня
организация		
организация	орган власти	
организация	орган власти	исполнительная власть
организация	орган власти	законодательная власть
организация	орган власти	судебная власть
организация	орган власти	координационный орган
организация	орган власти	совещательный орган
организация	политическая организация	
организация	торговая организация	

организация	финансово-экономическая организация	
организация	финансово-экономическая организация	банк
организация	финансово-экономическая организация	инвестиционная компания
организация	финансово-экономическая организация	экономическая организация
организация	финансово-экономическая организация	биржа
организация	военная организация	
организация	организация сферы услуг	
организация	энергетика, газо- и водоснабжение	
организация	организация сферы услуг	связь
организация	организация сферы услуг	аудит/консалтинг
организация	организация сферы услуг	юридические услуги
организация	организация сферы услуг	туризм
организация	организация сферы услуг	гостиничный бизнес
организация	организация сферы услуг	бытовые услуги
организация	организация сферы услуг	индустрия красоты
организация	организация сферы услуг	предприятие общественного питания
организация	организация сферы услуг	страховая компания
организация	религиозная организация	
организация	СМИ	
организация	производственная организация	
организация	производственная организация	добывающая промышленность
организация	производственная организация	металлургия
организация	производственная организация	химическая промышленность
организация	производственная организация	машиностроение
организация	производственная организация	лёгкая промышленность
организация	производственная организация	пищевая промышленность
организация	производственная организация	сельское хозяйство
организация	производственная организация	строительная компания
организация	IT-компания	
организация	транспортная организация	
организация	благотворительная организация	
организация	профессиональная организация	
организация	правоохранительная организация	
организация	медицинская организация	
организация	культурно-развлекательная организация	
организация	культурно-развлекательная организация	театр
организация	культурно-развлекательная организация	музей
организация	культурно-развлекательная организация	кино
организация	культурно-развлекательная организация	культурно-досуговый центр
организация	научно-исследовательская	

	организация	
организация	образовательная организация	
организация	спортивная организация	
организация	природоохранная организация	
организация	исправительное учреждение	
организация	преступная организация	
география		
география	административная единица	
геополитика	административная единица	
география	гора	
география	река	
география	водоем	
география	город	столица
геополитика	город	столица
география	населенный пункт	
геополитика	населенный пункт	
география	часть света	
геополитика	часть света	
организация	союз государств	
география	страна	
геополитика	страна	
география	пустыня	
география	остров/полуостров	
география	географическое название	
геополитика	географическое название	
география	местоположение	
география	местоположение	
документ		
документ	правовой документ	
документ	печатное издание	
организация	СМИ	
финансовый параметр	финансовый документ	
фильм		
музыкальное произведение		
художественное произведение		
литературное произведение		
программа/приложение		
документ	планы/программы	
графический документ		
документ	технический документ	
стандарт/регламент		
идентификатор		
текстовый элемент		
речевой элемент		
совокупность знаний		
персона	другое	
персона	профессия	
персона	должность	
персона	должность	
персона	должность	
персона	обращение	
персона	национальность	
персона	имя	
персона	имя	прозвище
персона	фамилия	

персона	отчество	
персона	частица	
персона	родственник	
персона	политическая принадлежность	
персона	религиозная принадлежность	
персона	звание	
география		
география	остров/полуостров	
география	населенный пункт	
геополитика	населенный пункт	
география	город	столица
геополитика	город	столица
география	гора	
география	водоем	
география	река	
география	пустыня	
организация	союз государств	
география	географическое название	
геополитика	географическое название	
география	страна	
геополитика	страна	
география	административная единица	
геополитика	административная единица	
география	континент	
геополитика	континент	
география	местоположение	
география	местоположение	
география		
геополитика		
единица измерения		
единица измерения	объем	
единица измерения	объем памяти	
единица измерения	баллы	
единица измерения	метры	
время		
время	день недели	
качество		
качество	характеристика	
качество	взгляды	
качество	статус	
качество	состояние	
мера		
финансовый параметр	денежная единица	
деятельность		
деятельность	действие	
деятельность	однократное действие	
деятельность	военное или судебно-оперативное действие	
деятельность	преступление	
деятельность	перемещение транспортом	
деятельность	тесты-испытания-экспертизы	
деятельность	процесс	
область деятельности		
деятельность	медицинская процедура	
деятельность	физическая активность	

деятельность	мыслительная деятельность	
деятельность	хобби	
устройство		
область знаний		
область знаний	наука	
область знаний	философия	
область знаний	идеология	
область знаний	религия	
область знаний	культура/искусство	
одежда		
пищевые продукты		
время года		
вещество		
язык		
животное		
место		
число		
ёмкость		
география	часть света	
геополитика	часть света	
болезнь		
количественное существительное		
действие		
время		
возраст		
время		
погодное явление		
приём пищи		
праздник		
тип правления и социальная формация		
этап жизни		
событие		
событие	спортивное мероприятие	
событие	религиозное мероприятие	
событие	праздничное мероприятие	
событие	политическое и/или деловое мероприятие	
событие	научно-исследовательское мероприятие	
событие	акция	
событие	учебное или экзаменационное мероприятие	
событие	игры и развлечения	
событие	историческое событие	
транспорт		
месяц		
часть тела		
музыкальный инструмент		
растение		
оружие		
группа животных		
организация		
финансы		

инструмент-запчасти		
мебель		
набор отдельных предметов		
посуда		
продукция		
вид спорта		
налоги и сборы		
другие сущности		
жанр в искусстве		
чувство		

- **Подтип сущности** (subtype)

Подтип зависит от типа сущности. В подтип заносится подтип 1 уровня и подтип 2 уровня (см. таблицу выше), разделенные запятой

```
{
  "subtype": "город, столица"
}
```

- **Атрибуты сущности** (attributes)

Массив элементов. Для найденных сущностей «Анализатор» выделяет ряд атрибутов. Атрибуты – это неименованные сущности, которые при разборе были связаны с описываемой именованной сущностью, а именно являются ее предикатом или приложением к ней. Значением типа атрибута является подтип неименованной сущности 1 уровня (см. таблицу выше), а для организаций будет написано, что тип «организация»

Примеры атрибутов:

```
{
  "type": "должность",
  "name": "пресс-секретарь"
}
```

- **Число ссылок** (count)

Количество упоминаний сущности во входном тексте.

```
{
  "count": 3
}
```

- **Список ссылок на позиции в тексте** (references)

Массив элементов. Список ссылок описывает позиции в тексте всех упоминаний сущности и имеет следующие параметры:

- позиция в тексте (position)
- длина в символах (length)

```
{
  "position": 0,
  "length": 23,
}
```

- **Части имени** (personnames)

Массив элементов. Имя персоны разбивается на части. Каждая часть имеет следующие поля:

- имя (name)



- тип (type).

Тип (type) может принимать следующие значения:

- Имя персоны (NamePart)
- Отчество (PatronymicPart)
- Фамилия (SurnamePart)

```
"personnames": [
  {
    "name": "Дмитрий",
    "type": "NamePart"
  },
  {
    "name": "Песков",
    "type": "SurnamePart"
  }
]
```

- **Факты (facts)**

Массив элементов. Описание факта см. параграф Свойства факта

### Свойства факта

Для каждой сущности «Анализатор» анализирует и выделяет факты, в которых зафиксировано участие сущности. Факт имеет следующие свойства:

- **Имя факта (name)**

Имя факта в тексте

- **Тип факта (type)**

Может принимать следующие значения:

- действие (action)
- событие (event)
- эмоция (emotion)
- высказывание (saying)
- упоминание (mentioning)

- **Тип актанта сущности, которой принадлежит факт (atype)**

Описание типа актанта см. Свойства актанта

- **Тональность факта (tonality)**

Для высказываний и упоминаний указывается тональность. Может принимать значения:

- высказывание с отрицательной тональностью
- высказывание с положительной тональностью
- тонально маркированное высказывание

- **Актанты (actants)**

Массив элементов. Описание актанта см. Свойства актанта

- **Семантика глагола (semantic)**

Обозначается символом из следующей таблицы:

Обозначение семантики	Обозначение подсемантики	Описание семантики	Описание подсемантики
А	0	Глагол движения	
А	а	Глагол движения	Изменение положения тела
А	о	Глагол движения	Помещение объекта
Б	0	Глагол местонахождения	

В	0	Фазисный глагол	
Д	0	Глагол чувства	
Д	э	Глагол чувства	Эмоция
Д	в	Глагол чувства	Воля
Ж	0	Глагол речи и мышления	
Ж	м	Глагол речи и мышления	Ментальная сфера
Л	0	Глагол обладания	
Н	0	Бытийный глагол	
Н	н	Бытийный глагол	Начало существования
Н	к	Бытийный глагол	Прекращения существования
З	0	Глагол названия	
Ф	0	Факитивный глагол	
П	0	Глагол природного явления	
Р	0	Глагол воздействия	
Р	с	Глагол воздействия	Создание объекта
Р	у	Глагол воздействия	Уничтожение объекта
С	0	Глагол изменения состояния	
Т	0	Глагол физиологии	
Г	0	Глагол свойства	
Г	з	Глагол свойства	Звук
Г	х	Глагол свойства	Запах
Г	ц	Глагол свойства	Цвет

- **Подсемантика глагола** (subsemantic)

См. таблицу в предыдущем пункте

- **Позиции в тексте** (references)

Массив элементов. Позиции для факта — это позиция и длина его главного слова в тексте - обычно одна, но если факт состоит из 2-х слов, например, "хочу спать", то тогда 2 - для первого и второго слова. Каждый элемент массива имеет следующие параметры:

- позиция в тексте (position)
- длина в символах (length)

### Свойства актанта

- **Сущность** (entity)

См. описание **Свойства сущности**

- **Тип актанта** (atype)

Это название члена предложения в синтаксическом разборе. Принимает значения:

- Подлежащее (Subject)
- Дополнение в винительном падеже (Object)
- Дополнение в дательном падеже (Adressee)
- Дополнение в творительном падеже (Instrum)
- Предложное дополнение (IndirectObject)
- Наречная группа места (AdverbialOfPlace)
- Наречная группа времени (AdverbialOfTime)
- Наречная группа цели (AdverbialOfPurpose)
- Наречная группа образа действия (AdverbialOfManner)

- Наречная группа уступки (AdverbialOfConcession)
- Наречная группа условия (AdverbialOfCondition)
- Наречная группа причины (AdverbialOfCause)
- Вводное слово (Disjunct)
- Сказуемое (Predicate)
- Событие (Event)
- Эмоция (Emotion)
- Пустое значение - для субъекта эмоции или события

### Пример разбора (JSON)

В качестве примера для разбора использовался следующий текст:

***Вчера пресс-секретарь Дмитрий Песков сказал в Москве господину Франсуа Олланду, что подобные угрозы не могут повлиять на последовательную политику России и президента Путина.***

В процессе разбора были выделены следующие 5 сущностей (часть полей не показана, см. ниже):

```
[
  {
    "id": 0,
    "name": "Дмитрий Песков",
    "type": "персона"
  },
  {
    "id": 1,
    "name": "Франсуа Олланд",
    "type": "персона"
  },
  {
    "id": 2,
    "name": "Путин",
    "type": "персона"
  },
  {
    "id": 3,
    "name": "Москва",
    "subtype": "город, столица",
    "type": "география"
  },
  {
    "id": 4,
    "name": "Россия",
    "subtype": "страна",
    "type": "геополитика"
  }
]
```

Рассмотрим более подробно описание сущности "Дмитрий Песков". У сущности есть атрибуты:

```
"attributes": [
  {
    "name": "пресс-секретарь",
    "type": "должность"
```

```
}  
],
```

#### Имя и фамилия:

```
"personnames": [  
  {  
    "name": "Дмитрий",  
    "type": "имя"  
  },  
  {  
    "name": "Песков",  
    "type": "фамилия"  
  }  
]
```

#### Ссылки на позицию сущности в тексте:

```
"references": [  
  {  
    "length": 14,  
    "position": 22  
  }  
]
```

#### Описание факта (часть полей не показана):

```
"facts": [  
  {  
    "actants": [  
      {  
        "atype": "Subject",  
        "entity": {  
          "name": "пресс-секретарь Дмитрий Песков",  
          "type": "персона"  
        }  
      },  
      {  
        "atype": "Predicate",  
        "entity": {  
          "name": "сказал",  
          "type": "Ж"  
        }  
      }  
    ],  
    "atype": "Addressee",  
    "entity": {  
      "name": "господину Франсуа Олланду",  
      "type": "персона"  
    }  
  },  
  {  
    "actants": [  
      {  
        "atype": "Subject",  
        "entity": {  
          "name": "пресс-секретарь Дмитрий Песков",  
          "type": "персона"  
        }  
      },  
      {  
        "atype": "Predicate",  
        "entity": {  
          "name": "сказал",  
          "type": "Ж"  
        }  
      }  
    ],  
    "atype": "Addressee",  
    "entity": {  
      "name": "господину Франсуа Олланду",  
      "type": "персона"  
    }  
  }  
]
```

```

"atype": "AdverbialOfPlace",
"entity": {
  "name": "в Москве",
  "type": "география"
}
},
{
"atype": "AdverbialOfTime",
"entity": {
  "name": "Вчера",
  "type": ""
}
},
{
"atype": "Высказывание",
"entity": {
  "name": "подобные угрозы не могут повлиять на последовательную политику
России и президента Путина",
  "type": "высказывание с отрицательной тональностью"
}
},
{
"name": "сказал",
"references": [
{
"length": 6,
"position": 37
}
],
"tonality": "высказывание с отрицательной тональностью",
"type": "Saying"
}
]

```

## Аутентификация

Если на сервере используется аутентификация Forms, то необходимо выполнить следующие действия:

1. Вызвать метод **Login**. Метод **Login** реализован у специального сервиса аутентификации
2. Если метод **Login** отработал успешно, то сервер присылает куки аутентификации. Полученные куки необходимо передавать при вызове всех методов сервиса извлечения сущностей

## Метод Login (специальный сервис аутентификации)

### Описание

Метод используется для аутентификации пользователя

### URI

URI	HTTP метод
/login	POST

## Входные данные

### JSON

```
{
  "username": "имя пользователя",
  "password": "пароль",
  "isPersistent": "true|false"
}
```

username - имя пользователя

password - пароль

isPersistent - true - создается кука с заданным временем хранения, false - создается сессионная кука

### Методы веб-сервиса

Доступны следующие методы, описанные ниже более подробно:

- `Types` - список возможных типов сущностей
- `SubTypes` - список возможных подтипов сущности
- `Languages` - список языков, для которых поддерживается извлечение сущностей
- `Profiles` - список доступных профилей для каждого языка
- `Localizations` - список доступных локализаций для названий типов и подтипов
- `ExtractEntities` - извлечение сущностей из текста

### Метод `Types`

#### Описание

Метод возвращает список возможных типов для сущностей.

#### URI

URI	HTTP метод
<code>/types?language=&amp;localization=</code>	Get

## Входные данные

`Language` - (опциональный параметр) язык разбора. Может принимать одно из значений, возвращаемых методом `Languages`. По умолчанию используется русский язык

`Localization` - (опциональный параметр) язык локализации. Может принимать одно из значений, возвращаемых методом `Localizations`. По умолчанию используется русская локализация

## Выходные данные

Список типов. Пример типов:

```
[
  "персона",
  "организация",
  "география",
  "геополитика",
  ...
]
```

## Метод Subtypes

### Описание

Метод возвращает список возможных подтипов, в зависимости от языка и типа сущности.

### URI

URI	HTTP метод
/subtypes?type=&language=&localization=	Get

### Входные данные

`Type` - тип сущности. Может принимать одно из значений, возвращаемых методом `Types`  
`Language` - (опциональный параметр) язык разбора. Может принимать одно из значений, возвращаемых методом `Languages`. По умолчанию используется русский язык  
`Localization` - (опциональный параметр) язык локализации. Может принимать одно из значений, возвращаемых методом `Localizations`. По умолчанию используется русская локализация

### Выходные данные

Список подтипов. Пример подтипов для типа "персона":

```
[
  "профессия",
  "должность",
  "обращение",
  "национальность",
  "родственник",
  ...
]
```

## Метод Profiles

### Описание

Метод возвращает список доступных профилей для каждого языка.

### URI

URI	HTTP метод
/profiles?language=&localization=	Get

### Входные данные

`Language` - язык разбора. Может принимать одно из значений, возвращаемых методом `Languages`  
`Localization` - (опциональный параметр) язык локализации. Может принимать одно из значений, возвращаемых методом `Localizations`. По умолчанию используется русская локализация

### Выходные данные

Список профилей перевода. Пример:

```
[
  "Универсальный",
  "Личная переписка",
  "Спорт",
  ...
]
```

## Метод Languages

### Описание

Метод возвращает список языков, для которых может быть выполнено извлечение сущностей. В качестве идентификатора языка используется обозначение языка в соответствии с RFC 5646 (<http://www.rfc-editor.org/rfc/rfc5646.txt>).

### URI

URI	HTTP метод
/languages	Get

### Входные данные

Отсутствуют

### Выходные данные

Список доступных языков. Пример:

```
[  
  "ru",  
  "en"  
]
```

## Метод Localizations

### Описание

Метод возвращает список языков, на которые могут быть локализованы predetermined свойства сущностей. В качестве идентификатора языка используется обозначение языка в соответствии с RFC 5646 (<http://www.rfc-editor.org/rfc/rfc5646.txt>).

### URI

URI	HTTP метод
/localizations	Get

### Входные данные

Отсутствуют

### Выходные данные

Список доступных языков

```
[  
  "ru"  
]
```

## Метод ExtractEntities

### Описание

Метод извлекает сущности, найденные во входном тексте.

### URI

URI	HTTP метод
/extractentities	POST



## Входные данные

### **JSON**

```
{
  "text": "входной текст",
  "language": "язык текста",
  "profile": "профиль",
  "localization": "язык локализации",
  "type": "тип сущностей",
  "actions": "флаг генерации описания фактов"
}
```

### **XML**

```
<?xml version="1.0" encoding="utf-8"?>
<ExtractEntities>
  <text></text>
  <language></language>
  <profile></profile>
  <localization></localization>
  <type></type>
  <actions></actions>
</ExtractEntities>
```

**Text** - входной текст

**Language** - (опциональный параметр) язык анализируемого текста. Может принимать одно из значений, возвращаемых методом `Languages`. Если язык не задан, то он будет определен автоматически

**Localization** - (опциональный параметр) язык локализации. Может принимать одно из значений, возвращаемых методом `Localizations`. По умолчанию используется русская локализация

**Profile** - (опциональный параметр) профиль настроек для разбора. Может принимать одно из значений, возвращаемых методом `Profiles`. По умолчанию используется универсальный профиль

**Type** - (опциональный параметр) тип сущности. Может принимать одно из значений, возвращаемых методом `Types`. Если задан, то извлекаются только сущности заданного типа. По умолчанию извлекаются все сущности

**Actions** - (опциональный параметр) булевский флаг для генерации описания фактов. Может принимать значения `true` (генерировать) или `false` (не генерировать). По умолчанию - `true`

## Выходные данные

На выходе получается список извлеченных сущностей в формате JSON или XML (Подробное описание свойств сущности с примерами см. в параграфе "Общие сведения о сущности")

## **Метод Analyze**

### Описание

Метод анализирует входной текст, извлекая сущности и определяя тональность текста.

### URI

URI	HTTP метод
/analyze	POST

## Входные данные

### **JSON**

```
{
  "text": "входной текст",
  "language": "язык текста",
  "profile": "профиль",
  "localization": "язык локализации",
  "type": "тип сущностей",
  "facts": "флаг генерации описания фактов"
}
```

**Text** - входной текст

**Language** - (опциональный параметр) язык анализируемого текста. Может принимать одно из значений, возвращаемых методом `Languages`. Если язык не задан, то он будет определен автоматически

**Localization** - (опциональный параметр) язык локализации. Может принимать одно из значений, возвращаемых методом `Localizations`. По умолчанию используется русская локализация

**Profile** - (опциональный параметр) профиль настроек для разбора. Может принимать одно из значений, возвращаемых методом `Profiles`. По умолчанию используется универсальный профиль

**Type** - (опциональный параметр) тип сущности. Может принимать одно из значений, возвращаемых методом `Types`. Если задан, то извлекаются только сущности заданного типа. По умолчанию извлекаются все сущности

**Facts** - (опциональный параметр) булевский флаг для генерации описания фактов. Может принимать значения `true` (генерировать) или `false` (не генерировать). По умолчанию - `true`

## Выходные данные

На выходе получается массив извлеченных сущностей, а также тональность текста:

```
{
  "entities": [],
  "tonality": ""
}
```

Подробное описание свойств сущности с примерами см. в параграфе Свойства сущности. Поле тональность текста может принимать следующие значения:

- отрицательная
- положительная
- смешанная.