

# СОДЕРЖАНИЕ

<b>ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК «ФАКТОРА»</b> .....	<b>2</b>
Модули стандартизации и обогащения .....	3
Возможности модуля «Адреса» .....	4
Возможности модуля «Телефоны» .....	8
Возможности модуля «Паспорта» .....	9
Возможности модуля «ФИО и пол» .....	10
Возможности модуля «Даты» .....	11
Возможности модуля email .....	12
Возможности модуля «ЕГРЮЛ и ЕГРИП» .....	31
Возможности модуля «Компании» .....	32
Возможности модуля «Модели автомобилей» .....	33
Возможности модуля «Общий» .....	34
Модули идентификации .....	36
Новости «Фактора» .....	38

# ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК «ФАКТОРА»

## **МОДУЛИ СТАНДАРТИЗАЦИИ И ОБОГАЩЕНИЯ**

Модули очистки и стандартизации распознают, обрабатывают и приводят к стандартному представлению адреса, телефоны, ФИО, email, даты, реквизиты компаний, документы и модели автомобилей.

Модули используют базы знаний в несколько миллионов единиц и уникальные алгоритмы для обработки данных, которые учитывают опечатки, неполноту, дублирование и разный формат данных.

Ниже для каждого модуля описаны возможности, алгоритмы, ограничения, используемые нормативные акты и справочники.

## **ВОЗМОЖНОСТИ МОДУЛЯ «АДРЕСА»**

## ПОЧТОВЫЕ АДРЕСА

- Разбиваем адреса на компоненты.
- Исправляем опечатки и сокращения.
- Приводим компоненты к единому формату.
- Переводим с латиницы на кириллицу.
- Восстанавливаем пропущенные компоненты: «190000 УЛИЦА ДОЛГООЗЁРНАЯ 12» «190000, г Санкт-Петербург, ул Долгоозёрная, д 12»
- Заменяем старые названия на новые: «Ленинград» «Санкт-Петербург». По запросу можем [оставлять устаревшие значения](#). В режиме работы «без актуализации» московская улица Б. Коммунистическая не переименуется в улицу Солженицына, а Ленинград останется Ленинградом. Применяется для сохранения адреса, как в паспорте.
- Понимаем адрес, если регион записан в виде цифрового кода: «05 Мажалис Батырая дом 14» «Респ Дагестан, село Маджалис, ул Батырая, д 14»
- Разбираем адрес, когда тип и наименование склеились: «МОСКВА УЛКРАСНОГОМАЯКА Д, 10» «г Москва, ул Красного Маяка, д 10»
- Используем набор городов «по умолчанию», если в адресах пропущен город, но при этом известно местоположение клиентов (например, Москва и Мурманск).
- Используем ФИАС и собственные справочники: дополнения ФИАС, распространённые улицы.
- Выявляем ошибки в ФИАС. Очевидные исправляем, об остальных сообщаем в налоговую.
- В случае неоднозначности предлагаем [несколько вариантов адреса](#).
- Почтовые индексы ФИАСа проверяем и исправляем по данным Почты России.
- Определяем:
  - код КЛАДР и код ФИАС;
  - идентификатор ФИАС для [всех уровней адреса](#);
  - численность населения;
  - код ИФНС для юрлиц и физлиц из ФИАС;
  - признак состояния дома из ФИАС;
  - [район внутри города](#). Район внутри города также учитываем при разборе адреса, даже если его нет в ФИАС. Например, для Москвы, ни одного района которой нет в ФИАС, адрес «Москва, Хамовники р-н, ул Остоженка» найдётся и получит статус «гарантированный»;
  - ОКТМО и ОКАТО по адресу, [проверяем и корректируем ОКТМО](#) по данным Росстата для отчётности по форме 260 в ЦБ.
- Проставляем коды качества для распознанного адреса:
  - [код полноты адреса](#);
  - [статус проверки](#);
  - [признак актуальности](#);
  - [код качества сделанных изменений](#);
  - [код соответствия ФИАС](#);

наполнение адреса по ФИАС.

- Записываем адреса в формате для Почты России.
- Разделяем адреса, пригодные для курьерской доставки и для почтовой рассылки. **Отдельные** коды полноты присваиваются адресам для доставки писем, но по которым человека не найти лично.
- По идентификатору ФИАС получаем все компоненты адреса.
- Приводим в порядок место рождения согласно требованиям [Указания ЦБ N 4936-У](#).

## **Ограничения**

- Сопоставляем с ФИАСом до уровня улицы. Если дома нет в ФИАС, то он будет разобран, потому что ФИАС не полон по домам. Для проверки, есть ли дом в ФИАС, проставляем **код соответствия ФИАС**.
- Распознавание адресов вне Российской Федерации до уровня города.
- Одна улица может находиться в двух-трёх районах города, поэтому внутригородские районы не всегда определяем однозначно.
- Если района города нет в ФИАС, то он не влияет на результат разбора: например, оба адреса «Москва, Хамовники р-н, ул Остоженка» «Москва, Останкинский р-н, ул Остоженка» разберутся в «Москва, ул Остоженка» со статусом «гарантированный».

## **Используемые материалы**

- ФИАС ФНС. Можно обновлять без перезапуска Фактора.
- Численность населения (Росстат).
- Если у вас есть собственная полная база адресов, Фактор **может использовать** её вместо ФИАС.
- Справочник **эталонных индексов от Почты России**. Можно обновлять без перезапуска Фактора.
- Коды **ОКТМО Росстата**.

## **Площади и стоимости квартир**

По адресу определяем:

- Площадь квартиры и частного дома по данным Росреестра.
- **Среднюю стоимость** квадратного метра в доме (по объявлениям в Интернете).

## *Ограничения*

- Площади квартир и частных домов были выложены в свободный доступ Росреестром в 2013 году, после не обновлялись, поэтому новые данные не добавляются. Из этого источника мы получили порядка 60% квартир городов-миллионников и 23% квартир для остальных территорий РФ.
- Справочник стоимости квадратного метра в доме мы собираем по объявлениям о продаже в Интернете в нескольких источниках, обновляем раз в год.

## *Используемые материалы*

ЕГРП Росреестра (площади квартир). Можно [обновлять без перезапуска Фактора](#).

## **Ближайшее отделение Пенсионного фонда России**

В процессе перевода пенсии на накопительный счет банка сотрудники и клиенты заполняют анкету. В анкете требуют указать номер отделения пенсионного фонда.

Номер отделения найти непросто, поэтому научились по адресу [определять ближайшее отделение ПФР](#). Можем вернуть несколько вариантов.

## *Используемые материалы*

Справочник отделений Пенсионного фонда России.

# ВОЗМОЖНОСТИ МОДУЛЯ «ТЕЛЕФОНЫ»

## Телефоны

- Выделяем код города из телефона.
- Приводим компоненты телефона к единому формату.
- Проверяем и восстанавливаем телефонные коды на основании почтового адреса.
- Преобразуем региональные телефонные коды в федеральные, если известен адрес (8-2-7915216, Москва 8 495 7915216).
- Определяем тип телефона по его коду и ключевым словам в строке («факс», «моб», «дом», «раб» и пр).
- Выделяем мобильные.
- Выявляем несуществующие (например, «1111111»).
- Определяем оператора с учётом базы данных перенесённых номеров.
- Определяем регион оператора для мобильных номеров, адрес до города для стационарных.
- Проверяем соответствие длины телефона системе нумерации населенного пункта.
- Разделяем множества телефонов, заданных одной строкой, например, «моб 916 1510679, дом 320-78-10, рабочий 80951128912 доб. 3342» с учетом их типа.
- Заменяем телефонные коды устаревшим номерам.
- Присваиваем коды качества и подробные маркеры каждому номеру.
- Проверяем на существование номера в справочнике номерной емкости Россвязи.
- Распознаём номера телефонов в формате: <телефон>!<код города>, например, 7914216!495.
- Расклеиваем номера телефонов 90590518194957295200 +79059051819, +74957295200.

### Ограничения

- Иностраные номера разбираем, если указана страна или код страны.
- Определяем оператора и регион только для российских номеров.
- Прямые мобильные не идентифицируются.
- Расклеиваем номера только если все части разобрали в корректные российские номера.

### Используемые материалы

- Справочник номерной ёмкости (Россвязь). Горячее обновление.
- База данных перенесённых номеров. Горячее обновление.
- Информация о перенумерациях на АТС с сайтов региональных провайдеров связи.
- Справочник кодов стран и длины номера.



# ВОЗМОЖНОСТИ МОДУЛЯ «ПАСПОРТА»

## Паспорта

### *Возможности*

- Проверяем
  - Формат номера и серии документов, удостоверяющих личность (ДУЛ).
  - Подскажем, каких полей, важных для этого типа ДУЛ, не хватает.
  - Подскажем, почему серия или номер считаются некорректными.
  - Дату выдачи и окончания срока действия ДУЛ с учетом даты рождения.
  - Паспорта по перечню недействительных паспортов от МВД. Перечень можно обновлять без перезапуска Фактора.
- Проверка СНИЛС.

# ВОЗМОЖНОСТИ МОДУЛЯ «ФИО И ПОЛ»

## ФИО и пол

- Разбиваем по компонентам (фамилия, имя, отчество).
- Определяем пол по фамилии, имени и отчеству.
- Выявляем инициалы.
- Исправляем опечатки.
- Проводим трансграфику: одиночные латинские буквы меняем на такие же кириллические (например, Сергей—Сергей).
- Исправляем окончания фамилии или отчества, если пол однозначно определён по другим полям.
- Заменяем уменьшительно-ласкательные имена полными.
- Понимаем ФИО, написанные латиницей. Алгоритм учитывает ГОСТы и то, как на самом деле люди могут писать имена.
- Выявляем несуществующие имена (например, «Не знаю»).
- Проставляем **коды качества** каждой компоненте.
- Перечисляем **характеристики качества ФИО**, и какие из них являются поводом для ручной проверки.
- **Склоняем**: из именительного падежа в родительный, дательный или творительный.

### *Ограничения*

Исходные ФИО должны быть в именительном падеже.

### *Используемые материалы*

Собственные выверенные справочники:

- Распространённых фамилий, имён, отчеств.
- Распространённых опечаток.
- Справочники для перевода в родительный, дательный, творительный падежи.

# ВОЗМОЖНОСТИ МОДУЛЯ «ДАТЫ»

## Даты

- Распознаем даты в различных форматах и приводим их к единому формату.
- Подскажем, что именно неверно в дате: дата рождения из будущего или значится в списке дат по умолчанию (например, «01.01.1970»).
- Проверим соответствие дня рождения и возраста.

# ВОЗМОЖНОСТИ МОДУЛЯ EMAIL

## Стандартные проверки

- *Извлекаем несколько email из одной строки.*
- *Производим замену часто встречающихся опечаток в синтаксисе и написании доменов.*
- *Исправляем частые случаи трансграфики:*
  1. локальная часть — кириллическая С С латиницей, когда все остальные символы на латинице;
  2. доменная часть — аналогично предыдущему пункту + если доменная часть записана в punycode.
- *Проверка формата по правилам RFC. E-mail состоит из трёх частей:*
  1. часть до @:
    - a. не обрамленные в двойные кавычки ("), разделенные точкой, наборы из букв, цифр и символов `!#$%&'*+/-=?^_`{|}~ , ;`;
    - b. обрамленные в двойные кавычки: все печатные символы, кроме двойных кавычек (") и обратного слэша (\).
  2. знак @;
  3. доменная часть после @:
    - a. состоит из символов базовой латиницы, цифр и дефиса, разделенные точками;
    - b. дефис не должен быть первым и последним символом в элементе доменной части, после конвертации в punycode;
    - c. проверяем существования домена первого уровня по справочнику IANA.
- *Национальная письменность в доменах второго уровня для доменов первого уровня (IDN):*
  1. В доменах ru, int, edu, gov, mil, aero, coop, eco, jobs, ooo, post, pro, travel, wed, au, am ,cz, ge, nl, so стандарт idn не поддерживается, то есть запрещена любая письменность, отличная от латиницы (в том числе кириллица).
  2. В доменах rf, rus, москва запрещена латинская письменность.
  3. В доменах com, net, tel, name, org, su допускается любая письменность (в том числе кириллица).

- **Валидация по правилам популярных почтовых сервисов, отличным от RFC:**

yandex.ru, ya.ru

mail.ru, list.ru, bk.ru, inbox.ru

hotmail.com, outlook.com

yahoo.com

gmail.com

icloud.com

aol.com

rambler.ru

письмо.рф, е-письмо.рф (не работает ориентировочно с 2019 года). Применяются общие правила проверки email.

- **Проверяем длину домена второго уровня.**
- **Вычисляем реальный адрес с учетом синонимов — для корректных поиска дубликатов и проверок.**
- **Проставляем подробные характеристики качества. Подробнее см. Матрицы email.**
- **Черные и белые списки email заказчика.**

## **Подключаем по запросу**

- Проверка по справочнику «одноразовых» email (10minutemail.com, getairmail.com, temp-mail.ru и т.п.).
- Проверка части до @ по справочнику стандартных адресов.

## **Ограничения**

Не проверяем существование email адреса и почтового сервера.


Подробнее: не делаем пинг домена, не проверяем пробное соединение и не отправляем проверочные письма, потому что:

- эти проверки требуют подключение к Интернет;
- скорость ответа на пробное соединение низкая, часто больше минуты;
- точность ответов на пробное соединение 50%, есть ложно положительные и ложно отрицательные результаты.

## ***Используемые материалы***

- Справочник доменов первого уровня. Источник: <https://data.iana.org>.
- Справочник одноразовых email. Источник: [disposable-email-domains](#).
- Правила популярных почтовых доменов. Источник: **HFLabs**.
- Справочник [стандартных адресов](#) (info, support, etc). Источник: **HFLabs**.
- Распространённые опечатки в написании доменов второго уровня. Источник: **HFLabs**.

# ВЫЧИСЛЕНИЕ РЕАЛЬНОГО EMAIL

 Доступно с версии «Фактора» 20.10+

Один и тот же email можно записать по-разному. Если пользователь оставляет разные email в качестве своего контакта, то отправитель напишет на разные адреса, но на самом деле письма попадут в один и тот же «ящик», одному человеку.

Классические примеры:

- *lena@ya.ru* — то же самое, что *lena@yandex.ru*
- *ivanIvanovich@googlemail.com* — то же самое, что *ivanIvanovich@gmail.com*

Мы вычисляем реальный email по правилам почтовых серверов. Реальный email — новое поле для аналитических задач, в основном поле email останется указанное пользователем значение.

Реальный email помогает при поиске дубликатов и идентификации пользователей. Например, строки *ivanIvanovich@googlemail.com* и *ivanIvanovich@gmail.com* разные, но реальный email у них совпадает.

Если реальный email отличается от указанного пользователем, то значение получит один из [маркеров](#):

- LOCAL\_SUB\_ADDRESSING — локальная часть содержит тег, который будет проигнорирован почтовым сервером;
- LOCAL\_ALIAS — локальная часть содержит символы, которые будут проигнорированы или изменены почтовым сервером;
- DOMAIN\_ALIAS — синоним в доменной части.

Ниже подробнее о правилах.

## Синонимы доменной части (после @)

Домен	Синонимы	Исходный email	Поле «Реальный email»	Маркер
@gmail.com	@googlemail.com	ivanIvanovich@googlemail.com	ivanIvanovich@gmail.com	DOMAIN_ALIAS
@yandex.ru	@ya.ru @yandex.kz @yandex.ua @yandex.com @yandex.by	lena@ya.ru	lena@yandex.ru	DOMAIN_ALIAS

## Синонимы локальной части

Домены по-разному работают со знаками точки и дефиса.

Все правила только для части до @:

Домен	Правило	Исходный email	Поле «Реальный email»	Маркер
@gmail.com	Все точки удаляем	lena.ivan@gmail.com	lenaivan@gmail.com	LOCAL_ALIAS
@yandex.ru	Точки заменяем на дефис	lena.ivan@yandex.ru	lena-ivan@yandex.ru	LOCAL_ALIAS
все домены	Удаляем двойные кавычки	"lenaivan"@yahoo.com	lenaivan@yahoo.com	LOCAL_ALIAS

## Плюс-адресация

Подадресация (плюс-адресация) — возможность написать в части до @ слова, от которых не изменится фактический адрес. Эти слова отделены от основной локальной части спецсимволом.

Домен	Спецсимвол/комментарий	Исходный email	Поле «Реальный email»	Маркер
Для всех доменов, кроме yahoo.com	плюс	lena+ivan@yandex.ru	lena@yandex.ru	LOCAL_SUB_ADDRESSING
yahoo.com	у сервиса нет подадресации, есть временные емейлы, см <a href="#">Исключения по ссылке</a>	lena+ivan@yahoo.com	пусто	—
Для всех доменов	если email начинается со спецсимвола, то ничего не делаем.	+79779683249@yandex.ru	+79779683249@yandex.ru	—





## ПРАВИЛА ДОМЕНОВ EMAIL

- yandex.ru, yandex.by, yandex.kz, yandex.com, yandex.ua, ya.ru, ya.by, ya.kz, ya.com, ya.ua
- mail.ru, list.ru, bk.ru, inbox.ru
- gmail.com
- hotmail.com, outlook.com
- icloud.com
- yahoo.com
- aol.com
- rambler.ru
- e-pismorf.com, pismorf.com
- e-письмо.рф, письмо.рф — не работает с начала 2019 года. Применяются общие правила проверки email.

Отключить проверку по правилам почтовых серверов

**i** С версии «Фактора» 20.10 правила доменов проверяем после вычисления реального email.






**i** Если емейл не проходит проверку, то ставим маркер LOCAL\_BAD\_SPECIFIC (см. Маркеры email).




# YANDEX.RU, YANDEX.BY, YANDEX.KZ, YANDEX.COM, YANDEX.UA, YA.RU, YA.BY, YA.KZ, YA.COM, YA.UA

Логин — количество символов от 4 до 30 включительно

## Символы:

- латинские буквы (a-z, A-Z);
- цифры (0-9);
- дефис;
- точка.

Правила	 Валидные почтовые адреса	Комментарии	 Невалидные почтовые адреса	Комментарии
 первый знак — буква	Ivanov. Ivan23@yandex.ru	<ul style="list-style-type: none"> <li>• заканчивается на цифру</li> <li>• двойных знаков нет</li> </ul>	Ivanov.- Ivan@yandex.ru	2 разных знака идут подряд
 последний знак — буква или цифра	Ivanov. Ivan23@ya.ru		Ivanov.- Ivan@ya.ru	
 недопустимо 2 и более знака (дефис /точка) любой комбинации подряд	Ivanov.Ivan. Ivanovich@yandex.ru	<ul style="list-style-type: none"> <li>• двойных знаков нет</li> <li>• начинается и заканчивается на букву</li> </ul>	Ivanov-- Ivan@yandex.ru	2 одинаковых знака идут подряд
	Ivanov.Ivan. Ivanovich@ya.ru		Ivanov-- Ivan@ya.ru	
	eIena+togetherMoreThan20Symbols@ya.ru	Изначально логин превышает максимальную длину. Но после учёта плюс-адресации понимаем, что реальный email корректный: elena@yandex.ru.		

 Исключение	 Валидные почтовые адреса	Комментарии	 Невалидные почтовые адреса	Комментарии
номер телефона до @	+79871234567@ya.ru		+89086329825@yandex.ru	после знака идет не цифра 7
Возможны варианты:				
<ul style="list-style-type: none"> <li>• номер начинается с 8, длина 11 цифр</li> <li>• номер начинается с +7, длина 11 цифр</li> <li>• номер начинается с 7, длина 11 цифр</li> <li>• номер не начинается с 8,+7 или 7, длина 10 цифр</li> </ul>	+79871234567@yandex.ru	<p>начинается со знака «+»</p> <p>первая цифра 7</p> <p>количество цифр 11</p>	+89086329825@ya.ru	

Описание в документации Яндекса.	89871234567@yandex.ru	первая цифра 8 количество цифр 11	890863298255@yandex.ru	длина 12 цифр
	89871234567@ya.ru		890863298255@ya.ru	
	79086329825@ya.ru	первая цифра 7 количество цифр 11	790863298255@yandex.ru	длина 12 цифр
	79086329825@yandex.ru		790863298255@ya.ru	
	9086329825@yandex.ru	количество цифр 10	908632982@yandex.ru	длина 9 цифр

## MAIL.RU, LIST.RU, VK.RU, INBOX.RU

Логин — количество символов от 4 до 31 включительно

### Символы:

- латинские буквы (a-z, A-Z);
- цифры (0-9);
- дефис;
- точка;
- нижнее подчеркивание.

Правила	✓ Валидные почтовые адреса	Комментарии	✗ Невалидные почтовые адреса	Комментарии
✓ первый знак — буква или цифра	79086329825_Ivanov.		Ivan_.	2 разных знака идут подряд
✓ последний знак — буква или цифра	Ivan@mail.ru	<ul style="list-style-type: none"><li>• начинается на цифру</li><li>• двойных знаков нет</li></ul>	Petrov@mail.ru	
✗ недопустимо 2 и более знака (дефис/точка/нижнее подчеркивание) <i>любой комбинации</i> подряд	Ivan_Ivanovich. Petrov@mail.ru	<ul style="list-style-type: none"><li>• двойных знаков нет</li><li>• начинается и заканчивается на букву</li></ul>	Ivan-- Petrov@mail.ru	2 одинаковых знака идут подряд

# GMAIL.COM

Логин — количество символов от 6 до 30 включительно

## Символы:

- латинские буквы (a-z, A-Z);
- цифры (0-9);
- точка.







Правила	✓ Валидные почтовые адреса	Комментарии	✗ Невалидные почтовые адреса	Комментарии
✓ первый знак — буква или цифра	79086329825. Ivanov.Ivan@gmail.com	<ul style="list-style-type: none"><li>• начинается на цифру</li></ul>	Ivan.. Petrov@gmail.com	2 точки идут подряд
✓ последний знак — буква или цифра		<ul style="list-style-type: none"><li>• двойных знаков нет</li></ul>		
✗ недопустимо 2 и более одинаковых знака (точка) подряд	Ivan.Ivanovich. Petrov@gmail.com	<ul style="list-style-type: none"><li>• двойных знаков нет</li><li>• начинается и заканчивается на букву</li></ul>	Ivan_Petrov@gmail.com	нижнее подчеркивание запрещено

## HOTMAIL.COM, OUTLOOK.COM

Логин — количество символов от 1 до 64 включительно

### Символы:

- латинские буквы (a-z, A-Z);
- цифры (0-9);
- дефис;
- нижнее подчеркивание;
- точка.







Правила	 Валидные почтовые адреса	Комментарии	 Невалидные почтовые адреса	Комментарии
 первый знак — буква	Ivanov_Ivan@hotmail.com	<ul style="list-style-type: none"> <li>• начинается на цифру</li> <li>• двойных знаков нет</li> </ul>	Ivan..Petrov@hotmail.com	2 точки идут подряд
 последний знак — буква/цифра/дефис /нижнее подчеркивание				
 разрешено 2 и более знака (дефис /нижнее подчеркивание) любой комбинации подряд				
 недопустимо 2 и более одинаковых знака (точка) подряд	Ivanovich--@hotmail.com	<ul style="list-style-type: none"> <li>• начинается на букву</li> <li>• заканчивается на дефис</li> <li>• двойной дефис</li> </ul>	.-Ivanovich@hotmail.com	недопустимое сочетание символов
	Ivanovich_@hotmail.com	<ul style="list-style-type: none"> <li>• начинается на букву или цифру</li> <li>• заканчивается на нижнее подчеркивание</li> <li>• двойное нижнее подчеркивание</li> </ul>		
	79086329825Ivanov@hotmail.com			

# ICLOUD.COM

Логин — количество символов от 3 до 20 включительно

## Символы:

- латинские буквы (a-z, A-Z);
- цифры (0-9);
- нижнее подчеркивание;
- точка;
- плюс.

Правила	 Валидные почтовые адреса	Комментарии	 Неважные почтовые адреса	Комментарии
 первый знак — буква	Ivanov_Ivan2		Ivanovich_@i	2 одинаковых знака подряд (нижнее подчеркивание)
 последний знак — буква/цифра	3@icloud.com	<ul style="list-style-type: none"><li>• начинается на букву</li><li>• двойных знаков нет</li><li>• заканчивается на цифру</li></ul>	cloud.com	
 недопустимо 2 и более знака (точка/нижнее подчеркивание/дефис) <i>любой комбинации</i> подряд				
 недопустим символ дефис	Ivanov. Ivan@icloud.com	<ul style="list-style-type: none"><li>• начинается и заканчивается на букву</li><li>• двойных знаков нет</li></ul>	Ivan_- Petrov@icloud.com 79086329825Ivanov@icloud.com	2 разных знака подряд (дефис и нижнее подчеркивание) начинается на цифру



# YAHOO.COM

Логин — количество символов от 4 до 32 включительно

## Символы:


- латинские буквы (a-z, A-Z)
- цифры (0-9)
- нижнее подчеркивание
- точка
- дефис

Правила	✓ Валидные почтовые адреса	Комментарии	✗ Невалидные почтовые адреса	Комментарии
✓ первый знак — буква	Ivanov.		Ivanov.Ivan.	2 точки идут подряд
✓ последний знак — буква или цифра	Ivan23@yahoo.com	<ul style="list-style-type: none"> <li>• заканчивается на цифру</li> <li>• двойных знаков нет</li> </ul>	Ivanovich@yahoo.com	
✗ недопустимо 2 и более знаков (точка /нижнее подчеркивание) <i>любой комбинации</i> подряд				
✗ недопустимо больше одной точки	Ivanov_Ivan_Ivanovich@yahoo.com	<ul style="list-style-type: none"> <li>• двойных знаков подряд нет</li> <li>• начинается и заканчивается на букву</li> </ul>	Ivanov_Ivan@yahoo.com	2 одинаковых знака идут подряд
			Ivanov_Ivan@yahoo.com	

У Yahoo даёт возможность пользователям в личном кабинете создавать дополнительные адреса. Они всегда содержат дефис и работают как самостоятельные адреса. Пользователь сам решает, сколько будет действовать его дополнительный адрес.

⚠ Исключение	✓ Валидные почтовые адреса	Комментарии	✗ Невалидные почтовые адреса	Комментарии
Допускается логин с дефисом (временный адрес):	Ivan-Ivanovich@yahoo.com	<ul style="list-style-type: none"> <li>• временный адрес</li> <li>• первый и последний знак буква</li> </ul>	Ivanov.Ivan-Ivanovich@yahoo.com	есть точка
✓ каждая из частей 1-32 символа				
✓ первый знак — буква				
✓ последний знак — буква или цифра				

---

 недопустимы точки и  
нижнее подчеркивание

Ivan02-  
Ivanovich16@yahoo.  
com

- временный адрес
- первый знак  
буква
- последний знак  
цифра

Ivanov\_Ivan-  
Ivanovich@yahoo.  
com






---

# AOL.COM

Логин — количество символов от 3 до 32 включительно

## Символы:

- латинские буквы (a-z, A-Z)
- цифры (0-9)
- нижнее подчеркивание
- точка





Правила	 Валидные почтовые адреса	Комментарии	 Невалидные почтовые адреса	Комментарии
 первый знак — буква	Ivanov. Ivan23@aol. com	<ul style="list-style-type: none"><li>• заканчивается на цифру</li><li>• двойных знаков нет</li></ul>	Ivanov. _Ivan@aol. com	любая последовательная комбинация точки и нижнего подчеркивания недопустима
 последний знак — буква или цифра				
 недопустимо 2 и более знаков (точка/нижнее подчеркивание) <i>любой комбинации</i> подряд	Ivanov_Ivan_ Ivanovich@ aol.com	<ul style="list-style-type: none"><li>• двойных знаков подряд нет</li><li>• начинается и заканчивается на букву</li></ul>	Ivanov_Ivan@ aol.com	2 одинаковых знака идут подряд

# RAMBLER.RU

Логин — количество символов от 3 до 32 включительно

## Символы:

- латинские буквы (a-z, A-Z)
- цифры (0-9)
- нижнее подчеркивание
- дефис
- точка




Правила	 Валидные почтовые адреса	Комментарии	 Неважные почтовые адреса	Комментарии
 первый и последний знак — буква или цифра	Ivanov. Ivan23@rambler.ru	<ul style="list-style-type: none"><li>• заканчивается на цифру</li></ul>	Ivanov. _Ivan@rambleg.ru	любая последовательная комбинация символов (точка/нижнее подчеркивание/дефис) недопустима
 недопустимо 2 и более одинаковых знаков (точка/нижнее подчеркивание/дефис) <i>любой комбинации</i> подряд	Ivanov_Ivan_Ivanovich@rambler.ru	<ul style="list-style-type: none"><li>• двойных знаков подряд нет</li><li>• начинается и заканчивается на букву</li></ul>	Ivanov_Ivan@rambler.ru	2 одинаковых знака идут подряд

## E-PISMORF.COM, PISMORF.COM

Логин — количество символов от от 1 до 80 включительно

### Символы:

- латинские буквы
- арабские цифры
- тире (дефис)
- символ подчеркивания

	 Валидные почтовые адреса	Комментарии	 Невалидные почтовые адреса	Комментарии
 любой	Ivanov.Ivan. Ivanovich@pismorf.com	<ul style="list-style-type: none"><li>• двойные знаки разрешены</li></ul>	Ivanov. Ivan+Ivanovich@pismorf.com	знак «+» запрещены 2 точки подряд
	Ivanov.-Ivan. Ivanovich@pismorf.com		Ivanov..Ivan. Ivanovich@pismorf.com	
	-Ivanov.Ivan. Ivanovich_@pismorf.com	<ul style="list-style-type: none"><li>• начинается с любого разрешенного знака</li></ul>	ИвановИван@pismorf.com	кириллица запрещена
	16Ivanov.Ivan. Ivanovich.@pismorf.com	<ul style="list-style-type: none"><li>• заканчивается на любой разрешенный знак</li></ul>	.Ivanov.Ivan. Ivanovich-@pismorf.com	запрещается символ точка в начале и конце ( не может быть разрешен в доменах правила, которые запрещены в RFC)

**Е-ПИСЬМО.РФ, ПИСЬМО.РФ — НЕ РАБОТАЕТ С НАЧАЛА  
2019 ГОДА. ПРИМЕНЯЮТСЯ ОБЩИЕ ПРАВИЛА ПРОВЕРКИ  
EMAIL.**

# ВОЗМОЖНОСТИ МОДУЛЯ «ЕГРЮЛ И ЕГРИП»

## Поиск компании в ЕГРЮЛ и предпринимателя в ЕГРИП

Проверяем реквизиты юридических лиц и индивидуальных предпринимателей по [ЕГРЮЛ](#) и [ЕГРИП](#).

При поиске по ИНН/ОГРН восстанавливаем атрибуты:

- Полное наименование из ЕГРЮЛ
- Краткое наименование из ЕГРЮЛ
- Полное наименование на английском языке
- КПП, ИНН
- ОГРН или НЗА для филиалов иностранных
- Код ОКПОФ
- Юридический адрес для ЮЛ и город регистрации для ИП.

Проверяем актуальность данных руководителя или учредителя компании по ФИО и ИНН.

### *Используемые материалы*

ЕГРЮЛ и ЕГРИП по данным от [ФНС](#). Обновляем раз в месяц. Можно обновлять без перезапуска Фактора.

# ВОЗМОЖНОСТИ МОДУЛЯ «КОМПАНИИ»

## Компании

- Выделяем организационно-правовую форму (ОПФ).
- Удаляем «мусор» из названия компании (мусорные слова, ИНН, КПП, почтовые адреса).
- Формируем название компании-ключа для последующей идентификации дубликатов.
- Заменяем старые ОКОГУ на ОКОГУ, действующие с 1 января 2012 года.
- Проверяем:
  - ИНН с учетом ОПФ по контрольной сумме, формату и справочнику кодов налоговых органов (СОУН). Дополняем справочник СОУН недостающими значениями из ЕГРЮЛ.
  - КПП по формату и справочнику СОУН. Дополняем справочник СОУН недостающими значениями из ЕГРЮЛ.
  - ОГРН по контрольной сумме, формату и справочнику кодов субъектов РФ.
  - ОКПО по контрольной сумме.
  - ОКВЭД по формату и допустимым значениям классификатора.
  - ОКАТО по контрольной сумме и формату.
  - SWIFT по формату.
  - КИО по длине и сравнению с ИНН.
  - ОКОГУ по справочнику.
  - БИК по формату.
  - Синтаксис URL в веб-сайте компании.

## Используемые материалы

- Коды СОУН от ФНС — для проверки ИНН и КПП.
- Справочники ОПФ для стандартизации наименований и наш справочник опечаток в ОПФ.
- Справочник ОКВЭД.
- Справочник ОКОГУ.



# ВОЗМОЖНОСТИ МОДУЛЯ «МОДЕЛИ АВТОМОБИЛЕЙ»

## Модели автомобилей

- Распознавание моделей, заданных в неструктурированном формате с учетом опечаток, избыточной или недостаточной информации, синонимов:
  - Alfa Romeo 156 ALFA-ROMEO, 156
  - Chery Amulet CHERY, AMULET/A15
  - Chery Tiggo 2.0 CHERY, TIGGO/T11
  - Kia Cee'd ED KIA, CEED
- Варианты распознавания моделей для случаев негарантированного распознавания:
  - BMW 3 BMW M3, BMW X3, BMW Z3
  - Kia K-серии KIA K2500, KIA K2700

### *Используемые материалы*

Справочник моделей автомобилей, составленный из нескольких источников:

- [auto.ru](http://auto.ru)
- [auto.basebuy.ru](http://auto.basebuy.ru)

## **ВОЗМОЖНОСТИ МОДУЛЯ «ОБЩИЙ»**

## **БАЗОВЫЕ ПРАВИЛА**

Правила, входящие в базовую поставку Фактор. Помогают преобразовать данные для миграции.

- Объединение двух и более полей в одно.
- Условный оператор.
- Копирование.
- Удаление или маркировка значения, содержащего «мусор».
- Замена по справочнику.
- Определение часового пояса РФ.

## **Обратная транслитерация**

- Распознавание адресов, написанных на английском языке и транслите.
- Распознавание ФИО, написанных на английском языке и транслите.

### ***Ограничения***

- Нет литературного перевода с английского языка на русский.
- Распознавание компаний не включено (только некоторые правовые формы).

## **Подробное логирование для мониторинга**

Фактор собирает статистику работы оффлайн задач, статистику обработки онлайн запросов всех сервисов, следит за состоянием сервера.

Всё это помогает мониторингу нагрузки, а также быстрому диагностированию проблем.

# МОДУЛИ ИДЕНТИФИКАЦИИ

**i** Сведения о функциональности модулей идентификации дубликатов и домохозяйств «Фактора»

- Идентификация дубликатов
  - Термины
  - Возможности
  - Ограничения
- Проверка по черным спискам
  - Возможности
  - Ограничения

Модули идентификации дубликатов позволяют идентифицировать похожие друг на друга записи. Для этого необходимо сравнить несколько полей, например, ФИО, серия и номер паспорта и дата рождения.

## Идентификация дубликатов

### **Термины**

Идентификация дубликатов в Факторе описывается в виде группы *сценариев*.

*Сценарий* — правило, определяющее какие поля должны совпасть и с какой степенью схожести.

Для определения степени схожести записей используется набор *компараторов*, сравнивающий отдельные поля записи.

*Компаратор* сравнивает одно или несколько полей одного типа. Например, даты рождения двух физических лиц или 10 телефонов одного клиента и 5 телефонов другого.

### **Возможности**

- Специальные компараторы для адресов и имен, с учетом специфики сравнения данных компонентов
- Компараторы для сравнения групп телефонов и адресов.
- Компараторы для сравнения строк на схожесть.
- Нечёткое сравнение **ДУЛ** и **дат** с учётом распространенных опечаток.

- Высокая скорость идентификации дубликатов.
- Полный поиск дубликатов на всей базе.
- Инкрементальный поиск на объеме обновлённых за день данных.
- Онлайн-сервис для нечеткого поиска по всему объему.

## ***Ограничения***

Для запуска идентификации дубликатов данные должны быть стандартизированы.

## **Проверка по черным спискам**

Поиск данных клиента (ЮЛ или ФЛ) на вхождение в справочники:

- террористов;
- иностранных публичных должностных лиц;
- любых иных списков, например, недобросовестных заемщиков.

Помогает выполнить требования ФЗ-115 о противодействии финансированию терроризма, а также идентифицировать неблагонадёжного клиента.

## ***Возможности***

- Возможность подключить любые справочники различных форматов как черные списки;
- Гибкие правила поиска по схожести (на основе идентификации дубликатов);
- Высокая скорость поиска по черным спискам:
  - онлайн: 1/4 секунды при миллионном справочнике;
  - пакетная обработка: 5 млн записей в час.

## ***Ограничения***

Справочники для поиска предоставляются Заказчиком.

## НОВОСТИ «ФАКТОРА»

## СДЕЛАНО В 2020

- Вычисляем реальный email
- Учитываем реальный email при проверке правил доменов
- Проверяем актуальность учредителей и директоров компаний по ЕГРЮЛ
- Код ОКТМО от Росстата для отчёта в ЦБ по форме 260
- Изменили сокращение для типа «городок»
- Выпускаем новый справочник геокоординат чаще
- Подключили новый ФИАС с «городами-районами»
  - Для заказчиков с EssenceCleanService
  - Изменение в определении уровня по ФИАС для «городов-районов»
  - Изменение кода полноты для адресов в «городах-районах» с планировочной структурой
  - Недоступно горячее обновление на новый ФИАС для «Фактора» 20.07 и ранее
- Изменили формат ФИАС для «Фактора»
- Понимаем адреса с административными округами, которых нет в ФИАС
- Версии ФИАС теперь проходят строгий карантин
- Вернули КЛАДР-коды планировочным структурам
- Новые типы адресных объектов
- Логирование
- Проверяем выполнение системных требований
- Унифицированная структура каталогов
- Проверяем оставшееся место на диске
  - Сообщаем о нехватке свободного места на диске
  - Динамически записываем количество свободного места в factor-stats-avg.log
- Логирование при срабатывании таймаута обработки, заданного для типа данных
- Логирование при срабатывании таймаута обработки запроса целиком

### 20.10 – Email

#### ***Вычисляем реальный email***

Технические подробности, не видимые заказчику

---

PGC-15621

Работает только в новом emailFilter

Один и тот же email можно записать по-разному. Если пользователь оставляет разные email в качестве своего контакта, то отправитель напишет на разные адреса, но на самом деле письма попадут в один и тот же «ящик», одному человеку.

Классические примеры:

- *lena@ya.ru* — то же самое, что *lena@yandex.ru*
- *ivanIvanovich@googlemail.com* — то же самое, что *ivanIvanovich@gmail.com*

Мы вычисляем реальный email по правилам почтовых серверов. Реальный email — новое поле для аналитических задач, в основном поле email останется указанное пользователем значение.

Реальный email помогает при поиске дубликатов и идентификации пользователей. Например, строки *ivanIvanovich@googlemail.com* и *ivanIvanovich@gmail.com* разные, но реальный email у них совпадает.

Учитываем синонимы доменной и локальной частей email, а также плюс-адресацию. Подробнее о правилах по ссылке: [вычисление реального email](#).

Если реальный email отличается от исходного, проставляем один из трёх новых маркеров:

- LOCAL\_SUB\_ADDRESSING — локальная часть содержит тег, который будет проигнорирован почтовым сервером;
- LOCAL\_ALIAS — локальная часть содержит символы, которые будут проигнорированы или изменены почтовым сервером;
- DOMAIN\_ALIAS — синоним в доменной части.

<i>ivanivanov+subaddr@gmail.com</i>	<i>ivanivanov@gmail.com</i>	LOCAL_SUB_ADDRESSING	После знака + стоит подадресация, которую игнорируют почтовые сервера.
<i>i.v.a.n.ivanov@gmail.com</i>	<i>ivanivanov@gmail.com</i>	LOCAL_ALIAS	Для домена <i>gmail.com</i> и его синонимов логин с точками и без — равнозначны.
<i>ivanivanov@googlemail.com</i>	<i>ivanivanov@gmail.com</i>	DOMAIN_ALIAS	Домены <i>googlemail.com</i> и <i>gmail.com</i> равнозначны.

Новое поле «реальный email» подключаем в рамках доработки.

## ***Учитываем реальный email при проверке правил доменов***

Лучше проверяем, какой email можно регистрировать.

Теперь проверяем по [правилам доменов email](#) только после вычисления реального email.



---

`e.l.e.n.a@gmail.com` Изначально при подсчете длины логина учитывали точки, хотя на самом деле: `e.l.e.n.a@gmail.com` и `elena@gmail.com` — это один и тот же email.

- Раньше получал **EDI TED**
  - Теперь **NOT\_SURE**
- После учёта плюс-адресации понимаем, что реальный email — `elena@gmail.com` — некорректный: длина логина в домене `gmail.com` от 6 символов.

---

`elena+togetherMoreT  
han20Symbols@ya.ru` Изначально логин превышает максимальную длину. Но после учёта плюс-адресации и синонимов домена понимаем, что реальный email корректный: `elena@yandex.ru`.

- Раньше получал **NO T\_SURE**
- Теперь **EDITED**

---

Работает при использовании новых алгоритмов, обратитесь в поддержку HFLabs для подключения.

Технические подробности, не видимые заказчику

---

Если уже подключен новый `emailFilter`, то рекомендуем переобработать следующие email-адреса:

- статус `NOT_SURE`;
- содержат `@ya`, но не `@yandex.ru`;
- содержат `@gmail`;
- содержат в локальной части до `@` знаки: кавычки (`"`), точку (`.`), плюс (`+`).

## 20.02 — ЕГРЮЛ и ЕГРИП

### ***Проверяем актуальность учредителей и директоров компаний по ЕГРЮЛ***

Раньше «Фактор» проверял только основные реквизиты компании по ЕГРЮЛ.

Теперь научили «Фактор» по ФИО учредителя и ИНН/ОГРН организации проверять, входит ли всё ещё человек в состав учредителей этой организации.

Аналогично для руководителей.

Например, у [Станичного казачьего общества «Новокосино»](#) есть руководитель с должностью «атаман»: Расновский Василий Степанович, в открытых данных ЕГРЮЛ хранится и его ИНН.

«Фактору» можно передать на проверку ИНН организации и ФИО атамана, а в ответе получить должность Василия Степановича и коды качества: какие атрибуты совпали с ЕГРЮЛ, а что удалось восстановить.

Когда общество сменил атамана, «Фактор» скажет, что такой руководитель не найден. Это будет сигналом для обновления данных из ЕГРЮЛ в исходной системе.

На [хорошем сервере](#) проверка работает со скоростью 1000 записей в секунду.

Подробности в документации: [Проверка связанных ФЛ \(egrulSflFilter\)](#).

## 20.09 — Адреса

### ***Код ОКТМО от Росстата для отчёта в ЦБ по форме 260***

Технические подробности, не видимые заказчику

---

[PGC-15685](#) В отчёте о банкоматах по форме 260 **ЦБ РФ** требует код ОКТМО адреса банкомата. Код ОКТМО должен соответствовать [данным Росстата](#), ведь это первоисточник.

В ФИАС тоже есть поле «код ОКТМО», но проблема в том, что изменения поступают в базу ФИАС с опозданием на 3-6 месяцев. Когда Росстат присваивает новые коды ОКТМО, ФИАС не сразу подходит для подготовки отчётности: многие адреса не проходят проверку. Весной 2020 года порядка 4% адресов не проходило проверку из-за неактуальных данных в ФИАС.

Наш новый алгоритм проверяет актуальность кода ОКТМО из ФИАС при помощи справочника Росстата. Помимо актуальности проверяет принадлежность кода соответствующему муниципальному району, сельскому поселению, если адрес однозначно найден в справочнике Росстата.

Если код ОКТМО в ФИАС некорректный, то берём ОКТМО из Росстата.

Проставляем код качества полученного ОКТМО.

Подключаем в рамках доработки.

### ***Изменили сокращение для типа «городок»***

Технические подробности, не видимые заказчику

---

[PGC-15856](#) Раньше выводили сокращенное название такое же, как и полное — «городок».

В ФИАС от 07.08.2020 таких объектов 1170, и у них используется другое сокращение.

Мы сделали как в ФИАС, теперь выводим сокращенное название «г-к».

20.08 — Адреса

## ***Выпускаем новый справочник геокоординат чаще***

Технические подробности, не видимые заказчику

---

PGC-15822 Пользователи наших продуктов постоянно делятся с нами новой информацией по геокоординатам. Так, всем миром, мы улучшаем справочник.

Раньше мы обновляли справочник геокоординат нечасто, раз в квартал: процесс был ручной, и тестирование кропотливым.

Мы автоматизировали этот процесс. Теперь справочник выпускаем по мере появления новых данных, примерно раз в месяц.

## ***Подключили новый ФИАС с «городами-районами»***

Технические подробности, не видимые заказчику

---

PGC-15841, PGC-15863, PGC-15866

### **Для заказчиков с [EssenceCleanService](#)**

geoname\_id для переехавших в районы городов теперь будет от области.

Например, *Московская обл, г Наро-Фоминск*: geoname\_id был 523426, теперь 524925.

С 5 июня 2020 года в ФИАС произошли необычные изменения. Девять городов Московской области перешли в областное подчинение и теперь находятся на уровне «район».

Список «городов-районов»:

- Воскресенск;
- Дмитров;
- Истра;

- Клин;
- Наро-Фоминск;
- Ногинск;
- Одинцово;
- Орехово-Зуево;
- Сергиев Посад.

Пример: *Московская обл, г Одинцово, ул Акуловская*

<b>Регион</b>	<i>Московская обл</i>	<i>Московская обл</i>
<b>Район</b>		<i>г Одинцово</i>
<b>Город</b>	<i>г Одинцово</i>	
<b>Улица</b>	<i>ул Акуловская</i>	<i>ул Акуловская</i>

Сергиево-Посадского, Орехово-Зуевского, Одинцовского, Ногинского, Наро-Фоминского, Клинского, Истринского, Дмитровского и Воскресенского районов больше не существует.

Города и населенные пункты, подчинявшиеся району, теперь подчиняются городу.

Пример: *Московская обл, Одинцовский р-н, г Кубинка* *Московская обл, г Одинцово, г Кубинка*

<b>Регион</b>	<i>Московская обл</i>	<i>Московская обл</i>
<b>Район</b>	<i>Одинцовский р-н</i>	<i>г Одинцово</i>
<b>Город</b>	<i>г Кубинка</i>	<i>г Кубинка</i>
<b>Улица</b>	<i>ул Кленовая</i>	<i>ул Кленовая</i>

## **Изменение в определении уровня по ФИАС для «городов-районов»**

Теперь адрес *Московская обл, г Истра* получит **уровень по ФИАС — район** (FIAS\_DISTRICT). Раньше получал — **город** (FIAS\_CITY).

Также работает и для других «городов-районов».

## **Изменение кода полноты для адресов в «городах-районах» с планировочной структурой**

Изменилось значение кода полноты для адресов в «городах-районах» с планировочной структурой.

Пример: *Московская обл, г Дмитров, мкр им Владимира Махалина, д 3*

- Раньше этот адрес и похожие (до дома, с планировочной структурой в городе) получали **код полноты — полный** (GOOD).
- Теперь они получают код полноты — **не хватает квартиры** (UNDEF\_05).

## Недоступно горячее обновление на новый ФИАС для «Фактора» 20.07 и ранее

Горячее обновление на новый ФИАС доступно только для «Фактора» 20.08+.

С предыдущими версиями «Фактора» новый ФИАС несовместим.

### 20.01 – Адреса

## **Изменили формат ФИАС для «Фактора»**

Изменили внутренний формат ФИАС для «Фактора», он не совместим с предыдущими версиями «Фактора».

При скачивании ФИАС нового формата будет на 10% тяжелее, это нормально.

Блок не виден заказчику и не выводится при печати

---

Если справочник ФИАС поставляется отдельно от сборки «Фактора» (то есть сборка «Фактора» м ного меньше 800 Мб), то перед установкой «Фактора» 20.01+ скачайте актуальный ФИАС соответствующей версии и подложите его в директорию для обновления.

Для заказчиков, у которых подключен свой ФИАС, нужна будет регенерация справочника (sovzond, dellin).

## **Понимаем адреса с административными округами, которых нет в ФИАС**

Части адреса, которых нет в ФИАС, обычно мешают гарантированном разбору адреса.

Внутригородских районов и административных округов в ФИАС нет. Мы учли это в «Факторе» и это помогло избежать ручного разбора для таких адресов:

г Москва, ЦАО, Хамовники, турчанинов 6 стр 2 119034 Россия Москва г пер Турчанинов д 6 стр 2

- Было: **На ручной разбор**
- Стало: **Гарантированный**

Г. МОСКВА, ТиНАО, ВОРОНОВСКОЕ, САХАРОВО ДЕР. д.16 108830 Россия Москва г Вороновское п д Сахарово д 16

- Было: **На ручной разбор**
- Стало: **Гарантированный**

Санкт-Петербург метро Спортивная 190000 Россия Санкт-Петербург г

- Было: **На ручной разбор**
- Стало: **Гарантированный**

г Каменск-Уральский, **Синарский р-н**, победы 64 623428 Россия Свердловская обл г Каменск-Уральский пр-кт Победы д 64

- Было: **На ручной разбор**
- Стало: **Гарантированный**

Учтены 12 административных округов Москвы, а также и округа и районы других городов.

## 20.04 – Адреса

### ***Версии ФИАС теперь проходят строгий карантин***

В январе и феврале 2020 года обновлений ФИАС не было. С марта налоговая снова выкладывает обновления ФИАС <https://fias.nalog.ru/Updates>. Сейчас раз в неделю.

Как и раньше, выгружают версии с датой «минус три дня» от даты публикации. Если на сайте обновление от 20 апреля, то появилось оно 23 апреля.

В каждой новой выгрузке ФИАС есть свои проблемы. Поэтому мы не публикуем версии ФИАС в формате «Фактора» в автоматическом режиме, как прежде. Так будет до стабилизации ситуации. Сейчас новые версии мы проверяем автотестами, затем подозрительные области исследуем вручную. По возможности исправляем критичные ошибки и только после этого выкладываем в общий доступ.

Из мартовских версий мы выложили только ФИАС от 23 марта. Он совместим с прежними версиями «Фактора», но разбор некоторых адресов будет хуже. Чтобы сохранить уровень качества разбора адресов, мы учитываем новые особенности ФИАС в алгоритмах «Фактора». В ближайшие месяцы не рекомендуем обновлять ФИАС на горячую, а делать это вместе с обновлением версии «Фактора». Обновлять на горячую можно, если в новой версии есть что-то важное для бизнес процессов. Например, обновлённые коды ОКТМО. Но нужно быть готовым к тому, что большие группы адресов станут разбираться негарантированно, как случилось в ФИАС от 13 апреля с Адлерским районом и всеми его посёлками.

Чтобы качество адресов не просело, потребуется обновить «Фактор» на версию 20.04+.

Новое в формате ФИАС:

- Больше нет признака административного центра . Это приведет к изменению форматирования результата адреса у пользователей сервиса EssenceClean.

Было: **г Волгоград, пер им Кирова**

Стало: **Волгоградская обл, г Волгоград, пер им Кирова**

- Больше нет признака состояния дома (Состояние дома).
- Нет КЛАДР-кодов для 749 планировочных структур и 106 улиц. КЛАДР-код будет пустой для этих записей.

По возможности мы исправим ошибки в ФИАСе. О выходе новых версий ФИАС для «Фактора» сообщим в документации и в поддержке по запросу.

## 20.05 — Адреса

### ***Вернули КЛАДР-коды планировочным структурам***

В ФИАС код КЛАДР для каждой записи хранится в виде полного кода одной строкой и по частям — код каждого компонента в отдельном поле.

Раньше мы формировали код КЛАДР из отдельных частей. Полный код одной строкой не использовали.

В 2020 году в ФИАС начали теряться КЛАДР-коды для объектов, перенесенных с одного уровня на другой. В результате получилось, что для 1000 планировочных структур код КЛАДР планировочной структуры не заполнен в отдельном поле.

Мы изменили логику на своей стороне. Теперь, в случае ошибки в ФИАС, берем полный код КЛАДР одной строкой.

Порядка 150 адресов из 1000 не удалось исправить, необходимо исправление в ФИАС. Мы сообщили ФНС о проблеме.

### ***Новые типы адресных объектов***

В ФИАС от 18 мая 2020 и далее появились новые типы адресных объектов.

ЖЕЛЕЗНОДОРОЖНЫЙ БЛОКПОСТ	ж/д бл-ст	ФИАС	Амурская обл, Магдагачинский р-н, железнодорожный блокпост Красная Падь
САМОУПРАВЛЕНИЕ	с/упр	дополнительные справочники иностраннх адресов	Литва, Вильнюсский уезд, Электренское самоуправление
ГОРОДСКОЕ САМОУПРАВЛЕНИЕ	г. с /упр	дополнительные справочники иностраннх адресов	Литва, Клайпедский уезд, Палангское городское самоуправление, п Швянтойи

## 20.03 — заметки к релизу

### ***Логирование***

С версии «Фактора» 20.02+ добавили новые подробные логи для мониторинга работы «Фактора».

Их формат может меняться, поэтому не стоит настраивать автоматическое считывание данных.

Для подключения к системам мониторинга мы добавим API со стабильным контактом.

Детальное описание логов здесь: [Журналы первичной диагностики](#).

## 20.02 — заметки к релизу

### ***Проверяем выполнение системных требований***

У нас есть требования к настройке ОС linux: [Настройка параметров ОС \(20.01 и раньше\)](#). Когда они не выполнены или значения отличаются от требований, «Фактор» работает медленнее под высокой нагрузкой.

Чтобы предупредить такие ошибки, начали проверять настройки при старте «Фактора»:

- При несоответствии будет ошибка в логе.
- Начиная с релиза 20.05, «Фактор» не запустится, если требования не выполнены.

Рекомендуем после установки обновления «Фактора» 20.02+ проверить логи на наличие ошибок.

### ***Унифицированная структура каталогов***

С этой версии все продукты HFLabs — «Фактор», «Подсказки» и «Единый клиент» — получили единую структуру каталогов. В будущем это упростит жизнь как техническим специалистам заказчика, так и техподдержке.

Поскольку инструкции до версии 20.02 использовали собственную структуру каталогов, в инструкцию по обновлению на 20.02 мы добавили пункты, которые переведут «Фактор» на новую унифицированную структуру. Не пропустите их, потому что начиная с 20.02 вся документация будет построена на новой структуре каталогов.

## 20.04 — заметки к релизу

### ***Проверяем оставшееся место на диске***

Когда на сервере с «Фактором» заканчивается место на жёстком диске, то возникают различные ошибки. Чаще всего не работает поиск дубликатов, потому что ему требуется больше всего места. Обычно такие ошибки непросто диагностировать. Чтобы причину проблем найти как можно раньше, мы добавили в «Фактор» проверки на объем оставшегося места на диске. Проверки по умолчанию выключены. Поможем настроить в рамках поддержки.



## Сообщаем о нехватке свободного места на диске

Добавили возможность выводить предупреждения и ошибки в `factor.log` при переходе пороговых значений для отслеживания свободного места на диске.

Есть 2 режима настройки:

- выводить WARN при переходе значения, например, 7Гб:  
`[WARN] (DiskSpaceHealthWatcher) Free disk space on 'tmp' by path '/tmp/factor' below WARN threshold. Available: 6 GB (threshold: 7 GB)`
- выводить ERROR при переходе значения, например, 5Гб:  
`[ERROR] (DiskSpaceHealthWatcher) Free disk space on 'tmp' by path '/tmp/factor' below ERROR threshold. Available: 2 GB (threshold: 5 GB)`

Когда место освободили:

- сообщения об ошибках и предупреждения перестанут выводиться в лог;
- «Фактор» запишет сообщение, что места достаточно:  
`[INFO] (DiskSpaceHealthWatcher) Free disk space on 'tmp' by path '/tmp/factor' normalized. Available: 7 GB`

Директории, для которых можно настроить отслеживание места:

- для справочников черных списков;
- для работы с хешами дедупликации;
- для индексов;
- общая для обновления справочников;
- для временных файлов работы «Фактора».

По умолчанию поведение выключено, обратитесь в службу поддержки для подключения.

## Динамически записываем количество свободного места в `factor-stats-avg.log`

Для динамического мониторинга свободного места на диске добавили в лог статистики `factor-stats-avg.log` показатели *свободного/общего* размера директорий:

- `blacklist` — справочники модуля Черные списки;
- `dedup` — директория работы с хешами дедупликации;
- `idx` — директория индексов;
- `dct` — директория подключаемых справочников;
- `tmp` — временная директория.

## ***Логирование при срабатывании таймаута обработки, заданного для типа данных***

Когда запись на обработку содержит несколько типов данных, можно настроить свой таймаут на каждый из этих типов: адреса, телефоны или компании.

Мы улучшили логирование ситуаций, когда срабатывает таймаут на обработку части записи.

Раньше:

- в `factor-stats-api.log` в колонке `timeout` выводили число превышений таймаутов, заданных через параметры: `clean.rt.timeout`, `factor.clean.timeout.record`, `factor.clean.timeout.single`.
- сообщения о срабатывании таймаута выводились в `factor.log`

Сейчас:

- в `timeout` учитывается еще и количество запросов на которых отработал таймаут-фильтр.
- выводим сообщение о сработавшем таймауте в `factor-longops.log`:  
**factor-longops.log:**

Полностью переписали `timeoutfilter`. Также добавили возможность при срабатывании основного таймаута применить более простой и быстрый алгоритм.

## ***Логирование при срабатывании таймаута обработки запроса целиком***

Ограничение на обработку всего запроса можно настроить только для стандартизации данных в сервисах `CleanService` и `SiebelConnectorToFactor`.

Раньше сообщение о срабатывании таймаута при выставленных параметрах:

- `factor.clean.timeout.record` , `factor.clean.timeout.single` — и в `CleanService`, записывали в `factor.log`.
- `clean.rt.timeout` — интерфейс `SiebelConnectorToFactor`, в `factor-siebel-online-НазваниеСущности.log`.

Сейчас для обоих интерфейсов информация о срабатывании таймаута записывается в `factor-longops.log`:

**factor-longops.log:**

В ответе отдаем статус `500Internal Server Error` и сообщение об ошибке:

## **Статистика**

- Обновили 64 справочников, не считая ежедневных ЕГРЮЛ и ЕГРИП.
- Закрыли 301 запрос в поддержке Фактора.

## Проверяем e-mail по правилам их создания на mail.ru, gmail.com и других популярных сервисах

Раньше Фактор проверял формат e-mail только на соответствие стандарту RFC 2822.

Многие почтовые сервисы дополняют стандартные ограничения своими правилами, поэтому не все e-mail правильно получали код качества *Корректный*.

Теперь отмечаем e-mail на ручную проверку, если он не соответствует правилам почтового сервиса.

Мы добавили:

### 1. Проверку e-mail по правилам популярных почтовых сервисов:

- yandex.ru, ya.ru
- mail.ru, list.ru, bk.ru, inbox.ru
- hotmail.com
- yahoo.com
- gmail.com
- icloud.com

### 2. Распознавание кириллических логинов для доменов:

- письмо.рф
- е-письмо.рф

Исходный e-mail	Код качества		Комментарий
	Был	Стал	
Иванов.И.И@письмо.рф	Некорректный	Корректный	письмо.рф разрешает кириллицу и точки
Ivanov_I_I@gmail.com	Корректный	Некорректный	gmail запрещает нижнее подчеркивание

Если вы уже проверяете e-mail в Факторе, то проверка заработает автоматически после обновления на 8.9.

## Фактор проверяет существование имени, когда переводит имена с латиницы

Полностью переделали алгоритм перевода ФИО с латиницы на кириллицу.

Задача перевода в общем случае нерешаема, потому что одна и та же латинская буква может означать разные русские, например G — это Г или Ж, IA — это либо Я, либо ИА.

Одно и то же имя можно написать разными способами. Знакомьтесь, Юлия Юрьевна:

- Uliа Ur'evna
- Iuliia Iurevna
- Yulia Yurevna

Раньше мы сложные случаи добавляли в словарь перевода. Всех случаев в словаре не учесть, поэтому часто получалось неправильно.

Теперь мы одновременно переводим всевозможными способами и проверяем результат по справочникам русских фамилий, имён и отчеств. Если попали однозначно, результат отдаём. Если возможны несколько вариантов, то выбираем один и ставим специальный [код качества](#).

Новый алгоритм учитывает [12 стандартов](#) транслитерации, их комбинации и то, как на самом деле люди пишут свои имена.

Если вы уже использовали старый вариант перевода ФИО — обновим в рамках поддержки. Если не использовали, то подключим как доработку. Обращайтесь к сотруднику HFLabs.

## Поможем сэкономить на рассылке sms

Наши заказчики рассылают SMS через посредников либо напрямую через операторов связи.

За услуги посредников надо доплатить, поэтому напрямую через операторов рассылать дешевле.

Определить оператора номера телефона можно по данным [Россвязи](#). Но Россвязь не сообщает, когда человек сменил оператора с сохранением номера. Смена операторов отражается в распространяемой ЦНИИ Связи базе данных перенесённых номеров ([БДПН](#)).

Мы подключили БДПН и теперь Фактор знает обо всех 7 млн переносах.

Отдаём обновлённые данные в поле "оператор":

- Если вы его получаете, то правильный оператор определится сразу после установки Фактор 8.7.
- Если поле "оператор" ещё не используете, то подключим как доработку.

Также можем выводить признак, менял ли абонент оператора.

Справочник **выкладываем каждый день**, обновляйте без перезапуска Фактора.

## **Переобработайте адреса — станет больше хороших**

В релизах 8.3, 8.4 и 8.6 доработали алгоритмы адресов для учёта часто встречающихся особенностей.

Чтобы стало больше гарантированных адресов, необходимо обработать адреса со статусами `Not_validated_has_ambi` и `Not_validated_has_unparsed_parts`.

На обработку обязательно отправлять исходные строки с адресами, а не текущий результат разбора.

## **Определяем ближайшие станции метро**

Для курьерских служб мы добавили определение трёх ближайших станций метро.

Исходный адрес: г Москва, Стремянный пер, д 38.

Координаты, полученные Фактором на основе адреса: `GEO_LAT = 55.7275357`, `GEO_LNG = 37.6270583`.

Информация о метро получаем по координатам, дополняем названием ветки:

0,2км до м. Серпуховская (Серпуховско-Тимирязевская)

0,3км до м. Добрынинская (Кольцевая)

0,7км до м. Павелецкая (Кольцевая)

Расстояние до метро определяем по прямой.

Сделано для всех городов России, где есть метро: Москвы, Санкт-Петербурга, Казани, Самары, Новосибирска, Екатеринбурга, Нижнего Новгорода.

## Улучшения адресов

### ***Улучшаем почтовый индекс***

Индекс для почтовой рассылки нужен для точной доставки писем и посылок.

Когда индекс не удаётся определить по данным Почты России и ФИАСу, оставляем исходный индекс. Но исходный индекс может противоречить разобранному адресу, поэтому теперь, если регион индекса отличается от региона адреса, то исходный индекс не попадёт в результирующий адрес.

Значения кода качества индекса для рассылки изменились! Нужно переделать интеграцию, если используете список кодов качества индекса. Новые значения соответствуют реальным уровням ФИАСа: [сравнение старых и новых значений](#).

### ***Внутригородские районы в адресах***

Учли, что в адресе люди могут написать реальный район города, которого пока нет в ФИАС. Теперь даже если исходно указан район города не из ФИАС, он не мешает разбору.

Адрес *Санкт-Петербург, Центральный микрорайон, ул Мира, 12* станет *г Санкт-Петербург, ул Мира, д 12* со статусом *гарантированный*.

### ***Улучшили разбор неполных адресов***

Раньше если город пропущен и есть улица, адрес мог быть гарантированным только в одном случае: исходный индекс точно указывает на пропущенный город.

Теперь для адресов с пропущенным городом мы учитываем исходный индекс, даже если он неточный. Определяем регион по первым трём цифрам индекса и ищем варианты уже внутри региона. Если внутри региона такая улица одна, то адрес разберётся гарантированно. Если таких улиц несколько, Фактор отдаст несколько вариантов внутри одного региона.

Примеры гарантированных адресов, когда в регионе есть лишь одна такая улица:

- 197379 УЛИЦА ДОЛГООЗЁРНАЯ 12 197373, г Санкт-Петербург, ул Долгоозерная, д 12

Примеры неоднозначных адресов:

184209, КНЯЖЕГУБСКАЯ — в Мурманской области улица Княжегубская есть в двух населённых пунктах:

- Мурманская обл, Кандалакшский р-н, нп Зареченск, ул Княжегубская
- Мурманская обл, Кандалакшский р-н, пгт Зеленобор Княжегубская

108999 УЛИЦА БОГОРОДСКАЯ — Фактор предлагает 4 варианта разбора:

- Московская обл, г Ногинск, ул Богородская
- Московская обл, Щелковский р-н, деревня Назимиha, ул Богородская
- г Москва, г Троицк, ул Богородская
- г Москва, поселение Первомайское, деревня Пучково, ул Богородская

Не работает для популярных названий.

## ***Добавили безопасный учёт опечаток***

москва тулчанинов 119034, г Москва, пер Турчанинов

## ***Понимаем коды регионов в адресах***

Регион адреса хранится в виде кода во многих системах. Из них в Фактор на обработку приходят адреса с цифрами вместо слова.

? Неизвестное вложение

Раньше, чтобы перевести код региона в наименование, требовалось подключить специальную настройку. Иначе адрес уходил на ручную проверку.

Теперь разбираем гарантированно из коробки:

- ,191028,78, УЛ. МОХОВАЯ,ДОМ 78 г Санкт-Петербург, ул Моховая, д 78
- 05 Мажалис Батырая дом 14 Респ Дагестан, село Мажалис, ул Батырая, д 14

## ***Расклеиваем типы в адресах***

Регулярно мы встречаем адреса, в которых тип склеен с наименованием: *гМосква, улМоховая*. Так происходит из-за ошибок при перекладывании из одной системы в другую. Ошибку давно исправили, а слипшийся адрес остался в хранилище и пришёл на обработку в Фактор.

? Неизвестное вложение

Раньше мы разбирали адрес неполностью и отправляли его на ручную проверку.

Сейчас разбираем полностью и гарантированно:

- МОСКВА УЛКРАСНОГОМАЯКА Д,10 г Москва, ул Красного Маяка, д 10
- респБашкортостан гИшимбай улБогданаХмельницкого д16 кв35 Респ Башкортостан, г Ишимбай, ул Богдана Хмельницкого, д 16, кв 35

Не будет работать, если одновременно допущена опечатка или тип передан в нестандартном виде.

## **Новые поля по адресу из ФИАС**

### ***Получаем код ИФНС для физлиц из ФИАС***

Раньше по адресу получали из ФИАС только код ИФНС для юрлиц. Для отправки налоговой отчётности в правильную налоговую добавили ещё и ИФНС для физлиц.

Логика осталась прежней: если для объекта код есть, то возвращаем его, если нет — берём от "родителя".

Подключаем по запросу как доработку.

### ***Собираем классификационный код ФИАС***

Это цифровой код, похожий на код КЛАДР.

Нужен для:

- понимания адресообразования;
- интеграции с ведомствами.

Код выглядит так:

СС+PPP+ГГГ+ППП+СССС+УУУУ+ДДДД

- СС — код региона;
- PPP — код района;
- ГГГ — код города;
- ППП — код населенного пункта;
- СССС — код элемента планировочной структуры;
- УУУУ — код улицы;
- ДДДД — счетчик домов.

Например, *Московская обл, Ленинский р-н, с/п Развилковское, п Развилка, снт Поляна, ул Вишневая, д 34* 50014005002001300260001.

Подключаем по запросу как доработку.



## ***Получаем состояние дома из ФИАС***

Определяем состояние дома.

Состояние дома пригодится для понимания, является ли дом новостройкой или нежилым зданием. Например, в ФИАС 120 тысяч домов (0,5%) — новостройки.

**Заполненность** в ФИАС оставляет желать лучшего: 97,4% домов имеют статус без особого состояния.

Подключаем по запросу как доработку.

## **Стандартизация компаний**

### ***Дополняем проверки ИНН***

Первые четыре цифры ИНН содержат код региона и номер налогового органа, проверяем их по перечню **СОУН**.

В ЕГРЮЛ и ЕГРИП есть ИНН, которые не проходят проверку по СОУН. В налоговой мы узнали, что это корректные ИНН и такие номера налоговых существуют.

Мы расширили справочник СОУН в Факторе дополнительными кодами СОУН: 9701, 9710, 9715, 9717, 9718, 9721, 9723, 9729.

ООО "СМАРТ ГРУПП", ИНН 9721048647

- Был код качества ИНН *NOT\_VALID\_TAX\_CODE* — некорректный код СОУН.
- Стал код качества *GOOD* — хороший ИНН.

Добавили только массовые случаи. Например, с кодом 9721 в ЕГРЮЛ 4877 компаний.

В ЕГРЮЛ/ЕГРИП есть и ИНН, начинающиеся на другие «коды налоговых», которых нет в СОУН, например, 2040. Но их не добавили, потому что таких в ЕГРЮЛ по 1-2 компании и они все уже ликвидированы.

### ***Выводим подробный статус компании из ЕГРЮЛ и ЕГРИП***

Раньше Фактор отдавал только статус: действующая / в стадии ликвидации / ликвидирована.

Сейчас ставим дату ликвидации и в дополнение передаём подробный статус из ЕГРЮЛ и ЕГРИП.

Филиалам ставим статус головной компании.

Примеры:

МУНИЦИПАЛЬНОЕ ПРЕДПРИЯТИЕ "ДЕТАЛИ БЫТОВОЙ ТЕХНИКИ"

- ОГРН: 1086315007151
- Статус: Ликвидирована
- Дата ликвидации: 2008-07-25
- Подробный статус: "Прекращение деятельности юридического лица в связи с исключением из ЕГРЮЛ на основании п.2 ст.21.1 Федерального закона от 08.08.2001 №129-ФЗ".

Индивидуальный предприниматель ВИСОТИНА ОКСАНА СЕРГЕЕВНА

- ОГРНИП: 304244710300031
- Дата прекращения деятельности: 2017-03-09
- Подробный статус: Индивидуальный предприниматель прекратил деятельность в связи с принятием им соответствующего решения

## ***Учитываем больше адресов при поиске юрлиц в ЕГРЮЛ***

Счастливым обладателям [ЕГРЮЛ-фильтра](#).

Раньше при поиске учитывали только один адрес, а теперь сколько угодно. Это поможет найти больше компаний, потому что:

- в данных клиентов-юрлиц фактический и юридический адрес часто перепутаны местами.
- в ЕГРЮЛ у филиалов два адреса: местонахождения филиала и юридический адрес головной компании.

Чем больше адресов передано, тем ниже скорость: адреса стандартизируются перед поиском. Поэтому лучше ограничиться двумя или тремя адресами.

Настраиваем как доработку по запросу.

## ***Находим ИП даже с опечаткой в имени или после смены фамилии***

Раньше Фактор находил предпринимателя в ЕГРИП только если в дополнение к ИНН/ОГРН точно совпало ФИО.

Не все ИП находились по таким правилам, потому что:

- В реальных данных есть опечатки и инициалы.
- При смене фамилии в ЕГРИП — новая, а в базе контрагентов — старая.

Мы добавили сравнение ФИО по частям, теперь Фактор находит ИП при совпадении только фамилии или имени и отчества:

- ***Иванов, ОГРН 315344300056295 ИП Иванов Евгений Александрович, ИНН 344309531858, ОГРН 315344300056295***

- *Сергеева Антонина Петровна, ИНН 790105092140 ИП Аниканова Антонина Петровна, ИНН 790105092140, ОГРН 309272115600013*

Настраиваем как доработку по запросу.

## **Поддержка изменений в справочниках налоговой**

### ***Защитились от сбоев налоговой, или Новый формат выгрузок ЕГРЮЛ и ЕГРИП***

Зимой было два типа сбоев в сервисе налоговой службы для выгрузки данных из ЕГРЮЛ и ЕГРИП:

Во второй половине января совсем не было обновлений для ЕГРЮЛ и ЕГРИП. По телефону нам объясняли, что неполадки связаны с проведением технических работ. Возможно, это была часть подготовки к переходу на новый формат выгрузки: с 1 апреля ФНС переходит на новый формат файлов с данными ЕГРЮЛ и ЕГРИП ([приказ от 12 января 2017](#)). На самом деле данные уже выгружаются в новом формате: по ЕГРЮЛ — с 31.01.2017, а ЕГРИП — с 10.03.2017.

Несколько раз обновления дописывали в папку со вчерашними данными. Налоговая выкладывает изменения в директории с датами. Каждый день — новая директория. Например, для ЕГРИП: 21 февраля была информация 12 тысячах изменений, а 22 февраля — уже о 30 тысячах. Это значит, что, если не посмотреть в директории за прошедшие дни, то можно было потерять изменения по 18 000 индивидуальных предпринимателей.

Мы в Факторе поддержали новый формат и настроили полную синхронизацию с каталогами налоговой, чтобы получать даже те обновления, которые выложены задним числом.

### ***Поддержка нового формата ФИАС***

С 13 февраля налоговая изменила формат выгрузки **ФИАС в виде dbf**. Теперь вместо одного файла addrobj.dbf в архиве 86 файлов вида addrobj1.dbf, addrobj2.dbf ... .

Мы поддержали изменение. ФИАС для Фактора можно скачать по [ссылке](#) и обновить по [инструкции](#). Неважно, стоит у вас 8.1 или 7.0 — везде будет работать.

### ***Изменения в ФИАСе в апреле***

В ФИАСе появился уровень зданий под номером 8, пока пустой. А ещё поменялись названия файлов, выкладываемых налоговой: ADDROBJ\*ADDROB\*

Поддержали эти изменения при генерации ФИАСа для Фактора.

## Обнаружили неверные индексы в старых ФИАС

В апрельских ФИАС были массовые ошибки в почтовых индексах. Если используете ФИАС от 10 апреля, то лучше [обновить](#) на майский или июньский.

Или просто обновите Фактор на версию 8.4, в ней корректный ФИАС.

## Различаем дубли в ФИАС

Для 3000+ населённых пунктов и улиц в ФИАС есть дубликаты: их названия, типы и родительские объекты полностью совпадают.

Проблема была в том, что Фактор возвращал неоднозначность и предлагал два одинаковых варианта — выглядит как полный бред, потому что при ручной обработке невозможно выбрать из двух одинаковых адресов.

Оказалось, что только в ~20% случаев в ФИАС задублирована одна и та же улица или деревня. Остальные записи только выглядят одинаково, но относятся к физически разным объектам. Мы ещё научимся отдавать уточнения по физически разным объектам, чтобы было проще. Сейчас мы научились выделять точные дубликаты и перестали ставить им статус *неоднозначный*. Вот ключевые правила, по которым мы считаем адреса абсолютными дублями:

- Индексы и коды ОКАТО совпадают — выбираем любую запись.
- Оба индекса пустые, коды ОКАТО совпадают — выбираем любую запись.
- Один из индексов пустой, коды ОКАТО совпадают — выбираем запись с индексом:

*г Калининград, ул Заводская*

В ФИАСе две записи с одинаковым ОКАТО:

Адрес	Индекс	Код КЛАДР	Код ФИАС	ОКАТО
Россия, обл Калининградская, г Калининград, ул Заводская	236028	39000001000028000	4d7b1772-3d91-4b17-b8ee-c17e48ddf330	27401373000
Россия, обл Калининградская, г Калининград, ул Заводская	-	39000001000120300	a1455f59-ca57-4104-ba11-e45b234f391d	27401373000

Выбираем запись с индексом, ставим статус проверки **VALIDATED** — гарантированный, все идентификаторы из ФИАС берём от верхней записи.

Если индексы либо коды ОКАТО **не** совпадают — перед нами 2 разных объекта.

Почему так? Отличающиеся ОКАТО означают, что улицы относятся к разным объектам территориального деления. Это одинаковые улицы в разных районах города, либо улицы в деревнях, которые полностью влились в город.

Подробнее на странице с [новостями 8.3](#).

## **Поддержали новую структуру ФИАС**

ФИАС — эталонный справочник адресов, в нём адреса хранятся в структурированном виде. Каждой части адреса соответствует свой уровень: регион, район, город, населённый пункт, улица, дом.

Рассмотрим пример: *188691 Ленинградская обл, д Кудрово, мкр Новый Оккервиль.*

В этом адресе есть:

- регион — *Ленинградская обл,*
- населённый пункт — *д Кудрово,*
- и микрорайон — *мкр Новый Оккервиль.* Стоп. У нас же нет уровня «микрорайон», есть «улица». Но какой-то необычный у неё тип — микрорайон?

В ФИАС *мкр Новый Оккервиль* относится к дополнительному уровню — «планировочной структуре».

На этом уровне располагаются дачные товарищества и микрорайоны. Сейчас 146 тысяч таких объектов.

Мы поддержали уровень планировочной структуры в Факторе, чтобы понимать объекты на этом уровне.

- Для адресов в формате КЛАДР, где не так много уровней, выводим планировочные структуры в населённый пункт или улицу — зависит от того, к какому объекту принадлежит.  
Пример: *188691 Ленинградская обл, д Кудрово, мкр Новый Оккервиль.* Мкр Новый Оккервиль — планировочная структура. Кудрово — деревня на уровне населённых пунктов. Выводим мкр Новый Оккервиль в поле для улицы.
- Если вы храните адреса в формате ФИАС, можем выводить новый уровень в отдельное поле.

Новый ФИАС не совместим со старыми версиями Фактора, потому что изменилась структура полей. Если самостоятельно обновляете ФИАС, то проверьте совместимость версий по [инструкции](#).

## **КЛАДР превратится в тыкву в новогоднюю ночь**

Налоговая опубликовала [объявление](#).

С 1 января 2018 года КЛАДР исчезнет и останется только ФИАС. Фактор работает с ФИАС с 2015 года. Если ваши системы ещё используют КЛАДР, удалите его и используйте Фактор и Подсказки для приведения адресов к ФИАС. Наша статья о переходе с КЛАДР на ФИАС: <https://habrahabr.ru/company/hflabs/blog/333736>

## Для разработчиков

### ***Упростили автоматизацию скачивания ФИАС для Фактора***

Стабилизировали пути, по которым надо скачивать ФИАС для Фактора.

В инструкции по [обновлению справочника ФИАС](#) см. примечание "Разработчикам"

### ***Новые параметры сервиса поиска адреса***

В Факторе есть способ вручную поискать адреса в ФИАСе. Он используется в [РМАД](#) и в [Едином клиенте](#).

Также сервис поиска можно вызывать из других систем. Он поможет, если нужно найти город /населённый пункт по точному совпадению и без подчинённых объектов. В остальных случаях используйте [Подсказки](#).

С версии 8.1 можно сократить число адресов в ответе:

- выводить только актуальные адреса,
- не выводить подчинённые объекты.

Новые параметры и примеры вызова сервиса поиска адреса — в [документации](#).

## Для поддержки

### ***Перешли на java 8***

Обновили [требования к платформе](#). Для Фактора 8.8 и далее требуется java 8.

### ***Упростили скачивание справочника ФИАС***

Если вы самостоятельно обновляете ФИАС для Фактора, то теперь будет проще найти нужную версию.

Раньше для двух и более версий Фактора подходил один и тот же ФИАС. Теперь для каждой версии Фактора есть свой ФИАС.

Инструкция [обновление справочника ФИАС](#) стала проще.

## ***Пополнили базу знаний о проблемах интеграции с Сибель***

Проблемы при настройке коннектора

## ***Добавили версию чёрных списков***

На [web-страницу](#) с версиями справочников добавили дату каждого чёрного списка:

Версии справочников	
phone	2017-02-16
kladr	2017-02-16
geoCodes	2017-01-26
fias	2017-02-09
BlackList / terrorists.csv	2014-11-27
BlackList / Factiva.csv	2015-12-24

## ***Переделали логирование Фактора***

1. Появилась возможность записывать долгие запросы в отдельный файл. По умолчанию выключено.
2. Для расследования сложных проблем можно временно понизить уровень логирования без перезапуска Фактора.