СОДЕРЖАНИЕ

ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК «ФАКТОРА» 2

Модули стандартизации и обогащения 4
Возможности модуля «Адреса» 18
Возможности модуля «Телефоны» 24
Возможности модуля «ФИО и пол» 27
Возможности модуля «Паспорта» 28
Возможности модуля «E-mail» 30
Возможности модуля «Даты» 33
Возможности модуля «ЕГРЮЛ и ЕГРИП» 34
Возможности модуля «Компании» 35
Возможности модуля «Автомобили» 37
Возможности модуля «Общий» 38

Модули идентификации 40
Модуль «идентификация дубликатов» 42
Модуль «проверка по чёрным спискам» 43

Новости «Фактора» 44

ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК «ФАКТОРА»

Описание функциональных возможностей «Фактора»

«Фактор» — серверный программный продукт. «Фактор» помогает автоматизировать процессы проверки и очистки информации о физических и юридических лицах. Обновляет и исправляет некорректные контактные данные с учётом опечаток. Приводит данные к единому виду и находит дубликаты физических и юридических лиц.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ОСНОВНЫЕ ФУНКЦИИ И ОСОБЕННОСТИ

- Обрабатывает контактные данные, содержащие ошибки, опечатки и неактуальные компоненты.
- Проверяет, очищает, разбивает по компонентам и приводит к единому формату контактные данные (почтовые адреса, ФИО, телефоны и другие).
- Выявляет дубликаты физических и юридических лиц на основании контактной информации.
- Использует настраиваемые эвристики и алгоритмы, пополняемые справочники и словари для обработки различных видов данных.

Модули стандартизации и обогащения

Возможности, алгоритмы, ограничения, нормативные акты и справочники, используемые в модулях.

- Возможности модуля «Адреса»
- Возможности модуля «Телефоны»
- Возможности модуля «ФИО и пол»
- Возможности модуля «Паспорта»
- Возможности модуля «E-mail»
- Возможности модуля «Даты»
- Возможности модуля «ЕГРЮЛ и ЕГРИП»
- Возможности модуля «Компании»
- Возможности модуля «Автомобили»
- Возможности модуля «Общий»

Модули идентификации

Сведения о функциональности модулей выявления похожих записей.

- Модуль «идентификация дубликатов»
- Модуль «проверка по чёрным спискам»

Новости «Фактора»

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

МОДУЛИ СТАНДАРТИЗАЦИИ И ОБОГАЩЕНИЯ

- Адреса
- Другие типы данных
    - Компании
    - Поиск компании в ЕГРЮЛ и предпринимателя в ЕГРИП
    - Телефоны
    - ФИО и пол
    - Паспорта
    - Даты
    - E-mail
    - Марка, модель и тип автомобиля
    - Формат VIN-номера
    - Формат госномера
- Базовые правила
    - Обратная транслитерация
    - Подробное логирование для мониторинга

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

АДРЕСА

Описание

- Разбиваем адреса на компоненты.
- Исправляем опечатки и сокращения.
- Приводим компоненты к единому формату.
- Переводим с латиницы на кириллицу.
- Восстанавливаем пропущенные компоненты: «197373 УЛИЦА ДОЛГООЗЁРНАЯ 12» «г Санкт-Петербург, ул Долгоозерная, 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» «117519, 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 стандартных уровней адреса;

  идентификатор ФИАС земельных участков и квартир;

  код ИФНС для юрлиц и физлиц 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»;

  коды ОКСМ, ОКТМО и ОКАТО по адресу, проверяем и корректируем ОКТМО по данным Росстата для отчётности по форме 260 в ЦБ.

  кадастровый номер дома, квартиры, участка из ФИАС.

- Проставляем коды качества для распознанного адреса:

  код полноты адреса;

  статус проверки;

  признак актуальности;

  код качества сделанных изменений;

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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 место рождения к требованиям Указания Банка России N 4936-У.
- Приводим адрес к требованиям Указания Банка России N 4990-У о формате реестра вкладчиков.
- Выводим дату изменения адреса по ФИАС/ГАР.

Ограничения

- Иностранные адреса поймем максимум до уровня города.
- Российские адреса обязательно проверяем по справочнику адресов до уровня улицы, а дома уже может не быть в ФИАС/ГАР. Если дом не из ФИАС/ГАР, то он будет разобран. Для проверки, есть ли дом в ФИАС, проставляем код соответствия ФИАС.
- 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».
- Не поддерживаем в домовой части признак владения «домовладение» — в ответе вернем «дом».

Используемые материалы

- ФИАС ФНС. Можно обновлять без перезапуска Фактора.
- 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.

Описание

- Конвертируем адреса в геокоординаты с указанием точности.
- Находим ближайшие к адресу клиента филиалы.
- Вычисляем положение объекта относительно МКАД и КАД: внутри или снаружи.
- Определяем ближайшие станции метро. Сделано для всех городов России, где есть метро: Москвы, Санкт-Петербурга, Казани, Самары, Новосибирска, Екатеринбурга, Нижнего Новгорода.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Можно обновлять справочник геокоординат без перезапуска «Фактора».

Ограничения

- Геокодируем только те адреса, которые есть в ФИАС.
- Покрытие адресов из ФИАС геокоординатами не полное.
- Положение относительно МКАД и КАД — только для адресов Москвы и области, Санкт-Петербурга и области.
- Справочник границ МКАД и КАД, а также справочник метро нельзя обновить нагорячую.
- Расстояние до метро определяем по прямой.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ДРУГИЕ ТИПЫ ДАННЫХ

Компании

Наименования компаний

- Формируем название компании-ключа для последующей идентификации дубликатов.

- Извлекаем ОПФ из наименования.
- Исправляем часто встречающиеся опечатки в ОПФ.
- Расшифровываем аббревиатуры ОПФ.
- Находим ОПФ в начале, середине и конце наименования.
- Разделяем слитно написанные ОПФ и наименование.
- Если находим несколько ОПФ, проверяем их на конфликт.

  Допустимые сочетания нескольких ОПФ отмечаем отдельным маркером, выбираем итоговую ОПФ по настройке.

  Неоднозначные сочетания ОПФ отмечаем отдельным маркером, ставим статус NOT_SURE.

- Определяем, что компания относится к судам или государственным учреждениям.

  Отмечаем такие наименования специальными маркерами.

  Распознаем такие наименования как не требующие ОПФ.

- Отличаем иностранные наименования от российских: распознаем такие наименования как не требующие ОПФ.
- Извлекаем из наименования реквизиты компании:

  ИНН;

  КПП;

  ОГРН;

  Номер лицевого счета.

- Удаляем мусор.
- Ставим маркеры.
- Удаляем кавычки, если нет ОПФ (доступно с версии «Фактора» 24.6+).

Ограничение

Фильтр не стандартизирует наименование компании до того вида, как оно записано в ЕГРЮЛ-е.

Другие атрибуты

- Заменяем старые ОКОГУ на ОКОГУ, действующие с 1 января 2012 года.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- 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 с учетом ОПФ, 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 сравнение 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.

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
- Дату регистрации.
- Статус: активный/ликвидированный.
- Код ОКОПФ
- Юридический адрес для ЮЛ и город регистрации для ИП.

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

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Используемые материалы

ЕГРЮЛ и ЕГРИП по данным от ФНС. Обновляем раз в месяц. Можно обновлять без перезапуска «Фактора».

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, «11111111»).
- 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, характеризующие качество номера:

    Перенесённые номера.

    Красивые номера.

    Изменённые «Фактором».

    Иностранные.

    Неоднозначные.

    Слишком длинные или слишком короткие.

- 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

- Иностранные 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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Используемые материалы

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

ФИО и пол

- Раскладываем ФИО, пришедшее одной строкой, на фамилию, имя и отчество.
- Определяем пол.
- Исправляем опечатки по похожести и по справочникам замен.
- Разворачиваем уменьшительно-ласкательное имя в полное.
- Разбираем тюркские имена и отчества.
- Понимаем фамилии, похожие на отчество и отчества, похожие на фамилии.
- Разбираем ФИО, пришедшие и одной строкой, и гранулярно одновременно. Отдаем приоритет наиболее полному ФИО.
- Понимаем, что у иностранцев может не быть отчества.
- Переставляем ФИО, пришедшее в другом порядке.
- Удаляем однозначные дубли и не удаляем дубли в иностранных ФИО.
- Исправляем ошибку раскладки.
- Удаляем инициалы, если есть имя и отчество.
- Удаляем мусор.
- Склеиваем разделенные пробелом ФИО и расклеиваем ФИО без пробела.
- Можем не менять ФИО, но подсветить, что хотели бы в нем исправить.
- Ставим статус и маркеры.
- Склоняем: из именительного падежа в родительный, дательный или творительный.
- Проверяем на соответствие требованиям ФНС.

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Паспорта

Возможности

Выявит скрытые плохие документы — подходящие по формату, но просроченные или которые в принципе не могут существовать, т.к. не бывает такой серии документа.

Проверяем:

1. Формат номера и серии для 20 типов документов ФЛ.

1. Заполненность полей: серия, номер, дата выдачи, дата окончания срока действия, кем выдан, код подразделения.
2. Формат серии и номера.
3. Формат даты выдачи.
4. Формат даты истечения срока действия — при наличии такого поля.
5. Формат и существование кода подразделения.
6. Дату выдачи на корректность:
   a. дата из прошлого,
   b. возраст человека на момент выдачи документа позволял получить этот документ,
   c. указанный тип документа существовал в стране на указанную дату выдачи.
7. Актуальность документа:
   a. Указанная в документе дата истечения срока действия еще не наступила.
   b. Срок действия документа, определенный по его типу, не закончился (водительское удостоверение и загранпаспорт действуют 10 лет, временное удостоверение личности — 2 года).
   c. Паспорт РФ не просрочен по возрасту (граждане обязаны менять паспорт в 20 и 45 лет).
8. Соответствие даты выдачи и даты окончания.

Дополнительно для паспортов РФ проверяем:

1. Существование серии (первые две цифры — это код региона РФ, вторые две — год печати бланка).
2. Расхождение между годом печати бланка из серии и датой выдачи (разницу более чем в три года считаем подозрительной).

Ограничения

Не можем проверить принадлежность документа именно этому клиенту (соответствие паспорта указанному ФИО и дате рождения), т.к. данных нет в открытых источниках. Такая проверка осуществляется только через СМЭВ.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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.

E-mail

Стандартные проверки

- Извлекаем несколько email из одной строки.

- Исправляем часто встречающиеся опечатки 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 (исключение — домены, где разрешена другая письменность);

     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):

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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 рф, рус, 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.
- Проставляем тип домена: публичный сервис, одноразовый или корпоративный (при подключенном внутреннем справочнике).
- Вычисляем реальный адрес с учетом синонимов — для корректных поиска дубликатов и проверок.
- Проставляем подробные характеристики качества. Подробнее см. Маркеры 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 и адреса почтового сервера.

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

- эти проверки требуют подключение к Интернет;

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- скорость ответа низкая: единицы запросов в секунду;
- при любой проверке есть ложно положительные и ложно отрицательные результаты, потому что многие почтовые сервера блокирует подобные проверки.

Существуют сервисы, которые делают запрос к почтовому серверу и говорят, найден ли искомый email.

Почему мы не делаем так, на примере:

1. gmail.com и yandex.ru позволяет отправить запрос на проверку реально ли существует email. Отвечает честно.
   Если почта находится на gmail.com, можно проверить реальная она или нет.
2. mail.ru позволяет отправить запрос на проверку. Отвечает не честно, дает много ложно положительных срабатываний.
3. Свой почтовый сервер — используется у крупных компаний или у частных лиц, может отвечать на запрос по-разному в зависимости от настройки.

То есть, такая проверка работает медленно, но не дает никаких гарантий.

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**.
- Список однобуквенных доменов второго уровня. Источник: **HFLabs**.

Марка, модель и тип автомобиля

Возможности

1. Стандартизируем марки и модели транспортных средств, заданные в неструктурированном формате с учетом опечаток, избыточной или недостаточной информации, синонимов.
2. Распознаем популярный вариант модели и предлагаем другие, если модель встречается у нескольких брендов и задана без указания конкретной марки.
3. Подскажем тип транспортного средства.

Формат VIN-номера

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Возможности

1. Проверяем VIN-номер на соответствие шаблону, которому должен удовлетворять каждый идентификационный номер.
2. Заменяем символы кириллицы на такие же символы латиницы (трансграфика).

Формат госномера

Проверяем автомобильные регистрационные номерные знаки на формат.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «АДРЕСА»

Описание

- Разбиваем адреса на компоненты.
- Исправляем опечатки и сокращения.
- Приводим компоненты к единому формату.
- Переводим с латиницы на кириллицу.
- Восстанавливаем пропущенные компоненты: «197373 УЛИЦА ДОЛГООЗЁРНАЯ 12» «г Санкт-Петербург, ул Долгоозерная, д 12»
- Заменяем старые названия на новые: «Ленинград» «Санкт-Петербург». По запросу можем о ставлять 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» «117519, 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 стандартных уровней адреса;

  идентификатор ФИАС земельных участков и квартир;

  код ИФНС для юрлиц и физлиц 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»;

  коды ОКСМ, ОКТМО и ОКАТО по адресу, проверяем и корректируем ОКТМО по данным Росстата для отчётности по форме 260 в ЦБ.

  кадастровый номер дома, квартиры, участка из ФИАС.

- Проставляем коды качества для распознанного адреса:

  код полноты адреса;

  статус проверки;

  признак актуальности;

  код качества сделанных изменений;

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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 место рождения к требованиям Указания Банка России N 4936-У.
- Приводим адрес к требованиям Указания Банка России N 4990-У о формате реестра вкладчиков.
- Выводим дату изменения адреса по ФИАС/ГАР.

Ограничения

- Иностранные адреса поймем максимум до уровня города.
- Российские адреса обязательно проверяем по справочнику адресов до уровня улицы, а дома уже может не быть в ФИАС/ГАР. Если дом не из ФИАС/ГАР, то он будет разобран. Для проверки, есть ли дом в ФИАС, проставляем код соответствия ФИАС.
- 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».
- Не поддерживаем в домовой части признак владения «домовладение» — в ответе вернем «дом».

Используемые материалы

- ФИАС ФНС. Можно обновлять без перезапуска Фактора.
- 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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ГЕОКОДИРОВАНИЕ

Описание

- Конвертируем адреса в геокоординаты с указанием точности.
- Находим ближайшие к адресу клиента филиалы.
- Вычисляем положение объекта относительно МКАД и КАД: внутри или снаружи.
- Определяем ближайшие станции метро. Сделано для всех городов России, где есть метро: Москвы, Санкт-Петербурга, Казани, Самары, Новосибирска, Екатеринбурга, Нижнего Новгорода.

Можно обновлять справочник геокоординат без перезапуска «Фактора».

Ограничения

- Геокодируем только те адреса, которые есть в ФИАС.
- Покрытие адресов из ФИАС геокоординатами не полное.
- Положение относительно МКАД и КАД — только для адресов Москвы и области, Санкт-Петербурга и области.
- Справочник границ МКАД и КАД, а также справочник метро нельзя обновить нагорячую.
- Расстояние до метро определяем по прямой.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ПЛОЩАДИ И СТОИМОСТИ КВАРТИР

Описание

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

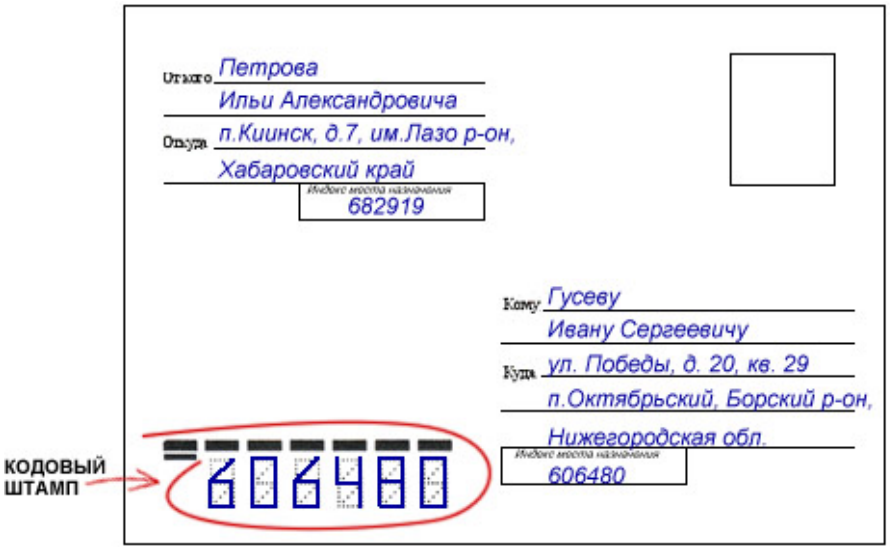
АДРЕСА В ФОРМАТЕ ПОЧТЫ РОССИИ

Представления адреса в формате Почты России сделаны для того, чтобы можно было взять адрес после обработки его «Фактором» и отправить на печать на конверте.
Очень удобно для почтовых рассылок!

На конвертах обычно указано 3 строчки адреса + ФИО в дательном падеже.

Адрес делится на следующие части

- POST_STREET_OTHER — тип улицы, улица, дом, расширение дома;
- POST_CITY_SETTLEMENT_DISTRICT — тип города, город, тип НП, НП, район, тип района;
- POST_SUBJECT — субъект, тип субъекта.



© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

БЛИЖАЙШЕЕ ОТДЕЛЕНИЕ ПЕНСИОННОГО ФОНДА РОССИИ

Описание

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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, «11111111»).
- 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, характеризующие качество номера:
  - Перенесённые номера.
  - Красивые номера.
  - Изменённые «Фактором».
  - Иностранные.
  - Неоднозначные.
  - Слишком длинные или слишком короткие.
- 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

- Иностранные 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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Используемые материалы

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

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ПРОВЕРКА ТЕЛЕФОНА ПО ПЛАНУ НУМЕРАЦИИ

При стандартизации все телефоны проверяются по плану нумерации. При этом определяется оператор, которому принадлежит номерная емкость телефона, и регион этого оператора.

Корректные по формату номера телефонов, если их нет в плане нумерации, получают код качества *OUT_OF_RANGE* (Нет в плане нумерации).

Пример 1:

Исходный телефон: 9653288406

Стандартизованный телефон:

- Номер: +7 965 328-84-06
- Оператор: ОАО "Вымпел-Коммуникации"
- Регион: Москва и Московская область
- Код качества: Корректный телефон

Пример 2:

Исходный телефон: 4996997777

Стандартизованный телефон:

- Номер: +7 499 699-77-77
- Оператор: нет
- Регион: нет
- Код качества: Нет в плане нумерации

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «ФИО И ПОЛ»

ФИО и пол

- Раскладываем ФИО, пришедшее одной строкой, на фамилию, имя и отчество.
- Определяем пол.
- Исправляем опечатки по похожести и по справочникам замен.
- Разворачиваем уменьшительно-ласкательное имя в полное.
- Разбираем тюркские имена и отчества.
- Понимаем фамилии, похожие на отчество и отчества, похожие на фамилии.
- Разбираем ФИО, пришедшие и одной строкой, и гранулярно одновременно. Отдаем приоритет наиболее полному ФИО.
- Понимаем, что у иностранцев может не быть отчества.
- Переставляем ФИО, пришедшее в другом порядке.
- Удаляем однозначные дубли и не удаляем дубли в иностранных ФИО.
- Исправляем ошибку раскладки.
- Удаляем инициалы, если есть имя и отчество.
- Удаляем мусор.
- Склеиваем разделенные пробелом ФИО и расклеиваем ФИО без пробела.
- Можем не менять ФИО, но подсветить, что хотели бы в нем исправить.
- Ставим статус и маркеры.
- Склоняем: из именительного падежа в родительный, дательный или творительный.
- Проверяем на соответствие требованиям ФНС.

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «ПАСПОРТА»

Паспорта

Возможности

Выявит скрытые плохие документы — подходящие по формату, но просроченные или которые в принципе не могут существовать, т.к. не бывает такой серии документа.

Проверяем:

1. Формат номера и серии для 20 типов документов ФЛ.

1. Заполненность полей: серия, номер, дата выдачи, дата окончания срока действия, кем выдан, код подразделения.
2. Формат серии и номера.
3. Формат даты выдачи.
4. Формат даты истечения срока действия — при наличии такого поля.
5. Формат и существование кода подразделения.
6. Дату выдачи на корректность:
    a. дата из прошлого,
    b. возраст человека на момент выдачи документа позволял получить этот документ,
    c. указанный тип документа существовал в стране на указанную дату выдачи.
7. Актуальность документа:
    a. Указанная в документе дата истечения срока действия еще не наступила.
    b. Срок действия документа, определенный по его типу, не закончился (водительское удостоверение и загранпаспорт действуют 10 лет, временное удостоверение личности — 2 года).
    c. Паспорт РФ не просрочен по возрасту (граждане обязаны менять паспорт в 20 и 45 лет).
8. Соответствие даты выдачи и даты окончания.

Дополнительно для паспортов РФ проверяем:

1. Существование серии (первые две цифры — это код региона РФ, вторые две — год печати бланка).
2. Расхождение между годом печати бланка из серии и датой выдачи (разницу более чем в три года считаем подозрительной).

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Ограничения

Не можем проверить принадлежность документа именно этому клиенту (соответствие паспорта указанному ФИО и дате рождения), т.к. данных нет в открытых источниках. Такая проверка осуществляется только через СМЭВ.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «E-MAIL»

E-mail

Стандартные проверки

- Извлекаем несколько email из одной строки.

- Исправляем часто встречающиеся опечатки 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 (исключение — домены, где разрешена другая письменность);
    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 рф, рус, москва запрещена латинская письменность.
3. В доменах com, net, tel, name, org, su допускается любая письменность (в том числе кириллица).

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- 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.
- Проставляем тип домена: публичный сервис, одноразовый или корпоративный (при подключенном внутреннем справочнике).
- Вычисляем реальный адрес с учетом синонимов — для корректных поиска дубликатов и проверок.
- Проставляем подробные характеристики качества. Подробнее см. Маркеры 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 и адреса 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 низкая: единицы запросов в секунду;
- при любой проверке есть ложно положительные и ложно отрицательные результаты, потому что многие почтовые сервера блокирует подобные проверки.

Существуют сервисы, которые делают запрос к почтовому серверу и говорят, найден ли искомый email.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Почему мы не делаем так, на примере:

1. gmail.com и yandex.ru позволяет отправить запрос на проверку реально ли существует email. Отвечает честно.
   Если почта находится на gmail.com, можно проверить реальная она или нет.
2. mail.ru позволяет отправить запрос на проверку. Отвечает не честно, дает много ложно положительных срабатываний.
3. Свой почтовый сервер — используется у крупных компаний или у частных лиц, может отвечать на запрос по-разному в зависимости от настройки.

То есть, такая проверка работает медленно, но не дает никаких гарантий.

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**.
- Список однобуквенных доменов второго уровня. Источник: **HFLabs**.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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
- Дату регистрации.
- Статус: активный/ликвидированный.
- Код ОКОПФ
- 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».

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «КОМПАНИИ»

Компании

Наименования компаний

- Формируем название компании-ключа для последующей идентификации дубликатов.

- Извлекаем ОПФ из наименования.
- Исправляем часто встречающиеся опечатки в ОПФ.
- Расшифровываем аббревиатуры ОПФ.
- Находим ОПФ в начале, середине и конце наименования.
- Разделяем слитно написанные ОПФ и наименование.
- Если находим несколько ОПФ, проверяем их на конфликт.

  Допустимые сочетания нескольких ОПФ отмечаем отдельным маркером, выбираем итоговую ОПФ по настройке.

  Неоднозначные сочетания ОПФ отмечаем отдельным маркером, ставим статус NOT_SURE.

- Определяем, что компания относится к судам или государственным учреждениям.

  Отмечаем такие наименования специальными маркерами.

  Распознаем такие наименования как не требующие ОПФ.

- Отличаем иностранные наименования от российских: распознаем такие наименования как не требующие ОПФ.
- Извлекаем из наименования реквизиты компании:

  ИНН;

  КПП;

  ОГРН;

  Номер лицевого счета.

- Удаляем мусор.
- Ставим маркеры.
- Удаляем кавычки, если нет ОПФ (доступно с версии «Фактора» 24.6+).

Ограничение

Фильтр не стандартизирует наименование компании до того вида, как оно записано в ЕГРЮЛ-е.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Другие атрибуты

- 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 с учетом ОПФ, 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 сравнение 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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «АВТОМОБИЛИ»

Марка, модель и тип автомобиля

Возможности

1. Стандартизируем марки и модели транспортных средств, заданные в неструктурированном формате с учетом опечаток, избыточной или недостаточной информации, синонимов.
2. Распознаем популярный вариант модели и предлагаем другие, если модель встречается у нескольких брендов и задана без указания конкретной марки.
3. Подскажем тип транспортного средства.

Формат VIN-номера

Возможности

1. Проверяем VIN-номер на соответствие шаблону, которому должен удовлетворять каждый идентификационный номер.
2. Заменяем символы кириллицы на такие же символы латиницы (трансграфика).

Формат госномера

Проверяем автомобильные регистрационные номерные знаки на формат.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

ВОЗМОЖНОСТИ МОДУЛЯ «ОБЩИЙ»

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

МОДУЛИ ИДЕНТИФИКАЦИИ

Сведения о функциональности модулей выявления похожих записей.

- Идентификация дубликатов
  - Термины
  - Возможности
  - Ограничения
- Проверка по чёрным 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 похожести.

Для определения степени похожести 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 похожесть.
- Нечёткое сравнение Документ ФЛ и дат с учётом распространенных опечаток.
- Высокая скорость идентификации дубликатов.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- Полный поиск дубликатов на всей базе.
- Инкрементальный поиск на объеме обновлённых за день данных.
- Онлайн-сервис для нечеткого поиска по всему объему.
- Кластерный режим поиска дубликатов в сервисах «Единого Клиента».

Ограничения

Для запуска идентификации дубликатов данные должны быть стандартизированы.

Проверка по чёрным 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 похожести (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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

МОДУЛЬ «ИДЕНТИФИКАЦИЯ ДУБЛИКАТОВ»

Термины

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

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

Для определения степени похожести 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 похожесть.
- Нечёткое сравнение Документ ФЛ 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.
- Кластерный режим поиска дубликатов в сервисах «Единого Клиента».

Ограничения

Для запуска идентификации дубликатов данные должны быть стандартизированы.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

МОДУЛЬ «ПРОВЕРКА ПО ЧЁРНЫМ 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 похожести (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.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

НОВОСТИ «ФАКТОРА»

Что нового в 2025 году:

- API «Фактора»
  - Включили расширенный ответ healthCheck
  - Удалили FactorService из сборок без РМАД
- Модуль «Адреса»
  - Лучше разбираем квартиры с дробью
  - Исправили адреса с кавычками и тире
- Модуль «Телефоны»
  - Привели к единому формату мобильные номера белорусских операторов
- Модуль «Компании»
  - Пополнили справочник синонимов для страховых
  - Актуализировали требования для валидации БИК
  - Исправили ошибку с удалением пробела перед кавычкой в середине строки
- Модуль «ФИО и пол»
  - Добавили правила склонения фамилий на -цын/-цына
- Модуль «ФИО и пол» (новый алгоритм)
  - Склеиваем разделенные пробелом ФИО и расклеиваем ФИО без пробела
- Модуль «Email»
  - Добавили в перечень доменов xmail.ru
- Модуль «Паспорта»
- Технические новости

API «Фактора»

Включили расширенный ответ healthCheck

> Если настроена интеграция на ответ /api/manage/health , то она может сломаться, так как изменилась структура ответа.

Раньше endpoint «Фактора» http://[]:[]/factor-service-[]/api/manage/health возвращал в ответе только общий статус приложения:

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

/api/manage/health
{"status":"UP"}

Теперь по умолчанию этот endpoint возвращает расширенный ответ, содержащий общий статус приложения и статусы подключенных компонент.

Пример расширенного ответа:

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

/api/manage/health


```
{
  "status": "UP",
  "components": {
    "applicationContext": {
      "status": "UP"
    },
    "diskSpaceHealth.deduplication": {
      "status": "UP",
      "details": {
        "name": "dedup",
        "path": "file:///opt/factor/dedup/",
        "total": 214642462720,
        "free": 165604425728,
        "warn_threshold": -1,
        "error_threshold": -1
      }
    },
    "diskSpaceHealth.dictionaries": {
      "status": "UP",
      "details": {
        "name": "dct",
        "path": "file:///opt/factor/dictionaries/",
        "total": 214642462720,
        "free": 165604425728,
        "warn_threshold": -1,
        "error_threshold": -1
      }
    },
    "diskSpaceHealth.indexes": {
      "status": "UP",
      "details": {
        "name": "idx",
        "path": "file:///opt/factor/indexes/",
        "total": 214642462720,
        "free": 165604425728,
        "warn_threshold": -1,
        "error_threshold": -1
      }
    },
    "diskSpaceHealth.tmp": {
      "status": "UP",
      "details": {
        "name": "tmp",
        "path": "file:///opt/factor/tmp/",
        "total": 214642462720,
        "free": 165604425728,
        "warn_threshold": -1,
        "error_threshold": -1
      }
    }
  }
}
```


Также теперь доступны endpoint, возвращающие в ответе статус и информацию по каждому компоненту в отдельности — http://[]:[]/factor-service-[]/api/manage/health/[_].

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Для мониторинга доступности приложения «Фактор» рекомендуем использовать endpoint http://[]:[]/factor-service-[]/api/manage/health/ applicationContext.

Пример ответа:

/api/manage/health/applicationContext
{"status":"UP"}

Подробности: Мониторинг доступности «Фактора»

Удалили FactorService из сборок без РМАД

SOAP/REST интерфейс *FactorService* первым появился в «Факторе», был предназначен для пользовательского интерфейса РМАД, но имел метод, который вызывался для проверки работоспособности «Фактора»: *getBuildInfo*. Этот метод часто использовался независимо от того, был ли подключен РМАД.

Метод имеет аналоги:

1. *api/manage/health/applicationContext* – легковесный метод, который возвращает статус приложения;
2. *api/manage/info* – легковесный метод, который возвращает информацию о сборке;
3. *api/services/rs/ManagementDictionaryService* – методы для работы со справочниками.

Метод *getBuildInfo* более недоступен, пожалуйста, воспользуйтесь указанными выше альтернативами.

В сборках с РМАД *FactorService* останется.

Модуль «Адреса»

Лучше разбираем квартиры с дробью

Раньше было ограничение: в номере квартиры не могло быть значения с дробью. Номер после дроби определяли как вторую квартиру.

Сняли это ограничение.

Пример: 404102, ВОЛГОГРАДСКАЯ ОБЛАСТЬ, Г. ВОЛЖСКИЙ, Б-Р ПРОФСОЮЗОВ, Д.32, КВ.200 /202

- Было: Волгоградская обл, г Волжский, б-р Профсоюзов, д 32, кв 200, кв2 202
- Стало: Волгоградская обл, г Волжский, б-р Профсоюзов, д 32, кв 200/202

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Кроме того, внутри «Фактора» был запрет на одинаковые номера в номере квартиры. Сняли это ограничение.

Пример: 404102, ВОЛГОГРАДСКАЯ ОБЛАСТЬ, Г. ВОЛЖСКИЙ, Б-Р ПРОФСОЮЗОВ, Д.32, КВ.200/200А

- Было: Волгоградская обл, г Волжский, б-р Профсоюзов, д 32, кв 200, кв2 А
- Стало: Волгоградская обл, г Волжский, б-р Профсоюзов, д 32, кв 200/200А

Рекомендуем переобработать адреса, где в результате «Фактор» заполнил второй номер квартиры/офиса/помещения/комнаты, если он выводится, либо если в исходном или результирующем номере есть / или -.

Исправили адреса с кавычками и тире

Раньше справочник адресов для «Фактора» делали из файлов DBF-формата, в котором есть ограничения на некоторые символы.

В результате часть адресов отличалась от первоначального написания в ГАР БД.

Отказались от этого формата и теперь адреса отдаём корректными.

Изменение повлияло на 1 населенный пункт и 114 планировочных структур. Со 2 мая 2025 выпускаем ФИАС без этих ошибок.

Пример: Ульяновская обл, Майнский р-н, поселок Рыбхоза «Пионер»

- Было: Ульяновская обл, Майнский р-н, поселок Рыбхоза <Пионер>
- Стало: Ульяновская обл, Майнский р-н, поселок Рыбхоза «Пионер»

Рекомендуем переобработать адреса, у которых название населённого пункта, планировочной структуры или улицы содержат символы: < , ? или >.

Модуль «Телефоны»

Привели к единому формату мобильные номера белорусских операторов

Раньше мы не ограничивали длину кода оператора для Беларуси. Из-за этого возникали проблемы с кодом оператора 29: для него «Фактор» мог забирать следующую цифру из кода телефона в код оператора.

В новой версии для мобильных номеров Беларуси установлен единый стандарт:

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- Код страны — +375.
- Код оператора — всегда из двух цифр.
- Номер телефона — 7 цифр.

Пример: *+375 29 2309-32-8*

- Было: 375/292/309328
- Стало: 375/29/2309328

Модуль «Компании»

Пополнили справочник синонимов для страховых

Научились удалять ОПФ из страховых ЗАО, АО и ПАО.

Пример: Страховое акционерное общество "Ресо-Гарантия".

- было (*name_no_opf*): Страховое "Ресо-Гарантия";
- стало (*name_no_opf*): Ресо-Гарантия.

Актуализировали требования для валидации БИК

Столкнулись с проблемой: существующий валидный БИК получал статус NOT_SURE. Проблема возникала из-за того, что ранее БИК должен был начинаться на 04, но в какой-то момент это правило утратило актуальность.

Обновили алгоритм валидации БИК в соответствии с положением N732-П «О платежной системе Банка России». Теперь фильтр ищет в исходной строке последовательность из 9 цифр и возвращает статус EDITED, если выполнены условия:

- 1 разряд принимает значения 0, 1 или 2;
- 2 - 9 разряды принимают значения от "00000001" до "99999999".

Пример: 012520101

- было: NOT_SURE
- стало: EDITED

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Исправили ошибку с удалением пробела перед кавычкой в середине строки

Ранее при обработке названий компаний пробел с кавычкой в середине строки удалялся. Это приводило к склеиванию слов.

Исправили эту проблему, теперь пробел не удаляется.

Пример: Агрофирма "Колобок

- было (*name_no_opf*): Агрофирма"Колобок
- стало (*name_no_opf*): Агрофирма "Колобок

Модуль «ФИО и пол»

Добавили правила склонения фамилий на -цын/-цына

Столкнулись с тем, что фильтр склонения ФИО по падежам не умеет обрабатывать фамилии, оканчивающиеся на цын/цына.

Добавили новое правило в фильтр declensionFilter.

Пол	Фамилия в именит. падеже	Фамилия в родит. падеже	Фамилия в дат. падеже	Фамилия в творит. падеже
муж	Трубицын	Трубицына	Трубицыну	Трубицыным
жен	Пшеницына	Пшеницыной	Пшеницыной	Пшеницыной

Модуль «ФИО и пол» (новый алгоритм)

Склеиваем разделенные пробелом ФИО и расклеиваем ФИО без пробела

Исходное ФИО могут записать без пробелов или с лишними пробелами:

- П о п о в А л е к с а н д р В е н и а м и н о в и ч
- Алексеенко Клавди я Петровна
- БЕДЕЛЬ **ВЛАДИМИР**АЛЕКСАНДРОВИЧ
- ИзмайловаВероникаВалентиновна
- ИЗМАЙЛОВА**ВЕРОНИКА**ВАЛЕНТИНОВНА

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

- **ГКФХ**АЗАРНЫХ ЕВГЕНИЙ ИВАНОВИЧ

Раньше «Фактор» мог справиться только с самыми простыми случаями разделённых или склеенных ФИО — например, в некоторых случаях мог добавить пробел перед прописной буквой в склеенном слове.

Теперь научили «Фактор» обрабатывать большинство ФИО с лишними или недостающими пробелами. Наш новый алгоритм пробует разрезать или объединить части ФИО, если исходных слов нет в справочниках «Фактора». Разделение или объединение считаем успешным, если в результате:

- получились справочные фамилия, имя, отчество;
- ФИО состоит из трёх или четырёх слов;
- нашли только один способ получить корректное ФИО.

В противном случае разделение или объединение не применяем. Также разделение не применяем для ФИО на латинице.

Исходное ФИО	Результат	Комментарий
ПоповАлександрВениаминович	Попов Александр Вениаминович	• маркер ФИО: SPLIT. Такие ФИО сначала объединяем в одну строку, затем разделяем; • маркер фамилии: JOIN; • маркер имени: JOIN; • маркер отчества: JOIN.
Алексеенко Клавди я Петровна	Алексеенко Клавдия Петровна	маркер имени: JOIN.
СМЕЛОВ С.ЕРГЕЙ АЛЕКСЕЕВИЧ	Смелов Сергей Алексеевич	маркер имени: JOIN.
БЕДЕЛЬ ВЛАДИМИРАЛЕКСАНДРОВИЧ	Бедель Владимир Александрович	маркер ФИО: SPLIT.
ИзмайловаВероникаВалентиновна ИЗМАЙЛОВАВЕРОНИКАВАЛЕНТИНОВНА	Измайлова Вероника Валентиновна	• маркер ФИО SPLIT; • для обоих примеров одинаковый результат вне зависимости от исходного регистра, исходный регистр не учитываем.
ГКФХАЗАРНЫХ ЕВГЕНИЙ ИВАНОВИЧ	Азарных Евгений Иванович	• в мусор унесли строку «ГКФХ»; • маркер ФИО: SPLIT.
Исключение:	Акинтьев	Все исходные слова справочные, алгоритм склеивания не сработает и поведение не изменится.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Акинтьев Сергей Алекс андрович	Сергей Алекс Андрович

Маркеры

Добавили новые маркеры:

Название	Экстрактер	Описание
SPLIT	all_markers	Расклеили компоненты ФИО.
WTF_SPLIT		Ставится в режиме *return_source=true* (все изменения отключены, но возможно расклеивание компонентов).
JOIN	last_markers first_markers middle_markers	Убрали разрыв в компоненте ФИО.
WTF_JOIN		Ставится в режиме *return_source=true* (все изменения отключены, но возможно расклеивание компонентов).

Скорость

На среднестатистической базе данных, где не много склеенных или разделенных ФИО, скорость обработки ФИО станет на 2% ниже.

На склеенных или расклеенных ФИО скорость может стать заметно ниже.

Модуль «Email»

Добавили в перечень доменов xmail.ru

Mail.ru предоставляет возможность перенести почту с Gmail на Mail.ru. Для этого они добавили специальный вспомогательный домен xmail.ru. Раньше все емейлы с таким доменом мы считали валидными и не проверяли по доменным правилам. Сейчас проверяем по правилам, действующим для gmail.com.

Пример: abcd_@xmail.ru. Нужно проверять емейл по правилам домена gmail.com. Последний символ для gmail.com должен быть буквой или цифрой. В исходном емейле последний символ — нижнее подчеркивание:

- Было: abcd_@xmail.ru — GOOD_EMAIL;EDITED — ошибочно помечали такие емейлы как корректные.
- Стало: abcd_@xmail.ru — NOT_VALID_EMAIL;NOT_SURE;LOCAL_BAD_SPECIFIC — исправили ошибку.

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.

Модуль «Паспорта»

Технические новости

© ООО «ХФ Лабс». Данный документ содержит конфиденциальную коммерческую информацию. Никакая часть документа не может быть воспроизведена любым способом без письменного разрешения ООО «ХФ Лабс». Персональные данные, приведенные в качестве примеров, своей совокупностью не идентифицируют какую-либо конкретную частную персону.