ЗУМ ПЕРВЫЙ

Время и дата: 12:42 06.11.2021

Тип события: Общее событие

# Тема: Не определен

Теги: текст, модель, ошибка, фраза, распознавание, версия, само дело, файл, станислав, сегмент, репрезентация, помощь, план, следующий этап, следующее этап, облако тэгов, искусственный интеллект, доработка сервиса, данный момент, экстракт, текст, модель, ошибка, само дело, фраза, файл, топики, теги, станислав, слова, распознавание, помощь, лайты, версия, частота встречаемости, следующее этап, облако тегов, наше решение, искусственный интеллект, доработка сервиса

Текст:

Добрый день, ещё раз уважаемые коллеги, рад вас приветствовать на рассуждениях национального. кодекса по искусственному  
интеллекту, который. Как я понимаю, подготовлен совместными усилиями центров. при правительстве аналитической альянс  
искусственный интеллект, конечно, разработка этических норм в сфере.  
  
 искусственного интеллекта представляется крайне  
важной проблемой, давай, дворянин шага пока все это. встречается, мог быть, кто то ожесточается и рассказать немножко  
про вашу команду,. вы работаете как реальная компания, которая занимается уменьшением серия работ с основанием. речи.  
  
   
Или это ваше хобби, вы собрались специально для конкурса месяца, здравствуйте,. пожалуй, я начну, мы собрались на самом  
деле месте для участия в. подобных конкурсах, плюс у нас есть небольшой опыт по созданию подобного сова,. в частности,  
вот наш текущий проект, это некая эволюция того, что мы.  
  
 делали ранее уже в рамках одного из мероприятий решался  
подобный кейс, и,. соответственно, то, что сделали сейчас, это некое развитие собственно, мы отдали в. дальнейшем  
надеемся на при помощи опять же механизмов и катона да, может. быть, мы с кем то познакомимся, может быть, мы кому то  
понравимся,.  
  
 и, в конце концов, мы очень сильно хотим вывести наше решение уже. в продуктовую стадию, спасибо большое,  
что не отвечает название изначально исторически у. него совсем другой смысл сейчас это просто красиво выглядящая  
аббревиатура, изначально это. было комплексное обеспечение безопасности, и чего то там чего то чего то.  
  
 уже история  
этого не помнит на самом деле, а если славная департамент. завтра захочется купить решение вас есть понимание, как, по  
его готовы продавать. как сервис, и точная модель как передач про на самом деле мы. рассматривали уже в принципе  
несколько моделей монетизации.  
  
 В том числе и как. сервис, и как подписку, ну будет, скажем так, захотите нас купить,  
мы. обязательно предложим варианты, рекламодатель времени на доработку сервиса, пожалуй, на доработку сервиса. может  
уйти порядка трёх четырёх месяцев, что касается до обучения моделей, то.  
  
 тут, наверное, лучше станислав ответить,  
станислав, помоги, пожалуйста, да может быть, ещё. более детальные вопросы коллег типа модели расскажет, как она  
работает, здравствуйте, меня. зовут станислав, очень приятно в основе нашего движка распознавания используются  
алгоритмы воск. .  
  
 Вот и их модели доступны в общем доступе, две модели доступны, это. лайк версия и версия слух,  
соответственно, мы используем их обоих, для чего. это делается на версия лайт позволяет уложиться в первый критерий,  
пять минут. на часовое видео, предварительно распознавания мы корректируем ошибки,.  
  
 С помощью нескольких алгоритмов. .   
Я чуть позже об этом. Скажу вот после того, как первичная, лайт,.  
  
 версия текста готова, нана сразу вводится во фронт  
пользователя и далее на. кэмден сразу начинается более глубинное либо извините, я вас перебью показывать так. хорошо,  
а у кого же. Сайт как инструмент может быть свой свой.  
  
 экран, отправиться я стас, я уже более от и комментируй надо на.   
библии захаров тогда да ладно, да, собственно,. Это как бы нашего решения,. но вся магия происходит на сервере фбр  
условно.  
  
 Пайп лайн на языке. . Питания, а можно бока показать, откройте, пожалуйста, доступно. У кого так бег.  
  
 руслан,  
я отключаюсь вот сейчас, когда я вкратце смогу пробежаться пока как. давид, да потому. Что у нас происходит  
соответственно, у нас в докеров,. .  
  
 Разные модели первичной у нас используются лайт модель воска, вот она принимает.   
на. Себя ваш файл, если файл изначально заливается,. Не авто, у нас.  
  
 система проверяет, какой файл это если это  
портрет, то мы с помощью. специального. Кода к его преобразован клаву, причём он должен быть шестнадцатый же. битрейт.  
  
   
Ему на канальные то. Есть для лучшего распознавания она ход подходит,. . Любой файл, то.  
  
 Есть мы их обрабатываем, уже  
умеем. С разными файлами. работать. Сразу же начинается распознавания вот вот наступит.  
  
 Через докер это можно.   
параллельно. Все на разных серверах. Да вы не на одном, а связывать.  
  
 цепи фермеров вот после того, как лайк модель  
отработала. В нас, причём. не просто текстовая репрезентация, выводятся адже сам структура, которая по каждому слову.   
показывает вероятность ошибки, для чего это нужно, это позволяет нам обрабатывать каждое.  
  
 . Слово и целую в разрезе  
контекста нечиста по одному. Слову об этом. дальше пламя текстовом описании руслан то, что я тебе секунду, просто вот,.  
  
   
собственно,. Да вот наш флот модель функцию первичную загружают. Сначала лайт модель,. потом эта функция проверяется  
файл вас или нет и любой другой кривой.  
  
 глаз канал на. Следующую процессора, то у нас войск текст, мы берём. джейсон,  
структуру, вытаскиваем этот слова, вытаскиваем потенциальная ошибка, которому потом работаем, да. экстракт секса  
выводит непосредственно тексту репрезентацию.  
  
 Следующим этапом, так как идеальная модель. не бывает, да, безусловно,  
доступны в паблике, нам приходится различным алгоритмам и. корректировать ошибки первичном используем так называемого  
алгоритм. Семь спал с миллиард, то.  
  
 есть мы с помощью алгоритма дистанции левенштейн мы проверяем контексте следующего  
слова. какой может быть предыдущем, это используется с помощью чистоты грамм. На русском. языке базы вот она  
несколько.  
  
 Гигабайт весит и мы проверяем по частоте. вероятность встречаемости слова, следующего за предыдущ ищем и  
корректируем условное. Слово, например,. там раз поздно человек, человек так произнёс.  
  
 И модель так это всё. записала,  
мои слова объединим, уменьшим соответственно, процент ошибок, следующий. Этап это я. тут чуть не указал, ещё одна  
библиотека чжон.  
  
 Спал, соответственно, тоже грамматической. ошибки. Если модель какие то ошибки развела в. Словах при  
распознавании мы.  
  
 их тоже корректируем,. По большой базе данных там тоже используются. Би граммы. и несёт другие  
алгоритмы так,.  
  
 Следующим этапом у нас специальная. Модель модель,. мы её. Сами обучили через дип сегмент на русском  
базе, на русской.  
  
 базе вектора просто. Слов она позволяет примерно представить, где предложен заканчивается, но. в  
большинстве случаев она угадывает и в принципе неплохо делит текст на. предложение.  
  
 Следующее так у нас пользуются  
библиотеками ираком. А моё чудо обучили,. она используется для проставления пунктуации, секция конкретно запитай от  
работы, тоже очень,. даже хорошо.  
  
 Следуя так у нас процесс характер джон спел просто читаем. . Шара в потоке так так  
так, что это происходит мы берём, смотрим. частоты встречаемости файл после того, как убираем стоп слова, которые.  
  
 Не  
влияют. на логику,. И вот эти вот теги нам позволяют понять в принципе,. о чем идёт разговор,.  
  
 Теги у нас выводятся в  
виде облака. Следующее. квантовый генератор. Покажу вам, как это выглядит примерно а для чего это.  
  
 нужно, это нужно  
для того, чтобы человек. Даже будучи ни за что. не зашёл в текст, он уже видел какие примерно какие примерно фразы. .  
  
   
Передаются фразы и слова, например, здравоохранение,. И плюс по частоте встречаемости они. становятся больше или меньше  
в зависимости от того, как часто они встречаются. .  
  
 этого текста тобой репрезентация, но мы ещё не внедрили. На фронтон  
там. уже более красивое облако тегов происходит я сейчас куда. Нибудь к ну.  
  
 ладно,. Это мы можем позже показать уже  
рядом вот в табличке будут. визуальная репрезентация с облаком тегов, дальше по плану открой, пожалуйста, а потом. так  
мы укладываем тайминги, да сколько у нас времени, пожалуйста, гриш корректирования.  
  
 не успеваю, план открой,  
пожалуйста, откажу, минуты осталось, но. Успеем так, как. мы изменили облако тегов,. И сами торговые презентации  
вытащили, мы отправляемся на.  
  
 бокам, соответственно, и вы вон тот вытаскиваем топик, топики вытаскиваются на уровне.   
эвристических правил не статистически составил вот эти вот слова, которые из тегов. чаще всего встречаются на про  
медицина, пациент, здравоохранения и государство, скорее всего,. программа там знаю, там цифровизации области  
здравоохранения, амелия же топик, где там.  
  
 используются там законодательства государства, закон, реформа,  
ответственность крысу, реформ законодательства,. И так. далее, то. Есть у нас.  
  
 Список там большее чем из ста правил. .   
Вот как я называю выбросить кому он примерно помогает отказать в какой. топик.  
  
 О чем идёт речь,. В данном в данной  
записи дальше экстракт. . Саммари у нас есть две различные модели первой выбили библиотеку,.  
  
 Тел дыры. вот она  
позволяет сгруппировать несколько быть точнее, я группировал по несколько предложений. и скармливая вот. Так вот эти  
предложения модели на выходе нам краткую.  
  
 презентацию того, краткую выжимку, о чем говорили, высказывает. И есть вторая  
моделям. бартеру она чуть медленнее работает. Но не думаю, что мы сможем использовать.  
  
 обе и смотреть, как конкретно,  
в каких предложениях лучше представлена репрезентация самолёта,. краткая выжимка сообщения, последний этап это  
протекает, текст,. Это то, что мы. текст, чтобы не являлся, едим на простыне, мы его удобно разделяем на.  
  
 параграфы,  
добавляем абзацы дни ширину, ограничен. Сто двадцать символ это удобно читать. визуально,. Пока фронтон там ещё не все  
добавлено из того, что могли.  
  
 не получилось, завершайте покажи пород сейчас. Фронт буквальным, там торчком, да,  
собственно,. есть таблица с распознано. Нами и вот как выглядит хлеба.  
  
 С исходного. текста и сегментирования текст уже  
с пунктуацией с исправленными ошибками, плюс если. мы нажмём проиграть. Нажми проиграть нас будет, по словам, эти  
подсветка, что.  
  
 . В данный момент тщательно пожалуйста, поиграй, но комп нажал на пожалуйста, конкретный. текст не  
отдаст, например,. Долго начала да термин.  
  
 Версия был уже открыт. дай то есть наш фронтон, позволяет подсвечивать  
данный момент такие фразы произносятся. не то, что подсвечено жёлтым, это те слова,. И фразы, в которых.  
  
 . Так мы не.   
Слышим то, что она. Говорит, но на.  
  
 . .