

हे YouTube, इस वीडियो में मैं आपको दिखाने जा रहा हूँ कि कैसे आप जल्दी से किसी भी ऑडियो को पाठ में बदल सकते हैं, जो कि फुसफुसाते हुए पायथन में मुफ्त ओपन सोर्स पैकेज का उपयोग कर के पाठ में बदल सकते हैं। मैं यह दिखाने जा रहा हूँ कि मैंने इसे स्थापित किया है, एक उदाहरण दिखा रहा हूँ कि मैं इसे कैसे चलाता हूँ और इसकी तुलना किसी मौजूदा लाइब्रेरी से करता हूँ। इसलिए शुरू करने में, आप शायद फुसफुसाते हुए हब रपिजिटी में जाना चाहते हैं जो हम यहां देख रहे हैं और वे निर्देश देते हैं कि आप इसे कैसे स्थापित कर सकते हैं। अब एक बात ध्यान में रखें जब आप पपि को केवल नाम फुसफुसाते हुए स्थापित करते हैं तो यह सही संस्करण स्थापित करने वाला नहीं है। हम इस गिट रपिजिटी से इंस्टॉल करना चाहते हैं। तो बस इस पपि को कमांड इंस्टॉल करें और इसे अपने वातावरण में चलाएं कि आप पायथन चला रहे हैं। और उन्होंने यहां यह भी उल्लेख किया कि आपको FFMPEG स्थापित करने की आवश्यकता है। इसे करने के लिए कुछ निर्देश हैं, लेकिन मैं पहले से ही अपने कंप्यूटर पर स्थापित था। अब जब मेरे पास कानाफूसी है, तो चलो बस कुछ ऑडियो बनाएं जो मैं इस पर परीक्षण कर सकता हूँ। तो मैं कुछ मुहावरों को कहने वाला हूँ। मॉडल आमतौर पर मॉडल को समझने के लिए कठिन होते हैं। हालांकि यह सिर्फ पाठ के लिए भाषण है। यह एक तरह का मजा होगा। मैं एक ट्रकिटटू के रूप में क्लाउड 9 पर रहना पसंद करूंगा जो एक मक्खी को चोट नहीं पहुंचाएगा। मैं पानी से बाहर एक मछली की तरह होगा और मौसम के नीचे होने के लिए एक बेला के रूप में फटि होगा। चलो इसे बचाओ। चलो इसे एक लहर के रूप में सहेजते हैं। उनके पास निर्देश हैं कि कैसे हम इसे स्थापित होने के बाद कमांड लाइन से सीधे कैसे चला सकते हैं। मैं आपको दिखाने वाला हूँ कि पायथन एपीआई का उपयोग कैसे करें, जो वे यहां दिखाते हैं। तो यह वास्तव में सरल है। हम सिर्फ कानाफूसी आयात करते हैं। फिर हम अपना मॉडल बनाने जा रहे हैं, जो कि हम लोड करने वाले हैं। मॉडल जैसी आधार कहा जाता है। और फिर बस इस मॉडल ऑब्जेक्ट का उपयोग करके, हम अपनी ऑडियो फाइल पर ट्रांसक्राइब चलाते हैं। इसलिए मैंने इसे मुहावरों का नाम दिया। आइए वेब संस्करण का उपयोग करें। हम चाहते हैं कि यह परिणाम वापस कर दे। अब, मैंने देखा कि जब मैंने इसे पहले दौड़ाया, तो मुझे यह त्रुटि मिली है कि योकिडुआ के आधे टेसर और फ्लोट टेसर के कारण। मैं इसे हल करने में सक्षम था। तो यह ध्यान में रखने के लिए कुछ है। यदि यह आपके लिए काम नहीं करता है, तो आपको गरिने के लिए फ्लोटिंग प्वाइंट 16 सेट करने की आवश्यकता हो सकती है। और आप देख सकते हैं कि यह यहां चलाने के बाद, इसने भाषा को पहले से ही अंग्रेजी के रूप में पाया है और फिर इस परिणाम ऑब्जेक्ट में उनमें कुछ अलग तरीके हैं, लेकिन हम जो चाहते हैं, वह सिर्फ पाठ है और हम देख सकते हैं कि यह दिखाता है जैसे परिणाम अच्छा है मैं एक ट्रकि पोनी के रूप में बादल नौ पर रहना पसंद करूंगा, जो एक मक्खी को चोट नहीं पहुंचाएगा, मैं पानी से बाहर एक मछली की तरह होगा और इसने इस मछली को पानी से बाहर कर दिया। एक फडिल के रूप में फटि है और शायद मैंने यह नहीं कहा कि यह स्पष्ट रूप से एक और बात है कि जब आप पहली बार इसे चलाते हैं, तो यह आधार मॉडल डाउनलोड करना होगा। तो आप एक प्रगत बार को पार करते हुए देख सकते हैं और आपको उस मॉडल को डाउनलोड करना होगा। और यह कहता है कि जब आप इस ट्रांसक्राइब को चलाते हैं, तो यह वास्तव में आपकी ऑडियो फाइल के 30 सेकंड के चंक्स ले रहा है और उस पर भविष्यवाणियां चला रहा है। अब एक और दृष्टिकोण भी है जैसी आप ले सकते हैं, जो एक नचिले स्तर का दृष्टिकोण है, जहां आप वास्तव में मॉडल बनाते हैं और फिर आप ऑडियो ऑब्जेक्ट बनाते हैं और पैटर्न ट्रिम करते हैं। तो आप बस यह सुनिश्चित करते हैं कि यह ऑडियो चंक् केवल 30 सेकंड है। सेकंड या यह इसे 30 सेकंड के साथ थपथपाएगा क्योंकि यह लंबाई है जो मॉडल इनपुट के रूप में होने की उम्मीद करता है। फिर यह एक लॉग माउस स्पेक्ट्रोग्राम बना रहा है। यह भाषा का पता लगा रहा है और हम यहां डिकोड कर सकते हैं और यदि हम चाहते थे तो बहुत अधिक विकल्प प्रदान कर सकते हैं। यदि मैं इस सेल को चलाता हूँ, तो फिर से यह त्रुटि प्राप्त करें, जैसी मैं अब डिकोडिंग विकल्पों में सेट कर सकता हूँ, FP16 दोषों के बराबर है। और वास्तव में इस बार ऐसा लगता है कि यह सब कुछ सही हो गया। मैं पानी से बाहर एक मछली की तरह होगा। और एक बेला के रू

प में फटि है।तो यह कानाफूसी के लिए है।मैं सर्फ इसकी तुलना मौजूदा प्रकार के मॉडल से करना चाहता हूं।और ऐसा करने के लिए एक लोकप्रिय पुस्तकालय भाषण मान्यता पुस्तकालय है।जिस तरह से हम स्पीच रकिग्नशिन लाइब्रेरी को चलाते हैं, वह है कहिम इसे आयात करते हैं और फिर इस पहचानकर्ता ऑब्जेक्ट को बनाते हैं, जिससे हम तब अपनी ऑडियो फ़ाइल के साथ लोड कर सकते हैं।उसके बाद, आप पहचानकर्ता वस्तु ले सकते हैं और उसके लिए कुछ अलग पहचान के तरीके हैं।और हम Google पहचान का उपयोग करने जा रहे हैं और देखते हैं कि परिणाम क्या है।तो ऐसा लगता है कि यह कोई वरिम चहिन नहीं मिला है, और क्लाउड नौ अलग है।मैं एक ट्रकि पोनी के रूप में क्लाउड नाइन पर रहना पसंद करूंगा जो एक मक्खी को चोट नहीं पहुंचाएगा।लेकिन एक बात ध्यान में रखें कि यह वास्तव में Google भाषण मान्यता एपीआई का उपयोग कर रहा है।व्हिस्पर लाइब्रेरी, आपके पास वास्तव में मॉडल डाउनलोड है और यह आपका उपयोग करना है।मैं यह भी सलाह देता हूं कि आप व्हिस्पर पेपर पर एक नज़र डालें, जो इस कोड के साथ जारी किया गया था।वे इस बारे में भी विस्तार से जाते हैं कि मॉडल को कैसे प्रशिक्षित किया गया था और वास्तुकला का उपयोग किया जाता है।व्हिस्पर विभिन्न भाषाओं के एक समूह पर काम करता है।वे जो प्रदर्शन कहते हैं वह भाषा के आधार पर भिन्न होता है।तो आप यहां GitHub Repo पर जा सकते हैं, जहां उनके पास एक प्लॉट है जो दिखाता है कि कौन सी भाषाएं वास्तव में यहां सलाखों के लिए सबसे अच्छा प्रदर्शन करती हैं।छोटा बेहतर है और बड़ा मतलब है कि यह खराब प्रदर्शन करता है।तो अभी भी बहुत प्रभावशाली भाषाओं की संख्या जो इस मॉडल पर काम करती है।