

ちょっとYouTube、このビデオでは、Whisperと呼ばれるPythonの無料オープンソースパッケージを使用して、任意のオーディオをテキストにすばやく変換する方法を紹介します。私はそれをインストールしたことを示し、それをどのように実行したかの例を示し、それを既存のライブラリと比較します。したがって、最初から、ここで見ているHubリポジトリを入手して、インストール方法について指示を出したいと思うでしょう。ここで、ピップをインストールするときに留意すべきことは、名前のささやきだけをインストールするだけで、適切なバージョンをインストールしません。このgitリポジトリからインストールしたいと思います。したがって、このPIPインストールコマンドを使用して、Pythonを実行している環境で実行してください。また、ここでは、FFMPEGがインストールされる必要があると述べました。それを行うための指示はいくつかありますが、私はすでにコンピューターにそれをインストールしていました。ささやきインストールがあるので、これをテストできるオーディオを作成しましょう。だから私はいくつかのイディオムと言うつもりです。通常、イディオムはモデルが理解するのが難しいです。これは単なるテキストのスピーチですが。これはちょっと楽しいでしょう。ハエを傷つけない1つのトリックポニーとしてクラウド9に参加したいと思います。私は水から出た魚のように、そして天候の下にあるフィドルのようにフィットするでしょう。これを救いましょう。波として保存しましょう。彼らは、インストールされたら、コマンドラインからこれを直接実行する方法についての指示を持っています。Python APIの使用法を紹介します。だから本当に簡単です。ささやきをインポートするだけです。その後、モデルを作成します。これはロードされます。ベースと呼ばれるモデル。そして、このモデルオブジェクトを使用するだけで、オーディオファイルで転写されます。だから私はそれをイディオムと名付けました。Waveバージョンを使用しましょう。これを結果を返してほしい。今、私はこれを以前に実行したときに気づきました、クーダのハーフテンソルとフロートテンソルのためにこのエラーが発生しました。私はこれを解決することができました。それは心に留めておくべきことです。それがあなたのためにうまくいかない場合は、浮かぶポイント16を落ちて落ちる必要があるかもしれません。そして、あなたはそれがここで実行された後に見ることができます、それはすでに英語として言語を検出しました、そして、この結果のオブジェクトにはそれらにいくつかの異なる方法がありますが、私たちがこれの中に入りたいのは単なるテキストであり、それが見た目であることがわかりました結果が良いように、私はフライを傷つけない1つのトリックポニーとしてクラウドナインになりたいと思います私は水から魚のようになり、これはこの魚を少し混乱させましたフィドルのようにフィットし、多分私はそれを明確に言わなかったかもしれません。そのため、進行状況バーが渡っているのを見るかもしれません。そのモデルをダウンロードする必要があります。そして、この転写を実行すると、実際にオーディオファイルの30秒のチャンクを取得し、予測を実行していると述べています。これで、より低いレベルのアプローチである別のアプローチもあります。実際にモデルを作成し、オーディオオブジェクトとパターントリミングを作成します。したがって、このオーディオチャンク

がわずか30秒であることを確認してください。秒またはそれは、モデルが入力として期待する長さであるため、30秒でそれを軽くたくでしょう。次に、ログマウススペクトログラムを作成します。それは言語を検出しているので、ここでデコードして、望むならもっと多くのオプションを提供できます。このセルを実行すると、このエラーが再び取得されます。これをデコードオプションに設定できるようになりました。FP16は障害に等しくなります。そして実際、今回はすべてが正しいように見えます。私は水から魚のようになります。フィドルとしてフィットしています。ささやきのためです。既存のタイプのモデルと比較したいだけです。そして、これを行うための人気のライブラリは、音声認識ライブラリです。音声認識ライブラリの実行方法は、それをインポートしてからこの認識器オブジェクトを作成し、オーディオファイルをロードできることです。その後、認識器オブジェクトを取得でき、そのためのいくつかの異なる認識方法があります。そして、Googleの認識を使用して、結果が何であるかを見てみましょう。したがって、句読点を追加しなかったように見え、クラウドナインは異なります。ハエを傷つけない1つのトリックポニーとしてクラウドナインに参加したいと思います。しかし、心に留めておくべきことの1つは、これが実際にGoogle Speech Septunion APIを使用していることです。ささやきライブラリ、あなたは実際にモデルをダウンロードしていて、それはあなたのものです。また、このコードでリリースされたウィスパーペーパーをご覧ください。また、モデルがどのように訓練されたか、および使用されているアーキテクチャについて詳しく説明します。ささやきは、さまざまな言語で動作します。彼らが言うパフォーマンスは、言語によって異なります。したがって、Github Repoでここに行くことができます。ここでは、このバーで実際にどの言語が最適かを示すプロットがあります。小さい方が優れており、大きいということは、パフォーマンスが悪化することを意味します。このモデルが機能する言語の数は、まだかなり印象的です。