Wstęp

Obecnie coraz więcej dokumentów przechowujemy w sposób elektroniczny. Od 2020 roku pojawiła się nawet możliwość załatwienia spraw urzędowych przez Internet i składanie wniosków bez wychodzenia z domu. Trzymanie dokumentów na komputerze ubezpiecza nas od możliwej straty informacji z powodu fizycznych szkód, a w przypadku ich dużej ilości wielką zaletą jest szybkość, podczas sortowania, filtracji i wyszukiwania danych. Danymi, które te dokumenty mogą zawierać, są tekst i obrazy. Tekst - jest zbiorem zdań, zdania jest zbiorem słów, a słowo składa się z liter.

Wprowadzanie tekstu, zawartego w dokumencie do komputera, odbywa się za pomocą klawisz klawiatury, każda z których jest oznakowana pojedynczymi literami, cyframi lub symbolami interpunkcyjnymi. Komputer traktuje wciśnięta klawisze jako unikalny kod binarny lub liczbowy, a następnie używa wybranego przez system pliku czcionki, w celu wyświetlenia powiązanego z nim znaku na ekranie monitora. Stworzenie elektronicznej kopii dokumentu poprzez ręczne wpisywania każdego znaku jest dość czasochłonne w przypadku, gdy składa się on z więcej niż jednej strony, ale takie rozwiązanie powoduje, że w wyniku dostajemy stworzony edytowalny tekst.

Szybszym sposobem na przekazanie dokumentu komputerowi jest stworzenie jego wersji cyfrowej za pomocą skanera lub kamery. Taki dokument zostanie umieszczony w pamięci komputera jako plik graficzny, zapisany w odpowiednim formacie, najczęściej którymi są: JPEG, PDF, PNG albo TIFF. Tego typu plik przedstawia obraz na ekranie za pomocą zawartej informacji o ilości pikseli oraz ich położenia i koloru. Z tego wynika, że taka wersja dokumentu elektronicznego, służy tylko do podglądu jego treści.

Od lat 1950 ciągle się rozwija rozwiązanie, za pomocą którego w dzisiejszych czasach można uzyskać tekst z każdego pliku graficznego. Jest to oprogramowanie komputerowe, które się nazywa „OCR”. W swojej pracy pokażę jego zastosowanie we własnej aplikacji, cel której, jest odczytywanie danych z najczęściej spotykanych wśród ludzi dokumentów. Jest to dokument potwierdzający sprzedaż – „Paragon Fiskalny”.

**Rozdział I**

1. Optyczne rozpoznawanie znaków

Optyczne rozpoznawanie znaków (Optical Character Recognition, OCR), jest to technologia, która automatycznie identyfikuję liczby, litery oraz znaki interpunkcyjne w plikach graficznych i konwertuje ich do postaci danych, zrozumiałych komputerowi. Na wejściu podawany jest wcześniej przygotowany plik rastrowy, który poddaje się analizie, w celu odnajdywania w nim zbioru pikseli, wizualizujących symbol, a następnie do każdego znalezionego, przypisuje się kod binarny lub liczbowy z tablicy znaków. W wyniku, zwracany jest stworzony, edytowalny tekst. W zależności od wybranego oprogramowania OCR, rozpoznawanie tekstu może odbyć się w trakcie generowania pliku graficznego, przy użyciu skanera lub urządzenia wyposażanego w kamerę. W innym przypadku, możliwe jest przekazanie przygotowanej wcześniej grafiki zawierającej tekst do systemu OCR.

Obecnie nie istnieję idealnego oprogramowania, które zawsze potrafi rozpoznać tekst o dowolnej czcionce lub pismo ręczne z dokładnością do 100%, lecz istnieje możliwość dostosować go do odpowiedniego problemu. Prawdopodobieństwo tego, że OCR odniesie sukces podczas pracy, zależy w pierwszej kolejności od jakości materiału wejściowego.

Technologia OCR składa się, z oprogramowania i pliku rastrowego. Plik może być stworzony za pomocą takich urządzeń jak skaner, faks, aparat fotograficzny lub przy użyciu programów służąc do tworzenia/edycji grafiki. W przypadku skanera, rozpoznawanie tekstu może zostać wykonane podczas skanowania, wtedy wynik zostanie zapisany do pliku tekstowego. W innych sytuacjach, mając już gotowy plik graficzny, rozpoznanie zaczyna się po tym, jak przekażemy go na wejście do wybranej aplikacji.

Oprogramowanie OCR nie jest idealnym rozwiązaniem, ponieważ …, lecz … . W dzisiejszych czasach większość programów stworzonych do realizacji tego celu, używają sztucznej inteligencji co zwieksza… . (o sieciach nieronowych)

**1.1. Proces rozpoznawania znaków**

Konwertowanie pliku graficznego do postaci edytowalnego tekstu , podzielone jest na kilka kroków. Każdy krok to zbiór odnośnych algorytmów wykonujących część pracy OCR.

**1.1.1 Pre-processing**

Istotną rolę, w procesie rozpoznawania znaków, odgrywa jakość materiału wejściowego. Obrazy o małej rozdzielczości, z niską kontrastowością lub z szumem są trudniejsze w rozpoznaniu programem OCR. Dlatego, przed procesem identyfikacji znaków, implementuje się do obrazu metody programowe, które mają na celu polepszyć jakość pliku rastrowego, a to, spowoduję zwiększenie współczynnika rozpoznawania tekstu. Następnie, wymienione są, kilka z podstawowych technik, wykonywanych podczas Preprocess’ingu:

**Binaryzacja(ang. binarization)**

Algorytm binaryzacji konwertuje obraz kolorowy lub z odcieniem szarości na obraz binarny, w którym pikseli mogą przyjmować jedną z dwóch wartości. Celem tego procesu jest oddzielenie pierwszego planu, czyli tekstu, od tła.

**Usuwanie szumów(ang. De-speckle)**

Jedną z najczęściej pojawiających się problemów, powiązanych z jakością obrazu jest powstanie różnego rodzaju szumów. Jednym z powodów ich pojawienia, może być zanieczyszczenia, znajdujące się na obiektywie kamery cyfrowej lub szubie skanera. Brud lub rysy znajdujące się na dokumencie też sprzyjają pojawieniu szumów na obrazie.

**Poziomowanie(ang. de-skew)**

Podawane na wejściu dokumenty, czasami mogą być zeskanowane lub sfotografowane ukośnie. Tekst w podobnych materiałach jest trudniejszy do identyfikacji programami OCR. W związku z tym, żeby uniknąć podobnych problemów i zwiększyć szanse na prawidłowe rozpoznanie znaków, do obrazu stosują się algorytmy które wyprostowują linii tekstu.

**Analiza układu dokumentu(ang. document layout analysis)**

Zanim program zacznie rozpoznawać znaki, najpierw istotnie jest zidentyfikować i sklasyfikować elementy występujące w dokumencie. Zazwyczaj takimi elementami mogą być zdjęcia, grafika, tabeli, linii, tekst, tytuł , akapity, łamy. Bez tej analizy kolejność zdań w rozpoznanym tekście może się różnić od dokumentu wejściowego, a wyrazy przenoszone mogą zostać podzielone na pól.

**1.1.2 Character recognition**

Oprogramowanie OCR rozpoznaje tekst konsekwentnie. Najpierw łączy pikseli w ewentualne znaki, a znaki – w ewentualne słowa. Następnie, program sprawdza czy jest odpowiednik tego słowa w słowniku. Jeżeli tak, to zostaje ono oznaczone jako rozpoznane. W innym przypadku, program zaproponuje najbardziej prawdopodobny wynik, tym samym zmniejszy współczynnik rozpoznawania tekstu.

**1.1.3 Post Processing**

**1.2 Użycie sieci niejonowej do rozpoznawania znaków**

Нейронная сеть єто программа создана в подобии к человеческой но очень отличаеться. Нейро сеть можна обучить принимать правильные решения, но только к конкретной теме. Отвечая правильно нейросеть ничего не знает о теме и о вопросе, для нейросети существуют только весы числа и тд(обяснить каждый елемент: вход внутренность и выход). Подавая ей одни данные и то какой должен быть результат она меняет соответсвенно значение весов и других. В результате, подавая ей данные о которых она не знает она будет испольховать правильно расставленные весы и выдавать ответ который может нас устроитьт а может не устроить. Так мы пределяем нужно ли ей доучиваться или нет. РАСКАЗАТЬ О ГЕЙТАХ  
  
**КАК ПРОИСХОДИТ ОБУЧЕНИЕ**

**Шаг-за-шагом вперед  
  
И назад (рассказать про гардиент)**

**Раскзать о сетях рекуренцийних**

**ЛСТМ – Это рекуренцийная сеть. Зачем нудны рекуренцийные сети. В чем их плюч и в чем минус. Начать рассказать о том что внутри о гейтах.**

**Тессеракт – это балбалабла он, работал так и так, но теперь от версиии такой то работает с сетями рекуренци ЛСТМ, как происходит распознание текста. Рассказать о том что это от гугла , поискать и попереписывать информацию с офф странички.**

**Конец.   
Начать с того как я обучал тессеракт.**