

PLAIN TEXT , IDEAL SITUATION ANALYSIS

RAW IMAGE

CHAPTER ONE

The Wake-Up Call

He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on.

Ground Truth

"CHAPTER ONE The Wake-Up Call He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on."

WORD ERROR RATE

The WER calculation is based on a measurement called the "Levenshtein distance."

The Levenshtein distance is a measurement of the differences between two "strings."

To put it in a simple formula,

Word Error Rate = (Substitutions + Insertions + Deletions) / Number of Words Spoken

But how do you add up those factors? Let's look at each one:

A substitution occurs when a word gets replaced (for example, "noose" is transcribed as "moose")

An insertion is when a word is added that wasn't said (for example, "SAT" becomes "essay tea")

A deletion happens when a word is left out of the transcript completely (for example, "turn it around" becomes "turn around")

Character Error Rate

This value indicates the percentage of characters that were incorrectly predicted.

The lower the value, the better the performance of the ASR system with a CharErrorRate of 0 being a perfect score.

The Character Error Rate (CER) compares, for a given page, the total number of characters (n), including spaces, to the minimum number of insertions (i),

substitutions (s) and deletions (d) of characters that are required to obtain the Ground Truth result.

The formula to calculate CER is as follows: $CER = [(i + s + d) / n] * 100$

TESSERACT OCR

EXTRACTION -

CHAPTER ONE The Wake-Up Call He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on.

#Detect Language other than English

IT considers the special symbols,

it is case sensitive as well

Word_Error_Rate: 0.018691588785046728

Char_Error_Rate: 0.006546644844517185

CHAPTER ONE CHAPTER ONE

The Wake-Up Call

He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on.

KERAS OCR

Word_Error_Rate: 0.6635514018691588

Char_Error_Rate: 0.46808510638297873

EXTRACTION -

chapter one the wakeup call middle collapsed right of packed he in the he courtroom a was of
this distinguished trial lawyers he also country most s one was threethousandollar italian
who well known for the man was a as which draped his wellfed frame for his remarkable
of suits string as by the shock of what legal victories t simply stood theres paralyzed had
witnessed the julian mantle had been reduced l just great to helpless victim and
squirming the ground like in was on a a now shaking shivering like everything fant and
and sweating maniac a seemed in slow motion from that point to move on

Case sensitivity is lost

new lines is not detected

Does not consider the special chars and punctuations

Keras OCR Result Example

CHAPTER ONE

The Wake-Up Call

He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on.

chapter
wake up
the
middle
right
collapsed
he
in
the
of
this
country
most
ones
who
known
well
for
man
was
as
well fed
draped
which
his
frame
suits
there
stood
legal
victories
simply
witnessed
had
the
just
great
victim
squirring
and
was
now
a
shivering
shaking
sweating
and
fant
and
motion
from
seemed
in
slow
to
move

one
call
packed
of
he
courtroom
a
was
trial
lawyers
he
also
was
three thousand dollar
Italian
the
string
remarkable
for
his
of
as
by
shock
what
paralyzed
the
of
been
reduced
mantle
had
to
helpless
ground
the
like
in
on
a
like
everything
maniac
a
point
that
on

EASY OCR

Word_Error_Rate: 0.04672897196261682

Char_Error_Rate: 0.011456628477905073

Extraction -

CHAPTER ONE The Wake-Up Call He collapsed right in the middle of a packed courtroom He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to victim and was now squirming on the ground like a helpless infant; shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on

either paragraph false or true, the prediction is found as line or paragraph

special characters are also detected

It is case sensitive

provides confidence too.

CHAPTER ONE

CHAPTER ONE

The Wake-Up Call

The Wake-Up Call

He collapsed right in the middle of a packed courtroom.

He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything seemed to move in slow motion from that point on.

PADDLEOCR (Parallel Distributed Deep Learning)

Word_Error_Rate: 0.07476635514018691

Char_Error_Rate: 0.011456628477905073

Extraction -

CHAPTER ONE The Wake-Up Call He collapsed right in the middle of a packed courtroom. He was. one of this country's most distinguished trial lawyers. He was also. a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of. legal victories. I simply stood there, paralyzed by the shock of what. I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless in fant, shaking and shivering and sweating like a maniac. Everything. seemed to move in slow motion from that point on..

it does acknowledge special characters

it does acknoledge case sensitivity

requires GPU

it extracts sentences and not words

CHAPTER ONE
CHAPTER ONE

The Wake-Up Call

The Wake-Up Call

He collapsed right in the middle of a packed courtroom. He was.

He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also.

one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian

a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of

suits which draped his well-fed frame as for his remarkable string of legal victories. I simply stood there, paralyzed by the shock of what

legal victories. I simply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to

I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless in-

a victim and was now squirming on the ground like a helpless infant, shaking and shivering and sweating like a maniac. Everything

seemed to move in slow motion from that point on.

seemed to move in slow motion from that point on.

DONUTOOCR

Extraction –

<s_cord-v2><s_menu><s_nm> The Wake-Up Call</s_nm><s_num> He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also</s_nm><s_num> a man who was as well known for the three-thousand-dollar Italian</s_nm><s_num> suits which draped his well-fed frame as for his remarkable string of</s_nm><s_num> legal victories. Isimply stood there, paralyzed by the shock of what</s_nm><s_num> I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless in-</s_nm><s_num> a</s_nm><s_price> a</s_price><sep/><s_nm> fant, shaking and shivering and sweating like a manac. Everything seemed to move in slow motion from that point on.</s_menutype_cnt></s_total></s>

After post - processing

The Wake-Up Call He collapsed right in the middle of a packed courtroom. He was one of this country's most distinguished trial lawyers. He was also a man who was as well known for the three-thousand-dollar Italian suits which draped his well-fed frame as for his remarkable string of legal victories. Isimply stood there, paralyzed by the shock of what I had just witnessed. The great Julian Mantle had been reduced to a victim and was now squirming on the ground like a helpless in- a a fant, shaking and shivering and sweating like a manac. Everything seemed to move in slow motion from that point on.

Word_Error_Rate: 0.08411214953271028

Char_Error_Rate: 0.03273322422258593

it does acknowledge special characters

it does acknoledge case sensitivity

it extracts sentences and not words

NOTE – TESSERACT AND DONUT OCR CANNOT TO TOGETHER IN ONE ENV, THEY FACE SOME DEPENDENCY ISSUE