

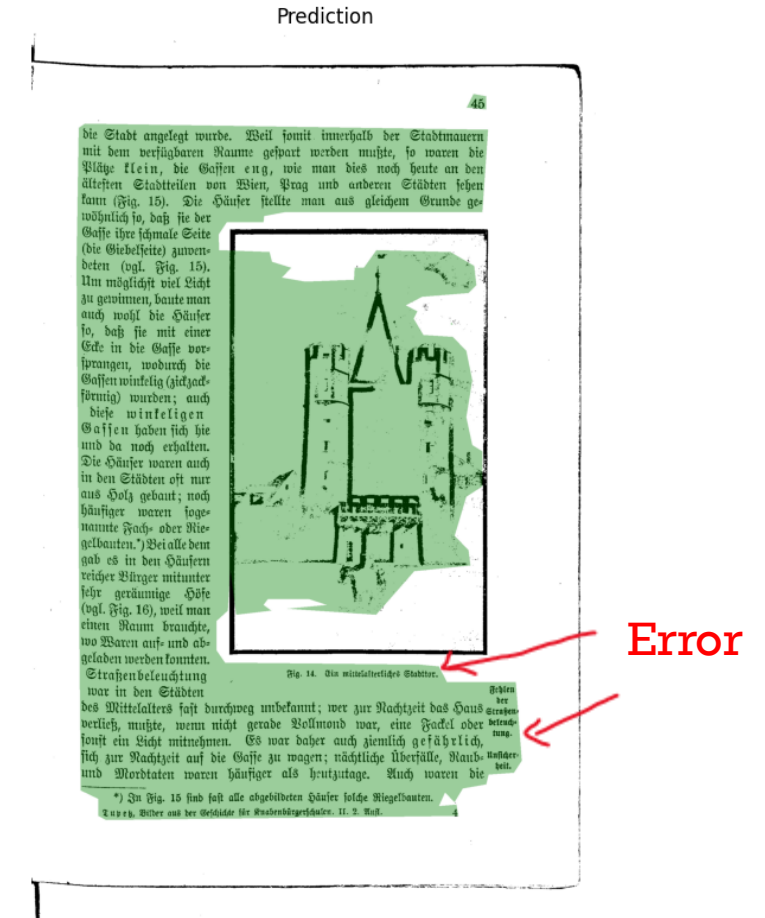
Chunk-based approach

NLP - Praktikum

Alzbeta Hrabosova

Manual data labeling

- Error score per page
error words / total words



Ground Truth

194

ganz Europa, auf welcher die durch Napoleon in Verwirrung gebrachten staatlichen Verhältnisse neu geordnet wurden. Die Oberleitung hatte bei diesen Verhandlungen der einflußreichste Staatsmann, den Europa damals besaß, der österreichische Minister Metternich, der kurz zuvor zur Belohnung für seine Verdienste um die Machtstellung Österreichs in den Fürstenstand erhoben worden war. Österreich erhielt durch den Wiener Kongreß im wesentlichen den Umfang, den es heute noch besitzt; von den Erwerbungen, die es zur Zeit Napoleons gemacht hatte, blieben ihm insbesondere Salzburg und Venezien, wodurch die Monarchie, da auch die Lombardei österreichisch war, in zweckmäßiger Weise abgerundet (arrondiert) wurde. Im deutschen Bunde, der an die Stelle des deutschen Reiches trat, führte Österreich den Vorsitz.

Zu Ehren der in Wien versammelten Monarchen und Würdenträger fanden rauschende, glanzvolle Festlichkeiten statt, die lange im Gedächtnisse der Zeitgenossen fortlebten. Viele derselben wurden in Laxenburg gefeiert, dem Lieblingsitze des Kaisers Franz, der sich daselbst in getreuer Nachbildung mittelalterlicher Burgen die „Franzensburg“ hatte erbauen lassen (s. die Abbildung Fig. 59). Eines dieser Feste verherrlichte auch der größte Meister der Töne, den die Welt damals besaß, Ludwig van Beethoven, durch die Aufführung eines seiner Tonwerke, der „Schlacht bei Vittoria“. (Bei Vittoria in Spanien hatte der englische Feldherr Wellington, welcher ebenfalls dem Kongresse beiwohnte, 1812 einen glänzenden Sieg über die Franzosen erfochten.)

Unter der väterlichen Regierung des „guten Kaisers Franz“ erholte sich Österreich rasch von den Wunden, welche ihm die Napoleonischen Kriege geschlagen. Ackerbau und Viehzucht, Handel und Gewerbe, Wissenschaft und Kunst gelangten zu neuer Blüte.

Für die Landwirtschaft war besonders die Gründung von landwirtschaftlichen Schulen, dann von Ackerbaugesellschaften und landwirtschaftlichen Vereinen, für das Gewerbe die Errichtung der technischen Institute in Wien, Prag, Graz u. s. w. und die Veranstaltung von Gewerbeausstellungen wichtig. Erzherzog Johann, der Bruder des Kaisers, selbst Eigentümer mehrerer Eisenwerke in Steiermark, bewog die Besitzer der Eisenwerke, welche die unerschöpflichen Spateisensteinlager des steirischen Erzberges ausbeuteten, die noch bestehende Vordernberger Eisenwerksgesellschaft zu gründen, so daß nun die Gewinnung und Verwendung des Eisens im großen betrieben werden konnte. Der Kohlenbergbau, im Zusammenhang damit das Maschinen- und Eisenbahnwesen begannen sich in dieser Zeit zu entwickeln. Die erste Eisenbahn in Österreich war die von Budweis nach Linz (1825), welche aber anfangs nur eine Pferdebahn war. Auch die großartigen Kunststraßen über den Loibl-Paß und das Stifiser Joch entstanden in dieser Zeit. Kaiser

Prediction

194

ganz Europa, auf welcher die durch Napoleon in Verwirrung gebrachten staatlichen Verhältnisse neu geordnet wurden. Die Oberleitung hatte bei diesen Verhandlungen der einflußreichste Staatsmann, den Europa damals besaß, der österreichische Minister Metternich, der kurz zuvor zur Belohnung für seine Verdienste um die Machtstellung Österreichs in den Fürstenstand erhoben worden war. Österreich erhielt durch den Wiener Kongreß im wesentlichen den Umfang, den es heute noch besitzt; von den Erwerbungen, die es zur Zeit Napoleons gemacht hatte, blieben ihm insbesondere Salzburg und Venezien, wodurch die Monarchie, da auch die Lombardei österreichisch war, in zweckmäßiger Weise abgerundet (arrondiert) wurde. Im deutschen Bunde, der an die Stelle des deutschen Reiches trat, führte Österreich den Vorsitz.

Zu Ehren der in Wien versammelten Monarchen und Würdenträger fanden rauschende, glanzvolle Festlichkeiten statt, die lange im Gedächtnisse der Zeitgenossen fortlebten. Viele derselben wurden in Laxenburg gefeiert, dem Lieblingsitze des Kaisers Franz, der sich daselbst in getreuer Nachbildung mittelalterlicher Burgen die „Franzensburg“ hatte erbauen lassen (s. die Abbildung Fig. 59). Eines dieser Feste verherrlichte auch der größte Meister der Töne, den die Welt damals besaß, Ludwig van Beethoven, durch die Aufführung eines seiner Tonwerke, der „Schlacht bei Vittoria“. (Bei Vittoria in Spanien hatte der englische Feldherr Wellington, welcher ebenfalls dem Kongresse beiwohnte, 1812 einen glänzenden Sieg über die Franzosen erfochten.)

Unter der väterlichen Regierung des „guten Kaisers Franz“ erholte sich Österreich rasch von den Wunden, welche ihm die Napoleonischen Kriege geschlagen. Ackerbau und Viehzucht, Handel und Gewerbe, Wissenschaft und Kunst gelangten zu neuer Blüte.

Für die Landwirtschaft war besonders die Gründung von landwirtschaftlichen Schulen, dann von Ackerbaugesellschaften und landwirtschaftlichen Vereinen, für das Gewerbe die Errichtung der technischen Institute in Wien, Prag, Graz u. s. w. und die Veranstaltung von Gewerbeausstellungen wichtig. Erzherzog Johann, der Bruder des Kaisers, selbst Eigentümer mehrerer Eisenwerke in Steiermark, bewog die Besitzer der Eisenwerke, welche die unerschöpflichen Spateisensteinlager des steirischen Erzberges ausbeuteten, die noch bestehende Vordernberger Eisenwerksgesellschaft zu gründen, so daß nun die Gewinnung und Verwendung des Eisens im großen betrieben werden konnte. Der Kohlenbergbau, im Zusammenhang damit das Maschinen- und Eisenbahnwesen begannen sich in dieser Zeit zu entwickeln. Die erste Eisenbahn in Österreich war die von Budweis nach Linz (1825), welche aber anfangs nur eine Pferdebahn war. Auch die großartigen Kunststraßen über den Loibl-Paß und das Stifiser Joch entstanden in dieser Zeit. Kaiser

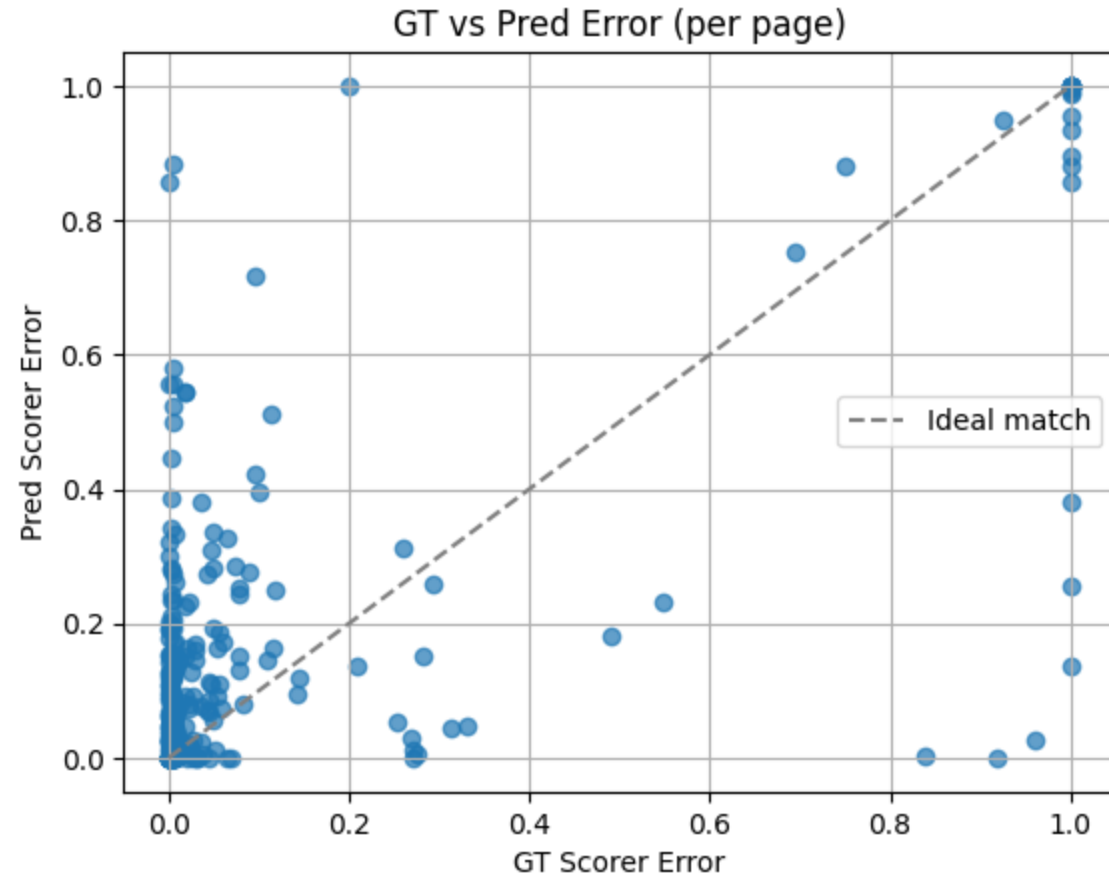
No Errors

Chunks-size/Overlap-size experiments

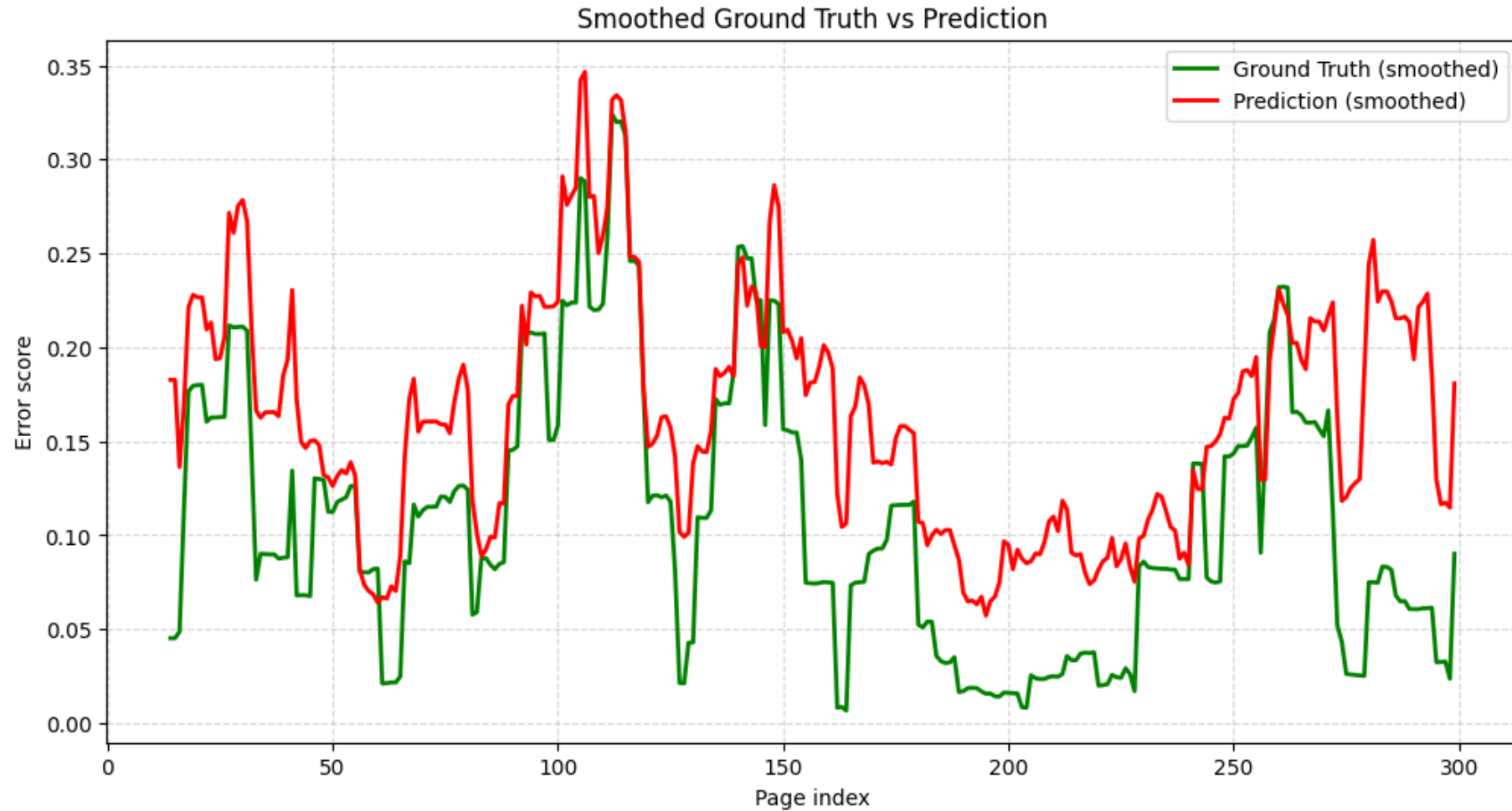
Chunk size	Overlap size	Mean		Std. dev.		Correlation
		GT	PR	GT	PR	
10	5	0.227	0.503	0.371	0.308	0.81
10	7	0.227	0.514	0.371	0.309	0.799
15	7	0.227	0.328	0.371	0.386	0.851
15	10	0.227	0.328	0.371	0.402	0.833
20	15	0.227	0.278	0.371	0.398	0.879
20	15	0.227	0.284	0.371	0.400	0.877

Evaluation

Metric	GT Scorer	Pred Scorer
Count	300	300.0000
Mean	0.1078	0.1657
Median	0.0048	0.0676
Stdev	0.2685	0.2532
Min	0	0.0000
Max	1	1.0000
MAE	0.1078	0.1657
Correlation	0.7088	—
GT vs Pred RMSE	0.2075	—



Evaluation



Supervised OCR Error Detection

Presented by: Arslan Ahmed

Approach



Word-level labeling

Label each word in input line as “error”, “correct”, or “unmatched”



Tokenization and Model Training

Train a token classification model to predict error/correct labels



Evaluation

Assess model performance on ManualGTScorer

How to label each word?

Overview:

- For each input file, we compare each line to its corresponding ground truth (GT) line.
- If we find an exact line match, all words in that line are labeled as “correct”.
- For lines without exact matches, we proceed with contextual (partial) matching to assign word-level labels.

Labeling Rules:

1. Exact Marginalia or Paragraph Match:

- All words in an input line that exactly matches a marginalia or paragraph line in GT are labeled correct.

How to label each word?

2. Contextual Alignment (Partial Match):

- For lines without an exact match, we perform word-by-word comparison using context from the ground truth paragraph.
- Context: An input word is labeled 'correct' if it has an exact match in the ground truth line, along with a match in either its previous or next word.

3. Marginalia Phrase Inside Line:

- Words in input lines that match a marginalia phrase (from GT) within a paragraph line are labeled error.

4. Unmatched Words:

- Words not aligned to any GT or marginalia are labeled unmatched.

Word label - Example

l_0001: We(unmatched) | te(unmatched)
l_0002: Q(unmatched)
l_0003: S(unmatched)
l_0004: 1(unmatched)
l_0005: 16(correct)
l_0006: Rückzug(error) | der(error) | wichen(correct) | zurück(correct) | und(correct) | ließen(correct) | sich(correct) | in(correct) | keine(correct) | offene(correct) | Schlacht(correct) | ein(correct) | sondern(correct)
l_0007: Skythen(correct)
l_0008: verwüsteten(correct) | nur(correct) | das(correct) | Land(correct) | und(correct) | verschütteten(correct) | die(correct) | Brunnen(correct) | bis(correct) | den(correct)
l_0009: Persern(correct) | die(correct) | Lebensmittel(correct) | ausgingen(correct) | und(correct) | Darius(correct) | den(correct) | Befehl(correct) | zum(correct)
l_0010: Rückzug(error) | der(error) | Rückzug(correct) | gab(correct) | Jetzt(correct) | begannen(correct) | die(correct) | Skythen(correct) | die(correct) | Verfolgung(correct) | um(correct)
l_0011: Perser(correct)
l_0012: schwärmten(correct) | das(correct) | Heer(correct) | von(correct) | allen(correct) | Seiten(correct) | und(correct) | bedrängten(correct) | es(correct) | so(correct) | sehr(correct)
l_0013: Die(error) | Griechen(error) | daß(correct) | es(correct) | froh(correct) | sein(correct) | mußte(correct) | wieder(correct) | den(correct) | Ister(correct) | zu(correct) | erreichen(correct) | Zum(correct) | Glück(correct)
l_0014: bei(correct) | der(correct)
l_0015: Brücke(correct)
l_0017: Fig(correct) | 8(correct) | Darius(correct) | mit(correct) | Dienern(correct) | uber(error) | ihm(correct) | Ormuzd(correct)
l_0018: Wandbild(correct) | im(correct) | Palast(correct) | zu(correct) | Persepolis(correct)
l_0019: fand(correct) | man(correct) | dort(correct) | die(correct)
l_0020: Brücken(correct) | noch(correct) | vor(correct)

Word label - Example

Ground Truth (Regions)

Prediction (Lines)

Prediction (Line coloring by word-level labels)

- 16 -

- 16 -

- 16 -

Rückzug der Skythen. wichen zurück und ließen sich in keine offene Schlacht ein, sondern verwüsteten nur das Land und verschütteten die Brunnen, bis den Persern die Lebensmittel ausgingen und Darius den Befehl zum Rückzug gab. Jetzt begannen die Skythen die Verfolgung, umschwärmten das Heer von allen Seiten und bedrängten es so sehr, daß es froh sein mußte, wieder den Ister zu erreichen. Zum Glück fand man dort die Brücken noch vor; der Mann, dem Darius seine Rettung verdankte, war Histäus, der Stadtherrscher (Tyrrann) von Milet, einer der griechischen Städte an der Westküste Kleasiens. Es waren nämlich Skythen an den Fluß gekommen und hatten den Griechen geraten, die Brücken abzubrechen, dann würde das ganze persische Heer zugrunde gehen; Miltiades, ein Griechenfürst vom Hellespont, hatte dem beige stimmt, da dann die Griechen Kleasiens sich vom persischen Joch befreien könnten, aber Histäus widersprach und meinte umgekehrt, die Stadtherrscher hätten alle Ursache, die Rettung des persischen Königs und Heeres zu wünschen, da sie nur durch des Königs Macht in ihrer Stellung geschützt würden. Ohne diesen Schutz hätten die Bürgerschaften ihrer Städte sie schon längst vertrieben. So blieben die Brücken erhalten und Darius war gerettet. Histäus wurde belohnt, Miltiades floh in seine griechische Heimat.



Fig. 8. Darius mit Dienern (über ihm Ormuzd). Wandbild im Palast zu Persepolis.



Fig. 8. Darius mit Dienern (über ihm Ormuzd). Wandbild im Palast zu Persepolis.



Fig. 8. Darius mit Dienern (über ihm Ormuzd). Wandbild im Palast zu Persepolis.

Word label - Example

Ground Truth (Regions)

welchem Wald und Wiese vorherrschen, die Region des böhmischen Landes und der Hochalpen. Hier gewahren wir Gegenstände weit ausgedehnter Massen (A/jes), inselartig zerstreuter Berggruppen (Böhmerwald, Erzgebirge, Riesengebirge, Sudeten, Böhm.-näh. Höhenzug, Karst) und Verbindungen beider Formas (Karpathen).

2. Bestimmt auf ähnliche Art die Lage der Bergketten und Gipfel, die ihr auf der Karte von Mittel-Europa schon kennen lerntet, und vervollständigt die Kenntnis durch das Hineinnehmen der fehlenden Namen aus den Kronländerkarten. Auf diese Art werdet ihr nach und nach von sehr vielen, selbst kleinen Berggipfeln Rechenschaft geben können.

3. That dasselbe Kinderweise oder nach Flugschritten, oder in irgend einer andern (immerhin aber in einer gewissen) Ordnung mit den Orten und besonders dabei die Lage (an Flüssen, Kanälen, Seen, auf der Ebene, im Gebirge, an Escalen, an Meere u. s. f.), die Bevölkerungsklasse, die Verbindung mit andern Orten durch Straßen, Bahnen

Nur wenige bewohnte Orte finden wir in den Hochtäälern der Alpen. Unsere Karte zeigt nur sehr wenige an, die (weil als kleine Orte mit schwarzen Punkten bezeichnet) von den Schneegipfeln sorgsam unterschieden werden müssen. Ein Vergleich der Karten 7 und 8 wird jeden entscheidenden Zweifel bannen.

nicht ermüdet, wenn die guten Früchte nicht lange auf sich warten lassen. Der werlet auf diese Weise dahin kommen, die topische Geographie eines Landes, soweit sie von der Karte abgelesen werden kann, selbst zu machen, und dies wird auch ungleich größerer Nutzen bringen, als jedes noch so feine Aus-

wendigeren aus einem Bache. Um dieses Ziel zu erreichen, muss unser Bestreben zuerst dahin gehen, eine Karte richtig lesen zu lernen, das Uebrige findet sich dann leicht.

Mittel-Europa hier schon bekannt sind. Suchet aus den Krallenderkarten die

Prediction (Lines)

Welchen Wald und Wiese vorherrschen, die Region des höheren Landes und der Hochtäler. Hier gewahren wir Gegenätze weit ausgedehnter Masses (Alpen), isolierend zerstreuter Berggruppen (Böhmerwald, Erzgebirge, Riesengebirge, Sudeten, Böhmer-mähr. Höhenzug, Karst) und Verbindungen beider Formen (Karpaten).

2. Bestimmt auf ähnliche Art die Lage der Bergketten und Giegel, die hier auf der Karte von Mitteleuropa schon kennen gelernt, und vervollständigt das Kenntnis durch das Herüberziehen der fehlenden Nummern aus den Kreisländerkarten. Auf diese Art werdet ihr nach und nach von sehr vielen, selbst kleinen Bergzügen Rechenschaft geben können.

3. Tritt dasselbe Kinderwies oder nach Fungestrichen, oder in irgend einem andern (sonst nicht aber in einer gewissen) Ordnung mit den Orten und bezeichnet dabei die Lage (an Flüssen, Kanälen, Seen, auf der Ebene, im Gebirge, an Eisenbahnen, am Meere u. s. f.), die Bevölkerungsklasse, die Verbindung mit andern Orten durch Straßen, Bahnen.

Nur wenige bewohnte Orte finden wir in den Hochtälern der Alpen. Unsere Karte zeigt nur sehr wenige an, die (weil als kleine Orte mit schwarzen Punkten bezeichnet) von den Schöngipfeln vorwiegend unterschieden werden müssen. Ein Vergleich der Karten 7 und 8 wird jeden entscheidenden Zweifel bannen.

nicht erwidert, werden die guten Früchte nicht lange auf sich warten lassen. Der
verleiht auf diese Weise dahin kommen, die topische Geographie eines Landes,
soweit sie von der Karte abgelesen werden kann, selbst zu machen, und diese
wird auch ungleich grössere Nutzen bringen, als jedes noch so schöne Aus-

1. Bestimmt die Namen aller Ozeane, die auch von der Karte von Mittel-Europa her schon bekannt sind. Sucht aus den Kreisländerkarten die

weiches Wald und Wiese vorherrschen, die Region des höheren Landes und der Hochalpen. Hier gewahren wir Gegensätze weit ausgedehnter Massen (Alpen), isolierend stehender Berggruppen (Böhmerwald, Erzgebirge, Riesengebirge, Sudeten, Böhmis. mähr. Höhenzug, Karst) und Verbindungen beider Formen (Karpathen).

2. Bestimmt auf ähnliche Art die Lage der Bergketten und Gipfel, die in der Karte von Mittel-Europa schon kennen gelernt, und vervollständigt die Kenntnisse durch das Herüberreichen der fehlenden Namen aus den Kreisläufen heraus. Auf diese Art werdet ihr nach und nach von sehr vielen, selbst kleine Bergnamen Rechenenschaft geben können.

3. Tritt diese Verteilung oder nach Flächheiten, oder in irgend einer andern (numerisch aber in einer gewissen) Ordnung teil des Ortes und bemerkt dabei die Lage (an Flüssen, Kanälen, Seen, auf der Ebene, im Gebirge, an Einsehlungen, am Meere u. s. f.), die Bevölkerungsdichte, die Verteilung der andern Orte durch Systeme, Bahnen

Nur wenige bewohnte Orte finden wir in den Hochtälern der Alpen. Unsere Karte zeigt nur sehr wenige an, die (weil als kleine Orte mit schwarzen Punkten bezeichnet) von den Schneegipfeln vermagt überschritten werden müssen. Ein Vergleich der Karten 7 und 8 wird jeden entscheidenden Zweifel bannen.

nicht erstellt, werden die guten Früchte nicht lange auf sich warten lassen. Ich
verleiht auf diese Weise dahin kommen, die typische Geographie eines Landes
soweit sie von der Karte abgelesen werden kann, selbst zu machen, und damit
wird auch ungleich grössere Nutzen bringen, als jedes noch so schöne Aus-

1. Bestimmt die Namen aller Gewässer, die auch von der Karte von Mittel-Europa her schon bekannt sind. Suchst aus den Kreisländerkarten die

Word label - Example

Ground Truth (Regions)

Prediction (Lines)

Prediction (Line coloring by word-level labels)

94

das besetzte Lager des Entsatpheres an und erschot einen glänzenden Sieg. Einige Tage später ergab sich auch die Festung Belgrad. Unter den vielen Siegen, die der Prinz erschoten hat, ist der bei Belgrad der berühmteste. Noch heute singt man das damals entstandene Soldatenlied von „Prinz Eugenius, dem edlen Ritter“, der die „Brücke (über die Donau) schlugen“ ließ, um dem Kaiser „Stadt und Festung Belgrad“ zurückzuerobern; noch heute ertönt beim Marsche der österreichischen Reiterei die gleichfalls altentimliche Weise, nach der das Lied gesungen wurde.

Prinzen
von
Wallenstein
1799

220
Grafen
1799

Infolge dieses glänzenden Sieges mußte die Türkei im Frieden zu Passarowitz (Hr. Passarowitz) neuerdings ausgedehnte Gebiete an Österreich abtreten, nämlich das Banat, einen Teil Bosniens, der später wieder verloren ging und erst in unsern Tagen wieder mit Österreich vereinigt wurde, und sogar große Teile der heutigen Königreiche Serbien und Rumänien. Nach einem ruhmvollen Leben, wie es wenig Sterblichen zu teil geworden ist, starb der Prinz im Jahre 1786, tiefbetrauert von seinem dankbaren Monarchen, aber auch von dem ganzen österreichischen Volke. Kaiser Franz Josef I. hat dem „Mitter ohne Furcht und Tadel“ auf dem Plage vor der Hofburg in Wien (durch den Bildhauer Fernow) ein herrliches Reiterdenkmal errichten lassen.

Nur Wiederholung:

Die pragmatische Sanction. Ausserhalb des habsburgischen Mannstammes. Erhebung der weiblichen Thronfolge durch Kaiser Karl VI. (Pragmatische Sanction). Anerkennung derselben durch die Fürsten Europas.

Maria Theresia. Ihre Erziehung. Vermählung mit dem Herzoge Franz von Lothringen. Thronbesteigung Maria Theresias 1740. Bartenstein.

Angriff des Königs Friedrich II. von Preußen. Die übrigen Feinde Österreichs (Bayern, Frankreich, Spanien). Schlechter Zustand des österreichischen Heeres. Weltmangel.

Maria Theresia in Ungarn. Die ungarische Krönungsfeier. Der Reichstag zu Pressburg. Siege der Österreicher über die Bayern. Nidung der Franzosen aus Böhmen. Abtretung Schleiens. Kaiserwahl und Krönung Franz I.

21. Der siebenjährige Krieg.

Kaiserin Maria Theresia vor dem Kriege. Im heldenmüthigen Kampfe um ihr gutes Recht hatte die große Kaiserin Maria Theresia den größten Teil ihrer Erblande gegen die Feinde fast aller europäischen Staaten behauptet; nur ein einziges größeres Land hatte sie ab-

94

das besetzte Lager des Entsatpheres an und erschot einen glänzenden Sieg. Einige Tage später ergab sich auch die Festung Belgrad. Unter den vielen Siegen, die der Prinz erschoten hat, ist der bei Belgrad der berühmteste. Noch heute singt man das damals entstandene Soldatenlied von „Prinz Eugenius, dem edlen Ritter“, der die „Brücke (über die Donau) schlugen“ ließ, um dem Kaiser „Stadt und Festung Belgrad“ zurückzuerobern; noch heute ertönt beim Marsche der österreichischen Reiterei die gleichfalls altentimliche Weise, nach der das Lied gesungen wurde.

Prinzen
von
Wallenstein
1799

220
Grafen
1799

Infolge dieses glänzenden Sieges mußte die Türkei im Frieden zu Passarowitz (Hr. Passarowitz) neuerdings ausgedehnte Gebiete an Österreich abtreten, nämlich das Banat, einen Teil Bosniens, der später wieder verloren ging und erst in unsern Tagen wieder mit Österreich vereinigt wurde, und sogar große Teile der heutigen Königreiche Serbien und Rumänien. Nach einem ruhmvollen Leben, wie es wenig Sterblichen zu teil geworden ist, starb der Prinz im Jahre 1786, tiefbetrauert von seinem dankbaren Monarchen, aber auch von dem ganzen österreichischen Volke. Kaiser Franz Josef I. hat dem „Mitter ohne Furcht und Tadel“ auf dem Plage vor der Hofburg in Wien (durch den Bildhauer Fernow) ein herrliches Reiterdenkmal errichten lassen.

Nur Wiederholung:

Die pragmatische Sanction. Ausserhalb des habsburgischen Mannstammes. Erhebung der weiblichen Thronfolge durch Kaiser Karl VI. (Pragmatische Sanction). Anerkennung derselben durch die Fürsten Europas.

Maria Theresia. Ihre Erziehung. Vermählung mit dem Herzoge Franz von Lothringen. Thronbesteigung Maria Theresias 1740. Bartenstein.

Angriff des Königs Friedrich II. von Preußen. Die übrigen Feinde Österreichs (Bayern, Frankreich, Spanien). Schlechter Zustand des österreichischen Heeres. Weltmangel.

Maria Theresia in Ungarn. Die ungarische Krönungsfeier. Der Reichstag zu Pressburg. Siege der Österreicher über die Bayern. Nidung der Franzosen aus Böhmen. Abtretung Schleiens. Kaiserwahl und Krönung Franz I.

21. Der siebenjährige Krieg.

Kaiserin Maria Theresia vor dem Kriege. Im heldenmüthigen Kampfe um ihr gutes Recht hatte die große Kaiserin Maria Theresia den größten Teil ihrer Erblande gegen die Feinde fast aller europäischen Staaten behauptet; nur ein einziges größeres Land hatte sie ab-

94

das besetzte Lager des Entsatpheres an und erschot einen glänzenden Sieg. Einige Tage später ergab sich auch die Festung Belgrad. Unter den vielen Siegen, die der Prinz erschoten hat, ist der bei Belgrad der berühmteste. Noch heute singt man das damals entstandene Soldatenlied von „Prinz Eugenius, dem edlen Ritter“, der die „Brücke (über die Donau) schlugen“ ließ, um dem Kaiser „Stadt und Festung Belgrad“ zurückzuerobern; noch heute ertönt beim Marsche der österreichischen Reiterei die gleichfalls altentimliche Weise, nach der das Lied gesungen wurde.

Prinzen
von
Wallenstein
1799

220
Grafen
1799

Infolge dieses glänzenden Sieges mußte die Türkei im Frieden zu Passarowitz (Hr. Passarowitz) neuerdings ausgedehnte Gebiete an Österreich abtreten, nämlich das Banat, einen Teil Bosniens, der später wieder verloren ging und erst in unsern Tagen wieder mit Österreich vereinigt wurde, und sogar große Teile der heutigen Königreiche Serbien und Rumänien. Nach einem ruhmvollen Leben, wie es wenig Sterblichen zu teil geworden ist, starb der Prinz im Jahre 1786, tiefbetrauert von seinem dankbaren Monarchen, aber auch von dem ganzen österreichischen Volke. Kaiser Franz Josef I. hat dem „Mitter ohne Furcht und Tadel“ auf dem Plage vor der Hofburg in Wien (durch den Bildhauer Fernow) ein herrliches Reiterdenkmal errichten lassen.

Nur Wiederholung:

Die pragmatische Sanction. Ausserhalb des habsburgischen Mannstammes. Erhebung der weiblichen Thronfolge durch Kaiser Karl VI. (Pragmatische Sanction). Anerkennung derselben durch die Fürsten Europas.

Maria Theresia. Ihre Erziehung. Vermählung mit dem Herzoge Franz von Lothringen. Thronbesteigung Maria Theresias 1740. Bartenstein.

Angriff des Königs Friedrich II. von Preußen. Die übrigen Feinde Österreichs (Bayern, Frankreich, Spanien). Schlechter Zustand des österreichischen Heeres. Weltmangel.

Maria Theresia in Ungarn. Die ungarische Krönungsfeier. Der Reichstag zu Pressburg. Siege der Österreicher über die Bayern. Nidung der Franzosen aus Böhmen. Abtretung Schleiens. Kaiserwahl und Krönung Franz I.

21. Der siebenjährige Krieg.

Kaiserin Maria Theresia vor dem Kriege. Im heldenmüthigen Kampfe um ihr gutes Recht hatte die große Kaiserin Maria Theresia den größten Teil ihrer Erblande gegen die Feinde fast aller europäischen Staaten behauptet; nur ein einziges größeres Land hatte sie ab-

Tokenization and Model Training

Tokenization

- Use a pre-trained transformer tokenizer (e.g., DistilBERT for German).
- Each line is represented as a sequence of words and their assigned labels.
- For tokenization, each word is split into one or more sub-tokens as required by the tokenizer.
- Only the first token of each word is assigned the word's label; all subsequent sub-tokens are assigned a special label (-100) so they are ignored during model training.
- This approach maintains word-level supervision even when words are split into sub-tokens, allowing the model to learn from the original labeled data.

Tokenization and Model Training

Model Training:

- Fine-tune a pre-trained transformer (e.g., DistilBERT) for token classification (labels: correct/error)
- **Input:** Tokenized sequences of words per line, with aligned labels.
- **Output:** The model predicts the label (correct/error) for each word in the input sequence.
- **Training parameters:**
 - Learning rate: $2e-5$
 - Batch size: 8 (train & eval)
 - Epochs: 10
 - Weight decay: 0.01

Tokenization and Model Training

Model Training:

- **Class Imbalance Handling:**

- Compute class weights inversely proportional to the frequency of each class (correct/error).
- Use a custom training loop that applies these weights in the loss function to handle class imbalance and improve error detection.

- **Custom Trainer**

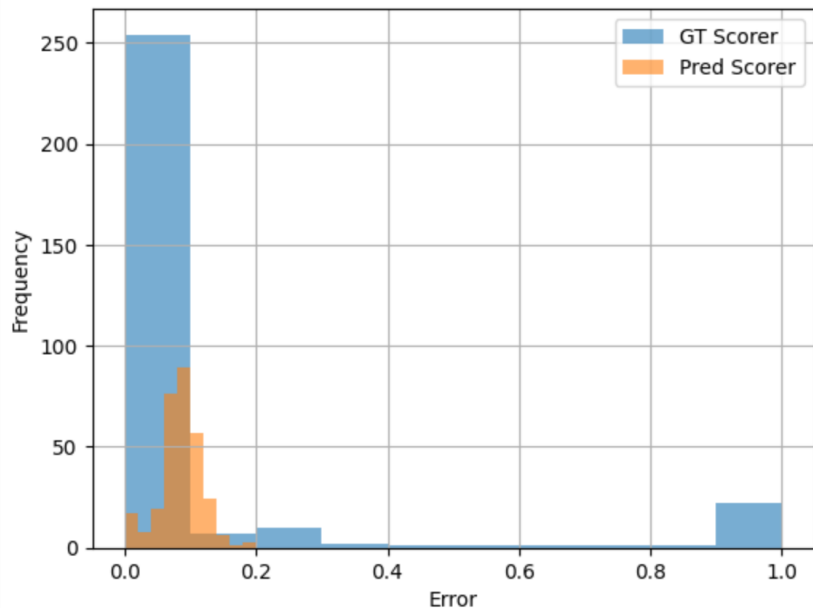
- Use a custom Trainer that applies weighted cross-entropy loss, giving more importance to rare classes (like errors).
- During loss calculation, we ignore all sub-tokens (i.e., tokens with label -100), so only the main word tokens contribute to model training.

Evaluation

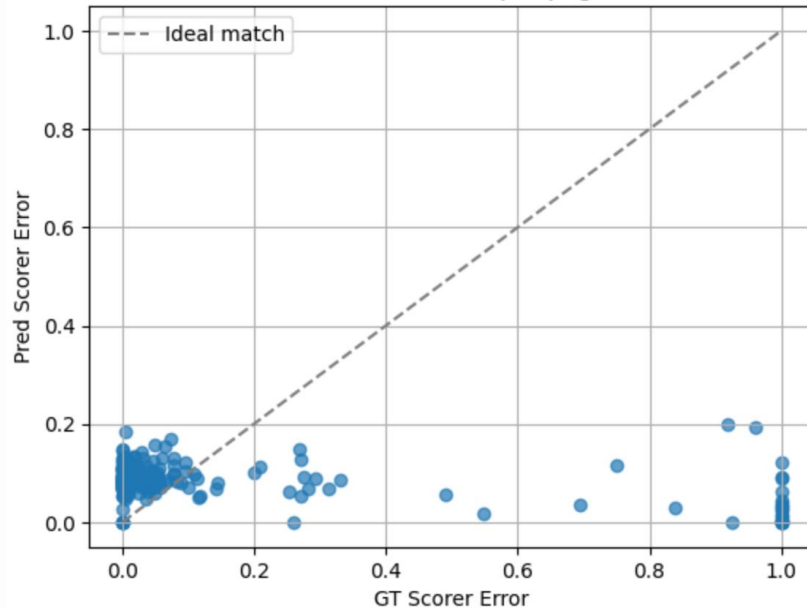
Metric	GT Scorer	Pred Scorer	
+=====+	+=====+	+=====+	+=====+
Count	300	300.0000	
+-----+	+-----+	+-----+	+-----+
Mean	0.1078	0.0850	
+-----+	+-----+	+-----+	+-----+
Median	0.0048	0.0859	
+-----+	+-----+	+-----+	+-----+
Stdev	0.2685	0.0328	
+-----+	+-----+	+-----+	+-----+
Min	0	0.0000	
+-----+	+-----+	+-----+	+-----+
Max	1	0.1991	
+-----+	+-----+	+-----+	+-----+
MAE	0.1078	0.0850	
+-----+	+-----+	+-----+	+-----+
Correlation	-0.3678	-	
+-----+	+-----+	+-----+	+-----+
GT vs Pred RMSE	0.2827	-	
+-----+	+-----+	+-----+	+-----+

Evaluation

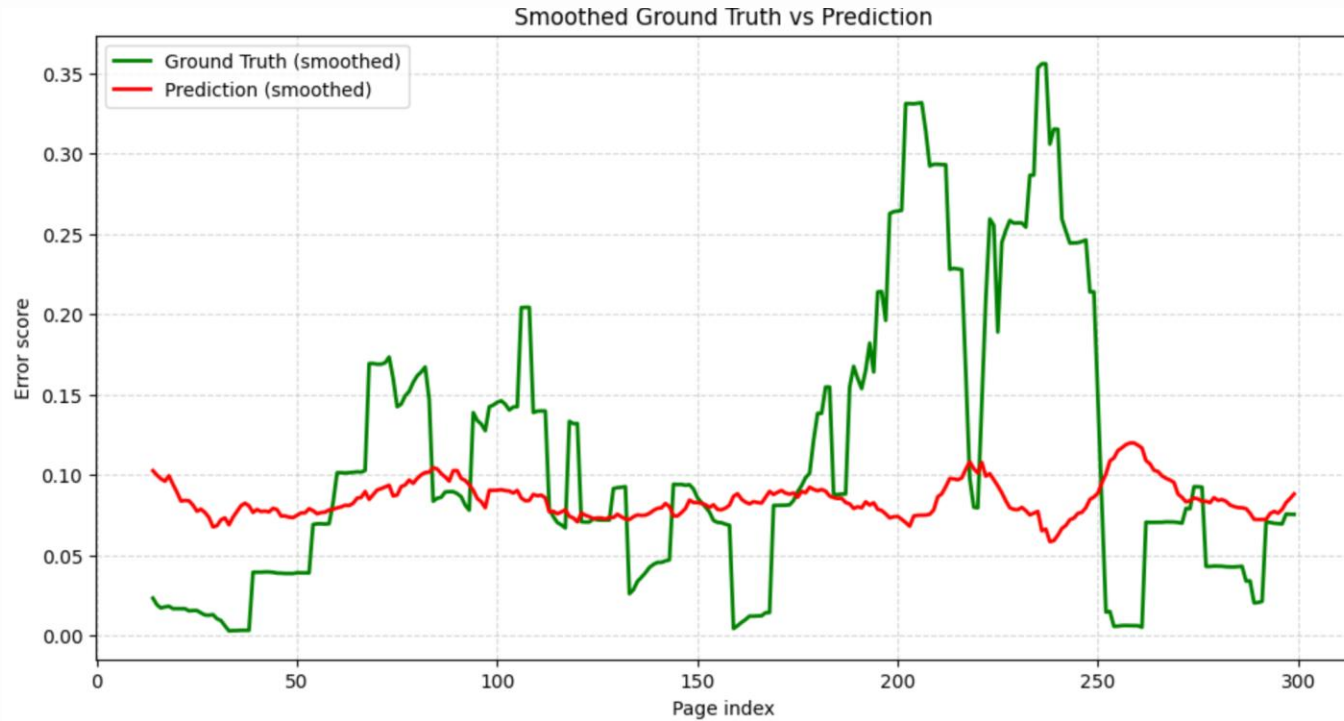
Error Distributions



GT vs Pred Error (per page)



Evaluation



Evaluation

--- top-30 GT pages by relative error ---

Rank	Page	RelErr
1	1737364816_00000001	1.0000
2	1737377780_00000004	1.0000
3	1737379171_00000132	1.0000
4	1737379627_00000003	1.0000
5	1737386917_00000003	1.0000
6	1737389673_00000080	1.0000
7	1737393700_00000004	1.0000
8	1738471039_00000010	1.0000
9	173857282X_00000390	1.0000
10	1738584399_00000150	1.0000
11	1738584852_00000167	1.0000
12	1739039688_00000224	1.0000
13	1740257367_00000080	1.0000
14	1740257367_00000087	1.0000
15	1740384288_00000002	1.0000
16	1740388542_00000002	1.0000
17	1740391969_00000066	1.0000
18	1740391969_00000080	1.0000
19	1740399145_00000435	1.0000
20	1740399145_00000436	1.0000
21	1740399404_00000002	1.0000
22	1740399404_00000010	1.0000
23	1740401204_00000020	1.0000
24	1740402545_00000002	1.0000
25	1740402723_00000010	1.0000
26	1740404564_00000001	1.0000
27	1740406060_00000003	1.0000
28	1740826949_00000062	1.0000
29	1741190975_00000026	1.0000
30	1741190975_00000043	1.0000

--- top-30 Predicted pages by relative error ---

Rank	Page	RelErr
1	1741701600_00000024	0.4360
2	174170569X_00000012	0.2914
3	1744303657_00000355	0.2758
4	1740392299_00000023	0.2722
5	1744303657_00000369	0.2681
6	1740404564_00000112	0.2066
7	1741193575_00000027	0.2000
8	1737386917_00000328	0.1991
9	1739039688_00000203	0.1923
10	1740405188_00000036	0.1867
11	1740379381_00000113	0.1855
12	1740399145_00000128	0.1813
13	1740400666_00000080	0.1727
14	1740395999_00000052	0.1719
15	1737381346_00000170	0.1687
16	1740392299_00000018	0.1671
17	1741701600_00000061	0.1663
18	1740405188_00000011	0.1604
19	173736378X_00000120	0.1560
20	1741704316_00000001	0.1552
21	1737381346_00000249	0.1537
22	1740391551_00000049	0.1533
23	174429397X_00000011	0.1528
24	1739037960_00000033	0.1493
25	1737359839_00000247	0.1489
26	174170569X_00000008	0.1469
27	1744303878_00000108	0.1467
28	1740388542_00000038	0.1461
29	1744303878_00000068	0.1454
30	1740404947_00000238	0.1451

Evaluation

--- page-level error statistics ---

GT pages : $\mu = 0.2443$

Pred pages : $\mu = 0.0811$

RMSE : 0.4507