



# OCR-BW

*Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen*



## Altbestandserschließung mit KI-Anwendungen

-

## ein Werkstattbericht

05.06.2024, Kristina Stöbener und Dorothee Huff



---

# Agenda

1. Hintergrund
2. Erschließung mit ChatGPT
  - Inhaltsangabe
  - Korrektur von OCR-Daten
  - Sacherschließung
3. Grenzen
4. Fazit



---

## Hintergrund Projekt OCR-BW (2019-2022)

- Ziel: Aufbau eines Kompetenzzentrums für Volltexterschließung von handschriftlichen und gedruckten Werken mit der UB Mannheim
- Aufgabe: Evaluierung von Transkribus anhand verschiedener Textkorpora aus dem Bestand der UB Tübingen (Erstellung von Ground-Truth-Daten und Modelltraining mit einer angestrebten Zeichenfehlerrate (CER)  $< 5\%$ )
- ursprüngliche Überlegung: durchsuchbare Volltexte können als Ersatz für eine ressourcenaufwändige Tiefenerschließung dienen



- 
- Frage: können die erzeugten automatischen Transkriptionen für Handschriften und historische Druckwerke mit Hilfe weiterer KI-Anwendungen für die Tiefenerschließung und Aufwertung der Daten nachgenutzt werden?
  - Stand: automatisch erzeugte Transkriptionen mit Fehlerraten im Bereich von durchschnittlich 2-5%
    - Ergebnis nicht fehlerfrei, was den Erfolg einer Volltextsuche beeinträchtigt



---

## Ausgangspunkt schwach strukturiertes Material

- historischen Dokumenten fehlen oftmals Inhaltsverzeichnis, Inhaltsangabe, Register oder Index, die einen schnellen Einstieg in das Dokument ermöglichen
- Beispiele für Anwendungsfälle: Tagebücher, Korrespondenzen, Protokolle, Konsilien, Akten
- Volltexte erleichtern den Zugang zu historischen Dokumenten, aber man muss bereits wissen, wonach man sucht bzw. auf gut Glück suchen
  - Lösung?: Nachnutzung der Volltexte mit weiteren KI-Anwendungen wie ChatGPT, um mit geringem Ressourcenaufwand eine Tiefenerschließung zu ermöglichen



---

## Erschließung mit ChatGPT

- Ermittlung von Erschließungsdaten
  - Inhaltsangabe
  - Schlagworte
- Grundlage: Transkriptionsdaten (Ground-Truth-Daten im Vergleich zu automatisch erzeugten Transkriptionen)
- Ziel: Erschließung soll auswertungsoffen sein



## Expeditions-Tagebuch Edwin Hennig (UAT 407/80)

- 171 Seiten
- 35.334 Wörter
- 233.795 Zeichen



- 74.713 Tokens (berechnet mit dem OpenAI-Tokenizer für GPT-3.5 und GPT-4)
- ChatGPT gibt als Verarbeitungsgrenze 8192 Tokens an (1.2x weniger Tokens mit dem hier verwendeten ChatGPT 4o)

Dr. Edw. Hennig  
Berlin N. 4. Invalidenstr. 43.



## Inhaltsangabe GT-Daten

*Gib mir bitte eine präzise und sachliche Zusammenfassung des Inhalts.*

Der Bericht dokumentiert die Tendaguru-Expedition vom 11. März 1909 bis 29. April 1910, die von Dr. Edwin Hennig und Dr. Janensch geleitet wurde. Die Expedition begann in Berlin und führte die Forscher über verschiedene europäische Städte nach Afrika, wo sie in Dar-es-Salaam, Lindi und schließlich im Tendaguru-Gebiet ankamen.

Die Hauptaktivität der Expedition bestand in der Ausgrabung und Dokumentation von Dinosaurierfossilien. Der Tendaguru-Hügel erwies sich als äußerst reich an Fossilien, insbesondere von Sauriern. Mehrere Fundstellen wurden entdeckt, an denen eine Vielzahl von Knochen und anderen Fossilien freigelegt wurden. Zu den bemerkenswerten Funden gehörten gut erhaltene Beinknochen, Wirbel und Schädelreste.

Die Tagebücher beschreiben detailliert die geologischen und paläontologischen Untersuchungen, die täglichen Arbeiten und Herausforderungen, wie die Organisation der Grabungen, die Versorgung der Arbeiter, klimatische Bedingungen und gesundheitliche Probleme. Trotz der widrigen Umstände, einschließlich heftiger Regenfälle und gesundheitlicher Beschwerden, konnten die Forscher bedeutende wissenschaftliche Erfolge erzielen.

Neben der wissenschaftlichen Arbeit enthält das Tagebuch auch Berichte über Begegnungen mit der lokalen Bevölkerung und deren Kultur sowie Jagdausflüge und Beobachtungen der Flora und Fauna der Region.

Die Expedition war ein bedeutender Beitrag zur Paläontologie und zur Kenntnis der Dinosaurierfauna Ostafrikas. (189 Wörter)



## Inhaltsangabe GT-Daten

*Reduziere die Inhaltsangabe bitte um die Hälfte.*

Die Tendaguru-Expedition (11. März 1909 - 29. April 1910) unter der Leitung von Dr. Edwin Hennig und Dr. Janensch begann in Berlin und führte über Europa nach Afrika, wo sie in Dar-es-Salaam und Lindi ankamen.

Hauptziel war die Ausgrabung von Dinosaurierfossilien im Tendaguru-Gebiet.

Die Expedition entdeckte zahlreiche Fundstellen mit gut erhaltenen Saurierknochen, darunter Beckenknochen, Wirbel und Schädel.

Die Tagebücher dokumentieren die geologischen und paläontologischen Arbeiten, die Organisation der Grabungen, die Versorgung der Arbeiter und die klimatischen Herausforderungen. Trotz widriger Bedingungen wie starkem Regen und gesundheitlichen Problemen erzielten die Forscher bedeutende wissenschaftliche Erfolge.

Zusätzlich berichten die Tagebücher über Begegnungen mit der lokalen Bevölkerung, Jagdausflüge und Beobachtungen der regionalen Flora und Fauna. Die Expedition trug wesentlich zur Paläontologie und zum Verständnis der Dinosaurierfauna in Ostafrika bei. (124 Wörter)



---

## Inhaltsangabe GT-Daten

### *Manuelle Modifikation*

Die Tendaguru-Expedition (11. März 1909 - 29. April 1910) unter der Leitung von Dr. Edwin Hennig und Dr. Werner Janensch begann in Berlin und führte über Europa nach Afrika, wo sie in Daressalam und Lindi ankamen.

Hauptziel war die Ausgrabung von Dinosaurierfossilien im Tendaguru-Gebiet. Die Expedition entdeckte zahlreiche Fundstellen mit Saurierknochen und anderen Fossilien. Die Tagebücher dokumentieren die geologischen und paläontologischen Arbeiten, die Organisation der Grabungen, die Versorgung der Arbeiter und die klimatischen Herausforderungen. Zusätzlich berichten die Tagebücher über Begegnungen mit der lokalen Bevölkerung, Jagdausflüge und Beobachtungen der regionalen Flora und Fauna. Die Expedition trug wesentlich zur Paläontologie und zum Verständnis der Dinosaurierfauna in Ostafrika bei. (106 Wörter)



## Inhaltsangabe Modell „German Giant“ (4,30% CER)

Das Dokument beschreibt die Tendaguru-Expedition, die vom 11. März 1909 bis zum 4. Juli 1910 unter der Leitung von Dr. Eow. Renni stattfand. Es handelt sich um ein detailliertes Tagebuch der Expedition, das die Reise von Berlin nach Ostafrika beschreibt, einschließlich Natur- und Wetterbeobachtungen, geologischer und paläontologischer Funde, insbesondere Dinosaurierfossilien am Tendaguru. Das Team interagiert mit Einheimischen und bewältigt Herausforderungen wie Krankheiten und Logistikprobleme. Die tägliche Routine umfasst wissenschaftliche Arbeiten, Jagdausflüge und Lagerleben. Die Expedition liefert wertvolle geologische und paläontologische Erkenntnisse und dokumentiert sorgfältig ihre Entdeckungen und Erlebnisse.



---

## Inhaltsangabe „German Giant“ (korrigiert von ChatGPT)

Das Dokument beschreibt die Tendaguru-Expedition von 1909-1910 unter der Leitung von Dr. Eow. Renni. Es handelt sich um ein Tagebuch, das die Reise von Berlin nach Ostafrika detailliert darstellt. Beobachtungen umfassen Flora, Fauna, Wetterbedingungen und geologische Merkmale. Besondere Aufmerksamkeit gilt den Dinosaurierfossilien am Tendaguru, die sorgfältig dokumentiert und analysiert werden. Das Team interagiert mit Einheimischen, erlebt lokale Kultur und bewältigt Herausforderungen wie Krankheiten und Logistikprobleme. Die tägliche Routine umfasst wissenschaftliche Arbeiten, Jagdausflüge und Lagerleben. Die Expedition legt den Grundstein für wichtige geologische und paläontologische Erkenntnisse.



## Inhaltsangabe Modell „German\_Kurrent\_XIX\_pylaia“ (14,07% CER)

Das Dokument enthält eine Reihe von Tagebucheinträgen und Berichten über verschiedene Ereignisse und Beobachtungen, verfasst von Dr. Janensch. Diese Einträge dokumentieren das tägliche Leben, Wetterbedingungen, Begegnungen mit der Natur und Tieren sowie persönliche Reflexionen und Beschreibungen von Krankheiten und deren Auswirkungen während einer Expedition im Jahr 1926. Es gibt auch detaillierte Beschreibungen von Fossilienfunden und anderen wissenschaftlichen Beobachtungen. Einige Passagen reflektieren die Beziehung des Autors zu anderen Personen und deren Charaktereigenschaften. Die Texte bieten einen Einblick in die Herausforderungen und Erlebnisse des Autors in einem vermutlich kolonialen Kontext, einschließlich Jagderlebnissen und dem Umgang mit der einheimischen Bevölkerung.



## Exkurs: Korrektur von OCR durch ChatGPT I

Der Akide hat sein sehr stattliches, mit breiter, erhöhter Barasa  
versehenes Haus neuerdings auf eine kleine Erhöhung verlegt,  
man hat nun in seinem Schatten sitzend einen hübschen Blick auf  
Namunda u. Namyura; sowie durch

die Bäume hindurch auf den Tendaguoueder Akide hat selbst  
keine Schambe, doch sah ich eine Ganze Schar recht  
gedrungener kräftiger Zingen, die offenbar in guter Vermehrung  
begriffen waren. Ich wurde äußerlich sehr freudlich empfangen u.  
wir scherzten viel miteinander. d. Akide ist ein hübscher,  
selbstbewußter Mann, ob er wollstes Vertrauen verdient, steht  
auf einem andern Blatt. Ich schenkte ihm statt des begehrten  
Antilopenbratens ein Perlhuhe zum Entsetzen meiner Leute, die  
es gern als Zukost gehabt hätten. Hierschien reger Briefverkehr zu  
herrschen, obwohl keine Hoheit selbst nur arabisch lesen  
können, auch stand Tinte u. Feder bereit. Für d. Tier nahm ich 2  
weitere Träger an, die aber den Weg so wenig wußte tem, wie der  
Akide selbst; ein über Nanundo führenden it offenbar  
eingegregen. Nach manchem Ein u. HHer fanden wir den rechten  
über k dope, der mir vom Millbenkern an selbst bekannt war.  
Jenseits schoß ich wiederum zweimal auf ein Mbawala, gab dann  
d. Verfolgung auf, doch war es ebenfalls nicht geflosen. Eiging  
nach u. schoß es mit der Fliete. Nu.d d. Trägers wieder  
ausgesöhnt; ich mußte de dope nach2 Mann zurückschicken.  
Gegen 12 im Lager.

Der Artikel hat sein sehr stattliches, mit breiter erhöhter Barasa  
versehenes Haus neuerdings auf eine kleine Erhöhung verlegt. Man hat nun in seinem  
Schatten sitzend einen hübschen Blick auf Namunda und Namyura sowie durch

hohes Gras hindurch, das nach ihrer Angabe bei der Ernte 6 m Höhe erreichen  
soll. Wie in Wamachinga fiel mir auch hier ein hell gefärbtes Gras auf. Auch  
stand auf dem Gipfel der kleinen Anhöhe eine Gruppe alter, bereits  
getrockneter und abgestorbener, mäßig hoher Büsche wie Orangen, aber an  
den Stacheln erkannte ich die Aloe. Mittagessen sehr gut und die Küche steht in  
einem reizenden Widerspruch zur mangelnden Ausdehnung. Recht gewürztes  
Huhn und Wild, Gemüsesauce, Reis, Hirsebieb und die übrigen Getränke, also  
ganz untropisch europäisch, außer dass ich vor der Hitze während des  
Mittagsschlafs floh. (...) Hielt mich eine Weile fest und wollte mich dann auf den  
Rückweg begeben, als ich aus der Nähe des Baches donnernden Schall hörte. In  
der Ahnung, einen vorzeitlichen Mosasaurus oder dergleichen vor mir zu haben,  
hielt ich mich zurück, bis das Geräusch weiter ertönte und immer deutlicher  
wurde, es hörte sich wie Donnern an, als ein riesiges, mir unbekanntes Tier mit  
Schwanz und Kopf durch den Bach stapfte und mir einige Zeit entgegengestellt  
blieb. Dabei stieß es in den fast schwarzen Wasserschlamm eine Welle von  
blauen Schaumblasen hoch. Schließlich kam ich aus der Deckung hervor,  
zitternd und – das Ungeheuer war ein uralter, halb verwachsener, zitternder  
Affe, der mit ganzer Kraft ein Rohr schlug und der sich verächtlich über mein  
Weglaufen auslachte. Ich empfand eine Mischung von Furcht, Entsetzen, Zorn  
und Scham und hätte das verwünschte Biest am liebsten erschlagen, wenn nicht  
Ali dazu gekommen wäre, der offenbar viel mehr Nerven hatte und mich über  
meine Torheit belächelte.



## Exkurs: Korrektur von OCR durch ChatGPT II

8

mit Tongesteinen gleicher Farbe fortwährend zu wechseln oder zu solchen zu verwittere. Wegen starker Löwengefahr Gewehre geladen u. Falle gestellt.

14. IV. 36 24 Marsch auf d. fast ebenen Noto-Plateau bis zum Yambenst. Lager). Auf d. Wege sehr viel Spuren, darunter Leopard u. kl. Wildkatzen, sonst Antilopem. Kurz vor d. Junken stärkerer Anstieg. Nachm. Schießübungen nach ~~FlaSchichten~~ Flaschen. Riege kämpfe d. Schwarzen D. Leute essen nur einmal am Tage ordentlich Mtama, sind imd. Mehrzahl stets vergnügt, siegen unterwegs u. gehen mit d. schweren Lasten (auf Schulter od. Kropf) so schnell wie wir. Abends kühl.

15. IV. Abstieg vom Noto. Schöner Blick auf d. Binnenland. ~~DeutSchichten~~ Deutschen Mittelgebirgscharakter Vegetation für geolog. Beobachtungen zu stark, Abweichen vom Wege kaum möglich. Fünftes Lager um 7t vorm. im Dorf kwakueihn. Die groben Konglomerate schienen sich aus d. Landschaft herauszuheben u. außerdem din allettwärts vnebruiteten Guwölle von Quarz u. kistallinem Materdal zu liefern.



## Exkurs: Korrektur von OCR durch ChatGPT III

8

mit Tongesteinen gleicher Farbe fortwährend zu wechseln oder zu solchen zu ~~verwittere~~-~~verwittern~~. Wegen starker Löwengefahr Gewehre geladen u. ~~Falle~~ und ~~Fallen~~ gestellt.

14. IV. ~~36-24~~ Marsch auf ~~d.~~-~~dem~~ fast ebenen Noto-Plateau bis zum Yambenst. ~~Lager~~-Lager. Auf ~~d.~~-~~dem~~ Wege sehr ~~viel~~-~~viele~~ Spuren, darunter Leopard u. ~~kl.~~ und kleine Wildkatzen, sonst ~~Antilopen~~-~~Antilopen~~. Kurz vor ~~d.~~-~~dem~~ Junken ~~stärkerer~~-~~stärkerer~~ Anstieg. Nachm. Schießübungen nach Flaschen. ~~Riege~~-~~Ring~~-~~kämpfe~~ ~~d.~~-~~Schwarzen~~-~~D.~~-~~der~~ Schwarzen. Die Leute essen nur einmal am ~~Tage~~-~~Tag~~ ordentlich Mtama, sind ~~imd.~~-~~in~~ der Mehrzahl stets vergnügt, ~~siegen~~-~~singen~~ ~~unterwegs~~ u. ~~unterwegs~~ und gehen mit ~~d.~~-~~den~~ schweren Lasten (auf Schulter od. ~~Kropf~~) oder Kopf) so schnell wie wir. Abends kühl.

15. IV. ~~Abstieg~~-~~IV.~~ Abstieg vom Noto. Schöner Blick auf ~~d.~~-~~das~~ Binnenland. ~~Deutschen~~-~~Deutscher~~ ~~Mittelgebirgscharakter~~-~~Mittelgebirgscharakter~~. Vegetation für ~~geolog.~~-~~geologische~~ Beobachtungen zu stark, Abweichen vom ~~Wege~~-~~Weg~~ kaum möglich. Fünftes Lager um ~~7~~ ~~vorm.~~-~~7~~ Uhr vormittags im Dorf ~~kwakueihn~~-~~Kwakueihn~~. Die groben Konglomerate schienen sich aus ~~d.~~-~~der~~ Landschaft herauszuheben u. ~~und~~ außerdem ~~in~~ ~~alle~~ ~~Richtungen~~ ~~verbreitetes~~ ~~Gerölle~~ von Quarz u. ~~kristallinem~~ ~~Materdal~~ und kristallinem Material zu liefern.

16. IV. IV. Marsch zum ~~Tendagurn~~-~~J.~~ 104. Tendaguru. Viel Dickicht, ~~Bamber~~-~~Bambus~~-~~gestrüpp~~ u. ~~und~~ rechten Urwald, sehr interessant, aber ~~f.~~ für Träger ~~unbequamer~~-~~unbequemer~~ als ~~d.~~-~~die~~ parkähnliche Barra-Barra auf dem Noto. 4 ~~Std.~~-~~Stunden~~ vom Lager Korallen am Wege, gleich darauf guter, wichtiger ~~Aufschluß~~-~~Aufschluss~~ am Matarawe-Bach: Solith unterlagert von dunklem ~~ton~~ (~~Grundwasserspiegel~~ ~~Plares~~ ~~Wasse~~ ~~Ton~~ (~~Grundwasserspiegel~~)). Stankenweis durch das Wasser des Mtshingivi, kurz vorher molluskenreicher Kalk (~~Trigenien~~ ~~Schicht?~~) (~~Trigonia~~-~~Schicht?~~) 4 ~~Std.~~-~~Stunden~~ vor ~~d.~~-~~Tager~~, ~~dem~~ Lagerplatz ~~platz~~ erste ~~Saurierkroehen~~-~~Saurierknochen~~ am ~~Wege~~-~~s.~~ ~~Wege~~, 1 ~~Std.~~-~~Stunde~~ vorher ~~Abstieg~~-~~Abstieg~~ mit prächtigem Blick auf ~~d.~~-~~die~~ Teselburg ~~Tendagurn~~ u. ~~d.~~-~~Tendaguru~~ und die ~~höhen~~-~~Höhen~~ von Namwivange.



---

# Inhaltsangabe Juristische Konsilien (UAT 84/205)

## Unterschlagung des Hillerischen Stiftungskapitals

- **Hintergrund:** Veruntreuung des Stiftungskapitals durch den Consistorial-Secretarius Weickersreuter.
- **Entscheidung:** Weickersreuter wurde zur Rückzahlung des unterschlagenen Kapitals verpflichtet. Die Verwaltung der Stiftung wurde neu geregelt, um weitere Unterschlagungen zu verhindern.

## Appellationssache Jacob Jetter

- **Hintergrund:** Schadensersatzforderung gegen den Chirurgen Conrad Schulz wegen eines schlecht behandelten Fußes.
- **Entscheidung:** Schulz wurde zur Zahlung von Schadensersatz verurteilt, da die Behandlung als grob fahrlässig eingestuft wurde.



---

## Sacherschließung: Beispiel Personen

- 1. Schritt: Ermittlung von Personennamen (die fünf am häufigsten genannten)
  - Zuverlässigkeit der Findung von Namen als Qualitätsmarker
  - Grundlage: manuell getaggttes Dokument
  - Test anhand verschiedener CER-Stufen und von ChatGPT korrigierten automatischen Transkriptionsdaten
- 2. Schritt: Identifizierung der Personen
  - Vervollständigung der Namen
  - Lebensdaten
  - GND-Nummern



---

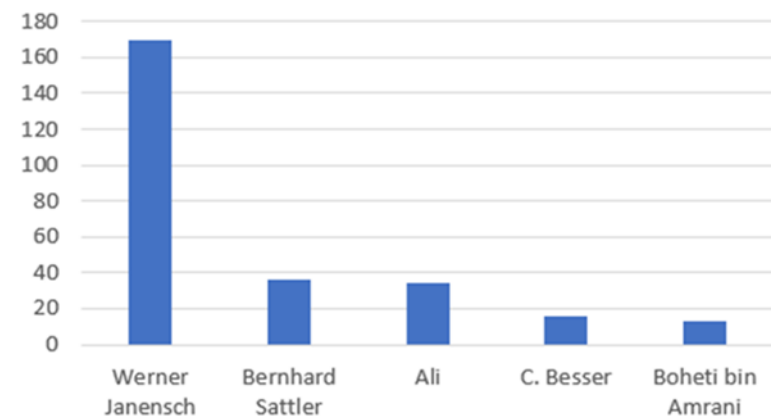
## Ergebnis manuelles Tagging

1. Werner Janensch (J, J., Jan., Janensch, Dr. J, Dr. J., Dr. Janensch)
  - Expeditionsleiter, Paläontologe → 170 Nennungen
2. Bernhard Sattler (Sattler, Hr./Herr Sattler)
  - Ingenieur, Vertreter der DOAG → 36 Nennungen
3. Ali (Ali, Boy Ali)
  - Boy (persönlicher Diener) → 34 Nennungen
4. C. Besser (Hr./Herr Besser)
  - Vertreter der DOAG → 16 Nennungen
5. Boheti bin Amrani (Boheti)
  - Aufseher und Präparator → 13 Nennungen



## Ergebnis GT-Daten

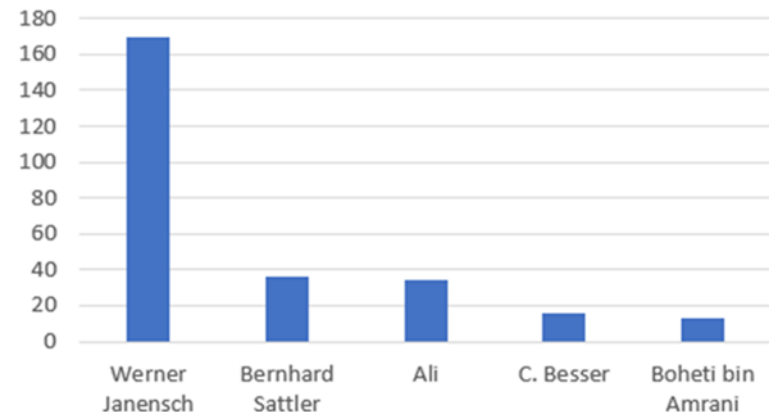
1. Dr. Janensch → 74 Nennungen (-96, aber nur 73 Nennungen der Instanz „Janensch“)
  2. Ali → 35 Nennungen (+1)
  3. Boheti → 8 Nennungen (-5)
  4. Bornhardt → 6 Nennungen (5 Nennungen, getaggt als „Literatur“)
  5. Sattler → 4 Nennungen (-32)
- C. Besser wird nicht gefunden





## Ergebnis „M6“ (2,43% CER)

1. Dr. Janensch → 30 Nennungen (-140)
2. Herr Sattler → 13 Nennungen (-23)
3. Herr Besser → 8 Nennungen (-8)
4. Boheti → 6 Nennungen (-7)
5. Salim → 5 Nennungen (+3)
6. Ali → 5 Nennungen (-29)

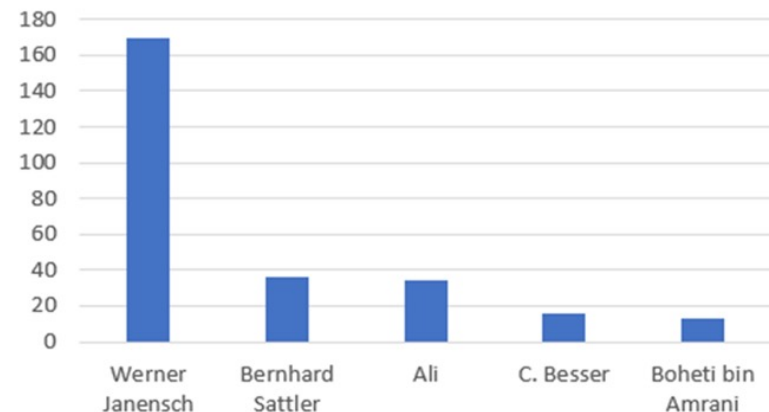




## Ergebnis „German Giant“ (4,30% CER)

1. Herrn Besser - 10 Nennungen (-6)
2. Dr. Janensch - 8 Nennungen (-162)
3. Herr Schwartz - 5 Nennungen (+/- 0)
4. Dr. Eow. Renni - 4 Nennungen (+3)
5. Herr Jakler - 3 Nennungen (gemeint Sattler? -33)

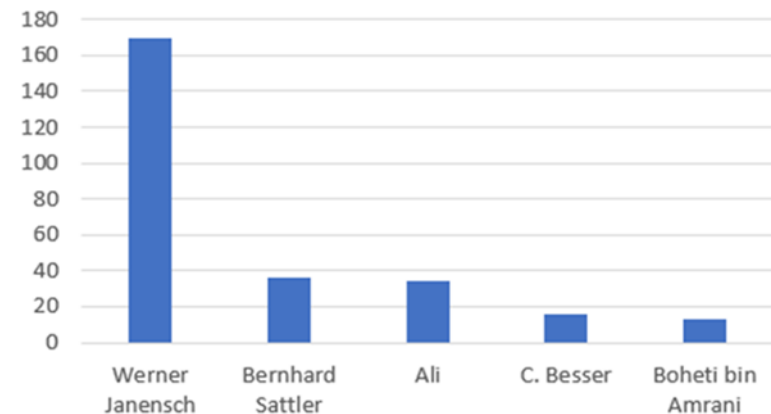
- Ali und Boheti bin Amrani werden nicht gefunden





## Ergebnis „German Giant“ tokenisiert

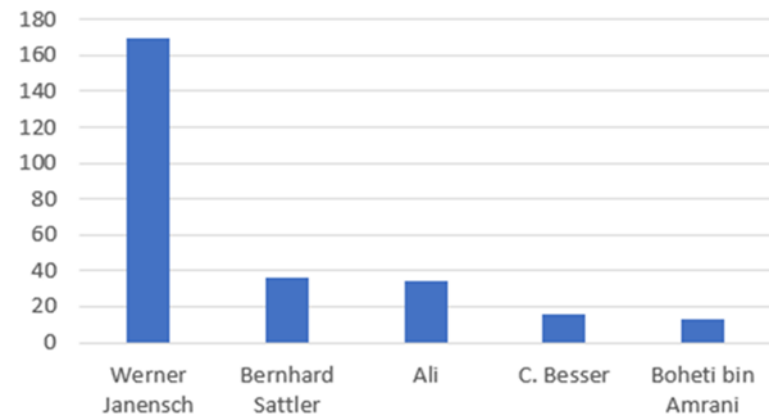
1. Dr. Janensch → 124 Nennungen (-46)
2. Herr Sattler → 24 Nennungen (-12)
3. Ali → 22 Nennungen (-12)
4. Herr Besser → 11 Nennungen (-5)
5. Boheti → 1 Nennungen (-12)





## Ergebnis „German Giant“ korrigiert

1. Dr. Janensch → 203 Nennungen (+33)
2. Herr Sattler → 29 Nennungen (-7)
3. Boheti → 12 Nennungen (-1)
4. Herr Kaiser → 7 Nennungen (+6, ein Hr. Kaiser wird einmal erwähnt, ansonsten fällt der Begriff noch zweimal)
5. Dr. Renni → 1 Nennung (+/-0 für Hennig)





Welche Personen werden in den Dokumenten genannt? Gib mir mit eine Liste absteigend nach Erwähnung mit der Anzahl der Nennungen aus und fasse verschiedene Versionen eines Namens zusammen.

< 2/2 >

Hier ist eine Liste der im Dokument erwähnten Personen, absteigend nach der Anzahl ihrer Nennungen, zusammengefasst in verschiedene Versionen eines Namens:

1. **Dr. Janensch** - 14 Nennungen
2. **Sattler (Herrn Sattler)** - 12 Nennungen
3. **Wertmann (Herrn Wertmann)** - 3 Nennungen
4. **Rechenberg (Gouverneur v. Rechenberg)** - 2 Nennungen
5. **Dr. Eow. Renni** - 1 Nennung



Welche Personen werden in den Dokumenten genannt? Gib mir eine Liste absteigend nach Erwähnung mit der Anzahl der Nennungen aus und fasse verschiedene Versionen eines Namens zusammen.

Namen Nach Erwähnung Sortiert			
	Name	Anzahl der Nennungen	
1	Ich	20	
2	Dez	19	
3	Janensch	12	
4	Knall	9	
5	Herde	8	
6	Tal	7	



---

## Orte nach Häufigkeit sortiert

In dem Dokument sind folgende Orte nach der Häufigkeit ihrer Erwähnungen sortiert:

1. **Tendaguru** - Der Ort, an dem die Expedition hauptsächlich stattfand, wird oft erwähnt, insbesondere im Zusammenhang mit den Ausgrabungen von Dinosaurierfossilien.
2. **Berlin** - Als Arbeitsort von Werner Janensch und als Standort des Museums für Naturkunde, das die Expedition organisierte, wird Berlin häufig genannt.
3. **Dar-es-Salaam** und **Zanzibar** - Diese Orte werden im Zusammenhang mit der Route der Expedition und als wichtige geografische Punkte in Ostafrika erwähnt.
4. **Lindi** - Wird mehrfach im Kontext der Reiseroute und lokaler Aktivitäten während der Expedition genannt.
5. **Aden** - Wird im Kontext der Reiseroute und geografischen Beobachtungen erwähnt.



# Dokumentstruktur

Der Textausschnitt umfasst mehrere Tagebuchdaten, die die Chronologie der Ereignisse während der Expedition dokumentieren. Hier sind die Daten, die in den bereitgestellten Teilen des Dokuments erwähnt werden:

1. 11. März 1909
2. 13. März 1909
3. 15. März 1909
4. 20. März 1909
5. 22. März 1909
6. 23. März 1909
7. 25. März 1909
8. 27. März 1909
9. 28. März 1909
10. 31. März 1909





---

# Organisationen

In dem Textausschnitt werden mehrere Organisationen und institutionelle Strukturen erwähnt, die in Bezug auf die Expedition von Bedeutung sind. Hier sind die hauptsächlich genannten Organisationen:

1. **Naturkundemuseum Berlin** (Museum für Naturkunde) - Die Hauptorganisation hinter der Expedition, die die Forschungsreise finanzierte und organisierte.
2. **Deutsche Ost-Afrika-Gesellschaft (D.O.A.G.)** - Diese Gesellschaft wird erwähnt, insbesondere im Zusammenhang mit der Logistik und Unterstützung der Expedition vor Ort.
3. **Paläontologische Gesellschaft** - Werner Janensch war Mitglied und später Ehrenmitglied dieser wissenschaftlichen Gesellschaft, die sich der Erforschung von Fossilien widmet.



# Schlagworte allgemein

Um die inhaltliche Relevanz des Textes über die Expedition nach Tendaguru zu erfassen, können folgende Schlagworte als besonders relevant betrachtet werden:

1. **Tendaguru-Expedition** - Zentraler Kontext des Dokuments, wichtige paläontologische Expedition in Ostafrika.
2. **Fossilienfund** - Hauptziel der Expedition, umfasst die Entdeckung und Sammlung von Dinosaurierfossilien.
3. **Dr. Werner Janensch** - Leiter der Expedition und eine zentrale Figur in der Paläontologie und Geologie.
4. **Geologische Untersuchungen** - Wissenschaftliche Aktivitäten, die während der Expedition durchgeführt wurden.
5. **Interaktion mit Einheimischen** - Beschreibt die sozialen Begegnungen und kulturellen Erfahrungen der Expeditionsteilnehmer mit der lokalen Bevölkerung.
6. **Jurazeit** - Zeitliche Einordnung der Fossilien, die in der Tendaguru-Formation gefunden wurden.
7. **Deutsches Kolonialreich** - Historischer und politischer Kontext der Expedition, beeinflusst durch die deutsche Kolonialpolitik in Ostafrika.
8. **Naturkundemuseum Berlin** - Institution, die die Expedition finanzierte und organisierte.
9. **Paläontologie** - Wissenschaftliche Disziplin, die die Erforschung prähistorischer Lebensformen durch Fossilien umfasst.
10. **Biodiversität der Kreidezeit** - Wissenschaftliches Interesse an der Vielfalt und Ökologie der Lebensformen aus der Kreidezeit, zu der die Tendaguru-Betten gehören.



---

## Erkenntnisse Schlagwortsuche

- wenn das Ergebnis der automatischen Texterkennung nicht zu schlecht ist, werden die relevanten Personen und Begriffe gefunden (meist jedoch nicht in der tatsächlich vorhandenen Häufigkeit)
- das Ergebnis wird besser, wenn man sich an die Tokensobergrenze hält und das Dokument entsprechend aufteilt
- derselbe Prompt kann entweder die gewünschte Namensliste zum Ergebnis haben, aber auch Listen von Substantiven oder sogar eine Zählung der häufigsten Worte insgesamt
- Lösung zur Absicherung: die gefundenen Entitäten durch eine Volltextsuche bestätigen



## GND-Nummern

Finde bitte diese Namen in der GND.



Die Gemeinsame Normdatei (GND) ist eine große Autoritätsdatei, die von der Deutschen Nationalbibliothek verwaltet wird und vor allem zur Normierung von Personennamen, Körperschaften und geographischen Begriffen dient. Ich kann auf diese externe Datenbank nicht direkt zugreifen, um die Informationen zu überprüfen.

Finde bitte diese Namen online

3 Websites durchsucht

Werner Janensch ist in der Gemeinsamen Normdatei (GND) unter der ID 117078166 registriert. Er war ein bedeutender deutscher Paläontologe und Geologe, geboren am 11. November 1878 in Herzberg (Elster) und verstorben am 20. Oktober 1969 in Berlin. Janensch ist vor allem für seine Führungsrolle bei der Tendaguru-Expedition in das heutige Tansania bekannt, bei der zahlreiche spätjurassische Dinosaurierfossilien, einschließlich mehrerer vollständiger Brachiosaurus-Skelette,



---

# GND- und Wikidata-Nummern für Orte

Hier sind die GND- und Wikidata-Einträge für die Orte entlang der Reiseroute der Expedition:

## 1. Berlin

- GND: 4005727-1
- Wikidata: [Q64](#)

## 2. Neapel

- GND: 4042535-8
- Wikidata: [Q2634](#)

## 3. Port Said

- GND: 4045910-1
- Wikidata: [Q202184](#)

## 4. Suez-Kanal

- GND: 4059644-5
- Wikidata: [Q12516](#)



---

## Ergebnis GND-Nummern

- Identifizierung von GND-Nummern durch ChatGPT unzuverlässig  
→ oftmals Halluzinationen
- Verbesserung durch Prompt, der auf einschlägige Websites verweist (z.B. Deutsche Biographie)
- halbautomatische Lösung:  
→ Ausgabe von Personen + manuelle Nachrecherche in GND



---

Ergebnis für UAT 407/80 in Open Digi

[https://opendigi.ub.uni-tuebingen.de/opendigi/UAT\\_407\\_080](https://opendigi.ub.uni-tuebingen.de/opendigi/UAT_407_080)



Info Inhalt OCR-Volltext

Signatur

UAT 407/80

Titel

Edwin Hennig (1882-1977), Nachlass

Umfangsgabe

164 Seiten

Untertitel

Tagebuch, Teil 1

Person(en)

Hennig, Edwin

Ort(e)

Afrika

Laufzeit

11.3.1909-29.4.1910

Schlagwörter

Dinosaurier [\[O\]](#)

Familienarchiv [\[O\]](#)

Geologie [\[O\]](#)

Nachlass [\[O\]](#)

Paläontologie [\[O\]](#)

Swahili [\[O\]](#)

Tendaguru-Expedition (1909-1913) [\[O\]](#)

Daressalam [\[O\]](#)

Deutsch-Ostafrika [\[O\]](#)

Lindi (Tansania) [\[O\]](#)

Tansania [\[O\]](#)

Tendaguru-Schichten [\[O\]](#)

Deutsch-Ostafrikanische Gesellschaft [\[O\]](#)

Hennig, Edwin [\[O\]](#)

Janensch, Werner [\[O\]](#)

Sattler, Bernhard [\[O\]](#)

Zusammenfassung

Inhaltsangabe: Die Tendaguru-Expedition (11. März 1909 - 29. April 1910) unter der Leitung von Dr. Edwin Hennig und Dr. Werner Janensch begann in Berlin und führte über Europa nach Afrika, wo sie in Daressalam und Lindi ankam. Hauptziel war die Ausgrabung von Dinosaurierfossilien im Tendaguru-Gebiet. Die Expedition entdeckte zahlreiche Fundstellen mit Saurierknochen und anderen Fossilien. Die Tagebücher dokumentieren die geologischen und paläontologischen Arbeiten, die Organisation der Grabungen, die Versorgung der Arbeiter und die klimatischen Herausforderungen. Zusätzlich berichten die Tagebücher über Begegnungen mit der lokalen Bevölkerung, Jagdausflüge und Beobachtungen der regionalen Flora und Fauna. Die Expedition trug wesentlich zur Paläontologie und zum Verständnis der Dinosaurierfauna in Ostafrika bei. (erstellt unter Anwendung von ChatGPT 4o)

Weitere Informationen

Tagebuch, Teil 2

URN

urn:nbn:de:bsz:21-dt-210688

DOI

10.20345/digitue.21068 [\[O\]](#)

Digitale Repräsentationen

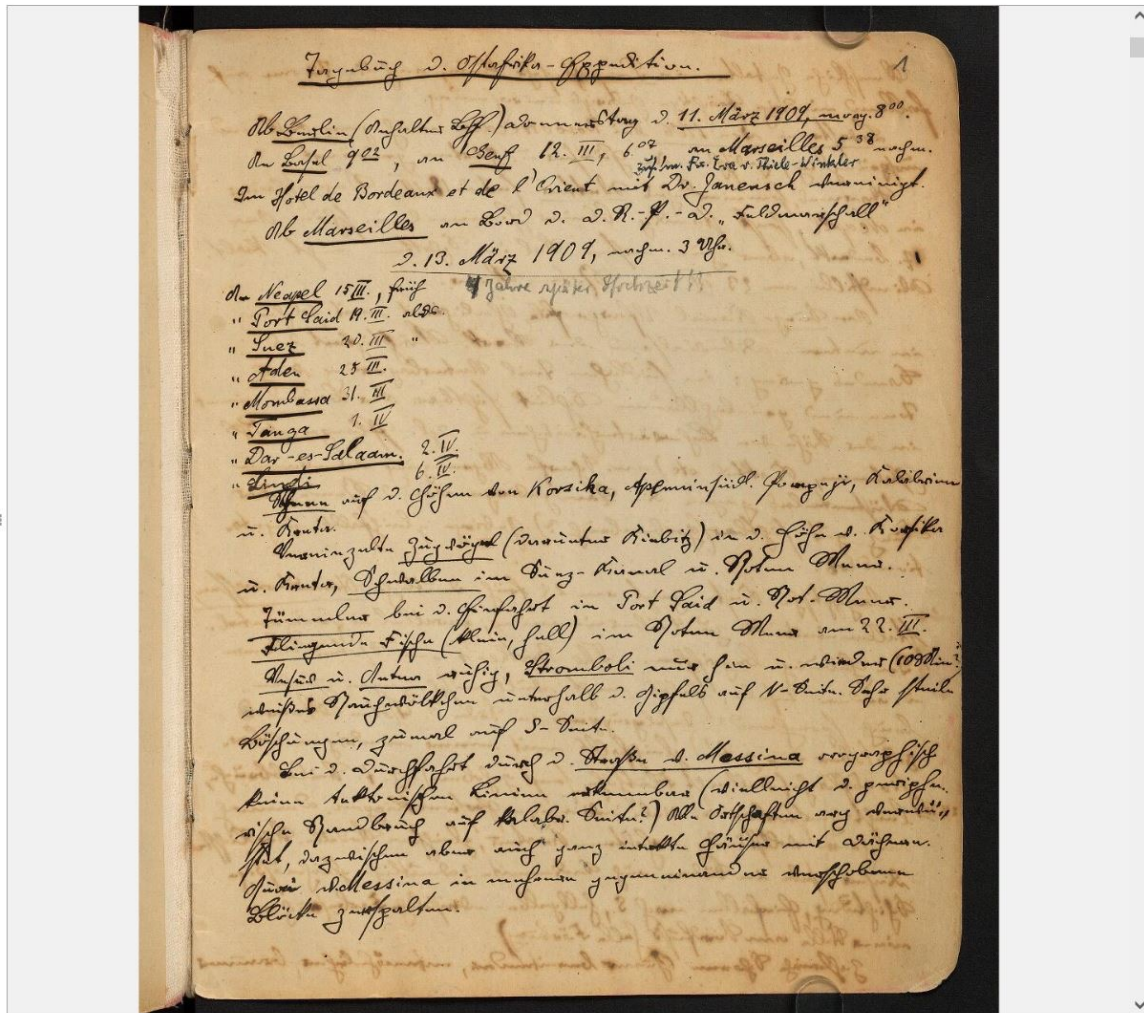
METS

IIIF-Manifest

Downloads

PDF: Ganzes Werk, Aktuelle Seite

JPEG: Aktuelle Seite





### Schlagwörter

Dinosaurier [GND](#)

Familienarchiv [GND](#)

Geologie [GND](#)

Nachlass [GND](#)

Paläontologie [GND](#)

Swahili [GND](#)

Tendaguru-Expedition (1909-1913) [GND](#)

Daressalam [GND](#)

Deutsch-Ostafrika [GND](#)

Lindi (Tansania) [GND](#)

Tansania [GND](#)

Tendaguru-Schichten [GND](#)

Deutsch-Ostafrikanische Gesellschaft [GND](#)

Eberhard Karls Universität Tübingen. Geologisch-Paläontologisches Institut [GND](#)

Hennig, Edwin [GND](#)

Janensch, Werner [GND](#)

Sattler, Bernhard [GND](#)

### Zusammenfassung

Inhaltsangabe: Die Tendaguru-Expedition (11. März 1909 - 29. April 1910) unter der Leitung von Dr. Edwin Hennig und Dr. Werner Janensch begann in Berlin und führte über Europa nach Afrika, wo sie in Daressalam und Lindi ankamen. Hauptziel war die Ausgrabung von Dinosaurierfossilien im Tendaguru-Gebiet. Die Expedition entdeckte zahlreiche Fundstellen mit Saurierknochen und anderen Fossilien. Die Tagebücher dokumentieren die geologischen und paläontologischen Arbeiten, die Organisation der Grabungen, die Versorgung der Arbeiter und die klimatischen Herausforderungen. Zusätzlich berichten die Tagebücher über Begegnungen mit der lokalen Bevölkerung, Jagdausflüge und Beobachtungen der regionalen Flora und Fauna. Die Expedition trug wesentlich zur Paläontologie und zum Verständnis der Dinosaurierfauna in Ostafrika bei. (erstellt unter Anwendung von ChatGPT 4o)



---

## Grenzen

- Zuverlässigkeit bzw. Halluzinationen
- Reproduzierbarkeit: selbst mit demselben Prompt variieren die Ergebnisse für dasselbe Dokument
- Beibehaltung der Formatierung nach z.B. Datenkorrektur wurde nicht immer beachtet → Problem, für Nachnutzung der Daten in einer Transkriptionsumgebung oder Präsentationsplattform
- Nichteuropäische Sprachen: inhaltliche Zuordnung von Begriffen in Swahili schwierig (Frage nach einheimischen Bevölkerungsgruppen ergab eine Mischung aus entsprechenden Bezeichnungen, aber auch solchen für Tiere und Landschaften)



## Fazit

- CER muss nicht bei Null liegen → automatische Transkriptionsdaten können für Erschließung genutzt werden
  - Inhaltsangabe funktioniert gut, auch bei Dokumenten, die die Tokengrenze überschreiten
  - bei der Schlagwortsuche werden nicht immer alle Instanzen gefunden (Namen sind oftmals schwieriger für OCR/HTR), aber ausreichend für die Schlagwortvergabe
  - Textkorrektur und Schlagwortsuche funktioniert besser innerhalb der Tokengrenze → längere Texte aufteilen
- Korrektur scheint besser auf unformatiertem Fließtext als auf zeilengetreuen Transkriptionen zu funktionieren → dies für Sacherschließung nutzen
- man kann ChatGPT fragen, wie sich Anfragen verbessern lassen
  - ein Prompt ergibt jedoch nicht immer dasselbe Ergebnis
  - Prompts sollten so ausführlich und ausbuchstabiert wie möglich sein → mitunter muss man mehrfach nachfragen (zäher Prozess)



# Danke.

## **Kontakt:**

Universitätsbibliothek Tübingen  
Wilhelmstraße 32, 72074 Tübingen

Kristina Stöbener

Telefon: +49 7071 29-72834

[kristina.stoebener@uni-tuebingen.de](mailto:kristina.stoebener@uni-tuebingen.de)

Dorothee Huff

Telefon: +49 7071 29-72852

[dorothee.huff@uni-tuebingen.de](mailto:dorothee.huff@uni-tuebingen.de)