

Probleme bei der Qualitätskontrolle zum digitalen Deutschen Kolonial-Lexikon

Kathleen Thiele, Dresden

Bei den ca. 3000 digitalisierten und bearbeiteten Seiten des Deutschen Kolonial-Lexikons waren Fehler nicht zu vermeiden. Die auftretenden Fehler hatten nicht nur menschlichen, sondern auch technischen Ursprung. In allen Arbeitsschritten bei der Umsetzung von der analogen in die digitale Form traten kontinuierlich Probleme auf, die zu Fehlern führten. Daher war Fehleranalyse ein wichtiges Instrument der Qualitätskontrolle.

Fehler bei der Texterkennung (OCR)

Erste Probleme traten bereits bei der Texterkennung auf, da es auf Grund der 1920 gedruckten Vorlage zur falschen Interpretation von Zeichen kam.

Das lag einerseits an teilweise schlechten Originalvorlagen in denen der Druck zu schwach oder der Zeilenabstand zu eng gesetzt war. Durch die Maserung des Papiers entstanden beim Scannen Phantomstriche und -flecken, die sich durch die Papierqualität des Originals nicht vermeiden ließen.

Die Verzerrungen, die durch die Bindung entstehen würden, wurden durch die Zerlegung des Originals vermieden.

Die häufigsten Fehler, die bei der Texterkennung durch eine problematische Vorlage entstehen, sind Fehlinterpretationen von Zeichen. So konnte es passieren, daß aus einem „S“ eine „8“, aus einem „I“ oder „l“ eine „1“, aus einer „5“ eine „6“ oder aus einem „m“ die Lautfolge „un“ wurde. Ebenfalls wurden die Sonderzeichen für die Deutsche Mark und den Deutschen Pfennig grundverschieden interpretiert und mußten manuell ersetzt werden.

Fehler bei der Bearbeitung

Trotz intensiver und gewissenhafter Arbeit unterlaufen auch den Mitarbeitern Fehler, die in der Menge des bearbeiteten Materials liegen. So konnte es in bestimmten Fällen passieren, daß Fehler der OCR-Erkennung

Angra Pequena, der ursprüngliche Name für Lüderitzbucht (s. d.), Deutsch-Südwestafrika, der nur noch historisches Interesse hat. Er ist bereits mit der Geschichte der großen portugiesischen Entdeckerfahrten verknüpft, denn der Name wurde ihr von dem großen Bartolomeo Dias (s. d.) bereits im Jahre 1486 verliehen im Hinblick auf die ziemlich enge Einfahrt in die Bucht. Der Seefahrer errichtete ein Marmorkreuz mit dem Wappen Portugals, das aber, nachdem es drei Jahrhunderte auf einem Felsen gestanden hatte, von schätzesuchenden Abenteurern zerstört wurde. Die vergessene Bucht wurde zu einem der am meisten genannten Punkte des Erdballs, als durch die Depesche des Fürsten Bismarck nach Kapstadt vom 24. April 1884 das Land nördlich vom Oranjefluß offiziell als unter dem Schutze des Reiches stehend erklärt wurde (s. Kolonialgeschichte). Während der ersten Zeit wurde der Name in Deutschland kurzerhand für das ganze Hinterland gebraucht. Wenn er auch später mehr und mehr außer Gebrauch kam und der dem Erwerber der Bucht zu Ehren gewählte ihn schließlich völlig verdrängte, so verdient er doch schon um deswillen nicht vergessen zu werden, weil mit ihm die Erinnerung an den Beginn einer aktiven Kolonialpolitik von seiten des Deutschen Reiches untrennbar verbunden ist. Dove.

Rohscan aus den Kolonial-Lexikon

übersehen, Textteile abgeschnitten oder Absatz- und Schriftformatierungen bei der HTML-Programmierung vergessen wurden. Diese Fehler wurden bei der anschließenden Korrekturlesung durch unsere Studenten und Mitarbeiter überprüft und behoben.

Fehler bei der automatischen Verlinkung

Um eine schnelle und komfortable Nutzung des Kolonial-Lexikons zu ermöglichen, war eine Verlinkung der im Lexikon vorkommenden Verweise sinnvoll. Die Verlinkungen mußten automatisch erzeugt werden, da eine manuelle Verlinkung bei der großen Anzahl von Artikeln (ca. 10.000) zu Zeitaufwendig gewesen wäre.

Bedingt durch die Komplexität der Schlagworte konnte die automatische Verlinkung nicht fehlerfrei durchgeführt werden. So war es nötig, defekte oder fehlerhafte Verlinkungen zu suchen und manuell zu korrigieren. Dazu wurde das Hyperlinkcheckprogramm XENU verwendet.

lenumbruch ausgeführt, der zur Folge hatte, daß diese Links zum Teil nicht funktionierten.

Ein anderes Problem war die inhaltlich falsche Verlinkung von Artikeln durch folgende drei Grundprobleme:

1. Synonyme Wörter:

- die Fallen – das Fallen
- der Verkehr (Mobilität) – der Verkehr (Beziehungen) – der Verkehr (Geschlecht)
- der Stengel – Stengel (Name einer Person)
- der Kaiser (Staatsoberhaupt) – Kaiser (Name einer Person)
- der Fang – die Fang (Volk)
- die Hörner (vom Tier) – die Hörner (Musikinstrument)

```

http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/H/Hamburgisches_Kolonialinstitut.html
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/%D6/%D6ffentliches_Recht.html
\_____ error code: 404 (not found)
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/H/Harze.html
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/%C4/%C4therische_%D6le.html
\_____ error code: 404 (not found)
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/H/Hamiten.html
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/D/Deutsch-%20Ostafrika.html
\_____ error code: 404 (not found)
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/H/Handel.html
http://www.stub.bildarchiv-dkg.uni-frankfurt.de/dfg-projekt/Lexikon-Texte/K/Kaiserin-%20Augustafluß.html
\_____ error code: 404 (not found)

```

Fehlerausgabe von XENU

Eine häufige Fehlerquelle war dabei die falsche Verlinkung von Artikeln, deren Anfangsbuchstaben Umlaute sind. Diese Artikel liegen wegen ihrer geringen Anzahl nicht in separaten Ordnern, sondern sind ihrem ursprünglichen Stamm zugeordnet. Das heißt, der Artikel *Ätherische Öle* liegt im Ordner A, der Artikel *Übertretung* liegt im Ordner U, *Ölpalme* unter O, usw.. Die Verlinkung wurde jeweils auf die nicht existierenden Ordner Ä, Ö und Ü gesetzt.

Des weiteren kam es zu Fehlern durch ungewollt auftretende Zeilenumbrüche im Quelltext der Verlinkungen. So wurde jeweils nach dem Bindestrich in zusammengesetzten Wörtern, wie *Deutsch-Ostafrika*, *Kaiserin-Augustafluß* und *Reichs-Marineamt*, ein Zei-

2. Unterschiedliche Schreibweisen:

- Reichsmarineamt – Reichs-Marineamt
- Mkatasenke – Mkata-Senke
- Kahnfluß – Kahn
- Oertzen – Örtzen

3. Leerzeichen:

- Deutsch- Südwestafrika, nur Südwestafrika verlinkt statt Deutsch-Südwestafrika
- Deutsch- Ostafrikanische Gesellschaft, nur Ostafrikanische Gesellschaft verlinkt
- Jaluit –Gesellschaft, nur Jaluit verlinkt

Trotz intensiver Korrekturlesung und vieler Korrekturläufe sind bei der Menge der Schlagworte sicher noch Fehler enthalten. Für entsprechende Hinweise ist die Arbeitsgruppe dankbar.