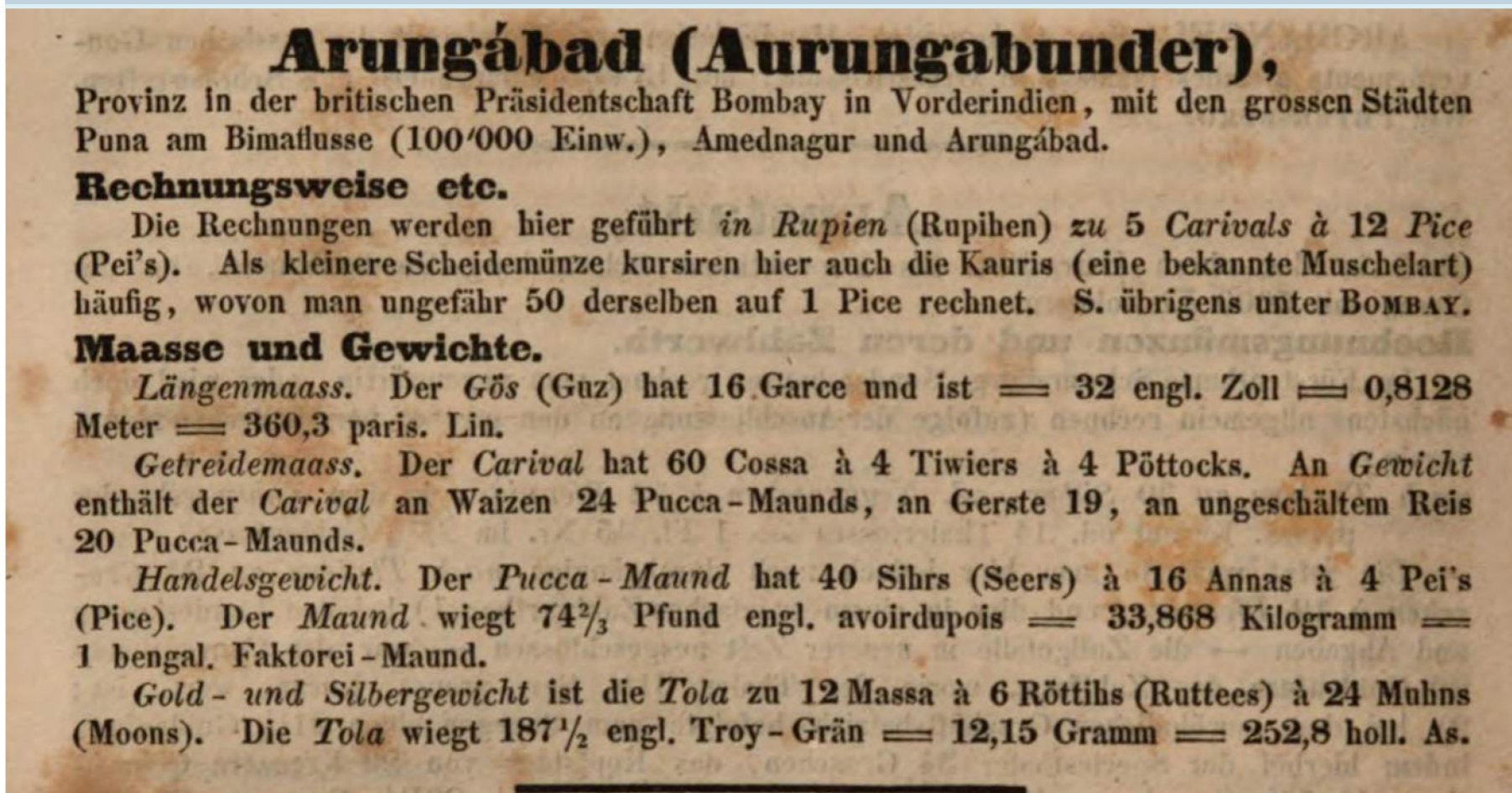


Szenario-basierte Planung eines semantischen Digitalisierungsvorhabens in der digitalen Geschichtswissenschaft

Werner Scheltjens  0000-0002-5209-9052

Christoph Schlieder  0000-0002-7226-8204



(a) Digitalisat Noback-Handbuch

Arungabad (Aurangabad) ,
 Provinz in der britischen Präsidentschaft Bombay in Vorderindien , mit den grossen Städten
 Puna am Bimafusse (100'000 Einw.) , Amednagur und Arungabad .
 Rechnungsweise etc.
 Die Rechnungen werden hier geführt in Rupien (Rupihen) zu 5 Carivals à 12 Pice
 (Pei's) . Als kleinere Scheidemünze kursiren hier auch die Kauris (eine bekannte Muschelart)
 häufig , wovon man ungefähr 50 derselben auf 1 Pice rechnet . S. übrigens unter BOMBAY .
 Maasse und Gewichte .
 Längenmaass . Der Gös (Guz) hat 16 Garce und ist = 32 engl. Zoll = 0,8128
 Meter = 360,3 paris. Lin.
 Getreidemaass . Der Carival hat 60 Cossa à 4 Tiwiers à 4 Pöttocks . An Gewicht
 enthält der Carival an Waizen 24 Pucca - Maunds , an Gerste 19 , an ungeschältem Reis
 20 Pucca - Maunds .
 Handelsgewicht . Der Pucca - Maund bat 40 Sihrs (Seers) à 16 Annas à 4 Pei's
 (Pice) . Der Maund wiegt 74 ²/₃ Pfund engl. avoirdupois = 33,868 Kilogramm =
 1 bengal. Faktorei - Maund .
 Gold - und Silbergewicht ist die Tola zu 12 Massa à 6 Rötthi (Ruttees) à 24 Muhns
 (Moons) . Die Tola wiegt 187 ¹/₂ engl. Troy - Grän = 12,15 Gramm = 252,8 holl. As

(b) Beurteilung der OCR-Qualität

```

2 <article>
3 <lemma>Arungabad (Aurangabad) ,</lemma>
4 <search-area-0>
5 <search-area-1>Provinz in der britischen Präsidentschaft Bombay in Vorderindien , mit den grossen Städten
6 Puna am Bimafusse ( 100'000 Einw. ) , Amednagur und Arungabad.</search-area-1>
7 <indicator-1>Rechnungsweise etc.</indicator-1>
8 <search-area-1>Die Rechnungen werden hier geführt in Rupien ( Rupihen ) zu 5 Carivals à 12 Pice
9 ( Pei's ) . Als kleinere Scheidemünze kursiren hier auch die Kauris ( eine bekannte Muschelart )
10 häufig , wovon man ungefähr 50 derselben auf 1 Pice rechnet . S. übrigens unter BOMBAY.</search-area-1>
11 <indicator-1>Maasse und Gewichte.</indicator-1>
12 <search-area-1>
13 <indicator-2>Längenmaass .</indicator-2>
14 <search-area-2>Der Gös ( Guz ) hat 16 Garce und ist = 32 engl. Zoll = 0,8128
15 Meter = 360,3 paris. Lin.</search-area-2>
16 <indicator-2>Getreidemaass .</indicator-2>
17 <search-area-2>Der Carival hat 60 Cossa à 4 Tiwiers à 4 Pöttocks . An Gewicht
18 enthält der Carival an Waizen 24 Pucca - Maunds , an Gerste 19 , an ungeschältem Reis
19 20 Pucca - Maunds .</search-area-2>
20 <indicator-2>Handelsgewicht .</indicator-2>
21 <search-area-2>Der Pucca - Maund bat 40 Sihrs ( Seers ) à 16 Annas à 4 Pei's
22 ( Pice ) . Der Maund wiegt 74 2/3 Pfund engl. avoirdupois = 33,868 Kilogramm
23 1 bengal. Faktorei - Maund.</search-area-2>
24 <indicator-2>Gold - und Silbergewicht.</indicator-2>
25 <search-area-2>ist die Tola zu 12 Massa à 6 Rötthi ( Ruttees ) à 24 Muhns
26 ( Moons ) . Die Tola wiegt 187 1/2 engl. Troy - Grän = 12,15 Gramm = 252,8 holl. As.</search-area-2>
27 </search-area-0>
28 </article>
    
```

(c) Test der Auszeichnung mit der Lexikographie-Ontologie

Semantische Digitalisierung

Die Retrodigitalisierung von nicht urheberrechtlich geschützten Bibliotheks-sammlungen hat Lexika, Handbücher und Enzyklopädien als digitale Quellen verfügbar gemacht. Als Ergebnis dieser „ersten Digitalisierung“ liegt meist ein Scan (a) und der oft mit Fehlern behaftete Volltext (b) vor.

Für bestimmte, konkret identifizierbare historische Forschungsfragen ist eine zweite, semantische Digitalisierung unabdingbar (c). Diese strebt die Extraktion und explizite Modellierung der semantischen Struktur des enzyklopädischen Wissens an.

Planung der semantischen Modellierung

Für die Planung der auf die Explizierung semantischer Beziehungen zielenden Digitalisierung historischer Quellen steht im Prinzip das allgemeine Methodeninventar der ontologischen Modellierung zur Verfügung.

An erster Stelle sind hier szenario-basierte Methoden zu nennen, die den Planungsprozess an sogenannten Kompetenzfragen orientieren, d.h. einem Katalog der Fragen, die Fachanwender:innen anhand der Modellierung untersuchen und beantworten wollen (Kendall, McGuinness, 2019; Lodi et al., 2017, Carriero et al. 2021).

Beispiele für bei der Planung betrachtete Kompetenzfragen

Modellierungsebene	1	2	3
Frage	Welche Typen von Maßen führt Noback (1850) zum Lemma Arungabad auf?	Welche Maßeinheiten für Längen sind unter Arungabad definiert?	An welchen Handelsplätzen "Vorderindiens", sind lokale Maßeinheiten an das Maßsystem einer Kolonialmacht angeglichen worden?
Antwort	"Längenmaass", "Getreidemaass", ...	"EN Guz, DE Gös", "EN DE Garce"	"Benares", "Bombay", ... , "Ajinga"
Algorithmisierung	Kann durch Blättern im Digitalisat beantwortet werden. Keine semantische Modellierung erforderlich.	Die NER darf die deutsche Transkription "Gös" nicht als von "Guz" unterschiedene Maßeinheit identifizieren.	Anwender:innen müssen definieren können, was Angleichung der Maßeinheit bedeutet. Die Umsetzung erfolgt über Inferenzdienste auf der semantischen Modellierung.

Ergebnisse der Planung

- 15 historisch-metrologische Kompetenzfragen
- 3 Modellierungsebenen unterschiedlicher Komplexität
- Selbst auf Modellierungsebene 1 lassen sich nur wenige Fragen mit vertretbarem Aufwand händisch mit dem Digitalisat beantworten.
- Die OCR-Qualität ist hinreichend für die Beantwortung nur weniger Kompetenzfragen der Modellierungsebene 1.
- Zur Beantwortung der Fragen wird eine Domänenontologie zur historischen Metrologie sowie zum Handel und zur historischen Geographie benötigt.
- Kompetenzfragen der Ebene 3 erfordern eine lexikographische Ontologie, die z.B. die Auswahlentscheidungen und die Verweisstruktur des Handbuchs abbilden kann.

Digitalisat

Noback, Christian, Noback, Friedrich (1850): Vollständiges Taschenbuch der Münz-, Maass-, und Gewichtsverhältnisse, der Staatspapiere, des Wechsels- und Bankwesens, und der Usanzen aller Länder und Handelsplätze. Leipzig: F.A. Brockhaus.
 Bayerische Staatsbibliothek, Münchner Digitalisierungszentrum, Digitale Bibliothek

Literatur

- Carriero, V. et al. (2021): "Pattern-based Design Applied to Cultural Heritage Knowledge Graphs", in: Semantic Web 12: 313 – 357.
- Kendall, E., McGuinness, D.(2019): Ontology engineering. (= Synthesis Lectures on The Semantic Web: Theory and Technology, Lecture 18). Morgan and Claypool.
- Lodi, G. et al. (2017): "Semantic Web for Cultural Heritage Valorisation", in: Hai-Jew, Shalin (ed.): Data Analytics in Digital Humanities. Multimedia Systems and Applications. Springer: Cham 3-37.

Kontakt

Prof. Dr. Werner Scheltjens, Professur für Digitale Geschichtswissenschaft, Otto-Friedrich-Universität Bamberg, werner.scheltjens@uni-bamberg.de