



Florian Hermann/ Susanne Gütter

Regex und Grafische Modellierung in QGIS

QGIS-Tipps zum Nutzen Regulärer Ausdrücke und eigener Modell-Werkzeuge am Beispiel der CAD-GIS-Konvertierung von Grabungsplänen

Im alltäglichen Umgang mit digitalen Grabungsdaten stehen häufig zeitraubende, sich wiederholende Arbeitsschritte an. Beispielsweise, wenn im Zuge der Umstellung von CAD- zu GIS-gestützter Grabungsdokumentation bestehende digitale Grabungspläne aus DWG- in GIS-Formate zu überführen sind. Dabei sollen sowohl topologisch saubere Geometrien erzeugt, als auch bereits in den CAD-Plänen zu den Objekten vorhandene Daten in die Attributtabelle übernommen werden. Bei der Konvertierung größerer Komplexe wird schnell der Bedarf an effizienten Arbeitsabläufen deutlich. Mit Hilfe einer genau an die vorliegende Datenstruktur angepassten Prozessierung kann dabei die händische Bearbeitung der einzelnen Pläne wesentlich reduziert werden.

An dieser Stelle sollen nun zwei zwar voneinander unabhängige, aber flexibel miteinander kombinierbare Helfer für derartige repetitive Arbeiten vorgestellt werden: Sogenannte **Reguläre Ausdrücke** zur Extraktion von Informationen aus Texten und die **grafische Modellierung** in QGIS, mit der einzelne Bearbeitungsschritte zu einem automatisierten Prozess verkettet werden können.

Am konkreten Beispiel wird gezeigt, wie der QGIS-Feldrechner dazu genutzt werden kann, um aus den Layerbezeichnungen Informationen in die Attributtabelle zu übertragen und wie sich über die Grafische Modellierung in QGIS beliebige Algorithmen zu selbst erstellten Werkzeugen verknüpfen lassen. Als Ausgangsbasis dient die umfangreiche Grabungsdokumentation einer komplexen Grabung, die mithilfe von Tachy-CAD (als AutoCAD-Applikation) erstellt worden ist und welche die folgenden Kriterien weitgehend erfüllt:

- Die CAD-Layerbezeichnungen beinhalten die Bezugsdaten (wie Planum/Dokumentationseinheit/Befund-Nr./Fund-Nr.).
- Die CAD-Layerbezeichnungen folgen einem konsequenten Benennungsschema mit definierten Trennzeichen zwischen den einzelnen Informationsblöcken (z.B. Unterstriche).
- Jeder Befund liegt auf einem eigenen CAD-Layer.

Nach dem Import der CAD-Daten nach QGIS (siehe Exkurs) liegen im QGIS-Projekt die Linien- und Punktlayer des Grabungsplanes vor, wobei in einer Spalte der zugehörigen Attributtabelle zu jedem Punkt- und Linienobjekt die ursprüngliche Layerbezeichnung des CAD-Plans gelistet ist.

Exkurs: CAD-Import nach QGIS

Sachdaten liegen in den digitalen Grabungsplänen auf verschiedene Weise vor: Vorteilhaft zur Migration ist ein CAD-Schema, mit nach Befunden getrennten Layern, in denen die Layerbezeichnungen die nötigen Bezugsdaten (wie Planum, Dokumentationseinheit, Befund-Nr., Fund-Nr. o. ä.) beinhalten. Informationen liegen daneben auch häufig als Blöcke, die Punktgeometrien zugeordnet sind (Nivellierhöhen, Befundmitten, Fund- oder Proben-Nr., o. ä.) oder platzierte Texte (ohne Geometriebezug) vor.

Der Import von Liniengeometrie-Layern kann direkt über den QGIS eigenen DWG-Importer erfolgen (Option *Blockreferenzen expandieren* abwählen). Der Plan wird als Geopackage gespeichert. Die beim Import automatisch erzeugte Attributtabelle besitzt, neben zahlreichen anderen CAD-generierten Einträgen, eine Spalte „Layer“, die den Layernamen enthält.

Punktgeometrien mit Blockattributen lassen sich günstiger aus dem CAD-Programm heraus über Datenextraktion (DATENEXTRAKT) in eine CSV-Textdatei exportieren, wobei sichergestellt werden muss, dass beim Export neben den gewünschten Attributen auch die Geometriespalten (XYZ) und der Layername mit übergeben werden. Auf diesem Weg lassen sich Probleme mit Blöcken beim DWG-Import umgehen. Die Punkte werden anschließend in QGIS über die Funktion *getrennter Text* als Layer hinzugefügt (Strg+Umschalt+T).

Zur weiteren Verarbeitung in QGIS und zur Erzeugung von Polygonen sind saubere, *gültige* Geometrien vorausgesetzt, die weder Überschneidungen, doppelte Stützpunkte noch doppelte Objekte enthalten. Zudem sollten Linien ohne Klaffungen anschließen und geschlossene Objekte als geschlossene Polylinien vorliegen. Für diese Bereinigung steht in QGIS eine Reihe an Werkzeugen zur Verfügung, auf die hier nicht näher eingegangen werden soll.

Reguläre Ausdrücke

Reguläre Ausdrücke (regular expressions, kurz: **Regex)** sind, abstrakt gesprochen, Zeichenketten (engl. strings) mit syntaktischen Regeln, mit welchen (Teil-)Mengen anderer Zeichenketten beschrieben werden können. Sie dienen also als Platzhalter, mit welchen komplexere Sachverhalte wie: suche „eine zwischen **1 und 3 Stellen lange Zahl** nach dem Zeichen **X** und vor einem **Y**“, oder „alle **Wörter nach einer dreistelligen Zahl** und nach einem **_** (Unterstrich), die mit **P** beginnen“ ausgedrückt werden können. Sie sind in zahlreichen Bereichen der theoretischen wie praktischen Informatik in verschiedenen Ausprägungen implementiert und können daher auch in vielen in der Archäologie eingesetzten Softwareprogrammen genutzt werden¹. Zwei populäre Beispiele hierfür sind der vielfältig nutzbare Texteditor Notepad++, dessen Suche-Dialog auf **Reguläre Ausdrücke** umgestellt werden kann sowie QGIS, wo sie insbesondere in den *Feldrechnern* von Nutzen sind. Im Internet finden sich viele Anleitungen zum Thema. Besonders empfehlenswert sind Seiten, die eine „live“-Testumgebung anbieten, in der das Ergebnis der Anwendung der eigenen Ausdrücke direkt an eigenen Testdatensätzen geprüft werden kann (siehe Link-Tipps).

An dieser Stelle soll für unseren Zweck das Funktionsprinzip der **Regulären Ausdrücke** anhand des Beispieldatensatzes in Tabelle 1 erläutert werden.

Im Beispiel sind im Layernamen das Freilegungsniveau („DE-NR“), die Kategorie der dargestellten Objekte (als Code und Text), die Befundnummer und weitere Informationen nach einem festen, von Trennzeichen geordneten Schema angeordnet:

100.DE_500_Befundgrenze_3333-OK-unsicher

[Nummer der Dokumentationseinheit].DE_[Code]_[CodeBeschreibung]_[Befundnummer]-[Lagebezeichner]-[Kommentar]

Nun gilt es ein Schema zu entwickeln, nach dem die gewünschten Daten innerhalb der Zeichenkette des Layernamens zur Extraktion erkannt werden. Da die einzelnen Informationsblöcke nicht gleich lang sind und weder die Benennung noch die Reihenfolge der Parameter während der Vermessungsarbeiten immer stringent eingehalten wurden (Tippfehler!), lassen sie sich rein nach ihrer Position (z.B. Schnittnummer immer an Stellen 2-3) nicht auftrennen². Um die im Layernamen enthaltenen Informationen trotzdem - ohne aufwändige Vorab- Bereinigung der Layerbezeichnungen in den CAD-Plänen - innerhalb QGIS in einzelne Spalten zu extrahieren, kann die Funktion „`regex_substr`“ genutzt werden. Sie findet sich im Feldrechner der Attributtabelle unter der Rubrik Zeichenkette. Diese Funktion gibt aus einer Zeichenkette in einem bestimmten Feld einen gefilterten Wert aus, welcher den definierten Auswahlkriterien entspricht.

Die Funktion ist nach folgendem Muster aufgebaut:

`regex_substr("SpaltennameLayerbezeichnung", 'DefinitionAuswahl')`

¹ In QGIS werden Reguläre Ausdrücke nach der Spezifikation PCRE (Perl Compatible Regular Expressions) verwendet.

² Als gewachsene Struktur ist dieser Namensaufbau der Layer vielleicht typisch für eine langjährige Grabung mit sich wandelnden Dokumentationssystemen, jedoch ist seine Struktur nicht mit dem Ziel einer automatisierten Weiterverarbeitung entworfen worden.

ID	Layer
1	100.DE_500_Befundgrenze_3333-OK-unsicher
2	100.DE_400_Grabungsgrenze
3	101.DE_500_Befundgrenze_3333-UK
4	100.DE_500_Befundgrenze_605-OK-unsicher, entspricht ok Befund 2212
5	125.DE_500_Befundgrenze_3333-OK-unsicher

Tabelle 1: Beispieldatensatz der QGIS-Attributtabelle eines importierten CAD-Datensatzes. Für jedes Geometrie-Objekt (Punkt, Linie, Polygon) wird beim Import ein Datensatz/eine Zeile generiert. Die ursprüngliche, in CAD geführte Layerbezeichnung wird in einer Spalte *Layer* als Textstring (Zeichenkette) wiedergegeben.

Exkurs: Layerbezeichnungen

Die Nutzung eines eindeutigen Trennzeichens (z. B. Unterstrich) erleichtert die Bearbeitung ungemein. Zudem sind eine festgelegte Reihenfolge der Informationen und standardisierte Kürzel und Bezeichner hilfreich.

[]	Eckige Klammern definieren gesuchte Zeichen innerhalb der gesamten Zeichenkette. Es kann z. B. eine Zeichenfolge [DE], einzelne vorkommende Zeichen [2,3] oder auch ein Bereich von Buchstaben oder Ziffern genannt werden, wie [1-3], [A-ZäöüÄÖÜ]. Anwendungsbeispiel: [D]e Es werden nur jene großen „D“ ausgefiltert, auf die ein „e“ folgt
{ }	In geschweiften Klammern wird genannt, wie oft das voranstehende Zeichen/Gruppe vorkommen muss, z. B. {2} = genau 2-mal, {2,4} = mindestens 2-, höchstens 4-mal, {2,} = mindestens 2-mal
()	In einfachen Klammern werden mehrere Ausdrücke gruppiert (Anwendungsbeispiele in Rot in Tab. 3)
?	Der voranstehende Ausdruck ist optional, er kann also einmal vorkommen, braucht es aber nicht (entspricht {0,1})
+	Der voranstehende Ausdruck muss mindestens einmal, darf aber auch mehrfach vorkommen (entspricht {1,})
*	Der voranstehende Ausdruck darf beliebig oft (auch keinmal) vorkommen (entspricht {0,})
.	Punkt; ein sog. Metazeichen. Es entspricht praktisch jedem Zeichen mit Ausnahme von Zeilenumbrüchen und kann daher gut als Platzhalter benutzt werden
^	Durchsuche die Zeichenkette von vorne (gilt für die in QGIS verwendete Regex-Spezifikation! Sonst steht ^ häufig für Negation des folgenden Ausdrucks!)
\$	Suche die Zeichenkette vom Ende her. Steht am Ende der Eingabe
\d	Steht für eine Ziffer (entspricht [0-9])
\D	Steht für ein Zeichen, das keine Ziffer ist
\w	Steht für einen Buchstaben, eine Ziffer oder einen Unterstrich (entspricht [a-zA-Z_0-9]); umfasst keine Umlaute!
\W	Steht für ein Zeichen, das weder Buchstabe noch Zahl noch Unterstrich ist
\"	Zur Maskierung. QGIS verlangt vor einem Backslash ein sog. Maskierungszeichen („\“, siehe dazu Fußnote 1). Daher müssen diese im Feldrechner doppelt angegeben werden.

Es gibt auch die Möglichkeit, Bezüge innerhalb der Auswahl zu definieren, so sorgt z.B. AusdruckA(?!AusdruckB) dafür, dass AusdruckB nicht auf den vorgenannten AusdruckA folgen darf.

Tabelle 2: Detailsprache der in Tab. 3 genutzten Regex-Ausdrücke.

Extraktion von:	Ausdruck:	sucht im Feld „Layer“ nach:
Nummer der Dokumentationseinheit	<code>regexp_substr(„layer“, „(\d{2,3}).DE“)</code>	2-3-stelliger Zahl vor „.DE“
Code (Objektklasse)	<code>regexp_substr(„layer“, „DE_{(\d{3})}“)</code>	3-stelliger Zahl nach „DE_“
Befund-Nummer	<code>regexp_substr(„layer“, „_{(\d{2,4})(?!.*)}“)</code>	2-4-stelliger Zahl nach „_“ und (an beliebiger Stelle) nicht vor „_“
Lagebezug (Oberkante, Unterkante, Zwischenniveau)	<code>regexp_substr(„layer“, „\d{2}((OK;UK;ZN)){2})“)</code>	2-stelligem Textstring „OK/UK/ZN“ nach Zahl und „_“
Beschreibung Objektklasse	<code>regexp_substr(„layer“, „\d{3}_(\d{3}([a-zA-Z äöüÄÖÜ])*_)“)</code>	Textstring mit mögl. Leerzeichen (im String zwischen „Z“ und „ä“) und beliebig oft vorkommenden Buchstaben nach 3-stelliger Zahl und „_“, jedoch vor „_“
Kommentar	<code>regexp_substr(„layer“, „_{(?!([OK;UK;ZN]))(0-9a-zA-Z äöüÄÖÜ)*}\$“)</code>	n-stelligem alphanumerischem String am Ende, mit Leerzeichen (im String zwischen „Z“ und „ä“) und beliebig oft vorkommenden Buchstaben, welcher auf „_“ folgt und nicht „OK/ZN/UK“ entspricht

Tabelle 3: Regex-Ausdrücke zur Bearbeitung der in Tabelle 1 gelisteten Beispieldatensätze. Gruppen werden in runde Klammern gesetzt, sie sind hier rot hervorgehoben.

In der Klammer nach der Funktion wird also zuerst (in Anführungszeichen) das Feld der Attributtabelle genannt, in welcher der Layername steht. Nach einem Komma folgt dann (in einfache Anführungszeichen/Apostroph gesetzt) die Beschreibung, welche Stellen in der Zeichenfolge des Layernamens die gewünschte Information zur Ausgabe in eine neue Spalte der Attributtabelle enthält.

Für die Definition der gewünschten Zeichenauswahl stehen zahlreiche Optionen zur Verfügung³. Hier kommen u. a. die in Tabelle 2 aufgezeigten Ausdrücke zum Einsatz.

Für die Extraktion der einzelnen Informationsbestandteile im Layernamen in neue Spalten der Attributtabelle sind nun nach und nach die in Tabelle 3 aufgelisteten Beispielausdrücke auszuführen. Durch diese Befehle wird der **Regex**-Ausdruck auf den jeweiligen Eingabestring im Feld „Layer“ angewandt. Der Feldrechner zeigt unterhalb der Eingabe eine Vorschau des Ergebnisses an. Nur dieses wird separat in einer neuen Attributspalte abgelegt.

Auf diese Weise können also Einträge in die Felder „Dokumentationseinheit“, „Befundnummer“, „Planumsnummer“, etc. bequem für große Datenmengen extrahiert werden. Dies kann im vorliegenden Beispiel sowohl auf die Liniengeometrien des DWG-Importes als auch auf die Punktlayer angewandt werden, die über die CAD-Datenextraktion via CSV-Import geladen wurden.

³ Vgl. https://en.wikipedia.org/wiki/Perl-Compatible_Regular_Expressions [05.02.2022] und https://dewiki.de/Lexikon/Regul%C3%A4rer_Ausdruck [05.02.2022]

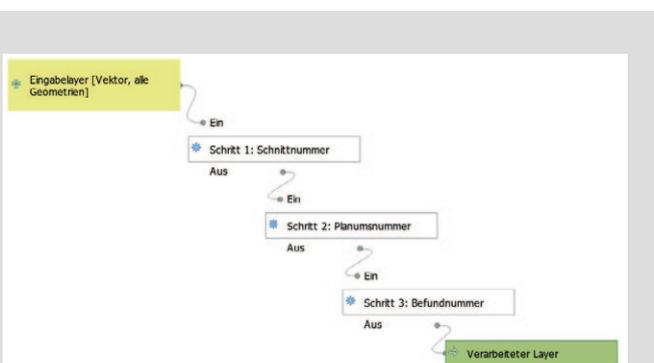


Abb. 1: Beispiel für ein einfaches lineares Modell mit einer Eingabe (gelb), drei Verarbeitungsschritten (weiß) und einer Ausgabe (grün).

Grafische Modellierung in QGIS

Sollen mehrere aufeinander folgende Bearbeitungsschritte wiederholt angewendet werden, so lassen sie sich in QGIS zu einem maßgeschneiderten Werkzeug verketteten. Dabei können sowohl selbst angepasste Befehle, wie die oben vorgestellte Spaltentrennung mit Regex-Ausdrücken, als auch bereits in QGIS implementierte Werkzeuge verbunden werden.

Erstellt wird eine solche Verkettung mit Hilfe der Grafischen Modellierung (im QGIS-Menü *Verarbeitung*), welche in Grundzügen bereits seit der Version 2.0 in QGIS implementiert ist. Wie die Bezeichnung bereits andeutet, werden die Arbeitsschritte (Prozesse) grafisch organisiert (Abb. 1), wofür weder Skript- noch Programmierkenntnisse notwendig sind. Eine solche nutzerspezifizierte Abfolge einzelner Arbeitsschritte wird *Modell* (engl. *model*) genannt. Einmal erstellt, steht es als weiteres Werkzeug in QGIS (unter *Verarbeitung* > *Werkzeugkiste* > *Modelle*) zur Verfügung. Die neu definierten Modelle können auf andere Rechner übertragen werden, indem sie in das entsprechende Verzeichnis kopiert werden. Voraussetzung zur Übertragung ist eine QGIS-Installation in einer zum Modell passenden Version⁴.

(Speicherort: C:\Users\[Benutzername]\AppData\Roaming\QGIS\QGIS3\profiles\default\processing\models).

Arbeitsablauf der Grafischen Modellierung

Am oben eingeleiteten Beispiel wird nun die grundsätzliche Arbeitsweise der **Grafischen Modellierung** demonstriert. Es soll erneut die Attributtabelle eines Vektorlayers neu strukturiert werden. Der AutoCAD-Import führt eine Spalte „Layer“ mit, in der der ursprüngliche Layername als Textstring (Zeichenkette) angegeben ist. Anstatt wie oben für jeden Bestandteil des Namensstrings den jeweiligen **Regex**-Befehl einzeln auszuführen, sollen die Einzelbefehle verbunden und so durch einen Klick ausführbar werden.

Unter dem Reiter *Verarbeitung* der QGIS-Menüleiste wird unter *Grafische Modellierung* die Entwicklungsumgebung aufgerufen (STRG + ALT + G). Innerhalb des Fensters *Modellentwurf* können nun beliebige aufeinanderfolgende Arbeitsschritte definiert werden. Dabei steht auch der Großteil der auch sonst in QGIS verfügbaren Algorithmen und Eingabeformen

⁴ Alle Beispiele sind in der QGIS Version 3.16 umgesetzt. In früheren QGIS Versionen ab Version 2.0 ist die Grafische Modellierung zwar grundsätzlich bereits implementiert, jedoch fehlen ggf. einige der hier vorgestellten Werkzeuge und Funktionalitäten. Das schränkt die Austauschbarkeit der Modelle zwischen den QGIS-Versionen mitunter stark ein. Grundsätzlich agiert QGIS in der Versionenkompatibilität aufwärts vergleichsweise tolerant, während abwärts sehr häufig mit Inkompatibilität zu rechnen ist. Als problemfrei hat sich der Austausch innerhalb derselben Version erwiesen.

zur Auswahl.

Die horizontale obere Menüleiste gibt die Möglichkeit zum Laden und Speichern von Modellen sowie Exportfunktionen nach Python und in grafische Formate (Raster, PDF, SVG). Außerdem findet sich dort die zum Modell gehörige Hilfedatei, in der Erläuterungen zum eigenen Modellwerkzeug formuliert werden können, die dann bei der Ausführung des Modells angezeigt werden. Ganz rechts im Menü dient ein grüner Pfeil zum Ausführen und Testen der Modelle.

Modelle bestehen stets aus drei Teilen: der **Eingabe**, den **Verarbeitungsschritten (Algorithmen)** sowie der **Ausgabe**.

Zunächst muss im Fenster *Modellentwurf* eine Eingabe definiert werden, welche im Modell als Platzhalter für den später tatsächlich zu verarbeiteten Ausgangs-layer dient. Hierfür

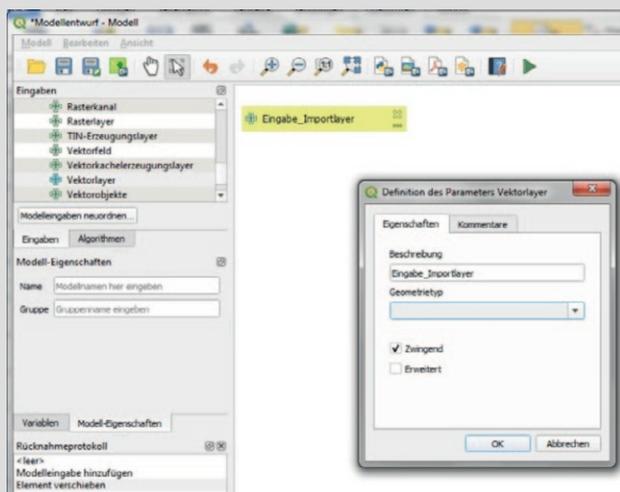


Abb. 2: QGIS-Eingabemaske zum Modellentwurf, Auswahl der Parameter für den Eingabelayer.

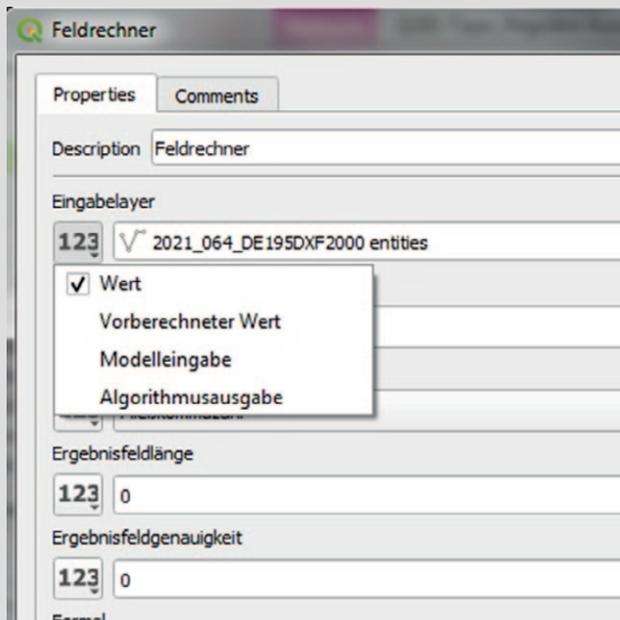


Abb. 3: Auswahl der Eingabekategorien für den Layer, auf den der Verarbeitungsschritt des Modells angewendet werden soll.

stehen im Tab *Eingaben* (i.d.R. auf der linken Seite) zahlreiche Eingabeparameter zur Verfügung. Die Palette reicht von allgemeinen Eingabemöglichkeiten wie „Vektorobjekte“, „Vektorlayer“, „Rasterlayer“ zu spezielleren wie „Zeitpunkt“, „Rasterkanal“, „Datenbankschema“, etc.

1. **Eingabe:** Im vorliegenden Fall sollen Vektorlayer verarbeitet werden, daher wird der entsprechende Eintrag aus dem Listenfeld entweder per Drag&Drop oder Doppelklick dem noch leeren weißen Bereich rechts hinzugefügt. Nun sind für den Eingabetyp spezifische Parameter zu definieren, hier u. a. der Name, den das Eingabefeld im Modell tragen soll sowie eine Definition der erlaubten Geometrietypen. Damit ist eine Eingabe definiert. An dieser Stelle könnten auch weitere Eingaben, auf welche die folgenden Bearbeitungsschritte ebenfalls angewendet werden sollen, hinzugefügt werden.

2. Nun muss der erste gewünschte **Verarbeitungsschritt** eingebracht werden. Die zur Verfügung stehenden Algorithmen finden sich unter dem gleichnamigen Reiter. Die Palette an Werkzeugen schließt neben den QGIS nativen Algorithmen, auch jene der GDAL-, GRASS- und SAGA-Pakete mit ein. Darüber hinaus kann auch auf nutzerdefinierte Modelle zugegriffen werden, was die Erstellung verschachtelter Modelle erlaubt. Im konkreten Fall fügen wir dem Modell den Algorithmus „Feldrechner“ (unter *Vektortabelle*) hinzu.

Im Fenster *Feldrechner* ist neben dem Namen des Verarbeitungsschritts (*Description*; möglichst klare Funktionsbezeichnung wählen, z. B. „Schritt 1: DE-Nr“) zunächst der *Eingabelayer* von Bedeutung. Ein Klick auf den *123-Button* links erlaubt die Auswahl aus verschiedenen Eingabekategorien. Die Kategorie „Wert“ bezieht sich auf außerhalb des Modells vorhandene Eingabelayer (also i. d. R. im QGIS-Projekt geladene Vektorlayer), bei der Auswahl „Vorberechneter Wert“ kann die Eingabe per Ausdruck (Skript) definiert werden. „Algorithmusausgaben“ beziehen sich auf die Zwischenergebnisse einzelner Verarbeitungsschritte des Modells; diese Option wird später im Prozesse benötigt. Der erste Verarbeitungsschritt dieses Modells wird dagegen mit einer „Modelleingabe“, eben der in Schritt 1 definierten, gefüttert.

Die Angaben zu *Feldname*, *Ergebnisfeldtyp*, *-länge* und *-genauigkeit* beziehen sich auf die Spalte in der Attributtabelle des Ergebnislayers, in die das Ergebnis dieses Verarbeitungsschrittes geschrieben werden soll.

Im Reiter *Ausdruck* bzw. *Funktionseditor* ist der eigentliche Verarbeitungsschritt anzugeben, im Beispiel eben der Reguläre Ausdruck. Es ist zunächst die „Nummer der Dokumentationseinheit“ aus der Spalte „Layer“ der Eingabe zu extrahieren (Tab. 3), daher lautet der Regex-Ausdruck: `regexp_substr(„layer“;(\d{2,3}).DE)`. Es wird der von oben bekannte Befehl `regexp_substr` genutzt, welcher den Teil der Zeichenkette liefert, die dem gegebenen Regulären Ausdruck entspricht. Achtung: der *Feldrechner* im *Modelleditor* zeigt, anders als sein normales Pendant, keine Vorschau des Ergebnisses an – schließlich ist der konkrete *Eingabelayer*, auf die sich der Ausdruck beziehen könnte, noch nicht geladen!

Schließlich ist zu definieren, wie das Ergebnis ausgegeben werden soll (*Berechnet*, am unteren Ende der Maske). Soll das Ergebnis im Modell weiter als Eingabe Verwendung finden, ist an dieser Stelle nichts weiter zu tun. Handelt es sich um das Endergebnis, ist in der entsprechenden Leiste ein frei wählbarer Name zu vergeben.

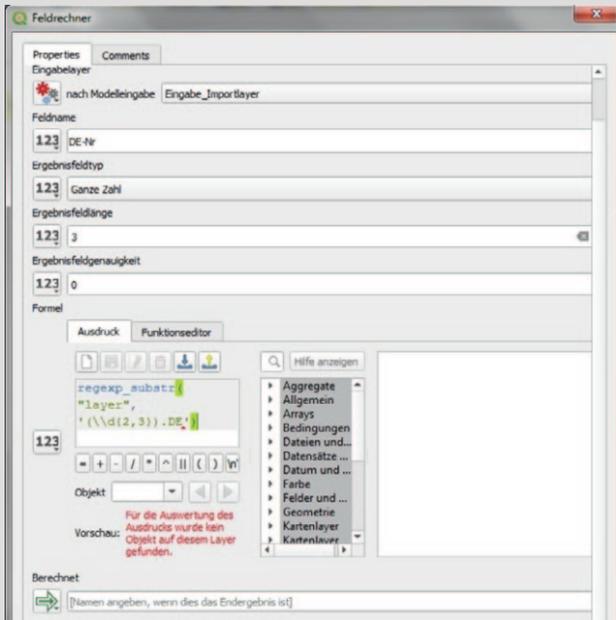


Abb. 4: Zur Definition von Verarbeitungsschritten in der Modellerstellung können Regex-Ausdrücke eingesetzt werden (QGIS-Feldrechner).

Nach einem Klick auf *OK* schließt sich das Fenster und ein neues, nun weißes, Kästchen erscheint. Dieses ist mit der oben definierten Eingabe über den Knoten „Ein“ verbunden (vgl. Abb. 1)⁵. Der Knoten „Aus“ verbleibt zunächst ohne Anschluss.

3. Für den nächsten Schritt im Beispiel, die „Extraktion des Codes bzw. der Objektklasse“ aus der Layer-Zeichenkette, kann bis auf ein Detail zunächst alles wie im obigen Schritt beschrieben mit dem passenden Regex-Befehl wiederholt werden. Anstatt jedoch als *Eingabelayer* eine *Modelleingabe* zu definieren, muss nun auf die Ausgabe des vorherigen Verarbeitungsschritts, also eine *Algorithmusausgabe*, zugegriffen werden – im konkreten Fall auf den Wert „Berechnet“ aus dem Algorithmus „Schritt 1: DE-Nr“.

Nach Schließen dieses Fensters erscheint ein weiteres weißes Kästchen mit dem Algorithmus in der Modellgrafik. Es speist sich über den Knoten „Ein“ mit der Ausgabe von Schritt 1 (vgl. Abb. 1).

Für die folgenden Einträge des Beispiels „Befund-Nummer“, „Lagebezug (Oberkante, Unterkante, Zwischenniveau)“ und „Beschreibung Objektklasse“ wird alles wie im vorangegangenen Schritt wiederholt.

4. Beim finalen Verarbeitungsschritt, im Beispiel der Extraktion der Kommentarspalte, muss eine **Ausgabe** definiert werden. Das bedeutet lediglich, dass unter *Berechnet* ein Name für den Ausgabebayer anzugeben ist, etwa „Verarbeiteter Layer“. Diese Ausgabe erzeugt ein eigenes grünes Kästchen im Modellplan (vgl. Abb. 1).

Das Modell ist damit einsatzfähig, sollte allerdings zunächst einmal getestet werden. *Modell ausführen* (F5) öffnet eine Eingabemaske, welche die eben für das Modell definierten Eingaben (hier: Eingabelayer) abfragt. Trägt man bei der Ausgabe nichts ein, wird standardmäßig ein Temporärlayer erzeugt. Alternativ kann das Ergebnis direkt in ein Vektorformat gespeichert werden. Beim Durchlauf des Modells lässt sich im Protokoll nun der Ablauf beobachten:

⁵ Eine Eingabe kann auch mit verschiedenen Verarbeitungsschritten verbunden werden (1:n-Beziehung).

Zunächst wird die Eingabedatei geladen, dann werden nacheinander alle definierten Arbeitsschritte ausgeführt, bevor das Resultat geladen wird.

Der nach Ablauf des Tests neu erzeugte (Temporär)Layer besitzt die Geometrien des Eingabelayers. Seine Attributtabelle ist jedoch in unserem Beispiel um drei neue Spalten erweitert. Sollten Fehler in der Verarbeitung auftreten, wird das entsprechend im Protokoll des Modells vermerkt.

Nach erfolgreichem Test sollte das Modell unter *Modelleigenschaften* im Entwurfswindow mit einem griffigen Namen benannt und abgespeichert werden. Es wird dadurch als Kopie auf dem Rechner abgelegt – unabhängig vom QGIS-Projekt, in dem es erstellt worden ist. Künftig wird es damit als Werkzeug in den Verarbeitungswerkzeugen von QGIS-Projekten des Rechners unter den *Modellen* zur Verfügung stehen. Über das Kopieren der Modelldatei aus dem QGIS-Ordnerverzeichnis kann das Modell auch auf andere Rechner übertragen werden. Bei Bedarf kann es auch im Projekt selbst hinterlegt werden, etwa wenn das Projekt einschließlich der neuen Werkzeuge an Arbeits- oder Projektpartner weitergeleitet werden soll.

Über den *Modellentwurf* lassen sich im Beispiel des Konvertierungsprojektes auch weitere Bearbeitungsschritte in einen halbautomatisierten Prozess umsetzen. Dies können etwa die Reparatur von Geometrien oder die Polygonerzeugung mit anschließender Übertragung der Attribute auf die erzeugten Polygone sein.

Fazit

Mit den vorgestellten Funktionen der Regex-Befehle und der Grafischen Modellierung ist es mit Bordmitteln von QGIS möglich, passgenau auf die eigene Datenstruktur abgestimmte Modellwerkzeuge zu kreieren, die eine erhebliche Hilfe bei vielen Bearbeitungsschritten an digitalen Grabungsdokumentationen bieten können.

Je umfangreicher die Grabungsdaten, desto mehr wird es sich lohnen, sich durch diese strukturelle Vorarbeit zu tüfteln – ohne Scheu vor ungewohnten Ausdrücken und grafischen Verkettungen.

Links

Überblick über die QGIS QRegularExpression (PCRE):
https://en.wikipedia.org/wiki/Perl-Compatible_Regular_Expressions

Deutschsprachiger Überblick zu Regulären Ausdrücken:
https://dewiki.de/Lexikon/Regul%C3%A4rer_Ausdruck

Englischsprachige Einführung zu Regulären Ausdrücken:
<https://www.regular-expressions.info/>

Zum Austesten der Regulären Ausdrücke:
<https://regex101.com/>
<https://regexpr.com/>
 (Spezifikation jeweils auf PCRE umschalten!)