

Integration von Python in T_EX am Beispiel von Katalogeinträgen

Lukas C. Bossert, Uwe Ziegenhagen

Viele Dissertationen in der Archäologie enthalten am Ende der Arbeit einen Katalog, in dem die untersuchten Daten in einem bestimmten System aufgeschlüsselt präsentiert werden. Ein solcher Katalog kann aus Bildern, Bohrproben, Architekturelementen etc. bestehen.

Eine händische Erstellung der einzelnen Einträge bspw. über `\section` oder `\subsection` und anschließend in einer `items`-Umgebung ist nicht effizient, fehleranfällig und nur bei wenigen Katalogeinträgen einsetzbar. Es muss zudem berücksichtigt werden, dass die vorgegebenen Kategorien nicht für jeden Katalogeintrag passend sind, sodass Kategorien leer bleiben und dann im Katalogeintrag nicht auftauchen sollen. Dabei soll der Code des Katalogeintrags möglichst viel redundante Tipp-Arbeit abnehmen, wie sie bspw. bei Maßeinheiten vorkommt. Darüber hinaus sollen alle Einträge immer gleich formatiert sein und ihr Aussehen global verändert werden können.

Nachdem eine Lösung gefunden wurde, die allen bisher genannten Anforderungen entspricht (siehe unten), sollten in einer Kategorie alle Seitenzahlen enthalten sein, auf denen im Haupttext auf den Katalogeintrag verwiesen wird. Ein Lösungsansatz sah die Nutzung von `glossaries` vor: Dafür musste allerdings für jeden Katalogeintrag ein eigenes Glossar angelegt werden, was nicht nur viel händische Arbeit bedeutet, sondern auch die Zählerkapazitäten von T_EX überforderte.

Es musste also eine andere Lösung her und ich (Lukas) erinnerte mich an einen Beitrag, den Uwe Ziegenhagen in `dtk` 1/2015 vorgestellt hatte: Damals ging es darum, bei einem Werkkatalog die Erwähnung des Stückes im Haupttext anzugeben. Prima, genau das, was ich gesucht habe und den damals verwendeten Code fand ich auf seiner Webseite.¹ Allerdings sah ich nun, dass die Aufgabe damals darin bestand, nur *eine* Erwähnung im Haupttext anzugeben, was mit `label` und `ref` bewerkstelligt werden konnte. Ich brauche jedoch alle Erwähnungen im Haupttext. Meine E-Mail-Anfrage bei Uwe nach einer Möglichkeit, mehr als nur eine Erwähnung anzeigen zu lassen, beantwortete er mit dem Vorschlag, mittels Python das Problem zu lösen.

Um gemeinsam an dem hybriden Konstrukt von X_YL^AT_EX und Python arbeiten zu können, wurde ein Repository auf `github`² erstellt, einer Plattform, die unabhängig von der Programmiersprache einen exzellenten Austausch und eine detaillierte Versionskontrolle ermöglicht.

¹ <http://uweziegenhagen.de/?p=3020>

² <https://github.com/LukasCBossert/DTK-TeX-Python>

Während des Schreibprozesses an diesem Artikel und am Python-Code haben wir Schützenhilfe von Herbert Voß erhalten, der uns eine T_EX-immanente Lösung erarbeitet hat, die als weitere Möglichkeit an unserem konkreten Beispiel präsentiert wird.

Katalogeintrag

Als Aufgabe soll ein Katalog zu den Häusern in Pompeji angelegt werden, der Auskunft über den Namen des Hauses, dessen Verortung und Grundstücksgröße geben soll. Zudem soll eine kurze Beschreibung enthalten sein und die Angabe über die Innenausstattung, die wiederum aufgeschlüsselt werden kann auf die Untergruppen Mosaik, Wandgemälde und Skulptur. Zum Schluss soll im Katalogeintrag angezeigt werden, auf welchen Seiten des Fließtextes das Haus genannt wird.

Die gefundene Lösung für die Umsetzung der Anforderungen funktioniert mit Hilfe des `keyval` Pakets.³ Dafür werden in der Präambel verschiedene Schlüssel (`keys`) definiert. In der Grundversion sieht die Definition eines Eintrages wie folgt aus:⁴

```
\define@key{family}{key}{#1}
```

Für das konkrete Beispiel wird die `key`-Familie (`family`) mit `catalogue` angegeben, der Schlüssel (`key`), was einer Kategorie im Katalog entspricht, als `house` bezeichnet und als Resultat soll der Wert im Makro `\KVhouse` gespeichert werden:

```
\define@key{catalogue}{house}{\def\KVhouse{#1}}
```

Dem Beispiel entsprechend können alle Kategorien als Schlüssel angelegt werden:

```
1 \makeatletter
2 \define@key{catalogue}{house}{\def\KVhouse{#1}}
3 \define@key{catalogue}{label}{\def\KVlabel{#1}}
4 \define@key{catalogue}{description}{\def\KVdescription{#1}}
5 \define@key{catalogue}{location}{\def\KVlocation{#1}}
6 \define@key{catalogue}{size}{\def\KVsize{#1}}
7 \define@key{catalogue}{interior}{\def\KVinterior{#1}}
8 \define@key{catalogue}{interiorM}{\def\KVinteriorM{#1}}
9 \define@key{catalogue}{interiorW}{\def\KVinteriorW{#1}}
10 \define@key{catalogue}{interiorS}{\def\KVinteriorS{#1}}
11 \makeatother
```

Das Aussehen eines Katalogeintrages wird separat mit dem Makro `\newcommand\catalogueentry[1]...` definiert: Darin wird zunächst festgelegt, dass eine neue

³ Basierend auf der Idee vorgestellt auf: <http://tex.stackexchange.com/a/254336/98739>

⁴ Vgl. <http://www.tug.org/tugboat/tb30-1/tb94wright-keyval.pdf>

Gruppe beginnt (`\begingroup`), sodass es zu keinen Problemen mit den jeweils definierten Schlüsseln kommt, da diese für jeden Katalogeintrag neu definiert werden. Dann sollen die Einträge im Flattersatz gesetzt werden (`\RaggedRight`) und schließlich die Angabe, welche Schlüsselfamilie (hier `catalogue`) auszulesen ist.

```
1 \newcommand\catalogueentry[1]{%
2 \begingroup
3 \RaggedRight
4 \setkeys{catalogue}{#1}
5 ...
```

Es folgt in der Definition die Verarbeitung der einzelnen Schlüssel: Da nur die Ausgabe einer Kategorie erfolgen soll, wenn diese auch mit Informationen versehen ist, wird dies über die Abfrage `\ifdef` erledigt. Der Inhalt der Kategorie `house` wird als `\section`-Titel verwendet und, wenn vorhanden, mit einem `\label` versehen. Die weiteren Kategorien sollen in einer `labeling`-Umgebung aufgelistet werden. Die Definition wird mit `\endgroup` geschlossen.

```
1 ...
2 \ifdef{\KVhouse}{\section{\KVhouse
3   \ifdef{\KVlabel}{\label{\KVlabel}}{}}{}}
4 \begin{labeling}{Beschreibung}
5   \ifdef{\KVdescription}{\item[Beschreibung] \KVdescription}{}
6   \ifdef{\KVlocation}{\item[Verortung] \KVlocation}{}
7   \ifdef{\KVinterior}{\item[Ausstattung] \KVinterior
8     \begin{labeling}{Wandgemälde}
9       \ifdef{\KVinteriorM}{\item[Mosaike:] \KVinteriorM}{}
10      \ifdef{\KVinteriorW}{\item[Wandgemälde:] \KVinteriorW}{}
11      \ifdef{\KVinteriorS}{\item[Statuen:] \KVinteriorS}{}
12    \end{labeling}
13    }{}
14   \ifdef{\KVsize}{\item[Größe] \SI{\KVsize}{\meter\squared}}{}}
15 \end{labeling}
16 \endgroup
17 }
```

Mit dieser Einstellung lassen sich die Katalogeinträge schon sehr gut darstellen. Im Fließtext des Hauptdokuments kann man an gewünschter Stelle mit `\catalogueentry` einen Katalogeintrag eintragen. So wird aus den folgenden Angaben

```
\catalogueentry{%
  house={Haus des M. Fabius Rufus},
  label={haus:M-Fabius-Rufus},
  size={172},
  description={Haus besteht aus mehreren Einzelgebäuden.},
```

```

location={Regio VII, Insula 16, Eingang 17--22.},
interior={Reicher Fundkomplex.},
interiorM={S/W-Mosaik},
interiorW={Dionysius mit einer Mänade, Narzissus und ein Cupido, Hercules und
↪Deinira, etc.},
interiorS={Bronzene Statue eines Epheben},
}

```

ein Katalogeintrag, der so aussieht:

Haus des M. Fabius Rufus

Beschreibung	Haus besteht aus mehreren Einzelgebäuden.
Verortung	Regio VII, Insula 16, Eingang 17--22.
Ausstattung	Reicher Fundkomplex.
	Mosaik: S/W-Mosaik
	Wandgemälde: Dionysius mit einer Mänade, Narzissus und ein Cupido, Hercules und Deinira, etc.
	Statuen: Bronzene Statue eines Epheben
Größe	172 m ²

Wie man bei diesem Beispiel sieht, ist die Reihenfolge, in der die Kategorien angegeben werden, irrelevant, da die Definition in der Präambel entscheidend ist. Der Wert bei `size` wird intern sogleich an das vordefinierte `\SI`-Makro mit entsprechender Einheit (m²) übergeben. Ähnlich kann auch mit anderen Angaben verfahren werden (bspw. bei Abbildungen können die `\label` an ein vordefiniertes Macro `\cref` übergeben werden).

Bei dem oben beschriebenen Katalogeintrag zum ›Haus des M. Fabius Rufus‹ sind alle Kategorien ausgefüllt. Wenn eine Kategorie nicht ausgefüllt ist, wird sie nicht ausgegeben. Allerdings besteht die Kategorie `interior` [Ausstattung] zusätzlich aus drei Unterkategorien (`interiorM` [Mosaik], `interiorW` [Wandgemälde], `interiorS` [Statuen]). Diese sollen auch dann angezeigt werden, wenn `interior` selbst nicht definiert ist.

Ein solcher Fall tritt beim ›Haus des Wilden Ebers‹ auf. Im Fließtext ist der Katalogeintrag wie folgt ausgefüllt:

```

\catalogueentry{%
  house={Haus des Wilden Ebers},
  label={Haus-des-Wilden-Ebers2},

```

```
size={54},
description={Renovierung nach Erdbeben 62\,n.\,Chr.},
location={Regio VII, Insula 4, Eingang 48, 43},
interiorM={S/W-Mosaik},
interiorW={Venus, Leda und der Schwan, Ariadne und Theseus},
}
```

So wird daraus:

Haus des Wilden Ebers

Beschreibung	Renovierung nach Erdbeben 62 n. Chr.
Verortung	Regio VII, Insula 4, Eingang 48, 43
Ausstattung	Mosaik: S/W-Mosaik Wandgemälde: Venus, Leda und der Schwan, Ariadne und Theseus
Größe	54 m ²

Um zu diesem Ergebnis zu kommen, ist es notwendig, mit Booleschen Operatoren zu arbeiten. Dafür müssen bei den key-Definitionen drei Operatoren eingeführt werden:

```
1 \newbool{@KVinteriorM}%Mosaik
2 \newbool{@KVinteriorW}%Wandgemälde
3 \newbool{@KVinteriorS}%Statue
```

Diese Booleschen Operatoren werden bei der Definition der einzelnen Einträge eingebaut und als true gesetzt, wenn dieser Eintrag mit Informationen versehen wird. Konkret sieht die Modifikation so aus:

```
1 \define@key{catalogue}{interiorM}{\def\KVinteriorM{#1}\booltrue{@KVinteriorM}}
2 \define@key{catalogue}{interiorW}{\def\KVinteriorW{#1}\booltrue{@KVinteriorW}}
3 \define@key{catalogue}{interiorS}{\def\KVinteriorS{#1}\booltrue{@KVinteriorS}}
```

Zudem müssen die Booleschen Operatoren bei der Ausgabe der Katalogeinträge eingesetzt werden, sodass geprüft wird (`\ifbool{expr}`), ob einer oder mehrere der Operatoren auf true gesetzt ist.

```
1 \ifdef{\KVinterior}{%
2   \item[Ausstattung] \KVinterior
3   \ifbool{bool{@KVinteriorM}}
4     or bool{@KVinteriorW}
5     or bool{@KVinteriorS}}{%
```

```

6      \begin{labeling}{Wandgemälde}
7          \ifdef{\KVinteriorM}{\item[Mosaike] \KVinteriorM}{}
8          \ifdef{\KVinteriorW}{\item[Wandgemälde] \KVinteriorW}{}
9          \ifdef{\KVinteriorS}{\item[Statuen] \KVinteriorS}{}
10         \end{labeling}
11     }{}
12     {\ifboolexpr{bool{@KVinteriorM}
13         or bool{@KVinteriorW}
14         or bool{@KVinteriorS}}{%
15         \item[Ausstattung]%
16         \begin{labeling}{Wandgemälde}
17             \ifdef{\KVinteriorM}{\item[Mosaike] \KVinteriorM}{}
18             \ifdef{\KVinteriorW}{\item[Wandgemälde] \KVinteriorW}{}
19             \ifdef{\KVinteriorS}{\item[Statuen] \KVinteriorS}{}
20         \end{labeling}
21     }{}

```

Mit dieser modifizierten Ergänzung der Katalogeintragsdefinition werden die Unterkategorien von `interior` ausgegeben, auch wenn `interior` selbst nicht definiert ist.

Was noch fehlt ist die Ausgabe der Seiten, auf denen der Katalogeintrag erwähnt wird. Dies erfolgt zunächst über Python, anschließend wird die T_EX-Variante vorgestellt. Lösungen über andere Programmiersprachen sind auch denkbar.

Problemlösung mit Python

Python ist eine interpretierte Programmiersprache, die den Python-Neuling durch weitestgehende Nicht-Nutzung von Klammern verwirrt. Stattdessen nutzt Python Tabulator-Einrückungen, um zusammengehörige Code-Teile zu kennzeichnen. Was erst einmal gewöhnungsbedürftig klingt, hat in der Praxis Vorteile, denn selbst fremder Python-Code ist erstaunlich gut lesbar. Im aktuellen Anwendungsfall beschränken wir uns jedoch nicht auf Python, sondern nutzen mit `pandas` eine Erweiterung zur Datenanalyse. `pandas`, ursprünglich zur Analyse von Finanzdaten entwickelt, erleichtert die Arbeit mit strukturierten Daten aller Art. Das Einlesen, Filtern und Aggregieren von Daten jeder Art geht damit erstaunlich einfach, dem interessierten Leser sei eine nähere Beschäftigung mit `pandas` daher ans Herz gelegt.

Da Python auf Basis des L^AT_EX-Quelltextes keine Information hat, auf welchen Seiten ein Stichwort erwähnt wird, muss dies auf L^AT_EX-Ebene weggeschrieben werden.

Dazu bedienen wir uns des `\newwrite` Befehls, der eine neue Ausgabedatei öffnet. In der Präambel des T_EX-Dokuments muss aufgeführt werden, dass eine neue Datei geschrieben wird, die den gleichen Namen hat, aber die Endung `.sti`. In diese Datei

werden alle Stichwörter und die dazu gehörende Seitenzahl geschrieben, die mit dem Makro `\eintrag{Stichwort}` definiert sind.

```
\newwrite\myfile
\immediate\openout\myfile=\jobname.sti
\newcommand{\eintrag}[1]{\immediate\write\myfile{#1:\thepage}}
```

Für die spätere gesammelte Anzeige der Stichwörter und die entsprechende Seitenzahl müssen zwei weitere Befehle definiert werden:

```
\newcommand{\ausgabe}[2]{#1 was referenced on page(s): #2}
\newcommand{\stichworttabelle}{\IfFileExists{\jobname.tab}{\input{\jobname.tab
↵}}{\Tab file not found}}
```

Die erste sorgt dafür, dass bei dem `\ausgabe{Stichwort}{ }` Makro das Stichwort sowie der Text *was referenced on page(s):* mit den Seitenzahlen gesetzt wird. Der zweite Befehl gibt die Stichwörter und die Seitenzahlen als Tabelle aus.

Zusammengesetzt sieht der Text in der Präambel wie folgt aus, wobei die Ausgabe des `\ausgabe` Makros zugleich für das konkrete Beispiel abgeändert wurde.

```
1 \newwrite\myfile
2 \immediate\openout\myfile=\jobname.sti
3 \newcommand{\eintrag}[1]{\immediate\write\myfile{#1:\thepage}}
4 \newcommand{\ausgabe}[2]{#2}
5 \newcommand{\stichworttabelle}{\IfFileExists{\jobname.tab}{\input{\jobname.tab
↵}}{\Tab file not found}}
```

Schauen wir uns nun den Python-Code dazu an:

Integration

Im L^AT_EX-Dokument muss nun an allen Stellen, an denen das Stichwort auftaucht oder der Rückverweis stattfinden soll, der Befehl `\eintrag{Stichwort}` gesetzt werden. Wenn alle Stichwörter gesetzt sind, kann man die Art der Ausgabe wählen. Entweder werden alle Seitenzahlen zu einem Stichwort angezeigt, dies erfolgt über `\ausgabe{Stichwort}{ }` oder Stichwörter und Seiten werden mit `\stichworttabelle` tabellarisch aufgelistet.

Für unser konkretes Beispiel der Katalogeinträge bedeutet dies, dass das Makro `\ausgabe` in die Katalogeintragdefinition integriert werden muss.

```
1 \define@key{catalogue}{literature}{\def\KVliterature{#1}}
```

Damit wird der Wert, der im `\catalogueentry` unter `literature` angegeben ist (hier *Haus*), an den Befehl `\ausgabe` übergeben:

```

1 \catalogueentry{%
2   literature={\ausgabe{Haus}{}},
3 }

```

Das bedeutet selbstverständlich, dass im Fließtext entsprechend `\eintrag{Haus}` gesetzt werden muss.

Wenn etwas in den Katalogeintrag `literature` steht, dann wird dieser Eintrag übernommen, ansonsten wird geschaut, ob das `label` definiert ist und wenn dies nicht der Fall ist, bleibt dieser Katalogeintrag undefiniert und es findet auch keine Weiterverarbeitung statt.

Für die Darstellung des Katalogeintrages muss ebenfalls eine Neudefinition eingeführt werden:

```

\ifdef{\KVliterature}{\item[Erwähnungen] \KVliterature}{}

```

Ein vollständiger Katalogeintrag:

```

1 \catalogueentry{%
2   house={Haus des M. Fabius Rufus},
3   label={haus:M-Fabius-Rufus},
4   size={172},
5   description={Haus besteht aus mehreren Einzelgebäuden.},
6   location={Regio VII, Insula 16, Eingang 17--22.},
7   interior={Reicher Fundkomplex.},
8   literature={\ausgabe{Haus-Fabius}{}},
9 }

```

Nun wird *Haus* an `\ausgabe` übergeben. Im Folgenden Fall ist dies der Wert vom Katalogeintrag `label`:

```

1 \catalogueentry{%
2   house={Haus des Wilden Ebers},
3   label={Haus-des-Wilden-Ebers},
4   size={54},
5   description={Renovierung nach Erdbeben 62\,n.\,Chr.},
6   location={Regio VII, Insula 4, Eingang 48, 43},
7 }

```

Während im ersten Fall im Fließtext an gewünschter Stelle `\eintrag{Haus-Fabius}` gesetzt werden muss, gilt für die zweite Variante `\eintrag{Haus-des-Wilden-Ebers}`.
4, 9

Haus des M. Fabius Rufus

Beschreibung Haus besteht aus mehreren Einzelgebäuden.
Verortung Regio VII, Insula 16, Eingang 17–22.
Ausstattung Reicher Fundkomplex.
Größe 172 m²
Erwähnungen S.

Haus des Wilden Ebers

Beschreibung Renovierung nach Erdbeben 62 n. Chr.
Verortung Regio VII, Insula 4, Eingang 48, 43
Größe 54 m²
Erwähnungen S.

Sobald diese Vorarbeiten getan sind, alle Stichwörter entsprechend mit `\eintrag` gesetzt sind, dann muss einmal (bspw. mit X_Y-L^AT_EX) kompiliert werden, anschließend der Python-Code im entsprechenden externen Programm ausgeführt und schließlich zurück im T_EX-Editor wiederum einmal kompiliert werden.

	Seite
Stichwort	
Haus-Fabius	8
Haus-Rufus	4, 8
Haus-des-Wilden-Ebers	4, 5, 8

4, 9

4, 5, 8

Problemlösung mit T_EX

Die andere, T_EX-immanente Lösung arbeitet über `\labelcref`. Das heißt, dass man im Gegensatz zur Python-Variante die Rückverweise nicht manuell setzen kann, sondern sie werden automatisch über die im Fließtext gesetzten `\labelcref` erstellt. Dafür muss der Code des Katalogeintrages `literature` geändert werden:

```

1 \ifKVmark \item[Erwähnungen] S.\~\expandafter\csname\KVlabel-pages\endcsname \
  ↳\fi
2 ...
3 \define@boolkey{catalogue}{KV}{literature}{true}{}

```

Zudem muss das Makro `\labelcref` bearbeitet werden, sodass \TeX die gesetzten Einträge auslesen kann.

```

1 \let\LabelCref\labelcref
2 \renewcommand\labelcref[1]{\expandafter\label@cref#1,,\@nil}
3 \def\label@cref#1,#2,#3\@nil{%
4   \expandafter\ifx\csname#1-pages\endcsname\relax % schon definiert??
5     \expandafter\xdef\csname#1-pages\endcsname{\arabic{page}}%
6   \else
7     \expandafter\xdef\csname#1-pages\endcsname{\csname#1-pages\endcsname, \
  ↳\arabic{page}}%
8   \fi
9   \LabelCref{#1}%
10  \ifx\relax#2\relax % kein label mehr da
11    \def\next{}%
12  \else
13    \def\next{\label@cref#2,#3\@nil}%
14  \fi
15  \next}

```

Bei den Katalogeinträgen reicht es, am Ende `literature` einzutragen:

```

1 \catalogueentry{%
2   house={Haus des M. Fabius Rufus},
3   label={haus:M-Fabius-Rufus},
4   size={172},
5   description={Haus besteht aus mehreren Einzelgebäuden.},
6   location={Regio VII, Insula 16, Eingang 17--22.},
7   interior={Reicher Fundkomplex.},
8   literature,
9 }

```