

SiGMaMat, SIS-Visualizer und KonSys

Die Zukunft der Konstanzer Sacherschließung

Ralph Hafner

SiGMaMat, SIS-Visualizer und KonSys sind erstens ein wichtiger Beitrag zur Bibliotheks-KryptoAkronymoManie¹ und zweitens drei Bausteine auf dem Weg der Konstanzer Sacherschließung hin zu einer Vernetzung mit anderen Sacherschließungssystemen.

Der SIS-Visualizer und der SiGMaMat sind Werkzeuge, die dazu dienen, die etwas unsaubereren „Altdaten“ des bisherigen Werkzeugs SIS², mit dem die Konstanzer Systematik derzeit gepflegt und durchsucht werden kann, in interoperable Daten umzuwandeln. Eine Systematik ist dann interoperabel, wenn ihre Strukturen klare Hierarchien ohne Widersprüche enthalten und ihre Verbalisierungen (Schlagwörter) eindeutig identifizierbar sind. Das Gesamtziel, das wir verfolgen und zu dem diese beiden Werkzeuge die ersten Schritte darstellen, ist, die Sacherschließung teilautomatisieren zu können, d. h. aus von anderen Einrichtungen bereits geleisteter Sacherschließung - z. B. verbaler Sacherschließung nach der GND³ oder klassifikatorischer Sacherschließung nach DDC⁴ oder RVK⁵ - Notationen für die Konstanzer Systematik ableiten zu können⁶.

SIS-Visualizer

Der SIS-Visualizer ist ein in der Bibliothek der Universität Konstanz bereits 2013 entwickeltes Tool zur Visuali-

sierung des hierarchischen Aufbaus der Konstanzer Systematik⁷.

Der SIS-Visualizer erlaubt eine hierarchische Navigation in der Konstanzer Systematik und bietet damit einen übersichtlichen Einblick in ihren Aufbau. Man kann sich von der obersten

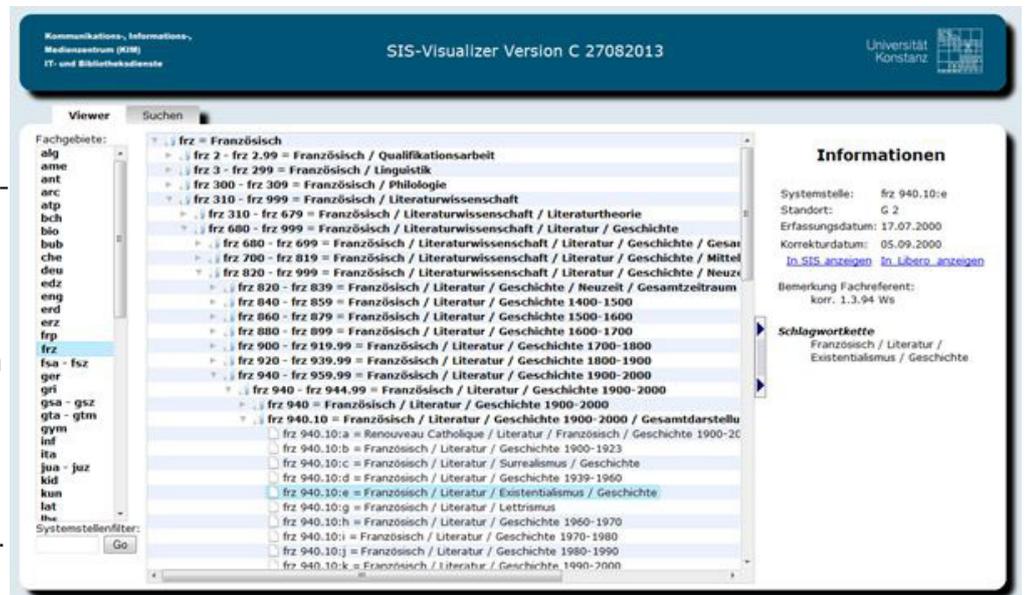


Abbildung 1: SIS-Visualizer: Browsing in der Systematik

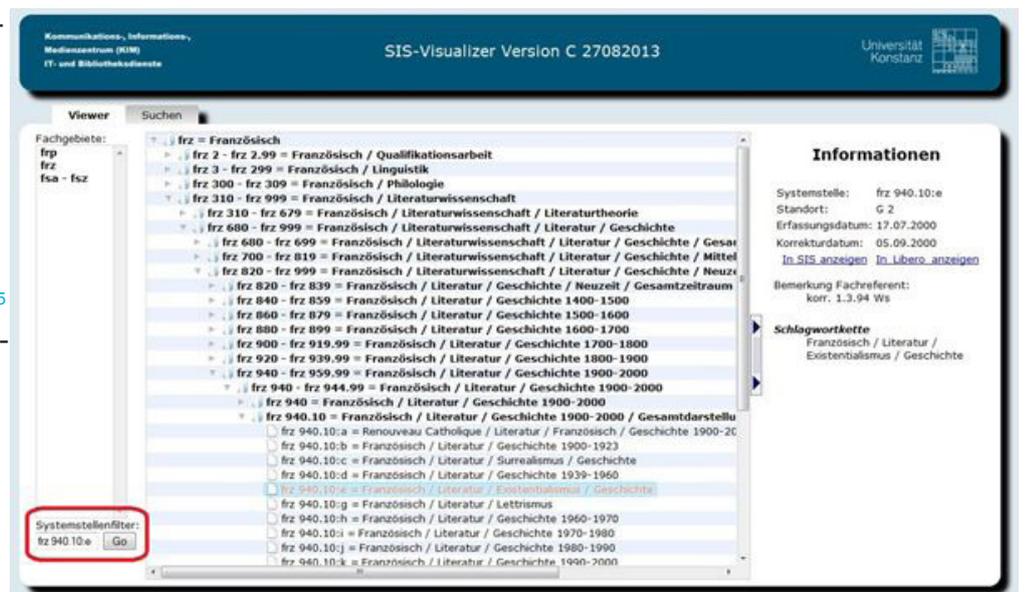


Abbildung 2: SIS-Visualizer: Einstieg in die Systematik über den Systemstellenfilter

Ebene der Systematik, einem Fach (beispielsweise spr = Linguistik (= Sprachwissenschaft)), nach unten durchklicken zu seinen Teilgebieten (spr 140 - spr 174.99 = Morphologie <Linguistik>) und weiter zu einzelnen speziellen Themen (spr 173.10 = Wortbildung / Nomen)⁸. Oder man steigt durch Eingabe einer Systemstelle an einem gewünschten Punkt in die Systematik ein.⁹ Einen weiteren Einstieg in die Konstanzer Systematik ermöglicht die verbale Suche im SIS-Visualizer.

Der SIS-Visualizer zeigt aber auch falsche oder noch fehlende Hierarchiestrukturen auf, die in unserem aktuellen System zur Anzeige und Bearbeitung der Konstanzer Systematik, dem Schlagwort-Informations-System nur schwer zu erkennen waren. Damit hilft der SIS-Visualizer diese strukturellen Fehler in der Systematik zu beheben. Eine korrekte hierarchische Struktur ohne Widersprüche und ohne Lücken ist notwendig, um die Systematik mit anderen vergleichen zu können.

SIGMaMat

Der SiGMaMat ist eine Art Schlagwortwaschmaschine, die aus den unsauberen, mehrdeutigen Daten im Konstanzer Schlagwort-Informations-System (SIS) saubere, eindeutige Schlagwörter mit einer Referenz-ID aus einer Normdatei macht. Das Akronym steht für SIS-GND-Matching-Automat.

Das Ziel ist, künftig möglichst nur noch eindeutig dereferenzierte Begriffe in unserer Konstanzer Systematik zu haben.

Der SiGMaMat arbeitet wie folgt:

Er vergleicht die Zeichenketten der Begriffe aus der Konstanzer Systematik mit Zeichenketten aus den Normdateien GND und VIAF¹⁰. Unsere Hauptreferenzdatei ist die GND. Die GND liegt uns als RDF-Datendump¹¹ vor. Der Vergleich läuft über SPARQL-Anfragen. Beim Vergleich werden Normalisierungen vorgenommen. Wenn keine exakten Übereinstimmungen zwischen den Zeichenketten in den verschiedenen Systemen gefunden werden können, wird eine Fuzzy Search¹² durchgeführt. Wenn der SiGMaMat in der GND keine Referenz finden kann, startet er eine Anfrage über die SRU-Schnittstelle¹³ im VIAF. Bei dem Vergleich der Daten werden neben der Übereinstimmung

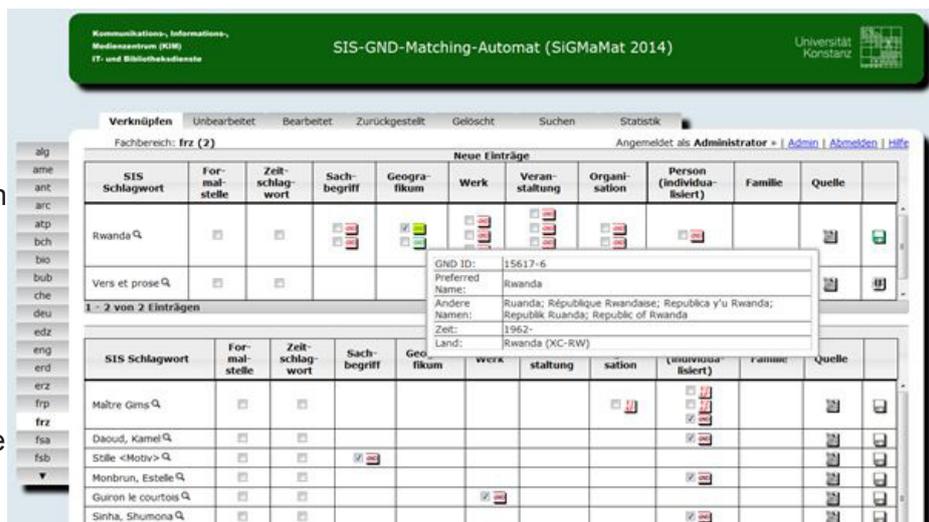


Abbildung 3: SiGMaMat: Verknüpfen der Konstanzer Schlagwörter mit Begriffen aus den Normdateien

der Zeichenkette auch die Übereinstimmung der Schlagwortkategorie überprüft. Weichen die Kategorien voneinander ab, ist das Schlagwort in der Konstanzer Systematik z. B. ein Sachschlagwort und in der GND ein Geographikum, dann mappt der SiGMaMat die Begriffe nicht. Werden hinreichend Kriterien der Übereinstimmung¹⁴ zwischen der Zeichenkette aus der Konstanzer Systematik und der Normdatei gefunden, dann wird das Konstanzer Schlagwort auf den Normbegriff gemappt.

Mithilfe des SiGMaMats konnten ca. 70 % unserer rund 180.000 Schlagwörter automatisch den Schlagwörtern der Deutschen Normdatei GND und der internationalen Normdatei VIAF eindeutig zugeordnet werden. Die restlichen 30 % (ca. 54.000 Schlagwörter) müssen von Hand gemappt werden.

Den zuständigen FachreferentInnen wurden im SiGMaMat ihre Fächer zugeordnet. Nach Anmeldung erhalten sie lesenden und schreibenden Zugriff auf die von ihnen betreuten Fächer und können dort die Verknüpfungen zwischen den Konstanzer Schlagwörtern, die nicht automatisch gemappt werden konnten, und den Schlagwörtern der Normdateien herstellen. Bei Bedarf konnten sich die FachreferentInnen von Hilfskräften unterstützen lassen.

KonSys

Ein weiterer Baustein des Projekts ist KonSys¹⁵, das neue Editor- und Recherchetool für die Konstanzer Systematik. KonSys wird also der Nachfolger von SIS. Die mithilfe von SIS-Visualizer und SiGMaMat bereinigten Daten

aus dem SIS werden nach KonSys migriert werden. In einem nächsten Schritt sollen dann noch die Schlagwortfolgen, die den Notationen der Systematik zugeordnet sind, bereinigt werden: In manchen Systematikzweigen wird in der Schlagwortfolge die komplette Hierarchie der Systematik bis hin zu einem speziellen Problem ausgedrückt, wobei zuweilen die Reihenfolge der Schlagwörter durcheinander geraten und ein Teil der Begriffe verloren gegangen ist; in anderen Systematikzweigen besteht die Schlagwortfolge nur aus ein bis zwei Schlagwörtern, die die Spezifizierung gegenüber der direkt hierarchisch darüber liegenden Ebene ausdrückt. Das bedeutet aktuell auch, dass man in den unterschiedlichen Fächern unterschiedlich suchen muss. So wird man in den Fächern mit den ein-bis-zwei-Begriffs-Schlagwortfolgen mit einer Suche mit mehreren Begriffen nicht fündig. Die bereinigte Version soll zunächst wieder Einheitlichkeit für alle Fächer herstellen. Sie soll zudem - angefangen beim Stamm des

Systematikzweigs, dem Fach - bis hin zu den Verästelungen der Systematik alle Begriffe enthalten, die der Spezifizierung bis zu dem Punkt der Systematik entspricht, der Übersichtlichkeit halber aber immer nur den letzten spezifizierenden Begriff anzeigen, da es sonst zum Teil sehr lange Schlagwortfolgen würden¹⁶. Der KonSys-Editor, das KonSys-Recherche-tool und der Algorithmus zur Bereinigung der Schlagwortfolgen werden derzeit programmiert, das Migrationsszenario ist in Vorbereitung.

Über den Fortgang des Projekts und wie der Analyzer Agent und der Recommender Agent die in anderen bibliothekarischen Einrichtungen bereits geleistete Sacherschließung analysieren, und aus diesen Daten dann Empfehlungen z.B. für eine Konstanzer Notation für ein bei uns neuerworbenes Buch machen und damit die Sacherschließungsarbeit im Fachreferat erleichtern, darüber wird an dieser Stelle wieder berichtet werden.

Literatur:

Hafner, Ralph; Schelling, Bernd: Automatisierung der Sacherschließung mit Semantic-Web-Technologie. In: o-bib. Das offene Bibliotheksjournal 2 (2015), S. 161–175.
Kasprzik, Anna: Projektbericht: Implementierung eines Hierarchisierungsalgorithmus' für die Konstanzer Systematik. <http://kops.uni-konstanz.de/handle/123456789/24166> (18.01.2017).

Fußnoten:

- 1 Ein in bibliothekarischen Einrichtungen verbreiteter Wahn, unverständliche Abkürzungen zu verwenden.
- 2 SIS ist das Schlagwort-Informationssystem für die Konstanzer Systematik. http://sis.ub.uni-konstanz.de/sis/stammsatz_verbal.html
- 3 GND = Gemeinsame Normdatei
- 4 DDC = Dewey-Dezimalklassifikation
- 5 RVK = Regensburger Verbundklassifikation
- 6 Eine ausführlichere Beschreibung des Konstanzer Projekts zur Teilautomatisierung der Sacherschließung s. Artikel: Hafner, Ralph; Schelling, Bernd: Automatisierung der Sacherschließung mit Semantic-Web-Technologie. In: o-bib. Das offene Bibliotheksjournal 2 (2015), S. 161–175
- 7 Näheres zur Entwicklung des SIS-Visualizers s.: Kasprzik, Anna: Projektbericht: Implementierung eines Hierarchisierungsalgorithmus' für die Konstanzer Systematik. <http://kops.uni-konstanz.de/handle/123456789/24166> (18.01.2017)
- 8 s. Abb. 1.
- 9 s. Abb. 2.
- 10 Neben GND und VIAF (= Virtual International Authority File) lassen wir auch andere Quellen wie Wikipedia und Fachlexika als Referenz zu.
- 11 RDF (engl.) = Resource Description Framework (System zur Beschreibung von Ressourcen)
Dump (engl.) = Auszug eines Speicherinhalts
- 12 Fuzzy Search = Unscharfe Suche
- 13 SRU = Search/Retrieve via URL
- 14 Kriterien für automatisches Mappen waren z.B. die hundertprozentige Übereinstimmung mit „preferred name“ oder „variant name“ aus der Normdatei, wenn gleichzeitig diese Zeichenkette nicht mehrfach in der Normdatei vorkam.
- 15 Noch ein schönes neues Akronym: KonSys = Konstanzer Systematik
- 16 Üblicherweise besteht die Schlagwortfolge der obersten Hierarchiestufe eines Systematikzweiges aus einem Schlagwort und mit jeder Hierarchieebene darunter kommt ein Schlagwort dazu. Auf Hierarchieebene 4 besteht die Schlagwortfolge also aus 4 Schlagwörtern.
Beispiel: Schlagwortfolge der 1. Ebene: Französisch.
Beispiel: Schlagwortfolge der 4. Ebene: Französisch / Literaturwissenschaft / Literaturtheorie / Poetik