

Gerd Simon

Wie finde ich einen Namen, einen Begriff oder ein Thema in der Sammlung von Archivalien und Rara im GIFT-Archiv?

Jede Sammlung dieser Art hat ihre eigene Geschichte. Das GIFT-Archiv hat seine Anfänge schon lange vor der Entstehung der >Gesellschaft für interdisziplinäre Forschung Tübingen e.V.< (GIFT 1996-2012), zunächst ab 1971 beschränkt auf den Themenbereich >Sprachwissenschaften im 3. Reich<, dann allmählich mehr oder weniger bedingt durch die Verhältnisse in den größeren Archiven (von mir früh beschrieben als „wie auf Opas Dachkammer“)- ausgeweitet auf andere Kultur- und Geschichtswissenschaften, v.a. einzelner Philologien wie die Literaturwissenschaften und ihre Forschungsgegenstände (z.B Dichter), sowie Philosophie (ab 1990 zentral zu Hans Vaihinger), schließlich auch einige Personen und Themen aus den MINT-Wissenschaften („Dt. Mathematik“, Rassenkunde) sowie, veranlasst durch manche Zufallsfunde, auch aus Jura und Medizin.

Zuerst erstellten wir Findmittel, v.a. alphabetisch geordnete und z.T mit Regesten oder auch Kopien versehene Zettelkästen, die sich noch heute, zusammengefasst in 2 Schränken mit 40 Zettelkästen (jeweils ca. 200 Zettel) im GIFT-Archiv Nehren finden. Dann machte mich Anfang der 90er Jahre das spätere GIFT-Mitglied George LEAMAN (heute Leiter des >Philosophy Documentation Centers< USA) auf das Internet aufmerksam. Von da ab sammelten wir Neues im PC. Erste Versuche, die Informationen der Zettelkästen in den PC zu bringen, ergab das niederschmetternde Hochrechnungsergebnis: dafür brauchen wir mindestens 30 Jahre. Die wären heute vorbei. Aber wir hätten vieles andere nicht machen können, zumal uns immer weniger Finanzen gewährt wurden.

Als Anfang unseres Jahrhunderts die Digitalisierung die direkte Übertragung von Kopien der Archivalien und Rara vorlagengetreu in unseren PC einigermaßen erschwinglich machte, leistete ich mir die 20.000 Euro für mehr als 200.000 Digitalisate. Zu diesem >Imperium< (wie ich es nannte) stellte ich einen groben Überblick (mittlerweile über 1250 Seiten mit durchschnittlich 20 Informationen) her. Ein Mitarbeiter, der mir heute leider nicht mehr zur Verfügung steht, erkannte sehr schnell, dass herkömmliche PC-interne Suchmaschinen wie copernic oder docfetcher den Bedürfnissen von Historikern nicht sonderlich gerecht werden, konstruierte eine Suchmaschine, die mehr leistete und die ich bis heute benutze.

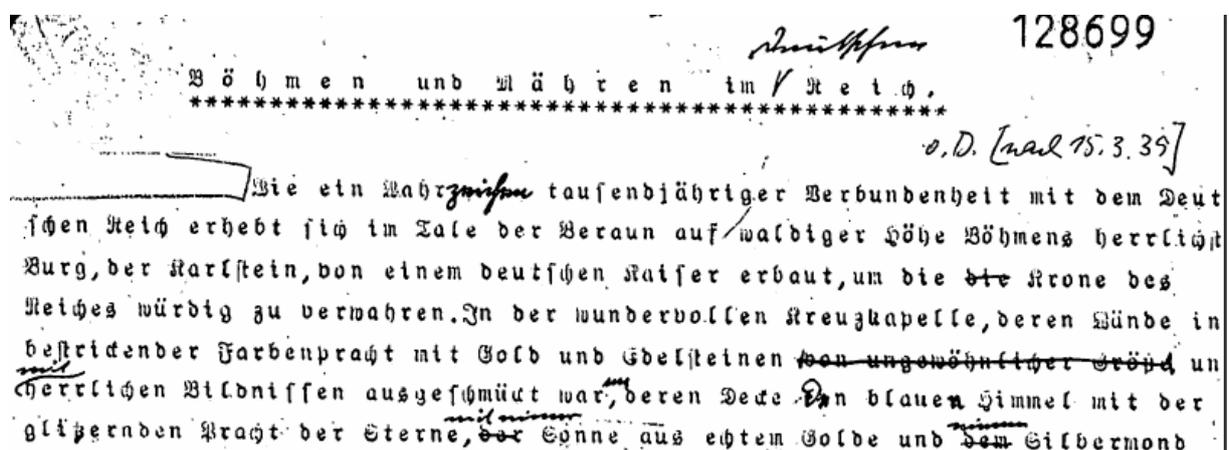
Von dieser Suchmaschine profitieren nicht nur ich und meine Mitarbeiterinnen, sondern auch alle Forscher, die mich bitten, in meinem „Imperium“ nach bestimmten Stichworten (Personen, Begriffen + Themen) zu suchen und ihnen die Funde zu mailen (Bisher übriggens kostenfrei).

Im Februar 2017 fing ich mir einen Trojaner, der große Teile meines „Imperiums“ vernichtete. Glücklicherweise hatte ich die Digitalisate kurz zuvor einem luxemburger Kollegen zukommen lassen, der sie mir danach wieder retournierte. Was ich ihm leider nicht zuschickte, waren die damals ca. 250 Chronologien die wir zu wichtigen Personen und Themen angefangen, manchmal zu einem gewissen Abschluss gebracht hatten. Das betraf auch den Überblick und nahezu alle Artikel, die wir auf Grund der Vorarbeiten in den Chronologien schon hergestellt hatten. Zumindest die Endfassungen waren verloren. Wir sind auch heute noch dabei, den Schaden zu reparieren. Ich schätze allerdings, dass wir bisher vier Fünftel in einen besseren Zustand als vorher gebracht haben.

So häufig wie mein „Imperium“ unterschätzt wird, wenn man gar nicht weiß, was es erfasst, verwandelt sich diese Unter- in eine Überschätzung, nach nur einem Einblick. Darum hier der Versuch darzulegen, welche Anfragen ich zu welchem Grade beantworten kann.

Jede pdf- oder doc-Datei hat eine eigene Suchmaschine. Diese ermittelt nur die Stellen in der jeweiligen Datei. Die 6-stellige Anzahl der Dateien in meinem „Imperium“ erfordern aber eine Suchmaschine, die mir sagt: in dieser oder jener Datei findet sich etwas zu dem gesuchten Namen, Begriff oder Thema.

Dabei setzt man fälschlicherweise voraus, dass Suchmaschinen alles lesen könnten. Historiker haben es z.B. mit unterschiedlichen Schriftarten zu tun. Eines ist klar: Handschriften wird eine Suchmaschine nie brauchbar entziffern können. Dabei muss es sich nicht einmal um so schwer entzifferbare Handschriften handeln wie die des Philosophen Hans Vaihinger. Noch im 3. Reich gab es neben der lateinischen Schrift, die heute am verbreitetsten ist, noch die Fraktur. Das betraf nicht nur Gedrucktes. Etliche Schreibmaschinen besetzte man damals mit Typen in Fraktur. Zum Beispiel hier ein Textanfang aus dem Nachlass des damals am meisten geschätzten Germanisten Erich Gierach:



Suchmaschinen sind nur brauchbar, wenn sie Texte auf bestimmte allgemeine Merkmale absuchen. Sie müssen also alle Varianten erfassen, die unterschiedliche Schriften einer Schriftart aufweisen. Je allgemeiner die Merkmale sind, desto mehr Funde ermöglichen sie, desto mehr Fehlfunde sind aber auch zu erwarten. Je detaillierter die Merkmale sind, desto weniger Funde sind möglich, Abweichungen werden dann also nicht erfasst. In der Regel lassen sich aber auch eine Menge Fehlfunde ausschließen.

Die Suchmaschine meines „Imperium“ ist auf allgemeinste Merkmale reduziert, erzielt bei lateinischen Schriften über 90% Treffer, bei der Fraktur aber deutlich weniger als 70%. (Das liegt u.a. daran, dass die Normierung von Großbuchstaben erst sehr spät erfolgte). Sie erfasst nicht nur groß und klein geschriebene Wörter, sondern auch gesperrte und unterstrichene Wörter. Da man im 3. Reich nicht selten Namen und wichtige Begriffe `g e s p e r r t` schrieb, tippte oder druckte bzw. mit Unterstreichungen versah, entgehen Historikern durch die gängigen Suchmaschinen (copernic oder docfetcher) gerade diese Hervorhebungen, die mehr und wichtigere Treffer versprechen als andere Wörter. Auch die pdf- und die doc-Dateien erfassen diese Hervorhebungen häufig nicht.

Als erstes sollte man lernen, die Suchmaschine für das „Imperium“ klar zu unterscheiden von den Suchmaschinen in den einzelnen Dateien. Erstere erzeugen sehr viele Fehlfunde, nicht nur weil sie auch „unpassende“ Varianten der Schreibweise erfassen, sondern auch weil in der Sprache viele Wörter mehrere, manchmal nicht zusammenhängende Bedeutungen haben (Beispiel: Das Verbum kommen). Letztere finden dagegen manche Treffer nicht, weil diese gesperrt oder unterstrichen sind.

Alle Treffer zu finden, kann keine Suchmaschine versprechen. Für mich am effektivsten war ein Verfahren, das die Stärken beider Suchmaschinenarten kombiniert. Am einfachsten zu finden sind Wörter und Begriffe bzw. Namen und Themen, deren Bedeutungsspektrum keine oder nur wenige Abweichungen kennt. Am schwierigsten sind Allerweltsnamen. (s.dazu www.gerd-simon.de/Allerweltsnamen)

Ich erinnere noch, dass ich im Bundesarchiv (seinerzeit noch in Koblenz) einen hochinteressanten Text fand, dessen Verfasser sich Werner MÜLLER nannte. (Heute nenne ich ihn nach seinem Hauptwerk „Kreuz-und Kreis-MÜLLER“). Es kostete mich viel Schweiß und Zeit, bis ich die Identität dieses ursprünglichen Ortungs- und Symbolforschers mit dem Werner MÜLLER identifizieren konnte, den ich in der Unibibliothek Tübingen persönlich kennengelernt hatte. Allein im >Berlin Document Center<, in dem die Amerikaner zig Millionen Personalakten gesammelt haben, gab es mehrere 1000 Werner MÜLLERs.

In meinem vergleichsweise kleinen „Imperium“ meldet meine Suchmaschine 5.526 MÜLLERs und 2.973 Werner MÜLLERs. Da diese Suchmaschine aber die Schwäche hat, nicht wie GOOGLE für das ganze Internet “beieinander stehende“ Wörter wie ein Wort zu behandeln, sind in letzteren „Treffern“ auch alle Texte ermittelt, in denen beide oder meistens auch nur eines der beiden Wörter vorkommen. Es betrifft dann natürlich auch manche anderen, die Werner heißen. Hinzu kommt, dass man z. T. noch heute auch Allerweltsnamen nicht mit Vornamen, sondern mit Vorliebe mit dem Dr. –Titel erwähnt, den wiederum Tausende MÜLLERs erwarben. Es blieb also nichts anderes übrig, als alle Texte, in dem die 5.526 MÜLLERs vorkommen, aufzumachen und das gesuchte Wort mit der dateiinternen Suchmaschine aufzufinden und aus dem Zusammenhang zu ermitteln, ob es sich überhaupt um unseren Werner Müller handeln kann. Manchmal muss dann immer noch offen bleiben, ob der gefundene MÜLLER mit dem gesuchten MÜLLER identisch ist. Im Falle meines Kollegen Werner MÜLLER waren solche Suchen auch dadurch erschwert, dass er auf mehreren Gebieten als Wissenschaftler tätig war (Ortungs- und Symbolforschung, Hakenkreuzforschung, Indianderforschung usw.), im 3. Reich an der Uni Straßburg Professor war, in Tübingen aber vor meinen Hinweisen kaum eine Handvoll Kollegen wussten, dass er einmal Professor war, weil er nach 1945 nur noch einen Dokortitel führte. Diese Kollegen waren zu diesem Zeitpunkt, als ich die Identitätssuche abgeschlossen hatte, gestorben (oder in einem Falle dement) wie inzwischen auch Werner MÜLLER selbst.

Über die Allerweltsnamen hinaus gibt es auch Namen, die manchmal sehr selten vorkommen, die aber gleich geschrieben werden wie Nichtnamen, z. B. Gegenstände (z.B. Fels) oder Eigenschaften (z. B. Schön). Hier muss man fast noch mehr mit Fehlfunden rechnen als bei den Allerweltsnamen.

Fazit: Ich habe nicht nur in den Zettelkastenschränken und im PC nach einem Stichwort (Name oder Begriff bzw Thema) zu suchen, sondern muss auch die Schwächen der Suchmaschinen kennen und darf auch die Erwartungen nicht zu hoch schrauben. 100% Treffer dürfte man bei historischen Texten vermutlich nie erreichen.