Historisch
Kulturwissenschaftliche
Informationsverarbeitung

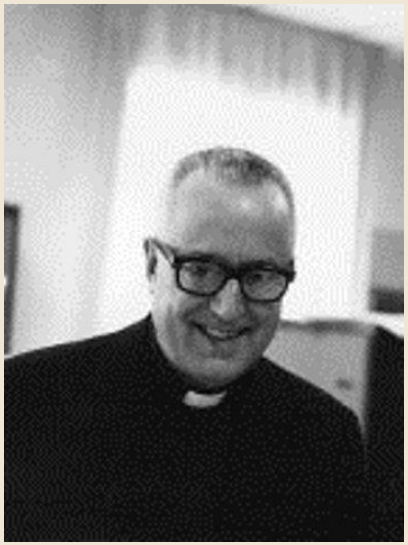
Wenn die Quellen überfliessen. Spitzweg und Big Data

Manfred Thaller,
Universität zu Köln

Graz, 27. Februar 2015

DHd2015

… Gelehrte

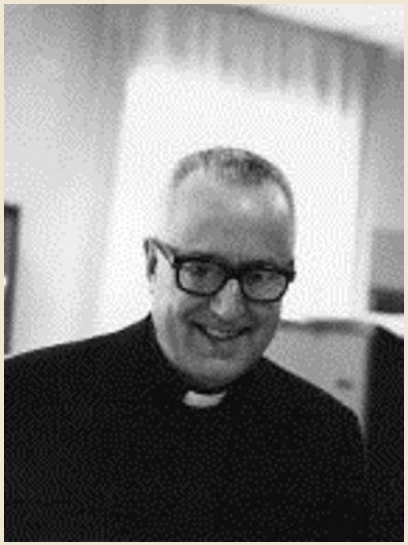


Spitzweg, Gelehrter Mönch am Schreibtisch

Padre Roberto Busa, 1913 -2011

Science Fiction Autoren und Gelehrte

E. E. „Doc“ Smith, 1890 - 1966



Padre Roberto Busa, 1913 -2011

Häresie 1

Es ist für Geisteswissenschaftlerinnen *nicht* zwingend notwendig, fantasielos und intellektuell ängstlich zu sein.

„Big Data“ und Gelehrte

Computer: neue Prosopographien

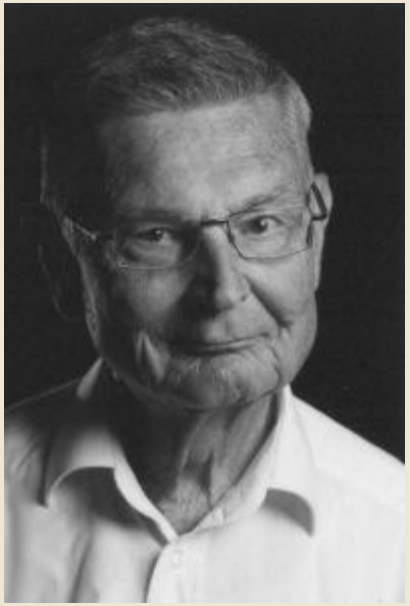
Arnold J. Toynbee, 1889-1975

„Big Data“ und Gelehrte

[Computer:] neue Prosopographien

Prosopographia Regnorum Orbis Latini
Karl Ferdinand Werner, 1924 - 2008

„Big Data“ und Gelehrte



Computer: (?) neue Prosopographien

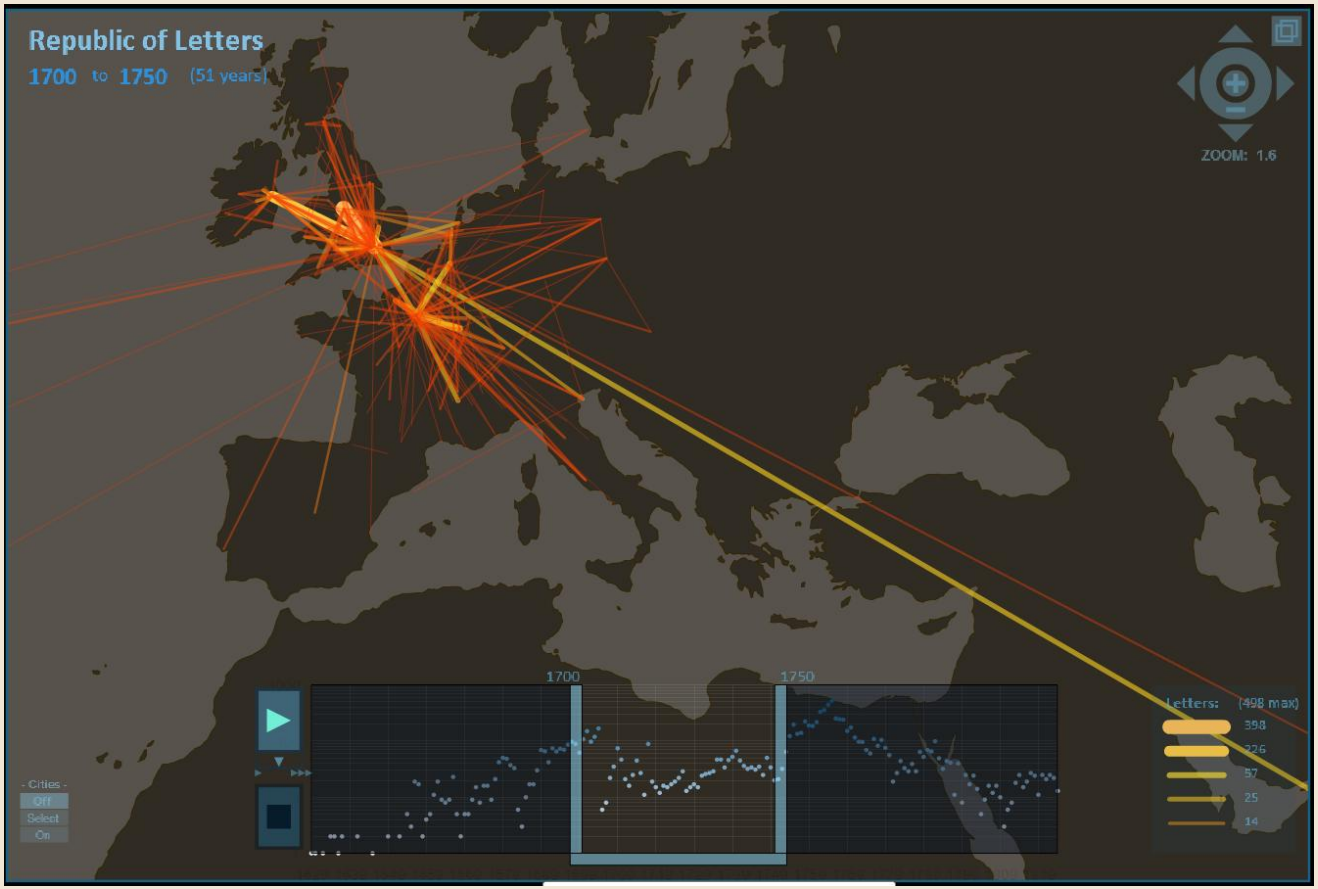
Memorialüberlieferung, 1974 ff.
[Karl Schmid,] Joachim Wollasch

Häresie 2

Auch methodisch sehr konservative Grundhaltungen sind mit dem Wunsch vereinbar, grundsätzlich Neues zu wagen.

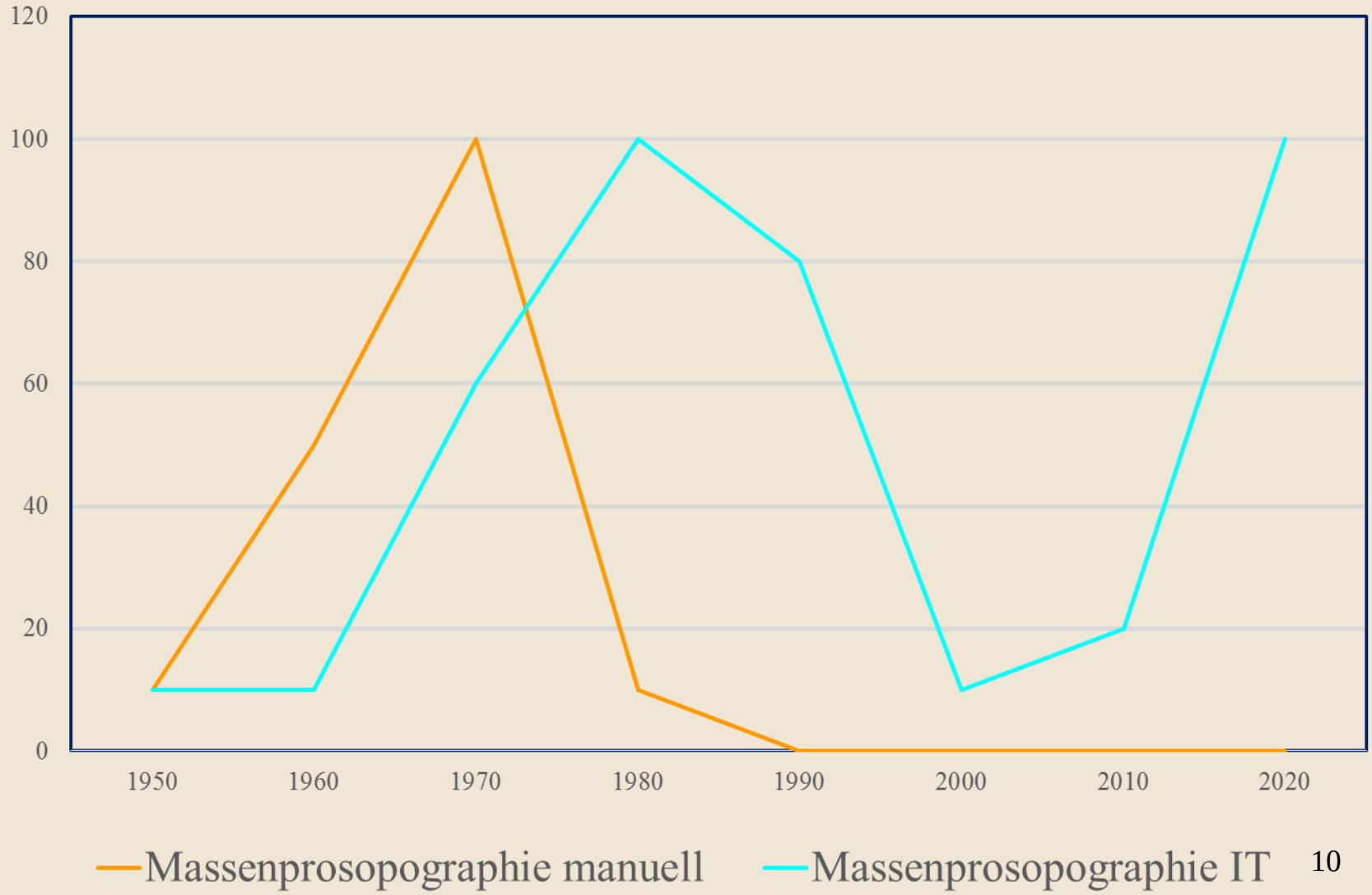
Das kann auch *ohne* IT gründlich schief gehen.

„Big Data“ und Gelehrte



http://republicofletters.stanford.edu/

Massenprosopographien



… Gelehrte

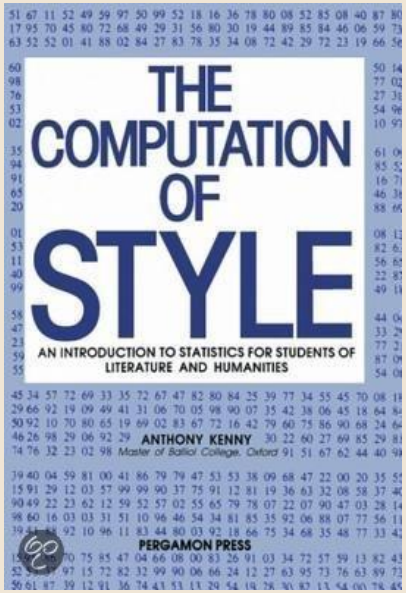
Sir Anthony Kenny

… Gelehrte

Sir Anthony Kenny

Geschichte der abendländischen Philosophie, 2014

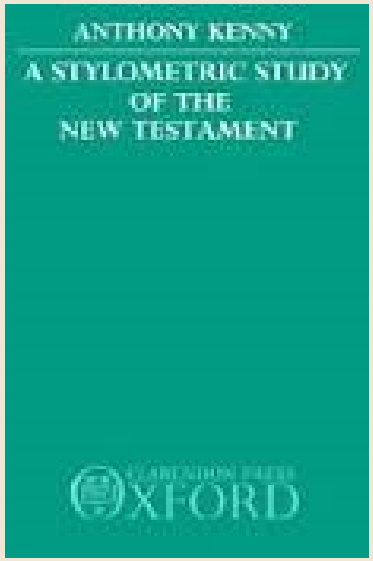
… Gelehrte



Sir Anthony Kenny

The computation of style, 1981

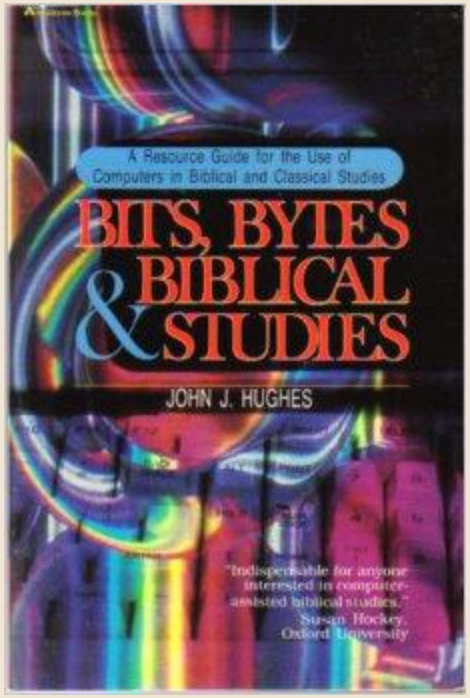
… Gelehrte



Sir Anthony Kenny

A Stylometric Study of the New Testament, 1986

Gelehrte im Kontext



John J. Hughes: Bits, Bytes and Biblical Studies: A Resource Guide for the Use of Computers in Biblical and Classical Studies, 1987

Lexikonformat, 643 pages

Ein Blick zurück: 1993

Neben den Vorteilen enormer Speicherkapazitäten historischer Fakten und Daten rückt nun immer stärker auch die Möglichkeit präziser Dokumentationen von bildlichen Dokumenten durch EDV-Verarbeitungen in den Vordergrund der moderner Geschichtsforschung – und natürlich die rasche Verfügbarkeit und Überprüfung dieser Dokumente durch die internationale Kommunikationsvernetzung.

Der Standard, 28./29.8.1993, p. 13

Ein Blick zurück: 1993

Geschichtsforschung per Computer

Die modernen Historiker arbeiten mit weltweiten Netzwerken

Walter Müller

Graz – Bis in die frühen Sechziger Jahre blieb die historische Wissenschaft beschränkt auf einige wenige monolithische Blöcke wie Wirtschafts-, Rechts- oder zuletzt Sozialgeschichte. In den folgenden Jahren verästelten sich die Geschichtswissenschaften in rund 40 bis 50 unterschiedliche Disziplinen. Ein Teil der Wissenschafter begann, verstärkt elektronische Datenverarbeitung in ihren Forschungen, die immer vernetzter wurden, einzusetzen. Dabei entwickelten sich die Historiker von reinen Computeranwendern zu Programmentwicklern.

Neben den Vorteilen enormer Speicherkapazitäten historischer Fakten und Daten rückt nun immer stärker auch die Möglichkeit präziser Dokumentationen von bildlichen Dokumenten durch EDV-Verarbeitungen in den Vordergrund der moderner Geschichtsforschung – und natürlich die rasche Verfügbarkeit und Überprüfung dieser Dokumente durch die internationale Kommunikationsvernetzung.

150 Historiker aus 23 Ländern trafen sich dieser Tage in Graz zur 8. Jahrestagung der internationalen „Association for History and Computing" (AHC), um Erfahrungen dieser neuen computerunterstützten Forschung auszutauschen. Erstmals wurden auch Wissenschafter aus dem südosteuropäischen Raum verstärkt ins wissenschaftliche Gespräch eingebunden.

Leonid Borodkin von der Moscow State University erinnerte daran, daß die osteuropäischen Universitäten nach wie vor unter einem Mangel an elektronischen Ressourcen leiden. Westeuropäische Wissenschafter versuchen daher seit einiger Zeit, östliche Universitäten mit EDV-Anlagen auszustatten. Erst kürzlich sei ein Rechenpool für Geisteswissenschaftliche Institute in Estland von norwegischen Universitäten finanziert worden, ergänzte AHC-Chef Manfred Thaller.

Es sei im Osten gegenwärtig klimatisch eine günstige Zeit für die Geisteswissenschaften, so Thaller. In Polen etwa stehe momentan eine Sprach- und EDV- Ausbildung weit höher im Kurs als die gängigen betriebswirtschaftlichen Studien.

Eine der zentralen Aufgaben der künftigen EDV-unterstützten Geschichtswissenschaften: Die Weiterentwicklung international einheitlicher Kommunikationsebenen und Parameter.

ARD THEMEN

puter-
ologien

Der Standard, 28./29.8.1993, p. 13

Ein Blick zurück: 1993

HISTORIKERTAGUNG

Schwerpunkt Ostkontakte

Graz. — „Die Kunst der Kommunikation" steht derzeit in Graz im Mittelpunkt der Jahrestagung der „Association for History and Computing", die von den Universitätsinstituten für Historische Grundwissenschaften und Wirtschafts- und Sozialgeschichte veranstaltet wird. Besonderer Schwerpunkt in der Nähe des ehemaligen Eisernen Vorhangs ist der Wissensaustausch mit osteuropäischen Kollegen.

Kleine Zeitung, 27.8.1993, p. 10

Ein Blick zurück: 1993

10 KLEINE ZEITUNG NACHRICHTEN FREITAG 27. AUGUST 1993

VORSORGEMEDIZIN

Grazer im Kampf gegen Brustkrebs

Graz. — Jede neunte Österreicherin erkrankt an Brustkrebs, trotzdem werden die Möglichkeiten der Früherkennung kaum genützt.

60 österreichische Mediziner haben sich an einer Initiative des Grazer Krebsspezialisten Michael Smola beteiligt und im „Consensus-Bericht" den Stand der Wissenschaft zum Thema dargelegt. Ihre Empfehlungen werden demnächst veröffentlicht.

HIMMELSATLAS

Neuer Vorstoß in die Galaxis

Wien, Innsbruck. — Größe, Form und Geschichte unserer Galaxis sind noch lange nicht erforscht. Jetzt soll die südliche Hemisphäre mit einem neuen Aufnahmeverfahren im Infrarotbereich des Lichts abgetastet werden.

DENIS (Deep Near Infrared Survey of the Southern Sky) heißt das internationale Himmelsatlasprojekt, das auch Innsbrucker und Wiener

HISTORIKERTAGUNG

Schwerpunkt Ostkontakte

Graz. — „Die Kunst der Kommunikation" steht derzeit in Graz im Mittelpunkt der Jahrestagung der „Association for History and Computing", die von den Universitätsinstituten für Historische Grundwissenschaften und Wirtschafts- und Sozialgeschichte veranstaltet wird. Besonderer Schwerpunkt in der Nähe des ehemaligen Eisernen Vorhangs ist der Wissensaustausch mit osteuropäischen Kollegen.

Auch die Bären durften am Mauterner Jubeltag mitnaschen FOTO: SCHÖLLENBAUER

MAUTERN JUBILIERT

Bärig ist so eine Honigtorte

Mautern. — Der 1,000.000ste Besucher, das 25jährige Betriebsjubiläum und der erste Braunbären-Jahrestag — Mautern, was willst du mehr an Jubiläen.

Käthe und Dirk Walcker aus Schleswig-Holstein in der Bundesrepublik kamen zur rechten Zeit in den Tierpark Mautern. Das Geschenk für die Einmillionsten Besucher kann sich sehen und erleben lassen — ein einwöchiger Aufenthalt in Mautern. Familie Walcker will die Woche gleich nun, im Anschluß an ihren bereits dritten Urlaub in der Grünen Mark, genießen.

Zum Feiern hatten auch die Braunbären aus der ehemaligen DDR, die vor genau einem Jahr in die Obersteiermark kamen. Sie erhielten eine aus Bienenwaben geformte Honigtorte, die den zwei pelzigen Damen und dem Herrn sichtlich schmeckte.

Wohl fühlen sich auch die vielen kleinen Gäste in Mautern. Beim Ausbau des 60 Hektar großen Tierparks entstand der größte Holzspielplatz in Österreich. Apropos Gäste: Allein heuer erwartet Heinrich Reuss, Besitzer der Anlage, an die 60.000 Besucher.

Mysteriöse Lkw-Ladungen in Italien beschlagnahmt

Maschinenteile für den Iran. Gerüchte über militärische Zwecke.

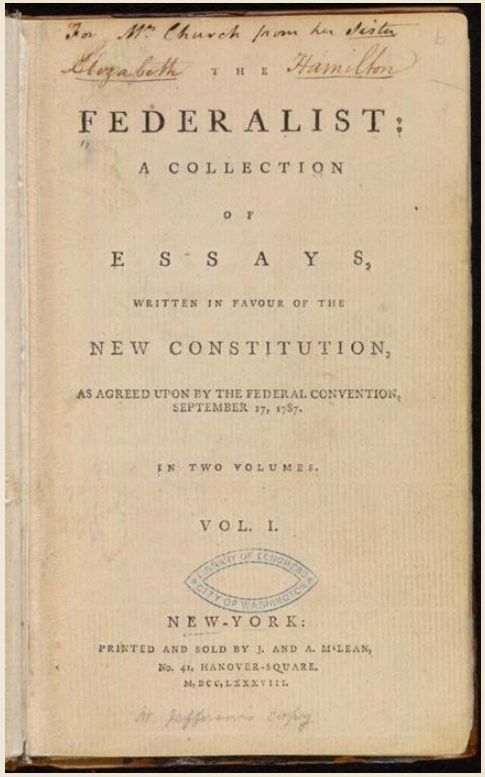
Rätsel um den Inhalt von insgesamt zwölf riesigen Holzkisten, die, verladen auf vier Sattelschleppern mit Tiroler Kennzeichen, vom italienischen Zoll beschlagnahmt wurden. Die Sattelauflieger samt Ladung stehen seit einigen Tagen auf dem Zollamtsplatz Sadobre-Freienfeld bei Sterzing. Wo sich die Zugmaschinen und die Fahrer aufhalten, ist nicht bekannt. Die Chauffeure konnten den Zollbehörden für die Ladung, angeblich Maschinenteile zur Herstellung von Automobilkolben, keine Transitgenehmigungen und keine Erklärung, daß es sich um Transitware handelt, vorweisen. Die Maschinenteile sollten im Auftrag einer in der Nähe von Köln ansässigen Firma in den Hafen von Venedig gebracht und dort verschifft werden. Angeblicher Abnehmer ist eine Firma in einer Kleinstadt in der Nähe Teherans.

Richter Dr. Christian Mayer vom Amtsgericht Bozen hat jetzt beim Lieferanten die ausständigen Papiere angefordert. Dann wird entschieden, ob die Kisten geöffnet werden. Mittlerweile ranken sich wilde Gerüchte um die Ladung. Es gibt Spekulationen, wonach es sich um Maschinenteile, die auch zur Herstellung von Kriegsgerät verwendet werden können, handelt. Der Iran gilt als sogenannte „Schwellenmacht", also als Staat, dem die Entwicklung von Nuklearwaffen gelinge könnte.

Peter Kimeswenger

Kleine Zeitung, 27.8.1993, p. 10

… Gelehrte


For Mr. Church from his sister
Elizabeth Hamilton

THE
FEDERALIST;
A COLLECTION
OF
ESSAYS,
WRITTEN IN FAVOUR OF THE
NEW CONSTITUTION,
AS AGREED UPON BY THE FEDERAL CONVENTION,
SEPTEMBER 17, 1787.

IN TWO VOLUMES.

VOL. I.

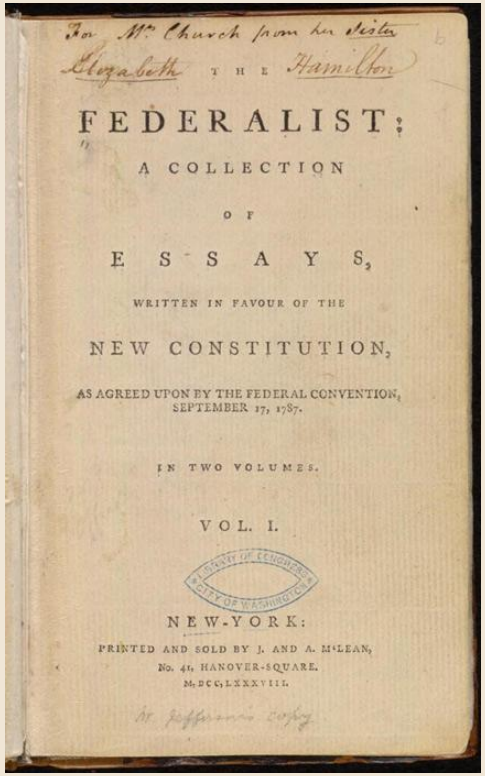
NEW-YORK:
PRINTED AND SOLD BY J. AND A. M'LEAN,
No. 41, HANOVER-SQUARE.
M, DCC, LXXXVIII.

Mr. Jefferson's copy

Frederick Williams & Frederick Mosteller, 1941

The Federalist Papers, 1786

Anfänge



For Mr. Church from his sister
Elizabeth Hamilton

THE

FEDERALIST;

A COLLECTION

OF

ESSAYS,

WRITTEN IN FAVOUR OF THE

NEW CONSTITUTION,

AS AGREED UPON BY THE FEDERAL CONVENTION,
SEPTEMBER 17, 1787.

IN TWO VOLUMES.

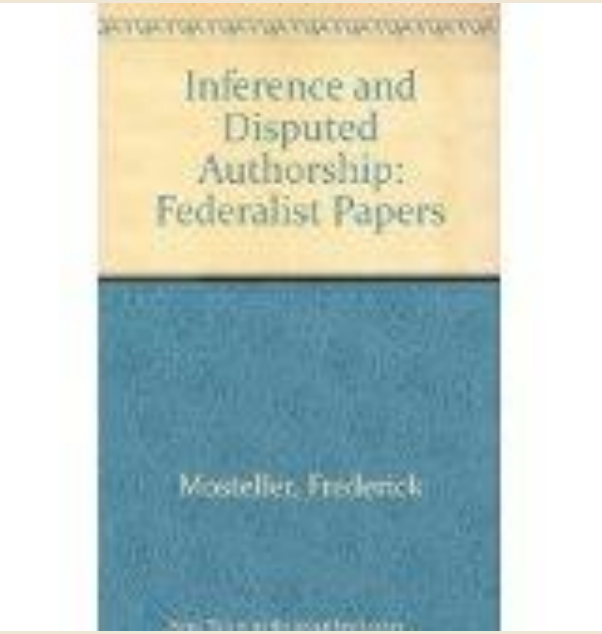
VOL. I.

NEW-YORK:

PRINTED AND SOLD BY J. AND A. M'LEAN,
No. 41, HANOVER-SQUARE.
M, DCC, LXXXVIII.

Mr. Jefferson's copy

The Federalist Papers, 1786



Frederick Mosteller and David L. Wallace, Inference and Disputed Authorship: *The Federalist*, 1964

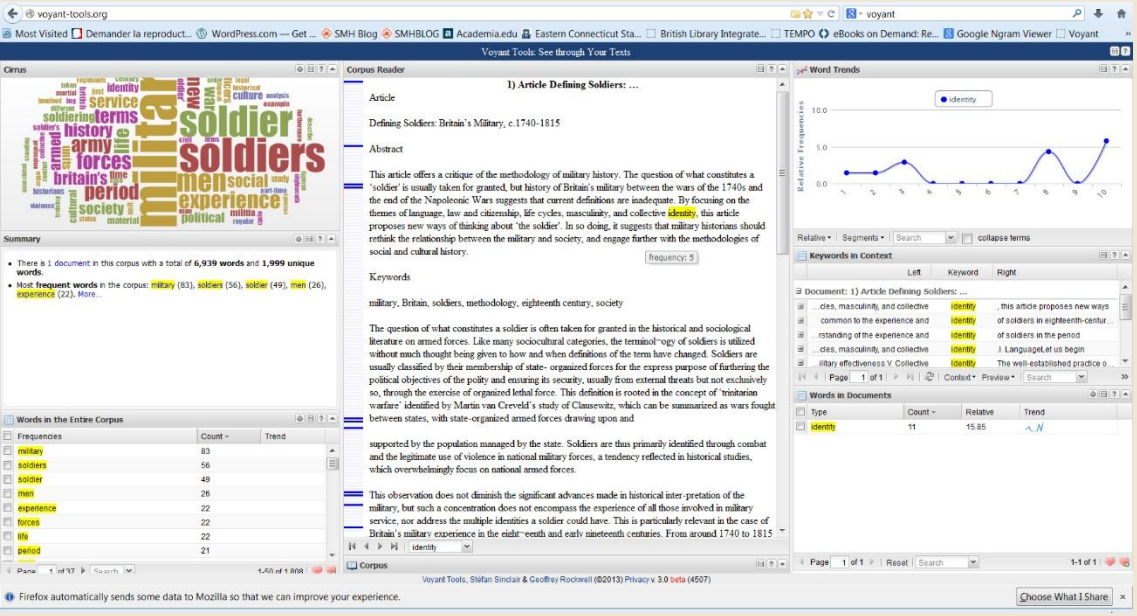
Anfänge

Philip J. Stone:
The General Inquirer, 1961

IBM 7090, 1960er

2015: http://www.wjh.harvard.edu/~inquirer/

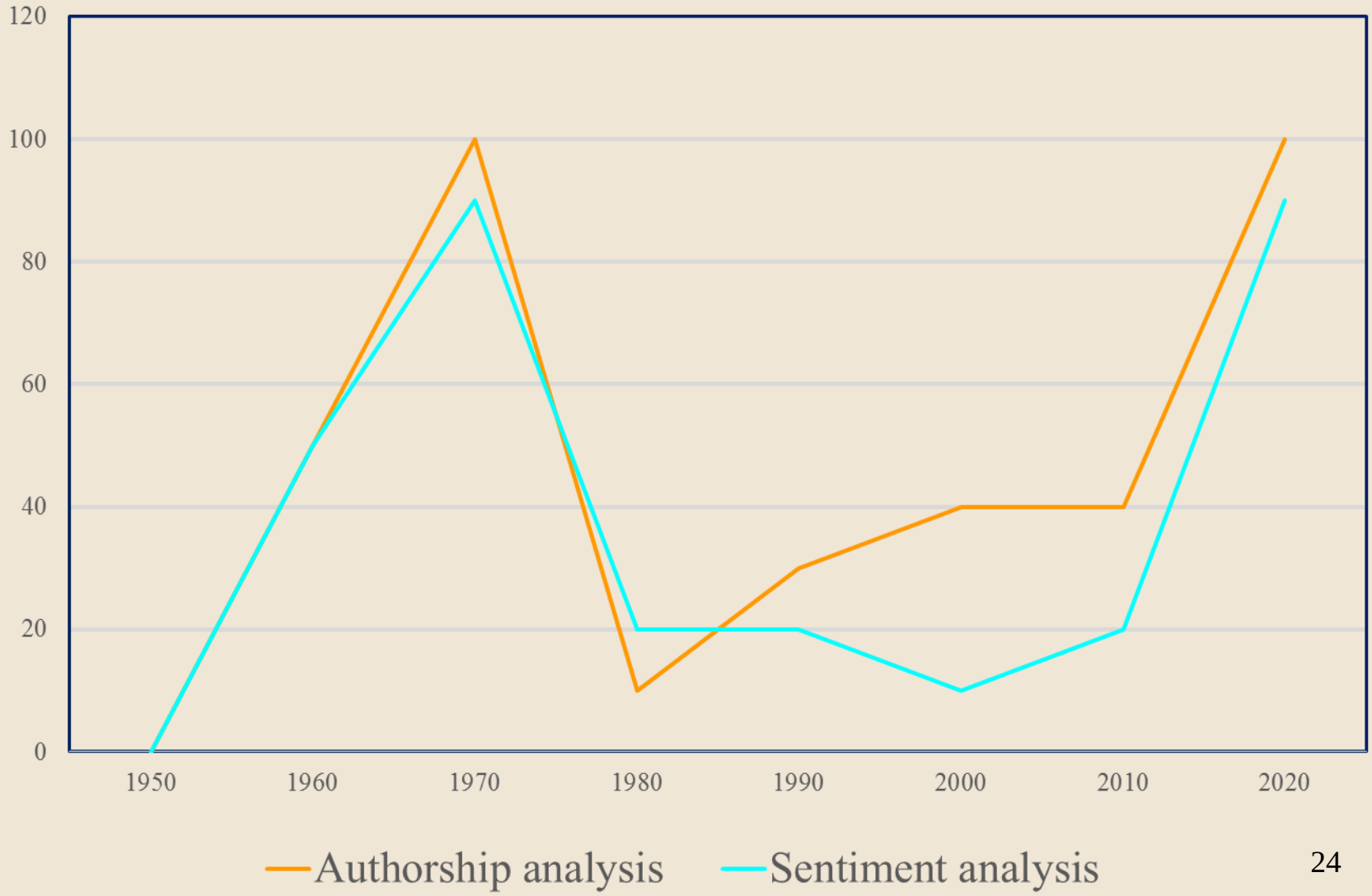
Status quo



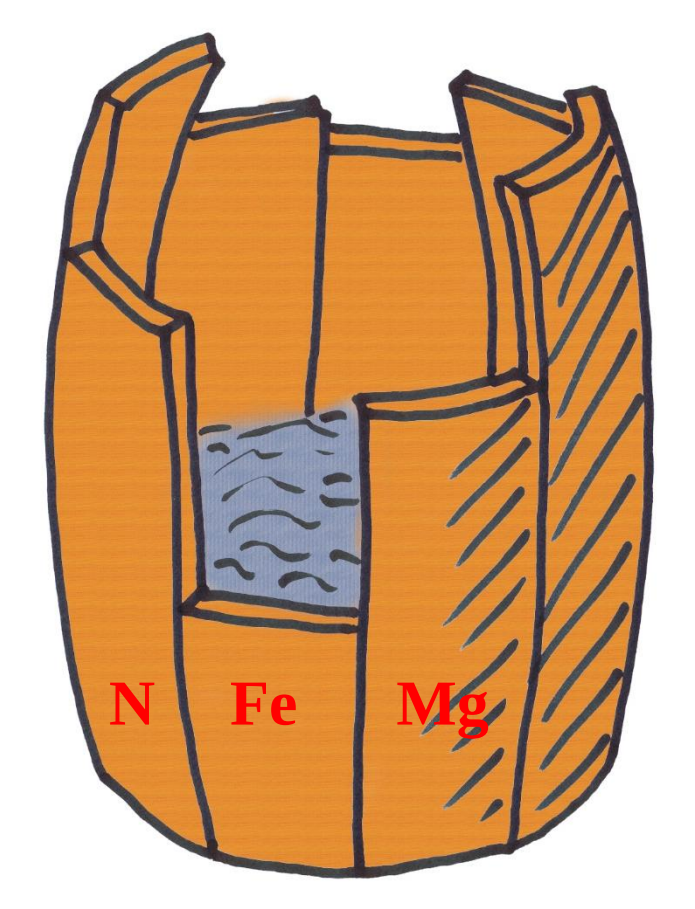
Stéfan Sinclair & Geoffrey Rockwell

http://hermeneuti.ca/

Verläufe


120
100
80
60
40
20
0
1950
1960
1970
1980
1990
2000
2010
2020
Authorship analysis
Sentiment analysis

Versuch einer Verallgemeinerung



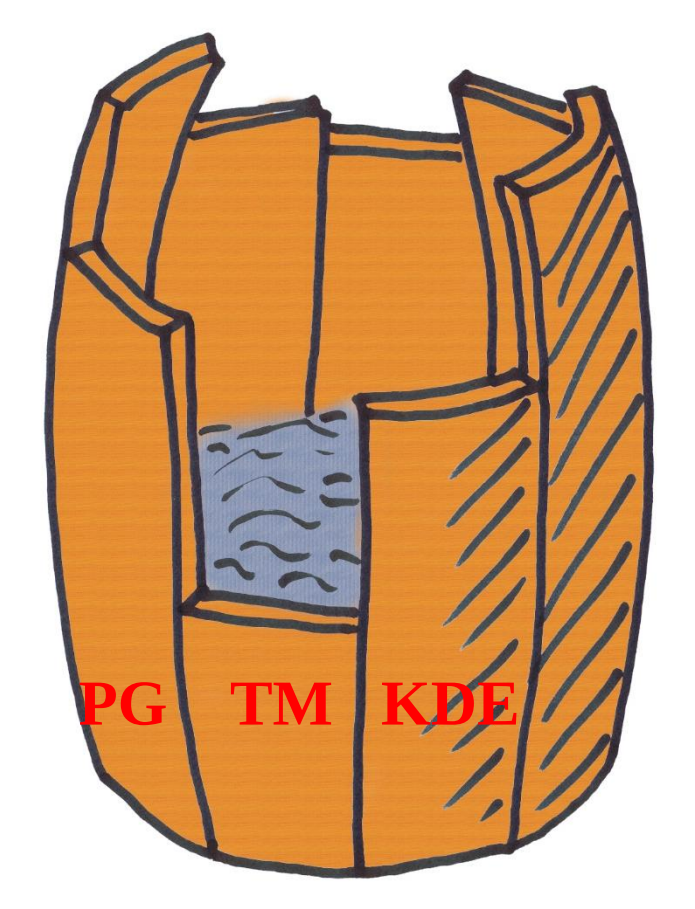
N = Stickstoff

Fe = Eisen

Mg = Magnesium

Sprengels Minimumgesetz / Liebigs Fass: Pflanzliche Ernährung

Versuch einer Verallgemeinerung



PG = Prozessorgeschwindigkeit

TM = Theoretische Modelle

KDE = Aufwand für die Datenerfassung

Big data …

Eine Gasturbine von General Electric produziert an einem Tag 588 Gigabyte Daten.

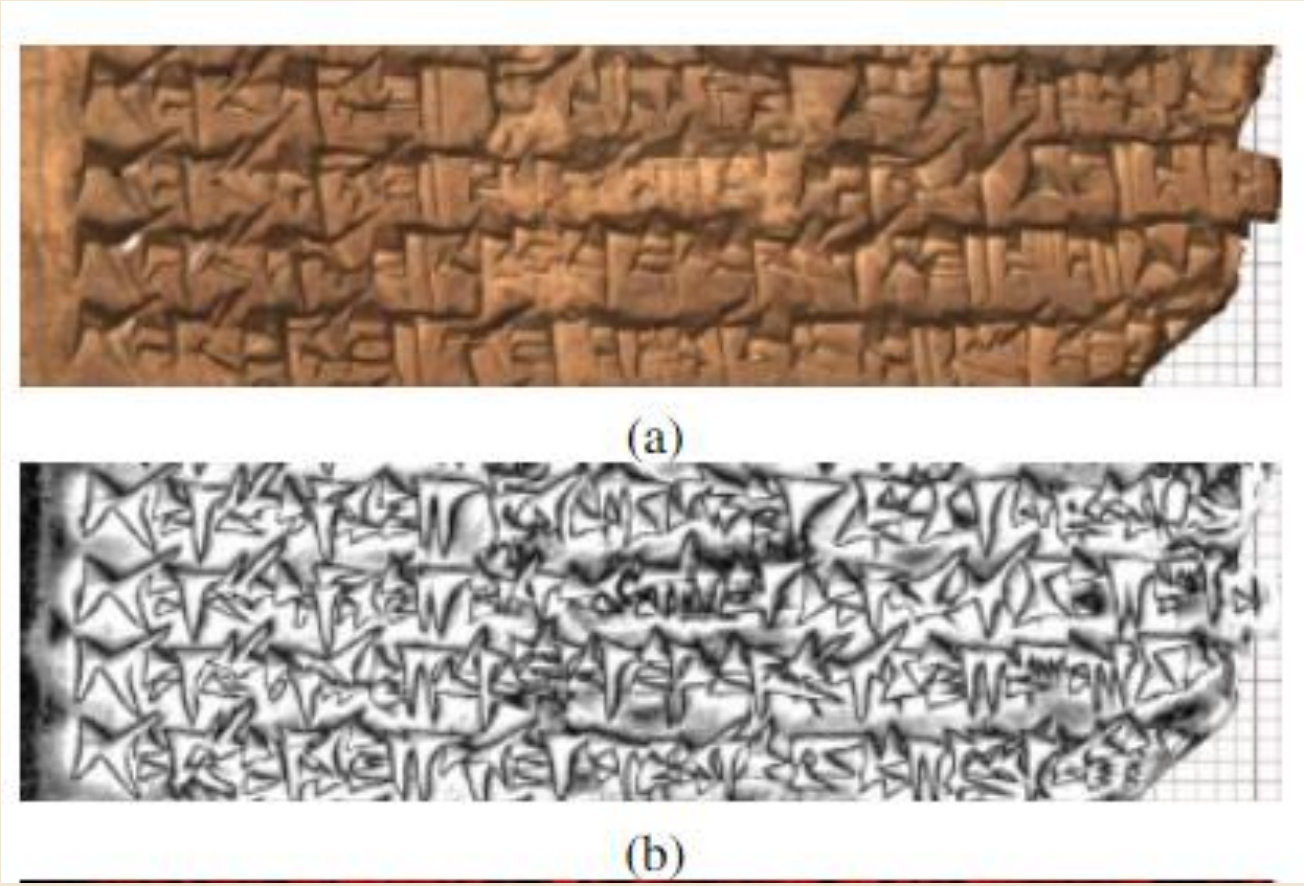
Süddeutsche Zeitung, 41 / 19. Februar 2015, p. 21, Max Hägler: Transparenz ist zu wenig

Big data …

Big data == n * FK* VK

FK = Filterkomplexität
VK = Verarbeitungskomplexität

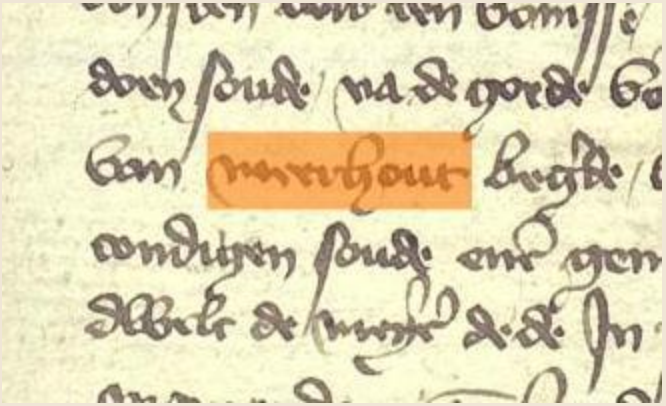
Filterkomplexität I

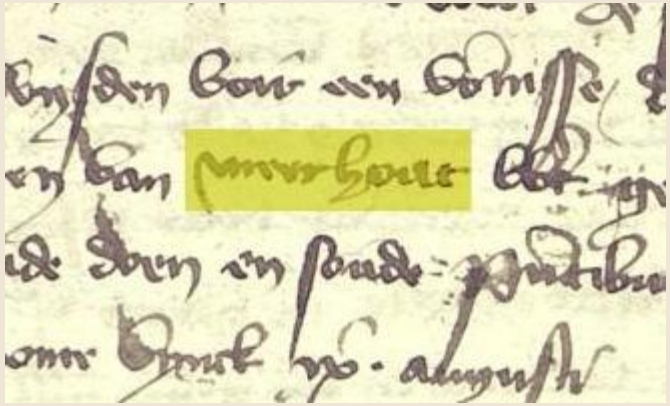


H. Mara - http://www.academia.edu/1217620/GigaMesh_and_Gilgamesh_-_3D_Multiscale_Integral_Invariant_Cuneiform_Character_Extraction

Filterkomplexität I

Meerhout in the *scheppenregister* of the city of Leuven:

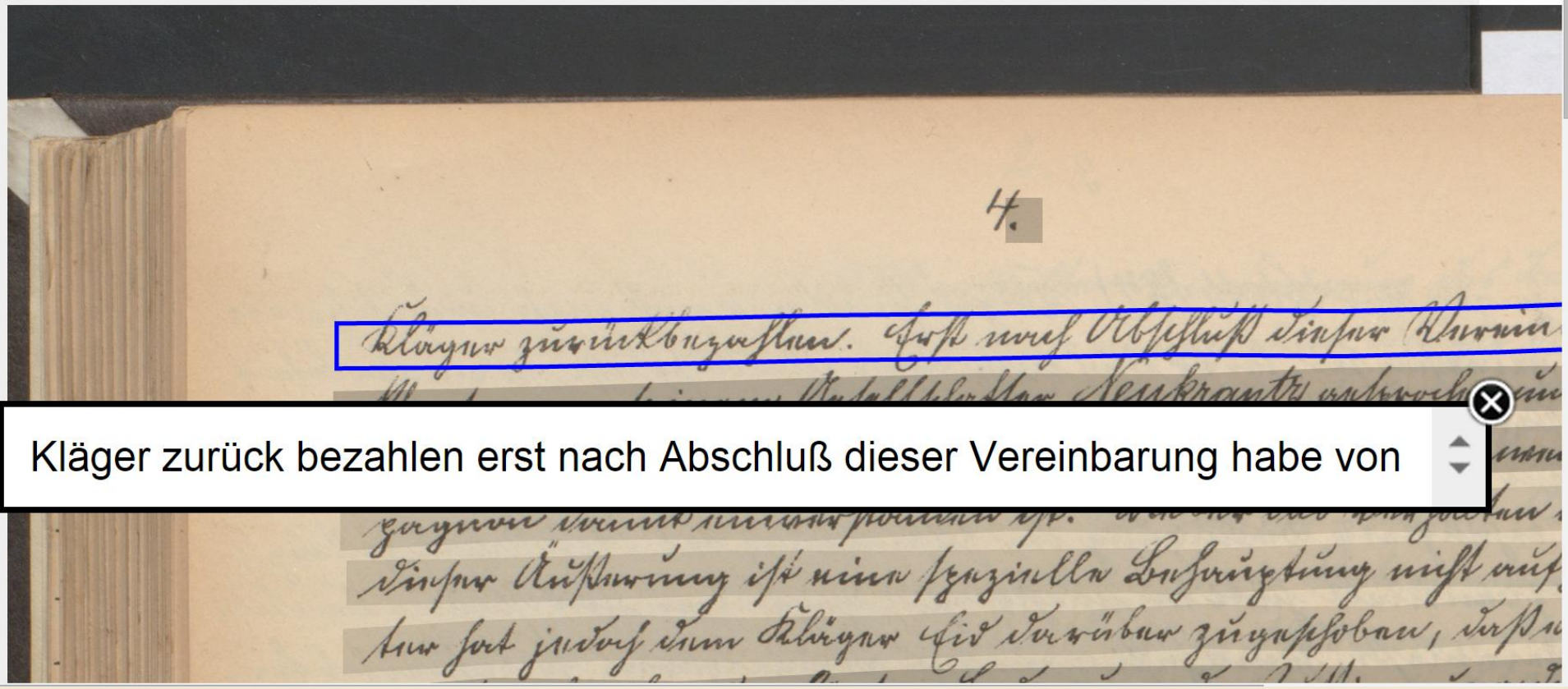




L. Schomaker, MONK: http://monk.target.rug.nl/

Filterkomplexität I

You are transcribing page 1 of chapter 40 of **Reichsgericht** collection.



http://transcriptorium.eu/

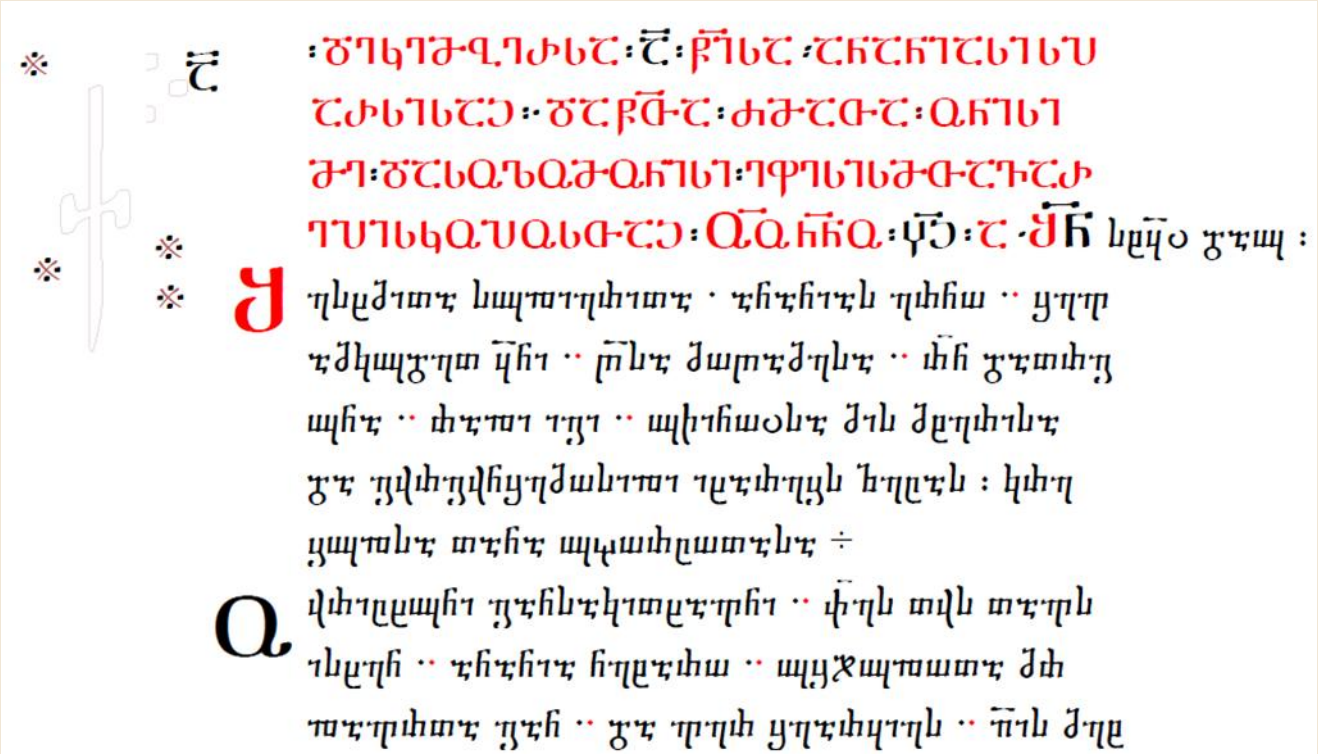
Filterkomplexität I

http://www.adfontes.uzh.ch

Filterkomplexität II

J. Gippert, http://www1.uni-hamburg.de/COMST/handbookonline.html

Filterkomplexität II



J. Gippert, http://www1.uni-hamburg.de/COMST/handbookonline.html

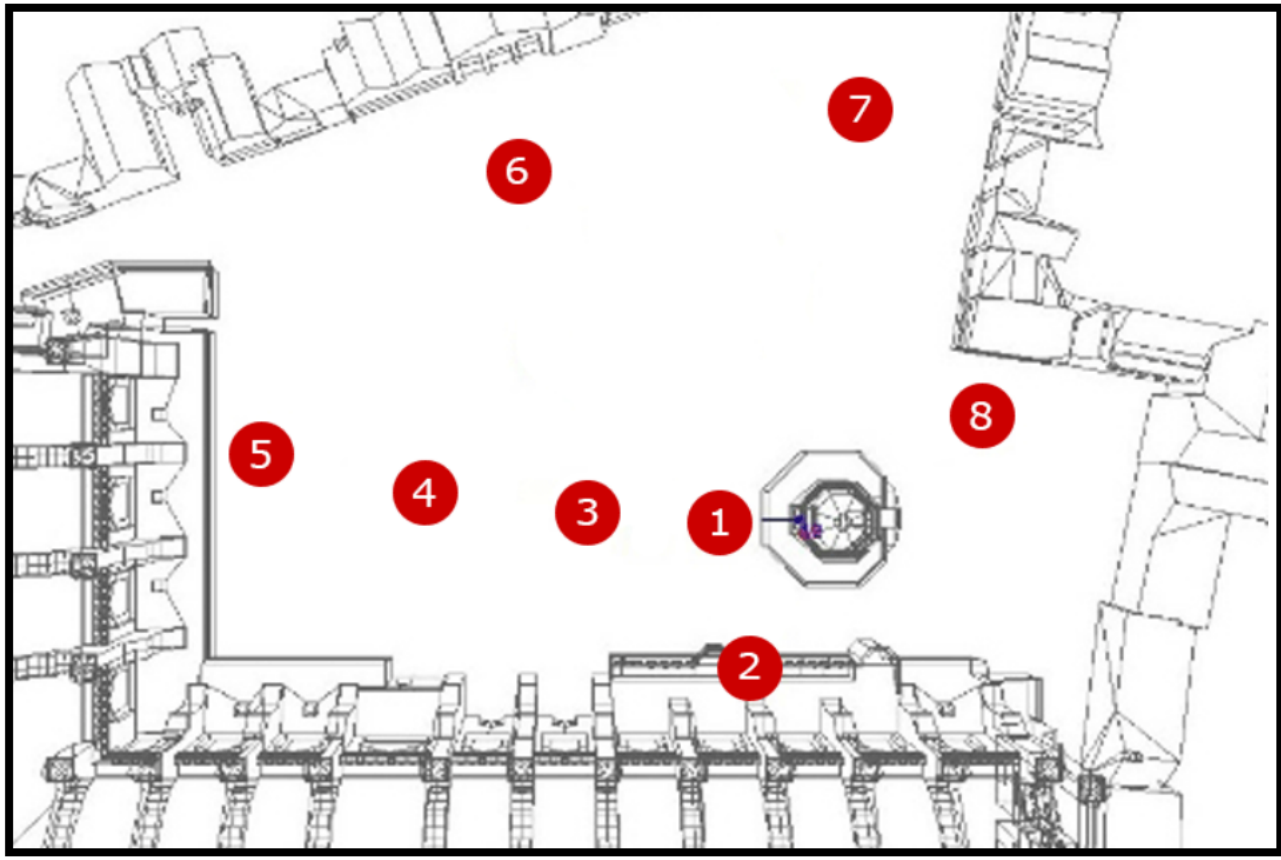
Verarbeitungskomplexität

http://vpcp.chass.ncsu.edu/ --- „John Donne“

Verarbeitungskomplexität

http://vpcp.chass.ncsu.edu/ --- „John Donne“

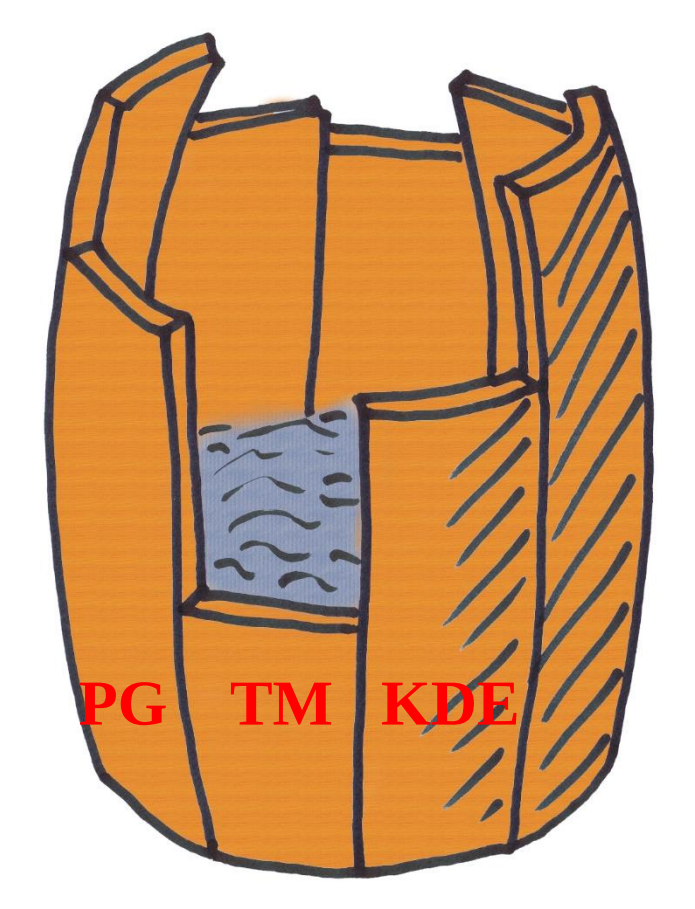
Verarbeitungskomplexität



Choose a Location:	
500 People	2500 people
Location 1	Location 1
Location 2	Location 2
Location 3	Location 3
Location 4	Location 4
Location 5	Location 5
Location 6	Location 6
Location 7	Location 7
Location 8	Location 8
1200 people	5000 people
Location 1	Location 1
Location 2	Location 2
Location 3	Location 3
Location 4	Location 4
Location 5	Location 5
Location 6	Location 6
Location 7	Location 7
Location 8	Location 8

http://vpcp.chass.ncsu.edu/ --- „John Donne“

Versuch einer Verallgemeinerung



PG = Prozessorgeschwindigkeit

TM = Theoretische Modelle

KDE = Aufwand für die Datenerfassung

Big data …

Big data without big theory is big bullshit!

„Sie hören ja nichts anderes mehr: *big data*. Aber big data without big theory is big bullshit!“ Ohne Theorie, ohne Verständnis des Menschen und seiner Psyche komme man nicht weiter. Management sei eine *liberal art*, eine Geisteswissenschaft.

Süddeutsche Zeitung, 261 / 13. November 2014, p. 19, „Nahaufnahme“: Interview mit Richard Straub, ehemaliger Chief Learning Officer (Leiter der Personalentwicklung) bei IBM.

Big data …

What Do You Do with a Million Books?
Greg Crane, 2006

Ja, abgesehen von den 1000, die klassische Texte enthalten, was eigentlich?

Was um alles in der Welt machen wir dann mit 20.000 Archivkilometern?

Eine Definition von eScience

Naturwissenschaft in einer Arbeitsumgebung, in der

(1) der Zugang zur für eine Forschungsfrage benötigten Information,

(2) die Analyse dieser Information

(3) und die Publikation der gewonnenen Ergebnisse

gleichermaßen gut durch die (verteilte) Informationstechnologie unterstützt wird.

Eine Definition von eHumanities

Geisteswissenschaft in einer Arbeitsumgebung, in der

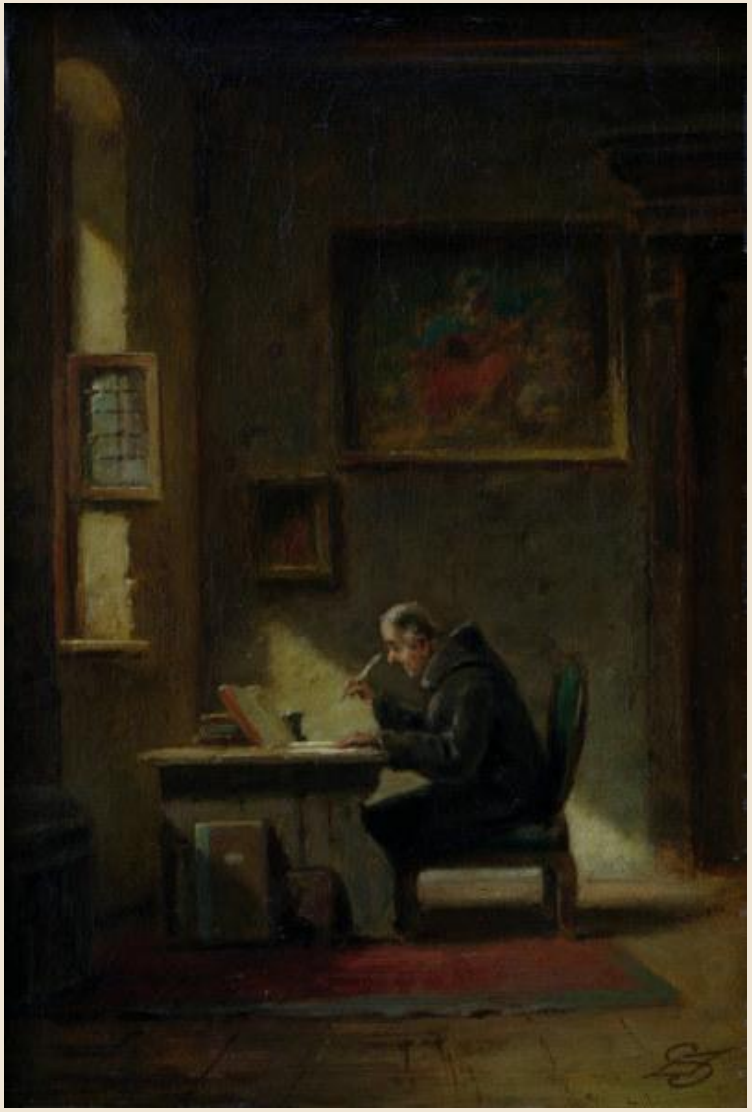
(1) der Zugang zur für eine Forschungsfrage benötigten Information,

(2) die Analyse dieser Information

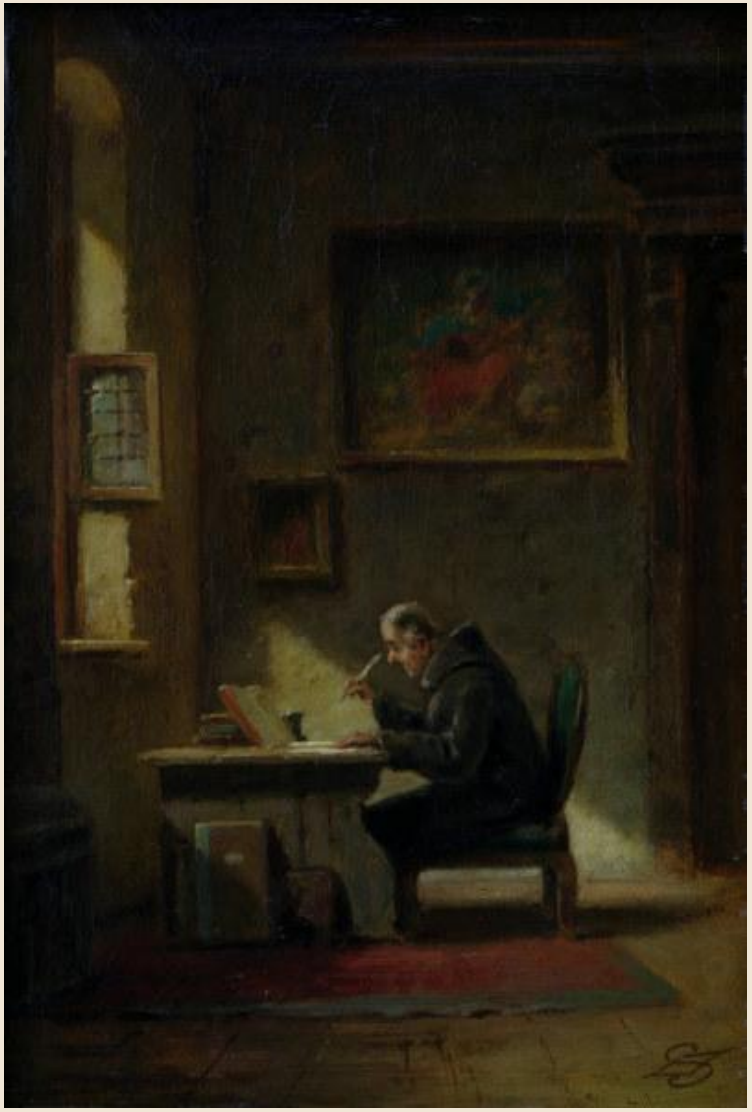
(3) und die Publikation der gewonnenen Ergebnisse

gleichermaßen gut durch die (verteilte) Informationstechnologie unterstützt wird.

Wenn die Quellen überfließen …



Wenn die Quellen überfließen …



Kritik

Konzentration auf Infrastruktur verhindert Diskussion, was in den Geisteswissenschaften mit IT erreicht werden soll.

(W. McCarty: http://digitalhumanities.org/humanist/)

Mechanistische Übernahme aus den Hard Sciences; zu starke Konzentration auf Werkzeuge trivialisiert die methodisch konzeptuelle Diskussion.

(P. Svensson http://digitalhumanities.org/dhq/vol/5/1/000090/000090.html im Kontext von: .../vol/3/3/000065/000065.html und …/vol/4/1/000080/000080.html)

Kritik & Angst

The digital humanities are at a rhetorical and institutional crossroads, having just enough critical mass of public attention to warrant a torrent of speculation.

Shawna Ross, “In Praise of Overstating the Case: A review of Franco Moretti, *Distant Reading*”, in: DDQ 8/1 (2014)

http://www.digitalhumanities.org/dhq/vol/8/1/000171/000171.html

Such worries have highlighted the conflict rhetoric characterizing DH meta-discourse — with its familiar debates about distant versus close reading, about tool builders versus tool users, about rejecting versus incorporating theory, about using DH skills to be helpful around the office versus being more protective of our time. What unites these debates is the common tension between, on the one hand, desiring the fertility and excitement created by the

Kritik & Angst

Such worries have highlighted the conflict rhetoric characterizing DH meta-discourse — with its familiar debates about distant versus close reading, about tool builders versus tool users, about rejecting versus incorporating theory, about using DH skills to be helpful around the office versus being more protective of our time.

… on the one hand, desiring the fertility and excitement created by the coexistence of a diversity of approaches and interactions among very differently-minded scholars and, on the other, fearing that collaboration-related compromises could devalue DH work, subsequently re-marginalizing DH workers.

Shawna Ross, “In Praise of Overstating the Case: A review of Franco Moretti, *Distant Reading*”, in: DDQ 8/1 (2014)

http://www.digitalhumanities.org/dhq/vol/8/1/000171/000171.html

Häresie 3

Mir ist völlig egal, was die „Digital Humanities“ sind.

Mich interessiert, welche epistemischen Möglichkeiten sich aus der Anwendung von Algorithmen auf geisteswissenschaftliche Quellen ergeben.

Nennt das Ding doch „Isabella“.

Häresie 4

Jede geisteswissenschaftliche Disziplin gehört zu den Digital Humanities.

Digital Humanities sind jedoch ein intellektuelles Abenteuer und ein anspruchsvolles Unterfangen.

Die korrekte Bedienung eines Smartphones qualifiziert nicht als Digital Humanist.

Thallers Ungleichheiten

Die Digital Humanities sind nicht identisch

(1) Mit digitalen Bibliotheken.
(2) Mit der Korpuslinguistik.
(3) Mit dem Textmining.
(4) …

Sie *nutzen* diese für epistemische Erweiterungen der Geisteswissenschaften souverän.

PS: Souverän genug um keine Angst vor der Informatik zu haben.

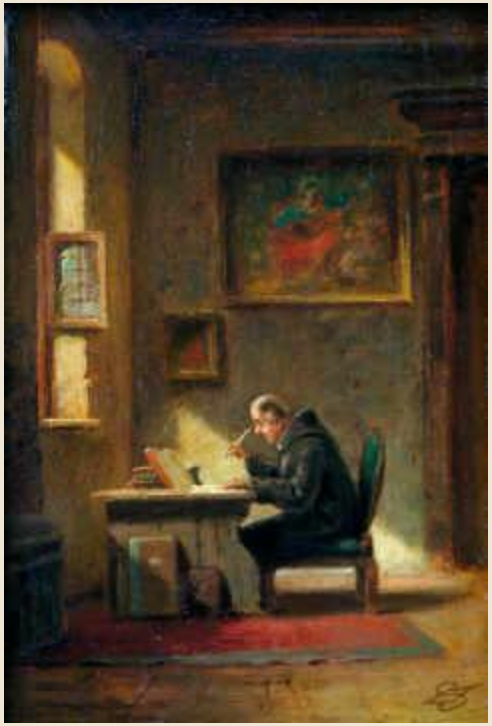
Häresie 5

Die modernen Informationssysteme haben die Bedingungen für die Geisteswissenschaften so gründlich verändert, dass wir den Mut haben müssen, die wissenschaftliche Praxis zu ändern.

Dazu brauchen wir eine ernsthaft geführte epistemische Debatte, in voller Kenntnis der Konzepte der Digital Humanities.

PS: Und keine darüber, wer im Sandkasten mitspielen darf.

Spitzwegs Welt



Spitzwegs Welt

Unsere Welt

Unsere Welt

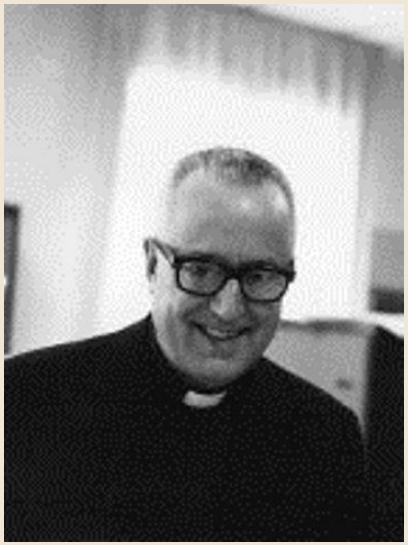
Dies ist die Welt der Informationssysteme.

Sie wartet nicht auf uns.

Wenn die Geisteswissenschaften die IT nicht nutzen, tun es eben andere.

Bitte, bitte: Mut zur Fantasie …

E. E. „Doc“ Smith, 1890 - 1966



Padre Roberto Busa, 1913 -2011

Danke!

manfred.thaller@uni-koeln.de