

Libraries as 'epistemological agents' in Digital Humanities research

Prof. Dr. Jan Christoph Meister
Universität Hamburg

*Reviving the past and keeping up with the future –
the libraries' role in preserving and providing access
to newspapers and news media*



Beth Holland, George Lucas Educational Foundation (2015)

21st-Century Libraries: The Learning Commons

- role of physical libraries may become even more important. They are not just **a place** to house resources, but one **in which to create meaning** from them
- libraries of the 21st century provide a welcoming common **space that encourages exploration, creation, and collaboration**
- they bring together the **best of the physical and digital** to create learning hubs“

*Kate Wittenberg, Council on Library
and Information Resources (2014)*

The Role of the Library in 21st-Century Scholarly Publishing (ctd.)

- librarians have the potential to **play a leading role** in moving forward with new models of scholarly communication
- research library is in a position to **provide both the platform** and
- many of **the skills needed** to enable the creation of new forms of scholarship and to disseminate the resulting content to a wide audience of users

*Kate Wittenberg, Council on Library
and Information Resources (2014)*

The Role of the Library in 21st-Century Scholarly Publishing

- focus on users → on **disseminating information and services that respond to users' needs in whatever form** seems most appropriate to the content
- information literacy → are users actually creating **new ways of evaluating content** that are different from those with which publishers are comfortable?
- role of information professionals → rapid development and change in the current digital environment ... **new models for academic publishing** are needed

Reviving the past and keeping up with the future – the libraries' role in preserving and providing access to newspapers and news media

- “libraries are also agents of media transformation by digitizing their historical holdings of, i.a., newspapers and journals as cultural heritage”
- “how can libraries cooperatively support preparation of the digital raw-material for historical research, text and data mining, other digital humanities or educational purposes?”

Digital Humanities approaches in news and media studies

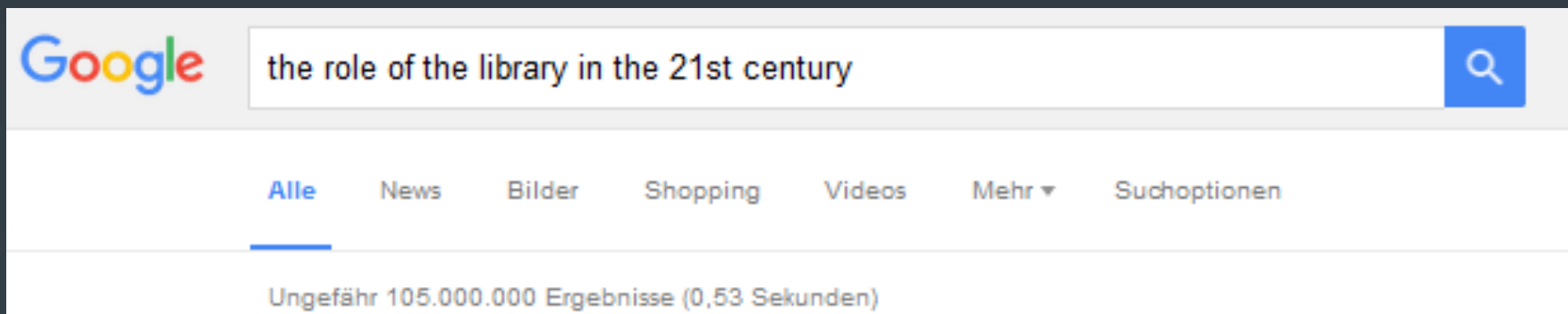
„Smart data“ approaches → indepth *close reading* of small & exemplary samples relies on interactive procedures:

- digitization & creation of a relevant digital corpus
- categorization & annotation of corpus elements via meta-data
- hypothesis driven iterative analysis/interpretation-cycles

„Big data“ approaches → bottom-up *distant reading* relies on automated procedures

- creation of large, indiscriminate collections
- automatic segmentation and structuring of collection items (parsing, normalisation, named entity recognition etc.)
- automatic statistical & ML-based exploration (pattern recognition, **topic modelling** etc.)

Small data/close reading vs. big data/distant reading: an example



https://www.google.de/search?q=future+of+the+library&ie=utf-8&oe=utf-8&gws_rd=cr&ei=ApEUV-qkHoGgsgHHI7LQAQ#q=the+role+of+the+library+in+the+21st+century [18.4.2016]

The Role of the Library in 21st-Century Scholarly Publishing ...

www.clir.org > Publications > Reports > pub142 ▼ Diese Seite übersetzen

The changing landscape of libraries, publishers, and scholarly societies; university views on tenure and digital scholarship; the emerging role of search engines; ...

Du hast diese Seite am 17.04.16 besucht.

21st-Century Libraries: The Learning Commons | Edutopia

www.edutopia.org/.../21st-century-libraries-learning... ▼ Diese Seite übersetzen

14.01.2015 - Printed books still play a critical role in supporting learners, but digital ...

The libraries of the 21st century provide a welcoming common space ...

Du hast diese Seite am 17.04.16 besucht.

The Role of Librarians in the 21st Century - presentations

de.slideshare.net/.../the-role-of-librarians-in-the-21st... ▼ Diese Seite übersetzen

13.06.2007 - The Role of Librarians in the 21st Century. 1. The Role of ... The

Ongoing Debate on the Role of Libraries: 2 Haunting Questions 1.

Libraries Driving Access to Knowledge in the 21st Century ...

www.webpages.uidaho.edu/.../fagbola-uzoigwe-ajeg... ▼ Diese Seite übersetzen

von O Fagbola - Zitiert von: 6 - Ähnliche Artikel

The new role of libraries in the 21st century is to be a learning and knowledge center for their users as well as the intellectual commons for their respective ...

Exploring the role of the 21st century library in an age of e ...

knightfoundation.org/.../exploring-role-21st-centur... ▼ Diese Seite übersetzen

25.02.2012 - Photo Credit: Flickr user Jeff Wilcox. In the age of e-books and online content, what's the role of the 21st century library? That question brought ...

Libraries reinvent themselves for the 21st century ...

articles.chicagotribune.com/.../chi-library-future-20... ▼ Diese Seite übersetzen

12.12.2013 - For centuries, the defining role of the library has been as a repository of books. Now, in the 21st century, the library faces perhaps its most ...

Beyond books: what it takes to be a 21st century librarian ...

www.theguardian.com > ... > All sectors ▼ Diese Seite übersetzen

31.01.2011 - Beyond books: what it takes to be a 21st century librarian ... In all library roles customer service and communication skills are important.

Role of Libraries in the 21st Century - The Stationers ...

<https://stationers.org/.../193-role-of-libraries-in-the-...> ▼ Diese Seite übersetzen

Role of Libraries in the 21st Century. 21 November 2012. david-pearson Libraries have a future, writes David Pearson, but only if they adapt to changing readers ...

21st Century Library Blog

<https://21stcenturylibrary.com/> ▼ Diese Seite übersetzen

24.03.2016 - A 21st Century Library Discussion Forum. ... projects in public libraries, they will continue to balance the importance of funding innovation that ...

[PDF] Community building for public libraries in the 21st century ...

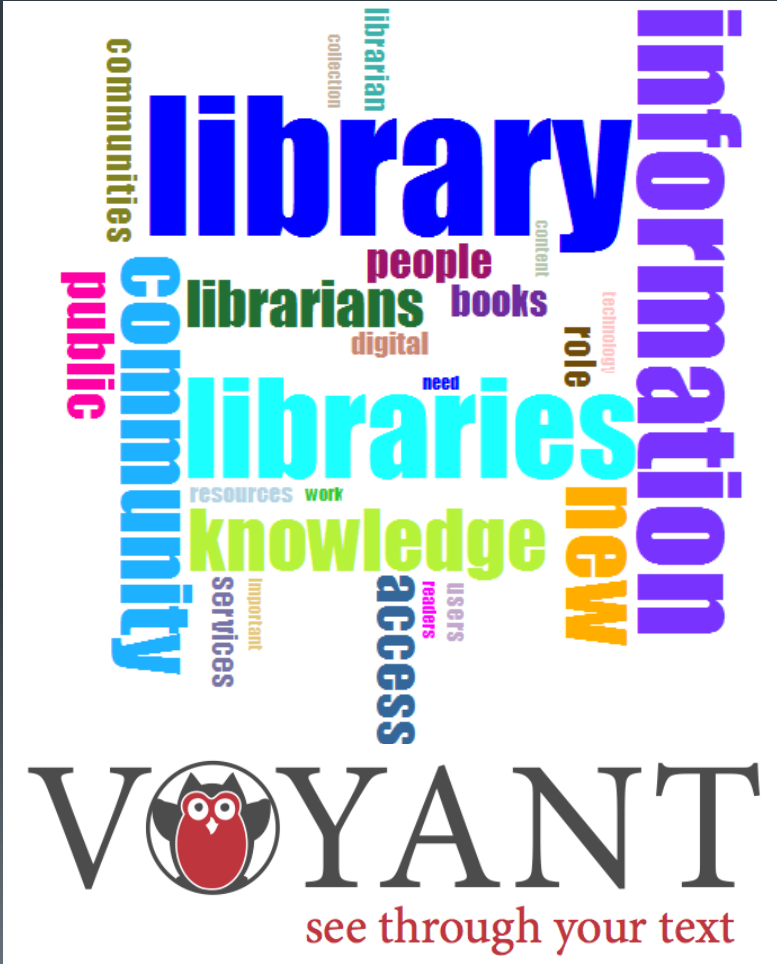
www.ifla.org/...libraries/.../community_building_for... ▼ Diese Seite übersetzen

von H van Duijnhoven - Ähnliche Artikel

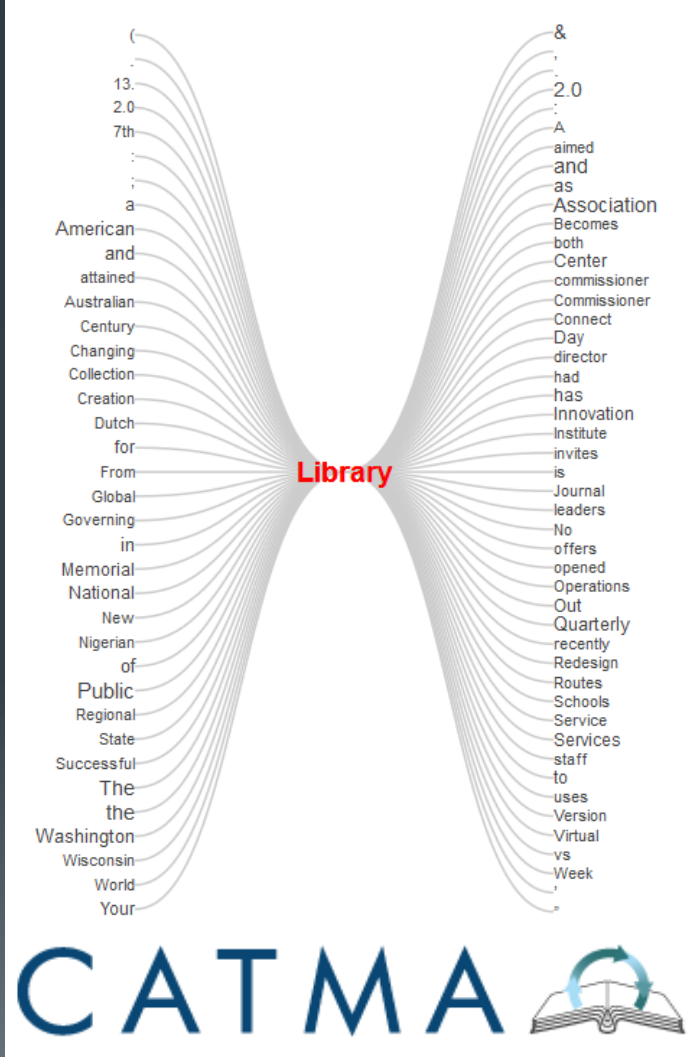
will examine the important role of the librarian in building these communities. In section ... legitimates the function of the public library in the 21st century.

... plus approx. another 104.999.990 documents related to our topic

Interactive *close reading* of the corpus



<http://voyant-tools.org>

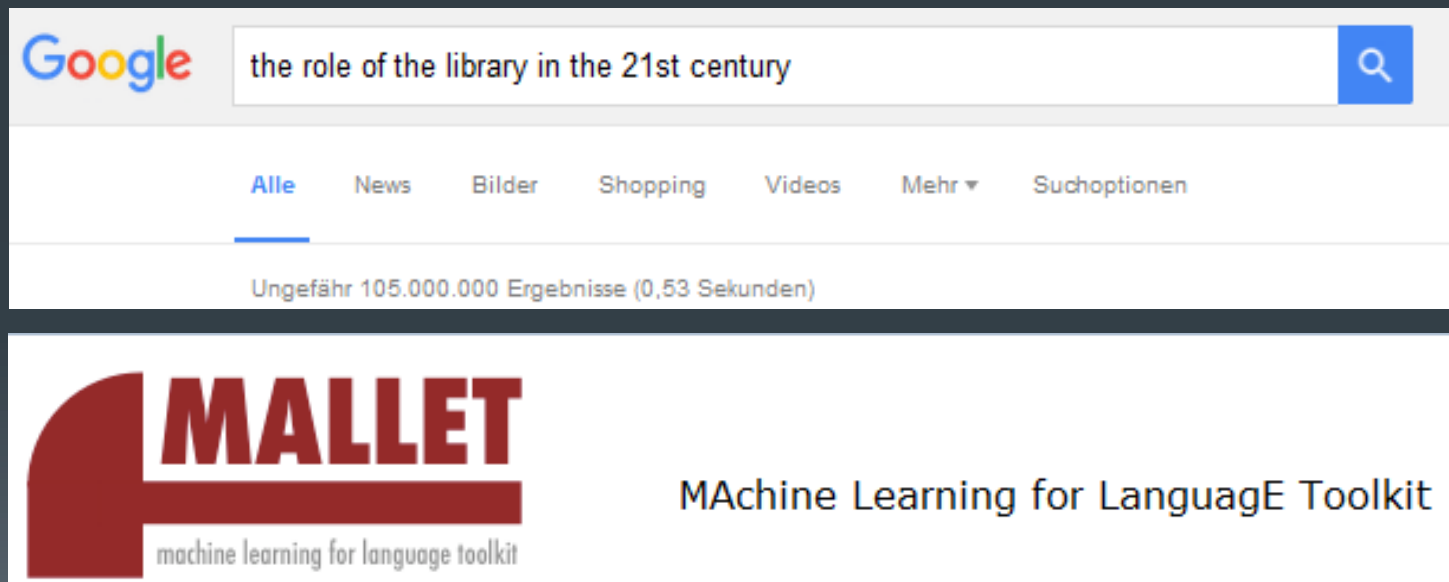


<http://catma.de>

Topic Modelling: a *Distant Reading* method for algorithmic analyses of large corpora

- fully automatic: does not require annotated training data
- purely statistical: does not analyze linguistic structure, but random „bag of words“
- scalable: simple to run and robust to variation in language and data quality
- clustering based: uses contextual clues to group related words and to distinguish between uses of ambiguous words
- synthetic: reduces the dimensionality of a corpus to several hundred clusters or **topics, represented as distributions over the full vocabulary**

Topic Modelling analysis of the first ten documents in a collection of news articles on



<http://mallet.cs.umass.edu/about.php> [20.04.2016]

Results 1

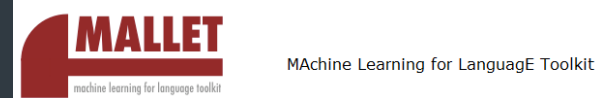


MAchine Learning for LanguagE Toolkit

TopicID	words..
1	resources technology staff traditional management access development social communication provide
2	space digital learning books physical students create printed school collaboration
3	public st brecher century philanthropy center foundation directors arlington importance
4	library services online skills technologies librarian technology project people patrons
5	librarians books people training open health teaching city life day
6	information publishing users publishers scholarly work models based scholars process
7	content community role research digital environment make form changing tools
8	librarians important reference internet change search ability students good stakeholders
9	information knowledge access services user century st users centre critical
10	library libraries share computer age free literacy year formats individual

- remove English stopwords
- 10 topics
- 200 iterations (random „bags of words“)
- output: show first ten words per topic

Results 2



topicId	words..	
1	st century internet books online printed today computer read growth room organize journal stories communities	Digital Era
2	printers arlington row heights studios kuhl memorial impact american chicago directors adults back begun marketplace	
3	space offer circulation longer commons hub design environment role words called transparent opened learners ll	
4	information open journals web understand students found paper pay vision format serves lot levels set	Open Access
5	content community share places communities year free media centers credit million computer miami museum spots	Open Data
6	role based users creating electronic world partnerships increasingly collaborative understanding bring engines disseminating current structure	Users
7	library librarians online project staff librarian find job working good provide jobs teaching librarianship html	Self-Conception
8	future city london world major directors history examples lending country president support outreach previously evolving	Past ∃ Present ∃ Future
9	work resources social academic tools ways focus projects works business include literacy search active vast	Business
10	people training health general central life emma book teams events hoyle michael answer birkwood katie	Work ∃ Life -Balance
11	learning students physical school collaboration meet teachers wanted parker francis photo transformed student comments meaning	Students as users
12	information users librarians communication created institutions participation continue data wide conversation processes introduction facilitate maintained	
13	services important skills patrons time service learn change basic part stakeholders question evaluate areas work	Management
14	access university important librarian databases researchers years recent consult catalogues print real abstracting titles business	Relationship to Universities

Topic 14: *access university important librarian databases researchers years recent consult catalogues print real abstracting titles business*

→ Relationship to Universities

24 technologies reference technology ability schools things skills implement market school scanner troubleshoot classes fact im	Technology
25 publishing publishers scholarly models content research scholars environment process kinds authors scholarship publications peer readers	Publication Issues

- remove English stopwords
- 25 topics
- 500 iterations (random „bags of words“)
- output: show first 15 words per topic
- possibility to **assign interpretative labels to 15-word topics**



Results 3



MAchine Learning for Language Toolkit

topicId	words..	
1	community free share content places social directors opportunities media computer centers physical explore access year	Situation Today
2	access experience place developing collections national making area brought staff position ensure tend searchable human	History and Future
3	work make traditional develop authors focus include quality readers effective evaluating understanding order engines funds	Technology
4	technologies technology ability change patrons time learn things schools basic meet implement management scanner troubleshoot	History and Future
5	future city room director year turn pearson back began train bar reflect rooms digitally money	Students as Users
6	students school project areas good jobs central evaluate problems student working graduate profession comments includes	Social Parametres
7	content environment community creating projects innovative environments credibility social fields collaboration tools players networked active	Librairies as Cultural Institutions (of the past?)
8	libraries books training teaching group children million businesses view research history lost latest expanded end	
9	role based communication librarians databases staff organizations researchers continue relevant critical complex computer large require	Learning
10	space learning physical teachers wanted accessible parker francis commons credit stacks study photo meaning westlake	Librarian's Work
11	library online librarians provide service librarian search work education members long importance groups customer bibliographic	Work - Life - Balance
12	librarians people find health librarianship life day job general technological emma january staff kate events	Opportunities and Problems
13	libraries technology century st called shift innovation poor foster functional cooperation mission types poverty maintained	Societal Role
14	knowledge management user services sharing materials ict networking made supply power institution society full nigeria	Future Concepts
15	library virtual challenges operations growth circulation member societies storage driving transforming lack country emphasis efficiently	Services
16	services skills reference people important reading stakeholders part marketing medium staff class writing traditional ago	History and Future
17	digital create play internet communities past key traditional strong expertise strategies longer indexing structure processes	Future Concepts
18	important changing creation electronic individual literacy created commercial provided years read set increasingly cultural ideas	Paper
19	open printed st paper means hub mail maintain due click opened format page kind younger	Local Examples
20	university world business journals london stay teams librarian conversation major lending facilitate local oldham book	Implementation
21	information resources users development system web variety form institutions issue participation works professional catalog developments	Universities and Scholars
22	publishing publishers scholarly models users research scholars process academic kinds peer publication scholarship publications collaborati	
23	centre internet formats report clients oecd today material providing articles sites consult catalogues documentation centuries	
24	public age brecher philanthropy center foundation support asked private role paid cragg computers angeles los	
25	library spaces space design offer arlington printers chicago row heights studios kuhl memorial cities teens	Space and Location

- remove English stopwords
- 25 topics
- 750 iterations (random „bags of words“)
- output: show first 15 words per topic
- possibility to **assign interpretative labels to topics** increases



...so: what is the libraries' role in this (and in other) Digital Humanities research?

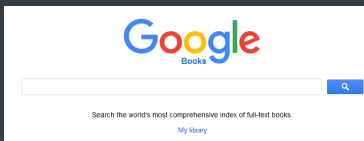
„Topic models are powerful and scale to large data sets, but their ease of use can be deceptive.“

Matthew Jockers & David Mimno (2013)

large data sets: the good, the not-so good and the fake & useless



Deutsches Textarchiv: discipline- and genre independent collection of (digital full) texts in German 1600-1900, approx. 2400 items [4/2016] – excellent, high-quality collection of digital resources



Google Books: currently approx. 30 mio digitized volumes. A powerful resource, but with methodological short comings – lack of transparency, proprietary & copyright issues, cultural bias



Deutsche Digitale Bibliothek: portal linking to content provided by approx. 30.000 institutions - politically motivated, practically useless for DH research

Opportunities vs. risks of digital collection based research into news and media

- large data sets held or made accessible by libraries, such as newspapers & newsmedia archives, invite the application of Digital Humanities methods
- methods & approaches like Topic Modelling will over the next 5-10 years become *business as usual* and increasingly complement qualitative research
- universities and research institutions without a dedicated DH department/center lack the expertise to protect against the risks of unreflected, simplistic „datafication“ research

→ Libraries as *Epistemological Agents*
can provide *data management & analysis expertise*

data management

- aggregation & quality control of large digital corpora
- documentation of rules & procedures followed in corpus composition (reproducibility of results!)

data analysis

- offer *best practice* oriented consultation to researchers
- provide, monitor and maintain a basic suite of robust „tried and tested“ analytical tools that can be reliably used by non-DH specialists (e.g., parsers, segmenters, topic modelling & sentiment analysis tools, visual and sound analysis tools, markup tools, visualizers, etc.)
- provide a **DH helpdesk** for the non-specialist user of DH applications

The library's role as *epistemological agent* in DH research

Conceptual DH: theory, concepts, methods, tools development

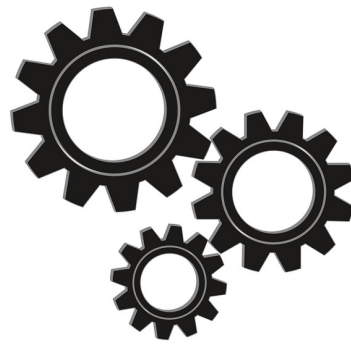
the library's *DH helpdesk service*

Curated digital collections

Non-curated digital collections (Google Books)

Dynamic ad-hoc collections (web crawling)

DH methods & procedures



(quantitativ/qualitative, statistical/hermeneutic, automatic/interactive, etc.)

Applied DH

