Digital Humanities and the Study of the Web and Web Archives

PhD-seminar organised by the Digital Humanities Lab/NetLab

Aarhus University, Denmark

3-5 December 2012

Today the Internet is the medium which holds the most multifaceted set of materials documenting contemporary social, cultural, and political life. It has become the fulcrum for the general development of media, including mass media and a growing variety of digital devices. If the communicative infrastructure of society in the late 20th century was centered on television, it is today centered on the Internet.

As a variety of digital media penetrates all spheres in society, they also play a still more dominant role for the social sciences, humanities and arts. They do so in three respects: as archives for contemporary life, as a toolbox for the study of all sorts of digital collections, including digitized collections of non-digital materials (often labeled cultural heritage), and as a means for enforced communication within all spheres of society.

In this course we will concentrate on the first of these aspects, the role of the Internet and particularly on online and archived web materials for scholars studying contemporary political, social, and cultural phenomena.

The course format

There will be four lectures, and paper presentations. Each PhD scholar will deliver a short paper (6-8 pages, max. 19.200 type units) and a short video-presentation. Each PhD scholar will present the main ideas of their videos and papers during the course (2-3 tracks depending on the number of participants). The videos will be distributed in a Facebook group for all participants on beforehand.

Costs

Participation, meals, and accommodation are offered free of charge. However, beverages must be paid for by the participants. In addition, participants must pay for travel themselves.

Dates and venue

The PhD-seminar will take place 3-5 December 2012 at Hotel Vejlefjord in Denmark (http://www.hotelvejlefjord.dk/en).

Number of participants

20 (maximum).

ECTS-points

4 ECTS.

Seminar enrolment and application deadline

The seminar application, including a one-page abstract, must be sent by email no later than 1 October to Niels Brügger, nb@imv.au.dk. The abstract can be either a project description or an expression of interest in the field. Notification about enrolment will be sent out no later than 5 October.

Participants who are enroled must submit the video and the paper no later than 12 November.

For questions about course content please contact Niels Ole Finnemann, finneman@imv.au.dk.

Seminar readings

Seminar readings will be made available one month before the course, with required reading and suggested literature for the seminar. Students are expected to have read the literature before the beginning of the seminar.

The lectures and the lecturers

Situating Digital Methods in Digital Humanities, E-Sciences, Web Science, Data Science and the Computational Turn

Richard Rogers, Prof. dr., Media Studies, University of Amsterdam

How productive is the ontological distinction between the natively digital and the digitized, that is, between the objects, content, devices and environments that are "born" in the new medium, as opposed to those that have "migrated" to it? Should the current methods of study change, however slightly or wholesale, given the focus on objects and content of the medium? The research program put forward here engages with "virtual methods" that import standard methods from the social sciences and the humanities. That is, the distinction between the natively digital and the digitized also could apply to current research methods. What kind of Internet research may be performed with methods that have been digitized vis-à-vis those that are natively digital (such as recommendation systems and folksonomy)? I strive to shift the attention from the opportunities afforded by transforming ink into bits, and instead inquire into how research with the Internet may move beyond the study of online culture only. How to capture and analyze hyperlinks, tags, search engine results, archived Websites, and other digital objects? How may one learn from how online devices (e.g., engines and recommendation systems) make use of the objects, and how may such uses be repurposed for social and cultural research? Ultimately, I propose a research practice that grounds claims about cultural change and societal conditions in online dynamics, introducing the term "online groundedness."

Prof. dr. Richard Rogers is University Professor and holds the Chair in New Media & Digital Culture at the University of Amsterdam. He is Director of Govcom.org, the group responsible for the Issue Crawler and other info-political tools, and the Digital Methods Initiative, which seeks to rework method for Internet research. Among other books, Rogers is author of *Information Politics on the Web* (MIT Press, 2004), awarded the best book of the year by the American Society of Information Science & Technology (ASIS&T). His latest book, *Digital Methods*, is also to be published by MIT Press (February, 2013). Rogers is a three-time Ford Fellow and has received research grants from the Soros Foundation, Open Society Institute, Mondriaan Foundation and, most recently, the MacArthur Foundation.

Digital Footprints: A method for studying user behavior on Facebook

Anja Bechmann, Assistant Professor at the Centre for Internet Research and Head of Digital Footprints Research Group, Aarhus University

A lot of the data on Facebook, such as wall posts, news feeds, friend-lists, interests etc, is fenced off from the outside world and are only accessible by the owner and the owner's Facebook friends. This has made it nearly impossible to analyze what is going on inside Facebook in terms of data (structures). This talk presents Digital Footprints (digitalfootprints.dk), a newly developed system to help collect and analyze Facebook user data. Digital Footprints is a research group (at the Centre for Internet Research, Aarhus University, Denmark) interested in the data that users share, expose or trade when communicating through the internet. We will present and discuss the system as method and participants will be given a login to the

software and try to use it on small research cases. The workshop will work as a test bed of the software on different research topics. We will end the workshop with a plenum discussion of future developments.

Anja Bechmann is Assistant Professor at the Centre for Internet Research and Head of Digital Footprints Research Group at Aarhus University. Anja Bechmann's current research examines interoperability and personal information data flow in social media across digital media platforms and services. She is initiator and co-developer of Digital Footprints Analytics with lead-programmer Peter Vahlstrup. She is also currently studying design/production processes in selected internet companies with a specific emphasis on virtual production environments and the role of the (participant) user in the internet software development process. She has published in, amongst other places, *New Media & Society* and *Information, Communication & Society*. Her overall research interests are internet culture and industry and include research in new economy business models, virtual production environments, strategies of user relevance through data mining, mobile internet/smartphones, cloud computing, open APIs, interoperability, personal data exchange and privacy regulation on the internet. She is both a board member of the Danish Council for Greater IT Security and the Centre for Internet Research.

Historical Network Analysis of the Web

Niels Brügger, Associate Professor and Head of the Centre for Internet Studies, Aarhus University

This presentation discusses some of the fundamental methodological challenges related to doing historical network analyses of the web based on material in web archives. Since the late '90s many countries have established extensive national web archives, and software supported network analysis of the online web has for a number of years gained currency within internet studies. However, the combination of these two phenomena — historical network analysis of material in web archives — can at best be characterized as an emerging new area of study. Most of the methodological challenges within this new area revolve around the specific nature of archived web material. On the basis of an introduction to the processes involved in web archiving as well as of the characteristics of archived web material, the presentation outlines and scrutinizes some of the major challenges which may arise when doing network analysis in web archives, among others such issues as completeness, construction of a corpus, temporal and spatial inconsistencies, and cross-archive analyses. The presentation uses an ongoing case study to feed the methodological discussion, namely the political network on the web which was available to a voter at the Danish parliamentary elections in 2011, 2007, and 2001.

Niels Brügger is Associate Professor and Head of the Centre for Internet Studies at Aarhus University. His primary research interest is web archiving and web history, and he is writing a history of the Danish Broadcasting Corporation's website www.dr.dk. Within these fields he has published monographs and a number of edited books as well as articles in international journals and edited volumes. He has edited the book *Web History* (Peter Lang, 2010) which puts the emerging field of web history on the agenda of internet research, and his most recent book is *Histories of Public Service Broadcasters on the Web* (co-edited with Maureen Burns, Peter Lang 2012). Since 2010 he is a member of the national Danish Internet Archive, netarkivet.dk's editorial board, and since 2012 he has been a member of the international editorial board of the journal *New Media & Society*. He is co-director of NetLab (netlab.dk), the internet

focused working group within the Digital Humanities Lab, the national Danish research infrastructure within the humanities (http://dighumlab.dk/).

Methodological issues in the study of Danish Internet Favourites in 2009

Niels Ole Finnemann, Professor, MSO, Dr. Phil. Department of Aesthetics and Communication. Aarhus University, Denmark

The project "Danish Internet Favourites 2009" aims to analyze a corpus of favourite sites, identified in a representative national survey conducted in 2009 (a report of the survey can be found at www.changingborders.au.dk). The survey is used to establish a list of favourite sites as identified by 1150+ respondents, but since the sites themselves often have been changed or even disappeared the content can only be found in the official Danish Internet archive, netarkivet.dk (netarchive.dk) if at all. The list of favourite URLs will be used to retrieve the sites from the archive. The lecture will discuss issues related to the establishment of the corpus and to three different perspectives for the analysis:

1) To validate the so far more or less unverified archive which was established in 2005 as a result of a parliamentary decision. The validation includes variety of parameters for completeness which will be discussed. It is assumed that the archive is incomplete both in well-known respects (e.g. no Flash) and in many less predictable respects (see Brügger and Finnemann, 2012, in review). Finally this part aims also at validating the official Danish archiving strategy based on 'snapshot' harvesting of .dk + so-called Danica, selective harvesting and event harvesting.

2) To further analyze "the long tail" pattern of favourite sites which turned up in the survey data and to look into the thematic content (possibly low level semantic analysis of data, meme-tracking, content analysis and analysis of typical sets of favourites) — all of it depending on the character of the materials in the archive.

3) To contribute to the development of a research infrastructure for the study of web materials and archived web materials.

The presentation will conclude with further reflections on the particular character of web materials and archived web materials, the concept of digital media in the plural and the relations to "new media " and internet studies on the one hand and digital humanities on the other.

Niels Ole Finnemann Nielsen is co-director of Netlab (netlab.dk), part of the / or a theme / or a theme leader of the Danish Digital Humanities Lab (http://dighumlab.dk/), focusing on the study of online as well as archived web materials in collaboration with the official Danish Internet archive, netarkivet.dk. He was formerly Director for Center for Cultural Studies at the Faculty of Arts, Aarhus University (1996-1999) and Head of the Centre for Internet Studies, Aarhus University (2000-2010). He has been a member of the steering committees of a number of national research projects and EU COST projects. He has served as expert evaluator for the Research Councils of The Netherlands and Norway, and for the EU FP7-programme. He is currently member of "Fagkomiteen for samfundsvetenskap", FRISAM, The Norwegian

Research Council for Social Sciences. He has published a number of articles on media theory and the concept of digital media, in international journals and anthologies.

The seminar is organized by Professor Niels Ole Finnemann, Center for Internet Studies & Ass. Professor Niels Brügger, Head of Center for Internet Studies, Aarhus University. Niels Ole Finnemann and Niels Brügger are also joint leaders of NetLab - a research infrastructure project which is a part of Danish Digital Humanities Lab.