International Symposium on Digital Humanities,
Växjö 7-8 November 2016

Book of Abstracts

Linnæus University
Sweden

ORGANIZATION

Conference Organization

Conference Chairs:
Koraljka Golub, Linnaeus University
Marcelo Milrad, Linnaeus University

Programme Committee Members:
Mats Dahlström, University of Borås
Elisabet Göransson, Lund University and Stockholm University
Jutta Haider, Lund University
Isto Huvila, Uppsala University
Annemarie Lloyd, University of Borås
Maija Paavolainen, University of Helsinki Library
Franjo Pehar, University of Zadar
Michael Pidd, University of Sheffield
Annika Rockenberger, University of Oslo
Henriette Roued-Cunliffe, University of Copenhagen
Nevena Skrbic Alempijevic, University of Zagreb
Kim Tallerås, Oslo and Akershus University College
Mikko Tolonen, University of Helsinki
Francesca Tomasi, University of Bologna
Marijana Tomic, University of Zadar
Johan Åhlfeldt, Lund University and Umeå University

Organizing Committee:
Börje Björkman, Linnaeus University
Robert Bunjaku, Växjö City Library
Tamara Laketic, Linnaeus University

Session chairs
Aris Alissandrakis, Linnaeus University
Joacim Hansson, Linnaeus University
Isto Huvila, Uppsala University
Johan Höglund, Linnaeus University
Koraljka Golub, Linnaeus University
Marcelo Milrad, Linnaeus University

Small cover image: "Devil Codex Gigas", copyright: Kungl. biblioteket, via Wikimedia Commons).

Preface

Digital Humanities (DH) is an emerging field that lies at the intersection of Information Technologies (IT) and Humanities. The field includes both scholars and practitioners in a wide range of disciplines that comprise diverse but complementary areas like data mining, interactive visualization, GIS (Geographical Information Systems), multimedia games, digital story telling, library and information science, history and archeology to name a few. More recently, DH is starting to draw attention from different IT companies, the public sector and commercial actors. These latest developments are starting to pave new business opportunities for these companies and at the same time bringing new funding possibilities to the humanities.

The unique cross-sector and cross-disciplinary Digital Humanities Initiative at Linnaeus University[1] is proud to organize the *first International Symposium in Digital Humanities* to be held in Sweden, on November 7th-8th 2016 in Växjö[2]. Linnaeus University has become the first Swedish university (thereby making Sweden the second Nordic country) to join DARIAH-EU[3], Europe's largest initiative on DH spanning 18 countries, which is therefore supporting the Symposium. Linnaeus University aims to act as a strong driving force in the region and boasts well-established, close collaborations with the surrounding business sphere and the community at large. Due to this specific orientation, this Symposium is co-sponsored by Växjö City Library, Växjo Kommun, the Big Data Consortium with funding from Region Kronoberg & Regionförbundet i Kalmar län, and IEC – Information Engineering Centre.

In the call for papers we have invited researchers and practitioners in related disciplines to Digital Humanities to present, discuss and demonstrate different possibilities, current efforts and upcoming trends in this emergent field. The responses have been very encouraging and we have received 26 different contributions. Each submission has been peer-reviewed by at least two reviewers. The Programme Committee (PC) consists of 16 members coming from the UK, Italy, Croatia, Finland, Denmark, Norway, and Sweden. The final program boasts 2 invited speakers, 1 panel, 6 sessions with a total of 17 presentations, and a poster session with 4 submissions. The authors come from 10 countries on 3 continents; while the majority are from universities and related academic institutions, there are also 3 representatives from the external public sector (National Library of Norway, Det Fria Ordets Hus (The House of Free Speech from Växjö) and Musik i Syd (a Swedish music institution in the regions of Skåne and Kronoberg).

Of the invited speakers, Marianne Ping Huang, Development Coordinator at Aarhus University and the representative for Denmark in DARIAH-EU, will talk about Digital Humanities and DARIAH-EU. Theresa Anderson, Head of the Connected Intelligence Centre at University of Technology, Sidney will deliver an invited talk entitled "*Keeping the Human in the Digital: Challenges of Technology and Transdisciplinary in a Data-Intensive World*".

The panel entitled "*Digital Humanities in the Nordic Countries: Current Efforts, Perspectives and Challenges*" will be delivered by representatives of the four Nordic countries. Presentations are organized into six sessions: 1) Digital Humanities and Linguistics, 2) GLAM and Textual Scholarship, 3) Interactive Tools, 4) DH across Disciplines, 5) DH across Sectors, and 6) GIS and Visualisation.

[1] http://lnu.se/en/digihum/
[2] https://lnu.se/en/research/conferences/international-digital-humanities-symposium/
[3] http://dariah.eu

We strongly believe that this Symposium and the rich discussions that will arise will be thought provoking for the participants and will play a key role in fostering a larger community of Nordic DH researchers and practitioners in the coming years.

Special thanks to the invited speakers and all the authors for making this Symposium on Digital Humanities such a rich and diverse event in terms of content and future collaborative possibilities. Our sincere gratitude also to the PC members who have provided their valuable reviews and have done so on a tight schedule! And not the least, we highly appreciate the important support of our sponsors – they have made it possible to arrange a registration-free two-day event with all meals and coffee breaks included.

Koraljka Golub and Marcelo Milrad, Linnaeus University

Växjö, November 2nd 2016

Posters

Keeping the human in the digital: challenges of technology and transdisciplinarity in data-intensive worlds

Theresa Dirndorfer Anderson
Connected Intelligence Centre (CIC)
University of Technology Sydney (UTS)
Email: theresa.anderson@uts.edu.au

In this presentation, Anderson will share some of the design principles that have informed a data science program that draws on theory and practice straddling the humanities, social science and computational disciplines. The first two years of this program's operation reveal insights about opportunities for transdisciplinary approaches to tackle 'wicked' problems. In the MDSI, students tackle large, complex data challenges within a human-centred data science program. Students are encouraged to develop a critical mind that thinks ethically and systemically about the uses of data and analytics and to embrace the ambiguities and uncertainties (both present and emerging) that working with data "in the wild" entails. Engaging with the analytical, ethical and creative challenges of data science practice foregrounds social concerns about data and technology.

Digital Humanities in DARIAH EU

Marianne Ping Huang
Development Coordinator, Arhus University

How do Digital Humanities and digital research infrastructures fit? How much tech, how many researchers, how many partners does it take to forge a dynamic and scalable network, with a fitting layer of services? Digital Research Infrastructure for the Arts and Humanities, DARIAH-EU is built for researchers, by researchers – the presentation will give an impression of mission, scope, dynamics, and challenges of DARIAH-EU. And suggest how to collaborate on European strategies for digital transformation in the arts and humanities.

Digital Humanities in the Nordic Countries: Current efforts, perspectives and challenges

Koraljka Golub
Department of Cultural Sciences
Linnaeus University, Sweden
koraljka.golub@lnu.se

Marcelo Milrad
Department of Media Technology
Linnaeus University, Sweden
marcelo.milrad@lnu.se

Marianne Ping Huang
School for Communication & Culture
Faculty of Arts, Aarhus University, Denmark
mph@cc.au.dk

Mikko Tolonen
Department of Modern Languages / Heldig, Helsinki Digital Humanities Centre
University of Helsinki, Finland
mikko.tolonen@helsinki.fi

Inés Matres
Department of Philosophy, History, Culture and Art Studies
University of Helsinki, Finland
ines.matres@helsinki.fi

Andreas Bergsland
Department of Music
Norwegian University for Science and Technology (NTNU), Norway
andreas.bergsland@ntnu.no

BACKGROUND AND MOTIVATION

The particular exploration of new ways of interactions between society and Information Communication Technologies (ICT) with a focus on the Humanities has the potential to become a key success factor for the values and competitiveness of the Nordic region, having in mind recent EU and regional political discussions in the field of Digital Humanities (European Commission, 2016; Vetenskapsrådet's Rådet för forskningens infrastrukturer, 2014). Digital Humanities (DH) is a diverse and still emerging field that lies at the intersection of ICT and Humanities, which is being continually formulated by scholars and practitioners in a range of disciplines (see, for example,

Svensson & Goldberg, 2015; Gardiner & Musto, 2015; Schreibman, Siemens, & Unsworth, 2016). The following are examples of current areas of fields and topics: text-analytic techniques, categorization, data mining; Social Network Analysis (SNA) and bibliometrics; metadata and tagging; Geographic Information Systems (GIS); multimedia and interactive games; visualisation; media.

DARIAH-EU (http://dariah.eu), is Europe's largest initiative on DH, comprising over 300 researchers in 18 countries, thereby opening up opportunities for international collaboration and projects. Of Nordic countries, Denmark is the full partner with four universities, Copenhagen, Aarhus, Aalborg and University of Southern Denmark. Danish DARIAH activities are facilitated by the national DH Infrastructure DIGHUMLAB, hosted at the DARIAH-DK coordinating institution, Aarhus University. Sweden's first university, Linnaeus University, joined in May 2016 as a collaborative partner. Finland (University of Helsinki) and Norway (Norwegian University of Science and Technology) have taken action to join by the end of 2016 also, as collaborative partners.

The idea of forming the Nordic Hub of DARIAH-EU has been recently brought forward by Mike Mertens, the Head of DARIAH-EU. It was met by enthusiasm among the representatives of the listed universities, who would at this panel like to present and discuss possible joint opportunities and challenges in Nordic DH. With its tradition in supporting the Humanities research and development, Nordic countries may serve as a bastion for (Digital) Humanities. The Nordic Hub of DARIAH-EU may pay the way forward towards reaching that aim.

CURRENT EFFORTS (PROJECTS, R&D AND EDUCATION)

Koraljka Golub and Marcelo Milrad will present and analyse the cross-sector and cross-disciplinary Digital Humanities Initiative at Linnaeus University along the axes of its strengths, weaknesses, opportunities and threats. Their long-term vision is to create a leading education in this field and to establish a leading research regional centre that combines in novel ways already existing expertise from different departments and faculties working in close collaboration and co-creation with people and different organizations (both public and private sector) from the surrounding society.

Marianne Ping Huang will present the Danish DIGHUMLAB, with specific focus on the DARIAH-related activities with the national DH research infrastructure, which is now bidding for its second period on the DK road map for research infrastructures. The presentation will include a presentation of the Danish participation for DH and open cross sectorial innovation with DARIAH Humanities at Scale (2015-2017) and how this intersects with Aarhus being European Capital of Culture 2017.

Mikko Tolonen and Inés Matres will present the ongoing development at the University of Helsinki (and in Finland) regarding digital humanities. This includes the recently launched Heldig (digital humanities centre https://www.helsinki.fi/en/researchgroups/helsinki-digital-humanities) and how it can relate to collaboration in Dariah. Matres will also discuss a survey about Finnish needs about networking within Dariah that she has been mapping before Växjö conference.

Andreas Bergsland will discuss the role of digital humanities (DH) at NTNU, a technology and science university with an explicit mission of public humanities. He will present several ongoing initiatives: the ARTEC interdisciplinary task force, DH projects in linguistics, print and sound cultures, digital storytelling and performance, archives, and pedagogy. Bergsland will conclude by reflecting on the opportunities and challenges of promoting critical, sustainable, ethical, and interdisciplinary DH at NTNU in partnership with academic and nonacademic actors and local and global networks such as DARIAH.

PROSPECTS AND CHALLENGES

Major opportunities in DH in Nordic countries lie in the collaborative democratic tradition that supports re-combining already existing expertise and resources encompassing 1) different universities, 2) various disciplines, and 3) the wider community through input from related public and private sectors. These points serve to unite and consolidate already existing expertise in order to create new constellations for collaboration leading to new knowledge and products (expertise, education, research, public and relevant commercial services). Possibilities to collaborate across Nordic countries can take place at a number of levels, including joint research and innovation, education efforts, expertise and experience exchange, bringing in international views to address more regional challenges. Ensuing important value for the general public could be a (re)-affirmation of the value of humanities in particular, and academic practices in general.

Challenges would be discussed in terms of the emerging job market, the low number of students pursuing

carriers in humanities at the Master level (e.g., in Sweden), and the fact that DH as a field is still in its infancy, leading to it being quite difficult to get funding and grants to carry out long-term research that sustain our efforts over time. Not the least, epistemological, conceptual and terminological differences in approaches by the different disciplines and sectors may present further challenges and therefore may require additional resources to reach an understanding. Further, while there is a strong collaborative spirit across Nordic countries, there will certainly be administrative issues with cross-university collaboration as the current working structures are based on individual units.

REFERENCES

European Commission. (2016). *Horizon 2020: Social Sciences & Humanities*. Available at https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-humanities

Gardiner, E. and Musto, R. G. (2015). *The Digital Humanities: A Primer for Students and Scholars.* Cambridge: Cambridge University Press.

Vetenskapsrådet's Rådet för forskningens infrastrukturer. (2014). *Områdesöverikt för forskningens infrastrukturer*. Available at http://www.vr.se/download/18.2302fa711489c4798d4a35fa/1411461229423/Samtliga+områden+infrastruktur.pdf

Schreibman, S., Siemens, R., and Unsworth, J. (2016). *A New Companion to Digital Humanities.* (2nd ed.). Malden, MA; Chichester, West Sussex, UK: Wiley-Blackwell.

Svensson, P., and Goldberg, D. T. (Eds.). (2015), *Between Humanities and the Digital*. Cambridge, Ma.: MIT Press.

PEAR 4 VXO: A case study using an Augmented Reality framework to facilitate public engagement

Aris Alissandrakis and Nico Reski
Department of Media Technology
Linnæus University, Växjo
aris.alissandrakis,nico.reski@.lnu.se

1 Introduction

Although the concept of Augmented Reality (AR) has been around for some years, particularly taking advantage of the availability of smartphone and tablet devices, it has not quite reached the mainstream[1]. However, the recent success of the mobile game *Pokemon GO* has considerably contributed to the public's interest in mobile AR.

This paper presents our initial efforts towards a framework for facilitating public engagement through mobile AR, that fall under the overall project title "Augmented Reality for Public Engagement" (PEAR).

Althoughboth Augmented and Virtual Reality technologies are becoming more accepted, their adoption for interdisciplinary purposes, especially in relation to the Humanities, is usually focused on the visual aspect (e.g. recreating a historical location, or digitising and displaying artefacts)[2] and thus not taking advantage of another very important aspect which is the possibility to *interact* with the virtual content.

We position our work in relation to the field of Digital Humanities as providing a framework that allows people to engage and interact in-situ with virtual content that can be the visualization of people's consensus, but also an aggregation of multiple sources of data.

2 Concept and Interaction Design

The purpose of the PEAR framework is to visualize data from (and for) the public discussion about a certain issue, in an interactive and engaging way, thus creating additional value. It allows the public to both

- be informed (by viewing the visualization using the AR mobile app), and also
- participate in the discussion by using social networks, or by interacting with the AR visualization.[3]

The data visualization is continuously updated, based on the public interaction. More specifically, as users debate about a topic on social media networks, such as e.g. Twitter

[1] Mainstream adoption for AR technologies is predicted within the next 5-10 years, see *Gartner Hype Cycle for Emerging Technologies Report 2016* (http://www.gartner.com/newsroom/id/3412017).

[2] See for example the two projects http://www.dukewired.org/projects/visualizing-venice and https://augmentedpalimpsest.wordpress.com.

[3] Currently, only participation via social networks (Twitter) is implemented in the prototype. Direct interaction with the AR visualization within the mobil app is a feature under development.

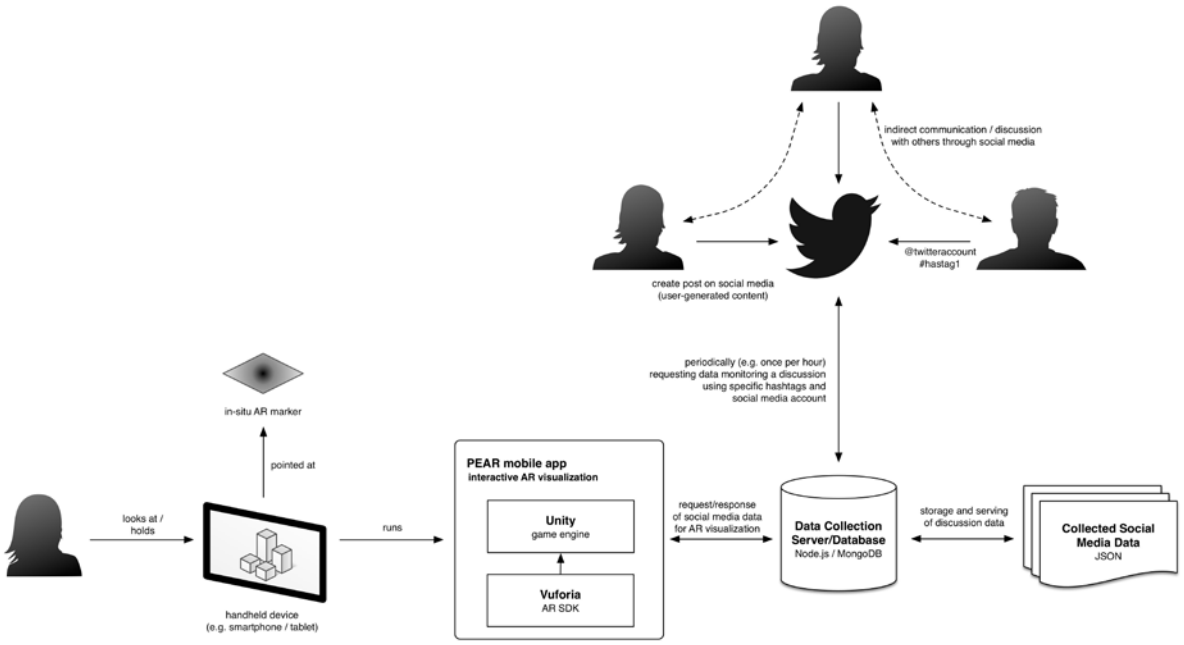


Figure 1: Overview of the PEAR architecture.

or Facebook, public data can be collected using certain keywords or hashtags. These data can be analysed and ranked (according to the frequency of different hashtags), and a visualization (representative of the current discussion) can be created. The data visualization can be designed to additionally highlight any recent changes (between previous and current updates), providing further insights.

An essential part behind the idea of PEAR is the strategic placement of the AR marker (needed to show the visualization), ideally leading the user to a real-world location that is connected to the discussed topic, e.g. a place of interest, a historical site, a café in town, etc. The encouragement of the user to physically move to the real-world location in order to observe (via the AR visualization) the latest updates about the discussion aims to facilitate the users engagement in the discussion as new thoughts, ideas or opinions may be formed through the visit of the site. It can be argued that users could obtain such information at home, online; however, they would be at a (both geographical and psychological) distance. Providing access to "live" information in-situ may encourage participation and involvement, and also allow them to reflect from a closer perspective.

The implemented prototype of the framework is divided into two parts: a data collection (server-side) part, and an interactive AR visualization (client-side) part (see Figure 1).

3 PEAR 4 VXO: A case study

In collaboration with the *Växjö kommun*, an implementation of the PEAR framework was deployed related to an ongoing public debate regarding the future development of the *Ringsberg/Kristinebergs* area in the city of Växjö, Sweden. The public was invited to vote about this issue by composing a tweet that included `@vaxjokommun` (the kommun's Twitter account) and specific hashtags, such as e.g. `#parkRK` or `#matRK` (indicating a preference to either develop a park or a restaurant/café at the site).

Figure 2 (left) illustrates the overall workflow. Figure 2 (right) shows the deployed on-site stand. The top side of the stand, besides the AR marker, includes information about

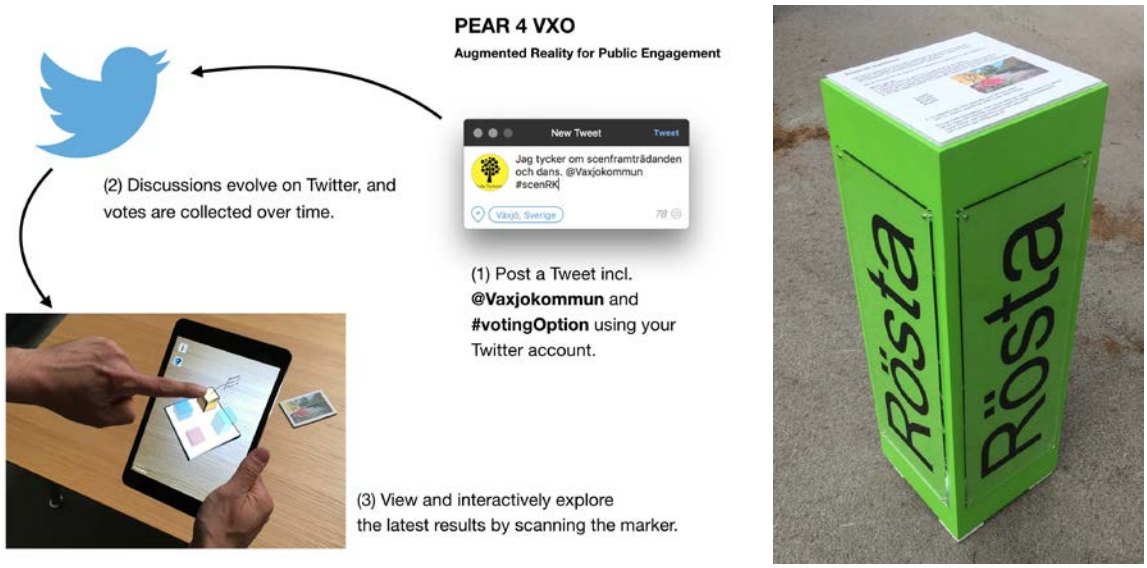


Figure 2: Overall workflow of the deployed PEAR 4 VXO campaign (left), and photo of the on-site stand with info and AR marker (right).

the campaign (it invites the public to visit, download the AR app, scan the AR marker to view the latest results, as well as obtain overall impressions of the discussed area while they are there). Figure 3 gives a visual impression of using the app.

The PEAR 4 VXO campaign ran from end of May to end of August 2016. There was a kickoff event during the "Växjö Värstad" town festival on May 28. The overall campaign was advertised in the April and May newsletters of the kommun. On May 26, two days before the kickoff, two news items[4,5] were released by the kommun. An article in the local newspaper about the Växjö Värstad festival mentioned the campaign. On June 7 there was an event related to the Digital Humanities initiative[6], where the PEAR 4 VXO app and overall PEAR concept were presented. On June 17, there was a press release[7] by Linnæus University (LNU). On June 20, there was a post[8] on the kommun's Instagram account.

Although the campaign was advertised in a number of ways, the outreach and engagement with the public (in relation to the PEAR 4 VXO aspect) fell short of the expectations. A number of possible reasons for this are discussed below.

Although a version of the mobile app was also made for iOS devices, it could not be available online[9] on the App Store in time for the campaign launch. As the market share of iOS devices in Sweden is greater than for Android devices, this was a limitation.

Regarding the participation, it is possible that using Twitter was not familiar, at least for the Växjö public. Given the growing popularity of the social network in Sweden[10] this

[4] https://www.mynewsdesk.com/se/vaxjo_kommun/pressreleases/nu-kan-du-paaverka-utvecklingen-av-ringsberg-kristineberg-via-din-smartphone-1416605

[5] https://monicaskagne.wordpress.com/2016/05/26/chans-att-paverka-hur-vill-du-att-ringsbergkristineberg-ska-utvecklas

[6] https://lnu.se/en/research/searchresearch/digital-humanities

[7] https://lnu.se/en/meet-linnaeus-university/current/news/2016/new-smartphone-app-aims-to-increase-public-engagement-in-local-issues

[8] https://www.instagram.com/p/BG35Tf-oJy8/?hl=en

[9] It was possible to install the app given physical access to an iOS device, but this was not a viable option for the general public.

[10] In Sweden, 22% of internet users use Twitter, see *Svenskarna och internet 2015: En årlig studie av svenska folkets internetvanor* (http://www.soi2015.se).

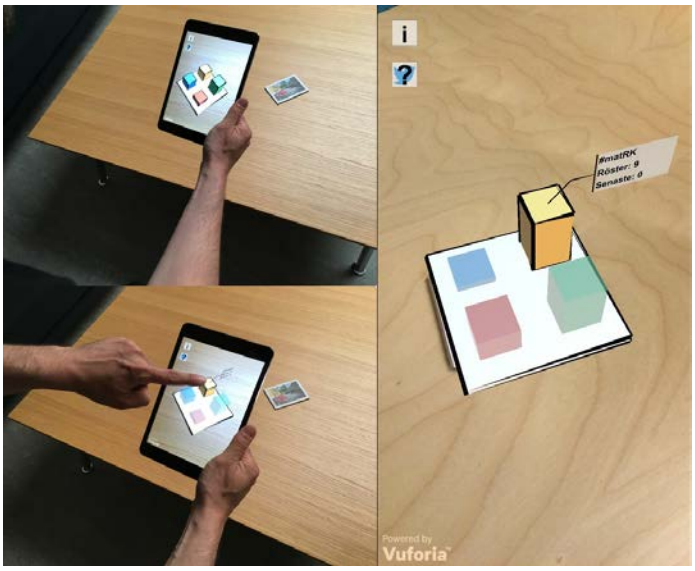


Figure 3: Photos from the prototype PEAR 4 VXO app in action.

hopefully will be less of an issue in the future.

Additionally, and perhaps more critically, although the campaign by the kommun mentioned the existence of the mobile app, they did not provide direct links and information on how to download it, or shared the documents showing how to use Twitter to vote (as seen in Figure 2, left). Besides the kickoff, two "bumps" in the activity happened around a) the Digital Humanities event (where a link to the Google Store was provided), and b) the press release by LNU (which again included this information) and the Instagram post (that possibly made people to visit the site, and interact with the stand/AR marker).

Nevertheless, there was at least a low level of user activity throughout the campaign, that visited the on-site marker in the Ringsberg/Kristinaberg area to check the latest voting results. Furthermore, the collected data and server logs point towards an overall successful technical implementation and deployment, as well as testing of both the server- and client-side parts of the PEAR framework, over the course of a three month campaign.

4 Conclusion

In this paper we introduced the PEAR framework, an approach of using mobile AR to facilitate public engagement in (online) debates and discussions. While there was a limited response and engagement from the public during the PEAR 4 VXO deployment, the results indicate a successful deployment and functionality – and thus serve as overall extended technical validation of the PEAR framework's concept and implementation.

As these were initial efforts, more studies (deployments and data collections) need to be conducted in the future in order to address the potential facilitation of the public's engagement using the PEAR framework. We also see potential in the extension of the implementation, including data visualization and collection from multiple sources at the same time, e.g. Twitter, Facebook, online databases, and sensors. Adapting, or rather extending, the architecture to additionally visualize sensor data in-situ is a future research direction.

Digital Humanities: An Exploration of a New Digital Humanities Programme in Higher Education and its Meaning Making by Community Partners

Soniya Billore
Department of Marketing
School of Business and Economics, Linnaeus University, Sweden
soniya.billore@lnu.se

Koraljka Golub
Department of Library and Information Science
School of Cultural Sciences, Linnaeus University, Sweden
koraljka.golub@lnu.se

INTRODUCTION

Today we live in a rapidly evolving environment supported by high-quality technology and fast communication, which has impacted professional, commercial and academic aspects of the society. In higher education there is a constant need to keep respective fields up-to-date with the changing needs of the society. A comprehensive approach to address this challenge is the creation of interdisciplinary or cross-disciplinary programmes. The field of Digital Humanities (DH), lying at the intersection of humanities and computing is such an example which may be instrumental in creating the change by serving as a platform for novel interdisciplinary connections to address societal challenges. DH has been identified as a tool that assists in furthering research, studies and collaboration between distantly related fields which have not been connected conventionally in the discipline of humanities (e.g., Kirshenbaum, 2010).

Further, higher education and its array of offerings is influenced by market economies. For example, Newman, Couturier and Scurry (2010) study the dynamics of market influences on higher education and point to the inevitable transformations due to agendas of political parties, public policies, corporate goals and market structures. Recent EU political discussions emphasize the need for addressing societal needs through fields like DH, where Humanities and Social Sciences are viewed as "essential to maximise the returns to society from investment in science and technology" (European Commission, 2016).

Higher education is viewed as central to the objectives of public and private institutions and related economic development, employment, skills development and talent acquisitions. The development of new courses with a unique course offering helps to position the universities and higher education institutions differently from the others and helps counter global competition. The local community stakeholders can provide assistance in identifying the required set of skills for the job market. They can guide on how interdisciplinary connections can be built up so as to expand the possibilities. Not the least, courses and programmes informed by actual needs will enable the highly skilled professionals, whose education has been markedly enhanced by practice-informed education and joint, cross-sector innovation, to address future societal challenges. All this is applicable to the field of humanities and the newly evolved Digital Humanities.

METHODOLOGY

The paper reports on a pilot study undertaken with the purpose of informing the developments of DH courses at Linnaeus University (LNU) in the way that reflects actual societal needs, based on input from the relevant stakeholders in the LNU region encompassing south-eastern Sweden. To this purpose, a focus group interview of representatives from relevant public institutions and organizations was conducted. While the invitation was sent out to 9 representatives from the LNU's DH Initiative network (https://lnu.se/en/research/searchresearch/digital-humanities/), 4 were able to attend the focus group interview, and 1 provided his views via an email interview. The five participants represented:

1. Kulturparken Småland (encompassing the biggest group of local cultural venues, http://www.kulturparkensmaland.se/1.0.1.0/108/1/) in Växjö,
2. Kalmar Castle (http://www.kalmarslott.se) in Kalmar,

3. Det fria ordets hus (The House of Free Speech, http://www.vaxjo.se/-Det-fria-ordetshus-/Om-Det-fria-ordet-hus/) in Växjö, and
4. AV Media Region Kronoberg (publically financed institution for media and IT in schools, http://avmedia.kronoberg.se).

The focus group interview and the email interview were of a structured type and comprised 11 questions, grouped around the following 4 subsections:

1. Views about DH as a general topic and as the specific LNU initiative;
2. Relevance of DH to the Linnaeus University region;
3. The skills and talent pool in the market in relation to DH; and,
4. Points of attention for the DH project to aid needs of the society and local industry.

RESULTS

The participants largely agreed that DH is a very broad topic that needs concrete references to applications in order to relate it to various industries. DH can be both a method and a way of thinking and can be used in optimal ways to engage people and customers in their work and social engagement; for the latter, it is important to devise new modes of participation of end users in order to collect their ideas and interests, and create new content. The LNU DH Initiative was hailed as an excellent idea allowing for cross-sector collaboration, and for humanities to connect to computing and business disciplines more substantially.

The overall perception was that the Linnaeus University region needed to be developed in many sectors, ranging from business, culture, education, innovation etc. The project could contribute to and engage with society through schools, museums and other public platforms. Sustainability of the DH Initiative would need to be addressed. The respondents had a consensus that strategic thinking, though very important, was largely lacking in the overall societal development and in the related plans in the local municipal bodies. The DH programme was therefore recommended to include education related to strategy and strategic thinking. The DH programme is also expected to include communication skills building and working in multidisciplinary teams. Finally, the respondents pointed out that as a higher education provider in Sweden it was important to remember that the DH programme needs to keep the three pillars of Education, Research and Engagement with Society as primary in their course development.

REFERENCES

European Commission. (2016). Horizon 2020: Social sciences & humanities. Available at https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-humanities.

Kirschenbaum, M., 2010. What is digital humanities and what's it doing in English departments? *Debates in the digital humanities*, ADE Bulletin, 150, pp.1-7.

Newman, F., Couturier, L., & Scurry, J. (2010). *The future of higher education: Rhetoric, reality, and the risks of the market*. San Francisco: John Wiley & Sons

A Cross-Cultural Study of Attitudes to Digital Tools Among Students and Teachers in the European Language Classroom

Soniya Billore
School of Business and Economics
Linneaus university
Soniya.Billore@lnu.se

Christina Rosén
Department of Languages
Linneaus university
Christina.rosen@lnu.se

BACKGROUND

What role does information technology and digital media play in the language classroom? In Sweden, a survey of adults' and young people's computer usage showed that 69% of the 9 to 14-year olds used the computer for computer games an average week (Svensson 2008, pp.21). However, the use of digital media in language teaching is still underutilized (Kuang-wu Lee, 2000). In a study conducted by Sundqvist & Wikström (2015, pp.74) their findings "indicate a positive relation between gameplay and L2 English, at least for boys". Their explanation for boys being better was that much fewer girls were gamers. Al-Jarf (2004) elaborated on the use of web based lessons and when used as a supplement to class room teaching this was found to be more effective than the traditional pedagogical style, dependent only on the text book. In 2006 the European Commission stated eight common key competences for life-long learning (European Commission 2010). Five of these key competences are: Communication in a foreign language; Digital competence; Learning to learn; Sense of initiative and entrepreneurship, and Cultural awareness and creativity, all important for the digitization in education.

In recent times, the Swedish school authorities (Skolverket, 2016) have suggested a new national strategy to better exploit the potential of IT in schools. The strategy contains proposals for actions aimed at supporting all students and teachers to develop the digital skills they need to improve results and to prepare students for an increasingly digitalized society. In Sweden every student gets a laptop from their school. Nevertheless, digital media are underused in the classroom. Fredriksson (2011) showed in her pilot study of Swedish upper secondary students' use of computers in the classroom that computers were used to a very small degree in German teaching, although the majority was in favor of this. She also found a need for targeted educational efforts to change students'computer use and that both teachers and students should be involved in this process.

In this study we focus on the use of digital tools in the teaching and learning of English in Sweden and Germany. English is the first compulsory foreign language in both countries. Sweden and Germany are closely related, even so the situation concerning digitalization and the status of English differs (cf. Burchard et al. 2016). German students do not get a computer from their school and the input of English outside school is more limited than in Sweden, where English is considered to be a 'second' and not a foreign language. In a recently conducted interview with teachers in Germany (Weller 2016), 48% said, that they would like to use digital tools in the classroom. Although Sweden is far ahead in the use of computers in education, teachers' lack of training in the use of digital media has been reported in Sweden as well (cf. Fredriksson 2011, Fredholm, 2016). Rosenberg (2010) emphasizes the use of the learner as a creativity resource to provide teachers with an unending source of ideas and study material. Therefore, it is essential to 'create more situations in which the learners can contribute, initiate, control and create what happens in the classroom' (Deller, 1990).

In accordance with Rosenberg and the key competences, stated by the European Commission, this pilot study will present first results from interviews with teachers and questionnaires among students in the two countries. The next step will be to involve local entrepreneurs in the creation and management of digital tools for language learning. It is believed that a collaborative approach and Co-creation between teachers, students and entrepreneurs can prove to be a fitting solution to enhance the effectiveness and applicability of digital media for language learning.

RESEARCH OBJECTIVE

In Sweden and in Germany English is a compulsory part of the curricula, but there are differences in the intensity of input outside school. In both countries there is increasing importance given to the use of digital tools in schools, but Sweden is far ahead, as stated above. The purpose of our pilot study is to compare how the use of

digital tools in language teaching differs between the two countries and to compare the attitudes to digital tools between students and teachers. The hypothesis is that there will be differences in the overall attitude to digital tools between the students and teachers and also between the two countries. It is of interest to know the nature of the existing gaps between the two clusters.

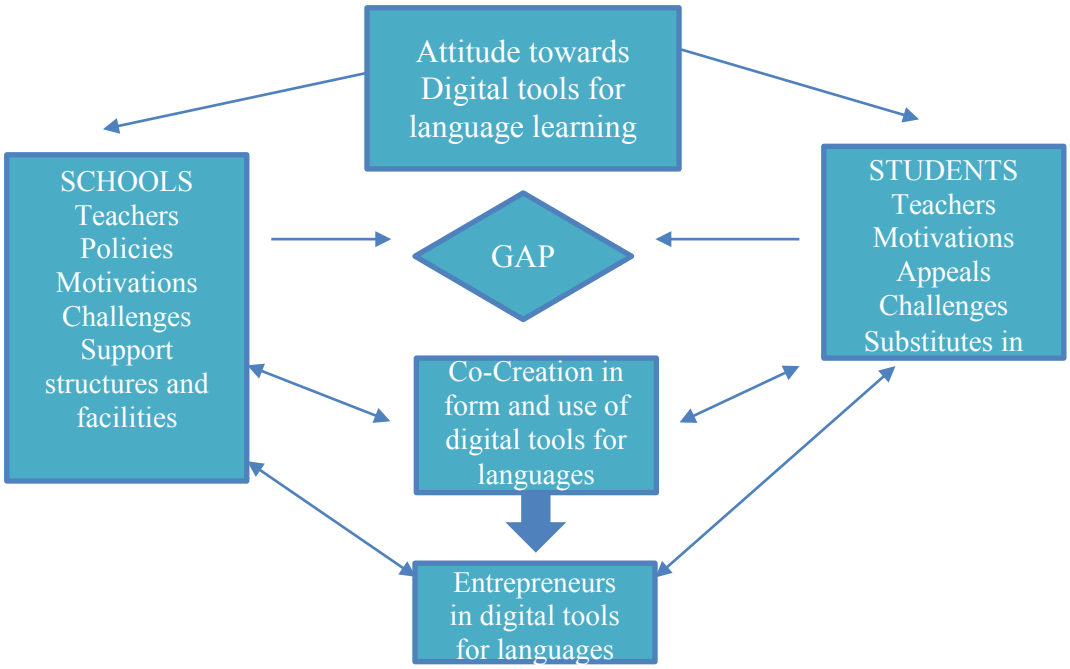
METHODOLOGY

The teachers' use of digital tools, their motivations, challenges and associated support systems for digital pedagogy is examined through interviews with 10 teachers of English in Sweden and 10 in Germany (ongoing). The parameters used to interview the teachers included 1. affect (liking), 2. perceived usefulness, 3. perceived control, and 4. behavioral intention (Buabeng-Andoh,2012) to use the digital tools for English language teaching. The students' use of and attitudes to digital media for learning English are examined through questionnaires, answered by 15 Swedish and 30 German students. The questionnaire was based on the following parameters: (i) The use of computers at home, (ii) The use of computers in school, (iii) The use of digital tools for learning English. All informants also evaluated an online program to have one controlled variable. Identification of existing gaps will be used as a premise to motivate a collaborative and co-creative process wherein a better environment for digital learning of languages can be built. We assume that students from different cultures will come up with different emotions, motivation and experiences that can also add immense value in the teaching and learning process. As stated above we will present first results of work in progress.

Extension of the pilot project: In a follow up of the results from the pilot, local entrepreneurs in the business of digital tools for language learning will be involved. The following holistic research framework will be used to connect the various elements of the study into one larger frame (See figure 1). It is expected that these stakeholders will benefit from the co-creative and collaborative activities between teachers and students.

Digital technology in languages is one of the most developing areas in recent times. A great deal of research is still needed in this area, though. Hence it is believed that this study will help to fill the gaps and work towards knowledge creation that can be used both academically and commercially.

Figure 1
Research framework



REFERENCES

Al Jarf, R. (2004). The effects of web-based learning on struggling EFL college writers. *Foreign Language Annals*, 37(1), 49 –57.

Buabeng-Andoh, C. (2012). Factors influencing teachers' adoption and integration of information and communication technology into teaching: A review of the literature. International Journal of Education and Development using Information and Communication Technology, 8(1), 136.

Burchard et al. (2016). Digitalisierung an Deutschlands Schulen. Milliarden für das Ende der Kreidezeit. http://www.tagesspiegel.de/wissen/digitalisierung-an-deutschlands-schulen-milliarden-fuer-das-ende-der-kreidezeit/14673782.html (accessed on 14.10.2016)

Deller, S. (1990) *Lessons from the learner*. Harlow: Longman

Weller, J. (2016). Woran-die-Digitalisierung-der-Schulen-scheitert https://www.welt.de/politik/deutschland/article153777823/Woran-die-Digitalisierung-der-Schulen-scheitert.html (accessed on 13.10.2016)

European Commission (2010). Education and Training 2010, Work programme. Cluster Key Competences – Curriculum Reform. Synthesis Report on Peer Learning Activities in 2007. https://ec.europa.eu/education/policy/school/doc/peer07_en.pdf (accessed on 12.09.2016)

Fredholm, K. (2016). *Eleverna, datorn och språket Studier av skoldatoriseringens effekter på elevers attityder, skrivstrategier och textproduktion i spanskundervisningen på gymnasiet*. Licentiatuppsats. Studier i språkdidaktik 12, Stockholms universitet.

Fredriksson, C. (2011). Datorn i skolans tyskundervisning ur ett elevperspektiv. Om svenska gymnasieelevers syn på datormedierad kommunikation och lärandet. *Forskning om undervisning och lärande*, 11.

Kuang-wu Lee, J. (2000). English Teachers' Barriers to the Use of Computer-assisted Language Learning. *The Internet TESL Journal*, VI (12). http://iteslj.org/ (accessed on 03.02.2015)

Langer de Ramirez, L. (2010). *Empower English Language Learners with Tools from the Web*. Corwin.

Motteram, G. (2013). *Innovations in learning technologies for English language teaching*. British Council 2013 Brand and Design/C607.

Rosenberg, M. (2010). 'The Learner as a creativity resource', Creativity in the English language classroom, Alan Maley and Nik Peachey (Eds.). Available at https://englishagenda.britishcouncil.org/sites/ec/files/F004_ELT_Creativity_FINAL_v2%20WEB.pdf. (accessed on 12.09.2016)

Skolverket (2016). Ny nationell strategi för att bättre ta tillvara möjligheterna med it i skolan. http://www.skolverket.se/om-skolverket/press/pressmeddelanden/2016/pressmeddelanden-2016-1.244816/ny-nationell-strategi-for-att-battre-ta-tillvara-mojligheterna-med-it-i-skolan-1.248733 (accessed on 12.09.2016)

Sundqvist, P. & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *Elsevier*, 51, 65–76. http://www.sciencedirect.com/science/article/pii/S0346251X15000433 (accessed on 12.09.2016)

Svensson, P. (2008). *Språkutbildning i en digital värld*. Norstedts Akademiska förlag.

Swe-Clarin: Language Resources and Technology for Digital Humanities

Lars Borin[1] **Nina Tahmasebi**[1] **Elena Volodina**[1] **Stefan Ekman**[2] **Caspar Jordan**[2]
[1]Språkbanken, Department of Swedish
University of Gothenburg, Sweden
firstname.lastname@svenska.gu.se
sam@sweclarin.se, info@sweclarin.se
[2]Swedish National Data Service
University of Gothenburg, Sweden
firstname.lastname@snd.gu.se

CLARIN AND SWE-CLARIN

CLARIN (Common Language Resources and Technology Infrastructure) is a European Research Infrastructure Consortium (ERIC), an ESFRI (European Strategy Forum on Research Infrastructures) initiative which aims at (a) making extensive language-based materials available as primary research data to the Humanities and Social Sciences (HSS) research communities; and (b) offering state-of-the-art language technology (LT) as an e-research tool for this purpose, positioning CLARIN centrally in what is often referred to as the *digital humanities* (DH).

Swe-Clarin as the Swedish CLARIN node was established in 2015 with funding from the Swedish Research Council by a consortium consisting of 9 members – so-called Swe-Clarin centers – representing the Swedish academic community as well as public memory institutions. The academic members are well balanced over the LT field, covering existing and possible research areas and user groups, and the memory institutions provide access to many of the language-based materials of interest to the users. Swe-Clarin is coordinated by Språkbanken, University of Gothenburg.

From the start, Swe-Clarin has aimed to establish good relations to the HSS fields and open the door for all researchers who wish to work with DH research using text and speech as primary research data. To avoid being a project by language technologists for linguists, we strive to include the HSS researchers in the process as early as possible. Our preferred way of doing this has been to establish small pilot projects with at least one member from the HSS field and at least one Swe-Clarin consortium member, together formulating a research question the addressing of which requires working with large language-based materials. Ideally, the collaboration should additionally always include a data owner, a person or persons representing the institution where the text or speech data is kept – typically a memory institution.

The pilot projects aim to spread the word of Swe-Clarin, show the potential of using language technology in DH research, create a user base for the tools and resources developed and maintained by Swe-Clarin, and last but not least, having this development being informed by input from users in the earliest possible stages of the project. Some pilot projects are already underway (see below).

In addition to the pilot projects, we have arranged workshops and user days and published newsletters and a blog. The workshops held so far have been on topics such as: general introduction to Swe-Clarin, our tools and resources; historical resources and tools; making cultural heritage text data available for research; and HSS research on digitized speech data, such as those of the Swedish Media Archive. We have started a series of workshops called *Swe-Clarin on tour* where Språkbanken's widely used Korp corpus infrastructure (Borin et al. 2012) is used to explore previously unexplored materials in a hands-on manner, giving researchers of LT and HSS the opportunity to meet and discuss research questions and the potentials of using LT for DH. The experience from working with HSS researchers will help reveal the limitations of existing tools and hopefully also engender general methodological discussion, thus setting the stage for future development of tools more appropriate for DH research. The first such workshop was held at Stockholm University in the spring of 2016. It featured the ethnographic questionnaires collected by the Nordic Museum since the late 1920s and now digitized by them, and it was attended mainly by ethnologists. The next workshop in the series will be held in Umeå in conjunction with the Swedish Language Technology Conference in November 2016. There the material in focus will be the *Swedish Government Official Reports* (Statens offentliga utredningar, SOU), in the version digitized by the National Library of Sweden, comprising more than 400 million words covering the years 1922–1998.

SOME SWE-CLARIN PILOT PROJECTS

Attitudes Toward Rhetoric Over Time

In this pilot project, a historian of rhetoric at Uppsala University together with the Swe-Clarin center Språkbanken explored how Språkbanken's Korp infrastructure could be applied to the research question of how the attitudes to rhetoric expressed in Swedish public discourse have changed over the last 200 years. The focus in the pilot project was on a large (almost 1 billion words) digitized historical newspaper material provided by

the National Library, but some preliminary studies of modern social media were also included for comparison. (Viklund and Borin 2016)

A Text Analysis Toolbox for Learner Language

The Swe-Clarin center at Uppsala University has developed SWEGRAM, a web service that provides automatic linguistic annotation at word and sentence level, which can subsequently be used to derive statistics on different linguistic characteristics of the texts, for example, the number of words and sentences in a text, the average length of a word, the distribution of word classes or different measures of readability. In a collaboration with researchers at the Department of Scandinavian Languages at Uppsala University, SWEGRAM has been made the basis for a web-based tool for annotation and quantitative analysis of student essays for the national exam in Swedish and Swedish as a second language for different grades (3rd, 6th, 9th grade). (Megyesi et al. 2016)

The Annotated Strindberg Corpus

The Swe-Clarin center at Stockholm University in collaboration with the Swedish Literature Bank (Litteraturbanken) and the editorial team of the National Edition of August Strindberg's Collected Works aim to construct a linguistically annotated corpus of Strindberg's collected works. The National Edition consists of 72 volumes with about 6 million words published between 1981 and 2012. The annotated version of the corpus will enable new kinds of research to be conducted on this material, as well as pave the way for even deeper annotation in the future. (Nilsson Björkenstam et al. 2014)

LAST BUT NOT LEAST

We strongly encourage you to contact us if you are interested in any of our resources, in conducting a pilot study with us or if you have any ideas or questions regarding digital humanities research with respect to language technology and resources: <info@sweclarin.se>. See also <https://sweclarin.se>.

REFERENCES

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012* (pp. 474–478). Istanbul: ELRA.

Megyesi, B., Näsman, J., & Palmér, A. (2016). The uppsala corpus of student writings: Corpus creation, annotation, and analysis. In *Proceedings of LREC 2016* (pp. 3192–3199). Portorož: ELRA.

Nilsson Björkenstam, K., Gustafson Capková, S., & Wirén, M. (2014). The Stockholm University Strindberg Corpus: Content and possibilities. In R. Lysell (Ed.), *Strindberg on international stages/Strindberg in translation*. Cambridge: Cambridge Scholars Publishing.

Viklund, J., & Borin, L. (2016). How can big data help us study rhetorical history? In *Selected Papers from the CLARIN Annual Conference 2015* (pp. 79–93). Linköping: LiU EP.

Using Enkapsis Theory for Unravelling Societal Complexities: the case of Uber

Christine Boshuijzen- van Burken
Department of Informatics
Linnaeus University, Sweden
christinevanburken@gmail.com

Darek M. Haftor
Department of Informatics
Linnaeus University, Sweden
darek.haftor@lnu.se

INTRODUCTION

Societal challenges may be evoked by novel digital technologies that connect different stakeholders in society. An example of a societal challenge is the arrival of smartphone applications that create peer-to-peer businesses, such as the transportation networking application Uber or accommodation network application Airbnb. In particular Uber has caused unrest in the traditional taxi sector, it has evoked legal debates and numerous newspaper articles have been devoted to this so-called 'disruptive technology' (see for example Weisse and Guynn 2014; Suster 2014; Adhikari 2015). Uber directly or indirectly connects drivers to passengers, credit card companies to developers of geographical maps, mobile network providers to car manufacturers, legal authorities to international investors, etcetera.

A way of understanding the complex relationships between the entities and stakeholders in digitally enabled constellations such as Uber, is in terms of enkaptic interlacements. The theory of enkapsis is a philosophical tool, based on a specific view of reality, which may guide a novel understanding of the relationship between artefacts and entities and between social structures that exist in reality. We contribute to interdisciplinary research by using insights from philosophy to understand societal complexities caused by digital technologies. It falls in the Digital Humanities project, since it connects both the academic and public and private sectors in interdisciplinary research and innovation to tackle societal challenges.

ENKAPSIS THEORY

The theory of enkapsis is unique to Dooyeweerd, a Dutch philosopher, who introduced his theory in his book "A New Critique of Theoretical Thought" (1953). It is part of a larger philosophical framework of individuality structures and modal structures which we will not address in detail. In short, the theory of individuality structures (and modal structures) is a project to understand the nature of specific things and events in reality as well as grasp the identity of entities in a non-reductionist manner. It takes as a basis that reality presents itself according to a number of irreducible yet closely interrelated aspects (Dooyeweerd distinguished 15 aspects in total, such as numerical, spatial, psychological, ethical, juridical, etc). The theory of enkapsis is only a small part of this project.

Furthermore, Dooyeweerd distinguishes between enkaptic interlacements and part-whole relationships in order to explain how individuality structures cohere amidst their differences (Chaplin, 2011). We have added a third relationship, namely juxtapositional relationships.

An *enkaptic interlacement* pre-supposes that the structures of things and events, or those of societal relationships functioning in it, have an independent internal leading function and an internal structural principle of their own. (Dooyeweerd, 1953, Vol. III, 637).

A *part-whole* relationship can be defined as follows: "In all those things whose structure is not that of a homogeneous aggregate, a *part* is essentially qualified by the structure of the whole. In this case the structure of a whole can never be construed by means of its parts, because the parts, as such, are entirely dependent on the whole. The question what is a part of a non-homogeneous whole cannot be decided by a functional mathematical-physical analysis, but only by an inquiry into the internal individuality-structure of this whole. This fact has always been lost sight of on the functionalist standpoint." (Dooyeweerd 1953, Vol III p. 638-639).

We call a constellation a juxtapositional relationship when two wholes function independent of each other and cooperate on a temporary, non-necessary basis. Two wholes can be taken apart without disrupting or intervening their respective qualifying functions.

ANALYSIS

If we apply this rather abstract theory to the complex case of Uber, our initial findings are that firstly Uber has a part-whole relationship with the information technology infrastructure: without digital technologies, Uber loses its meaning and will not function according to its leading function, namely to connect drivers to passengers. Secondly, Uber has an enkaptic interlacement with the credit card company that takes care of the automatic payments, since Uber can exist independent of the credit card company system (it can handle payments in different ways, even in cash, as it does in some countries where credit cards are a rarity) and the credit card payment system does not depend on Uber for its existence. Thirdly, Uber has a juxtapositional relationship with regular taxi companies and legal authorities (although there are cases where legal authorities have entered in an enkaptic interlacement with Uber, such as in the Philippines). In this paper we will further explore how the theory of enkapsis can explain a multiplicity of complex relationships and furthermore, how these relationships relate to different responses to Uber in different countries, cultural settings and legal systems.

REFERENCES

Adhikari, S. (2015) (16 February 2015). Taxi industry slams uber's 20,000 jobs stunt. *Technology Spectator.*

Chaplin, J. (2011) *Herman Dooyeweerd. Christian Philosopher of State and Civil Society* .University of Notre Dame, Notre Dame.

Dooyeweerd, H. (1953) *A New Critique of Theoretical Thought*, Vol III. Presbyterian & Reformed Publishing Co, Phillipsburg, NJ.

Suster, M. (22 November 2014). In defense of Uber: An objective opinion. Retrieved from http://www.bothsidesofthetable.com/2014/11/22/in-defense-of-uber-an-unbiased-opinion/

Weisse, E., & Guynn, J. (2014) (19 November 2014). Uber tracking raises privacy concerns. *USA Today.*

This document is prepared for a poster presentation at the Digital Humanities Conference, 7&8 November 2016, LNU, Vaxjo

Walking in My Shoes: a Case Study from a Born-Digital Archive

Emmanuela Carbé PAD –
Pavia Archivi Digitali
University of Pavia, Italy
emmanuela.carbe@unipv.it

Ricky Erway concisely explained scenarios and issues regarding the preservation of digital materials (Erway 2010), and with Barrera-Gomez proposed certain fundamental steps for the preservation of born digital contents extracted from physical media (Erway – Barrera-Gomez 2013). They suggest to "walk before run", a valuable advice for those who work in projects related to digital humanities, which relay on architectures based on scalability and interoperability.

The vulnerability of bits and the obsolescence of media also raise new challenges with respect to the preservation of cultural heritage produced in the last decades. The availability of great amounts of digital material of various kinds poses questions on the role of digital curators and memory institutions in physical preservation of digital material and accessibility to documents.

In 2009 a research team at the University of Pavia decided to develop the "PAD – Pavia Archivi Digitali" project, aiming at long-time preservation of digital papers from Italian writers and journalists, and their accessibility to the research community. Pavia seemed to be a good location to build a Born-Digital Archive, also because of the long-standing archivistic tradition of its "Centre for Research in the Manuscript Tradition of Modern and Contemporary Authors". PAD in its beginning consisted in a long walk and yet, despite the experience with six authors and the improvement of all the procedures, every case is characterized by new problems which are always different and unique.

A few international institutions have been working on projects for the preservation of born-digital papers of writers, including the Harry Ramsom Center, which preserves some collections such as that of Michael Joyce (Stollar Peters 2006). An other significant example is the collection of the Salman Rushdie digital archive, preserved by the Emory University's Manuscript, Archives and Rare Book Library (Carroll – Farr – Hornsby – Ranker 2011).

The aim of PAD is to try to be as flexible as possible in terms of the amount of material types, authors and archive dimensions: its main feature is an integrated quality control system that manages each single phase of a bestowal almost in real time, allowing the ingestion, classification and validation of virtually every file type under a strict and accurate supervision. The locally developed Quality Control Software, dubbed QUANDO (Quality control for Archiving and Networking Digital Objects), is used to check all the important aspects of an archive's life, integrating information manually entered with data that has been gathered automatically using a PAD-developed application suite, which performs several actions on every single archive (checksumming, virus control, metadata extraction, synchronization, etc.).

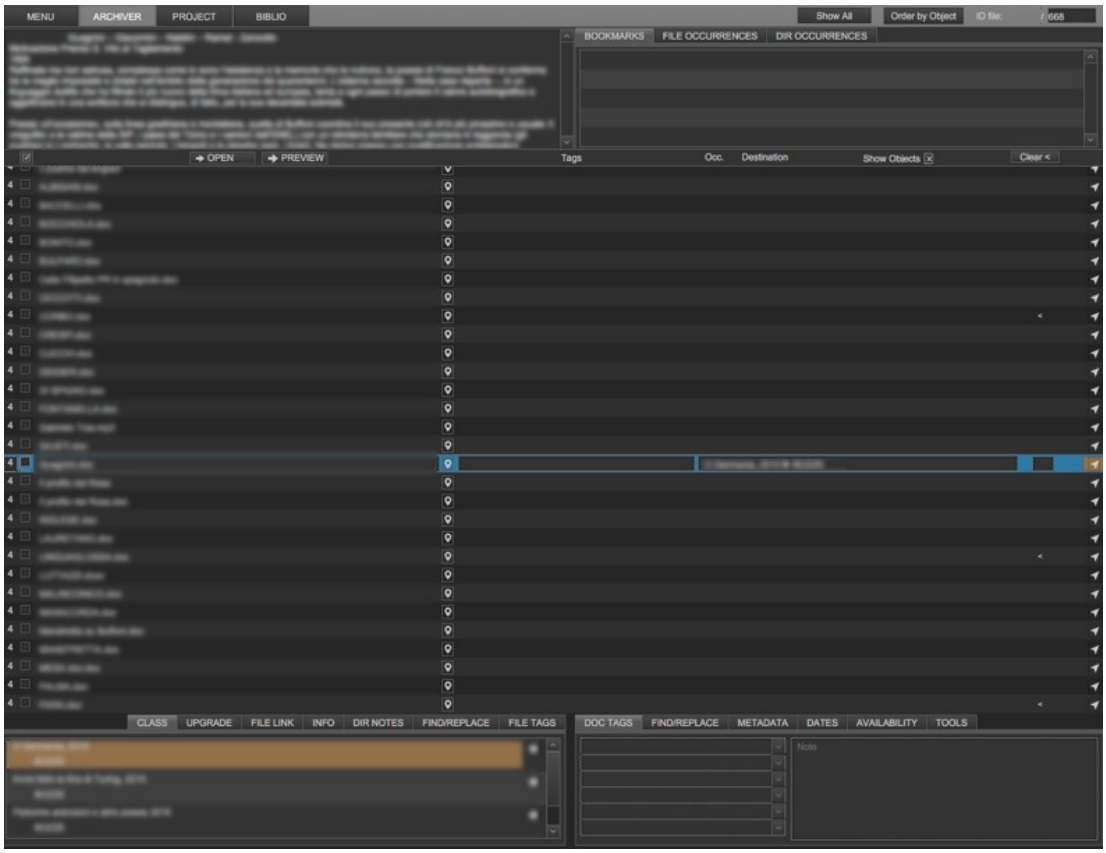
The most difficult acquisition for PAD has been that of Francesco Pecoraro's archive. It has been the best test case for our procedures and workflow, and helped us to reexamine many challenges, such as the ingestion of files from different media and of the materials published on the blog and on social networks. The author, who is also an architect, was very popular for his blog "Tash-tego", active from 2005 to 2011; thereafter he has been quite active on Facebook until April 2015. His most popular book is the novel *La vita in tempo di pace* (Pecoraro 2013), which became in 2014 a relevant literary case. In 2015 the author transferred more than 43.000 files to PAD, plus the materials coming from old floppy disks and a DVD. He gave the indication to keep some private correspondence undisclosed for 30 years. Before the bestowal, as stated in the workflow procedure for every ingestion, the author is asked to fill in an informational survey consisting of 15 main areas of questions. Pecoraro stated that he first used a PC for writing in the '80s. He used to work with a Windows 7 based desktop workstation at the time of the bestowal and uses Dropbox for the majority of his writings. He also makes use of two external hard disks to keep materials. The author backed up his work several times, especially (but not limited to) upon workstation substitution. With his archive PAD faced in fact a chinese-box styled form of organization and a lot of problems in the first validation step, such as checking whether sensitive information was present in the archive and determining all files that should have been kept undisclosed.

The architecture for preservation and handling of data has been designed following the OAIS Reference Model indications. The archival system is based on five areas: staging, deposit, permanent, work, and info. Two copies of each archive, including documentation, metadata, and file format conversion, are kept in Pavia and another one is stored more than 90 kilometers away from PAD's main site.

Upon arrival, the PAD preservation plan prescribes that materials are stored into the temporary area, where they are preserved while waiting for the availability of an operator. In the deposit area the archive integrity is

checked, as well as possible viruses. In case any malware is found, the author is notified immediately and, in case of need, assistance is offered. SHA-1 hashes are generated. An application generates a list of unique files that have been transferred, which is sent to the author for validation. In case of afterthoughts the author can decide to remove a file or a set of files. Attached to the list, a summary is sent indicating the total amount of transferred files, the number of unique files and the size of the entire archive. In the case of Pecoraro's archive many problems have been faced: how was it possible to check in more than 43.000 files, in terms of pointing out undisclosed items or very private files that have to be removed or kept unavailable, and then visually check complex cases such as this one? This very difficult experience helped us focus on an application that could manage the file from the deposit to the permanent area, which would add metadata that could be helpful for the fourth step of the process, in which the files are transferred to be catalogued. All Pecoraro's files are now being checked through this software, where it is possible to choose actions for all files and folder: such actions include checking with the author, determine sensitive and undiscloseable files, find files with technical problems and files to ask the author about.

In this way the author is be able to check and decide about all the archive files. At the same time it is possibile to add metadata, tag each files in two different ways (the single file or all the file duplicated), to add notes to files and directories, to review dates, to check the occurrences of files and directories, to preview each file and to directly open any file.



PAD software, developed using FileMaker

REFERENCES

Stollar Peters C., When Not All Papers are Paper: A Case Study in Digital Archivy (2006), *Journal of the Society of Georgia Archivists*, **24**, 22-34.

Carroll L., Farr E., Hornsby P., Ranker B. (2011), A Comprehensive Approach to Born-Digital Archives, *Archiviara*, **72**, 61-92.

Erway, R. (2012) You've got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Recevied on Physical Media. Dublin, Ohio: OCLC Research.

Barrera-Gomez, J. and Erway, R. (2013) Walk this Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house. Dublin, Ohio: OCLC Research.

Pecoraro F. (2013) La vita in tempo di pace, Ponte Alle Grazie, Roma.

Social media, cultural heritage and migrant communities in a globalized world

Laia Colomer*, Nuno Otero[§♯], and Julia Schmidt[§]
[§]Department of Media Technology, Linnaeus University, Sweden
*Department of Social Sciences, Linnaeus University, Sweden
[♯]CIS-IUL, Portugal

laia.colomer@lnu.se
nuno.otero@lnu.se

THIRD CULTURE KIDS (TCKs) AND THEIR USE OF SOCIAL MEDIA

The open nature of social media enables the potential of new decentralized and less hierarchical social structures and promotes new dynamics of social and cultural practices. These new technologies can foster the perception among people that they belong to a larger community by virtue of the identity they share online rather than to where they physically lived or culturally belong. In this sense, social media is increasingly playing a key role in enabling *collective identity*, a *sense of community*, and supporting *collective cultural creation* among citizens across the globe.

The potential for online community creativity is relevant for migrant communities to whom social media is actually becoming a distinctive arena of social life. Accordingly, we argue that it serves as sources of community building among people with social and cultural affinities but with restricted possibilities of offline meetings. Considering this framing, we are exploring these circumstances among one of the most globalized migrant community: the Third Cultural Kids (TCKs), a migrant community spread globally who are not culturally defined by their passport, family background, origin and nationality, or their cultural affiliations, but by their multicultural rootless and restlessness regarding personal traditional personal identity issues (Pollock and Reken, 2009; Bell-Villada and Sichel, 2011). As a global nomad community, they take fully advantage of Web 2.0 media, not only to "feel connected to the world" and to "be connected between them", but basically to acknowledge and build their sense of community.

The material posted in social media is thus a source of information on TCK's personal and social self-perceptions, interesting enough to be analyzed by any researcher who wishes to relate their personal and collective experiences into cultural identity issues. Most of these materials are personal narrations/accounts and but by applying qualitative and quantitative analysis techniques we are exploring the possibilities of identifying familiarities, similarities and patterns from which we can create a vivid picture of the on-going construction of this community identity. This initial step seems fundamental to us if one thinks of creating digital tools with features that are in line with the characteristics of this community. More specifically, our current research is framed in accordance with the following research questions:

- How can we help reinforce the sense of identity of this group by fostering ways to build a stronger online community using social media? What particular features should the digital networking tools possess in order to support the creation and maintaining of this community?
- What public spaces are relevant for this particular community and in what ways can these be enhanced with digital technologies to promote communication? Are location-based services a possible technological solution? How can location based services be designed in order to connect TCKs and their meaningful places?

We expect that our understanding of this community identity can extend the notions of meaningful public spaces and consequently cultural heritage.

INITIAL EXPLORATORY INVESTIGATIONS

In order to start exploring the problem space suggested by the research questions referred to above we have conducted two pilot studies. In the first study, a discourse analysis of an unique Twitter chat on "global citizens" was run to explore how cross-cultural and global identities are constructed among a community of TCKs (Colomer & Schmidt 2015). More specifically, we analysed the twitter feed of #TCKchat for two weeks in 2015. TCKchat is organized by BateConsulting and is a biweekly event with questions aiming to start discussions and experience exchange between TCKs. The topic within our two-week span was "Global Citizenship Explored".

Our data collection for the initial study consisted of 832 tweets, generated by 51 contributors. Using discourse analysis we explored the contributors opinion on the term “Global Citizen” and how the community works together to create terms to describe themselves. “Global citizen” does not have a positive meaning to most contributors. From the additional hashtags in the posts we can see that “(global) nomad” and “expat” are often used. However, there was no deeper discussion on the terms which might be due to the limitation of 140 characters per tweet on Twitter. It is important to note though, that terms defined by researchers or experts do not always resonate as well as expected with the groups they are meant to describe.

In the second study, a wider analysis (using text categorization software packages NVivo 11[1] and Semantria[2]) of 24 open Facebook groups dedicated to TCKs was performed to identify what is important for members to share and discuss within the community. The findings show sharing experience, community and identity as the most important topics being addressed within the groups. Words like “moved” and “community” were among the most used in the texts written by TCKs. The main themes (so called nodes) identified by NVivo were Group, TCK, and Community. Semantria identified Identity, Globalization and Passport as the most mentioned categories out of a custom category list. The custom query results show TCK, Terms, and Culture as the main discussed topics, followed by Education, Globalization, and Identity. 225 places and countries from all over the world appear in the texts with France being far in the lead. (Schmidt 2016)

The results suggest that TCKs have a strong interest in building their community identity, sharing their experiences, and discussing the terms used to describe themselves, and that they do use social media in the process of creating community and identity.

CONCLUSIONS AND FUTURE WORK

Our preliminary research efforts gave us a glimpse of the main topics and points of discussion that TCKs seem to be engaged with when taking advantage of social media channels.

We aim to understand the relevance of social media in the process of both meaning construction and community building for this community of global migrants. We will do it by further analyzing quantitatively and qualitatively the uses of social media among TCKs and other onward migrants. This analysis will be the starting point for our future explorations concerning the design of digital tools that will support this community. More specifically we will:

- Explore the design space of digital networking tools to help create and maintain a thriving TCKs online community.
- Explore to what extent it is possible to infuse public places with appropriate digital technologies that foster further engagement with this community and between members of the community.

TCKs might help us understand how to create communalities across places and cultures that will foster cultural mutual understanding and multi-cultural practices. In other words, TCKs can teach us something about the value of multiculturalism and how to promote digital technology based activities that support such value.

REFERENCES

Bell-Villada, G.H. and Sichel, N. (Eds.) (2011). Writing out of Limbo. International Childhoods, Global Nomads, and Third Culture Kids. Cambridge Scholars Publishing. Newcastle upon Tyne.

Colomer, L. and Schmidt, J. (2015). *How Third Culture Kids Build up their Community Identity Through Social Media*. Unpublished report, part of the requirements concerning the course "Network Society and Internet Cultures", Social Media and Web Technologies Masters' program. Department of Media Technology. Linnaeus University (for a copy of the report please contact the authors).

Pollock, D.C. and Reken, R.E. (2009). Third Culture Kids. Growing Up Among Worlds. Nicholas Brealey publishing. Boston/London.

Schmidt, J. 2016 *Categorization of Facebook Group Messages.* Unpublished report, part of the requirements concerning the course "Adaptive and Semantic Web", Social Media and Web Technologies Masters' program. Department of Media Technology. Linnaeus University (for a copy of the report please contact the authors).

[1] NVivo 11 by QSR International (www.qsrinternational.com)

[2] Semantria Excel Plugin by Lexalytics (www.lexalytics.com/semantria/excel)

Sketch Layer Separation in Multi-Spectral Historical Document Images

AmirAbbas Davari[1], Armin Häberle[2], Christian Riess[1]
1. Pattern Recognition Lab, Computer science Department,
Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
[amir.davari, christian.riess]@fau.de
2. Bibliotheca Hertziana - Max-Planck-Institut für Kunstgeschichte, Rome, Italy
Haeberle@biblhertz.it

MOTIVATION

Recently, the progress in computer science made links between this field and art history possible. The examination of cultural artifacts particularly gained new insights in materials and object structure from new possibilities in high-resolution image acquisition and data processing. However, although there is a large number of specialized examination methods (for example IR/UV reflectography, fluorescence, micro-XRF, Raman spectroscopy), the research on hand drawings still suffers from major diagnostic gaps: due to the particular properties of red chalk, it is currently not possible to visualize preparatory drawings from red chalk if they have been overlaid by inks.

Up until the end of the 19th century, hand drawings were typically created with different materials in several working steps. In the final drawing, multiple layers of different materials overlap, such that the lower, older layers can oftentimes not be identified visually. However, these layers represent the various steps of the genesis of the respective work. Thus, separating these layers by means of imaging and image processing promises a direct look into the artistic creation process of the work. Therefore, such a technical approach may supply answers to key questions of art history on the object and the artist, and may help for objectivation of attribution and authenticity.

In this work, we investigate an approach to close this diagnostic gap. In contrast to image acquisitions with a limited spectral window (like infrared or ultraviolet light), we propose to image an object using a multispectral camera, and to process the acquired data with methods from the field of pattern recognition. A multispectral camera operates primarily in the range of visible light, but subdivides the light into much more channels than red, green, and blue. This allows to differentiate materials of different physical compositions from their reflection patterns after processing the data with a computational algorithm. We evaluate the proposed method on drawings that were created to exactly mimic the original work process, using the same materials and papers. The controlled creation process provides knowledge about the drawing layers, which allows to assess the accuracy of the method. This approach has the potential to avoid radiation damage to the work of art under examination, while still being able to provide meaningful information. Thus, it appears as a method that can have broad applications in the domain of art history and conservation of art, that extend beyond the range of currently used approaches (Cosentino, 2014) (Pronti, 2015).

WORKFLOW

The proposed workflow for analyzing drawings is shown in Figure 1, and explained in several steps below.

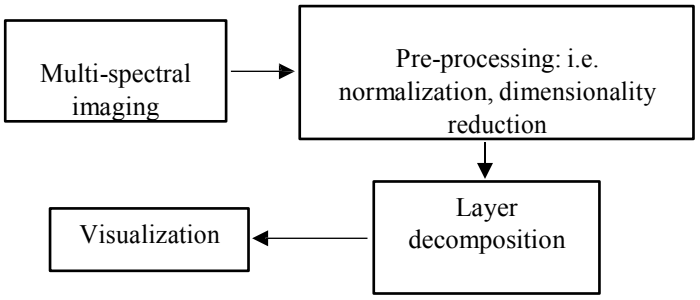
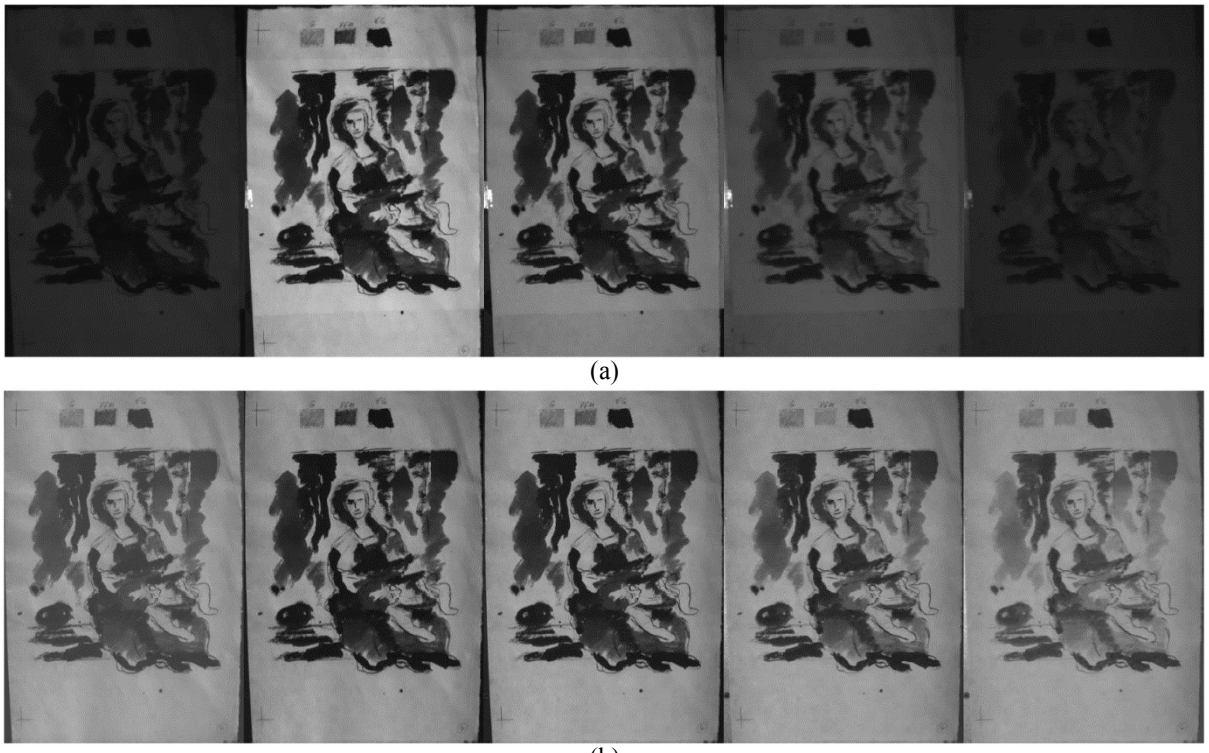


Figure 1: Proposed workflow

Color Channel Normalization

The sensitivity of the multispectral sensor varies across the captured spectrum. As a result, different channels exhibit varying brightness, which makes analysis of the images more difficult. To avoid these brightness differences, we capture an image of a homogeneous white surface before scanning the drawing. In the scan of the

drawing, each channel is corrected with a multiplicative factor, such that the sensor response on the white surface is homogeneous; Figure 2.



(a)

(b)

Figure 2: Five sample spectral channels of a 258-band multispectral image. Images from left to right correspond to channels 31, 76, 100, 173 and 205 respectively. a) Before normalization b) After normalization.

Dimensionality Reduction on the Spectral Bands

The 1040 spectral bands make it challenging to analyze the multispectral image directly. Therefore, a common approach in other communities working with multispectral images is to greatly reduce the dimensionality before further processing (Harsanyi, 1994). During dimensionality reduction, most of the spectral bands are either discarded (because they are found to be non-informative) or summarized using a linear or non-linear mapping function.

Principle component analysis (PCA), introduced in (Hotelling, 1933), is probably the most popular tool for this task. We first average every four neighbored channels, which reduces the dimensionality to 258, and then use PCA to summarize this information in only 5 channels (mathematically, this corresponds to preserving 99.5% of the image spectrum variance on our data). The representation of the image information in such a much lower number of channels makes the subsequent processing step more robust.

Clustering

Drawing regions of identical chalk or ink show a similar reflectance behavior. To group image points of similar reflectance, we experimented with clustering algorithms. Specifically, we compared the performance of the K-means algorithm (Lloyd, 1982) and Gaussian mixture models (GMM) (Friedman, 1997). In both cases, we expect that each chalk or ink is sorted into an individual cluster.

EXPERIMENTAL SETUP, RESULTS AND EVALUATION

Data

We created a set of sketches with multiple layers of chalk and ink. We used papers and inks/chalks of the same chemical composition as it was commonly used in ancient drawings. After drawing each layer, the picture has been scanned with a standard flatbed scanner. This allows to compute the true layer composition for evaluating the layer separation algorithm by subtracting two subsequent scanned images. A sample sketch from this data is shown in Figure 3.

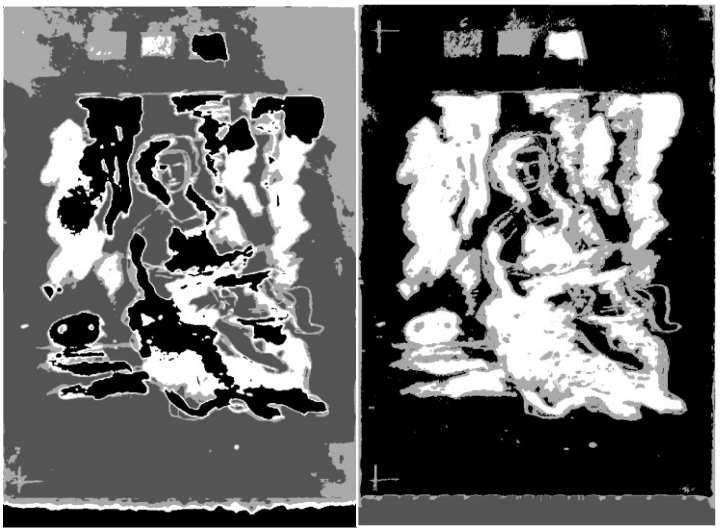
Figure 3: Sample layers from the evaluation data

For imaging, we use a multi-spectral camera equipped with a CMOS sensor, capable of capturing the spectrum in a wavelength range of 400nm to 1000nm, subdivided into 1040 so-called spectral bands.

Evaluation

For the evaluation of the data, we first compute the layer separation on the multispectral data. Then, we compare the result of this computation to the individual (known) layers from the data creation process. A result image is shown in Figure 4 (computed on the input data in Figure 3). On the left, the separation result of the K-means algorithm is shown, on the right the result for GMM clustering. Identical grayscale values indicate identical cluster membership. K-means grouped the background into two clusters (most likely due to inhomogeneous illumination), while GMM isolated the background into a single cluster. It also turned out that GMM performed better in separating diluted inks, which can be best seen in the top right of the sitting person in the picture. Thus, we tentatively conclude that GMM might be a better suited method for isolating sketch layers.

In future work, we aim to expand this study to a larger dataset, and also to quantitatively evaluate the performance of sketch layer separation. For quantitative evaluation, it is important to map the multispectral picture pixel by pixel to the known layers, which can be achieved by using an image registration algorithm.



K-means result GMM result

Figure 4: Clustering results

REFERENCES

Cosentino, A. (2014). Identification of pigments by multispectral imaging; a flowchart method. Heritage Science, 2(1), 1.

Friedman, N., & Russell, S. (1997, August). Image segmentation in video sequences: A probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence* (pp. 175-181). Morgan Kaufmann Publishers Inc.

Harsanyi, J. C., & Chang, C. I. (1994). Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. IEEE Transactions on geoscience and remote sensing, 32(4), 779-785.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, *28*(2), 129-137.

Pronti, L., Ferrara, P., Uccheddu, F., Pelagotti, A., & Piva, A. (2015, November). Identification of pictorial materials by means of optimized multispectral reflectance image processing. In Information Forensics and Security (WIFS), 2015 IEEE International Workshop on (pp. 1-6). IEEE.

Working with Digital Humanities: Sectoral Concerns and Cross-sectoral Collaborations

Isto Huvila
Uppsala University

August 24, 2016

As an explicitly transdisciplinary idea, digital humanities provides opportunities to bring together people and interests across the sectors, from a variety of scholarly and practical disciplines, and from the society at large. It may not be crucial that everyone has an identical understanding what digital humanities is if the stakeholders have shared or compatible concerns, and practical respect each others priorities. This applies to research questions and practical inter- ests as well as their projected significance. The compatibility of concerns and mutual understanding of each others priorities is not, however, something that would be given in a collaboration or dependency relation that crosses different disciplines and sectors. To give a few examples, problems may occur when de- veloping and borrowing digital tools from across contexts to address research questions from the humanities disciplines, when the limits of digital approaches to address specific questions are negotiated within and between contexts, and increasingly, when digital humanities researchers are using data provided by and originally produced in other sectors and situations. Even if the digital human- ities literature in general and, to a verying degree, individual research projects have emphasised the importance of being critical and sensitive to the implica- tions of using borrowed and newly developed technologies and understanding the data, so far, there is relatively little empirical research on the implications of cross-sectoral collaborations.

The aim of the presentation is to systematise observations on various problems relating to cross-sectoral collaborations in the context of digital humani- ties. It draws from an empirical study of archaeological documentation practices in Sweden and an analysis of the perspectives the stakeholders of archaeolog- ical information. Archaeology is an example of a discipline within which the cross-sectoral collaboration has always been significant and has increased in the post-war years due to the heritage legislation that mandates archaeological investigations before land use [1]. At the present, these investigations are pro- ducing unprecedented amounts of digital documentation data with a significant scholarly potential. In a large number of countries the majority of archaeologi- cal fieldwork is currently financed by land developers as an obligatory exercise regulated by the law. Even if the priorities and wordings vary from one country

to another (e.g. [2]), the principal purpose of archaeological fieldwork is to produce an adequate documentation of an archaeological site for future research. In Sweden, the purpose is three-fold (in this particular order) as to document an archaeological site, take care of finds, to report and communicate (mediate) the results. Moreover, the documentation material and finds shall be preserved for the future, be scholarly interpreted and placed in a cultural historical con- text [4]. What is crucial, however, is that the fieldwork itself and where it is conducted is not initiated by a scholarly interest but a need or want to develop land. Further, the financing that is coming from the land developers has an inevitable influence on the priorities of conducting fieldwork, and an interest in a part of the results.

According to the analysis of the empirical material, a major question for contemporary archaeological practices is how well the current information process is capable providing meaningful information for the different stakeholders and even more importantly, what are its implications to the usability and usefulness of the information, and collaborations between stakeholders in the different sectors. Are the matters of concern [3] of the other parties understood by the individual stakeholders and what are the consequences of understanding and not understanding them? Even if the principal conclusion of the analysis so far is that there is much to be done to help the different parties to understand each other and each others priorities in cross-sectoral collaborations, it is equally ap- parent that many of the significant issues are relatively common organisational and collaborative challenges documented in the literature. They are not specific to archaeology or even to digital humanities.

References

[1] Jean-Paul Demoule. Rescue Archaeology: A European View. Annual Review of Anthropology, 41:611–626, 2012.

[2] Kristian Kristiansen. Contract archaeology in Europe: an experiment in diversity. World Archaeology, 41(4):641–648, December 2009.

[3] Bruno Latour. Why has critique run out of steam? from matters of fact to matters of concern. Critical Inquiry, 30(2):225–248, 2004.

[4] Riksantikvarieämbetet. Riksantikvarieämbetets föreskrifter och allmänna råd om uppdragsarkeologi. Kulturrådets författningssamling KRFS 2015:1, Stockholm, 2015.

Digital Humanities and Games Research Across the Disciplines

Johan Höglund
Department of Languages
Linnaeus University, Kalmar and Växjö, Sweden
Johan.hoglund@lnu.se

Abstract

The effect of violent computer games on individuals and on society has been the object of a great number of studies reaching across different disciplines, including traditional Humanities, International Relations Studies, and Psychology. Unfortunately, studies conducted within one discipline pay very limited attention to research conducted in other fields. Thus, important research data is rarely shared. The reasons for this lack of cross-disciplinary consideration can be attributed to many different factors. Humanities oriented research is often published in journals other than IR studies, or psychological studies. The various fields engaged in this type of research also employ different methodologies that highlight different aspect while obscuring others. Finally, the research is funded by different agencies, with different agendas.

This presentation first describes the current situation through studies belonging to the Humanities, International Relations Studies and Psychology. These studies share an interest in the computer game genre commonly known as the First Person Shooter (FPS), a violent game genre where the gamer controls an armed avatar and observes the game world through a first-person perspective. The presentation discusses how the general research context (funding body, audience, problem formulation), the theoretical framework, and the methodologies of the different studies inform the research. Here, it is noted that Humanities research is often state-sponsored and conducted within Humanities departments or by one of few DH research centres that exist globally. Since the late 1990s, Humanities research has either focussed on discussing how participatory digital games function differently from other forms of culture such as literature or film (see Juul 2005, Malliet 2007), or it has conducted an often Foucauldian or Baudrillardian interrogation of the games, discussing them as deeply ideological spaces (Wark 2007). The methodological tools employed by this research are virtually always qualitative and hermeneutic. International Relations research also comes out of state-sponsored or private universities, but is sometimes connected to organisations such as the Institute of World Politics. Following the cultural turn of IR during the last two decades (Van Veeren 2009), this research has become increasingly attentive to the way that military games engage with global politics and future military conflict. The focus of game studies conducted within the confines of IR studies is thus the way in which the FPS imagines future global conflict. This research is often qualitative and does discuss the narratives and discourses of the games, but it also employs interviews and quantitative methods to investigate how gamers's ideas about global relations are affected by the games (Zamaróczy 2016). Finally, psychological research into violent games comes from a large number of funding bodies, from state-run universities to private foundations, the health care sector, and the US Department of Defence (DoD) (Höglund 2008). The research produced by these various agencies focuses primarily on to what extent violent games produce violent behaviour or not (Anderson et al., 2002), but it also includes studies on how games can train soldiers before combat or help treat veterans suffering from post traumatic stress disorder (Rizzo et al 2006). The link between violent computer games and aggressive behaviour is notoriously difficult to study in laboratory experiments, and a few alternative ways of assessing the relationship have been suggested (Sauer and Nova 2015). Even so, this research is firmly quantitative and often disregards the qualitative aspect.

The question that the presentation will address in relation to these studies is how these different fields may benefit from cross-disciplinary exchange. The presentation suggests that by considering results gained in psychological studies, and by making some use of the quantitative and laboratory methods common in this discipline, the humanities or IR researcher would be in a considerably better position to discuss the effect that the FPS has on the individual. In other words, broadening the disciplinary perspective would make it possible to consider not only the ideological, political and aesthetic content of the material, but also how gamers actually respond to the material. Similarly, humanities and IR related research could help researchers working in the field of psychology to ask more relevant and precise questions that take into consideration the qualitative content of a particular game before examining its effects in a laboratory setting. In other words, by considering humanities and IR research, the simple question if games encourage aggression in gamers may be rephrased into the more

complex question if games encourage aggression against particular groups in society, or support state aggression against certain nationalities. This discussion may be of interest to scholars conducting research on digital games, but it may also be of general interest to Digital Humanities since the formation of games research takes place in the crossroads of several different disciplines.

REFERENCES

Anderson, C. A, B. J. Bushman. (2002) Violent Video Games and Hostile Expectations: A Test of the General Aggression Model. *Personality and Social Psychology Bulletin,* 28.12): 1679-1686.

Höglund, J. (2008). Electronic Empire: Orientalism Revisited in the Military Shooter. *Game Studies*. 8.1. http://gamestudies.org/0801/articles/hoeglund

Juul, J. (2005). *Half-Real: Video Games between Real Rules and Fictional Worlds*, Cambridge: The MIT Press,

Malliet, Steven. (2007). Adapting the Principles of Ludology to the Method of Video Game Content Analysis. *Game Studies* 7.1. http://gamestudies.org/0701/articles/malliet

Rizzo. A, J, et al. (2006). A Virtual Reality Exposure Therapy Application for Iraq War Military Personnel with Post Traumatic Stress Disorder: From Training to Toy to Treatment. *NATO Advanced Research Workshop on Novel Approaches to the Diagnosis and Treatment of Posttraumatic Stress Disorder.* IOS Press, Washington D.C., 235-250

Sauer, J. D, A Drummond, and N. Nova. (2015). Violent video games: The effects of narrative context and reward structure on in-game and postgame aggression. *Journal of Experimental Psychology Applied*. 21.3. 205-214.

Van Veeren, Es. (2009). The 'Cultural Turn' in International Relations: Making Sense of World Politics. *E-International Relations*. May 10. http://www.e-ir.info/2009/05/10/the-'cultural-turn'-in-international-relations-making-sense-of-world-politics/.

Wark, M. (2007). *Gamer Theory.* Cambridge, Harvard University Press.

Zamaróczy, N de. (2016). Are We What We Play? Global Politics in Historical Strategy Computer Games. *International Studies Perspectives.* 0.1, 1–20.

Graph Analysis of Word Networks

Lars G Bagøien Johnsen
Department of Research and Dissemination
National Library of Norway
lars.johnsen@nb.no

AIMS

In this talk we present how semantic connections between words can be visualized and represented as graphs created from word networks (e.g. Turney and Pantel 2010), using concepts and methods from graph theory (e.g Chakrabarti and Faloutsos 2012). Graphs have a visual as well a formal structure, and we explore how clustering seen in visualizations can be analyzed and represented. For that purpose we will look at clustering mechanisms for graphs, and discuss their interpretation in the context of word meanings.

RESEARCH QUESTIONS

One question that is addressed is how textual raw data can be transformed into structures that somehow represent knowledge of language, and partially also reflect the external significance of language, i.e. how structures are mapped to the world around us. In particular, the search for how graphs may be interpreted and put to use in the analysis of language and literary works, ranging from the disambiguation of particular words to semantic fields represented as a collection of words.

METHODS

Graphs are constructed from word vectors. While there are several ways of constructing words that goes together, we consider vectors made from coordinative construction in Norwegian, constructions like *ost og kjeks* (cheese and biscuits). Coordination structures introduce a certain semantics right from the start – two words are coordinated if they share something in the context in which they are uttered (or written). Each coordination is assigned a particular weight computed from its frequency, in the form of pointwise mutual information (PMI). The computation is done with basis in the corpus (the digitized texts from the Norwegian National Library). Using a relevance measure like PMI ensures to a certain degree that the selected trigrams of conjunctions are full phrasal words, i.e. that the words seen in the coordination are the words coordinated and not the start or end of a phrase.

A word vector V for a given word w, then contains all the words w is coordinated with, and yields a set of word pairs (w, v) for v in V. These word pairs constitute the basis graph for w, where the whole graph is created by collecting similar word pairs for each v in V. A graph for w is then constructed from edges like $w \rightarrow v$ where the process is repeated for each v. Which word pairs to use are determined by their PMI, so that all words coordinated with *ost* (cheese) for example, satisfying the cutoff point, constitute word vector for *ost*. An example graph is provided in figure 1 for the word *kirsebær* (cherry).

Graphs are created using the Python programming language using the module networkx (NetworkX (2016)).

For the analysis, we rely on two key properties of graphs, (1) their clique structure, which consists of subsets of the graph where all nodes are connected to each other, and (2), communities which consist of a partitioning of the graph. Each of these carries information about the semantics of the words that make up the graph.

MAIN FINDINGS

Consider the graph[i] depicted in Figure 1, constructed as described above from the word form *kirsebær* (cherry), displaying some of words related to it. From its visual appearance a couple of word groups are apparent. There is an upper area with inter-

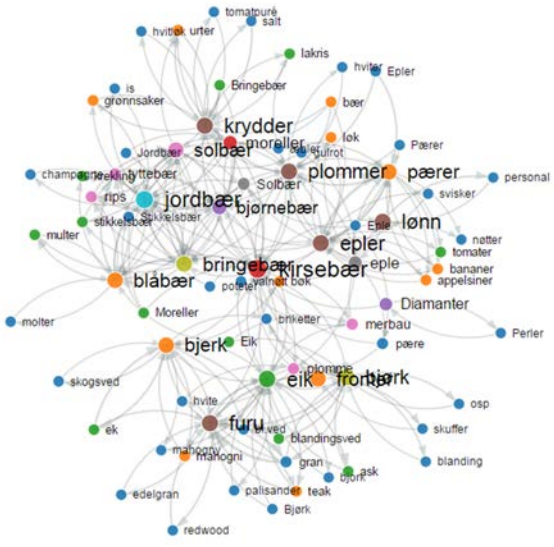


Fig 1: Coordination graph for kirsebær (cherry)

connected elements containing berries, and the lower area containing wood types, in addition to other accidental readings and connections for some of the words.

Clustering from clique structure, using the k-cliques (Chakrabarti and Faloutsos ibid), creates a hierarchy of word sets, ordered by set inclusion. The actual clustering corresponds to the denseness seen in the visual layout. The words building the hierarchy are typically well interconnected and the method may not give any result if the graph is sparsely populated, or for those parts of the graph which do not have interconnected nodes. However when there are enough nodes, the methods finds different readings of words and separate them into groups of similar words. For example, a 4-clique from the graph in figure 1 is yielding one five element list *bjerk, bjørk, furu, kirsebær,* and *eik*, illustrating the wood reading (different kind of trees), while the berry (or fruit) reading comes from a 7-clique (as well as lower n-cliques) with the seven element list *kirsebær, plommer, jordbær, bringebær, moreller, solbær* and *bjørnebær*. All the k-cliques elaborate one of these readings in the hierarchy of sets.

A community analysis on the other hand is using the distance measure between nodes and usis the whole graph. The main difference from k-cliques is that the community analysis creates a partition of all the nodes instead of a hierarchy. However, the different partitions typically contain related words. We illustrate with parts of two sets that complement the k-clique analysis, with a berry reading (containing 22 entries): {*bringebær, jordbær, moreller, solbær, kirsebær*, … }, and a wood reading (25 entries): {*bjørk, eik, ask, teak*, …}.

The main finding is that both k-clique clusters and community detection may be used to find different meaning levels for words, and that k-cliques are in general more conservative with a high precision, while community detection in general creates partitions that covers the whole graphs.

RELEVANCE TO WORKSHOP THEMES

The work described here is especially relevant in the themes of defining Digital Humanities, in particular how computational methods change the way we study and form qualitative knowledge out of quantitative information (see also Turney and Pantel ibid.), as well as the interdisciplinary aspects of using methods from different fields like linguistics (word meanings), mathematics (graph theory) and social sciences (network methods on graphs).

REFERENCES

Deepayan Chakrabarti and Christos Faloutsos 2012. *Graph Mining*, Morgan & Claypool Publishers.
NetworkX (2016). https://networkx.github.io/
Turney P.D. and Pantel P. 2010. From frequency to meaning: vector space models of semantics. Journal of Artificial Intelligence Research, 37:141–188.

[i] This graph and others like it can be studied interactively online, together with a list of k-cliques, at: http://www.nb.no/sp_tjenester/beta/ngram_1/galaxies#terms=kirseb%C3%A6r

Network Visualization for Digital Humanities: Two Case Studies of Visual Analyses for Text Analytics

Ilir Jusufi Department of Media Technology Linnaeus University, Sweden
ilir.jusufi@lnu.se

Andreas Kerren Department of Computer Science Linnaeus University, Sweden
andreas.kerren@lnu.se

INTRODUCTION

Much of the data created nowadays in fields such as Digital Humanities (DH) is of relational nature, such as social or semantic networks. Researchers often decide to depict networks as node-link diagrams to make a better sense of the complex nature of data (cf. Figure 1). Understanding the topology of such a network can be very important. For instance, if we show our friends as network nodes and their friendship as edges between the nodes, it becomes easy to identify groups of friends from different social settings (work friends, high school friends, etc.).

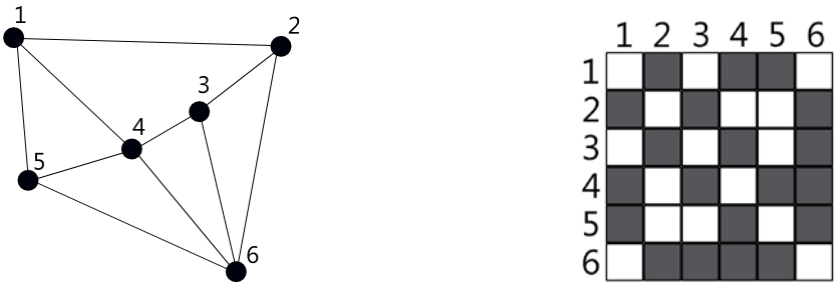


Figure 1. The same (undirected) network depicted by using two different visual representations: as node-link metaphor on the left hand side, and as a matrix metaphor on the right.

Networks usually have additional attributes attached to their elements (Kerren et al., 2014). For instance, we can model a number of documents in a repository as nodes and use edges to describe co-authorship. Additionally, we might want to explore other aspects of such a corpus, like the keywords for each document, its genre, and various other data associated. Here, it is often desirable to get an overview about the network structure and how different data values relate to this structure. For the more general case of visual text analytics, we refer to the surveys of (Kucher and Kerren, 2015; Kerren et al., 2014). In this paper, we present two case studies for visualizations in DH with a focus on publication networks. But first, we will introduce our data sets used in these studies.

Data

Both approaches presented here use the so-called *Jigsaw data set* (D1) containing metadata for every IEEE InfoVis and VAST conference paper (Stasko et al., 2014). Here, nodes represent the papers. The edges between nodes represent co-authorship, i.e., if two papers share an author, then their node representations are connected with an edge. This creates a co-authorship network. Each paper has additional metadata attached to it, such as concept terms for describing the content of the paper or publication and conference data.

An additional data set (D2) has been used for our first case study. This data set is very similar to the first one, but instead of concept terms, we have computed and attached the most used keywords for each document published in conferences within HCI and Information Visualization venues.

CASE STUDY 1

There are several ways to visually encode additional attributes to the nodes in a network (Kerren et al., 2014). For instance, we can position specific nodes close to appropriate regions in the display that hold some specific semantics. One example is shown in Figure 2. In this screenshot we see the *JauntyNets tool* visualizing the D2 data set (Jusufi et al., 2013). The most used keywords, i.e., the attributes, are shown as rectangles in the circular

layout. Each paper with a certain number of occurrences of the given keyword has an invisible link to it. This invisible link acts as a spring and pulls the papers towards itself. The higher the number of occurrences of the keyword, the stronger the pull. In consequence, nodes move to those attributes with the highest pull and may build clusters. Users can additionally group certain attributes according to a desired semantics. In the figure, for instance, we have created three groups of attributes (called *interaction*, *mobile*, *graph*).

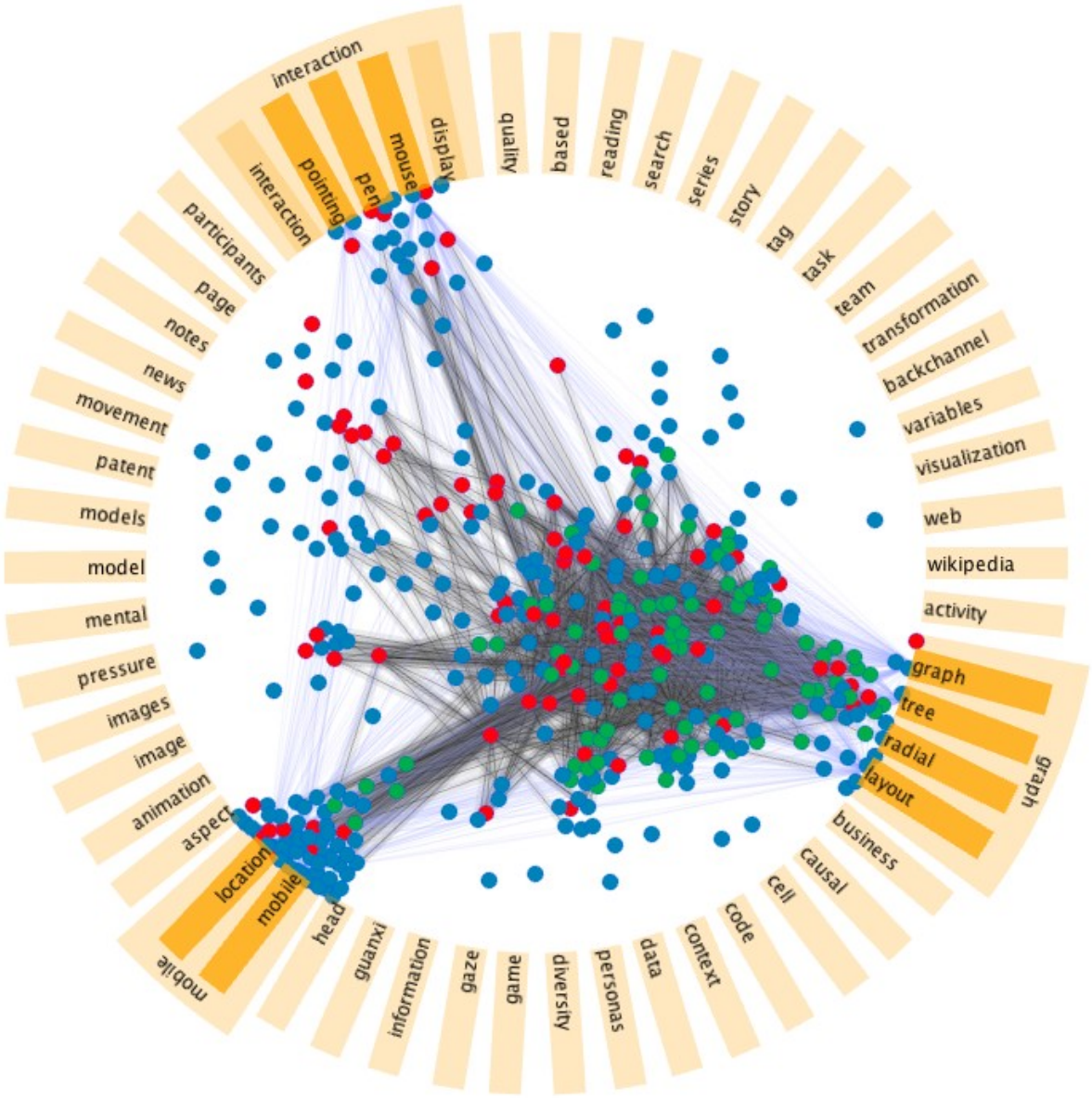


Figure 2. The screenshot displays a network of 421 nodes with 55 attributes (data set D2). The nodes are colored differently, because they have been clustered based on their attribute values. Taken from (Jusufi et al., 2013).

Figure 3 shows the D1 data set. Two attributes *InfoVis* and *VAST* specify the conference where the papers were published. InfoVis papers are shown by blue nodes, while VAST papers are represented by red nodes. The attribute group *geovisualization* has been created manually. Interesting structural features can be noticed immediately, such as a lot of unconnected nodes or subgraphs. Other noticeable structures are a number of cliques. Upon the closer examination of the two cliques marked as A and B, we are able to understand the following short story: there was only one shared author in group A, and he published articles related to geovisualization in both venues; in contrast to group B where three distinct authors collaborated and published exclusively at the InfoVis conference.

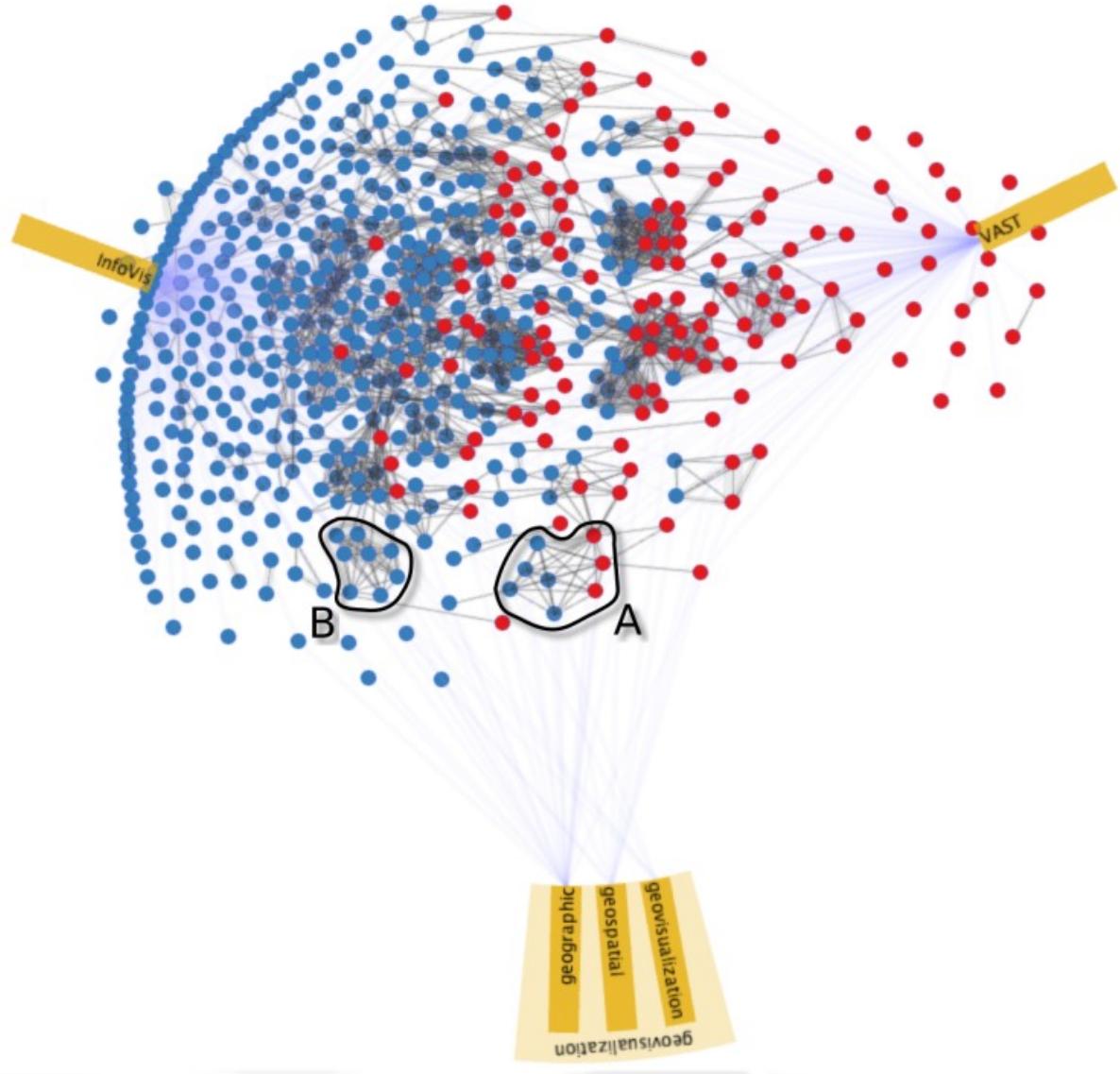


Figure 3. The screenshot displays papers from the D1 data set. Taken from (Jusufi et al., 2013).

CASE STUDY 2

Now, we present a different approach for exploring papers and co-authorship (Jusufi et al., 2014). A k-means clustering algorithm has been used to cluster the papers based on the concept terms from data set D1 (cf. Figure 4). Nodes are placed in different clusters in a circular layout. The edges are shown in two ways: internal co-authorship within a single cluster and external co-authorship going to other clusters. Additionally, co-authorship patterns regarding publication years within clusters can be observed as well by a green color gradient. The most often-occurring concept terms are drawn as tag cloud into the center of each cluster. In Figure 4, we show one of the graphical layouts we implemented. Here, the edges have been bundled together to avoid clutter.

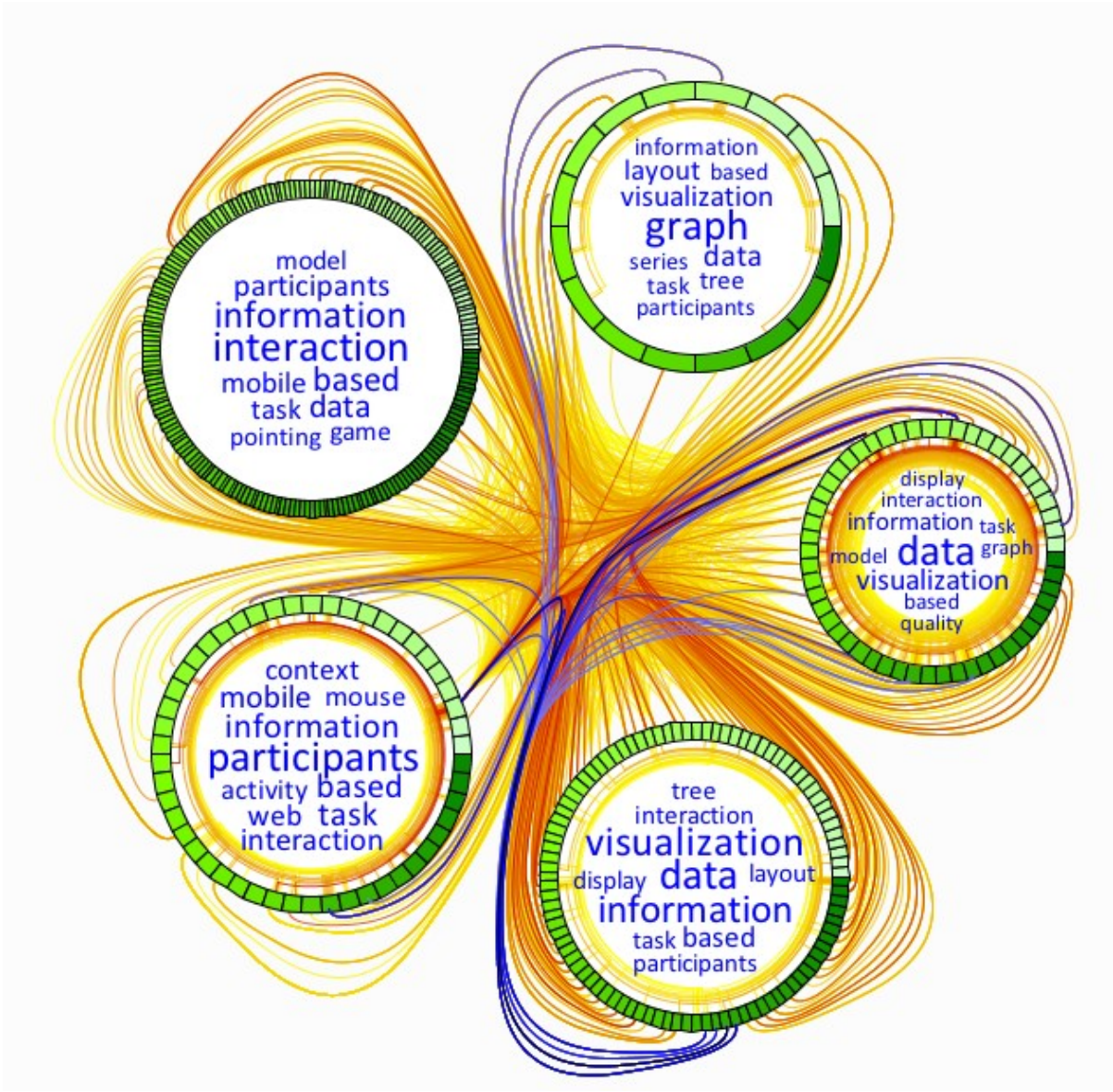


Figure 4. A screenshot of our approach representing five distinct clusters. Nodes (i.e., papers) are represented around each cluster with a green color saturation specifying the publication date. The edge saturation represents the number of shared co-authors. Selected edges are highlighted in blue. Taken from (Jusufi et al., 2014).

As an example how interaction supports the analysis, we might ask if people publish within the same topic, or if the papers are topically distributed within different clusters, see Figure 5. After selecting the only paper in cluster A (highlighted in orange) that has edges to other clusters with more than three co-authors, we can immediately identify related terms, because the concept terms unrelated to the paper are grayed-out. Upon selecting the other two papers in the clusters B+C, we can learn that these papers have almost no similarities.

CONCLUSION

The challenge of analyzing many hundreds or thousands of related text documents increases with the daily growing amount of new data. We have presented two different visualization approaches that could help researchers in DH to investigate the content and metadata of publication corpora. Both approaches are highly interactive. Some insights that were gained by using the presented approaches are not possible with traditional computational methods, especially when users are not aware of what they want to find, i.e., for exploratory analyses.

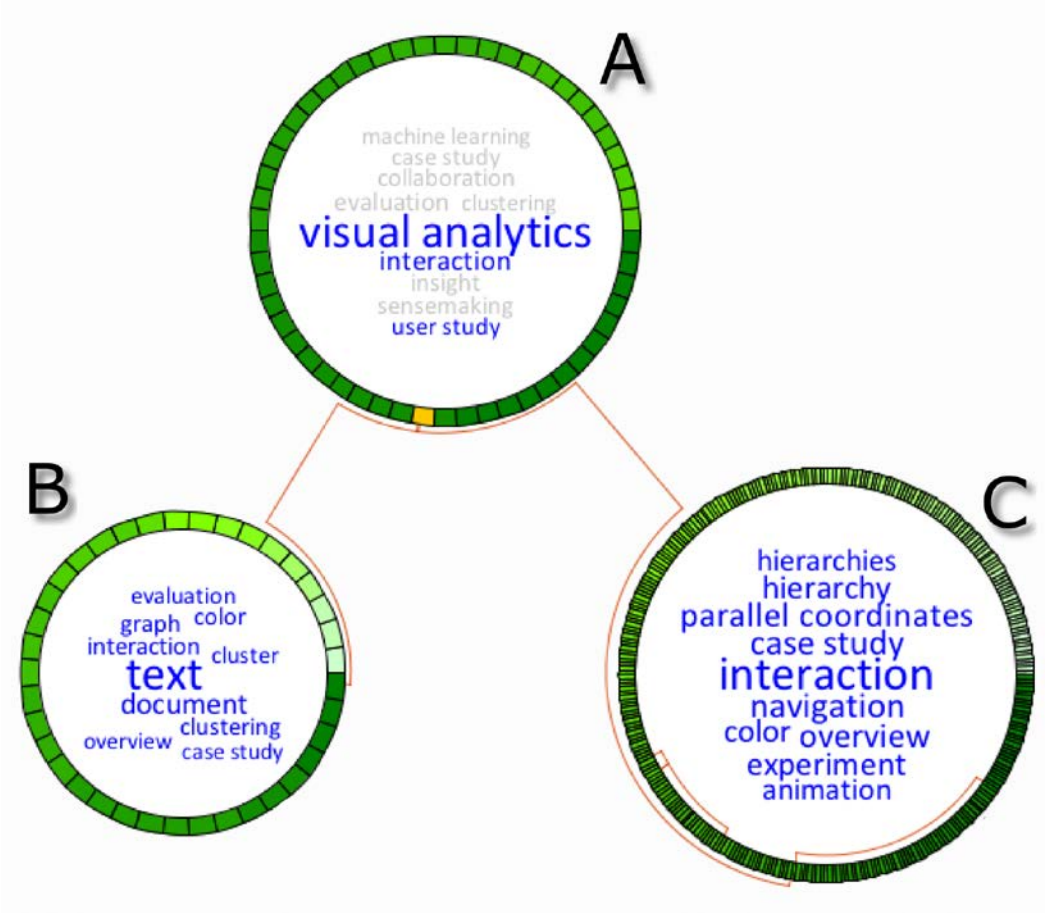


Figure 5. A group of authors who published the selected paper (highlighted in orange) wrote papers that again were grouped into separate clusters. Taken from (Jusufi et al., 2014).

REFERENCES

Jusufi, I., Kerren, A., and Zimmer, B. (2013) Multivariate Network Exploration with JauntyNets. In Proceedings of the 17th International Conference on Information Visualisation (IV '13). IEEE Computer Society, 19–27.

Jusufi, I., Kerren, A., Liu, J., and Zimmer, B. (2014) Visual Exploration of Relationships Between Document Clusters. In Proceedings of the 2014 International Conference on Information Visualization Theory and Applications (IVAPP '14). SciTePress, 195–203.

Kerren, A., Purchase, H. C., and Ward, M. O. (Eds.) (2014) *Multivariate Network Visualization*. Volume 8380 of Lecture Notes in Computer Science (LNCS) State-of-the-Art Survey, Springer.

Kucher, K., and Kerren, A. (2015) Text visualization techniques: Taxonomy, visual survey, and community insights. In Proceedings of the 2015 IEEE Pacific Visualization Symposium (PacificVis '15). IEEE Computer Society, 117–121.

Stasko, J., Görg, C., Liu, Z., Pratap, S., and Sainath, A. (2014, Jan 1) Jigsaw: Visual Analytics for Exploring and Understanding Document Collections. Retrieved from http://www.cc.gatech.edu/gvu/ii/jigsaw/.

Re-approaching new stemmatics: Choice of relationship revealing readings for cladistic analysis.

Katarzyna Anna Kapitan
Department of Nordic Research
University of Copenhagen
kak@hum.ku.dk

AIM

This paper presents preliminary results of my research in the methodological field of computer-assisted stemmatics. It explores the possibilities of applying phylogenetic systematics for textual traditions, as well as discusses the controversies regarding the methodological principles within the field of new stemmatics, such as choice of relationship revealing readings. Additionally, it presents the results of the experiments I have conducted employing the PHYLIP package — a free package of programs for inferring phylogenies, developed by Felsenstein (2005) — aimed to asses an influence of a sample size and types of variants on the results of cladistic analysis. The oldest manuscripts of "Hrómundar saga Gripssonar" — a post-medieval Icelandic saga preserved in 36 manuscripts — served as a case study for the second part of my presentation.

METHODS

Similarities between theoretical assumptions behind cladistics and stemmatology have been explicitly presented by Howe, Barbrook, Mooney and Robinson (2004) in the second volume of "Studies in Stemmatology". An application of computer-assisted methods, originating from phylogenetics, to answer the questions of textual criticism, has been recognized in an academic discourse as a powerful tool in revealing manuscripts' filiation. Yet, there is a disagreement regarding the fundamental question: what kind of textual variation can, or should, be used for analysis? Salemans (1996) suggested a set of strictly systematized classification of parsimony informative variants, while Robinson (1996) has claimed that all types of variants should be analyzed, including linguistic variants.

Another question that arose recently is a role of small samples in the tree-building process, as discussed by Hall and Parsons (2013). It seems obvious that philologists for centuries used some sort of sampling to reveal manuscript filiation, however they usually did not discuss their sampling procedure, nor publish underlying samples. This attitude does not belong to the age of open scholarship, which requires access to data to allow replication of experiments.

This paper takes an experimental approach towards the mentioned problems, thus a number of tests have been conducted to support the claims made by the author, discussed below in the "Main Findings" section.

RESEARCH QUESTIONS

The experiments I have conducted sought answers for the following questions:

- How does linguistic variation influence the manuscripts' filiation?
- How to choose a relationship revealing variants?
- Are complete transcriptions of all witnesses necessary to conduct a computer-assisted analysis?
- If not, how small can a sample be? Can analysis be based only on *loci critici*?

MAIN FINDINGS AND UNDERLYING WORK

In my paper I present the results of the experiments I have conducted with use of Pars — a general parsimony program — and DrawTree and Consense — tree-plotting programs, included in the PHYLIP package. The aim of the experiments was to assess the influence of linguistic and minor variants on the results of a cladistics analysis. Additionally I conducted experiments, which aimed to examine an influence of a sample size on the results. For that purpose I employed a sequential analysis for all the characters I collected from the entire saga. The results of each experiment were plotted into Consense software, in order to achieve a consensus tree on which the stemma can be based.

The results of my experiments suggest that cladistics can be employed in traditional textual criticism, and that computer assisted methods improve efficiency of the analysis by decreasing the time necessary for traditional data interpretation. I claim that an input dataset can be based only on traditional *loci critici,* and the achieved results are as equally valid as the ones based on complete transcriptions, which include all sorts of

variants. Moreover, the presence of linguistic variants seemed to introduce noise to the analysis.

RELEVANCE

My paper aims to present possibilities given to a textual critic by an application of tools developed within evolutionary biology. Even though the discussion on applications of cladistics in philological research has been present in the academic discourse for around thirty years now, as presented by Salemans (1987), the computer-assisted methods seem to still face resistance from traditional philologists, most recently expressed by Trovato (2014). The possible reasons for this lack of trust in cladistics will be briefly addressed in this paper.

REFERENCES

Felsenstein, C. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Hall, A. Parsons, K. (2013). Making stemmas with small samples, and digital approaches to publishing them: testing the stemma of Konráðs saga keisarasonar. *Digital Medievalist,* 9. Retrieved from http://digitalmedievalist.org/journal/9/hall/

Howe, C., Barbrook, A., Mooney, L., Robinson, P. (2004). Parallels between stemmatology and phylogenetics. In. P. van Reenen, A. den Hollander, M. van Mulken (Eds). *Studies in Stemmatology,* vol. II (pp. 3-11). Amsterdam: John Benjamins Publishing Company.

Robinson, P. (1996). Computer-Assisted Stemmatic Analysis and 'Best-Text' Historical Editing. In P. van Reenen, M. van Mulken, J. Dyk (Eds). *Studies in Stemmatology* (pp. 71-104). Amsterdam: John Benjamins Publishing Company.

Salemans, B. (1987). Van Lachmann tot Hennig: cladistische tekstkritiek. *Gramma,* 11. Parts translated in: Salemans B. (1996).

Salemans B. (1996). Cladistics or the Resurrection of the Method of Lachmann: On Building the Stemma of Yvain. In P. van Reenen, M. van Mulken, J. Dyk (Eds). *Studies in Stemmatology* (pp. 3-70). Amsterdam: John Benjamins Publishing Company.

Trovato, P. (2014). *Everything You Always Wanted to Know about Lachmann's Method. A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. Padova:Libreriauniversitaria edizioni.

Folklore Tracks: hGIS and Folklore Collection in 19th Century Denmark

Ida Storm, Holly Nicol, Georgia Broughton, Timothy R. Tangherlini
University of California, Los Angeles
tango@humnet.ucla.edu

INTRODUCTION

Folklore has played a significant role in the "imagining of the nation" since the inception of the field in the late 18th century. In Scandinavia, the "golden age" of folklore collection of the 19th century coincided with rapid changes in political, economic, and social organization. Although some later folklorists have expressed skepticism about these collections, this skepticism is often based on perceived notions of how these collections came to be, rather than a deep exploration of the actual practices of the collectors themselves. We show how techniques from GIS wedded to time tested archival research methods can reveal how a folklore collection came into being. By detailing the routes taken by a single folklore collector over the course of his fifty-year career, we trace not only his selection biases for geographic areas (and by extension, social and economic classes), but also the impact that intellectual currents, political developments and changes to transportation infrastructure have on his collecting. This work considerably extends qualitative assessments of Tang Kristensen's collecting (Christiansen 2013) and is a key contribution toward the development of the "Folklore Macroscope" (Tangherlini 2013).

Our target corpus is the folklore collections of the Dane, Evald Tang Kristensen (1843-1929) who, over the course of his sixty year career, traveled over 10,000 km, largely on foot, visiting ~4,500 storytellers in 4,281 places, recording these stories in 24,000 field diary pages. In this work, we focus on determining how and where Tang Kristensen traveled in Denmark as he created his collection. We develop detailed route maps projected onto appropriate historical base maps showing his movement through the countryside. In all, we map 267 fieldtrips, starting in 1868 and ending in 1916.

METHODS AND RESULTS

Data Extraction

When we began this work, there was no existing catalog of Tang Kristensen's field collecting routes—we had to devise this by coordinating annotations in his hand written field diaries with his four volume memoir *Minder og Oplevelser* (1923-1927). The memoir is based largely on letters he wrote home detailing all of his stops while out collecting, and includes information on means of transportation as well as travel dates. Our team began by making "proto-routes." We extracted trip start and end dates, as well as all stops and stop order for each trip by hand, and aligned these proto-routes with the field diaries (Fig. 1). In later work, we will also align field stops with our electronic catalog of informants.

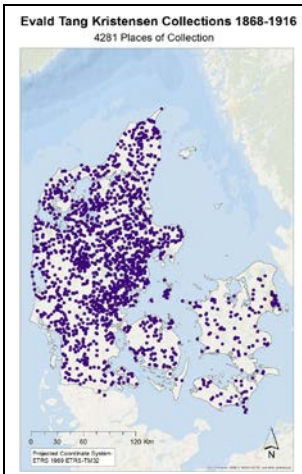


Figure 1: aggregated stops geocoded with the address locator.

Figure 2: Example of a route.

Address Locator

Finding the locations for the stops we extracted in our "proto-routes" was a significant challenge. As with most historical data, places can be difficult to locate: some are very small, names have changed, and some places have disappeared. Contemporary gazetteers are inadequate to the task and often confound, rather than solve, queries. To address this problem, we downloaded the historical place name database developed by the Afdeling for Navneforskning, Københavns Universitet, and used it to generate a customized "Address Locator" for 19th century Denmark. We matched the stops from each fieldtrip with the address locator, generating a "best guess" for each fieldtrip. Multiple places with the same name were resolved through the "Interactive Rematch" interface. To derive the final fieldtrip stops, each trip was inspected individually.

Routes

With the stops in a provisional sequential order, we created the "most likely route" for each trip. A basic assumption was that, unless otherwise specified, Tang Kristensen would take the shortest path between two points, an assumption that aligns with the underlying "Network Analyst" algorithm in ArcMap. We used a transportation network from OpenStreetMaps pruned against the cadastral survey maps of ~1880, the highest resolution historical maps from the era. Since Tang Kristensen occasionally traveled by boat, ferry lines based on ferry schedules and close study of historical maps were also added. By feeding the provisional sequential stops to the network analyst, we were able to create the most likely routes for each trip as a single line record. These routes were then visualized as a line with sequentially numbered stops (Fig 2). Simple statistics, such as route length, as well as descriptors from our database, such as dates of collection, field diary pages, and modes of transportation augment the visualization.

Animations

Animations provide a dynamic representation of Tang Kristensen's movement through the countryside than static route maps that augment the static representations. These animations reveal, for example, the numerous times where he backtracked. To allow for sequential animations of all fieldtrips, we devised an additional "absolute order" field, and split all routes into inter-stop segments.

Travel Statistics

By splitting routes into inter-stop segments, we could develop more detailed statistics regarding segment length, speed of travel, and travel mode. More importantly, we can now aggregate segment statistics, and align this information with other data, allowing us to address a broad range of questions. For example, we can see how far he traveled when he lived in a specific place, his travel distances at different times of year, and his travel distances in different parts of the country. Furthermore, we can consider changes in average travel segment or fieldtrip distance over time. Future work will align stops with storytellers, allowing us to include story statistics with the fieldtrip statistics. Population data and transportation data will further add to this picture.

CONCLUSIONS

Our work reveals the shifting parameters of Tang Kristensen's field collecting, from his intensely local focus early on to his more expansive and confident travels at the end of his career, when his collecting was no longer aligned with Romantic nationalist goals, but more in tune with a thick descriptive approach to Jutlandic rural life. By using hGIS techniques, we can provide a degree of detail about his travels missing in earlier studies. Our approach enables a truly macroscopic approach to folklore collecting, allowing us to interrogate Tang Kristensen's field collecting at varying levels of resolution. For example, we can move from the micro-consideration of a single fieldtrip, to a meso-consideration of all trips that included a particular parish, to a macro-consideration of all his trips taken as a whole.

REFERENCES

Christiansen, Palle Ove. *Tang Kristensen og tidlig feltforskning i Danmark. National etnografi og folklore 1850-1920*. Copenhagen: The Royal Danish Academy of Sciences and Letters, 2013.

Tang Kristensen, Evald. *Minder og Oplevelser.* Volumes 1-4. Viborg: Forfatterens forlag, 1923-27.

Tangherlini, Timothy R. "The Folklore Macroscope." *Western Folklore* 72.1 (2013): 7-27.

Digitization of Material Written in Three Scripts and Three Languages: A Croatian Example of Cross-Institutional Collaboration in DH

Assistant Professor Marijana Tomić, PhD
Department of Information Sciences
University od Zadar, Croatia
mtomic@unizd.hr

INTRODUCTION

Zadar is a small town situated at the middle of Croatian part of east Adriatic coast. Its convenient geopolitical position, as well as its indented coast contributed to its imposition as an important cultural and political center of Dalmatia throughout the rich history of Croatia. Its university tradition is centuries-old and begins with ecclesiastical education first mentioned in 10th century. In 2002, the University of Zadar was founded and a base of contemporary University is faculty of Philosophy, Teaching College, as well as Dominican university.

Due to its geographical position on the crossroads of Western and Eastern Europe, Croatian had a significant role in European history and culture, with a strong influences of Latin as well as Byzantine cultures. The corpus of Croatian Mediaeval literature is therefore written on three scripts – Glagolitic, Latin and Bosnian variant of Cyrillic scripts, called *bosanica*, and in three languages – Church Slavonic, Croatian redaction of the Church Slavonic language and Latin (Hercigonja, 1999, 2009). Although Glagolitic script was in use in several other countries, it is only in Croatia that it took angular form, used also for the first Croatian printed books, as early as in 1483. The most important Mediaeval written heritage in Croatia is written on Glagolitic and Latin scripts, and an important part of that corpus is nowadays part of collections of heritage institutions in Zadar.

In accordance with current interdisciplinary trends in cataloguing, research and communication of cultural heritage in digital age, the interdisciplinary scientific project *Digitization, bibliographic description and research of texts written on Glagolitic, Croatian Cyrillic and Latin scripts until the end of 19th century in Zadar and Šibenik area* is being carried out at the University of Zadar in co-operation with Vestigia Manuscript Research Centre of University of Graz, Austria. The goals of the project are: (1) **digitization** of old and rare books written on Glagolitic, Croatian Cyrillic and Latin scripts, primarily manuscripts (2) **Machine readable cataloguing** based on existing printed catalogues, **making of inventory lists, registration and inclusion** of Croatian manuscript and early print, primarily Glagolitic, collections in Croatian and European portals of written heritage. This goal includes the research in the field of information sciences, by focusing on research of standardization of bibliographic description of manuscripts and early prints and their digitization, and in the fields of digital humanities and humanities user needs. In order to achieve this goal, the project connects also to national project *Production, publishing and maintaining national cataloguing rules: 2014-2016*. This goal became the major goal, as research in semantic web, linked open data and other standards of publication of data, as well as research data are

conducted here. FRBR-LRM, CIDOC-CRM and other conceptual models, as well as machine-readable formats are being mapped and researched together with standards for description of manuscripts in different communities, in order to achieve appropriate metadata scheme for the description of manuscripts coming from different institutions. (Varniene-Janssen, Juškys, 2013) (3) **Gathering scientists** from various scientific fields, especially humanities, interested in research of Croatian written heritage in digital environment, to enable their research of written heritage, to research their needs and to create proper conditions for their research, to digitize the material of their interest, as well as to implement and develop tools for manuscript research in digital environment (tools for transliteration and transcription, visual tagging, etc.) (Holm, Jarrick, Scott, 20015) Some tools are designed specially for palaeographic research of Glagolitic material and will be tested on the project (visual tagging of digitized pictures of Glagolitic and Cyrillic manuscripts) as well as tools for visualization of research data on manuscripts and collections (Essert, et al.); (4) **communication of the heritage** is an important part of a project, respectively presenting heritage in a contemporary creative and innovative manner to various groups of users, using information technology, web portal, virtual exhibitions, presentations, mapping locations on touristic and other geographic maps as well as in the wider touristic offer. In this goal, inclusion of public is planned, particularly of associations of citizens interested in Glagolitic heritage, the most important segment of Croatian written heritage, following the EU recommendations connected with open science, citizen engagement and citizen researchers in the age of digital culture. As the project supports educational activities at university, it involves students in each project phase and enables conditions for laboratory and field education. In order to achieve those goals, several working groups are formed and workflow with indicated activities and methodology for the activities of each group are designed, and will be presented in this presentation, together with the assumptions for cross-institutional cooperation in the field of digital humanities will be listed on the example of Zadar University and main research goals and expected results will be presented.

REFERENCES

Herciognja, E. (2006) Tropismena i trojezična kultura hrvatskoga srednjovjekovlja. Matica hrvatska, Zagreb.

Hercigonja, E. (1999) Glagolitism and Glagolism. Ivan Supičić (ed.). *Croatia in the Early Middle Ages*. London: Philip Wilson, pp. 369-400.

Hercigonja, Eduard. 2008. Glagolism in the High Middle Ages. Ivan Supičić (ed.). *Croatia in the Late Middle Ages and the Renaissance*.London: Philip Wilson, 171-226

Essert, M., et al. (2013) XML to visual tags migration: proposed methodologies for the research of digitized Croatian Mediaeval Glagolitic material. *Summer School in the Study of Historical Manuscripts: Proceedings*. University of Zadar.

Varniene-Janssen, R., Juškys, J. (2013) Strategic, methodological and technical solutions for the creation of seamless content of the digital cultural heritage: Lithuanian approach. *Summer School in the Study of Historical Manuscripts: Proceedings*. University of Zadar.

Clemens, R., Graham, T. (2007) Introduction to manuscript studies. Cornell University Press, Ithaca & London

Savenije, B., Beunen, A (2012) Cultural Heritage and the Public Domain. // Liber Querterly : the journal of the Association of European Research Libraries, vol. 22, no. 2, pp. 80-97

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Available at: https://openaccess.mpg.de/Berlin-Declaration

Terras, M., Nyhan, J. Vanhoutte, E. ed. (2013) Defining digital humanities: a reader. Ashgate, Burlington.

Warwich, C. (2012) Studying users in digital humanities. *Digital humanities in practice.* Facet Publishing, London.

Holm, P., Jarrick, A., Scott, D. (2015) Humanities World Report. Palgrave Macmillan

Warwick, C. (2012) Institutional models for digital humanities *Digital humanities in practice.* Facet Publishing, London.

When Big(gish) Data Goes Online: Cross-disciplinary opportunities and challenges

Jukka Tyrkkö
Linnéuniversitet

The last three decades have witnessed a remarkable increase in the number and volume of linguistic corpora available to the research community. Structured corpora comprising hundreds of millions or even billions of words of data are no longer unusual, and unstructured data sets such as Google Books, which are increasingly used in a very corpus-like manner, can encompass over a hundred billion words. Many of these large datasets are also made available online, and server-side query tools such as CQPweb, SketchEngine and the Brigham Young front-end to MySQL make it easy for anyone to use very large corpora both quickly and efficiently. While these corpora may fall short of criteria used to define 'big data' in some disciplines, the volume of text available is typically far beyond anything a single researcher or a research team could ever hope to process either manually or with the help of rudimentary search tools. However, while online corpora do open up new worlds of discovery, they also typically impose considerable limits to the types of queries available, provide quantitative data in difficult to process and sometimes misleading manner, and generally do not allow the researcher direct access to the underlying full datasets, more often than not for reasons of copyright and publishing agreements.

Although many of these large text collections and corpora were primarily designed with the linguist in mind, scholars from a wide variety of fields within the humanities and social sciences are also increasingly turning to these data sets for both qualitative and quantitative evidence, such as finding illustrative quotes or indications of diachronic trends that support theoretical arguments. Instead of extrapolating arguments from small and necessarily anecdotal evidence, humanities scholars are increasingly open to the idea of studying cultural, societal and political questions using 'big data' and methodologies such as *culturomics* (Michel et al 2011; Nunberg 2010) and *distant reading* (Moretti 2005). As the conceptual and methodological worlds of qualitative and quantitative research collide, the new challenge is how to operationalize joint research endeavors in the most beneficial fashion (see, e.g., McEnery and Baker 2016).

In this paper, I will discuss some of the opportunities and challenges that these large data sets and online interfaces can bring about, drawing examples from a collaborative project involving a team of social scientists and a corpus linguist. Using the *British Hansard Corpus*, a computer-readable, richly annotated edition of British Parliamentary debates (1803-2005), our objective has been to challenge certain claims made in political science about country references in historical political discourse, namely, that references to foreign nation states as examples to be followed only emerged as a major discursive strategy of policy-making around the time of the Second World War (Meyer et al 1997). The 1.6-billion-word dataset, which includes 7.6 million speeches delivered by over 40,000 MPs, is a new kind of historical corpus: not a sample drawn from an amorphous population, but an exhaustive and arguable complete record of a specific well-defined register of language use. Fully annotated both for standard linguistic variables and semantically tagged using data from and the conceptual network developed for the *Historical Thesaurus of the Oxford English Dictionary* and the Samuels semantic tagger (Alexander et al in press), the Hansard corpus has proven extremely useful and informative, but the data has also coughed up various surprises and potential problems, particularly if one were to rely solely on the online interface. In the present paper, the pros and cons of the online version and the standalone corpus are discussed and evaluated with

particular reference to their usefulness in cross-disciplinary (digital) humanities projects, where efficient data management and ease of accessibility have to be balanced with the inherent complexity of textual accounts of ideas and concepts.

References

Alexander, Marc and Mark Davies. 2015-. Hansard Corpus 1803-2005. Available online at http://www.hansard-corpus.org.

Alexander, Marc, Fraser Dallachy, Scott Piao, Alistair Baron, Paul Rayson. In Press. Metaphor, Popular Science and Semantic Tagging: Distant reading with the Historical Thesaurus of English'. In *Digital Scholarship in the Humanities (DSH)*.

Alasuutari, Pertti, Marjaana Rautalin and Jukka Tyrkkö. Accepted. The formation of interdependent decision-making: The case of British Parliament, 1803-2005. Presentation to be delivered at The Australian Sociological Association conference. Melbourne. 28.11-1.12.2016.

McEnery, Anthony and Helen Baker. 2016. *Corpus Linguistics and 17th-century Prostitution: Computational Linguistics and History*. (Corpus and Discourse). London: Bloomsmury Academic.

Meyer John W, John Boli, George M. Thomas and Francisco O Ramirez. 1997. World Society and the Nation-State. In *American Journal of Sociology*. 103(1): 144–181.

Michel, Jean-Baptiste, Erez Lieberman Aiden et al. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. In *Science* 331, 176-182 (published online ahead of print: 12/16/2010). <http://www.sciencemag.org/content/331/6014/176.full.pdf>.

Moretti, Franco. 2005. *Graphs, maps, trees: Abstract models for a literary history*. London: Verso.

Nunberg, Geoff. 2010. Humanties research with the Google Books corpus. http://languagelog.ldc.upenn.edu/nll/?p=2847.

The ‘Rise of the Eurosceptics’ in 2014: A Europeanized Media Discourse about Euroscepticism

Abstract:
2014 can be considered a ’breakout year’ for Euroscepticism. Largely on account of the 2014 European Parliament (EP) elections, where over one quarter of the available seats went to members of Eurosceptic parties (Treib, 2014), national media outlets in many Member States reported about the ‘earthquake’, ‘virus’, or ‘rising tide’ of Euroscepticism sweeping across the continent. Although Eurosceptics tend to oppose European integration, the media discourse about them may – rather paradoxically – drive a process of Europeanization that grants saliency to European issues and actors across different national cases. The present study investigates whether national media discourses about Euroscepticism can be considered *Europeanized*, that is, when an issue involving non-national or EU level actors is discussed in a similar fashion across different Member States. We ask:

To what extent are national media discourses about Euroscepticism Europeanized?

What factors explain the presence of a Europeanized media discourse about Euroscepticism?

To answer the first question, we use LDA topic modelling in R to expound the main topics associated with the word ‘Euroscepticism’ in 2014 national print media articles (N=1545) across six countries: Sweden, Denmark, Ireland, the United Kingdom, France, and Spain. LDA topic modelling is a digital method suitable for digital humanities (Blei, 2014), since it identifies recurring themes across a large corpus of texts. However, topic modelling is rarely applied to cross-national (and thus cross-language) data corpora. We create six LDA models and represent the results in a Gephi visualization. The findings demonstrate a strong convergence of European topics associated with Euroscepticism in five of the six cases, with the United Kingdom as the exception, and we conclude the media discourse about Euroscepticism is Europeanized.

In the second phase, we test which factors explain this Europeanization through using bivariate logistical regressions. We find that the type of newspaper (tabloid or broadsheet), the existence of a successful Eurosceptic party, and the country’s relationship to the EU budget each can predict whether Euroscepticism is discussed in a national or European scope.

While our study focuses on print media articles, we know that contemporary media systems are of a ‘hybrid’ nature (Chadwick, 2013). Thus, the print media articles in this study have also been disseminated via online channels like mainstream news websites and social media. Our poster will focus on the digital methods employed to operationalize our research questions across disparate national cases. The methodology, and aim of the

study, connect to the themes of the workshop by using digital methods, via computational tools, to address a pan-European societal challenge through an interdisciplinary approach at the nexus of media studies and political communication.

References:

Blei, David M. "Topic modeling and digital humanities." Journal of Digital Humanities 2.1 (2012): 8-11.

Chadwick, Andrew. *The hybrid media system: Politics and power*. Oxford University Press, 2013.

Treib, O. (2014). The voter says no, but nobody listens: causes and consequences of the Eurosceptic vote in the 2014 European elections. *Journal of European Public Policy*, *21*(10), 1541-1554.

Authors:

Michael Bossetta
PhD student at the Centre for European Politics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark. Phone: +45 35 33 12 26. Email: mjb@ifs.ku.dk

Anamaria Dutceac Segesten
Assistant Professor in European Studies, Lund University
PO Box 201, 22100 Lund, Sweden. Phone: +46 46 222 8821
Email: anamaria.dutceac_segesten@eu.lu.se

Poster Proposal:
Digital (hi)storytelling – counterfactuals and fiction in digital culture

Ylva Grufstedt & Cecilia Trenter
University of Helsinki Linnaeus University
ylva.grufstedt@helsinki.fi
cecilia.trenter@lnu.se

Digital (hi)storytelling – counterfactuals and fiction in popular culture

Complex and heterogeneous storyworlds are mediated through various media by means of several aesthetic expressions, and digital storytelling is a relevant aspect of mediation. This abstract gives an overview of two research projects on digital storytelling that are closely related and focus on the same category of questions, using different methods and materials. The purpose is to highlight one of many current international research alliances on the topics of social memory, adaptation, movie and game studies within the context of digital storytelling.

Counterfactuals in game design and models of historical change

As part of her ongoing PhD project, Ylva Grufstedt has taken an interest in game design and models of causality and historical change in strategy games like *Europa Universalis* and *Civilization*. In this type of simulatory games, players may change history by manipulating various factors based on the rule set and framework created by the developers. Her research focuses on the way these games engineer and facilitate models of historical change by enabling alternative, i.e. counterfactual, (hi)storytelling through gameplay, by way of mapping the specificities of video games as an interactive form of historical modelling, world building and storytelling. The expressed aim is to better understand the interplay of agency, counterfactuals and historical consciousness by analyzing gameplay videos of counterfactual play.

Gameplay and social memory

In Cecilia Trenter´s part of the project *Medeltidens form- och tankevärld i Dragon Age: Origins* (Medievalism in Dragon Age: Origins), she examined the use of fictitious history by studying Dragon Age: Origins and Dragon Age II (2009-2011 BioWare) within the fantasy-medieval context. By reflecting the present in the past, people are not only creating images of who we are but also what it takes to reach our goals. Since the identity-process mobilizes action, it is of interest to investigate how gameplay employs social memory in the creation of an epical action-plot. The project is based on the assumption that understanding the projective identity through a PC as a part of a flow of time with a past, a present and a future in the game, is a crucial part of the gameplay. [1]

[1] Gameplay and historical consciousness in Dragon Age: Origins and Dragon Age II (Bioware) (published at Meaningful Play 2012 at State University of East Lansing http://meaningfulplay.msu.edu/proceedings2012/).

Methodology and Applications of Visual Stance Analysis: an Interactive Demo

Kostiantyn Kucher and Andreas Kerren
Department of Computer Science Linnaeus University, Sweden
{kostiantyn.kucher,andreas.kerren}@lnu.se

Carita Paradis
Centre for Languages and Literature
Lund University, Sweden
carita.paradis@englund.lu.se

Magnus Sahlgren
Gavagai AB, Sweden
mange@gavagai.se

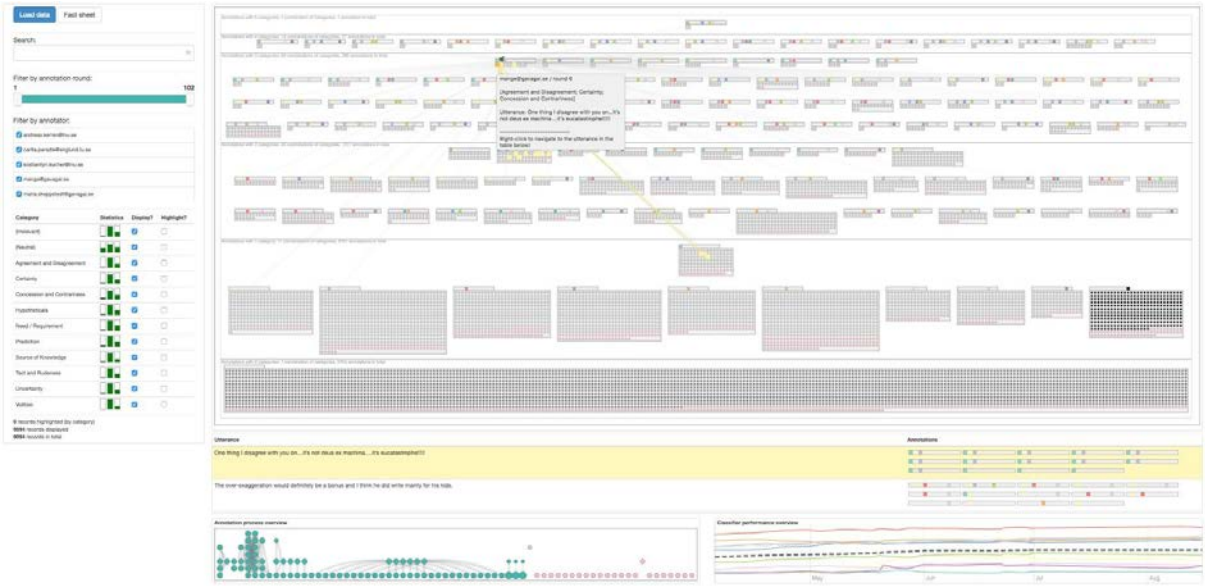


Figure 1: Visualization of text annotation data in our visual analytics system ALVA.

ABSTRACT

Analysis of stance in textual data can reveal the attitudes of speakers, ranging from general agreement/disagreement with other speakers to fine-grained indications of wishes and emotions. The implementation of an automatic stance classifier and corresponding visualization techniques facilitates the analysis of human communication and social media texts. Furthermore, scholars in Digital Humanities could also benefit from such an approach by applying it for literature studies. For example, a researcher could explore the usage of such stance categories as certainty or prediction in a novel. Analysis of such abstract categories in longer texts would be complicated or even impossible with simpler tools such as regular expression search.

Our research on automatic and visual stance analysis is concerned with multiple theoretical and practical challenges in linguistics, computational linguistics, and information visualization. In this interactive demo, we demonstrate our web-based visual analytics system called ALVA, which is designed to support the text data annotation and stance classifier training stages (Kucher, Kerren, Paradis, & Sahlgren, 2016). In contrast to existing tools used for similar tasks, ALVA combines multi-label annotation, active learning, and visualization of annotated data. ALVA supports separate user roles of data annotators (for instance, linguists without prior training in information visualization) and analysts to facilitate the annotation and training processes. It provides annotators with a clean, simple interface to label utterances with multiple stance categories. It also provides the analysts with several visualizations to support exploratory visual analysis of collected annotation data and facilitate classification improvements. Figure 1 demonstrates how our visualization representation based on the semantic substrates principle (Shneiderman & Aris, 2006) groups text annotations by combination of tagged

stance categories. We are currently using text data in English collected from blogs with our previous tool uVSAT (Kucher et al., 2016), but our approach could also be used to train the classifier based on another text genre, language, or set of categories.

We also demonstrate how visual stance analysis could be practically applied to literature studies by combining the automatic stance classifier with text visualization principles (Kucher & Kerren, 2015). Our prototype depicted in Figure 2 provides an overview of stance classification results for a fiction text (divided into utterances). The overview consists of scatter plots for individual stance categories, resembling the document overview in uVSAT (Kucher et al., 2016). Each positively classified utterance is represented by a dot marker in the corresponding plot, and its position in text is mapped to the dot's position. The overview supports details on demand and navigation over the text. The prototype also provides a detailed text view with stance category labels and details on demand, thus supporting both distant and close reading approaches (Jänicke, Franzini, Faisal, & Scheuermann, 2015). Furthermore, classification confidence values reported by the classifier are mapped to the opacity of overview markers. They are also used for filtering to focus only on more reliable results. The prototype can be used to estimate the number of utterances with detected stance in a given text, compare the results for several stance categories, and explore the text in detail. With the stance classification accuracy improving over time, we believe such an approach could be useful for scholars in Digital Humanities.

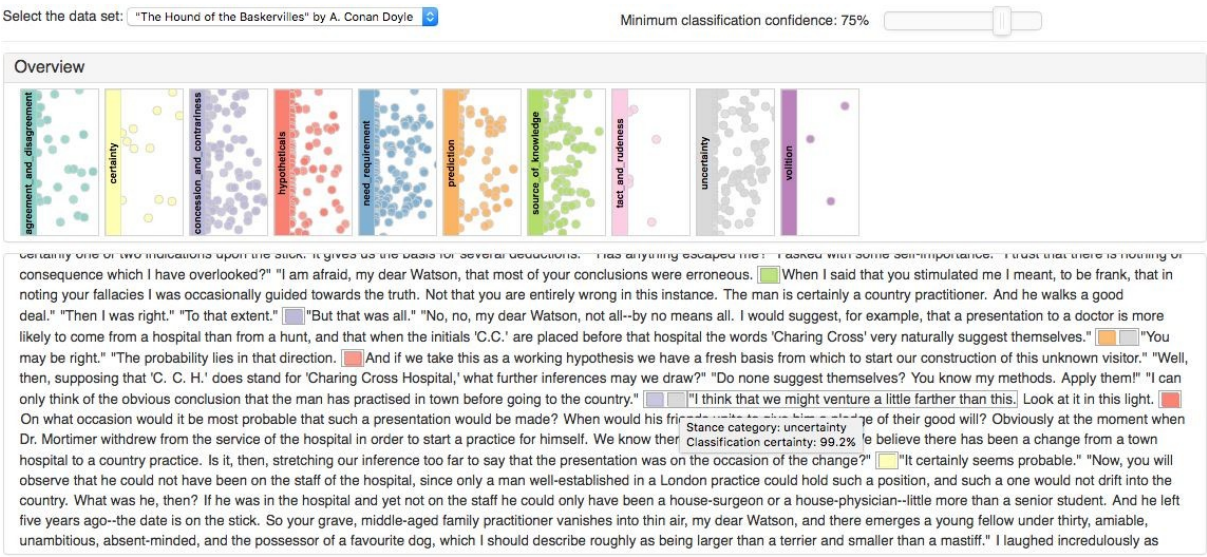


Figure 2: A prototype visualization of stance classification results for literature.

REFERENCES

Jänicke, S., Franzini, G., Faisal, C., & Scheuermann, G. (2015). On close and distant reading in digital humanities: A survey and future challenges. *Proceedings of the EG/VGTC Conference on Visualization (EuroVis '15) — STARs*, DOI 10.2312/eurovisstar.20151113.

Kucher, K., & Kerren, A. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. *Proceedings of the 8th IEEE Pacific Visualization Symposium (PacificVis '15)*, 117–121, DOI 10.1109/PACIFICVIS.2015.7156366.

Kucher, K., Kerren, A., Paradis, C., & Sahlgren, M. (2016). Visual analysis of text annotations for stance classification with ALVA. *Poster Abstracts of the EG/VGTC Conference on Visualization (EuroVis '16)*, 49–51, DOI 10.2312/eurp.20161139.

Kucher, K., Schamp-Bjerede, T., Kerren, A., Paradis, C., & Sahlgren, M. (2016). Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization*, **15,** 2, 93-116, DOI 10.1177/1473871615575079.

Shneiderman, B., & Aris, A. (2006). Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, **12**, 5, 733-740, DOI 10.1109/TVCG.2006.166.

Växjö Go! - a Digital Humanities Poster Project Proposal to Support Inclusion and Promote Local Culture

Alexandra Stiernspetz-Nylén
Det Fria Ordets Hus, Växjö
Alexandra.StiernspetzNyhlen@vaxjo.se

Susanna Nordmark
Linnaeus University, Växjö
susanna.nordmark@lnu.se

Mathias Boström
Musik i Syd, Växjö
mathias.bostrom@musikisyd.se

INTRODUCTION

The Digital Humanities (DH) project at Linnaeus University (LNU) have provided the external partners of this proposal with the opportunity to each pinpoint their needs for expressing and sharing their respective organizations' everyday activities through mobile digital technologies. It is the main object for each external partner to promote societal inclusion in Växjö, and to increase the overall knowledge of what they each can offer the public regarding local literature and music. This project proposal mainly applies to DH Symposium theme 3, and to subthemes 1, 4, 7 & 8.

Inspired by the world-renowned mobile game "Pokemon Go!" we are aiming at developing a "Växjö Go!" mobile application that aspires to inspire activity, inclusion and collaboration, by providing users with access to some of Växjö's cultural points of interest (POI). Each POI can be collected as rewards (either bronze, silver, gold or diamond icon quills (for literature POI:s) or ditto G clefs (for music POI:s). Each POI will offer additional information (e.g. an author's portrait, a text example, a piece of music, etc.) connected to the author or composer. You can also share your daily rewards score along with your opinion of the provided additional info by using emoticons.

Additionally, for Växjö as a city with rapid growth in multiple sectors, the application can serve as a tool for increasing inclusion and kinship of the citizens of Växjö. With the application, users can express their daily mood with the "How is Växjö Today?" feature by choosing a color and one of a select number of emoticons. The mood votes will be GPS mediated so that each Växjö district get its own vote, alongside the grand total for all districts. The results will be accessible on a Växjö map in the app where you can zoom in on each district or see the overall Växjö result, hour by hour. The mood results can also be visualized on the public displays of Växjö, such as on the buses, in the libraries, at schools, at the municipality head offices, etc.

REALIZATION MODE AND POSSIBLE RESEARCH AREAS

This DH project proposal builds on two basic ideas. One is to present some of the cultural points of interests of Växjö everyday life, and the other is to try to increase the overall interconnection between different Växjö districts by asking and visualizing "How is Växjö today?"

The application will offer several modes of gathering and sharing POI:s, of which two are the most prominent: Either you go for an active "culture POI hunt" and receive rewards at each POI, or you will be duly prompted when just passing a POI (provided that you have activated this function in the application). Regardless of method, users have the possibility to gather POI:s, learn more through the additional information provided, achieve rewards (quills for literary rewards, G clefs for music rewards), and share results with others. The quill/G clef is clickable for the user to retrieve additional information (e.g. a short portrait, a piece of music, or an URL to further info).

The “How is Växjö today” part of the application will allow everyone with the app to participate, however you need to allow the app to use location services on your mobile device. When voting, the applications automatically includes GPS coordinates, time of day, and emoticon + color chosen by the user. The voting procedure can be fulfilled twice a day per user. The results will be analyzed and visualized on a Växjö map in the application, and presented on public displays in Växjö, e.g. on buses, at the university, in schools, at the libraries (schools, town and university), etc.

The design and test group will involve external project partners (which will have the possibility to further invite participants from ongoing related projects in their respective organizations), LNU researchers, pupils from upper secondary schools in Växjö, and other representatives of the Växjö citizens.

Proposed research areas for an upcoming study could be:

- In what ways can initiatives like this make a difference? Will this particular combination of “the digital” and “the humanities” really make people more interested of cultural elements in their immediate surroundings? And, in which ways will it (if at all) encourage people to meet over the invisible but still obvious borders between different parts of Växjö?
- What are the requirements (and on what levels do they apply) for inspiring further cultural institutions to collaborate and share their various materials through the proposed technology?

We can already now see that the Växjö Go! application could easily include more partners, e.g. Kulturparken Småland (with “Historical Växjö”), the local Facebook group “Images of Old Växjö”, and the Tourist Bureau (with “Current Events” etc.). The initial thought has been and still is to create a tool that could serve multiple cultural Växjö interests.

However, a limitation at the very end - fulfilling this project proposal is one thing, but “who” (i.e. what Växjö “entity”) would take over the everyday management and support of the application? We see that application ownership and everyday maintenance needs to be defined as one of the first steps of this proposal in order to avoid extensive delays, or even project failure.

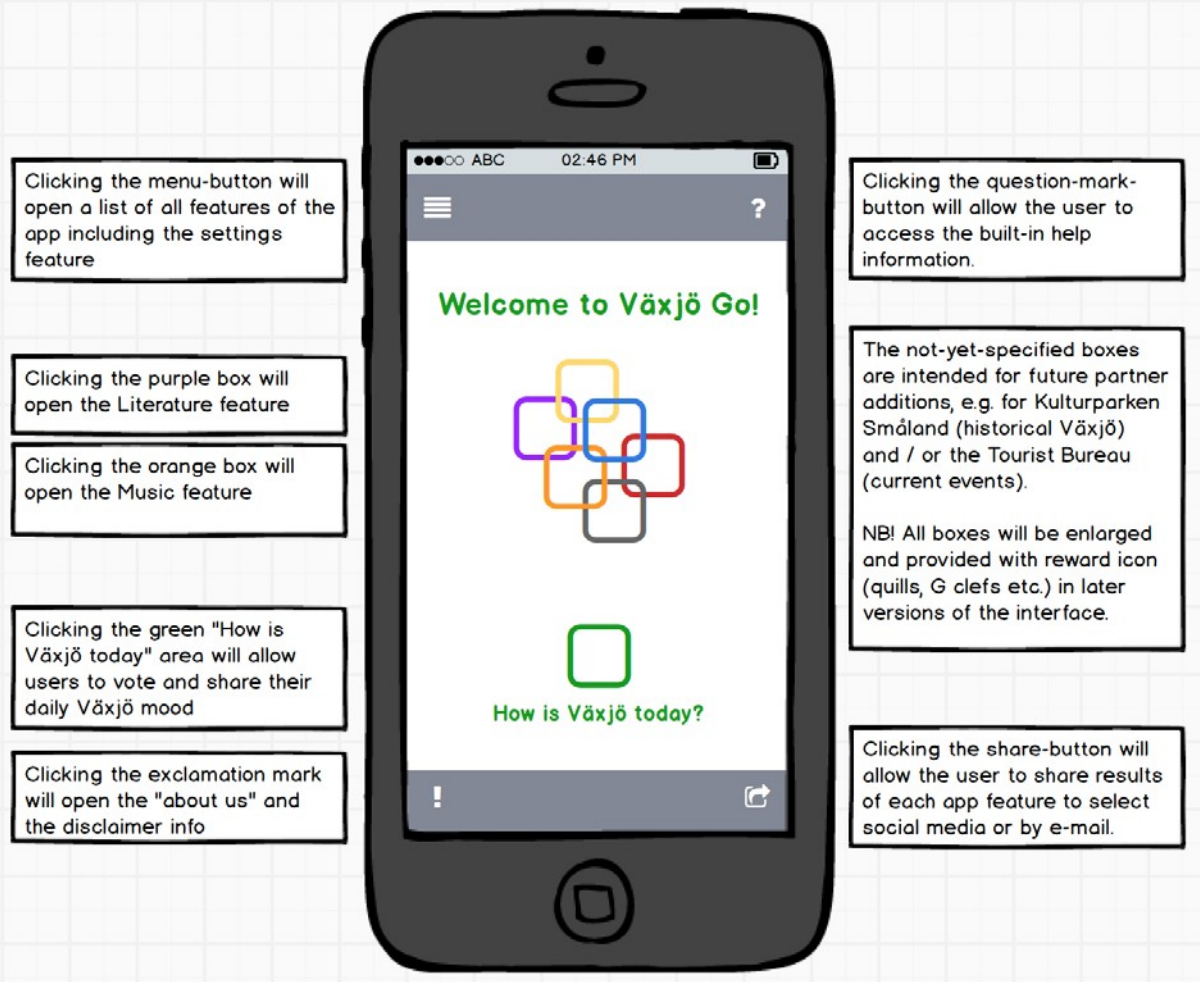
REFERENCES

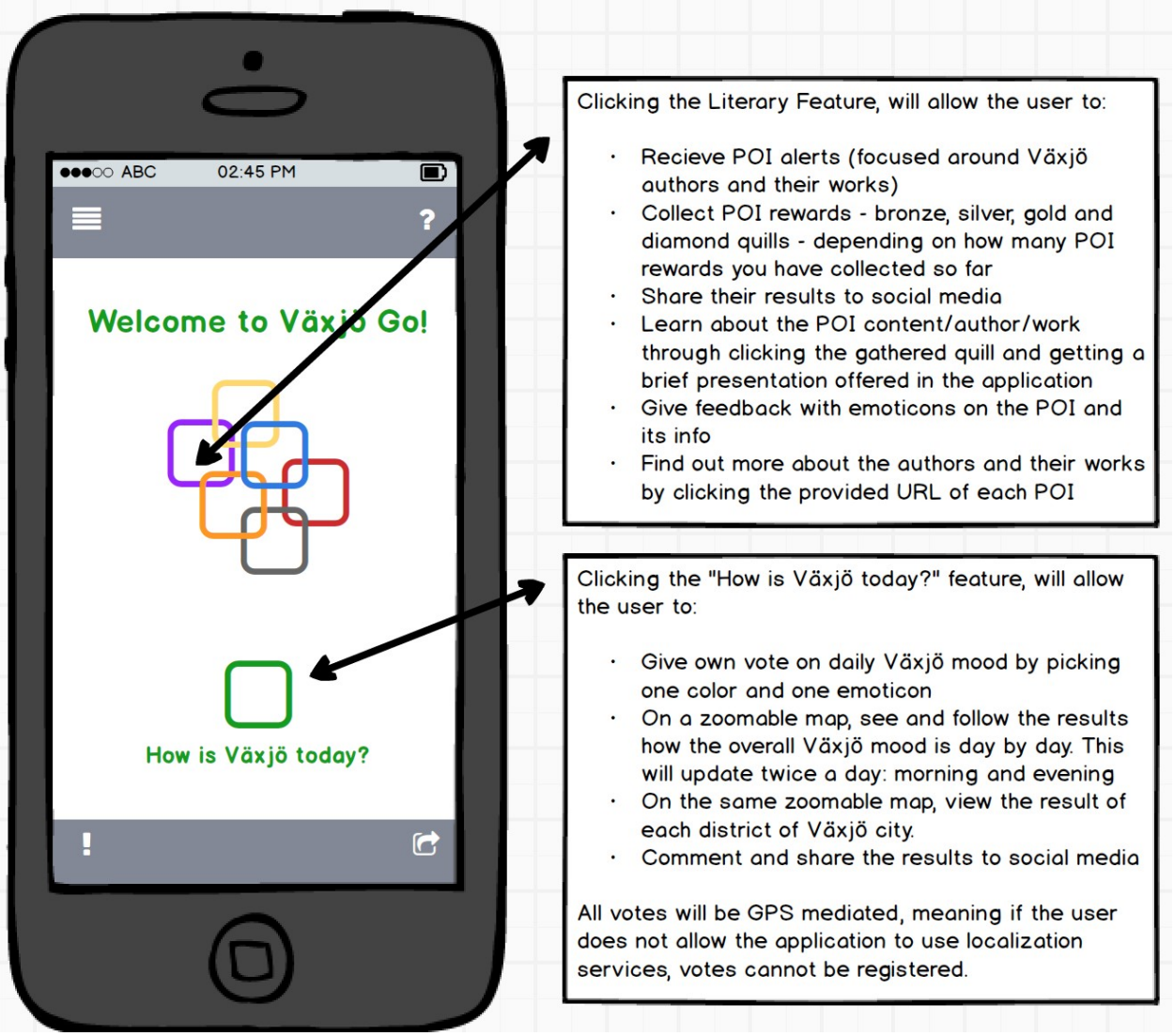
Franz, B. (2015). Immigration Youth, Hip Hop and Online Games. Alternative Approaches to the Inclusion of Working Class and Second Generation Migrant Teens. London: Lexington Books.

Gudiksen, S., and Svabo, C. (2014). Making and Playing with Customer Journeys. In Simonsen et al. (eds) *Situated Design Methods*, (2014) p. 139-159.

de Souza e Silva, A., and Frith, J. (2012). Mobile Interfaces in Public Spaces. Locational Privacy, Control and Urban Sociability. New York: Routledge.

APPENDIX (2 IMAGES)




●●●○○ ABC
02:45 PM
?
Welcome to Växjö Go!
How is Växjö today?
!
Clicking the Literary Feature, will allow the user to:
· Recieve POI alerts (focused around Växjö authors and their works)
· Collect POI rewards - bronze, silver, gold and diamond quills - depending on how many POI rewards you have collected so far
· Share their results to social media
· Learn about the POI content/author/work through clicking the gathered quill and getting a brief presentation offered in the application
· Give feedback with emoticons on the POI and its info
· Find out more about the authors and their works by clicking the provided URL of each POI
Clicking the "How is Växjö today?" feature, will allow the user to:
· Give own vote on daily Växjö mood by picking one color and one emoticon
· On a zoomable map, see and follow the results how the overall Växjö mood is day by day. This will update twice a day: morning and evening
· On the same zoomable map, view the result of each district of Växjö city.
· Comment and share the results to social media
All votes will be GPS mediated, meaning if the user does not allow the application to use localization services, votes cannot be registered.